Nanog-dependent function of Tet1 and Tet2 in establishment of pluripotency

Yael Costa[1]*, Junjun Ding[2]*, Thorold W. Theunissen[1,3]*, Francesco Faiola[2]*, Timothy A. Hore[4], Pavel V. Shliaha[5], Miguel Fidalgo[2], Arven Saunders[2], Moyra Lawrence[1,3], Sabine Dietmann[1], Satyabrata Das[6], Dana N. Levasseur[6], Zhe Li[7], Mingjiang Xu[7], Wolf Reik[4,8], José C. R. Silva[1,3#], Jianlong Wang[2#]

Supplemental Information

Contents

Supplementary Methods

General experimental procedures

Generation of ^{FLBIO}Nanog expressing ES cells

To increase the level of functional ^{FLBIO}Nanog for efficient affinity purification, we set up the in vivo biotinylation of Nanog in a Nanog conditional knockout cell line (*Nanog*$^{flox/flox}$)[1]. In these cells, the endogenous *Nanog* alleles have been sequentially targeted by homologous recombination and replaced by floxed Nanog alleles. A Flag- and biotin-tagged Nanog transgene was then introduced into these cells by electroporation followed by selection with puromycin (1 μg/mL). GFP-Cre was used to remove floxed *Nanog* and to create *Nanog*$^{-/-}$ alleles to establish the stable ES cell line NGA2 expressing a ^{FLBIO}Nanog transgene in a *Nanog*$^{-/-}$ background. The BirA-V5 transgene was then introduced into NGA2 ES cells followed by G418 (350 μg/mL) selection and several clones (NGB ES lines) were established (see Fig. 1). Expression of ^{FLBIO}Nanog in these lines was confirmed by western blotting (WB) using an anti-Nanog antibody (Millipore) and streptavidin–HRP (GE Healthcare) (see Supplementary Fig. 1). Standard procedures for ES cell culture, western blotting, and lentivirus infection were followed and have been described elsewhere[2-3].

Nuclear extract preparation from ES cells

ES cells containing ^{FLBIO}Nanog with (NGB) and without (NGA) the BirA-V5 transgene were expanded to five large square dishes (245 x 245 mm), washed with PBS, and scraped off to prepare nuclear extracts as previously described[4-6]. Briefly, cells were first resuspended in Buffer A [10 mM HEPES (pH 7.6), 1.5 mM $MgCl_2$, 10 mM KCl] to remove cytoplasmic proteins and then in Buffer C [20 mM HEPES (pH 7.6), 25% glycerol (v/v), 0.42 M NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA] to obtain nuclear extracts. The salt concentration was subsequently decreased to 100 mM by dialyzing to Buffer D [20 mM HEPES (pH 7.6), 0.2 mM EDTA, 1.5 mM $MgCl_2$, 100 mM KCl, 20% glycerol] at 4^{o}C for 3 hrs, and precipitated proteins were removed by centrifugation. Freshly prepared nuclear extracts were then subjected to affinity purification as described below. All buffers were supplemented with PMSF and a protease inhibitor cocktail (SIGMA) before use.

Affinity purification of Nanog protein complexes

Three independent affinity purification (AP) approaches were employed to isolate Nanog protein complexes for MS identification as described below.

First, streptavidin (SA) agarose affinity purification was performed as described[3-4,7-8] with modifications. Briefly, nuclear extracts from five large square dishes (245 x 245 mm) were pre-cleared with 0.5 mL of Protein G agarose beads (Roche Diagnostics) for 1 hr at 4°C in the presence of 750 units of Benzonase (Novagen), incubated with 0.5 mL of SA agarose beads (Invitrogen), and rotated for 6 hrs at 4°C. Subsequently, captured complexes were subjected to five 15-min washes in Buffer D [20 mM HEPES (pH 7.6), 0.2 mM EDTA, 1.5 mM $MgCl_2$, 100 mM KCl, 20% glycerol] supplemented with 0.02% NP-40, and bound material was eluted by boiling for 5 min in Laemmli sample buffer. Samples were then fractionated on a 10% SDS-polyacrylamide gel, stained with GelCode™ Blue Safe Protein Stain buffer (Thermo), and subjected to whole lane LC–MS/MS sequencing and data analysis. Experiments were performed on three NGB ES cell lines. NGA2 ES cells were used as a control AP sample (Fig. 1a).

Second, Flag agarose (M2, Sigma) affinity purification was performed as described above for SA agarose purification with the following modifications. Pre-cleared nuclear extracts were incubated with 500 µL of α-Flag M2 agarose beads, and immuno-complexes were then eluted four times for 1 h each at 4°C with 0.3 mg/mL Flag peptide in Buffer D (supplemented with 0.02% NP-40). Flag AP was performed on NGB19 ES cells. $Nanog^{flox/flox}$ ES cells were used as a control AP sample.

Third, for affinity purification of endogenous Nanog-interacting proteins, nuclear extracts from two large square dishes of J1 ES cells were prepared as described above. Prior to immunoprecipitation, 50 µg of Nanog (Bethyl Laboratories, cat#A300-397A) or control IgG (Millipore, cat#PP64) antibodies were pre-bound to, and nuclear extracts pre-cleared with, 200 µL of protein G agarose beads (Roche, cat#11243233001) overnight at 4°C with gentle rotation. Afterwards, pre-bound antibodies were incubated with pre-cleared nuclear extracts for 4 hrs at 4°C with gentle rotation. Immunoprecipitates were then washed four times, eluted from the beads, concentrated, and separated by SDS-PAGE. Finally, whole lanes were excised from the gel and subjected to LC-MS/MS mass spectrometry analysis.

Streptavidin (SA) pull-down, co-immunoprecipitation and western blot analysis

Pull-down using SA beads was performed as described for affinity purification and mass spectrometry in the main text, except that the eluate from SA capture was used for western blot analysis. To validate Tet1-Nanog interaction in ES cells, nuclear extracts were prepared from NGB or J1 ES cells and adjusted to 150 mM salt by dialysis. Endogenous Tet1 was then

immunoprecipitated with 20 μg of pre-bound Tet1 antibody (Millipore, 09-872), and co-immunoprecipitated Nanog was identified by western blotting with Streptavidin-HRP (NGB cells) or anti-Nanog (J1 ES cells) antibodies. For the control IP, we used 20 μg of rabbit IgG beads (Millipore, PP64). Additional antibodies used in western blot and coIP/IP analyses are listed below: anti-Tet1 (Millipore 09-872), anti-Nanog (Millipore AB5731), anti-Oct4 (Santa Cruz sc5279), anti-Sox2 (Santa Cruz sc17320), anti-β-Actin (Sigma A5441), anti-Tubulin (Abcam ab6064-100), anti-Mta1/2 (Bethyl A300-911A), anti-Hdac2 (Bethyl A300-705A), anti-Bptf (Santa Cruz sc-98404), anti-Brca2 (Santa Cruz sc-1819), anti-Emsy (Santa Cruz sc-34995), anti-Sgol2 (Santa Cruz sc-161222), and anti-Zfp609 (Santa Cruz sc-132181). For coIP in HEK293T cells, we transiently co-transfected cells with plasmids expressing Flag/biotin (FLBIO)-tagged Tet1 or Tet2 and V5his-tagged Nanog. The BirA-V5his construct was used as a negative control. Two days after transfection, total lysates or nuclear extracts were prepared and incubated with anti-M2 Flag agarose overnight. On day 2, unbound material was washed away, and bound material eluted by boiling in Laemmli buffer and subjected to western blot analyses.

Immunodepletion of Tet1 in ES cells

Two milligrams (mg) of ES cell nuclear extracts were used for sequential IPs with 3 μg of anti-Tet1 antibody conjugated to Protein G agarose beads (Roche Diagnostics) (100 μL beads per 3 μg antibody). Each IP was performed at 4ºC for 6 hrs. Beads were washed six times with 900 μL buffer [20 mM HEPES (pH 7.9), 25% glycerol (v/v), 0.15 M NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.02% NP-40, and 0.5% BSA] per wash, and eluted three times by boiling with 30 μL each of Laemmli sample buffer. One third (30 μL) of each IP sample was subjected to western blotting with the indicated antibodies. Ten percent of nuclear extract (NE) input and final non-bound NE were also loaded for positive and negative controls, respectively.

Effect of Nanog and Tet1 suppression in iPS cell generation from MEFs

For lentiviral production, STEMCCA plasmid[9], pLKO.1-puro empty vector, and pLKO.1-shRNA plasmids targeting Nanog and Tet1 were co-transfected with packaging vectors into HEK293T cells. Viral supernatants were harvested after 48 hrs and virus titer was determined by QuickTiter Lentivirus Kit (Cell Biolabs).

Functional studies of Nanog and Tet1 suppression by shRNAs in reprogramming were performed according to a published procedure with modifications[10]. Briefly, *Oct4*-GFP

mouse embryonic fibroblasts (MEFs) were seeded at 5 x 10^4 cells per well of a 6-well plate. The next day, MEFs were incubated with concentrated lentiviruses containing four reprogramming factors (4F) and shRNAs against Nanog or Tet1 in the presence of 6 μg/mL polybrene for 16 hrs. One day after transduction, cultures were switched to ES cell medium supplemented with LIF. Three days after transduction, MEFs were counted and 5000 cells were then re-seeded onto 6-well plates with irradiated, puromycin-resistant DR4 feeder cells. Infected cells were selected with puromycin (1.5 μg/mL) at day five after transduction and selection was maintained for 10 days. iPS cell colonies were scored 16 days after transduction and stained for alkaline phosphatase (AP) activity using a commercial kit (Sigma). Flow cytometric analysis for *Oct4*-GFP fluorescence in reprogrammed cells was performed on an LSR-II Flow Cytometer System (BD Biosciences).

Effect of Tet1 or Tet2 suppression in iPS cell generation from neural stem (NS) cells

Clonal lines of reprogramming intermediates transgenic for PB-flox-Nanog-Pgk-Hygro or PB-flox-Nanog-Pgk-Hygro and PB-flox-HsaTet1Mut-IRES-Blast were generated as described in the Methods accompanying the main text. After selection, 2 x 10^4 cells were seeded per well in a 12-well plate in serum/LIF medium and transfected in suspension with siTet1 or siTet2 or a negative non-targeting control siRNA (siNT) (FlexiTube GeneSolution and AllStars Negative control, QIAGEN) using Lipofectamine RNAiMAX Reagent (Life Technologies). Medium was switched to N2B27/2i/LIF[11] the following day and the cells were transfected a second time with the respective siRNAs after two more days. Puromycin selection for an *Oct4*-GFP-IRES-puro reporter transgene was applied from day 6 of 2i/LIF treatment. GFP-positive colonies were scored at day 10.

Reprogramming assays in adult neural stem (NS) cells

To investigate the consequences of Nanog and Tet1 co-expression during reprogramming, adult NS cells were infected with pMX-based retroviral reprogramming factors Oct4, Klf4 and c-Myc (OKM)[12]. Cultures were switched to ES cell medium (serum/LIF) at day 3 post-transduction. A clonal line of NS-derived proliferative, *Oct4*-GFP negative cells (reprogramming intermediates) was transfected using nucleofection (Amaxa) with various combinations of the following transgenes: 1 μg of PB-flox-Nanog-Pgk-Hygromycin, PB-flox-HsaTet1-IRES-Blasticidin, PB-flox-HsaTet1Mutant-IRES-Blasticidin, PB-flox-Empty-Pgk-Hygromycin or PB-flox-Empty-IRES-Blasticidin plus 2 μg of the PBase expression vector pCAGPBase[13]. Dual hygromycin and blasticidin selection was applied to transfectants

for a minimum of 12 days and maintained until medium switch to 2i/LIF. Stable transgene expression was confirmed by qRT-PCR. In three independent experiments, $1.0x10^3$, $5.0x10^4$, or $2.5x10^5$ stable NS-derived transfectants were seeded per well in a 6-well plate in serum/LIF medium. After 1 ($5.0x10^4$), 2 ($2.5x10^5$), or 4 ($1.0x10^3$) days, medium was switched to 2i/LIF. Puromycin selection for an *Oct4*-GFP-IRES-puro reporter transgene was applied from day 6 of 2i/LIF treatment. GFP-positive colonies were scored at day 10.

Reprogramming assays in mouse embryonic fibroblasts (MEFs)

Nanog-GFP MEFs were infected in ES medium (serum/LIF) with pMX-based retroviral reprogramming factors (Oct4, Klf4, cMyc, and Sox2; OKMS)[12]. They were then grown for seven days before a clonal line of proliferative, *Nanog*-GFP-negative cells (reprogramming intermediates) was transfected using nucleofection (Amaxa) with various combinations of 1 μg of PB-flox-Nanog-Pgk-Hygro, PB-flox-HsaTet1-IRES-Blast, PB-flox-Empty-Pgk-Hygro or PB-flox-Empty-IRES-Blast plus 2 μg of the PBase expression vector, pCAG-PBase[13]. Dual hygromycin and blasticidin selection was applied to transfectants for a minimum of 12 days and maintained until medium switch to 2i/LIF. Stable transgene expression was confirmed by qRT-PCR. In two independent experiments, 1 x 10^4 stable transfectants were seeded per well in a 6-well plate in either serum/LIF (experiment 1) or directly into 2i/LIF medium (experiment 2). In the case of experiment 1, medium was switched to 2i/LIF after 1 day. Puromycin selection for a *Nanog*-GFP-IRES-puro reporter transgene was applied from day 6 of 2i/LIF treatment. GFP-positive colonies were scored at day 10.

Heterokaryon-based fusion reprogramming

Heterokaryons were generated by fusing mouse ES cells and human B (hB) lymphocytes following a published protocol[14]. Briefly, the *Tet1*$^{+/+}$, *Tet1*$^{-/-}$ or *Tet1*$^{-/-}$ Rescue (*Tet1*$^{-/-}$R) ES cells and hB lymphocytes were mixed in a 1:1 ratio and fused with 50% polyethylene glycol (PEG-1500; Roche Diagnostics). Non-fused hB cells were eliminated by addition of Ouabain (10^{-5} M; Sigma) 6 hrs after fusion. To eliminate proliferating mouse ES cells, Ara-C (cytosine β-D arabino furanoside; 10^{-5} M; Sigma) was added to the medium 6 hrs after cell fusion and then removed after 16 hrs. The reprogramming of hB cells was monitored for three days by quantitative real-time PCR analyses of human-specific gene expression. The sequences of human gene-specific primers are provided in Supplementary Table 5.

Chromatin immunoprecipitation (ChIP) coupled with quantitative real-time PCR (ChIP-qPCR)

ChIP was performed as previously described[15]. Primer sequences were designed according to overlapping peaks as previously published[16-17] and are provided in Supplementary Table 5. Real-time PCR was run with a LightCycler 480 (Roche) instrument with LightCycler DNA master SYBR Green I reagents. Differences between samples and controls were calculated based on the $2^{-\Delta\Delta CT}$ method and normalized to GAPDH. Measurements were performed in duplicate or in triplicate.

qRT-PCR analysis

Total RNA was extracted using the RNeasy kit (Qiagen) and cDNA was generated using Superscript III (Invitrogen). Relative expression levels were determined using the TaqMan Fast Universal PCR Master Mix (Applied Biosystems) with FAM-labeled TaqMan gene expression assays (see Supplementary Table 5). Average threshold cycles were determined from triplicate reactions and the levels of gene expression were normalized to GAPDH (VIC-labeled endogenous control assay). Error bars indicate standard deviations or ranges of fold change relative to the reference sample, as indicated in the legend. qRT-PCR experiments were performed on a StepOnePlus Real Time PCR System (Applied Biosystems).

Immunofluorescence staining

For immunofluorescence analysis, cells were cultured on glass slides and permeabilised directly in 0.4% Triton X-100, followed by fixation in 4% PFA. The cells were stained with a rabbit polyclonal primary antibody against trimethylated H3K27 (1:500) from Upstate (cat#07-449).

Blastocyst injection

The capacity to contribute to the germ lineage was assessed at E12.5 in genital ridges from embryos generated by blastocyst injection of iPS cells transgenic for Nanog and Tet1WT or Tet1Mut. These cells contain an *Oct4*-GFP reporter transgene. Prior to blastocyst injection, floxed transgenes were excised by transient co-transfection with Cre recombinase and a Cherry reporter plasmid, and sorted 24 hours later for the latter. All mice procedures were performed in accordance with Home Office guidelines and regulations in University of Cambridge, UK.

Mass spectrometry analysis of 5hmC and 5mC levels

One microgram (μg) of genomic DNA was prepared for mass spectrometry by digestion with the DNA Degradase Plus kit (Zymo Research). The resulting nucleosides were analysed by

LC-MS/MS on a LTQ Orbitrap Velos mass spectrometer (Thermo Scientific) fitted with a nanoelectrospray ion-source (Proxeon). Mass spectral data were acquired in selected reaction monitoring (SRM) mode, monitoring the transitions 258 → 142.0611 (5hmC), 242 → 126.0662 (5mC) and 228 → 112.0505 (C). Parent ions were selected for SRM with a 4 mass unit isolation window and fragmented by HCD with a relative collision energy of 35%, with a resolution >14,000 for the fragment ions.

Peak areas from extracted ion chromatograms of the relevant fragment ions were used for quantification, by external calibration, relative to standards obtained by digestion of nucleotide triphosphates. All samples were run in triplicate.

Hydroxymethyl-sensitive qPCR

Preparation of 2 μg of genomic DNA for glucosylated hydroxymethyl-sensitive qPCR (glucMS-qPCR) was performed as stipulated in the EpiMark kit (NEB). Primers were designed to amplify over a single MspI site from within gene promoters, introns, and intragenic regions (Supplementary Table 6). 5mC and 5hmC levels calculated from glucMS-qPCR were expressed as hmC/C and mC/C ratios.

Analysis of overlapping Nanog and Tet1 genomic binding sites

ChIP-Seq datasets for Nanog and Tet1 were downloaded from GEO (Accession numbers: GSE11431, GSE11724, GSE24841, GSE26832), reads were aligned to the mouse reference (mm9) genome using Bowtie (http://bowtie-bio.sourceforge.net)[18] and peaks were called with MACS (http://liulab.dfci.harvard.edu/MACS)[19]. Nanog and Tet1 genomic binding sites that were within 400 bp distance of the inferred peak summits were counted for all pairs of Nanog and Tet1 datasets to provide a ranking measure. Since we compared the number of overlapping peaks between three Tet1 versus two Nanog ChIP-Seq studies this number includes some redundant peaks. To assess the statistical significance, 10,000 random permutation trials were performed, where in each trial Nanog and Tet1 peaks were randomly re-assigned and genomic locations within a 1-kb distance were counted. Gene ontology (GO) analysis for the Nanog and Tet1 shared targets was performed with the David online database as previously described[4,20].

Mouse ES cell ChIP-seq data sets for H3K4me1 and H3K27ac were downloaded from the GEO database (GSE24164)[21]. ChIP-seq data were aligned with Bowtie to the mouse reference genome requiring unique matches and analyzed with MACS to determine peaks.

Nanog, H3K4me1, and H3K27ac ChIP-seq data sets in H1 ES cells from ENCODE were downloaded from the UCSC database and coordinates were mapped to hg18 using the UCSC *liftover* tool. Nanog binding sites were classified according to the presence of histone modification peaks within +/- 800 nt of the Nanog ChIP-seq peak summit. 5hmC data for mouse and human were downloaded from the GEO database (GSE36173)[22]. 5hmC profiles were centered at Nanog motifs within +/- 200 bp of the Nanog peak summit. The profiles are averaged over 500 bp windows. The number of binding sites shown for each category in Supplementary Fig. 16b are comparable (promoter-proximal sites, n=489; distal H3K4me1 sites, n=482; distal H3K4me1 and H3K27ac sites, n=455; distal no H3K4me1/H3K27ac sites, n=440). Microarray data was downloaded from the GEO database (GSE2972)[23], and was re-processed and normalized by quantile normalization using the affy/limma Bioconductor packages. Normalized expression values are shown in log2 scale (Supplementary Fig. 16c).

Interactomics analysis

An interactomics analysis experiment always requires a control to discriminate between the true interactors and the contaminants[24-25]. For that reason different cell lines expressing different types of baits (one that contains a tag and one that does not) or different beads (ones that bind prey and ones that do not) are used for affinity purification (AP) and its control. The assumption is that the control experiment will identify contaminants. However, using different cell lines or different beads might still result in some contaminants being reproducibly identified at higher spectral counts (SPC) in AP, but not in its control. Particularly, 1) when the bait contains a tag in the experimental, but not in the control cell line a contaminant might interact with the tag of the bait, but not with the bait itself[26]. Tagging or overexpressing baits might also cause their aggregation or misfolding leading to interactions not exhibited by non-tagged bait[27]. Another approach is to add a tag to a bait, that promotes its modification (biotinylation) *in vivo* by an enzyme (BirA) expressed in experimental, but not in control lines. BirA has been shown to possess remarkable specificity *in vivo* by modifying only a single protein in *E.coli*[28]. However, when the enzyme is overexpressed it might non-specifically biotinylate itself and other proteins[29]. If the enzyme non-specifically modifies a contaminant it might also interact with beads, but will appear as an interactor of the bait as a result of LC-MS/MS analysis. Finally, if experimental and control lines are sufficiently different, a prey might be expressed differently, which would also skew the results[24]. 2) When the beads used in AP and its control are different, they might

bind contaminants differently and contaminants preferentially bound in AP will appear as true interactors[24].

Thus, clearly distinguishing between true interactors and contaminants might not be feasible when a single bait modified by a single tag is used for the reasons outlined above. However, these effects could be accounted for to a certain degree in large-scale analysis, when several proteins are used as baits and promiscuous preys are penalized by score (for example in SAINT[30] and CompPASS[31] algorithms) or are removed[31]. This is not feasible, however, with experiments identifying the interactome of a single protein. Rees et al. suggested the iPAC method[26], whereby a bait is modified with a tag that contains two different epitope sites: Flag and strep-tagII. Then two different APs are performed targeting each epitope site (with strep-Tactin beads and anti-Flag beads). A prey needs to be identified in both APs, but not in control experiments to be called a true interactor. This approach is similar to tandem affinity purification (TAP), but offers the benefit of conserving weaker and more transient interactions, which might be disrupted by TAP[32].

We performed an experiment similar to iPAC, but involving three independent APs with their corresponding controls to find high-confidence Nanog interactors (as described in Supplementary Table 1). Similarly to iPAC, two epitope sites (Flag tag and peptide tag for *in vivo* biotinylation by *BirA*[7]) were introduced to Nanog within a single tag and two independent APs were performed targeting each epitope site (with anti-Flag beads and streptavidin-conjugated beads respectively). The third affinity purification targeted wild-type Nanog with a polyclonal antibody against it, bound to protein G agarose beads.

The APs and their controls described above resulted in three tables containing SPC for preys in APs and their controls (the numbers in the tables below are made up for illustrative purposes).

Streptavidin beads, biotinylated Nanog		
Prey#	SPC AP	SPC control
Prey 1	10	5
Prey 2	1	0
…		
Prey n	6	9

Anti-Flag beads, Flag-tagged Nanog		
Prey#	SPC AP	SPC control
Prey 1	3	0
Prey 2	4	9
…		
Prey n	7	4

IgG anti-Nanog beads, wild type Nanog		
Prey#	SPC AP	SPC control
Prey 1	17	1
Prey 2	0	2
…		
Prey n	3	4

For the first three steps of analysis the results for the three APs were treated independently.

First step. A single spectral count was added to all preys in all three APs and their controls.

Streptavidin beads,	Anti-Flag beads, Flag-tagged	IgG anti-Nanog beads,

biotinylated Nanog		
Prey#	SPC AP	SPC control
Prey 1	11	6
Prey 2	2	1
…		
Prey 1500	7	10

Nanog		
Prey#	SPC AP	SPC control
Prey 1	4	1
Prey 2	5	10
…		
Prey 1000	8	5

wild type Nanog		
Prey#	SPC AP	SPC control
Prey 1	18	2
Prey 2	1	3
…		
Prey 700	4	5

Second step. A ratio of SPC between AP and control purification (AP/control SPC ratio) was calculated for every prey in all three APs. Thus, at this point every prey in an AP is represented by a single number: its AP/control SPC ratio.

Streptavidin beads, biotinylated Nanog	
Prey#	AP/control SPC ratio
Prey 1	1.83
Prey 2	2
…	
Prey 1500	0.7

Anti-Flag beads, Flag-tagged Nanog	
Prey#	AP/control SPC ratio
Prey 1	4
Prey 2	0.5
…	
Prey 1000	1.6

IgG anti-Nanog beads, wild type Nanog	
Prey#	AP/control SPC ratio
Prey 1	9
Prey 2	0.33
…	
Prey 700	0.8

Adding a value of one to SPC of all APs and their controls (step one) serves a dual purpose. First, it allowed calculating a ratio of SPC even when a protein had no SPC in the control. Second, it biased the ratios in favor of proteins that had a high number of SPC, e.g. a protein with 2 SPC in the AP and 1 spectral count in the control will end up with a ratio of 1.5 (3/2), while a protein with 10 SPC in AP and 5 SPC in the control will end up with a ratio of 1.83 (11/6).

Third step. The AP/control SPC ratios for the preys were then converted to cumulative probabilities (CPs) (for each of the three APs independently). CPs were computed in the following way[33]: preys within a given AP were rank ordered based on the magnitude of AP/control SPC ratio, then equation 1 was applied:

$$CP = 1 - \textit{prey rank/number of preys} \quad \text{(Eq. 1)}$$

Streptavidin-conjugated beads, biotinylated Nanog			
prey	ratio	rank	CP
prey 1	1.83	150	0.90
prey 2	2	140	0.906
…	…	…	…
prey 1500	0.7	1470	0.02

Anti-Flag beads, Flag-tagged Nanog			
prey	ratio	rank	CP
prey 1	4	50	0.95
prey 2	0.5	1000	0
…	…	…	…
Prey 1000	1.6	130	0.87

IgG anti-Nanog beads, wild type Nanog			
prey	ratio	rank	CP
Prey 1	9	7	0.99
Prey 2	0.33	690	0.015
…	…	…	…
Prey 700	0.8	600	0.142

Thus, CP is the proportion of AP/control SPC ratios that are as low or lower than a given AP/control SPC ratio within a given AP. Inversely 1 – CP is the probability of observing an AP/control SPC ratio as high or higher than a given AP/control SPC ratio within a given AP[34-35]. All the AP/control SPC ratios and their associated CPs within a given AP can be visualized by plotting empirical cumulative distribution functions. Supplementary Figure 3 demonstrates empirical cumulative distribution functions of AP/control SPC ratios for the three APs in this study. It also demonstrates AP/control SPC ratios and CP for 27 proteins identified as true interactors (see below).

The smaller the 1-CP of an AP/control SPC ratio, the less AP/control SPC ratios are higher than it within a given AP. This means the lower 1-CP for a given prey, the better it performed compared to other preys within a given AP. It is important to note that since CPs are computed for each of the three APs individually, the same AP/control SPC ratio will result in different CPs in different APs.

After CPs were computed from AP/control SPC ratios, the results of the three APs were combined to identify high-confidence interactors. We required a prey to be identified in all three APs to be considered a candidate for a true interactor. We found 233 of such preys. Combined cumulative probability (CCP) was calculated for them using equation 2:

$$CCP_i = -log_2 ((1-CPFLAG_i) * (1- CPSA_i) * (1- CPWT_i)) \text{ (Eq.2)}$$

CCP_i denotes combined cumulative probability for prey i. $CPFLAG_i$, $CPSA_i$, and $CPWT_i$ are CPs for $prey_i$ for the three APs. We multiplied the probabilities to observe AP/control SPC ratios as high as or higher than that found for a prey in the three APs and took a log_2 of the product, which resulted in better performing preys having a higher overall score.

To estimate CCP distribution that would occur by chance (for proteins not interacting with Nanog) we simulated CCPs for 23,300 (one hundred times the actual dataset) preys in a method similar to estimating distribution of random D scores in CompPASS algorithm[31]. For each of the simulated preys a CP was selected randomly from each of the three APs. The three selected CPs were then used to compute the score as shown in equation 2. We then calculated for every CCP how many among the 233 preys are expected to have a given or higher CCP by chance (number of preys among the 23,300 simulated preys divided by 100) and how many preys are actually observed to have a given or higher CCP. Figure 1b in the main text represents the result of this calculation.

A total of 42 preys from the experimental dataset and 8.74 preys from the simulated dataset had a score higher than 8.5, which was selected as the threshold score (false discovery rate of 17%)[36]. However, this analysis is incomplete, since a prey might have a very low 1-CP in one AP only, but still have a high CCP. For example, the table below gives SPC and associated 1-CPs for three tubulin proteins in the three APs. These proteins are obviously contaminants, due to their cytoplasmic localization. It is noteworthy that all three of them have high 1-CP in APs with streptavidin-conjugated beads and anti-Flag beads, but very low 1-CP in AP with IgG anti-Nanog beads. Clearly, IgG anti-Nanog beads interact with the representatives of tubulin family members with higher affinity than control IgG beads.

Summary of the AP-MS analysis for three tubulin proteins:

		SPC, control in brackets			1-CP		
	CCP	Streptavidin-conjugated beads	anti-Flag beads	IgG anti-Nanog beads	Streptavidin-conjugated beads	anti-Flag beads	IgG anti-Nanog beads
Tubb2b	12.4237	45(25)	50(37)	38(1)	0.19401	0.36963	0.00254
Tuba1a	9.22772	40(25)	48(43)	35(2)	0.29815	0.44084	0.01269
Tubb5	9.02729	12(7)	11(8)	9(0)	0.29101	0.37068	0.01777

High AP/control SPC ratio in one, but not in the other APs is expected from a contaminant. To remove such preys from the final list of high-confidence interactors we applied a rule that a prey should have a 1-CP of 0.15 or lower in at least two out of three APs. Applying this additional criterion resulted in 27 proteins in our final list of high-confidence interactors as opposed to 3.28 in our simulated dataset. This reduced FDR to approximately 10% (the proportion of null features that are called significant).

Supplementary Table 2 gives the list of 27 total true interactors with their CCPs, SPCs in AP and in the control (in brackets), and 1-CPs observed in all of the three APs. In addition, for each of the 27 preys we provide three cumulative distribution functions of AP/control SPC ratios, with the ratio of the prey shown as a vertical line (Supplementary Fig. 3). This allows us to visually compare how much better a prey performed than other preys in the three APs. We also compute q-values for each of the interactors as the minimum FDR at which the interactor is called significant (for a detailed explanation of q-values and FDR see Ref[37]).

We believe that our analysis as well as the list of interactors is quite conservative. First, we require a prey to be identified in all three APs. Second, though choosing a score of 8.5 and 1-CP of 0.15 in two out of three APs as a filter when selecting interactors might seem

somewhat arbitrary, the interactors in our final list are expected to be significant since we provide a false discovery rate for our analysis. Furthermore, the actual FDR is probably lower than the 10% we specified since we used 1-CP values chosen from the three APs at random to estimate CCP distribution expected for preys not interacting with Nanog (contaminants). However, some of these values belong to true interactors and thus, the expected scores calculated for the preys not interacting with Nanog are probably over-estimated.

Finally, Rees et al.[26] and Trinkle-Mulcahy et al.[38] have suggested that some interactions might be true, but not necessarily biologically interesting. For example, they identified that many of their baits interacted with ribosomal and heat shock proteins, which is what is expected since all proteins interact with translation and folding machinery at the time of their synthesis. We did not identify any proteins that serve functions of protein synthesis and turnover among the list of our interactors.

Supplementary References

1 Das, S., Jena, S. & Levasseur, D. N. Alternative Splicing Produces Nanog Protein Variants with Different Capacities for Self-Renewal and Pluripotency in Embryonic Stem Cells. *The Journal of biological chemistry*, doi:M111.290189 [pii]10.1074/jbc.M111.290189 (2011).

2 Ivanova, N. *et al.* Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533-538 (2006).

3 Wang, J. *et al.* A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364-368, doi:nature05284 [pii] 10.1038/nature05284 (2006).

4 Ding, J., Xu, H., Faiola, F., Ma'ayan, A. & Wang, J. Oct4 links multiple epigenetic pathways to the pluripotency network. *Cell research*, doi:cr2011179 [pii]10.1038/cr.2011.179 (2011).

5 Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic acids research* **11**, 1475-1489 (1983).

6 van den Berg, D. L. *et al.* An Oct4-centered protein interaction network in embryonic stem cells. *Cell stem cell* **6**, 369-381, doi:S1934-5909(10)00091-3 [pii] 10.1016/j.stem.2010.02.014 (2010).

7 de Boer, E. *et al.* Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 7480-7485 (2003).

8 Kim, J., Cantor, A. B., Orkin, S. H. & Wang, J. Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. *Nature protocols* **4**, 506-517 (2009).

9 Sommer, C. A. *et al.* Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem cells (Dayton, Ohio)* **27**, 543-549 (2009).

10 Hong, H. *et al.* Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature* (2009).

11 Silva, J. *et al.* Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol* **6**, e253 (2008).

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

13 Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722-737, doi:S0092-8674(09)00969-6 [pii] 10.1016/j.cell.2009.07.039 (2009).

14 Pereira, C. F. & Fisher, A. G. Heterokaryon-based reprogramming for pluripotency. *Current protocols in stem cell biology* **Chapter 4**, Unit 4B 1 (2009).

15 Lee, T. I., Johnstone, S. E. & Young, R. A. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nature protocols* **1**, 729-748 (2006).

16 Marson, A. *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521-533, doi:S0092-8674(08)00938-0 [pii] 10.1016/j.cell.2008.07.020 (2008).

17 Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343-348, doi:nature10066 [pii]10.1038/nature10066 (2011).

18 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25, doi:gb-2009-10-3-r25 [pii]10.1186/gb-2009-10-3-r25 (2009).

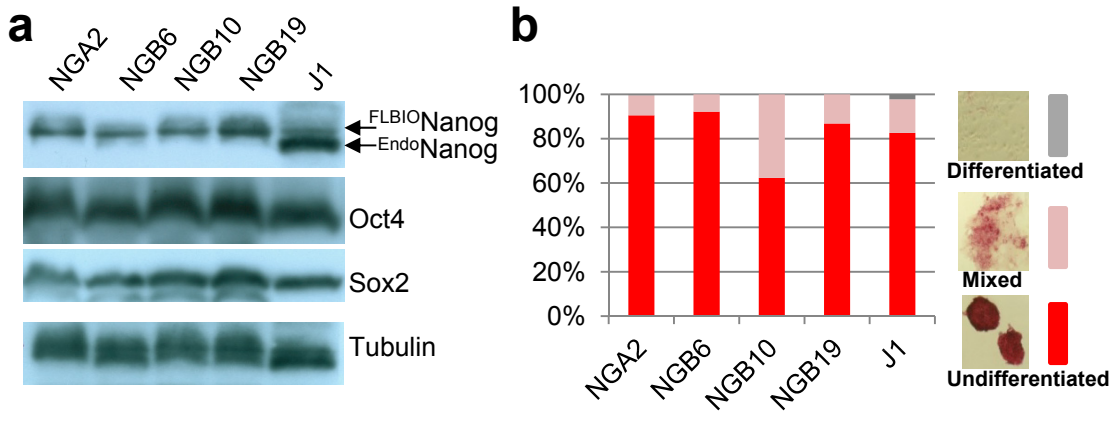
19 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137, doi:gb-2008-9-9-r137 [pii]10.1186/gb-2008-9-9-r137 (2008).

20 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57, doi:nprot.2008.211 [pii]10.1038/nprot.2008.211 (2009).

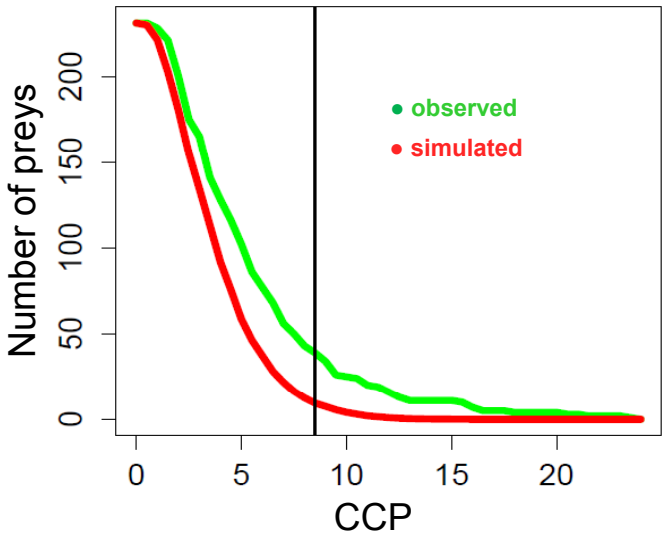
21 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936, doi:1016071107 [pii]10.1073/pnas.1016071107 (2010).

22 Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368-1380, doi:S0092-8674(12)00534-X [pii]10.1016/j.cell.2012.04.027 (2012).
23 Hailesellasse Sene, K. *et al.* Gene function in early mouse embryonic stem cell differentiation. *BMC Genomics* **8**, 85, doi:1471-2164-8-85 [pii]10.1186/1471-2164-8-85 (2007).
24 Dunham, W. H., Mullin, M. & Gingras, A. C. Affinity-purification coupled to mass spectrometry: Basic principles and strategies. *Proteomics* **12**, 1576-1590, doi:10.1002/pmic.201100523 (2012).
25 Pardo, M. & Choudhary, J. S. Assignment of protein interactions from affinity purification/mass spectrometry data. *Journal of proteome research* **11**, 1462-1474, doi:10.1021/pr2011632 (2012).
26 Rees, J. S. *et al.* In vivo analysis of proteomes and interactomes using Parallel Affinity Capture (iPAC) coupled to mass spectrometry. *Mol Cell Proteomics* **10**, M110 002386, doi:M110.002386 [pii]10.1074/mcp.M110.002386 (2011).
27 Rumlova, M., Benedikova, J., Cubinkova, R., Pichova, I. & Ruml, T. Comparison of classical and affinity purification techniques of Mason-Pfizer monkey virus capsid protein: the alteration of the product by an affinity tag. *Protein Expr Purif* **23**, 75-83, doi:10.1006/prep.2001.1488S1046592801914883 [pii] (2001).
28 Schatz, P. J. Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in Escherichia coli. *Biotechnology (N Y)* **11**, 1138-1143 (1993).
29 Beckett, D., Kovaleva, E. & Schatz, P. J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci* **8**, 921-929 (1999).
30 Choi, H. *et al.* SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nature methods* **8**, 70-73, doi:nmeth.1541 [pii]10.1038/nmeth.1541 (2011).
31 Sowa, M. E., Bennett, E. J., Gygi, S. P. & Harper, J. W. Defining the human deubiquitinating enzyme interaction landscape. *Cell* **138**, 389-403, doi:S0092-8674(09)00503-0 [pii]10.1016/j.cell.2009.04.042 (2009).
32 Chen, G. I. *et al.* PP4R4/KIAA1622 forms a novel stable cytosolic complex with phosphoprotein phosphatase 4. *The Journal of biological chemistry* **283**, 29273-29284, doi:M803443200 [pii]10.1074/jbc.M803443200 (2008).
33 Dalgaard, P. in *Introductory Statistics with R: Statistics and Computing* (ed Peter Dalgaard) 364 (2008).
34 Rosner, B. in *Fundamentals of Biostatistics* (Duxbury Pr, 2010).
35 Rice, J. A. in *Mathematical Statistics and Data Analysis* 672 (Duxbury Press, 1994).
36 Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9440-9445, doi:10.1073/pnas.1530509100 1530509100 [pii] (2003).
37 Kall, L., Storey, J. D., MacCoss, M. J. & Noble, W. S. Posterior error probabilities and false discovery rates: two sides of the same coin. *Journal of proteome research* **7**, 40-44, doi:10.1021/pr700739d (2008).
38 Trinkle-Mulcahy, L. *et al.* Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *The Journal of cell biology* **183**, 223-239, doi:jcb.200805092 [pii]10.1083/jcb.200805092 (2008).

Supplementary Figures



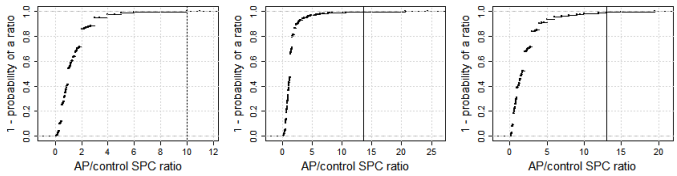
Supplementary Figure 1 | Characterization of ES cells expressing ^{FLBIO}Nanog. a, Normal expression levels of Oct4, Nanog, and Sox2 in NGA2, NGB, and J1 ES cells by western blot analyses of total cell lysates using anti-Nanog, anti-Oct4, anti-Sox2, and anti-tubulin antibodies. **b**, Normal clonogenicity of NGA2 and NGB ES cells compared with wild-type J1 ES cells in a colony formation assay. Individual colonies were stained for AP activity and scored into three categories (uniformly undifferentiated, partially differentiated or mixed, and differentiated) as indicated.



Supplementary Figure 2 | Distributions of combined cumulative probability (CCP) scores for preys observed in all three affinity purifications (APs) (green) and for simulated preys (red). Vertical line represents the cut-off threshold of 8.5 (false discovery rate of 10%).

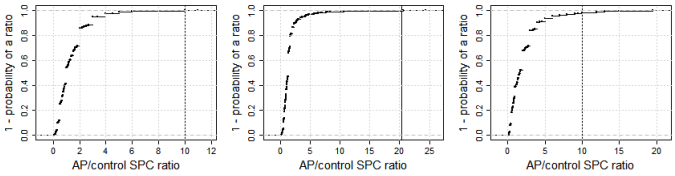
Nanog

anti-FLAG | streptavidin | anti-Nanog



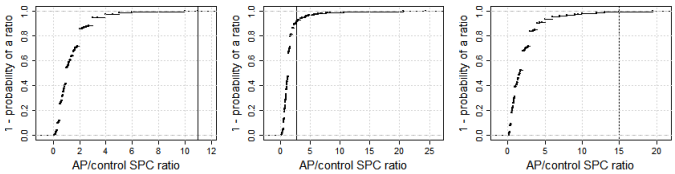
Zfp609

anti-FLAG | streptavidin | anti-Nanog



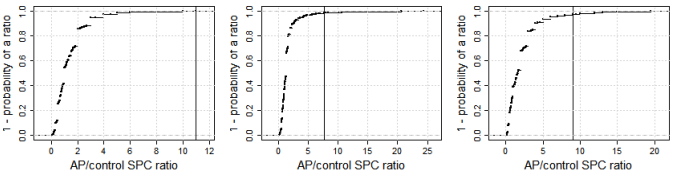
Mki67

anti-FLAG | streptavidin | anti-Nanog



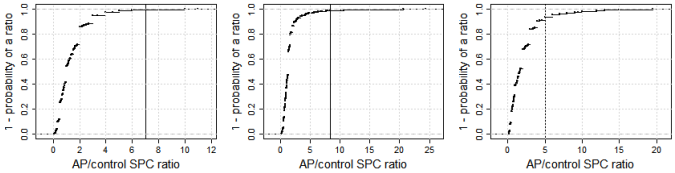
Emsy

anti-FLAG | streptavidin | anti-Nanog



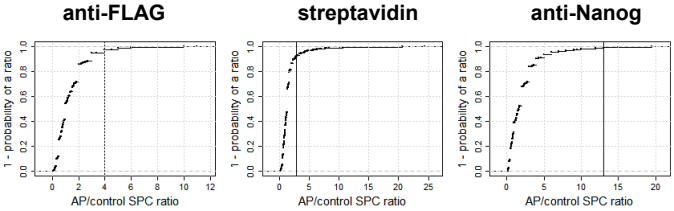
Nacc1

anti-FLAG | streptavidin | anti-Nanog



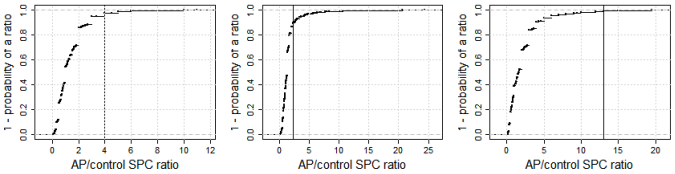
Jmjd1c

anti-FLAG | streptavidin | anti-Nanog



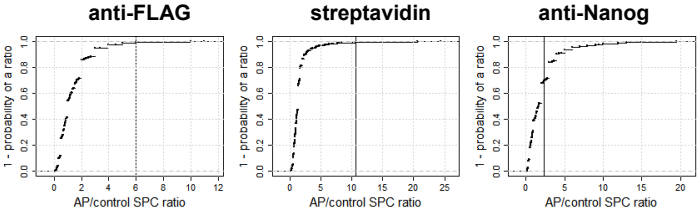
Brca2

anti-FLAG | streptavidin | anti-Nanog



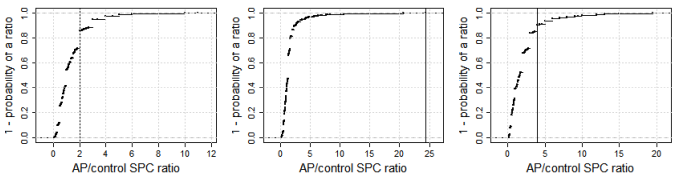
Tet1

anti-FLAG | streptavidin | anti-Nanog



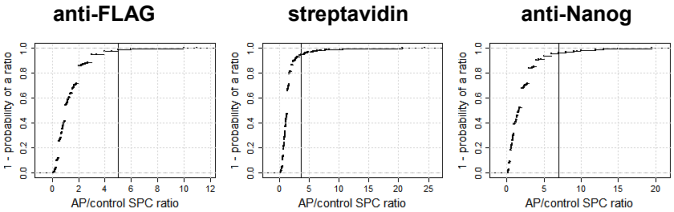
Ncor2

anti-FLAG | streptavidin | anti-Nanog



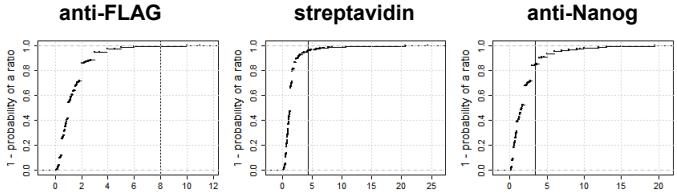
Sgol2

anti-FLAG | streptavidin | anti-Nanog



Hnrnpm

anti-FLAG | streptavidin | anti-Nanog



Qser1

anti-FLAG | streptavidin | anti-Nanog

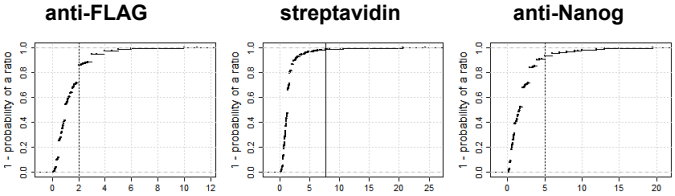
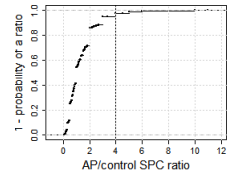


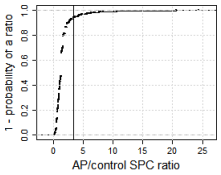
FIGURE S3 (part 1)

Mga

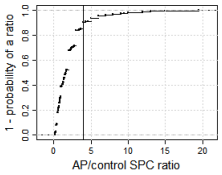
anti-FLAG



streptavidin

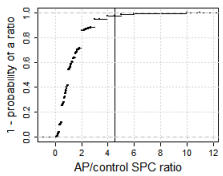


anti-Nanog

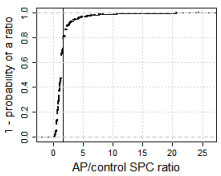


Gatad2a

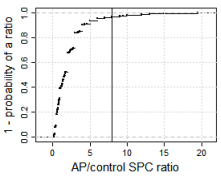
anti-FLAG



streptavidin

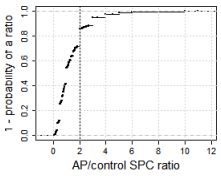


anti-Nanog

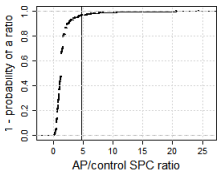


Rbm14

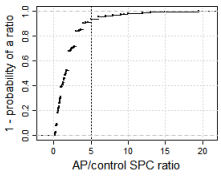
anti-FLAG



streptavidin

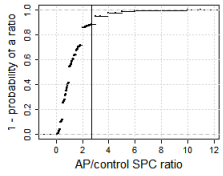


anti-Nanog

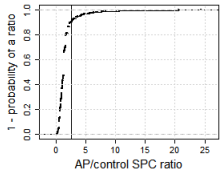


Skiv2l2

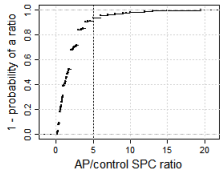
anti-FLAG



streptavidin

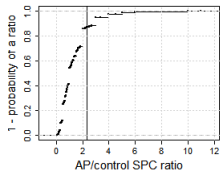


anti-Nanog

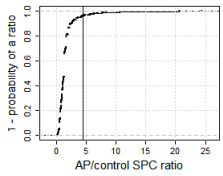


Lmnb1

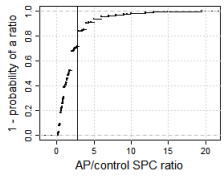
anti-FLAG



streptavidin

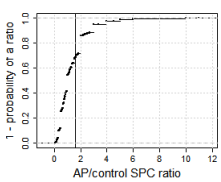


anti-Nanog

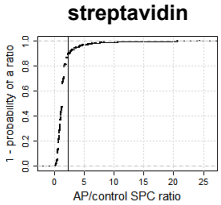


Rfc1

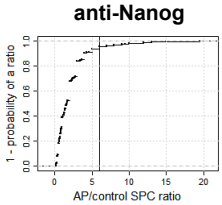
anti-FLAG



streptavidin

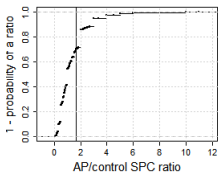


anti-Nanog

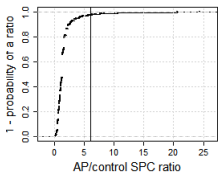


Bptf

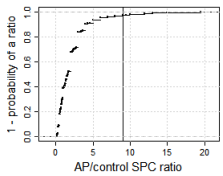
anti-FLAG



streptavidin

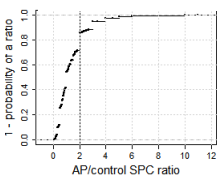


anti-Nanog

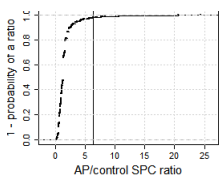


Arid3b

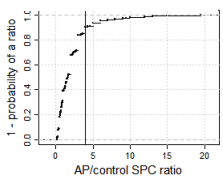
anti-FLAG



streptavidin

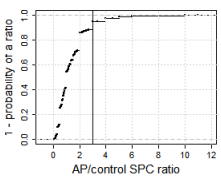


anti-Nanog

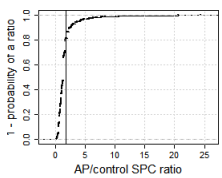


Hdac2

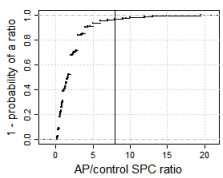
anti-FLAG



streptavidin

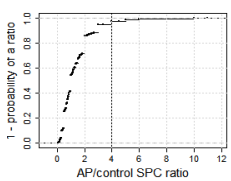


anti-Nanog

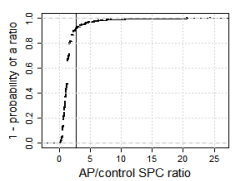


Pou5f1 (Oct4)

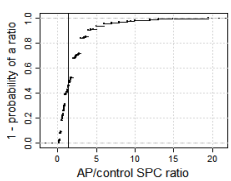
anti-FLAG



streptavidin

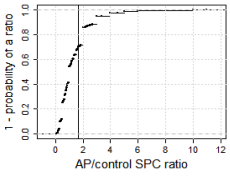


anti-Nanog

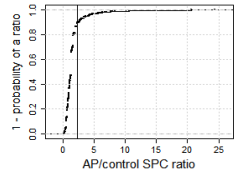


Mta2

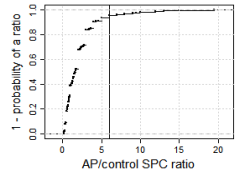
anti-FLAG



streptavidin

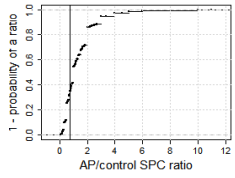


anti-Nanog

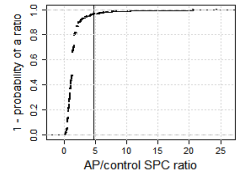


Trrap

anti-FLAG



streptavidin



anti-Nanog

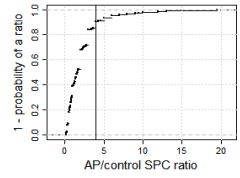


FIGURE S3 (part 2)

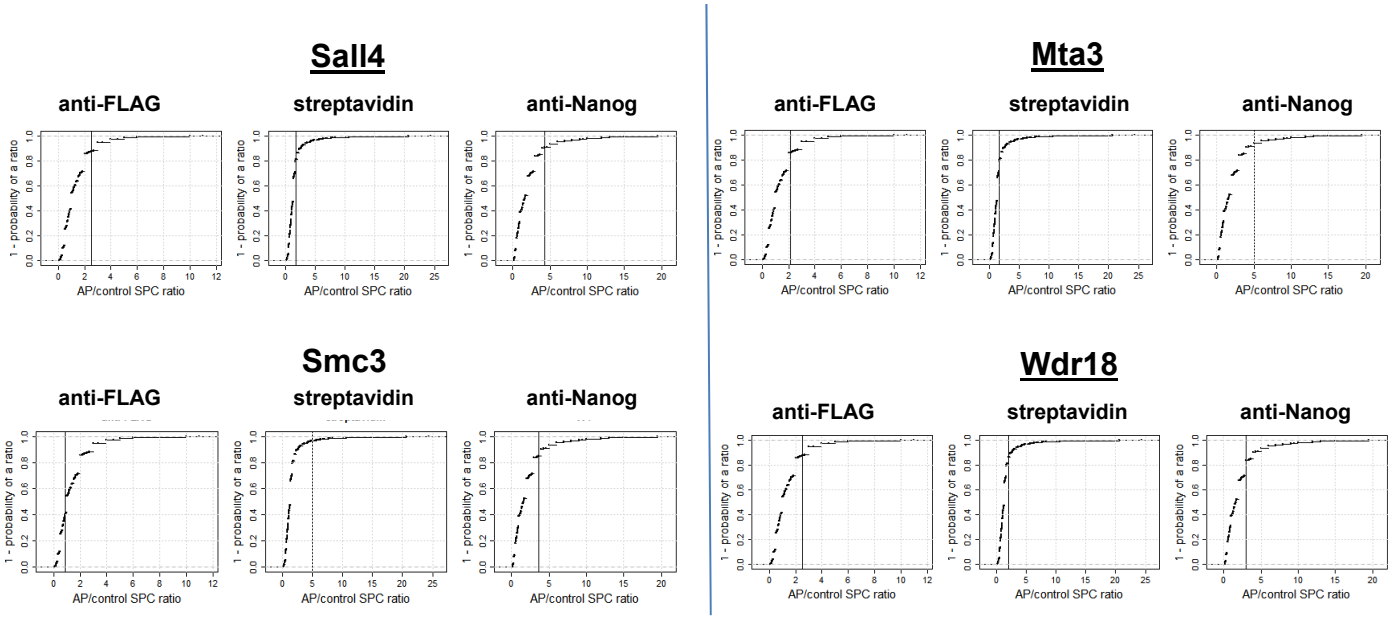
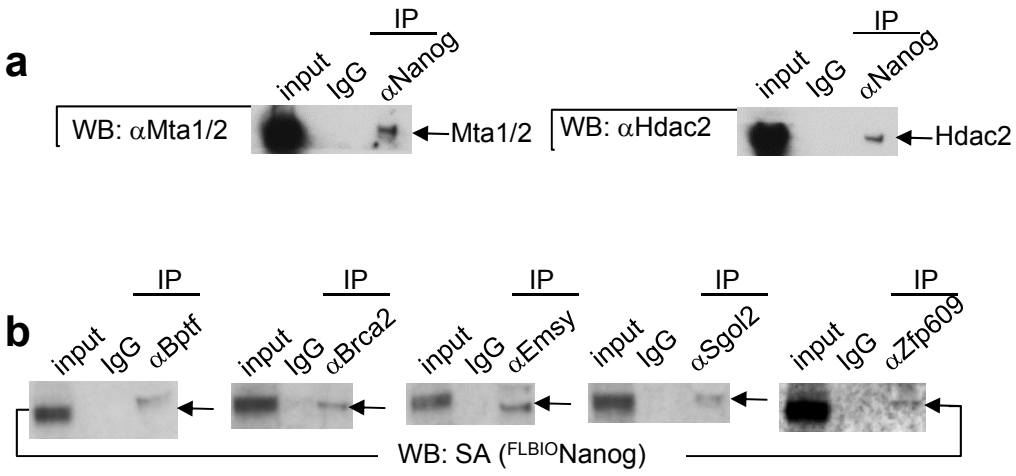


FIGURE S3 (part 3)

Supplementary Figure 3 | Empirical distribution functions of AP/control spectral count (SPC) ratios for 27 high-confidence Nanog interacting proteins from APs with anti-FLAG, streptavidin-conjugated, and IgG anti-Nanog beads. The lines represent the ratio of SPC between AP and control for the 27 high-confidence candidates listed in Fig. 1b from the three APs.



Supplementary Figure 4 | Validation of association of Nanog with known and novel factors. a, Validation of endogenous association of Nanog with Mta2 and Hdac2. IPs were performed with an anti-Nanog antibody using nuclear extracts from J1 ES cells followed by western blots with indicated antibodies. **b,** Validation of association of Nanog with 5 novel factors as indicated.

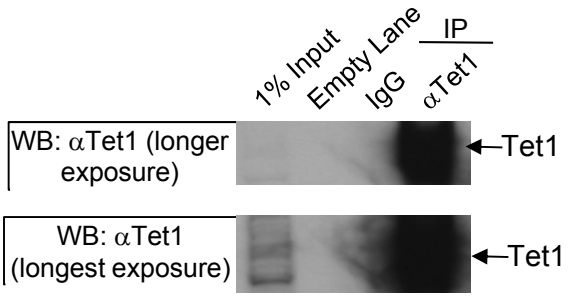
a

Pulldown/IP	Nr0b1 (Dax1)	Zfp281
^{FLBIO}Nanog SA-IP#1	4(0)	4(0)
^{FLBIO}Nanog SA-IP#2	5(0)	2(0)
^{FLBIO}Nanog SA-IP#3	12(0)	1(0)
^{FLBIO}Nanog FLAG-IP	2(0)	1(0)
^{end}Nanog Ab-IP	0(0)	0(0)

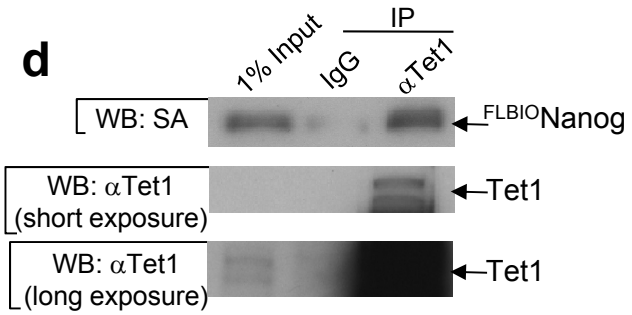
b

Pulldown/IP	Tet1
^{FLBIO}Nanog SA-IP#1	6(0)
^{FLBIO}Nanog SA-IP#2	12(0)
^{FLBIO}Nanog SA-IP#3	11(0)
^{FLBIO}Nanog FLAG-IP	5(0)
^{end}Nanog Ab-IP	6(2)

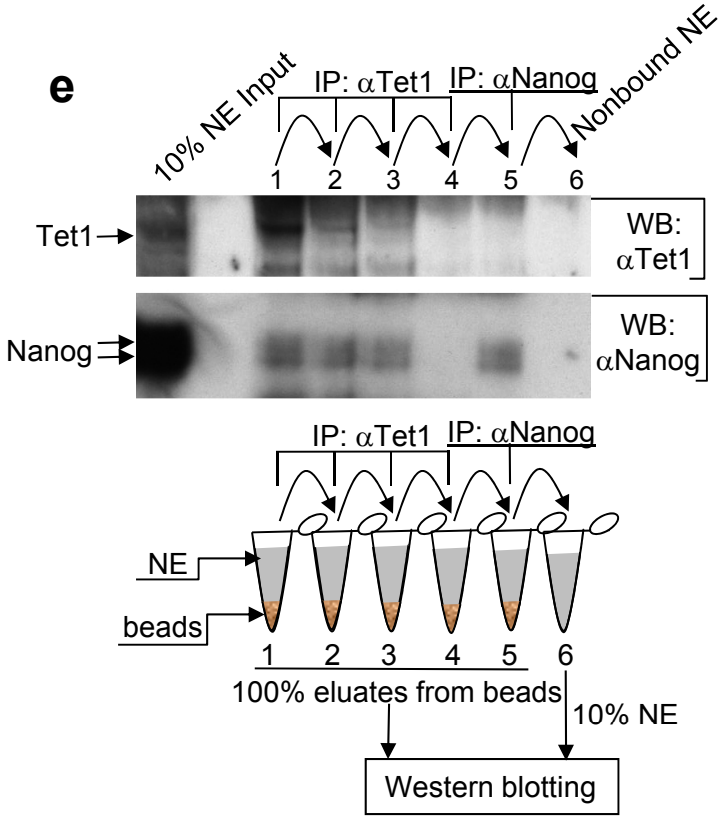
c



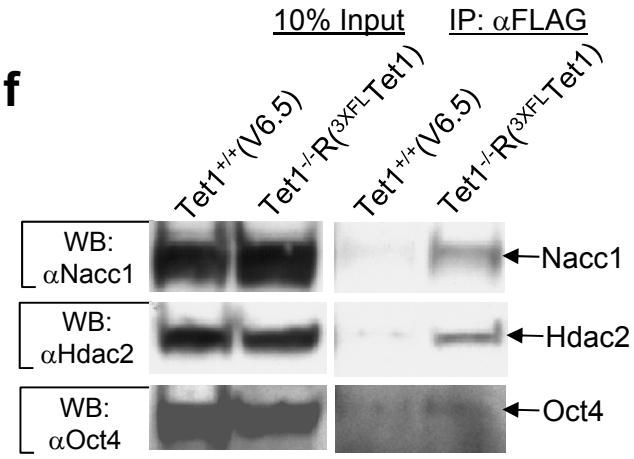
d



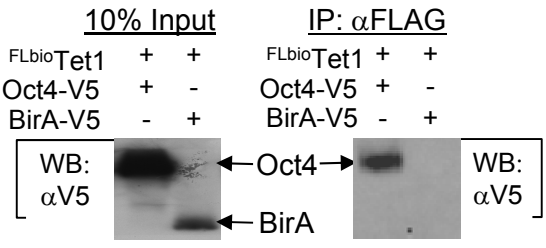
e

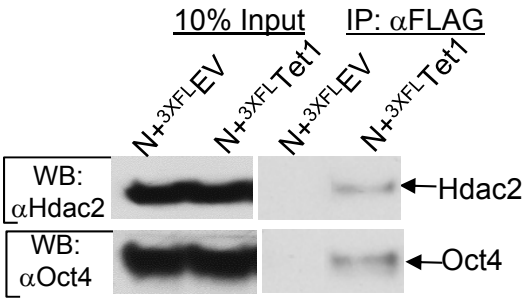


f



g





h

Protein ID	No. of Peptide
Tet1	**57(0)**
Sgol2	3(0)
Qser1	10(0)
Hdac2	6(4)
Oct4	3(1)

Supplementary Figure 5 | Validation of Nanog partners. a, List of peptides for two known Nanog interacting proteins (Nr0b1/Dax1 and Zfp281) in our three independent AP-MS studies. The number within brackets indicates the background peptide(s) identified by MS in control pull-down samples. **b,** List of peptides for Tet1 in our three independent AP-MS studies. **c**, Different exposures of the gel image shown in Fig. 1c to indicate the presence of Tet1 in the input. **d**, Validation of Nanog-Tet1 interaction by immunoprecipitation (IP) in NGB19 ES cells. These ES cells are the V6.5 line as opposed to the J1 line used in Fig. 1c. Nuclear extracts were subjected to IP with IgG or anti-Tet1 antibody followed by western blotting with Streptavidin-HRP (top panel) and anti-Tet1 (middle and bottom panels) to detect ^{FLBIO}Nanog and Tet1, respectively. The nuclear extract (NE) input is also included. **e**, Immunodepletion of Tet1 reveals that not all of Nanog protein is associated with all Tet1 protein in ES cells. Endogenous Tet1 was depleted by serial IPs with anti-Tet1 antibody (lanes 1-4). The non-bound nuclear extracts (NE) from step 4 were subjected to IP with anti-Nanog antibody. Note the presence of a large amount of Nanog but no Tet1 in the IP samples after three rounds of Tet1 immunodepletion using anti-Tet1 antibody (lane 3). The absence of Tet1 in the 4th Tet1 IP sample indicates complete depletion of Tet1 in NE, which accordingly results in no detection of Nanog in the same IP sample (lane 4). However, when the non-bound NE from the 4th Tet1 IP was subjected to Nanog IP, a fair amount of Nanog is still detected (lane 5). A diagram depicting the procedure of the experiment is also presented below the western blots. **f**, Association of Nanog partner proteins Nacc1, Hdac2, and Oct4 with Tet1 in ES cells. Wild-type (V6.5) and Tet1 null (Tet1$^{-/-}$) ES cells rescued with 3xFLAG tagged Tet1 (^{3xFL}Tet1) were used for IP with an anti-FLAG antibody followed by western blotting with the indicated antibodies. **g**, Association of the Nanog partner protein Oct4 with Tet1 in heterologous HEK293T cells. Cells were transiently transfected with the indicated constructs, and IP was performed with anti-FLAG antibody followed by western blotting with anti-V5 antibody. **h**, Association of Nanog partner proteins Sgol2, Qser1, Hdac2, and Oct4 with Tet1 in reprogramming intermediate cells under serum/LIF culture. The reprogramming intermediate cells expressing Nanog (N) and 3xFL empty vector (^{3xFL}EV) or Nanog (N) and ^{3xFL}Tet1 were subjected to IP with anti-FLAG antibody followed by MS identification of Tet1-interacting proteins (left table) and by western blotting with the indicated antibodies (right). The number within brackets in the top table indicates the background peptide(s) identified by MS in control N+^{3xFL}EV pull-down samples. Although background binding to the beads was observed for Hdac2 and Oct4, their interactions with Tet1 were validated by IP with an anti-FLAG antibody followed by western blotting with the corresponding antibodies.

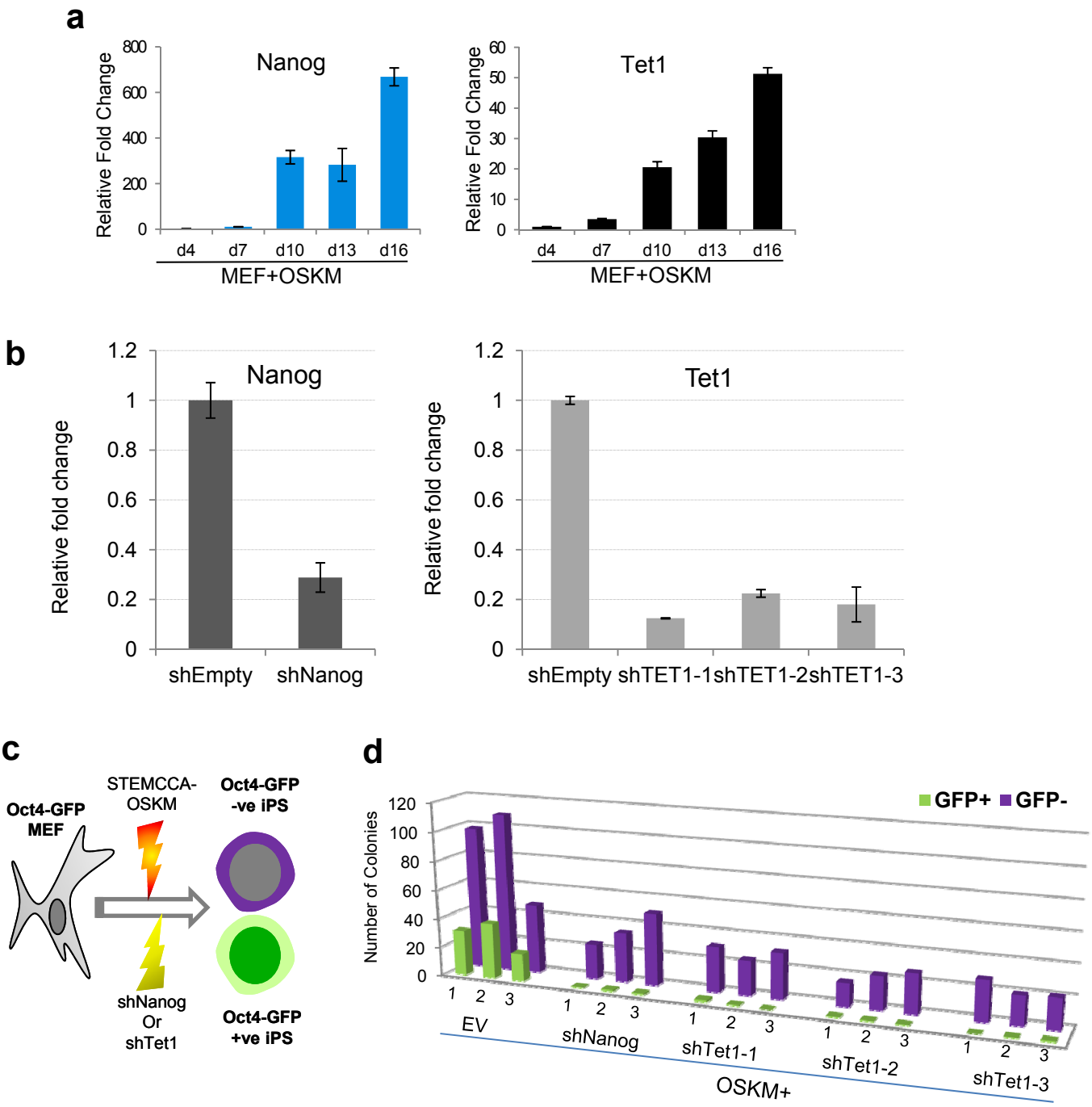


FIGURE S6 (part 1)

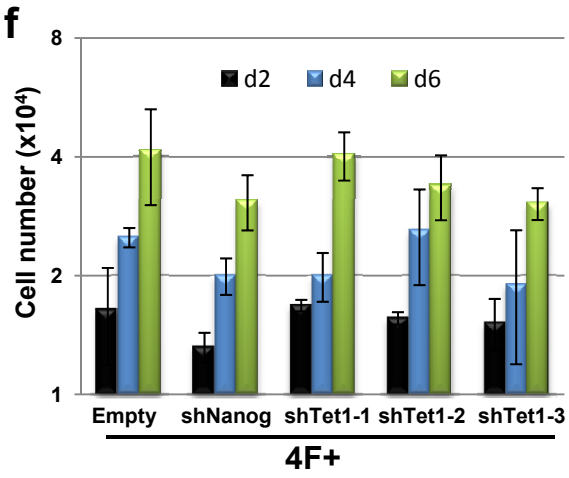
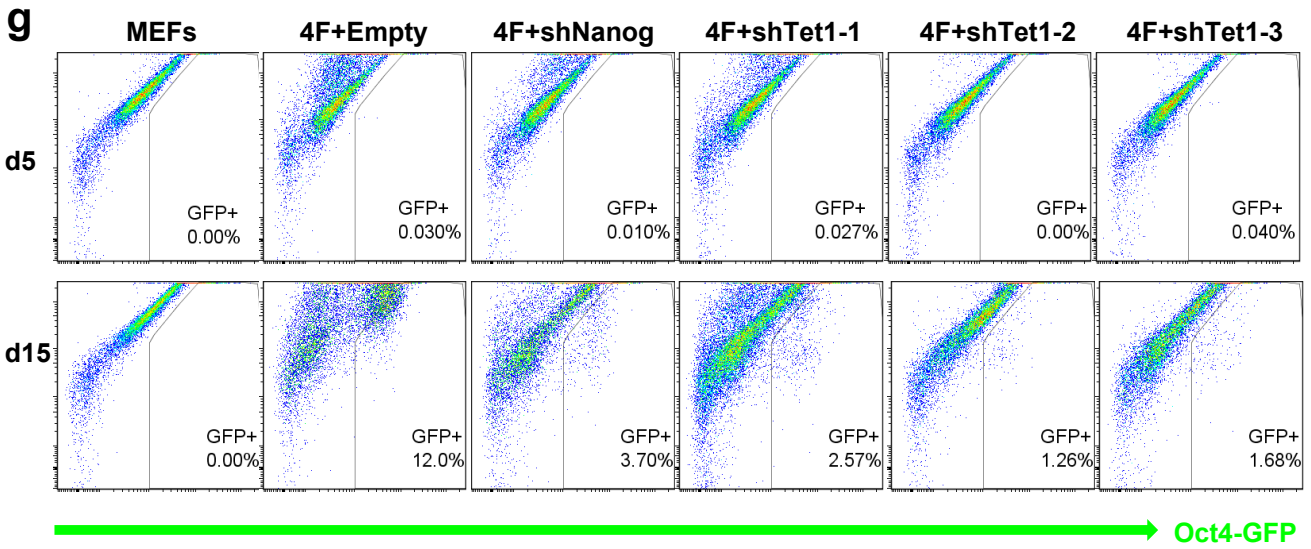
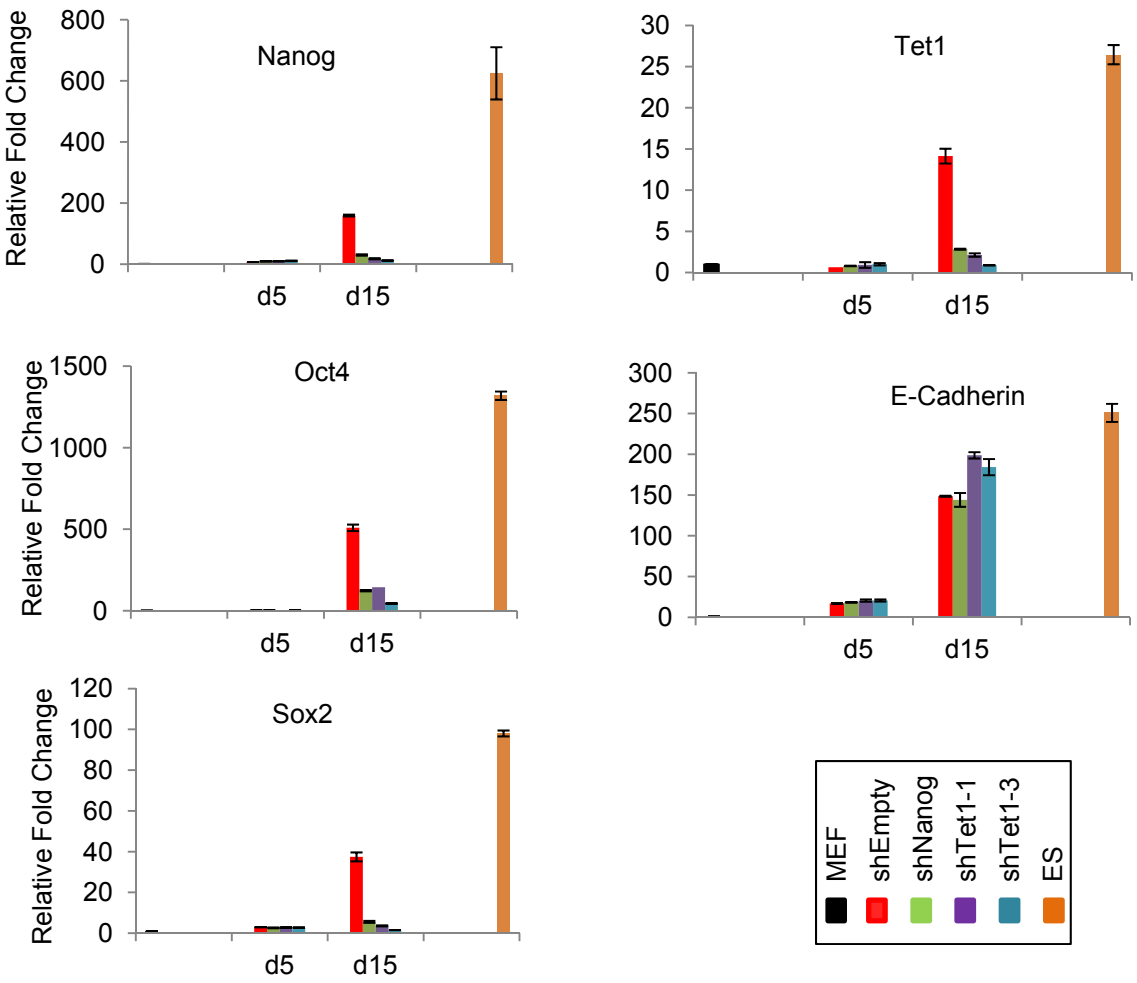


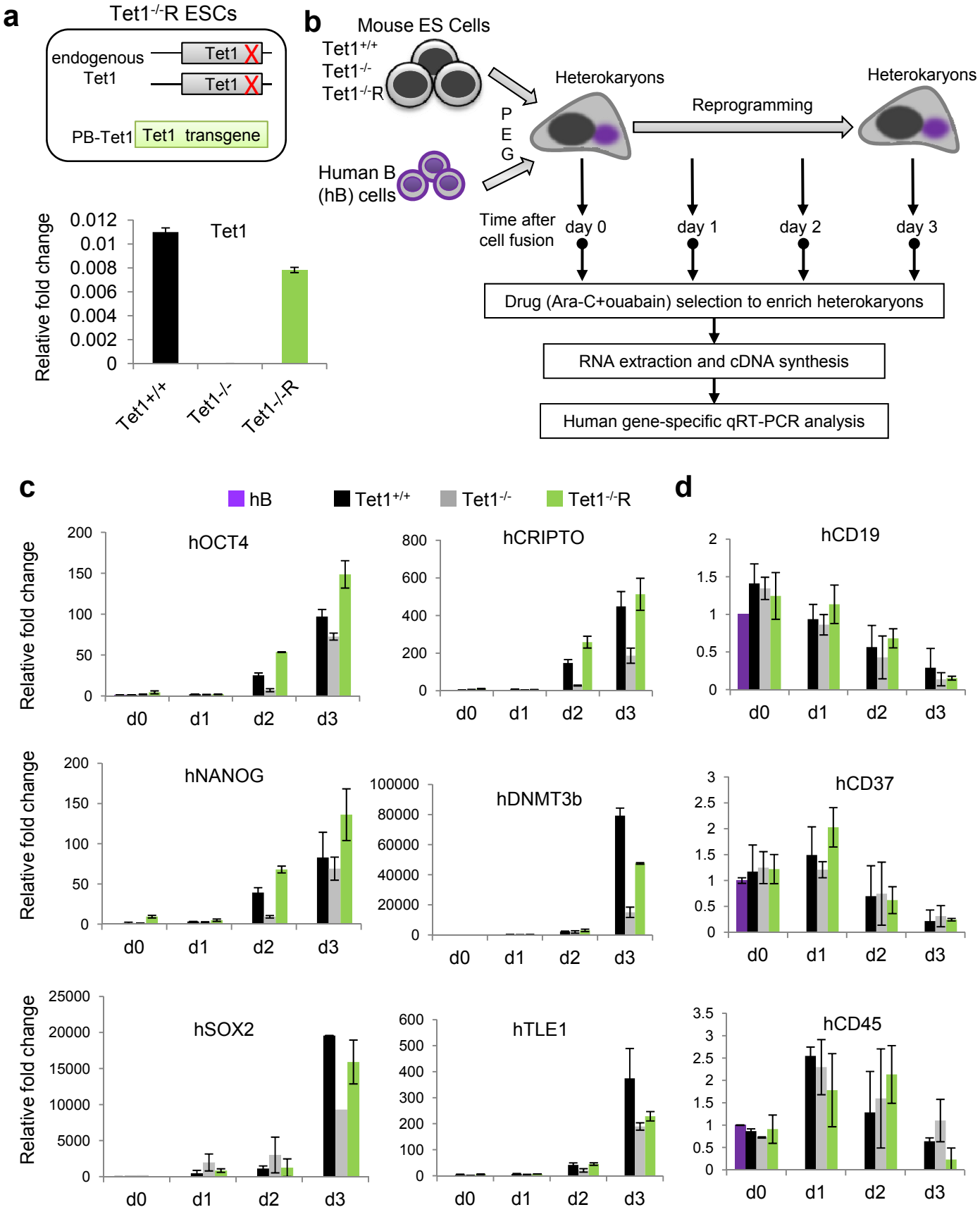
FIGURE S6 (part 2)



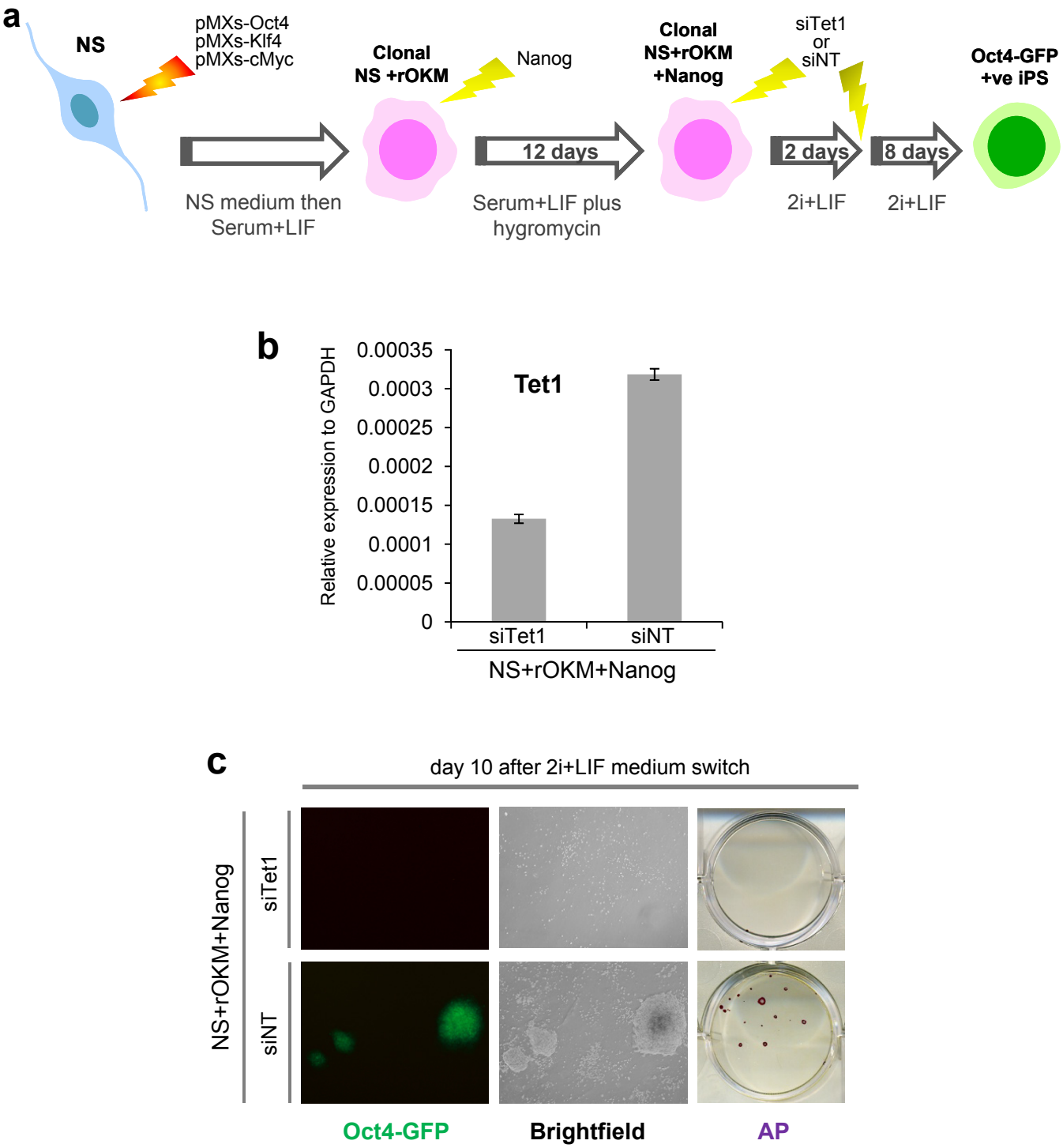
Supplementary Figure 6 | Both Nanog and Tet1 are required for efficient MEF reprogramming. a, Expression profiles of Nanog and Tet1 during reprogramming of MEFs. Error bars indicate standard deviation (n=3). **b,** Relative expression levels of Nanog and Tet1 upon corresponding shRNA treatment. Lentiviral pLKO.1 viruses expressing shRNA against Nanog (shNanog), or three independent shRNAs against Tet1 (shTet1-1/2/3) were used to infect CJ7 ES cells followed by selection with puromycin (1.0 mg/mL) for 6 days. RNAs were extracted for qRT-PCR analysis. Error bars indicate standard deviation (n=3). **c**, Strategy to test the requirement of Nanog and Tet1 for efficient reprogramming by RNAi. **d**, Down-regulation of Nanog or Tet1 compromises reprogramming efficiency. Data from three independent experiments (1-3) are presented. **e**, Alkaline phosphatase (AP) staining of iPSC colonies. *Oct4*-GFP reporter MEFs were co-transduced with STEMCCA lentivirus expressing 4 reprogramming factors (4F) (Oct4, Sox2, Klf4, and c-Myc) and pLKO.1 viruses, empty or expressing shRNA against Nanog (shNanog), or three independent shRNAs against Tet1 (shTet1-1/2/3). iPS cell colonies were stained for AP activity 16 days after transduction. **f**, MEF growth is minimally affected by co-transfection of reprogramming factors and shRNAs. MEFs were transduced with pLKO.1 empty virus or viruses expressing shNanog and shTet1-1/2/3, and cells were counted at each time point as indicated. Y-axis values represent cell number. Error bars indicate standard deviation (n=3). **g**, Flow cytometric analysis of *Oct4*-GFP cells during iPSC reprogramming. *Oct4*-GFP fluorescence was analyzed on an LSR-II Flow Cytometer System (BD Biosciences) at day 5 and day 15 after viral infection.



Supplementary Figure 7 | Quantitative PCR analysis of marker gene expression during the time course of MEF reprogramming with OSKM. Down-regulation of Nanog and Tet1 with respective shRNA impairs reactivation of the pluripotency genes *Oct4* and *Sox2*, but does not affect reprogramming intermediate marker *E-cadherin* expression during the reprogramming process. RNAs from day 5 and day 15 MEFs infected with viruses expressing OSKM reprogramming factors as well as the indicated shRNAs were subjected to RT-qPCR analyses. The respective gene expression levels in MEFs (black bars) and in ES cells (orange bars) are also included. The error bars indicate standard deviation (n=3).


a
Tet1-/-R ESCs
endogenous Tet1
Tet1
Tet1
PB-Tet1
Tet1 transgene
Relative fold change
Tet1
0.012
0.01
0.008
0.006
0.004
0.002
0
Tet1+/+
Tet1-/-
Tet1-/-R
b
Mouse ES Cells
Tet1+/+
Tet1-/-
Tet1-/-R
Human B (hB) cells
P E G
Heterokaryons
Reprogramming
Heterokaryons
Time after cell fusion
day 0
day 1
day 2
day 3
Drug (Ara-C+ouabain) selection to enrich heterokaryons
RNA extraction and cDNA synthesis
Human gene-specific qRT-PCR analysis
c
hB
Tet1+/+
Tet1-/-
Tet1-/-R
Relative fold change
hOCT4
hCRIPTO
hNANOG
hDNMT3b
hSOX2
hTLE1
d0
d1
d2
d3
d
hCD19
hCD37
hCD45

Supplementary Figure 8 | Tet1 depletion in ES cells compromises heterokaryon-based fusion reprogramming of human B cells. a, Top, depiction of the transgenic rescue *Tet1*$^{-/-}$ (*Tet1*$^{-/-}$R) ES cells expressing a Tet1 transgene from a PiggyBac vector (PB-Tet1). The parental *Tet1*$^{-/-}$ ES cells are described elsewhere[39]. Bottom, relative Tet1 transcript levels in *Tet1*$^{+/+}$, *Tet1*$^{-/-}$, and *Tet1*$^{-/-}$R ES cells analyzed by RT-qPCR. Error bars indicate standard deviation (n=3). **b**, Schematic illustration of the heterokaryon-based fusion reprogramming strategy, which has been described previously[14]. The two drugs Ara-C and ouabain are used to specifically eliminate proliferating mouse ES cells and human cells, respectively. **c**, RT-qPCR analyses of human pluripotency gene expression in heterokaryons at the indicated time points. The compromised reactivation of these genes in fusions with *Tet1*$^{-/-}$ ES cells (grey bars) and rescue of these genes' expression in fusions with *Tet1*$^{-/-}$R ES cells (green bars) relative to their expression in fusions with *Tet1*$^{+/+}$ ES cells (black bars) during the time course of the reprogramming process unambiguously establish a critical role of Tet1 in reprogramming. Error bars indicate standard deviation (n=3). **d**, RT-qPCR analyses of human B (hB) cell marker gene expression. Down-regulation of hB cell lineage marker gene expression in all the fusions indicates successful reprogramming of hB cells in heterokaryons. Error bars indicate standard deviation (n=3).



Supplementary Figure 9 | Tet1 is required for efficient NS cell reprogramming by Nanog. a, Schematic overview of the reprogramming system using stable reprogramming intermediates and medium switch from serum+LIF to 2i/LIF. **b**, Expression of endogenous Tet1 in NS+rOKM+Nanog transgenic reprogramming intermediates two days after transfection with siTet1 or non-targeting siRNA (siNT) control. Error bars indicate standard deviation (n=3). **c**, Tet1 knockdown significantly reduces the efficiency of reprogramming in intermediates transgenic for Nanog.

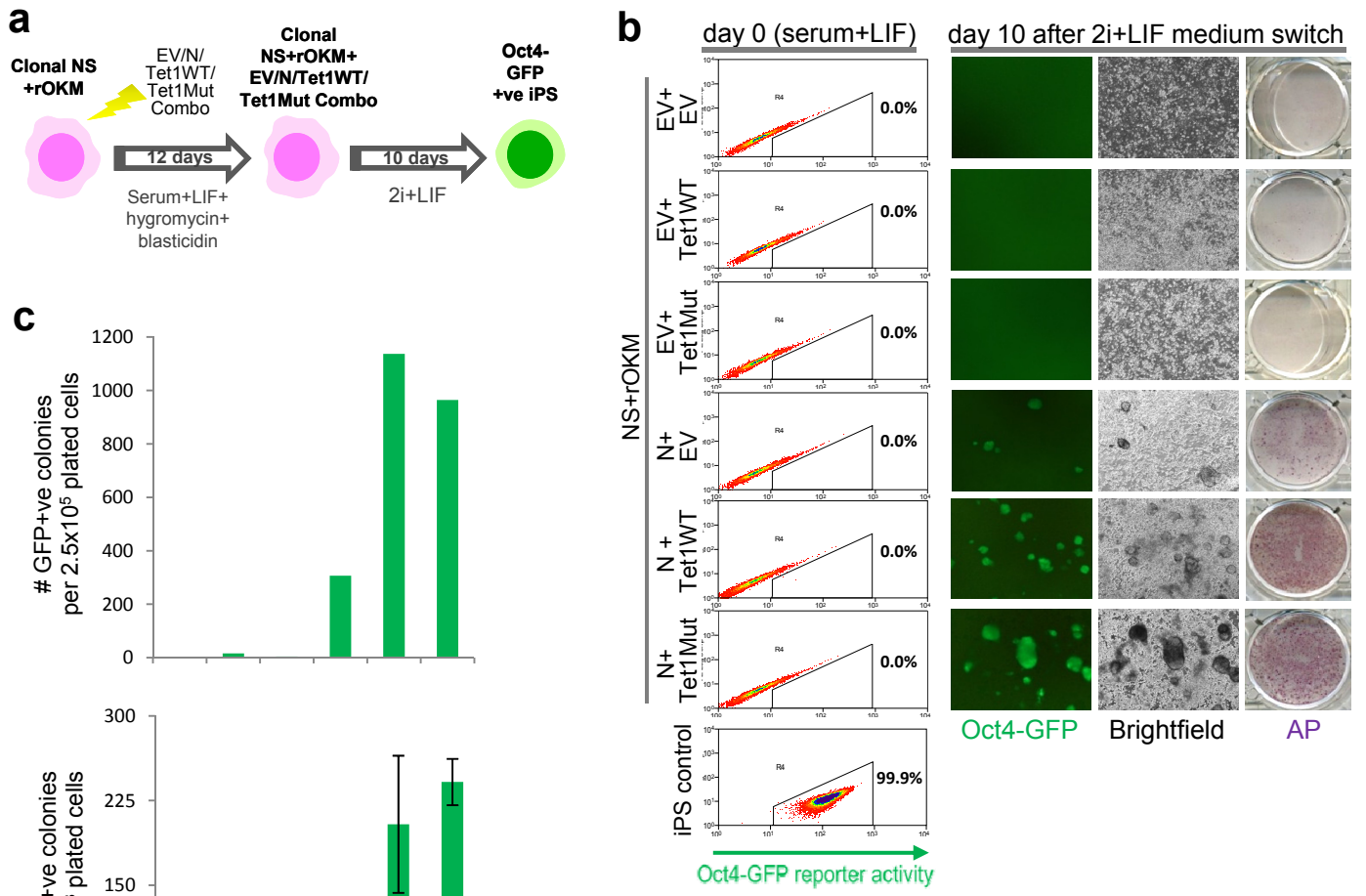


FIGURE S10 (part 1)

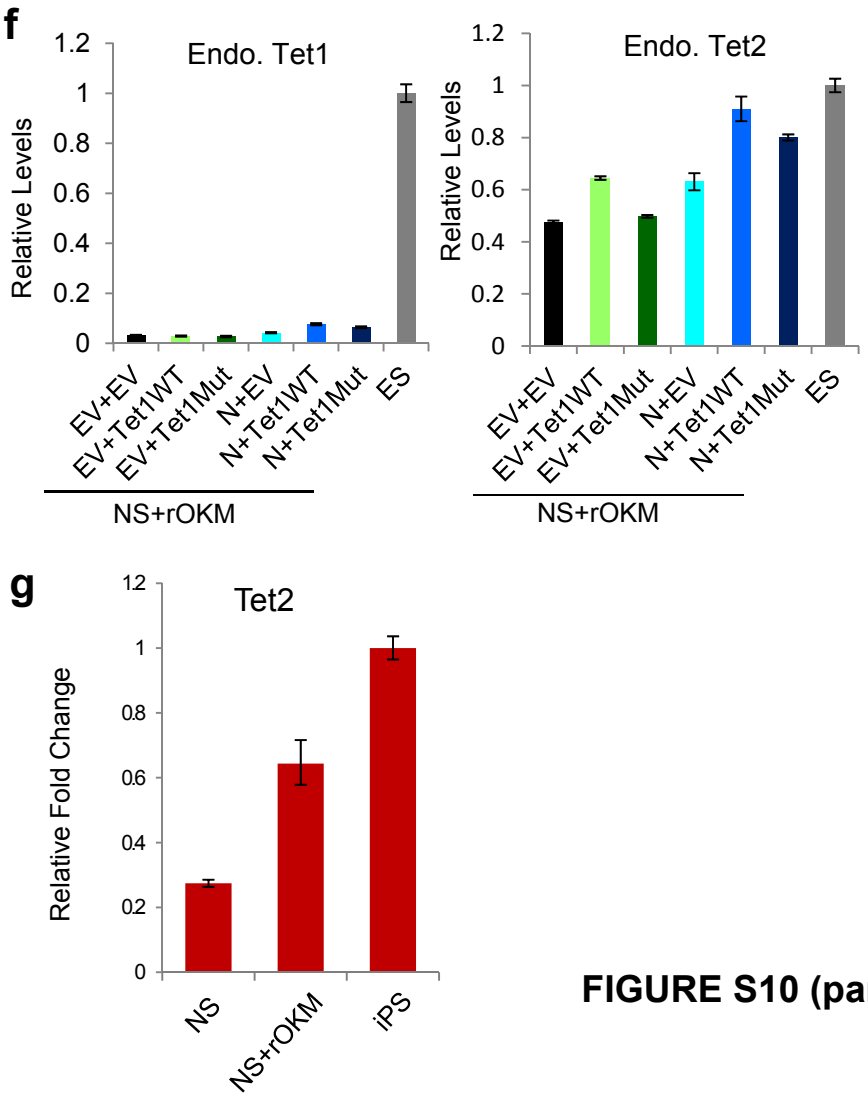
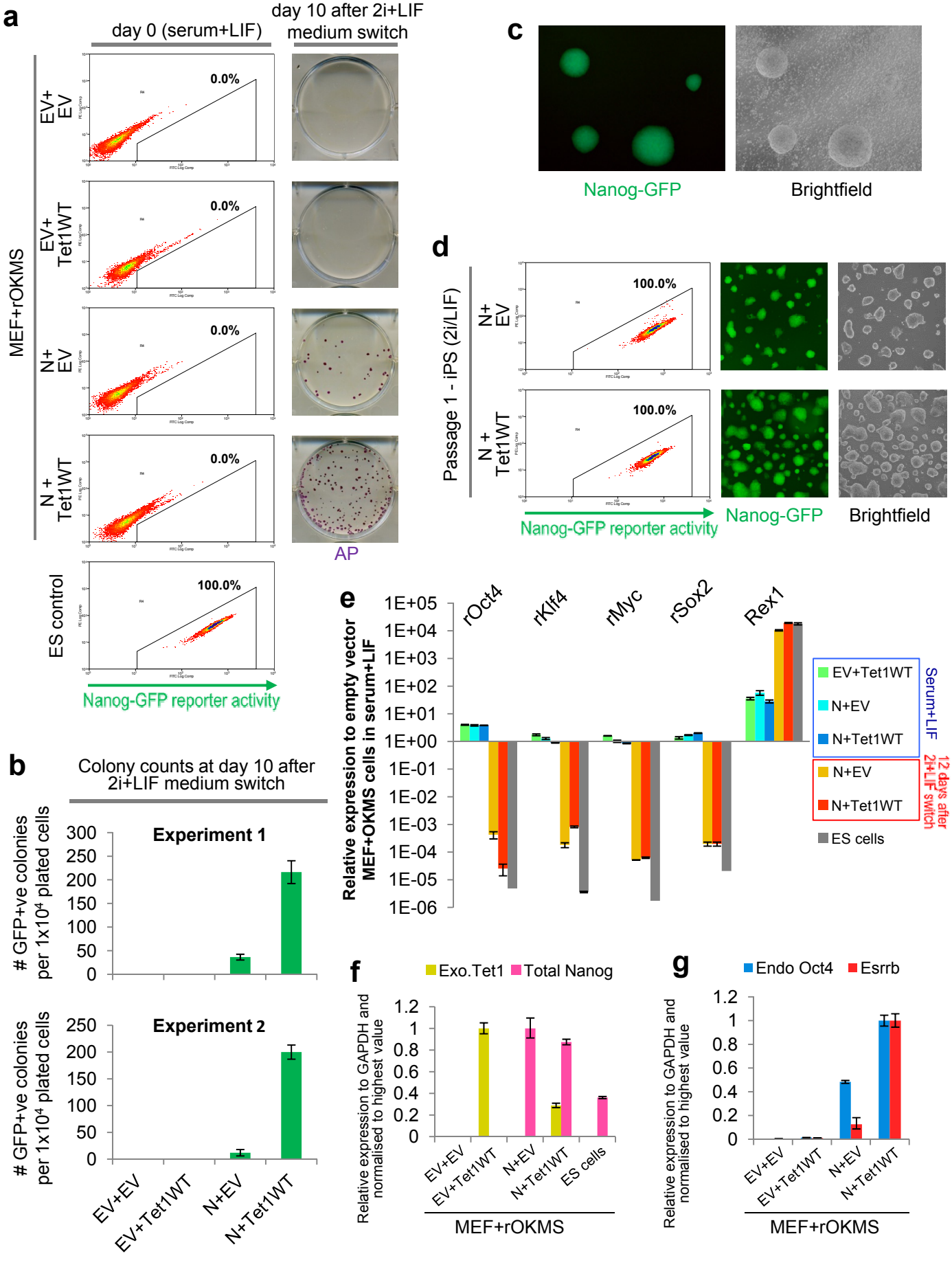


FIGURE S10 (part 2)

Supplementary Figure 10 | Synergy between Nanog and Tet1WT/Tet1Mut during NS cell reprogramming. **a**, Depiction of the strategy to test the Nanog and Tet1 partnership in reprogramming efficiency. Various combinations of empty vector (EV), Nanog (N), wild-type Tet1 (Tet1WT), and mutant Tet1 (Tet1Mut) PB transgenes were used to generate stable clones before medium switch to 2i/LIF. **b,** Both wild-type and mutant Tet1 enhance Nanog-dependent reprogramming. *Left* Absence of *Oct4*-GFP reporter activity in reprogramming intermediates cultured in serum+LIF. iPS cell control was cultured in 2i/LIF. *Right* Fluorescence images and alkaline phosphatase (AP) staining at day 10 of 2i/LIF treatment. **c,** Quantification of the number of iPS cell colonies at day 10 of 2i/LIF treatment in two additional independent experiments to Fig. 2c and Supplementary Fig. 10b. Error bars indicate standard deviation (n=3). **d**, Expression of Nanog, endogenous and exogenous Tet1, pluripotency markers, and retroviral transgenes in NS+rOKM transfectants (cold colors), derivative iPS cells (warm colors), and control ES cells (grey). Error bars indicate the standard deviation of fold changes relative to empty vector (EV+EV) NS+rOKM transfectants (n=3). **e**, Immunofluorescence analysis with an antibody against trimethylated lysine 27 on histone 3 (H3K27me3) in female reprogramming cell intermediates (Serum+LIF) and iPS cells (2i+LIF). **f**, Endogenous Tet2, but not endogenous Tet1, is upregulated in NS+rOKM transfectants exogenously expressing Nanog and Tet1WT/Tet1Mut. The color codes correspond to those in **d**. Error bars indicate standard deviation (n=3). **g,** Relative Tet2 expression in NS, NS+rOKM, and iPS cells. Error bars indicate standard deviation (n=3).

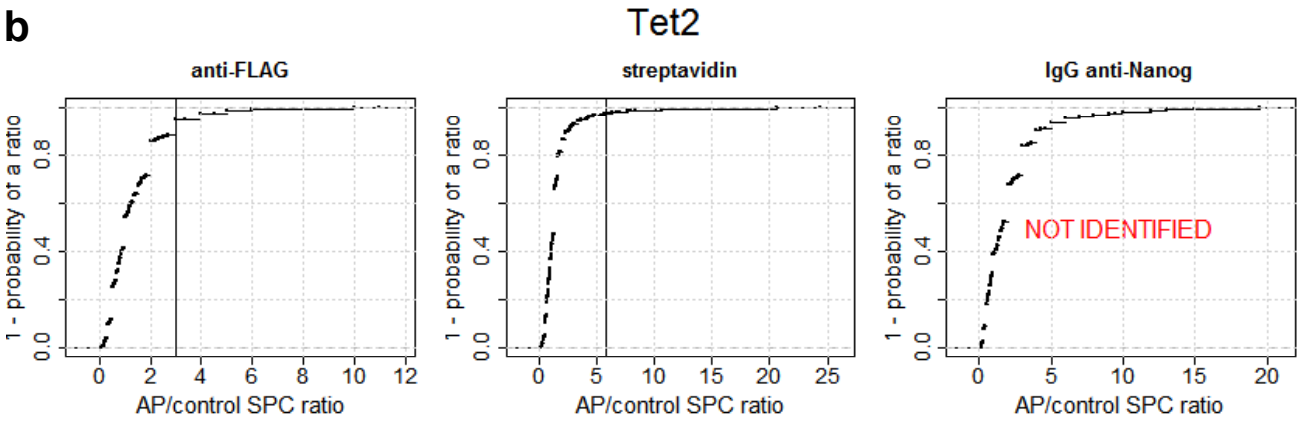

a
day 0 (serum+LIF)
day 10 after 2i+LIF medium switch
MEF+rOKMS
EV+ EV
EV+ Tet1WT
N+ EV
N + Tet1WT
ES control
0.0%
100.0%
AP
Nanog-GFP reporter activity
b
Colony counts at day 10 after 2i+LIF medium switch
GFP+ve colonies per 1x10⁴ plated cells
Experiment 1
Experiment 2
EV+EV
EV+Tet1WT
N+EV
N+Tet1WT
c
Nanog-GFP
Brightfield
d
Passage 1 - iPS (2i/LIF)
N+ EV
N + Tet1WT
100.0%
Nanog-GFP reporter activity
Nanog-GFP
Brightfield
e
Relative expression to empty vector MEF+OKMS cells in serum+LIF
rOct4
rKlf4
rMyc
rSox2
Rex1
EV+Tet1WT
N+EV
N+Tet1WT
ES cells
Serum+LIF
12 days after 2i+LIF switch
f
Exo.Tet1
Total Nanog
Relative expression to GAPDH and normalised to highest value
MEF+rOKMS
g
Endo Oct4
Esrrb
Relative expression to GAPDH and normalised to highest value
MEF+rOKMS

Supplementary Figure 11 | Nanog and Tet1 synergistically enhance the reprogramming efficiency of MEFs. a, Tet1 enhances Nanog-dependent reprogramming. *Left,* Absence of *Nanog*-GFP reporter activity in reprogramming cell intermediates (rOKMS)[11,40] transfected with various combinations of two of the following: Nanog (N), Tet1 wild-type (Tet1WT) and empty vector (EV) piggyBac transgenes, and cultured in serum plus LIF. ES cell control was cultured in 2i/LIF. *Right,* Alkaline phosphatase (AP) staining at day 10 of 2i/LIF treatment. **b**, Quantification of the number of iPS cell colonies at day 10 of 2i/LIF treatment in two independent experiments. Error bars indicate standard deviation (n=3). **c**, Fluorescence and bright-field images of N+Tet1WT transgenic colonies at day 10 of 2i/LIF treatment. **d,** Passage 1 N+EV and N+Tet1WT iPS cells generated from induction in **a**. *Left* panels show flow cytometry analysis of iPS cells cultured in 2i/LIF with 100% activity of the *Nanog*-GFP reporter. *Middle* and *Right* panels show fluorescence and bright-field images of transgenic iPS cell lines. **e**, Expression of retroviral transgenes and pluripotency marker Rex1 in MEF+rOKMS transfectants (cold colors), derivative iPS cells (warm colors), and control ES cells (grey). Error bars indicate the standard deviation of fold changes relative to empty vector (EV+EV) MEF+rOKMS transfectants (n=3). **f,** Expression of Nanog and exogenous Tet1 in MEF+rOKMS transfectants and in control ES cells. Error bars indicate standard deviation (n=3). **g,** Transcriptional priming of *Esrrb* and *Oct4* in N+Tet1WT in MEF+rOKMS transfectants. Error bars indicate standard deviation (n=3).

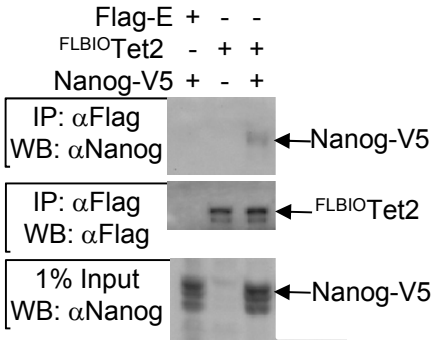
a

Pulldown/IP	Tet2
^{FLBIO}Nanog (ESCs)/SA#1	13(2)
^{FLBIO}Nanog (ESCs)/SA#2	16(2)
^{FLBIO}Nanog (ESCs)/SA#3	20(2)
^{FLBIO}Nanog (ESCs)/FLAG	2(0)
^{end}Nanog (ESCs)/Ab	0(0)

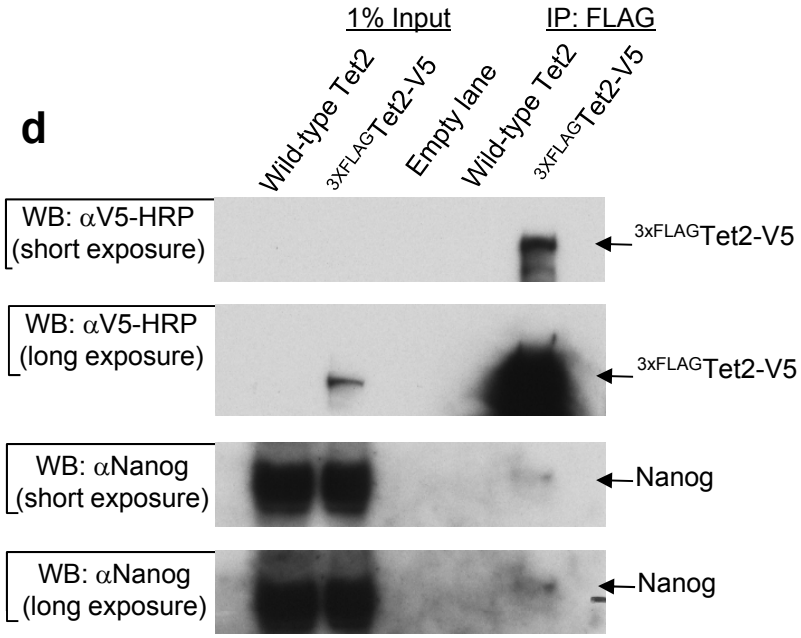
b



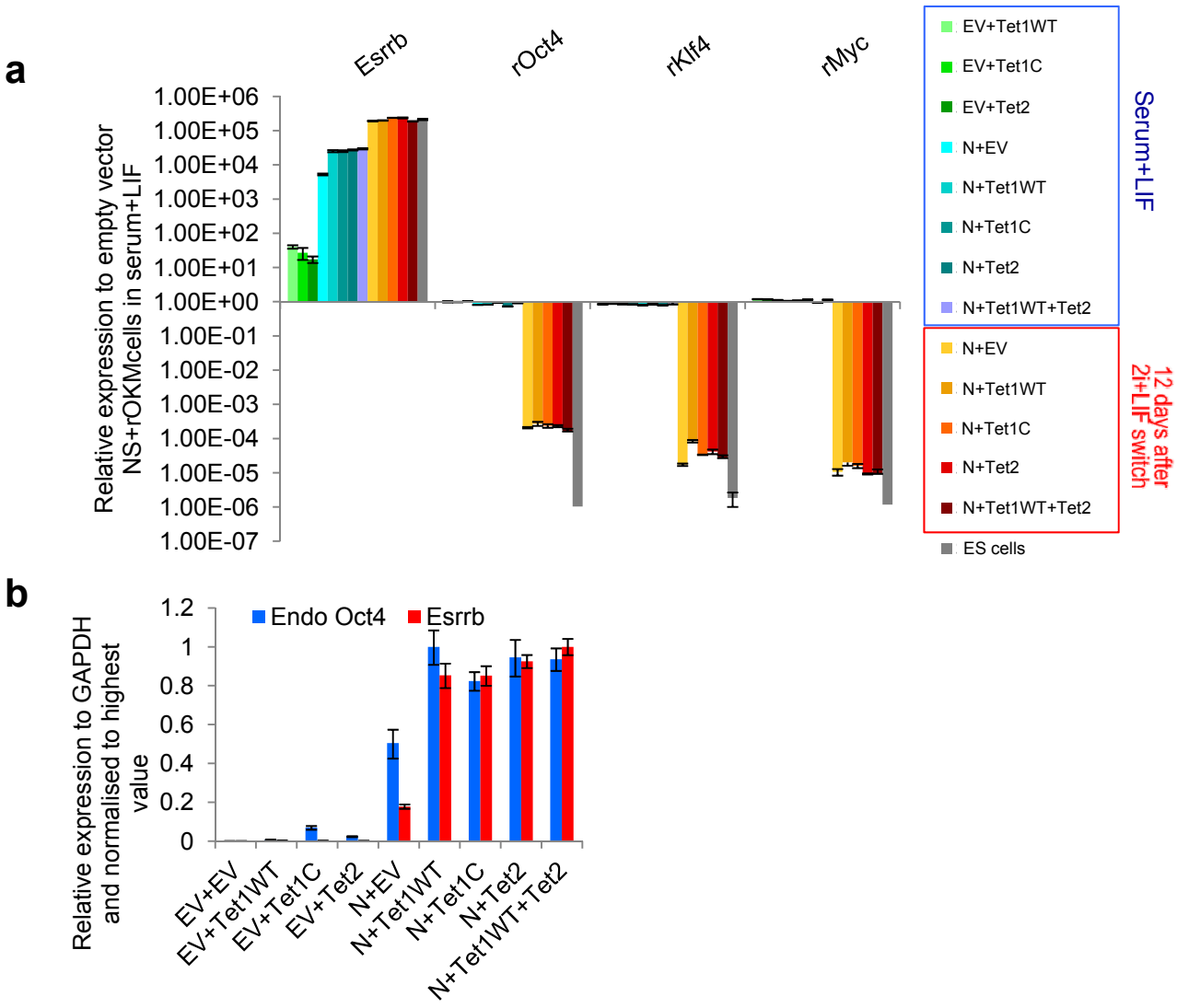
c



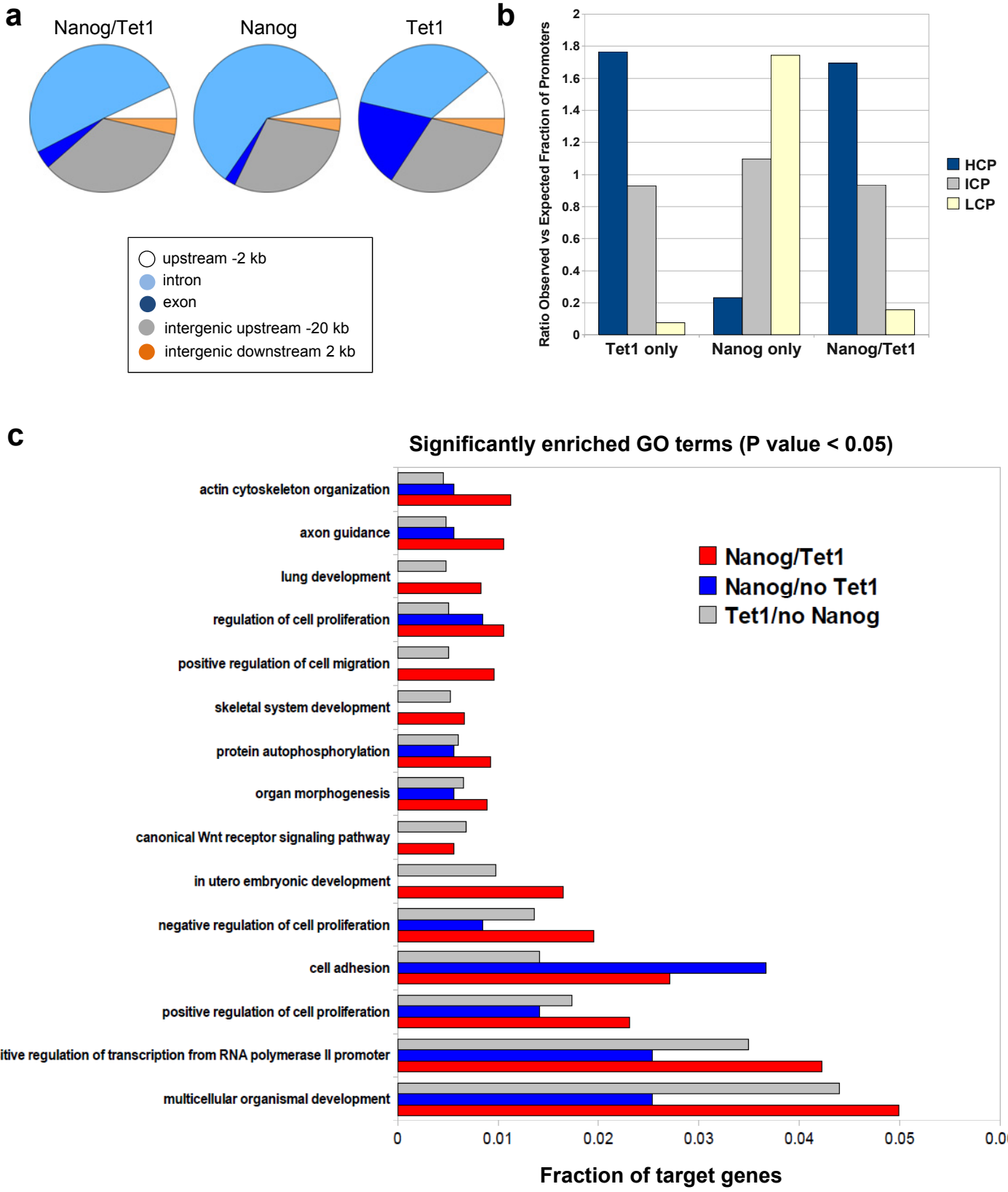
d



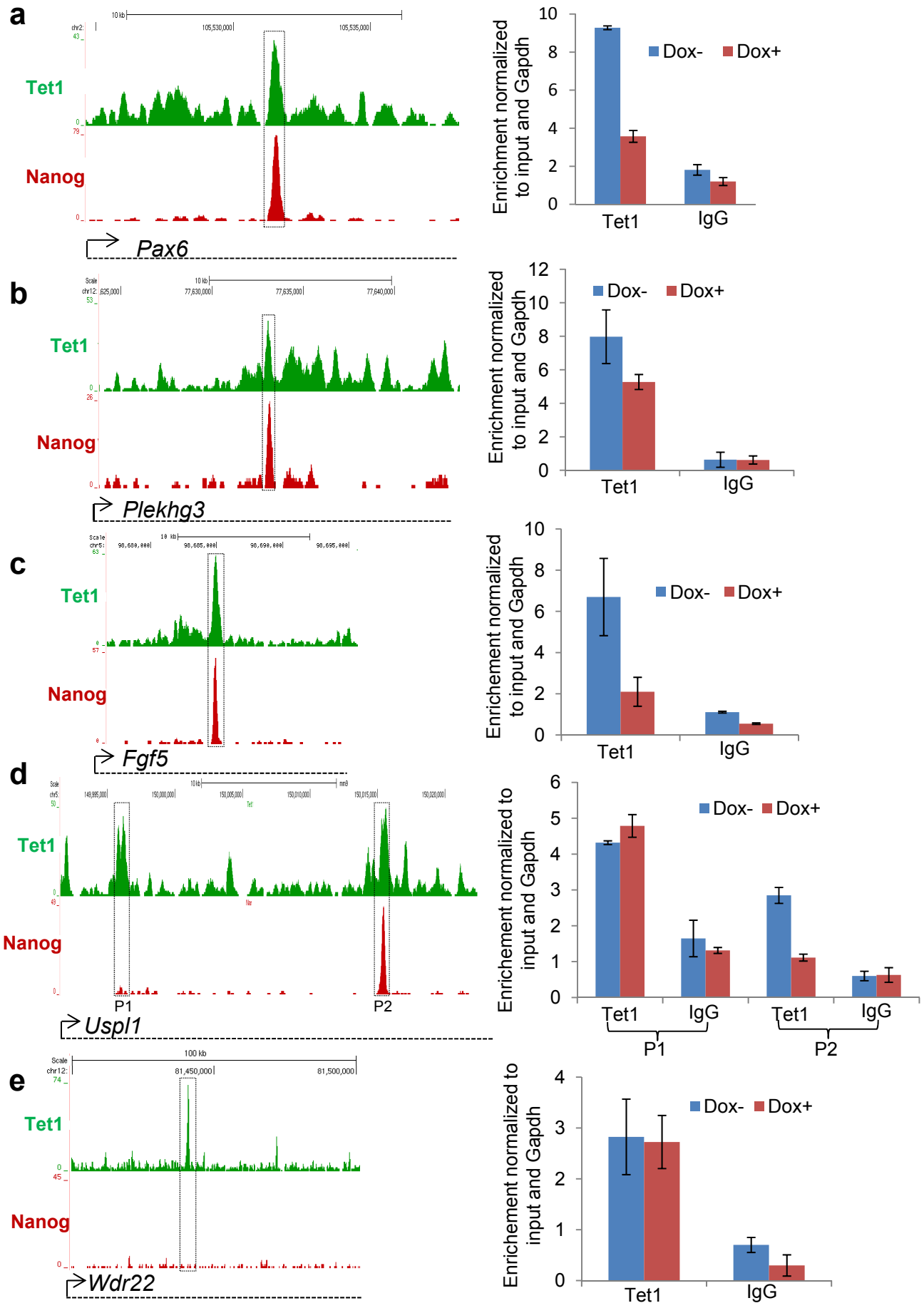
Supplementary Figure 12 | Validation of the Nanog-Tet2 interaction. a, List of peptides for Tet2 in our three independent AP-MS studies. The number within brackets indicates the background peptide(s) identified by MS in control pull-down samples. **b,** Empirical distribution functions of AP/control spectral counts (SPC) ratios for the three APs with anti-FLAG, streptavidin-conjugated, and IgG anti-Nanog beads, respectively. Each point represents a separate protein. The vertical lines represent the AP/control SPC ratios for Tet2. **c**, Validation of the Nanog-Tet2 interaction by coIP in HEK293T cells. **d,** Validation of the Nanog-Tet2 interaction by coIP in ES cells stably expressing a ^{3xFLAG}Tet2-V5 transgene under the endogenous promoter.



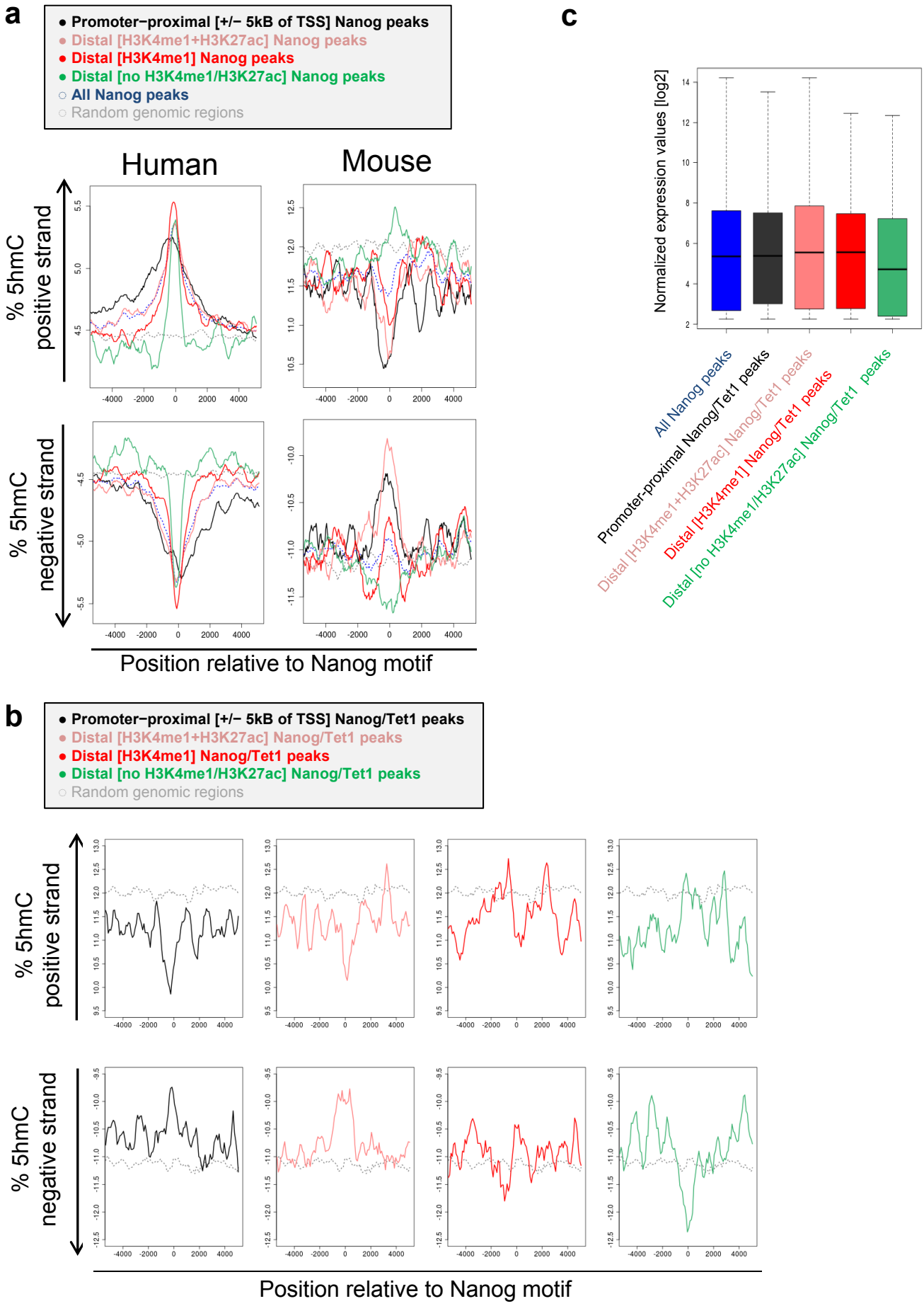
Supplementary Figure 13 | Synergy between Nanog and Tet2/Tet1C during somatic cell reprogramming. a, Expression of the pluripotent marker *Esrrb* and retroviral transgenes in NS+rOKM transfectants (cold colors), derivative iPS cells (warm colors), and control ES cells (grey). Error bars indicate standard deviation of fold changes relative to empty vector (EV+EV) NS+rOKM transfectants (n=3). **b**, Transcriptional priming of *Oct4* and *Esrrb* during reprogramming by various combinations of empty vector (EV), Nanog, Tet1WT, Tet1C, and Tet2. Error bars indicate standard deviation (n=3).


a
Nanog/Tet1
Nanog
Tet1
upstream -2 kb
intron
exon
intergenic upstream -20 kb
intergenic downstream 2 kb
b
Ratio Observed vs Expected Fraction of Promoters
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Tet1 only
Nanog only
Nanog/Tet1
HCP
ICP
LCP
c
Significantly enriched GO terms (P value < 0.05)
actin cytoskeleton organization
axon guidance
lung development
regulation of cell proliferation
positive regulation of cell migration
skeletal system development
protein autophosphorylation
organ morphogenesis
canonical Wnt receptor signaling pathway
in utero embryonic development
negative regulation of cell proliferation
cell adhesion
positive regulation of cell proliferation
positive regulation of transcription from RNA polymerase II promoter
multicellular organismal development
Nanog/Tet1
Nanog/no Tet1
Tet1/no Nanog
0
0.01
0.02
0.03
0.04
0.05
0.06
Fraction of target genes

Supplementary Figure 14 | Genomic distribution of Nanog and Tet1 overlapping binding sites. a, Common binding sites of Nanog and Tet1 are preferentially associated with gene bodies (introns+exons). Overlapping ChIP-seq peaks within 20 kb upstream and 2 kb downstream of the TSS were assigned to genes. 20-2 kb upstream are annotated as intergenic (upstream), 2-0 kb upstream as promoter, and 0-2 kb downstream as intron, exon, or intergenic (downstream). **b**, Promoters of common targets of Nanog and Tet1 correspond predominantly to those with high CpG density (HCP) (p-value = 2.770166e-141, hypergeometric distribution), rather than low CpG content (LCP) or intermediate CpG content (ICP). For a definition of the different classes of CpG content promoters see Ref[41]. **c**, Gene ontology (GO) analysis of the common targets of Nanog and Tet1. GO terms with a p-value < 0.05 and their corresponding fractions of target genes are shown.


a
Tet1
Nanog
Pax6
Enrichment normalized to input and Gapdh
Dox-
Dox+
Tet1
IgG
b
Tet1
Nanog
Plekhg3
Enrichment normalized to input and Gapdh
Dox-
Dox+
Tet1
IgG
c
Tet1
Nanog
Fgf5
Enrichement normalized to input and Gapdh
Dox-
Dox+
Tet1
IgG
d
Tet1
Nanog
P1
P2
Uspl1
Enrichement normalized to input and Gapdh
Dox-
Dox+
Tet1
IgG
Tet1
IgG
P1
P2
e
Tet1
Nanog
Wdr22
Enrichement normalized to input and Gapdh
Dox-
Dox+
Tet1
IgG

Supplementary Figure 15 | Nanog recruits Tet1 to common target genes in mouse ES cells. a-d, Tet1 and Nanog share many common genomic loci of a subset of target genes as shown in main Fig. 4**a-c**. Here we show that Tet1 binding to the shared loci of *Pax6* (**a**), *Plekhg3* (**b**), *Fgf5* (**c**), and *Uspl1* (**d**) is appreciably reduced upon Nanog depletion. *Left*, Tet1 and Nanog overlapping peaks. *Right*, ChIP-qPCR analysis of Tet1 binding to the overlapping peaks shown on the left panel in NgcKO cells before (Dox-) and after (Dox+) Nanog removal. **e,** Tet1 binding to the non-overlapping peak of the *Wdr22* locus is Nanog-independent. Note that a similar non-overlapping peak was also present in the *Uspl1* locus (P1 in **d**) where Tet1 binding is also independent of Nanog. ChIP was performed using anti-Tet1 (Millipore) in NgcKO cells in the absence (-) and presence (+) of doxycycline (Dox). IgG serves as a control. Error bars indicate standard deviation (n=3).


a
Promoter−proximal [+/− 5kB of TSS] Nanog peaks
Distal [H3K4me1+H3K27ac] Nanog peaks
Distal [H3K4me1] Nanog peaks
Distal [no H3K4me1/H3K27ac] Nanog peaks
All Nanog peaks
Random genomic regions
Human
Mouse
% 5hmC positive strand
% 5hmC negative strand
Position relative to Nanog motif
c
Normalized expression values [log2]
All Nanog peaks
Promoter-proximal Nanog/Tet1 peaks
Distal [H3K4me1+H3K27ac] Nanog/Tet1 peaks
Distal [H3K4me1] Nanog/Tet1 peaks
Distal [no H3K4me1/H3K27ac] Nanog/Tet1 peaks
b
Promoter−proximal [+/− 5kB of TSS] Nanog/Tet1 peaks
Distal [H3K4me1+H3K27ac] Nanog/Tet1 peaks
Distal [H3K4me1] Nanog/Tet1 peaks
Distal [no H3K4me1/H3K27ac] Nanog/Tet1 peaks
Random genomic regions
% 5hmC positive strand
% 5hmC negative strand
Position relative to Nanog motif

Supplementary Figure 16 | Comparison of Nanog/Tet1 peaks with base-resolution analysis of 5hmC in the ES cell genome. a, Nanog binding sites in mouse and human (H1) ES cells were classified according to (i) their proximity to gene promoters with distal sites defined as those positioned at least 5 kb from the transcription start site (TSS), and (ii) the presence of histone modification peaks within +/-800 nt of the Nanog ChIP-seq peak summit. The combined presence of H3K27ac and H3K4me1 is considered to mark an active enhancer chromatin signature, while the presence of H3K4me1 alone marks poised enhancers[21]. 5hmC profiles were obtained from genome-wide Tet-assisted bisulfite sequencing (TAB-Seq) in mouse and human (H1) ES cells[22], and are centered at all Nanog motifs within +/-200 bp of the Nanog peak summit. Note that 5hmC enrichment on the negative strand increases in the negative direction. **b**, 5hmC profiles at Nanog/Tet1 binding sites in mouse ES cells classified according to the presence of H3K4me1 and H3K27ac (see **a**). **c**, Absolute expression of genes proximal to the enhancer categories shown in (**b**) was obtained from microarray analysis in mouse ES cells[23]. Normalized expression values are shown in $\log_2$ scale in boxplots. Error bars represent one standard deviation above and below the mean of the data.

Supplementary Tables

Supplementary Table 1

Summary of all cell lines and beads used in the three AP-MS studies performed.

Nanog affinity epitope	Cell line AP	Cell line control	Beads AP	Beads control
Flag	NGB19 (Nanog knockout with ^{FLBIO}Nanog and BirA transgenes)	The parental line of NGA2 (i.e., Nanog$^{flox/flox}$) before Cre excision and no ^{FLBIO}Nanog transgene introduction	Anti-Flag agarose (M2, Sigma) beads	Anti-Flag agarose (M2, Sigma) beads
Biotin	NGB6/10/19 lines (Nanog knockout with ^{FLBIO}Nanog and BirA transgenes)	NGA2 (Nanog knockout with ^{FLBIO}Nanog transgene)	Streptavidin agarose beads (Invitrogen)	Streptavidin agarose beads (Invitrogen)
Native antibody	J1 ES cells	J1 ES cells	Protein G agarose beads (Roche) with anti-Nanog IgG (Bethyl Laboratories)	Protein G agarose beads (Roche) with control IgG (Millipore)

Supplementary Table 2

Summary of MS data on 27 high-confidence candidate proteins (Nanog is the bait).

	CCP score	q-value	SPC, control in brackets			1-CP		
			streptavidin beads	anti-Flag beads	IgG anti-Nanog beads	streptavidin beads	anti-Flag beads	IgG anti-Nanog beads
Nanog	23.6	0.00	12.7(0)	9(0)	12(0)	0.007	0.002	0.005
Zfp609	23.2	0.00	19.3(0)	9(0)	9(0)	0.003	0.002	0.018
Mki67	21.1	0.00	1.7(0)	10(0)	14(0)	0.083	0.002	0.003
Emsy	20.3	0.00	6.7(0)	10(0)	8(0)	0.014	0.002	0.026
Nacc1	18.0	0.00	7.3(0)	6(0)	4(0)	0.011	0.005	0.064
Jmjd1c	16.5	0.00	36.3(12)	3(0)	12(0)	0.079	0.027	0.005
Brca2	16.1	0.00	1.3(0)	3(0)	12(0)	0.106	0.027	0.005
Tet1	15.9	0.00	9.7(0)	5(0)	6(2)	0.009	0.006	0.306
Ncor2	15.7	0.00	23.3(0)	1(0)	3(0)	0.001	0.140	0.094
Sgol2	15.5	0.00	2.7(0)	4(0)	6(0)	0.053	0.010	0.038
Hnrnpm	15.0	0.01	59.3(13)	23(2)	33(9)	0.046	0.004	0.156
Qser1	12.9	0.03	6.7(0)	1(0)	4(0)	0.014	0.140	0.064
Mga	12.7	0.03	2.3(0)	3(0)	3(0)	0.060	0.027	0.094
Bptf	12.4	0.04	5(0)	4(2)	8(0)	0.024	0.297	0.026
Gatad2a	12.2	0.04	15(9)	17(3)	7(0)	0.292	0.024	0.031
Arid3b	11.9	0.04	5.3(0)	1(0)	3(0)	0.020	0.140	0.094
Rbm14	11.8	0.04	8.3(1)	1(0)	4(0)	0.031	0.140	0.064
Hdac2	11.6	0.04	11.3(6)	8(2)	7(0)	0.190	0.054	0.031
Skiv2l2	10.4	0.07	4(1)	7(2)	4(0)	0.099	0.115	0.064
Pou5f1/Oct4	9.6	0.09	1.7(0)	3(0)	3(2)	0.083	0.027	0.554
Lmnb1	9.4	0.10	12.3(2)	6(2)	10(3)	0.039	0.130	0.288
Mta2	9.4	0.10	15.3(6)	12(7)	5(0)	0.104	0.316	0.046
Rfc1	9.3	0.09	1.3(0)	7(4)	5(0)	0.106	0.317	0.046
Trrap	8.9	0.11	3.7(0)	5(7)	3(0)	0.033	0.653	0.094
Sall4	8.9	0.11	52.3(29)	39(15)	25(5)	0.189	0.121	0.092
Mta3	8.6	0.11	16.3(10)	16(7)	4(0)	0.295	0.139	0.064
Wdr18	8.5	0.11	3(1)	4(1)	2(0)	0.137	0.121	0.163

Supplementary Table 3

Observed versus expected overlapping genomic binding sites of Nanog and Tet1 in ES cells. Randomization trials were performed for comparisons between three ChIP-Seq datasets for Tet1[18-19] and two for Nanog[16-17]. N and C refer to the use of Tet1 antibodies raised against N-terminal or C-terminal peptides of Tet1 protein, respectively.

	Nanog ChIP-seq dataset	# of peaks	Tet1 ChIP-seq dataset	# of peaks	# of Nanog peaks in overlap	# of peaks in overlap expected
1	Chen, 2008	2971	Williams, 2011 (N)	13246	320	95.8
2	Chen, 2008	2971	Wu, 2011	20780	606	133.3
3	Chen, 2008	2971	Williams, 2011 (C)	21522	622	132.4
4	Marson, 2008	10310	Williams, 2011 (N)	13246	802	394.7
5	Marson, 2008	10310	Wu 2011	20780	1269	564.1
6	Marson, 2008	10310	Williams, 2011 (C)	21522	1770	775.3

Supplementary Table 4

Genomic coordinates of Nanog and Tet1 co-bound sites. See the separate spreadsheet document.

Supplementary Table 5

List of Taqman probes, qRT-PCR/ChIP-qPCR primers, shRNA oligos, and siRNA target sequences.

Primer sequence	Primer name
Taqman® qRT-PCR analysis	
Nanog	Applied Biosystems ID: Mm02384862_g1
Tet1 (human)	Applied Biosystems ID: Hs00286756_m1
Tet1 (murine)	Applied Biosystems ID: Mm01169087_m1
Klf2	Applied Biosystems ID: Mm01244979_g1
Esrrb	Applied Biosystems ID: Mm00442411_m1
Rex1	Applied Biosystems ID: Mm03053975_g1
Retr. cMyc-F	TGGTACGGGAAATCACAAGTTTGTA
Retr. cMyc-R	GGTCATAGTTCCTGTTGGTGAAGTT
Retr. cMyc-probe	FAM-CCCTTCACCATGCCCC-MGB
Retr. Oct4-F	TGGTACGGGAAATCACAAGTTTGTA
Retr. Oct4-R	GGTGAGAAGGCGAAGTCTGAAG
Retr. Oct4-probe	FAM-CACCTTCCCCATGGCTG-MGB
Retr. Klf4-F	TGGTACGGGAAATCACAAGTTTGTA
Retr. Klf4-R	GAGCAGAGCGTCGCTGA
Retr. Klf4-probe	FAM-CCCCTTCACCATGGCTG-MGB
GAPDH	Applied Biosystems ID: 4352339E
SYBR® Green qRT-PCR analysis	
Nanog-F	TTGCTCTTTCTGTGGGAAGG
Nanog-R	CCAGGAAGACCCACACTCAT
Tet1-F (murine)	ATTGAGGTGGAGAAGTGGG
Tet1-R (murine)	GGAGAAGGGTTGGTTTGC
GAPDH-F	ACCCAGAAGACTGTGGATGG
GAPDH-R	CACATTGGGGGTAGGAACAC
CD19-F (human)	GCTCAAGACGCTGGAAAGTATTATT
CD19-R (human)	GATAAGCCAAAGTCACAGCTGAGA
CD37-F (human)	GTGGCTGCACAACAACCTTATTT
CD37-R (human)	GCCTAACGGTATCGAGCGAG
CD45-F (human)	CCCCATGAACGTTACCATTTG
CD45-R (human)	GATAGTCTCCATTGTGAAAATAGGCC
CRIPTO-F (human)	AGAAGTGTTCCCTGTGTAAATGCTG
CRIPTO-R (human)	CACGAGGTGCTCATCCATCA
DNMT3b-F (human)	GTCAAGCTACACACAGGACTTGACAG
DNMT3b-R (human)	AGTTCGGACAGCTGGGCTTT
NANOG-F (human)	CCAACATCCTGAACCTCAGCTAC
NANOG-R (human)	GCCTTCTGCGTCACACCATT
OCT4-F (human)	TCGAGAACCGAGTGAGAGGC
OCT4-R (human)	CACACTCGGACCACATCCTTC
SOX2-F (human)	CACACTGCCCCTCTCACACAT
SOX2-R (human)	CATTTCCCTCGTTTTTCTTTGAA
TLE1-F (human)	TGTCTCCCAGCTCGACTGTCT
TLE1-R (human)	AAGTACTGGCTTCCCCTCCC
GAPDH-F (human)	TCTGCTCCTCCTGTTCGACA
GAPDH-R (human)	AAAAGCAGCCCTGGTGACC
SYBR® Green ChIP-qPCR analysis	
Gapdh-F	AAGCTCATGAGGCACAGAATGGTC
Gapdh-R	TGGGTACATGGTGACTTTCCTAGGC
Esrrb peak-F	GAAGAACTGAATTGCTTGGG
Esrrb peak-R	GGACAGGATGCACTTTGGA
Pax6 peak-F	AGGAAGGCTTTGTGGAGGC

Pax6 peak-R	CCGAGGACTGGGTAATCTGC
Plekhg3 peak-F	TGTCCTCTAGTCTGCCGCTCTT
Plekhg3 peak-R	ATGGCTTTCAGGTGACTTTGA
Fgf5 peak-F	ACCTGGCTCCTCTTGTCTTCC
Fgf5 peak-R	CCATCCCACCTGTGCTTGA
Uspl1 Peak1-F	GAAGTGCTCTGGTCTTCCCCTTCC
Uspl1 Peak1-R	GCCCAAGAGCTACCCCACAAACA
Uspl1 Peak 2-F	GGGGTGGCCCTGAAATAAC
Uspl1 Peak 2-R	CACAGAAGGGTGACCAGCAGA
WDR22 Peak-F	CCCCAGGTGGAAGGCGTTGT
WDR22 Peak-R	TTGGTGCGTCGTGAAATTGAG
shRNA sequences	
shNanog	GACAGTGAGGTGCATATAC
shTet1-1	GCAGATGGCCGTGACACAAAT
shTet1-2	GCTCATGGAGACTAGGTTTGG
shTet1-3	AGAAGCAGTGTACACATAATA
siRNA target sequences	
siTet1 pool	CAGGTGGGTTTGCAGAAACAA, AAGGTTGGATTTGATCACACA, AAGACAGACTTTAACAACAAA, CTCGAGTTGCATCAACCTTAA
siTet2 pool	ATGCCTCGGGTTCATATTTGA, TACCGTGACTACATCACCATA, CATGCAGTATTTCCCGAATAA, TCCGAAGGATGCAAACGGGAA
AllStars Negative Control (siNT)	Cat number 1027281

Supplementary Table 6

Primers used for hydroxymethyl-sensitive qPCR.

Esrrb Intron FW	ACTTTAGATGGGACCGCCATTATC	Intronic region of the Esrrb gene
Esrrb Intron RV	GGGATTTGCTTCAATAGGACTTCA	
Esrrb Promoter FW	TTTTCCTTCACAGGGTCAGG	Promoter of the Esrrb gene
Esrrb Promoter RV	ATGGCTTCTTGAGTGGCGTA	
Oct4_FW	ACAGGCTTTGTGGTGCGATG	Promoter of the Oct4 gene
Oct4_RV	GGTGGGTGGAGGAGCAGAG	