nature cell biology

ARTICLES

https://doi.org/10.1038/s41556-018-0047-x

NCoR/SMRT co-repressors cooperate with c-MYC to create an epigenetic barrier to somatic cell reprogramming

Qiang Zhuang[1,2,3,5,16], Wenjuan Li[2,3,4,5,16], Christina Benda[2,3,5,16], Zhijian Huang[2,3,5], Tanveer Ahmed[3,5,6,7], Ping Liu[2,3,5,8], Xiangpeng Guo[1,2,3,5], David P. Ibañez[2,3,4,5], Zhiwei Luo[2,3,4,5], Meng Zhang[2,3,4,5], Mazid Md. Abdul[2,3,4,5], Zhongzhou Yang[3], Jiayin Yang[9], Yinghua Huang[1,3,5,7], Hui Zhang[3,5,7], Dehao Huang[2,3,4], Jianguo Zhou[2,3,5], Xiaofen Zhong[2,3], Xihua Zhu[2,3,5], Xiuling Fu[10], Wenxia Fan[2,3,5], Yulin Liu[11], Yan Xu[2,3,5], Carl Ward[2,3,5], Muhammad Jadoon Khan[2,3,5], Shahzina Kanwal[2,3,5], Bushra Mirza[6], Micky D. Tortorella[12], Hung-Fat Tse[9,13,14], Jiayu Chen[15], Baoming Qin[1,3,5,7,13], Xichen Bao[1,2,3,5], Shaorong Gao[15], Andrew P. Hutchins[10]* and Miguel A. Esteban[1,2,3,5,13]*

Somatic cell reprogramming by exogenous factors requires cooperation with transcriptional co-activators and co-repressors to effectively remodel the epigenetic environment. How this interplay is regulated remains poorly understood. Here, we demonstrate that NCoR/SMRT co-repressors bind to pluripotency loci to create a barrier to reprogramming with the four Yamanaka factors (OCT4, SOX2, KLF4 and c-MYC), and consequently, suppressing NCoR/SMRT significantly enhances reprogramming efficiency and kinetics. The core epigenetic subunit of the NCoR/SMRT complex, histone deacetylase 3 (HDAC3), contributes to the effects of NCoR/SMRT by inducing histone deacetylation at pluripotency loci. Among the Yamanaka factors, recruitment of NCoR/SMRT–HDAC3 to genomic loci is mostly facilitated by c-MYC. Hence, we describe how c-MYC is beneficial for the early phase of reprogramming but deleterious later. Overall, we uncover a role for NCoR/SMRT co-repressors in reprogramming and propose a dual function for c-MYC in this process.

Somatic cells can be reprogrammed to induced pluripotent stem cells (iPSCs) by enforced expression of defined exogenous factors, originally OCT4, SOX2, KLF4 and c-MYC (OSKM)[1]. At the onset of reprogramming, the exogenous OSKM bind to DNA across the genome and induce successive rounds of chromatin reorganization to allow the activation of the entire pluripotency gene network[2–4]. However, OSKM do not operate in isolation and need co-regulators to modify the local epigenetic environment[5–8]. Despite the growing evidence regarding transcriptional and epigenetic responses in reprogramming, it remains unclear how OSKM and different transcriptional co-regulators (co-activators and co-repressors) work with or antagonize each other to induce a pluripotent state[9].

In this report, we explored the function of two well-known co-repressors, nuclear receptor co-repressor (NCoR) and silencing mediator of retinoid and thyroid hormone receptor (SMRT)[10,11], both of which have fundamental roles in preserving cellular identity and tissue homeostasis, in reprogramming.

Results

NCoR/SMRT co-repressors create a barrier to OSKM reprogramming. First, we assessed *Ncor1* (encoding NCoR) and *Ncor2* (encoding SMRT) expression in mouse embryonic fibroblasts (MEFs), embryonic stem cells (ESCs) and OSKM reprogramming by quantitative PCR with reverse transcription (RT–qPCR). Both co-repressors were expressed in all three cell types, with an increase in the levels of *Ncor1* during reprogramming and in ESCs compared to MEFs (Supplementary Fig. 1a). We then knocked down *Ncor1/2* in OG2 MEFs[12–15] transduced with OSKM retroviruses (Supplementary Fig. 1b–d). Knocking down either co-repressor significantly enhanced the number of *Oct4*–green fluorescent protein-positive (*Oct4*-GFP$^+$) colonies in both serum-based and

[1]Key Laboratory of Regenerative Biology of the Chinese Academy of Sciences, Joint School of Life Sciences, Guangzhou Medical University and Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China. [2]Laboratory of RNA, Chromatin, and Human Disease, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China. [3]Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China. [4]University of Chinese Academy of Sciences, Beijing, China. [5]Guangzhou Regenerative Medicine and Health Guangdong Laboratory (GRMH-GDL), Guangzhou, China. [6]Department of Biochemistry, Quaid-i-Azam University, Islamabad, Pakistan. [7]Laboratory of Metabolism and Cell Fate, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China. [8]Institute of Health Sciences, Anhui University, Hefei, China. [9]Department of Medicine, The University of Hong Kong, Hong Kong, China. [10]Department of Biology, Southern University of Science and Technology of China, Shenzhen, China. [11]Guangzhou FitGene Biotechnology Co. Ltd., Guangzhou, China. [12]Drug Discovery Pipeline, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China. [13]Hong Kong-Guangdong Joint Laboratory of Stem Cells and Regenerative Medicine, The University of Hong Kong and Guangzhou Institutes of Biomedicine and Health, Hong Kong, China. [14]Shenzhen Institutes of Research and Innovation, The University of Hong Kong, Hong Kong, China. [15]School of Life Sciences and Technology, Tongji University, Shanghai, China. [16]These authors contributed equally: Qiang Zhuang, Wenjuan Li and Christina Benda. *e-mail: andrewh@sustc.edu.cn; miguel@gibh.ac.cn

serum + vitamin C (Vc) culture conditions[16] (Fig. 1a,b) and the colonies appeared quicker (Fig. 1c). This effect was not limited to the *Oct4*-GFP reporter, as immunofluorescent detection showed a comparable increase in the number of NANOG-positive (NANOG$^+$) colonies in OG2 MEFs and ICR background MEFs (Supplementary Fig. 1e–h). Besides, *Ncor1/2* knockdown could similarly enhance reprogramming in secondary OG2 MEFs[17] and secondary MEFs with a GFP knock-in into the endogenous *Oct4* locus[18], either with OSKM retroviral transduction (without doxycycline) or with doxycycline (without OSKM) (Supplementary Fig. 1i,j). Thus, suppressing *Ncor1/2* enhances OSKM reprogramming irrespective of the transduction method, reporter system or the type of MEFs.

The increase in reprogramming efficiency with *Ncor1/2* knockdown was unrelated to higher proliferation[19], as the knockdowns actually reduced cell numbers (Supplementary Fig. 1k). Moreover, *Ncor1/2* knockdown enhanced OSKM reprogramming with knockout serum replacement (KSR) medium[20], producing >50% of *Oct4*-GFP$^+$ cells by day 14, and *Ncor1* + *Ncor2* produced ~80% (Fig. 1d and Supplementary Fig. 1l). We also used iCD1 medium, which is typically used in OSK reprogramming[21]. *Ncor1/2* knockdown in iCD1 had a modest improvement when using OSK, but there was a significant enhancement using OSKM (Supplementary Fig. 1m). In addition, we characterized several colonies produced with *Ncor1/2* knockdown, confirming that they are bona fide pluripotent cells (Supplementary Fig. 2a–f). Hence, suppression of NCoR/SMRT enhances OSKM reprogramming using different media, including high-efficiency media, and the resulting iPSC lines are fully reprogrammed.

To understand how NCoR and SMRT block reprogramming, we did RNA sequencing (RNA-seq) at day 5 and day 9 of OSKM reprogramming in *Ncor1/2*-depleted cells in serum + Vc (Fig. 1e, Supplementary Fig. 3a and Supplementary Table 1). Differentially expressed genes in either knockdown were similar (Fig. 1e and Supplementary Fig. 3a), but the overall effect of *Ncor2* knockdown was more potent (Fig. 1e). Unexpectedly, many MEF-enriched somatic genes were upregulated by *Ncor1/2* knockdown at day 5, which became less noticeable at day 9 (Fig. 1f and Supplementary Fig. 3b–d). Genes related to the mesenchymal-to-epithelial transition[22,23] were unaffected in *Ncor1/2*-depleted cells at both time points (Supplementary Fig. 3e). Importantly, particularly at day 9, there was upregulation of pluripotency-related genes, including multiple components of the previously identified 'second-wave' genes[24], with *Ncor1/2* knockdown (Fig. 1g and Supplementary Fig. 3f). These results were validated by RT–qPCR (Supplementary Fig. 3g–i). Therefore, suppressing *Ncor1/2* in OSKM reprogramming first leads to a transient defect in the repression of somatic genes, and then an accelerated and more potent activation of pluripotency genes.

NCoR/SMRT co-repressors require HDAC3 to impair reprogramming. NCoR and SMRT modulate chromatin by acting as a docking platform to bridge transcription factors[11] and epigenetic modifiers, including different histone deacetylase (HDAC) family members[10,11,25]. HDACs seemed to be good candidates to mediate the detrimental effects of NCoR/SMRT on OSKM reprogramming because pan-HDAC inhibitors such as valproic acid (VPA) or trichostatin (TSA) enhance reprogramming[26]. In many contexts, HDAC3 is responsible for the deacetylase activity of the NCoR/SMRT co-repressor complex, and, similarly, HDAC3 requires NCoR/SMRT[27,28]. Yet, HDAC1, 2, 4, 5 and 7 can also interact with NCoR/SMRT[11]. Thus, we first looked at the expression of *Hdac1* to *Hdac11* in MEFs, ESCs and during OSKM-mediated reprogramming. *Hdac8*, *9*, *10* and *11* were not expressed in any of the three cell types, whereas *Hdac7* was only expressed in MEFs and during reprogramming and was downregulated in ESCs, and *Hdac1* to *Hdac6* were expressed in all three cell types (Fig. 2a). We then knocked down the expressed HDACs in OSKM reprogramming in serum (Fig. 2b and Supplementary Fig. 4a). *Hdac3* knockdown significantly increased reprogramming, whereas among the other HDACs, only *Hdac2* knockdown could enhance reprogramming (Fig. 2b). Conversely, *Hdac4*, *5* and *7* knockdowns impaired reprogramming (Fig. 2b), in agreement with our previous report demonstrating that they are important for the mesenchymal-to-epithelial transition phase[29]. Two additional short hairpin RNAs (shRNAs) against *Hdac3* (Supplementary Fig. 4a) also increased reprogramming efficiency in serum and serum + Vc (Fig. 2c). Importantly, VPA did not synergize with *Hdac3* knockdown to promote reprogramming (Fig. 2c), suggesting that pan-HDAC inhibitors enhance OSKM reprogramming through HDAC3 inhibition.

We prepared a deacetylase-null mutant form of HDAC3 (HDAC3-YF; Y298F) that acts as a dominant negative, and a double-mutant HDAC3 (HDAC3-YF-KA; K25A/Y298F) (Fig. 2d and Supplementary Fig. 4b) that also has reduced ability to interact with NCoR/SMRT[28] (Supplementary Fig. 4c). As additional controls, we produced dominant-negative versions of HDAC1 (HDAC1-YF; Y303F) and HDAC2 (HDAC2-YF; Y304F)[30]. Overexpression of HDAC1-YF and HDAC2-YF moderately enhanced OSKM reprogramming in serum (Fig. 2e), whereas HDAC3-YF potently enhanced it in serum and serum + Vc (Fig. 2f), and, as with HDAC3 depletion, there was no synergy with VPA (Fig. 2f). Importantly, the enhancing effect of HDAC3-YF on reprogramming disappeared in the HDAC3-YF-KA mutant, demonstrating that HDAC3-YF needs to interact with NCoR/SMRT to act as a dominant negative for HDAC3 (Fig. 2d,e). HDAC3-YF also induced faster *Oct4*-GFP$^+$ colony formation (Supplementary Fig. 4d,e), and enabled very high reprogramming efficiency (~76% at day 14) using KSR (Supplementary Fig. 4f). In addition, as with NCoR/SMRT depletion, HDAC3-YF reduced cell proliferation of early reprogramming intermediates (Supplementary Fig. 4g), whereas increased pluripotency gene expression in the late phase of reprogramming (Supplementary Fig. 4h). However, reprogramming with HDAC3-YF only increased somatic genes moderately at day 5 (Supplementary Fig. 4i).

We then performed single-cell RT–qPCR of a time course of OSKM reprogramming in serum + Vc with primers against a panel of both pluripotency and somatic genes (Supplementary Fig. 4j and Supplementary Tables 2,3). Multiple pluripotency genes were upregulated in the HDAC3-YF overexpressing cells at day 9 and more obviously at day 13, and, accordingly, principal component analysis (PCA) showed a substantial acceleration in the commitment to pluripotency (Fig. 2g,h and Supplementary Fig. 4k,l). As with bulk RT–qPCR, there was only a mild increase of somatic genes at day 5 of reprogramming with HDAC3-YF (Fig. 2i and Supplementary Fig. 4m).

We concluded that HDAC3 mediates the negative effects of NCoR/SMRT on OSKM reprogramming, and this function requires deacetylase activity, but the effects of suppressing NCoR/SMRT or HDAC3 on somatic genes are not identical. The latter can perhaps be explained by the fact that depletion of NCoR/SMRT at target sites removes other enzymatic activities[25] and may also allow the recruitment of co-activators[10].

HDAC3 induces histone deacetylation at restricted genomic loci in reprogramming. To gain insight into how HDAC3 impairs OSKM reprogramming, we analysed the total levels of histone H3 acetylation (AcH3) and histone H4 acetylation (AcH4) by immunoblotting in MEFs, ESCs and OSKM reprogramming intermediates in serum + Vc. As previously reported[31], the levels of AcH3 and AcH4 were higher in ESCs than in MEFs (Fig. 3a). We also observed an increase in the levels of AcH3 and AcH4 in OSKM reprogramming that was potentiated with VPA or TSA (Fig. 3a,b). Yet, depletion of NCoR/SMRT or overexpressing HDAC3-YF did not further increase the levels of AcH3 or AcH4, and the same occurred with

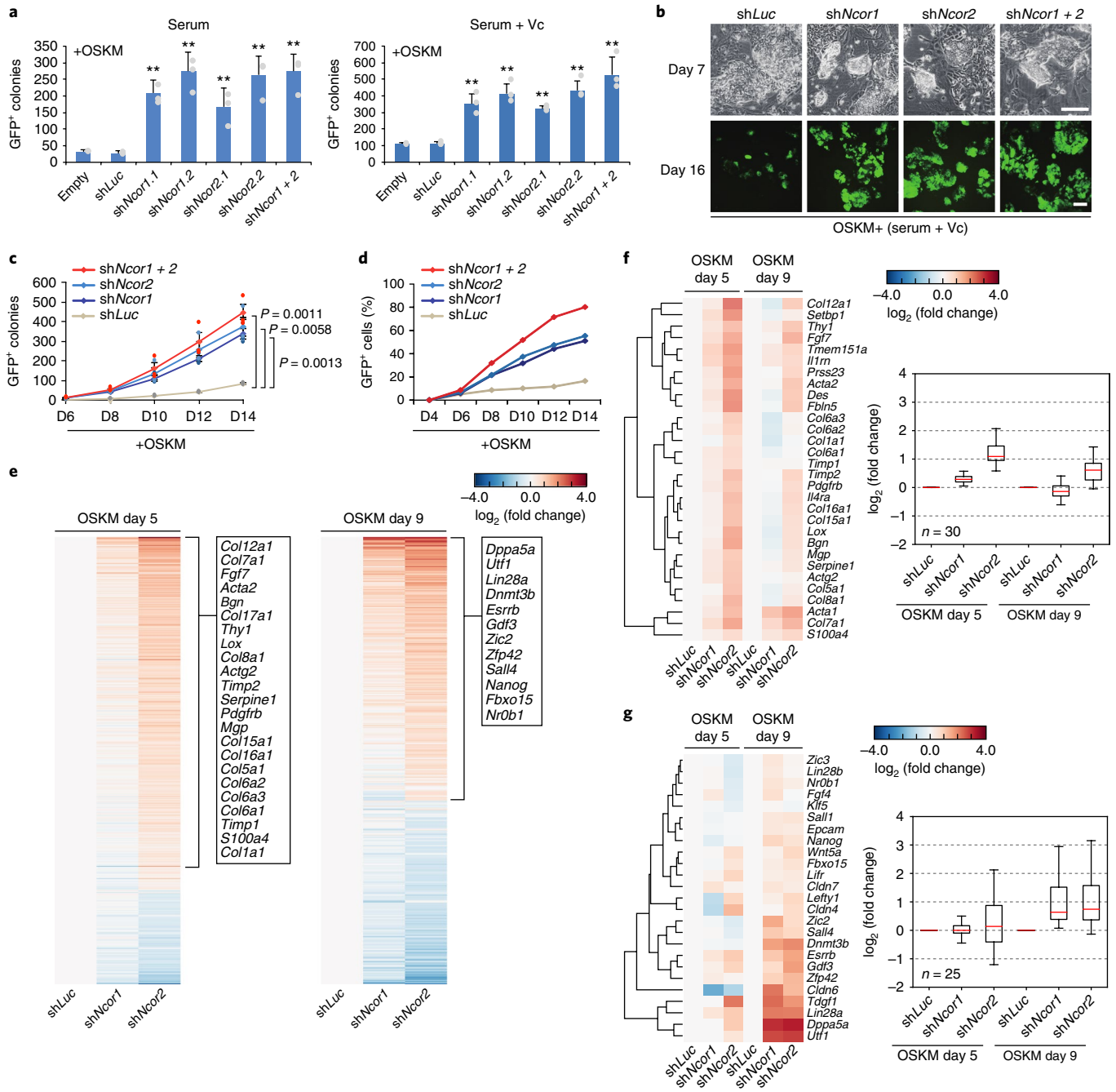

Fig. 1 | Depletion of NCoR/SMRT co-repressors enhances OSKM reprogramming. a, The number of *Oct4*-GFP$^+$ colonies on day 18 (serum) or day 16 (serum + Vc) in MEFs transduced with OSKM and shRNAs against *Luciferase* (sh*Luc*), *Ncor1*, *Ncor2*, or combined *Ncor1* and *Ncor2* (sh*Ncor1* + *2*) in serum and serum + Vc. Independent experimental data are plotted as individual points (also for all subsequent figures). Two shRNAs were used per gene; for all subsequent experiments, sh*Ncor1.1* and sh*Ncor2.2* shRNAs were used unless otherwise indicated. Data are the mean ± s.e.m. ($n = 3$ biological replicates with 3 technical replicates each). The *P* value was calculated using two-tailed unpaired Student's *t*-test, $^{**}P < 0.01$. 'Empty' is the empty vector. For this and all subsequent data, the source data are provided in Supplementary Table 6 unless otherwise mentioned. **b**, Phase-contrast (top) and *Oct4*-GFP (bottom) images of MEFs transduced with OSKM and shRNAs against *Luciferase*, *Ncor1*, *Ncor2*, or combined *Ncor1* and *Ncor2* on day 7 and day 16 in serum + Vc. Scale bar, 100 µm. **c**, Time course of the appearance of *Oct4*-GFP$^+$ colonies in MEFs transduced with OSKM and shRNAs against *Luciferase*, *Ncor1*, *Ncor2*, or combined *Ncor1* and *Ncor2* in serum + Vc. Data are the mean ± s.e.m. ($n = 3$ biological replicates with 3 technical replicates each). The *P* value was calculated using two-tailed unpaired Student's *t*-test. D, day. **d**, As in panel **c**, but shows the percentage of GFP$^+$ cells in KSR as measured by flow cytometry. Data are represented as the mean of three technical replicates from one experiment. **e**, RNA-seq analysis of MEFs transduced with OSKM and shRNAs against *Luciferase*, *Ncor1*, or *Ncor2* in serum + Vc. Heatmaps showing significantly upregulated and downregulated genes on day 5 and day 9 relative to sh*Luc*. **f**, Heatmap and boxplots showing the fold change of selected day 5 upregulated somatic genes on day 5 and day 9. The number of genes is indicated by '*n*'. For boxplots, the red central line is the median, the boxes indicate the upper and lower quartiles, the whiskers indicate the 1.5 interquartile range. **g**, As in panel **f**, but shows the fold change of selected day 9 upregulated pluripotency genes.

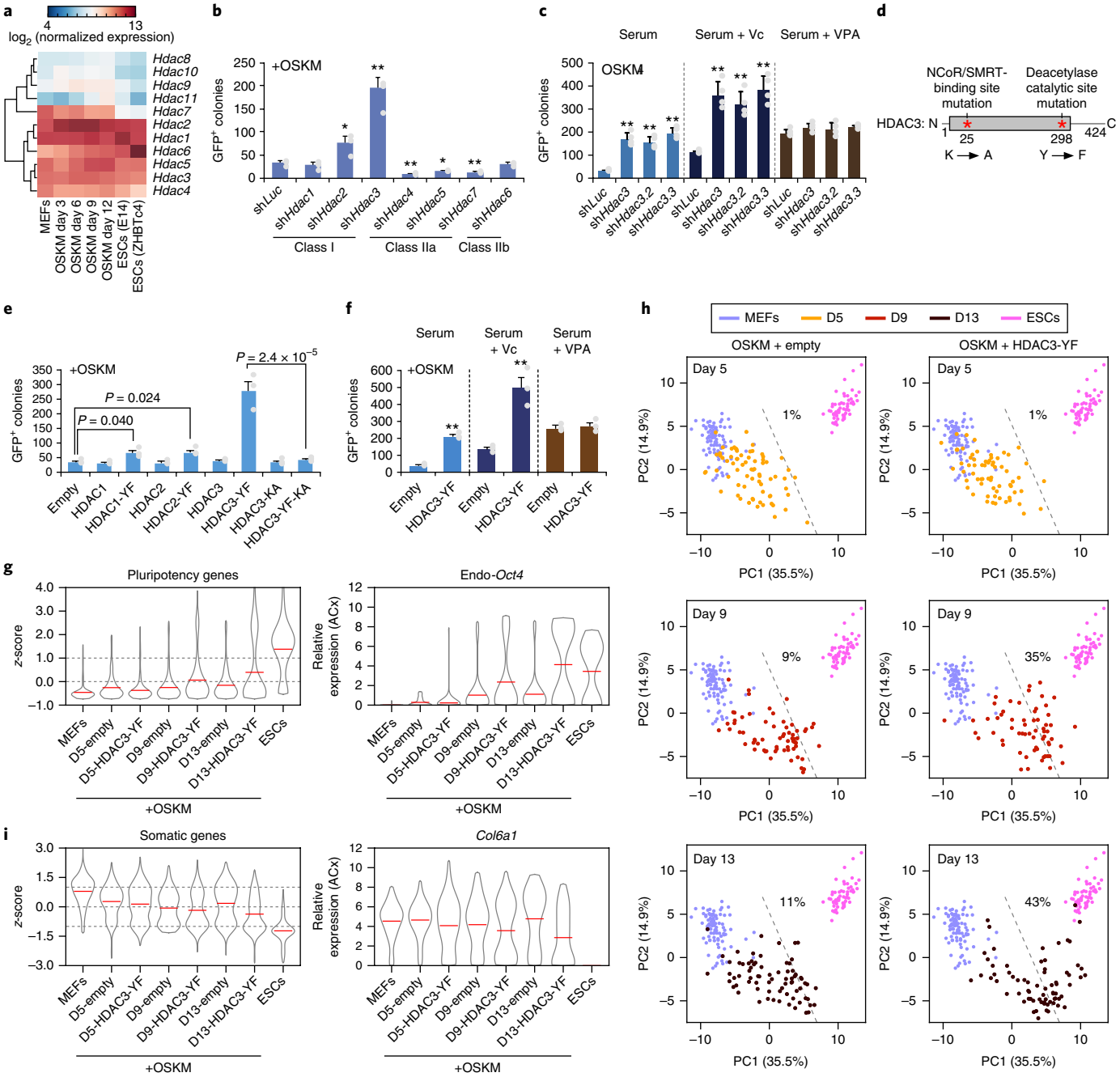

Fig. 2 | HDAC3 executes the negative effects of NCoR/SMRT co-repressors on OSKM reprogramming. a, RNA-seq analysis of *Hdac* family members during reprogramming and in MEFs and ESCs. Data were taken from GSE57967, GSE29278, GSE44288 and GSE20851. **b**, The number of *Oct4*-GFP$^+$ colonies on day 18 in MEFs transduced with OSKM and shRNAs against *Hdac1* to *Hdac7* in serum. Data are the mean ± s.e.m. ($n = 3$ biological replicates with 3 technical replicates each). The *P* value was calculated using two-tailed unpaired Student's *t*-test, $^*P < 0.05$ and $^{**}P < 0.01$. **c**, The number of *Oct4*-GFP$^+$ colonies in MEFs transduced with OSKM and shRNAs against *Luciferase* or *Hdac3* in serum (day 18), serum + Vc (day 16) and serum + VPA (day 16). Data are the mean ± s.e.m. ($n = 4$ biological replicates with 3 technical replicates each). The *P* value was calculated using two-tailed unpaired Student's *t*-test, $^{**}P < 0.01$. **d**, Schematic of HDAC3 and the mutants generated. **e**, The number of *Oct4*-GFP$^+$ colonies on day 18 in MEFs transduced with OSKM and the indicated HDACs (or their mutants) and empty vector in serum. Data are the mean ± s.e.m. ($n = 3$ biological replicates with 3 technical replicates each). The *P* value was calculated using two-tailed unpaired Student's *t*-test. **f**, The number of *Oct4*-GFP$^+$ colonies in MEFs transduced with OSKM and HDAC3-YF or empty vector in serum (day 18), serum + Vc (day 16) and serum + VPA (day 16). Data are the mean ± s.e.m. ($n = 3$ biological repeats with 3 technical replicates each). The *P* value was calculated using two-tailed unpaired Student's *t*-test, $^{**}P < 0.01$. **g**, Single-cell RT-qPCR for MEFs, MEFs transduced with OSKM and HDAC3-YF or empty vector in serum + Vc, and OG2 ESCs. Violin plots of single-cell RT-qPCR expression of the *z*-score for a set of merged pluripotent genes (*Utf1*, Endo-*Oct4*, *Dnmt3b* and *Sall4*) and Endo-*Oct4* shown individually. Red line indicates the mean. Endo-*Oct4*, endogenous *Oct4*. **h**, As in panel **g**, but showing the PCA. Each point indicates a single cell. **i**, As in panel **g**, but for violin plots of merged somatic genes (*Col6a1*, *Acta2* and *S100a4*) and *Col6a1* shown individually.

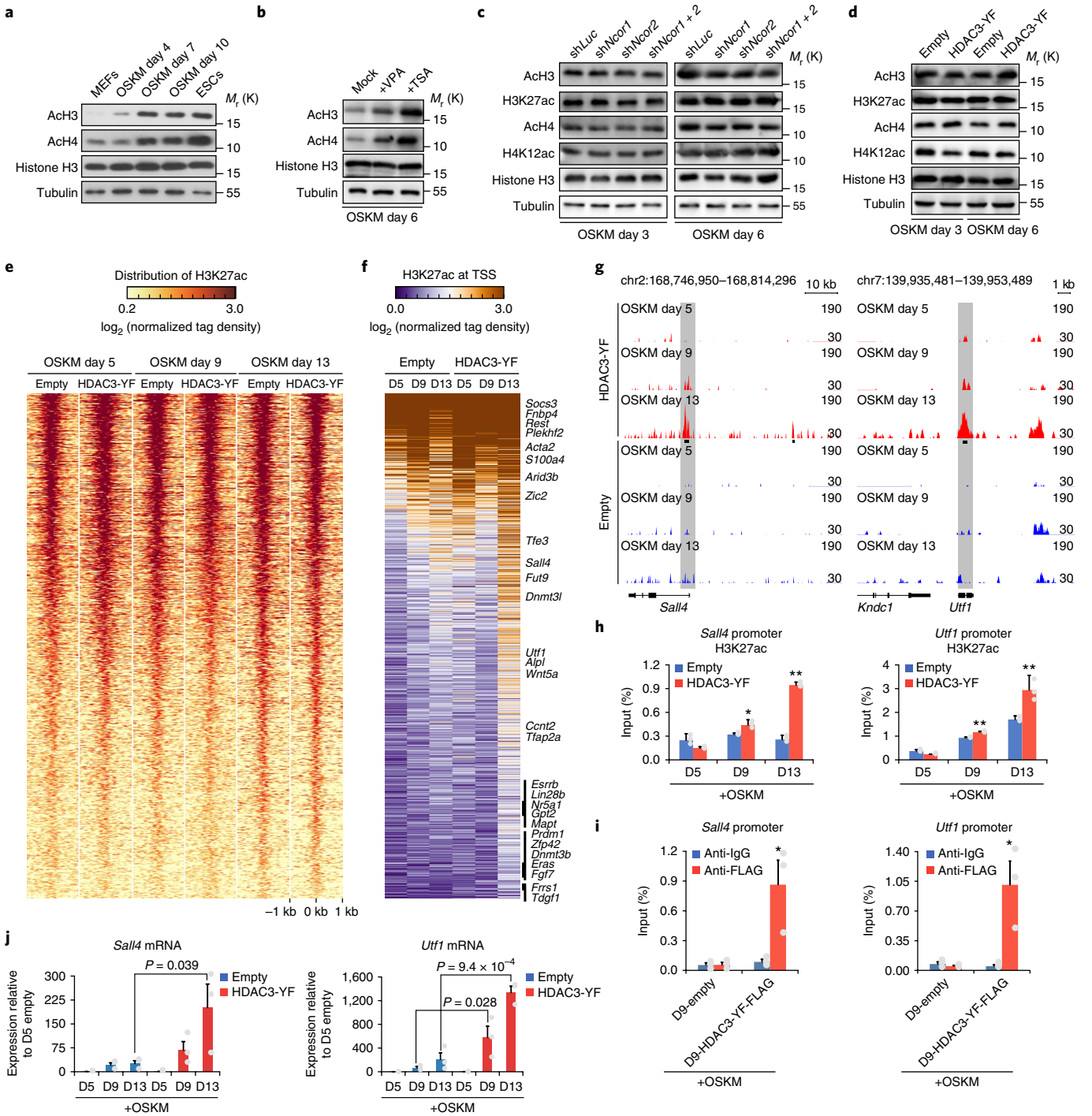

Fig. 3 | NCoR/SMRT–HDAC3 complex induces histone deacetylation at pluripotency loci in OSKM reprogramming. a, Immunoblotting for the indicated proteins using lysates from ICR MEFs, ICR MEFs transduced with OSKM in serum + Vc at different time points and in OG2 ESCs. Histone H3 and tubulin were the loading controls. All unprocessed scans of blots in this and the subsequent figures are shown in Supplementary Fig. 9. **b**, Immunoblotting for the indicated proteins using lysates from MEFs transduced with OSKM in serum mock treated or treated with VPA or TSA from days 3 to 5 on day 6. **c**,**d**, Immunoblotting for the indicated proteins using lysates from MEFs transduced with OSKM control (empty vector or sh*Luc*), HDAC3-YF, or shRNAs against *Ncor1*, *Ncor2* or combined *Ncor1* and *Ncor2* in serum + Vc on day 3 and day 6. **e**, Tag density pileups of H3K27ac peaks at the indicated time points of MEFs transduced with OSKM and HDAC3-YF or empty vector in serum + Vc. **f**, Global view of H3K27ac levels at genes with >1.5-fold upregulated normalized tag density near the TSS at any reprogramming time point. **g**, Genome views of H3K27ac tag density at *Sall4* and *Utf1*; black bars under the tracks indicate H3K27ac peaks (also for all subsequent figures). chr, chromosome. **h**, ChIP–qPCR analysis of H3K27ac at *Sall4* and *Utf1* on days 5, 9 and 13 in MEFs transduced with OSKM and HDAC3-YF or empty vector in serum + Vc. Data are the mean ± s.e.m. ($n = 3$ biological replicates with 3 technical replicates each). The *P* value was calculated using two-tailed unpaired Student's *t*-test, $^*P < 0.05$ and $^{**}P < 0.01$. **i**, ChIP–qPCR for HDAC3-YF binding at *Sall4* and *Utf1* in MEFs transduced with OSKM and HDAC3-YF or empty vector on day 9. Data are the mean ± s.e.m. ($n = 3$ biological replicates with 3 technical replicates). The *P* value was calculated using two-tailed unpaired Student's *t*-test, $^*P < 0.05$. **j**, RT–qPCR of *Sall4* and *Utf1* in MEFs transduced with OSKM and HDAC3-YF or empty vector. Data are the mean ± s.e.m. ($n = 3$ biological replicates with 3 technical replicates each). *P* values were calculated using one-tailed unpaired Student's *t*-test.

two representative acetylated histone marks that are known to be regulated by NCoR/SMRT–HDAC3: H3K27 acetylation (H3K27ac) and H4K12 acetylation (H4K12ac)[32,33] (Fig. 3c,d). The lack of a global effect on histone acetylation was validated using chromatin immunoprecipitation followed by sequencing (ChIP–seq) for H3K27ac in a time course (days 5, 9 and 13) of OSKM reprogramming in serum + Vc with HDAC3-YF or empty vector (Fig. 3e). This assay also showed that, in OSKM reprogramming, most H3K27ac is located near the transcription start site (TSS) (Supplementary Fig. 5a). These results are consistent with multiple studies showing that the NCoR/SMRT–HDAC3 complex targets a small fraction of the total genome in a particular cellular context[28].

We then measured the levels of H3K27ac at all TSS and collected those showing >1.5-fold upregulation with HDAC3-YF (Fig. 3f). Interestingly, the list of genes was biased for pluripotency-related genes, including 'second-wave' genes. The genome views for H3K27ac tag density for four pluripotency loci (*Sall4*, *Utf1*, *Nanog* and *Zic2*) showed an increase in H3K27ac at day 9 and, particularly, at day 13 of OSKM reprogramming with HDAC3-YF (Fig. 3g and Supplementary Fig. 5b). The H3K27ac ChIP–seq results were confirmed by ChIP–qPCR of the same loci (Fig. 3h and Supplementary Fig. 5c). ChIP–qPCR for HDAC3-YF also showed significant enrichment at these loci at day 9 of OSKM reprogramming (Fig. 3i and Supplementary Fig. 5d), indicating that the effect of HDAC3-YF overexpression on H3K27ac levels in OSKM reprogramming is direct. Likewise, there was good temporal correlation between the changes in H3K27ac and increased gene expression measured by RT–qPCR (Fig. 3j and Supplementary Fig. 5e). One caveat of these experiments is that the data were generated from bulk populations, and it is likely that the changes at day 9 (or earlier) are greater in a subset of cells, as noticed above with the single-cell RT–qPCR (Fig. 2h). Our ChIP–seq analysis also detected moderately higher levels of H3K27ac at somatic loci in cells reprogrammed with OSKM and HDAC3-YF, but only in a few cases and exclusively at day 5 (Fig. 3f and Supplementary Fig. 5f). In summary, HDAC3, probably acting as part of the NCoR/SMRT complex, impairs OSKM reprogramming by inducing histone deacetylation at restricted loci including pluripotency loci.

Context-specific recruitment of NCoR/SMRT to pluripotency loci in reprogramming. We performed ChIP–seq of day 9 OSKM reprogramming in serum + Vc samples to study whether NCoR/SMRT are recruited to pluripotency loci to induce histone deacetylation through HDAC3. We included ESCs for comparison because they also express *Ncor1/2* (Supplementary Fig. 1a). As expected, a big proportion of NCoR/SMRT-binding peaks were shared in reprogramming (76%) and in ESCs (56%) (Fig. 4a,b and Supplementary Table 4). These peaks were predominantly located near the TSS (Supplementary Fig. 6a), consistent with our H3K27ac ChIP–seq data.

Further analysis showed substantial differences between reprogramming cells and ESCs, so we defined context-specific binding groups for NCoR/SMRT (Fig. 4a): group 5 (reprogramming specific), group 7 (ESC specific) and group 15 (common between reprogramming and ESCs). Gene Ontology (GO) analysis showed that NCoR/SMRT-binding sites are enriched for stem cell-related terms in group 5 and group 15, particularly group 5, whereas in group 7, NCoR/SMRT-binding sites are enriched for differentiation-related terms (Fig. 4c and Supplementary Fig. 6b). For example, NCoR/SMRT are recruited to the pluripotency loci *Sox2* and *Prdm4* [34] only during reprogramming and, consistent with the idea that NCoR/SMRT impair reprogramming through HDAC3, these loci display lower levels of H3K27ac in reprogramming than in ESCs (Fig. 4d). Conversely, NCoR/SMRT binds to *Klf2* both during reprogramming and in ESCs, but this locus has low H3K27ac levels in reprogramming and high levels in ESCs (Supplementary Fig. 6c). The latter correlates with the high expression of *Klf2* in ESCs, which implies that the NCoR/SMRT complex is inactive when recruited to pluripotency loci in ESCs or, alternatively, that additional mechanisms counteract its negative effects. This finding fits well with the observation that other co-repressor complexes are recruited to actively transcribed loci in ESCs[35,36]. We confirmed NCoR/SMRT binding to the *Nanog* and *Utf1* promoters in reprogramming and to a lesser extent in ESCs by ChIP–qPCR (Fig. 4e), whereas there was little or no NCoR/SMRT bound in MEFs. As for group 7 (ESC specific), we observed NCoR/SMRT binding to differentiation-related loci including *Sox7* (endoderm), *Tcf15* (mesoderm) and *Pou4f2* (ectoderm) (Supplementary Fig. 6d), suggesting that NCoR/SMRT prevents the expression of lineage specifiers in ESCs. Yet, *Ncor1/2* knockdown in ESCs did not induce differentiation (Supplementary Fig. 6e–h). These data support that NCoR/SMRT recruitment to pluripotency loci derails reprogramming, whereas in ESCs, these co-repressors have a different function potentially related to tissue specification during development[11].

The above specific examples supported an anti-correlation between NCoR/SMRT recruitment in reprogramming and H3K27ac levels, and an increase in H3K27ac levels when reprogramming cells become pluripotent. We then studied whether this observation applies to all NCoR/SMRT peaks in reprogramming, and whether HDAC3-YF overexpression helps to increase H3K27ac levels at NCoR/SMRT target loci to mimic the pluripotent state. As there was no difference in the mean normalized levels of all H3K27ac peaks (Supplementary Fig. 7a), we measured the levels of H3K27ac at days 5, 9 and 13 of reprogramming (with HDAC3-YF or empty vector), and in ESCs, for all NCoR/SMRT-binding sites at day 9 of reprogramming. We observed that for all loci bound by NCoR/SMRT, the early reprogramming time points have low levels of H3K27ac that are elevated at day 13 (less so at day 9), particularly with HDAC3-YF, and in ESCs (Fig. 5a). We detected similar results when measuring H3K27ac levels in reprogramming for the NCoR/SMRT-binding groups 5 and 15 (Fig. 5b,c and Supplementary Fig. 7b) and for the 'second-wave' genes bound by NCoR/SMRT in reprogramming (Fig. 5d). Moreover, to study whether the enhancing effect of HDAC3-YF on H3K27ac levels at NCoR/SMRT-binding sites is associated with increased gene expression, we performed RT–qPCR for a panel of pluripotency-related genes (including many 'second-wave' genes) bound by NCoR/SMRT within 2 kb of their TSS. There was good correlation between NCoR/SMRT binding, elevated H3K27ac levels at day 13 (less so at day 9) of reprogramming and increased gene expression when HDAC3-YF is overexpressed (Fig. 5e), and similar results were obtained with *Ncor1/2* knockdown. By contrast, a panel of 'second-wave' genes with higher expression in ESCs than in MEFs and without NCoR/SMRT binding within 20 kb of the TSS at day 9 of reprogramming, showed little or no increase in H3K27ac levels at day 9 or day 13 with HDAC3-YF. This effect was visible at, for example, the *Gdf3* locus (Supplementary Fig. 7c,d). Of note, we also found that for NCoR/SMRT-binding group 7 (ESC specific), H3K27ac levels were low in reprogramming, but there was a rise in levels in ESCs (Supplementary Fig. 7b,e). However, H3K27ac levels at these loci in ESCs are significantly lower than for group 5 and group 15 (Fig. 5b,c). In this regard, it is known that ESCs display modest levels of H3K27ac at some developmental genes, which poises them for rapid activation during differentiation[37]. Indeed, H3K27ac levels in group 7 genes increased at day 13 of reprogramming with empty vector alone (Supplementary Fig. 7e), suggesting that during reprogramming these loci are not directly regulated by NCoR/SMRT–HDAC3. Overall, these results confirm that NCoR/SMRT suppress pluripotency gene reactivation through histone deacetylation mediated by HDAC3, although we cannot exclude the participation of other mechanisms.

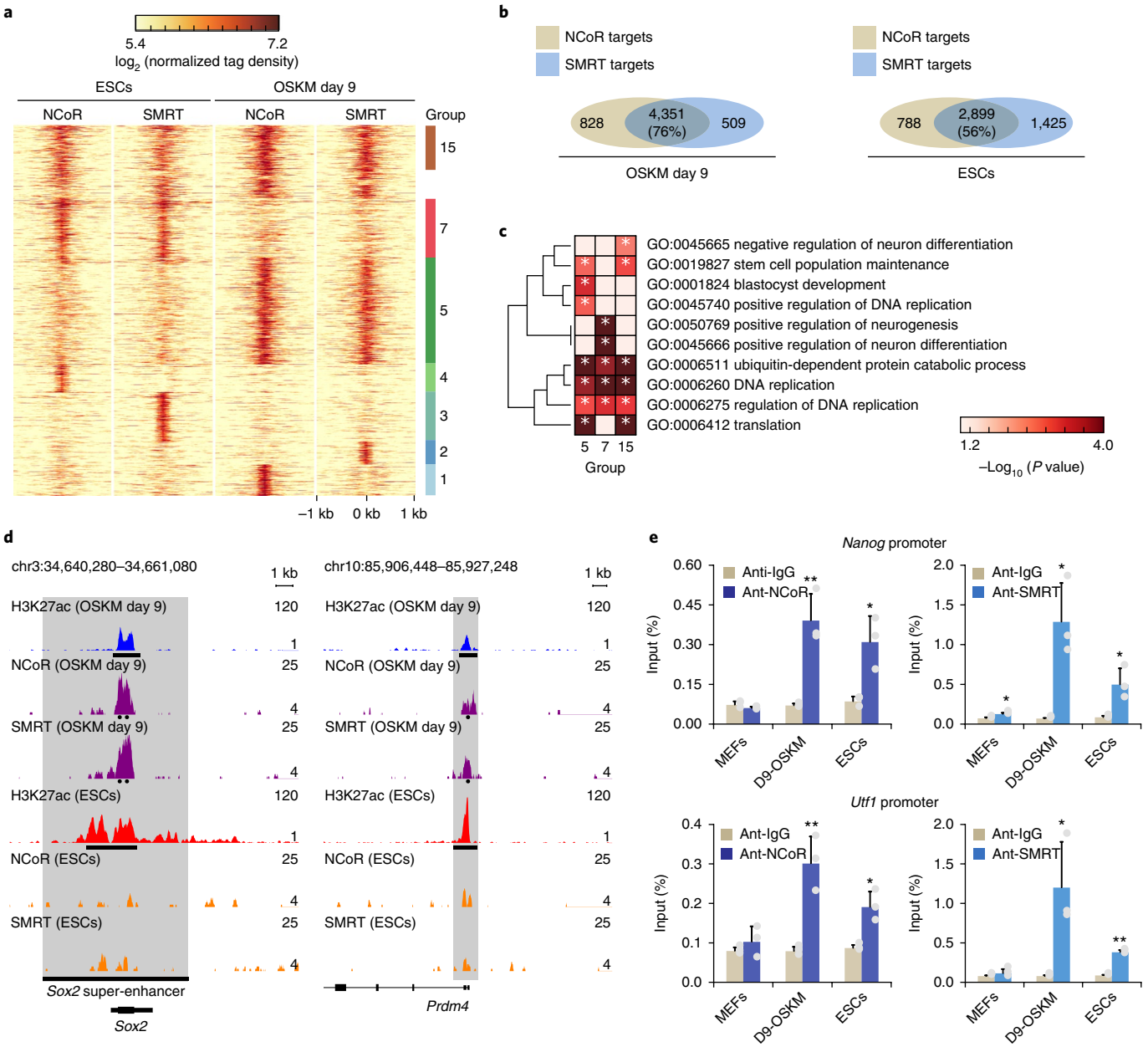

Fig. 4 | Context-specific binding of NCoR/SMRT co-repressors in OSKM reprogramming and ESCs. a, Tag density pileups of NCoR or SMRT peaks in OG2 ESCs or on day 9 of MEFs transduced with OSKM in serum + Vc. Peaks were merged into the same genomic loci if the summits were within 400 bp, and the resulting merged peaks were allocated to groups depending on their binding pattern in the different conditions. The three major groups of NCoR/SMRT binding are: group 5 (reprogramming specific), group 7 (ESC specific) and group 15 (common between reprogramming and ESCs). For clarity, groups 6 and 8–14 are not labelled. **b**, Overlap of NCoR-binding and SMRT-binding sites on day 9 of OSKM reprogramming and in ESCs. NCoR and SMRT were considered as overlapping if their peak summits were within 200 bp of each other. **c**, GO analysis of groups 5, 7 and 15. NCoR/SMRT-binding sites were annotated to all genes within 2 kb and the closest gene within 20 kb; duplicate genes were removed. An extended list is in Supplementary Fig. 6b. *Benjamini-Hochberg corrected $P < 0.05$. **d**, Genome views of NCoR and SMRT binding at *Sox2* and *Prdm4*. Black dots under the tracks indicate that NCoR or SMRT was detected as bound at that location. The *Sox2* super-enhancer[54] is indicated with a black bar below. Note that, in ESCs, the H3K27ac signal extends along the *Sox2* super-enhancer. **e**, ChIP–qPCR analysis of NCoR and SMRT binding at *Nanog* and *Utf1* on day 9 of MEFs transduced with OSKM in serum + Vc; OG2 ESCs and MEFs are also included. Data are the mean ± s.e.m. ($n = 3$ biological replicates with 3 technical replicates each). The *P* value was calculated using two-tailed unpaired Student's *t*-test, *$P < 0.05$ and **$P < 0.01$.

NCoR/SMRT–HDAC3 complex is recruited by c-MYC in reprogramming. The exogenous OSKM factors were strong candidates for recruiting NCoR/SMRT–HDAC3 to DNA. Co-immunoprecipitation in HEK293T cells confirmed the interaction of OSK with NCoR/SMRT, but binding was stronger with c-MYC (Fig. 6a). Notably, this c-MYC–NCoR/SMRT interaction required the amino-terminal domain of NCoR and SMRT (Fig. 6b,c), which is necessary for their repressive function[11]. Likewise, we could co-immunoprecipitate the N-terminals of NCoR and SMRT with both exogenous c-MYC and endogenous HDAC3 at day 9 of OSKM reprogramming in serum + Vc (Fig. 6d), and, similarly, exogenous c-MYC could co-precipitate endogenous HDAC3 (Fig. 6e), all supportive of an interaction between c-MYC and the NCoR/SMRT–HDAC3 complex. Motif discovery in the NCoR/SMRT ChIP–seq

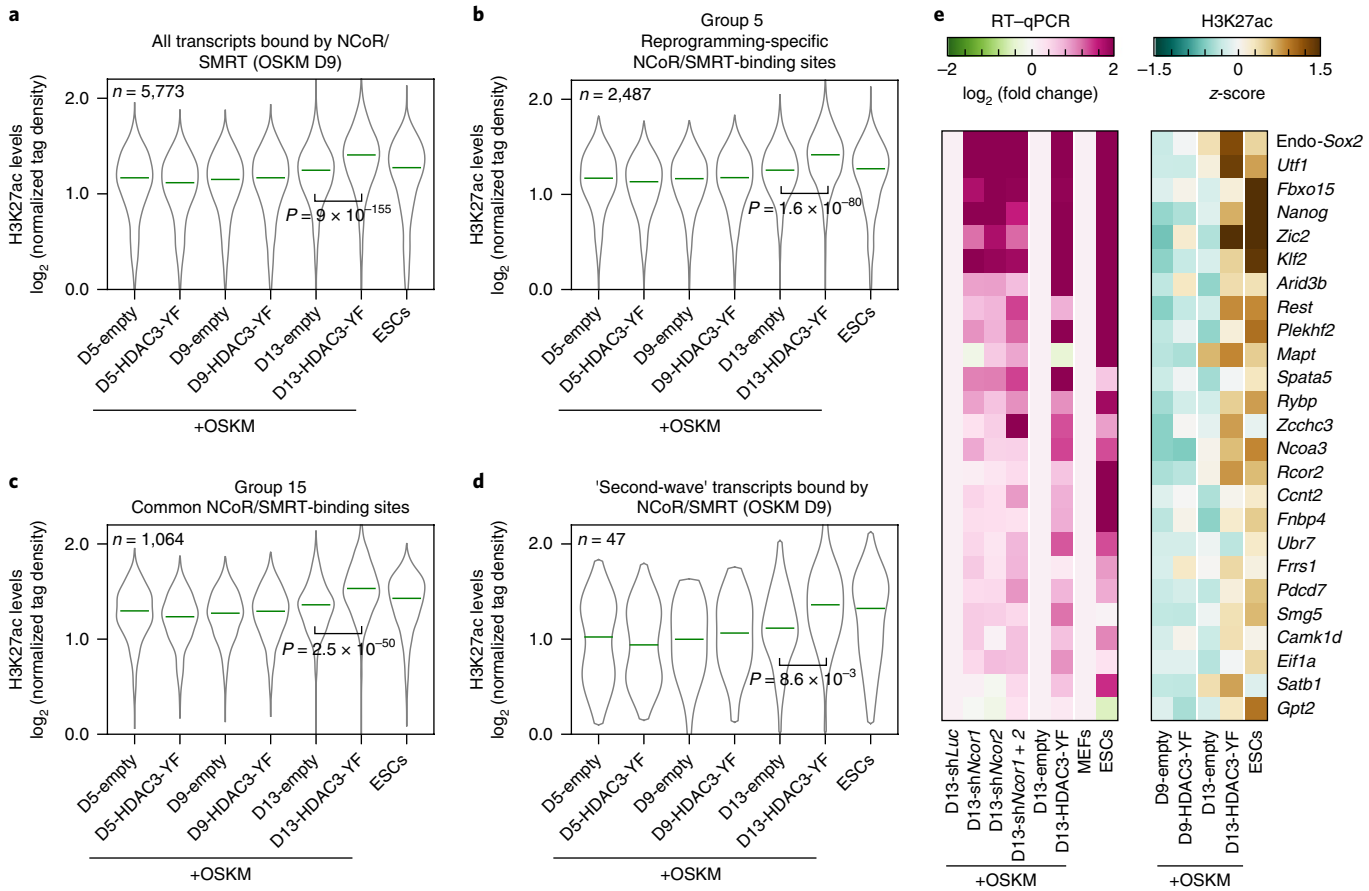

Fig. 5 | Interference of HDAC3 leads to an increase of H3K27ac at NCoR/SMRT-binding sites. a, Violin plots of the levels of H3K27ac within a 1 kb window centred on all OSKM day 9 NCoR/SMRT-binding peaks. Values were normalized to the mean of all peak intensities to account for differences in ChIP-seq efficiency (Supplementary Fig. 7a). The green bar shows the median of the data; the number of transcripts is indicated by '*n*'. The *P* value was calculated using two-tailed Mann–Whitney *U*-test, and only empty vector versus HDAC3-YF significance on day 13 is shown for clarity (also in subsequent violin plots). **b**,**c**, As in panel **a**, but containing reprogramming-specific peaks (group 5; **b**) and reprogramming and ESC common peaks (group 15; **c**) only (see also Supplementary Fig. 7b). The green bar shows the median of the data; the number of transcripts is indicated by '*n*'. The *P* value was calculated using two-tailed Mann–Whitney *U*-test. **d**, As in panel **a**, but showing the 'second-wave' genes with an NCoR/SMRT-binding site within 2 kb of the TSS (105 transcripts). The green bar shows the median of the data; the number of transcripts is indicated by '*n*'. The *P* value was calculated using two-tailed Mann–Whitney *U*-test. **e**, Twenty-five selected pluripotency-related genes including 21 'second-wave' genes bound by NCoR/SMRT at their TSS were analysed by RT–qPCR (green–pink) during a reprogramming time course of MEFs transduced with OSKM and shRNAs against *Luciferase*, *Ncor1*, *Ncor2*, combined *Ncor1* and *Ncor2*, empty vector, or HDAC3-YF, and in E14 ESCs and untransduced MEFs. sh*Ncor1/2* experiments are relative to sh*Luc*, HDAC3-YF are relative to the empty vector and ESCs are shown relative to MEFs. In addition, the levels of H3K27ac were measured (blue–brown) on day 9 or day 13 of an OSKM-mediated reprogramming time course for MEFs transduced with HDAC3-YF or empty vector. Tag density of H3K27ac was measured in a 1 kb window centred on the TSS of the indicated gene, and scores were converted to a *z*-score based on the total variance of all peaks.

data also discovered a putative c-MYC-binding motif besides several typical motifs associated with well-known transcription factor partners of NCoR/SMRT[38] (Fig. 6f). Moreover, *Ncor1/2* knockdown failed to enhance reprogramming efficiency in serum or serum + Vc with OSK (Fig. 6g), as we previously saw in the iCD1 reprogramming system (Supplementary Fig. 1m), and, likewise, expression of HDAC3-YF had no beneficial effect on reprogramming with OSK (Fig. 6h). Consistent with these results, ChIP–qPCR for NCoR/SMRT showed low enrichment at the *Nanog* and *Utf1* loci in OSK reprogramming compared to OSKM (Fig. 6i).

To study the genome-wide correlation between sites bound by NCoR/SMRT and c-MYC, we compared NCoR/SMRT binding at day 9 OSKM reprogramming with published ChIP–seq data[3] for OCT4, SOX2, KLF4 and c-MYC in pre-iPSCs, 48-h OSKM reprogramming and ESCs (Fig. 7a). Pre-iPSCs are stable clonal cell lines that appear frequently in OSKM reprogramming[39–41]; they have gone through the initial stages of reprogramming but failed to activate the pluripotency network. Notably, the strongest overlap of NCoR/SMRT with OSKM in pre-iPSCs was with c-MYC (72% of all NCoR/SMRT peaks), then with KLF4 (44%) and OCT4/SOX2 (both ~5%) (Fig. 7b,c). Yet, most of the overlapping sites between NCoR/SMRT and KLF4 in pre-iPSCs were co-bound by c-MYC (Fig. 7d), suggesting that c-MYC recruits NCoR/SMRT to those sites. There was also good overlap of NCoR/SMRT with OCT4 (75%), SOX2 (61%) and c-MYC (66%) in 48-h OSKM reprogramming, but a large proportion of those sites were shared between these three reprogramming factors (Fig. 7e–g). Many (86%) of the c-MYC-binding sites in 48-h OSKM reprogramming that overlapped with NCoR/SMRT sites were preserved in pre-iPSCs (Supplementary Fig. 8a). This suggests that c-MYC attracts NCoR/SMRT to those sites early in reprogramming. GO analysis of NCoR/SMRT and c-MYC co-bound sites in 48-h OSKM reprogramming or pre-iPSCs was enriched for stem cell-related and blastocyst-related terms (Supplementary Fig. 8b). Blastocyst-related terms were reflected in individual genes relevant for the late phase of reprogramming (for example, *Sox2*, *Prdm4* and *Klf2*)

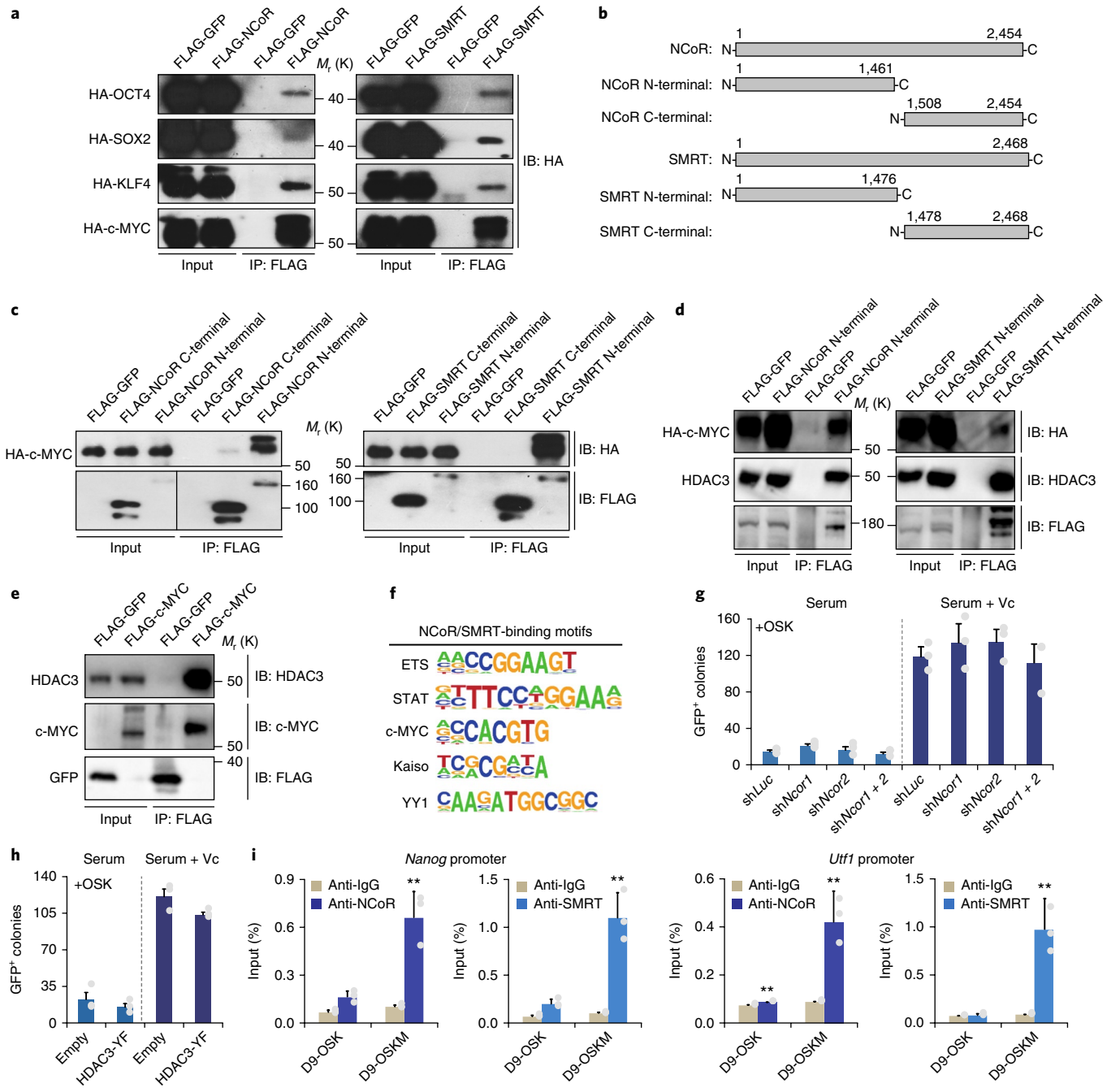

Fig. 6 | c-MYC mediates the effects of NCoR/SMRT–HDAC3 co-repressor complex on OSKM reprogramming. a, Immunoprecipitation (IP) of nuclear extracts from HEK293T cells transfected with HA-tagged individual OCT4, SOX2, KLF4 and c-MYC in combination with FLAG-tagged GFP, NCoR or SMRT, followed by immunoblotting (IB). **b**, Schematic maps of the NCoR/SMRT truncations used in this study; numbers are for amino acids. **c**, Immunoprecipitation of nuclear extracts from HEK293T cells transfected with HA-tagged c-MYC and FLAG-tagged GFP, NCoR, NCoR carboxy-terminal or NCoR N-terminal domain, SMRT, SMRT N-terminal or SMRT C-terminal domain, followed by immunoblotting. **d**, Immunoprecipitation of nuclear extracts from day 9 OSKM reprogramming in serum + Vc. MEFs were transduced with OSK and HA-c-MYC, along with FLAG-GFP, FLAG-NCoR N-terminal or FLAG-SMRT N-terminal domains. Samples were immunoprecipitated with FLAG antibody, and immunoblotting was performed with antibodies against the indicated proteins. **e**, Immunoprecipitation of nuclear extracts from day 9 OSKM reprogramming in serum + Vc. MEFs were transduced with OSK and FLAG-GFP or OSK and FLAG-c-MYC. Samples were immunoprecipitated with FLAG antibody, and immunoblotting was performed with antibodies against the indicated proteins. **f**, Selected, significantly enriched, DNA-binding motifs found in NCoR/SMRT-binding sites detected using HOMER. STAT, signal transducer and activator of transcription. **g**, The number of *Oct4*-GFP$^+$ colonies in MEFs transduced with OSK and shRNAs against *Luciferase*, *Ncor1*, *Ncor2*, or combined *Ncor1* and *Ncor2* in serum (day 21) or serum + Vc (day 18). Data are the mean ± s.e.m. ($n = 3$ biological replicates with 3 technical replicates each). **h**, The number of *Oct4*-GFP$^+$ colonies in MEFs transduced with OSK and co-transduced with HDAC3-YF or empty vector in serum (day 21) or serum + Vc (day 18). Data are the mean ± s.e.m. ($n = 3$ biological replicates with 3 technical replicates each). **i**, ChIP–qPCR for NCoR and SMRT binding at *Nanog* and *Utf1* in MEFs transduced with OSK or OSKM on day 9 in serum + Vc. Data are the mean ± s.e.m. ($n = 3$ biological replicates with 3 technical replicates). The *P* value was calculated using two-tailed unpaired Student's *t*-test, $^{**}P < 0.01$.

(Fig. 7h and Supplementary Fig. 8c). Conversely, the overlap of NCoR/SMRT at day 9 reprogramming with c-MYC (or OSK) in ESCs was low compared to pre-iPSCs (Supplementary Fig. 8d). NCoR/SMRT in ESCs did not overlap well either with c-MYC (or OSK) in ESCs, further supporting that NCoR/SMRT have different functions in reprogramming and in ESCs (Supplementary Fig. 8e).

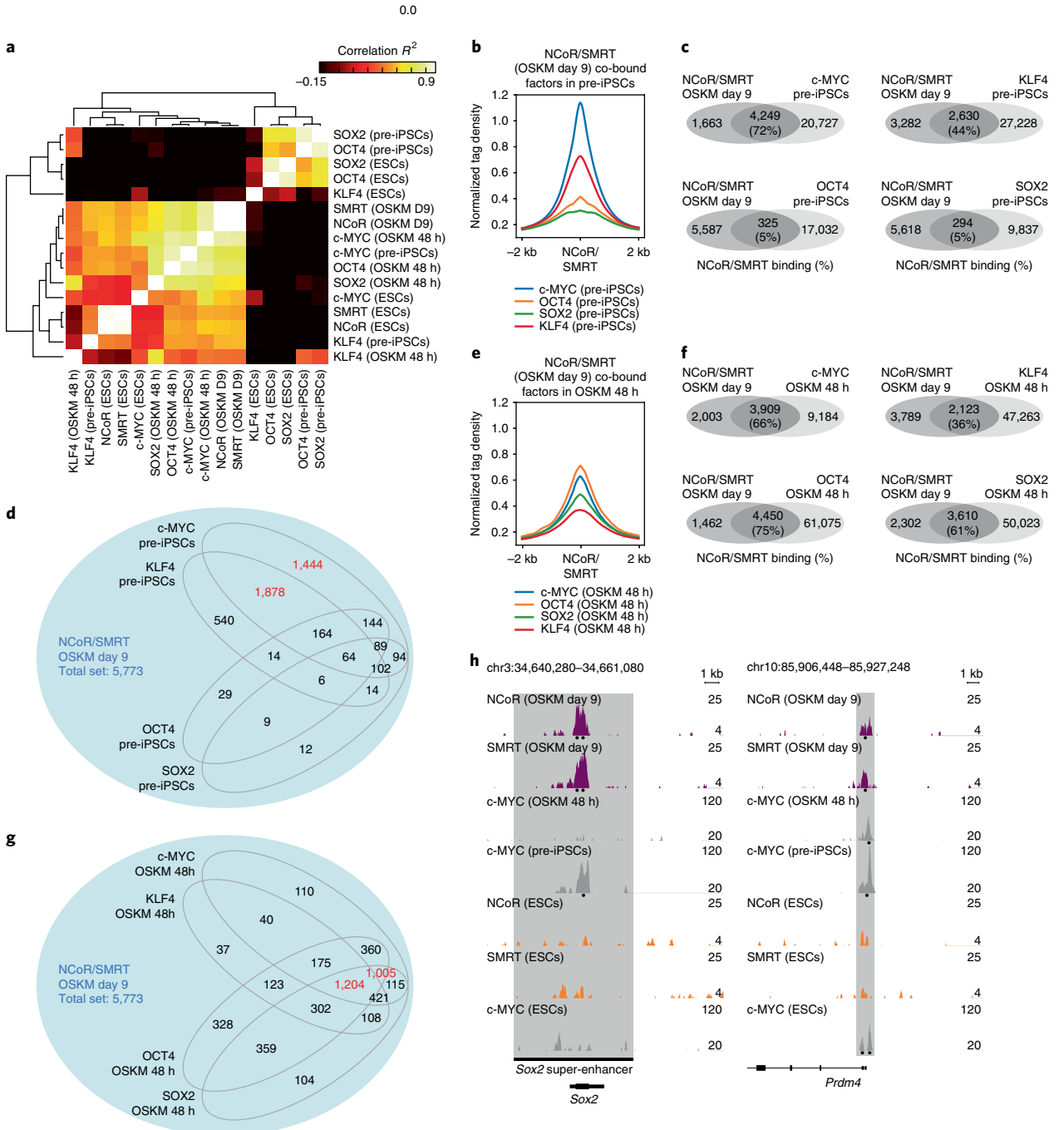

Fig. 7 | NCoR/SMRT cooperate with c-MYC to impair OSKM reprogramming. a, Correlation heatmap of the indicated ChIP–seq data. Peak data were taken from GSE90895 (the same for panels **b**–**h**). All ChIP–seq peaks were pair-wise overlapped, and the correlation pattern of overlaps (R^2) was measured. **b**, Pileup tag density plots for NCoR/SMRT-bound sites on day 9 of OSKM reprogramming versus the binding sites for OCT4, SOX2, KLF4 and c-MYC in pre-iPSCs. Plots are centred on NCoR/SMRT-binding sites, and the tag density of the indicated OSKM factors is shown. Data are normalized to library size. **c**, Pair-wise Venn overlap for NCoR/SMRT-bound sites on day 9 OSKM reprogramming versus OSKM in pre-iPSCs. Peaks were considered overlapping if their summits were within 400 bp. **d**, Four-way Venn diagram of the number of overlapping peaks between NCoR/SMRT binding on day 9 of OSKM reprogramming versus OCT4, SOX2, KLF4 and c-MYC binding in pre-iPSCs. Peaks were considered overlapping if the summits were within 400 bp of each other. Genes co-bound by NCoR/SMRT and both KLF4 and c-MYC, but not OCT4 and SOX2, are highlighted in red. **e**, As in panel **b**, but for 48 h OSKM reprogramming. **f**, As in panel **c**, but for 48 h OSKM reprogramming. **g**, As in panel **d**, but for 48 h OSKM reprogramming. Genes co-bound by NCoR/SMRT and OCT4, SOX2 and c-MYC are highlighted in red. **h**, Genome views for *Sox2* and *Prdm4*. The location of the *Sox2* super-enhancer[54] is indicated. Black dots under the tracks indicate that NCoR/SMRT or c-MYC was detected as bound at that location.

Thus, among the four Yamanaka factors, c-MYC mostly helps to recruit NCoR/SMRT to genomic loci (including pluripotency loci) to induce a negative effect on reprogramming.

Dual role of exogenous c-MYC in reprogramming. To further assess the relationship between c-MYC and NCoR/SMRT in the repression of pluripotency genes in reprogramming, we constructed a DsRed lentiviral reporter driven by the *Oct4* distal enhancer (*Oct4*-DE)[42] (Fig. 8a). This reporter integrates into the genome and can be directly activated by the binding of pluripotency transcription factors without the need to undergo full reprogramming. Individual overexpression of OCT4, SOX2 or KLF4 in MEFs modestly activated the reporter, but their combination was synergistic (Fig. 8b,c). By contrast, c-MYC alone had no activating effect, but combined with OSK, impaired the increase in reporter activity (Fig. 8b,c and Supplementary Fig. 8f). *Ncor1/2* knockdown partly relieved the negative effect of c-MYC on the reporter (Fig. 8b). These results implied that recruitment of NCoR/SMRT co-repressors to pluripotency loci by c-MYC is a negative force for reprogramming, but this is inconsistent with the observation that c-MYC added to OSK enhances reprogramming efficiency[43,44]. To investigate this further, we reprogrammed MEFs with OSK and a doxycycline-inducible c-MYC in serum + Vc (Fig. 8d). Induction of c-MYC from days 3–9 caused a peak in the number of colonies that were positive for the stem cell marker alkaline phosphatase, which appears relatively early in reprogramming (Fig. 8e). This increase was not further altered if c-MYC was induced during the entire experiment (Fig. 8e), consistent with previous reports[24,40]. The number of *Oct4*-GFP$^+$ colonies also reached a peak when c-MYC was induced from day 3 to day 9, but this number was significantly reduced when c-MYC was induced for a longer time (Fig. 8f). Conversely, inducible versions of OCT4, SOX2 and KLF4 were beneficial at all stages of reprogramming (Fig. 8g).

Next, to understand the temporal requirements of HDAC3 activity in mediating c-MYC effects, we treated cells with TSA or activated the expression of an inducible HDAC3-YF construct in an OSKM reprogramming time course in serum + Vc. TSA enhanced reprogramming efficiency less significantly when added in the late phase

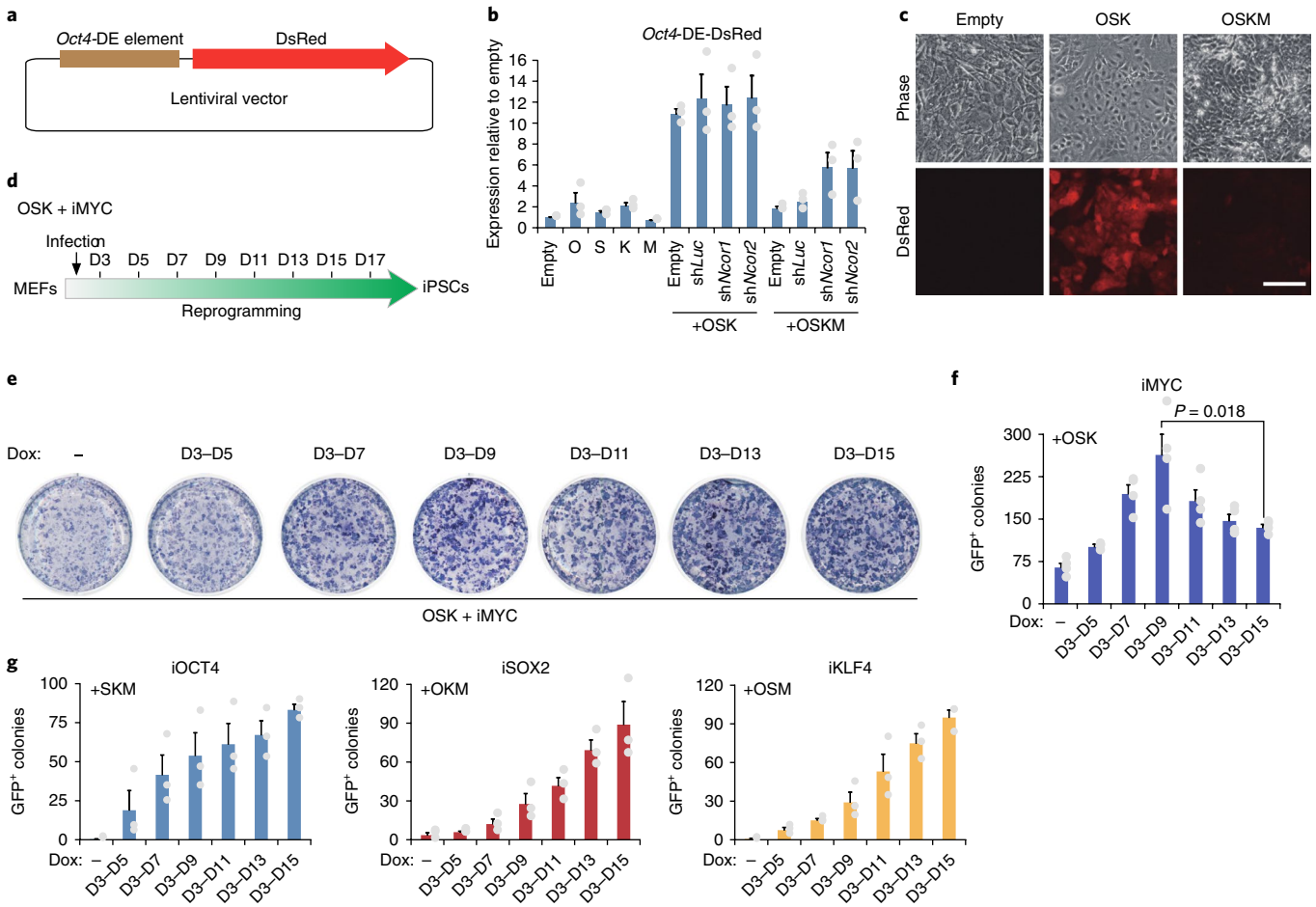

Fig. 8 | The deleterious effect of exogenous c-MYC in the late phase of reprogramming. a, Schematic of the *Oct4* reporter element. *Oct4*-DE was cloned in front of DsRed in a lentiviral expression vector (Oct4-DE-DsRed). **b**, RT–qPCR for DsRed 6 days after co-transducing ICR MEFs with the DsRed reporter and the indicated individual OCT4, SOX2, KLF4 and c-MYC, or combined OSK and OSKM, and shRNAs against *Luciferase*, *Ncor1* and *Ncor2* in serum + Vc. Data are the mean ± s.e.m. ($n = 3$ biological replicates with 3 technical replicates). **c**, Phase-contrast and fluorescence images detecting DsRed signal in MEFs co-transduced with the DsRed reporter and OSK, OSKM or empty vector in serum + Vc on day 6. Scale bar, 100 μm. **d**, Schematic depicting the treatment windows of doxycycline for the inducible transgenes in reprogramming. **e**, Alkaline phosphatase staining of MEFs transduced with OSK and inducible c-MYC (iMYC) in serum + Vc and treated with doxycycline (Dox) to induce c-MYC expression for the indicated time windows. Images were collected on day 16. **f**, As in panel **e**, but shows the number of *Oct4*-GFP$^+$ colonies counted on day 16. Data are the mean ± s.e.m. ($n = 4$ biological replicates with 3 technical replicates each). The *P* value was calculated using two-tailed unpaired Student's *t*-test. **g**, As in panel **f**, but for inducible OCT4 (iOCT4; GFP$^+$ colonies counted on day 20), inducible SOX2 (iSOX2; GFP$^+$ colonies counted on day 16) or inducible KLF4 (iKLF4; GFP$^+$ colonies counted on day 16). Data are the mean ± s.e.m. ($n = 3$ biological replicates with 3 technical replicates each).

(day 7–15) than in the early phase or all the time (Supplementary Fig. 8g). By contrast, the enhancing effect of activating HDAC3-YF in the late phase of reprogramming was comparable to its activation at all time points (Supplementary Fig. 8h). This discrepancy is probably related to the simultaneous inhibition of multiple HDACs by TSA[45]. Overall, these experiments show that repression of pluripotency loci by exogenous c-MYC is detrimental for the late phase of reprogramming through recruitment of NCoR/SMRT–HDAC3.

Several reports have shown that adding transactivator domains (for example, VP16), which partly work by recruiting histone acetyltransferases[46], to OCT4 and SOX2 greatly enhances OSK reprogramming efficiency[13]. Because OSK often co-bind together with c-MYC to target sites in OSKM reprogramming[4,47], we envisaged that OCT4-VP16 and SOX2-VP16 might counteract the repressive effect of NCoR/SMRT–HDAC3 on OSKM reprogramming. Indeed, we observed a significant increase of DsRed activity when OCT4-VP16 or SOX2-VP16 were overexpressed in combination with KLF4 and c-MYC, which was associated with an improvement in the number of *Oct4*-GFP$^+$ colonies in serum + Vc (Supplementary Fig. 8i,j). This suggests that, at least partly, one mechanism by which VP16-engineered transcription factors enhance reprogramming efficiency is by counteracting the negative effect of NCoR/SMRT co-repressors recruited by OSKM, especially by c-MYC, to pluripotency loci.

Discussion

The NCoR/SMRT–HDAC3 co-repressor complex interacts with all of OSKM to create an epigenetic barrier to reprogramming, but c-MYC is the major partner. This finding is paradoxical because exogenous c-MYC has traditionally been considered beneficial for reprogramming efficiency[43,44]. Yet, it helps to explain why reprogramming with OSKM but not OSK is prone to induce pre-iPSCs[21,40], and why HDAC inhibitors have a more robust effect on reprogramming with OSKM than with OSK[21,26,48]. It may also explain why high and sustained expression of OSKM can lead to a divergent reprogramming route (F-class cells) amenable to iPSC conversion with HDAC inhibitors[49], and why c-MYC negatively influences mouse[44,50,51] and human[52] iPSC quality. Because *Ncor1/2* knockdown also enhances somatic gene expression in the early phase of reprogramming and c-MYC contributes to shutting down the somatic cell programme[40], it is tempting to speculate that c-MYC similarly recruits NCoR/SMRT to repress somatic loci, although this effect may be HDAC3 independent.

Another co-repressor complex with a major role in reprogramming is the MBD3–NuRD complex[53], which contains HDAC1 and HDAC2 but not HDAC3. Thus, one could envisage that the simultaneous, yet optimized to balance somatic cell dedifferentiation/proliferation and pluripotency gene activation, suppression of components from both MBD3–NuRD and NCoR/SMRT–HDAC3 complexes, could allow universal and ultra-fast deterministic reprogramming.

Altogether, our work extends the repertoire of functions of NCoR/SMRT–HDAC3 co-repressor complex to somatic cell reprogramming and sheds light onto the intricate dual role of c-MYC in this process (see schematic in Supplementary Fig. 8k).

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41556-018-0047-x.

Received: 13 July 2017; Accepted: 29 January 2018;
Published online: 12 March 2018

References

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
2. Apostolou, E. & Hochedlinger, K. Chromatin dynamics during cellular reprogramming. *Nature* **502**, 462–471 (2013).
3. Chronis, C. et al. Cooperative binding of transcription factors orchestrates reprogramming. *Cell* **168**, 442–459 (2017).
4. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994–1004 (2012).
5. Cheloufi, S. et al. The histone chaperone CAF-1 safeguards somatic cell identity. *Nature* **528**, 218–224 (2015).
6. Mansour, A. A. et al. The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature* **488**, 409–413 (2012).
7. Singhal, N. et al. Chromatin-remodeling components of the BAF complex facilitate reprogramming. *Cell* **141**, 943–955 (2010).
8. Di Stefano, B. et al. C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature* **506**, 235–239 (2014).
9. Xu, Y. et al. Transcriptional control of somatic cell reprogramming. *Trends Cell Biol.* **26**, 272–288 (2016).
10. Perissi, V., Jepsen, K., Glass, C. K. & Rosenfeld, M. G. Deconstructing repression: evolving models of co-repressor action. *Nat. Rev. Genet.* **11**, 109–123 (2010).
11. Mottis, A., Mouchiroud, L. & Auwerx, J. Emerging roles of the corepressors NCoR1 and SMRT in homeostasis. *Genes Dev.* **27**, 819–835 (2013).
12. Szabó, P. E., Hübner, K., Schöler, H. & Mann, J. R. Allele-specific expression of imprinted genes in mouse migratory primordial germ cells. *Mech. Dev.* **115**, 157–160 (2002).
13. Wang, Y. et al. Reprogramming of mouse and human somatic cells by high-performance engineered factors. *EMBO Rep.* **12**, 373–378 (2011).
14. Pijnappel, W. W. et al. A central role for TFIID in the pluripotent transcription circuitry. *Nature* **495**, 516–519 (2013).
15. Di Stefano, B. et al. C/EBPα creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4. *Nat. Cell Biol.* **18**, 371–381 (2016).
16. Esteban, M. A. et al. Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell* **6**, 71–79 (2010).
17. Gao, Y. et al. Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell* **12**, 453–469 (2013).
18. Stadtfeld, M., Maherali, N., Borkent, M. & Hochedlinger, K. A reprogrammable mouse strain from gene-targeted embryonic stem cells. *Nat. Methods* **7**, 53–55 (2010).
19. Hanna, J. et al. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**, 595–601 (2009).
20. Blelloch, R., Venere, M., Yen, J. & Ramalho-Santos, M. Generation of induced pluripotent stem cells in the absence of drug selection. *Cell Stem Cell* **1**, 245–247 (2007).
21. Chen, J. et al. Rational optimization of reprogramming culture conditions for the generation of induced pluripotent stem cells with ultra-high efficiency and fast kinetics. *Cell Res.* **21**, 884–894 (2011).
22. Samavarchi-Tehrani, P. et al. Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7**, 64–77 (2010).
23. Li, R. et al. A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7**, 51–63 (2010).
24. Polo, J. M. et al. A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151**, 1617–1632 (2012).
25. Stender, J. D. et al. Control of proinflammatory gene programs by regulated trimethylation and demethylation of histone H4K20. *Mol. Cell* **48**, 28–38 (2012).
26. Huangfu, D. et al. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat. Biotechnol.* **26**, 795–797 (2008).
27. Guenther, M. G., Barak, O. & Lazar, M. A. The SMRT and N-cor corepressors are activating cofactors for histone deacetylase 3. *Mol. Cell Biol.* **21**, 6091–6101 (2001).
28. Sun, Z. et al. Deacetylase-independent function of HDAC3 in transcription and metabolism requires nuclear receptor corepressor. *Mol. Cell* **52**, 769–782 (2013).
29. Zhuang, Q. et al. Class IIa histone deacetylases and myocyte enhancer factor 2 proteins regulate the mesenchymal-to-epithelial transition of somatic cell reprogramming. *J. Biol. Chem.* **288**, 12022–12031 (2013).
30. Lahm, A. et al. Unraveling the hidden catalytic activity of vertebrate class IIa histone deacetylases. *Proc. Natl Acad. Sci. USA* **104**, 17335–17340 (2007).
31. Gaspar-Maia, A., Alajem, A., Meshorer, E. & Ramalho-Santos, M. Open chromatin in pluripotency and reprogramming. *Nat. Rev. Mol. Cell Biol.* **12**, 36–47 (2011).
32. Hatzi, K. et al. A hybrid mechanism of action for BCL6 in B cells defined by formation of functionally distinct complexes at enhancers and promoters. *Cell Rep.* **4**, 578–588 (2013).

33. Bhaskara, S. et al. Hdac3 is essential for the maintenance of chromatin structure and genome stability. *Cancer Cell* **18**, 436–447 (2010).
34. Bogani, D. et al. The PR/SET domain zinc finger protein Prdm4 regulates gene expression in embryonic stem cells but plays a nonessential role in the developing mouse embryo. *Mol. Cell Biol.* **33**, 3936–3950 (2013).
35. Liang, J. et al. Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat. Cell Biol.* **10**, 731–739 (2008).
36. Saunders, A. et al. The SIN3A/HDAC corepressor complex functionally cooperates with NANOG to promote pluripotency. *Cell Rep.* **18**, 1713–1726 (2017).
37. Wang, A. et al. Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* **16**, 386–399 (2015).
38. Barish, G. D. et al. The Bcl6-SMRT/NCoR cistrome represses inflammation to attenuate atherosclerosis. *Cell Metab.* **15**, 554–562 (2012).
39. Chen, J. et al. H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* **45**, 34–42 (2013).
40. Sridharan, R. et al. Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136**, 364–377 (2009).
41. Silva, J. et al. Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol.* **6**, e253 (2008).
42. Chew, J. L. et al. Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol. Cell Biol.* **25**, 6031–6046 (2005).
43. Wernig, M., Meissner, A., Cassady, J. P. & Jaenisch, R. c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell Stem Cell* **2**, 10–12 (2008).
44. Nakagawa, M. et al. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat. Biotechnol.* **26**, 101–106 (2008).
45. Haberland, M., Montgomery, R. L. & Olson, E. N. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat. Rev. Genet.* **10**, 32–42 (2009).
46. Hirai, H., Tani, T. & Kikyo, N. Structure and functions of powerful transactivators: VP16, MyoD and FoxA. *Int. J. Dev. Biol.* **54**, 1589–1596 (2010).
47. Soufi, A. et al. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555–568 (2015).
48. Liang, G., Taranova, O., Xia, K. & Zhang, Y. Butyrate promotes induced pluripotent stem cell generation. *J. Biol. Chem.* **285**, 25516–25521 (2010).
49. Tonge, P. D. et al. Divergent reprogramming routes lead to alternative stem-cell states. *Nature* **516**, 192–197 (2014).
50. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317 (2007).
51. Buganim, Y. et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209–1222 (2012).
52. Panopoulos, A. D. et al. Aberrant DNA methylation in human iPSCs associates with MYC-binding motifs in a clone-specific manner independent of genetics. *Cell Stem Cell* **20**, 505–517 (2017).
53. Rais, Y. et al. Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502**, 65–70 (2013).
54. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).

Acknowledgements

We thank all members of the Laboratory of RNA, Chromatin, and Human Disease for their comments, and Duanqing Pei and Jiekai Chen for their support, advice and/or technical assistance. We also thank J. Auwerx and R. M. Evans for providing materials. This work was supported by the National Key Research and Development Program of China (2016YFA0100102 and 2016YFA0100300), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16030502), the National Natural Science Foundation of China (31371513, 31471242, 81571238, 31671537, 31370995 and 31571524), the National Natural Science Foundation of China–Israel Science Foundation Joint Research Scheme (31461143011), the National Natural Science Foundation of China/Research Grants Council Joint Research Scheme (81261160506), the Guangdong Province Science and Technology Program (2013B050800010, 2014A030312001, 2014A030310327, 2016B030229007, 2016A050503037 and 2015A030308007), the Guangzhou Science and Technology Program (201508030027), the Queensland–Chinese Academy of Sciences (Q-CAS) project (GJHZ1242), the Shenzhen Science and Technology Council Basic Research program (JCYJ20150331142757383 and JCYJ20170307110638890) and the Chinese Academy of Sciences-Croucher Funding Scheme for Joint Laboratories (CAS16HU01/CAS16902). The Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine is supported by the Science and Technology Planning Project of Guangdong Province, China (2017B030314056). C.W. is supported by Zhujiang Talents-Overseas Young Talents Postdoctoral Fellowship, M.J.K. and S.K. are supported by the CAS President's International Fellowships, and D.P.I and M.M.A. are supported by the CAS-TWAS President's PhD Fellowship.

Author contributions

M.A.E., Q.Z. and C.B. conceived the original idea, and W.L. and A.P.H. contributed to the idea. Q.Z., W.L., C.B. and Z.H. conducted most of the experiments. A.P.H. and Z.Y. conducted all the bioinformatics analysis. Q.Z., W.L., C.B., A.P.H. and M.A.E. analysed the data. T.A., P.L., X.G., D.P.I., Z.L., M.Z., M.M.A., Z.Y., J.Y., Y.H., H.Z., D.H., J.Z., X.Zhong, X.Zhu, X.F., W.F., Y.L., Y.X., C.W., M.J.K. and S.K. contributed to the experiments and/or the analysis. B.M., M.D.T., H.-F.T., J.C., B.Q., X.B. and S.G. provided relevant advice and/or infrastructural support. M.A.E. supervised the study, with help from A.P.H., and provided most of the financial support. M.A.E. and A.P.H. wrote the manuscript with help from W.L.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41556-018-0047-x.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.P.H. or M.A.E.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Animal study. Animal experiments in this paper are compliant with all relevant ethical regulations regarding animal research, and were conducted under the approval of the Animal Care and Use committee of the Guangzhou Institutes of Biomedicine and Health under licence number 2013014. Chimeras were produced by injecting iPSCs into blastocysts derived from ICR mice, followed by implantation into pseudopregnant ICR mice.

Cell culture and reprogramming experiments. PlatE cells, HEK293T cells, ICR MEFs, OG2 MEFs[7,12–14,16,17] and the two types of secondary MEFs[17,18] were maintained in DMEM-high glucose (HyClone) supplemented with 10% FBS (Biowest). OG2 MEFs were used for all MEF experiments unless otherwise indicated; they have multiple copies of a transgenic *Oct4*-GFP reporter (Supplementary Fig. 1b–d). Primers for qPCR amplification of the *Oct4*-GFP reporter cassette[55] using genomic DNA are listed in Supplementary Table 5. Maintenance of mouse ESCs (OG2 and E14) and iPSCs was conducted on 0.1% gelatin (Millipore)-coated plates in chemically defined N2B27-based medium: DMEM/F12 (HyClone) and Neurobasal (Gibco) mixed 1:1, supplemented with N2 (Gibco), B27 (Gibco), non-essential amino acids (Gibco), GlutaMAX (Gibco), sodium pyruvate (Cellgro), penicillin/streptomycin (HyClone), 0.1 mM β-mercaptoethanol (Gibco), 1,000 U per ml leukaemia inhibitory factor (LIF; Millipore), CHIR99021 (3 μM; Selleck) and PD0325901 (1 μM; Selleck). For mouse reprogramming experiments, 15,000 MEFs were transduced in 12-well plates using retrovirus-containing supernatants generated by PlatE and lentivirus-containing supernatants generated by HEK293T cells. First, we transduced MEFs once with OSK or OSKM retroviruses, and after 12 h with the additional shRNA or overexpression viruses, as indicated, for another 24 h; *Ncor1* + *Ncor2* knockdown was executed by transducing with each virus for 12 h. HDAC3-YF retroviruses were administered diluted (one part virus media and two parts fresh media, the empty vector too) to avoid excessive reduction of proliferation. After the infections, the medium was changed to mouse ESC serum medium (DMEM-high glucose supplemented with 15% FBS, non-essential amino acids, GlutaMAX, sodium pyruvate, penicillin/streptomycin, 0.1 mM β-mercaptoethanol and 1,000 U per ml LIF), mouse ESC KSR medium (DMEM-high glucose supplemented with 15% KSR (Gibco), non-essential amino acids, GlutaMAX, sodium- pyruvate, penicillin/streptomycin, 0.1 mM β-mercaptoethanol, N2 (Gibco), 5 ng per ml basic fibroblast growth factor (bFGF; Peprotech) and 1,000 U per ml LIF), or iCD1 medium (prepared as described elsewhere[21]), and renewed daily. For the reprogramming of secondary MEFs without exogenous OSKM transduction, we transduced MEFs with shRNA retroviruses and 24 h later changed to serum + Vc medium with doxycycline. For the reprogramming with inducible c-MYC/OCT4/SOX2/KLF4, MEFs were first transduced with retroviruses of three reprogramming factors, and then transduced with inducible lentiviruses of the fourth reprogramming factor, and finally transduced with rtTA (reverse tetracycline-controlled transactivator) retroviruses. Reprogramming cells were not split at any time before extracting lysates or colony counting. Vc (A4034, Sigma) was used at 50 μg per ml, VPA (Sigma) at 1 mM, TSA (Sigma) at 20 nM and doxycycline (Sigma) at 1 μg per ml for the indicated times. ESCs were infected with shRNA viruses and selected with puromycin for 2 days after infection. For the reporter experiments, we first transduced ICR MEFs with the lentivirus reporter (viruses were administered diluted, one part virus media and two parts fresh media, to avoid high background), split into different wells and then transduced with the indicated retroviruses.

Plasmids. pMXs retroviral vectors separately expressing OSKM were purchased from Addgene. Plasmid pCMX-FLAG-NCoR was a kind gift from J. Auwerx (Ecole Polytechnique Federale de Lausanne, Switzerland) and pCMX-FLAG-SMRT was from R. M. Evans (Salk Institute for Biological Studies, California, USA); short forms of both genes were subcloned into pMXs. All other constructs, including the VP16-related plasmids, were made by us using complementary DNA (cDNA) obtained from MEFs, purchased from Addgene, or amplified from other plasmids, and cloned into pMXs or the lentiviral vectors pW-TRE (for the inducible system) or pRlenti (for the *Oct4*-DE reporter). DNA mutagenesis or deletion was produced using suitable oligos and a PCR-based method. shRNA inserts were cloned into pRetroSuper retroviral[56] or pLKO lentiviral (for the shRNAs in E14 ESCs) vectors. shRNA target sequences are listed in Supplementary Table 5.

Immunoprecipitation, immunoblotting, immunofluorescence and flow cytometry. For immunoprecipitation, HEK293T cells were transfected with each cDNA using jetPEI (Polyplus transfection) and lysed 48–72 h later in lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5% NP-40 and 1 mM EDTA). For reprogramming samples, 2×10^7 cells were used for each immunoprecipitation. Cells at day 9 of reprogramming were collected and washed twice with PBS, then lysed on ice with occasional vortexing in lysis buffer (with 1 mM fresh dithiothreitol (DTT) in the HDAC3 pull-down experiment). Anti-FLAG M2 magnetic beads (35 μl; M8823, Sigma) were incubated overnight with cell lysate fractions. Samples were then loaded on an Invitrogen magnetic separator, and beads were washed six times with TBS washing buffer (20 mM Tris-HCl pH 7.6 and 140 mM NaCl). Proteins were eluted with 0.5 mg per ml of FLAG peptide (Sigma) buffer and analysed by SDS–PAGE. Immunoblotting was performed following standard principles and blots were developed manually or with FUSION SOLO 4M (Vilber Lourmat; using the software Fusion-CAP v16.07). All unprocessed scans of blots are shown in Supplementary Fig. 9. Immunofluorescence was assessed using a Leica TCS SP2 spectral confocal microscope. The following primary antibodies were used for immunoprecipitation, immunoblotting or immunofluorescence: anti-FLAG (Sigma F7425; 1:1,500), anti-haemagglutinin (HA; Sigma H6908; 1:1,000), anti-stage-specific embryonic antigen 1 (SSEA-1; Cell Signaling MC480; 1:1,000), anti-c-MYC (Santa Cruz sc-764; 1:1,000), anti-histone H3 (Abcam ab1791; 1:5,000), anti-AcH3 (Millipore 06-599; 1:5,000), anti-H3K27ac (Abcam ab4729; 1:5,000), anti-AcH4 (Millipore 06-866; 1:1,000), anti-H4K12Ac (Abcam ab46983; 1:2,000), anti-HDAC3 (Abcam ab7030; 1:2,000) and anti-NANOG (Bethyl A300-397A; 1:500). Flow cytometry data were collected on a BD Accuri C6 machine and analysed with Flowjo (v10.4) software.

RNA isolation, RT–qPCR and RNA-seq. Total RNA was isolated following standard procedure. RT–qPCR analysis was performed using SYBR Green (Takara) and an ABI 7500 real-time PCR machine. Samples were run in triplicate or duplicate, and the values were normalized on the basis of *Gapdh*. Primers are listed in Supplementary Table 5. RNA-seq library preparation and sequencing were performed by RiboBio.

ChIP–qPCR and ChIP–seq. ChIP was performed as previously described[57] with slight modifications. Briefly, cells were crosslinked in freshly prepared formaldehyde solution (1% final concentration for 10 min at room temperature) and then quenched with 125 mM glycine (for 5 min at room temperature). Fixed cells were washed with PBS, harvested, flash-frozen in liquid nitrogen and stored at −80 °C for further use. Samples were lysed in nuclei extraction buffer (50 mM HEPES KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40 alternative and 0.25% Triton X-100) supplemented with protease inhibitor cocktail (Roche) for 10 min at 4 °C. Pellets were resuspended in protein extraction buffer (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA and 0.5 mM EGTA) supplemented with protease inhibitor cocktail and incubated for 10 min at room temperature. For sonication, pellets were resuspended in sonication buffer (10 mM Tris-HCl pH 8.0, 0.5 mM EGTA, 1% Triton X-100 and protease inhibitor cocktail), and then fragmented with a bioruptor (Diagenode) sonicator at 4 °C using high amplitude and 30 s on and 30 s off cycles to produce size ranges between 200 base pairs (bp) and 500 bp. 2 μg of each antibody was prebound by incubating with Protein A + G Dynabeads (Invitrogen 100-07D) in blocking buffer (PBS supplemented with 0.5% Tween) for 6 h at 4 °C. Washed beads were added to the chromatin lysate and incubated overnight. Samples were washed twice with low-salt washing buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS and 1% Triton X-100), twice with high-salt washing buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, 0.1% SDS and 1% Triton X-100), twice with LiCl buffer (10 mM Tris-HCl, 250 mM LiCl, 1 mM EDTA, 1% NP-40 and 1% Na-deoxycholate), twice with TE buffer (10 mM Tris-HCl pH 8.0 and 1 mM EDTA) supplemented with 50 mM NaCl and eluted in elution buffer (50 mM Tris-HCl, 10 mM EDTA and 1% SDS). Eluates were incubated at 65 °C for 20 min, followed by de-crosslinking at 65 °C for 6–15 h. Samples were diluted in TE buffer and then treated with RNase A (Roche) for 60 min at 37 °C, followed by incubation with proteinase K (Sigma) for 45 min at 56 °C. DNA was purified with phenol:chloroform:isoamyl alcohol and used for RT–qPCR or ChIP–seq. ChIP–qPCR primers are listed in Supplementary Table 5. ChIP–seq library preparation and sequencing were performed by RiboBio. The following antibodies were used for ChIP experiments: control anti-IgG (A7016, Beyotime), anti-NCoR (ABE251, Millipore), anti-SMRT (ab24551, Abcam) and anti-H3K27ac (ab4729, Abcam).

For HDAC3 ChIP–qPCR, the protocol was slightly modified as follows: cells were double fixed with disuccinimidyl glutarate (20593, Thermo) for 40 min and 1% formaldehyde for 10 min at room temperature. Then, 2×10^7 fixed cells were lysed in 1 ml ChIP buffer (50 mM Tris-HCl pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100 and 0.1% sodium deoxycholate) for 15 min on ice. Lysates were then centrifuged at 4,000 r.p.m. for 5 min at 4 °C and supernatants discarded. Sonication was processed in sonication buffer (50 mM Tris-HCl pH 7.5, 0.1% SDS and 10 mM EDTA). Additional experimental procedures are the same as above.

RNA-seq and ChIP–seq analysis. RNA-seq reads were aligned to the mouse Ensembl transcriptome (mm10, v76) using bowtie2 (v2.2.1)[58] and RSEM (v1.2.18)[59], GC-normalized using EDASeq (v2.0.0)[60]. Differential expression was called using DESeq2 (v1.6.9)[61] using a q value of 0.1 and fold change of >1.5. ChIP–seq reads were aligned to the mm10 genome using bowtie2, and peaks were called using MACS1 [62] for the NCoR/SMRT ChIP–seq and DFilter (v1.6)[63] for the H3K27ac ChIP–seq. A summary of the discovered peaks for the major NCoR/SMRT-binding groups is in Supplementary Table 4. Motif discovery was performed using HOMER[64]. All other analyses were performed using glbase[65].

Single-cell RT–qPCR and analysis. Single cells were collected from their respective time points using a 5–10-μm or 10–17-μm Fluidigm C1 capture plate according to the manufacturer's instructions. RT–qPCR was performed on a Fluidigm BioMark using TaqMan primer probes purchased from Life Technology (Supplementary Table 2). Note that TaqMan primers for *Oct4* target the 3′

untranslated region of endogenous *Oct4*. A total of 1,560 independent cells were used for the analysis. The resulting Ct (cycle threshold) values were transformed to average control at x (ACx, in this case, x is 25) values (Supplementary Table 3), essentially as described before[51], except a baseline Ct of 25 was chosen. Single-cell RT–qPCR analysis was performed using the 'realtime' and 'pca' modules of glbase[65]. The scatter plots in Fig. 2h and Supplementary Fig. 4k all share the same PCA projection. The primers detecting the exogenous transgenes (*Oct4*, *Sox2*, *Klf4* and *c-Myc*) were removed prior to performing PCA, as were the two control primers *Actb* and *Gapdh*.

Extended bioinformatics details. GO analysis for the RNA-seq and ChIP–seq data was performed using GOseq[66]. The ChIP–seq grouped pileup heatmaps were generated using the glbase function 'glglob.chip_seq_cluster_heatmap'. Briefly, the input ChIP–seq peaks were merged into a redundant set of genomic coordinates; all peaks within 400 bp were merged into the same unequally distributed bins and were then assigned to the appropriate binary classification group based on the binding pattern in the original ChIP–seq regions. Sequence tag pileup data were extracted for each genomic coordinate and the subsequent matrix was sorted from most complex group (top) to the least complex groups (bottom). RNA-seq and ChIP–seq data from previous publications were uniformly processed using the pipeline in the study or the peak files were taken directly from the publication and lifted over to mm10.

Statistics and reproducibility. Data of bar charts are represented as mean or the mean ± standard error of the mean (s.e.m.). For boxplots, the central line is the median, the boxes indicate the upper and lower quartiles, the whiskers indicate the 1.5 interquartile range and the circles are outliers. For the quantifications shown, the provided *n* values refer to independent experiments or sample sizes as indicated in each figure or figure legend. Significance was tested using two-tailed unpaired Student's *t*-tests, unless otherwise stated, or with a two-tailed Mann–Whitney *U*-test as indicated in the respective figure legends. The related *P* values or *P* value range are either shown in the figure or the figure legend. No statistical method was used to predetermine sample size. RNA-seq data in Fig. 1e–g and Supplementary Fig. 3a,c–f were performed in biological duplicates. Single-cell RT–qPCR data in Fig. 2g–i and Supplementary Fig. 4j–m were performed once with 1,560 independent cells. ChIP–seq data in Figs. 3e–g,4a–d,5a–e,6f and 7a–h and Supplementary Figs. 5a,b,f,6a–d,7a–e and 8a–e were performed once. For reproducibility of representative figures: Fig. 1b is representative of three independent experiments; Fig. 3a–d is representative of three independent experiments; Fig. 6a is representative of three independent experiments; Fig. 6c–e is representative of two independent experiments; Fig. 8c,e is representative of three independent experiments; Supplementary Fig. 1e,h is representative of three independent experiments; Supplementary Fig. 1m is representative of two independent experiments; Supplementary Fig. 2a–f was performed once with two or four iPSC clones depending on the experiment; Supplementary Fig. 4b,c is representative of two independent experiments; Supplementary Fig. 4d is representative of three independent experiments; Supplementary Fig. 6f is representative of three independent experiments; Supplementary Fig. 8f,i is representative of three independent experiments; and Supplementary Fig. 8g is representative of two independent experiments.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability. RNA-seq and ChIP–seq data are available in the Gene Expression Omnibus (GEO) database under the accession number GSE70740. Previously published RNA-seq data that were re-analysed for the HDAC expression in Fig. 2a are available under accession codes GSE57967 [67], GSE29278 [68], GSE44288 [69] and GSE20851 [70]. Previously published ChIP–seq data that were re-analysed here are available under the accession code GSE90895 [3]. Source data have been provided as Supplementary Table 6. All other data supporting the findings of this study are available from the corresponding authors on reasonable request.

References

55. Yeom, Y. I. et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881–894 (1996).
56. Grez, M., Akgun, E., Hilbergt, F. & Ostertag, W. Embryonic stem cell virus, a recombinant murine retrovirus with expression in embryonic stem cells. *Proc. Natl Acad. Sci. USA* **87**, 9202–9206 (1990).
57. Raghav, S. K. et al. Integrative genomics identifies the corepressor SMRT as a gatekeeper of adipogenesis through the transcription factors C/EBPβ and KAISO. *Mol. Cell* **46**, 335–350 (2012).
58. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
59. Li, B., & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
60. Risso, D., Schwartz, K., Sherlock, G. & Dudoit, S. GC-content normalization for RNA-seq data. *BMC Bioinformatics* **12**, 480 (2011).
61. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
62. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
63. Kumar, V. et al. Uniform, optimal signal processing of mapped deep-sequencing data. *Nat. Biotechnol.* **31**, 615–622 (2013).
64. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
65. Hutchins, A. P., Jauch, R., Dyla, M. & Miranda-Saavedra, D. glbase: a framework for combining, analyzing and displaying heterogeneous genomic and high-throughput sequencing data. *Cell Regen.* **3**, 1 (2014).
66. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene Ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).
67. Bao, X. et al. The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters. *Cell Res.* **25**, 80–92 (2015).
68. Shen, Y. et al. A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
69. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
70. Guttman, M. et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* **28**, 503–510 (2010).

nature research

Corresponding author(s): Miguel A. Esteban and Andrew P. Hutchin

Initial submission ☐ Revised version ☐ Final submission ☒

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   No statistical methods were used to predetermine sample sizes.

2. Data exclusions

   Describe any data exclusions.

   No data were excluded.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   Fig. 1a,1c,2b,2e,2f,3h-j,4e,5e (left column),6g-i,8b,8g, and Supplementary Fig. 1a,1c,1d,1f,1g,1i,1j,3g,3i,4a,4e,4h,4i,5c-e,6e,8h,8j were repeated 3 times independently (biological replicates with 3 technical replicates each); Fig. 2c,8f, and Supplementary Fig.3h were repeated 4 times independently (biological replicates with 3 technical replicates each); Supplementary Fig. 1k,4g were repeated 3 times (biological replicates with 2 technical replicates each); Fig. 1d and Supplementary Fig. 1l,2a-f,4f,6g,6h were performed once in the lab (3 technical replicates).

   The RNA-seq in Fig.1e and other related figures were performed in biological duplicates; the single-cell RT-qPCR in Fig.2h and related figures was performed once using 1,560 cells; the ChIP-seq of H3K27ac in Fig.3e and related figures was performed once; the ChIP-seq of NCoR/SMRT in Fig.4a and related figures was performed once.

   For reproducibility of representative figures: Fig. 1b,3a-d,6a,8c,8e, and Supplementary Figure 1e,1h,4d,6f,8f,8i are representative of 3 independent experiments; Fig. 6c-e, and Supplementary Fig. 1m,4b,4c,8g are representative of 2 independent experiments; Supplementary Fig. 2a-f were performed once with 2 or 4 iPSC clones depending on the experiment.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   Samples were not randomized for the experiments.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   The investigators were not blinded to group allocation during data collection and/or analysis. There is no need of blinding because most of experiments have been done by 2 or 3 researchers independently and are highly reproducible, besides, the ChIP-seq or RNA-seq results can cover whole genome or transcriptome and do not cause any bias.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☒	☐ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	All the qPCR or clone counting data were analyzed by Excel 2010; flow cytometry data were analyzed by Flowjo (v10.4); RNA-seq data were analyzed using bowtie2 (v2.2.1), RESM (v1.2.18), EDASeq, (v2.0.0) DESeq2 (v1.6.9), GOseq (1.22.0); ChIP-seq analysis was performed using glbase (commits 1101-1399; Hutchins, A.P. et al, Cell Regen., 2014) MACS (v1.4), and DFilter (v1.6); HOMER (v4.9.1) was used for motif discovery; single-cell RT-qPCR analysis was performed using the 'realtime' and 'pca' modules of glbase.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	Unique materials, such as plasmids or cell lines, will be made available to researchers upon reasonable request. Other reagents used in this study are all available from different commercial companies.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Western blot and immunofluorescence antibodies:
anti-AcH3 (Millipore 06-599; 1:5000),
anti-AcH4 (Millipore 06-866; 1:1000),
anti-histone H3 (Abcam ab1791; 1:5000),
anti-H3K27ac (Abcam ab4729; 1:5000),
anti-H4K12Ac (Abcam ab46983; 1:2000),
anti-FLAG (Sigma F7425; 1:1500),
anti-HA (Sigma H6908; 1:1000),
anti-HDAC3 (Abcam ab7030; 1:2000),
anti-c-MYC (Santa Cruz sc-764; 1:1000),
anti-NANOG (Bethyl A300-397A; 1:500),
anti-SSEA-1 (Cell Signaling MC480; 1:1000).

ChIP-seq antibodies:
anti-IgG (Beyotime, A7016, 2 µg/sample),
anti-H3K27ac (Abcam, ab4729, 2 µg/sample),
anti-NCoR (Millipore, ABE251, 2 µg/sample),
anti-SMRT (Abcam, ab24551, 2 µg/sample).

All the antibodies used in the manuscript were bought from commercial companies and are widely used for similar experiments by other researchers worldwide. The utility of these antibodies is stated on the websites of the corresponding suppliers.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

Plat-E (ATCC; RRID: CVCL_0063),
HEK293T cells (ATCC; RRID: CVCL_B488),
OG2 MEFs (generated in our institute),
ICR MEFs (generated in our institute),
OG2 secondary MEFs (a gift from Dr. Shaorong Gao, Tongji University, Shanghai, China),
Oct4-GFP knock-in secondary MEFs (a gift from Dr. Jose Polo, Monash University, Melbourne, Australia),
E14 mouse ESCs (ATCC, CRL-1821).

b. Describe the method of cell line authentication used.

Gene expression by RT-qPCR of MEF and ESC lysates was used to confirm the expression of specific somatic or pluripotency genes. No other cell line authentication was performed.

c. Report whether the cell lines were tested for mycoplasma contamination.

Yes, they were all monthly tested and devoid of mycoplasma contamination.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

To our knowledge, no cell lines used in this study are listed in the database of commonly misidentified cell lines maintained by ICLAC.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

The OG2 MEFs used in this study are isolated from E13.5 of the crossed offsprings of 129 female mice (purchased from Beijing Vital River Laboratory Animal Technology Co., Ltd.) and OG2 transgenic male mice (purchased from the Jackson Laboratory). The resulting iPSCs were injected into male ICR mouse blastocysts to form chimeras. ICR MEFs were isolated from E13.5 embryos of ICR mice.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Not applicable.

nature research

Corresponding author(s): Miguel A. Esteban and Andrew P. Hutchins

☐ Initial submission ☐ Revised version ☒ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	The reprogramming samples are digested with 0.05% trypsin at indicated time points and filtered with a membrane to remove cell clusters.
6. Identify the instrument used for data collection.	A BD Accuri C6 machine was used for the data collection.
7. Describe the software used to collect and analyze the flow cytometry data.	The flow cytometry data were collected with BD Accuri C6 software, and were analyzed by Flowjo (v10.4).
8. Describe the abundance of the relevant cell populations within post-sort fractions.	Most of the cells (around 80%) were within the post-sort fractions in the preliminary FSC/SSC gating.
9. Describe the gating strategy used.	We set the preliminary FSC/SSC gates based on cell size and complexity to remove debris and other events of non-interest, then we set the gate for the GFP positive populations based on GFP negative ICR MEFs (control).

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒

nature research

Corresponding author(s): Miguel A. Esteban and Andrew P. Hutchins

☐ Initial submission ☐ Revised version ☒ Final submission

ChIP-seq Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data deposition

1. For all ChIP-seq data:

☒ a. Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ b. Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

2. Provide all necessary reviewer access links.
The entry may remain private before publication.

This token: 'knybiqsqtnwllqr' can be used to provide access to GSE70736.

3. Provide a list of all files available in the database submission.

SRA files containing the raw FASTQ reads for all samples (not shown, as SRA numbers are not available until the entry is made public, the SRP number is SRP060656), and additionally:

Peak BED files:
esc_h3k27ac_peaks.bed.gz
esc_ncor_peaks.bed.gz
esc_smrt_peaks.bed.gz
mef_h3k27ac_peaks.bed.gz
oskm_d13_flag_h3k27ac_peaks.bed.gz
oskm_d13_hdac3yf_h3k27ac_peaks.bed.gz
oskm_d5_flag_h3k27ac_peaks.bed.gz
oskm_d5_hdac3yf_h3k27ac_peaks.bed.gz
oskm_d9_flag_h3k27ac_peaks.bed.gz
oskm_d9_hdac3yf_h3k27ac_peaks.bed.gz
oskm_d9_ncor_peaks.bed.gz
oskm_d9_smrt_peaks.bed.gz

BigWig Files:
esc_h3k27ac.bw
esc_input.bw
esc_ncor.bw
esc_smrt.bw
mef_h3k27ac.bw
oskm_d13_flag_h3k27ac.bw
oskm_d13_hdac3yf_h3k27ac.bw
oskm_d5_flag_h3k27ac.bw
oskm_d5_hdac3yf_h3k27ac.bw
oskm_d9_flag_h3k27ac.bw
oskm_d9_hdac3yf_h3k27ac.bw
oskm_d9_ncor.bw
oskm_d9_smrt.bw

4. If available, provide a link to an anonymized genome browser session (e.g. UCSC).

Not applicable.

Methodological details

5. Describe the experimental replicates.

There are no biological replicates.

6. Describe the sequencing depth for each experiment.

Reads are single end, 51 bp reads, with an estimated fragment length of 70-120 bp in size.

Sequencing statistics are: (total tags/uniquely mapped tags) was :

esc_h3k27ac (19375856/13114502)
esc_input (24187917/15609335)
esc_ncor (20127564/12395104)
esc_smrt (17554926/11617732)
mef_h3k27ac (17487231/13330315)
oskm_d13_flag_h3k27ac (16467364/11678424)
oskm_d13_hdac3yf_h3k27ac (17460337/12482034)
oskm_d5_flag_h3k27ac (20044440/13983183)
oskm_d5_hdac3yf_h3k27ac (19238964/12769070)
oskm_d9_flag_h3k27ac (17577074/10584681)
oskm_d9_hdac3yf_h3k27ac (18611448/12755085)
oskm_d9_ncor (20081348/13433924)
oskm_d9_smrt (20566884/13991188)

7. Describe the antibodies used for the ChIP-seq experiments.

The following antibodies were used for ChIP-seq experiments:
ChIP-seq antibodies:
anti-IgG (Beyotime, A7016, 2 µg/sample),
anti-H3K27ac (Abcam, ab4729, 2 µg/sample),
anti-NCoR (Millipore, ABE251, 2 µg/sample),
anti-SMRT (Abcam, ab24551, 2 µg/sample).

8. Describe the peak calling parameters.

Alignment:
bowtie2 -p 6 --very-sensitive --end-to-end --no-unal -U input.fastq.gz -x mm10 -S out.sam

For NCoR/SMRT ChIP-seq:
macs14 -g mm -bw=200 -m 6,40 -n name -t bam_file

For H3K27ac peaks:
run_dfilter.sh -lpval=6 -d=h3k27ac_file.bed -c=input.bed' -bs=100 -ks=60 -refine -o=oskmd13_flag_h3k27ac.bed &>results.err

9. Describe the methods used to ensure data quality.

No H3K27ac peaks are at an estimated FDR level of >5% and all reported peaks are >5 fold, as reported by DFilter.

For NCoR/SMRT peaks, >96% of peaks have >5 fold enrichment. No peaks are at an FDR level >5%, as reported by Macs.

10. Describe the software used to collect and analyze the ChIP-seq data.

ChIP-seq analysis was performed using glbase (commits 1101-1399; Hutchins, A.P. et al, Cell Regen., 2014), MACS (v1.4), DFilter (v1.6), and HOMER (v4.9.1).