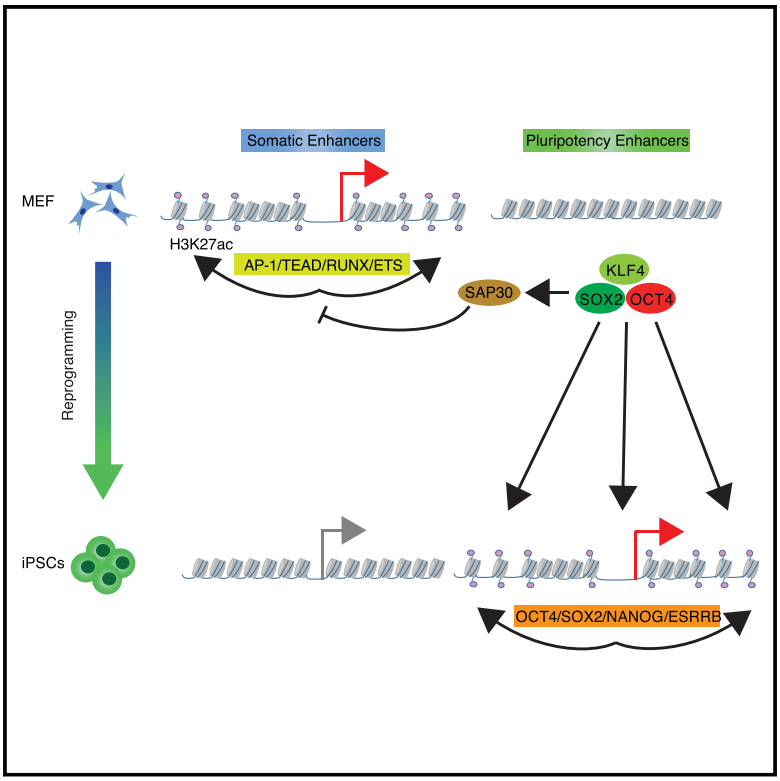
Article

Cell Stem Cell

Chromatin Accessibility Dynamics during iPSC Reprogramming

Graphical Abstract



Highlights

- ATAC-seq reveals chromatin accessibility dynamics during reprogramming
- TFs associated with initial chromatin closing are barriers for reprogramming
- OSK opens the pluripotent loci gradually through a direct process
- OSK closes the somatic loci indirectly in part through SAP30

Authors

Dongwei Li, Jing Liu, Xuejie Yang, ..., Andrew P. Hutchins, Jiekai Chen, Duanqing Pei

Correspondence

aphutchins@icloud.com (A.P.H.), chen_jiekai@gibh.ac.cn (J.C.), pei_duanqing@gibh.ac.cn (D.P.)

In Brief

Li et al. show that Yamanaka factors remodel the nuclear architecture of MEFs following a binary logic that may guide further improvement in reprograming technology and be applicable for other cell-fate decisions.

Li et al., 2017, Cell Stem Cell *21*, 819–833
December 7, 2017
https://doi.org/10.1016/j.stem.2017.10.012

Cell Stem Cell
Article

Chromatin Accessibility Dynamics during iPSC Reprogramming

Dongwei Li,[1,2,3,4,6] Jing Liu,[1,2,3,6] Xuejie Yang,[1,2,3,4,6] Chunhua Zhou,[1,2,3,4] Jing Guo,[1,2,3] Chuman Wu,[1,2,3] Yue Qin,[1,2,3,4] Lin Guo,[1,2,3] Jiangping He,[1,2,3,4] Shenyong Yu,[1,2,3] He Liu,[1,2,3,4] Xiaoshan Wang,[1,2,3,4] Fang Wu,[1,2,3,4] Junqi Kuang,[1,2,3,4] Andrew P. Hutchins,[1,2,7,*] Jiekai Chen,[1,2,3,7,*] and Duanqing Pei[1,2,3,4,5,6,7,8,*]
[1]CAS Key Laboratory of Regenerative Biology, South China Institute for Stem Cell Biology and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China
[2]Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, South China Institute for Stem Cell Biology and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China
[3]Joint School of Life Sciences, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou Medical University, Guangzhou 511436, China
[4]University of Chinese Academy of Sciences, Beijing 100049, China
[5]GUANGZHOU Regenerative Medicine and Health Laboratory, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China
[6]These authors contributed equally
[7]Senior author
[8]Lead Contact
*Correspondence: aphutchins@icloud.com (A.P.H.), chen_jiekai@gibh.ac.cn (J.C.), pei_duanqing@gibh.ac.cn (D.P.)
https://doi.org/10.1016/j.stem.2017.10.012

SUMMARY

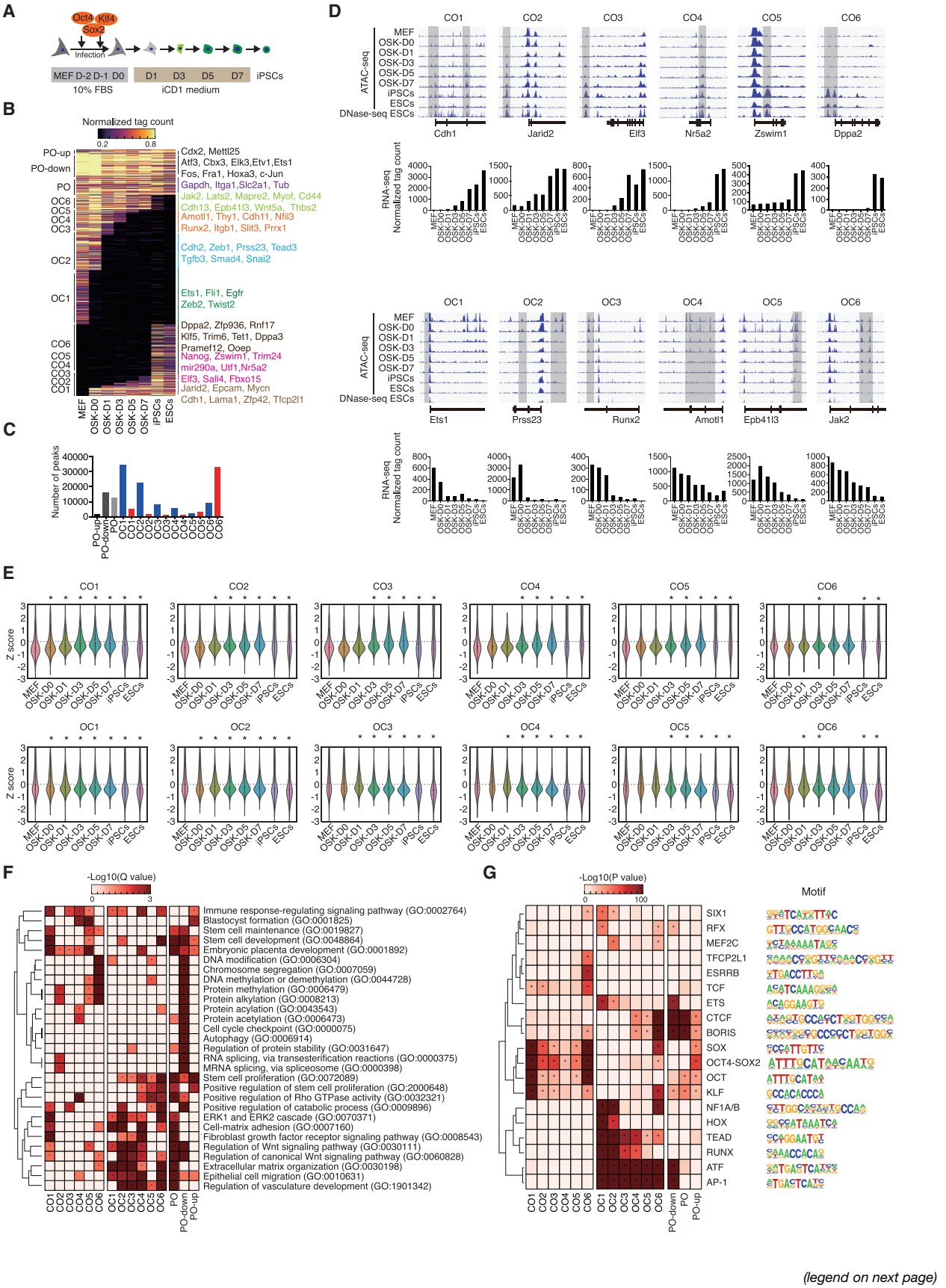
Cell-fate decisions remain poorly understood at the chromatin level. Here, we map chromatin remodeling dynamics during induction of pluripotent stem cells. ATAC-seq profiling of MEFs expressing Oct4-Sox2-Klf4 (OSK) reveals dynamic changes in chromatin states shifting from open to closed (OC) and closed to open (CO), with an initial burst of OC and an ending surge of CO. The OC loci are largely composed of genes associated with a somatic fate, while the CO loci are associated with pluripotency. Factors/conditions known to impede reprogramming prevent OSK-driven OC and skew OC-CO dynamics. While the CO loci are enriched for OSK motifs, the OC loci are not, suggesting alternative mechanisms for chromatin closing. Sap30, a Sin3A corepressor complex component, is required for the OC shift and facilitates reduced H3K27ac deposition at OC loci. These results reveal a chromatin accessibility logic during reprogramming that may apply to other cell-fate decisions.

INTRODUCTION

The induction of pluripotent stem cells (iPSCs) from fibroblasts by Yamanaka factors represents a unique system to understand the logic of cell-fate decisions (Takahashi and Yamanaka, 2006). The pathways from somatic cells to pluripotency have been well described by high-throughput methods such as microarrays and RNA sequencing (RNA-seq) that have provided a comprehensive molecular roadmap for gradual fate changes (Cacchiarelli et al., 2015; Hussein et al., 2014; Polo et al., 2012; Stadtfeld et al., 2008; Takahashi and Yamanaka, 2006; Toh et al., 2016). Broadly, the reprogramming process can be divided into three phases, the initial, middle, and maturation phases that are coupled to the ordered execution of multiple biological processes (Hussein et al., 2014; Polo et al., 2012), one of the earliest of which is a mesenchymal-to-epithelial transition (MET) (Li et al., 2010; Samavarchi-Tehrani et al., 2010).

It is clear that chromatin is dynamically remodeled at all of these stages in a specific manner (Koche et al., 2011), but the mechanisms are complex and incompletely understood (Smith et al., 2016). Recent studies have provided mechanistic insights at the molecular level. For example, Plath and colleagues reported that three reprogramming transcription factors (TFs) OCT4, SOX2, and KLF4 (OSK) first engage active somatic enhancers to initiate their silencing by redistributing somatic TFs and decommissioning enhancers (Chronis et al., 2017). They concluded that OSK collaborate among themselves and also with stage-specific TFs to orchestrate somatic-enhancer inactivation and pluripotent-enhancer activation (Chronis et al., 2017). Given that OSK are mainly transcriptional activators (Chen et al., 2008, 2016; Sridharan et al., 2009), it is not clear how they orchestrate the silencing of somatic enhancers directly.

Here, we report the global chromatin accessibility dynamics as cells are reprogramed from a somatic to a pluripotent state upon transfection of MEFs with OSK. We show that during reprogramming chromatin changes are defined by a fast initial wave of open to closed (OC), which is followed by a slower opening up of chromatin from closed to open (CO), and ends with a climactic wave of CO. The initial OC wave includes loci enriched with motifs for AP-1, TEAD, RUNX, ETS, and MAD-family TFs. Overexpression of TFs that belong to those families, such as the AP-1-family members c-Jun, c-Fos, and Fosl1, ETS-family member Ets1, TEAD-family member Tead3, RUNX-family members Runx1 and Runx2, and the MAD-family member Mef2c, all inhibited reprogramming. Of these factors, c-Jun inhibits reprogramming by preventing OSK-driven OC and skewing the overall OC/CO dynamics. We show that SOX2 and KLF4, and to a lesser



(legend on next page)

extent OCT4, are primarily drivers of CO, but not OC, suggesting that they must first activate potent (indirect) OC factors to remodel chromatin. Indeed, we show that SAP30, a frequent member of the SIN3 co-repressor complex (Silverstein and Ekwall, 2005) that is important for promoting pluripotency with NANOG (Saunders et al., 2017), acts in reprogramming to directly suppress key somatic genes. Together, our results reveal the chromatin accessibility logic during reprogramming and identify SAP30 as one of the first responders (Pei, 2009) that help silence somatic genes.

RESULTS

Global Chromatin Dynamics during OSK-Driven Reprograming of MEFs

Reprogramming of somatic cells with OSKM (OSK+c-MYC) is considered a slow and inefficient process (Esteban et al., 2010; Takahashi and Yamanaka, 2006), making comprehensive molecular analysis challenging (Chronis et al., 2017). We addressed this problem by taking advantage of OG2 MEFs (mouse embryonic fibroblasts), which bear an Oct4-GFP reporter whose expression reflects the establishment of pluripotency (Esteban et al., 2010; Liu et al., 2015b; Szabó et al., 2002) and the iCD1 reprogramming system (Chen et al., 2011), that is a chemically defined medium lacking FBS and c-MYC from the reprogramming cocktail, in which up to 40% of cells are GFP^+ by day 7 (Figures 1A, S1A, and S1B). We then harvested cells at D0, 1, 3, 5, and 7, along with MEFs, iPSCs, and embryonic stem cells (ESCs), as controls, and processed them for ATAC sequencing (ATAC-seq) (Buenrostro et al., 2013). Peaks of open chromatin were identified in the ATAC-seq results (Figure S1C; STAR Methods) and were highly consistent between independent experiments ($R > 0.8$; Figures S1D and S1E). Similarly, ATAC-seq data between ESCs and iPSCs data were also highly consistent $R = 0.93$ (Figure S1F), with the same chromatin patterns (Figure 1B). By comparing the peaks at each locus between MEFs and ESCs, the dataset reveals three basic groups, closed in MEFs but open in ESCs (CO), open in MEFs but closed in ESCs (OC), permanently open in both MEFs and ESCs (PO), which can be further subdivided into open in both but declining (PO-down) or increasing in intensity (PO-up) (Figure 1B; Table S1). The CO and OC peaks can be further divided into 6 subgroups (CO1–6; OC1–6), based on the day of opening and closing, covering the changes in chromatin between MEFs and iPSCs/ESCs. Counting these peaks further showed that OCs outnumber COs in the initial phases of reprogramming, but the pattern is reversed at the end of reprogramming as COs outnumber OCs (Figures 1B and 1C). This is consistent with earlier observations that reprogramming factors target different chromatin loci at the beginning and end of reprogramming (Sridharan et al., 2009) and is reminiscent of chromatin dynamics observed with specific epigenetic marks (Mikkelsen et al., 2008). CO peaks tended to be widely distributed among genomic features, suggesting both promoter and enhancer remodeling, while OC peaks tended to be gene distal suggesting mainly enhancer remodeling, and PO peaks were strongly enriched for promoters (Figure S1G). Among the peaks closed or opened, we identified genes specific to each OC or CO group, such as the pluripotency-related genes Cdh1, Jarid2, Elf3, Nr5a2, Zswim1, and Dppa2 acquiring open chromatin at specific stages of reprogramming, and conversely the somatic genes Ets1, Prss23, Runx2, Amotl1, Epb41l3, and Jak2 being progressively closed (Figure 1D). Importantly, OC and CO dynamics correlates with changes in gene expression, indicating that these are also functional changes (Figures 1D and 1E), and that chromatin accessibility is an important precursor for gene expression changes during reprogramming. Gene ontology (GO) analysis of the genes associated with OC and CO loci showed that the OC loci correspond to genes related to the somatic state (e.g., extracellular matrix organization), while the CO loci are largely associated with the pluripotent state (blastocyst formation; stem cell maintenance) (Figure 1F), which matches earlier reports that somatic genes are rapidly downregulated, followed by a slower activation of pluripotency genes (Buganim et al., 2013; Soufi et al., 2012; Sridharan et al., 2009).

To gain mechanistic insight into the drivers managing these chromatin dynamics, we analyzed the TF motifs associated with OC and CO peaks and showed that the CO loci have binding motifs for the exogenous transgenes OCT4, SOX2 and KLF4, which, agrees with other work indicating a direct binding-activation

Figure 1. ATAC-Seq Reveals Chromatin Dynamics during Reprogramming

(A) Schematic for the reprogramming time course and data collection. MEFs were infected with retroviral vectors containing Oct4, Sox2, and Klf4 and switched into iCD1 reprogramming medium and then harvested for ATAC-seq at the time points indicated in the lower row. D indicates day of the reprogramming.

(B) Loci of open chromatin were arranged into groups depending upon the day of reprogramming they changed from closed to open (CO) or open to closed (OC) or when they were permanently open (PO). PO was subdivided into those loci that declined (but remained open) as PO-down, those that increased (PO-up), and those that were unchanged (PO). Only serial (not transient) groups are shown here (see Figure S2C). Units are in normalized sequence tag count (see Figure S1C).

(C) The number of peaks defined in each of the OC/CO and PO categories.

(D) Selected genomic views of the ATAC-seq data, and comparison with DNase sequencing (DNase-seq) data from ESCs (GSM1014187) (Yue et al., 2014), are shown for the indicated OC/CO groups: Cdh1 (chr8:106,583,813–106,654,710), Jarid2 (chr13:44,716,254–44,747,836), Elf3 (chr1:135,248,174–135,261,330), Nr5a2 (chr1:136,926,019–136,977,558), Zswim1 (chr2:164,821,225–164,828,214), Dppa2 (chr16:48,306,779–48,320,018), Ets1 (chr9:32,689,014–32,721,691), Prss23 (chr7:89,504,809–89,525,973), Runx2 (chr17:44,755,486–44,821,809), Amotl1 (chr9:14,561,384–14,620,554), Epb41l3 (chr17:69,137,961–69,248,544), and Jak2 (chr19:29,243,094–29,278,873). OC/CO loci are marked with a gray box. All genome views are to the same vertical scale (0–20, normalized tag count), except DNase-seq (0 to 80, normalized tag count), and for all subsequent genome views. The RNA-seq expression values for the respective genes is shown in the bar plot below the genome view. RNA-seq expression units are in normalized tag counts.

(E) Violin plots of the expression level for all genes with a TSS within 10 kb of an ATAC-seq peak for each OC/CO category. Data were converted to a *Z* score based on the row-wise SD for each peak. *p value <0.01 Mann-Whitney U test versus the levels in MEFs.

(F) Gene ontology (GO) analysis for all genes within 10 kb of an ATAC-seq peak. *Benjamini-Hochberg corrected p value (Q value) <0.05 from a Wallenius noncentral hypergeometric distribution implemented in goseq. Data were clustered based on a Euclidean distance matrix and complete linkage.

(G) TF motifs significantly enriched at least >1.5-fold for CO/OC/PO categories of ATAC-seq peaks. The motifs for TFs are indicated on the right of the heatmap. *p value <1e-20. Data were clustered based on a Euclidean distance matrix and complete linkage.

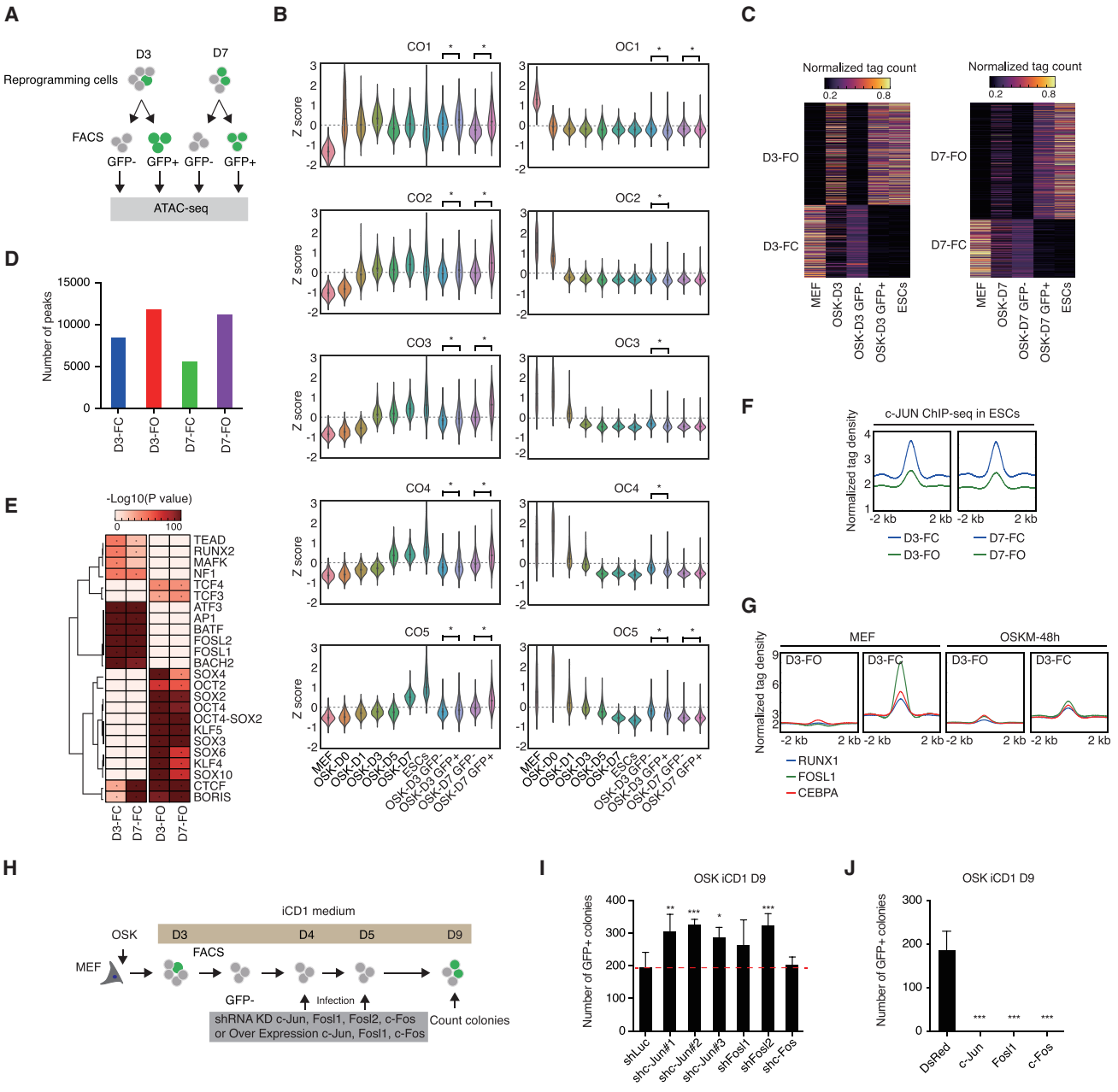


Figure 2. AP-1 TFs Impair Chromatin Remodeling

(A) Schematic of the flow cytometry sorting of the Oct4-GFP positive or negative cells at D3 and D7 for ATAC-seq. The GFP is driven by the OG2 reporter, which is a reporter for Pou5f1/Oct4 gene expression.

(B) Violin plots for the normalized ATAC-seq tag density for all peaks within the indicated OC/CO groups at the indicated time points. Data were converted to a *Z* score based on the row-wise SD for each peak. *$p < 0.01$ Mann-Whitney U test.

(C) Peaks were divided into those peaks open in MEFs and closed in ESCs and vice versa. Peaks were then further divided into those peaks that failed-to-close (FC) and those that failed-to-open (FO) for both day 3 (left heatmap) and day 7 (right heatmap).

(D) The total number of peaks in FC and FO for day 3 and day 7 cells.

(E) TF motif discovery within the day 3 or day 7 FC or FO peaks. *p value <1e-20. Data were clustered based on a Euclidean distance matrix and complete linkage.

(F) Average ChIP-seq read density of c-JUN data in ESCs (GSM1587320) (Liu et al., 2015b); the pileups are centered on FC/FO peaks for day 3 or day 7 cells.

(G) Average ChIP-seq read density for the somatic genes RUNX1, FOSL1, and CEBPA in MEFs or in OSKM 48-hr reprogramming cells (Chronis et al., 2017); the pileups are centered on FC/FO peaks for day 3 or day 7 cells.

(H) Schematic of the experimental design. Cells were sorted for Oct4-GFP negative cells at day 3, and the cells were then infected with small hairpin RNAs (shRNAs) targeting Luciferase as a control, and c-Jun, Fosl1, Fosl2, c-Fos, or cells were transfected with an overexpression construct containing DsRed, as a control, or c-Jun, Fosl1, and c-Fos. The cells were then cultured, and the number of Oct4-GFP$^+$ colonies were counted at day 9.

(legend continued on next page)

mechanism (Chen et al., 2008, 2016; Sridharan et al., 2009) (Figures 1G and S1H). We took advantage of a recent comprehensive analysis of TF binding during reprogramming (albeit in a reprogramming system that includes c-MYC, which we omit in our system) to explore OSK binding to these OC/CO groups. In the early stages of reprogramming (as represented by OSKM 48-hr samples), OSK binds to the CO1 group peaks, potentially directly mediating their opening (Figure S2A). Intriguingly, OSK are also recruited to OC3/4/5 (as defined in Figure 1B), which would be future sites of chromatin closing; however, in the later stage of reprogramming (represented by pre-iPSCs), that OSK binding is lost (Figure S2A), suggesting that OSK may, in addition to activating genes found at CO peaks, be promiscuously binding to chromatin during the early stages of reprogramming, in agreement with previous reports (Buganim et al., 2013; Chronis et al., 2017; Soufi et al., 2012; Sridharan et al., 2009). However, here we suggest that this promiscuous binding at OC peaks is reduced in the later stages of reprogramming and then lost completely in ESCs/iPSCs (Figure S2A). Why OSK is recruited to these OC loci is unclear as most of the OC loci do not have OSK motifs but are instead enriched with motifs for TEAD, RUNX, ETS, ATF, MAD, and AP-1-family TFs (Figure 1G), and, indeed, these OC peaks are enriched for the somatic TFs RUNX1, FOSL1, and CEBPA (Figure S2B), suggesting that OSK may either be passive, or indirectly silencing these regions of chromatin, although it has been suggested OSK act to directly displace key somatic TFs by promiscuous binding to DNA in the early stages of reprogramming (Chronis et al., 2017). Interestingly, there is a category of loci that are transiently opened during reprogramming (Figure S2C). These close-open-close (COC) peaks can be further classified into 10 different groups (COC1–10) (Figures S2C and S2D). As in the CO groups, the COC groups contain canonical SOX-OCT and KLF motifs, but COC peaks also contain enriched motifs for somatic TFs, such as AP-1, ATF, and developmental motifs, such as the endoderm-specific OCT4-SOX17 compressed motif (Figure S2E) (Aksoy et al., 2013). The enrichment of these motifs may help explain why a primitive streak-like stage is seen during reprogramming (Cacchiarelli et al., 2015; Polo et al., 2012; Takahashi et al., 2014), as chromatin becomes transiently permissive for the activity of alternative TFs and cell fates. Indeed, we observe many enriched GO terms for developmental genes near to open chromatin (Figure S2F).

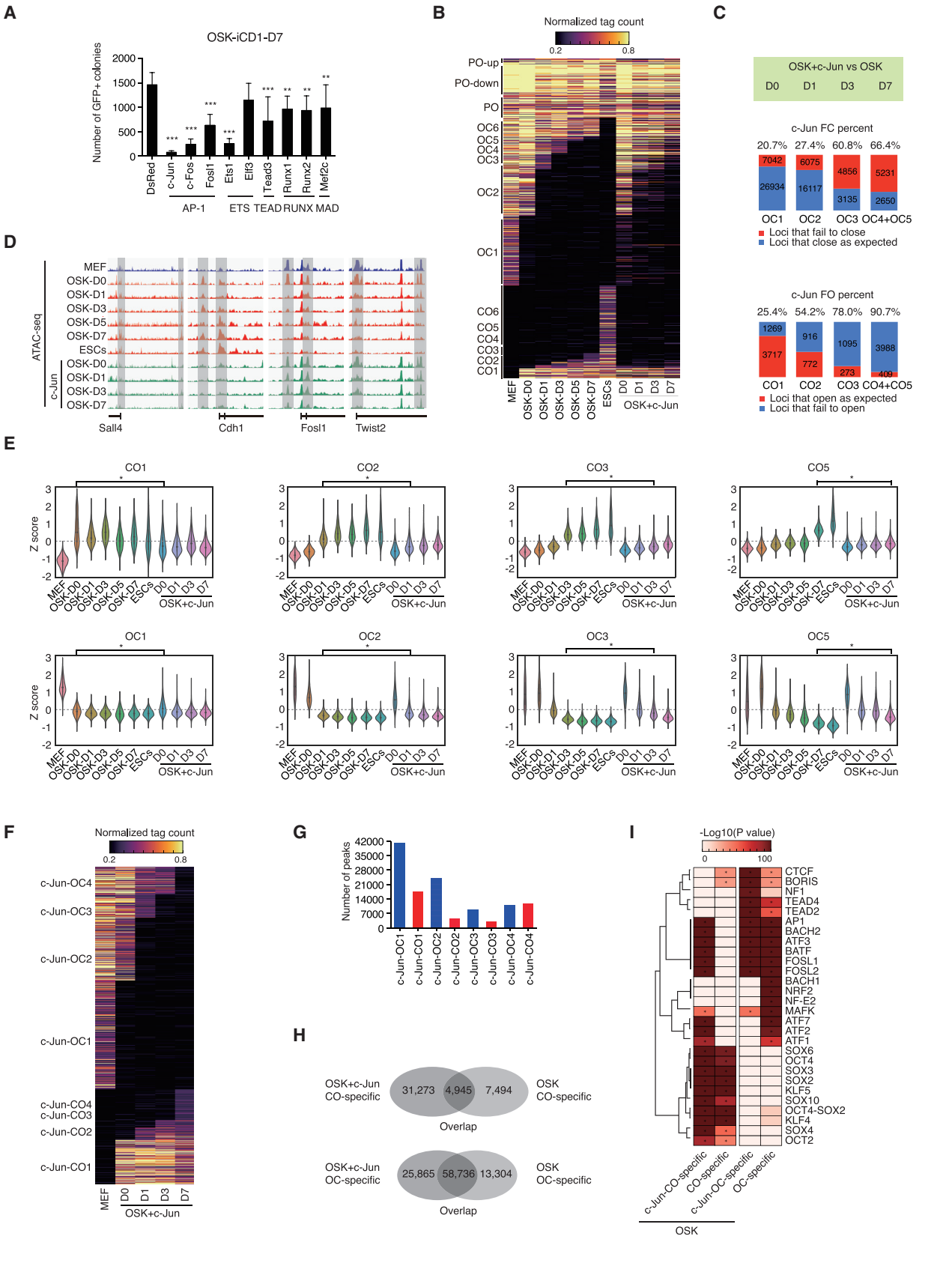
To see whether this same logic also exists in other reprogramming systems, we reanalyzed three ATAC-seq datasets for pluripotent reprogramming: two extended OSKM datasets where the starting cells were MEFs (Cheloufi et al., 2015; Chronis et al., 2017), and an OSKM system using B cells instead of MEFs as the starting cell type (Di Stefano et al., 2016). Principal component analysis (PCA) of gene expression of the three OSKM-based reprogramming systems, compared to our OSK iCD1 MEF system, showed that the three MEF-based systems all traverse similar trajectories to iPSCs, while the B cells arrived at the same place from a different starting point (Figure S3A). PCA of ATAC-seq data from those systems suggested a similar pattern, albeit chromatin seemed to be in advance of gene expression (Figure S3B). Intriguingly, all three systems follow a very similar OC/CO logic as our iCD1 system, albeit at different genomic loci, and with different numbers of OC/CO both early and late, likely due to differences in starting cell types, and sample coverage across the reprogramming time course (Figures S3C–S3E). Importantly, independent of the reprogramming system used, we detected the same somatic TF motifs in the OC peaks of all three systems: AP-1, ETS, RUNX, and TEAD family TFs (Figures S3F–S3H), highlighting the generality of these TFs. There were some important differences though, AP-1 motifs also appeared in the CO peaks in the B cell reprogramming system (Figure S3F), and in OC peaks the B cell system showed enrichment for the immune-related PU.1 and IRF-family TF motifs. Overall, these analyses reveal a binary logic regarding the opening and closing of genomic loci as cells reprogram and emphasize the importance of somatic TFs in the initial stages of reprogramming.

Somatic TFs as Reprogramming Barriers

To further test the open/close dynamic logic in reprogramming, we wished to see whether it can help explain why a significant portion of the MEFs cannot be reprogrammed, a common problem in reprogramming experiments. We reasoned that MEFs that are resistant to reprogramming fail to open or close loci appropriately for correct reprograming. In the iCD1 reprogramming system used here small numbers of GFP^+ cells can be detected as early as day 3; hence, we sorted D3 and D7 GFP^+ and GFP^- cells and then compared their chromatin accessibility by ATAC-seq (Figure 2A). Analysis of the overall levels of chromatin for the OC/CO groups indicates that the GFP^- cells have defects in both CO and OC (Figure 2B). To look closer at the defects, we calculated those loci that failed to close (FC), and failed to open (FO) (Figures 2C and 2D), and found that while D3 GFP^+ cells have a chromatin accessibility pattern much closer to ESCs and D7 GFP^+ cells, all GFP^- cells harbor a similar pattern (Figure 2C, right panel). As expected the FO peaks were dominated by motifs for the reprogramming factors, OCT4, SOX2 and KLF4, while FC peaks contained significantly enriched motifs at both D3 and D7 for AP-1 family TFs including ATF3, BATF, FOSL2, and BACH2 (Figure 2E). Interestingly, NF1, RUNX, and TEAD-family motifs are also present at FC loci (Figure 2E). Additionally, these FC peaks tend to be bound by somatic TFs; for example, when c-Jun was overexpressed in ESCs, it is preferentially recruited to these FC loci (Figure 2F), indicating they are target loci for c-JUN. A similar observation was seen in RUNX1, FOSL1, and CEBPA chromatin immunoprecipitation sequencing (ChIP-seq) data in MEFs and OSKM 48-hr (Chronis et al., 2017); those somatic TFs were bound primarily to FC peaks (Figure 2G). The D3 GFP^- cells can still proceed to become GFP^+ cells; hence, to understand the relative importance of chromatin opening and closing and the function of the somatic TFs, we sorted

(I) Effect of KD of c-Jun, Fosl1, Fosl2, and c-Fos in day 3 GFP^- cells (see schematic in H). After two rounds of infection, GFP^+ clones were counted at day 9. For iPSC generation, 3×10^4 D3 GFP^- cells were plated per well in 24-well plates. D, day; Luc, luciferase. Data are from 6 biological replicates in 2 independent experiments and are shown as the mean ± SEM. *p value <0.05, **p value <0.01, ***p value <0.001, one-way ANOVA with Dunnett's test.

(J) Overexpression c-Jun, Fosl1, c-Fos in day 3 GFP^- cells. After 2 rounds infection, GFP^+ clones were counted at day 9. For iPSC generation, 3×10^4 day 3 GFP^- cells were plated per well in 24-well plates. No GFP^+ colonies were observed when c-Jun, Fosl1, and c-Fos were overexpressed. D, day. Data are from 6 biological replicates in 2 independent experiments and are shown as the mean ± SEM. ***p value <0.001, one-way ANOVA with Dunnett's test.



(legend on next page)

GFP$^-$ cells at day 3 (Figure 2H) and then either knocked down AP-1 factors (Figures 2I and S4A) or overexpressed AP-1 factors (Figure 2J) and then allowed the cells to continue reprogramming and counted the number of resulting GFP$^+$ colonies. Knockdown of c-Jun and Fosl2 modestly but significantly increased the number of resulting GFP$^+$ colonies (Figure 2I), while conversely overexpression of c-Jun, Fosl1, and c-Fos completely ablated the ability of the GFP$^-$ cells to become GFP$^+$ cells (Figure 2J). This suggests that somatic TFs are barriers for reprogramming.

c-JUN Impedes Reprogramming by Disrupting Chromatin Dynamics

OSK are TFs known to activate gene expression by opening chromatin, i.e., the CO loci (Figure S2A), a process studied extensively and relatively well understood (Buganim et al., 2013; Gaspar-Maia et al., 2011; Koche et al., 2011; Mikkelsen et al., 2008; Polo et al., 2012; Sridharan et al., 2009). Indeed, opening of chromatin often improves reprogramming, for example, inhibition of histone deacetylases improves reprogramming (Huangfu et al., 2008), and the loss of many epigenetic repressors also helps (Chen et al., 2013; Schmidt and Plath, 2012; Xu et al., 2016). What is less well understood is the requirement for chromatin closing in reprogramming (Chronis et al., 2017), particularly at key somatic and fibroblastic genes (Buganim et al., 2013). We began by hypothesizing that TFs corresponding to the motifs in the OC loci (Figure S1H) may act negatively during reprogramming as we have already shown that the AP-1-family TFs c-Jun, c-Fos, Fosl1 (Fra1), and Fosl2 (Fra2) all impair reprogramming (Liu et al., 2015b); here, we show that in addition the ETS-family Ets1, TEAD-family Tead3, RUNX-family Runx1, Runx2, and MAD-family Mef2c can all significantly decrease the formation of GFP$^+$ colonies (Figure 3A). As overexpression of c-Jun was the most efficient at inhibiting reprogramming, we performed ATAC-seq on an OSK+c-Jun time course to determine where the defect in chromatin dynamics occurs. Western blot analysis indicated that, while c-JUN declines as cells are reprogrammed and is lost in ESCs, overexpression of c-JUN substantially increases the protein level of c-JUN (Figure S4B). ATAC-seq data show that c-Jun blocks reprogramming by perturbing the OSK-driven OC/CO dynamics (Figure 3B), and if we measure the peaks for the OC/CO sites we can see that the defect is mainly at OC3, as about 60% of the loci that should close fail to close (Figure 3C). This was matched by a near total failure to correctly open CO peaks. This effect can be seen at specific loci; for example, Sall4 and Cdh1 both failed to open, and Fosl1 and Twist2 failed to close (Figure 3D), crucially the chromatin status matches their gene expression pattern: Fosl1 and Twist2 both fail to be correctly downregulated, while Sall4 and Cdh1 fail to upregulate when c-Jun is overexpressed with OSK (Figure S4C). We can extend this pattern to all OC/CO groups independently of the binary threshold approach we use in Figures 3B and 3C. The levels of ATAC-seq signal for all groups indicated that even as early as OC1 the loci show significantly impaired closing, but at all days of reprogramming the OC groups were significantly down (Figure 3E).

Although c-Jun expression derailed normal OC/CO reprogramming dynamics, it still had a strong effect on chromatin, so we re-clustered the OSK+c-Jun datasets without ESCs and a new OC/CO pattern emerged (Figure 3F). Curiously, OSK+c-Jun led to the opening and closing of more loci than OSK alone (Figures 1C and 3G). There was little overlap in the pattern of CO peaks between OSK and OSK+c-Jun, and the vast majority of peaks were specific to either condition (Figure 3H). OSK+c-Jun opened 31,273 specific loci, suggesting that the cells are heading to an alternate cell fate (Figure 3H, top). For OC peaks, there was both a substantial overlap of 58,736 peaks, but also 25,865 and 13,304 specific peaks (Figure 3H, bottom). We then classified the peaks into those that were specific to OSK alone or OSK+c-Jun specific (Figure 3H). Motif discovery indicated that, while both sets of OC peaks were enriched for c-JUN motifs, the OSK+c-Jun-specific CO peaks were enriched for c-JUN, while the OSK-alone-specific CO peaks were not (Figure 3I). Additionally, those genes that were upregulated during OSK+c-Jun reprogramming were related to somatic functions, such as angiogenesis or lung development (Figures S4D and S4E). These results suggest that c-Jun inhibits reprogramming by altering the normal balance of OC/CO dynamics and instead initiating its own CO cascade that alters cell fate.

Figure 3. c-JUN Impedes Reprogramming by Impairing Chromatin Closing and Opening

(A) OSK and one of the AP-1-family TFs (c-Jun, c-Fos, Fosl1), ETS-family TFs (Ets1, Elf3), TEAD-family TF (Tead3), RUNX-family TFs (Runx1, Runx2), or MAD-family (Mef2c) were co-overexpressed during reprogramming, and the number of Oct4-GFP colonies was counted at day 7. For iPSCs generation, 2×10^4 OG2-MEFs were plated per well in 12-well plates. Data were from four independent experiments and are shown as the mean ± SEM. *p value <0.05, **p value <0.01, ***p value <0.001, one-way ANOVA with Dunnett's test.

(B) Heatmap of OC/CO loci, as defined in Figure 1B, this time including OSK+c-Jun reprogramming data. D indicates day.

(C) The OC/CO OSK-mediated regions were assessed for the impact of the co-expression of c-Jun. Bar chart shows the number of peaks misregulated by c-Jun. Peaks were counted as either peaks that should open but fail-to-open (FO) when c-Jun is present or peaks that should close but fail-to-close (FC) in the presence of c-Jun. The percentage of peaks that FC or FO are indicated. Data are generated for each day of the time course by comparing the OSK versus OSK+c-Jun for the respective days.

(D) Genome views of the ATAC-seq data for the pluripotency genes Sall4 (chr2:168,747,667–168,843,575), the epithelial gene Cdh1 (chr8:106,596,163–106,612,996), and the key somatic gene and AP-1 TF Fosl1 (chr19:5,440,934–5,457,109) and the mesenchymal gene Twist2 (chr1:91,798,164–91,825,398). Regions of open chromatin are marked with a gray box.

(E) Violin plots for the normalized ATAC-seq tag density for all peaks within the indicated OC/CO groups at the indicated time points. Data were converted to a *Z* score to emphasis change. Data were converted to a *Z* score based on the row-wise SD for each peak. *p < 0.01 Mann-Whitney U test.

(F) OC/CO peaks redefined from the ATAC-seq data when OSK and c-Jun are co-expressed in MEFs.

(G) Bar chart showing the number of ATAC-seq peaks for the c-Jun OC/CO-defined peaks in (F).

(H) Venn diagrams showing the overlap of OSK OC/CO peaks and OSK+c-Jun OC/CO peaks. Specific categories are defined as peaks exclusive to OSK or OSK+c-Jun ATAC-seq experiments; the overlap indicates peaks that CO or OC in both OSK and OSK+c-Jun.

(I) Motif discovery within the categories of OSK and OSK+c-Jun-specific peaks defined in (H). *p value <1e-20. Data were clustered based on a Euclidean distance matrix and complete linkage.

Since fetal bovine serum (FBS) inhibits reprogramming (Chen et al., 2011), we tested the effect of FBS and showed that FBS, like c-Jun, causes a similar fail-to-close and fail-to-open effect that leads to a failure to close 49% of OC3 and open 60% of CO3 loci (Figures S4F and S4G). This suggests that these FC/FO patterns provide a rational explanation for the observed impairments in reprogramming.

Oct4/Sox2/Klf4 Open Chromatin Loci

OSK are pioneering factors that can engage chromatin directly (Soufi et al., 2015). However, reprogramming is a cooperative process with OSK, the impact of each factor may be masked in our previous analysis. To resolve the role of each factor, we took advantage of the fact that ATAC-seq provides information about the physical status of chromatin at open or close states and performed ATAC-seq on MEFs infected with O, S, K, alone or in combination, at the early stages of reprogramming (Figure 4A). Although the individual factors are opening a lot of chromatin loci, it is only the OSK-D0/D1 (CO1/CO2, OC1/OC2) loci that are important for reprogramming, so we looked at the responses of the OSK CO1/2 or OC1/2 loci as defined in Figure 1B. We show that O, S, and K alone or in combination open chromatin in specific patterns, and no single factor or pair of factors can open/close chromatin in the same pattern as OSK (Figure 4B). We then categorized the CO1/2 or OC1/2 OSK peaks into the simplest tier of O, S, or K or combinations that could open/close a locus; i.e., if a locus could be opened by O and S alone, but not K, then we would count that locus as O and S dependent; hence, loci can appear in more than one list (Figures 4C and 4D). It became apparent that, while all three of OSK can open loci and are thus acting as pioneer TFs to open chromatin (Soufi et al., 2015), S and K could open nearly twice as many loci as O alone, and only 257 O peaks were exclusively opened by O, as many of the O-specific peaks could also be opened by S or K alone (Figures 4C and 4D), Interestingly, the S-containing double factors OS and SK both also opened more peaks than OK, suggesting that S has a strong role in opening chromatin (Figure 4C). It should be pointed out that we excluded c-MYC from our reprograming cocktail, which has interesting functions in both opening and closing chromatin (Buganim et al., 2013; Soufi et al., 2015; Sridharan et al., 2009). So, in the absence of c-MYC, we can conclude that SOX2 and KLF4 are major drivers of chromatin opening. By analyzing the motifs within those dependent-open and dependent-closed loci, we show that all dependent-open clusters contain OCT, SOX, KLF motifs as expected (Figure 4E). However, the dependent-closed peaks do not contain any OCT, SOX, and KLF motifs and instead contain the same collection of AP-1, RUNX, TEAD motifs we have seen previously (Figures 4E and S1H). These results suggest that OSK open chromatin by directly binding, while they close chromatin indirectly.

SAP30 Is Required for Reprogramming

The fact that OC loci contain no motifs for the reprogramming factors suggests that O, S, and K may activate other factors to perform the closing of chromatin. Additionally, an alternative mechanism has also been recently proposed that OSK may impair the activity of somatic TFs by promiscuous binding to active enhancers and displacing them (Chronis et al., 2017). To test the hypothesis that an indirect mechanism may be responsible for the closing of chromatin loci, we selected genes with a transcription start site (TSS) within 10 kb of a CO1–5 peak and intersected that with all genes upregulated >1.5-fold between MEFs and ESCs (Table S2) that were annotated as an epigenetic factor in the EpiFactors database (Medvedeva et al., 2015). This resulted in 18 candidate genes (Figures 5A and S5A), of which Sap30 appeared interesting as its expression was upregulated early at day 0, when most of the chromatin closing is taking place (Figure S5B). SAP30 is an adaptor for the SIN3A and NCOR corepressor complexes and can recruit HDACs to deacetylate histones (Laherty et al., 1998) and is part of the SIN3A complex upregulated during reprogramming, along with SIN3A itself (Figure S5B). In MEFs, Sap30 is expressed only at a basal level, as measured by RT-PCR, even though the chromatin at the TSS of Sap30 was open (Figures 5B and 6C). ATAC-seq data indicated two regions of chromatin become open as the cells reprogram and may represent enhancer regions (En1 and En2; Figures 5B and S5C). Importantly En1 responds only to OSK, while En2 responds to S, OS, SK, and OSK (Figure 5B), and only En1 recruits EP300, OCT4, and SOX2 in ESCs (Figure S5C), and finally only En1 chromatin matches Sap30 gene expression for the combinations of O, S, and K factors (Figure 5C) and is not upregulated when c-Jun is added to the reprogramming cocktail (Figure S5B). To test whether the two putative enhancers En1 and En2 at the Sap30 locus are responsible for its activation by the reprograming factors, we constructed reporters for these putative enhancers and showed that in a luciferase assay En1 alone is active as an enhancer (Figure 5D). We then show that exogenously expressed Sap30 significantly enhanced reprogramming (Figure S5D), albeit the improvement was small in the efficient iCD1 reprogramming system. Conversely, knocking down Sap30 substantially impairs reprogramming, as measured by the number of GFP^+ colonies (Figures 5E and S5E). To understand the mechanism by which Sap30 knockdown derails reprogramming, we performed ATAC-seq on D3 and D7 cells undergoing reprogramming with Sap30 knocked down and show that many peaks failed to close at both D3 and D7 (Figures 5F and S5F). At D7, while the early closing peaks (OC1/OC2) had mostly closed as expected, of the OC3 closing peaks 44% had failed to close at D7 with Sap30 knockdown (KD), and for the OC4 and OC5 peaks >70% of these peaks had failed to close, indicating a severe disruption in chromatin closing from the OC3 (day 1) stage onward (Figure 5G). Looking at the overall levels of chromatin indicates the defect in chromatin remodeling is evident by day 3 (OC3) and the cells do not recover (OC5) (Figure 5H). To explore the time requirement of Sap30 during reprogramming, we knocked down Sap30 at different time points, and this result indicated that Sap30 is mainly required at the earlier stages (D1–2) as KD from D5–7 had little effect on reprogramming while at D3–7 KD, although significant, was relatively modest in impairing reprogramming (Figures 5I and S5G). These results suggest that SAP30 is required for reprogramming by impacting chromatin remodeling.

SAP30 Is a Master Regulator of Chromatin Dynamics at Key Somatic Genes

To understand how SAP30 regulates reprogramming, we performed ChIP-seq of both SAP30 and SIN3A at D3 and D7 of an

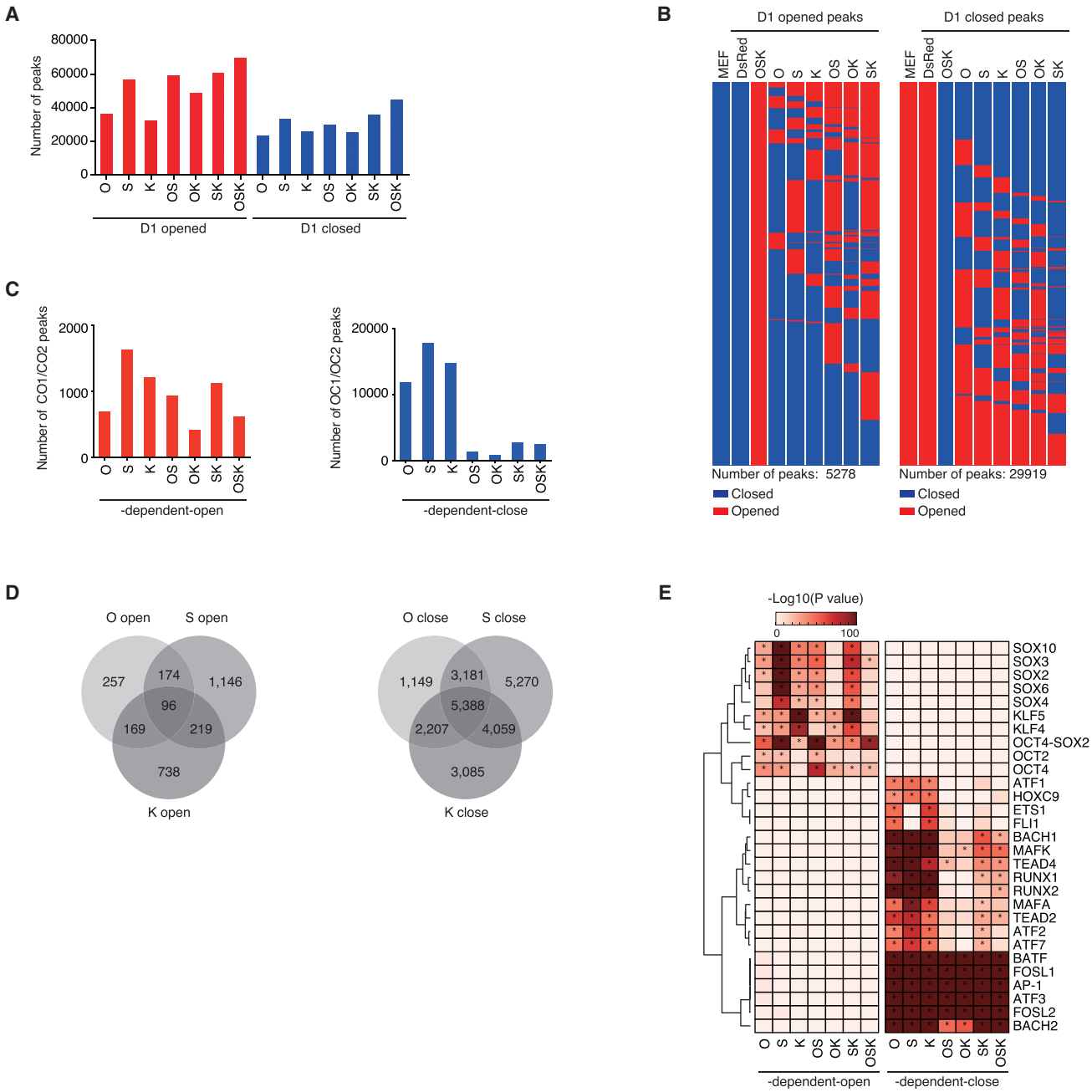


Figure 4. Oct4, Sox2, and Klf4 Act to Open Chromatin

(A) MEFs infected with Oct4 (O), Sox2 (S), or Klf4 (K) alone or in combination and the total numbers of chromatin regions (peaks) that become open or closed at day 1.

(B) Heatmap of the CO1/2 peaks (left) or OC1/2 peaks (right), as defined in Figure 1B, and their response when either single O, S, K, or pairs of O, S, K factors are expressed in a reprogramming time course. Any influence of the medium or cell-culture conditions was removed by only considering peaks that specifically opened in OSK and not DsRed D1 control or closed in OSK and remained open in DsRed D1 control.

(C) Number of peaks in the early stage of the OC1/2 and CO1/2 that can be opened or closed by single or pairs of reprogramming factors, derived from (B). An OSK peak was considered dependent-open or dependent-closed if it could open/close that peak by itself or if not by a single factor, by a combination of factors, for the simplest "tier" of factors. For example, if S, K, SK, and OS could open/close a peak, then it would be considered as S and K dependent, while an OSK peak that that could be opened/closed by OS and OK (but not O, S, or K alone) would be considered as both OS and OK dependent. Peaks can thus appear in more than one list, and the sum of peaks is greater than the total number of OSK peaks.

(D) Venn plots for the number of peaks for overlapping between each dependent class, derived from (C).

(E) TF motifs significantly enriched at least 1.5-fold for each defined group of ATAC-seq peaks, derived in (C). The TF family motifs are indicated on the right of the heatmap. *p value <1e-20. Note that the OSK motifs only appear in the dependent-open peaks, while the AP-1, TEAD, RUNX, and ETS somatic motifs are only enriched in the dependent-closed peaks. Data were clustered based on a Euclidean distance matrix and complete linkage.

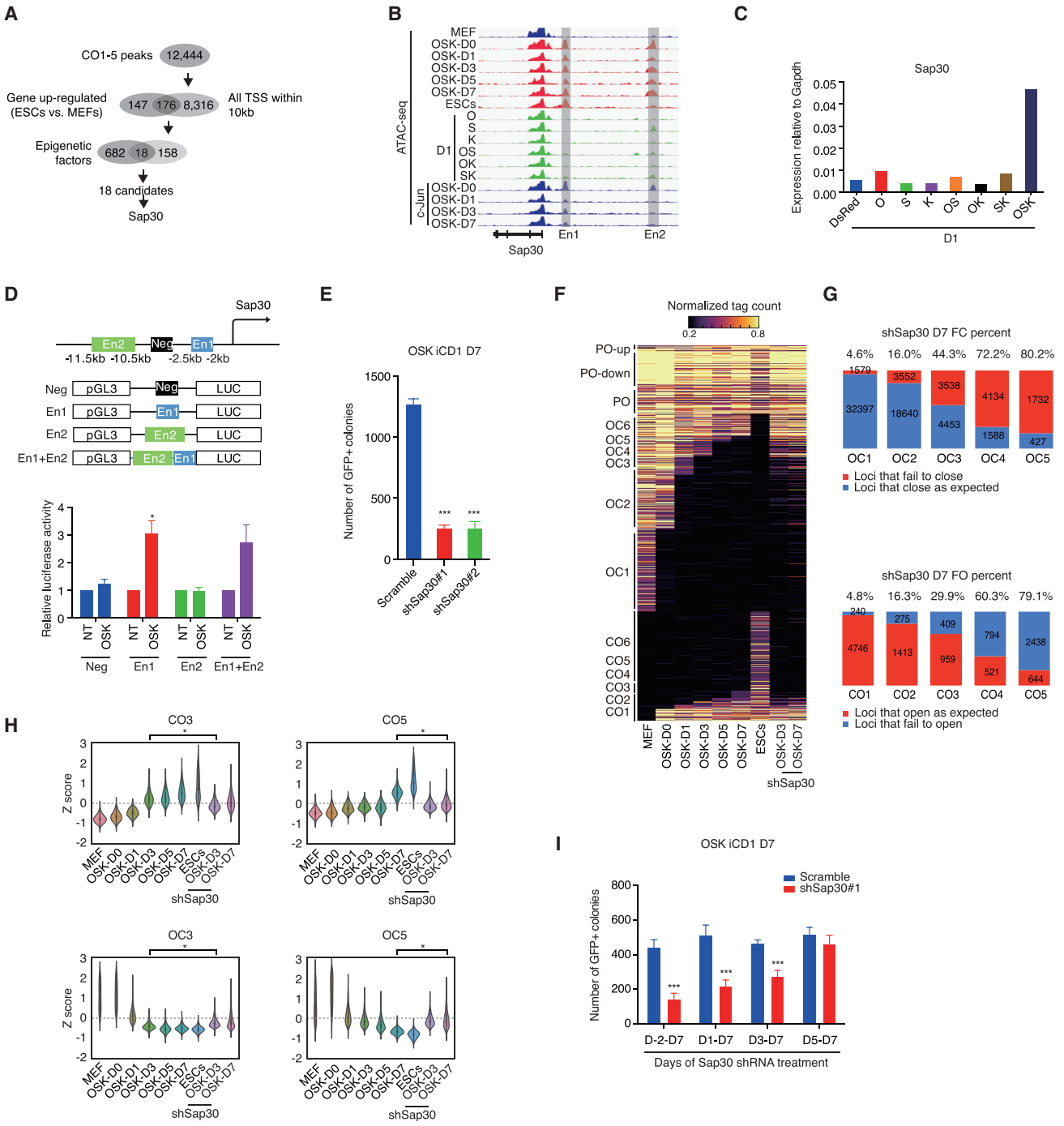


Figure 5. Sap30 Is Required for Efficient Reprogramming

(A) Schematic of the strategy used to identify Sap30. The CO1–5 peaks (Table S1) were annotated to 8,492 TSSs within 10 kb, of which 176 genes were also upregulated in ESCs compared to MEFs (Table S2). Of these 176 genes, 18 candidate genes were selected as potential epigenetic regulators as they overlapped with the EpiFactors database (Figure S6A) (Medvedeva et al., 2015).

(B) Genome view of the ATAC-seq data at the Sap30 locus (chr8:57,480,773–57,505,323). Chromatin is open at the TSS of Sap30 at all stages of reprogramming, but two putative enhancer regions (En1, En2) show dynamic opening during reprogramming.

(C) qRT-PCR of Sap30 relative to Gapdh at day 1 of a reprogramming time course in cells transduced with O, S, K alone, or their combinations. Data are from two biological replicates and are shown as the mean.

(D) Enhancer-luciferase assays for the regions of open chromatin near to the Sap30 gene. Schematic of the luciferase constructs used (left) and luciferase results (right). Neg, negative control region; En1/2, enhancer 1/2; OSK, co-transfection of Oct4, Sox2, and Klf4; NT, no transfection of reprogramming factors. Data are from three biological replicates and are shown as the mean ± SEM. *p value <0.05 from a one-way ANOVA with Dunnett's test.

(legend continued on next page)

OSK reprogramming time course. We show a substantial overlap of SAP30 and SIN3A, but each also had unique binding patterns (Figure 6A). We further show that both SIN3A and SAP30 are bound close to TSSs (Figure 6B), in agreement with previous reports that SIN3A binds closely to TSSs (Saunders et al., 2017; van Oevelen et al., 2010). As SAP30 and SIN3A are known to recruit HDACs (Laherty et al., 1998), we also mapped global H3K27ac by ChIP-seq and show that, in addition to being co-localized at TSSs with SAP30/SIN3A, H3K27ac had a broader distribution, extending 50–500 kb on either side of the TSS (Figure 6B). Although we originally hypothesized that SAP30 directly bind to OC loci, we were surprised that there was almost no association of SAP30 with OC peaks (Figure 6C). Instead, there was a close association with a subset of PO ATAC-seq regions that we called "PO-down" (Figures 1B, 6C, 6D and S6A), which are regions that are open, and remain open, but are reduced in intensity as the cells reprogram (Figure 1B). We noticed that within the set of PO-down genes we could identify many of the same key somatic TFs that we have previously shown to impair reprogramming, for example c-Jun, Fosl1, c-Fos, Ets1, and Runx2 (Figures 1B, 3A, and 6E). Crucially, all of these loci are occupied by SAP30, and KD of Sap30 leads to a large upregulation of H3K27ac levels at both day 3 and day 7 (Figures 6F and 6G). Overall, knocking down of Sap30 prevented the silencing of 507 somatic genes, including the key somatic genes Ets1, Fosl1, Fosl2, Bach1, and Zeb2 and many other AP-1 family TFs and other factors that GO analysis indicates are involved in developmental processes, such as dendrite morphogenesis, brain development, and blood vessel development (Figures 6H, 6I, and S6B; Table S3). Finally, we show that, of these 507 somatic genes, SAP30 does not alter chromatin accessibility, but it does lead to an increase in H3K27ac levels (Figure 6J). These results indicate that SAP30 works to decrease the levels of H3K27ac at key somatic genes. Loss of SAP30 during reprogramming leads to inappropriately high levels of H3K27ac which leads to upregulated somatic genes and a consequent derailing of reprogramming. These results further confirm SAP30 as a master regulator of chromatin closing by engaging the PO-down loci.

DISCUSSION

In this report, we mapped the chromatin accessibility dynamics associated with cell-fate changes during the conversion of MEFs to iPSCs by OSK and show that the dynamics has a binary open-close logic. Remarkably, this dynamic pattern can help explain how c-JUN and FBS impede reprogramming, as these factors lead to a failure to close the appropriate genomic loci at the correct time. Our ATAC-seq data were enriched for motifs for AP-1 TFs such as c-JUN, thus providing an explanation of our earlier work on c-JUN being a guardian of the somatic cell fate (Liu et al., 2015b). Our current work also extends the list beyond c-JUN to include TFs from the ETS, TEAD, RUNX, and MAD families, that all impair reprogramming when overexpressed. Additionally, we argue that the closing of open somatic loci, rather than the opening of pluripotent loci, is a pre-requisite for iPSCs, an effect also seen recently in other systems (Chronis et al., 2017), indicating that OSK must overcome powerful somatic TFs to alter cell fate and reprogram cells.

An attractive model for how OSK acts to bring about reprogramming is the pioneer model, where OSK act as pioneer factors to displace nucleosomes in closed chromatin and directly initiate the opening up of chromatin (Soufi et al., 2012, 2015). And indeed, we show that of S, K, and to a lesser extent O appear to be major chromatin openers (Figure 4), at least early in reprogramming. However, how chromatin closing occurs is less clear. OSK are not generally known as transcriptional repressors, but they do promiscuously bind to regions of open chromatin in the early stages of reprogramming (Buganim et al., 2013; Chronis et al., 2017; Sridharan et al., 2009), although this effect is lost rapidly by day 1 in our accelerated system (Figure S2A). Therefore, the pioneer model is not applicable to chromatin loci undergoing closing at the early phases of reprogramming.

Key somatic genes need to be downregulated for reprogramming to proceed, in a process that has previously been unclear because OSK are not generally known as transcriptional repressors (Buganim et al., 2013). It seems likely that OSK instead rely on "1st responders" that indirectly close chromatin and silence somatic genes (Pei, 2009), although it has been argued that OSK directly acts to displace somatic TFs away from enhancers (Chronis et al., 2017). In addition to the displacement of somatic TFs, we show that OSK activates SAP30 and, along with SIN3A complexes, acts to deacetylate chromatin at key somatic TFs, causing their downregulation.

Although SAP30 and SIN3A co-occupy many loci during reprogramming, 63% of SAP30 binding sites are not co-occupied by SIN3A (Figure 6A), indicating in addition to SAP30 and SIN3A common activity, they also have substantial context-specific functions, for example, SAP30 can also cooperate with the NCOR co-repressor complex for context-specific activity (Laherty et al., 1998). Similarly, we also observed enrichment of SAP30/SIN3A at a class of "PO-up" sites that have elevated

(E) KD of Sap30 impairs somatic cell reprogramming. Two shRNAs targeting Sap30 were co-transfected along with OSK and the number of GFP$^+$ colonies was assessed at day 7. For iPSCs generation, 2 × 104 OG2-MEFs were plated per well in 12-well plates. Data are from 12 replicates from 5 independent experiments and are shown as the mean ± SEM. ***p value <0.001 from a one-way ANOVA with Dunnett's test.

(F) Heatmap of OC/CO groups (as defined in Figure 1B), including ATAC-seq data when Sap30 was knocked down.

(G) Bar chart of the number and percentage of peaks that FO or FC when Sap30 is knocked down for the indicated OC/CO groups at day 7. Numbers and percentage were generated using only the state of chromatin at day 7 in OSK+shSap30.

(H) Violin plots for the normalized tag count for the ATAC-seq data for each OC or CO category (as indicated), for all time points including the shSap30 ATAC-seq data. Data were converted to a *Z* score based on the row-wise SD for each peak. *p value <0.01, Mann-Whitney U test.

(I) Effect of KD of Sap30 in reprogramming at specific time point windows. Sap30 was knocked down starting at the indicated day and proceeding for the remainder of the reprogramming time course. GFP$^+$ colonies were counted on day 7. Vectors containing an shRNA targeting Sap30 were transfected at the specified time points of reprogramming. For iPSC generation, 1 × 10^4 OG2-MEFs were plated per well in 24-well plates. D, day. Data are from 9 biological replicates in 3 independent experiments and are shown as the mean ± SEM. ***p value <0.001, two-way ANOVA with Sidak correction between the scramble control and the shSap30#1 KD.

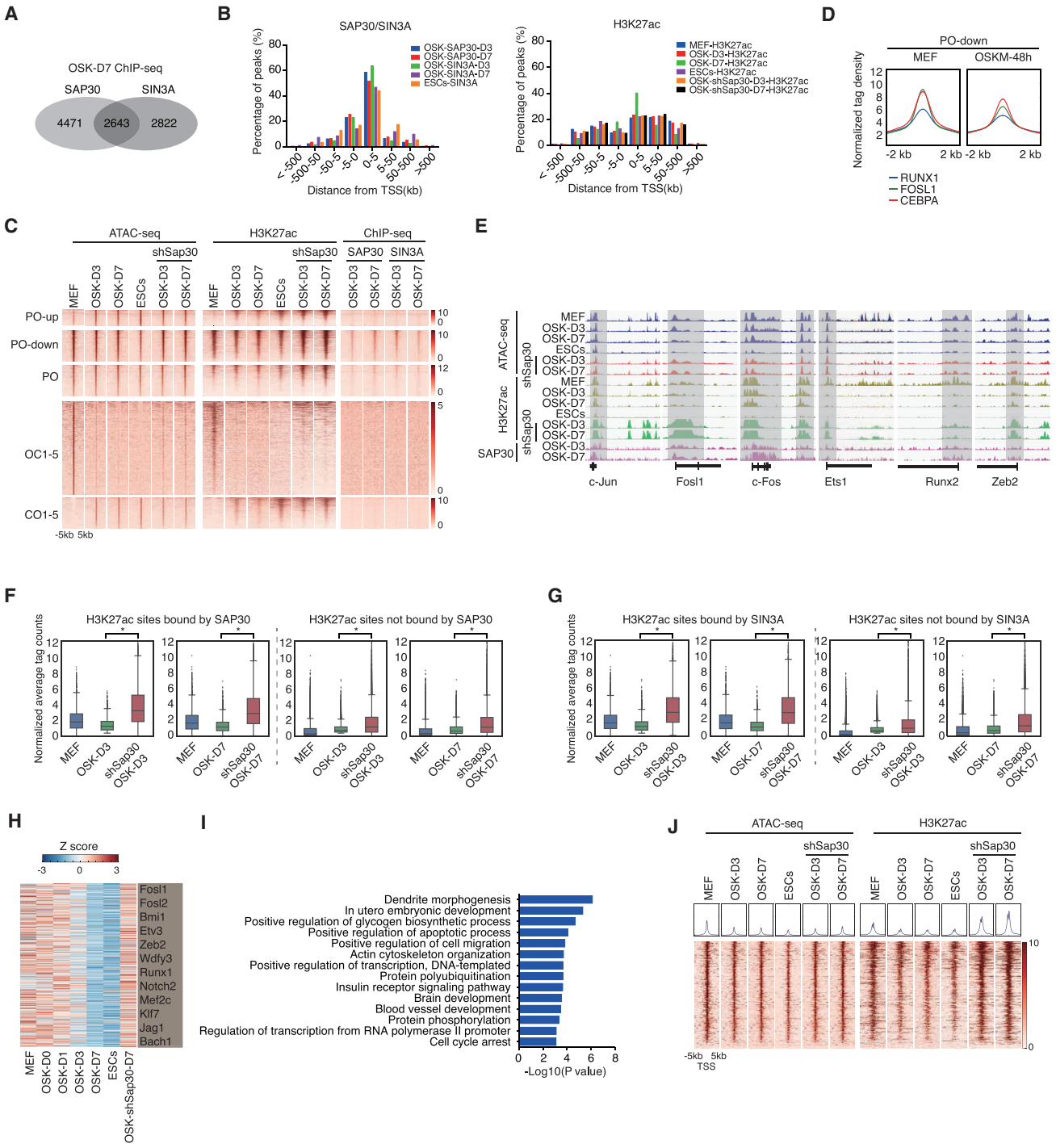


Figure 6. Sap30 Specifically Regulates Key Somatic Genes and Promotes Reprogramming by Removing H3K27ac

(A) SIN3A, SAP30 ChIP-seq overlap Venn diagrams. SAP30 was fused with a 3xFlag tag, co-expressed with OSK during reprogramming and precipitated using an anti-Flag antibody; SIN3A was precipitated using a anit-SIN3A antibody. SAP30 and SIN3A were subjected to ChIP-seq at day 3 and day 7 of reprogramming. SAP30 or SIN3A were considered overlapping if their peak spans overlapped by 1 bp or more, the Venn indicates the overlap of SAP30 and SIN3A at day 7.

(B) Distribution of SAP30 and SIN3A binding (left) or the distribution of H3K27ac peaks (right), with respect to the nearest TSS, for the indicated reprogramming conditions. The percentage of peaks in each category is indicated.

(C) Heatmap of sequence read density for ATAC-seq and H3K27ac, SAP30, SIN3A data, in MEFs, OSK-D3, OSK-D7, ESCs, and Sap30 KD at OSK-D3, OSK-D7 signal on CO/OC/PO category, ranked by mean signal strength. Windows are centered on the ATAC-seq peak summit. Each row of the heatmap is a genomic locus.

(D) Pileup of the somatic TFs RUNX1, FOSL1, and CEBPA at the PO-down ATAC-seq peaks in MEFs or OSKM 48-hr cells. Raw data was reanalyzed from Chronis et al. (2017).

(legend continued on next page)

levels of H3K27ac as reprogramming progresses. This is in agreement with reports that SIN3A can activate genes during muscle differentiation (Cheng et al., 2014; van Oevelen et al., 2010), and it has also been reported that a SIN3A/HDAC2 complex cooperates with NANOG to activate gene expression of pluripotent genes in the late stages of reprogramming (Saunders et al., 2017). Similar context-specific activation and repression is likely for SAP30. Consequently, it will be important to tease apart these differences in co-repressor complex activity, especially considering the common observation that the loss or inhibition of epigenetic co-repressors often improves reprogramming efficiency (Buganim et al., 2013; Chen et al., 2013; Huangfu et al., 2008; Xu et al., 2016).

One caveat in this study is the use of pooled populations of cells. At the single-cell level, there is indeed a binary logic (as a locus can only be open or closed); however, when analyzing a pool of cells this may not be the case, as the ATAC signal shows levels of chromatin opening and closing. Additionally, we use pooled populations at different stages of reprogramming, and this heterogeneity may mask cell states. Consequently, it will be critically important to apply new single-cell technologies to the understanding of chromatin dynamics (Buenrostro et al., 2015b; Cusanovich et al., 2015). Finally, the binary close/open dynamics described here may be applicable to other cell-fate decisions including those for differentiation and trans-differentiation. Interrogation of the chromatin dynamics may thus reveal strategies for altering the epigenetic environment and improving *in vitro* differentiation and transdifferentiation protocols. However, which targets to alter will be complex, and the logic described here provides the basis to design novel strategies to modulate cell fates with precision.

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
  - Reprogramming and transgenic experiments
  - Fluorescence activated cell sorting
  - Luciferase assay
  - qRT-PCR analyses
  - Immunoblotting
  - ATAC-seq
  - ChIP-seq
  - ChIP-seq library preparation
  - RNA-seq and gene expression analysis
  - ATAC-seq bioinformatic analysis and peak calling
  - Strategy for re-calling peaks from the ATAC-seq data
  - SAP30 and SIN3A ChIP-seq analysis and peaks calling
  - Transcription factor motif discovery and gene ontology
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and five tables and can be found with this article online at https://doi.org/10.1016/j.stem.2017.10.012.

AUTHOR CONTRIBUTIONS

D.L., J.L., A.P.H., and J.C. initiated the project. D.L., J.L., and J.C. designed the experiments. D.L. conducted and performed ATAC-seq experiments. D.L. and A.P.H. performed the bioinformatics analysis of the data and wrote the methods. X.Y. performed the ChIP-seq experiments. X.Y., L.G., S.Y., and C.Z. performed the reprogramming experiments. J.G. performed the luciferase assay. C.W. performed the western blot experiments. C.W., Y.Q., C.Z., and F.W. prepared the shRNA plasmid and performed qPCR experiments. H.L. and X.Y. prepared the RNA-seq libraries. X.W. and J.H. prepared RNA-seq data and ChIP-seq data. J.G. and J.K. constructed plasmids. D.P. conceptualized and supervised the project and wrote the manuscript with help from D.L. and A.P.H.

ACKNOWLEDGMENTS

This work was supported in part by the National Key R&D Program of China (2017YFJC040188, 2014CB965200, 2016YFA0101800, and 2016YFA0100400), National Natural Science Foundation of China (31471242, 31522033, 31550110206, 31421004, 31771424, 31401084, 31501185, and 31530038), "Frontier Science Research Program" of the CAS (QYZDJ-SSW-SMC009), Science and Technology Planning Project of Guangdong Province (2014B050504008, 2014B050502012, 2014B020225002, 2015B020228003), Guangdong Natural Science Funds for Distinguished Young Scholars (S2013050014040), and Youth Innovation Promotion Association of CAS. The authors also gratefully thank the support from the Guangzhou Branch of the Supercomputing Center of the Chinese Academy of Sciences.

Received: April 13, 2017
Revised: August 30, 2017
Accepted: October 23, 2017
Published: December 7, 2017

(E) Selected genome views for the ATAC-seq and H3K27ac, Sap30 ChIP-seq data in selected reprogramming time day3 and day7(OSK) versus OSK day3 and day7 when Sap30 was knocked down. Genome views: c-Jun (chr4:95,037,392–95,116,649), Fosl1 (chr19:5,446,411–5,454,187), c-Fos (chr12:85,467,708–85,491,253), Ets1 (chr9:32,691,918–32,723,378), Runx2 (chr17:44,724,979–44,740,035), and Zeb2 (chr2:45,100,690–45,118,632).
(F) All H3K27ac peaks were divided into those bound by SAP30 (left pair of boxplots) or those not bound by SAP30 (right pair of boxplots). SAP30 was considered bound if 1 bp of its peak span overlapped with any base pair of the H3K27ac marked region. The H3K27ac levels were measured for MEFs, OSK-D3, OSK-D7, and partially reprogrammed cells where Sap30 was knocked down. Boxplots were for SAP30 binding at day 3 or at day 7. *p value <0.01 Mann-Whitney U test.
(G) As in (F), but for SIN3A.
(H) Heatmap of all SAP30-bound genes expressed in MEFs and ESCs that were >1.5-fold upregulated when Sap30 was knocked down, compared to the control OSK-D7 reprogramming sample. SAP30 was considered as bound using annotatePeaks function from HOMER with default settings. A selection of specific genes is indicated on the right of the heatmap.
(I) Gene ontology analysis of the genes as defined in the heatmap. Analysis was performed using DAVID (Huang et al., 2009).
(J) Heatmap and pileup of ATAC-seq and H3K27ac ChIP-seq signal for MEF, OSK-D3, OSK-D7, ESCs, and knock down Sap30 in OSK-D3, OSK-D7 cells. The heatmap and pileups are centered on the target gene's TSS (upstream 5 kb and downstream 5 kb of the TSS).

REFERENCES

Aksoy, I., Jauch, R., Chen, J., Dyla, M., Divakar, U., Bogu, G.K., Teo, R., Leng Ng, C.K., Herath, W., Lili, S., et al. (2013). Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. EMBO J. *32*, 938–953.

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

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015a). ATAC-seq: A method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. *109*, 21–29.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015b). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490.

Buganim, Y., Faddah, D.A., and Jaenisch, R. (2013). Mechanisms and models of somatic cell reprogramming. Nat. Rev. Genet. *14*, 427–439.

Cacchiarelli, D., Trapnell, C., Ziller, M.J., Soumillon, M., Cesana, M., Karnik, R., Donaghey, J., Smith, Z.D., Ratanasirintrawoot, S., Zhang, X., et al. (2015). Integrative analyses of human reprogramming reveal dynamic nature of induced pluripotency. Cell *162*, 412–424.

Cheloufi, S., Elling, U., Hopfgartner, B., Jung, Y.L., Murn, J., Ninova, M., Hubmann, M., Badeaux, A.I., Euong Ang, C., Tenen, D., et al. (2015). The histone chaperone CAF-1 safeguards somatic cell identity. Nature *528*, 218–224.

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

Chen, J., Liu, J., Chen, Y., Yang, J., Chen, J., Liu, H., Zhao, X., Mo, K., Song, H., Guo, L., et al. (2011). Rational optimization of reprogramming culture conditions for the generation of induced pluripotent stem cells with ultra-high efficiency and fast kinetics. Cell Res. *21*, 884–894.

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

Chen, J., Chen, X., Li, M., Liu, X., Gao, Y., Kou, X., Zhao, Y., Zheng, W., Zhang, X., Huo, Y., et al. (2016). Hierarchical Oct4 binding in concert with primed epigenetic rearrangements during somatic cell reprogramming. Cell Rep. *14*, 1540–1554.

Cheng, J., Blum, R., Bowman, C., Hu, D., Shilatifard, A., Shen, S., and Dynlacht, B.D. (2014). A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. Mol. Cell *53*, 979–992.

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.

Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science *348*, 910–914.

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

Esteban, M.A., Wang, T., Qin, B., Yang, J., Qin, D., Cai, J., Li, W., Weng, Z., Chen, J., Ni, S., et al. (2010). Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. Cell Stem Cell *6*, 71–79.

Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. Nat. Rev. Mol. Cell Biol. *12*, 36–47.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A.E., and Melton, D.A. (2008). Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat. Biotechnol. *26*, 795–797.

Hussein, S.M., 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.

Hutchins, A.P., Jauch, R., Dyla, M., and Miranda-Saavedra, D. (2014). glbase: A framework for combining, analyzing and displaying heterogeneous genomic and high-throughput sequencing data. Cell Regen. (Lond.) *3*, 1.

Hutchins, A.P., Yang, Z., Li, Y., He, F., Fu, X., Wang, X., Li, D., Liu, K., He, J., Wang, Y., et al. (2017). Models of global gene expression define major domains of cell type and tissue identity. Nucleic Acids Res. *45*, 2354–2367.

Koche, R.P., Smith, Z.D., Adli, M., Gu, H., Ku, M., Gnirke, A., Bernstein, B.E., and Meissner, A. (2011). Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell Stem Cell *8*, 96–105.

Kumar, V., Muratani, M., Rayan, N.A., Kraus, P., Lufkin, T., Ng, H.H., and Prabhakar, S. (2013). Uniform, optimal signal processing of mapped deep-sequencing data. Nat. Biotechnol. *31*, 615–622.

Laherty, C.D., Billin, A.N., Lavinsky, R.M., Yochum, G.S., Bush, A.C., Sun, J.M., Mullen, T.M., Davie, J.R., Rose, D.W., Glass, C.K., et al. (1998). SAP30, a component of the mSin3 corepressor complex involved in N-CoR-mediated repression by specific transcription factors. Mol. Cell *2*, 33–42.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Li, B., and Dewey, C.N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., et al. (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell Stem Cell *7*, 51–63.

Liu, J., Chen, J., and Pei, D. (2015a). Reprogramming Mouse Embryonic Fibroblasts Using Different Reprogramming Factors (Nature Protocol Exchange).

Liu, J., Han, Q., Peng, T., Peng, M., Wei, B., Li, D., Wang, X., Yu, S., Yang, J., Cao, S., et al. (2015b). The oncogene c-Jun impedes somatic cell reprogramming. Nat. Cell Biol. *17*, 856–867.

Medvedeva, Y.A., Lennartsson, A., Ehsani, R., Kulakovskiy, I.V., Vorontsov, I.E., Panahandeh, P., Khimulya, G., Kasukawa, T., and Drabløs, F.; FANTOM Consortium (2015). EpiFactors: A comprehensive database of human epigenetic factors and complexes. Database (Oxford) *2015*, bav067.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49–55.

Pei, D. (2009). Regulation of pluripotency and reprogramming by transcription factors. J. Biol. Chem. *284*, 3365–3369.

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.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: A flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

Risso, D., Schwartz, K., Sherlock, G., and Dudoit, S. (2011). GC-content normalization for RNA-Seq data. BMC Bioinformatics *12*, 480.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.K., Beyer, T.A., Datti, A., Woltjen, K., Nagy, A., and Wrana, J.L. (2010). Functional genomics reveals

a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell *7*, 64–77.

Saunders, A., Huang, X., Fidalgo, M., Reimer, M.H., Jr., Faiola, F., Ding, J., Sánchez-Priego, C., Guallar, D., Sáenz, C., Li, D., and Wang, J. (2017). The SIN3A/HDAC corepressor complex functionally cooperates with NANOG to promote pluripotency. Cell Rep. *18*, 1713–1726.

Schmidt, R., and Plath, K. (2012). The roles of the reprogramming factors Oct4, Sox2 and Klf4 in resetting the somatic cell epigenome during induced pluripotent stem cell generation. Genome Biol. *13*, 251.

Silverstein, R.A., and Ekwall, K. (2005). Sin3: A flexible regulator of global gene expression and genome stability. Curr. Genet. *47*, 1–17.

Smith, Z.D., Sindhu, C., and Meissner, A. (2016). Molecular features of cellular reprogramming and development. Nat. Rev. Mol. Cell Biol. *17*, 139–154.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994–1004.

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.

Sridharan, R., Tchieu, J., Mason, M.J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q., and Plath, K. (2009). Role of the murine reprogramming factors in the induction of pluripotency. Cell *136*, 364–377.

Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell Stem Cell *2*, 230–240.

Szabó, P.E., Hübner, K., Schöler, H., and Mann, J.R. (2002). Allele-specific expression of imprinted genes in mouse migratory primordial germ cells. Mech. Dev. *115*, 157–160.

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.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Sasaki, A., Yamamoto, M., Nakamura, M., Sutou, K., Osafune, K., and Yamanaka, S. (2014). Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. Nat. Commun. *5*, 3678.

Toh, C.X., Chan, J.W., Chong, Z.S., Wang, H.F., Guo, H.C., Satapathy, S., Ma, D., Goh, G.Y., Khattar, E., Yang, L., et al. (2016). RNAi Reveals Phase-Specific Global Regulators of Human Somatic Cell Reprogramming. Cell Rep. *15*, 2597–2607.

van Oevelen, C., Bowman, C., Pellegrino, J., Asp, P., Cheng, J., Parisi, F., Micsinai, M., Kluger, Y., Chu, A., Blais, A., et al. (2010). The mammalian Sin3 proteins are required for muscle development and sarcomere specification. Mol. Cell. Biol. *30*, 5686–5697.

Xu, Y., Zhang, M., Li, W., Zhu, X., Bao, X., Qin, B., Hutchins, A.P., and Esteban, M.A. (2016). Transcriptional Control of Somatic Cell Reprogramming. Trends Cell Biol. *26*, 272–288.

Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: Accounting for selection bias. Genome Biol. *11*, R14.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al.; Mouse ENCODE Consortium (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355–364.

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

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti-c-JUN (1:2000)	Cell Signaling Technology	Cat#9165; RRID: AB_2130165
HRP antibody anti-GAPDH (1:4000)	KangChen Bio-tech	Cat#KC-5G5; RRID: AB_2631280
Rabbit polyclonal anti-H3K27ac	Abcam	Cat#ab4729; RRID: AB_2118291
Rabbit polyclonal anti-mSIN3A	Abcam	Cat#ab3479; RRID: AB_303839
Mouse monoclonal anti-FLAG	Sigma	Cat#m1804; RRID: AB_2637089
Rabbit polyclonal anti-IgG	Abcam	Cat#ab37415; RRID: AB_2631996
Chemicals, Peptides, and Recombinant Proteins		
Leukemia Inhibitory Factor (LIF)	Millipore	Cat#ESGE107
CHIR99021	Sigma	Cat#SML1046-5MG
PD0325901	Sigma	Cat#PZ0162-5MG
FBS	Moregate	Batch#67301120
DMEM-Dulbecco's Modified Eagle Medium, High Glucose	Hyclone	Cat#SH30022-2B
GlutaMAX	GIBCO	Cat#35050079
Non-Essential Amino Acids Solution	GIBCO	Cat#11140076
2-Mercaptoethoethanol	GIBCO	Cat#21985-023
Sodium Pyruvate Solution	GIBCO	Cat#11360070
Penicillin-Streptomycin Solution	Hyclone	Cat#SV30010
Phosphate-Buffer Saline (PBS)	GIBCO	Cat#C14190500BT
Trypsin-EDTA (0.25%), phenol red	GIBCO	Cat#25200114
0.1% Gelatin Solution	Millipore	Cat#ES-006-B
TRI Reagent	MRC	Cat#TR118-200
Puromycin Dihydrochloride	Thermo Fisher Scientific	Cat#A1113803
Oligo-dT	TaKaRa	Cat#D511
RRI	TaKaRa	Cat#D2313A
SsoAdvanced Universal SYBR Green Supermix	Bio-Rad	Cat#172-5274
Lipofectamine 3000 Reagent	Thermo Fisher	Cat#L3000001
IGEPAL CA-630	Sigma	Cat#198596
DNase/RNase Free Deionized Water	TIANGEN	Cat#RT121
Critical Commercial Assays		
Nextera DNA library preparation kit	Illumina	Cat#FC-121-1031
QIAGEN MinElute kit	QIAGEN	Cat#28006
KAPA Library Quantification kit	KAPA BIOSYSTEMS	Cat#KK4824
TruSeq RNA Sample Prep kit	Illumina	Cat#RS-122-2001
VAHTS Turbo DNA Library Prep kit	Vazyme Biotech	Cat#TD-503
NextSeq500 High output 150 cycles	Illumina	Cat#FC-404-2002
NextSeq500 Mid output 150 cycles	Illumina	Cat#FC-404-2001
VAHTSTM Library Quantification Kit	Vazyme Biotech	Cat# NQ101 - NQ106
AMPure XP beads	BEECKMAN COULTER	Cat#A63882
Deposited Data		
ATAC-seq data, ChIP-seq data, RNA-seq data	This paper	GEO:GSE93029

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Cell Lines		
OG2 Mouse Embryonic Fibroblasts:CBA/CaJ x C57BL/6J	This paper	NA
OG2 Mouse Embryonic Stem cells:CBA/CaJ x C57BL/6J	This paper	NA
Platinum-E (Plat-E)	A gift from The Fourth Military Medical University	N/A
HEK293T	ATCC	Cat#CRL-1126
Experimental Models: Organisms/Strains		
OG2 transgenic mice:CBA/CaJ x C57BL/6J	The Jackson Laboratory	Mouse strain datasheet: 004654
Oligonucleotides		
Oligonucleotides are summarized in Tables S4 and S5	N/A	N/A
Recombinant DNA		
pMX-Oct3/4	(Takahashi and Yamanaka, 2006)	Addgene Cat#13366
pMX-Sox2	(Takahashi and Yamanaka, 2006)	Addgene Cat#13367
pMX-Klf4	(Takahashi and Yamanaka, 2006)	Addgene Cat#13370
pMX-DsRed	Laboratory of D. Pei (Liu et al., 2015b)	N/A
pMX-c-Jun	Laboratory of D. Pei (Liu et al., 2015b)	N/A
pMX-c-Fos	Laboratory of D. Pei (Liu et al., 2015b)	N/A
pMX-Fosl1	Laboratory of D. Pei (Liu et al., 2015b)	N/A
pMX-Ets1	This paper	N/A
pMX-Elf3	This paper	N/A
pMX-Tead3	This paper	N/A
pMX-Mst1	This paper	N/A
pMX-Runx1	This paper	N/A
pMX-Runx2	This paper	N/A
pMX-Mef2c	This paper	N/A
pMX-Sin3a	This paper	N/A
pMX-Sin3b	This paper	N/A
pMX-Rbbp7	This paper	N/A
pMX-Rbbp4	This paper	N/A
pMX-Sap18	This paper	N/A
pMX-Sap30	This paper	N/A
pMX-Hdac1	This paper	N/A
pMX-Hdac2	This paper	N/A
M69 pCMV pGL3 Luciferase	was a gift from Randall Moon	Addgene Cat#17186
Software and Algorithms		
EDAseq	(Risso et al., 2011)	https://bioconductor.org/packages/devel/bioc/html/EDASeq.html
Macs14 v1.4.2	(Zhang et al., 2008)	http://liulab.dfci.harvard.edu/MACS/00README.html
Dfilter v1.5	(Kumar et al., 2013)	http://collaborations.gis.a-star.edu.sg/~cmb6/kumarv1/dfilter/tutorial.html
Bowtie v2	(Langmead and Salzberg, 2012)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Rsem v1.2.22	(Li and Dewey, 2011)	https://deweylab.github.io/RSEM/
Glbase	(Hutchins et al., 2014)	https://bitbucket.org/oaxiom/glbase/wiki/Home
FASTQC	Babraham Bioinformatics	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Homer	(Heinz et al., 2010)	http://homer.ucsd.edu/homer/motif/
Bedtools	Bedtools	http://bedtools.readthedocs.io/en/latest/
deeptools	(Ramírez et al., 2014)	https://deeptools.readthedocs.io/en/latest/
Samtools	Samtools	http://www.htslib.org/
Goseq	(Young et al., 2010)	https://bioconductor.org/packages/release/bioc/html/goseq.html
DAVID	(Huang et al., 2009)	https://david.ncifcrf.gov/
Flowjo	FLOWJO LLC	https://www.flowjo.com/
GraphPad Prism 6.0	GraphPad Software	https://www.graphpad.com/scientific-software/prism/
Illustrator	Adobe System Software Ireland	http://www.adobe.com/cn/products/cs6/illustrator.html
Photoshop	Adobe System Software Ireland	http://www.adobe.com/cn/products/cs6/photoshop.html
Bio-RAD CFX Manager	BIO-RAD	http://www.bio-rad.com/en-us/product/cfx-manager-software?tab=Download
ZEN	Zeiss	https://www.zeiss.com/microscopy/int/software-cameras.html
Accuri C6 Plus	BD biosciences	http://www.bdbiosciences.com/us/instruments/research/cell-analyzers/bd-accuri/m/1294932/overview
BD FACSAria II	BD biosciences	https://www.bdbiosciences.com/cn/home
Dual-Luciferase Reporter Assay System	Promega	https://www.promega.com/products/reporter-assays-and-transfection/reporter-assays/dual_luciferase-reporter-assay-system/?catNum=E1910
AB Applied Biosystems	Gene Company Limited	http://www.genehk.com/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information, and requests for reagents will be fulfilled by the Lead Contact, Prof. Duanqing Pei (pei_duanqing@gibh.ac.cn).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

OG2 transgenic mice (CBA/CaJ x C57BL/6J) were purchased from the Jackson laboratories (Mouse strain datasheet: 004654). Animals were individually housed under a 12 hr light/dark cycle and provided with food and water *ad libitum*. Our studies followed the guidelines for the Care and Use of Laboratory Animals of the National Institutes of Health, and the protocols were approved by the Committee on the Ethics of Animal Experiments at the Guangzhou Institutes of Biomedicine and Health. All efforts were made to minimize animal discomfort.

Cell culture

Mouse embryonic fibroblasts (MEFs) were derived from 13.5 d.p.c mouse embryos by crossing male Oct4–GFP (OG2) reporter allele-carrying mice (Szabó et al., 2002) (CBA/CaJ X C57BL/6J) to 129Sv/Jae female mice. MEFs, HEK293T and PlatE cells were maintained in DMEM supplemented with 10% FBS (NATOCOR), GlutaMAX (GIBCO) and non-essential amino acids (NEAA, GIBCO), Mouse ESCs and iPSCs were cultured feeder-free with N2B27-2i medium (50% (v/v) DMEM (Hyclone), 50% (v/v) knock out DMEM (GIBCO), N2 (GIBCO), B27 (GIBCO), NEAA (GIBCO), GlutaMAX (GIBCO), PD0325901 (1 μM, In house-synthesized), CHIR99021 (3 μM, In house-synthesized), LIF (10 ng/ml, In house-synthesized)). HEK293T were purchased from ATCC (CRL-1126). mESCs were derived in-house. All of the cell lines have been confirmed as mycoplasma contamination free with the MycoAlert Mycoplasma Detection Kit (Lonza, LT07-318).

METHOD DETAILS

Reprogramming and transgenic experiments

MEFs were reprogrammed as previously described (Chen et al., 2011; Liu et al., 2015a). Briefly, MEFs within 2 passages were plated at $2*10^4$ per well (12 well plate) and then infected with retrovirus generated from PlatE cells. Infected cells were cultured in iCD1 medium (Chen et al., 2011), or ESC medium (DMEM, 15% (v/v) FBS, penicillin/streptomycin (GIBCO), NEAA (GIBCO), Pyruvate (GIBCO), 2-Mercaptoethanol (GIBCO), GlutaMAX (GIBCO), LIF (10 ng/ml)). iPSC colonies were picked and then maintained as described above for ESCs. A step-by-step protocol describing the reprogramming procedure in this study can be found at the Nature Protocol Exchange (Liu et al., 2015a). shRNAs were transfected into cells by lentivirus. All shRNA sequences used in this manuscript can be found in STAR Methods.

Fluorescence activated cell sorting

Cells were treated with 2.5% pancreatic enzymes (GIBCO) and were resuspended in PBS + 1% BSA. GFP was detected using an Accuri^{TM} C6 Plus (BD biosciences) and $GFP^{-/+}$ cell sorting was performed on a BD Aria II (BD biosciences).

Luciferase assay

HEK293T cells were seeded in 24-well plates at 1.3×10^5 cells/well the day before transfection. Oct4/Klf4/Sox2 expression vector (200 ng/well), pGL3-reporter (100 ng/well) and TK-Renilla (1 ng/well) were co-transfected into the cells with Lipo-3000 (Life Tech) according to the manufacturer's instructions. pcDNA-ctrl were added accordingly to make the total DNA amount to be 700 ng/well for each sample. Cells were harvested 48 hr post-transfection and luciferase activity was detected with the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions.

qRT-PCR analyses

qRT-PCR reactions were set up in technical triplicate with the SYBR Green QPCR Master Mix (Applied Biosystems), all qPCR primers used in this manuscript can be found in STAR Methods.

Immunoblotting

Immunoblotting was accomplished as previously described (Liu et al., 2015b). Cells were collected into 1.5 mL Eppendorf tubes and lysed in protein lysis buffer with protease inhibitor cocktail (Roche) on ice for 10 min, and then boiled the lysis product in 100°C for 10 min. After centrifugation, the cell lysis products were subjected to SDS-PAGE and incubated with the corresponding primary antibody and secondary antibodies. The following antibodies were used: anti-c-JUN (CST no. 9165, 1:1000), anti-GAPDH (Bioworld, AP2063, 1:5000).

ATAC-seq

ATAC-seq was performed as previously described (Buenrostro et al., 2013; Buenrostro et al., 2015a). In brief, a total of 50,000 cells were washed once with 50 μL of cold PBS and resuspended in 50 μL lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.2% (v/v) IGEPAL CA-630). The suspension of nuclei was then centrifuged for 10 min at 500 g at 4°C, followed by the addition of 50 μL transposition reaction mix (25 μL TD buffer, 2.5 μL Tn5 transposase and 22.5 μL nuclease-free H_2O) of Nextera DNA library Preparation Kit (96 samples) (FC-121-1031, Illumina). Samples were then PCR amplified and incubated at 37°C for 30min. DNA was isolated using a MinElute Kit (QIAGEN). ATAC-seq libraries were first subjected to 5 cycles of pre-amplification. To determine the suitable number of cycles required for the second round of PCR the library was assessed by quantitative PCR as described (Buenrostro et al., 2015a), and the library was then PCR amplified for the appropriate number of cycles. Libraries were purified with a Qiaquick PCR (QIAGEN) column. Library concentration was measured using a KAPA Library Quantification kit (KK4824) according to the manufacturer's instructions. Library integrity was checked by gel electrophoresis. Finally, the ATAC library was sequenced on a NextSeq 500 using a NextSeq 500 High Output Kit v2 (150 cycles) (FC-404-2002, Illumina) according to the manufacturer's instructions.

ChIP-seq

ChIP was performed as described previously (Liu et al., 2015b), Briefly, cells were fixed in 18 mL DMEM (Hyclone) containing 1% formaldehyde for 15 min at room temperature with rotation, the reaction was quenched by the addition of 2 mL of 0.125 M glycine. And then the cells were washed with PBS 3 times. Cells were lysed in ChIP buffer A (50mM HEPES-KOH, 140 mM NaCl, 1mM EDTA (pH 8.0), 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 50 mM Tris-HCl (pH 8.0), and protease inhibitor cocktail) for 10 min at 4°C. Pellets were lysed in ChIP buffer B (1% SDS, 50 nM Tris-HCl (pH 8.0), 10mM EDTA and protease inhibitor cocktail) for 5 min at 4°C. The DNA was fragmented to 100-500 bp by sonication, and then centrifuged at 12,000 g for 2 min. The supernatant was diluted with ChIP IP buffer (0.01% SDS, 1% Triton X-100, 2mM EDTA, 50mM Tris-HCl (pH 8.0), 150 mM NaCl and protease inhibitor cocktail). Immunoprecipitation was performed using 2 μg antibody added to protein A/G Dynabeads, and incubated overnight at 4°C. Beads were washed, eluted and reverse crosslinked. DNA was purified by using the MinElute Reaction Clean up Kit (QIAGEN).

ChIP-seq library preparation

The ChIP DNA library for NextSeq 500 sequencing was constructed with VAHTS Turbo DNA Library Prep Kit for Illumina (Vazyme Biotech) according to manufacturer's instructions. AMPure XP beads were used for purification steps. The library was quantified with VAHTS Library Quantification Kit for Illumina (Vazyme Biotech). Libraries were sequenced on an Illumina NextSeq 500 v2 using 50bp paired-end reads.

RNA-seq and gene expression analysis

Total RNA was prepared with TRIzol. For quantitative PCR, cDNAs were synthesized with ReverTra Ace (Toyobo) and oligo-dT (Takara), and then analyzed by qPCR with Premix Ex Taq (Takara). For RNA-seq, TruSeq RNA Sample Prep Kit (RS-122-2001, Illumina) was used for library construction and the sequencing was done using a NextSeq 500 High Output Kit v2 (75 cycles) (FC-404-1005, Illumina), according to the manufacturer's instructions. RNA-seq was processed as described in (Hutchins et al., 2017), briefly reads were using aligned to a transcriptome index generated from the Ensembl annotations (v79), using RSEM (Li and Dewey, 2011), bowtie2 (Langmead and Salzberg, 2012), and normalized using EDASeq (Risso et al., 2011). RNA-seq data is expressed in units of GC-normalized tag counts.

ATAC-seq bioinformatic analysis and peak calling

All sequencing data were mapped onto the mm10 mouse genome assembly using bowtie2 (–very-sensitive). Low quality mapped reads were removed using samtools (view –q 35) and only unique reads mapping to a single genomic location and strand were kept. We removed mitochondrial sequences using 'grep –v 'chrM'. Biological replicates were merged, and peaks were called using dfilter (Kumar et al., 2013) (with the settings: -bs = 100 –ks = 60 –refine). BigWig files were produced using genomeCoverageBed from bedtools (scale = 10^7/ < each_sample's_total_unique_reads >) and then bedGraphToBigWig. Gene ontology and gene expression measures were called by first collecting all TSSs within 10 kb of an ATAC-seq peak, and then performing GO analysis with goseq (Young et al., 2010), or measuring gene expression. Other analysis was performed using deeptools (Ramírez et al., 2014), or glbase (Hutchins et al., 2014).

Strategy for re-calling peaks from the ATAC-seq data

Using only the peaks called by dfilter leads to a large number of false negative peaks. Hence, we 're-call' all peaks by measuring the sequence tag density in all ATAC-seq libraries for all peaks, irrespective of which library the peak was originally called in by dfilter. However, to use this strategy, we need to estimate the background signal that occurs in the ATAC-seq by chance alone, and so discriminate peak from non-peak. To generate a background, we first collected a superset of all peaks by merging all peaks with centers less than 350 bp apart. We similarly merged all ATAC-seq unique reads into a superset sequence library, and then removed all of the reads overlapped by 1 bp with any 'open' peak in our superset of ATAC-seq peaks. This provides a list of 'sequencable' genomic regions that were not inside putative ATAC-seq peaks, and from which we can randomly draw a pseudo-background. We then randomly extracted 50,000,000 reads as a pseudo-input (analogous to the input used in ChIP-seq experiments) to stand in for a random background. We then measured the density of sequence reads within a 700 bp window centered on each peak in the pseudo-input, and did the same for the ATAC-seq experimental peaks (Figure S1C). This suggested that a suitable background to call peak from non-peak would be 0.2734, resulting in a 0.1% false positive rate based on calling peaks within the pseudo-input. We then measured the sequence tag density for all peaks in our superset of peaks, and then 're-called' a peak as a peak or non-peak for each ATAC-seq experiment based on the threshold of 0.2734. Ultimately this led to more accurate peak calling, as weak peaks rejected by dfilter could be recovered as a real peak. All down-stream analysis is based on this threshold value of 0.2734 if the ATAC-seq is below this value it is annotated as 'closed' and above 'open'.

SAP30 and SIN3A ChIP-seq analysis and peaks calling

Reads form ChIP-seq experiments were mapped to the mouse genome (mm10) using Bowtie2 (–very-sensitive), as described for ATAC-seq data, and only those reads that mapped once were retained for further analysis. Peaks were called using MACS2 (Zhang et al., 2008) software with the default parameters.

Transcription factor motif discovery and gene ontology

Motif analysis was performed using HOMER (Heinz et al., 2010) using default settings. Motifs were only kept if the P value was < 0.01 and (< percent of target > / < percent of background >) was > 1.5. Gene ontology enrichment was performed using goseq (Young et al., 2010), with default settings.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bar chart data are presented as the mean ± SEM. Sequence pileups and heatmaps indicate the mean tag count per 1e6 bp in the library. Violin plots indicate the kernel density estimation of the data distribution, the boxes inside indicate the first and third quartiles, the white dot indicates the median. Boxes in boxplots indicate the first and third quartiles, whiskers indicate 1.5* < inter quartile range >, fliers indicate data points outside of that range. Sample number (n) indicates the number of independent biological samples in each experiment. Statistical tests used in this manuscript include one-way ANOVA with Dunnett's test, two-way ANOVA with Sidak

correction, Student's t test and two-tailed Mann-Whitney U test. Tests were considered significant as indicated in the figure legends. Other statistical tests were implemented as part of the respective computational framework: Gene ontology analysis was performed using goseq (Young et al., 2010) with a Wallenius noncentral hypergeometric distribution, and DAVID (Huang et al., 2009) using the EASE score (a modified Fisher exact test). Motif discovery using HOMER (Heinz et al., 2010) used a binomial distribution.

DATA AND SOFTWARE AVAILABILITY

The ATAC-seq, ChIP-seq and RNA-seq data described in this study was deposited with the gene expression omnibus with the accession number GEO: GSE93029.