Stem Cell Reports

Genome-scale CRISPRa screen identifies novel factors for cellular reprogramming.

--Manuscript Draft--

Manuscript Number:	STEM-CELL-REPORTS-D-18-00530R1
Article Type:	Research Article
Keywords:	CRISPR/Cas9; CRISPR activation; Epiblast Stem cells; reprogramming; Genome-wide screen; Activation screen; Mouse; Gain-of-function; CRISPR screen; Reprogramming pathways; induced pluripotent stem cells; Embryonic stem cells
Corresponding Author:	Emmanouil Metzakopian, Ph.D Cambridge university Cambridge, UNITED KINGDOM
First Author:	Emmanouil Metzakopian, Ph.D
Order of Authors:	Emmanouil Metzakopian, Ph.D
	Jian Yang, Ph. D
	Sandeep Sundara Rajan, Ph. D
	Mathias Friedrich, M.D
	Guocheng Lan, Ph. D
	Xiangang Zou, Ph. D
	Hannes Ponstingl, Ph. D
	Dimitrios Garyfallos
	Pentao Liu, Ph. D
	Allan Bradley, Ph. D
Abstract:	Primed epiblast stem cells (EpiSCs) can be reverted to a pluripotent embryonic stem cell (ESC) like state by the expression of single reprogramming factors such as Klf4 and Nanog. To date, only a few genes have been shown to be able to return EpiSCs to pluripotency and their molecular mechanisms are not fully understood. We used CRISPRa to perform a genome-scale, gain of function (GoF) reprogramming screen in EpiSCs. We identified 142 candidates, amongst them known reprogramming factors such as Oct4, Nanog, Klf2 and Nr5a2 and validated 50 novel genes, of which we chose Sall1 for further investigation. We show that Sall1 augments reprogramming of mouse embryonic fibroblasts as well and that these induced pluripotent stem cells are indeed fully pluripotent including the formation of chimeric mice. In addition, we demonstrate that Sall1 synergises with Nanog in re-programming and that overexpression in ESCs delays their conversion from naïve ESCs to primed EpiSCs by keeping them in a formative state. Using RNA-seq, we identify and validate Klf5 and Fam189a2 as new downstream targets of Sall1 and Nanog. In summary, our study identifies a number of novel candidates that potentially play a role in EpiSC reprogramming and provides new insights into molecular mechanisms of the transition between naïve and primed states. This work demonstrates the power of using CRISPR technology in understanding complex processes such as reprogramming.

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Genome-scale CRISPRa screen identifies novel factors for cellular reprogramming

Jian Yang[‡⊥+], Sandeep S Rajan[‡†+], Mathias J Friedrich[‡+], Guocheng Lan[¶], Xiangang Zou[¶], Hannes Ponstingl[‡], Dimitrios A Garyfallos[‡], Pentao Liu[‡¢], Allan Bradley[‡], and Emmanouil Metzakopian[‡†*]

[‡]Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK

[†]UK Dementia Research Institute, Department of Clinical Neuroscience, University of Cambridge, CB2 0AH, UK

[¶]Cancer Research UK, Cambridge Institute, Li Ka Shing Centre, University of Cambridge, Cambridge, CB2 0RE, UK

[¢]School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, Stem Cell and Regenerative Medicine Consortium, University of Hong Kong, Hong Kong, China

[⊥]*Current address:* Key Laboratory of Arrhythmias, Ministry of Education, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China

[+]These authors contributed equally to the work.

*****Correspondence should be addressed to:

Emmanouil Metzakopian

UK Dementia Research Institute
Cambridge Biomedical Campus
University of Cambridge
Cambridge, CB2 0AH, U.K

Email: em698@medschl.cam.ac.uk

Abstract

Primed epiblast stem cells (EpiSCs) can be reverted to a pluripotent embryonic stem cell (ESC) like state by the expression of single reprogramming factors such as *Klf4* and *Nanog*. To date, only a few genes have been shown to be able to return EpiSCs to pluripotency and their molecular mechanisms are not fully understood. We used CRISPRa to perform a genome-scale, gain of function (GoF) reprogramming screen in EpiSCs. We identified 142 candidates, amongst them known reprogramming factors such as *Oct4*, *Nanog, Klf2* and *Nr5a2* and validated 50 novel genes, of which we chose *Sall1* for further investigation. We show that *Sall1* augments reprogramming of mouse embryonic fibroblasts as well and that these induced pluripotent stem cells are indeed fully pluripotent including the formation of chimeric mice. In addition, we demonstrate that *Sall1* synergises with *Nanog* in reprogramming and that overexpression in ESCs delays their conversion from naïve ESCs to primed EpiSCs by keeping them in a formative state. Using RNA-seq, we identify and validate *Klf5* and *Fam189a2* as new downstream targets of *Sall1* and *Nanog*. In summary, our study identifies a number of novel candidates that potentially play a role in EpiSC reprogramming and provides new insights into molecular mechanisms of the transition between naïve and primed states. This work demonstrates the power of using CRISPR technology in understanding complex processes such as reprogramming.

Introduction

The ability of pluripotent stem cells (PSCs) to self-renew and their potential to differentiate into multiple cell types makes them potentially useful for clinical applications [1,2]. PSCs can either be derived from early embryos or be induced by reprogramming somatic cells with Yamanaka factors i.e., *Oct4*, *Sox2*, *c-Myc* and *Klf4* amongst other transcription factors, mRNAs, microRNAs and small molecules [3-6]. These PSCs are referred to as induced pluripotent stem cells (iPSCs). During early mouse embryo development, at least two types of PSCs can be derived, naïve embryonic stem cells (ESCs) from the inner mass of the blastocyst and primed post-implantation epiblast stem cells (EpiSCs) [7-10]. While both have the potential to differentiate into multiple lineages, only ESCs can contribute extensively to chimeras, showing unbiased developmental potential. Both ESCs and EpiSCs express major pluripotent transcription factors such as *Oct4* and *Sox2* at similar levels. In EpiSCs however, reduced expression of pluripotency-associated factors such as *Rex1* and *Klf4* and elevated levels of early differentiation markers like *Fgf5, Gata6* and *Otx2* indicate their restricted developmental potential. Interestingly, EpiSCs cultured in fully defined ESC medium (with selective inhibition of mitogen-activated protein kinase and glycogen synthase kinase 3 and Leukemia Inhibitory Factor (LIF); 2i/LIF medium) can be reprogrammed into ESCs by over-expressing single gene such as *Nanog*, *Klf4* or *Nr5a2*[11]. Therefore, EpiSCs are a useful model to identify novel reprogramming factors through genetic screens.

Recently, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and its associated protein, Cas9, has gained importance by achieving simple, precise and rapid editing of the genome, enabling large scale experiments such as genetic

screening. While the RNA-programmable (single guide RNA, sgRNA) endonuclease Cas9 is used to induce double-strand breaks in defined genomic locations, its catalytically dead variant (dCas9) can be fused with transcriptional activators and directed towards promoter regions to increase gene expression (CRISPR activation, CRISPRa) [12,13].

Genome-wide screening is a powerful unbiased approach to discover genes and pathways that underlie biological processes. To date, the identification of key transcription factors and epigenetic modifiers within naïve and primed PSCs has been investigated by employing either gain-of-function (GoF) screens using cDNA libraries and *PiggyBac* transposons or loss-of-function (LoF) screens using or RNA interference [14-16].

Here, we describe the development and application of a genome-scale CRISPRa screen to identify genes that contribute to mouse EpiSC reprogramming. We show that our screening approach not only detects established reprogramming factors such as *Oct4* and *Nanog*, but also identifies novel candidate genes. We focus on the role of *Sall1*, a transcription factor belonging to the Spalt-like gene family, which has been implicated in cellular reprogramming in a number of studies but has not been sufficiently investigated[17-20]. Our work substantiates *Sall1* as a potent reprogramming gene candidate by demonstrating its ability to reprogram EpiSCs and mouse embryonic fibroblasts (MEFs) to iPSCs. In addition, we show that *Sall1* may exert its functions by interacting synergistically with *Nanog* to reprogram cells to ground state pluripotency.

Results

GoF CRISPRa screen identifies novel reprogramming genes

As a first step towards a genome-scale CRISPRa reprogramming screen, we sought to determine the optimal Cas9 transactivation system, as several variants have been published[21-26]. To that end, we created *PiggyBac*-transposable expression vectors with a Blasticidin-mCherry cassette for four different dCas9-CRISPRa systems: dCas:VP160, dCas9:SunTag, dCas9:VPR and dCas9:SAM (synergistic activation mediator, Supplementary Figure 1).

Furthermore, we designed a versatile sgRNA expression construct (pKLV-PB-U6-gRNA-PGK-Puro-T2A-TagBFP)[27] harbouring a Puromycin selection cassette with a blue fluorescent protein (BFP) marker (Supplementary Figure 1), which can be stably integrated into target genomes either as Lentivirus or via *PiggyBac* mediated transposition.

Using these vectors and single guides directed against the promoter region of two genes with low baseline expression, *Ascl1* and *Neurog2*, we sought to find the most potent CRISPR activation system. After stable integration of dCas9-CRISPRa and the sgRNA-vectors into HEK293 cells via transposition and antibiotic selection, RT-qPCR revealed that dCas9:SAM achieved the highest overexpression of both target genes and thus was chosen as CRISPRa system for all other experiments (Figure 1a).

To perform a genome-scale activation screen, we designed a pooled library of 87,863 sgRNAs targeting a 250 bp region upstream of the transcription start site (TSS) of 19,994 genes with an average of 4 guides each (Figure 1b and Supplementary Table 5).

We decided to use EpiSC derived from *Oct4*-GFP reporter transgenic mice as they have been used for this purpose before[28]. Characteristically for EpiSCs these cells already exhibit a baseline *Oct4* (and therefore GFP) expression. However, only cells successfully reprogrammed to the naïve pluripotent state are able to maintain and increase *Oct4* expression upon plating in medium containing Leukemia Inhibitory Factor (LIF), with selective inhibitors against mitogen-activated protein kinase and glycogen synthase kinase 3; (2i), hereafter 2i/LIF medium. Thus, successfully reprogrammed *Oct4*-GFP EpiSCs can be identified by their strong GFP expression (Supplementary Figure 2a) and the characteristic ESC-like morphology, and grow as distinct colonies, whereas EpiSCs failing to reprogram either detach and die or differentiate.

For the screen, we first generated *Oct4*-GFP EpiSCs expressing dCas9:SAM using *PiggyBac* transposase mediated stable integration of our dCas9:SAM-Blasticidine cassette and antibiotic selection. We then transduced 100×10^6 dCas9:SAM expressing EpiSCs with our sgRNA expression construct containing the library at a multiplicity of infection (MOI) of 0.3 (Supplementary Figure 2b). Two days later, we FACS-sorted 10×10^6 successfully transduced cells by BFP expression, giving a library coverage of around 114-fold. These BFP^+ cells were seeded in 2i/LIF medium to select for reprogramming cells. After 14-16 days of culture in 2i/LIF, 480 GFP^+ colonies were harvested for expansion (Figure 1c). Next generation sequencing revealed 146 sgRNAs targeting 142 different genes (Supplementary Table 6). These included known reprogramming factors *Nanog*[29,30], *Klf2*[31] and *Nr5a2*[11], confirming the specificity of the screen.

GOTERM analysis on these 142 genes identified an enrichment in pathways related to transcriptional activation, expression of various transcription factors and enrichment towards stem cell maintenance (Figure 1d).

To validate these candidate genes individually, we chose the highest performing sgRNA for each from the library, including *Nanog* as a positive control and again transduced dCas9:SAM expressing *Oct4*-GFP EpiSCs. We expected the validation rate to be no higher than 50%, as small scale single colony sub-sampling showed an average of 2 sgRNAs present in most colonies (*data not shown*), where one sgRNA presumably acts as the driver responsible for reprogramming while the other is co-amplified as a passenger. After plating and culture in 2i/LIF medium as before, GFP^+ ESC-like colonies could be observed for *Oct4*, *Nanog* and 52 of the candidate genes, resulting in a 36 % validation rate (Supplementary Table 7). The efficiency of reprogramming was gene dependent ranging from 5 to 165 colonies per 1 x 10^6 cells transfected (Figure 1e). Amongst the genes with the highest colony counts were our positive controls *Nanog* and one of the Yamanaka factors *Oct4*, as well as transcription factors *Klf2 and Nr5a2* with a known role in reprogramming, confirming the validity of our CRISPRa approach.

Gene dosage is critical for *Oct4* mediated reprogramming

We observed in our screen that CRISPRa mediated induction of the pluripotency marker *Oct4* produced a significant number of ESC-like colonies, contradicting previous studies which suggest that continuous expression of *Oct4* using cDNA is inefficient in EpiSC reprogramming[11,32]. We therefore overexpressed *Oct4* via cDNA in our EpiSCs and indeed were unable to generate any iPSC colony, while CRISPRa achieved robust reprogramming (Figure 2b).

We speculated that gene dosage might be the underlying issue and repeated the experiment using a tet-inducible *Oct4*-cDNA with the aim of titrating the *Oct4* amount. CRISPRa mediated induction of endogenous *Oct4* mRNA achieved roughly half the expression level found in ESCs, whereas total *Oct4* mRNA expression in EpiSCs transfected with constitutive *Oct4* cDNA slightly surpassed it (Supplementary Figure 2c). Titration of Doxycyclin (Dox) mostly resulted in *Oct4* mRNA expression amounts comparable to constitutive *Oct4* cDNA and only very low concentrations of Dox gave levels similar to CRISPRa. Nevertheless, all cDNA mediated overexpression conditions failed to reprogram EpiSCs, while CRISPRa again succeeded. When checking the expression of *Oct4* on the protein level via Western Blot, we found that *Oct4* cDNA derived protein reached disproportionally higher amounts than expected from the qPCR results (Supplementary Figure 2d, top panel). CRISPRa, on the other hand, achieved *Oct4* protein expression similar to that in ESCs (Supplementary Figure 2d, bottom panel). We suspect that differences in mRNA stability might be the reason for these results, as CRISPRa simply drives the endogenous mRNA which then will be physiologically regulated, while cDNA derived mRNA could, for example, be more stable due to its differing polyadenylation.

This indicates that CRISPRa could have some additional utility in hitting a goldilocks zone of induction for genes where artificially high expression from traditional cDNA constructs might prove detrimental. This agrees with our observation that although our screening library contained an average of 4 sgRNAs per gene, almost all candidate genes from our screen were derived from only one specific sgRNA per target. Indeed, when measuring transactivation efficiency of all sgRNAs for the candidate gene *Sall1* and the positive control gene *Nanog* via qPCR, sgRNAs showed vastly different activities in a distribution that suggests a dependency on the distance of the sgRNA

to the transcription start site (Supplementary Figure 2e). Indeed, this is also supported by a recent report by Liu et al.[33] which shows proof-of-principle MEF reprogramming using CRISPRa. In their experiments, only sgRNAs targeting the *Oct4* promoter in very specific locations (-71 and -127 bp from the transcription start site, TSS) achieved activation sufficient for reprogramming, while in our experiments, a sgRNA -101 bp from the TSS was successful.

Sall1 **facilitates EpiSC reprogramming in cooperation with** ***Nanog***

Umodl1 and *Sall1* were the two most potent validated candidates from our screen. We confirmed that *Umodl1* upregulates *Lifr, Essrb, Nanog and Sox2* and downregulated *Tgfbr1* as would be expected in iPSC reprogramming when media was switched from EpiSC to 2i/LIF (Supplementary Figure 2f). We decided to examine *Sall1* further because, as a member of the Spalt-family of transcription factors, it already had been reported to cooperate with *Nanog* to promote the maintenance of ESC stem cell state[34,35] and to play an important role in reprogramming and ESC differentiation [17,19,20,34]. However, the downstream targets of *Sall1* involved in reprogramming have not been sufficiently explored. Having found that *Sall1* is also able to independently reprogram EpiSCs, we set to investigate the underlying mechanisms.

First, we asked whether *Sall1* and *Nanog* also act synergistically in EpiSC reprograming by overexpressing them individually and combinatorically in *Oct4*-GFP EpiSCs. In order to further validate our experimental approach, we performed these experiments both with CRISPRa as well as cDNA mediated overexpression and also verified that the observed activity of the *Sall1* specific sgRNA was not due to cross-reactivity with *Sall4*, a known pluripotency factor (Supplementary Figure 2g).

We confirmed CRISPRa induction and cDNA mediated overexpression of *Sall1*, *Nanog* and *Oct4* via RT-qPCR 72h after transfection and found a 2.5 – 3.5 fold increase in expression by CRISPRa and a 10 - 20 fold increase in expression through cDNA (Supplementary Figure 3a). EpiSC reprogramming efficiencies were evaluated as *Oct4*-GFP^{+}, ESC-like colonies as described above and revealed a marked increase in colony number when *Sall1* and *Nanog* were co-expressed (Figure 2a). As mentioned above, *Oct4* induction via CRISPRa was successful in reprogramming EpiSCs in contrast to cDNA overexpression, but we did not observe its significant synergy with either *Sall1* or *Nanog*. The reprogrammed colonies exhibited strong GFP expression and had an ESC-like morphology (Figure 2b). Pluripotency markers examined by RT-qPCR (*Rex1*, *Sox2, Klf4* and *Essrb*) were markedly increased in all the reprogrammed colonies; concordantly, EpiSC markers *Gata6*, *Fgf5* and *Otx2* showed decreased expression (Figure 2c). *Sall1* reprogrammed EpiSCs (MF1 and C57BL/6 background) contributed significantly to chimeras when injected into C57BL/6 blastocysts (Figure 2d and Supplementary Figure 3e).

In order to exclude the possibility that the baseline GFP expression of the *Oct4*-GFP reporter EpiSCs might skew the correct identification of successfully reprogrammed EpiSCs, we repeated these experiments with *Nanog*-GFP reporter EpiSCs[28] which show strong GFP expression upon successfully entering the naive ESC state, but virtually none in the primed EpiSC state[11]. Both gene induction using CRISPRa and overexpression via cDNA confirmed the reprogramming capability of *Sall1* alone and in synergy with *Nanog* (Supplementary Figures 3b-d). Notably, colony formation assays in 2i/LIF recapitulated the results obtained with *Oct4*-GFP EpiSCs and the reprogramming kinetics as measured in time course experiments were comparable between the two reporter cell lines (Supplementary Figure 3f).

Sall1 and *Nanog* delay differentiation of ESCs into EpiSCs

ESCs readily differentiate into EpiSCs in culture medium containing the EpiSC self-renewal factors Activin and FGF2: after 3 passages, the cells begin to acquire the characteristics of EpiSCs[36]. Having confirmed the ability of *Sall1* and *Nanog* to reprogram EpiSC, we investigated whether higher levels of *Sall1* and *Nanog* can delay this conversion. We generated stable *Sall1*, *Nanog* or *Sall1+Nanog* overexpressing *Rex1*-GFP ESCs[37] via *PiggyBac* mediated transposition of the respective cDNAs. We cultured the cells in EpiSC media and used flow cytometry to quantify the *Rex1*-GFP$^+$ population as a measure of cells remaining in the ESC ground state. *Nanog* and *Sall1+Nanog* transfected cells maintained a significantly higher proportion of GFP$^+$ cells across a time course of 21 days (passaged every 2-3 days), whereas *Sall1* only cells did not (Figure 2e). The expression of naïve pluripotency and EpiSC markers analysed by RT-qPCR followed a similar pattern (Supplementary Figures 3g-i), although *Sall1* only cells delayed upregulating the differentiation markers *Fgf5* and *Otx2*. Concordantly, when plated in 2i/LIF medium, *Nanog* and *Sall1+Nanog* overexpressing cells retained the ability to form ESC colonies through most of the time course, and cells overexpressing *Sall1* did not show impaired colony formation capacity until after 6 days (Figure 2f). This shows that *Sall1* does not have the same capacity as *Nanog* to keep the cells at ESC ground state, but may indicate that *Sall1* confers a longer 'formative state'[38] during conversion.

Sall1 promotes MEF reprogramming and works synergistically with *Nanog*

We next tested whether *Sall1* can enhance somatic cell reprogramming as well. To this end, we stably transfected *Oct4*-GFP reporter mouse embryonic fibroblasts (*Oct4*-GFP-MEFs) with the Yamanaka factors under constant expression via the CAG promoter (CAG4F, Supplementary Figure 1) together with dCas9:SAM and *Sall1* and

Nanog sgRNAs either alone or in combination. Upon allowing them to reprogram in ESC media, *Sall1* sgRNA transfected MEFs produced a significantly higher number of *Oct4*-GFP$^+$ and Alkaline Phosphatase positive (AP$^+$) colonies (Figure 3a and Supplementary Figure 4a) with ESC-like morphology (Supplementary Figure 4b) compared to CAG4F alone, mirroring the results obtained from EpiSC reprogramming, including synergism between *Sall1* and *Nanog*.

In order to examine the dynamics of MEF reprogramming, we chose cDNA mediated tetracycline-inducible *Sall1* expression (TRESall1, Supplementary Figure 1 and Supplementary Table 2). Initially, we co-transfected *Oct4*-GFP-MEFs with TRESall1 and CAG4F. We cultured the transfected cells in ESC medium (M15) and induced *Sall1* expression with three different concentrations of Doxycycline (Dox, 0.1, 0.5 and 1.0 µg/ml) to find a suitable concentration to mediate reprogramming. After 18 days in culture, *Oct4*-GFP$^+$ colonies were counted and stained for AP to identify iPSC colonies. Doxycycline concentrations of 0.5 or 1.0 µg/ml resulted in a significant 2 to 3-fold increase in *Oct4*-GFP$^+$ and AP$^+$ colonies (Figure 3b and Supplementary Figure 4c) and we chose a Doxycycline concentration of 0.5 µg/ml for all subsequent experiments. To determine whether there is a time point during reprogramming beyond which *Sall1* will not improve the efficiency any further, we again co-transfected MEFs with CAG4F and TRESall1 and induced *Sall1* expression at 0, 2, 4, 6, 8, 10 and 12 days of reprogramming. As before, we counted *Oct4*-GFP$^+$ and AP$^+$ colonies after 18 days and observed that only *Sall1* expression during the first four days resulted in higher reprogramming efficiency, whereas *Sall1* induction at a later stage did not (Supplementary Figure 4d).

Nanog has been reported to promote MEF reprogramming[39] and in order to elucidate whether - similar to the effect on EpiSC reprogramming - *Sall1* can act synergistically here as well, we co-transfected MEFs with combinations of inducible expression vectors for the Yamanaka factors, *Nanog* and *Sall1* (TRE4F, TRENanog and TRESall1). We induced expression as before and counted reprogrammed colonies at day 18 according to *Oct4*-GFP$^+$ fluorescence and AP$^+$ staining. We observed that co-expression of *Sall1*+*Nanog*/4F led to a 1.5-fold increase in colony number over expression of either *Sall1*/4F or *Nanog*/4F (Figure 3a and Supplementary Figure 4a), indicating synergism between the two factors in MEF-reprogramming.

The *Sall1*-iPSCs derived from these experiments were morphologically similar to ESCs with a compact dome like structure and expressed the *Oct4*-GFP reporter. Moreover, immunofluorescent staining of these iPSCs showed protein expression of the ESC-markers SSEA-1 and Nanog (Figure 3c). When we cultured these iPSCs in differentiation medium (DMEM/10% FCS or N2B27 medium[40]), the colonies exited ground state pluripotency and differentiated into mesoderm, endoderm and ectoderm lineages as confirmed by immunofluorescent staining for expression of smooth muscle actin (SMA), alpha fetoprotein (AFP) and β-Tubulin III (Figure 3d). In addition, when we injected these iPSCs into blastocyst, live chimeras were born (Figure 3e). Both *in vitro* and *in vivo* data confirmed the pluripotency of these *Sall1*-iPSCs.

Female mESCs have two activated X chromosomes when maintained at ground state[41] and randomly inactivate one of them once they undergo differentiation. Staining with anti- H3K27me3 antibody detects this event as foci on the inactivated X chromosome[42]. We derived iPSCs from female MEFs by co-transfecting with 4F/*Sall1* as before and then differentiated them in DMEM/10% FCS for 5 days. Loss of *Oct4*

expression demonstrated successful differentiation and the presence of H3K27me3 foci indicated X chromosome silencing. In contrast iPSC cultured in 2i/LIF strongly expressed *Oct4* protein and lacked any H3K27me3 foci (Figure 3f). Together, this data demonstrates that *Sall1* can enhance 4F driven somatic cell reprogramming and that 4F/ *Sall1* reprogrammed iPSCs are naïve and pluripotent.

It was previously reported that only three Yamanaka factors (*Oct4*, *Sox2* and *Klf4*) are essential and sufficient for reprogramming, albeit at a lower efficiency than in conjunction with *c-Myc*. Moreover, the three essential factors can be replaced by other transcription factors or small molecules such as *Gata3*[43] or valproic acid[44-46]. We co-transfected MEFs with tet-inducible *Sall1* and three different expression constructs, each containing *c-Myc* and two out of three essential factors (*Klf4* and *Sox2*, CKS; *Oct4* and *Klf4*, OCK; *Oct4* and *Sox2*, OCS). After 18 days of culture in Doxycycline-containing 2i/LIF medium, we could not observe iPSC colonies in any of the combinations (Supplementary Figures 4e-g), indicating that *Sall1* cannot replace the function of either *Oct4, Sox2 or Klf4* in MEF reprogramming.

RNA-seq identifies potential mechanisms of cellular reprogramming mediated by *Sall1* and *Nanog*

In order to find downstream targets of *Sall1* and *Nanog*, we transfected *Oct4*-GFP EpiSCs with cDNA for either gene alone or in combination as described above and performed RNA-seq after 24 h of overexpression. Our analysis identified 372 genes that were differentially expressed specific to *Sall1* transfected cells compared to empty vector, and 307 genes specific to *Nanog* transfected cells. We observed a large overlap of 568 genes (45%) between both sets (Figure 4a, Supplementary Table 8)

and GOTERM analysis revealed that they are involved in a number of developmental processes and signalling cascades (Figure 4b, Supplementary Table 8).

Amongst those commonly regulated genes were *Myc*, *Mycn*, *Tet3*, *Tex10*, *Jarid2*, *Fgfr1*, *Mbd2*, *Lifr* and *Smad7* (Figure 4a) which have previously been implicated in the promotion of cellular reprogramming or inhibition of ESC differentiation [8,47-56]. Upregulation of the Lif receptor (*Lifr*) and downregulation of the Fibroblast Growth Factor Receptor (*Fgfr1*) is expected in EpiSC reprogramming and validates our RNA-seq and RT-qPCR data (Figure 4c). Furthermore, we found 215 genes which were only regulated when *Sall1* and *Nanog* were overexpressed together (Supplementary Figure 4h and Supplementary Table 8), such as *Dnmt3c* and *Hdac9*, reported to be involved in the epigenetic regulation of male germ cell maintenance[57] and muscle differentiation[58,59], respectively; as well as a modest upregulation of *Utf1*, a transcription factor known to synergize with the Yamanaka factors in reprogramming[60].

We had already independently identified the genes *Klf5* and *Fam189a2* in our GoF screen (Figure 1e, Supplementary Table 5) and RNA-seq showed them to be potentially regulated by *Sall1* and *Nanog*, respectively. We validated the RNA-seq results for these genes with RT-qPCR in EpiSCs 24 h after cDNA transfections (*Sall1*/*Nanog*/*Sall1*+*Nanog*) and found a good correlation between both methods (Figure 4c). While *Klf5* narrowly failed the stringent P value cut-off for the RNA-seq results in *Nanog* overexpressing cells, RT-qPCR indicated that *Klf5* may be regulated by *Nanog* as well, albeit to a lesser extent than by *Sall1*. *Fam189a2* on the other hand seemed to be regulated significantly stronger by *Nanog* than *Sall1*. When we co-expressed *Sall1* and *Nanog*, we did not observe a significant increase in expression for these downstream targets compared to *Sall1* or *Nanog* alone (Figure 4c); we did

however find synergistic elevation of expression for the genes *Myc*, *Mycn*[61] and *Arid2*, all of which have been shown to play a role in reprogramming and chromatin remodelling [62,63].

We used CRISPRa to induce expression of *Klf5*, *Fam189a2, Tex10* and *Tet3* in *Oct4*-GFP EpiSC and found that all were able to augment reprogramming into iPSCs (Figure 4d). Reprogramming by *Fam189a2* occurred in 10 days, while *Klf5*, *Tex10* and *Tet3* required between 14 and 20 days. In all cases, the number of reprogrammed colonies was significantly lower than compared to *Sall1* or *Nanog* (Figure 4g), which may indicate that multiple downstream targets of *Sall1* and *Nanog* participate in reprogramming.

We tested the regulatory relationship between *Sall1+Nanog* and *Klf5+Fam189a2* by transfecting EpiSCs with CRISPRa for *Klf5 and Fam189a2.* CRISPRa significantly increased *Klf5* and *Fam189a2* transcription, respectively, but not *Sall1* and *Nanog* expression, indicating *Klf5* and *Fam189a2* are respective downstream targets of *Sall1* and *Nanog* (Figure 4e). We then investigated some of the key genes which were differentially expressed in RNA-seq data. After culturing transfected EpiSCs in EpiSC medium, the cells were collected for RT-qPCR analysis. We observed that both *Sall1* and *Klf5* up-regulated *Smad7* (negative regulator of TGF-β signalling and activator of Stat3[64]), *Gp130* and *Lifr*, suggesting the repression of TGF-β signalling and activation of Jak/Stat3 signalling. *Nanog* and *Fam189a2* on the other hand down-regulated *Fgfr1*, *Tgfr1* and *Mbd2* and up-regulated *Esrrb* expression, indicating the repression of FGF and TGF-β signalling, inhibition of epigenetic repression and promotion of self-renewal and pluripotency (Figure 4f). Functionally, co-activation of both *Klf5* and *Fam189a2* generated significantly more *Oct4*-GFP^{+} colonies than *Klf5* or *Fam189a2*

alone. As expected, co-activation of either *Sall1* and its downstream target *Klf5* or *Nanog* and its downstream target *Fam189a2* showed no synergistic effects in *Oct4*-GFP$^+$ colony production, whereas co-activation of either *Sall1* and *Fam189a2* or *Nanog* and *Klf5* resulted in increased numbers of *Oct4*-GFP$^+$ colonies compared to *Klf5* and *Fam189a2* co-activation (but lower than *Sall1* and *Nanog* co-activation). Collectively these results suggest that *Klf5* and *Fam189a2* are situated downstream of *Sall1* and *Nanog,* respectively, and can synergise similar to *Sall1* and *Nanog* albeit with a lower efficiency than their upstream regulators (Figure 4g).

Discussion

With the advent of CRISPR/Cas9, genomic *in vitro* screening has become sufficiently accessible and affordable to enable experiments at a genome wide scale and is typically used in the context of loss of function by exploiting the ability of wildtype Cas9 to induce double strand breaks in the target DNA. Catalytically dead Cas9 fused to transcriptional activator proteins on the other hand opens up the possibility of driving gene expression. To date few CRISPR activation screens have been performed[65-67] using previously established gain-of-function libraries[21,68,69]. However, none of them targeted stem cell reprogramming and while some recent publications have used CRISPRa in this field of research, they have been restricted to a few genes to demonstrate proof-of-concept[33,70,71].

Our present study shows that a genome-scale CRISPRa screen, in conjunction with an experimental model such as epiblast stem cells in which a single overexpressed gene may mediate reprogramming to pluripotency, is a powerful tool for gene discovery. We identified 142 candidate reprogramming factors and found amongst them known pluripotency factors such as *Nanog* and *Oct4,* principally validating our

approach. *Nanog* overexpression can reprogram EpiSCs to iPSCs and deletion of *Nanog* almost completely impairs reprogramming in somatic cells (unless ascorbic acid is supplied in the ESC medium)[72-74]. *Oct4* is one of the Yamanaka factors in somatic reprogramming[3] and the level of *Oct4* seems critical for maintenance of ESCs and differentiation[75], as it has been reported that an artificially reduced level of *Oct4* maintains ESCs in a robust pluripotent state, whereas wild-type levels of *Oct4* enable differentiation[76,77]. Curiously, while in our current work CRISPRa induced *Oct4* readily and robustly reprogrammed EpiSCs into iPSCs, overexpression via cDNA in the past has failed to do so[11], a result we were able to reproduce here. We show that gene dosage is one critical aspect to explain this behaviour and it is conceivable that too high a level of *Oct4*, especially when supplied exogenously, is detrimental to pluripotency[78,79]. This has important implications for screening approaches similar to ours: while traditional systems like cDNA libraries only produce a singular expression level per gene, we and others[33] show that CRISPRa in conjunction with sgRNAs tiled throughout the regulatory regions of genes can provide a multitude of expression levels and thus a higher probability of matching the physiological gene dosage. This gives CRISPRa mediated approaches much more flexibility, to the point where smaller, targeted screens with many sgRNAs per gene may be used to probe the regulatory regions of genes of interest.

Along similar lines, the choice of CRISPRa system may well influence the outcome of a screen: We tested several systems and chose SAM for our screen as it achieved the highest expression levels and it is conceivable that repeating our screen with a different CRISPRa system at lower activation efficiencies could produce a non-redundant list of candidate genes.

The observation that sgRNA activities vary depending on their position with respect to the TSS certainly serves to explain why most of our candidate genes were only identified by a single sgRNA in our screen. However, we also acknowledge that reprogramming is inherently a very inefficient process and thus, a very large initial cell number may be required to deeply cover a genome-wide library and give a sufficient number of cells a chance to gain pluripotency. While we performed our GoF screen with 10 x 10^6 library-transduced cells (library coverage 114x), a deeper coverage or a more focused library promises to uncover reprogramming candidates the present screen might have missed.

Our screen identified *Sall1* as a potent EpiSC reprogramming factor. *Sall1* is a member of the Spalt-like gene family which in mice comprises 4 genes – *Sall1*, *2*, *3* and *4*. *Sall1* and *Sall4* have been implicated in the establishment of pluripotency[18] in studies showing that the action of demethylase *Utx* on *Sall1* and *Sall4* is required to enable MEF reprogramming[17]. Furthermore, it has been demonstrated that *Sall4* activates *Oct4* expression while *Sall1* is a direct binding partner of *Nanog*[34,80] and has been suggested to be required in *Nanog*-mediated open heterochromatin maintenance within ESCs and EpiSCs[35]. So far, it is unclear whether *Sall1* plays an active role in EpiSC reprogramming. In our work, we demonstrate that by either activation of endogenous *Sall1* expression via CRISPRa or by overexpression of transgenic *Sall1*, EpiSCs can be reprogrammed to iPSCs. We show by combined overexpression that *Sall1* and *Nanog* synergistically augment reprogramming of EpiSCs and MEFs. However, *Sall1* cannot replace *Oct4*, *Sox2* or *Klf4* in MEF reprogramming, suggesting that unlike these core pluripotency factors, *Sall1* alone cannot initiate the reprogramming machinery in more differentiated cells. It can be postulated that one of

its roles may be in facilitating epigenetic modification and nucleosome remodelling e.g. through interaction with *Nanog* and the deacetylase complex (NurD)[20,81].

The ability of *Sall1* to reprogram EpiSCs, however, is not sufficient to keep ESCs in the naive pluripotent state. Unlike *Nanog*, overexpressing *Sall1* in ESCs only marginally delayed loss of pluripotency in differentiation experiments. However, it slowed expression of EpiSC markers *Fgf5* and *Otx2* and preserved the ability to generate ESC-like colonies. In embryoid body differentiation of ESCs, overexpression of *Sall1* inhibited *Otx2* expression and a formative pluripotent phase between naïve and primed states was postulated when cells lost naïve pluripotency markers and gained post-implantation markers such as *Otx2* and *Oct6* among others[34,38]. Considering that even after 21 days in differentiation medium, some *Sall1* overexpressing cells still formed ESC like colonies in 2i/LIF, these cells may be stalled in a formative state.

We used RNA-seq to identify downstream targets of *Sall1* and *Nanog* in EpiSCs and found novel genes as well as factors previously implicated in pluripotency or stem cell maintenance. *Esrrb*, a downstream target of *Nanog*, plays an important role in maintaining ESCs pluripotency and reprogramming by interacting with the core pluripotency network via *Sox2*[82,83]. *Tex10* was recently reported to be a pluripotency factor and partner of *Sox2,* capable of promoting MEF reprogramming[84], a role we further extended to EpiSC reprogramming. *Tet3* is a member of the ten-eleven translocation (Tet) protein family, which regulate DNA methylation. *Tet1* and *Tet2* have already been implicated in somatic reprogramming and *Tet2* has been reported to promote EpiSCs to a naïve state[51,53]. Here, we show that *Tet3* can mediate EpiSC reprogramming as well.

The Krueppel-like factor family proteins *Klf2*, 4, 5 are also pluripotency factors and both *Klf2* and *Klf4* have been shown to facilitate reprogramming. The potential of *Klf5* however is unclear as it has been reported to be incapable of reprogramming EpiSCs in a study by Hall et al.[85], while Jeon et al. and recently Azami et al., both were able to derive iPSCs from EpiSCs by cDNA mediated *Klf5* overexpression[86,87].

Our RNA-seq analyses demonstrated synergistic pathways through which *Sall1* may function together with *Nanog* in increasing reprogramming efficiency. We identified *Klf5* in our GoF screen and confirmed its ability to reprogram EpiSCs via CRISPRa transcriptional activation. RT-qPCR validation also suggested that *Klf5* is one of the downstream targets of *Sall1*. Similar to our observations with *Oct4*, this may reflect a Goldilocks effect of gene expression levels and highlights the utility of different overexpression approaches to discover new pluripotency factors. LIF dependent activation of Jak/Stat3 and its role in ESC self-renewal and reprogramming has been widely studied to date[64,88]. cDNA overexpression of *Klf5* has been shown to compensate for the absence of LIF in maintaining pluripotency of ESCs and that it can also reprogram EpiSC via LIF-independent pathways[86,89]. Besides *Klf5*, our data also indicates that *Sall1* positively and negatively regulates the Jak/Stat3 and TGFβ pathways by upregulating the *Gp130, Lif* receptor and *Smad7* respectively, together providing a novel role of *Sall1* in EpiSC reprogramming.

Fam189a2 was identified as a new target of *Nanog* in EpiSC reprogramming and our data showed that both *Nanog* and *Fam189a2* downregulate the *Tgfbr1* and upregulate *Esrrb* expression. We postulate that the observed synergy between *Sall1* and *Nanog* as well as their downstream effectors *Klf5* and *Fam189a2* is partially due to the

combined activation of Jak/Stat3, suppression of TGFβ signalling and upregulation of pluripotent genes such as *Esrrb*.

In conclusion, using a genome-scale CRISPR activation screen in the well-established EpiSCs reprogramming model, we identify known and previously unknown genes that can mediate cellular reprogramming in EpiSC. We demonstrate that the transcription factor *Sall1* can effectively reprogram EpiSC and MEFs and provide new insights into the role of *Sall1* in promoting and maintaining pluripotency. Other novel reprogramming candidates such as transcription factors *Atf1 and Bhlha15*, kinases *Idnk and Has1*, several olfactory receptor genes (*Olfr*) and others with less known functions like *Umodl1 and Prr3* deserve further in-depth investigation. Our studies demonstrate the strengths of CRISPRa screens in the identification of novel factors in molecular reprogramming and in illuminating biological pathways.

Methods

Ethics Statement

All animal experimental procedures were carried out in accordance with Home Office UK regulations and the Animals (Scientific Procedures) Act 1986 (licence No. 70/8387 and 80/2552). All experimental protocols were approved by the Animal Welfare and Ethical Review Body (AWERB) of Wellcome Genome Campus and the University of Cambridge CRUK Cambridge Institute. At the end of the study, mice were euthanized in accordance with stated Home Office UK regulations.

Constructs

Guide RNAs (sgRNAs) were expressed under the U6 promoter in expression constructs (pKLV-PB-U6-gRNA-PGK-Puro-T2A-TagBFP, Supplementary Figure 1) harbouring *PiggyBac inverted terminal repeats to enable transposase-mediated genomic integration* (*PB* transposon) and HIV-1 long terminal repeats to allow lentiviral genomic integration (Supplementary Figure 1). The sgRNA scaffold used in conjunction with dCas9:SAM was adapted to contain two MS2-binding loops as required by the SAM CRISPRa system[69]. The constructs also included a puromycin antibiotic resistance and a TagBFP marker. The four dCas9 variants were cloned into PB transposons and included the mCherry fluorescent marker and Blasticidine antibiotic resistance. The dCas9 cDNAs in these constructs were driven by the Ef1-α promoter and multiple consecutive cDNAs were linked by the T2A self-separating peptide sequence.

cDNAs of *Oct4*, *c-Myc*, *Klf4*, *Sox2*, *Sall1* and *Nanog* were cloned into *PiggyBac* transposons under the control of the CAG promoter (PBCAG) or the Tet response element (PBTRE). Combinations of cDNAs were linked by the T2A self-separating

peptide sequence. PBEF1α-TET3G encoding the Tet-On-3G transactivator protein was co-transfected with PBTRE-cDNA to enable Doxycycline induction of the PBTRE constructs.

When stable integration by transposition of the transgene was required, a plasmid encoding *PiggyBac* transposase (*HyPBase*[90]) was co-transfected. Schematic representations of all the constructs used in the study are shown in Supplementary Figure 1.

Cell Culture

Oct4-GFP and *Nanog*-GFP Epiblast stem cells (EpiSCs) were generated as described previously[28]. Briefly, EpiSC were derived from post-implanted pregnant transgenic mice at E5.75. Cells from embryos were cultured on human fibronectin (Millipore) coated plates in complete EpiSC media based on N2B27 which comprised 50% Neurobasal media, 50% DMEM-F12 media, 0.1mM β-ME, penicillin (100 U/ml), streptomycin (100 µg/ml), L-glutamine (2 mM), 1X N2 and 1X B27 supplement (Invitrogen) and was supplemented with 20 ng/ml Activin-A (R & D Systems) and 12 ng/ml FGF2 (Peprotech). When confluent, the media was aspirated and the EpiSC were dissociated with Accutase (Millipore) for 3 min. Dissociated cells were spun down at 300g for 5 min and plated either at 1:6 or at 1:8 split ratio on human fibronectin coated plates in complete EpiSC media for maintenance.

For screening and reprogramming, EpiSCs were cultured in 2i/LIF medium comprising N2B27 media (as above) supplemented with 100 U/ml leukemia inhibitory factor (LIF, Millipore), 1 µM PD0325901 (Tocris) and 3 µM CHIR99021 (Tocris).

Rex1-GFP embryonic stem cells (ESCs) cells were generated as described previously[37]. ESCs were cultured on 0.1% gelatin coated plates. ES cells were

regularly maintained in ESC medium (M15) comprising knock-out DMEM containing 15% FBS, 0.1mM β-ME, penicillin (100 U/ml), streptomycin (100 µg/ml), L-glutamine (2 mM), and 100 U/ml LIF. Confluent ES cells were dissociated with 0.025% Trypsin-EDTA for 5 min. Detached cells were collected and spun down at 300g for 5 min and were plated at a split ratio of either 1:8 or at 1:10 on gelatin coated plates.

Mouse embryonic fibroblasts (MEFs) were cultured in M10 medium comprising knock-out DMEM containing 10% FBS, 0.1mM β-ME, penicillin (100 U/ml), streptomycin (100 µg/ml) and L-glutamine (2 mM). For reprogramming, after transfection, cells were cultured in ESC medium until the end of experiment or until the colonies were picked for iPSC line derivation.

GoF gRNA library design

The Gain of Function (GoF) library targeted the region of up to 250 base pairs upstream of the transcription start site (TSS) of each protein-coding gene. Up to 4 guides of 19 bp length were selected per gene. Protein coding genes and TSSs were obtained from the mouse reference assembly GRCm38 in the ENSEMBL database (version 78, http://www.ensembl.org,)[91] and TSSs were checked against CAGE data in the FANTOM data base (April 2015, http://fantom.gsc.riken.jp/)[92].

Potential guide sequences were identified on the reference assembly as the 19 bp sequence at the 5' end of each PAM motif (NGG), i.e. 2 to 20 bp upstream of each guanosine di-nucleotide in the reference. Guide sequences consisted of no more than 13 guanosine or cytosine bases (GC content < 70%). Guides in the 250 bp region on both DNA strands upstream of the TSS were then compared to all other potential guide sequences across the genome. Guide sequences with off-target sites exhibiting fewer than 3 mismatches over their 19 bp length were omitted from the design.

The remaining guides were sorted by a simple ad-hoc quality score intended to reflect a likely increased tolerance of mismatches distal of the PAM motif. For each potential off-target site a score was calculated that linearly increased with the number of mismatches and decreased with their distance to the PAM motif.

A selection algorithm was designed to spread high quality guides across the target region. To this end, the region upstream of the TSS was divided in quarters of roughly equal length. Starting with the quarter closest to the TSS the algorithm looped over quarters picking the best guide, by quality score, in each if available and adding it to the library until no more guide fitting the constraints could be found, or the target number of 5 guides per genes was reached. A constraint for the GC content of less than 55% was applied in the first loop and then relaxed to 70%.

GoF reprogramming screen

The GoF sgRNA library was synthesized by Custom Array, and the oligo pools were cloned into the lentiviral sgRNA expression plasmid via Gibson assembly as described by Shalem *et al*.[93], with minor modifications.

Oct4-GFP EpiSC cells were transfected with 200 ng of plasmid encoding *PiggyBac* transposase together with 1 µg dCas9:SAM to facilitate stable integration. Transfections were performed with Lipofectamine LTX (Invitrogen), according to manufacturer's instructions. Transfected cells were then selected by 10 µg/ml Blasticidine (Gibco) for 10 days. Post-selection, dCas9:SAM expressing *Oct4*-GFP EpiSC were expanded to 100 x 10^6 cells for lentiviral transduction.

Library transduction was carried out at a MOI (multiplicity of infection) of 0.3. After two days, 10 x 10^6 BFP^+ *Oct4*-GFP EpiSCs were sorted by flow cytometry. The sorted cells were seeded on fibronectin coated plates and allowed to recover in complete

EpiSC medium for 24 h. The medium was then changed to 2i/LIF in order to allow selection for reprogrammed cells. After 14-16 days in 2i/LIF, the individual reprogrammed colonies, verified by *Oct4*-GFP fluorescence, were picked and transferred to 96 well plates. Colonies were passaged twice in 2i/LIF before sequencing. Genomic DNA was extracted from each colony and PCR amplification across the stably integrated sgRNA was performed using primers described previously[94]. PCR amplicon libraries were pooled and Next Generation Sequencing was used to identify the distribution of sgRNA sequences.

CRISPRa transfections

Oct4-GFP EpiSC and *Nanog*-GFP EpiSC cells were grown to 70% confluence on fibronectin coated 6-well plates. The cells were dissociated with Accutase and re-suspended in EpiSC media for reverse transfections (approx. 1 X 10^6 cells in 250 µl per transfection).

Cells were transfected with 500 ng *PiggyBac* transposase together with 500 ng dCas9:SAM and 500 ng of sgRNA expression construct for one or more sgRNAs. Transfections were performed using Lipofectamine LTX (Invitrogen), according to manufacturer's instructions. Cells were cultured in EpiSC medium for 24 h. Stably-transfected cell lines were generated by selection with Blasticidine (10 µg/ml) for at least 10 days post-transfection. Integration of constructs was confirmed by PCR genotyping (Supplementary Table 1).

cDNA transfections

Oct4-GFP EpiSCs, *Nanog*-GFP EpiSCs and *Rex1*-GFP ESCs were transfected with 1 µg PBCAG expressing either *Sall1*, *Nanog* or a combination of both vectors via Lipofectamine LTX. Cells were co-transfected with plasmids encoding *PiggyBac*

transposase (500 ng) and mCherry Blasticidine (100 ng) selection markers to enable selection of transfected cells. Presence of PBCAG in the selected cells was confirmed with PCR (Supplementary Table 1).

EpiSC Reprogramming

Stable lines of *Oct4*-GFP and *Nanog*-GFP EpiSC generated either from CRISPRa or cDNA transfections were plated in triplicate on fibronectin coated 6-well plates in EpiSC medium. Medium was changed to 2i/LIF when cells reached 80% confluence and thereafter replaced every 2 days to select for reprogramming for up to 20 and 24 days for *Oct4*-GFP transfected EpiSCs and *Nanog*-GFP transfected EpiSCs, respectively. GFP^+ ESC-like colonies were counted and transferred onto gelatin coated plates for expansion. Expanded colonies were then phenotyped for gene expression by RT-qPCR (Supplementary Table 3).

To derive iPSCs for blastocyst injection, *Oct4*-GFP EpiSC were transfected with 1 µg PBTRESall1, 1 µg PBTET3G and 2 µg transposase using Lipofectamine LTX. Transgene expression was induced by supplementing the medium with 0.5 µg/ml Doxycycline after switching the cells to 2i/LIF.

Flow cytometry analysis of EpiSC reprogramming

Stably transfected CRISPRa *Oct4*-GFP and CRISPRa *Nanog*-GFP EpiSC were plated in triplicates on fibronectin coated 24-well plates at a density of 50,000 cells per well, in EpiSC media. Upon reaching confluence, the media was changed to 2i/LIF to select for reprogramming. Cells were harvested at regular time intervals and were analysed for GFP fluorescence in a flow cytometer (CytoFLEX, Beckman Coulter Life Sciences, Indianapolis US).

MEF Reprogramming

For sgRNA mediated reprogramming of MEFs, 1 x 10^6 *Oct4*-GFP MEFs were electroporated with 0.5 µg PBCAG4F, 1 µg gRNA Sall1 / 1 µg gRNA Nanog / 0.5 µg gRNA Sall1+0.5 µg gRNA Nanog, 1 µg dCas9:SAM and 0.5 µg *PiggyBac* transposase using the Amaxa Nucleofector (Amaxa, Lonza). The transfected cells were plated onto gelatin-coated 10 cm dishes in M15. After 24 h, the medium was replaced and was renewed every two days until day 18 when *Oct4*-GFP^+ and AP^+ cells were counted.

For inducible cDNA mediated reprogramming, 1 x 10^6 *Oct4*-GFP MEFs were electroporated with 1 µg PBTRE4F, 0.5 µg PBTRESall1, 2 µg PBEF-1αTet3G and 2 µg *PiggyBac* transposase using the Amaxa Nucleofector (Amaxa, Lonza). The transfected cells were plated in gelatin-coated 10 cm dishes in ESC medium. After 24 h, the medium was replaced and three different concentrations of Doxycycline (0.1, 0.5 and 1.0 µg/ml) were tested for induction of cDNA expression. At day 12, Doxycycline was withdrawn and the cells were cultured for 6 more days, after which *Oct4*-GFP^+ and alkaline phosphatase stained colonies were counted and single colonies picked for RT-qPCR.

To produce iPSCs for *in vitro* and *in vivo* assays, C57B/6J MEFs were transfected with the same amount of PBTRE4F, PBTRESall1, PBEF1α-TET3G and *PiggyBac* transposase as described above and induced with 0.5 µg/ml Doxycycline in ESC media.

All the combinations of cDNA transfections for MEF reprogramming are listed in Supplementary Table 2.

Western Blotting

Oct4-GFP EpiSCs transfected with CRISPRa *Oct4* and cDNA *Oct4* (both CAG and TRE) together with experimental controls and untransfected cells (EpiSCs and ESCs) were collected for Western blotting after 3 days 2i/LIF (Dox induction for 3 days). Protein amounts were determined using a Bradford assay and 30 µg of lysates were subjected to electrophoresis on 4–15% Mini-PROTEAN® TGX™ Precast Protein Gels (Biorad). Proteins were transferred onto PVDF membranes (Millipore) overnight at 30 V for 8 hours at 4°C. Transferred proteins were then immunoblotted for Oct4 (c-10, Santa Cruz, #sc-5279, dilution 1:800) and Gapdh (Sigma, #G8795, dilution 1:4000). All antibodies used are listed in Supplementary Table 4.

Conversion of ESCs to EpiSCs

Stable lines of cDNA transfected *Rex1*-GFP ESCs were cultured in 2i/LIF medium for 3 days prior to the conversion. Upon reaching 50% confluence, cells were dissociated and seeded on gelatin-coated plates in EpiSC medium for conversion for at least 3 passages. Cells were seeded on fibronectin-coated plates after the 2nd passage to promote differentiation. In addition, 600 cells were plated back on 0.1% gelatin coated plates in 2i/LIF medium for about 7 days to promote formation of iPS colonies. The colonies were assessed by AP staining. Lastly, at every passage, cells were assessed for *Rex1*-GFP fluorescence using flow cytometry and total RNA was extracted for RT-qPCR.

Quantitative RT-PCR (RT-qPCR)

Total RNA was extracted using the RNeasy Mini Kit (Qiagen) according to manufacturer's instructions. First strand cDNA was synthesized using qScript cDNA Supermix (Quantabio) according to manufacturer's protocol. All qPCR studies were

performed using Taqman Gene Expression Assays either in the 7900HT Fast Real-Time PCR system (Applied Biosystems) or the StepOnePlus™ Real-Time PCR System (Applied Biosystems). Samples were run in triplicate for both gene of interest and house-keeping genes. Expression levels were normalized to *Gapdh*. TaqMan probes used are listed in Supplementary Table 3.

RNA-sequencing

Oct4-GFP EpiSCs were transfected via Lipofectamine LTX with 1 µg PBCAGSall1, PBCAGNanog, PBCAGSall1+PBCAGNanog or empty vector (PBCAG:Empty). Cells were co-transfected with CAGmCherry (100 ng) to enable fluorescent marker selection. All transfections were performed in triplicates. 24 h later, cells were FACS sorted for $GFP^{+}mCherry^{+}$ and their RNA was extracted using the Rneasy Mini Kit (Qiagen) according to manufacturer's instructions. Illumina Truseq NGS libraries were prepared and sequenced on a Illumina HiSeq instrument.

Sequencing results were analysed by aligning reads to the mouse genome build GRCm38.p5 using the STAR aligner (Ver. 2.5.3a) and read counts were prepared with the TOPHAT package. Differential expression of genes was analysed with the DESeq2 package for the R statistical computing framework. We used a cut-off padj value of <0.001 to determine genes that were differentially regulated between experimental and control samples.

GO Term analysis

We used the GoToolBox platform (http://genome.crg.es/GOToolBox/) to perform enrichment analysis on the genes identified in the CRISPRa screen and via RNAseq as described previously[95]. In brief, a Data-Set was created using the Mouse Genome Informatics version (MGI) and then used for pathway enrichment analysis. Fold

changes for enriched gene sets were computed by dividing the frequency in our set by that of the reference. Representative pathways with a fold change of 2 or more and with a p-value of 0.05 or lower were graphically represented in the results.

Immunofluorescence

C57BL/6J MEF reprogrammed iPSCs were plated at 2 x 10^3 onto 24 well plates in 2i/LIF. After 24 hours, cells were washed with PBS and fixed in 4% paraformaldehyde (PFA), blocked and permeabilised with 1% bovine serum albumin (BSA) and 3% serum in PBS with 0.1%Triton X100. Samples were incubated with mouse anti-SSEA-1 (BD Biosciences) or rabbit anti-*Nanog* (Abcam) antibodies at 4℃ overnight, then rinsed and incubated with Alexa488-conjugated goat anti-Mouse IgM and Alexa594-conjugated goat anti-Rabbit IgG (Invitrogen), and counterstained with DAPI.

To examine the X chromosome status in female iPSCs, cells were plated at 5 x 10^3 on gelatin coated slides in 2i/LIF or in M10 medium for 5 days, then cells were fixed in 4% PFA, and immunofluorescence was performed as mentioned above. Slides were incubated with rabbit anti-H3K27me3 (Millipore) and mouse anti-*Oct4* antibodies (Santa Cruz), then Alexa594-conjugated goat anti-Rabbit IgG and Alexa488 goat anti-Mouse IgG and counterstained with DAPI. All antibodies used are listed in Supplementary Table 4.

***In Vitro* differentiation**

C57BL/6J MEF reprogrammed iPSCs were plated at 5 x 10^5 in a petri-dish in M10 medium for 4 days, then dissociated with 0.05% Trypsin/EDTA and plated at 1 x 10^5 in M10 on gelatinized plates for another 4 days. The cells were then fixed in 4% PFA and examined for mesoderm and endoderm markers using immunofluorescent staining with antibodies against smooth muscle antigen (SMA) (R&D Systems) and

alpha fetoprotein (AFP) (R&D Systems). For neuronal differentiation, cells were plated at 1.5 x 10^5 in N2B27 medium on gelatinized plates. The medium was changed every other day and at day 8, the cells were fixed and stained with anti-beta tubulin III (Tuj1) antibody (R&D Systems).

Chimeras

Chimeras were produced by a standard microinjection protocol. Chimerism was estimated based on coat colour since iPSCs derived from EpiSCs and MEFs are of MF1 and C57BL/6J genetic background (agouti and black furs) whereas the host blastocysts were from albino C57BL/6.

Statistical analyses

All statistical analyses were performed in GraphPad Prism (Version 6.01). Data are presented as means ± standard deviations. Statistical significance was determined either using a Student's unpaired *t*-test with two-tailed distribution or Two-way ANOVA. Students' t-test was used for comparison across two groups while Two-way ANOVA with multiple comparisons were performed on samples undergoing a time course experiment. *p*-values less than 0.05 were considered significant.

References

1 Martello, G. & Smith, A. The nature of embryonic stem cells. *Annu Rev Cell Dev Biol* **30:647-75.**, 10.1146/annurev-cellbio-100913-013116. (2014).

2 De Los Angeles, A. *et al.* Hallmarks of pluripotency. *Nature.* **525**, 469-478. doi: 410.1038/nature15515. (2015).

3 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell.* **126**, 663-676. Epub 2006 Aug 2010. (2006).

4 Warren, L. *et al.* Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell.* **7**, 618-630. doi: 610.1016/j.stem.2010.1008.1012. Epub 2010 Sep 1030. (2010).

5 Sandmaier, S. E. & Telugu, B. P. MicroRNA-Mediated Reprogramming of Somatic Cells into Induced Pluripotent Stem Cells. *Methods Mol Biol* **1330:29-36.**, 10.1007/1978-1001-4939-2848-1004_1003. (2015).

6 Hou, P. *et al.* Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science.* **341**, 651-654. doi: 610.1126/science.1239278. Epub 1232013 Jul 1239218. (2013).

7 Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell.* **4**, 487-492. doi: 410.1016/j.stem.2009.1005.1015. (2009).

8 Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature.* **453**, 519-523. doi: 510.1038/nature06968. (2008).

9 Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature.* **448**, 196-199. Epub 2007 Jun 2027. (2007).

10 Brons, I. G. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature.* **448**, 191-195. Epub 2007 Jun 2027. (2007).

11 Guo, G. & Smith, A. A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. *Development.* **137**, 3185-3192. doi: 3110.1242/dev.052753. (2010).

12 Gaj, T., Gersbach, C. A. & Barbas, C. F., 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**, 397-405. doi: 310.1016/j.tibtech.2013.1004.1004. Epub 2013 May 1019. (2013).

13 Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science.* **346**, 1258096. doi: 1258010.1251126/science.1258096. (2014).

14 Pritsker, M., Ford, N. R., Jenq, H. T. & Lemischka, I. R. Genomewide gain-of-function genetic screen identifies functionally active genes in mouse embryonic stem cells. *Proc Natl Acad Sci U S A.* **103**, 6946-6951. Epub 2006 Apr 6918. (2006).

15 Gayle, S., Pan, Y., Landrette, S. & Xu, T. piggyBac insertional mutagenesis screen identifies a role for nuclear RHOA in human ES cell differentiation. *Stem Cell Reports.* **4**, 926-938. doi: 910.1016/j.stemcr.2015.1003.1001. Epub 2015 Apr 1019. (2015).

16 Hu, G. *et al.* A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev.* **23**, 837-848. doi: 810.1101/gad.1769609. (2009).

17 Mansour, A. A. *et al.* The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature* **488**, 409-413, doi:10.1038/nature11272 (2012).

18 Gaspar-Maia, A. *et al.* MacroH2A histone variants act as a barrier upon reprogramming towards pluripotency. *Nat Commun* **4**, 1565, doi:10.1038/ncomms2582 (2013).

19 Sweetman, D. & Munsterberg, A. The vertebrate spalt genes in development and disease. *Dev Biol.* **293**, 285-293. Epub 2006 Mar 2020. (2006).

20 Basta, J. M., Robbins, L., Denner, D. R., Kolar, G. R. & Rauchman, M. A Sall1-NuRD interaction regulates multipotent nephron progenitors and is required for loop of Henle formation. *Development.* **144**, 3080-3094. doi: 3010.1242/dev.148692. Epub 142017 Jul 148631. (2017).

21 Joung, J. *et al.* Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature.* **548**, 343-346. doi: 310.1038/nature23451. Epub 22017 Aug 23459. (2017).
22 Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet.* **16**, 299-311. doi: 210.1038/nrg3899. Epub 2015 Apr 1039. (2015).
23 Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature.* **517**, 583-588. doi: 510.1038/nature14136. Epub 12014 Dec 14110. (2015).
24 Chavez, A. *et al.* Comparison of Cas9 activators in multiple species. *Nat Methods.* **13**, 563-567. doi: 510.1038/nmeth.3871. Epub 2016 May 1023. (2016).
25 Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell.* **159**, 635-646. doi: 610.1016/j.cell.2014.1009.1039. Epub 2014 Oct 1019. (2014).
26 Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol.* **33**, 139-142. doi: 110.1038/nbt.3149. (2015).
27 Metzakopian, E. *et al.* Enhancing the genome editing toolbox: genome wide CRISPR arrayed libraries. *Sci Rep* **7**, 2244, doi:10.1038/s41598-017-01766-5 (2017).
28 Yang, J. *et al.* Stat3 activation is limiting for reprogramming to ground state pluripotency. *Cell Stem Cell.* **7**, 319-328. doi: 310.1016/j.stem.2010.1006.1022. (2010).
29 Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643-655 (2003).
30 Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-642 (2003).
31 Qiu, D. *et al.* Klf2 and Tfcp2l1, Two Wnt/beta-Catenin Targets, Act Synergistically to Induce and Maintain Naive Pluripotency. *Stem Cell Reports* **5**, 314-322, doi:10.1016/j.stemcr.2015.07.014 (2015).
32 Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24**, 372-376, doi:10.1038/74199 (2000).
33 Liu, P., Chen, M., Liu, Y., Qi, L. S. & Ding, S. CRISPR-Based Chromatin Remodeling of the Endogenous Oct4 or Sox2 Locus Enables Reprogramming to Pluripotency. *Cell Stem Cell* **22**, 252-261 e254, doi:10.1016/j.stem.2017.12.001 (2018).
34 Karantzali, E. *et al.* Sall1 regulates embryonic stem cell differentiation in association with nanog. *J Biol Chem.* **286**, 1037-1045. doi: 1010.1074/jbc.M1110.170050. Epub 172010 Nov 170059. (2011).
35 Novo, C. L. *et al.* The pluripotency factor Nanog regulates pericentromeric heterochromatin organization in mouse embryonic stem cells. *Genes Dev.* **30**, 1101-1115. doi: 1110.1101/gad.275685.275115. Epub 272016 Apr 275628. (2016).
36 Guo, G. *et al.* Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development.* **136**, 1063-1069. doi: 1010.1242/dev.030957. Epub 032009 Feb 030918. (2009).
37 Wang, W. *et al.* Rapid and efficient reprogramming of somatic cells to induced pluripotent stem cells by retinoic acid receptor gamma and liver receptor homolog 1. *Proc Natl Acad Sci U S A.* **108**, 18283-18288. doi: 18210.11073/pnas.1100893108. Epub 1100892011 Oct 1100893111. (2011).
38 Smith, A. Formative pluripotency: the executive phase in a developmental continuum. *Development.* **144**, 365-373. doi: 310.1242/dev.142679. (2017).
39 Theunissen, T. W. *et al.* Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions. *Curr Biol.* **21**, 65-71. doi: 10.1016/j.cub.2010.1011.1074. Epub 2010 Dec 1030. (2011).

40 Ying, Q. L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol* **21**, 183-186, doi:10.1038/nbt780 (2003).
41 Lessing, D., Anguera, M. C. & Lee, J. T. X chromosome inactivation and epigenetic responses to cellular reprogramming. *Annu Rev Genomics Hum Genet* **14:85-110.**, 10.1146/annurev-genom-091212-153530. Epub 092013 May 091216. (2013).
42 Silva, J. *et al.* Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol.* **6**, e253. doi: 210.1371/journal.pbio.0060253. (2008).
43 Shu, J. *et al.* Induction of pluripotency in mouse somatic cells with lineage specifiers. *Cell* **153**, 963-975, doi:10.1016/j.cell.2013.05.001 (2013).
44 Huangfu, D. *et al.* Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nat Biotechnol* **26**, 1269-1275, doi:10.1038/nbt.1502 (2008).
45 Biswas, D. & Jiang, P. Chemically Induced Reprogramming of Somatic Cells to Pluripotent Stem Cells and Neural Cells. *Int J Mol Sci* **17**, 226, doi:10.3390/ijms17020226 (2016).
46 Feng, B., Ng, J. H., Heng, J. C. & Ng, H. H. Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells. *Cell Stem Cell* **4**, 301-312, doi:10.1016/j.stem.2009.03.005 (2009).
47 Ding, L. *et al.* Systems Analyses Reveal Shared and Diverse Attributes of Oct4 Regulation in Pluripotent Cells. *Cell Syst.* **1**, 141-151. doi: 110.1016/j.cels.2015.1008.1002. Epub 2015 Aug 1026. (2015).
48 Li, Q. *et al.* The Pluripotency Factor NANOG Binds to GLI Proteins and Represses Hedgehog-mediated Transcription. *J Biol Chem.* **291**, 7171-7182. doi: 7110.1074/jbc.M7116.714857. Epub 712016 Jan 714821. (2016).
49 De Silva, M. G. *et al.* Gene expression changes during step-wise differentiation of embryonic stem cells along the inner ear hair cell pathway. *Acta Otolaryngol.* **126**, 1148-1157. doi: 1110.1080/00016480600702118. (2006).
50 Iseki, H. *et al.* Combined Overexpression of JARID2, PRDM14, ESRRB, and SALL4A Dramatically Improves Efficiency and Kinetics of Reprogramming to Induced Pluripotent Stem Cells. *Stem Cells.* **34**, 322-333. doi: 310.1002/stem.2243. Epub 2015 Nov 1026. (2016).
51 Bagci, H. & Fisher, A. G. DNA demethylation in pluripotency and reprogramming: the role of tet proteins and cell division. *Cell Stem Cell.* **13**, 265-269. doi: 210.1016/j.stem.2013.1008.1005. (2013).
52 Festuccia, N., Owens, N. & Navarro, P. Esrrb, an estrogen-related receptor involved in early development, pluripotency, and reprogramming. *FEBS Lett* **23**, 1873-3468 (2017).
53 Fidalgo, M. *et al.* Zfp281 Coordinates Opposing Functions of Tet1 and Tet2 in Pluripotent States. *Cell Stem Cell.* **19**, 355-369. doi: 310.1016/j.stem.2016.1005.1025. Epub 2016 Jun 1023. (2016).
54 Hall, J. *et al.* Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal. *Cell Stem Cell.* **5**, 597-609. doi: 510.1016/j.stem.2009.1011.1003. (2009).
55 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* **337**, 816-821. doi: 810.1126/science.1225829. Epub 1222012 Jun 1225828. (2012).
56 Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.* **12**, 2048-2060. (1998).
57 Barau, J. *et al.* The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* **354**, 909-912, doi:10.1126/science.aah5143 (2016).
58 Haberland, M. *et al.* Regulation of HDAC9 gene expression by MEF2 establishes a negative-feedback loop in the transcriptional circuitry of muscle differentiation. *Mol Cell Biol* **27**, 518-525, doi:10.1128/MCB.01415-06 (2007).
59 Mihaylova, M. M. & Shaw, R. J. Metabolic reprogramming by class I and II histone deacetylases. *Trends Endocrinol Metab* **24**, 48-57, doi:10.1016/j.tem.2012.09.003 (2013).

60 Zhao, Y. *et al.* Two supporting factors greatly improve the efficiency of human iPSC generation. *Cell Stem Cell* **3**, 475-479, doi:10.1016/j.stem.2008.10.002 (2008).
61 Chappell, J. & Dalton, S. Roles for MYC in the establishment and maintenance of pluripotency. *Cold Spring Harb Perspect Med* **3**, a014381, doi:10.1101/cshperspect.a014381 (2013).
62 Awe, J. P. & Byrne, J. A. Identifying candidate oocyte reprogramming factors using cross-species global transcriptional analysis. *Cell Reprogram* **15**, 126-133, doi:10.1089/cell.2012.0060 (2013).
63 Singhal, N. *et al.* Chromatin-Remodeling Components of the BAF Complex Facilitate Reprogramming. *Cell* **141**, 943-955, doi:10.1016/j.cell.2010.04.037 (2010).
64 Yu, Y. *et al.* Smad7 enables STAT3 activation and promotes pluripotency independent of TGF-beta signaling. *Proc Natl Acad Sci U S A.* **114**, 10113-10118. doi: 10110.11073/pnas.1705755114. Epub 1705752017 Sep 1705755115. (2017).
65 Liu, Y. *et al.* CRISPR Activation Screens Systematically Identify Factors that Drive Neuronal Fate and Reprogramming. *Cell Stem Cell* **23**, 758-771 e758, doi:10.1016/j.stem.2018.09.003 (2018).
66 Heaton, B. E. *et al.* A CRISPR Activation Screen Identifies a Pan-avian Influenza Virus Inhibitory Host Factor. *Cell Rep* **20**, 1503-1512, doi:10.1016/j.celrep.2017.07.060 (2017).
67 Bester, A. C. *et al.* An Integrated Genome-wide CRISPRa Approach to Functionalize lncRNAs in Drug Resistance. *Cell* **173**, 649-664 e620, doi:10.1016/j.cell.2018.03.052 (2018).
68 Kampmann, M. CRISPRi and CRISPRa Screens in Mammalian Cells for Precision Biology and Medicine. *ACS Chem Biol* **13**, 406-416, doi:10.1021/acschembio.7b00657 (2018).
69 Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588, doi:10.1038/nature14136 (2015).
70 Weltner, J. *et al.* Human pluripotent reprogramming with CRISPR activators. *Nat Commun* **9**, 2643, doi:10.1038/s41467-018-05067-x (2018).
71 Guo, J. *et al.* An inducible CRISPR-ON system for controllable gene activation in human pluripotent stem cells. *Protein Cell* **8**, 379-393, doi:10.1007/s13238-016-0360-8 (2017).
72 Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell.* **138**, 722-737. doi: 710.1016/j.cell.2009.1007.1039. (2009).
73 Carter, A. C., Davis-Dusenbery, B. N., Koszka, K., Ichida, J. K. & Eggan, K. Nanog-independent reprogramming to iPSCs with canonical factors. *Stem Cell Reports.* **2**, 119-126. doi: 110.1016/j.stemcr.2013.1012.1010. eCollection 2014 Feb 1011. (2014).
74 Schwarz, B. A., Bar-Nur, O., Silva, J. C. & Hochedlinger, K. Nanog is dispensable for the generation of induced pluripotent stem cells. *Curr Biol.* **24**, 347-350. doi: 310.1016/j.cub.2013.1012.1050. Epub 2014 Jan 1023. (2014).
75 Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet.* **24**, 372-376. (2000).
76 Karwacki-Neisius, V. *et al.* Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. *Cell Stem Cell.* **12**, 531-545. doi: 510.1016/j.stem.2013.1004.1023. (2013).
77 Radzisheuskaya, A. *et al.* A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. *Nat Cell Biol.* **15**, 579-590. doi: 510.1038/ncb2742. Epub 2013 Apr 1030. (2013).
78 Radzisheuskaya, A. *et al.* A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. *Nat Cell Biol* **15**, 579-590, doi:10.1038/ncb2742 (2013).
79 Gao, X. *et al.* Reprogramming to pluripotency using designer TALE transcription factors targeting enhancers. *Stem Cell Reports* **1**, 183-197, doi:10.1016/j.stemcr.2013.06.002 (2013).
80 Zhang, J. *et al.* Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat Cell Biol.* **8**, 1114-1123. Epub 2006 Sep 1117. (2006).

81 Golipour, A. *et al.* A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* **11**, 769-782, doi:10.1016/j.stem.2012.11.008 (2012).
82 Festuccia, N., Owens, N. & Navarro, P. Esrrb, an estrogen-related receptor involved in early development, pluripotency, and reprogramming. *FEBS Lett*, doi:10.1002/1873-3468.12826 (2017).
83 Adachi, K. *et al.* Context-dependent wiring of Sox2 regulatory networks for self-renewal of embryonic and trophoblast stem cells. *Mol Cell* **52**, 380-392, doi:10.1016/j.molcel.2013.09.002 (2013).
84 Ding, J. *et al.* Tex10 Coordinates Epigenetic Control of Super-Enhancer Activity in Pluripotency and Reprogramming. *Cell Stem Cell.* **16**, 653-668. doi: 610.1016/j.stem.2015.1004.1001. Epub 2015 Apr 1030. (2015).
85 Hall, J. *et al.* Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal. *Cell Stem Cell* **5**, 597-609, doi:10.1016/j.stem.2009.11.003 (2009).
86 Jeon, H. *et al.* Comprehensive Identification of Kruppel-Like Factor Family Members Contributing to the Self-Renewal of Mouse Embryonic Stem Cells and Cellular Reprogramming. *PLoS One* **11**, e0150715, doi:10.1371/journal.pone.0150715 (2016).
87 Azami, T. *et al.* Klf5 suppresses ERK signaling in mouse pluripotent stem cells. *PLoS One* **13**, e0207321, doi:10.1371/journal.pone.0207321 (2018).
88 Tang, Y. & Tian, X. C. JAK-STAT3 and somatic cell reprogramming. *JAKSTAT* **2**, e24935, doi:10.4161/jkst.24935 (2013).
89 Ema, M. *et al.* Kruppel-like factor 5 is essential for blastocyst development and the normal self-renewal of mouse ESCs. *Cell Stem Cell* **3**, 555-567, doi:10.1016/j.stem.2008.09.003 (2008).
90 Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A hyperactive piggyBac transposase for mammalian applications. *Proc Natl Acad Sci U S A* **108**, 1531-1536, doi:10.1073/pnas.1008322108 (2011).
91 Cunningham, F. *et al.* Ensembl 2015. *Nucleic Acids Res* **43**, D662-669, doi:10.1093/nar/gku1010 (2015).
92 Ravasi, T. *et al.* An atlas of combinatorial transcriptional regulation in mouse and man. *Cell.* **140**, 744-752. doi: 710.1016/j.cell.2010.1001.1044. (2010).
93 Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science.* **343**, 84-87. doi: 10.1126/science.1247005. Epub 1242013 Dec 1247012. (2014).
94 Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera Mdel, C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* **32**, 267-273, doi:10.1038/nbt.2800 (2014).
95 Castro, D. S. *et al.* A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. *Genes Dev* **25**, 930-945, doi:10.1101/gad.627811 (2011).

Author Contributions

J.Y. initiated and designed the project, performed GoF screening and EpiSC, MEF reprogramming with cDNA, analysed the data and wrote the manuscript. S.S.R. performed EpiSC reprogramming, ESC conversion, MEF reprogramming with gRNA, RNA-seq, analysed the data and wrote the manuscript. M.F. performed EpiSC reprogramming and RNA-seq data analysis and wrote the manuscript. G.L. and X.Z. performed microinjection and analysed chimera data. H.P. designed the GoF sgRNA library and performed RNA-seq data analysis. D.G. performed experiments to determine various dCas9-mediated overexpression levels. P.L. and A.B. financially supported the project and interpreted the data. E.M. initiated, designed and supervised the project, performed GoF screening, validation, analysed and interpreted data, and wrote the manuscript.

Acknowledgements

We thank Caroline Sinclair, Michael Woods, Evelyn Grau and Brendan Doe from the research support facility (RSF) at the Sanger Institute for providing blastocysts. We also thank Prof. Austin Smith and Dr. Ge Guo from Wellcome Trust-Medical Research Council Cambridge Stem Cell Institute for the reagents. Jian Yang is funded by the National Science Foundation of China (NSFC) (31871491). This work was funded by the Wellcome Trust (WT098051). Emmanouil Metzakopian is funded by the UK Dementia Research Institute (RG86445) and Parkinson's Disease UK (F-1501).

Conflict of Interest

The authors declare no conflict of interest.

Figure Legends

Figure 1 | GoF EpiSC reprogramming screening with CRISPRa and sgRNA library

(a) Activation of *Ascl1* and *Neurog2* in HEK293 cells. Cells were transfected with one sgRNA per target and four different dCas9 versions. Gene expression was analysed by RT-qPCR and compared to *GAPDH* expression, fold change is expressed relative to dCas9:VP160 for each gene (* p<0.05; *** p<0.001). (**b)** Illustration of sgRNA design targeting gene promoters in the murine genome. **(c)** Screening strategy using *Oct4*-GFP EpiSCs with stable expression of dCas9:SAM and lentiviral transduction (MOI 0.3) of the sgRNA library. Reprogramming took place in 2i/LIF for 14-16 days, after sorting for transduced cells. *Oct4*-GFP$^+$ iPSC colonies were then picked and NGS identified candidate sgRNAs. **(d)** GOTOOLBox analysis of 142 genes identified in GoF screening. Pathways are presented along with fold change compared to reference; colours indicate p-values. **(e)** Individual validation of genes. *Oct4*-GFP EpiSCs were transfected with dCas9:SAM and single sgRNAs for 54 genes including *Nanog* and *Oct4* and reprogrammed in 2i/LIF for up to 16 days. The number of *Oct4*-GFP$^+$ iPSC colonies were counted (mean of triplicates ± s.d).

Figure 2 | *Sall1* and *Nanog* reprogram EpiSC and influence ESC differentiation

(a) Reprogramming efficiencies of *Sall1*, *Nanog* and *Oct4*. *Oct4*-GFP EpiSCs were stably transfected with CRISPRa or cDNA and cultured in 2i/LIF for 16 days without selection. *Oct4*-GFP$^+$ colonies were counted (mean of triplicates ± s.d.)–** p<0.01; *** p<0.001; **** p<0.0001. **(b)** Morphology of *Oct4*-GFP$^+$ colonies at day 20 in 2i/LIF. The reprogrammed colonies are morphologically similar to ESC colonies with *Oct4*-GFP$^+$ fluorescence. No colonies were observed in untransfected or mock transfected

EpiSCs. **(c)** RT-qPCR expression profiles of pluripotency markers and EpiSC markers in iPSC colonies normalized to *Gapdh* and relative to ESCs (mean of triplicates ± s.d). Cells were transfected with CRISPRa or cDNA as before. **(d)** Chimeric mouse produced with CRISPRa *Sall1* – induced PSCs injected into C57B/6 blastocyst. **(e)** Flow cytometric analysis of *Rex1*-GFP$^+$ cells cultured in EpiSC medium at the timepoints indicated. Cells were stably transfected with *Sall1* or *Nanog* cDNA, or empty vector and cultured in EpiSC medium (mean of triplicates ± s.d, * p<0.05; *** p<0.001, **** p<0.0001 vs. PBCAG:Empty). **(f)** Number of *Rex1*-GFP$^+$ ESC colonies recovered after ESCs were converted in EpiSC medium at indicated timepoints. 600 cells were plated at timepoint zero (mean of triplicates ± s.d., * p< 0.05; *** p< 0.001, **** p< 0.0001 vs. PBCAG:Empty).

Figure 3 | CRISPRa gene induction and cDNA mediated overexpression of *Sall1*, *Nanog* reprogrammed MEF to iPSCs

(a) (4F+CRISPRa) MEFs were stably transfected with CAG4F and gRNAs against *Sall1*/*Nanog*/*Sall1*+*Nanog* and reprogrammed in ESC medium. *Oct4*-GFP$^+$ colonies were counted after 18 days of reprogramming (mean of triplicates ± s.d., ** p<0.01, *** p<0.001). (4F+cDNA) MEFs stably transfected with TRE4F, TRENanog and TRESall1 (all co-transfected with PBEF-1αTet3G), induced with 0.5 µg/ml Doxycycline for 12 days and counted on day 18 (mean of triplicates ± s.d., ** p<0.01). **(b)** Alkaline phosphatase (AP$^+$) stained ESC colonies reprogrammed from MEFs by 4F alone and in combination with *Sall1* (induced with Dox at 0.5 µg/ml). **(c)** iPSCs reprogrammed from C57B/6J MEF with 4F/Sall1. *Oct4*-GFP expression and ESC like morphology (upper panel), immunofluorescent staining for pluripotency markers SSEA-1 and Nanog (lower panel). **(d)** *In vitro* differentiation of C57B/6 MEF reprogrammed iPSCs with 4F/Sall1. iPSCs were cultured either in N2B27 for neuronal differentiation,

immunofluorescence showed β-Tubulin III$^+$ neuron; or in M10 for mesoderm and endoderm differentiation as detected by α-SMA and AFP antibody staining **(e)** Chimeric mice produced with 4F/Sall1-iPSCs injected into CD1 blastocysts. **(f)** Activation of X chromosomes in female 4F/Sall1-iPSCs. iPSCs and the cells differentiated from iPSCs were co-immunostained for H3K27me3 and *Oct4*. Arrows indicate H3K27me3 foci.

Figure 4 | RNA-seq identifies potential mechanisms of reprogramming mediated by Sall1 and Nanog

(a) Venn diagram showing number and percentage of genes being differentially expressed in *Sall1* and *Nanog* overexpressing cells with a cut-off padj value < 0.001. Upregulated (green arrow) and downregulated (red arrow) genes for further experiments were chosen from the overlap between *Sall1* and *Nanog*, except *Klf5* and *Fam189a2*. **(b)** GOTOOLBox analysis of common regulated genes presented along with fold change compared to reference. Colours indicate p-values. **(c)** RT-qPCR validations for results obtained from RNA-seq analysis on *Sall1* and *Nanog* overexpressing EpiSCs, 24 h after transfection, normalized to *Gapdh* expression and relative to PBCAG:empty (mean of triplicates ± s.d.) **(d)** Reprogramming of *Oct4*-GFP EpiSCs via CRISPRa mediated gene induction of *Klf5*, *Fam189a2*, *Tex10* and *Tet3*. After transfection, cells were cultured in 2i/LIF for 20 days and *Oct4*-GFP$^+$ colonies were counted. (mean of triplicates ± s.d.). **(e)** RT-qPCR expression levels of *Klf5*, *Fam189a2*, *Sall1* and *Nanog* after CRISPRa mediated gene induction of *Klf5* and *Fam189a2*, flow-sorted for sgRNA expression 24 h after transfection, normalized to *Gapdh* and relative to dCas9:SAM (mean of triplicates ± s.d.). **(f)** RT-qPCR expression levels of key regulators in JAK/STAT3 and TGFβ signalling, flow-sorted for sgRNA expression 24 h after changing to 2i/LIF media (48 h after transfection), normalized to

Gapdh and relative to dCas9:SAM (mean of triplicates ± s.d.). **(g)** Reprogramming of *Oct4*-GFP EpiSCs via CRISPRa mediated gene induction of *Klf5*, *Fam189a2, Sall1* and *Nanog.* After transfection, cells were cultured in 2i/LIF for 20 days and *Oct4*-GFP^{+} colonies were counted (mean of triplicates ± s.d.).

Supplementary Figure 1 | Related to main Figure 1: Plasmids used in the study

LTR: long terminal repeat from HIV-1. **PB:** *PiggyBac* inverted terminal repeats to enable stable integration by *PBase* mediated transposition. **U6:** U6 promoter for sgRNA transcription. **sgRNA:** sgRNA scaffold. **sgRNA-MS2-MS2:** sgRNA scaffold with extended stem loops for dCas9:SAM. **PGK:** PGK promoter. **PuroR:** puromycin N-acetyltransferase. **T2A:** 2A peptide from *Thosea asigna* virus capsid protein. **BFP:** blue fluorescent protein. **Poly(A):** polyadenylation signal. **EF1-α:** human elongation factor 1 alpha promoter. **VP160-dCas9:** 10 tandem repeats of transcriptional activation domain of herpes simplex virus protein VP16 fused to dCas9. **mCherry:** mCherry fluorescent protein. **P2A:** 2A peptide from *porcine teschovirus-1* polyprotein. **BlastR:** *Aspergillus terreus* blasticidin S deaminase. **GCN4:** GCN4 single chain antibody. **VP64:** 4 tandem repeats of transcriptional activation domain of herpes simplex virus protein VP16. **dCas9-SunTag:** GCN4 peptide fused to dCas9. **dCas9-VPR:** VP64, P65 and RTA (transcriptional activation domain from the human herpesvirus 4) fused to dCas9. **MS2:** bacteriophage MS2 coat protein. **P65:** C-terminal portion of the p65 subunit of mouse NF-κB. **HSF1:** C-terminal activation domain from the human heat shock transcription factor. **E2A:** 2A peptide from equine rhinitis A virus polyprotein. **bpA:** bovine growth hormone polyadenylation signal. **CAG:** CMV enhancer, chicken β-actin and rabbit β-globin promoter. **TRE:** tet response element. **TET3G:** Tet-On-3G transactivator protein.

Supplementary Figure 2 | Related to main Figures 1 and 2: Supplementary results for GoF CRISPRa screen and gene dosage of *Oct4*

(a) Flow cytometry plots showing dynamics of GFP expression in *Oct4*-GFP reporter EpiSCs during reprogramming: *Oct4*-GFP EpiSCs transfected with dCas9:SAM + sgRNA against gene *Nanog*, dCas9:SAM only or untransfected. Over a time course of 14 days in selective 2i/LIF medium, the GFP positive cell population rapidly and completely disappears in the untransfected or dCas9:SAM only groups. However, cells reprogrammed by CRISPRa mediated overexpression of *Nanog* recover their *Oct4*-GFP levels by day 14. **(b)** Flow cytometry plots showing the percentage of BFP^{+} cells obtained upon transducing *Oct4*-GFP EpiSCs with lentiviral sgRNA library at an MOI of 0 (untransduced) and an MOI of 0.3. **(c)** RT-qPCR on *Oct4*-GFP EpiSCs transfected with CRISPRa and cDNA Oct4 showing levels of total *Oct4* in cells after 3 days in 2i/LIF. Levels of *Oct4* are normalized to *Gapdh* (mean of triplicates ± s.d.). **(d)** Western blot on *Oct4*-GFP EpiSCs transfected with CRISPRa and cDNA *Oct4* showing levels of total Oct4 in cells after 3 days in 2i/LIF (top panel). Western blot for Oct4 in gRNA and CAG cDNA transfected EpiSCs with a comparison to ESCs (bottom panel) *Gapdh* was used as loading control in both cases. **(e)** *Umodl1* regulates re-programming pathways: RT-qPCR expression levels of key regulators in JAK/STAT3 and TGFβ signalling, on flow-sorted for sgRNA expression either 24 h after transfection (EpiSC media) or 24 h after changing to 2i/LIF media (48 h after transfection), normalized to *Gapdh* and relative to dCas9:SAM (mean of triplicates ± s.d.).**(f)** Variable gene dosage with tiled guides: *Oct4*-GFP EpiSCs were transfected with dCas:SAM and all tiled sgRNAs for *Sall1* and *Nanog* available in our library. Numbers on top of bars indicate distance to the transcription start site (TSS). **(g)** RT-qPCR shows that *Sall4* is not overexpressed when *Oct4*-GFP+ve EpiSC are transfected with *Sall1* CRISPRa.

Supplementary Figure 3 | Related to main Figure 2: Reprogramming in *Nanog*-GFP reporter EpiSCs and supplementary results for ESC to EpiSC conversion

(a,b) Gene induction of *Sall1*, *Nanog* and *Oct4* in *Oct4*-GFP EpiSC **(a)** and *Nanog*-GFP EpiSCs **(b)**, respectively, with CRISPRa (single sgRNA per target) as well as cDNA mediated overexpression in Nanog-GFP EpiSCs. Expression levels were measured by RT-qPCR 72h after transfection and expressed in relation to dCas9:SAM-only or empty vector (mean of triplicates ± s.d). **(c)** Reprogramming efficiencies of *Sall1*, *Nanog* and *Oct4*. Nanog-GFP EpiSCs were stably transfected with CRISPRa or cDNA and cultured in 2i/LIF for 16 days without selection. Nanog-GFP$^+$ colonies were counted and are represented as mean of triplicates ± s.d. ** p<0.01; *** p<0.001; **** p<0.0001. **(d)** RT-qPCR expression profiles of pluripotency markers and EpiSC markers in iPSC colonies normalized to *Gapdh* and relative to ESCs (mean of triplicates ± s.d). Cells were transfected with cDNA as before. **(e)** Chimeric mouse produced with cDNA *Sall1* – induced PSCs injected into C57B/6 blastocyst. **(f)** Identical endpoints of *Oct4*-GFP and *Nanog*-GFP reporter EpiSCs in re-programming: *Oct4*-GFP EpiSCs (top panel) and *Nanog*-GFP EpiSCs (bottom panel) transfected with either dCas9:SAM alone or in combination with sgRNAs against *Sall1*, *Nanog* and *Oct4* and selected in 2i medium in a time course of 14 and 22 days. As before, *Oct4*-GFP EpiSCs lose initial GFP expression rapidly and recover it upon successful re-programming via CRISPRa mediated gene induction of *Sall1*, *Nanog* or *Oct4*, but not in cells transfected with dCas9:SAM only. *Nanog*-GFP reporter EpiSCs – showing no baseline GFP expression – nevertheless upregulate GFP with similar but slightly slower dynamics in sgRNA transfected groups only. On day 14 both reporter lines show comparable GFP positive cell populations and a clear synergistic effect when co-transfecting sgRNAs against *Sall1* and *Nanog*. **(g-i)** *Rex1*-

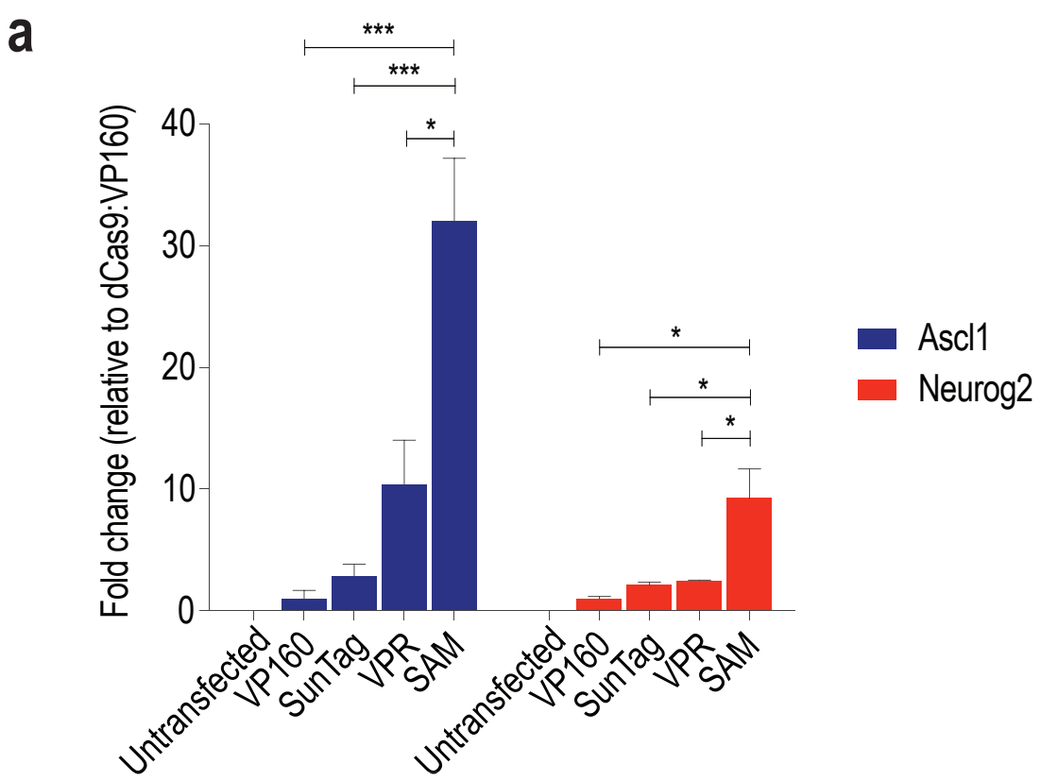
GFP$^+$ ES cells were stably transfected with *Sall1* or *Nanog* cDNA, or empty vector and cultured in EpiSC medium. Change of expression levels of pluripotency marker *Rex1* and differentiation markers *Fgf5* and *Otx2* measured by RT-qPCR normalized to *Gapdh* expression (mean of triplicates ± s.d)

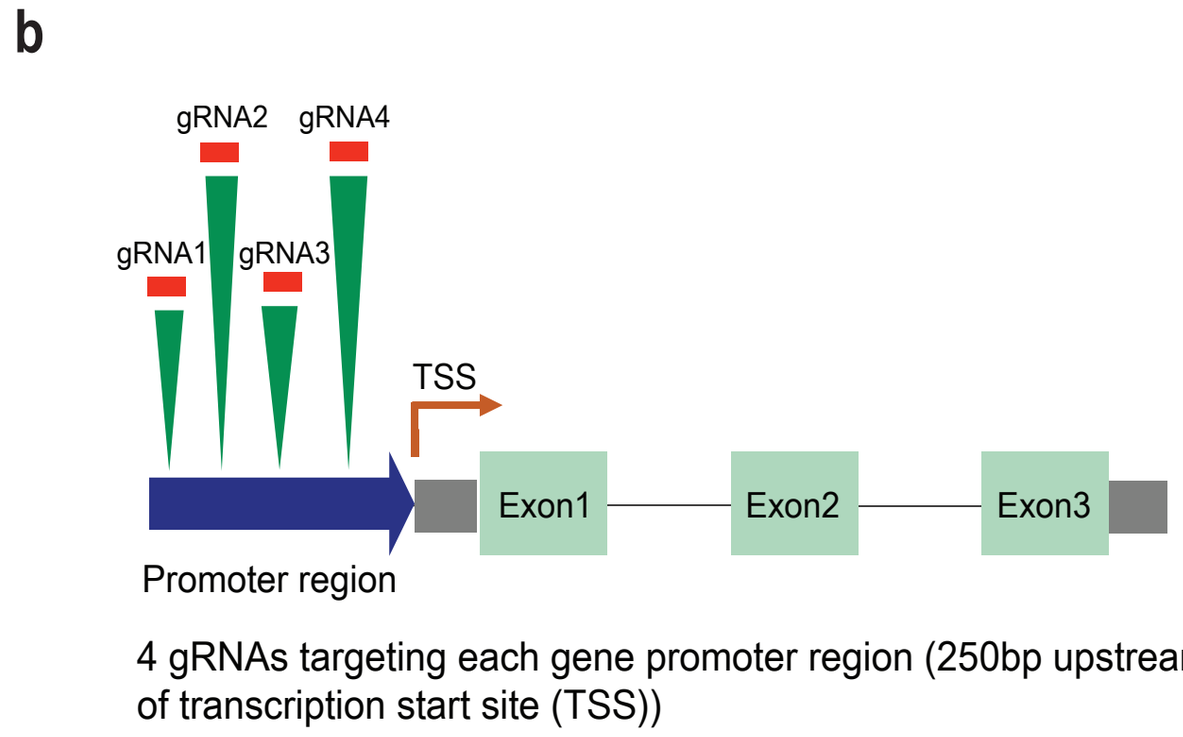
Supplementary Figure 4 | Related to main Figures 3 and 4: *Sall1* cannot replace *Oct4* in MEF reprogramming and supplementary results for RNA-seq

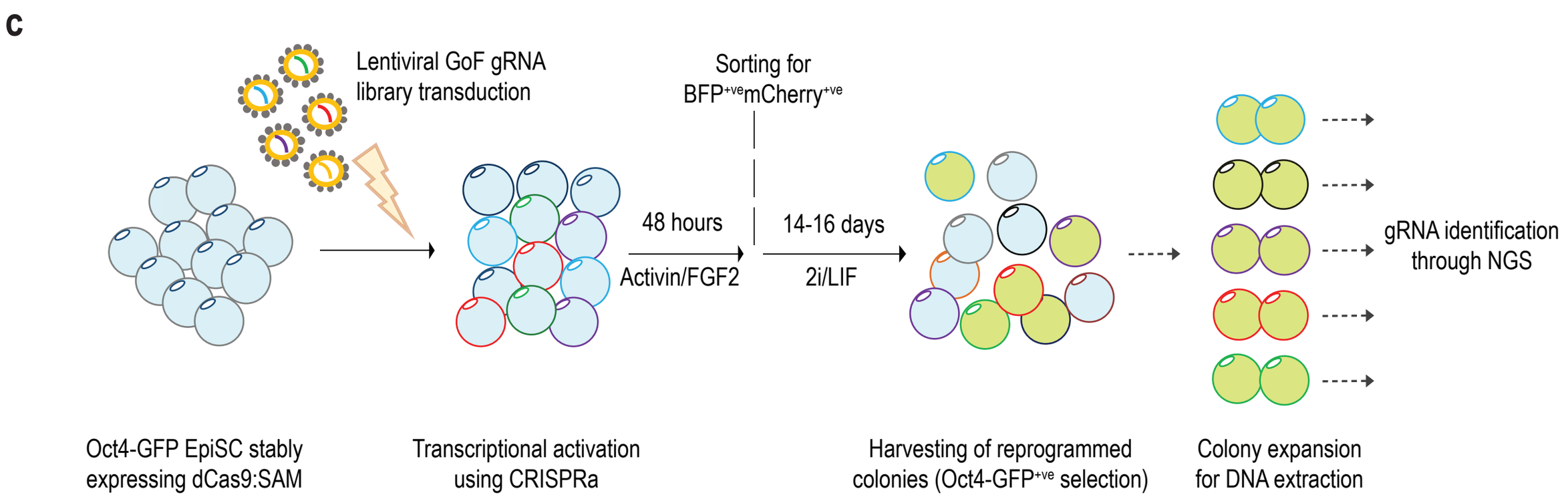
(a) (4F+CRISPRa) MEFs were stably transfected with CAG4F and gRNAs against *Sall1*/*Nanog*/*Sall1*+*Nanog* and reprogrammed in ESC medium. AP$^+$ colonies were counted after 18 days of reprogramming. Gene induction of *Sall1* produced more iPSC colonies compared to CAG4F alone and gene induction of both *Sall1* and *Nanog* produced significantly higher number of colonies compared to activation of either *Sall1* or *Nanog* alone (mean of triplicates ± s.d., ** $p<0.01$). (4F+cDNA) MEFs stably transfected with TRE4F, TRENanog and TRESall1 (all co-transfected with PBEF-1αTet3G) and induced with 0.5 µg/ml Doxycycline for 12 days. Overexpression of *Sall1* produced more iPSC colonies compared to CAG4F alone and overexpression of both *Sall1* and *Nanog* produced significantly higher number of colonies compared to activation of either *Sall1* or *Nanog* alone. AP$^+$ colonies were counted on day 18 (mean of triplicates ± s.d., *** $p<0.001$). **(b)** Morphology of *Oct4*-GFP$^+$ colonies at day 18 in ESC media. The reprogrammed colonies from *Oct4*-GFP$^+$ MEFs are morphologically similar to ESC colonies with *Oct4*-GFP$^+$ fluorescence. No colonies were observed in untransfected or dCas9:SAM only transfected MEFs. **(c)** MEF were stably transfected with CAG4F and TRESall1/PBEF-1αTet3G and reprogrammed in ESC medium. 24h after transfection, expression of Sall1 was induced with different concentration of Doxycycline for 12 days. After 18 days, both *Oct4*-GFP$^+$ and AP$^+$ colonies were counted. Overexpression of *Sall1* produced more iPSC colonies (mean

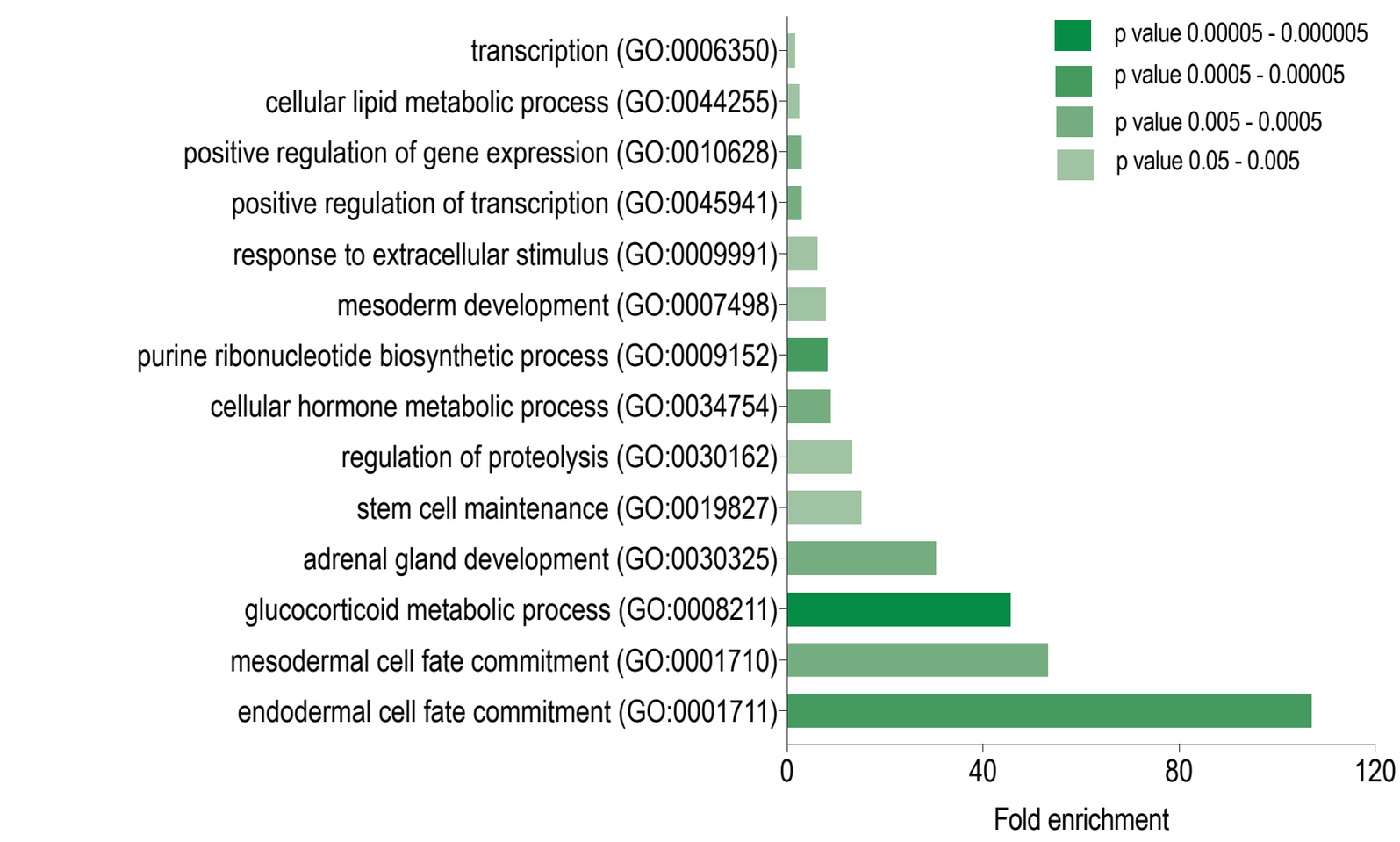
of triplicates ± s.d., ** p<0.01, *** p<0.001, $^{\Delta}$ p<0.05, $^{\Delta\Delta}$ p<0.01 vs. 4F). **(d)** MEFs stably transfected with CAG4F and TRESall1/PBEF-1αTet3G and reprogrammed in ESC medium. 0.5 µg/ml Doxycycline was added on the indicated days for the duration of reprogramming. *Oct4*-GFP^{+} and AP^{+} colonies were counted on D18 (mean of triplicates ± s.d., ** p<0.01, $^{\Delta\Delta\Delta}$ p<0.001 vs. 4F). **(e-g)** MEFs were stably transfected with combinations of *Oct4*, *C-myc*, *Klf4*, *Sox2* (OCKS) and *Sall1* cDNA, whereby *Sall1* replaced either **(e)** *Klf4*, **(f)** *Sox2* or **(g)** *Oct4*. Cells were reprogrammed in ESC medium for 18 days and AP^{+} colonies were counted (mean of triplicates ± s.d). **(h)** Venn diagram showing number and percentage of genes being differentially expressed in *Oct4*-GFP EpiSC transfected with cDNA for *Sall1*/*Nanog*/*Sall1+Nanog* for RNA-seq. Differentially expressed genes were identified with a cut-off padj value < 0.001. **(i)** CAG cDNA mediated overexpression of *Sall1*, *Nanog* in *Oct4*-GFP EpiSC transfected with either *Sall1*/*Nanog*/*Sall1+Nanog* for RNA-seq. Expression levels were measured by RT-qPCR 48h after transfection and expressed in relation to empty vector (mean of triplicates ± s.d).

FIGURE 1


a
Fold change (relative to dCas9:VP160)
0
10
20
30
40
***
***
*
*
*
*
Untransfected
VP160
SunTag
VPR
SAM
Untransfected
VP160
SunTag
VPR
SAM
Ascl1
Neurog2


b
gRNA2
gRNA4
gRNA1
gRNA3
TSS
Exon1
Exon2
Exon3
Promoter region
4 gRNAs targeting each gene promoter region (250bp upstream of transcription start site (TSS))


c
Lentiviral GoF gRNA library transduction
Sorting for BFP+veCherry+ve
48 hours
Activin/FGF2
14-16 days
2i/LIF
gRNA identification through NGS
Oct4-GFP EpiSC stably expressing dCas9:SAM
Transcriptional activation using CRISPRa
Harvesting of reprogrammed colonies (Oct4-GFP+ve selection)
Colony expansion for DNA extraction


d
transcription (GO:0006350)
cellular lipid metabolic process (GO:0044255)
positive regulation of gene expression (GO:0010628)
positive regulation of transcription (GO:0045941)
response to extracellular stimulus (GO:0009991)
mesoderm development (GO:0007498)
purine ribonucleotide biosynthetic process (GO:0009152)
cellular hormone metabolic process (GO:0034754)
regulation of proteolysis (GO:0030162)
stem cell maintenance (GO:0019827)
adrenal gland development (GO:0030325)
glucocorticoid metabolic process (GO:0008211)
mesodermal cell fate commitment (GO:0001710)
endodermal cell fate commitment (GO:0001711)
0
40
80
120
Fold enrichment
p value 0.00005 - 0.000005
p value 0.0005 - 0.00005
p value 0.005 - 0.0005
p value 0.05 - 0.005

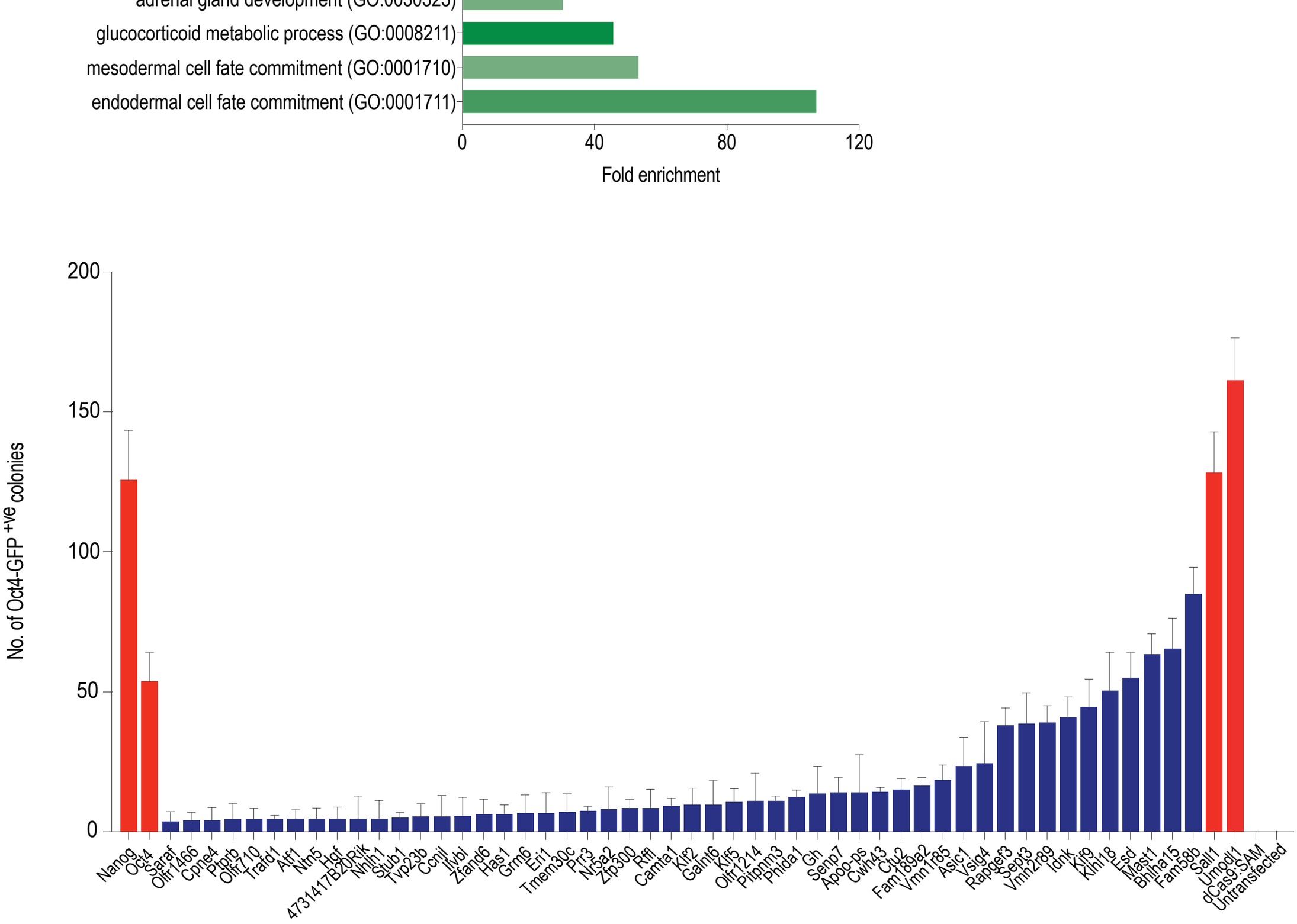
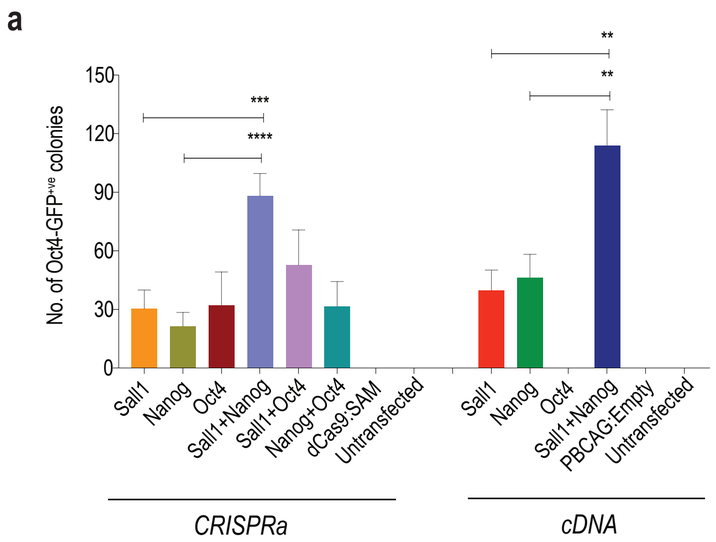
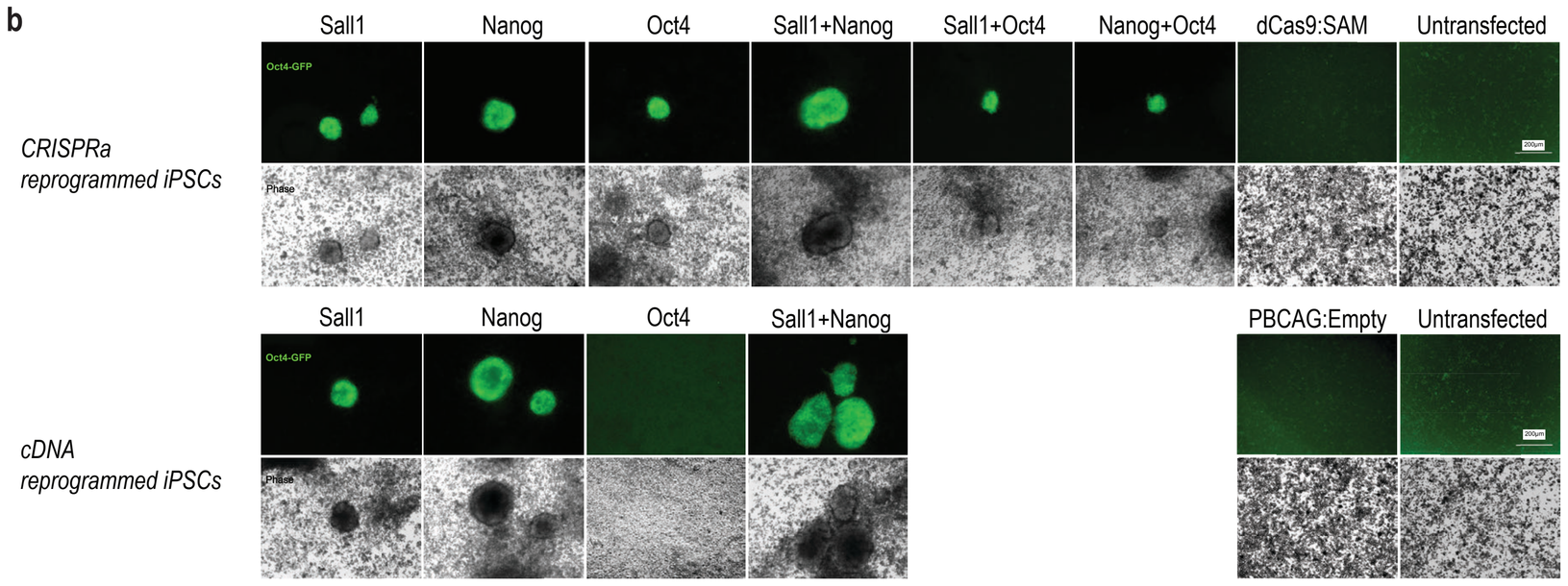
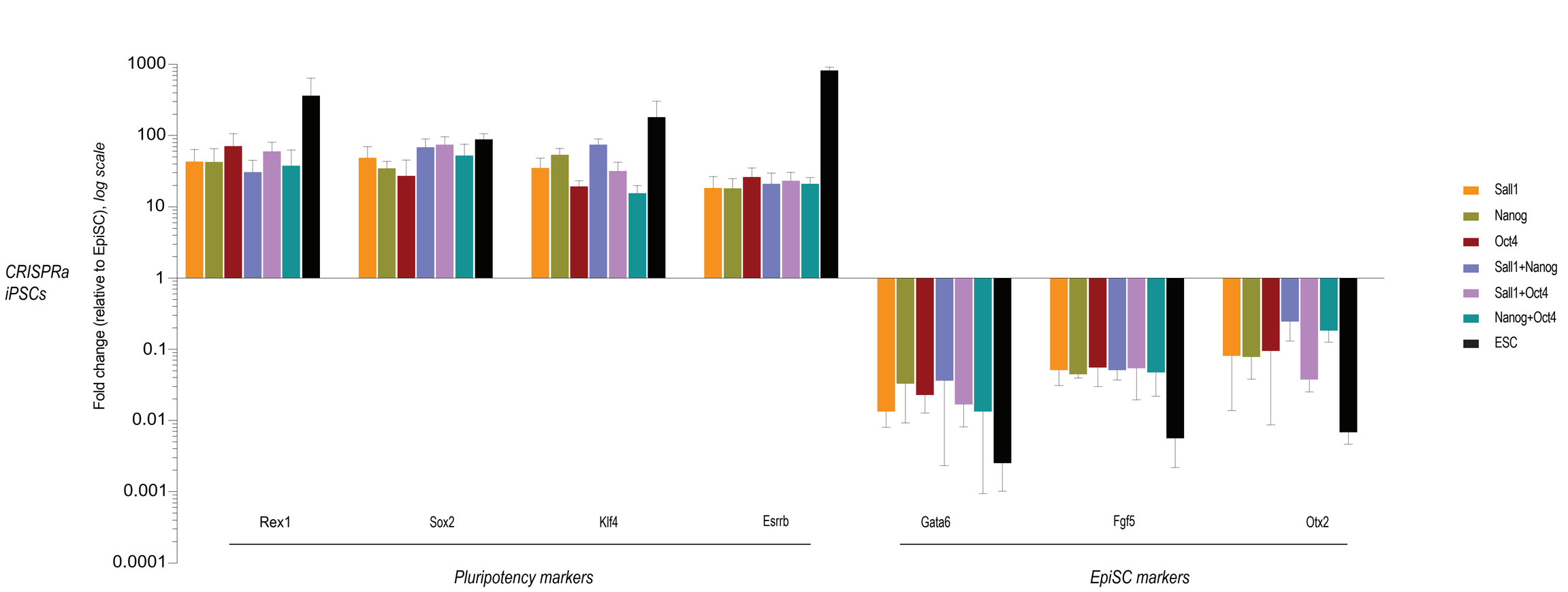
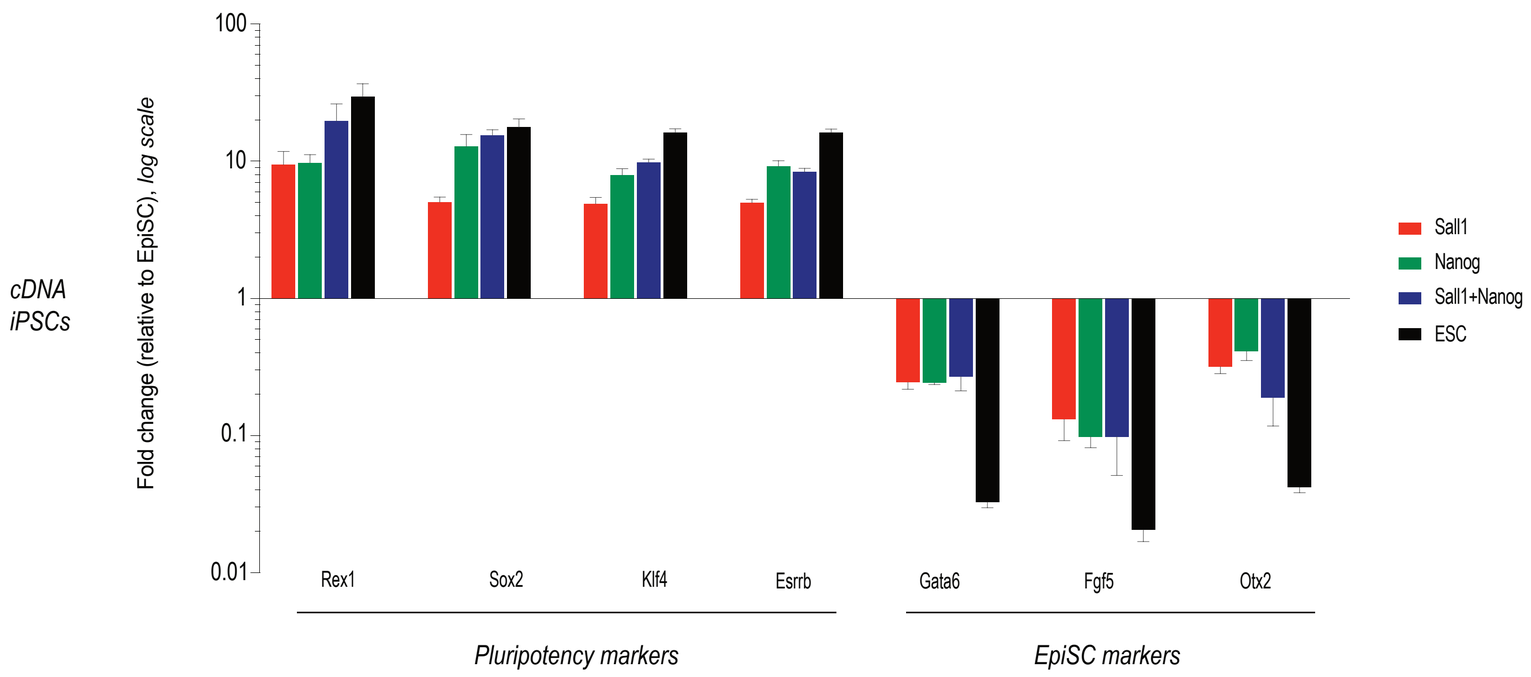

e
No. of Oct4-GFP +ve colonies
0
50
100
150
200
Nanog
Oct4
Saraf
Olfr1466
Cpne4
Ptprb
Olfr710
Trafd1
Atf1
Ntn5
Hgf
4731417B20Rik
Nhlh1
Stub1
Tvp23b
Ccnjl
Ilvbl
Zfand6
Has1
Grm6
Eri1
Tmem30c
Prr3
Nr5a2
Zfp300
Rffl
Camta1
Klf2
Galnt6
Klf5
Olfr1214
Pitpnm3
Phlda1
Gh
Senp7
Apoo-ps
Cwh43
Ctu2
Fam189a2
Vmn1r85
Asic1
Vsig4
Rapgef3
Sept3
Vmn2r89
Idnk
Klf9
Klhl18
Esd
Mast1
Bhlha15
Fam58b
Sall1
Umodl1
dCas9:SAM
Untransfected

FIGURE 2

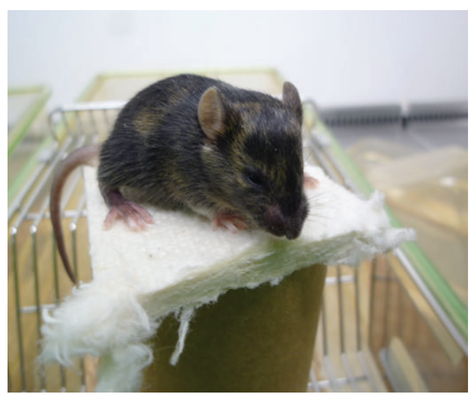




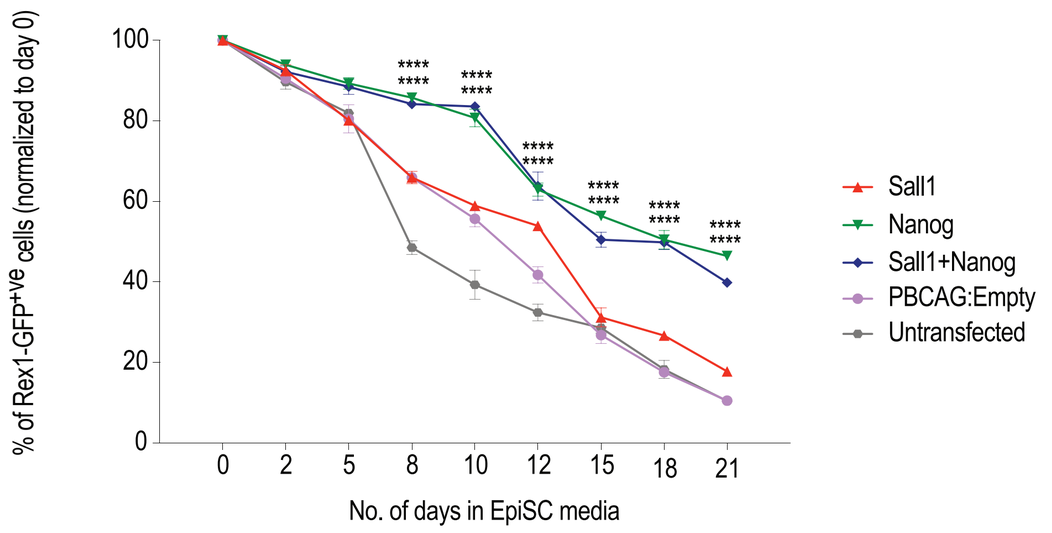




d



CRISPRa Sall1 iPSC chimeric mouse



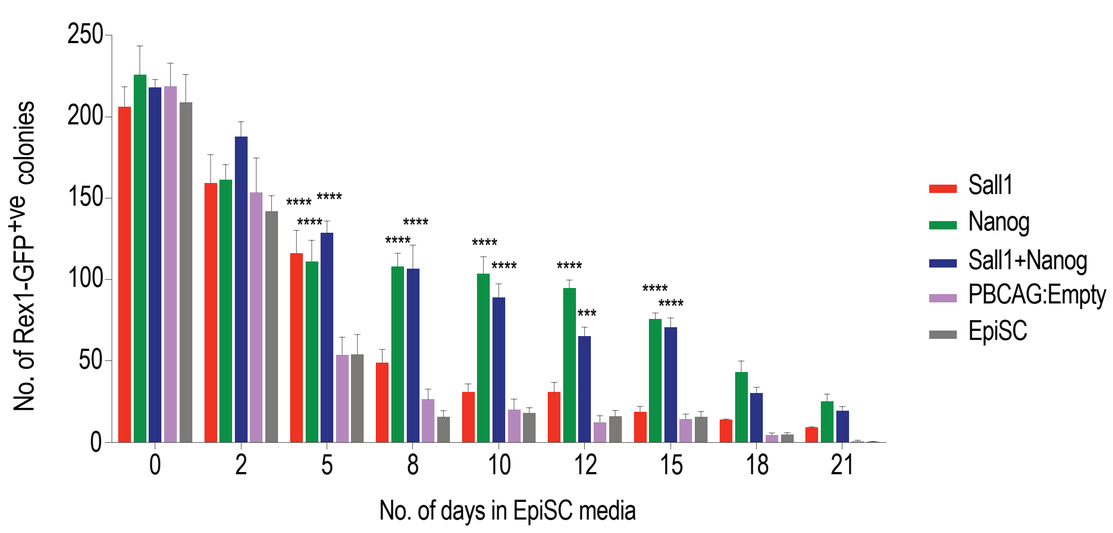
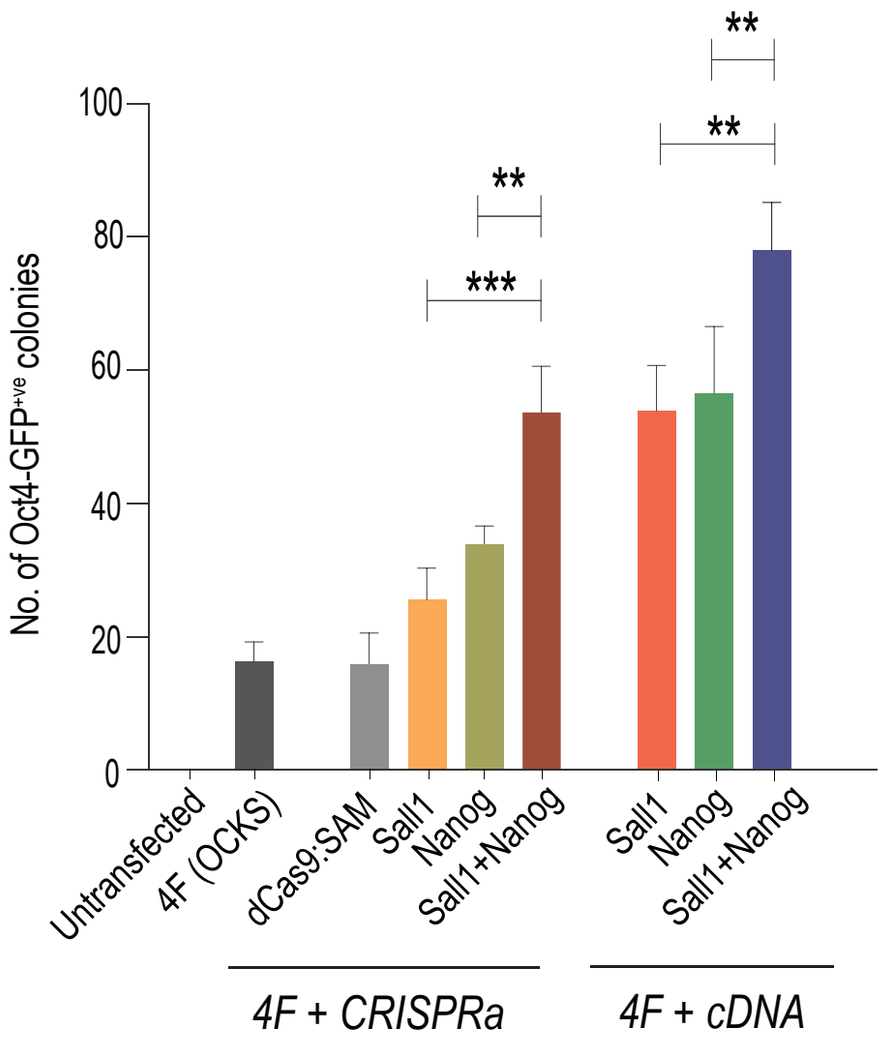
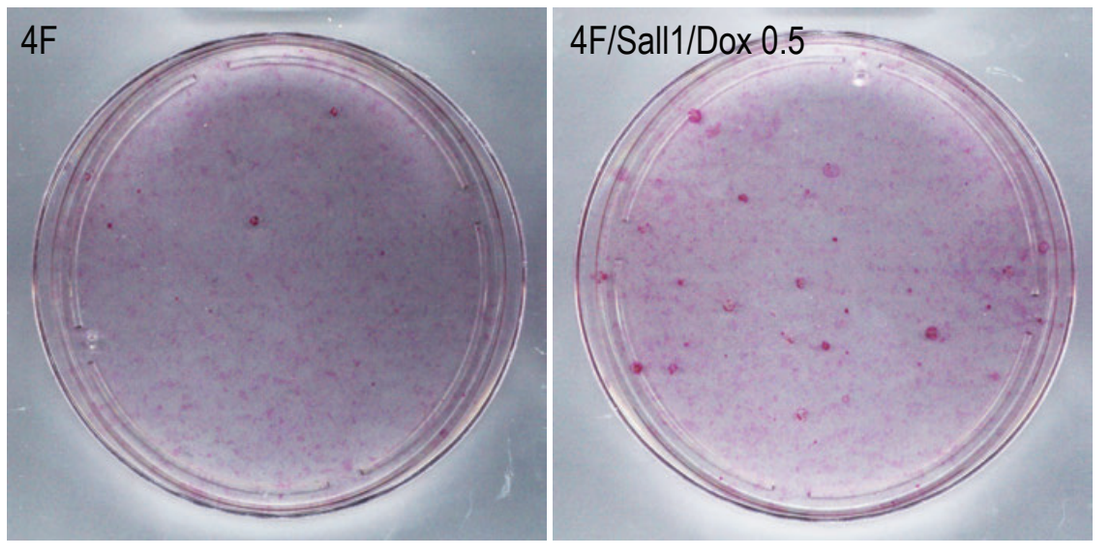


FIGURE 3

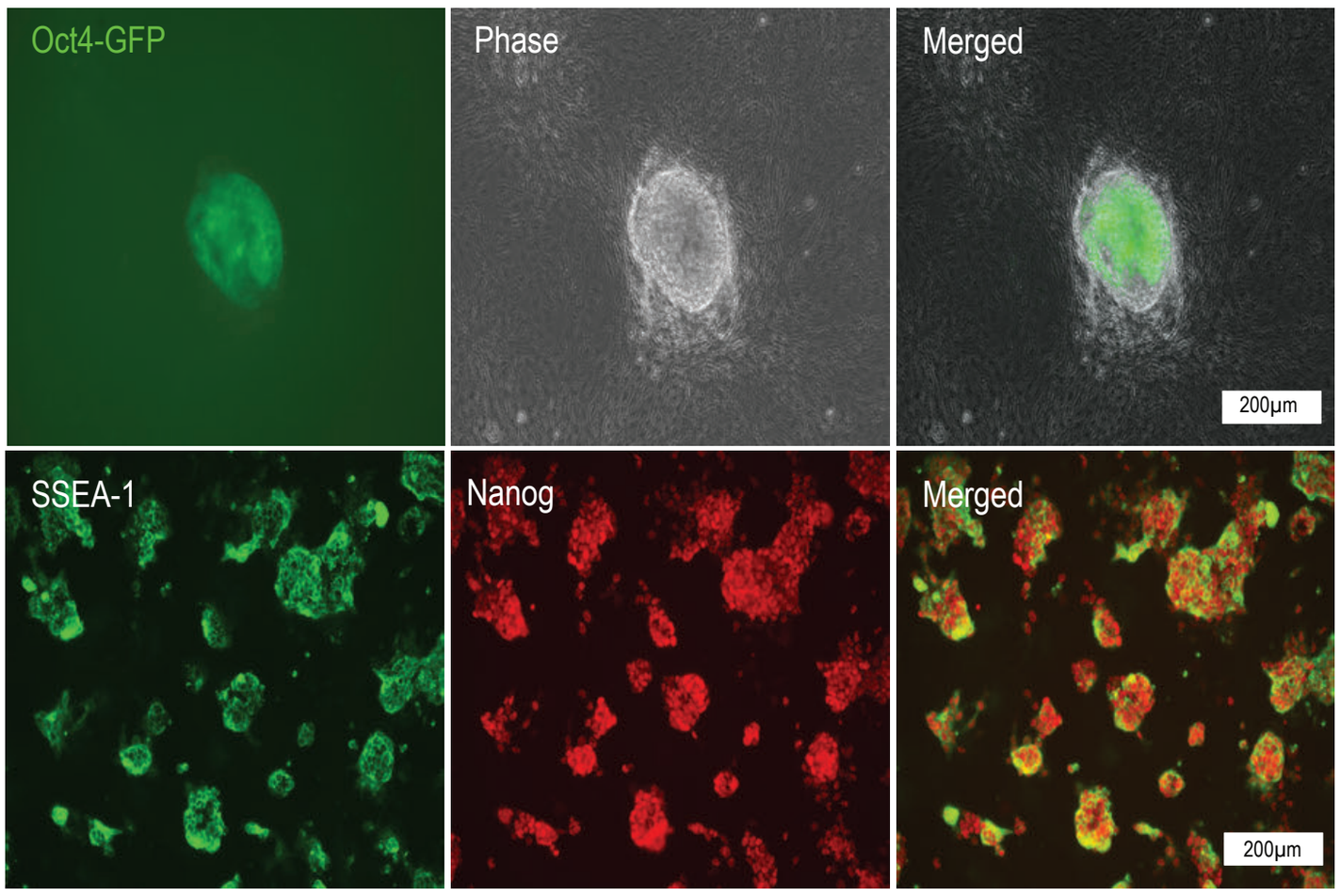
a



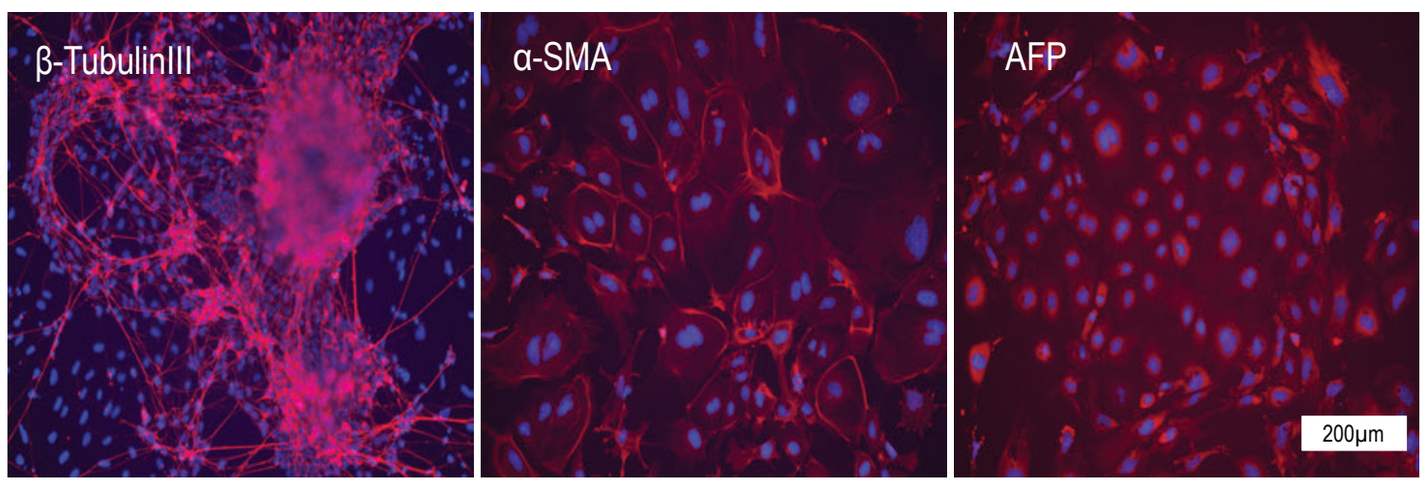
b



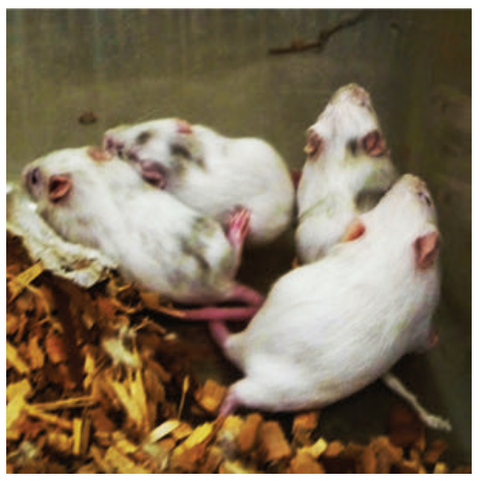
c



d

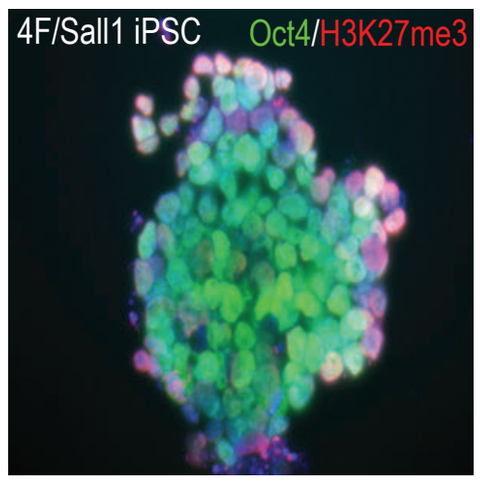


e



4F/Sall1 iPSC chimeric mice

f



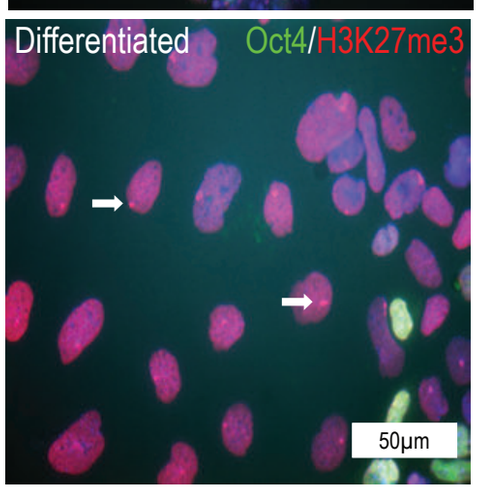
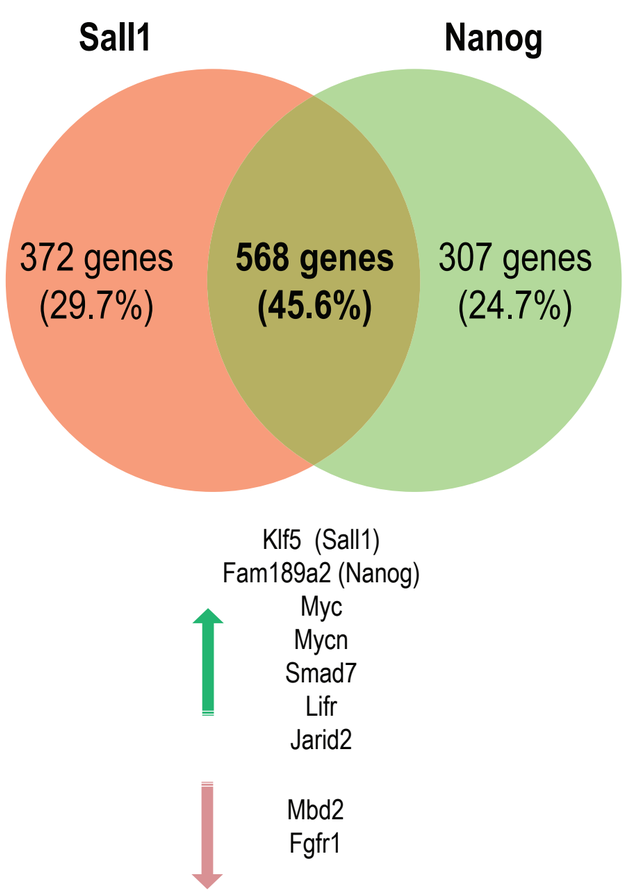
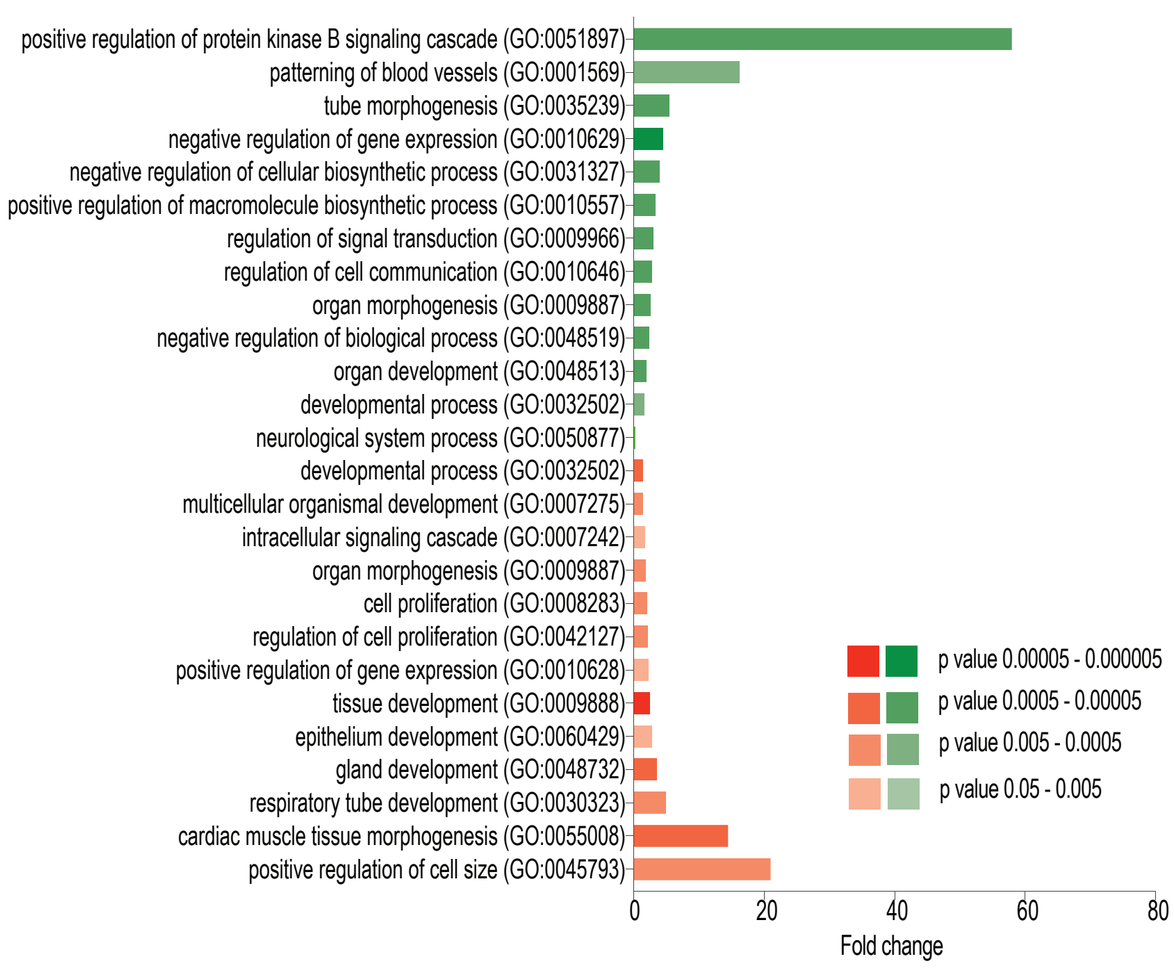
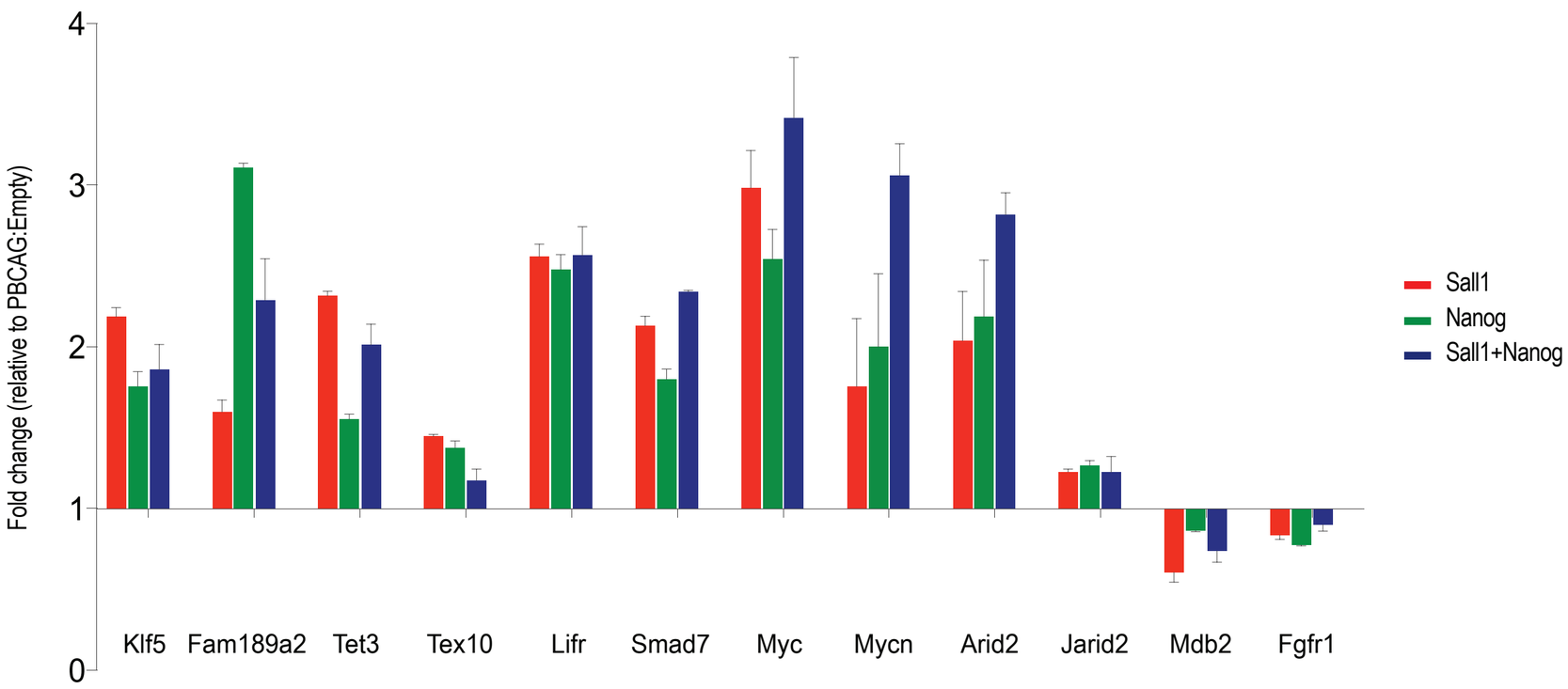
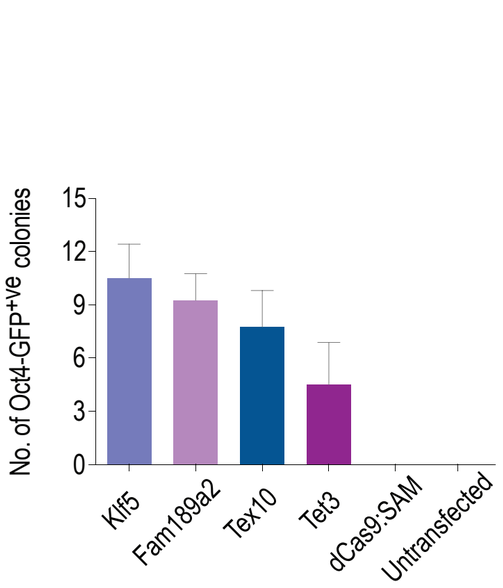


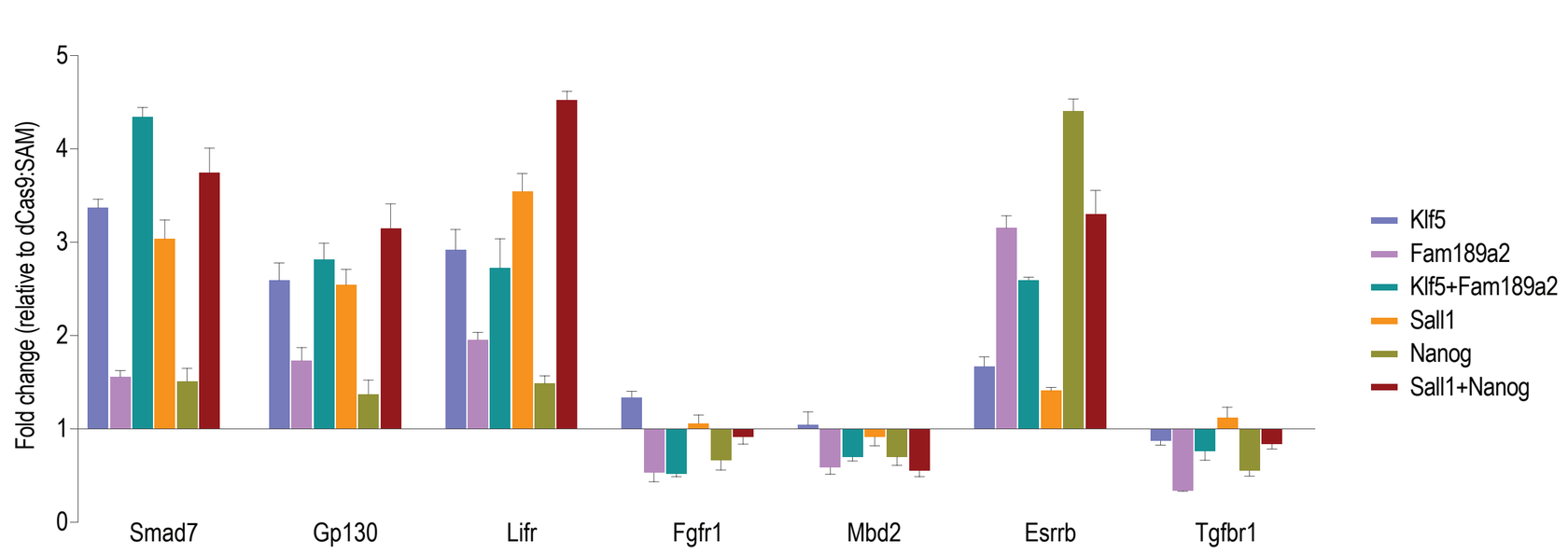
FIGURE 4

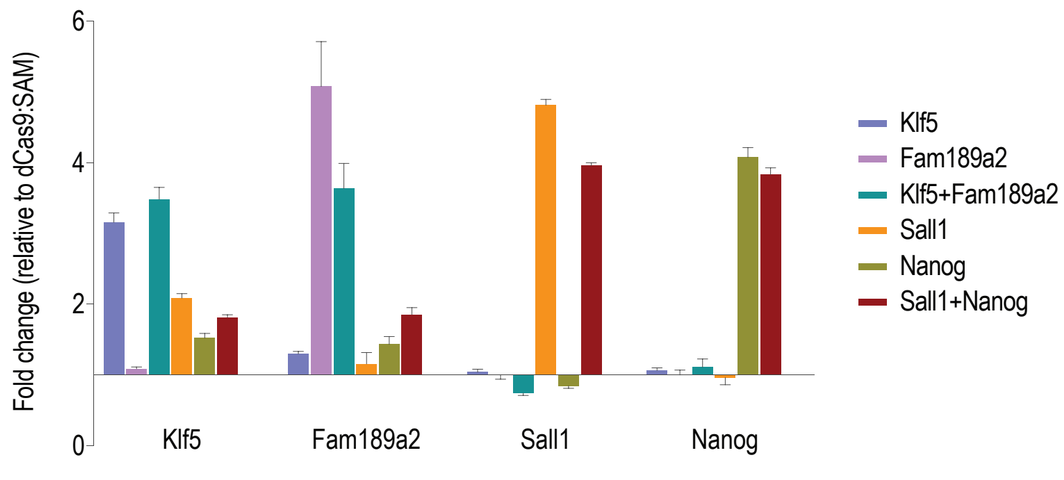

a
Sall1
Nanog
372 genes
(29.7%)
568 genes
(45.6%)
307 genes
(24.7%)
Klf5 (Sall1)
Fam189a2 (Nanog)
Myc
Mycn
Smad7
Lifr
Jarid2
Mbd2
Fgfr1

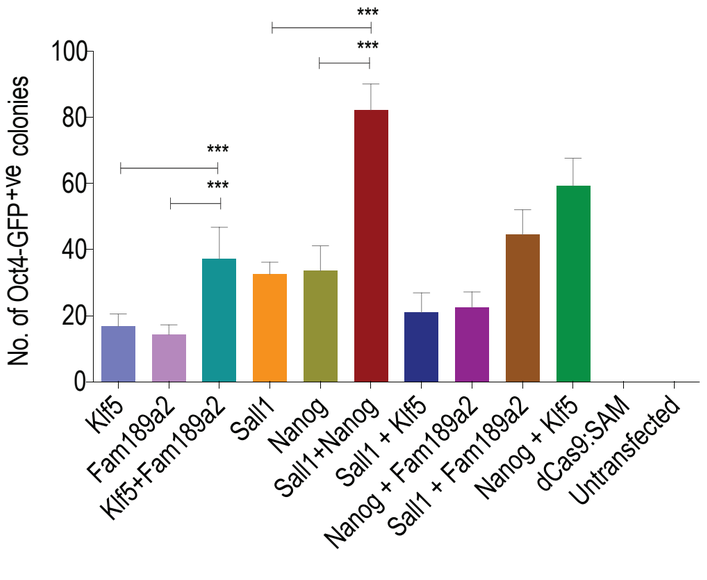

b
positive regulation of protein kinase B signaling cascade (GO:0051897)
patterning of blood vessels (GO:0001569)
tube morphogenesis (GO:0035239)
negative regulation of gene expression (GO:0010629)
negative regulation of cellular biosynthetic process (GO:0031327)
positive regulation of macromolecule biosynthetic process (GO:0010557)
regulation of signal transduction (GO:0009966)
regulation of cell communication (GO:0010646)
organ morphogenesis (GO:0009887)
negative regulation of biological process (GO:0048519)
organ development (GO:0048513)
developmental process (GO:0032502)
neurological system process (GO:0050877)
developmental process (GO:0032502)
multicellular organismal development (GO:0007275)
intracellular signaling cascade (GO:0007242)
organ morphogenesis (GO:0009887)
cell proliferation (GO:0008283)
regulation of cell proliferation (GO:0042127)
positive regulation of gene expression (GO:0010628)
tissue development (GO:0009888)
epithelium development (GO:0060429)
gland development (GO:0048732)
respiratory tube development (GO:0030323)
cardiac muscle tissue morphogenesis (GO:0055008)
positive regulation of cell size (GO:0045793)
p value 0.00005 - 0.000005
p value 0.0005 - 0.00005
p value 0.005 - 0.0005
p value 0.05 - 0.005
0
20
40
60
80
Fold change


c
Fold change (relative to PBCAG:Empty)
4
3
2
1
0
Klf5
Fam189a2
Tet3
Tex10
Lifr
Smad7
Myc
Mycn
Arid2
Jarid2
Mdb2
Fgfr1
Sall1
Nanog
Sall1+Nanog

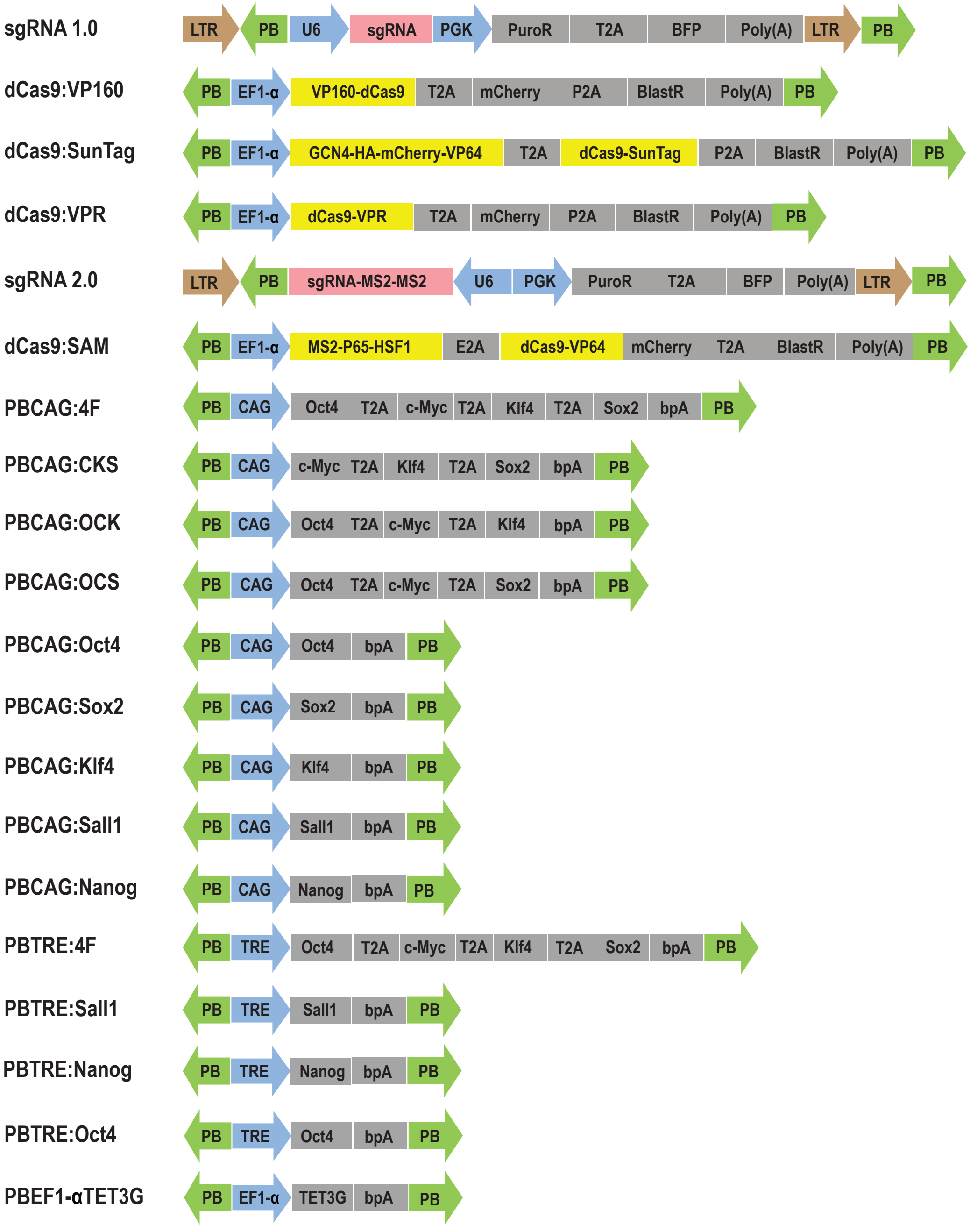

d
No. of Oct4-GFP+ve colonies
15
12
9
6
3
0
Klf5
Fam189a2
Tex10
Tet3
dCas9:SAM
Untransfected


f
Fold change (relative to dCas9:SAM)
5
4
3
2
1
0
Smad7
Gp130
Lifr
Fgfr1
Mbd2
Esrrb
Tgfbr1
Klf5
Fam189a2
Klf5+Fam189a2
Sall1
Nanog
Sall1+Nanog

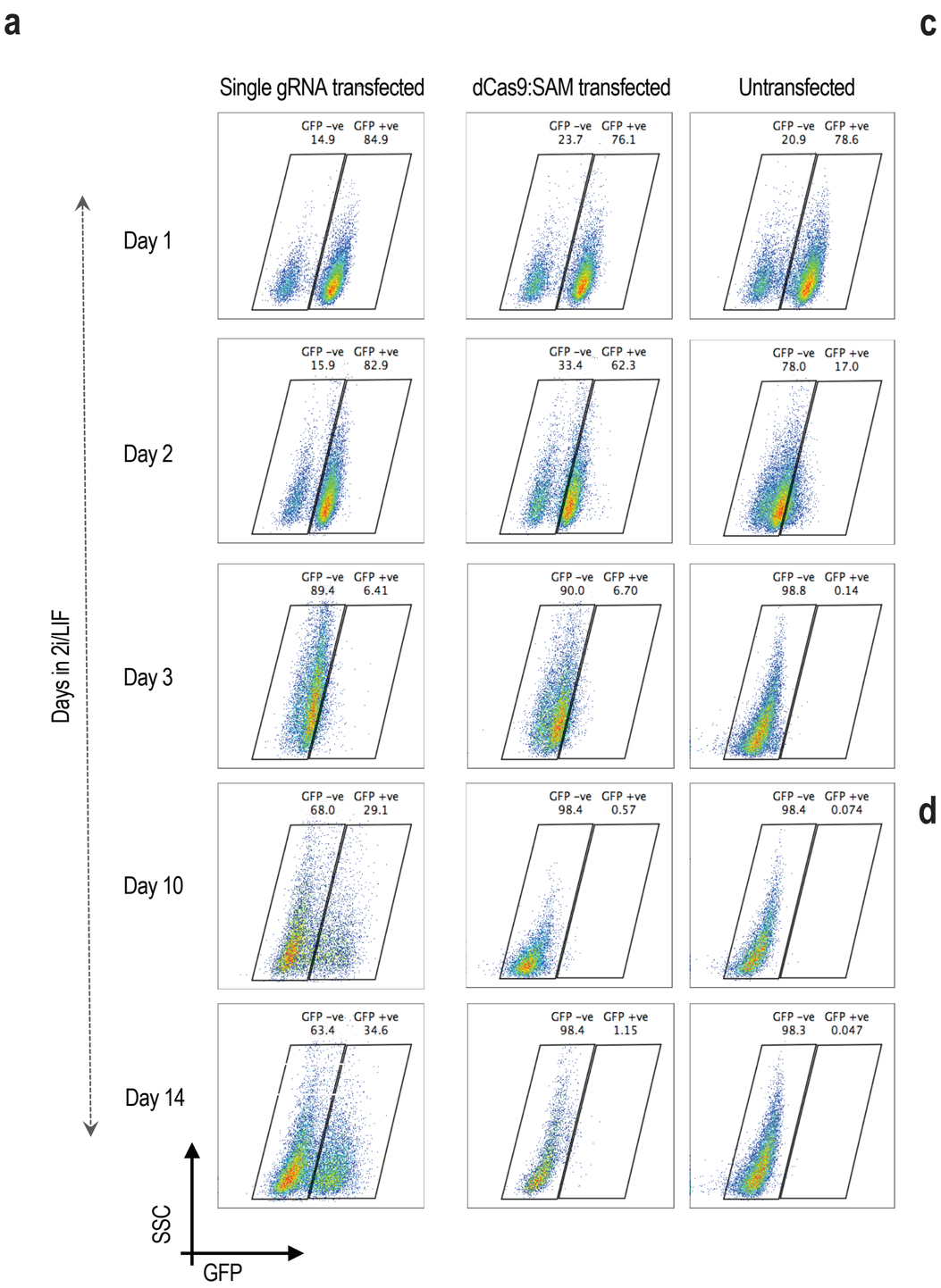

e
Fold change (relative to dCas9:SAM)
6
4
2
0
Klf5
Fam189a2
Sall1
Nanog
Klf5
Fam189a2
Klf5+Fam189a2
Sall1
Nanog
Sall1+Nanog

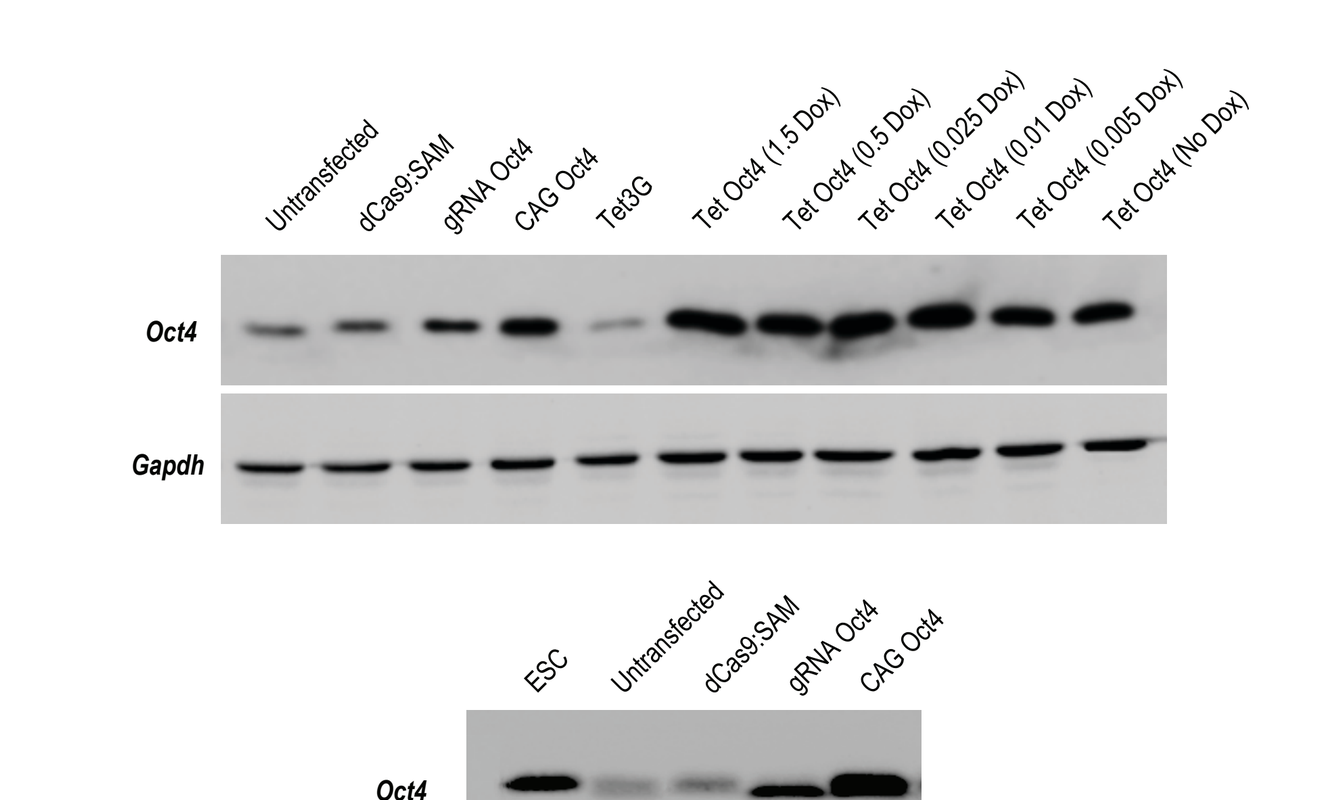

g
No. of Oct4-GFP+ve colonies
100
80
60
40
20
0
***
***
***
***
Klf5
Fam189a2
Klf5+Fam189a2
Sall1
Nanog
Sall1+Nanog
Sall1 + Klf5
Nanog + Fam189a2
Sall1 + Fam189a2
Nanog + Klf5
dCas9:SAM
Untransfected

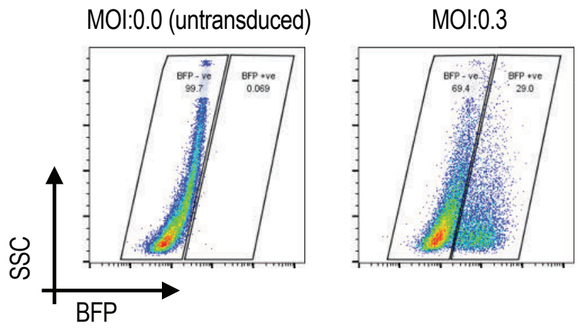
SUPPLEMENTARY FIGURE 1

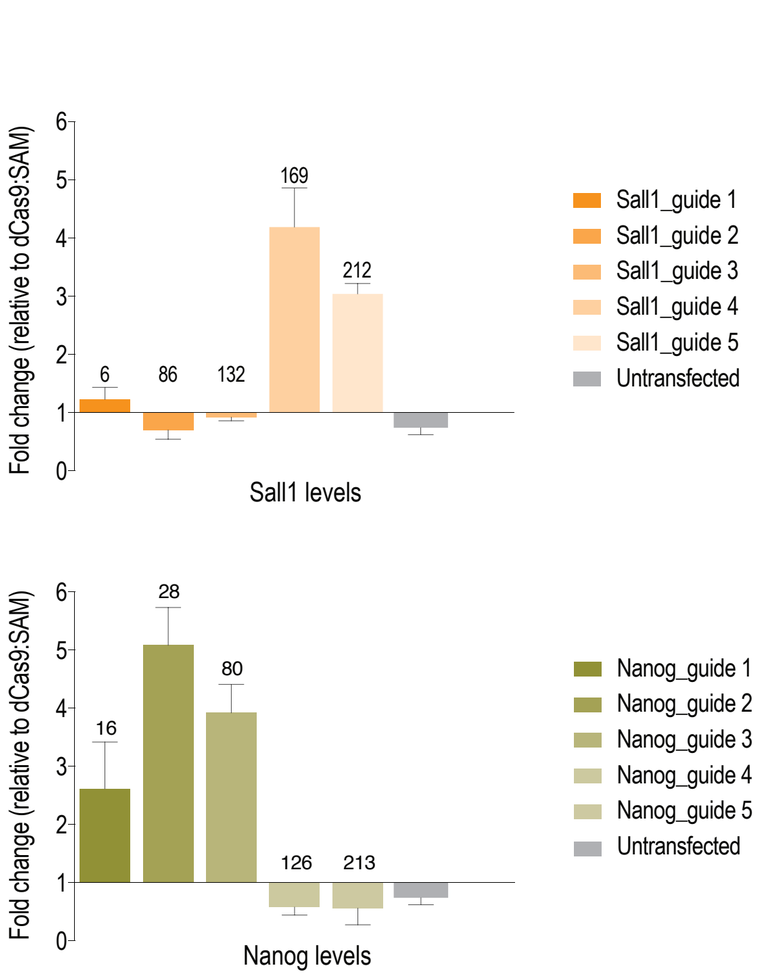


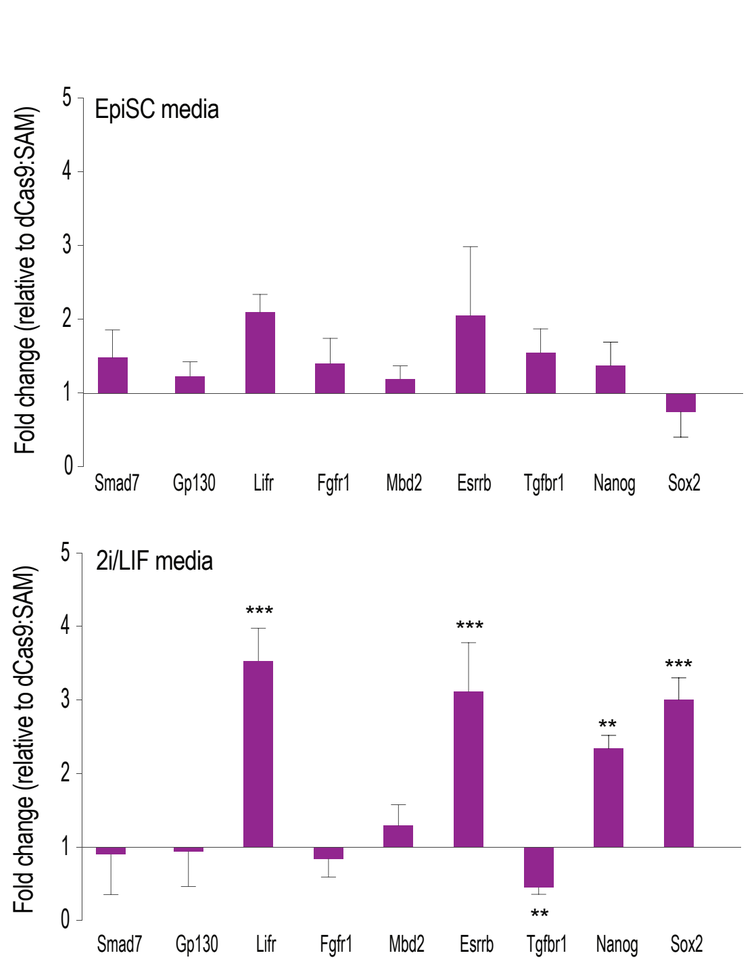
SUPPLEMENTARY FIGURE 2

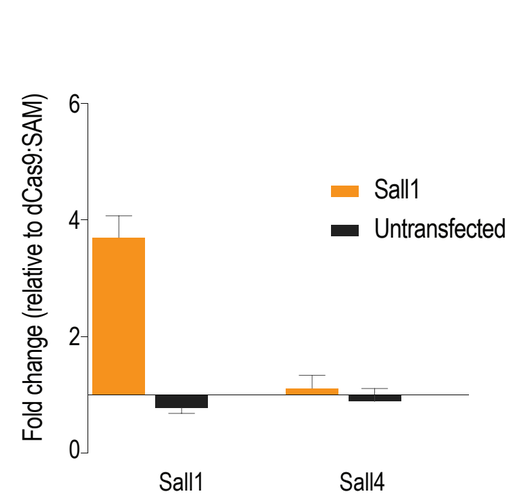








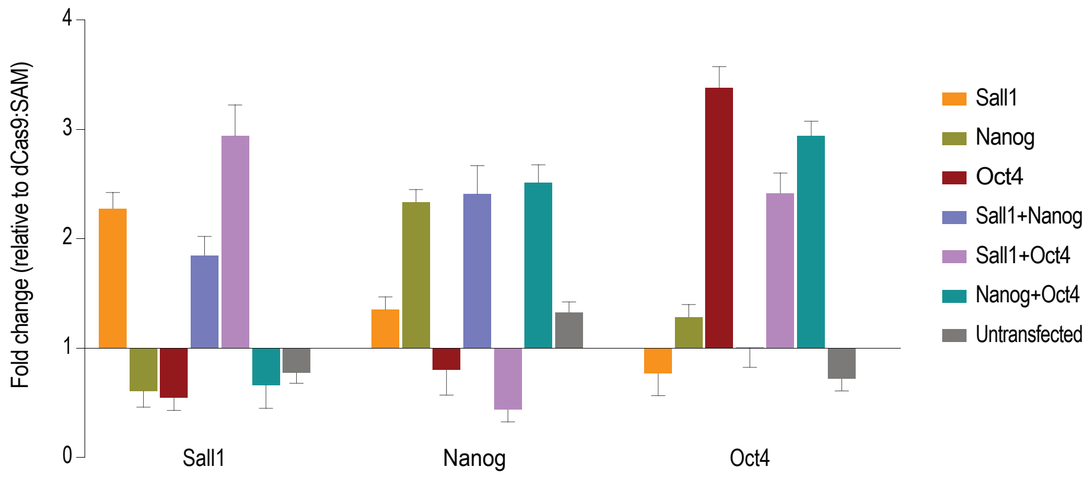




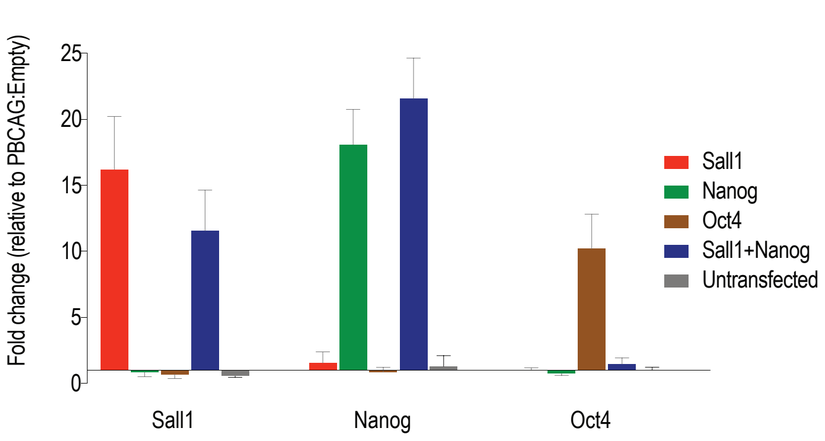
SUPPLEMENTARY FIGURE 3

a

CRISPRa transfected cells

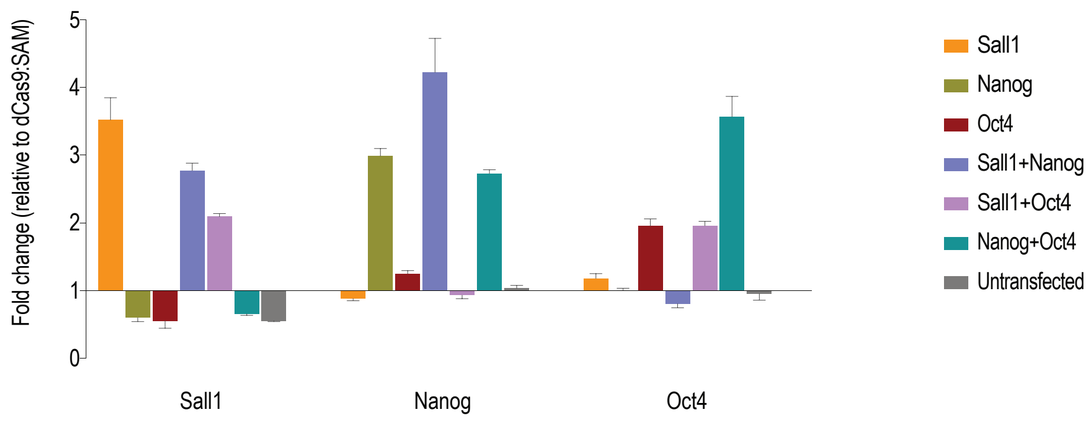


cDNA transfected cells

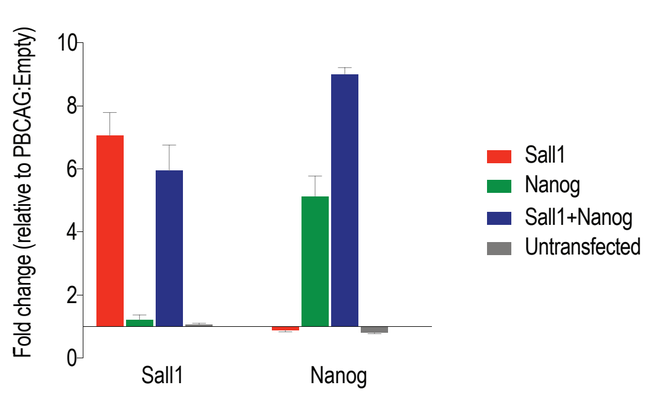


b

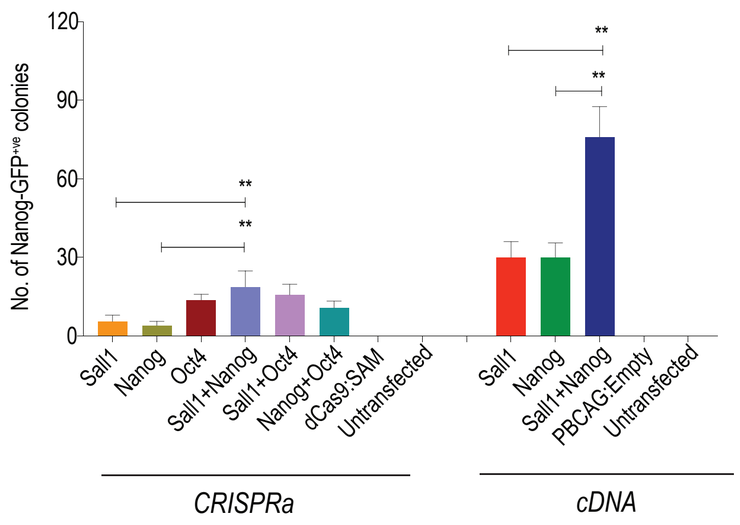
CRISPRa transfected cells



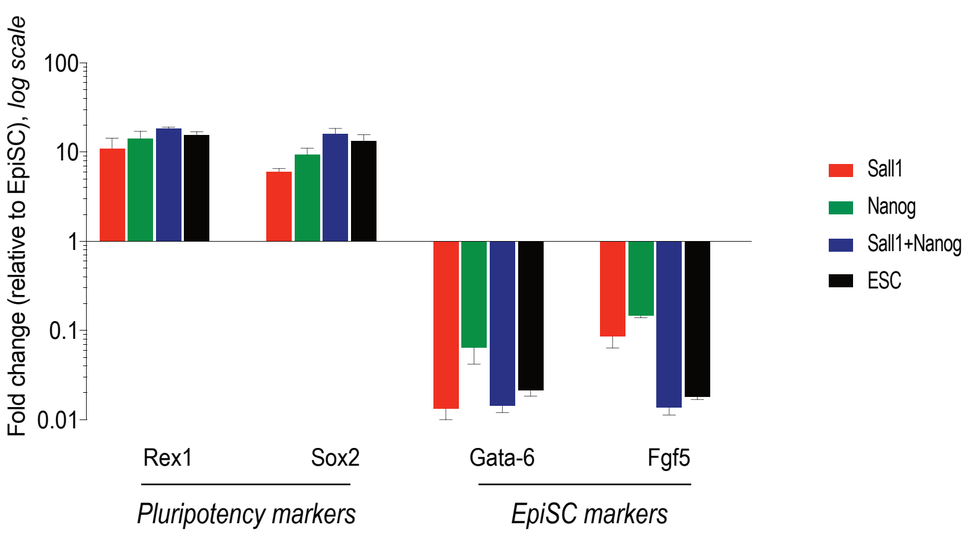
cDNA transfected cells



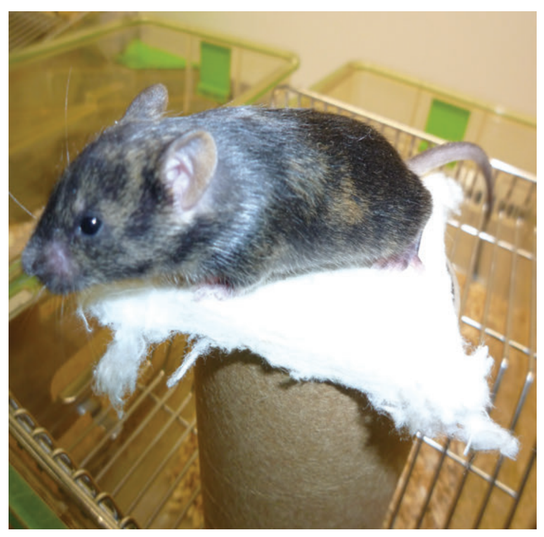
c



d

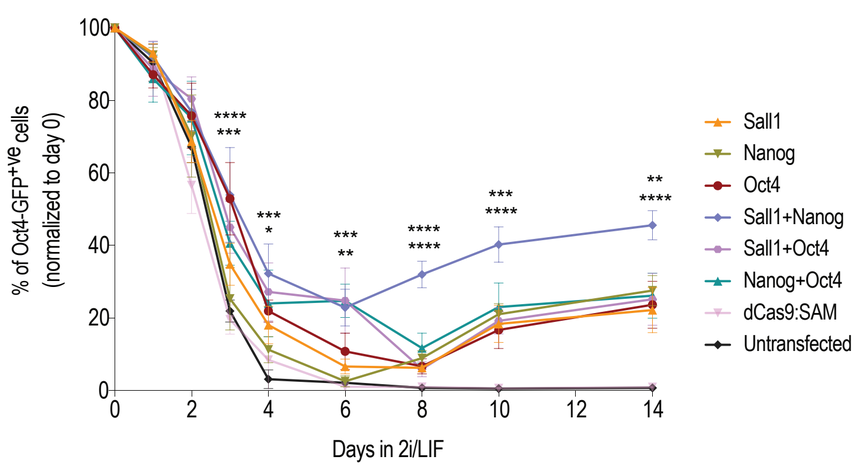


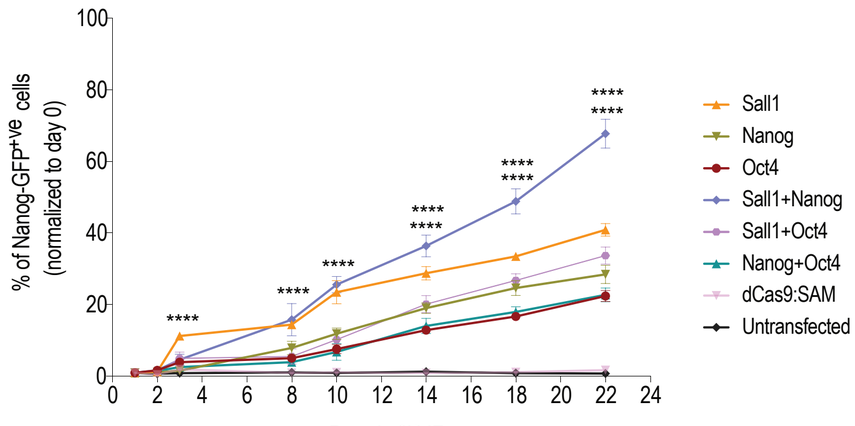
e



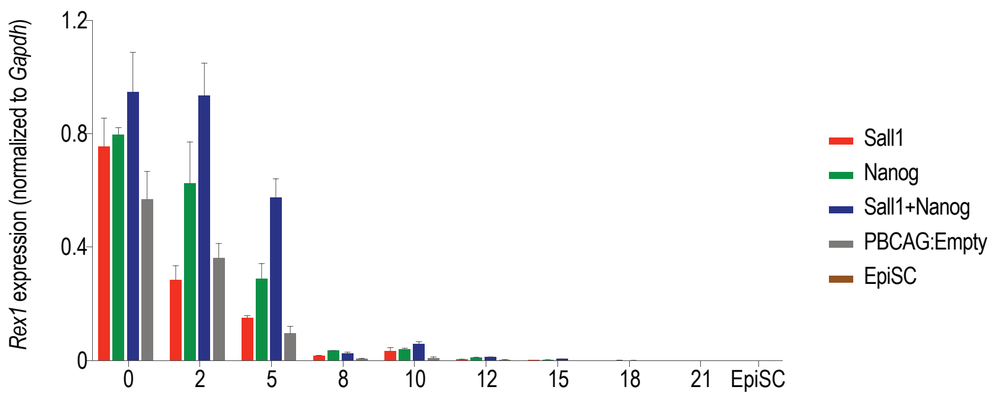
cDNA Sall1 iPSC chimeric mouse

f

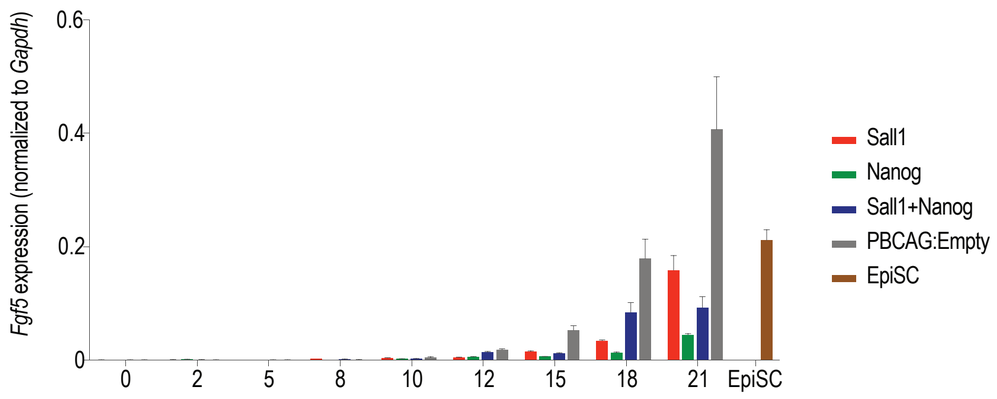




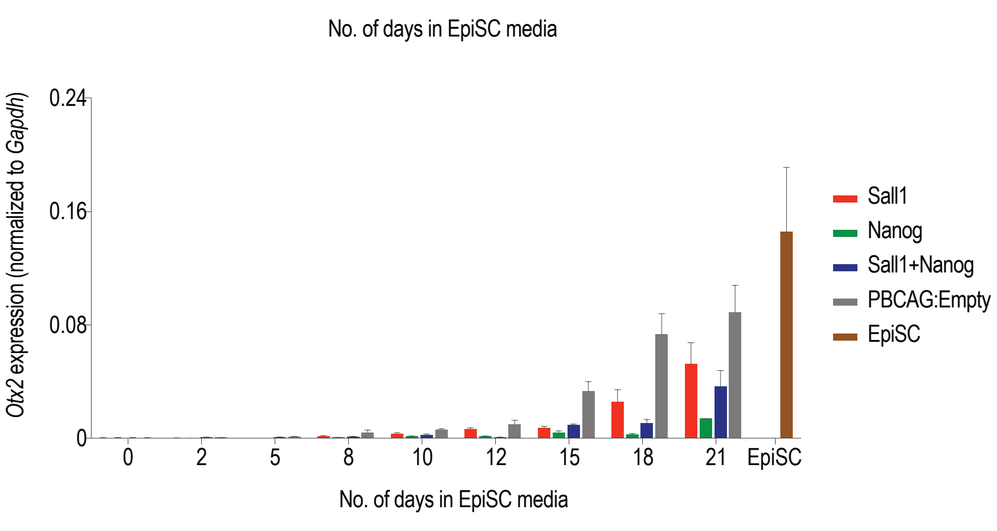
g



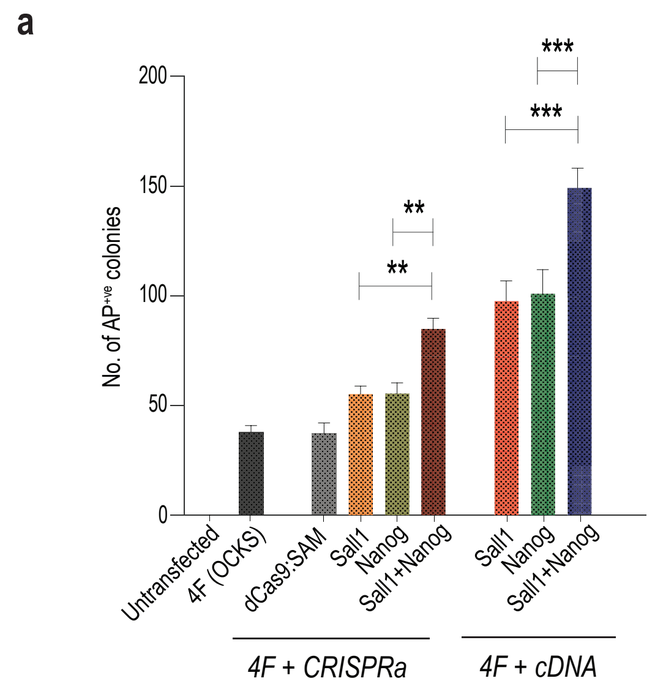
h

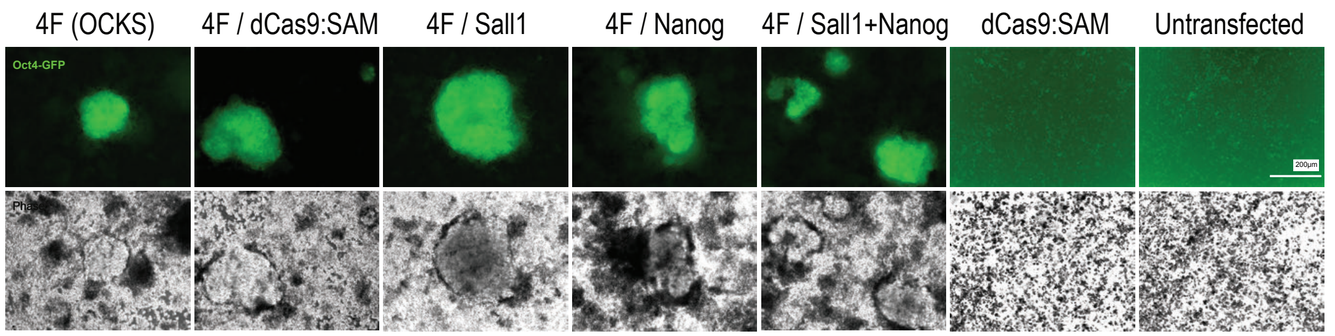


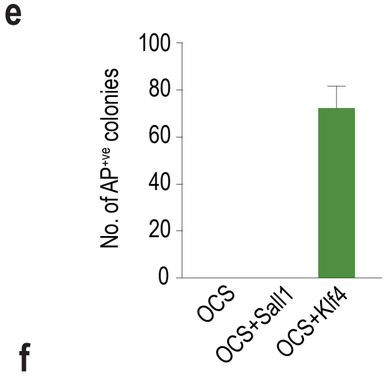
i

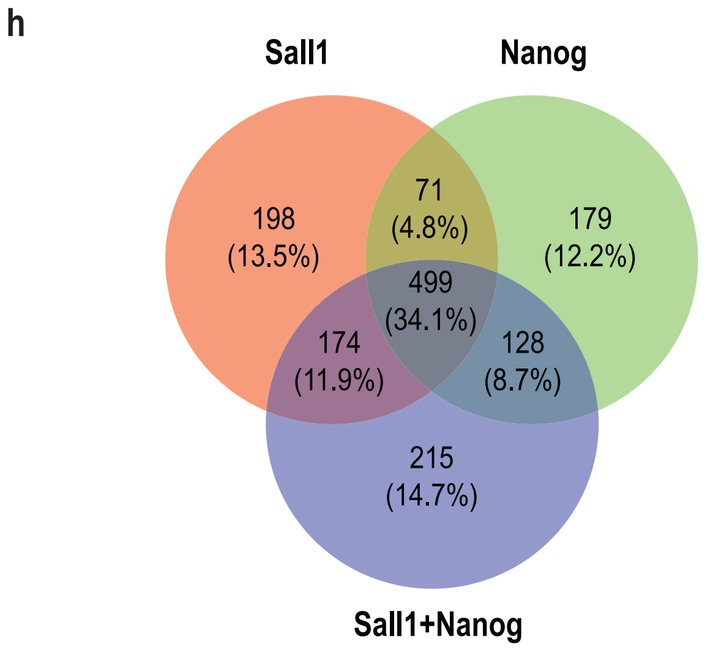


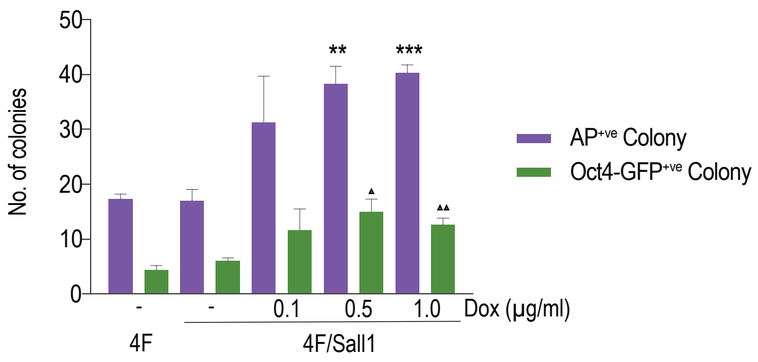
SUPPLEMENTARY FIGURE 4

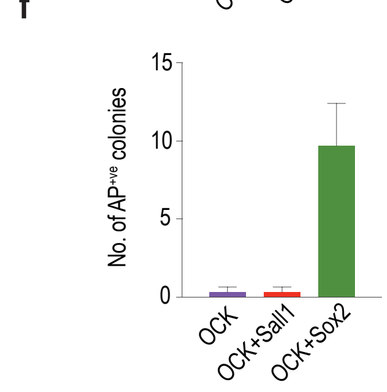


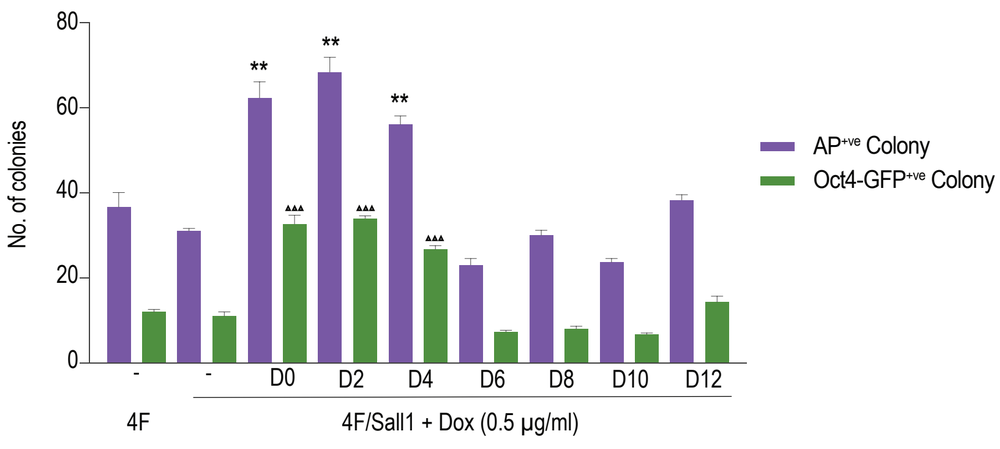


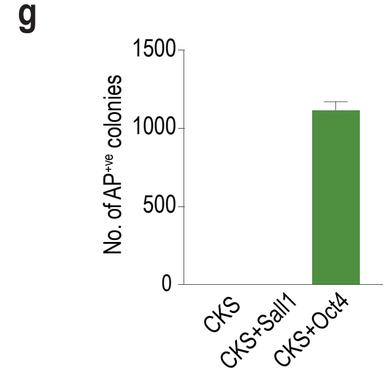


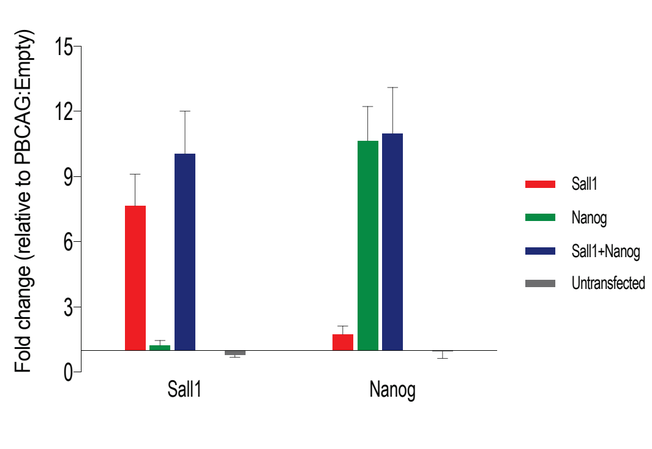












Supplementary Table 1

Primers for confirming *PiggyBac* mediated Gene integration (Genotyping)

Gene Name	Primer Name	Primer Sequence
Sall1-BPA	Sall1-BPA F	CAATCCTGTCAAGTTCCCAGAAAT
	Sall1-BPA R	CATCCCCAGCATGCCTGCTATT
Nanog-BPA	Nanog-BPA F	AGGGCTATCTGGTGAACGCATC
	Nanog-BPA R	AATCCTCCCCCTTGCTGTCCT
Oct4-c-Myc	Oct4-c-Myc F	GCCCCCAGGTCCCCACTTTG
	Oct4-c-Myc R	CCAGCTGATCGGCGGTGGAG
Klf4-Sox2	Klf4-Sox2 F	ACTATGCAGGCTGTGGCAA
	Klf4-Sox2 R	TTGCTGCGGGCCCGGCGGCT
Tet3G	Tet3G F	CCGTCCAGGCACCTCGATTAGTTC
	Tet3G R	GGTATGACTTGGCGTTGTTCC
Actb	Actb F	GTTTGAGACCTTCAACACCCC
	Actb R	GTGGCCATCTCCTGCTCGAAGTC
gRNA	pKLV_Flip_gRNA F	AGCAAAAAAAGCACCGACTCG
	pKLV_Flip_gRNA R	TAAAGCGCATGCTCCAGACTGC
dCas9:SAM	SamCas9 F	TTACTCAGTTCGTGCTCGTGGAC
	SamCas9 R	ATTGCCTTCACGATGAGTTCACA

Supplementary Table 2

Combinations of cDNA transfection used for MEF reprogramming

Constructs	Combinations (amount of DNA in μg)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
PBCAG4F	1.0	1.0													
PBCAGCKS							1.0			1.0			1.0		
PBCAGOCK								1.0			1.0			1.0	
PBCAGOCS									1.0			1.0			1.0
PBCAGOct4										1.0					
PBCAGKlf4												1.0			
PBCAGSox2											0.5				
PBTRE4F			1.0	1.0	1.0	1.0									
PBTRESall1		0.5		0.5		0.5							0.5	0.5	0.5
PBTRENanog					0.1	0.1									
PBEF1-αTET3G		1.0	1.0	1.0	1.0	1.0							1.0	1.0	1.0
PiggyBac Transposase	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0

Supplementary Table 3

Mouse RT-qPCR Probes for RNA Expression

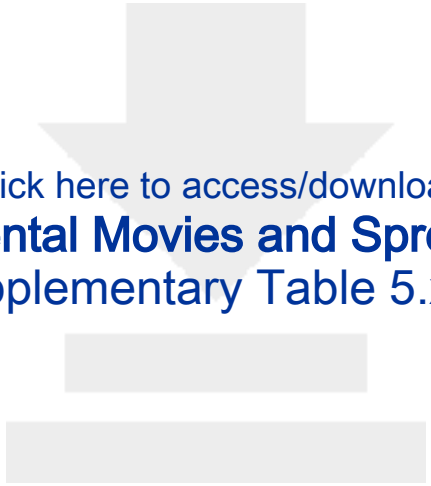
Gene Name	Applied Biosystems Catalogue Number
Esrrb	Mm00442411_m1
Fgf5	Mm00438615_m1
Gapdh	4352339E
Gata6	Mm00802636_m1
Id3	Mm00492575_m1
Klf4	Mm00516104_m1
Nanog	Mm02384862_g1
Pou5f1	Mm00658129_gH
Rex1	Mm03053975_g
Sox2	Mm03053810_s1
Id2	Mm00711781_m1
Mdb2	Mm00521967_m1
Jarid2	Mm00445574_m1
Tet3	Mm00805756_m1
Tex10	Mm06549480_m1
Klf5	Mm00456521_m1
Smad7	Mm00484742_m1
Gp130	Mm00439665_m1
Sall1	Mm00491266_m1
Otx2	Mm0046859_m1
Fgfr1	Mm00438930_m1
Fam189a2	Mm01194369_m1
Lifr	Mm00442942_m1
Myc	Mm00487804_m1
Mycn	Mm00476449_m1
Arid2	Mm00558381_m1

Supplementary Table 4

Antibodies for Western Blotting / Immunofluorescence staining

Antibody	Company	Catalogue Number	Dilution
SSEA-1 Clone MC480	BD Pharmingen	560079	1:200
Nanog	Abcam	Ab80892	1:150
β III Tubulin (Tuj1)	R & D Systems	MAB1159	1:150
α-Smooth Muscle Actin	R & D Systems	MAB1420	1:150
α-Fetoprotein	R & D Systems	MAB1368	1:150
H3K27me3	Millipore	07-449	1:1000
Oct4 (C10)	Santa-Cruz	SC-5279	1:150 (IF) 1:800 (WB)
Gapdh	Sigma	G8795	1:4000

Supplementary Table 5

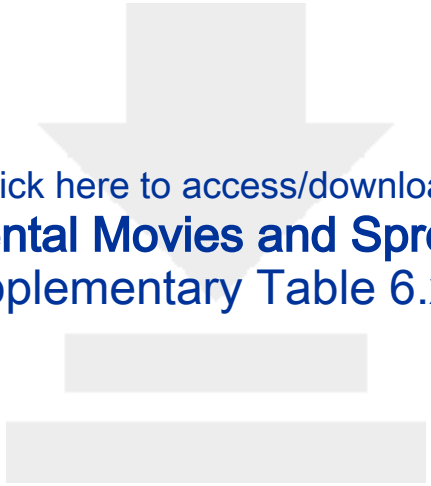


Click here to access/download

Supplemental Movies and Spreadsheets

Supplementary Table 5.xlsx

Supplementary Table 6

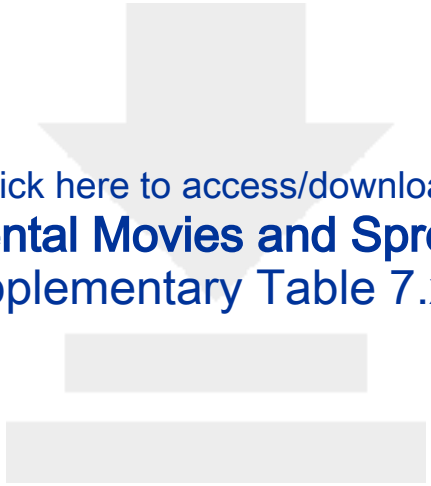


Click here to access/download

Supplemental Movies and Spreadsheets

Supplementary Table 6.xlsx

Supplementary Table 7

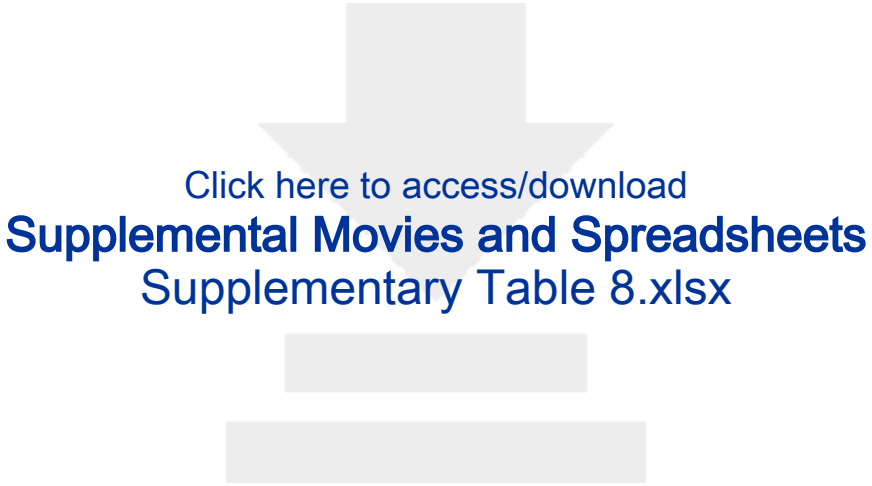


Click here to access/download

Supplemental Movies and Spreadsheets

Supplementary Table 7.xlsx

Supplementary Table 8



Click here to access/download

Supplemental Movies and Spreadsheets

Supplementary Table 8.xlsx