Published in final edited form as:

Nature. 2015 December 10; 528(7581): 218–224. doi:10.1038/nature15749.

The histone chaperone CAF-1 safeguards somatic cell identity

Sihem Cheloufi[1,2,3,*], **Ulrich Elling**[4,*], **Barbara Hopfgartner**[5,§], **Youngsook L Jung**[6,7,§], **Jernej Murn**[8,9], **Maria Ninova**[10], **Maria Hubmann**[4], **Aimee I Badeaux**[8,9], **Cheen Euong Ang**[11], **Danielle Tenen**[2,3,4,5,6,7,8,9,10,11,12], **Daniel J Wesche**[1,2,3], **Nadezhda Abazova**[1,2,3], **Max Hogue**[1,2,3], **Nilgun Tasdemir**[13], **Justin Brumbaugh**[1,2,3], **Philipp Rathert**[5], **Julian Jude**[5], **Francesco Ferrari**[6,7], **Andres Blanco**[8,9], **Michaela Fellner**[5], **Daniel Wenzel**[4], **Marietta Zinner**[4], **Simon E Vidal**[14], **Oliver Bell**[4], **Matthias Stadtfeld**[14], **Howard Y. Chang**[3,15], **Genevieve Almouzni**[16], **Scott W Lowe**[13], **John Rinn**[2,12], **Marius Wernig**[11], **Alexei Aravin**[10], **Yang Shi**[8,9], **Peter Park**[6,7], **Josef M Penninger**[4], **Johannes Zuber**[5,#], and **Konrad Hochedlinger**[1,2,3,#]

[1] Massachusetts General Hospital's Department of Molecular Biology, Cancer Center and Center for Regenerative Medicine, Boston, MA 02114, USA

[2] Department of Stem Cell and Regenerative Biology and Harvard Stem Cell Institute, Cambridge, MA 02138, USA

[3] Howard Hughes Medical Institute, Chevy Chase, MD, USA.

[4] Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna Biocenter (VBC), A-1030 Vienna, Austria

[5] Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Dr. Bohr-Gasse 7, A-1030 Vienna, Austria

[6] Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA.

[7] Division of Genetics, Brigham and Women's Hospital, Boston, MA, 02115, USA.

Correspondence and requests for materials should be addressed to KH (hochedlinger@research.mgh.harvard.edu) or JZ (johannes.zuber@imp.ac.at)..

*Co-first authors

§Co-second authors

#Co-corresponding authors

Author contributions

SC, KH, UE and JZ designed primary screens, analyzed and interpreted data. SC, JM and NA performed the arrayed screen and SC conducted follow-up cell biology and chromatin studies. UE and BH performed the multiplexed screen. UE performed validation experiments, genetic interaction assays and cell biology experiments with support from BH, MH, and DW. NT and SL assisted in the generation of inducible Col1a1::tetOP-Chaf1a shRNA cell lines, JB performed human reprogramming experiments. SC, AIB, AB and YS performed B cell to macrophage conversion experiments. CEA and MW conducted MEF to iN transdifferentiation experiments. YJ, MN, AA, FF and PP performed bioinformatics analyses. MH and UE conducted the CiA assay. DJW assisted with the SONO-seq experiments and HC helped with the ATAC-seq assay. JM, MH and MZ assisted with Western blot and chromatin studies. DT and JR conducted ChIP experiments and library construction. MS and SV provided secondary Oct4-tdTomato MEFs. JZ and SL provided the arrayed library. JZ and PR designed the extended chromatin library. MF, JJ and BH generated lentiviral vectors and RNAi reagents. JP and GA provided intellectual support and mentoring. KH, SC, JZ and UE wrote the paper with input from all co-authors.

All SONO-seq, ATAC-seq, ChIP-seq, RNA-seq and microarray data have been deposited in the Gene Expression Omnibus database under accession number GSE66534. The authors declare no competing financial interests.

Supplementary information

This manuscript contains 9 Extended Data Figures, 5 Supplementary Tables and 1 Supplementary Figure.

[8] Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA.

[9] Division of Newborn Medicine, Boston Children's Hospital, Boston, MA 02115, USA.

[10] California Institute of Technology, Division of Biology and Biological Engineering, Pasadena, CA 91125, USA.

[11] Institute for Stem Cell Biology and Regenerative Medicine, Department of Pathology and Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

[12] Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142, USA.

[13] Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

[14] The Helen L. and Martin S. Kimmel Center for Biology and Medicine, Skirball Institute of Biomolecular Medicine, Department of Cell Biology, NYU School of Medicine, New York, NY 10016, USA

[15] Center for Personal Dynamic Regulomes and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, CA 94305, USA.

[16] Centre de Recherche, Institut Curie, 75248 Paris, France

Abstract

Cellular differentiation involves profound remodeling of chromatic landscapes, yet the mechanisms by which somatic cell identity is subsequently maintained remain incompletely understood. To further elucidate regulatory pathways that safeguard the somatic state, we performed two comprehensive RNAi screens targeting chromatin factors during transcription factor-mediated reprogramming of mouse fibroblasts to induced pluripotent stem cells (iPSCs). Remarkably, subunits of the chromatin assembly factor-1 (CAF-1) complex emerged as the most prominent hits from both screens, followed by modulators of lysine sumoylation and heterochromatin maintenance. Optimal modulation of both CAF-1 and transcription factor levels increased reprogramming efficiency by several orders of magnitude and facilitated iPSC formation in as little as 4 days. Mechanistically, CAF-1 suppression led to a more accessible chromatin structure at enhancer elements early during reprogramming. These changes were accompanied by a decrease in somatic heterochromatin domains, increased binding of Sox2 to pluripotency-specific targets and activation of associated genes. Notably, suppression of CAF-1 also enhanced the direct conversion of B cells into macrophages and fibroblasts into neurons. Together, our findings reveal the histone chaperone CAF-1 as a novel regulator of somatic cell identity during transcription factor-induced cell fate transitions and provide a potential strategy to modulate cellular plasticity in a regenerative setting.

Introduction

Ectopic expression of transcription factors is sufficient to override stable epigenetic programs and hence alter cell fate[1]. For example, forced expression of the pluripotency-related transcription factors Oct4, Klf4, Sox2 and c-Myc (OKSM) in somatic cells yields induced pluripotent stem cells (iPSCs), which are molecularly and functionally equivalent to

embryonic stem cells (ESCs)[2]. Similarly, ectopic expression of lineage-specific transcription factors drives conversion of heterologous cells into cardiac, neuronal, myeloid and other specialized cell types[3]. However, the reprogramming process is generally slow and inefficient, suggesting that chromatin-associated mechanisms must be in place to safeguard somatic cell identity and confer resistance to cell fate change.

Previous efforts to identify chromatin modulators of iPSC formation included gain and loss of function screens, as well as transcriptional profiling of bulk or FACS-enriched cell populations undergoing reprogramming. However, iPSC modulators that do not change transcriptionally are typically overlooked when analyzing expression dynamics in reprogramming intermediates[4]. Moreover, known repressors of iPSC formation such as p53, Mbd3, Dot1l, and Dnmt1 were either predicted or identified from small candidate sets and some of these molecules appear to depend on specific cell contexts or culture conditions[5-7]. While large-scale RNAi screens have been used to systematically probe roadblocks to reprogramming[4,8,9], this approach remains technically challenging due to the lack of effective shRNAs, prevalent off-target effects, and biases in the library representation or the screening readout. We therefore hypothesized that additional barriers to iPSC formation remain to be discovered and should yield insights into mechanisms that safeguard somatic cell identity.

To systematically explore chromatin factors that resist transcription factor-induced cell fate transitions, we utilized custom microRNA-based shRNA libraries targeting known and predicted chromatin regulators in two independent screening strategies during the reprogramming of fibroblasts into iPSCs. Both screens validated previously implicated chromatin pathways and revealed novel, potent repressors of reprogramming. Through a series of cellular and molecular studies, we found that suppression of a histone chaperone complex markedly enhanced and accelerated iPSC formation by influencing local chromatin accessibility, transcription factor binding and histone H3K9 trimethylation (H3K9me3). We propose that this complex functions as a key determinant of cellular identity by resisting transcription-factor induced cell fate change.

Results

RNAi screens for chromatin barriers to reprogramming

We conceived two parallel strategies for screening chromatin-focused microRNA-based shRNA (shRNAmiR) libraries in transgenic (“reprogrammable”) mouse embryonic fibroblasts (MEF) harboring a doxycycline (dox)-inducible polycistronic *OKSM* cassette and a constitutive M2-rtTA driver[10]. We first designed an arrayed screening strategy using a previously described miR-30-based retroviral shRNA library targeting 243 genes[11] (1,071 shRNAmiRs in pLMN vector) introduced one-by-one into reprogrammable MEFs (Fig. 1a and Supplementary Table 1). Alkaline phosphatase-positive (AP^+), transgene-independent iPSC-like colonies were quantified using customized image analysis software after 12 days of dox exposure and 5 days of dox-independent growth. Reprogramming efficiency ratios were calculated relative to a control shRNA targeting Renilla luciferase (Ren.713).

In an independent multiplexed screen, we introduced an optimized miR-E-based[12] retroviral library targeting 615 known and predicted chromatin regulators (5,049 shRNAmiRs in pLENC) (Fig. 1b and Supplementary Table 2) into reprogrammable MEFs harboring an Oct4-GFP reporter[10]. To control for biases due to background reprogramming events, we conducted parallel multiplex-screens in a large number of biological replicates. Specifically, we transduced reprogrammable MEFs with the entire pool of 5,049 shRNAs in 48 biological replicates (>100 infected cells per shRNA and replicate) and induced OKSM expression 3 and 6 days after viral transduction to control for differences in target protein half-life, yielding 96 replicates in total (Fig. 1b). Library representation was then quantified by deep sequencing of transgene-independent Oct4-GFP$^+$ iPSCs isolated by FACS, and strong shRNAmiRs were identified using an additive score reflecting the consistency of shRNA enrichment across all replicates.

Nucleosome assembly is a major roadblock to iPSC formation

Remarkably, despite methodological differences, the most prominent hits that emerged from both screens were Chaf1a and Chaf1b, two subunits of the chromatin assembly factor complex CAF-1 involved in the deposition of canonical histones H3/H4 on newly synthesized DNA[13] (Fig. 1c-e and Supplementary Tables 1,2). Additional hits of interest included the SUMO-conjugating enzyme Ube2i, a novel repressor of iPSC formation, as well as the SET domain-containing H3K9 methyltransferase Setdb1, which has previously been shown to inhibit reprogramming[14-16]. Importantly, the top-scoring shRNAs targeting Chaf1a, Chaf1b, or Ube2i reduced the expression of their predicted target genes (Fig. 1f, Extended Data Fig. 1a-d and Supplementary Figure 1).

We validated enhanced reprogramming with top-scoring shRNAs from the multiplexed screen (Fig. 1g and Supplementary Table 2). While reprogramming efficiency increased from 0.5% in controls to 1-5% with Trp53 shRNAs, consistent with previous reports[17], Chaf1a, Chaf1b, or Ube2i suppression robustly enhanced reprogramming efficiency up to 20-60%, regardless of culture conditions (Fig. 1g and Extended Data Fig. 1e,f). shRNAs targeting Sae1 and Uba2a, which are components of the SUMO E1 ligase complex, mirrored the effects of Ube2i suppression on reprogramming (Extended Data Fig. 1g), further validating the sumoylation pathway as a major repressor of iPSC formation.

We next employed a dox-inducible shRNAmiR expression system to test whether transient suppression of Chaf1a, Chaf1b, or Ube2i during reprogramming affected the ability of the resultant iPSCs to contribute to normal development. To this end, shRNAs and OKSM were simultaneously expressed in reprogrammable MEFs for 7 days. Oct4-GFP$^+$ cells were then purified using FACS, followed by dox withdrawal to select for transgene-independent iPSCs. Injection of these iPSCs into blastocysts gave rise to adult high-grade chimeras that efficiently produced germline offspring (Fig. 2a and Extended Data Fig. 2a). These results indicate that silencing of Chaf1a, Chaf1b, or Ube2i during reprogramming does not compromise the potential of iPSCs to differentiate into somatic and germ cell lineages *in vivo*.

Functional interactions of CAF-1, Ube2i, and Setdb1

We next devised a combinatorial RNAi approach to assess whether the reprogramming barriers Chaf1a, Chaf1b, Ube2i and Setdb1 interact during iPSC generation. Constitutive miR-E-based shRNA vectors targeting different pairs of chromatin regulators were sequentially transduced into reprogrammable MEFs, which were subsequently induced with dox for 7 days. The transgene-independent fraction of reprogrammed cells at day 11 was then determined by flow cytometry and plotted as the ratio of Oct4-GFP$^+$ to Oct4-GFP$^-$ cells relative to an empty vector control (Fig. 2b, c and Extended Data Fig. 2b). Co-suppression of two different CAF-1 subunits, or of Setdb1 plus Chaf1a, Chaf1b or Ube2i impaired overall reprogramming efficiencies compared to suppression of each gene alone, although we cannot exclude that this phenotype is due to cellular toxicity elicited by potent suppression of these chromatin regulators. By contrast, simultaneous suppression of either CAF-1 subunit and Ube2i further increased reprogramming efficiency, suggesting that these factors may act in independent pathways or stages to repress iPSC formation.

Suppression of CAF-1 accelerates iPSC formation

To gain insights into the dynamics of reprogramming in the absence of the identified chromatin barriers, we followed the emergence of Oct4-GFP$^+$ cells over time. While suppression of Ube2i promoted Oct4-GFP activation slightly earlier than controls (day 6 with Ube2i shRNA vs. day 9 with Renilla shRNA), the suppression of either CAF-1 subunit triggered a dramatic acceleration of this process and consistently generated Oct4-GFP$^+$ cells as early as 5 days after OKSM expression (Fig. 2d). Similar effects were observed with analysis of Nanog expression (Fig. 2e and Extended Data Fig. 2c).

We next examined the ability of our candidate shRNAs to facilitate transgene-independent clonal growth, a hallmark of iPSCs. Suppression of either CAF-1 subunit or Ube2i gave rise to transgene-independent Oct4-GFP$^+$ cells after only 5 days of OKSM expression, compared to 9 days in control shRNA-treated cells (Fig. 2f). Examination of cell surface markers associated with different stages of reprogramming further showed that a larger fraction of Chaf1a shRNA vector-infected cells expressed the early intermediate marker Epcam at day 4 of iPSC formation compared to controls, with a subset of cells activating the Oct4-tomato reporter (Fig. 2g). Consistently, inhibition of Chaf1a supported the formation of AP$^+$, transgene-independent iPSCs after as little as 4 days of factor expression (Fig. 2h). Based on the dramatic effects of Chaf1a and Chaf1b suppression on the dynamics and efficiency of iPSC formation, we focused subsequent analyses on these two components of the CAF-1 complex.

Suppression of CAF-1 subunits did not significantly influence expression of the Col1a1::tetOP-OKSM; R26-M2rtTA system at the RNA or protein level, ruling out the possibility that the observed phenotype is due to direct modulation of the reprogramming transgenes (Extended Data Fig. 3a-c and Supplementary Figure 1). Moreover, the reprogramming increase elicited by Chaf1b shRNAs could be rescued by overexpression of an shRNA-resistant version of human CHAF1B cDNA, demonstrating specificity of the effect (Extended Data Fig. 3d). Lastly, knockdown of either Chaf1a or Chaf1b did not increase cell proliferation in the presence of OKSM induction, indicating a growth-

independent effect of CAF-1 suppression on reprogramming efficiency (Extended Data Fig. 3e,f).

Reprogramming phenotype requires optimal CAF-1 and OKSM dose

To determine whether the effect of CAF-1 on reprogramming depends on OKSM expression levels, we compared iPSC formation between reprogrammable MEFs carrying either one (heterozygous) or two (homozygous) copies of the *Col1a1::tetOP-OKSM* and *R26-M2rtTA* alleles[10,18]. While CAF-1 suppression in heterozygous MEFs enhanced iPSC formation by orders of magnitude, CAF-1 suppression in homozygous MEFs resulted in a more modest increase in iPSC numbers (Fig. 3a,b). Accordingly, we observed that CAF-1 knockdown had a stronger impact on iPSC derivation efficiency when infecting MEFs with viral vectors achieving moderate or transient OKSM expression compared to vectors achieving high OKSM expression (Fig. 3c,d and Extended Fig. 4a-e). These results show that the reprogramming phenotype induced by CAF-1 suppression is influenced by both the levels and the duration of OKSM expression.

CAF-1 is essential for embryonic growth and viability of cultured cells in the absence of exogenous OKSM expression[19-22], which we confirmed in NIH3T3 cells using the strongest CAF-1 shRNAs (Extended Data Fig. 5a). To test whether the duration and degree of CAF-1 suppression might affect reprogramming efficiency, we generated transgenic MEFs carrying a dox-inducible Chaf1a shRNA linked to an RFP reporter in the *Col1a1* locus (Fig. 3e). Infection of transgenic MEFs with a constitutive lentiviral vector expressing OKSM in the presence of low doses of dox for 2-9 days resulted in a progressive increase in the formation of $Nanog^+$ iPSC colonies, which plateaued by day 6 (Fig. 3e and Extended Data Fig. 5b-d). By contrast, exposure of replicate cultures to high doses of dox increased reprogramming efficiency until day 4 but decreased iPSC colony numbers thereafter. These data suggest that enhanced reprogramming is also dependent on CAF-1 dose, with early CAF-1 suppression being beneficial but long term, potent suppression being detrimental to iPSC derivation.

To investigate whether genomic perturbation of CAF-1 components mimics the phenotype elicited by shRNAs, we introduced mutations into the endogenous *Chaf1a* locus using CRISPR/Cas9 technology. Briefly, reprogrammable MEFs were transduced with lentiviral vectors expressing Cas9 and two independent single guide RNAs (sgRNAs) targeting the C-terminal PCNA interaction domain, which is essential for chromatin assembly[23]. Strikingly, treatment of MEFs with Cas9 and sgRNAs resulted in a similar increase in reprogramming efficiency as shRNA treatment, indicating that genetic perturbation of the endogenous *Chaf1a* locus phenocopies the effect of Chaf1a shRNAs (Fig. 3f). Sequencing of Chaf1a sgRNA-induced modifications revealed the presence of mostly biallelic (24/34) and to a lesser extent monoallelic (10/34) genome edits, suggesting that complex alterations of the *Chaf1a* coding region on one or both chromosomes can lead to hypomorphic alleles that promote iPSC generation (Extended Data Fig. 5e). Indeed, analysis of representative iPSC clones showed a consistent reduction of total Chaf1a and Chaf1b protein levels (Extended Data Fig. 3f and Supplementary Figure 1).

Collectively, these findings demonstrate that the observed reprogramming phenotype is dependent on the duration and levels of both CAF-1 and OKSM and that this effect can be recapitulated with multiple experimental paradigms.

CAF-1 depletion enhances direct lineage conversion

To investigate whether CAF-1 acts as a gatekeeper of cellular identity across different cell types, we tested the effect of Chaf1a knockdown on the reprogramming potential of hematopoietic stem and progenitors cells (HSPCs) isolated from fetal livers of reprogrammable mice (Fig. 4a). While HSPCs expressing Renilla shRNA gave rise to 56% and 76% Pecam$^+$ iPSC-like cells at day 4 and 6, respectively, suppression of Chaf1a using two independent shRNAs facilitated reprogramming towards a Pecam$^+$ state in 90% of cells at day 4 and 97% at day 6 (Extended Data Fig. 6a,b). Chaf1a knockdown cells also showed a striking elevation of Pecam expression levels at both time points, consistent with acquisition of a fully reprogrammed iPSC state (Fig. 4b,c). Accordingly, suppression of Chaf1a gave rise to more transgene-independent colonies compared to controls (Extended Data Fig. 6c).

To assess whether CAF-1 stabilizes somatic cell identity in cell fate conversion systems other than OKSM-mediated reprogramming, we first examined transdifferentiation of fibroblasts into induced neurons (iNs) upon overexpression of the transcription factor Ascl1 in MEFs[24]. Transgenic MEFs harboring dox-inducible shRNAs targeting Renilla luciferase or Chaf1a in the *Col1a1* locus (see Fig. 3e) were transduced with a dox-inducible Ascl1-expressing lentivirus and measured for iN formation at day 13 (Fig. 4d). CAF-1 knockdown consistently resulted in a two-fold increase (P=0.0075) in the number of MAP2$^+$ neurons (Fig. 4e,f and Extended Data Fig. 6d). We next tested the effect of CAF-1 suppression during the conversion of pre-B cells into macrophages upon overexpression of the myeloid transcription factor C/EBPα (Fig. 4g). Consistent with a previous study[25], we found that the myeloid markers Cd14 and Mac1 are activated in the majority of cells after 48 hours of C/EBPα expression (Extended Data Fig. 6e-g). Remarkably, shRNA suppression of CAF-1 markedly increased Cd14 and Mac1 expression levels at two different time points (Fig. 4h,i) as well as the fraction of Mac1$^+$ and Cd14$^+$ cells after 24 hours of C/EBPα induction (Extended Data Fig. 6e-g).

Together, these data indicate that CAF-1 suppression not only enhances the induction of pluripotency from different cell types but also facilitates cellular transdifferentiation, suggesting that CAF-1 may play a more general role in resisting transcription factor-induced cell fate conversions.

CAF-1 influences chromatin accessibility and Sox2 binding

Since CAF-1 functions as a histone chaperone[13], we reasoned that its reduction may result in a more accessible chromatin structure and thus facilitate transcription factor binding to their target loci. To test this possibility, we first performed Sonication of Cross-linked Chromatin-Sequencing (SONO-seq) analysis, which determines accessible chromatin regions based on their increased susceptibility to sonication[26]. We analyzed bulk cultures expressing OKSM for 3 days when no stable iPSCs are yet present (Extended Data Fig. 7a), focusing on ESC-specific regulatory elements. While ESC-specific promoter elements

showed no discernible difference in accessibility (*P* value = 0.51), we observed a significant enrichment of SONO-seq signal at ESC-specific enhancer elements in CAF-1 depleted cells at day 3 of OKSM expression (Extended Fig. 7b; *P* value $<2.6\times10^{-12}$).

To validate these observations with an independent, higher resolution method, we performed Assay of Transposase Accessible Chromatin using sequencing (ATAC-seq), which detects integration of the Tn5 transposase in open chromatin regions[27] (Extended Data Fig. 7c). Consistent with the SONO-seq data, ATAC-seq analysis of early reprogramming intermediates showed a more accessible chromatin configuration at regulatory regions including ESC-specific enhancers upon suppression of Chaf1a (Fig. 5a; *P* value $<10^{-15}$). Moreover, Chaf1a knockdown caused a significant increase in chromatin accessibility across ESC-specific super-enhancers at day 3 of iPSC formation (Supplementary Table 3 and Extended Data Fig. 7d,e; *P* value $<5.3\times10^{-16}$). Of note, super-enhancers linked to specialized cell types such as macrophages, lymphocytes and muscle cells were also significantly more accessible in Chaf1a depleted reprogramming intermediates compared to controls (Extended Data Fig. 7f). Taken together, these results suggest that CAF-1 suppression increases the permissiveness of cells to transcription factor-induced cell fate change by facilitating a more accessible local chromatin structure at enhancer elements.

Next, we performed ChIP-seq analysis for Sox2 at day 3 of OKSM expression in order to test our hypothesis that increased chromatin accessibility at enhancer elements influences reprogramming factor binding. Indeed, we detected an increase in Sox2 binding to ESC-specific regulatory elements in Chaf1a shRNA-treated cells compared to controls (Fig. 5b; Extended Data Fig. 8a). Approximately 90% of Sox2 binding sites were shared between Chaf1a knockdown and control cells whereas 10% were unique to cells expressing either Chaf1a shRNA or Renilla shRNA (Extended Data Fig. 8b). Although Chaf1a knockdown cells showed slightly fewer unique Sox2 binding sites than Renilla knockdown cells, these sites were enriched for ESC-specific Sox2 targets (Extended Data Fig. 8c) and ESC-specific super-enhancer elements (Supplementary Table 4). Of the Sox2-bound super-enhancers unique to CAF-1 knockdown cells, a subset also showed a more accessible chromatin structure by ATAC-seq analysis (e.g., *Sall1*; Fig. 5c and Supplementary Table 4). Notably, Sox2 binding was also increased across lineage-specific super-enhancers when comparing Chaf1a knockdown cells to control at day 3 of reprogramming, consistent with the observed increase in chromatin accessibility at these elements (Extended Data Fig. 7f and 8d).

Collectively, these results indicate that loss of CAF-1 contributes to reprogramming, at least in part, by increasing chromatin accessibility at pluripotency-specific enhancer elements and by promoting the binding of Sox2 to ESC-specific targets.

CAF-1 affects local heterochromatin and gene expression

Considering that CAF-1 plays crucial roles not only in histone exchange but also heterochromatin maintenance[19,21,28], we next examined the global distribution of the heterochromatin mark H3K9me3 during reprogramming in the presence of either Chaf1a and Renilla shRNAs. We did not detect significant differences in H3K9me3 levels across pluripotency-associated enhancers or transposable elements (Extended Data Fig. 9a,b). Likewise, RNA-seq analysis of the same intermediates failed to show differential expression

of transposable elements (Extended Data Fig. 9c). However, we detected a local depletion of H3K9me3 at a subset of somatic heterochromatin areas termed “reprogramming-resistant regions”, which have recently been linked to the low efficiency of somatic cell nuclear transfer (SCNT)[29] (Fig. 5d and Extended Data Fig. 9d,e). These data suggest that CAF-1 inhibition, in concert with OKSM expression, causes local changes in this key repressive histone modification, which may prime chromatin structure for efficient transcriptional activation.

MEFs expressing OKSM and Chaf1a shRNAs upregulated a number of pluripotency-related genes (e.g., *Utf1, Epcam, Nr0b1, Tdgf1, Sall4*) at day 6 of reprogramming relative to the Renilla shRNA control (Supplementary Table 5). Although these genes were not yet differentially expressed at day 3 of reprogramming (data not shown), chromatin associated with these genes was already more accessible, supporting the view that CAF-1 suppression may prime the genome for subsequent transcriptional activation (Fig. 5e). To exclude that transcriptional activation of these pluripotency genes is an indirect consequence of accelerated reprogramming following CAF-1 suppression, we utilized the “chromatin in-vivo assay” (CiA)[30]. Specifically, we introduced CAF-1 or control shRNAs into transgenic fibroblasts carrying an array of Gal4 binding sites (*UAS* elements) upstream of the endogenous *Oct4* promoter and a GFP reporter in place of the *Oct4* coding region. While expression of Gal4-VP16 alone or in combination with a control shRNA triggered weak Oct4-GFP activation, co-expression of Gal4-VP16 and independent shRNAs targeting Chaf1a or Chaf1b strongly enhanced GFP expression from the somatically silenced *Oct4* locus (Fig. 5f and Extended Data Fig. 9f). These results demonstrate that CAF-1 suppression enhances the direct transcriptional activation of the endogenous *Oct4* locus independently of OKSM-induced cell fate changes.

Discussion

We find that CAF-1 suppression not only enhances reprogramming towards pluripotency but also direct lineage conversion, suggesting that the study of iPSC formation may be a valuable approach to uncover general mechanisms that safeguard somatic cell identity. Importantly, enhanced reprogramming is influenced by the degree and duration of CAF-1 suppression, consistent with the essential role of CAF-1 during cellular growth[19,20,22]. The identification of dose-dependent regulators of reprogramming highlights the utility of RNAi screens to achieve hypomorphic gene expression states, allowing for the detection of phenotypes that might not have been observed with complete and permanent ablation of genes. However, suboptimal reduction of CAF-1 levels may also explain our inability to detect a consistent enhancement of iPSC formation upon CAF-1 suppression in preliminary human reprogramming experiments (data not shown). Similarly, variable CAF-1 gene dosage may account for phenotypic differences between our study and previous publications examining the function of CAF-1 in ESCs[21,31]. Suppression of CAF-1 in ESCs reportedly results in a decondensation of pericentric heterochromatin, the reactivation of transposable elements and cell cycle arrest and apoptosis, which differs from our observations in nascent iPSCs. An alternative explanation is that cell-intrinsic differences between early reprogramming intermediates and established pluripotent cells cause distinct phenotypes[21].

It is possible that other histone chaperones compensate for reduced CAF-1 levels, thereby contributing to the observed reprogramming phenotype. In support of this possibility, suppression of the CAF-1 subunit p60 in HeLa cells triggers alternative deposition of the histone variant H3.3[32]. Of interest, H3.3 deposition on chromatin has been associated with enhanced reprogramming in the context of SCNT[33,34], suggesting parallels between the processes of cloning and iPSC generation. Our finding that CAF-1 suppression leads to specific loss of H3K9me3 at reprogramming-resistant regions supports this notion. The recent insight that CAF-1 siRNA-treated ESCs revert towards a 2 cell-like embryonic state more amenable to reprogramming by SCNT is consistent with this conclusion and reinforces the view that CAF-1 may act as a general stabilizer of cell identity[31]. We propose a model whereby CAF-1 contributes to the maintenance of somatic cell identity by stabilizing chromatin patterns (Fig. 5g). Here, suppression of CAF-1 would trigger dilution of newly assembled nucleosomes at key enhancer elements and loosening of chromatin structure in conjunction with forced expression of cell type-specific transcription factors. These combined changes would generate an accessible chromatin landscape for efficient transcription factor binding and activation of key target genes.

Methods

Plasmids

For the pooled RNAi screen, shRNAs were expressed from the LENC vector, which has been described previously[12]. For the primary screen validation, timeline experiment and immunofluorescence staining, mouse shRNAs were cloned individually into LENC. For the double knockdown experiment, shRNAs were cloned into LEPC (MSCV-mirE-PGK-Puro-IRES-mCherry). For the reprogramming dynamics experiment and chimera mouse production, shRNAs were cloned into RT3CEPIN (TRE3G-mCherry-mirE-PGK-Puro-IRES-Neo). For reprogramming experiments with non-transgenic systems, previously published OKSM lentiviral vectors were modified to introduce promoters of different strength, which are described in the main figures.

Cell culture and media

Packaging cells (Platinum-E Retroviral Packaging Cell Line) for producing retroviral particles were cultured in DMEM supplemented with 15% FBS, 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin, sodium pyruvate (1 mM) and L-glutamine (4 mM) at 37°C with 5% CO_2. Mouse embryonic fibroblasts (MEF) were cultured in DMEM supplemented with 15% FBS, 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin, sodium pyruvate (1 mM), L-glutamine (4 mM), L-ascorbic acid (50 μM) at 37°C with lox oxygen (4.5% O_2). iPSCs were derived in DMEM supplemented with 15% FBS, 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin, sodium pyruvate (1 mM), L-glutamine (4 mM), 1000 U/ml LIF, 0.1 mM 2-mercaptoethanol, and 50 μg ml^{-1} ascorbic acid at 37°C with 5% CO_2 and 4.5% O_2. iPSCs for blastocyst injection were cultured on feeders in DMEM supplemented with 13% Knockout Serum Replacement (Gibco), 2% FBS, 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin, sodium pyruvate (1 mM), L-glutamine (4 mM), L-ascorbic acid (50uM), 1000 U/ml LIF, beta-mercaptoethanol, MEK inhibitor (PD0325901, 1μM) and GSK3 inhibitor (CHIR99021, 3μM) at 37°C with 5% CO_2. Conventional reprogramming media consisted of

DMEM supplemented with 15% FBS, 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin, sodium pyruvate (1 mM), L-glutamine (4 mM), 1000 U/ml LIF, 0.1 mM 2-mercaptoethanol unless otherwise noted. For some experiments, media was supplemented with MEK inhibitor (1μM), GSK3 inhibitor (3 μM), Dot1l inhibitor (1 μM) or ascorbate (50 μg/ml).

Reprogrammable MEFs containing either one or two copies of the Col1a1::tetOP-OKSM, Oct4-GFP and Rosa26 M2rtTA alleles[10] were derived from E13.5 embryos. MEFs were prepared after carefully excluding internal organs, heads, limbs and tails. Tissues were chopped into small clumps using scalpels and trypsin and subsequently expanded in MEF medium at low O_2 (4%). MEFs were frozen at passage 0 upon derivation and used at passages 1-3 for all downstream transduction and reprogramming experiments. MEFs were generally cultured at low O_2 (4%) and supplemented with ascorbate to prevent replicative senescence before OKSM induction. Reprogramming experiments were initiated at low oxygen levels during dox induction and completed at normal oxygen levels (20%) for experiments using miR-E vectors. MiR-30 assays were performed under normal oxygen levels.

To generate *Col1a1::tetOP-miR30-tRFP-Ren.713* and *Col1a1::tetOP-miR30-tRFP-Chaf1a.164* shRNA knock-in MEFs, miR30-based shRNAs targeting Chaf1a.164 or Ren.713 were cloned into a targeting vector as previously described[35] except that the GFP reporter was replaced with a turbo RFP reporter. ESCs harboring the R26-M2rtTA allele were targeted with these constructs and mice were generated by blastocyst injection. MEFs were harvested using standard protocols.

HSPCs were isolated from fetal livers of the same mid-gestation reprogrammable transgenic embryos used for MEFs derivation, dissociated by vigorous pipetting with a 1 ml tip, filtered using a 35 μm nylon mesh, followed by red blood cell lysis, and cultured in RPMI/FBS media supplemented with stem cell factor (SCF), IL3 and IL6 and transduced as indicated in the schematic (Fig. 4a).

Arrayed shRNA library preparation and screening

Single shRNA clones were picked from the master library at CSHL, arrayed in 12 × 96-well plates and sequence-verified individually using miR-30 backbone primers. An additional 200 unmatched clones were re-picked and sequenced to allow maximum coverage of the library. Reprogrammable MEFs carrying the OKSM inducible cassette and constitutive rtTA *(Col1a1::tetOP-OKSM; R26-M2rtTA)* were seeded at 10e4 cells per well in 96 well plates in duplicates and infected with the corresponding retroviral virus particles freshly produced and filtered. 48 hrs post transduction, MEFs from each row of the 96-well plate were trypsinized and transferred to 6-well dishes coated with 0.2% gelatin in standard reprogramming media supplemented with dox and G418 at 0.2 mg/ml for the first 6 days of OKSM expression. Dox was withdrawn at day 12, allowing stable iPSCs to form. iPSC colonies were then stained for alkaline phosphatase expression using the Vector Red Alkaline Phosphatase Substrate Kit (VectorLabs) according to the supplier's protocol, and plates were scanned using a Perfection V500 Photo scanner (Epson). To determine relative reprogramming efficiencies (Fig. 1c, Supplementary Table 1), automated counting of iPSC

colonies was performed using the image-processing software CL-Quant (Nikon) and a custom algorithm provided by NIKON. Data were normalized to Ren.713 control.

Pooled shRNA library preparation and screening

A miR-E based chromatin library comprising 5,049 sequenced-verified shRNAs targeting 615 known and predicted chromatin regulators was constructed by subcloning pools of sequence-verified miR-30 shRNAs into pLENC and combining them at equimolar concentrations into one pool[12]. This pool was transduced into MEFs carrying the *Col1a1::tetOP-OKSM* and *R26-M2rtTA* alleles, as well as a Pou5f1-EGFP reporter (termed Oct4-GFP) under conditions predominantly yielding a single retroviral integration in the genome. To generate a large number of independent biological replicates, primary MEFs from 4 triple transgenic embryos were transduced with the entire pool of 5,049 shRNAmiRs in 12 independent replicates at a representation of >100 cells per shRNA, yielding a total of 48 replicates (see Fig. 1 b). After 36 hrs, MEFs were treated with 0.5 mg ml^{-1} G418 for 3 days and 0.25 mg ml^{-1} G418 for an additionally 3 days. MEFs from each replicate were plated at densities of 500,000 cells per 15cm dish 3 or 6 days post transduction, and induced with dox for 7 days in serum/LIF containing media supplemented with ascorbate. After passaging for an additional 4 days in dox-free ESC media, Oct4-GFP expressing cells were sorted from each replicate using a FACSAriaIII (BD Bioscience).

Genomic DNA from infected MEFs (3d after infection) and sorted Oct4-GFP iPSC from each replicate was isolated using proteinase K lysis, followed by two rounds of phenol extraction using PhaseLock tubes (5prime) and isopropanol precipitation. Templates for deep-sequencing were generated by PCR amplification of shRNA guide strands using primers that tag the product with standard Illumina adapters (p7+loop, CAAGCAGAAGACGGCATACGA[INDEX]TAGTGAAGCCACAGATGT; p5+PGK, AATGATACGGCGACCACCGATGGATGTGGAATGTGTGCGAGG). For each sample, DNA was amplified in 12 parallel 50 μl PCR reactions using Encyclo Polymerase (Evrogen). PCR products were combined for each sample, precipitated and purified on a 2% agarose gel. Samples were analyzed on an Illumina High Seq 2500 and sequenced using a primer that reads in reverse into the guide strand (mirEEcoR1Seqprimer, TAGCCCCTTGAAGTCCGAGGCAGTAGGCA). Sequence processing was performed using a customized Galaxy platform. In all 96 iPSC samples (48 biological replicates, 3 or 6 days knockdown prior to OKSM expression) the normalized reads of each shRNA were divided by the normalized reads in MEFs 3 days after viral transduction, and the resulting ratio was used to calculate a score for each shRNA in each replicate (default score=0; score=1 if ratio>1, score=3 if ratio>10). Scores of each shRNA in 48 replicates were added separately for the day 3 and day 6 time point, yielding a sum score to estimate the overall enrichment of each shRNA over all replicates. All shRNA sequences and primary results from the arrayed and the multiplexed screen are provided in Supplementary Tables 1 and 2, respectively.

Retrovirus production, transduction of MEFs and derivation of iPSCs

Retroviral constructs were introduced into Platinium-E Retroviral Packaging cells using calcium phosphate transfection or lipofection as previously described[36]. shRNAs were

transduced into primary MEFs carrying single copies of the *Col1a1::tetOP-OKSM* and *R26-M2rtTA* alleles as well as the Oct4-EGFP reporter. For some experiments, Oct4-tomato knock-in MEFs were used (Stadtfeld et al., in preparation); the *Oct4-tomato* allele was generated equivalently to the *Oct4-GFP* allele[37]. For transduction, 180,000 cells were plated per well in a 6-well dish; all vectors were transduced in biological triplicate. After 36 hrs, transduced cells were selected with 0.5 mg ml^{-1} G418 for 3 days and 0.25 mg ml^{-1} G418 for an additional 3 days. 3 days after shRNA transduction, infected cells were washed with PBS (1x) and trypsinized with Trypsin-EDTA (1x) and 20,000 cells were plated into a 6-well. OKSM expression was induced for 7 days and cells were cultured in DMEM supplemented with 15% FBS, 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin, sodium pyruvate (1 mM), L-glutamine (4 mM), 1000 U/ml LIF, 0.1 mM 2-Mercaptoethanol, 50 μg ml^{-1} sodium ascorbate and 1 μg ml^{-1} doxycycline at 37°C with lox oxygen (4.5% O_2) and 5% CO_2. After 7 days of OKSM expression, cells were cultured for an additional 4 days without doxycycline to withdraw OKSM transgene expression at 37°C with 5% CO_2, ambient oxygen. Following trypsinization, cells were analyzed for Oct4-GFP expression using a FACS BD LSRFortessa (BD Biosciences), data were analyzed using FlowJo.

Phenotypic characterization of iPSCs

Alkaline phosphatase activity was measured using an enzymatic assay for alkaline phosphatase (VECTOR Red Alkaline Phosphatase (AP) Substrate Kit) according to the manufacturer's protocol. Nanog immunohistochemistry of iPSC colonies was performed as previously described[15] using anti-Nanog antibody (ab80892, abcam) at a dilution of 1:500. Cells were permeabilized with 0.2% Triton-X before blocking and antibody incubation.

Chimera production and germline transmission assays

Triple transgenic MEFs were reprogrammed as described, using a tetOP-inducible shRNAmiR expression vector RT3CEPIN (TRE3G-mCherry-mirE-PGK-Puro-IRES-Neo). Oct4-GFP^+ iPSCs generated with experimental shRNAs were sorted on day 7 of OKSM transgene expression. iPSCs were plated on feeders and cultured in DMEM supplemented with 13% Knockout Serum Replacement, 2% FBS, 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin, sodium pyruvate (1 mM), L-glutamine (4 mM), L-ascorbic acid (50 μM), 1000 U/ml LIF, beta-mercaptoethanol, MEK inhibitor (PD0325901, 1μM) and GSK3 inhibitor (CHIR99021, 3μM) at 37°C with 5% CO2. Polyclonally derived iPSCs were microinjected into B6 albino blastocysts to allow identification of chimeras based on coat color markers. Male chimeras were mated to B6 albino females to allow identification of germline transmission based on coat color.

Double knockdown assay

Triple transgenic reprogrammable MEFs were transduced with shRNA expressed from LEPC as previously described and cultured in MEF media. 3 days after retroviral infection, cells were sorted for mCherry expression and 40,000 cells were re-plated per well of a 6-well dish. On the next day, cells were infected with the corresponding second shRNA expressed from LENC. 24 hrs later, cells were cultured in DMEM supplemented with 15% FBS, 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin, sodium pyruvate (1 mM), L-glutamine (4 mM), 1000 U/ml LIF, 0.1 mM 2-Mercaptoethanol and 1 μg ml^{-1}, 50 μg ml^{-1}

sodium ascorbate, and doxycycline at 37°C with lox oxygen (4.5% O_2). 36 hrs after the second shRNA transduction, cell culture media was supplemented with 0.5 mg ml^{-1} G418 for 3 days and 0.25 mg ml^{-1} G418 for an additional 3 days to ensure double infection. After 7 days of OKSM transgene induction, cells were cultured in ESC medium for an additional 4 days without doxycycline to select for transgene-independent colonies at 37°C with 5% CO_2. Cells were analyzed for Oct4-GFP expression using a FACS BD LSRFortessa (BD Biosciences). The effect of double knockdown of targets on iPSC formation was determined by calculating the ratio of Oct4-GFP^+ to Oct4-GFP^- cells at day 11 relative to an empty vector control.

Analysis of reprogramming dynamics

Triple transgenic reprogrammable MEFs carrying the *Col1a1::tetOP-OKSM* and *R26-M2rtTA* alleles as well as a Oct4-GFP reporter were reprogrammed in replicate wells as previously described. Starting on day 4 of OKSM transgene expression, cells were analyzed for Oct-GFP expression in 24 hr intervals using a BD LSRFortessa (BD Biosciences). In addition, 20% of cells were replated and cultured under dox-free ESC conditions. After 13 days, cells were analyzed using a FACS BD LSRFortessa to determine the minimum time required for the establishment of transgene-independent iPSCs.

To determine Nanog expression dynamics, triple transgenic reprogrammable MEFs were reprogrammed in independent wells and analyzed every 24 hrs. Starting on day 4 of OKSM transgene expression, cells were trypsinized with Trypsin-EDTA (1x), washed with PBS (1x) and fixed with Paraformaldehyde (PFA) (4%) for 30 mins. Afterwards, cells were washed with PBS (1x) and stored at 4°C. After 11 days of reprogramming, cells were stained with anti-Nanog antibody (rabbit polyclonal, 1:400, Abcam) and analyzed using a FACS BD LSRFortessa (BD Biosciences).

Pecam staining of reprogramming intermediates was performed as previously described[18]. All samples were analyzed on a MACSQuant fluorescence cytometer (Miltenyi).

CRISPR/Cas9 editing of MEFs

sgRNAs targeting the *Chaf1a* locus were cloned into a lentiviral vector harboring the wild type Cas9 coding region, an sgRNA expression cassette, and a Thy 1.1 reporter transgene. Successfully transduced cells were purified by FACS using Thy1.1 expression, cultured for 7 days to allow for genome editing to occur and induced with dox for one week before measuring the fraction of Oct4-GFP^+ cells at day 11. Single Oct4-GFP^+ iPSCs were then plated to generate clonal iPSC cultures for PCR amplification of CRISPR-Cas9-induced genomic modifications, followed by Sanger sequencing. sgRNA were PCNA-1: GAAGCGCATTAAGGCAGAAA and PCNA-2: TTGGGAGCCTGCGGAGTCTT.

Transdifferentiation assays

Induced neurons were generated as described in the experimental scheme (Fig 4d). CAF-1 or Renilla RNAi inducible transgenic MEFs were transduced with Ascl1-inducible lentivirus, exposed to dox 24 hrs post induction, cultured in MEF media for the first 48 hrs and switched to serum-free neuronal media (N3B27) supplemented with dox for an additional 11

days. Cultures were fixed and stained for MAP2 as previously described[24]. Pre-B cells[25] (C10 line) were cultured in RPMI Medium, 10% charcoal stripped FBS (Invitrogen), 2 mM L-Glutamine, 100 U ml^{-1} penicillin, 1000 μg ml^{-1} streptomycin, 55 μM beta-mercaptoethanol. Pre-B cells were transduced 48 hrs before initiating macrophage transdifferentiation with estradiol (E2). For transdifferentiation assays, cultures were transduced with lentiviral pLKO vectors obtained from the Broad Insitute's RNAi consortium (empty vector “null control” or vector carrying stem-loop shRNAs targeting Chaf1a and Chaf1b subunits). Following selection of transduced cells with puromycin, cells were seeded at 1e6 cells ml^{-1} and supplemented with E2 and macrophage cytokines (IL3 and CSF) as previously described[25]. All time points were analyzed for Cd14 and Mac1 expression by flow cytometry on the same day.

Quantitative RT-PCR

RNA was extracted (Qiagen RNeasy mini kit) and reverse transcribed (GE illustra ready-to-go RT-PCR beads) according to the supplier's instruction. Quantitative PCR was performed using SybrGreen and a BIO-RAD CFX connect cycler. Primers used were:

b-Act-F: GCTGTATTCCCCTCCATCGTG;

b-Act-R: CACGGTTGGCCTTAGGGTTCAG;

Ube2i-R: GGCAAACTTCTTCGCTTGTGCTCGGAC;

Ube2i-F: ATCCTTCTGGCACAGTGTGCCTGTCC;

Chaf1b-R: GGCTCCTTGCTGTCATTCATCTTCCAC;

Chaf1b-F: CACCGCCGTCAGGATCTGGAAGTTGG;

Chaf1a-R: GTGTCTTCCTCAACTTTCTCCTTGG;

Chaf1a-F: CGCGGACAGCCGCGGCCGTGGATTGC.

SDS-PAGE and Western blot analysis

Whole-cell lysates from reprogramming intermediates were run on 4-20% gradient SDS-polyacrylamide gels and transferred to nitrocellulose membrane (Bio-Rad) by standard methods. Membranes were blocked for 1 h in 5% non-fat dry milk in 1 × TBS with 0.05% Tween-20 (TBST), rinsed, and incubated with primary antibody diluted in 3% BSA in TBST overnight at 4°C. The following primary antibodies were used: anti-Chaf1a (sc-10206, Santa cruz), anti-Chaf1b (sc-393662, Santa Cruz), anti-TBP (ab818, Abcam), anti-Ube2i (4786, Cell signaling), anti-PCNA (D3H8P, Cell signaling), HRP conjugate anti actin (AC-15, Sigma). Blots were washed in TBST, incubated with HRP-conjugated secondary antibodies for semi-quantitative Western blot analysis and IRdye 800CW or IRdye 680RD for quantitative westerns, as indicated. Secondary antibodies for both methods were incubated in 5% milk in TBST for 1 hour at room temperature (except for anti-β-ACTIN-Peroxidase antibody, which was incubated for 15min), and washed again. HRP signal was detected by

Enhanced ChemiLuminescence (Perkin Elmer). Fluorescent infrared signal was detected using LI-COR Odyssey imaging system.

ATAC-seq chromatin assay

To generate ATAC-seq libraries, 50,000 cells were used and libraries were constructed as previously described[27]. Briefly, cells were washed in PBS twice, counted and nuclei were isolated from 100,000 cells using 100 µl hypotonic buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% NP40) to generate two independent transposition reactions. Nuclei were split in half and treated with 2.5 µl Tn5 Transposase (Illumina) for 30 mins. at 37°C. DNA from transposed nuclei was then isolated and PCR-amplified using barcoded Nextera primers (illumina). Library QC was carried out using high sensitivity DNA biolanalyzer assay and qubit measurement and sequenced using paired-end sequencing (PE50) on the Illumina Hi-Seq 2500 platform.

SONO-seq and ChIP-seq chromatin assays

For all ChIP experiments, 10e7 reprogramming intermediates were collected per library. Chromatin precipitation assays were performed as previously described[38] using goat polyclonal anti-Sox2 antibody (AF2018, R&D). Briefly, cells were cross-linked on plate in 1% methanol-free formaldehyde and snap-frozen in liquid nitrogen until processed. Nuclei were isolated using 1 ml of cell lysis buffer (20 mM Tris pH8, 85 mM KCL, 0.5% NP40 and 1X HALT protease inhibitor cocktail), resuspended in nuclear lysis buffer (10 mM Tris-HCL pH7.5, 1% NP40, 0.5% Na deoxycholate, 0.1% SDS, 1X HALT protease inhibitor cocktail) and sonicated using optimized pulses of a Branson sonifier (1 min. ON/OFF pulses for 5 cycles) for ChIP-seq libraries and S220 Covaris sonicator (Settings: duty cycle 5%, intensity 6, cycles/burst 200, pulse length 60s, 20 cycles, 8°C) for SONO-seq input preparations. Sonications were verified for both methods using the 2100 Bioanalyzer. Immunoprecipitations were carried out by first adjusting salt concentration in sheared chromatin to 167mM NaCl and adding antibodies (6 µg of Sox2 antibody) and incubated for 3-4 hrs at 4°C. 50 µl Protein G Dynabeads (Invitrogen) were prepared for each IP reaction by washing 2-3 times in ChIP dilution buffer (16.7 mM Tris-HCl pH 8.1, 167 mM NaCl, 0.01% SDS, 1.1% Triton X-100, 1.2mM EDTA) and added for one additional hour to pull down bound chromatin. Bead complexes were washed 6 times in RIPA buffer (20 mM Tris-HCL pH8.1, 1 mM EDTA, 140 mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% Na deoxyhcholate), then twice with RIPA buffer with high salt concentration (500 mM), then twice in LiCL buffer (10 mM tris-HCL pH8.1, 1 mM EDTA, 1% DOC, 1% NP40, 250 mM LiCL) and twice in TE buffer. Complexes were then eluted and reverse cross-linked in 50 µl ChIP elution buffer (10 mM Tris-HCL pH8, 5 mM EDTA, 300 mM NaCl, 01% SDS) and 8 µl of reverse crosslinking buffer (250 mM Tris-HCl pH 6.5, 1.25 M NaCl, 62.5 mM EDTA, 5 mg ml^{-1} Proteinase K, 62.5 µg ml^{-1} RNAse A) by incubation at 65°C for 6 hrs. DNA was isolated using Ampure SPRI beads and yield quantified using Qubit fluorometer.

ChIP-seq libraries were constructed from 10 ng of immunoprecipitated DNA using the NEBNext ChIP-seq Library Prep Reagent Set for Illumina (New England Biolabs), following the supplier's protocol. Briefly, purified DNA was end-repaired and dA-tailed. Following subsequent ligation of sequencing adaptors, ligated DNA was size-selected to

isolate fragments in the range of 300-550 bp in length using Egels. Adaptor-ligated fragments were enriched in a 14-cycle PCR using Illumina multiplexing primers. Libraries were purified, analyzed for correct size distribution using dsDNA High Sensitivity Chips on a 2100 Bioanalyzer (Agilent), pooled and submitted for single-end 50 bp Illumina GAII high-throughput sequencing.

SONO-seq bioinformatics analysis

The reads were aligned to the mouse genome (mm9) using Bowtie with the unique mapping option[39] (Langmead et al 2009). The smoothed tag density profiles were generated using *get.smoothed.tag.density* function of the SPP R package[40] with a 100-bp Gaussian kernel, a 50 bp step and library size normalization. The positions of promoters and enhancers in ESCs were obtained from a publicly available data set[41]. To access the significance of the difference in the enrichment values between CAF-1 and Renilla knockdown samples, a paired Wilcoxon rank sum test was used.

ATAC-seq bioinformatics analysis

The reads were aligned to mm9 using BWA version 0.7.8 with -q 5 -l 32 -k 2 and paired option[42]. Non-primary mapping, failed QC, duplicates and non-paired reads were filtered. If one paired-end was mapped to one chromosome and the other end was mapped to a different chromosome, the read was not included. The reads aligned to chrM were also removed. Only uniquely mapped reads were used. The read density profiles were generated using 150 bp windows with a 20 bp step and were normalized by the library size. For the comparison between Chaf1 shRNA and Renilla shRNA samples, the read density profiles were further normalized using the mean values of all annotated promoters from mm9. For meta-analysis, the reads from Chaf1a.164 and Chaf1a.2120 knockdown samples were merged. The coordinates of promoters and enhancers in ESCs and MEFs were obtained from a publicly available data set[41]. The coordinates of the super-enhancers for the meta-gene plot were used from a recently published dataset[43]. Each super-enhancer region (with 5kb margins) was divided into 101 bins and the tag density signals were averaged in each bin. Significantly enriched regions were detected using Hotspot[44] with FDR=0.01. A one-sided paired Wilcoxon rank sum test was used for the comparison in the enrichment values between CAF-1 and Renilla knockdown samples. To classify the genomic locations of the peaks (promoters, coding exons, introns, intergenic regions, 5'UTR and 3'UTR), the annotations for mm9 were downloaded from UCSC (https://genome.ucsc.edu/cgi-bin/hgTables). The differential sites between CAF-1 and Renilla knockdown samples were identified using DiffBind with p=0.05 for the consensus ATAC-seq peaks after normalization with TMM (Trimmed mean normalization method)[45]. DiffBind uses statistical routines developed in edgeR[46]. A one-sided paired Wilcoxon rank sum test was used for the comparison in the enrichment values between CAF-1 and Renilla knockdown samples.

Sox2 ChIP-seq bioinformatics analysis

The reads were aligned and tag densities profiles were generated as in SONO-seq analyses. The log2 fold enrichment profiles were generated using *get.smoothed.enrichment.mle* in the SPP R package. The profiles were normalized by the background-scaling method using non-enriched regions. A paired Wilcoxon rank sum test was used for the comparison in the Sox2

enrichment values between Chaf1a and Renilla knockdown samples. For Sox2 peak comparison between CAF-1 and Renilla knockdown samples, reads were first subsampled to make the sequencing depth the same for each condition (number of peaks called tends to increase for greater sequencing depth). The significantly enriched peaks compared to input were detected using the SPP *find.binding.positions* function with e-value = 10. The overlapped peaks were compared with a margin of 200 bp. For unique peaks, we first identified the peaks that were present only in one condition (CAF-1 or Renilla knockdown) and compared the enrichment values (input-subtracted tag counts) between CAF-1 and Renilla knockdown. If the ratios between the enrichment values were >2 fold, we considered the peaks as "unique" for one of the conditions. We used Sox2 ChIP-seq data in ESCs from publicly available data sets[47] and analyzed data in the same as described above.

H3K9me3 ChIP-seq bioinformatics analyses

ChIP-seq data were mapped to the mouse genome (mm9) with Bowtie 0.12.7[39] allowing up to 3 mismatches, retaining uniquely mapping reads. To assess H3K9me3 signal distribution genome-wide, we divided the genome in 5 kb intervals, and for each interval, we calculated the ratio of RPM normalized signal in the IP and input samples. Intervals with less than 10 reads in the input samples (~10% of all) were excluded from further analyses due to low coverage. Intervals overlapping specific regions were extracted using the bedtools suite[48]. RRR annotations were obtained from Matoba et al[29], and signal across all included 5 kb intervals was averaged. For H3K9me3 enrichment over transposable element (TE) bodies, we used the mm10 genome version, as this release contains the most recent TE annotations. We extracted the genomic regions corresponding to TE families annotated in the mm10 RepeatMasker track in the UCSC genome browser (http://genome.ucsc.edu/), and calculated the normalized read counts in IP to input samples for each family. Due to the repetitive nature of TEs, we further validated all results considering reads that map to multiple (up to 10,000) positions in the genome, and scaling read counts by the number of valid alignments. This threshold for multiple mapping positions was chosen as it was previously shown to approximate results obtained allowing unlimited mapping positions, but at a significantly improved computation speed[49]. In all analyses, signal estimates based on uniquely mapping reads and based on reads mapping to multiple genomic positions produced similar results.

Gene expression analysis using microarrays

The microarray data were preprocessed using Affymetrix Expression Console version 1.3.0.187 and normalized by the RMA procedure. The limma Bioconductor package was used to select differentially expressed genes with false discover rate (FDR) ≤ 0.05 and at least 2 fold-change[50]. We performed functional analysis with Gene Set Enrichment Analysis (GSEA)[51] using the limma moderated t-statistic to rank the genes.

Association of ATAC-seq changes with transcriptional changes

ATAC-seq peaks were separately called for CAF-1 and Renilla knockdown at days 0, 3 and 6 as described above. To determine which genes from Supplementary Table 5 may be affected by altered ATAC-Seq signals, we incorporated long-range interaction data between promoters and enhancers based on ChIA-PET analysis in ESCs[52]. If there was no matched pair from the ChIA-PET tables, the regions proximal to the TSSs of genes (<4kb) were

taken. The regions were overlapped with the union ATAC-seq peaks of each conditions. For the overlapped peaks, the enrichment values (log2 tag counts) were compared between CAF-1 and Renilla knockdown samples with two-sided paired Wilcoxon rank sum test.

RNA-seq analysis of genes and transposable element bioinformatics analysis

RNA sequencing data was first pre-processed using Reaper[53] to remove any Illumina adapter sequences and computationally depleted of ribosomal RNA sequences (GenBank identifiers: 18S, NR_003278.3; 28S, NR_003279.1; 5S, D14832.1; and 5.8S, K01367.1) using Bowtie 0.12.7 allowing 3 mismatches[39]. For protein-coding gene expression analyses, pre-processed data was mapped to the mouse genome (mm10) using Bowtie 0.12.7 [39] allowing 3 mismatches, and retaining uniquely mapping reads. Mouse transcript annotations were obtained from RefSeq, and reads corresponding to the exonic regions of each gene were calculated using a custom phyton script. For overlapping genes, reads corresponding to overlapping regions were divided equally. Gene differential expression was analysed using the DESeq R package[54].

For TE expression analyses, data was mapped to the mm10 genome with 0 mismatches and considering reads that map to up to 10,000 genomic positions as in ChIP sequencing analyses. We then calculated the number of reads corresponding to TE regions annotated by the UCSC RepeatMasker track, scaling by the number of valid alignments for each read. Scaled reads for each TE family were summed, and normalized as RPM. Heatmaps were generated using the gplots R package, and differential expression analyses were performed using the DESeq R package[54]. Comparisons of RNA-sequencing results from analyses based on uniquely mapping reads, and based on reads mapping to multiple genomic positions, showed very similar results.

Statistical analyses

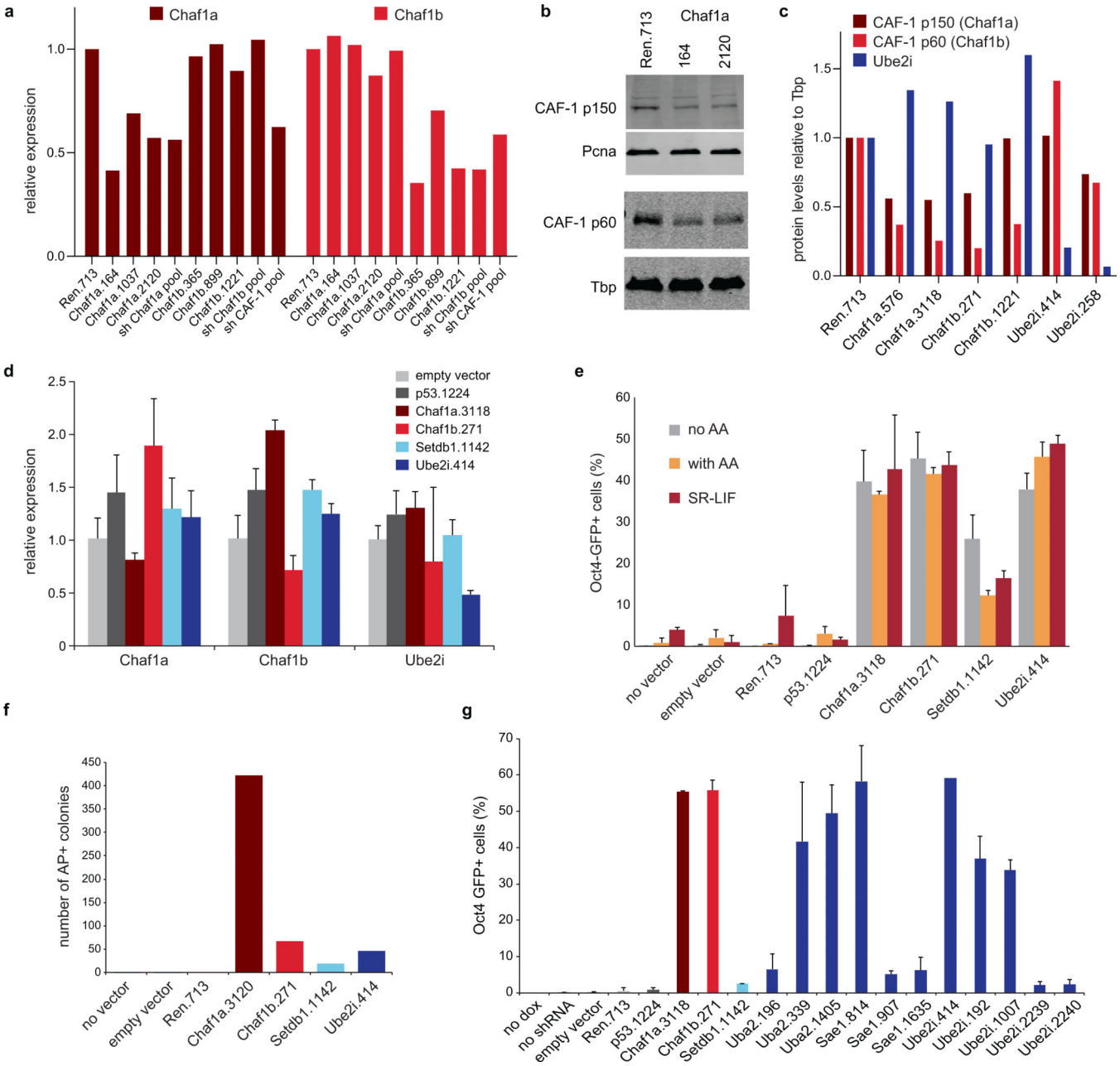
Unpaired Student t test was used for statistical analysis in replicates of cell biology experiments. All error bars represent means ± SEM or STDEV of independent biological replicates as indicated. A p-value of <0.05 was considered statistically significant. Numbers of replicate experiments (n) are shown in figure legends. All graphs with no error bars represent n=1. To assess significant differences in signal enrichment at ESC promoters, enhancers or super-enhancers by SONO-seq, ATAC-seq and ChIP-seq analysis upon CAF-1 knockdown or Renilla knockdown, a paired Wilcoxon rank sum test was used, where it is assumed that populations do not follow normal distributions. To identify differential ATAC-seq peaks between CAF-1 and Renilla knockdown samples, negative binomial models were used.

Chromatin in-vivo assay

CiA transgenic MEFs carrying an array of Gal4 binding sites (*UAS* elements) upstream of the endogenous *Oct4* promoter and a GFP reporter in place of the *Oct4* coding region[30] were immortalized using p53 shRNAs. Two clonal derivates of these MEFs were infected with retroviral vectors expressing Chaf1a, Chaf1b or Renilla shRNA (Figure 5 and data not shown). Cells were subsequently transduced with lentiviral vectors expressing either Gal4

alone or Gal4-VP16 in combination with a puromycin resistance cassette. Following drug selection, Oct4-GFP expression was measured by flow cytometry after 10 days.

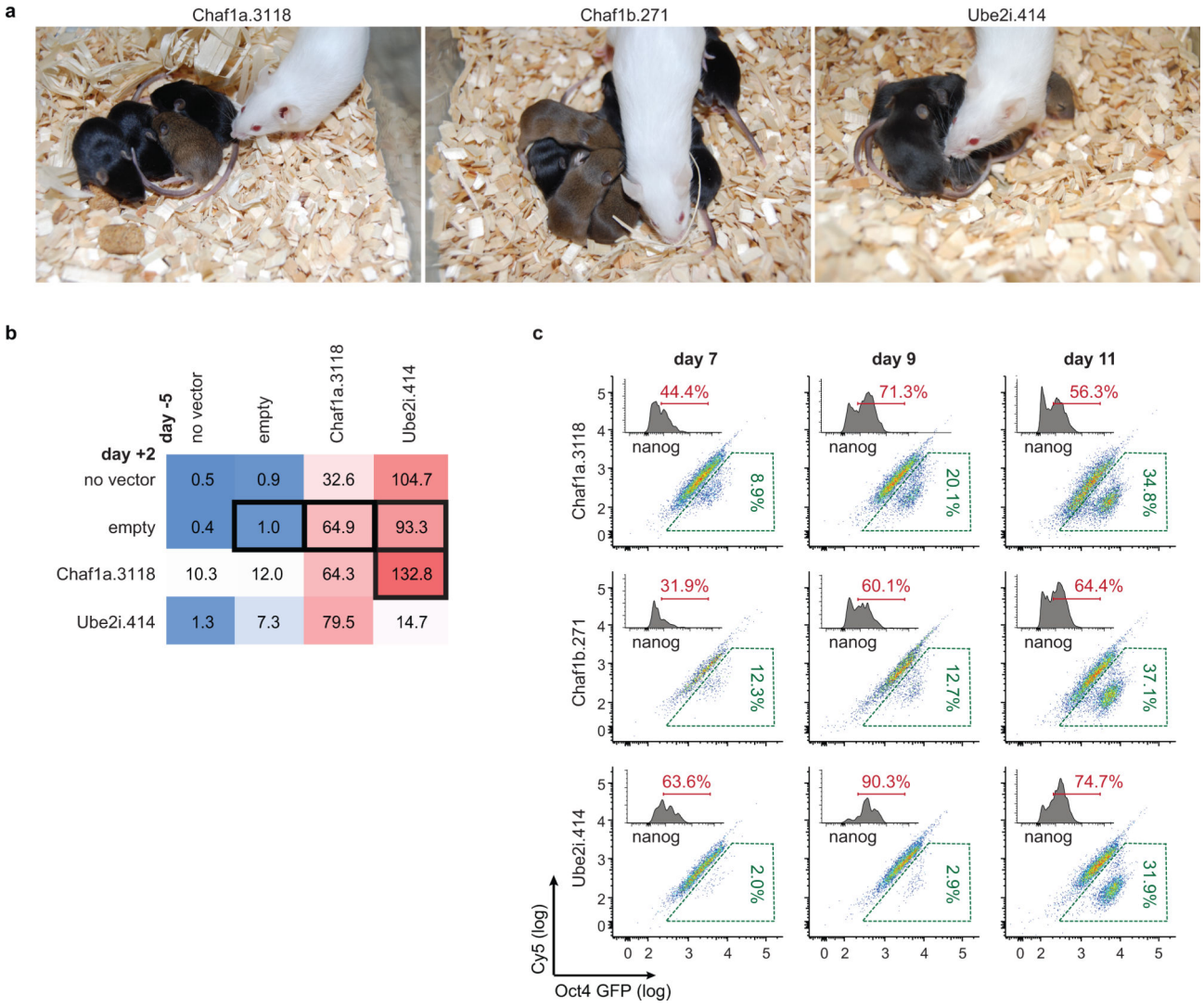
Extended Data



Extended Data Figure 1. Validation of hits from chromatin-focused shRNA screens

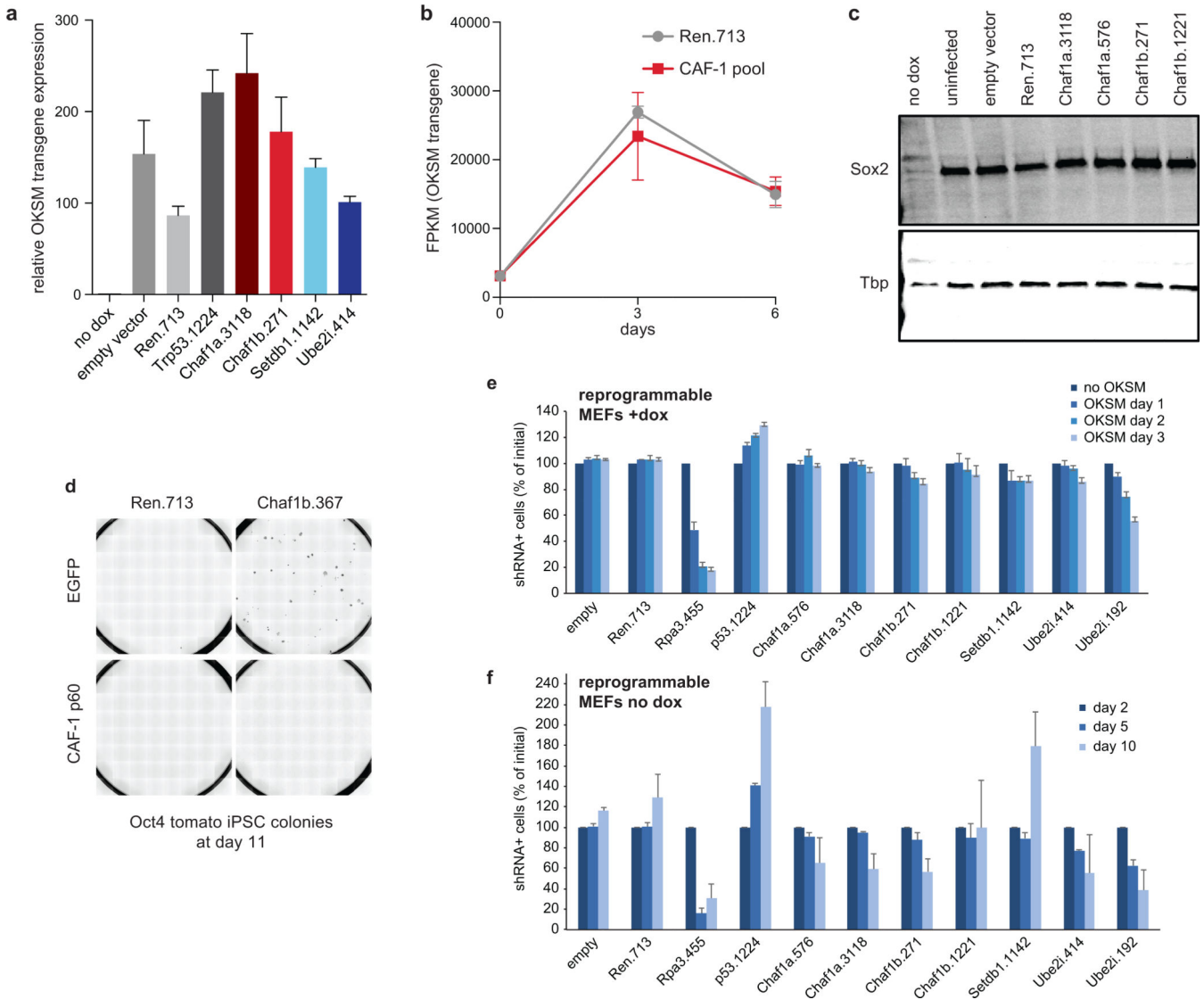
(a) Quantitative RT-PCR analysis to confirm suppression of Chaf1a and Chaf1b expression with miR-30-based vectors from arrayed screen. Sh Chaf1a pool, sh Chaf1b pool and sh CAF-1 pool denote pools of shRNAs targeting either Chaf1a, Chaf1b or both. **(b)** Western blot analysis to confirm knockdown of CAF-1 components using the top-scoring miR-30-

based shRNAs from arrayed screen (see Supplementary Figure 1 for full scans). **(c)** Quantification of data shown in Fig. 1f. **(d)** Quantitative RT-PCR analysis confirming knockdown with top-scoring miR-E-based shRNAmiRs targeting Chaf1a, Chaf1b or Ube2i from multiplexed screen. Error bars show standard deviation from biological triplicates. RNA and protein were extracted from reprogrammable MEFs 72 hours after dox induction in panels a-d. **(e)** Suppression of CAF-1 components, Ube2i and Setdb2 enhances reprogramming in the presence or absence of ascorbic acid (AA) as well as in serum replacement media containing LIF (SR-LIF). Oct4-GFP^{+} cells were scored by flow cytometry on day 11 after 7 days of OKSM induction and 4 days of transgene-independent growth. Error bars show standard deviation from biological triplicates. **(f)** Number of dox-independent, alkaline phosphatase (AP)-positive colonies emerging 2 weeks after plating 10,000 reprogrammable MEFs carrying shRNA vectors against indicated targets and cultured in serum replacement media containing 2i (SR-2i), n=1 experiment. **(g)** Effect of suppressing SUMO E2 ligase Ube2i, E1 ligases Sae1 and Uba2 on iPSC formation. Shown is fraction of Oct4-GFP^{+} cells at day 11 (7 days of OKSM induction, 4 days of transgene-independent growth). Error bars depict standard deviation from biological triplicates.



Extended Data Figure 2. Germline transmission of iPSCs, genetic interaction of shRNA hits and effect of CAF-1 or Ube2i suppression on reprogramming dynamics

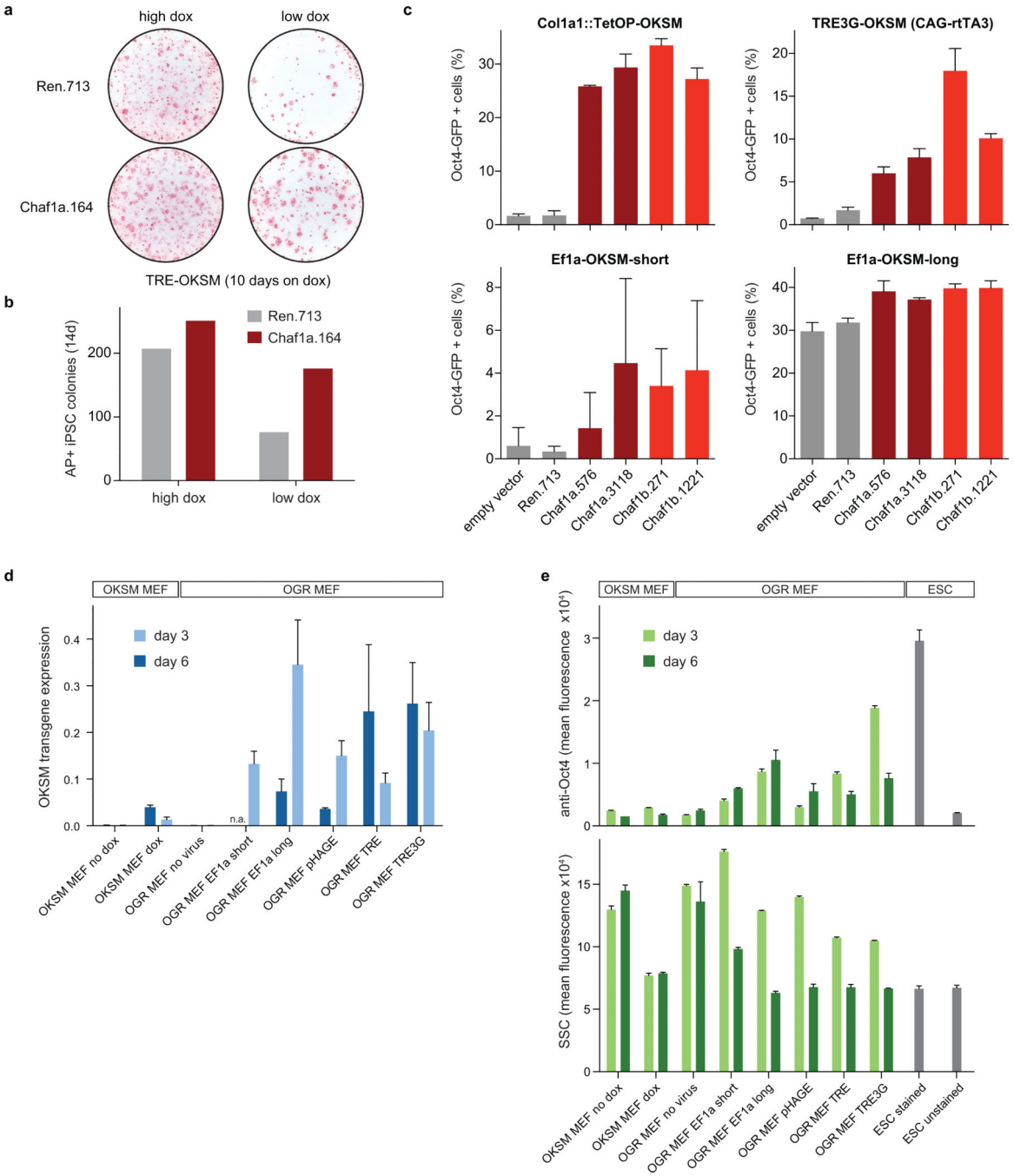
(a) Germline transmission of agouti chimeras generated from iPSCs using dox-inducible shRNA vectors targeting Chaf1a, Chaf1b, or Ube2i. Germline transmission was determined by scoring for agouti coat color offspring upon breeding chimeras with albino females. Germline transmission was observed in 8/8, 4/4 and 6/8 cases for Chaf1a iPSC-derived chimeras, in 7/7, 4/4, 7/7 and 9/9 cases for Chaf1b iPSC-derived chimeras, and in 5/5, 7/7 and 5/5 cases for Ube2i iPSC-derived chimeras. **(b)** Table summarizing effects of co-suppressing pairs of targets on emergence of Oct4-GFP$^+$ cells, shown as the ratio of Oct4-GFP$^+$ to Oct4-GFP$^-$ cells relative to an empty vector control. Experiment equivalent to Fig. 2b except that second shRNAs were transduced two days after induction of reprogramming. **(c)** Representative FACS plots showing effects of Chaf1a/b or Ube2i suppression on emergence of Oct4-GFP$^+$ cells at days 7, 9, and 11 of OKSM expression. Histogram plots show fraction of Nanog$^+$ cells within Oct4-GFP$^+$ cells.



Extended Data Figure 3. Effect of CAF-1 suppression on OKSM levels and cellular growth, and shRNA rescue experiment

(a) Quantitative RT-PCR for transgenic OKSM expression using reprogrammable MEFs transduced with indicated shRNA vectors. Error bars show standard deviation from biological triplicates. **(b)** RNA-seq analysis of OKSM transgene expression in reprogrammable MEFs transduced with Renilla and Chaf1a shRNAs and exposed to dox for 0, 3 or 6 days. Error bars indicate standard deviation from biological triplicates. **(c)** Western blot analysis for Sox2 and Tbp (loading control) in reprogrammable MEFs transduced with shRNA vectors targeting Renilla (Ren.713) or different CAF-1 components and exposed to dox for 3 days (see Supplementary Figure 1 for full scans). The same membrane was probed with anti-CAF-1 p150 and anti-CAF-1 p60 antibody to confirm knockdown (data not shown). **(d)** Rescue experiment to demonstrate specificity of Chaf1b.367 shRNA vector. Reprogrammable MEFs carrying Oct4-tomato knock-in reporter were infected with lentiviral vectors expressing either EGFP or human CAF-1 p60 (CHAF1B) before transducing cells with Renilla or Cha1fb.367 shRNAs and applying dox for 6 days. Colonies

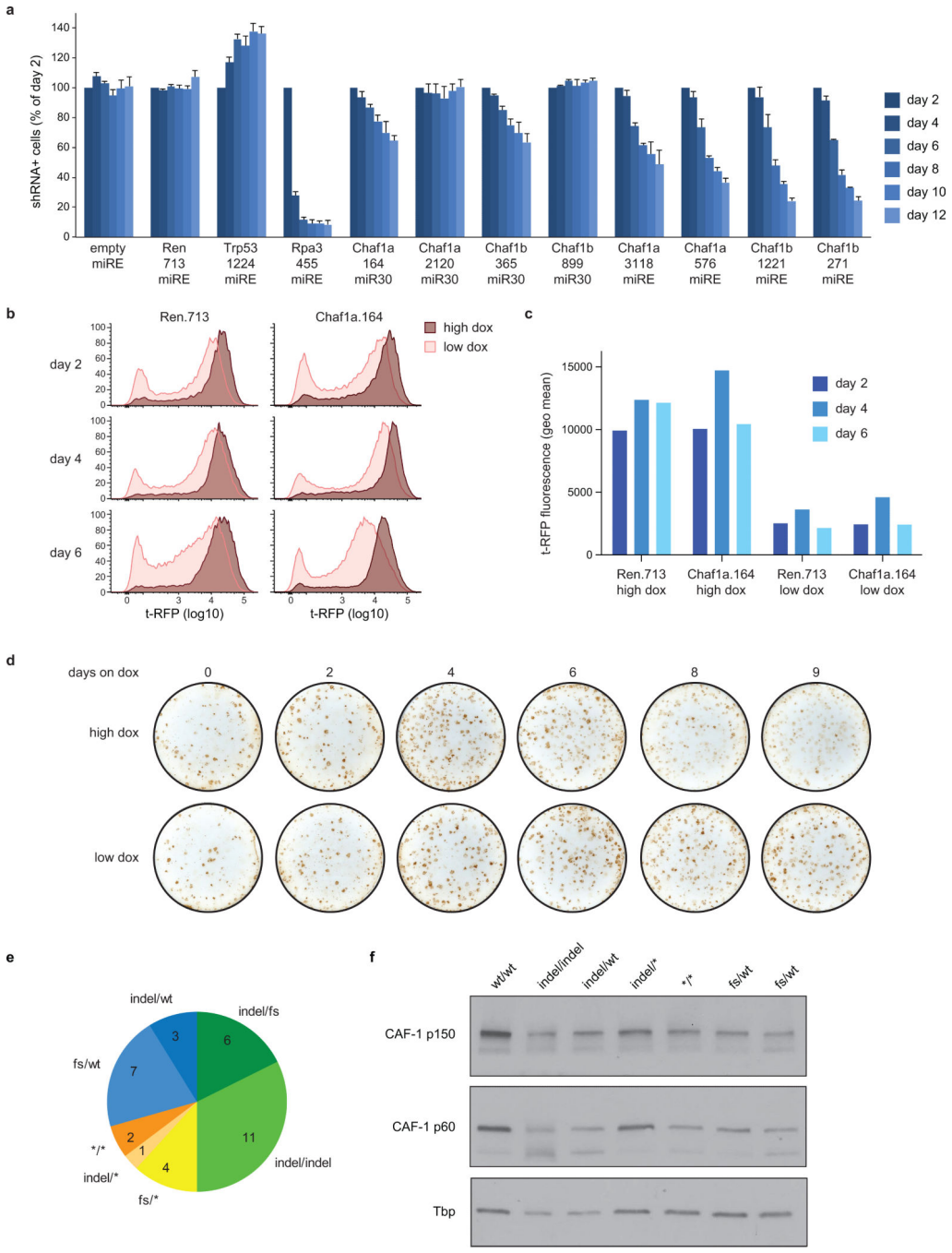
were counted at day 11. Note that CAF-1 p60 overexpression attenuates enhanced reprogramming elicited by Chaf1b suppression. **(e,f)** Competitive proliferation assay between shRNA vector-infected and non-infected reprogrammable cells using indicated shRNAs in the presence or absence of dox (OKSM expression). Note that CAF-1 suppression does not substantially affect the proliferation potential of reprogrammable MEFs after 1-3 days of dox (OKSM) induction while it impairs the long-term growth potential of uninduced MEFs. Data were normalized to cell counts in “no OKSM” condition for (e) and “day 2” time point for (f). Error bars show standard deviation from biological triplicates.



Extended Data Figure 4. Confirmation of CAF-1 reprogramming phenotype with alternative transgenic and non-transgenic vector systems

(a) Alkaline phosphatase (AP)-positive, transgene-independent iPSC colonies at day 14 following transduction of R26-M2rtTA MEFs with tetO-STEMCCA lentiviral OKSM expression vector and either Chaf1a.164 or Ren.713 shRNA vectors and treatment with high (2 μg/ml) or low (0.2 μg/ml) doses of dox for 10 days. **(b)** Quantification of data shown in (a). Experiment was performed at 3 different plating densities (n=1 experiment per density), representative data are shown. **(c)** Comparison of reprogramming efficiencies between *Col1a1::tetOP-OKSM; R26-M2rtTA* reprogrammable MEFs and wild type MEFs infected

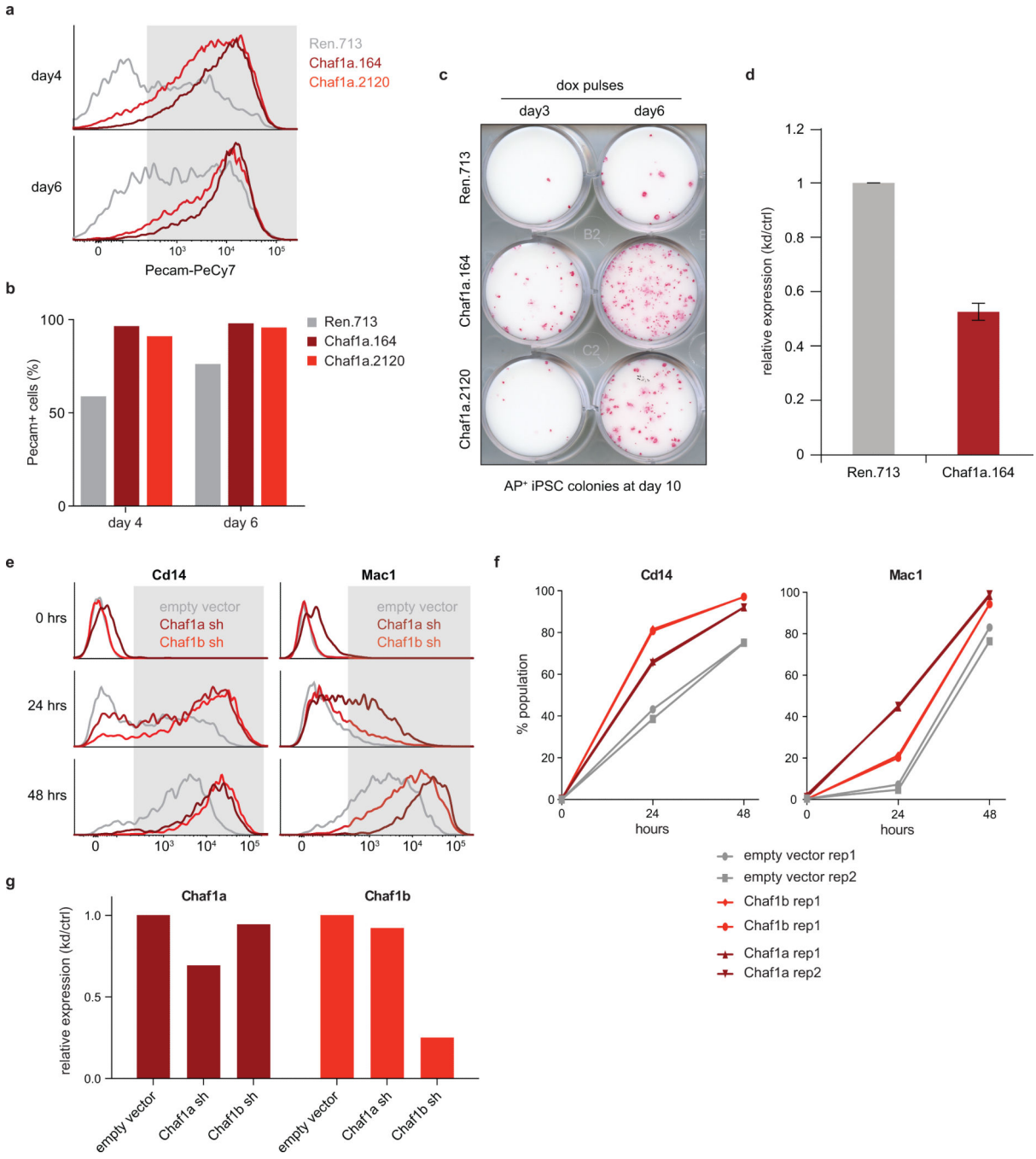
directly with OKSM-expressing lentiviral vectors containing either a strong Ef1a full-length promoter (Ef1a-OKSM long) or a weaker truncated promoter (Ef1a-OKSM short). TRE3G-OKSM is a lentiviral vector with a strong promoter, whose activity is downregulated over time upon infection of CAGS-rtTA3 transgenic MEFs (see below). Error bars show standard deviation from biological triplicates. **(d)** Quantitative RT-PCR data showing variability in OKSM expression levels over time using different vector systems. Cells were analyzed after 3 and 6 days of infection (lentiviral vectors) or dox exposure (reprogrammable MEFs). Error bars show standard deviation from biological triplicates. OGR MEF, transgenic MEFs carrying Oct4-GFP and CAGS-rtTA3 alleles. **(e)** Quantification of Oct4 protein levels by intracellular flow cytometry (top) and cellular granularity/complexity by side scatter (SSC) analysis of indicated samples (bottom). Error bars show standard deviation from biological triplicates.



Extended Data Figure 5. Effects of CAF-1 dose on NIH3T3 growth and reprogramming potential

(a) Competitive proliferation assay to determine effect of indicated Chaf1a and Chaf1b shRNA vectors on long-term growth potential of immortalized NIH3T3 cell line. Cells were infected with indicated constructs and the fraction of shRNA vector-positive cells was measured by flow cytometry at different time points. Data were normalized to cell counts at day 2 post transduction. Rpa3.455, control shRNA that induces apoptosis. Error bars show standard deviation from biological triplicates. **(b)** Histogram plots of MEFs harboring *R26-M2rtTA* allele and either *Col1a1::tetOP-miR30-tRFP-Ren.713* or *Col1a1::tetOP-miR30-*

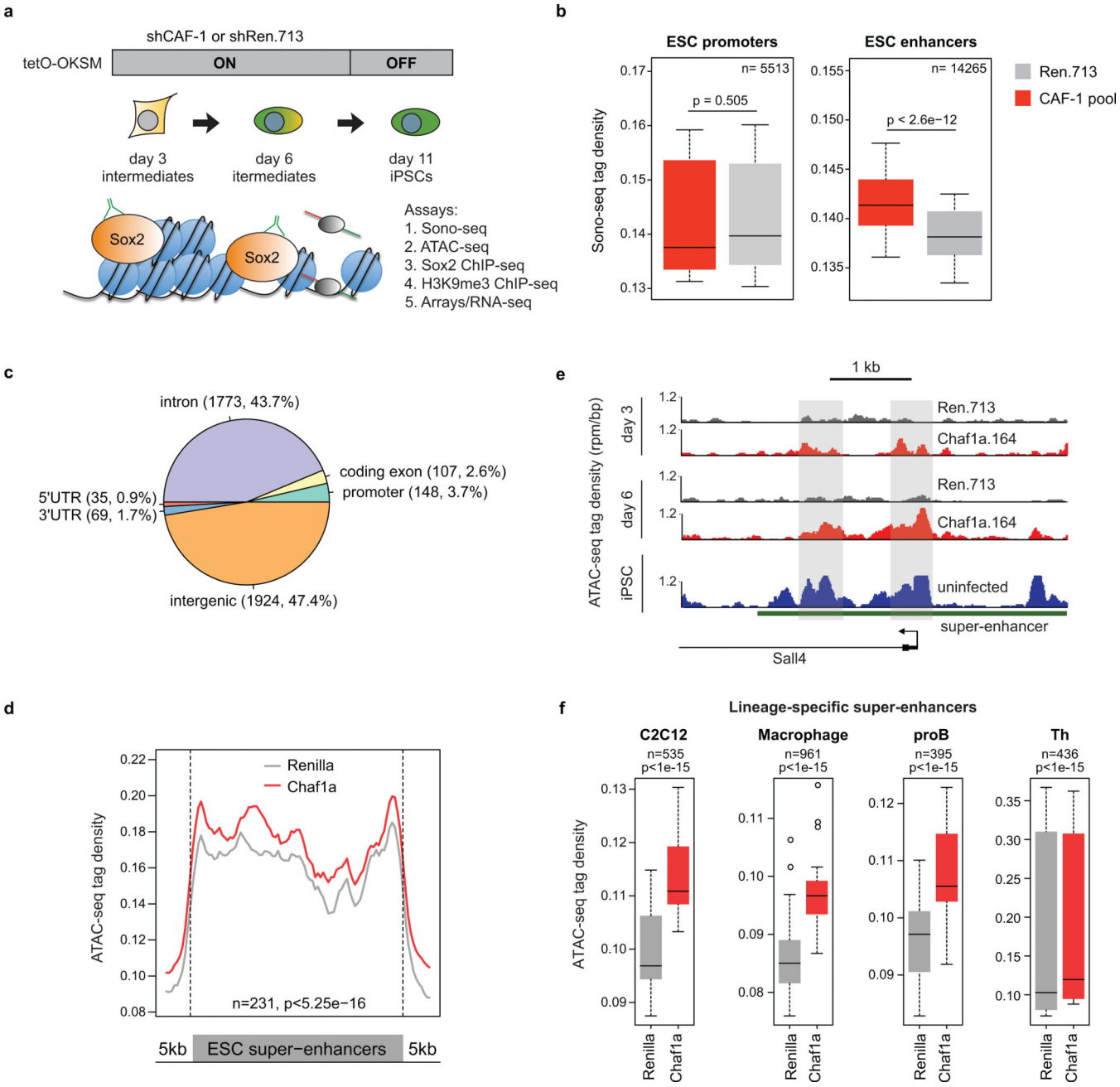
tRFP-Chaf1a.164 shRNA knock-in allele after transduction with pHAGE (Ef1a-OKSM) lentiviral vector and exposure of cells to different doses of dox for 2, 4 and 6 days. Low doses of dox result in lower expression of the shRNAmiR cassettes than high doses of dox. **(c)** Quantification of data shown in (b) using the geometric mean (n=1 experiment for 3 indicated time points). **d)** Reprogramming efficiency of *Col1a1::tetOP-miR30-tRFP-Chaf1a.164; R26-M2rtTA* MEFs infected with pHAGE (Ef1a-OKSM) vector and induced with high (2 μg/ml) or low (0.2 μg/ml) doses of dox for indicated number of days before scoring for $Nanog^+$ iPSCs by immunocytochemistry on day 9. **(e)** Classification of CRISPR/Cas9-induced mutations by sequence analysis of representative iPSC clones (wt, wild type; indel, insertion/deletion; fs, frameshift; *, point mutation). **(f)** Western blot analysis for CAF-1 subunits p150 and p60 in 6 representative iPSC clones after CRISPR/Cas9-induced modifications of the *Chaf1a* locus (see Supplementary Figure 1 for full scans). Wt/wt samples shows unmodified wild type control samples.



Extended Data Figure 6. Effect of CAF-1 suppression on HSPC reprogramming and transdifferentiation

(a) Gating strategy for determining Pecam$^+$ fraction (shaded area) in panel (b); data identical to Fig. 4c. **(b)** Quantification of the fraction of Pecam$^+$ cells at day 4 and day 6 of reprogramming. Data obtained from one experiment using 2 different Chaf1 shRNAs. **(c)** Transgene dependence assay during the reprogramming of hematopoietic stem and progenitor cells (HSPCs) into iPSCs in the presence of Chaf1a or Renilla shRNAs. Dox pulses were given for 3 or 6 days and alkaline phosphatase (AP)-positive colonies were

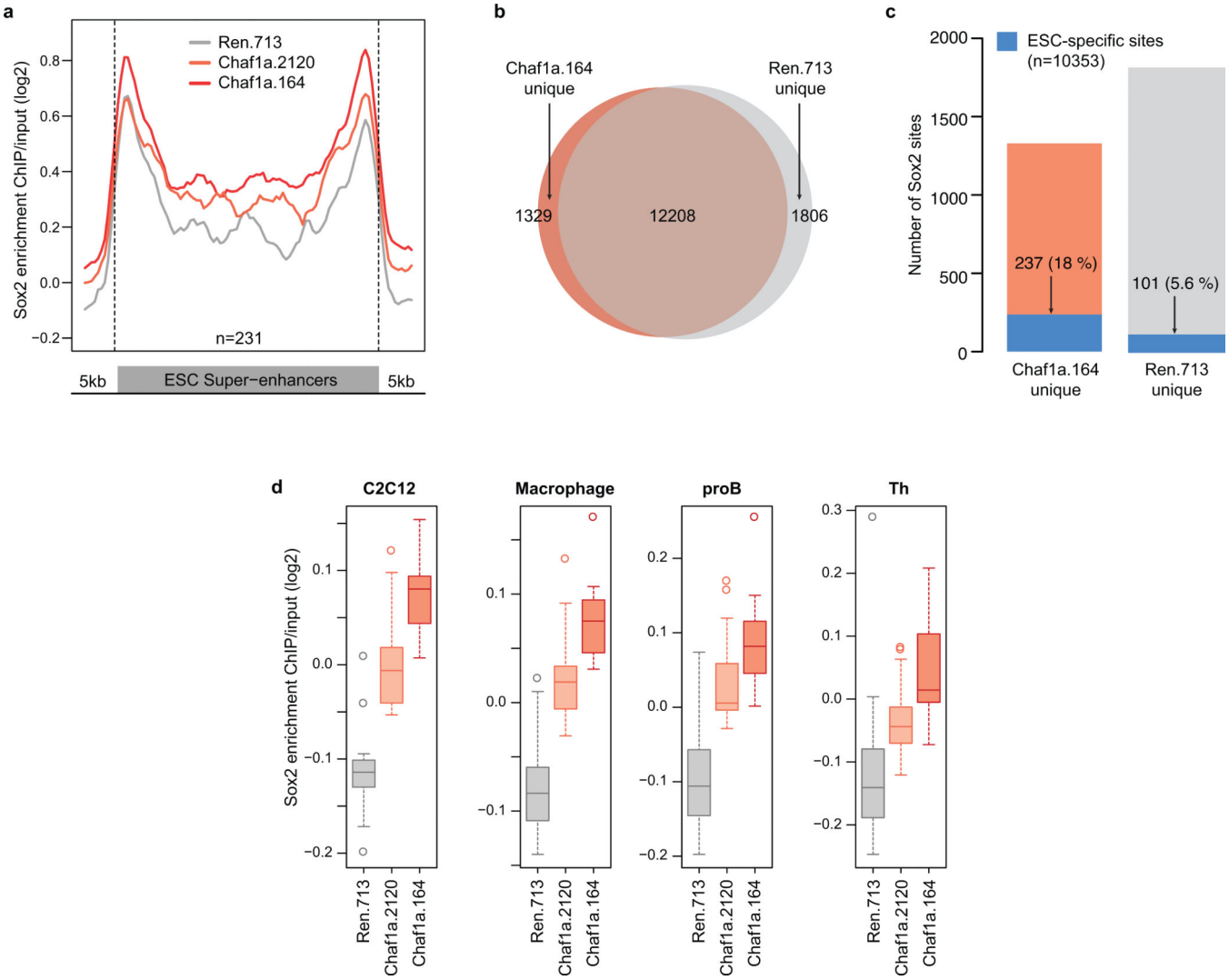
scored at day 10. **(d)** Quantitative RT-PCR analysis of Chaf1a expression to confirm knockdown after 3 days of dox induction, i.e. coexpression of shRNAmiR and Ascl1 (n=4 independent infections of the same Col1a1::tetOP-Chaf1a.164 shRNA MEF line; mean value +/− standard deviation). **(e)** Gating strategy for determining $Cd14^+$ and $Mac1^+$ fractions (shaded area) shown in (f); data identical to Fig. 4g. Positive gates were based on untreated (0 hour) control cells. **(f)** Quantification of the fraction of $Cd14^+$ and $Mac1^+$ cells at 0, 24 and 48 hours of transdifferentiation using indicated CAF-1 shRNA or empty control vector (n=2 independent infections; rep, replicate). **(g)** Quantitative RT-PCR analysis of Chaf1a and Chaf1b expression to confirm knockdown in transduced pre-B cell line prior to induction of transdifferentiation (kd/ctrl, knockdown/empty vector control; n=1 experiment, representative of 2 independent infections).



Extended Data Figure 7. CAF-1 suppression promotes chromatin accessibility at enhancer elements

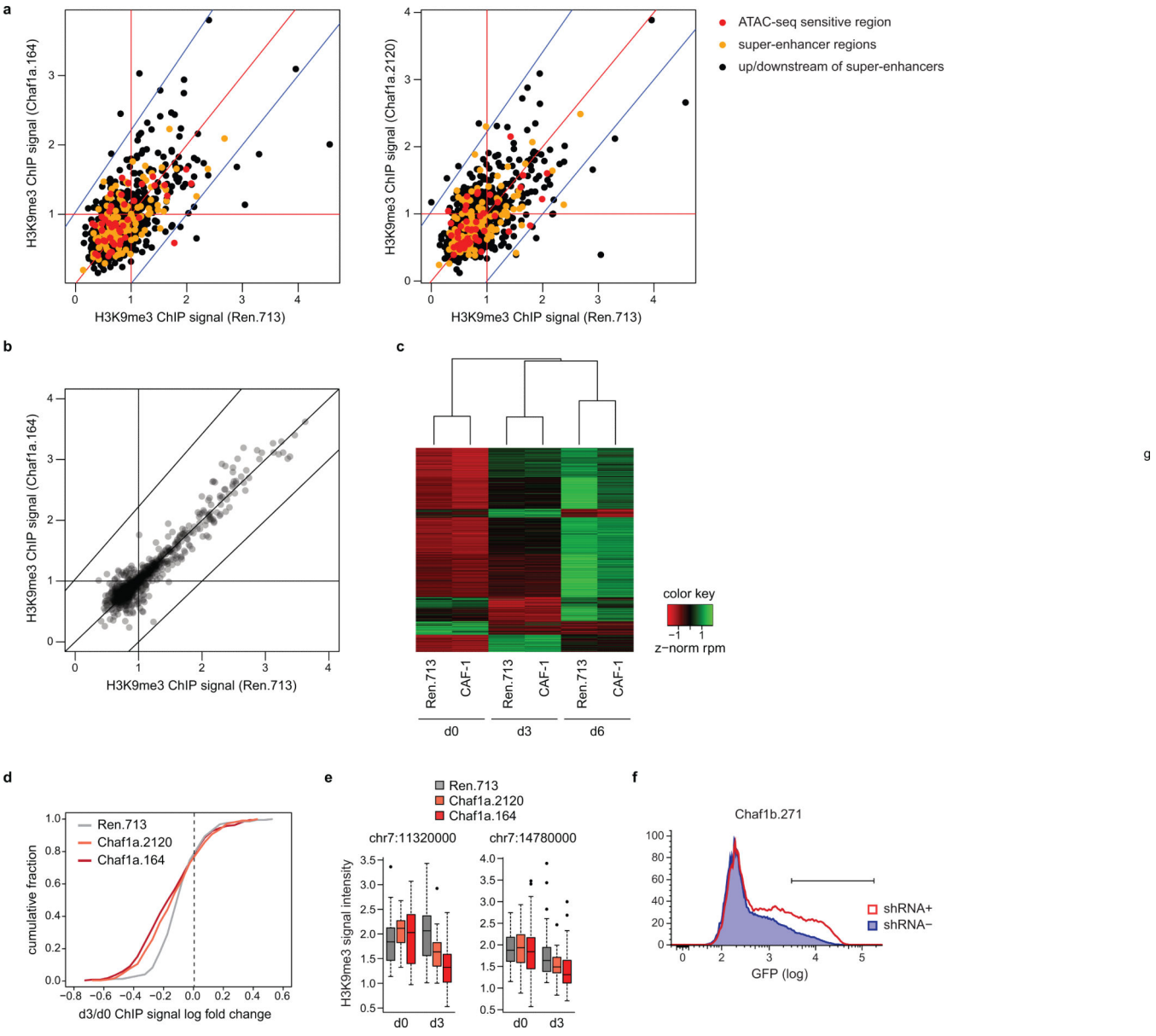
(a) Experimental outline and assays (SONO-seq, ATAC-seq, Sox2 ChIP-seq, H3K9me3 ChIP-seq, microarrays and RNA-seq) to dissect effect of CAF-1 suppression on chromatin accessibility, transcription factor binding, heterochromatin patterns and gene expression. Assays were performed either in early reprogramming intermediates (day 3) or throughout the reprogramming time course (ATAC-seq and gene expression). **(b)** SONO-seq analysis of CAF-1 knockdown and control cells at day 3 of reprogramming to determine accessible chromatin regions across promoters (n=5, 513) and ESC-specific enhancers (n=14, 265). CAF-1 shRNA vectors Chaf1a.164, Chaf1a.2120, Chaf1b.365 and Chaf1b.1221 were pooled

for this experiment. **(c)** ATAC-seq peak distribution across different genomic features. Shown is classification of peaks that are gained in CAF-1 knockdown cells compared to Renilla. **(d)** ATAC-seq analysis of Chaf1a and Renilla control cells at day 3 of reprogramming to measure global chromatin accessibility over pluripotency-specific super-enhancer elements. ATAC-seq data from Chaf1a.164 shRNA- and Chaf1a.2120 shRNA-transduced cells were merged for this analysis. **(e)** ATAC-seq accessibility maps at super-enhancer elements associated with the *Sall4* locus. Shaded grey bars highlight more accessible sites in Chaf1a knockdown samples at days 3 and 6 of reprogramming compared to Renilla shRNA controls. **(f)** ATAC-seq analysis of Chaf1a and control cells at day 3 of reprogramming to measure global chromatin accessibility over lineage-specific super-enhancer elements[43] (C2C12, myoblast cell line; proB, progenitor B cells; Th, T helper cells). N denotes number of examined enhancer elements for each cell type. ATAC-seq data from Chaf1a.164 shRNA- and Chaf1a.2120 shRNA-transduced cells were merged for this analysis.



Extended Data Figure 8. CAF-1 suppression facilitates Sox2 binding to chromatin

(a) Sox2 ChIP-seq enrichment across pluripotency-specific super-enhancer elements at day 3 of reprogramming in the presence of indicated shRNA vectors. **(b)** Venn diagram depicting shared and unique Sox2 targets in Chaf1a and Renilla knockdown cells. **(c)** Bar graph shows the number and fraction of ESC-specific Sox2 targets (blue color) among Sox2-bound sites that are unique to Chaf1a or Renilla knockdown cells at day 3 of OKSM expression. **(d)** Sox2 ChIP-seq analysis of Chaf1a and control shRNA-infected cells at day 3 of reprogramming to determine enrichment of Sox2 binding across lineage-specific super-enhancer elements[43] (C2C12, myoblast cell line; proB, progenitor B cells; Th, T helper cells; P value < 10^{-15} for all comparisons between Chaf1a knockdown cells and control).



Extended Data Figure 9. CAF-1 suppression induces specific depletion of H3K9me3 at somatic heterochromatin domains

(a) Scatter plots comparing H3K9me3 enrichment nearby ATAC-seq sensitive and super-enhancer regions between control (Ren.713) and Chaf1a knockdown cells (Chaf1a.164 and Chaf1a.2122) at day 3 of reprogramming. Values reflect normalized H3K9me3 ChIP signal (IP/Input) for 5 kb genomic regions overlapping ATAC-seq sensitive regions (red), super-enhancer regions (orange) and regions within 50 kb upstream and downstream of super-enhancers (black). **(b)** Scatter plots comparing H3K9me3 enrichment over transposable element (TE) families in control and Chaf1a knockdown cells at day 3 of reprogramming. Values reflect normalized H3K9me3 ChIP-seq signal (IP/Input) over families of TEs in the mouse genome. **(c)** Heatmap shows the relative changes (z-normalized) of TE family expression as estimated by RNA sequencing in control and Chaf1a knockdown cells at day 0, 3 and 6 of reprogramming. Data is clustered using the k-means algorithm. **(d)** Cumulative histogram showing the relative fraction of RRRs (x-axis) that display negative or positive enrichment (fold-change) of average H3K9me3 signal at day 3 of reprogramming in control and Chaf1a knockdown cells. Note that more RRR regions exhibit depletion of H3K9me3 in Chaf1a knockdown samples. **(e)** H3K9me3 ChIP-seq analysis of reprogramming-resistant regions (RRRs)[29] after 0 and 3 days of reprogramming. Box plots depict representative RRRs on chromosome 7 ($p<0.05$ for both shRNAs). See also Fig. 5d. **(f)** Histogram plot showing activation of UAS-Oct4-GFP transgene upon suppression of Chaf1b ($shRNA^+$ line) in the presence of Gal4-VP16 fusion protein. See Fig. 5f for quantification.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Drs. Bob Kingston, Chris Vakoc, Michael Tolstorukov and Greg Hannon for guidance and discussions, Drs. Brad Bernstein, Kathrin Plath, Konstantinos Chronis, Ying Shen and Oliver Tam for advice on the ATAC-seq analysis, Dr. Peter Brown for generously providing the Dot1l inhibitor and Dr. Thomas Graf for sharing the C10 cell line. We thank Chieko Nakada and Yasujiro Kiyota (Nikon) for providing software to quantify iPSC formation and Aaron Huebner for help with transdifferentiation experiments. We are grateful to Dr. Hanno Hock and the HSCI-CRM flow cytometry core for help with flow data analysis. We further thank Beicong Ma, Stefanie Muller, Martina Weissenboeck and the IMP/IMBA Biooptics and Transgenic core facility as well the CSF NGS lab for excellent technical assistance and all members of the Hochedlinger, Zuber, Penninger, Elling, Shi and Kingston labs for their feedback on various aspects of this project. We thank Alexander Stark, Aimee Deaton and Lindy Barrett for critical reading of the manuscript. S.C. was supported by the PRCRP at the Department of Defense (CA 120212). H.Y.C. by was supported by NIH P50-HG007735. J.Z. was funded by an ERC Starting Grant (n° 336860) and generous institutional funding from Boehringer Ingelheim. K.H. was supported by funds from the MGH, HHMI, NIH (R01 HD058013-06) and the Gerald and Darlene Jordan Chair in Regenerative Medicine.

References

1. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. Cell. 2013; 152:1237–1251. doi:10.1016/j.cell.2013.02.014. [PubMed: 23498934]
2. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006; 126:663–676. doi:10.1016/j.cell.2006.07.024. [PubMed: 16904174]
3. Vierbuchen T, Wernig M. Molecular roadblocks for cellular reprogramming. Molecular cell. 2012; 47:827–838. doi:10.1016/j.molcel.2012.09.008. [PubMed: 23020854]
4. Yang CS, Chang KY, Rana TM. Genome-wide functional analysis reveals factors needed at the transition steps of induced reprogramming. Cell reports. 2014; 8:327–337. doi:10.1016/j.celrep.2014.07.002. [PubMed: 25043178]

5. Mikkelsen TS, et al. Dissecting direct reprogramming through integrative genomic analysis. Nature. 2008; 454:49–55. doi:10.1038/nature07056. [PubMed: 18509334]
6. Onder TT, et al. Chromatin-modifying enzymes as modulators of reprogramming. Nature. 2012; 483:598–602. doi:10.1038/nature10953. [PubMed: 22388813]
7. Rais Y, et al. Deterministic direct reprogramming of somatic cells to pluripotency. Nature. 2013; 502:65–70. doi:10.1038/nature12587. [PubMed: 24048479]
8. Dejosez M, Ura H, Brandt VL, Zwaka TP. Safeguards for cell cooperation in mouse embryogenesis shown by genome-wide cheater screen. Science. 2013; 341:1511–1514. doi:10.1126/science.1241628. [PubMed: 24030493]
9. Qin H, et al. Systematic identification of barriers to human iPSC generation. Cell. 2014; 158:449–461. doi:10.1016/j.cell.2014.05.040. [PubMed: 25036638]
10. Stadtfeld M, Maherali N, Borkent M, Hochedlinger K. A reprogrammable mouse strain from gene-targeted embryonic stem cells. Nature methods. 2010; 7:53–55. doi:10.1038/nmeth.1409. [PubMed: 20010832]
11. Zuber J, et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature. 2011; 478:524–528. doi:10.1038/nature10334. [PubMed: 21814200]
12. Fellmann C, et al. An optimized microRNA backbone for effective single-copy RNAi. Cell reports. 2013; 5:1704–1713. doi:10.1016/j.celrep.2013.11.020. [PubMed: 24332856]
13. Smith S, Stillman B. Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication in vitro. Cell. 1989; 58:15–25. [PubMed: 2546672]
14. Chen J, et al. H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nature genetics. 2013; 45:34–42. doi:10.1038/ng.2491. [PubMed: 23202127]
15. Soufi A, Donahue G, Zaret KS. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell. 2012; 151:994–1004. doi:10.1016/j.cell.2012.09.045. [PubMed: 23159369]
16. Sridharan R, et al. Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1gamma in reprogramming to pluripotency. Nature cell biology. 2013; 15:872–882. doi:10.1038/ncb2768. [PubMed: 23748610]
17. Krizhanovsky V, Lowe SW. Stem cells: The promises and perils of p53. Nature. 2009; 460:1085–1086. doi:10.1038/4601085a. [PubMed: 19713919]
18. Polo JM, et al. A molecular roadmap of reprogramming somatic cells into iPS cells. Cell. 2012; 151:1617–1632. doi:10.1016/j.cell.2012.11.039. [PubMed: 23260147]
19. Quivy JP, Gerard A, Cook AJ, Roche D, Almouzni G. The HP1-p150/CAF-1 interaction is required for pericentric heterochromatin replication and S-phase progression in mouse cells. Nature structural & molecular biology. 2008; 15:972–979.
20. Hoek M, Stillman B. Chromatin assembly factor 1 is essential and couples chromatin assembly to DNA replication in vivo. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100:12183–12188. doi:10.1073/pnas.1635158100. [PubMed: 14519857]
21. Houlard M, et al. CAF-1 is essential for heterochromatin organization in pluripotent embryonic cells. PLoS genetics. 2006; 2:e181. doi:10.1371/journal.pgen.0020181. [PubMed: 17083276]
22. Ye X, et al. Defective S phase chromatin assembly causes DNA damage, activation of the S phase checkpoint, and S phase arrest. Molecular cell. 2003; 11:341–351. [PubMed: 12620223]
23. Rolef Ben-Shahar T, et al. Two fundamentally distinct PCNA interaction peptides contribute to chromatin assembly factor 1 function. Molecular and cellular biology. 2009; 29:6353–6365. doi:10.1128/MCB.01051-09. [PubMed: 19822659]
24. Chanda S, et al. Generation of induced neuronal cells by the single reprogramming factor ASCL1. Stem cell reports. 2014; 3:282–296. doi:10.1016/j.stemcr.2014.05.020. [PubMed: 25254342]
25. Bussmann LH, et al. A robust and highly efficient immune cell reprogramming system. Cell stem cell. 2009; 5:554–566. doi:10.1016/j.stem.2009.10.004. [PubMed: 19896445]
26. Auerbach RK, et al. Mapping accessible chromatin regions using Sono-Seq. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:14926–14931. doi:10.1073/pnas.0905443106. [PubMed: 19706456]

27. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature methods. 2013; 10:1213–1218. doi:10.1038/nmeth.2688. [PubMed: 24097267]
28. Huang H, et al. Drosophila CAF-1 regulates HP1-mediated epigenetic silencing and pericentric heterochromatin stability. Journal of cell science. 2010; 123:2853–2861. doi:10.1242/jcs.063610. [PubMed: 20663913]
29. Matoba S, et al. Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation. Cell. 2014; 159:884–895. doi:10.1016/j.cell.2014.09.055. [PubMed: 25417163]
30. Hathaway NA, et al. Dynamics and memory of heterochromatin in living cells. Cell. 2012; 149:1447–1460. doi:10.1016/j.cell.2012.03.052. [PubMed: 22704655]
31. Ishiuchi T, et al. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. Nature structural & molecular biology. 2015 doi:10.1038/nsmb.3066.
32. Ray-Gallet D, et al. Dynamics of histone H3 deposition in vivo reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. Molecular cell. 2011; 44:928–941. doi:10.1016/j.molcel.2011.12.006. [PubMed: 22195966]
33. Jullien J, et al. HIRA dependent H3.3 deposition is required for transcriptional reprogramming following nuclear transfer to Xenopus oocytes. Epigenetics & chromatin. 2012; 5:17. doi:10.1186/1756-8935-5-17. [PubMed: 23102146]
34. Wen D, Banaszynski LA, Rosenwaks Z, Allis CD, Rafii S. H3.3 replacement facilitates epigenetic reprogramming of donor nuclei in somatic cell nuclear transfer embryos. Nucleus. 2014; 5:369–375. doi:10.4161/nucl.36231. [PubMed: 25482190]
35. Premsrirut PK, et al. A rapid and scalable system for studying gene function in mice using conditional RNA interference. Cell. 2011; 145:145–158. doi:10.1016/j.cell.2011.03.012. [PubMed: 21458673]
36. Zuber J, et al. Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. Nature biotechnology. 2011; 29:79–83. doi:10.1038/nbt.1720.
37. Lengner CJ, et al. Oct4 expression is not required for mouse somatic stem cell self-renewal. Cell stem cell. 2007; 1:403–415. doi:10.1016/j.stem.2007.07.020. [PubMed: 18159219]
38. Bernstein BE, et al. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell. 2005; 120:169–181. doi:10.1016/j.cell.2005.01.001. [PubMed: 15680324]
39. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology. 2009; 10:R25. doi:10.1186/gb-2009-10-3-r25. [PubMed: 19261174]
40. Kharchenko PV, Tolstorukov MY, Park PJ. Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nature biotechnology. 2008; 26:1351–1359. doi:10.1038/nbt.1508.
41. Shen Y, et al. A map of the cis-regulatory sequences in the mouse genome. Nature. 2012; 488:116–120. doi:10.1038/nature11243. [PubMed: 22763441]
42. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. doi:10.1093/bioinformatics/btp324. [PubMed: 19451168]
43. Whyte WA, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013; 153:307–319. doi:10.1016/j.cell.2013.03.035. [PubMed: 23582322]
44. Sabo PJ, et al. Discovery of functional noncoding elements by digital analysis of chromatin structure. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101:16837–16842. doi:10.1073/pnas.0407387101. [PubMed: 15550541]
45. Ross-Innes CS, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature. 2012; 481:389–393. doi:10.1038/nature10730. [PubMed: 22217937]
46. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–140. doi:10.1093/bioinformatics/btp616. [PubMed: 19910308]
47. Marson A, et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell. 2008; 134:521–533. doi:10.1016/j.cell.2008.07.020. [PubMed: 18692474]

48. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. doi:10.1093/bioinformatics/btq033. [PubMed: 20110278]

49. Pezic D, Manakov SA, Sachidanandam R, Aravin AA. piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. Genes & development. 2014; 28:1410–1428. doi:10.1101/gad.240895.114. [PubMed: 24939875]

50. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology. 2004; 3 Article3, doi:10.2202/1544-6115.1027.

51. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:15545–15550. doi:10.1073/pnas.0506580102. [PubMed: 16199517]

52. Dowen JM, et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell. 2014; 159:374–387. doi:10.1016/j.cell.2014.09.030. [PubMed: 25303531]

53. Davis MP, van Dongen S, Abreu-Goodger C, Bartonicek N, Enright AJ. Kraken: a set of tools for quality control and analysis of high-throughput sequence data. Methods. 2013; 63:41–49. doi: 10.1016/j.ymeth.2013.06.027. [PubMed: 23816787]

54. Anders S, Huber W. Differential expression analysis for sequence count data. Genome biology. 2010; 11:R106. doi:10.1186/gb-2010-11-10-r106. [PubMed: 20979621]

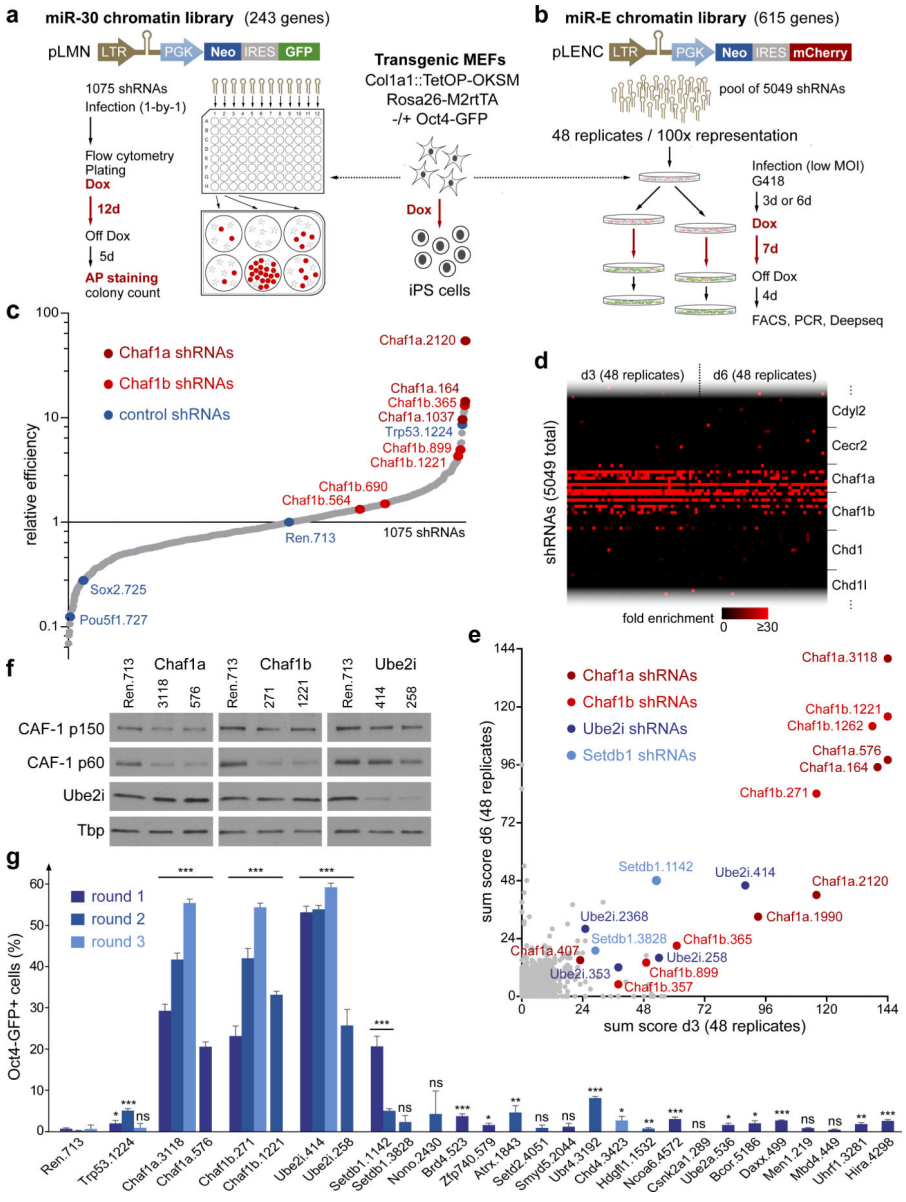


Figure 1. Arrayed and multiplexed shRNAmiR screening strategies to identify suppressors of reprogramming

(a,b) Schematic of arrayed (a) and multiplexed (b) RNAi screens. **(c)** Results from arrayed screen, depicting average reprogramming efficiency ratios of two biological replicates normalized to Renilla (Ren.713) shRNA control. **(d)** Heatmap depicting enrichment of selected shRNAs (shown in rows, ordered by gene symbol) over all 96 replicates (columns). **(e)** Scatter plot representing sum score of enriched shRNAs across all replicates. **(f)** Western blot analysis confirming shRNA suppression of CAF-1 p150 (Chaf1a), CAF-1 p60 (Chaf1b) and Ube2i at day 3 of reprogramming (see Supplementary Figure 1 for full scans). **(g)** Validation of hits from multiplex screen. Error bars indicate standard deviation (SD) from biological triplicates (*, p<0.05; **, p<0.01).

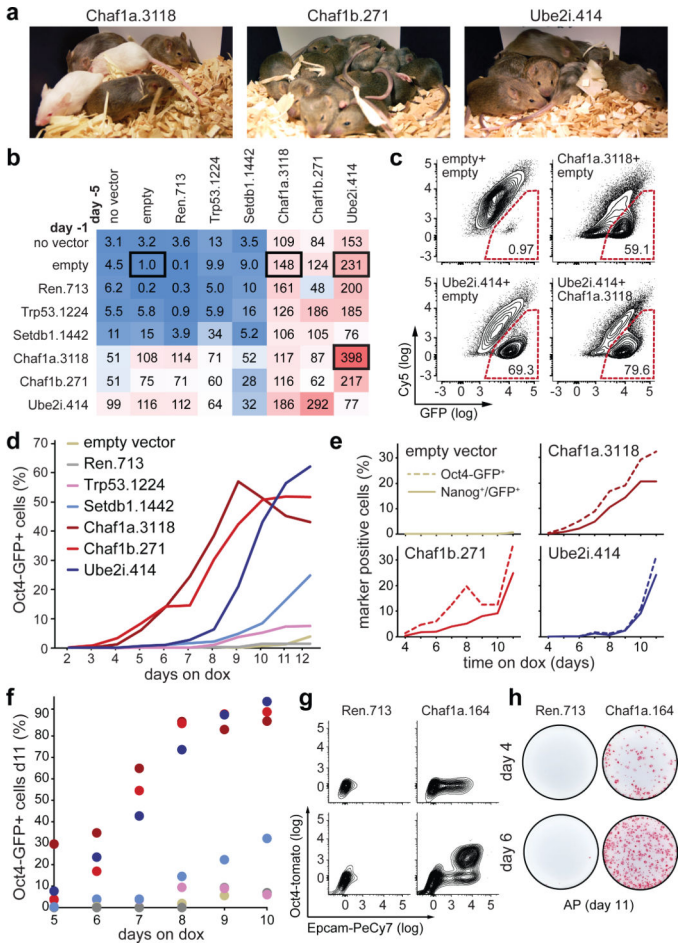


Figure 2. CAF-1 suppression accelerates reprogramming and yields developmentally competent iPSCs

(a) Generation of iPSC-derived chimeras using indicated shRNAmiRs. **(b)** Effect of co-suppression of indicated targets on reprogramming potential of MEFs, shown as ratio of Oct4-GFP^{+} to Oct4-GFP^{-} cells at day 11 relative to an empty vector control. **(c)** Flow cytometry plots of representative samples used for panel (b). **(d,e)** Time course analysis of Oct4-GFP (d) and Nanog (e) expression upon suppression of indicated targets in MEFs undergoing reprogramming. **(f)** Establishment of transgene-independent Oct4-GFP^{+} iPSCs in the presence of depicted shRNAs vectors. Samples were induced with dox for indicated number of days before dox withdrawal and analysis at day 13. **(g)** Expression dynamics of reprogramming markers Epcam and Oct4-tomato after 4 and 6 days of OKSM expression (media supplemented with 2i, ascorbate and Dot1l inhibitor). **(h)** Alkaline phosphatase (AP)-positive, transgene-independent iPSC colonies scored at day 11 after 4 or 6 days of OKSM expression (representative example from 2 biological replicates and 3 technical replicates).

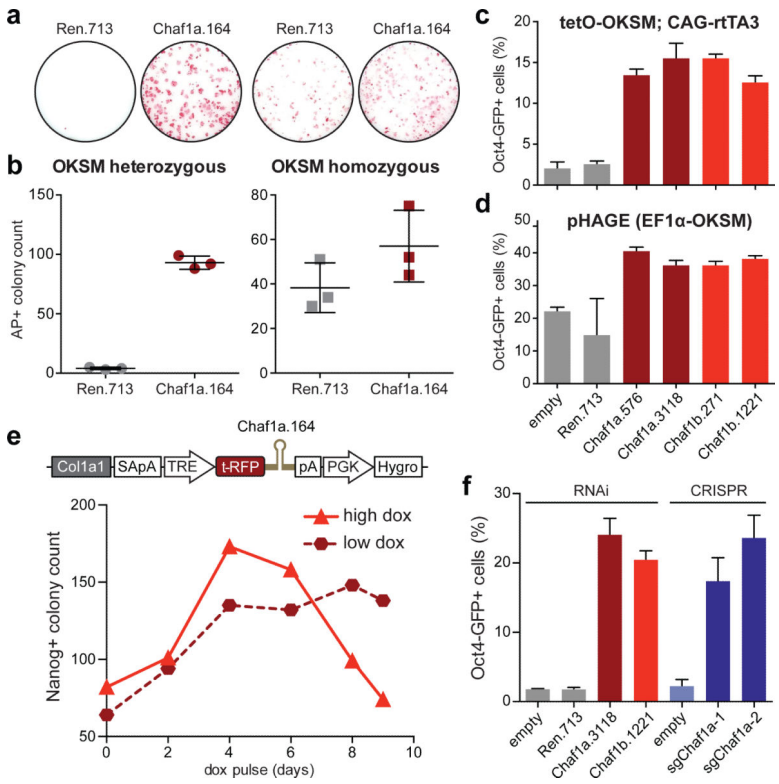


Figure 3. Reprogramming phenotype depends on optimal CAF-1 and OKSM dose

(a) Comparison of reprogramming efficiency upon Chaf1a knockdown using MEFs carrying one or two copies of *Col1a1::tetOP-OKSM* and *R26-M2rtTA*. Colonies were scored at day 10 after 6 days of dox exposure and 4 days of dox-independent growth. Error bars indicate standard deviation from biological triplicates. **(b)** Quantification of data shown in (a). **(c,d)** Effect of CAF-1 suppression on reprogramming efficiency when directly infecting MEFs with lentiviral vectors achieving medium (c) or high (d) OKSM expression levels, as determined by flow cytometry for Oct4-GFP at day 11. Error bars indicate standard deviation from biological triplicates. **(e)** Influence of duration and degree of Chaf1a suppression on reprogramming potential of MEFs carrying doxinducible shRNA cassette (top), as determined by immunocytochemistry for Nanog at day 9. Datapoints represent single experiment. **(f)** Comparison of reprogramming efficiencies when using shRNAs or sgRNAs targeting Chaf1a, as determined by flow cytometry for Oct4-GFP after 7 days of dox exposure and 4 days of dox-independent growth. Error bars indicate standard deviation from biological triplicates.

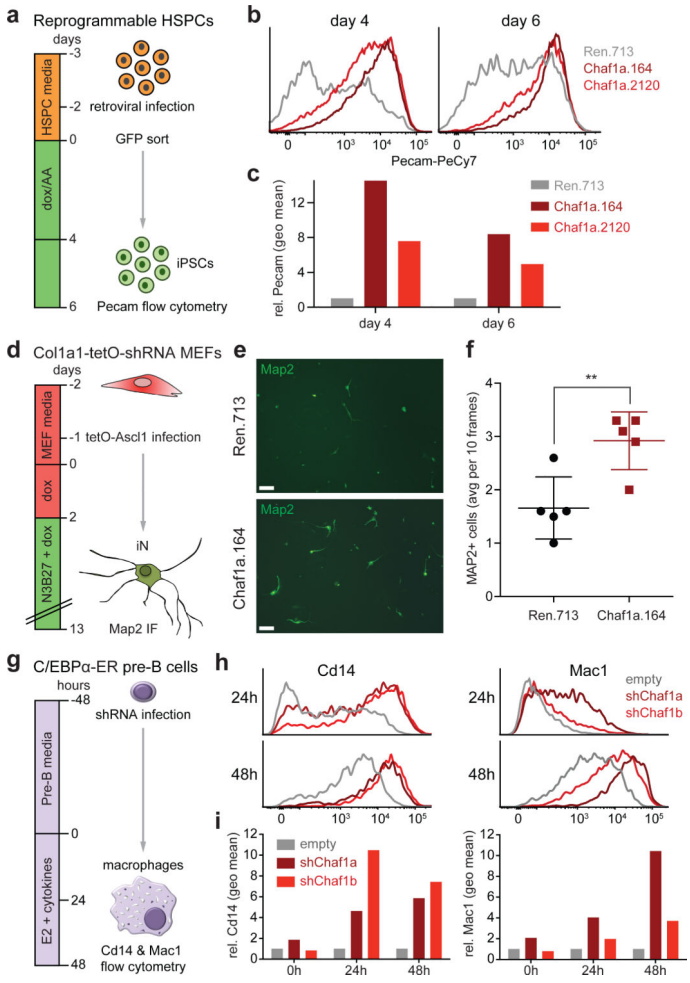


Figure 4. CAF-1 suppression enhances reprogramming in different cell conversion systems

(a) Reprogramming of fetal hematopoietic stem and progenitor cells (HSPCs) into iPSCs. **(b)** Flow cytometric analysis of Pecam expression during HSPC reprogramming using indicated shRNAs. **(c)** Quantification of data shown in (b). Values represent fold-change expression differences between experimental and control samples using geometric mean. Data were obtained from one experiment using 2 different Chaf1 shRNAs. **(d)** Transdifferentiation of MEFs into induced neurons (iNs). **(e)** Representative image of MAP2^{+} iNs after 13 days of transdifferentiation. Scale bars: 100 μm. **(f)** Quantification of transdifferentiation efficiency (n=5 independent experiments; values are mean +/− standard deviation; **, unpaired t-test; p=0.0075). **(g)** Transdifferentiation of pre-B cells into macrophages. **(h)** Activation of macrophage markers Cd14 and Mac1 in representative samples at indicated time points. **(i)** Cd14 and Mac1 expression levels in indicated samples (values represent fold-change expression differences between experimental and control samples using geometric mean; n=2 independent viral transductions).

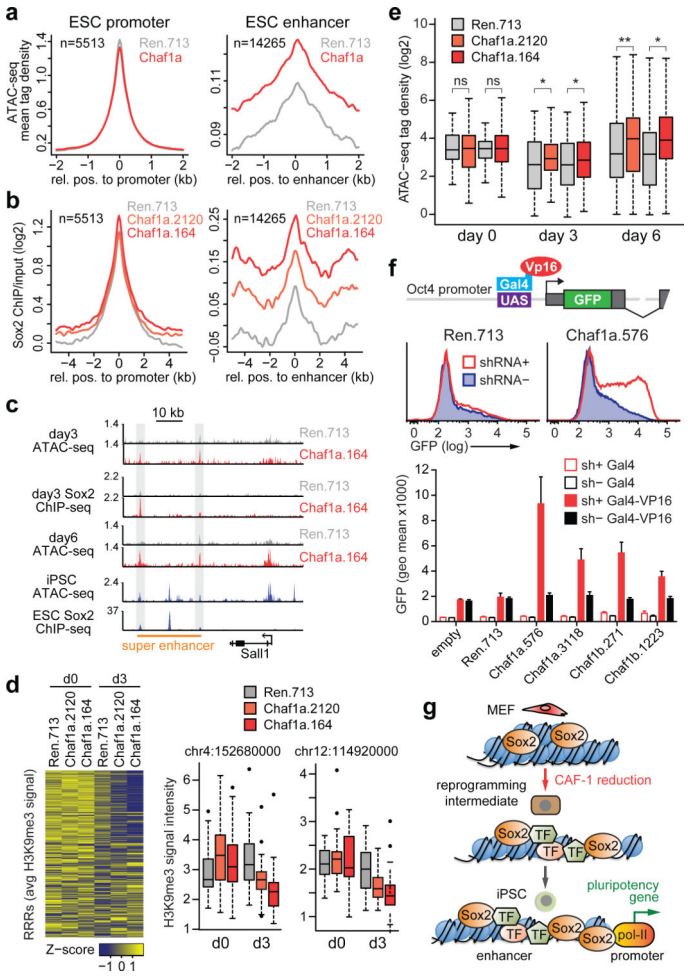


Figure 5. CAF-1 suppression facilitates chromatin accessibility, Sox2 binding and transcriptional activation of pluripotency genes

(a) ATAC-seq analysis of ESC-specific enhancers and promoters at day 3 of reprogramming. Shown are merged data for Chaf1a.164 and Chaf1a.2120 shRNA-infected cells ($p>0.5$ and $p<10^{-15}$ between Chaf1a and Renilla shRNAs for promoters and enhancers, respectively; n denotes number of examined promoter and enhancer elements). **(b)** Sox2 ChIP-seq analysis of ESC-specific enhancers and promoters at day 3 of reprogramming using weak (2120) and strong (164) Chaf1a hairpin ($p<1e-15$ for both shRNAs; see panel (a) for definition of n). **(c)** Representative ATAC-seq and Sox2 ChIP-seq peaks at the *Sall1* super-enhancer (y axis: tag density profiles). **(d)** H3K9me3 ChIP-seq analysis of reprogramming-resistant regions (RRRs)[29] after 0 and 3 days of OKSM expression. Heatmap shows all RRRs (rows); box plots show individual RRRs between day 0 and 3 in Chaf1a knockdown cells ($p<0.05$ for both shRNAs). **(e)** Chromatin accessibility at day 0, 3 and 6 for genes that become transcriptionally upregulated in Chaf1a shRNA-treated cells by day 6 (*, $p<0.05$; **, $p<0.01$). **(f)** Chromatin in-vivo assay (CiA) to directly measure effect of CAF-1 suppression on transcriptional activity of endogenous *Oct4* locus in fibroblasts upon overexpression of Gal4-VP16 fusion protein targeted to *Oct4* promoter. **(g)** Summary and model. TF, transcription factor; pol-II, RNA polymerase II.