Endogenous fluctuations of OCT4 and SOX2 bias pluripotent cell fate decisions

Daniel Strebinger [1], Elias T. Friman [1°], Cédric Deluz [1°], Subashika Govindan [1], Andrea B. Alber [1], David M. Suter [1*]

[1]Sponsored Stem Cells Research Chair (UPSUTER), The Institute of Bioengineering (IBI), School of Life Sciences, Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland

° These authors contributed equally to this work

* Corresponding author: david.suter@epfl.ch

Keywords: SOX2, OCT4, endogenous protein fluctuations, pluripotency, self-renewal, neuroectoderm, mesendoderm, cell fate, chromatin accessibility

Abstract

The SOX2 and OCT4 transcription factors are key regulators of embryonic stem (ES) cell self-renewal and differentiation, but how temporal fluctuations in their endogenous expression levels bias lineage commitment is unknown. We generated knock-in reporter fusion ES cell lines allowing to measure endogenous SOX2 and OCT4 protein fluctuations and determine their impact on mesendodermal and neuroectodermal commitment. Surprisingly, small differences in endogenous SOX2 and OCT4 levels impacted cell fate commitment in G1 but not in S phase. While SOX2 fluctuations had a minor impact on neuroectodermal commitment, elevated OCT4 levels at the onset of differentiation strongly biased ES cell towards both neuroectoderm and mesendoderm at the expense of self-renewal and primitive endoderm. Genome-wide measurements of chromatin accessibility revealed OCT4 level-dependent priming of differentiation-associated enhancers. Finally, CRISPR-Cas9 knock-out of an OCT4 binding site in a key *Eomes* enhancer abolished the ability of OCT4 to promote mesendodermal differentiation. Our study demonstrates how small endogenous fluctuations of transcription factors prime cell fate decisions in a cell cycle-specific manner by modulating

chromatin accessibility at regulatory regions, and thus represent a major source of heterogeneity in the ability of individual ES cells to respond to differentiation cues.

Main Text

Embryonic stem (ES) cells often exhibit both asynchrony and divergence in fate commitment when subjected to the same differentiation cues [1]. This obscures the effect of instructive signals on cell fate decisions and limits the generation of pure ES cell-derived cell populations for future regenerative medicine applications [1]. Intercellular variability in expression of cell fate regulators constitutes a potential source of bias for the differentiation potential of individual cells, however this remains poorly explored in the context of phenotypically homogeneous stem cell cultures. The transcription factors SOX2 and OCT4 (also known as POU5F1) are strictly required and collaborate to maintain ES cells in an undifferentiated state [2–6]. While they were also reported to play an antagonistic role in driving ES cells towards the neuroectodermal (NE) and mesendodermal (ME) fates [6,7], these conclusions were largely based on overexpression/knockdown or indirect correlations from fixed cells. How endogenous expression levels of SOX2 and OCT4 fluctuate over time, regulate each other and how this biases ES cell fate decisions is unknown.

To monitor SOX2 and OCT4 protein fluctuations in single living cells, we used CRISPR-Cas9 genome editing [8,9] to generate knock-in ES cell lines in which the endogenous SOX2 and OCT4 proteins are fused to luminescent or fluorescent tags. For fluorescence measurements, SOX2 and OCT4 were tagged with SNAP and HALO tags that bind to fluorescent ligands (Fig.1a,b, Supplementary Fig.1a-c). These knock-ins were generated in a previously established reporter cell line for ME and NE commitment, thus allowing us to monitor SOX2 and OCT4 protein levels in live cells as well as differentiation outcomes [10]. We confirmed that heterozygously tagged OCT4 is a good proxy for total OCT4 (Fig.1c and Supplementary Fig.1d). The luminescent cell line was generated by knock-in of Nanoluc [11] (NLUC) in fusion to both alleles of *Sox2* and P2A-Firefly Luciferase (FLUC) in fusion to one allele of *Sox1* to monitor NE commitment (Fig.1d,e and Supplementary Fig.1a-c). Quantitative

immunofluorescence analysis confirmed that the tags did not strongly alter SOX2 and OCT4 expression level distributions (Fig.1f), protein half-lives (Fig.1g and [12]), or growth rates (Fig.1h). SOX2 and OCT4 protein levels are positively correlated in single cells [13], however the mechanism underlying this correlation is unclear. To address this question, we established Doxycycline (dox)-inducible ES cell lines to overexpress fusions of SOX2 or OCT4 to YPet. Cells were treated with dox and fixed at various time points, and SOX2 and OCT4 were labelled by immunofluorescence. While upon YPet-OCT4 overexpression, SOX2 levels remained stable (Supplementary Fig.2a,b), SOX2 overexpression increased OCT4 levels already four hours after SOX2 induction (Supplementary Fig.2c,d). SOX2 levels decreased as a function of overexpressed YPet-SOX2 over time (Supplementary Fig.2e), suggesting negative regulation of endogenous SOX2 by YPet-SOX2. Overexpression of SOX2-SNAP but not a SOX2 mutant lacking its DNA binding domain (YPet-SOX2-delDBD) reduced SOX2 protein and mRNA levels, suggesting rapid negative transcriptional autoregulation (Fig.2a,b). Of note, as *Sox2* mRNA levels are very low at 6 hours after dox induction (Fig.2b), SOX2-NLUC decay is mainly driven by its half-life (8.2 +/- 1h (SE)), very close to the values obtained for SOX2-SNAP and published values for SOX2, suggesting that NLUC fusion does not perturb the stability of SOX2. In contrast, *Oct4* mRNA level was unaffected by SOX2 overexpression, thus we asked whether SOX2 overexpression could increase OCT4 stability. We measured OCT4 half-life by pulse-labelling OCT4-HALO with Halo-Sir 647 in the SBROS cell line after SOX2 overexpression, which increased the half-life of OCT4 by 50 % (Fig.2c). To determine whether endogenous variations in SOX2 and OCT4 levels regulate each other, we labelled SBROS cells with SNAP-SiR647 and the Halo-TMR dye. We then sorted cells in G1 based on DNA content, allowing minimizing the effects of cell cycle progression on differences in SOX2 and OCT4 levels. Cells were gated for intermediate SOX2 or OCT4 expression levels, and low/high OCT4 levels or low/high SOX2 levels, respectively (Fig.2d, Supplementary Fig.2f). In cells sorted for SOX2-high or SOX2-low levels, OCT4 levels 8 hours after sorting started to increase and decrease, respectively (Fig.2e, Supplementary Fig.2g), suggesting that endogenous SOX2 fluctuations regulate endogenous OCT4 levels. In contrast, SOX2 levels did

not change as a function of OCT4 levels (Fig.2f, Supplementary Fig.2g). Thus, SOX2 levels regulate *Sox2* transcription and OCT4 protein stability.

We then measured temporal fluctuations and intercellular variability of absolute protein copy numbers of SOX2-NLUC (SNSF cell line) by luminescence microscopy, using a signal calibration approach we reported previously [14]. As expected, SOX2 levels doubled over one cell cycle (Fig.2g, Supplementary Fig.3a), and SOX2 concentrations calculated after normalization on an inferred nuclear volume as described in [13] where constant on average (Supplementary Fig.3b). However, SOX2 concentrations fluctuated over a 2-3-fold range in single cells (Fig.2g). To measure the time scale of SOX2 concentration fluctuations, individual cells were assigned with a rank based on their initial SOX2 level (Fig.2h). We then used a rank-based autocorrelation function to determine the time scale of protein level fluctuations (the mixing time [15]) using data from either two or one full cell cycle, yielding a SOX2 mixing time on the order of one cell cycle (Fig.2h and Supplementary Fig.3c,d). Since < 2-fold changes in SOX2 expression were reported to alter pluripotency [16], our data suggest that rapid readjustment of SOX2 levels allows pluripotency maintenance despite fluctuation amplitudes of 2-3 fold. We performed analogous experiments using the SBROS cell line to monitor OCT4 levels over the cell cycle by live fluorescence microscopy (Fig.2j, Supplementary Fig.3e,f), and found similar OCT4 mixing times (Fig.2k,l). Thus, both SOX2 and OCT4 display 2-3 fold, rapid expression level fluctuations in the pluripotent state.

We next aimed to determine how endogenous SOX2 levels tune the probability of NE differentiation by monitoring SOX2-NLUC levels and SOX1-P2A-Fluc expression after removal of 2i and LIF (Fig.3a). Traces were then aligned and normalized for cell cycle duration *in silico* (using linear resampling of the time variable). During the cell cycle before Sox1+ cells appeared, higher SOX2 levels at the beginning and end, but not in the middle of the cell cycle were predictive of NE differentiation (Fig.3b, marked by *), suggesting a role for SOX2 in NE commitment specific to the M-G1 transition, in line with earlier findings from our laboratory [10].

We next investigated whether different SOX2 levels at the onset of differentiation impact NE and ME commitment. We sorted SBROS G1-gated cells stained with SNAP-Sir647 for low, medium and high SOX2 levels (Fig.3c and Supplementary Fig.4a). Cells were then released

from self-renewing conditions by seeding in the absence of 2i and LIF, and after four days NE and ME commitment was assessed by flow cytometry using the SOX1-P2A-eGFP and the BRA-P2A-mCherry reporters as readout. The fraction of eGFP+ cells scaled with initial SOX2 levels (Fig.3d), suggesting that SOX2 levels at the time of self-renewal release govern NE fate. In contrast, SOX2 levels had only a weak impact on ME commitment (Fig.3d).

Next, we sorted G1-gated cells stained with Halo-TMR in OCT4-low and OCT4-high subpopulations (Fig.3c) and cultured them for four days in the absence of LIF and 2i. Surprisingly, we found a large difference in NE and ME commitment between these populations (Fig.3e), even though average OCT4 levels differed by < 2-fold between them (Supplementary Fig.4b). To interrogate the potential causal relationship between high OCT4 levels and increased NE/ME commitment, a cell line allowing for inducible expression of SNAP-OCT4 in the SBR background was treated with dox for 12h to allow for a short pulse of OCT4 overexpression, followed by sorting of G1-gated cells and four days of culture without 2i and LIF (Fig.3f and Supplementary Fig.4c). Strikingly, this led to a ~ 2-fold increase in the fraction of mCherry+ cells, however the fraction of eGFP+ cells remained unchanged (Fig.3g and Supplementary Fig.4c). Given that NE commitment was strongly inhibited when we treated SNAP-OCT4 cells with dox throughout differentiation (Supplementary Fig.4d,e), this suggests that supraphysiological OCT4 levels at the onset of differentiation may counteract the impact of physiologically high OCT4 levels before release from self-renewing conditions.

We next measured the combinatorial impact of SOX2 and OCT4 on differentiation by sorting G1-gated cells into four different subpopulations (Fig.3h and Supplementary Fig.4f) followed by four days of differentiation (Fig.3i and Supplementary Fig.4g). As expected, SOX2-high/OCT4-high (SHOH) cells were the most efficient to differentiate towards both cell fates, SOX2-low/OCT4-high (SLOH) cells were less capable to differentiate towards NE, and OCT4-low (SHOL and SLOL) populations were strongly impaired in differentiating towards both fates. To ask whether endogenous SOX2 and OCT4 level variability impacts NE and ME commitment during later cell cycle stages, we performed the same experiments with S phase-gated cells. While overall differentiation efficiency was increased towards NE but not ME as compared to G1 phase-gated cells (Supplementary Fig.4h), the impact of SOX2 and OCT4 level variability

on NE and ME commitment was strongly decreased (Fig.3j and Supplementary Fig.4i). This suggests that SOX2 and OCT4 act in a cell cycle stage-specific manner to control NE and ME commitment.

We next aimed at identifying the earliest differences in cell fate commitment between cells differing in SOX2 and OCT4 levels. We thus measured changes in the expression of markers of the naïve state, primed state and primitive endoderm in SHOH, SHOL, SLOH and SLOL cells 24 hours after removal of 2i and LIF by RT-QPCR. Cells with low OCT4 levels displayed increased naïve pluripotency and primitive endoderm (PrE) marker expression, while cells with high OCT4 levels displayed increased primed pluripotency marker expression and higher expression of AscII, a master regulator of neuronal differentiation (Fig.3k). We next measured the expression levels of NANOG and GATA6 by immunofluorescence 24 hours after release from self-renewing conditions. In line with the QPCR results, cells with low OCT4 levels show a higher percentage of NANOG and GATA6 (Fig.3l). While PrE and definitive endoderm share most of their markers, definitive endoderm usually emerges only after 48-96 hours of differentiation [17], thus the expression of GATA4 and GATA6 we observed most likely reflects differentiation towards PrE. Taken together, these results suggest that cells with low OCT4 levels are biased towards the naïve state or the PrE fate as compared to cells with high levels of OCT4.

We then aimed to identify the molecular mechanisms by which small and transient endogenous fluctuations of SOX2 and OCT4 result in major biases in differentiation potential. As SOX2 and OCT4 were shown to regulate chromatin accessibility [18,19], we reasoned that variability in their expression level could alter the chromatin accessibility landscape and prime cells for different fates. We thus performed three biological replicates of ATAC-Seq in SHOH, SHOL, SLOH and SLOL cells. We quantified the accessibility in 81'132 open regions and compared High vs Low conditions for OCT4 and SOX2 as well as for SHOH vs SLOL (FDR < 10%). 3'914 loci (4.8%) were significantly up- or downregulated upon changes in OCT4 alone (538 loci), SOX2 alone (1'259 loci), or OCT4 and SOX2 together, i.e. SHOH vs SLOL (2'117 loci). We grouped the identified loci (Fig.4a) into OCT4-regulated (Groups 1 and 4), SOX2-regulated (Groups 2 and 5) and co-regulated (Groups 3 and 6) genes. All groups contained

differentiation-associated genes (Fig.4b) but Group 1 was the most enriched for genes involved in differentiation processes (Fig.4c, Supplementary Fig.5a-e). In contrast, pluripotency-associated super-enhancers were unaffected by OCT4 and SOX2 levels (Fig.4d). Loci in which accessibility was positively correlated to OCT4 and SOX2 levels were enriched for OCT4 and SOX2 binding, while those that were negatively correlated showed less overlap with OCT4 and SOX2 ChIP-Seq peaks and were enriched for H3K4me3, marking active promoters and transcription start sites (Fig.4e). Loci with increased accessibility in high OCT4 conditions (Group 1) lose accessibility upon OCT4 depletion, in line with OCT4 directly regulating accessibility at these sites (Supplementary Fig.5f). This suggests that fluctuations of endogenous OCT4 and SOX2 lead to temporal changes in chromatin accessibility, and that high OCT4 levels result in the opening of differentiation-associated enhancers.

Finally, we used CRISPR-Cas9 genome editing in the SBROS cell line to delete ~80 bp encompassing an OCT4 binding site in an enhancer critical for EOMES expression during differentiation of ES cells [20] and strongly dependent on OCT4 for its accessibility (Fig.4f and Supplementary Fig.5g,h). G1 phase-sorted OH and OL population were differentiated for 4 days and analysed by flow cytometry. Strikingly, ME commitment was strongly reduced and independent of OCT4 levels (Fig.4g) in this cell line. We also performed the same experiment with a second *Eomes* KO clone (clone B). While this cell line expressed sfGFP constitutively because of the stable sfGFP vector integration that happened during cell line generation, we could still estimate changes in the % of mCherry+ cells (online Methods). Similar to *Eomes* KO A cells, *Eomes* KO B cells were largely deficient and OCT4-independent in ME commitment (Supplementary Fig.5i), and unsorted *Eomes* KO A and B cells were also almost completely unable to commit to ME (Supplementary Fig.5j-l). This suggests that OCT4 priming of this critical *Eomes* enhancer is required for ME commitment of ES cells.

While gene expression fluctuations are increasingly being recognized as an important source of protein level variability in single cells, how these impact cellular functions remains largely unclear. Here we show that endogenous fluctuations in levels of pluripotency regulators have a major impact on ES cell differentiation potential. While Nanog displays prolonged, large

amplitude fluctuations that alter ES cell differentiation potential in serum + LIF [21], these are caused by transitions between naïve and primed ES cells [13] and thus reflect fluctuations between different phenotypic states. In contrast, ES cells maintained in a naïve state display small amplitude, transient fluctuations of SOX2 and OCT4, which nonetheless deeply impacts differentiation potential. Some of our findings such as enhancement of NE commitment by elevated SOX2 levels [6,7,22], or the impact of low OCT4 levels on increased self-renewal ability [23,24], or deficient progression to the epiblast stage of the embryo [25] are in line with earlier studies. However, the enhancement of NE commitment by high endogenous OCT4 levels and the absence of SOX2 impact on ME commitment contrasts with previous work [7]. Thus, overexpression or indirect correlations from fixed samples have to be interpreted with caution to assess the function of cell fate regulators. In Fig.5 we propose a revised model of the roles of SOX2 and OCT4 in pluripotency maintenance and germ layer commitment.

The large impact of small amplitude, transient OCT4 fluctuations on differentiation is surprising, suggesting a sensitive and rapid downstream mechanism modulating cell responsiveness to differentiation. While changes in the chromatin accessibility landscape as a function of OCT4 levels offer a convincing explanation for these observations, the reason for differential responses of pluripotency regulatory elements as compared with differentiation-related enhancers is unclear. The potential role for cooperativity with other pluripotency TFs or differential affinity of OCT4 binding sites will require further investigation. Finally, the fact that cells in G1 but not S phase are sensitive to SOX2 and OCT4 level variability raises the possibility that these mainly act shortly after mitosis to re-open closed enhancer regions, in line with their reported pioneer transcription factor activity [26] and their essential function in cell fate decisions at the Mitosis-G1 transition [10,18].

Acknowledgments

This work was supported by the Swiss National Science Foundation (PP00P3_1144828 and PP00P3_172905) to D.M.S, and the generous support of the Fondazione Teofil Rossi di Montelera e di Premuda and an anonymous donor advised by CARIGEST SA. We thank the Swiss Federal Institute of Technology (EPFL), the Biomolecular Screening Facility (EPFL-BSF)

and the EPFL Bioimaging and Optics Core Facility (EPFL-BIOP) for assistance in imaging and the EPFL Flow Cytometry Core Facility (EPFL-FCCF) for the fluorescence activated cell sorting.

Author contributions:

Conceptualization, D.S. and D.M.S.; Methodology, D.S., D.M.S., E.T.F., S.G., A.B.A.; Formal Analysis, D.S., D.M.S., E.T.F.., A.B.A.; Investigation, D.S., D.M.S., C.D., E.T.F., S.G., A.B.A.; Resources, D.M.S.; Writing – Original Draft, D.S. and D.M.S.; Writing – Review & Editing, D.S., D.M.S., C.D., E.T.F., S.G., A.B.A., Funding Acquisition, D.M.S.; Supervision, D.M.S.

Declaration of interests

The authors declare no competing interests.

Online Methods

Cell Culture and Cell Line Generation

Cell Culture

The E14 cell line (kindly provided by Didier Trono, EPFL) was used for all ES cell experiments, except for the SBR (Deluz et al. 2017) sub-cell lines that were generated from CGR8 ES cells (Sigma-Aldrich, Cat#07032901-1VL).

Cells were routinely cultured on dishes coated with 0.1% gelatin type B (Sigma G9391-100G), in GMEM (Sigma G5154-500ML) supplemented with 10% ES-cell qualified fetal bovine serum (Gibco 16141-079), nonessential amino acids (Gibco 11140-050), 2mM L-glutamine (Gibco 25030-024), sodium pyruvate (Sigma S8636-100ML), 100µM 2-mercaptoethanol (Sigma 63689-25ML-F), penicillin and streptomycin (BioConcept 4-01F00-H), homemade leukemia inhibitory factor (LIF), CHIR99021 (Merck 361559-5MG) at 3µM and PD184352 (Sigma

PZ0181-25MG) at 0.8µM. Cells were passaged by trypsinisation (Sigma T4049-100ML) every two to three days at a ratio of 1:10.

For imaging experiments, ES cells were cultured on dishes coated with 5µg/ml E-Cadherin, in N2B27 medium supplemented with LIF, CHIR99021 at 3µM and PD184352 at 0.8µM (N2B27+2iLIF). E-Cadherin coating was performed as previously described (Nagaoka et al. 2006). Briefly, 5µg/ml E-Cadherin (R&D 8875-EC or R&D 748-EC) in PBS (with Ca2+ and Mg2+; Amimed 3-05K00-I) were added to the culture vessel and incubated for 90 minutes at 37°C. Just before seeding, the E-Cadherin solution was removed, the surface of the vessel rinsed once with PBS and filled with the appropriate cell culture medium.

N2B27 medium was prepared by combining DMEM/F12 (Gibco 11320-033) + N2 supplement (Gibco 17502-001) medium with Neurobasal (Gibco 21103-049) + B27 supplement (Gibco 17504-001) medium, supplemented with penicillin (1000IU/ml) and streptomycin (1000mg/ml), 2mM L-Glutamine and 0.1mM 2-mercaptoethanol.

HEK 293T cells (ATCC) were cultured in DMEM (Gibco 41966-029) supplemented with 10% fetal bovine serum (Gibco 10270-106), penicillin and streptomycin and passaged every 2 days at a ratio of 1:8.

For the selection of transduced and transfected cells, the following antibiotic concentrations were used: 8µg/ml of Blasticidin (Gibco A11139-03), 2µg/ml of Puromycin (Gibco A11138-03) and 200µg/ml of Hygromycin B (Invitrogen 10687010). ES cells were transfected using X-treme gene 9 transfection reagent.

Cell Line	Reference
SBR	[10]
SBROS and derived cell lines (*Eomes* KO A and B)	This study
SNSF	This study
Calibration cells	[14]

Table 1: Cell Lines used in this study

Generation of Knock-in Cell lines

The SBROS and SNSF cell lines were generated using CRISPR-Cas9 mediated homology-directed repair (HDR). The repair templates were designed to contain a knock-in cassette flanked by homology arms (HAs) with the target sequence missing the endogenous STOP codon. Guide RNA sequences were designed to overlap with the endogenous STOP codon, and the repair templates contain mutations in the PAM sequence, thus ensuring that the repair plasmids are not cut (Supplementary Figure S1a).

The knock-in cassette (between the HAs) contains the coding sequence for the tag in frame with the protein of interest and a selection marker. For the SOX2 knock-ins, the cassette consisted of SNAP-IRES-Hygro (SBROS), or NLuc-loxP-P2A-Puro-sfGFP-loxP and NLuc-loxP-P2A-Bsd-sfGFP-loxP (SNSF). The knock-in cassette for OCT4 consists of a linker (WRAASRLTS)-Halo-IRES-Blasticidin (SBROS). The Sox1 knock-in cassette consists of P2A-FLuc-Stop-loxP-pGK-Hygro-Stop-loxP.

Guide RNAs targeting the Pou5f1, Sox2 and Sox1 loci were designed using the Zhang Lab toolbox (www.genome-engineering.org/crispr) and cloned into the pX330 vector (Sox2 and Oct4), expressing Cas9 and the guide RNA, or pX335 (Sox1), expressing Cas9n and the guide RNA [9]. The guide RNA sequences were 5'-GAC TGA GGC ACC AGC CCT CCC-3' for Pou5f1, 5'-GCA GCC CTC ACA TGT GCG ACA-3' for Sox2 and 5'-GAC GCA CAT CTA GCG CCG CG-3' for Sox1.

To generate the SBROS cell line, SBR cells [10] were transfected with pX330-Sox2 and pKI-SOX2-SNAP-IRES-Hygro at a 1:3 ratio. After two days, selection was started with Hygromycin B. After 11 days of selection, cells were stained with 12nM SNAP-SiR 647, and single cells were sorted for high SNAP expression into 96-well plates and grown out. One clone (SBRS) identified as homozygously targeted by PCR on genomic DNA was further validated by Western Blot and used to knock-in a Halo-Tag at the C-terminus of OCT4. To do so, this clone was transfected with pX330-Pou5f1 and pKI-OCT4-HALO-IRES-Bsd at a 1:3 ratio followed by blasticidin selection two days after transfection. Single colonies were then picked manually and

grown out, and one clone in which one allele was targeted as indicated by PCR on genomic DNA was further analysed by Western Blot.

To generate the SNSF cell line, E14 cells were co-transfected with pX330-Sox2 and pKI-Sox2-NLuc-loxP-P2A-Puro-sfGFP-loxP at a 1:3 ratio. After two days, selection with Puromycin was started. After six days of selection, cells were co-transfected with pX330-Sox2 and pKI-Sox2-NLuc-loxP-P2A-Bsd-eGFP-loxP, and selection with Blasticidin was initiated two days later. A homozygous blasticidin-resistant clone was identified by PCR on genomic DNA and subsequently recombined by transient transfection of a plasmid expressing Cre recombinase. Successful excision of the selection cassette was confirmed by PCR on genomic DNA. This resulting intermediate cell line (Sox2-NLuc cells) was co-transfected with pX335-Sox1 and pKI-Sox2-P2A-FLuc-loxP-pGK-Hygro-loxP, and selection with Hygromycin B was started two days later. Single clones were picked manually ten days later and the knock-in was confirmed using PCR on genomic DNA.

All knock-in and corresponding wild-type alleles were verified by Sanger sequencing of the PCR products. All sequences were preserved except for the presence of a single nucleotide insertion in the 3'UTR of the wild-type *Pou5f1* allele of the SBROS cell line.

Generation of the OCT4-binding site knockout in the *Eomes* enhancer

The SBROS *Eomes* enhancer knockout cell lines were generated using dual-guide CRISPR-Cas9-mediated excision of a ~80bp DNA fragment centered on an OCT4 binding site. SBROS cells were transfected with two pX330 plasmids expressing Cas9 and one sgRNA each (5'-GCT CCC ACC CCA CCC AAA CCG-3' and 5'- GGA AGG CAA TGC CAG GGT TT-3') as well as a plasmid expressing sfGFP. After two days, single sfGFP-positive cells were sorted into 96-well plates and expanded. Clones were screened by PCR (expected product size without deletion: 403bp) and two clones with deletion of the target sequence on both alleles were selected for further analysis. While Clone A was sfGFP-negative, clone B displayed stable sfGFP expression likely caused by the integration of the sfGFP expression vector after co-transfection. We then sequenced the PCR products of the genomic region encompassing the deleted region. Clone A yielded unclear sequencing results, suggesting that the deletions of

the two alleles were not identical. We thus ligated the PCR product in a plasmid pLVTRE3GMCS, transformed in competent E.Coli, picked 12 colonies and purified the plasmid. The 12 corresponding DNA purifications were sequenced and this resulted in two distinct sequences (6 clones each). In both cases, the target region was deleted and the two sequences differed only by a few nucleotides. Below we show the deleted sequences, and the predicted OCT4-binding motif is underlined:

GGTTTGGGTGGGGTGGGAGGAGGCCCTGGGAAAAACAGAATGCTAATGACCTTTTGAGTA GACGGAAGGCAATGCCAG

TTTGGGTGGGGTGGGAGGAGGCCCTGGGAAAAACAGAATGCTAATGACCTTTTGAGTAGA CGGAAGGCAATGCCAGGGTTT

For clone B, Sanger sequencing of the PCR product revealed a single unambiguous deletion:

GGTTTGGGTGGGGTGGGAGGAGGCCCTGGGAAAAACAGAATGCTAATGACCTTTTGAGTA GACGGAAGGCAATGCCAG

Lentiviral Vector Production and generation of stable cell lines

Lentiviral vectors were produced by Calcium Phosphate co-transfection of HEK 293T cells with the envelope (PAX2), packaging (MD2G), and lentiviral construct of interest. The viral vectors were concentrated 120-fold by ultracentrifugation as described previously [27]. Stable cell lines were generated by transducing 50,000 cells per well of a 24-well plate with 50μl of concentrated lentiviral vector particles. Antibiotic selection was started 48-72 hours later and maintained throughout passaging.

Molecular Biology Methods

DNA constructs and cloning

To generate the pKI Sox2-Nluc-loxP-P2A-Puro-sfGFP-loxP, two multiple cloning sites, the downstream one including a loxP site, were inserted into a pCMV backbone by Oligo annealing. Next, a P2A construct was inserted by oligo annealing between a ClaI and a BamHI site, and the Nluc or Fluc coding sequence fused to a loxP site was cloned between SpeI and

ClaI of the resulting plasmid. The selection cassette consisting of sfGFP-Puro or eGFP-Bsd was created using fusion PCR [28] and inserted between a BamHI and a XhoI site. Thereafter the Sox2 homology arms (HAs) were inserted. For the 5' HA, the PCR amplified 5' HA and the backbone were digested with BsmBI and the 3' HA was inserted by in-fusion cloning into an XbaI site. The pKI Sox1-P2A-Fluc-loxP-pGK-Hygro-loxP construct was based on a previously published plasmid pKI Sox1-P2A-loxP-eGFP-pGK-Hygro-loxP [10], in which eGFP was replaced by Fluc using restriction cloning with AclI and SacI. For all pX330 (Addgene #42230) and pX335 (Addgene #42335) constructs the vector was opened using BbsI and the guide RNAs were inserted using oligo annealing. The pKI Oct4-HALO-IRES-Bsd was based on pKI Oct4-Fluc-P2A-eGFP-Bsd, in which the Fluc-P2A-eGFP-Bsd cassette was replaced by Halo-IRESbsd using SpeI and NotI. To generate the pKI Sox2-SNAP-IRES-Hygro construct, pKI Sox2-Nluc-loxP-P2A-Puro-sfGFP-loxP was digested with NdeI and re-ligated, thus removing the Nluc-loxP-P2A-Puro-sfGFP cassette and resulting in pKI Sox2-NdeI. pLV pGK-rtTA3G-IRES-Hygro was digested with EcoRI and AgeI to remove the rtTA3G sequence, which was replaced with a P2A sequence by oligo annealing. The resulting vector pLV pGK-P2A-IRES-Hygro was digested with XbaI and the coding sequence for the SNAP-tag inserted, resulting in the pLV pGK-P2A-SNAP-IRES-Hygro construct. Next, the P2A-SNAP-IRES-Hygro cassette was amplified by PCR and used for in-fusion cloning with pKI-Sox2-NdeI linearized by PCR from the start of the 3' HA to the end of the 5' HA of SOX2 (excluding the stop codon). The resulting pKI-Sox2-P2A-SNAP-IRES-Hygro vector was amplified by inverse PCR to remove the P2A-SNAP cassette, which was then replaced by a SNAP-tag using in fusion cloning, resulting in the pKI SOX2-SNAP-IRES-Hygro construct. The different constructs used in this study are summarized in Table 2.

Name	Reference
pKI Sox2-Nluc-loxP-P2A-Puro-sfGFP-loxP	This study
pKI Sox2-Nluc-loxP-P2A-Bsd-eGFP-loxP	This study

pKI Sox2-SNAP-IRES-Hygro	This study
pKI Oct4-HALO-IRES-Bsd	This study
pKI Sox1-P2A-FLuc-loxP-pGK-Hygro-loxP	This study
pX330 Sox2	This study
pX330 Pou5f1	This study
pX335 Sox1	This study
pLV TRE3G-Ypet-Sox2	[10]
pLV TRE3G-Ypet-Oct4	[10]
pLV TRE3G-SOX2-SNAP	[10]
pLV TRE3G-YPet-Sox2-delDBD	[10]
pLV pGK-rtTA3G-IRES-Hygro	[14]
pLV pGK-rtTA3G-IRES-Bsd	[10]
pLV-pGK-Cre	[10]

Table 2: DNA constructs used in this study

Confirmation PCRs

For all knock-in and knock-out verification PCRs, genomic DNA was purified (Sigma G1N350-1KT) and subsequently used to identify clones with correctly targeted alleles. PCR was done using Phusion High Fidelity DNA Polymerase (ThermoScientific F530L). Primers are listed in Table 3.

Name	Sequence (5' -> 3')	Target
Sox1-KI_F2	GTG CCC CTG ACG CAC AT	In 5' HA
Sox1-KI_R2	CGC TGT GTG CCT CCT CTG	In 3' HA
Seq_Sox2_KI_fw	AGG TGC CGG AGC CCG	In 5'HA
Sox2_verif_rv_3'	GCA TGC TAG CCA CAA AGA AA	Downstream of 3' HA
COct4_3'_500bp_r	GAT CCA TAT GCC AGA ACT CCC AGA GTG ACA A	Downstream of 3' HA

(sf)GFP_verif_fw	CTC GGC ATG GAC GAG C	In sfGFP
Seqout_IRES_rv	AGA CAG GGC CAG GTT TCC	In IRES
COct4_5'_500bp_f	GAT CGT CGA CTA GCA CAA TCC CTT AGC GGT	Upstream of 3' HA
Eomes_KO_ver_f1	CAT AAG TAG ATC CGC GCT GC	Upstream of cut site
Eomes_KO_ver_r2	GGA TGG AGG GCA GGA ATT CT	Downstream of cut site

Table 3: Primers for Knock-in verifications

Western Blots

For Western Blotting, cells were trypsinised, collected by centrifugation and washed once in ice-cold PBS. 10 million cells were then resuspended in 500μl Hypotonic Buffer (20mM Tris-HCl pH7.4, 10mM NaCl, 3mM MgCl2), supplemented with 1mM PMSF (AppliChem A0999.0005) and Protease Inhibitors (Sigma P8340-5ML). After 15 minutes incubation on ice, cells were lysed by the addition of 25μl of 10% NP-40 (Roche 11.754599001) and subsequent vortexing for 15 seconds. Nuclei were collected by centrifugation (10 min, 4°C, 14.000 rpm) and lysed by resuspension in 30μl RIPA Buffer per million of cells (50mM Tris-HCl pH7.4, 1% NP-40, 0.5% Sodium Deoxycholate, 0.1% SDS, 2mM EDTA, 150mM NaCl), supplemented with PMSF (AppliChem A0999.0005) and Protease Inhibitors (Sigma P8340-5ML). Lysis was allowed to proceed for 30 minutes during which cells were incubated on ice. Every 10 minutes, the sample was vortexed and further incubated on ice. To separate the soluble nuclear protein from debris, lysed nuclei were centrifuged for 30 minutes (4°C, 14.000rpm). The protein concentration of the supernatant was determined by performing a Bicinchoninic acid assay (BCA) (ThermoFisher 23235) and 15μg of protein were mixed with Laemmli sample buffer (Invitrogen NP0007) and loaded on an SDS gel (BioRad 456-1094) for separation (SDS Running Buffer 25mM Tris, 190mM Glycine, 0.1%SDS). Proteins were subsequently transferred (Transfer Buffer 25mM Tris, 190mM Glycine, 20% Methanol, 0.1% SDS) from the gel onto a PVDF membrane (Merck IPVH07850) using a wet transfer system. The membrane was blocked with 5% milk (Roth T145.3) in PBS-T to reduce unspecific binding and incubated with the appropriate concentration of primary antibody overnight. The next day the membrane

was rinsed once with PBS-T, incubated with a secondary antibody in 5% milk in PBS-T and washed extensively with PBS-T. Finally, chemiluminescence was revealed using Clarity Western ECL Substrate (BioRad 170-5060) and the signal was detected on a Fusion FX 7 apparatus (Vilber). The antibodies and concentrations used are summarized in Table 4.

Target	Species	Dilution	Supplier
Primary Antibodies			
αSOX2	Rabbit	1:250	Invitrogen 48-1400
αOCT4	Mouse	1:200	Santa Cruz sc5279
Secondary Antibodies			
αMouse-IgG-HRP	Goat	1:10.000	Promega W402B
αRabbit-IgG-HRP	Goat	1:10.000	Promega W401B

Table 4: Antibodies used for Western Blotting

Imaging Methods

Immunofluorescence and Image Acquisition

ES cells were fixed for 15 to 30 min with ice-cold 2% PFA (AppliChem A0877,0500) in PBS, permeabilized and blocked with chilled PBS-Triton (0.5%, AppliChem A1388,0500) and 1% FBS for 30 - 60 min. Samples were incubated with the primary antibody in PBS and 1% FBS overnight at 4°C, washed twice in PBS, and incubated with the secondary antibody in PBS and 1% FBS for 45 - 60 min. Samples were then washed three times with 0.1% PBS-Tween (Fisher Scientific BP337-500), incubated with 1 μg/mL DAPI for 15 minutes, washed twice with 0.1% PBS-Tween and once with PBS. The antibodies used are listed in Table 5.

Immunofluorescence stainings were imaged using a 20x magnification objective (Olympus UPlanSApo 20x, NA 0.75) on an Olympus Cell XCellence or using a 20x magnification objective (Nikon PlanApo 20x, NA 0.75, CFI/60) on a GE InCell Analyzer 2200 apparatus. (GE #29027886, Biomolecular Screening Facility at EPFL)

Target	Species	Dilution	Supplier
Primary Antibodies			

αSOX2	Rabbit	1:200	Invitrogen 48-1400
αOCT4	Mouse	1:50	Cell Signaling #75463
αOCT4	Mouse	1:300	Santa Cruz sc5279
αNANOG	Rabbit	1:500	Abcam ab80892
αGATA6	Goat	1:20	AF1700
Secondary Antibodies			
αMouse-IgG-555	Donkey	1:1.000	ThermoFisher A31570
αMouse-IgG-488	Goat	1:1.000	ThermoFisher A11001
αMouse-IgG-647	Goat	1:1.000	ThermoFisher A21235
αRabbit-IgG-647	Chicken	1:1.000	ThermoFisher A21443
αGoat-IgG-647	Donkey	1:1.000	ThermoFisher A21447

Table 5: Antibodies used for Immunofluorescence staining

Live cell luminescence imaging

Time-lapse luminescence recordings were performed on an Olympus LuminoView LV200 microscope equipped with an EM-CCD cooled camera (Hamamatsu photonics, EM-CCD C9100-13), a 60x magnification objective (Olympus UPlanSApo 60x, NA 1.35, oil immersion) in controlled environment conditions (37°C, 5% CO2). One day before the experiment, cells were seeded on fluorodishes (WPI, FD35-100) coated with E-Cadherin in 2ml of N2B27+2iLIF. For quantitative NanoLuc imaging, knock-in cells were mixed with Calibration cells at a ratio of 1:10 as described previously [14]. The medium was supplemented with 1mM Luciferin (NanoLight Technology 306A) and 0.5µl of RealTime Glo Cell Viability Assay Substrate (Promega G9711). For imaging in pluripotency maintenance conditions, images were acquired every 299 s the NanoLuc channel and between 59 and 178 s in Firefly Luciferase channel with a cycle time of 8 to 17 minutes for up to 75 hours. For overexpression experiments of SOX2-SNAP and YPet-SOX2-delDBD, the exposure time in NLuc was 58 s and the cycle time 22 minutes. For imaging in differentiation conditions, images were acquired every 59 s in the NanoLuc channel and every 599 s in the Firefly Luciferase channel with a cycle time of 15 to 16 minutes for up to 70 hours.

For time-lapse imaging of SOX2-NLUC / SOX1-P2A-Fluc, we used an arbitrary threshold of 500 molecules per cell for at least 4 hours to classify cells as Sox1-positive.

Fluorescence Time-lapse microscopy

For time-lapse imaging of the induction kinetics of SOX2-SNAP and YPet-SOX2delDBD as well as for OCT4-HALO single cell imaging, cells were seeded in E-Cadherin-coated wells of an imaging-grade 96-well plate in N2B27+2iLIF. The next day, the medium was supplemented with SNAP-SiR647 dye (NEB S9102S) for SOX2-SNAP or HALO-SiR647 (gift from Suliana Manley, EPFL) at a concentration of 12nM for SOX2-SNAP imaging or 50nM for OCT4-HALO imaging. Cells were imaged using a 20x magnification objective (Olympus UPlanSApo 20x, NA 0.75) on an Olympus Cell XCellence in controlled conditions (37°C and 5% CO2) for 24 hours. For the induction experiments, doxycycline was added to a final concentration of 500ng/ml to induce transgene expression after one hour. For the OCT4-HALO imaging, since we observed a mild global fluorescence decrease in the cell population, we corrected for the loss of intensity on the population average to obtain single cell traces (Supplementary Fig. 3e,f).

Flow Cytometry and Fluorescence-Activated Cell Sorting (FACS)

Cell preparation for Fluorescence-Activated Cell Sorting (FACS)

SBROS cells were stained with the HaloTag TMR Ligand (Promega G8251) and SNAP-SiR647 dye (NEB S9102S) at a concentration of 100nM or 12nM, respectively, for one hour. Subsequently, cells were incubated in with Hoechst 33342 (Invitrogen H3570) at a concentration of 1.62μM for 15 minutes. Cells were then trypsinised, washed in PBS and resuspended in PBS/1% FBS for sorting on a BD FACSAria II.

Fluorescence Activated Cell Sorting and Flow Cytometry to assess cross-regulation of endogenous SOX2 and OCT4

To determine how endogenous SOX2 levels regulate OCT4 expression, we used the following sorting strategy: Cells were gated for G1 based on the Hoechst staining and on a narrow window of intermediate OCT4-HALO expression levels (25% of cells). This window was further subdivided into two windows defined by the highest or lowest 30% of SOX2-SNAP expression. (Supplementary Figure 2f). The converse strategy was used to determine how endogenous OCT4 levels regulate SOX2 expression. After FACS, cells were spun down, resuspended in N2B27 medium and seeded in a gelatinated 24-well plate in N2B27 or N2B27+2iLIF. After 7 hours, cells were incubated with both TMR and SNAP-SiR647 dye at a concentration of 100nM or 12nM, respectively, for one hour. Thereafter cells were again stained with Hoechst 33342 for 15 minutes and subsequently trypsinized and collected by centrifugation for Flow Cytometry Analysis on a BD Fortessa or BD LSR II, followed by analysis using the FlowJo software (Supplementary Figure 2g).

Fluorescence Activated Cell Sorting for SOX2, OCT4, and combinations of both

To evaluate the impact of SOX2 or OCT4 levels on differentiation, we gated cells in G1 based on their Hoechst profile and defined three sub-bins of 25% in SOX2-SNAP of all G1 cells (Supplementary Figure 4a) or into 25% of OCT4-HALO high and low cells (Supplementary Figure 4b), respectively.

For the quadruple sorts based on a combination of SOX2/OCT4 high and low cells, we gated in all G1 or S phase cells on four windows corresponding to 20% of the total cell population each (Supplementary Figure 4f).

Flow Cytometry Analysis of *in vitro* differentiation

Cells were washed once in PBS, trypsinised, collected by centrifugation and resuspended in PBS/1% FBS before flow cytometry analysis. All data acquisition was performed on a BD

Fortessa and analysis was performed using the FlowJo software. E14 cells were used as a negative control to gate for SOX1-eGFP and BRA-mCherry.

In-vitro differentiation Methods

In-vitro Differentiation

For live-cell luminescence microscopy, cells were cultured in N2B27+2iLIF for at least two passages before 30,000 cells were seeded on E-Cadherin in fluorodishes and incubated in N2B27+2iLIF overnight. The next day, the medium was changed to N2B27 supplemented with 1mM luciferin and 0.5*µ*l of RealTime Glo Cell Viability Assay Substrate, and image acquisition was started.

For differentiation assays after cell sorting, cells were seeded at a density of 60,000 cells/well of a 6-well plate on gelatin. Two days later, the medium was exchanged for fresh N2B27 and after four days differentiation outcomes were assessed by flow cytometry on a BD Fortessa.

ATAC-Seq

ATAC seq was performed on 50,000 cells for each condition as previously described [29]. Briefly, 50,000 cells were sorted by FACS, pelleted and washed with 1X ice cold PBS at 800g for 5 min. Cells were gently resuspended in 50 µl of ice-cold ATAC lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% NP40), and immediately pelleted at 800g for 10 min at 4°C. To transpose open chromatin regions, cells were resuspended in 50 µl of transposition reaction mix containing 0.5 µM of Tn5 transposase (gift from Prof. Bart Deplancke lab, EPFL) in TAPS-DMF buffer (10 mM TAPS-NaOH, 5 mM Mgcl2, 10% DMF) and incubated at 37°C for 30 min. The transposed DNA was purified using a DNA purification kit (Zymo Research #D4003) and eluted in 12 µl of water. A 65 µl PCR reaction was setup with 10 µl of transposed DNA, 0.5 µM of forward primer Ad1_noMX, 0.5 µM of multiplexing reverse primer Ad2.x (Buenrostro et al. 2013), 0.6x SYBR® Green I, and 1x PCR Master Mix (NEB #M0544). The samples were thermocycled at 72°C for 5 minutes, 98°C for 30 s, followed by 5 cycles of 98°C for 10 s, 63°C for 30 s and 72°C for 1 min. A 15 µl aliquot was analyzed by qPCR to determine

the number of additional cycles needed to avoid amplification saturation as described in Buenrostro et al., 2013. The amplified ATAC libraries were purified using a DNA purification kit (Zymo Research #D4003) and size selected using Agencourt AMPure beads (0.55X unbound fraction followed by 1.2X bound fraction). All libraries were sequenced with 75-nucleotide read length paired-end sequencing on a Illumina NextSeq 500 with 30-60 million reads being sequenced for each sample.

Data Analysis

Immunofluorescence Image Analysis

Immunofluorescence images were first background-corrected using the built-in function in the Fiji software. Semi-automated image analysis was then performed using a custom CellProfiler [30] pipeline. Images were segmented based on their DAPI signal and manually corrected for misidentified objects. Subsequently, fluorescence intensity was measured in the identified nuclei in all channels. The intensities were used to generate histograms of protein expression (NANOG, OCT4, and SOX2), to evaluate the effects of overexpression of OCT4 or SOX2 and to estimate the correlation between OCT4-HALO and total OCT4.

Cell Tracking and Single Cell Analysis

Cells were tracked manually using Fiji (ImageJ) by defining regions of interest (ROIs) throughout the movie. Next, all ROIs for a single cell were measured and the background (part of the image in the vicinity of the cell devoid of cells) was subtracted. We used a previously reported method to convert the observed light intensity to absolute molecule numbers [14].

To determine SOX2 levels in pluripotency conditions, cells were *in silico* synchronised for cell cycle progression using linear interpolation of the time variable, and absolute molecule numbers were converted to nuclear concentration, using a previously described model for the nuclear size increase during the cell cycle [13] and a reported estimate of the nuclear volume of ES cells [31]. To evaluate how cells readjust their SOX2 levels over time, we used a previously described rank-based autocorrelation [15] using data from cells tracked over one or two

consecutive cell cycles. To compare the autocorrelation function between data tracked for one and two cell cycles, we selected 100 random single cell traces from the SOX2 data and calculated the protein memory based on a conservative mixing time estimation [15]. As the results of one and two cell cycles were similar, we used a single cell cycle from the OCT4-HALO imaging to calculate the rank-based autocorrelation of OCT4-HALO.

To determine how SOX2 levels predict neuroectodermal differentiation, we classified tracked cell cycles based on their FLuc signal in four groups, using an arbitrary threshold of 500 AU in FLuc maintained for at least four hours: “negative cells” were defined as cells below the threshold throughout the movie; “before SOX1+” were defined as negative cells that pass the threshold in the next cell cycle; “turning SOX1+” were defined as cells passing the threshold in the current cell cycle; “SOX1+” were defined as cells with FLuc levels above the threshold. The “before SOX1+” cell population also contains traces that did not cover a full cell cycle before becoming Sox1-positive. All single cell traces were *in silico* synchronised using a linear interpolation of the time variable. A two-sample t-test with unequal variance (MatLab) was performed for the mean SOX2 levels in the cell cycle before cells turn SOX1 positive to evaluate the statistical significance.

To determine the induction kinetics in the YPet-SOX2-delDBD and SOX2-SNAP overexpressing cell lines, single cells were tracked over divisions in one daughter cell.

ATAC-seq analysis

ATAC-seq reads were aligned to the mouse reference genome mm10 using STAR with settings ‘--alignMatesGapMax 2000 --alignIntronMax 1 --alignEndsType EndtoEnd --outFilterMultimapNmax 1’. Duplicate reads were removed with Picard and reads not mapping to chromosomes 1-19, X, or Y were removed. For each sample, peaks were called with MACS2 with settings ‘-f BAMPE -q 0.01 -g mm’. Peaks from all samples were merged with BEDOPS. Peaks overlapping peaks called for ChIP-seq Input data from asynchronous mouse ES cells (GSE89599) were discarded. The HOMER2 function annotatePeaks.pl was used with settings ‘-noadj -len 0 -size given’ to count the number of reads for each sample in peaks.

Analysis of differentially abundant regions was done with edgeR and limma using TMM normalization. The analysis was done using three different contrasts; (i) SHOH vs SLOL, design ~0+Condition+Replicate, (ii) SOX2 high vs SOX2 low, design ~0+SOX2+OCT4+Replicate, (iii) OCT4 high vs OCT4 low, design ~0+OCT4+SOX2+Replicate. Regions with an adjusted p-value < 0.1 for at least one test were used in the analysis. Groupings were made according to fold-change direction and if loci were significantly different for OCT4 high vs OCT4 low only (OCT4 regulated), SOX2 high vs SOX2 low only (SOX2 regulated), or regulated by both OCT4 and SOX2 or OCT4 high/SOX2 high vs OCT4 low/SOX2 low (co-regulated). SOX2 and OCT4 peaks used to determine overlap were merged from two studies [18,19]. H3K4me3 peaks in ES-Bruce4 cells from ENCODE and ES-cell super-enhancers from Whyte et al. 2013 were converted to mm10 using liftOver. Gene ontology analysis was done using the closest UCSC-annotated gene to each peak with Fisher's exact test in topGO using genes closest to all peaks as background. bigWig files were generated by merging replicate bam files with samtools followed by the deepTools functions bamCoverage (with setting '--normalizeUsingRPKM'). Average lineplots were generated using deepTools computeMatrix (with setting 'reference-point'). Genome tracks were generated in the UCSC genome browser.

Half-life determination of SOX2 and OCT4[32]

SBROS cells were seeded at 30,000 cells/cm^2 on E-cadherin in N2B27 medium. After 24 h, cells were pulse labeled with 12 nM of SNAP-SiR 647 or 20 nM Halo-SiR ligand (gift from Suliana Manley, EPFL) for 30 min at 37 °C. Cells were then washed 3x with PBS and incubated in N2B27 medium for 15 min at 37 °C. This washing step was repeated once more, then cells were washed 2x with PBS and phenol-free N2B27 medium was added The fluorescence decay was imaged using an InCell Analyzer 2200 microscope (GE Healthcare Life Sciences) with a 20x Objective, 10% laser power, 300 ms exposure (Cy5: excitation filter 632/22 nm, emission filter 679/34 nm), and 2x2 binning for 12.5 h at intervals of 15 min. Images were analysed in FIJI, where the background was subtracted from all images in the stack (rolling ball radius = 50 pixels). The integrated fluorescence intensity was then quantified

by manual tracking. ROIs were drawn around each cell of interest at each time point and the integrated fluorescence intensity was calculated by multiplying the area of each ROI with its mean intensity. The local background was calculated by drawing a ROI close to the cell of interest in each time frame and by then multiplying its mean intensity with the area of the cellular ROI. The background intensity was subtracted from the cellular intensity for each time frame. In case of cell divisions, both daughter cells were tracked separately and their integrated intensities were summed. Each fluorescence intensity trace was normalized to the value of the first frame and the single cell decay rates (b) were determined by exponential curve fitting, using the curve fitting tool in Matlab (fitted equation: $f(t) = e^{-bt}$). Half-lives were then calculated as follows: $T_{1/2} = b/\ln(2)$. Single cell half-lives in 20 cells were quantified for SOX2-SNAP, OCT4-HALO and OCT4-HALO after 6-8 hours of SOX2 overexpression.

Statistical analysis

Here we list the different statistical tests that were used. The cell cycle times (Fig.1h) were compared using two-sided t-tests, showing no statistical significance. The delta Ct values in Figure 2b were analysed using two-sided t-tests comparing the *Sox2* or *Oct4* mRNA levels between 0 and 2 hours as well as 0 and 6 hours. The OCT4-HALO half-lives (Fig.2c) were assessed using a two-sided t-test with unequal variance. Differences between SOX2-low and SOX2-high or OCT4-low and OCT4-high were analysed using a one-sided t-test (Fig.2e,f). For the autocorrelation functions in Figures 2i,l and Supplementary Fig. 3d the error bars denote the SE estimated by bootstrapping. In Figure 3b, a two-sided t-test with unequal variance was performed. For panels d, e, g, i and j in Figure 3, and Figure S4d, all depicted tests are two-sided t-tests with unequal variance. Analysis in Figure 4g and Figure S5i was done using a Kruskall-Wallis test. Analysis in Figure S5j,k was performed with a two-sided t-test with unequal variance.

References

1. Cohen, D. E. & Melton, D. Turning straw into gold: directing cell fate for regenerative medicine. *Nat Rev Genet* **12,** 243–52 (2011).

2. Chew, J.-L. *et al.* Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4/Sox2 Complex in Embryonic Stem Cells. *Mol. Cell. Biol.* **25,** 6031–6046 (2005).
3. Okumura-Nakanishi, S., Saito, M., Niwa, H. & Ishikawa, F. Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. *J. Biol. Chem.* **280,** 5307–5317 (2005).
4. Tapia, N. *et al.* Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency. *Sci. Rep.* **5,** (2015).
5. van den Berg, D. L. C. *et al.* An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **6,** 369–381 (2010).
6. Zhao, S., Nichols, J., Smith, A. G. & Li, M. SoxB transcription factors specify neuroectodermal lineage choice in ES cells. *Mol. Cell. Neurosci.* **27,** 332–342 (2004).
7. Thomson, M. *et al.* Pluripotency Factors in Embryonic Stem Cells Regulate Differentiation into Germ Layers. *Cell* **145,** 875–889 (2011).
8. Yang, H. *et al.* One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell* **154,** 1370–1379 (2013).
9. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339,** 819–823 (2013).
10. Deluz, C. *et al.* A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. *Genes Dev.* **30,** 2538–2550 (2016).
11. Hall, M. P. *et al.* Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem Biol* **7,** 1848–57 (2012).
12. Liu, L. *et al.* G1 cyclins link proliferation, pluripotency and differentiation of embryonic stem cells. *Nat. Cell Biol.* **19,** 177–188 (2017).
13. Filipczyk, A. *et al.* Network plasticity of pluripotency transcription factors in embryonic stem cells. *Nat. Cell Biol.* **17,** 1235–1246 (2015).
14. Mandic, A., Strebinger, D., Regali, C., Phillips, N. E. & Suter, D. M. A novel method for quantitative measurements of gene expression in single living cells. *Methods San Diego Calif* **120,** 65–75 (2017).
15. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444,** 643–6 (2006).
16. Kopp, J. L., Ormsbee, B. D., Desler, M. & Rizzino, A. Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells Dayt. Ohio* **26,** 903–911 (2008).
17. Mfopou, J. K. *et al.* Efficient definitive endoderm induction from mouse embryonic stem cell adherent cultures: A rapid screening model for differentiation studies. *Stem Cell Res.* **12,** 166–177 (2014).

18. Liu, Y. *et al.* Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors in Pluripotent Stem Cells. *Cell Rep.* **19,** 1283–1293 (2017).

19. King, H. W. & Klose, R. J. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *eLife* **6,** (2017).

20. Kartikasari, A. E. R. *et al.* The histone demethylase Jmjd3 sequentially associates with the transcription factors Tbx3 and Eomes to drive endoderm differentiation. *EMBO J.* **32,** 1393–1408 (2013).

21. Kalmar, T. *et al.* Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol* **7,** e1000149 (2009).

22. Wang, Z., Oron, E., Nelson, B., Razis, S. & Ivanova, N. Distinct Lineage Specification Roles for NANOG, OCT4, and SOX2 in Human Embryonic Stem Cells. *Cell Stem Cell* **10,** 440–454 (2012).

23. 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 (2013).

24. 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 (2013).

25. Shahbazi, M. N. *et al.* Pluripotent state transitions coordinate morphogenesis in mouse and human embryos. *Nature* **552,** 239–243 (2017).

26. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161,** 555–568 (2015).

27. Suter, D. M. *et al.* Rapid generation of stable transgenic embryonic stem cell lines using modular lentivectors. *Stem Cells Dayt. Ohio* **24,** 615–623 (2006).

28. Hobert, O. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans. *BioTechniques* **32,** 728–730 (2002).

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

30. Kamentsky, L. *et al.* Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinforma. Oxf. Engl.* **27,** 1179–1180 (2011).

31. Chalut, K. J. *et al.* Chromatin Decondensation and Nuclear Softening Accompany Nanog Downregulation in Embryonic Stem Cells. *Biophys. J.* **103,** 2060–2070 (2012).

32. Alber, A. B. & Suter, D. M. Single-Cell Quantification of Protein Degradation Rates by Time-Lapse Fluorescence Microscopy in Adherent Cell Culture. *J. Vis. Exp. JoVE* (2018). doi:10.3791/56604

Figure 1: Knock-in cell lines

a Scheme of the knock-in alleles of the SOX1-BRA-Reporter-OCT4-HALO-SOX2-SNAP (SBROS) cell line. **b** Example images showing the localization of SOX2-SNAP and OCT4-HALO by fluorescence microscopy of SBROS cells. **c** Correlation between OCT4-HALO and total OCT4 levels determined by immunofluorescence (R=0.8). **d** Scheme of the knock-in alleles of the SOX2-NLUC-SOX1-FLUC (SNSF) cell line. **e** Luminescence microscopy images of differentiating SNSF cells showing the SOX2-NLUC and the SOX1-P2A-FLUC signal. **f** Distributions of SOX2, OCT4 and NANOG levels in wt, SBROS and SNSF cell lines as determined by quantitative immunofluorescence. **g** Protein half-lives of OCT4-HALO and SOX2-SNAP. Whiskers: lower and upper extreme; Box: lower and upper quartiles; Solid line: mean. **h** Cell cycle duration of wt and knock-in cell lines. Whiskers: lower and upper extreme; Box: lower and upper quartiles; Solid line: median; Outliers: solid points. Scale bars: 50μm.

Figure 2: Cross-regulation and single-cell expression level fluctuations of SOX2 and OCT4

a Endogenous SOX2 levels in single cells (turquoise) normalized to the value at t=0 in SNSF cells upon induction of WT SOX2 (SOX2-SNAP, n=40 from 3 movies) or a truncated SOX2 version missing the DNA binding domain (YPET-SOX2-delDBD, n=47 from 3 movies). Black line: kinetics of SOX2-SNAP levels after induction normalized on the maximum level (n=22 from 5 movies); Yellow line: kinetics of YPET-SOX2-delDBD levels after induction, normalized on the maximum level (n=20 from 1 movie). **b** *Sox2* and *Oct4* mRNA levels upon overexpression of YPET-SOX2 or YPET-SOX2-delDBD (n=3 experiments). **c** Changes of OCT4 half-life upon SOX2 overexpression (N=20 cells each, N=1 movie). **d** Sorting strategy to evaluate the impact of endogenous fluctuations of SOX2 and OCT4 on OCT4 and SOX2 protein levels. **e** Changes of SOX2 and OCT4 levels after sorting for high and low SOX2 levels (N=4). **f** Changes of SOX2 and OCT4 levels after sorting for high and low OCT4 levels (N=4). **g** Representative traces of the absolute number of SOX2 molecules and inferred concentration in nM. **h** Single cells were ranked according to their SOX2 expression at t=0 from low to high levels, and changes in ranks over time are shown. **i** Rank-based autocorrelation function of the SOX2 ranks (N=59). **j** Representative single cell traces of the integrated intensity of OCT4-HALO in single cells (top) and the corresponding inferred concentration (bottom). **k** Representation of the ranks as in g. **l** Rank-based autocorrelation function as in i (N=47). ** $p<0.01$; * $p < 0.05$; error bars: SE.

Figure 3: Endogenous SOX2 and OCT4 level fluctuations bias differentiation

a Snapshot from luminescence movie of SNSF cells in differentiation. White arrowheads: cells starting to express SOX1-P2A-FLUC; Orange arrowheads: cells staying negative throughout the whole experiment. **b** Single cell traces of cells in differentiation from 4 movies. Turquoise traces: cells becoming SOX1 positive in the middle cell cycle, average is shown in bold; Red traces: cells remaining negative throughout the experiment, average is shown in bold; Green line: average FLUC signal in cells that become SOX1-positive. **c** Strategies to sort cell populations with different OCT4 and SOX2 levels. **d** Outcome of differentiation for cells sorted for SOX2 (N=5 experiments for low and high, N=4 experiments for medium). **e** Outcome of differentiation for cells sorted for OCT4 (N=4 experiments). **f-g** Cells allowing for inducible expression of OCT4-SNAP were treated with or without 100ng/ml Dox for 12 hours before sorting for G1-cells (**f**), and differentiation outcome was assessed four days later (N=6 experiments) (**g**). **h** Sorting strategy for G1 or S phase and into the four following subpopulations: SOX2 high & OCT4 high (SHOH), SOX2 high & OCT4 low (SHOL), SOX2 low & OCT4 high (SLOH) or SOX2 low & OCT4 low (SLOL). **i** Outcome after four days of differentiation for cells sorted in G1-phase (N=5 experiments). **j** Outcome after four days of differentiation for cells sorted in S-phase (N=5 experiments). **k** Changes in mRNA expression of selected markers after 24 hours of differentiation in SHOH, SHOL, SLOH and SLOL cells populations sorted in G1 phase (N=3 experiments). **l** Immunofluoresence analysis of NANOG and GATA6 in OCT4 low and OCT4 high cells after 24 hours of differentiation. Histograms show distribution of expression levels; Insets show the fraction of NANOG/GATA6 positive cells in each population (n=1 experiment, N= at least 1000 cells per condition). **d**, **e**, **g**, **i** and **j**: the % of cells were normalized to the average across all samples. * $p < 0.05$; error bars: SE.

Figure 4: OCT4 and SOX2-dependent changes in chromatin accessibility

a Example tracks of the ATAC-Seq signal in SHOH, SHOL, SLOH and SLOL populations. Loci are grouped as “unaffected”, “OCT4 regulated” (Groups 1 and 4), “SOX2 regulated” (Groups 2 and 5) and “co-regulated” (Groups 3 and 6). **b** Heatmap of chromatin accessibility changes depending on SOX2 and OCT4 levels with examples of genes close to affected regulatory regions. **c** Gene Ontology (GO) term enrichment analysis for genes with close-by enhancers having higher accessibility in OCT4 high cells. **d** Chromatin accessibility ratio of SHOH vs SLOL cells in ES cell super-enhancer regions. **e** Percentage of regions in each group showing an overlap with OCT4, SOX2 and H3K4me3 ChIP-seq peaks. **f** Genome tracks showing the *Eomes* enhancer region depending on OCT4 for its accessibility and location of guide RNAs used for CRISPR/CAS9 mediated knock-out. **g** Outcome after four days of differentiation for cells sorted in G1-phase for SBROS cells (N=5 experiments) and *Eomes* KO A (N=4 experiments) cells. OL: OCT4-low cells; OH: OCT4-high cells. * $p<0.05$; **$p<0.01$; error bars: SE.

Figure 5: Model of ES cell differentiation biases caused by OCT4 and SOX2 levels

OCT4 level fluctuations are regulated by SOX2. Low OCT4 levels bias cells towards pluripotency maintenance or a primitive endodermal (PrE) cell fate, while high OCT4 levels license cells to acquire a mesendodermal (ME) or neuroectodermal (NE) cell fate caused by OCT4-dependent chromatin accessibility changes at differentiation-associated enhancers. High SOX2 levels increase the amount of NE progeny but not at the expense of ME.

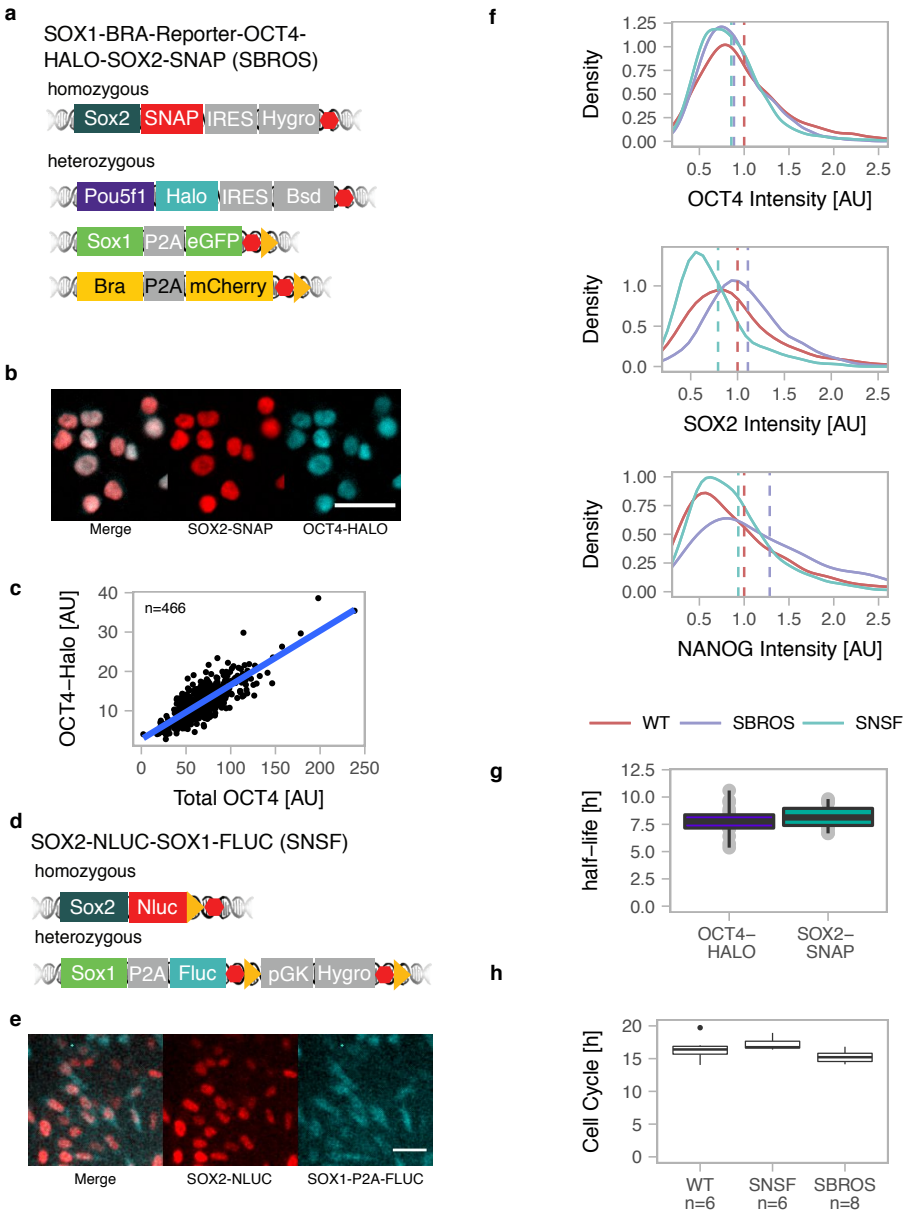

Figure 1

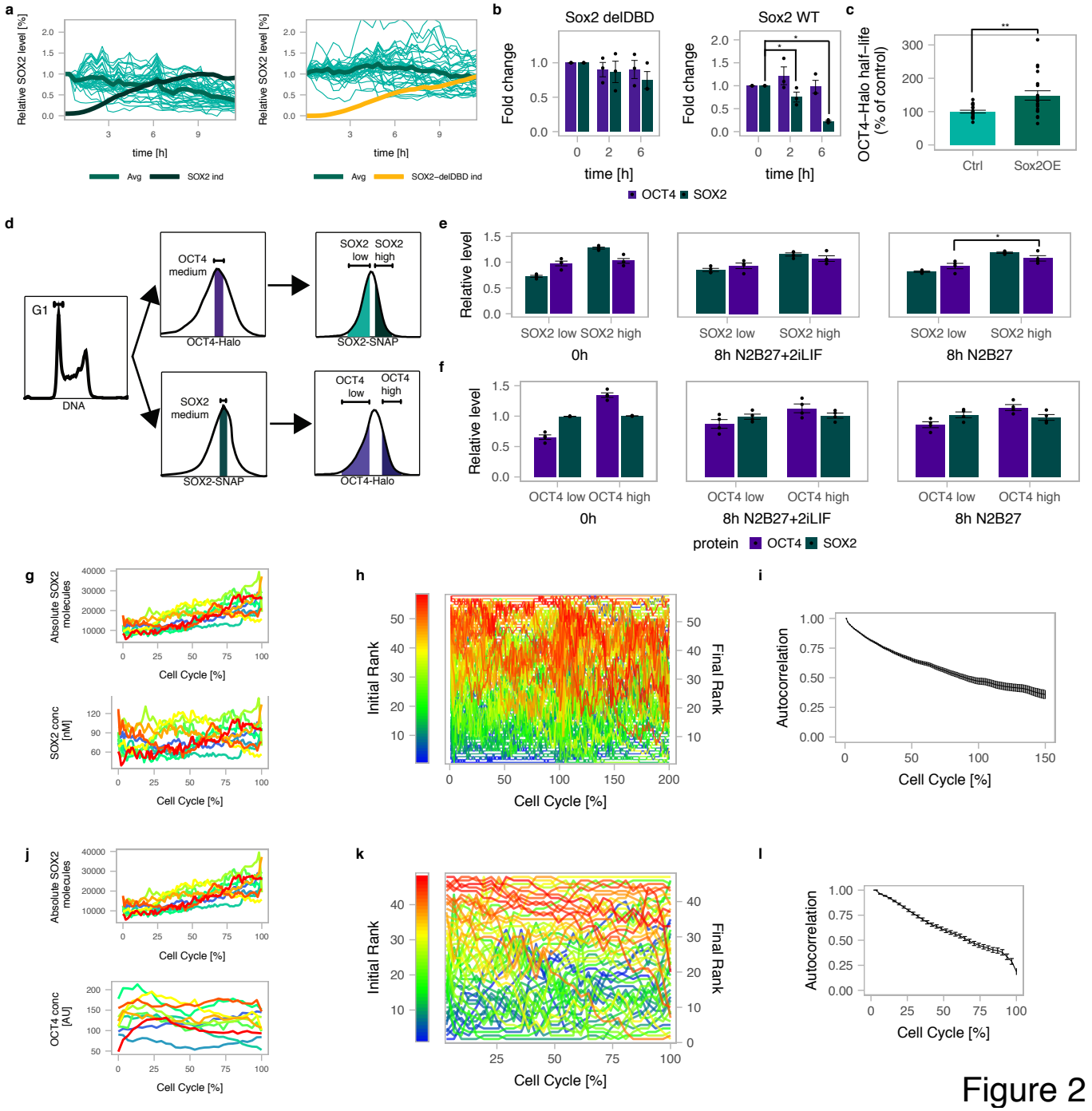

Figure 2

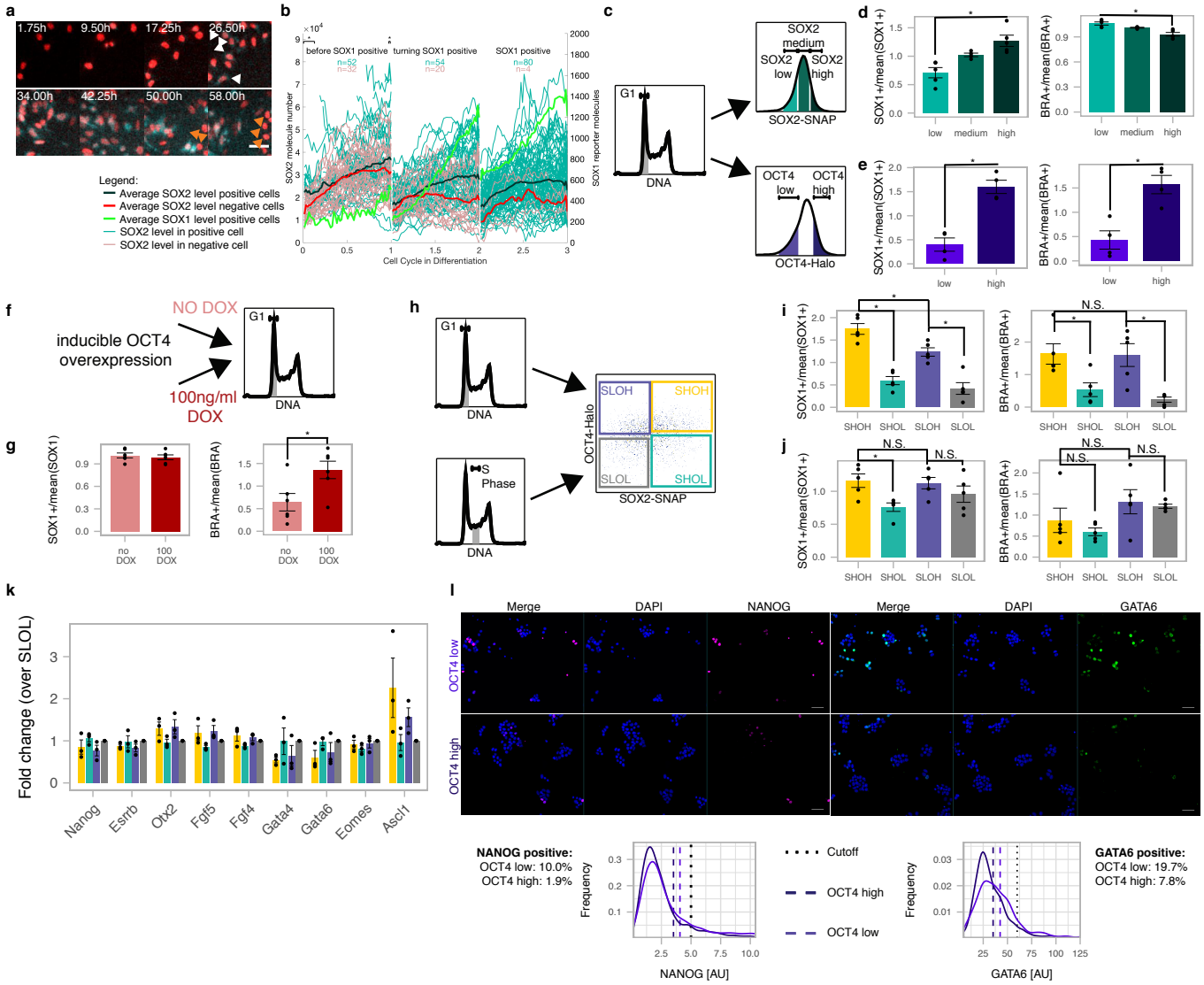

Figure 3

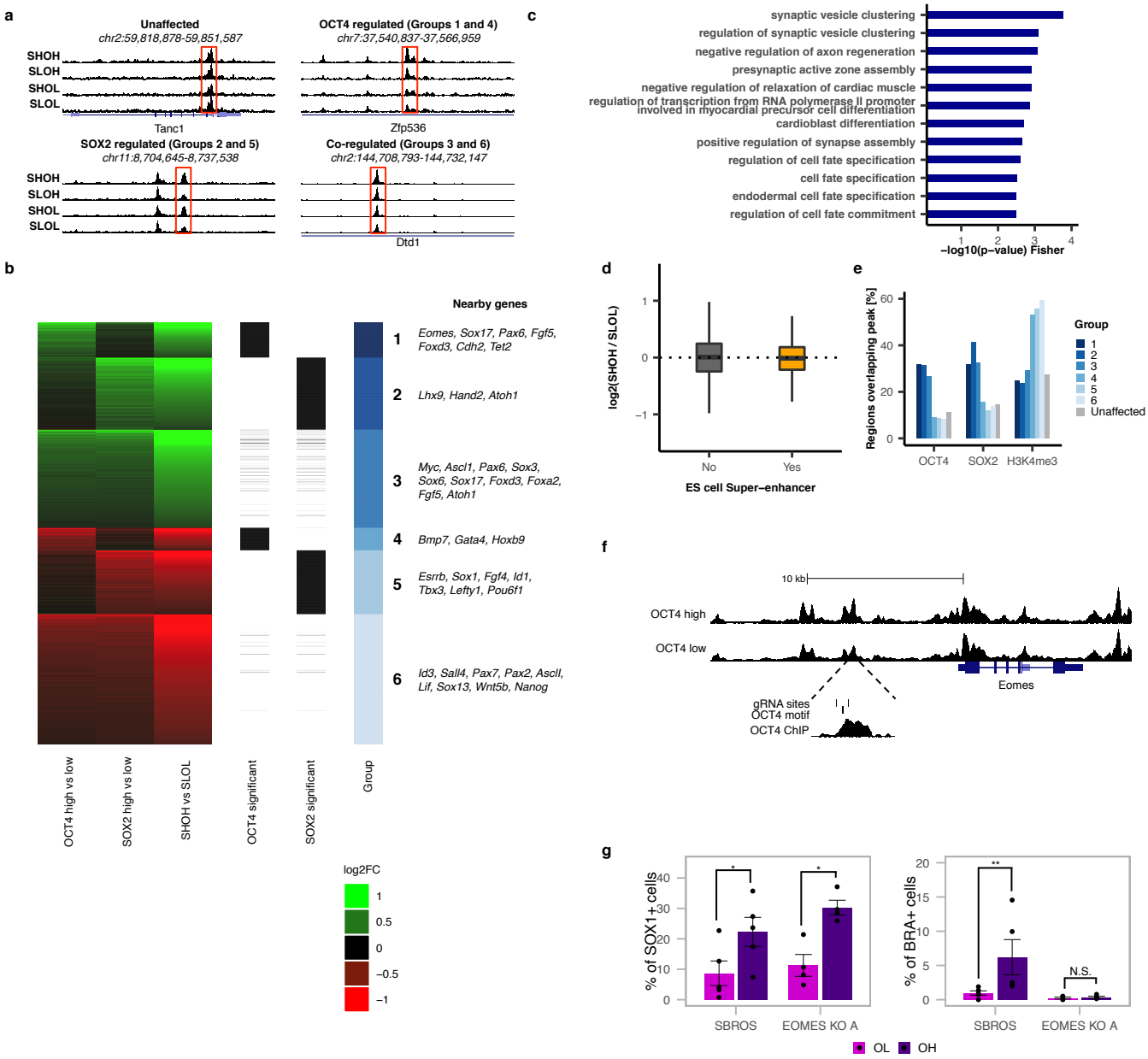
a
Unaffected
chr2:59,818,878-59,851,587
SHOH
SLOH
SHOL
SLOL
Tanc1
OCT4 regulated (Groups 1 and 4)
chr7:37,540,837-37,566,959
Zfp536
SOX2 regulated (Groups 2 and 5)
chr11:8,704,645-8,737,538
Co-regulated (Groups 3 and 6)
chr2:144,708,793-144,732,147
Dtd1
b
Nearby genes
1 Eomes, Sox17, Pax6, Fgf5, Foxd3, Cdh2, Tet2
2 Lhx9, Hand2, Atoh1
3 Myc, Ascl1, Pax6, Sox3, Sox6, Sox17, Foxd3, Foxa2, Fgf5, Atoh1
4 Bmp7, Gata4, Hoxb9
5 Esrrb, Sox1, Fgf4, Id1, Tbx3, Lefty1, Pou6f1
6 Id3, Sall4, Pax7, Pax2, Ascll, Lif, Sox13, Wnt5b, Nanog
OCT4 high vs low
SOX2 high vs low
SHOH vs SLOL
OCT4 significant
SOX2 significant
Group
log2FC
1
0.5
0
−0.5
−1
c
synaptic vesicle clustering
regulation of synaptic vesicle clustering
negative regulation of axon regeneration
presynaptic active zone assembly
negative regulation of relaxation of cardiac muscle
regulation of transcription from RNA polymerase II promoter involved in myocardial precursor cell differentiation
cardioblast differentiation
positive regulation of synapse assembly
regulation of cell fate specification
cell fate specification
endodermal cell fate specification
regulation of cell fate commitment
−log10(p−value) Fisher
d
log2(SHOH / SLOL)
No
Yes
ES cell Super−enhancer
e
Regions overlapping peak [%]
OCT4
SOX2
H3K4me3
Group
1
2
3
4
5
6
Unaffected
f
10 kb
OCT4 high
OCT4 low
Eomes
gRNA sites
OCT4 motif
OCT4 ChIP
g
% of SOX1+ cells
SBROS
EOMES KO A
% of BRA+ cells
N.S.
OL
OH

Figure 4

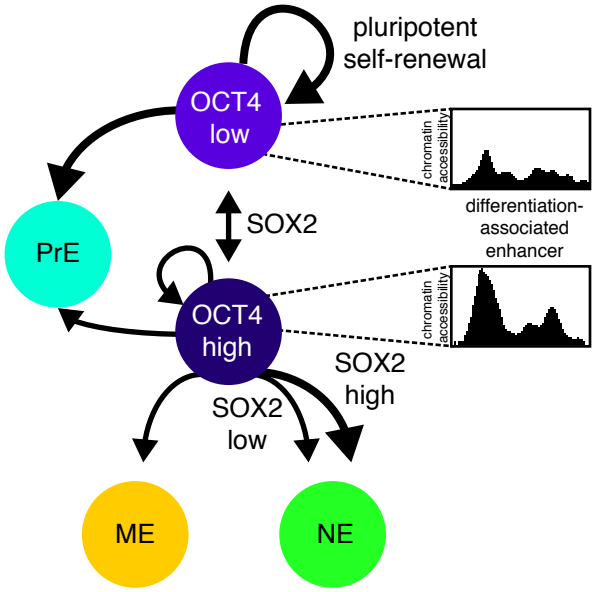
pluripotent
self-renewal
OCT4
low
chromatin
accessibility
differentiation-
associated
enhancer
SOX2
PrE
OCT4
high
chromatin
accessibility
SOX2
high
SOX2
low
ME
NE

Figure 5