

The role of OCT4 and SOX2 in regulating chromatin accessibility and cell fate across the cell cycle in embryonic stem cells

Thèse N° 9790

Présentée le 6 décembre 2019
à la Faculté des sciences de la vie
Unité du Prof. Suter
Programme doctoral en biotechnologie et génie biologique

pour l'obtention du grade de Docteur ès Sciences

par

Torgil Elias FRIMAN

Acceptée sur proposition du jury
Prof. B. E. Ferreira De Sousa Correia, président du jury
Prof. D. M. Suter, directeur de thèse
Prof. D. Schübeler, rapporteur
Prof. D. Gatfield, rapporteur
Prof. D. Trono, rapporteur

2019

Acknowledgements

Five years of work is a long time and therefore many people have been important both within the project but also as support. First of all, I would like to thank David Suter for giving me the opportunity to work in his lab and the advice, feedback, and freedom to allow me to train to be a scientist. Having a supportive supervisor is likely the most important factor for the success of a PhD and I am therefore particularly grateful. Cédric Deluz is a skilled technician who has been a close collaborator without whom this thesis would not have been possible. Daniel Strebinger was also crucial for a large part of the work presented here. I worked closely with Mahé Raccaud in the beginning of the thesis and for another project, and she is also a good friend. I want to thank all the other members of the lab over the years: Andrea Alber, Aleksandra Mandic, Onur Tidin, Laura Bischoff, Subashika Govindan, Armelle Tollenaere, Michael Sun, and Silja Placzek. I also want to thank the bachelor and master students I had the pleasure to supervise; Violaine Regard, Noémie Grouzmann, and David Maycotte Cervantes. I am also grateful to the collaborators that have helped in these projects, especially Vincent Gardeux and Marion Leleu, who both helped me with advice regarding bioinformatics when I was starting out from scratch. The core facilities at EPFL have also been essential for this work, including the Flow Cytometry Core Facility, Gene Expression Core Facility, and Bioimaging & Optics Platform. Other crucial services include the staff working at EPFL who take care of orders, supplies, cleaning, washing, management, IT, cluster computing, etc.

On a more personal note I would like to thank my family, who have been supporting me throughout this process; my mother Helena, my father Conny, my sisters Josefin and Sara, and their fantastic children Tilde, Stig, Elsi, and Lise. I also want to thank all my friends in Switzerland, Sweden, and abroad who make life outside the lab fun and interesting. In the last two years, I've had the privilege to spend a lot of time with my incredible partner, Olympe Peretz. Olympe, I don't think I could have managed the end of the thesis remotely as well without you and you have taught me so much. Thank you for everything. And finally, Maïs... You're the best (cat).

Abstract

Precise spatiotemporal regulation of gene expression is essential for development and homeostasis of complex organisms. This is achieved in large part by sequence-specific transcription factors (TF) that bind to genomic regulatory elements to activate or repress transcription. DNA in eukaryotic nuclei is wrapped around histones into nucleosomes, which are inaccessible to most proteins. However, certain TFs can access their binding sites in the presence of histones and promote nucleosome eviction to increase chromatin accessibility, allowing for binding of other protein. These TFs, called pioneer factors, include OCT4 and SOX2, which are master regulators of embryonic stem (ES) cells.

Dividing cells, including stem cells, must achieve proper gene regulation despite the progression of the cell cycle, which imposes several constraints. During S phase, the replication fork passes through the genome to make a copy of each chromosome, which leads to a loss of chromatin accessibility. Furthermore, the substantial chromatin compaction that occurs during mitosis to prepare for chromosome segregation is associated with eviction of most DNA-binding proteins and reduced accessibility at distal cis-regulatory elements including enhancers. How pioneer factors operate in the context of cell cycle progression is unclear. In this thesis, I will present work studying different features of OCT4 and SOX2, including how they interplay to regulate chromatin accessibility in ES cells, the impact of small endogenous fluctuations of their protein levels on cell fate and chromatin accessibility, and how their roles in regulating cell fate and chromatin accessibility is related to different stages of the cell cycle.

Using live-cell imaging of fluorescently labeled proteins, we discovered that OCT4 and SOX2 were associated to mitotic chromatin. Using cell-cycle specific degradation strategies, we could show that the presence of SOX2 and OCT4 at the mitosis-G1 (M-G1) transition contributes to maintain the pluripotency of ES cells. We further showed that ES cells with high levels of OCT4 in G1, but not in S phase, are more prone to differentiate towards neuroectoderm and mesendoderm cell fates. Cells with high levels of OCT4 displayed increased chromatin accessibility at enhancers of differentiation genes, suggesting accessibility fluctuates with OCT4 levels and can prime cells for differentiation.

To explore potential cell cycle-dependent regulation of chromatin accessibility, we degraded OCT4 at the M-G1 transition, which led to a reduction in accessibility at many OCT4-bound regions. When OCT4 returned later in the cell cycle, chromatin accessibility was recovered, although at many loci this recovery was incomplete, suggesting that OCT4 is required at the M-G1 transition to maintain these regions open. We then rapidly degraded OCT4 at other phases of the cell cycle. Surprisingly, loss of chromatin accessibility was highly similar at all cell cycle phases, showing that OCT4 is constantly required to maintain regulatory elements accessible. To explore the dynamic regulation of accessibility by OCT4, we performed time-course measurements of accessibility upon its degradation. This revealed that loss of accessibility was temporally correlated to the degradation of OCT4. These results suggest that chromatin accessibility is highly dynamic and continuously dependent on the presence of OCT4 across the cell cycle in ES cells.

Keywords

Transcription factors, embryonic stem cells, chromatin accessibility, pioneer factors, cell cycle, OCT4, SOX2, differentiation

Résumé

Une régulation spatiotemporelle précise de l'expression génique est essentielle au développement et à l'homéostasie des organismes complexes. Cela est assuré principalement par l'existence de facteurs de transcription (FT) séquence spécifiques qui se lient à des éléments génomiques régulateurs pour activer ou réprimer la transcription. Dans les noyaux eucaryotes, l'ADN est enroulé autour d'histones pour former des nucléosomes qui sont inaccessibles à la plupart des protéines. Cependant, certains FT peuvent accéder à leur site de liaison malgré la présence d'histones et favoriser l'éviction des nucléosomes augmentant l'accessibilité chromatinienne et permettant la liaison d'autres protéines. Ces FT, appelés facteurs pionniers, incluent OCT4 et SOX2, deux régulateurs majeurs des cellules souches embryonnaires (ES).

Pour les cellules capables de se diviser, dont les cellules souches, il est nécessaire d'assurer une régulation génique appropriée et ce malgré la progression dans le cycle cellulaire qui impose plusieurs contraintes. Pendant la phase S, la fourche de réplication progresse le long du génome pour copier chaque chromosome ce qui entraîne une perte de l'accessibilité chromatinienne. De plus, lors de la mitose, la compaction importante de la chromatine préparant la ségrégation des chromosomes, s'accompagne de l'éviction de la plupart des protéines se liant à l'ADN et d'une accessibilité réduite au niveau des éléments régulateurs en cis dont les séquences « enhancers ». La manière dont les facteurs pionniers agissent dans le contexte de la progression dans le cycle cellulaire reste inconnue. Dans cette thèse, je vais présenter l'étude de différentes caractéristiques d'OCT4 et SOX2, comme la manière dont ils interagissent pour réguler l'accessibilité chromatinienne dans les cellules ES, l'impact des faibles fluctuations endogènes du niveau de ces protéines sur le devenir cellulaire et l'accessibilité chromatinienne et l'influence des différentes phases du cycle cellulaire sur leurs rôles dans la régulation du devenir cellulaire et de l'accessibilité chromatinienne.

En utilisant la technique d'imagerie sur cellules vivantes de protéines marquées par fluorescence, nous avons montré qu'OCT4 et SOX2 étaient associés à la chromatine pendant la mitose. En utilisant des stratégies de dégradation dépendante du cycle cellulaire, nous avons pu montrer que la présence de SOX2 et OCT4 lors de la transition entre la mitose et la phase G1 (M-G1) contribuaient à maintenir la pluripotence des cellules ES. Nous avons de plus mis en évidence que les cellules ES présentant des niveaux plus élevés d'OCT4 en phase G1, mais pas en phase S, s'orientaient préférentiellement vers le neuroectoderme ou le mésendoderme lors de la différenciation. Les cellules avec des niveaux élevés d'OCT4 présentaient une accessibilité augmentée au niveau des séquences « enhancer » des gènes associés à la différenciation ce qui suggère que l'accessibilité fluctue avec les niveaux d'OCT4 et peut préparer les cellules à la différenciation.

Afin d'explorer une potentielle régulation de l'accessibilité chromatinienne dépendante du cycle cellulaire, nous avons dégradé OCT4 lors de la transition M-G1 ce qui a conduit à une réduction de l'accessibilité au niveau de plusieurs régions liées par OCT4. Lors du retour d'OCT4, plus tard dans le cycle cellulaire, l'accessibilité chromatinienne était retrouvée, bien qu'au niveau de certains loci, cette récupération reste incomplète, suggérant que la présence OCT4 est nécessaire lors de la transition M-G1 pour maintenir ces régions accessibles. Nous avons ensuite rapidement dégradé OCT4 lors d'autres phases du cycle cellulaire, ce qui a permis de montrer qu'OCT4 est requis de manière permanente pour maintenir l'accessibilité des éléments régulateurs. Afin d'explorer, la dynamique de la régulation de l'accessibilité par OCT4, nous avons mesuré l'accessibilité au cours du temps suite à sa dégradation. Cela a mis en évidence que la perte d'accessibilité était corrélée dans le temps à la dégradation d'OCT4. Ces résultats suggèrent que l'accessibilité chromatinienne est très dynamique et dépend continuellement de la présence d'OCT4 tout au long du cycle cellulaire dans les cellules ES.

Mots-clés

Facteurs de transcription, cellules souches embryonnaires, accessibilité chromatinienne, facteurs pionniers, cycle cellulaire, OCT4, SOX2, différenciation

Table of contents

Acknowledgements.....	2
Abstract.....	3
Résumé	4
Table of contents.....	5
List of figures.....	7
List of tables	9
Chapter 1 Introduction.....	10
1.1 General introduction.....	10
1.2 Regulation of cell identity	11
1.3 Early embryonic development.....	11
1.4 Embryonic stem cells	12
1.5 Transcriptional regulation.....	12
1.6 Chromatin in gene regulation	13
1.6.1 Chromatin dynamics	14
1.7 The cell cycle.....	14
1.7.1 Interplay between the cell cycle and cell fate	15
1.7.2 Impact of the cell cycle on chromatin and gene regulation.....	15
1.8 Transcription factors	16
1.8.1 Pioneer transcription factors	16
1.8.2 Transcription factor DNA-binding during mitosis	17
1.9 Regulation of pluripotency by transcription factors	17
1.9.1 The roles of OCT4 and SOX2 in pluripotency and differentiation.....	18
1.10 Heterogeneity in gene expression	19
1.11 Methods	19
1.12 Aims of the thesis	21
Chapter 2 A role for mitotic bookmarking of SOX2 in pluripotency and differentiation.....	22
2.1 Abstract.....	22
2.2 Introduction	22
2.3 Results	23
2.3.1 SOX2 and OCT4 have intrinsic mitotic chromosome-binding (MCB) properties that are independent of the pluripotency context.....	23
2.3.2 DNA-binding domain dependency and quantitative differences of SOX2 and OCT4 in MCB.....	24
2.3.3 SOX2 and OCT4 display distinct mobility but similar frequencies and residence times of long-lived DNA-binding events on mitotic chromosomes	26
2.3.4 SOX2 is retained on a small number of genomic loci during mitosis.....	28
2.3.5 Functional interrogation of mitotic bookmarking by degradation of SOX2 at the M–G1 transition	29
2.3.6 The presence of SOX2 at the M–G1 transition contributes to pluripotency maintenance but is dispensable for reprogramming to induced pluripotent stem (iPS) cells	31
2.3.7 SOX2 is required at the M–G1 transition to drive neuroectodermal differentiation	31
2.4 Discussion	33

2.5	Materials and methods	34
2.6	Supplementary figures	42
2.7	Tables	51
2.8	Acknowledgements	53
Chapter 3 Endogenous fluctuations of OCT4 and SOX2 bias pluripotent cell fate decisions		55
3.1	Abstract.....	55
3.2	Introduction	55
3.3	Results	56
3.3.1	Generation of a sox2-SNAP / oct4-halo knock-in ES cell line	56
3.3.2	SOX2 fluctuations regulate OCT4 levels	58
3.3.3	Characterisation of SOX2 and OCT4 fluctuations	61
3.3.4	SOX2 and OCT4 fluctuations impact NE and ME commitment.....	61
3.3.5	OCT4-high cells open differentiation enhancers	64
3.4	Discussion	65
3.5	Materials and methods	66
3.6	Supplementary figures	75
3.7	Tables	80
3.8	Acknowledgements	81
Chapter 4 Dynamic regulation of chromatin accessibility by pluripotency transcription factors across the cell cycle		83
4.1	Abstract.....	83
4.2	Introduction	83
4.3	Results	84
4.3.1	OCT4 and SOX2 regulate chromatin accessibility at mostly distinct loci	84
4.3.2	Cooperative binding between OCT4 and SOX2 is mainly mediated indirectly through changes in chromatin accessibility.....	87
4.3.3	OCT4 is required at the M-G1 transition to re-establish enhancer accessibility	88
4.3.4	OCT4 is required throughout the cell cycle to maintain enhancer accessibility.....	91
4.3.5	Dynamic relationship between OCT4 concentration and chromatin accessibility	93
4.4	Discussion	96
4.5	Materials and methods	96
4.6	Supplementary figures	102
4.7	Acknowledgements	110
Chapter 5 Conclusion		111
5.1	Discussion of achieved results	111
5.2	Potential future research avenues	114
References		116
Appendix		139
Curriculum Vitae		143

List of figures

Figure 2.1 SOX2 and OCT4 bind to chromosomes throughout mitosis.	23
Figure 2.2 MCB of SOX2 and OCT4 is mediated by their HMG and POUS domains, respectively. Fluorescence snapshots of living ES cells expressing H2B-CerFP and dox-inducible YPet fusion to truncated versions of SOX2 and OCT4. 25	
Figure 2.3 Differential retention and interdependence of SOX2 and OCT4 for MCB.	26
Figure 2.4 Dynamics of OCT4 and SOX2 in mitotic and interphase cells. .	27
Figure 2.5 ChIP-Seq analysis of SOX2 in unsynchronized and mitotic cells.	29
Figure 2.6 The role of SOX2 at the M–G1 transition in pluripotency maintenance and reprogramming to induced pluripotent stem (iPS) cells.....	30
Figure 2.7 SOX2 is required at the M–G1 transition to promote neuroectodermal differentiation.	32
Figure 3.1 Knock-in cell lines.....	57
Figure 3.2 Cross-regulation and single-cell fluctuations of SOX2 and OCT4.	60
Figure 3.3 Endogenous SOX2 and OCT4 level fluctuations bias differentiation.	63
Figure 3.4 SOX2 and OCT4-dependent changes in chromatin accessibility.	65
Figure 4.1 Interplay between OCT4 and SOX2 in regulating ES cell chromatin accessibility.	86
Figure 4.2 Characterization of OCT4/SOX2-dependent loci.	88
Figure 4.3 Mitotic degradation of OCT4 results in different patterns of accessibility loss.	91
Figure 4.4 Auxin-inducible degradation reveals pioneer activity of OCT4 at different cell cycle phases.	93
Figure 4.5 Time course analysis of chromatin accessibility changes during OCT4 degradation reveals its highly dynamic pioneer activity.....	95
Supplementary Figure 2.1 Interphase localization of YPet-fused transcription factors and additional control immunofluorescence and SNAP-tag fusion imaging, related to Figure 2.1.	42
Supplementary Figure 2.2 Confocal imaging of mitotic chromosome binding in ES cells, related to Figure 2.1 and Figure 2.2.	43
Supplementary Figure 2.3 Genomic analysis of Sox2-luciferase and Luciferase-Oct4 knock-in ES cells, related to Figure 2.1.	43
Supplementary Figure 2.4 DNA residence time extraction from single molecule imaging experiments, related to Figure 2.4.	44
Supplementary Figure 2.5 Quality control of mitotic cell sorting, related to Figure 2.5.....	45
Supplementary Figure 2.6 Peak calling and selection, related to Figure 2.5.	47
Supplementary Figure 2.7 Relative enrichment of called peaks, related to Figure 2.5.....	48

Supplementary Figure 2.8 ChIP-QPCR on selected peaks and negative control regions, related to Figure 2.5.....	48
Supplementary Figure 2.9 Genome-wide distribution of ChIP-seq peaks, related to Figure 2.5.	49
Supplementary Figure 2.10 De novo motif identification with MEME in the asynchronous sample, related to Figure 2.5.	49
Supplementary Figure 2.11 Sorting strategies for the Sox2 overexpressing cell lines, related to Figure 2.6 and Figure 2.7.....	50
Supplementary Figure 2.12 Genomic analysis of Sox1-P2A-eGFP / Brachyury-P2A-mCherry knock-in ES cells, related to Figure 2.7.	51
Supplementary Figure 3.1 Validation of knock-in cell lines.	75
Supplementary Figure 3.2 Cross regulation of SOX2 and OCT4.....	76
Supplementary Figure 3.3 Fluctuations of SOX2 and OCT4 over the cell cycle.	77
Supplementary Figure 3.4 Impact of SOX2 and OCT4 levels on differentiation outcomes.	78
Supplementary Figure 3.5 ATAC-seq analysis and characterization of differentially accessible regions.	79
Supplementary Figure 4.1	102
Supplementary Figure 4.2	103
Supplementary Figure 4.3	103
Supplementary Figure 4.4	104
Supplementary Figure 4.5	105
Supplementary Figure 4.6	106
Supplementary Figure 4.7	108
Supplementary Figure 4.8	109
Supplementary Figure 4.9	110

List of tables

Table 2.1: Primers used to verify genomic integration of the Sox2-Fluc and Fluc-Oct4 knock-in cassettes. HA: homology arms.....	51
Table 2.2: Primers used to verify genomic integration of the Sox1-P2A-eGFP and Brachyury-P2A-mCherry knock-in cassettes. HA: homology arms.	51
Table 2.3: Primers used for RT-QPCR.....	52
Table 2.4: Primers used for ChIP-QCPR.....	52
Table 3.1. Sequences of guide RNAs used to generate knock-in cell lines	80
Table 3.2 Primers for knock-in verification	80
Table 3.3 Antibodies used for Western Blotting	80
Table 3.4 Primers used for QPCR.....	81
Table 3.5 Antibodies used for Immunofluorescence staining	81
Appendix table 1	139
Appendix table 2.....	140
Appendix table 3.....	141
Appendix table 4.....	142

Chapter 1 Introduction

1.1 General introduction

In the first half of the 20th century, researchers discovered that the substrate of information inside cells is deoxyribonucleic acid (DNA). In the latter half, they mapped out the “genetic code”, describing how DNA is transcribed into the chemically similar molecule ribonucleic acid (RNA) and further translated into polypeptides of amino acids (proteins). While this principle is true for every living species on earth, it alone cannot fully account for the remarkable diversity of biological species and functions or the astounding organization that occurs during development of complex multicellular organisms. Two observations in particular highlight this. The first is from 1962, when John Gurdon demonstrated that the nucleus of a fully differentiated intestinal epithelium cell of a tadpole could be transplanted back into an oocyte without a nucleus and give rise to fully normal tadpoles (1). This was striking proof that all the genetic material needed to give rise to the organism must be retained in adult (somatic) cells. The second example came with the advent of advanced DNA sequencing technologies capable of mapping the entire genetic material of any organism. Comparison of the genetic material between species revealed several unexpected findings. While DNA sequence could clearly delineate evolutionary relationships, looking at protein-coding genes revealed a remarkably large similarity between species (2). Furthermore, both the number of genes and the size of genomes of different species is poorly correlated to what we would consider phenotypic complexity. Humans have half the number of genes compared to Rice, and two orders of magnitude less DNA than a marbled lungfish (3). From these observations, we can conclude that (i) cells of complex organisms must obtain their different identities without changing their genetic material and (ii) differences between species cannot be explained by looking at simple genetic features of our protein-coding genes.

The last decades have involved considerable efforts to understand how phenotypic differences between species, and different cell types of within species, arise. Some of the key discoveries that have yielded a clearer picture of this include: RNA molecules that can act as functional units, posttranscriptional mechanisms including alternative splicing that can yield different protein products from the same gene, posttranslational modifications that alter the functions of proteins, and the regulation of gene expression such that different genes are expressed in different cells. Together, these mechanisms provide at least part of an answer to the question of complexity in biology. The ensemble of molecules in a cell can react to their environment, including signals from neighboring cells, and change the expression of genes or the function of those gene products. If we consider that humans have around 25'000 protein-coding genes which can all possibly interact, that these can exist in multiple functional variants, and that they can have a large range of different concentrations, we arrive at the conclusion that there are virtually infinite combinations of those proteins (and this is not taking into account all the other molecules such as RNA and metabolites). While this complexity constitutes the basis of our current understanding of the problem of diversity in life, we are far away from being able to explain or model all the molecules and their interactions inside cells, and even further from understanding how life emerges out of them.

This thesis will address one of the mechanisms by which differences in gene expression are regulated, namely the action of transcription factors (TF); specifically, two TFs that are active in embryonic stem (ES) cells, OCT4 and SOX2. It will explore how these TFs can regulate the chromatin accessibility and how they affect cell fate, and how these two properties are affected by (i) the progression of the cell cycle and (ii) the concentration of OCT4 and SOX2.

This chapter will briefly introduce the main concepts under study in this thesis. Chapters two through four constitute three separate publications, which each have their own introduction section which goes more in depth in the topic in the respective study.

1.2 Regulation of cell identity

Each cell of an organism harbors exactly the same genetic material, excluding rare acquired mutations. The difference in phenotype thus arises from the combined expression of different parts of the genome and instructive signals from the environment, such as communication with neighboring or distant cells. I will refer to all these non-genetic molecular mechanisms that can alter gene expression as epigenetic, using the broad definition of the term (4). Epigenetic mechanisms serve two roles: (i) to maintain the correct cellular phenotype and (ii) to control differentiation of cells toward different lineages. Maintenance occurs primarily in somatic cells which have undergone terminal differentiation, but also in stem cells, which in the adult organism serve to regenerate the tissue, such as hematopoietic stem cells that generate red and white blood cells. Stem cells must maintain their cell identity through cell divisions and be responsive to signals that promote differentiation. During development, a single fertilized zygote divides and differentiates in an organized fashion to give rise to the whole organism. This requires extraordinarily robust regulation of differentiation and cell division, and relies on epigenetic mechanisms or “programs” that are executed in an ordered fashion. Development is one of the main fields of study in biology and the last decades have provided great advances in our understanding of epigenetic mechanisms that govern cell fate, especially in the embryo. Two of the major regulatory mechanisms includes the deployment of TFs (5) (see section 1.8) and modifications to DNA (6) and histone proteins (7) (see section 1.6) which control the expression of genes. These mechanisms also interplay with other regulatory components such as morphogenetic signaling (8), mechanical forces (9), post-transcriptional mechanisms including splicing (10), and translational regulation including translation rates (11). These different aspects come together to produce so-called “gene regulatory networks” that make up a cell identity (12) and provide the groundwork for further cell fate decisions, e.g. upon exposure to a signal. How these gene regulatory networks are wired to produce robust and reproducible cell fate decisions is an intense area of study that requires an integrative understanding of all the epigenetic mechanisms at play.

1.3 Early embryonic development

I will here describe the early developmental progression of mouse embryos, which are the most heavily used model system to study mammalian developmental biology. After fertilization, the zygote undergoes several cell divisions and eventually give rise to two separate lineages; the trophoectoderm (TE) and inner cell mass (ICM). This cell fate decision has been well-studied and includes multiple mechanisms including mechanical forces and concentration of TFs, including OCT4, CDX2, and TEAD4 (13). The TE lineage will give rise to extraembryonic tissues, including placenta. After formation of the blastocyst, where TE cells surround the ICM, the cells of the ICM will undergo repositioning and form an outer layer containing the primitive endoderm (PrE; which also gives rise to extraembryonic lineages including yolk sac) and epiblast. The epiblast contains pluripotent cells that will give rise to all cells of the adult organism, including germ cells. The cell fate decision between PrE and epiblast involves antagonistic expression of the two TFs NANOG and GATA6 (14). Around this time, implantation in the uterine wall occurs and the embryo undergoes largescale rearrangements to yield the epiblast in a cup-shape surrounded by the visceral endoderm (derived from the PrE) on the sides and the extraembryonic ectoderm on top (derived from the TE). The first symmetry breaking event of the epiblast occurs upon formation of the primitive streak, where cells at the proximal-posterior side of the epiblast begin to migrate through the streak and undergo an epithelial-to-mesenchymal transition, and give rise to cells of the mesoderm and endoderm lineages (referred to as the mesendoderm) (15). This requires the activity of multiple signaling pathways, including Bmp, Wnt, and Activin/Nodal. At the anterior-distal visceral endoderm, inhibitors of Bmp and Nodal signaling Cer1 and Lefty1 are secreted into the epiblast at the anterior side, blocking the inductive actions of these signaling molecules to prevent differentiation (16). These cells will later give rise to the ectoderm lineage, including the neuroectoderm which forms the central nervous system. These post-implantation differentiation events constitute what is called gastrulation, where the three germ layers mesoderm, endoderm, and ectoderm are formed. Cells of the mesendoderm (in the primitive streak) express the marker gene Brachyury (Bra), while cells of the neuroectoderm express Sox1 (13).

1.4 Embryonic stem cells

One of the problems with studying early development is defining lineage trajectories and cell types, as cells are constantly progressing through different stages. Furthermore, after implantation in the uterine wall, embryos are difficult to access which makes studying later cell fate decisions, including gastrulation, difficult. This became markedly easier with the advent of embryonic stem (ES) cells, which were first derived in 1981 from the ICM of mouse embryos (17, 18). Researchers could show that specific conditions could support the maintenance of these cells in culture and that they could be differentiated into cell types of the three germ layers in vitro, and were thus pluripotent. This was a turning point for research in developmental biology due to the fact that these cells could be expanded to easily perform experimental assays on large numbers of cells because they are “captured” in the, naturally transient, pluripotent state. Early culture conditions relied on irradiated feeder cells to provide the ES cells with unknown signaling molecules to maintain their pluripotency. Over time, culture conditions for ES cells improved and today chemically defined media compositions are available. Maintenance of pluripotency relies on different signaling molecules, including reinforcement of the STAT3 signaling by Leukemia Inhibitory Factor (LIF) (19). When grown using inhibitors of GSK-3 (mimicking Wnt signaling) and MEK (commonly called 2i for two inhibitors), ES cells enter into a “ground state” of pluripotency (20). These cells transcriptionally resemble the pre-implantation epiblast (21). In contrast, when grown under less controlled conditions such as those containing Serum and LIF (but no inhibitors), cells exist in a metastable configuration where they shuttle between the ground state and more developed or “primed” stages of pluripotency (22). Furthermore, other culture conditions have captured later developmental stages including so called EpiSCs which are primed and resemble the later post-implantation epiblast (23, 24). Some of the main advantages with culturing ES cells is the rapid expansion (they have a short cell cycle), the ability to study pluripotency and differentiation, and the ease of genetic manipulation. Indeed, ES cells were early on used to generate knock-out mice by reinjecting genetically modified cells into newly developing blastocysts to allow for the study of developmental phenotypes arising from removal of single genes (see section 1.11). Overall, ES cells are one of the most heavily used model systems in biological studies.

1.5 Transcriptional regulation

Gene expression is a multi-step cascade which requires the coordinated action of many different protein complexes. Transcription itself is catalyzed by RNA polymerase enzymes, and in eukaryotes messenger RNAs (mRNA) are transcribed by RNA polymerase II (Pol II), which consists of 12 subunits in mammals and makes up a large complex (25). The first regulatory step is recruitment of Pol II to promoters (upstream sequences) of genes, which requires the action of TFs (see section 1.8) and other co-activators. Once recruited, Pol II can stall on DNA (promoter-proximal pausing) or proceed to elongation (RNA synthesis) (26). Transcription can also pause during elongation. Finally, transcription is terminated and the RNA molecule is released from Pol II. Each step of this process is regulated by a large ensemble of molecules. Initiation of Pol II involves the preinitiation complex (PIC), consisting of multiple proteins that position Pol II at the transcription start site (27). Switching between initiation to elongation to termination involves different phosphorylation states of the C-terminal domain of the RPB1 subunit (28). Termination mostly proceeds after encountering poly(A) signals where an array of proteins regulates the termination of transcription and addition of poly(A) tails to the mRNA molecule (27). These are highly simplified descriptions of very complex processes, and does not include other essential steps such as pre-mRNA processing, splicing, and export. While these factors, together with properties of the protein product such as degradation rates, all contribute to the protein levels of genes, the main difference in gene expression between different cell types and cell states is the difference in transcription initiation (2).

While transcription is driven from the immediate upstream sequences from the transcription start site (TSS), called the core promoter, other sequences in the genome house regulatory elements responsible for controlling gene expression. These can be promoter-proximal (~1-2 kb upstream of TSS) or distal. Distal elements include enhancers, which were first discovered when a sequence on the SV40 virus genome could enhance the transcription of the β -globin gene when inserted at any

location or orientation in the same plasmid (29). Since then, it has become clear that animals rely heavily on enhancers to regulate gene expression (30). The favored model of enhancer function is looping to physically contact the promoter of the gene to deliver regulatory molecules promoting its transcription, although exactly how this works and whether looping is the only or predominant mode of enhancer function is still unclear (31). What is clear however is that animal genomes are littered with enhancers. It is estimated that the human genome houses hundreds of thousands of enhancer sequences (32) and that some genes may be regulated by over 20 enhancers (33). This enormous diversity provides a means for modular gene regulation, since which enhancers are active in a certain context will determine the transcription level of target genes.

1.6 Chromatin in gene regulation

Animal genomes are very large. This creates a spatial problem, as every human cell needs to package 2 m of linear DNA into a nucleus the size of a few picoliters. This packing is achieved by the hierarchical folding with the aid of proteins called histones. Four core histones (H2A, H2B, H3, and H4) form a core octamer of two tetramers around which DNA can wrap. This basic unit of DNA and histones is called a nucleosome, and comprises roughly 150 bp of DNA wrapped 1.65 times around the histones (34). The linker histone H1 sits between nucleosomes and promotes packing of nucleosomes into an ordered assembly, generating a 30 nm fiber (2). This fiber loops and folds onto itself to promote even higher packing. In metaphase, chromosomes obtain their highest level of compaction, being coiled into the well-known X-shaped structure of sister chromatids before being divided into two daughter cells containing one chromatid each. This mesh of DNA and proteins is called chromatin. Besides packing DNA, chromatin also functions to regulate genes. In general, compacted chromatin is inaccessible to binding by most regulatory proteins and the regulation and chromatin accessibility is thus an important parameter in gene activation and repression (35). Electron microscopy images of interphase nuclei reveals that some regions contain very dense regions of chromatin and some that are less dense. The dense regions are heterochromatin, which is enriched around the nuclear periphery and is generally transcriptionally inactive. The less dense regions, euchromatin, contain active regions of the genome (2). Modern methods using so-called 3C technologies (for chromosome conformation capture) has revealed that the genome is spatially organized and that euchromatin and heterochromatin regions tend to be in proximity to their respective chromatin type, in so-called A (euchromatin) and B (heterochromatin) compartments (36). Furthermore, within these compartments there is a further organizational structure at the megabase-scale called topologically associated domains (TAD), which are regions that interact preferentially within themselves compared to neighboring regions (37). Within these TADs, interactions between enhancers and promoters occur. This hierarchical organization of chromatin can play a functional role in the regulation of gene expression (38), although this area is still under heavy investigation.

Besides chromatin accessibility and chromatin structure, histone proteins themselves play a regulatory role and can be posttranslationally modified. For example, the lysine at position 9 in the tail of histone H3 modified with two or three methyl groups (abbreviated H3K9me2 and H3K9me3) is enriched in heterochromatin regions of the genome (39). Conversely, euchromatin is enriched for acetylation of lysine 16 of histone H4 (H4K16ac) (40). Promoters and enhancers display several posttranslational signatures at their local nucleosomes. For example, active promoters are typically decorated with the H3K4me3 modification, while enhancers house the H3K4me1 modification and both active promoters and enhancers display enrichment for H3K27ac (7). A plethora of such modifications exist but none can exhaustively identify every instance of a specific regulatory element (41). Many of these marks correlate with the transcriptional activity of the regulated gene. In some cases, altering histone modifications leads to functional changes in transcription (42). However, it is difficult to attribute which mechanism is responsible for this function due to the extensive feedback mechanisms between so called histone readers and writers, which regulate histone modifications, and the binding of other regulatory factors that may be influenced by these modifications (43). Nonetheless, histone marks are regulatory and also serve as good genome-wide identifiers of regulatory elements and their activity. For example, the functional definition of enhancers requires cloning the putative nucleotide sequence distally upstream or downstream of a reporter gene and observing its increased expression to a control sequence without enhancer function. When done in

embryos of e.g. mouse, fly, or zebrafish, patterns emerge that reveal the different tissues and stages the enhancer is active (32). While it is unfeasible to try hundreds of thousands of regulatory elements in all different possible contexts, bioinformatic analyses can identify regions enriched for e.g. H3K4me1 and H3K27ac and see the differential deposition of these marks across tissues, developmental stages, or upon experimental treatments.

1.6.1 Chromatin dynamics

Even within relatively sparsely packed euchromatin, DNA is wrapped around nucleosomes and inaccessible to proteins. Transcription requires that Pol II traverses the length of the gene to produce an mRNA molecule, meaning that histones must be evicted during the process. The same applies for proteins binding to promoters or enhancers, where a nucleosome-free region (NFR) must be displayed to allow for access of the molecular assembly required to initiate transcription. Chromatin must thus be highly dynamic to accommodate binding and transcription. The dynamic nature of chromatin is particularly evident at transcription sites, where a variety of mechanisms regulate the dynamic exchange of histones (44). Histone variants such as H2A.Z and H3.3 that are less stable are enriched at promoters and gene bodies to promote histone turnover during the transcription cycle. These variants are deposited by various histone chaperones, among which are FACT, a protein that recycles histones after Pol II passage (44). Chromatin modifiers deposit histone modifications which can affect the stability of histone-DNA interactions, such as acetylation of lysine residues which neutralize their positive charge (45). ATP-dependent chromatin remodelers such as the BAF (mammalian analog of SWI/SNF) complex can remove histones to produce a nucleosome-free region at regulatory regions (46). The combined action of these groups of proteins coordinate chromatin dynamics throughout the transcription cycle which includes initiation and binding by PIC followed by rapid elongation along the DNA template. Histone turnover has been measured using different methods including fluorescence recovery after photobleaching (FRAP), which has revealed that different histone variants display different exchange rates in the nucleus and that full recovery is not achieved on typical time scales measured by FRAP (tens of minutes) (47, 48). However, FRAP measures the global pool of molecules in undefined genomic regions. In contrast, pulse-chase experiments combined with high-throughput sequencing has allowed for estimation of histone turnover rates genome wide (49, 50). In ES cells, histone H3.3 exchanges gradually over the course of at least 12 hours at transcription start sites. Interestingly, active regulatory regions including enhancers were found to exhibit the highest turnover of H3.3, suggesting dynamic exchange of nucleosomes is a feature of active chromatin (50). Chromatin dynamics have also been studied in the context of the cell cycle (see section 1.7). For example, accessibility is almost completely lost on nascent chromatin after replication, and recovers over the course of two hours in mouse ES cells (51). Furthermore, histone marks are deposited on the chromatin of newly synthesized daughter strands with largely variable kinetics. For example, the repressive mark H3K27me3 recovers its pre-replication levels slowly (over 20 hours) (52), while H3K4me3 recovers after around 6 hours (53). ES cells have been suggested to be in a “hyperdynamic” chromatin state with higher histone exchange compared to differentiated cells, and that this is important for their ability to change cell fate (48). Whether this is related to the short cell cycle length of ES cells is unclear. In conclusion, chromatin accessibility, histone variants, and histone modifications are highly dynamic and these dynamics are context-dependent and regulated. Furthermore, which features of chromatin dynamics are important for gene regulation and which factors influence this is not fully understood.

1.7 The cell cycle

The doubling of genetic material and derivation of daughter cells, called cell division, is the basis for the continual existence of life on Earth. While the majority of cells of adult animals reside in a dormant non-dividing state, some cell types pass through the cell cycle to divide, including stem cells (2). During development, the zygote divides and gives rise to trillions of cells, requiring the tight regulation of which cells should divide and when. The cell cycle consists of four major phases, G1, S, G2, and M. In G1, short for first Gap phase, cells have just divided, are growing, and will soon enter S phase, for Synthesis. In S phase, replication proceeds to make a near-perfect copy of each chromosome. In the second gap phase, G2, cells are waiting to enter the final stage of cell division, mitosis (M). Mitosis consists of multiple stages where the chromosomes condense (prophase), the nucleus

breaks down (prometaphase), chromosomes get aligned in the center of the cell (metaphase), sister chromatids are segregated (anaphase), and two new nuclei form (telophase) before the final stage where the cellular material splits into two, each consisting of one nucleus (cytokinesis). The regulatory proteins responsible for progression through the cell cycle have been well studied and includes feedbacks between fluctuating levels of cyclins and their cyclin-dependent kinases that phosphorylate other proteins (54). Importantly, all cellular material needs to exactly doubled on average each cell cycle, meaning that transcription as well as translation and degradation need to be tightly regulated and coupled to cell cycle progression (55).

1.7.1 Interplay between the cell cycle and cell fate

The cell cycle has been linked to lineage decisions in multiple contexts, in particular with regards to pluripotency (54). It was early noted that ES cells display a short cell cycle with a particularly short G1 phase (56). However, cell cycle regulation has been shown to vary according to the state of pluripotency (57). Multiple studies in human ES cells have suggested that the G1 phase of the cell cycle is instructive for differentiation and that this is regulated in part by interaction with the cell cycle machinery (54, 58, 59). Increased propensity to differentiate in G1 has also been shown in mouse ES cells, but the differences were smaller than reported for human (60). Interestingly, inhibition of Myc or mTOR can lead ES cells to enter into a reversible dormant state, maintaining their pluripotency (61, 62). This is proposed to mimic diapause, where unfavorable conditions can induce a pause in the development of newly developing embryos (this does not occur in humans). These dormant cells divide at highly reduced rates and are transcriptionally suppressed. However, their ability to differentiate in the absence of cell cycle progression is not clear as these studies did not assess the impact of withdrawal of pluripotency signals or addition of differentiation signals. Another link between the cell cycle and cell fate decisions is upon reprogramming of somatic nuclei using SCNT, where mitotic nuclei reprogram at higher rates than interphase nuclei (63). These results suggest that mitosis may be a particularly permissive state for cell fate changes. Overall, the relationship between the cell cycle and cell fate is complex and not fully understood. More details on this can be found in sections 3.2 and 3.4.

1.7.2 Impact of the cell cycle on chromatin and gene regulation

One important aspect of cell cycle progression is its relationship with chromatin states and how chromatin perturbations imposed by DNA replication and cell division impact gene regulation. Active regions of the genome are replicated early in S phase compared to inactive regions (64) and the transmission of epigenetic status (i.e. nucleosome occupancy and histone modifications) to newly synthesized daughter strands is highly regulated (65). During mitosis, phosphorylation of histone tails induces chromatin compaction and most other interphase histone marks are retained while some are diminished, and this appears to be partially cell type-dependent (66–69). Furthermore, transcription is almost fully shut off during mitosis and recovers over the course of a few hours, called transcriptional reactivation (70). Several studies have suggested that histone marks are functional “bookmarks” involved in transcriptional reactivation after mitosis (66, 71). It is feasible that stable histone modifications throughout mitosis would enhance the binding of TFs after mitosis and thus contribute to maintenance of cell type-specific gene expression programs, although this has not been formally demonstrated. It should be noted that one study assessing global transcriptional reactivation found that housekeeping genes were more rapidly reactivated than cell type-specific genes (70). The different mechanisms at play in regulating gene reactivation are thus not fully understood. Chromatin accessibility is mostly retained at gene promoters during mitosis, while enhancers display diminished, but not completely abolished, accessibility (72–74). This indicates that mitotic chromatin is accessible to proteins despite its high level of compaction. Indeed, studies have shown multiple DNA-binding proteins that associate to mitotic chromatin, including transcription factors (see section 1.8.2) and epigenetic regulators (68). Whether the accessibility of mitotic chromatin plays a functional role and whether it is actively or passively maintained remains unclear. Furthermore, it is not currently known by which mechanisms chromatin accessibility is regulated during and after replication and mitosis.

1.8 Transcription factors

To express a gene at the correct level, Pol II must get recruited to the right place at the right time. The main players in this recruitment process are TFs. TFs have affinity for specific nucleotide sequences of DNA and this feature engenders them with a powerful mechanism to regulate gene expression, as they are the first molecules reaching DNA in the gene expression cascade (5). TFs were first discovered by biochemical extracts of proteins bound to specific DNA sequences (75). These proteins could also promote transcription when expressed in cells with a reporter gene harboring their sequence recognition site, known as motif, in the promoter (76). TFs consist of separate polypeptide domains with different functions including at least a DNA-binding domain that provides sequence specificity and a transactivation domain that promotes transcription. DNA binding domains have been well studied and can be separated into defined categories of homologous domains such as leucine zippers, zinc fingers, and homeodomains (76). In contrast, transactivation domains are less clearly delineated by amino acid sequence homology and tend to be enriched for low-complexity sequences (also known as intrinsically disordered regions) (75, 77). Notably, TFs can also repress transcription of genes by recruiting factors known as co-repressors (78).

TFs can be broadly categorized into general TFs and specific TFs (2). General TFs bind core promoter elements that exist at most or all genes and are required as part of the PIC for transcription. Specific TFs on the other hand bind sequences that can be scattered at different places in the genome. Specific TFs can be both ubiquitously expressed, expressed in some cell types, or be expressed in response to specific stimuli. They can also be expressed but only activated upon specific stimulation, as is the case for signaling effector proteins such as SMADs (79). In mammals, the number of TFs is estimated to range between around 1'500 to 2'500 (80, 81), thus comprising a substantial portion of protein-coding genes.

1.8.1 Pioneer transcription factors

TF binding was initially studied mainly in bacteria that do not have histones and using in vitro binding assays with naked DNA templates. How TFs access their binding sites in the context of eukaryotic chromatin was thus an unanswered question that remains incompletely answered. It was early suggested that there is a competition between TFs and nucleosomes based on the observation that part of the activating potential of TFs was mediated via chromatin-based mechanisms (82). It was later discovered that some factors, including FOXA1 and GATA4, can bind their sequence motifs on nucleosomal DNA and even promote to the eviction of histones in vitro, without the need for other regulatory factors (83). The TFs capable of binding nucleosomal DNA were dubbed “pioneer factors” due to their ability to bind sites in the genome before other regulatory proteins, for example upon differentiation (84). However, most uses of the term pioneer also include the ability to promote eviction of histones, either directly or by recruitment of other factors such as chromatin remodeler (see below). I will use this last definition throughout this thesis, i.e. pioneer factors are able to bind to and lead to the opening of compacted chromatin.

Since the initial discovery of pioneer factors, many TFs have been proposed to harbor this property (85). Pioneer factors are often able to drive cell fate transitions due to their ability to activate previously dormant regulatory elements. Furthermore, pioneering activity has been shown to be crucial for cellular reprogramming (86). For example, the pluripotency TFs OCT4, SOX2, and KLF4 can bind nucleosomal DNA in vitro and in vivo, which when overexpressed in somatic cell types can lead to the activation of embryonic regulatory elements (87). However, they are unable to access sites enriched for heterochromatin marks such as H3K9me3, showing that pioneering does not mean indifference to the chromatin status (88). Indeed, most TFs including pioneers bind a small subset of all their available binding sites in any given cell type, indicating that they do not intrinsically possess the ability to ubiquitously access the genome. Thus, while pioneer factors can access nucleosomal DNA and induce their opening, they are still subject to additional layers of regulation, including the expression of other TFs and co-regulators and features of the chromatin such as compaction, histone marks, and DNA methylation.

The ability to bind nucleosomal DNA has been shown to be conferred at least in part by structural properties of the proteins (89). After binding to nucleosomes, how do pioneers mechanistically evict histones? As mentioned above, FOXA1 and GATA4 can do this without the need for any other co-factors. In the case of FOXA1, its C-terminal end structurally resembles the linker histone H1, which enables histone eviction (90). However, neither OCT4, SOX2, or KLF4 can evict nucleosomes *in vitro*, although they are able to bind to them. It was recently shown that OCT4 occupancy is strongly correlated to binding of the protein BRG1. BRG1 is the ATPase subunit of the BAF chromatin remodeling complex (46). 24-hour depletion of OCT4 from embryonic stem (ES) cells leads to a loss in chromatin accessibility at thousands of OCT4 binding sites. This loss of accessibility is also seen upon depletion of BRG1. Thus, OCT4 accomplishes at least a large part of its pioneering function via recruitment of the BAF complex. Another pioneer GATA3 appears to use a similar mechanism (91), but the prevalence of recruitment of chromatin remodelers in pioneer factor function is underexplored. Furthermore, even for OCT4 which directly interacts with BRG1 in pull-down assays (92), it is difficult to prove that recruitment is direct. It is possible that OCT4 recruits other intermediary factors that then recruit BRG1. Nevertheless, it is the current understanding that pioneer factors are either structurally able to evict histones by competitive binding, or recruit chromatin remodelers.

1.8.2 Transcription factor DNA-binding during mitosis

Early studies revealed that TFs are displaced from chromosomes during mitosis and this was proposed to possibly serve as a mechanism to “reset” transcriptional programs after cell division (63, 93). However, later observations showed that some TFs remain associated to mitotic chromatin, dubbed “mitotic bookmarking” TFs (94). Many such TFs have since been identified, but the functional role of mitotic bookmarking is still elusive, in large part due to the difficulty in developing tools that ablate only the mitotic bookmarking ability and do not interfere with the function of the TF in other cell cycle phases. Furthermore, there is a lack of consistency in definitions as to what should classify as mitotic bookmarking. The question is in principle whether apparent co-localization with mitotic chromatin by microscopy is sufficient or whether site-specific binding has to be observed. This is further complicated by the many different methods to measure binding, including immunoprecipitation (IP; see ChIP-seq in section 1.11) and single-molecule imaging-based methods which each have their pros and cons. Nevertheless, it is clear that some fraction of TFs associate with mitotic chromatin as observed by fluorescence microscopy and sometimes IP. Two examples include FOXA1 (95) and GATA1 (96), which have been shown to bind mitotic chromosomes in liver and erythroid cells, respectively (although it should be noted that evidence suggests that association observed by fluorescence microscopy is cell type-independent (97)). Attempts were made to characterize the functional impact of mitotic binding by using systems to deplete the factor, synchronizing cells in mitosis, and then releasing them to observe transcriptional reactivation kinetics. FOXA1 depletion led to a slower reactivation of FOXA1-bound genes, but this included genes that did not appear to be bound by FOXA1 during mitosis (95). In contrast, GATA1-depletion led to a slower reactivation specifically of mitotically bound genes (96). However, subsequent genome-wide analyses showed that GATA1 mitotic occupancy was not predictive of transcriptional reactivation kinetics (71). The rapid binding and unbinding of many observed TFs, in the range of seconds to minutes (98), seems incompatible with stable site-specific occupancy in mitosis. While other mechanisms can be envisioned for how mitotic binding by TFs would contribute to epigenetic memory, such as functional binding events leading to chromatin remodeling or high local concentrations allowing for rapid re-binding in early G1, such scenarios have not been explored. In summary, TFs can associate with mitotic chromatin but the nature of the interactions and their functional relevance is unclear.

1.9 Regulation of pluripotency by transcription factors

The regulation of pluripotency has been extensively studied, in particular using the model system of mouse ES cells. Signaling pathways, including LIF, Wnt, and FGF/MEK, reinforce a network of TFs that together control the pluripotent state (99). Using correlative measures of expression levels across different pluripotency conditions, researchers have been able to model the interdependent pluripotency TF network and establish a minimal computational model containing 12 components and 16 interactions that can explain the experimental data (100). The TF nodes in this network

include STAT3, GBX2, KLF4, SALL4, KLF2, NANOG, ESRRB, TFCEP2L1, OCT4, and SOX2. However, the large diversity of TFs expressed in ES cells are not all required to maintain pluripotency and display a large array of phenotypes when depleted or overexpressed. For example, NANOG overexpression can sustain pluripotency in the absence of LIF signaling (101) but self-renewal does not strictly require NANOG (102) or ESRRB (103). However, cells devoid of NANOG or ESRRB die upon differentiation (102, 103). REX1, a marker of the ground state of pluripotency, is completely dispensable for embryonic development (104). KLF4 is required for self-renewal in ground state ES cells (105), but is not expressed in primed EpiSCs, which maintain pluripotency although resemble a later developmental stage (99). In contrast, SOX2 and OCT4 have been shown to be strictly required for pluripotency maintenance across all pluripotent cell types (99). These results highlight the complexity of the pluripotent regulatory network. More on this topic can be found in section 3.2.

1.9.1 The roles of OCT4 and SOX2 in pluripotency and differentiation

OCT4 is encoded by the *Pou5f1* gene and is part of the Pit, Oct, Unc (POU) family of TFs that bind to the octamer motif ATGCAAAT and variations thereof (106). It consists of a bipartite DNA-binding domain, with a specific (POUs) and homeodomain (POUh) region, which is flanked by N-terminal and C-terminal transactivation domains (107). SOX2 is member of the SRY-related-HMG-box family of TFs and consists of an HMG DNA-binding domain, which recognizes the CATTGTA motif, and a C-terminal transactivation domain (108). OCT4 and SOX2 bind to thousands of regions in pluripotent stem cells and can bind as a heterodimer on a composite OCT4-SOX2 motif (46, 109). In ES cells, binding of OCT4, SOX2, as well as NANOG measured by ChIP-seq is highly correlated, indicative of their cooperative interactions (46). OCT4 and SOX2 also interact with a large array of other regulatory proteins, such as transcription TFs including SALL4 and ESRRB as well as transcriptional regulators including members of the NuRD and PRC1 repressor complexes (92, 110, 111). OCT4 interacts with the BAF chromatin remodeling complex to assert its pioneering function, as mentioned above, and SOX2 also interacts with BRG1 (112) but it is unclear whether this is required to perform its pioneering function. Their interaction network enables OCT4 and SOX2 to both activate and repress target genes, although for OCT4 it has been shown that its activating function is sufficient to support pluripotency (113). The *Pou5f1* and *Sox2* genes are transcriptionally induced at the 4- to 8-cell stage during mouse embryogenesis, and OCT4- or SOX2-depleted embryos die at the blastocyst stage, where they fail to establish pluripotency (114). Conditional depletion of OCT4 or SOX2 in embryos and embryonic stem cells leads to aberrant differentiation into TE (114–117). Interestingly, it has been shown that pluripotency can be sustained upon depletion of SOX2 if OCT4 is overexpressed (116), although the mechanism behind this is unclear and may rely on other SOX family members that can substitute for SOX2 (118). While SOX2 overexpression can maintain pluripotency, OCT4 levels need to be within a controlled range to avoid differentiation (99). When overexpressed, OCT4 promotes differentiation into PrE and mesoderm while repression leads to TE differentiation (115). However, low levels of OCT4 (but not low enough to trigger TE differentiation) have been shown to support pluripotency to a greater extent than wildtype levels (119, 120). OCT4 and SOX2 are also involved in cell fate decisions after the establishment of pluripotency. OCT4 depletion post-implantation leads to developmental defects at the gastrulation stage (121). Furthermore, OCT4 expression is correlated to mesendoderm differentiation of ES cells, and overexpression of OCT4 represses neuroectodermal differentiation (122). SOX2 is expressed in the neuroectoderm as well as in neuromesodermal progenitors which reside in the caudal epiblast and give rise to somites and cells of the spinal cord (123–125). SOX2 is also expressed in neural progenitor cells (126, 127) and SOX2 overexpression in ES cells has been shown to inhibit differentiation into mesendoderm (122). Notably, OCT4 exists in different splice isoforms (128) and SOX2 and OCT4 can be posttranslationally modified (129). For example, OCT4 is phosphorylated at serine 229 by Aurora kinase b upon entry into mitosis (130), which has been proposed to inhibit its binding to mitotic chromatin. Importantly, OCT4 and SOX2 can support the reprogramming of somatic cells into induced pluripotent stem cells when expressed in combination with KLF4 and c-MYC (131).

1.10 Heterogeneity in gene expression

Most measurements of RNA and protein abundance constitute population averages of cells collected for biochemical extraction. In contrast, when measuring RNA at the single-cell level, heterogeneity between different cells can be found even within phenotypically homogeneous cell populations, i.e. the same gene may be expressed at high or low levels at any given moment. This is caused in part by the discontinuous process of transcription that produces RNA molecules in short bursts, with longer periods of inactivity (132, 133). This bursting is stochastic and can thus lead to variation in gene expression over time. The final protein concentration of a given gene is dependent on multiple additional factors, including mRNA processing, export, translation, and the half-lives of mRNA and protein. These layers of regulation buffer the highly stochastic transcription process to yield more stable protein levels, but these can also be heterogeneous (134). Protein expression heterogeneity has been linked to phenotypic consequences. This includes transitions between different phenotypic stages, as is the case for NANOG whose expression levels fluctuate in metastable ES cell cultures (102, 135), or pulsing expression of P53 in DNA damage repair (136). However, the impact of endogenous protein fluctuations around their mean levels in a phenotypically homogeneous population of cells has been less explored. More background to this topic can be found in section 3.2.

1.11 Methods

The studies presented in this thesis are largely based on genetic modification of ES cells, controlled protein depletion strategies, and high-throughput sequencing-based methods to quantify regulatory features and protein binding. I will explain these in turn.

Genetic modifications of mammalian cells

A major milestone in mammalian research came with the development of genetic tools to delete certain portions of the genome, generating knock-outs. Introducing DNA constructs into cells harboring regions overlapping a gene on both sides, with a replaced genetic element in the middle, could lead to the homologous recombination (HR) and replacement of a gene of interest (137). If replaced by an antibiotic resistance gene, successfully modified cells could be selected by antibiotics. After performing knock-outs in mouse ES cells, these could be injected into a new developing blastocyst and contribute to the development of a chimeric (mixed DNA) embryo. This litter was then further bred and crossed to generate isogenic mice heterozygous or homozygous for knockouts of specific genes, so called knock-out mice (138). The ease of culture and genetic manipulation made ES cells into one of the main tools used to study mammalian biology. The more recently developed CRISPR-Cas9 system allows for targeting virtually any genomic sequence using guide RNAs and inducing a DNA break, which further increases the probability of recombination by either non-homologous end joining (NHEJ) or HR (139). NHEJ can lead to a frame shift mutation, which effectively abolishes the gene, while the increased HR efficiency facilitates introduction of new genetic elements including fusion constructs to tag genes with other proteins (e.g. fluorescent proteins), so-called knock-in.

Controlled protein depletion strategies

Using knock-outs to study gene function has several drawbacks, including the inability to interrogate essential genes and to study the acute function of genes. Since genes can exhibit different functions in different cell types and through different developmental stages, simply deleting the gene can give misleading impressions about its role. Thus, several strategies are available to delete genes in a controlled manner. Inducible knock-outs utilize the Cre recombination system (140). The gene is modified to be flanked with loxP elements, which will recombine and excise the gene upon exposure to Cre recombinase. Cre recombinase can be expressed ubiquitously but retained in the cytoplasm by fusion to the ligand binding domain of the estrogen receptor, which is shuttled into the nucleus upon exposure to tamoxifen. Alternatively, Cre may be driven by a promoter of a cell type-specific gene and thus be expressed (and induce knockout) only in tissues of interest. Another depletion strategy involves replacing the gene promoter by one controlled by an inducible

system. For example, the tet repressor (tetR) promoter, together with its transactivator (tTA), can be used to either induce and repress transcription from the promoter upon exposure to tetracycline or doxycycline (115). The main downside of Cre and tet systems is that depletion depends on the inherent degradation kinetics (half-life) of the protein under study. For example, a protein with a half-life of 5 hours would still retain a quarter of its protein level after 10 hours of exposure, prohibiting studies of processes occurring at shorter time scales. More rapid systems have been developed that depend on the induced degradation of proteins rather than deactivated transcription, using so called degrons. Such systems include HaloTag (141), FKBP (142), and Auxin-inducible degron (AID) (143). For example, upon exposure to indole-3-acetic acid (also known as IAA or Auxin), proteins tagged with the AID domain get ubiquitinated by the Tir1 protein and targeted for protein degradation, decreasing their half-life. The advantage of the AID system is the rapid degradation kinetics, although there are several disadvantages, including the need to genetically modify the protein (which may impair its function), the need to exogenously express the Tir1 protein in the cell line, and “background” degradation levels which decreases the half-life of the protein (and thus its overall concentrations) even without exposure to Auxin. In addition to inducible degrons, there are also degrons that operate throughout the cell cycle to induce degradation. For example, the degron normally found on cyclin B1 gets targeted for rapid degradation at the end of mitosis by the anaphase-promoting complex (APC), and can be fused to other proteins to degrade them specifically at the metaphase-anaphase transition (96).

ChIP-seq

One crucially important parameter to measure to study gene regulation is the binding sites of DNA-binding proteins. Where a protein binds or does not bind limits its potential functional impact and can also inform about important regulatory regions, such as enhancers. The most widely used method for such studies is Chromatin Immunoprecipitation followed by high-throughput sequencing (ChIP-seq) (144). In this method, cells are crosslinked using chemicals (mainly formaldehyde), chromatin is extracted and fragmented (using sonication or enzymatic digestion), and DNA-fragments bound by (i.e. crosslinked to) a protein of interest are pulled down using specific antibodies. When coupled with high-throughput sequencing, this gives a quantitative genome-wide view of protein binding, where occupied sites are detected as “peaks” above the background signal and the height of the peak determines the relative binding. While extremely powerful, it typically requires large numbers of cells (hundreds of thousands or millions), and the non-uniform background signal can lead to difficulties in determining true but low-occupancy binding sites (144). Furthermore, it cannot differentiate between direct (protein-DNA) from indirect (e.g. protein-protein-DNA) interactions. The use of fixatives may also affect the results, as epitopes may be more or less preserved after fixation. Native ChIP-seq, i.e. done without fixation, is possible but typically requires even larger number of cells and has been mostly used to study stably bound proteins such as histones (145, 146). Nevertheless, ChIP-seq can be used to determine binding sites and relative binding strength across the genome.

ATAC-seq

Accessibility, or openness, of chromatin is a feature that is required for functionally (see above). Determining the relative chromatin accessibility across the genome is thus a powerful way to determine which regions may be functional, and how accessibility changes upon perturbations can reveal functional interactions. To this end, enzymes that cut naked DNA but not nucleosomal DNA can be used. DNase I digestion was for long the method of choice and can, coupled with high-throughput sequencing, reveal genome-wide accessibility profiles (33). However, this assay typically requires large amounts (millions) of cells and is experimentally cumbersome. A recently developed method, assay for transposase-accessible chromatin using sequencing (ATAC-seq) (147), provides equivalent information but facilitates this by requiring fewer cells (can even be done on single cells (148)) and a rapid workflow. This method relies on the Tn5 transposome, which when loaded with sequencing oligonucleotides can simultaneously cut DNA and tag the fragments with overhangs that can be directly amplified for sequencing.

1.12 Aims of the thesis

The goal of this thesis is to address several questions regarding the relationship between the transcription factors OCT4 and SOX2, their pioneering activity, their endogenous fluctuations, the cell cycle, and cell fate decisions. One of the overarching questions is how the dynamics of chromatin accessibility are affected by the action of these pioneer transcription factors, especially across the cell cycle. The studies presented here have attempted to address the following (non-exhaustive) list of questions:

- Are there cell cycle-phase specific functional roles of OCT4 and SOX2?
- How is chromatin accessibility regulated by OCT4 and SOX2 across the cell cycle?
- How do OCT4 and SOX2 concentrations fluctuate over time and are there functional consequences of this?
- What are the dynamics of chromatin accessibility upon changes in pioneer factor levels?

Chapter 2 A role for mitotic bookmarking of SOX2 in pluripotency and differentiation

This chapter represents the following publication which has been formatted to fit with the style of the thesis:

Deluz, C.*, **Friman, E.T.***, Strebinger, D.*, Benke, A., Raccaud, M., Callegari, A., Leleu, M., Manley, S. and Suter, D.M. (2016) A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. *Genes & Development*, 30(22), pp.2538-2550 *Equal contribution

Contribution: For this collaborative work, I generated and validated the knock-in reporter cell line (called SBR) to monitor ES cell differentiation towards neuroectoderm (Sox1/GFP) and mesendoderm (Brachyury/mCherry) cell fates. I was also involved in performing differentiation experiments. Furthermore, I performed ChIP-seq of SOX2 in asynchronous and mitotically enriched ES cell populations. Finally, I took part in the editing of the manuscript and performed control experiments for the revisions.

Note that supplementary movies are available on the publisher's website.

2.1 Abstract

Mitotic bookmarking transcription factors remain bound to chromosomes during mitosis and were proposed to regulate phenotypic maintenance of stem and progenitor cells at the mitosis-to-G1 (M–G1) transition. However, mitotic bookmarking remains largely unexplored in most stem cell types, and its functional relevance for cell fate decisions remains unclear. Here we screened for mitotic chromosome binding within the pluripotency network of embryonic stem (ES) cells and show that SOX2 and OCT4 remain bound to mitotic chromatin through their respective DNA-binding domains. Dynamic characterization using photobleaching-based methods and single-molecule imaging revealed quantitatively similar specific DNA interactions, but different nonspecific DNA interactions, of SOX2 and OCT4 with mitotic chromatin. Using ChIP-seq (chromatin immunoprecipitation [ChIP] combined with high-throughput sequencing) to assess the genome-wide distribution of SOX2 on mitotic chromatin, we demonstrate the bookmarking activity of SOX2 on a small set of genes. Finally, we investigated the function of SOX2 mitotic bookmarking in cell fate decisions and show that its absence at the M–G1 transition impairs pluripotency maintenance and abrogates its ability to induce neuroectodermal differentiation but does not affect reprogramming efficiency toward induced pluripotent stem cells. Our study demonstrates the mitotic bookmarking property of SOX2 and reveals its functional importance in pluripotency maintenance and ES cell differentiation.

2.2 Introduction

During mitosis, transcription is globally shut down, and RNA polymerases and most DNA-binding proteins are stripped off the chromosomes (149, 150). However, some transcription factors do not abide by this rule and remain bound to specific genes on mitotic chromosomes. These so-called mitotic bookmarking transcription factors are involved in physiological processes such as phenotypic maintenance (151–154) and ribosome biogenesis (155–157) as well as pathological events such as oncogenic transformation (154, 158, 159). However, the contribution of mitotic chromosome retention to these functions remains unknown. Transcription factors with mitotic bookmarking properties are often master regulators of cell fate, and some of them were reported to mediate rapid transcriptional reactivation after mitotic exit (158, 160, 161). This led to the suggestion that retention on mitotic chromosomes plays a role at the mitosis-to-G1 (M–G1) transition by ensuring proper restoration of the gene expression program and thereby cellular phenotype after cell division (162). This concept is particularly appealing in the context of self-renewing cells such as stem cells, which need to maintain their identity through a large number of cell divisions. However, to date, there is no direct experimental evidence substantiating this hypothesis.

Mammalian embryonic stem (ES) cells are maintained in a pluripotent state by a network of transcription factors (163) in which SOX2 and OCT4 play a central role. Both are strictly required for the maintenance of the pluripotent state (164, 165) and act mainly as a heterodimer that binds to a composite DNA motif to activate transcription of genes controlling pluripotency (166, 167). Together with KLF4, they also allow reprogramming of terminally differentiated cells to pluripotency (168, 169). SOX2 also plays a role in differentiation by favoring neuroectodermal commitment upon pluripotency exit (170, 171). Here we used live-cell imaging and biophysical techniques to demonstrate and characterize the retention of SOX2 and OCT4 on mitotic chromosomes of ES cells and performed ChIP-seq (chromatin immunoprecipitation [ChIP] combined with highthroughput sequencing) to show that SOX2 bookmarks specific genomic loci during mitosis. Using tools allowing us to specifically degrade SOX2 at the M–G1 transition and a new double knock-in reporter cell line to monitor neuroectodermal and mesendodermal differentiation, we dissect the role of SOX2 mitotic bookmarking in pluripotency maintenance, reprogramming, and differentiation.

2.3 Results

2.3.1 SOX2 and OCT4 have intrinsic mitotic chromosome-binding (MCB) properties that are independent of the pluripotency context

We first aimed to screen for pluripotency transcription factors with the ability to bind mitotic chromosomes. We engineered lentiviral constructs that allow doxycycline (dox)-inducible expression of 16 central pluripotency transcription factors (163), each of which is C-terminally fused to a yellow fluorescent protein (YPet) (Supplemental Methods). We then generated the 16 corresponding mouse ES cell lines that also constitutively expressed H2B-mCherry to identify mitotic cells and a reverse tetracycline transactivator (rtTA3G) to allow for inducible transcription factor-YPet expression and performed live-cell fluorescence imaging 16–24 h after dox treatment. All transcription factors were exclusively or mostly localized in the nucleus, suggesting that the YPet tag did not affect their subcellular localization (Supplementary figure 2.1A). Upon examination of metaphase cells, most transcription factors were either excluded or not clearly enriched on mitotic chromosomes, with the exception of SOX2, which was markedly colocalized with mitotic chromosomes (Figure 2.1A). To exclude a widespread effect of YPet in preventing mitotic chromosomal localization, for some of them, we also confirmed the absence of enrichment on mitotic chromosomes by immunofluorescence experiments (Supplementary figure 2.1B) and SNAP tag fusions (Supplementary figure 2.1C).

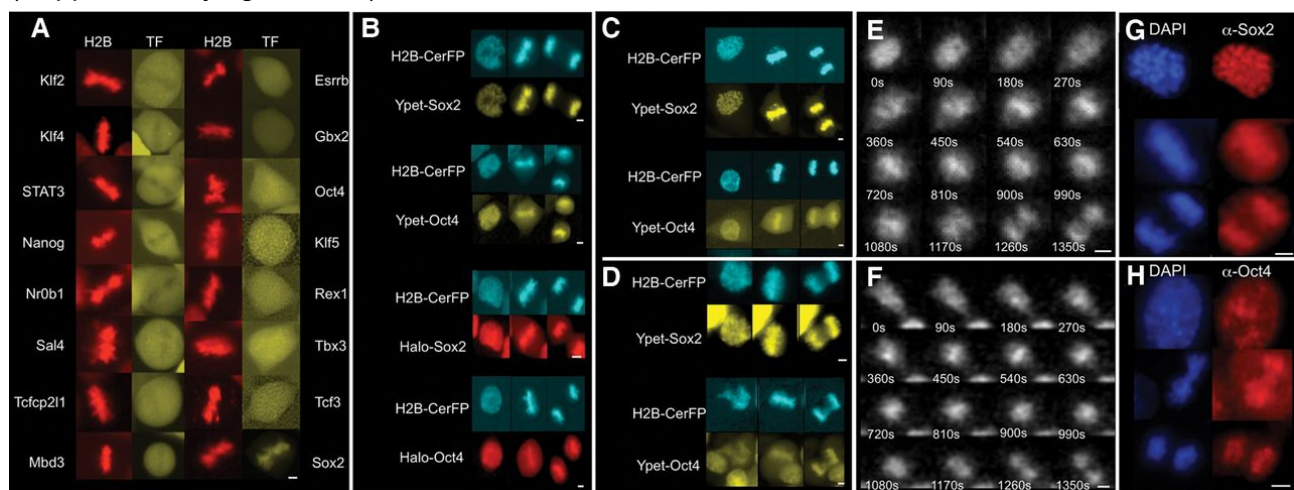


Figure 2.1 SOX2 and OCT4 bind to chromosomes throughout mitosis.

(A) Fluorescence snapshots of living ES cells in metaphase expressing H2B-mCherry and dox-inducible fusion proteins to pluripotency transcription factors. (B) Fluorescence snapshots of living E14 cells in prophase (left), metaphase (middle), and anaphase (right) expressing H2B-CerFP and different dox-inducible fusion proteins to SOX2 and OCT4. Cells expressing HaloTag fusions to SOX2 and OCT4 were labeled with HaloTag TMR ligand shortly before imaging. (C,D) Fluorescence snapshots of living NIH-3T3 cells (C) and HEK 293T cells (D) in prophase (left), metaphase (middle), and anaphase (right) expressing H2B-CerFP and YPet-SOX2 or YPet-OCT4. (E,F) Time series from luminescence time-lapse imaging of

SOX2-luciferase (SOX2-Luc) (E) and Luc-OCT4 (F) knock-in ES cell lines. (G,H) Immunofluorescence against Sox2 (G) and Oct4 (H) of E14 cells in prophase (top), metaphase (middle), and anaphase (bottom). Bars, 5 μ m.

Since OCT4 is a heterodimeric partner of SOX2 and appeared weakly enriched on mitotic chromosomes (Figure 2.1A), we generated four stable cell lines using higher lentiviral titers to allow broader expression ranges of N-terminal fusions of YPet or HaloTag to SOX2 and OCT4 under the control of a dox-inducible promoter (TRE3G) as well as constitutive expression of H2B-cerulean (H2B-CerFP) and rtTA3G. We then performed two-color time-lapse fluorescence imaging, revealing colocalization of H2BCerFP with YPet-tagged and Halo-tagged SOX2 and OCT4 throughout mitosis (Figure 2.1B; Supplementary figure 2.2; Supplemental Movies S1, S2). We then investigated MCB of SOX2 and OCT4 in NIH-3T3 and 293T cells, which are not pluripotent and do not express Sox2 and Oct4 endogenously. To do so, we engineered them for inducible expression of YPet-Sox2 or YPet-Oct4 and constitutive expression of H2B-CerFP. We observed that both transcription factors maintained their binding to mitotic chromosomes in these cell lines (Figure 2.1C,D), suggesting that this is an inherent property of SOX2 and OCT4 that is independent of the pluripotent context and that they can bind mitotic chromosomes independently of each other. To confirm MCB activity of endogenously expressed SOX2 and OCT4 in living cells, we used the CRISPR/ Cas9 technology to generate two heterozygous knock-in ES cell lines in which firefly luciferase (Luc) is fused to the C terminus of SOX2 and the N terminus of OCT4, respectively (Supplemental Methods; Supplementary figure 2.3). This enabled us to observe MCB by luminescence microscopy, allowing highly sensitive measurements that can be used to monitor the expression of endogenous genes in single cells (Suter et al. 2011). Approximately 15 min before cell division, the luminescence signal relocated to the midplane of the cell perpendicular to the cell division axis, consistent with a colocalization with metaphase chromosomes (Figure 2.1E,F; Supplemental Movies S3, S4). Finally, we performed immunofluorescence staining of endogenous Oct4 and Sox2, which also yielded consistent results (Figure 2.1G,H).

2.3.2 DNA-binding domain dependency and quantitative differences of SOX2 and OCT4 in MCB

To investigate the role of the DNA-binding domains of SOX2 and OCT4 in MCB, we generated different mutants that were N-terminally fused to YPet and expressed them in ES cells using our dox-inducible system. A deletion mutant of Sox2 devoid of its HMG DNA-binding domain lost its MCB activity (Figure 2.2A; Supplementary figure 2.2), whereas the HMG domain alone was sufficient for MCB (Figure 2.2B). Interestingly, removal of amino acids 133–135 of the nuclear localization signal (NLS) within the HMG domain reduced its MCB (Figure 2.2C). We generated two deletion mutants of Oct4, lacking either the POU_S or the POU_H DNA-binding domain. Deletion of the POU_S domain abrogated MCB (Figure 2.2D), in contrast to the POU_H domain deletion mutant, which retained a low MCB activity (Figure 2.2E; Supplementary figure 2.2). Furthermore, the POU_S domain alone displayed MCB (Figure 2.2F). This suggests that MCB of OCT4 is mediated mainly by the POU_S domain. Interestingly, the substitution of the POU_H domain with a NLS (RKRKR) rescued the MCB activity of OCT4 (Figure 2.2G). Taken together, these results suggest that the HMG domain of SOX2 and the POU_S domain of OCT4 are necessary and sufficient for their interaction with mitotic chromatin and that NLS sequences enhance their MCB.

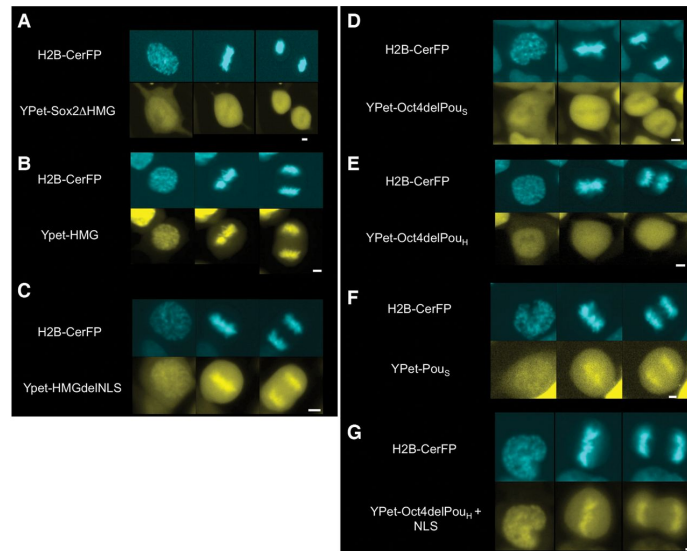


Figure 2.2 MCB of SOX2 and OCT4 is mediated by their HMG and POU5 domains, respectively. Fluorescence snapshots of living ES cells expressing H2B-CerFP and dox-inducible YPet fusion to truncated versions of SOX2 and OCT4.

(A) SOX2 with a deletion of its HMG DNA-binding domain. (B) HMG DNA-binding domain of SOX2. (C) HMG DNA-binding domain of SOX2 without its NLS. (D) OCT4 with a deletion of its POU5 DNA-binding domain. (E) OCT4 with a deletion of its POUH DNA-binding domain. (F) POU5 DNA-binding domain of OCT4. (G) OCT4 with a substitution of its POU5 DNA-binding domain with a NLS. In each panel, prophase is at the left, metaphase is in the middle, and anaphase is at the right. Bars, 5 μ m.

We next quantified the relative MCB strength of SOX2, OCT4, and their truncated versions by calculating the ratio of fluorescent signal on mitotic chromosomes to that in the cytoplasm in wide-field fluorescence microscopy images of cells in metaphase, a quantity that we define as the MCB index. We found a higher MCB index for SOX2 than for OCT4 in ES cells (Figure 2.3A), suggesting a stronger MCB activity for SOX2. To verify that the N-terminal fusion to the fluorescent tag did not bias the MCB efficiency of SOX2 and OCT4, we also generated cell lines expressing C-terminal fusions to YPet, which yielded consistent, although slightly higher, MCB indexes for both SOX2 and OCT4 (Figure 2.3A). We then asked whether these differences stem from intrinsic properties of SOX2 and OCT4 by measuring their MCB index in NIH-3T3 cells, which also yielded similar results (Figure 2.3B). To confirm the differences in MCB of endogenously expressed SOX2 and OCT4, we performed quantitative measurements on their MCB in the Sox2-Luc and Luc-Oct4 knock-in cell lines. Because of the lower spatial resolution of luminescence microscopy and the inability to simultaneously visualize fluorescently tagged H2B, we quantified the MCB properties of SOX2 and OCT4 by measuring their luminescence intensity profile across mitotic cells at metaphase, normalized over the average luminescence level four pixels from the center and on both sides of mitotic chromosomes (Figure 2.3C, inset). These measurements confirmed the weaker MCB activity of OCT4 as compared with SOX2 in the endogenous context (Figure 2.3C). Next, we investigated the interdependence of SOX2 and OCT4 in the relative strength of their MCB activity, which might be expected, since they can heterodimerize on DNA (167), and SOX2 was reported to assist DNA binding of OCT4 (172). To compare their MCB indexes in the presence or absence of each other, we generated four NIH-3T3 cell lines carrying an inducible construct for the expression of OCT4 fused to TagRFP-T (RFP-OCT4). This cell line (3T3-O) was further transduced to allow inducible expression of YPet-SOX2 (3T3-OS), YPet-SOX2 Δ HMG (3T3-OS Δ), or YPet-HMG (3T3-OH). Since coexpression of RFP-OCT4 and YPet-SOX2 decreased cell proliferation and thus the number of cell divisions that we could observe, we performed these experiments at lower doses of dox (20 ng/mL) on 3T3-O, 3T3-OS, 3T3-OS Δ and 3T3-OH cell lines. The MCB index of YPet-SOX2 was decreased only marginally in the presence of RFP-OCT4 (Figure 2.3D). We observed a similar MCB index of RFP-OCT4 in the 3T3-O, 3T3-OS Δ and 3T3-OH cell lines, but it was significantly increased in the 3T3-OS cell line (Figure 2.3D). Thus, SOX2 enhances the mitotic chromatin interaction of

OCT4, and the SOX2 HMG domain is required but not sufficient to increase the mitotic chromatin interaction of OCT4.

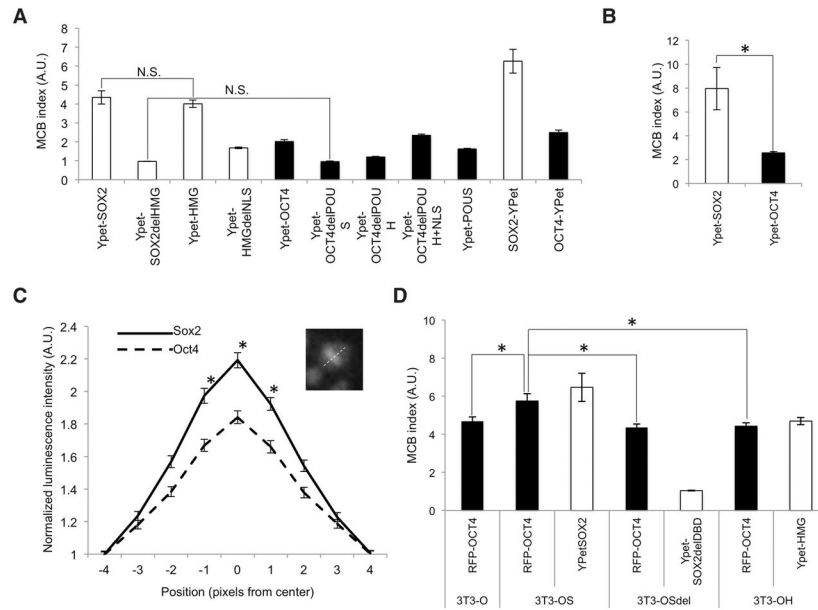


Figure 2.3 Differential retention and interdependence of SOX2 and OCT4 for MCB.

(A) MCB index of YPet fusions to wild type and mutants of SOX2 and OCT4 in ES cells. All samples were significantly different from each other ($P < 0.05$) unless specified. (N.S.) $P > 0.05$. $n \geq 20$. (B) MCB index of YPet fusions to SOX2 and OCT4 in NIH-3T3 cells. $n \geq 20$ (C) Quantification of luminescence signal across cells in metaphase for SOX2-Luc (solid line) and Luc-OCT4 (dashed line) knock-in cell lines. $n \geq 50$. The inset illustrates how the measurements were performed. (D) MCB index in 3T3-O, 3T3-OS, 3T3OSdel, and 3T3OH cell lines upon dox induction with 20 ng/mL. $n \geq 50$. (A.U.) Arbitrary units. (*) $P < 0.05$. Error bars indicate SEM.

2.3.3 SOX2 and OCT4 display distinct mobility but similar frequencies and residence times of long-lived DNA-binding events on mitotic chromosomes

To determine the residence times of SOX2 and OCT4 on mitotic chromatin, we performed single-molecule live-cell imaging experiments in ES cell lines that allow dox-inducible expression of Halo-SOX2 and Halo-OCT4 that we labeled with the Halo-TMR dye. Cells were treated with 50 ng/mL dox, allowing low Halo-tagged transgene expression levels for accurate identification of single DNA-bound molecules (173). We performed measurements on interphase and mitotic cells in the asynchronous population using highly inclined and laminated optical sheet (HILO) microscopy (174). To determine residence times on DNA ($1/k_{off}$), we used a previously described time-lapse imaging strategy (173) using imaging parameters that allowed us to measure long-lived specific DNA-binding events. The residence times that we measured in interphase were in close agreement with values described earlier for specific binding of SOX2 and OCT4 to DNA (172) and were only slightly shorter on mitotic chromatin; moreover, residence times were similar for both transcription factors (Figure 2.4A; Supplementary figure 2.4). We next investigated whether SOX2 and OCT4 have similar relative on rates of DNA binding. As $k_{on} = k_{off}/k_D$ and as SOX2 and OCT4 have similar residence times $1/k_{off}$, differences in k_{on} should manifest as differences in k_D (see the Materials and Methods). Since the equilibrium constant k_D is proportional to the ratio of unbound over bound molecules, it scales inversely with the number of binding events divided by the total fluorescence of the cell in a given individual image. The values obtained for this ratio were similar for SOX2 and OCT4 in interphase and mitosis (Figure 2.4B), suggesting no major difference in their on rates.

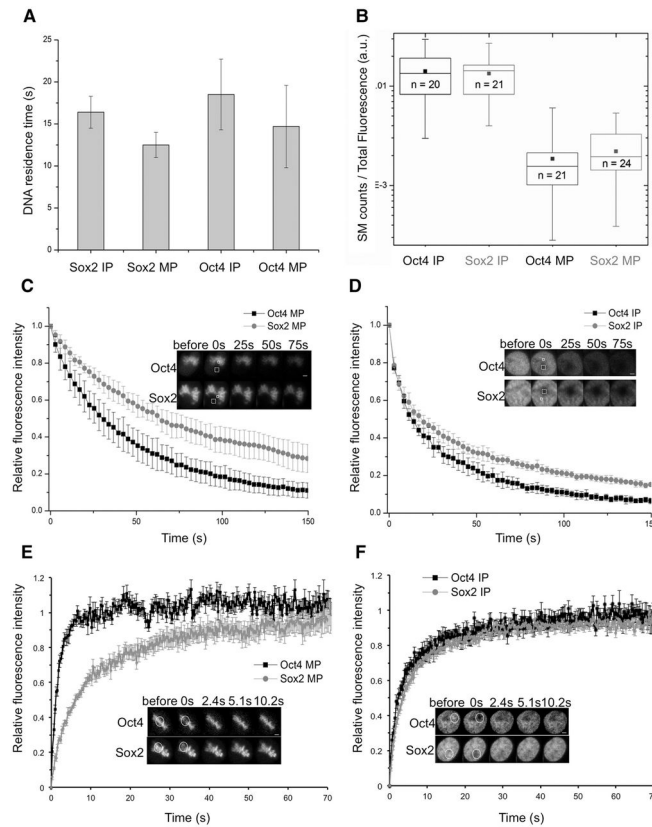


Figure 2.4 Dynamics of OCT4 and SOX2 in mitotic and interphase cells.

(A) SOX2 and OCT4 DNA residence times measured with time-lapse single-molecule microscopy. Error bars indicate the SE of linear fitting. (B) Relative bound fraction measured as single-molecule count per total cell fluorescence in a single frame (arbitrary units). Boxes represent 25–75 intervals, and whiskers indicate minimum and maximum values. The number of analyzed cells is indicated inside the boxes. (C,D) Fluorescence loss in photobleaching (FLIP) curves (arbitrary units) for OCT4 and SOX2 in M phase (MP) (C) or interphase (IP) (D). $n = 10$. Error bars indicate SD. The insets show examples of FLIP time series. (Dashed square) Bleaching area; (solid square) fluorescence recording area. Bars, 2 μm . (E,F) Fluorescence recovery after photobleaching (FRAP) curves (arbitrary units) for OCT4 and SOX2 in M phase (MP) (E) or interphase (IP) (F). $n = 10$. Error bars indicate SE. The insets show examples of FRAP time series. (Solid circle) Bleaching and fluorescence recording area. Bars, 2 μm .

We next performed fluorescence loss in photobleaching (FLIP) and fluorescence recovery after photobleaching (FRAP), which mainly reflect interactions with nonspecific binding sites (175), to measure the mobility of OCT4 and SOX2 in interphase and mitotic cells. YPet-SOX2 and, to a lesser extent, YPet-OCT4 displayed a slower fluorescence loss in mitosis ($t_{1/2}$ bleach SOX2 = 45.6 sec \pm 8.8 sec; $t_{1/2}$ bleach OCT4 = 29.0 sec \pm 8.8 sec) (Figure 2.4C) as compared with interphase ($t_{1/2}$ bleach SOX2 = 13.1 sec \pm 3.6 sec; $t_{1/2}$ bleach OCT4 = 12.9 sec \pm 2.2 sec) (Figure 2.4D), suggesting that SOX2 and OCT4 are exchanged less between mitotic chromosomes and the mitotic cytoplasm than within the interphase nucleus. In principle, this could be due to a lower mobility of molecules bound to mitotic chromosomes and/or a global retention within the mitotic chromosome environment. FRAP experiments revealed a slower recovery of YPet-SOX2 on mitotic chromosomes ($t_{1/2}$ recovery = 6.00 sec \pm 1.99 sec) (Figure 2.4E) as compared with interphase ($t_{1/2}$ recovery = 3.81 sec \pm 1.15 sec) (Figure 2.4F), thus confirming the lower mobility of SOX2 in mitotic cells. However, YPet-OCT4 showed the opposite behavior, with a faster recovery on mitotic chromosomes ($t_{1/2}$) as compared with interphase ($t_{1/2}$ recovery = 2.83 sec \pm 0.91 sec) (Figure 2.4F), suggesting that the slower FLIP half-time of mitotic OCT4 is not due to a lower mobility of OCT4 but rather its retention within the mitotic chromosome environment. Since mitotic chromatin consists mainly of highly compacted DNA wrapped around nucleosomes, our results could be explained by the higher affinity of SOX2 for nonspecific sequences on nucleosomal DNA as compared with OCT4 (176).

2.3.4 SOX2 is retained on a small number of genomic loci during mitosis

To investigate the genome-wide distribution of SOX2 during mitosis, we performed three independent ChIP-seq experiments on purified mitotic cells and asynchronous cells. Mitotic cells were obtained by 12 h of treatment with 200 ng/mL nocodazole followed by H3S10P labeling and sorting by FACS (Supplementary figure 2.6), as described previously (177). This resulted in 94.8% mitotic cells ($n = 1601$) in the sorted population as compared with 3.1% mitotic cells in the asynchronous samples ($n = 1029$), as assessed by inspection of DAPI staining of cell nuclei (Supplementary figure 2.5). We then performed Western blotting after Sox2 ChIP on mitotic and asynchronous cells, showing that Sox2 was pulled down in mitotic cells, although less efficiently than in asynchronous cells (Supplementary figure 2.6G). We performed ChIP-seq on SOX2 for both mitotic and unsynchronized samples and used model-based analysis of ChIP-seq (MACS2) (178) for peak calling on grouped triplicates from each condition. We included an additional filtering step to remove peaks previously identified as frequent artifacts in high-throughput sequencing data (excessive unstructured anomalous reads mapping) (Supplementary figure 2.6; (179)). High-amplitude peaks called in unsynchronized samples displayed either clear or no enrichment for SOX2 in mitotic samples, as assessed from sequence read visualization and ChIP-qPCR (ChIP combined with quantitative PCR) experiments (Figure 2.5A), thus excluding that peaks in mitotic cells are due to contaminating nonmitotic cells, confirming the purity of our mitotic cell preparation. MACS2 analysis yielded 10,523 peaks in asynchronous samples but only 84 peaks in mitotic samples (Figure 2.5B). While 35 out of 66 genes bound in mitosis were also bound in unsynchronized samples (Figure 2.5C), only a small number of called peaks overlapped between these two data sets (Figure 2.5B). Two factors may contribute to the low number of mitotic peaks: (1) the lower pull-down efficiency of SOX2 from mitotic samples, although it is unclear whether this is a general issue in the field, since differences in pull-down efficiency were not measured in other studies on mitotic bookmarking transcription factors, and (2) our stringent analytical peak calling; this is corroborated by the lower number of peaks that we found for GATA1 and FoxA1 when reanalyzing ChIP-seq data from (177) and (160) with our pipeline (Supplementary figure 2.6F,G). This raises the possibility that we might have missed a number of enriched loci, prompting us to perform visual track inspection of mitotic reads in regions where peaks were called only in the asynchronous samples. We indeed observed mitotic read enrichment in these regions and validated visually identified mitotic peaks located close to genes involved in pluripotency regulation (Rif1, Oct4, Nanog, and NFkB1) by ChIP-qPCR (Figure 2.5D). Conversely, peak-centered enrichment analysis (Supplementary figure 2.7) and ChIP-qPCR (Supplementary figure 2.8) revealed that peaks called only in mitotic samples were often enriched in the asynchronous samples as well, suggesting a larger overlap of bound loci than indicated by peak calling. While gene ontology analysis did not reveal any significantly enriched gene category in the mitotic samples, peak location analysis revealed a high proportion of mitotic peaks located within and downstream from genes, at the expense of promoters and upstream and intergenic regions (Supplementary figure 2.9). De novo motif analysis (MEME) revealed the known composite motif of Pou5f1::Sox2 as top-ranked in the asynchronous sample (Supplementary figure 2.10), but no motif could be identified from mitotic peaks. While this could be due to the low number of peaks called, it is also possible that DNA sequence-independent properties of mitotic chromatin mediating differential accessibility to DNA mediate SOX2 enrichment at certain genomic locations. Taken together, our data suggest that SOX2 is mostly bound nonspecifically to mitotic chromatin but is significantly enriched in the vicinity of a small number of genes, of which >50% overlap with genes also bound in asynchronous cells.

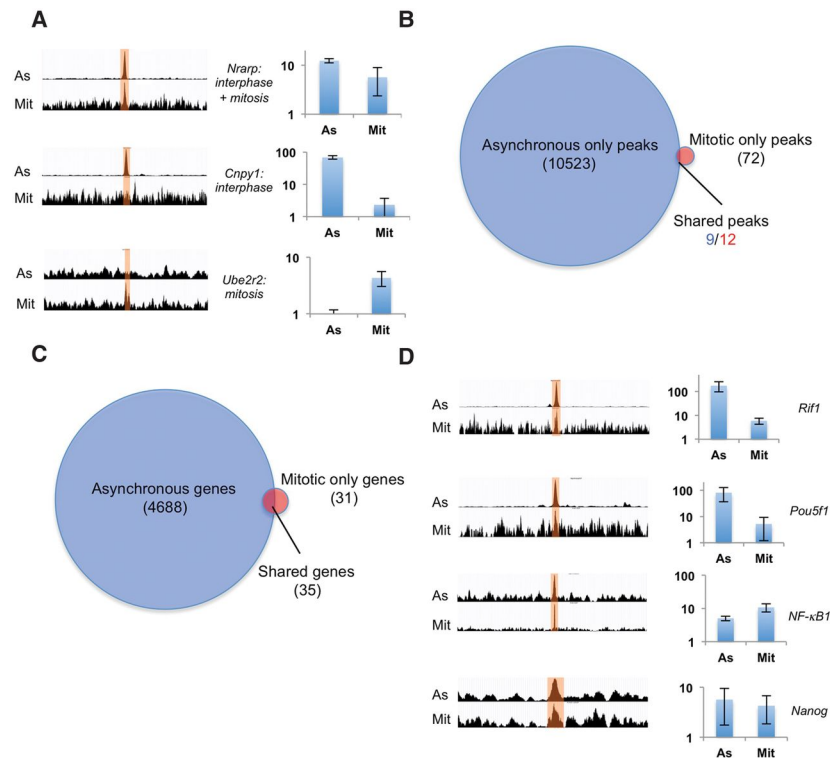


Figure 2.5 ChIP-Seq analysis of SOX2 in unsynchronized and mitotic cells.

(A) Examples of merged triplicate reads of peaks called in only the asynchronous samples, both the asynchronous and mitotic samples, or only the mitotic samples. For each example, asynchronous reads (As) are shown in the top row, and mitotic reads (Mit) are shown in the bottom row. Bar graphs show ChIP-qPCR results as relative fold enrichment over the input. $n = 2$. Error bars indicate SE. (B) Number of called peaks in asynchronous versus mitotic samples. The different numbers of shared peaks are due to the occasional occurrence of two peaks in the mitotic samples that match a single peak in the asynchronous sample. (C) Number of genes with SOX2 peaks 20 kb upstream of, 20 kb downstream from, or within genes. (D) Merged triplicate reads of peaks in the vicinity of pluripotency regulators that were not called but were identified as enriched by ChIP-qPCR. Bar graphs show ChIP-qPCR results represented as relative fold enrichment over the input. $n = 2$. Error bars indicate SE.

2.3.5 Functional interrogation of mitotic bookmarking by degradation of SOX2 at the M–G1 transition

We next asked whether sequestration of SOX2 on mitotic chromosomes and therefore its presence on or in the vicinity of target genes at the M–G1 transition are essential for its functions in cell fate decision making. To do so, we aimed at engineering tools that allowed us to (1) down-regulate SOX2 levels specifically at the M–G1 transition and (2) sort cell populations with defined average SOX2 expression levels to allow for meaningful comparisons with controls that are not degraded during mitosis. We thus fused SOX2 to a mitotic degron (MD) that allows degradation of SOX2 at the metaphase-to-anaphase transition or an inactive mutant thereof carrying a single R42A amino acid substitution that inactivates the destruction box (MD*) (177). Sox2-MD and Sox2-MD* were also tagged with either YPet or a SNAP tag, allowing us to sort populations according to their tagged SOX2 expression levels (Figure 2.6A). To assess the relative levels of these SOX2 fusion proteins in different phases of the cell cycle, we generated four ES cell lines expressing rtTA3G and the four respective TRE3G-driven fusion proteins, which were FACS-sorted 24 h after dox induction using windows spanning the same fluorescence intensities (Supplemental Methods). We then performed time-lapse fluorescence imaging of the YPet fusion and the SNAP-tag fusions labeled with the SNAP-SiR 647 dye (180). While the fluorescence signals for MD-fused SOX2 rapidly decreased during the metaphase-to-anaphase transition followed by an increase of the signal in newly divided cells, the MD*-fused SOX2 did not display observable changes in fluorescence signals during cell division (Figure 2.6B; Supplemental Movies S5–S8). We next generated two dox-inducible lentiviral vectors that allow expression of OCT4, KLF4, SOX2-YPet, and cMYC fused to either MD or MD* (OSKM-MD and OSKM-MD*, respectively) and transduced mouse embryonic fibroblasts (MEFs)

with either vector together with an expression vector for M2rtTA (181). We also observed a rapid decrease of MD-fused SOX2 that was not observable in the MD*-fused SOX2, but the recovery of the signal was slower than what we observed in ES cells (Figure 2.6B).

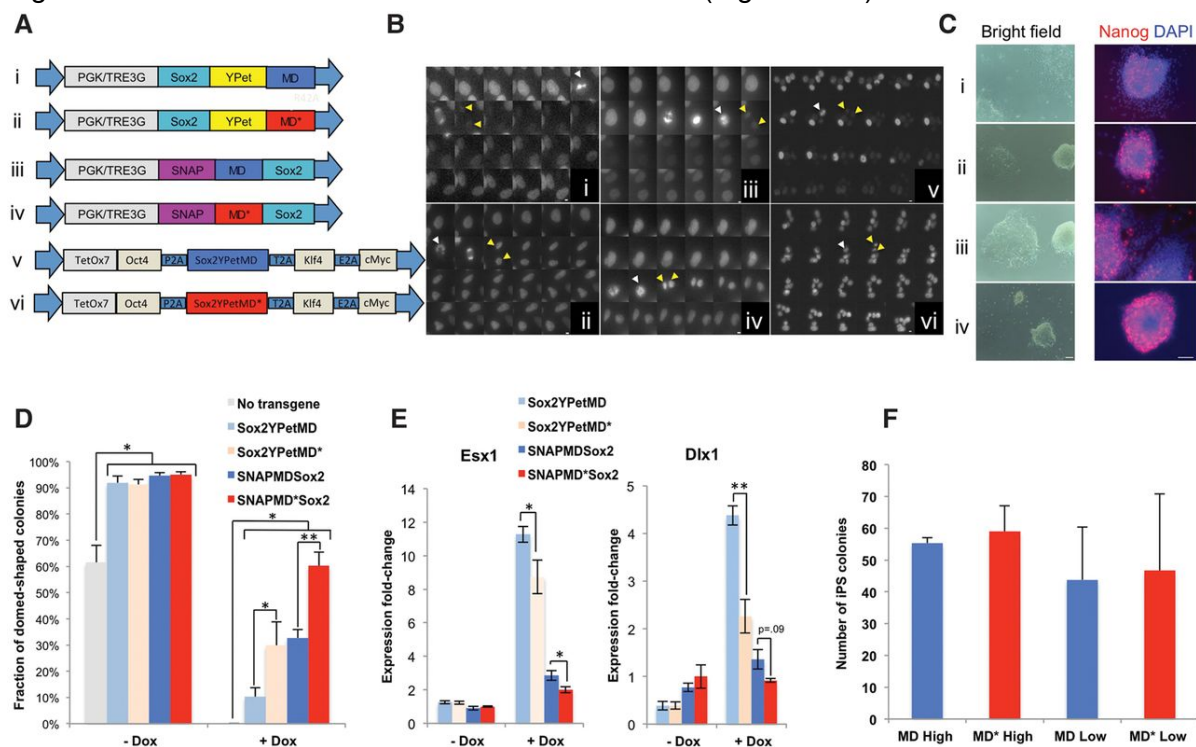


Figure 2.6 The role of SOX2 at the M-G1 transition in pluripotency maintenance and reprogramming to induced pluripotent stem (iPS) cells.

(A) Lentiviral constructs used to generate stable SOX2-MD and SOX2-MD* ES cell lines and for reprogramming to iPS cells. (B) Fluorescence time-lapse imaging of dividing ES cells (panels i-iv) or MEFs (panels v,vi) transduced with the corresponding lentiviral vectors (TRE3G promoter for panels i-iv) from A. Consecutive frames are separated by 15 min. Bars: panels i-iv, 5 μ m; panels v,vi, 10 μ m. White arrows indicate the cell just prior to division, and yellow arrows indicate the two daughter cells. (C) Representative bright-field images (left) and immunofluorescence staining against Nanog (red) and DAPI (blue) of the four 2T522C sub-cell lines transduced with the corresponding lentiviral vectors from A (PGK promoter) after 1 wk of culture in the presence of dox. (Red) Nanog; (blue) DAPI. Bars, 100 μ m. (D) Counting of dome-shaped and flattened colonies after 1 wk of culture with or without dox. n = 3 per condition for the original 2T522C cell line (no transgene); n = 6 per condition for all other cell lines. (E) RT-qPCR on Esx1 and Dlx1 after 1 wk of culture with or without dox, normalized against values obtained for SNAP-MD*-Sox2 without dox. n = 4 per condition. (F) Counting of iPS cell colonies generated from Oct4-GFP MEFs, as assessed by GFP fluorescence 20 d after transduction. High and low refer to different relative lentiviral titers. n = 3 high; n = 4 low. Error bars indicate SE. (*) P < 0.05; (**) P < 0.01.

We reasoned that there are two equally valid strategies for quantitative comparisons of MD and MD* cell lines: Either MD and MD* should have similar average SOX2 expression levels, at the expense of slightly higher SOX2 levels of the MD cell lines during cell cycle phases in which the MD is inactive, or both cell lines should have similar expression levels in cell cycle phases during which the MD is inactive but with slightly lower average expression levels for the MD cell lines. We were able to generate experimental conditions approximately matching these two strategies by sorting for narrow or large windows of fluorescence intensities, respectively (Supplemental Methods; Supplementary figure 2.11), and implemented these to assess the functional relevance of SOX2 at the M-G1 transition in pluripotency maintenance.

2.3.6 The presence of SOX2 at the M–G1 transition contributes to pluripotency maintenance but is dispensable for reprogramming to induced pluripotent stem (iPS) cells

To investigate the function of SOX2 at the M–G1 transition in pluripotency maintenance, we took advantage of the 2TS22C ES cell line, which allows turning off of Sox2 expression by addition of dox (182) and leads to rapid loss of pluripotency. We then generated four sub-cell lines that allow PGK-driven SOX2 fused to either YPet or a SNAP tag in combination with MD or MD* and sorted either narrow (PGK-Sox2-YPet fusion cell lines) or large (PGK-SNAP-Sox2 cell lines) fluorescence windows. We then compared the ability of the different SOX2 fusions to maintain pluripotency in the presence of dox. After 1 wk of dox treatment, we observed more flat colonies with uneven edges expressing lower and more heterogeneous Nanog levels in the MD cell lines when compared with the respective MD* cell lines (Figure 2.6C). We then quantified morphological changes by counting the number of flat colonies with uneven edges versus dome-shaped colonies with sharp edges in the four cell lines and the original 2TS22C cell line after 1 wk of culture in the presence or absence of dox. Although both MD-fused and MD*-fused versions of SOX2 partially rescued pluripotency as compared with the original 2TS22C cell line, this effect was significantly stronger in the MD* compared with the respective MD cell lines (Figure 2.6D). Using the same culture conditions, we also measured mRNA expression of *esx1* and *dlx1*, two trophodermal markers that have been shown to be up-regulated upon dox-mediated SOX2 knockdown in the 2TS22C cell line (182). We observed higher expression levels of both markers after 1 wk of dox in the MD cell lines as compared with the matched MD* cell lines, thus confirming the higher propensity of ES cells to differentiate in the absence of SOX2 during the M–G1 transition (Figure 2.6E). Interestingly, the SNAP-tagged versions of SOX2 were more potent in maintaining pluripotency than the YPet-tagged versions, which could be due to differences in expression levels between SNAP-tagged and YPet-tagged SOX2 cell lines and/or differences in the functionality of SOX2, depending on the tag and/or its location at the N terminus versus the C terminus. Taken together, our results suggest that the M–G1 transition is a key time window for the pluripotency-maintaining function of SOX2.

We next investigated whether the presence of SOX2 at the M–G1 transition is required for reprogramming to iPS cells. To do so, we generated lentiviral particles for the OSKM-MD and OSKM-MD* lentiviral vectors and measured the relative titers of both versions by transducing these vectors together with FuW-M2rtTA in 293T and quantifying YPet expression in the presence of dox by flow cytometry. We then used adjusted relative lentiviral titers of OSKM-MD and OSKM-MD* to transduce Oct4-GFP MEFs (183) using two different concentrations of lentiviral particles differing by a factor of 3. Dox was included in the cell culture medium from day 2–16 after transduction, and reprogrammed colonies were counted on day 20. We did not observe any difference in reprogramming efficiency between OSKM-MD and OSKM-MD*, suggesting that SOX2 expression at the M–G1 transition does not play a role in reprogramming toward iPS cells (Figure 2.6F).

2.3.7 SOX2 is required at the M–G1 transition to drive neuroectodermal differentiation

SOX2 has been shown to favor neuroectodermal over mesendodermal differentiation (170) and accelerate neuroectodermal commitment of ES cells upon overexpression (171). To study the role of mitotic SOX2 in neuroectodermal differentiation, we generated a double-reporter cell line to facilitate quantitative assessment of neuroectodermal and mesendodermal commitment. We used CRISPR–Cas9 to knock in a P2A-eGFP cassette downstream from the endogenous Sox1-coding sequence and a P2A-mCherry cassette downstream from the endogenous Brachyury-coding sequence (Supplemental Methods; Supplementary figure 2.12), allowing expression of eGFP in neuroectodermal cells and mCherry in mesendodermal cells (184, 185). We refer to this cell line as SBR (Sox1/Brachyury reporter). We confirmed the specificity of our reporters by differentiating the SBR cell line for 4 d and inspecting the colocalization of SOX1 and BRACHYURY immunoreactivity with eGFP and mCherry expression, respectively (Figure 2.7A). The SBR cell line was then used to generate two sub-cell lines expressing rtTA3G and either SNAP-MD-Sox2 or SNAP-MD*-Sox2. Since we intended to modulate average expression levels in both cell lines using different doses of dox, we used a single sorting strategy using a unique, narrow fluorescence intensity window using the SNAP-SiR-647 dye after overnight incubation with 100 ng/mL dox (Supplementary figure 2.11).

We then compared the potency of MD-Sox2 and MD*-Sox2 in enhancing neuroectodermal differentiation by culturing cells in N2B27 for 3 d in the presence of various doses of dox followed by flow cytometry to measure the fraction of eGFP+ cells. Whereas the SBR SNAP-MD*-Sox2 cell line displayed dose-dependent increased neuroectodermal commitment, differentiation of the SBR SNAP-MD-Sox2 cell line did not change upon dox addition (Figure 2.7B,C). We then compared the potency of MD-Sox2 and MD*-Sox2 in favoring neuroectodermal commitment over mesendodermal commitment by treating cells in N2B27 medium for 2 d followed by 2 d in N2B27 with 3- μ M CHIR, which enhances mesendodermal commitment (170). We observed a dox dose-dependent increase in the GFP+/mCherry+ ratio in the SBR SNAP-MD*-Sox2 cell line but not in the SBR SNAP-MD-Sox2 cell line (Figure 2.7D,E), suggesting that mitotic degradation of SOX2 suppresses its ability to favor neuroectodermal over mesendodermal commitment (Figure 2.7D,E). We thus conclude that the presence of SOX2 at the M–G1 transition is required for its ability to enhance neuroectodermal differentiation.

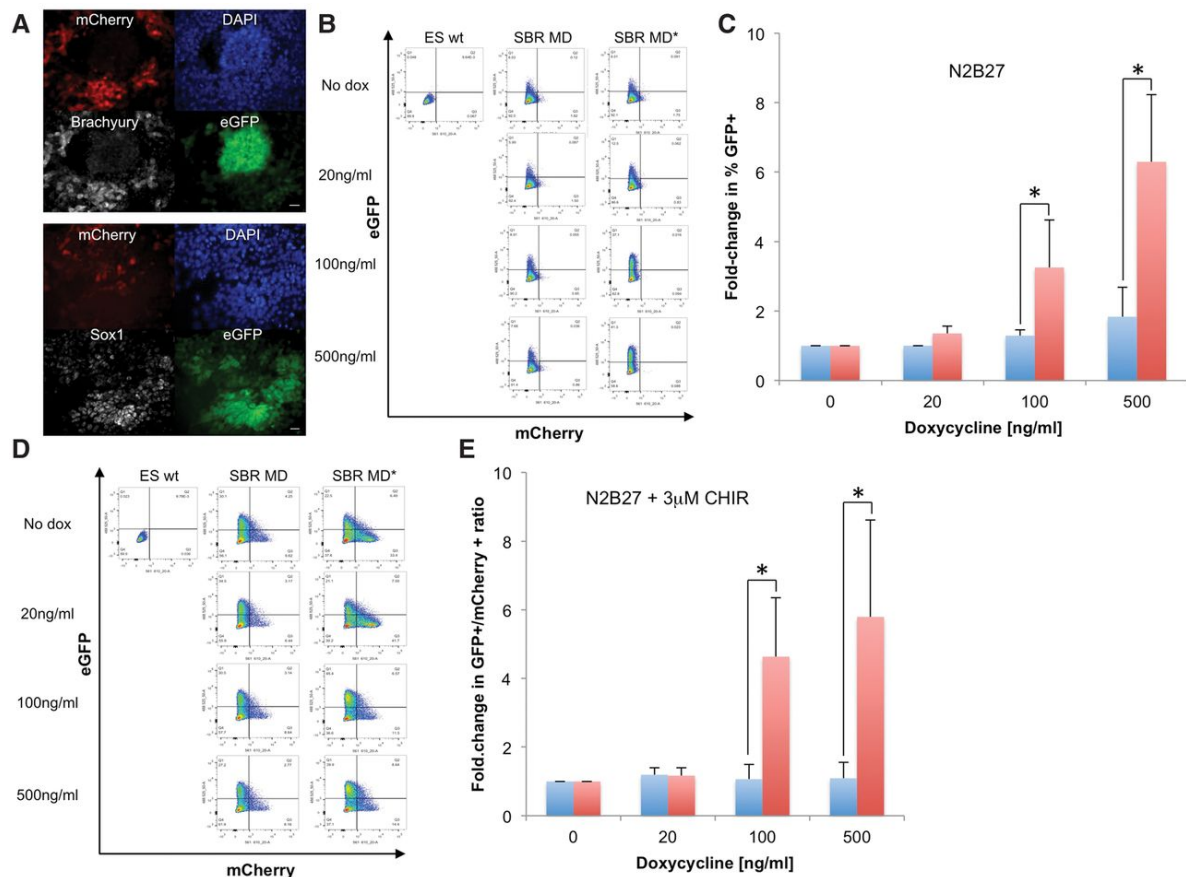


Figure 2.7 SOX2 is required at the M–G1 transition to promote neuroectodermal differentiation.

(A) Immunofluorescence analysis of BRACHYURY (top) and SOX1 (bottom) expression in the SBR cell line after 4 d of differentiation in N2B27. Bars, 20 μ m. (B) Flow cytometry of the SBR cell line expressing dox-inducible SNAP-MD-Sox2 (SBR MD) or SNAP-MD*-Sox2 (SBR MD*) after 3 d of differentiation in N2B27 with different doses of dox. (ES wt) Negative control. (C) Quantification of six independent flow cytometry experiments as shown in B, expressed as fold change in the fraction of eGFP+ cells and normalized over the values obtained without dox. (*) $P < 0.05$. (Blue bars) SNAP-MD-Sox2; (red bars) SNAP-MD*-Sox2. (D) Flow cytometry of the SBR cell line expressing dox-inducible SNAP-MD-Sox2 or SNAP-MD*-Sox2 after 4 d of differentiation with different doses of dox in N2B27 with the addition of 3 μ M CHIR after 48 h to induce mesendodermal differentiation. (ES wt) Negative control. (E) Quantification of seven independent flow cytometry experiments as shown in D, expressed as fold change in the eGFP+/mCherry+ ratio and normalized over the value of this ratio without dox. (Blue bars) SNAP-MD-Sox2; (red bars) SNAP-MD*-Sox2. (*) $P < 0.05$.

2.4 Discussion

Here we provide evidence of MCB of SOX2 and OCT4 and mitotic bookmarking by SOX2. Our results are in contrast to recent immunofluorescence-based experiments reporting that OCT4 does not bind to mitotic chromosomes (186). However, several studies reported unclear (160) or false negative results (155, 158, 187, 188) when using immunofluorescence to assess MCB, which was suggested to be due to limited accessibility of epitopes to antibodies within the dense mitotic environment (152, 155, 188) but also formaldehyde fixation artifacts (189, 190). Similar to previous reports on other transcription factors binding to mitotic chromosomes (189–191), we were unable to observe mitotic chromosome localization of SOX2 and OCT4 after formaldehyde fixation (data not shown). However, using a mixture of 95% methanol and 5% acetic acid previously shown to preserve the localization of transcription factors bound to mitotic chromosomes (189), we could observe MCB of endogenous SOX2 and OCT4 by immunofluorescence. Furthermore, here we used three unrelated tags to assess MCB of both overexpressed and endogenously expressed SOX2 and OCT4 in living cells and showed that their DNA-binding domains are necessary and sufficient for MCB. Interestingly, we observed a significant contribution of NLS sequences to the colocalization of SOX2 and OCT4 with mitotic chromosomes, in line with the reported sequence-independent electrostatic interactions of NLS with DNA (192, 193).

While mitotic bookmarking by sequence-specific transcription factors was reported >10 years ago (194), its biological function remains unclear. A prevailing idea in the field is that mitotic bookmarking could serve as a mechanism to maintain cell identity (154, 162, 177). This hypothesis was fueled by experimental evidence of mitotic bookmarking as a mechanism for rapid gene reactivation in G1 (158, 161, 177) and thus potentially contributing to robust maintenance of cell type-specific gene expression programs after cell division. However, recent studies have challenged the role of binding to specific DNA sites during mitosis for rapid transcriptional reactivation of targeted genes (160, 195), and, most importantly, the functional relevance of mitotic bookmarking in cell fate decision making has not been demonstrated so far. Our study reveals a key time window for the action of SOX2 on cell fate decisions, which temporally overlaps with its mitotic bookmarking activity and the re-establishment of transcriptional activity in newly divided cells (195). Interestingly, the functional importance of SOX2 at the M–G1 transition differed across the cell fate decisions that we studied. The impaired but not null pluripotency maintenance that we observed upon mitotic degradation of SOX2 could be due to incomplete SOX2 degradation at the M–G1 transition, but it is also likely that SOX2 expression in other cell cycle phases plays a role in pluripotency maintenance. Surprisingly, we did not observe any effect of mitotic SOX2 degradation on its ability to reprogram MEFs to iPS cells. This is unlikely to be due to differences in temporal pattern of SOX2 degradation in MEFs, since its increase in G1 was slower than in ES cells (Figure 2.6B), but we cannot exclude that the residual levels of SOX2 at the M–G1 transition were sufficient for its maximal potency in mediating reprogramming. The most striking impact on cell fate decision that we observed was the abrogation of the neuroectodermal differentiation-promoting activity of SOX2 when absent at the M–G1 transition. In ES cells, the decisions to commit to the mesendodermal versus neuroectodermal lineages take place in early and late G1, respectively (196), and are regulated notably by the cell cycle machinery (197). In our experiments with dox-inducible SOX2-MD, expression levels increase rapidly after cell division, and thus the absence of neuroectodermal fate enhancement is unlikely to result from insufficient expression at the end of the G1 phase but rather suggests that SOX2 acts during the M–G1 transition to enhance neuroectodermal commitment.

The M and G1 phases of the cell cycle were shown to be a window of opportunity for cellular differentiation and reprogramming (196, 198–200). However, the molecular mechanisms underlying the sensitivity of the M–G1 transition to the action of transcription factors on cell fate decisions remain unclear. Despite the clear enrichment of SOX2 on mitotic chromosomes observed by microscopy, we found a surprisingly small number of genes specifically bound during mitosis. In addition to the technical reasons invoked above, this might be due to mostly nonspecific DNA binding of SOX2 and/or DNA-independent binding of SOX2 to mitotic chromosomes. In contrast to GATA1 (177), RBPJ (201), or MLL (158) but similar to FOXA1 (160), we did not find enrichment of a particular class of genes bound by SOX2 during mitosis. The functional relevance of the sites specifically bound on mitotic chromosomes remains unclear, and there appears to be no quantitative difference in early G1 transcriptional activity of genes bound by FOXA1 or GATA1 in mitosis compared with genes

bound in interphase only (160, 195). Interestingly, nonspecific binding to mitotic chromosomes was also shown to regulate transcriptional reactivation in G1 (160). This raises the possibility that the high local concentration of mitotically bound transcription factors could mediate robust and fast occupancy of specific DNA sites in early G1, which could serve as a mechanism to ensure primacy against competing transcription factors in regulating target genes.

The role of mitotic bookmarking in stem cell regulatory networks remains largely unexplored. We found that OCT4 and potentially several other pluripotency transcription factors (Figure 2.1A) are also partially retained on mitotic chromosomes, suggesting that mitotic bookmarking could be widespread within the pluripotency network. The evidence that we present here on the functional role of mitotic bookmarking by SOX2 paves the way for future studies to identify and assess the functional relevance of other mitotic bookmarking factors in stem cell maintenance and differentiation.

2.5 Materials and methods

Live fluorescence microscopy

Live fluorescence imaging was performed at the biomolecular screening facility of the Swiss Federal Institute of Technology on an IN Cell Analyzer 2200 apparatus (GE healthcare) with controlled atmosphere (5% CO₂) and temperature (37°C) for long-term imaging and a 20× magnification objective. One day before imaging, cells were seeded in black-walled 96-well plates either uncoated (NIH-3T3 and 293T cells) or coated with E-Cadherin (R&D Systems, catalog no. 748-EC-050) as described (202), and transgene expression was induced with dox (100 ng/mL unless specified otherwise). To reduce background fluorescence, the cell culture medium was exchanged before imaging for Fluorobrite DMEM (Life Technologies) supplemented with the respective additives and dox concentrations for the different cell lines. Images were acquired for 16–24 h every 5 min using the following fluorescence channels: CFP channel for CerFP, YFP channel for YPet, TexasRed channel for TagRFP-T, Cy3 channel for the Halo-TMR dye, and Cy5 channel for the SNAP Sir-647 dye.

Immunofluorescence

E14 ES cells were fixed for 20 min with a chilled mix of 95% methanol/5% acetic acid that was maintained at –20°C, permeabilized with chilled PBS-Triton for 15 min, and blocked with PBS and 1% BSA for 30–60 min. Samples were incubated with primary antibodies rabbit anti-Sox2 (1:200; Life Technologies) and rabbit anti-Oct4 (1:500; Abcam, ab19857) in PBS and 1% BSA overnight at 4°C. Samples were washed twice in PBS and then incubated with Alexa 555 anti-rabbit (1:1000; Life Technologies, catalog no. 481400) in PBS and 1% BSA for 45–60 min followed by three washes with 0.1% PBS-Tween, incubation with 2 ng/mL DAPI, three washes with PBS and 0.1% Tween, and two washes with PBS.

MCB index calculation

We performed background subtraction on time-lapse data using Fiji software and manually measured mean fluorescence intensities of the SOX2 and OCT4 fusion proteins on metaphase chromatin divided by the intensity of a nearby area inside the cell.

Residence time measurements

To extract DNA residence times from raw data, we first extracted molecular localizations in each frame using Octane software. Peak detection thresholds were set individually for each sample set to account for variations in image background level. We identified molecule trajectories as the group of localizations in adjacent frames belonging to the same bound molecule using the following parameters: differences in position <160 nm (to avoid selection of moving nonbound molecules) and temporal distances less than two frames (to allow for missed localizations and fluorophore blinking). For each molecule trajectory, the length was assumed to be equal to the time that the molecule stays bound while unbleached. We fitted the distribution of trajectory lengths to an exponential curve to obtain an effective binding time for each gap time. Finally, we combined effective binding times

obtained from imaging with different gaps to disentangle bleaching rate from DNA off rates, allowing us to obtain DNA residence times. This analysis is described in more detail in Gebhardt et al. (2013).

FLIP and FRAP

FLIP and FRAP experiments were performed on a Zeiss LSM 700 confocal microscope equipped with a Plan-Apochromat 63×/1.40 oil DIC M27 objective (Zeiss) with controlled temperature and CO₂. Since a 488-nm laser (Zeiss) was used for imaging and bleaching of the sample regions, we used ES cell lines that allowed inducible YPet-SOX2 or YPet-OCT4 expression but were devoid of H2B-CerFP expression to avoid potential cross-talk between the two channels. Twenty-four hours before imaging, transgene expression was induced with 100 ng/mL dox. To reduce background fluorescence, the cell culture medium was exchanged before imaging for Fluorobrite DMEM (Life Technologies) supplemented as described above for ES cells and with dox.

FLIP In mitotic cells, we measured time traces of average fluorescence from a square region in the area containing condensed chromosomes while continuously bleaching a region in the cytoplasm. In interphase cells, both bleaching and measurement regions were inside the nucleus. For both mitotic and interphase cells, the size of the bleaching area was 4 μm², the size of the measurement area was 1 μm², and the distance between the bleaching and measurement regions was ~3 μm.

FRAP In mitotic cells, we pulse-bleached a round region in the area containing condensed chromosomes. We then measured time traces of average recovered fluorescence in the bleached region using 1.5% and 3% laser power for bleaching for Sox2 and Oct4, respectively. In interphase cells, the bleached and measured region was inside the nucleus. For both mitotic and interphase cells, the size of the bleaching and measurement area was 1.5 μm². Images were collected using a recording zoom of 2× at 0.3-sec intervals, including five prebleached images.

ChIP-seq analysis

Reads for each individual sample/replicate were mapped to the mouse genome (mm10 assembly) using the mapping module of HTSstation (based on Bowtie2) (203) with the option “discard PCR duplicates.” Triplicates were then combined to give a total of 1.53×10^8 mapped reads for the asynchronous samples, 8.3×10^7 mapped reads for the corresponding inputs, 1.2×10^8 mapped reads for the mitotic samples, and 1×10^8 mapped reads for the corresponding inputs. To avoid any potential bias due to unbalanced samples, ChIP samples were down-sampled preliminarily to the peak calling procedure. SOX2 peaks in the asynchronous and mitotic samples were identified with MACS2 (version 2.1) (204) from combined triplicates versus respective inputs. Since most reads fell into regions of non-specific binding in the mitotic samples, this made peak identification for this ChIP more difficult, and, similar to (160), we used a more lenient q-value threshold to define mitotic peaks. Therefore, a q-value threshold of 0.01 was used for the asynchronous samples, and a q-value threshold of 0.05 was used for the mitotic samples. Peaks overlapping any region found in the ENCODE blacklist (converted from mm9 to mm10 with LiftOver) (179) were excluded. BioScript (<http://bioscript.epfl.ch>) was used to annotate peaks to mm10 genes (operation Annotate) and quantify signals in regions surrounding peak summits (operation plotFeature). Motifs overrepresented in peaks were searched in sequences of 300 base pairs (centered on MACS2 summits), using MEME-ChIP (205) with default parameters and using asynchronous sequences as background.

Statistical tests

For Figure 2.3 (all panels), a two-tailed Student's t-test with unequal variance was used. For Figure 2.6, D–F, a two-tailed paired Student's t-test was used. For Figure 2.7, C and E, a Shapiro-Wilk test determined that the samples were nonnormally distributed; thus, a Wilcoxon-signed rank test was used to determine the statistical significance of the data.

Cell culture

The E14 cell line (kindly provided by Didier Trono, EPFL) was used for all ES cell experiments, except for the Luc-Oct4 knock-in and SBR cell and sub-cell lines that were generated from CGR8

ES cells (Sigma-Aldrich, Cat#07032901-1VL), and the 2TS22C cell line (obtained from Riken Bioresource Center, Ibaraki, 305-0074 Japan) that was used to generate corresponding sub-cell lines. Cells were cultured on dishes coated with 0.1% gelatin type B (Sigma), in GMEM (Sigma) supplemented with 10% ES-cell qualified fetal bovine serum, nonessential amino acids, L-glutamine, sodium pyruvate, 100µM 2 mercaptoethanol, penicillin and streptomycin, leukemia inhibitory factor, CHIR99021 at 3µM and PD184352 at 0.8µM. NIH-3T3 cells (kindly provided by Félix Naef, EPFL), and HEK 293T cells (ATCC) were cultured in DMEM supplemented with 10% fetal bovine serum, penicillin and streptomycin. Oct4-GFP MEFs (kindly provided by Matthias Lütolf, EPFL; (183) were cultured in DMEM supplemented with 10% fetal bovine serum, penicillin and streptomycin. For reprogramming experiments, MEFs were plated on matrigel coated 6-well plates at 2×10^5 cells per well 24 hours before transduction. Five days after transduction, cells from each well were trypsinized and replated on a 10cm cell culture dish coated with Matrigel.

DNA constructs

The Cre-recombinase (Cre) expression lentiviral vector was generated by replacing the CMV promoter and rtTA3G from pLenti CMV rtTA3 Hygro (Addgene plasmid #26730) with a PGK promoter upstream of the sequence encoding Cre recombinase. The plasmid encoding H2B-mCherry was obtained from Addgene (plasmid #21217). The vector for dox-inducible expression of each of the 16 pluripotency factors was generated as follow: A YPet sequence was PCR-amplified with primers encoding a 5' homology arm, a NdeI restriction site, a HA tag followed by two STOP codons, a LoxP site, a YPet fluorescent reporter and a 3'homology arm. This construct was inserted by infusion cloning (clontech) downstream of a gateway recombination cassette (GW). Subsequently, top and bottom oligonucleotides encoding loxP site and 2 more HA-tag sequences were annealed and ligated into the NdeI site. Finally, the cassette encoding GW-loxP-3xHA-2xSTOP-LoxP-Ypet was excised using restriction enzymes and transferred downstream of the TRE3G promoter in the pLentiCMVTRE3G Puro DEST lentiviral vector (Addgene plasmid # 27565). Pluripotency TFs were shuttled from a pENTR library generated in Deplancke lab, EPFL (206) into our final inducible vector by LR-Gateway recombination.

The H2B-CerFP construct was purchased from Addgene (plasmid#51006). The plasmid expressing rtTA3G was constructed by replacing tTR-KRAB and DsRed from pLV tTR-KRAB-Red (kindly provided by Didier Trono, EPFL), with rtTA3G that was PCR-amplified from Addgene plasmid # 26730, upstream of an internal ribosomal entry site and a blasticidin resistance. To construct the doxycycline-inducible expression lentiviral vectors, we first removed the Gateway cassette from pLenti CMVTRE3G Puro DEST (Addgene, plasmid#27565) and replaced it with a multicloning site by oligo annealing. We then inserted the different Halo-Tag, YPet, and TagRFP-T fusion proteins by restriction cloning. To generate the YPet-Sox2delHMG construct, 5' and 3' parts of the mouse Sox2 coding sequence (omitting amino acids 64 to 106 encoding the HMG DNA-binding domain) were PCR-amplified with primers flanked by restriction sites, ligated together, and cloned downstream and in frame to YPet. The following primers were used:

Sox2delHMG5'XbaIF: 5'-CACTCTAGAATGTATAACATGATGGAGACGGAG-3';

Sox2delHMG5'EcoRI: 5'-CAACTGAATTCCTGGGCCATCTTACGCC-3';

Sox2delHMG3'EcoRIF: 5'-CCAGAGAATTCCACCCGGATTATAAATACCGG-3';

Sox2delHMG3'AscIR: 5'-ACAGGCGCGCCTCACATGTGCGACAGGGG-3'.

To generate the YPet-HMG construct, the sequence of the Sox2 HMG domain was PCR-amplified with primers flanked by restriction sites, ligated together, and cloned downstream and in frame to YPet. We also generated a truncated YPet-HMG lacking the NLS by skipping the first three amino acids of the HMG domain. The following primers were used:

HMGXbaIF: 5' ACATCTAGAGTCAAGAGGCCCATGAACG-3';

HMGAscIR: 5' CACGGCGCGCCTTATTTATAATCCGGGTGCTCCT-3';

HMGNoNLSXbaIF: 5'-ACATCTAGACCCATGAACGCCTTCATG-3'.

To generate the YPet-Oct4delPouS construct, 5' and 3' parts of the mouse Oct4 coding sequence were PCR-amplified (omitting amino acids 131 to 205 encoding the PouS DNA binding domain) with

primers flanked by restriction sites, ligated together, and cloned downstream and in frame to YPet. The following primers were used:

Oct4delPH5'XbaIF: 5'-AAGTCTAGAATGGCTGGACACCTGGCT-3';

Oct4delP5'EcoRIR: 5'-GAGCAGAATTCCTGGGACTCCTCGGGAGT-3';

Oct4delP3'EcoRIF: 5'-AGACAGAATTCACAATGAGAACCTTCAGGAGATATG-3';

Oct4delPH3'AscIR: 5'-CCAGGCGCGCCTCAGTTTGAATGCATGGGAG-3'.

To generate the YPet-Oct4delPouH construct, the 5' and 3' parts of mouse Oct4 coding sequence were PCR-amplified (omitting amino acids 228 to 262 encoding the PouH DNA binding domain) with primers flanked by restriction sites, ligated together, and cloned downstream and in frame to YPet. We also generated a YPet-Oct4delPouH construct with an additional NLS by insertion of 15 nucleotides encoding the amino acid sequence RKRKR in place of the Pou_S domain sequence. The following primers were used:

Oct4delPH5'XbaIF: 5'-AAGTCTAGAATGGCTGGACACCTGGCT-3';

Oct4delH5'EcoRIR: 5'-ACATTGAATTCTCGCTTTCTCTTCCGGG-3';

Oct4delH3'EcoRIF: 5' ACATCGAATTCATTGAGTATTCCCAACGAGAAGAG-3';

Oct4delPH3'AscIR: 5'-CCAGGCGCGCCTCAGTTTGAATGCATGGGAG-3';

Oct4delH3'EcoRIF+NLS:

5'

ACATCGAATTCGGGAAGAGAAAGCGAATTGAGTATTCCCAACGAGAAGAG-3'.

To generate the YPet-PouS construct, the Pou_S domain was PCR-amplified with primers flanked by restriction sites, ligated together, and cloned downstream and in frame to YPet. The following primers were used:

PouSXbaIF: 5' GGATCTAGAGACATGAAAGCCCTGCAGA-3';

PouSAscIR: 5' ACAGGCGCGCCTTAGTCGGCTTCCTCCACC-3'.

The lentiviral vectors encoding SOX2 fusions to YPet or to a mitotic degron (MD) or point mutant thereof (MD*) were generated by PCR amplification with primers flanked by restriction sites, ligated together, and cloned downstream of either TRE3G promoter using pLV-TRE3G-MCS, or downstream of the PGK promoter of pLV tTR-KRAB-Red after excision of tTR-KRAB and DsRed. The MD coding sequence was synthesized by GeneArt and encodes an amino acid sequence identical to the one reported (177). The MD* coding sequence was generated by site-directed mutagenesis on the MD sequence to mutate the arginine at position 42 into an alanine (R42A), allowing to inactivate the MD as previously reported (177).

The lentiviral vectors allowing expression of Oct4, Klf4, cMyc and either Sox2-YPet-MD, Sox2-YPet-MD* were constructed by excising Sox2 from pFuW-TetO-OSKM (addgene plasmid # 20321), and ligating PCR-amplified Sox2-YPet-MD and Sox2-YPet-MD*.

All constructs were verified by Sanger sequencing.

Lentiviral vector production and generation of stable cell lines

Lentiviral vector production was carried out by calcium phosphate transfection of HEK 293T cells with the envelope (PAX2) and packaging (MD2G) constructs together with the lentiviral vector of interest, and concentrated 120-fold by ultracentrifugation as described previously (207), except for the initial screen of 16 pluripotency TFs, for which lentiviral particles were generated using XtremeGENE 9 DNA Transfection Reagent (Roche). Target E14 cells, NIH3T3 cells or HEK 293T cells were then seeded at 50,000 cells per well of a 24-well plate (E14 cells) or of a 6-well plate (NIH-3T3 and HEK 293T cells) and transduced with 50µl of concentrated lentiviral vector. Selection of transduced cells was performed by addition of the respective antibiotics 48-72 hours after transduction, and the antibiotics were maintained in the cell culture medium throughout passaging. For blasticidin selection, we used 8µg/ml (E14 cells) or 5µg/ml (NIH-3T3 and HEK 293T cells); for puromycin selection, we used 2µg/ml (all cell lines). For the initial screen for mitotic chromosome binding of 16 pluripotency factors, we first generated a cell line constitutively expressing H2B-mCherry and Cre recombinase. This cell line was then used to generate 16 sub-cell lines allowing dox-inducible expression of each of 16 transcription factors, by transducing 4000 cells in 96-well

plates twice (on two consecutive days) with non-concentrated lentiviral vector particles. Constitutive expression of Cre allowed excising a cassette containing 3 HA repeats and each transcription factor coding sequence to be in frame with a C-terminal YPet tag (Methods). Selection with puromycin was started 48hrs after the first transduction.

Confocal microscopy

The acquisition of confocal images (Supplementary figure 2.1) was performed on a Leica TCS SP8 confocal laser scanning microscope, with a 63x oil-immersion objective, using a 405nm laser to excite CerFP and a 514nm laser to excite YPet, and the pinhole diameter was set to obtain slices < 1µm thick in both channels.

Live luminescence microscopy

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) for 24-48 hours. One day before the experiment, cells were seeded on fluorodishes (WPI, FD35-100) coated with E-Cadherin as described (202). The next day, the medium was exchanged and supplemented with 1mM luciferin (NanoLight Technology, Cat#306A). Images were acquired every 90s with a 30s exposure time.

Single Molecule Microscopy

Single molecule imaging was performed on an inverted TIRF microscope (Axio Observer.D1, Zeiss) with a Plan-Apochromat, 100x, numerical aperture =1.46, oil objective (Zeiss). The microscope was equipped with a piezo objective scanner (N-725, Physik Instrumente), a motorized stage (MFC-2000, Applied Scientific Instrumentation) and with a live-cell environmental chamber providing constant 37°C temperature and 5% CO₂ concentration (Live Imaging Services). Samples were excited using a 561 nm laser (CrystaLaser) with irradiance of 1 kW/cm². To decrease out of focus fluorescence we performed imaging using highly inclined illumination (HILO) conditions (174). Fluorescent light was collected by the objective and directed through a dichroic mirror (89100BS, Chroma) and an emission filter (ET630/75m, Chroma) to an EMCCD camera (iXon 897+, Andor) with 160nm pixel size. Samples were incubated one hour before imaging with 5nM Halo-TMR substrate (Promega) for 30 minutes, washed and incubated again without substrate, and then the medium was changed to DMEM without phenol red to reduce background fluorescence, supplemented with 50ng/ml of doxycycline. To allow accurate measurements of long DNA residence times, we used a time-lapse illumination strategy that we described before (173). Briefly, this consists of directing the excitation laser light through an acousto-optic tunable filter (AOTF) device (AA Optoelectronics) acting as a shutter controlled by the LabView software (National Instruments), to introduce a variable dark time, or gap between images. We typically acquired between 200- 1000 image frames (2s exposure time) per cell for a given gap duration, with gaps between frames ranging between 0-8s (208).

Number of binding events/total fluorescence analysis

The bound fraction (BF) – that is, the ratio of bound versus free TF molecules – contains valuable biological information about the association and dissociation rates of the TF to chromatin. In fact, $BF \stackrel{\text{def}}{=} \frac{k_{on}}{k_{on} + k_{off}}$ where k_{on} and k_{off} are the pseudo-association and dissociation rates, respectively (209). The pseudo on-rate k_{on} is the only experimentally accessible parameter linked to the actual on-rate k_{on} by the expression $k_{on} = k_{on}[Seq]$, where Seq represents the concentration of free binding sites at the thermodynamic equilibrium. Since $KD = k_{off}/k_{on}$ is the pseudodissociation equilibrium constant of the TF, then $BF = \frac{1}{1 + KD}$. Also, for a given binding reaction where the TF binds DNA, at the equilibrium $KD = \frac{[TF] \cdot [DNA]}{[TF-DNA]}$, where brackets refer to the concentration of each species. In particular, $\frac{[TF]}{[TF-DNA]}$ is the ratio of unbound over specifically-bound TF molecules. Assuming that $[TF-DNA]$ is negligible as compared to the total amount of molecules $[TF]$ that are freely diffusing and non-specifically bound, we have $[TF] \gg [TF-DNA]$. Under this assumption, KD will scale inversely with the number of binding events n divided by the total fluorescence F of a given

image i and $BF \propto n/F$. We then estimated BF by computing the ratio n/F for a given image. Binding events (n) were localized running the Fiji plugin Octane and counted by visual inspection per each considered frame. The peak intensity threshold was set individually for each sample to account for variations in the image background level. Total nuclear fluorescence F was calculated by manual thresholding of the nucleus (interphase cells) or the whole cell (mitotic cells) using the Fiji software and corrected for the background noise, B , Irradiance and exposure time were kept constant for all samples as well as bleaching conditions to achieve single-molecule regime (~1 minute at maximum laser power). For each sample, we estimated the bound fraction (BF) by calculating the ratio of the number of individual binding molecules to total nuclear fluorescence in arbitrary units (a.u.). Ratios are averaged over 10 consecutive time points (frames 110-119; 2 minutes and 40 s after imaging started), so that the estimated bound fraction was given by: $BF = \frac{1}{10} \cdot \sum_{i=1}^{10} \frac{n_i}{F_i - B_i}$

Ratiometric BF's were calculated out of 20-21 different interphase or mitotic cells either expressing Sox2 or Oct4. For comparing distributions of BF values, 25th-75th percentile box-plots were used. To appreciate additional information about the spread of the data, whiskers were drawn from the upper and lower hinges of the box to the maximum and minimum values. Differences in data distributions between mitotic and interphase cells were scored by calculating the P-value through a Mann-Whitney non-parametric statistical test, assuming that distributions did not differ each other as null-hypothesis.

Mitotic cell sorting

E14 ES cells were synchronized using 200 ng/ml Nocodazole for 12 hours, trypsinized, fixed with 1% formaldehyde for 10 minutes and quenched with 200 mM Tris-HCl pH 8.0. Cells were permeabilized with cold 70% EtOH for 15 minutes, stained with 1:1000 anti-H3S10p antibody CMA312 (EMD Millipore) for 1 hour at 4°C, and subsequently stained with 1:80 anti-mouse IgG1 PE antibody (eBioscience) for 1 hour at 4°C. PE-positive cells were sorted, spun down, and stored at -80°C. For the DAPI control staining, sorted cells were stained with 1mg/ml DAPI for 15 minutes, imaged on a Zeiss Axioplan, and cells with condensed chromatin were counted.

ChIP-seq experiments

Asynchronous E14 ES cells (11×10^6 , 13×10^6 , and 14×10^6 cells per replicate) were fixed with 1% formaldehyde for 10 minutes at room temperature and quenched with 200 mM Tris-HCl pH 8.0, washed with PBS, spun down, and stored at -80°C. The cell pellet was resuspended and in 1.5 ml LB1 (50 mM HEPES-KOH pH 7.4, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10% Glycerol, 0.5% NP40, 0.25% Triton X-100), incubated 10 min at 4°C, spun down, and resuspended in 1.5 ml LB2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) incubated 10 min at 4°C. The pellet was spun down, rinsed twice with SDS shearing buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.15% SDS), and finally resuspended in 0.9 ml SDS shearing buffer. All buffers contain contain 1:100 diluted Protease Inhibitor Cocktail in DMSO (Sigma). Mitotically sorted cells were directly resuspended in SDS shearing buffer. The suspension was transferred to a milliTUBE 1 ml AFA fiber and sonicated on a E220 focused ultrasonicator (Covaris) using the following settings: 20 min, 200 cycles, 5% duty, 140W, and input sample aliquots were taken. To pull down SOX2-bound DNA, sonicated chromatin was incubated with 500 mg anti-SOX2 antibody Y-17 (Santa Cruz) per 1×10^6 cells overnight at 4°C. 2.5 ml Protein G Dynabeads (Thermo Fischer) per 1×10^6 cells were added to the chromatin and incubated 3 hours at 4°C. The chromatin was washed several times at 4°C with 5 min incubation between each wash and 2 min magnetization to collect beads; twice with Low Salt Wash Buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.15% SDS, 1 mM PMSF), once with High Salt Wash Buffer (10 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.15% SDS, 1 mM PMSF), once with LiCl Wash Buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 250 mM LiCl, 1% NP40, 1% sodium deoxycholate, 1 mM PMSF), and finally with TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM PMSF). Beads were finally resuspended in Elution buffer (TE buffer with 1% SDS and 150 mM NaCl), treated with 400 ng/ml Proteinase K and reverse crosslinked at 65°C 1100 rpm overnight. Input samples were treated with 100 mg/ml RNase A and 400 ng/ml Proteinase K and reverse crosslinked at 65°C 1100 rpm overnight. Sonicated chromatin samples were purified using Qiagen MinElute PCR purification kit,

and sequencing libraries were prepared using NEBNext ChIP-seq Library Prep Master Mix Set (NEB), either without size selection (replicate one) or with size selection to obtain average size chromatin fragments of ~400 bp (replicates two and three). Libraries were sequenced using 100 bp single end reads on an Illumina HiSeq 2500.

Western Blotting after Sox2 ChIP

9×10^6 asynchronous ES cells were fixed and 8.5×10^6 mitotic E14 cells were fixed, stained, and sorted as described in “ChIP-seq experiments”. Cells were lysed and sonicated, and SOX2 was pulled down as described in “ChIP-seq experiments” using 25 μ g antibody. The negative control sample contained SDS shearing buffer and 25 μ g anti-SOX2 antibody and was treated as samples during SOX2 IP. After washing with TE buffer, proteins were eluted from the Protein G beads by addition of 1X Laemmli buffer (62.5 mM Tris-HCl, pH 6.8, 10% glycerol, 1% LDS, 0.0045% Bromophenol blue, 10% beta-mercaptoethanol; BioRad) and incubation at 95°C for 5 min. 1/10 of each sample was loaded on a 4-20% Tris-glycine gel (BioRad) together with a PageRuler Plus protein ladder (Thermo). After gel separation, proteins were transferred to a MeOH-activated PVDF membrane (Immobilon) using semi-dry transfer. The membrane was blocked with 5% milk in PBS with 0.05% Tween-20 (Milk/PBS-T) for 30 minutes at RT, followed by incubation with Rabbit anti-SOX2 antibody (Cat. 48-14800; Thermo) in Milk/PBS-T at 1:250 overnight at 4°C. The next day, the membrane was washed once with PBS-T and then incubated with anti-Rabbit HRP Conjugate (Promega) at 1:10'000 for 45 minutes at room temperature. The membrane was washed five times in PBS-T, developed with Clarity Western ECL Substrate (BioRad), and detected on a Witec Fusion imaging system.

Generation of Sox2-Luc and Luc-Oct4 knock-in cell lines

The E14 Sox2-Luc and CGR8 Luc-Oct4 cell lines were generated using CRISPR-Cas9 mediated homology directed repair (HDR), as previously described (210). The repair templates for HDR were designed to contain a knock-in cassette flanked by 1kb homology arms with the target sequence. For the Luc-Oct4 knock-in, the repair template construct contained the 5' UTR of Oct4 upstream of a cassette encoding a blasticidin-TagRFP-T fusion protein followed by a F2A peptide allowing to separate the polypeptide chain encoding blasticidin-TagRFP-T from the downstream sequence by ribosomal skipping. Firefly luciferase was inserted downstream and in frame to this cassette and fused to the 5' part of the Oct4 coding sequence. For the Sox2-Luc knock-in, the repair template construct consisted of the 3' part of the Sox2 coding sequence fused to firefly luciferase, upstream of a cassette encoding a P2A peptide allowing ribosomal skipping followed by a fusion between blasticidin resistance and sfGFP. The guide RNAs were designed by using the Zhang Lab toolbox (www.genome-engineering.org/crispr) and identified guides were subsequently cloned into a pX330 backbone (Cong et al. 2013) to target the Start codon of Oct4 (guide sequence: 5'-GCA GGT GTC CAG CCA TGG GGA-3') or the Stop Codon of Sox2 (guide sequence: 5'-GCA GCC CTC ACA TGT GCG ACA-3'), respectively. Next, 2.4×10^6 ES cells were co-transfected with pX330 and the corresponding repair template in a 1:3 ratio (20 μ g total DNA), grown for two days before selection with blasticidin (10 μ g/ml) was started. 10 days later, colonies were picked manually and allowed to grow out. Genomic DNA was extracted from the candidate cell lines using a Genomic DNA extraction kit (Sigma). The knock-ins were verified by PCR of the target loci, using primers listed in Table 2.1.

Generation of the Sox1/eGFP Bra/mCherry (SBR) knock-in reporter cell line

Knock-in to express fluorescent proteins (FPs) from the Sox1 and Bra loci was performed with homology directed repair (HDR) as described previously (210). The HDR templates were designed to contain a knock-in cassette flanked by homology arms (HAs) spanning 1 kb surrounding the stop codons of Sox1 and Bra without the stop codon. The HDR templates were designed as: 5'-HA-P2A-FP-STOP-loxP-selection cassette-loxP-3'-HA. For Sox1, the FP was eGFP and the selection cassette PGK-Hygromycin resistance. For Bra, the FP was mCherry and the selection cassette PGK-Puromycin resistance-sfGFP to enable monitoring of loss of sfGFP expression following Cre recombination. Guide RNAs targeting the Sox1 and Bra loci were designed using the Zhang Lab toolbox (www.genome-engineering.org/crispr) and cloned into the pX335 vector expressing Cas9D10A (nickase) and the guide RNA. For pX335-Sox1, the guide RNA binds in a region spanning

the STOP codon, which is removed in the HDR template (guide sequence: 5'-GAC GCA CAT CTA GCG CCG CG). For pX335-Bra, the guide RNA binds after the STOP codon (guide sequence: 5'-GTG CTG AGA CTT GTA ACA AC-3'). To avoid cuts of the HDR template by Cas9, one nucleotide mismatch was incorporated in the untranslated 3'HA.

CGR8 mouse ES cells (Sigma Aldrich) were transfected with pX335-Sox1 and the Sox1 HDR template at a 1:3 ratio. After two days, selection was started with 200 ug/ml Hygromycin B. After 14 days of selection, the pool of cells was transfected with pX335-Bra and the Bra HDR template at a 1:3 ratio. After 2 days, selection was started with 2 ug/ml Puromycin. After 13 days of selection, cells were sorted for sfGFP expression and transfected with pLVEF1a-Cre constitutively expressing Cre recombinase. After 4 days, single sfGFP negative cells were sorted into wells of 96-well plates. Single colony outgrowths were split into four conditions, without selection, selection with either Hygromycin or Puromycin, and into N2B27 medium for differentiation. A clone that was sensitive to both antibiotics and expressed eGFP and mCherry upon differentiation was picked and verified using PCR of DNA extracted using a Genomic DNA extraction kit (Sigma) using the primers listed in Table 2.2.

Reverse transcription followed by Quantitative PCR (RT-QPCR)

Total RNA was extracted using a GenElute™ Mammalian Total RNA Miniprep Kit Q-PCR (Sigma-Aldrich), and reverse transcription was performed using an oligoDT primer using superscript II (Life Technologies). QPCR was performed on a 7900HT Fast Real-Time PCR System (Thermofischer) with SYBR green reagent (Roche). Rps9 cDNA was used for data normalization. Primers used for RT-QPCR are listed in Table 2.3.

ChIP-QPCR

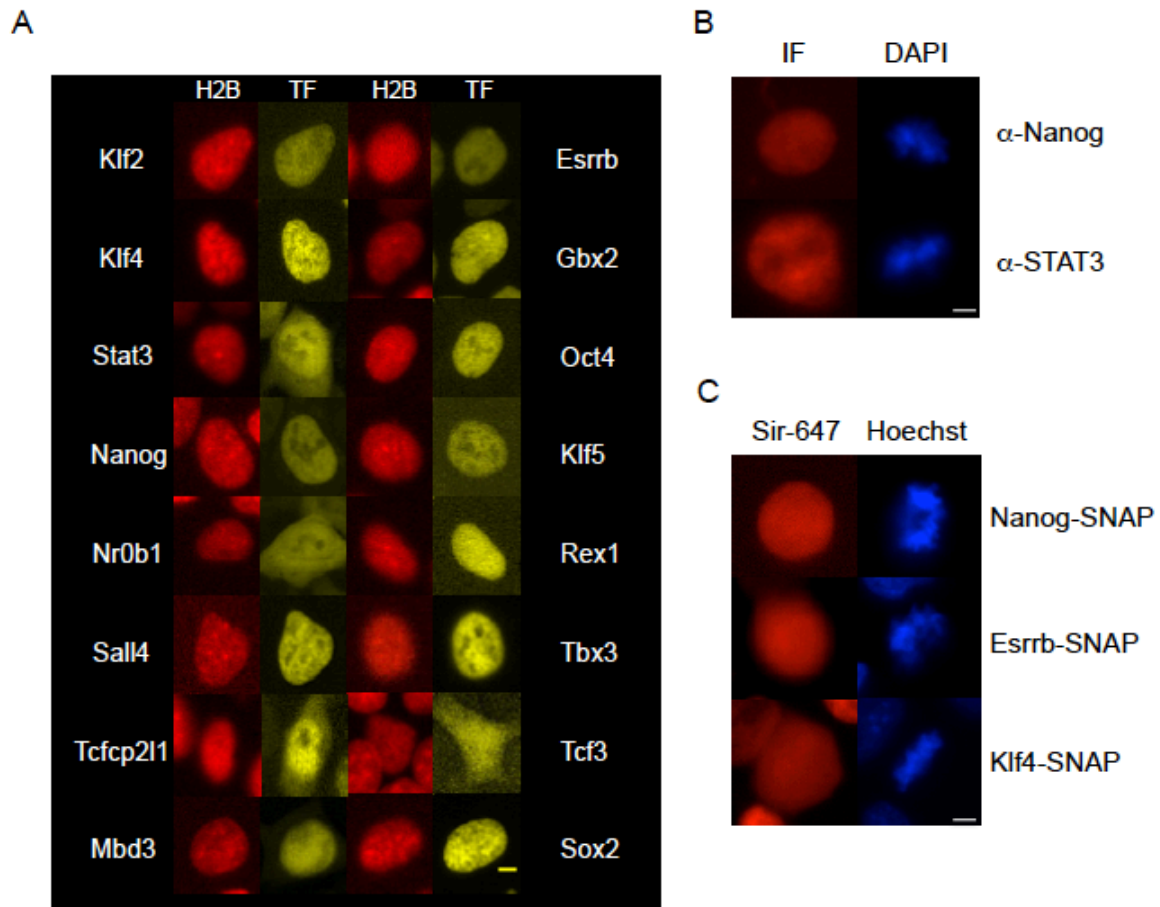
ChIP-QPCR was performed on library preparations of ChIP and input samples, using the primers listed in Table 2.4, on a 7900HT Fast Real-Time PCR System (Thermofischer) with SYBR green reagent (Roche).

Sorting of YPet-tagged and SNAP-tagged Sox2 cell lines

We developed two different strategies to sort for similar expression levels between MD and MD* cell lines. In the first strategy, we aimed at obtaining similar expression levels for the MD and MD* cell lines once expression levels have recovered from mitotic degradation. For convenient testing of this strategy, we used the dox-inducible cell lines, which allow expression levels that are high enough to be reliably measured by live cell imaging. We observed that sorting for the same, large fluorescence window (Supplementary figure 2.11a-b) of the TRE3G-Sox-YPet-MD and TRE3G-Sox2-YPet-MD* cell line, allowed obtaining cell populations which the median fluorescence intensity was slightly different between MD and MD* cell lines (about 1.4 fold higher for the TRE3G-Sox2-YPet-MD* as compared to the TRE3G-Sox2-YPet-MD cell line, Supplementary figure 2.11a-b). We then measured YPet fluorescence intensity from time-lapse series in single cells, 5 frames before observable chromatin condensation, and observed very similar average intensities (Supplementary figure 2.11i). Thus, even though these two cell lines have different median fluorescence intensities over the whole cell cycle, they have similar expression levels once the effect of mitotic degradation has disappeared. Conversely, we reasoned that sorting for a much narrower cell population should yield similar median fluorescence intensities, which should then be at the expense of slightly higher expression levels in the MD cell lines in cell cycle stages at which expression levels have recovered from mitotic degradation. We thus sorted a much narrower fluorescence window of the TRE3G-SNAP-MD-Sox2 and TRE3G-SNAP-MD*-Sox2 cell lines, which resulted in virtually identical median fluorescence intensities (Supplementary figure 2.11c-d). We then performed measurements of SNAP-SiR647 fluorescence intensities in cells 5 frames before observable chromatin condensation, and observed about 1.4-fold higher fluorescence intensities in the TRE3G-SNAP-MD-Sox2 than the TRE3G-SNAP-MD*-Sox2 cell line (Supplementary figure 2.11j). In that case, the two cell lines have similar median fluorescence intensities, but the SNAP-MD fluorescence intensity is higher when the effect of the MD has disappeared, thus compensating for the lower levels at the end of the M phase and in early G1. For sorting of the 2TS22C cell lines expressing the four different tagged versions of Sox2

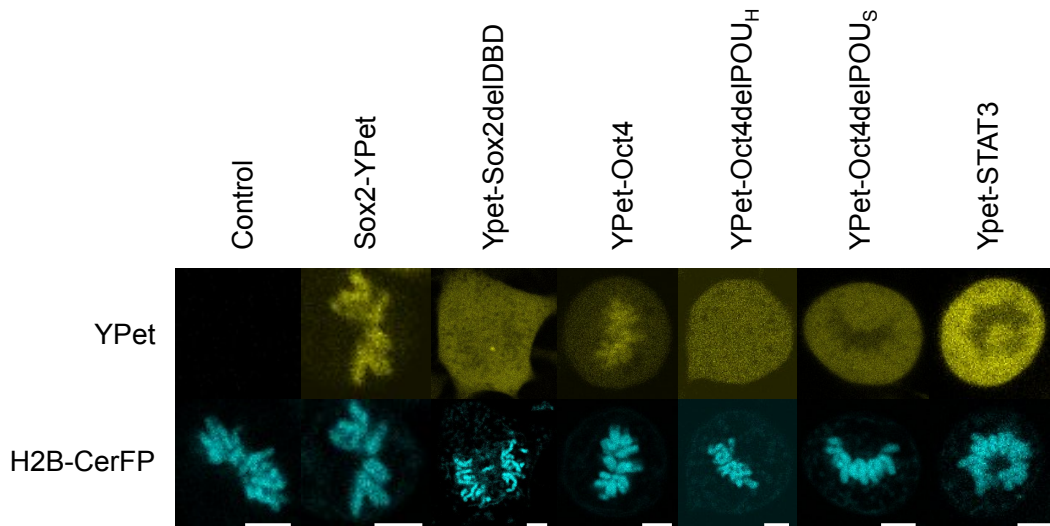
under the control of the PGK promoter, we used the “narrow window strategy” for the Sox2-YPet fusion (Supplementary figure 2.11).

2.6 Supplementary figures



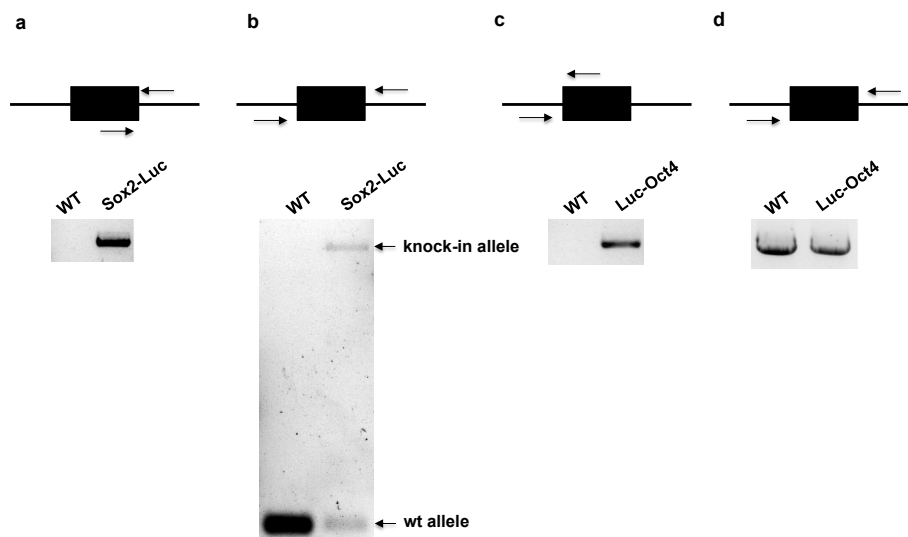
Supplementary Figure 2.1 Interphase localization of YPet-fused transcription factors and additional control immunofluorescence and SNAP-tag fusion imaging, related to Figure 2.1.

(A) Interphase localization of pluripotency transcription factors (TF) fused to Ypet. (B) Metaphase cells staining with a Nanog or STAT3 antibody. (C) Snapshots of metaphase cells of dox-inducible Esrrb-SNAP, Nanog-SNAP and Klf4-SNAP ES cell lines labeled with SNAP SiR-647 and Hoechst. Scale bars: 5 μ m. IF: immunofluorescence channel.



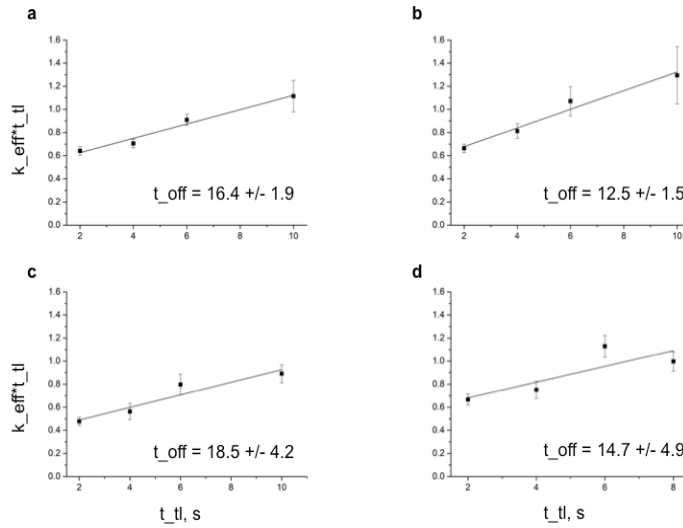
Supplementary Figure 2.2 Confocal imaging of mitotic chromosome binding in ES cells, related to Figure 2.1 and Figure 2.2.

ES cell lines expressing H2BCerFP and different doxycycline-inducible YPet fusion proteins were plated on E-cadherin coated dishes in medium containing 100ng/ml of doxycycline and imaged by confocal microscopy 24 hours later. Scale bars: 5 μ m



Supplementary Figure 2.3 Genomic analysis of Sox2-luciferase and Luciferase-Oct4 knock-in ES cells, related to Figure 2.1.

PCR analysis of the genomic insertion sites for the knock-in cassettes of Sox2-Luc (a-b) and Luc-Oct4 (c-d). The black boxes represent the knock-in cassette. The black line represents flanking genomic regions of the expected insertion sites for Sox2 and Oct4. Note that for Luc-Oct4, we were unable to amplify the whole knock-in cassette (expected product size: 5.2kb), thus only the PCR product for the wild type (wt) allele can be seen.



Supplementary Figure 2.4 DNA residence time extraction from single molecule imaging experiments, related to Figure 2.4.

The effective residence time (t_{eff}) that we measured in single molecule tracking experiments is a function of transcription factor residence time (t_{off}) and fluorophore bleaching (t'_{bleach}):

$$k_{eff} = k_{off} + k_{bleach};$$

$$\frac{1}{t_{eff}} = \frac{1}{t_{off}} + \frac{1}{t'_{bleach}};$$

As we vary the dark time, or gaps between images (t_{gap}), fluorophore bleaching (t'_{bleach}) and the effective residence time (t_{eff}) vary with a rate that depends on the time-lapse parameters (t_{int} - image acquisition time):

$$t_{tl} = t_{int} + t_{gap}, t_{int} = 2 \text{ s (const)};$$

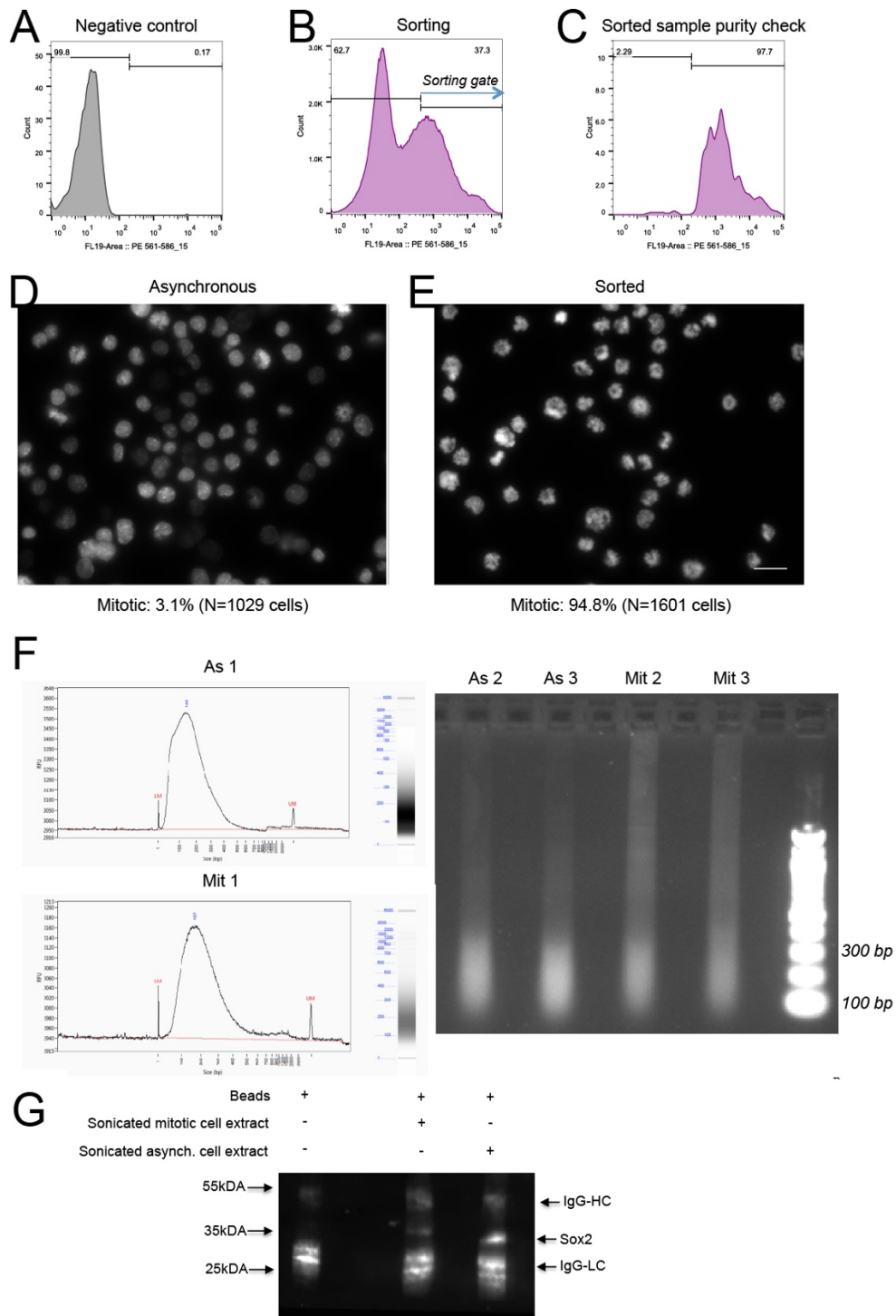
$$t'_{bleach} = \frac{t_{tl}}{t_{int}} t_{bleach}, t_{bleach} = (\text{const});$$

$$\frac{1}{t_{eff}} = \frac{1}{t_{off}} + \frac{t_{int}}{t_{tl}} \frac{1}{t_{bleach}};$$

Measuring the effective binding time for different gap conditions thus allows us to extract the true binding time t_{off} as a linear regression coefficient according to the equation below.

$$k_{eff} t_{tl} = \frac{1}{t_{off}} t_{tl} + \frac{1}{t_{bleach}} t_{int};$$

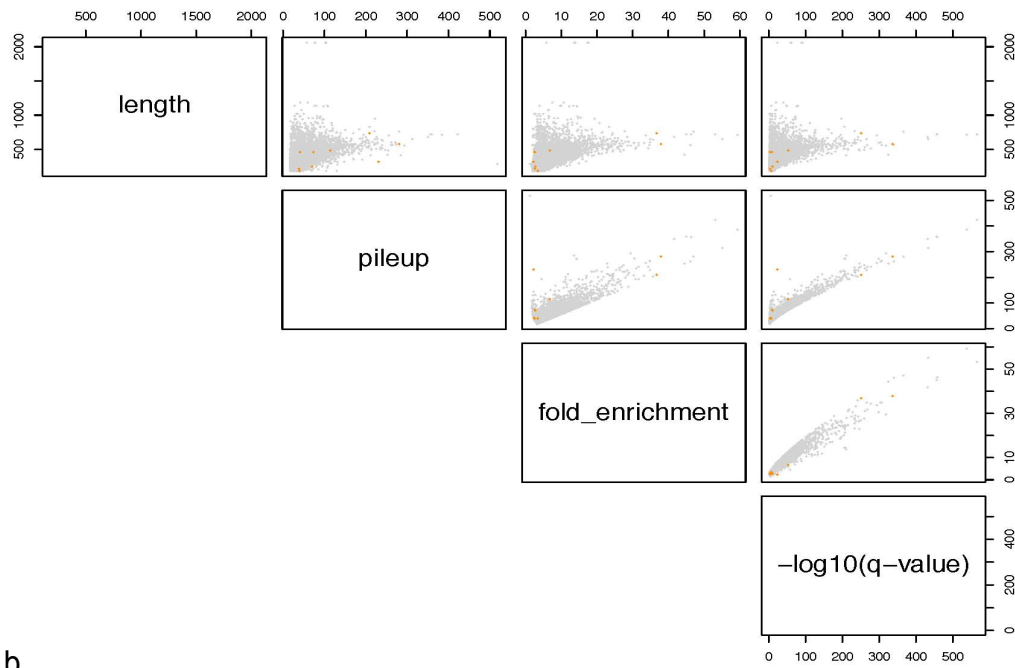
as shown here for (a) Sox2 in interphase; (b) Sox2 in M-phase; (c) Oct4 in interphase; (d) Oct4 in M-phase. $N \geq 3$ for each t_{tl} condition.



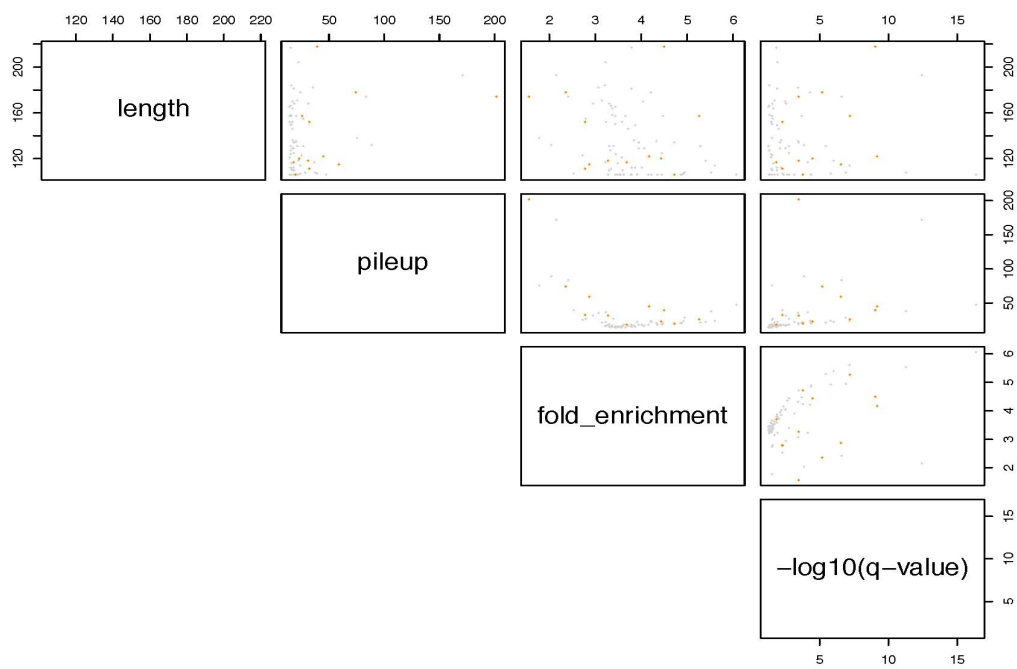
Supplementary Figure 2.5 Quality control of mitotic cell sorting, related to Figure 2.5.

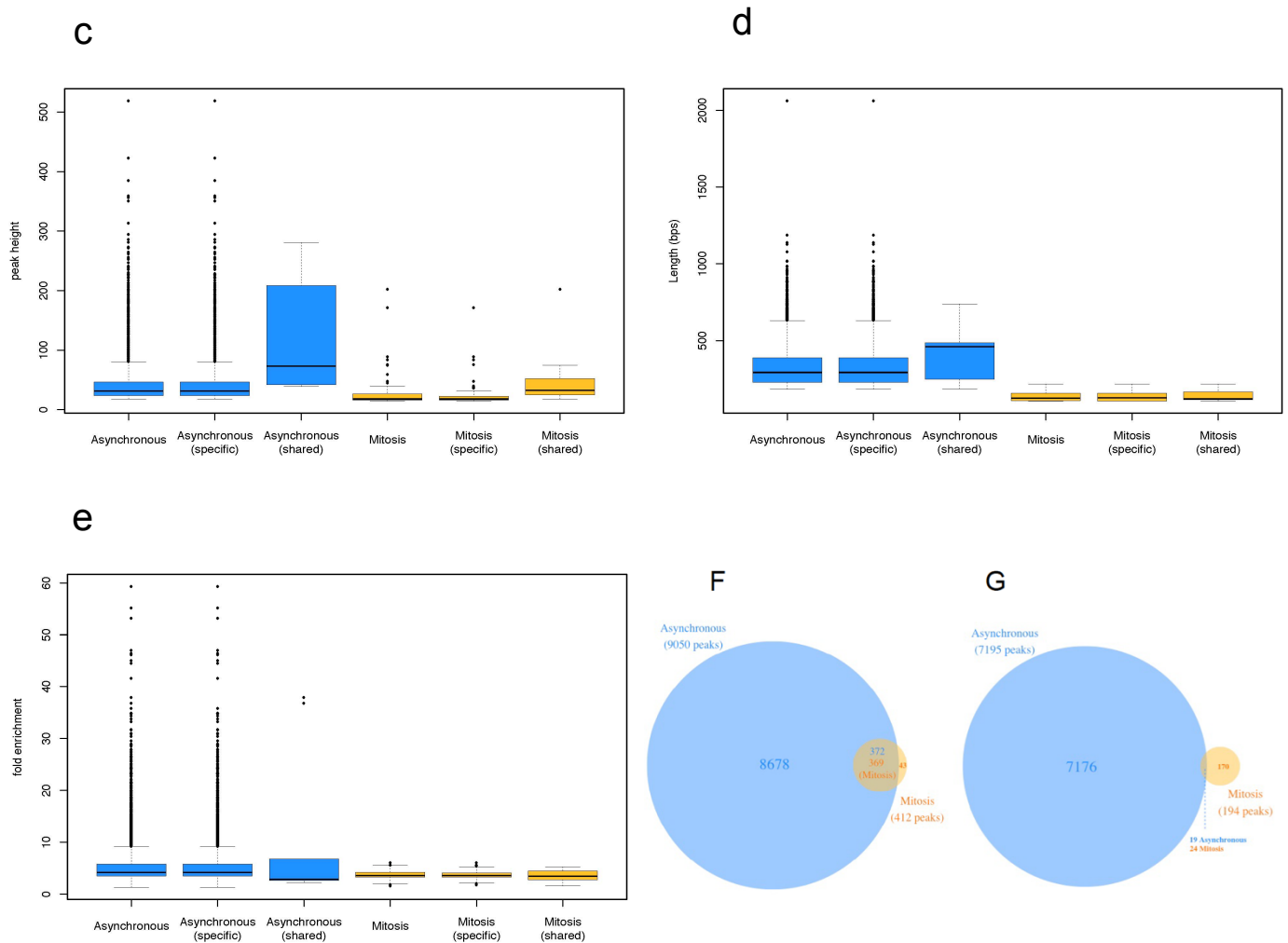
(A) Negative control cells stained with secondary antibody alone. (B) Sorting window for H3S10P-positive cells after nocodazole synchronization. (C) Reanalysis of sorted sample to determine the purity of H3S10P-positive cells (97.7%). (D-E) DAPI staining of asynchronous (D) and H3S10P-sorted cells (E) to quantify the fraction of mitotic cells with condensed chromatin. Scale bar: 20 μ m. (F) Sonication profiles of fragmented chromatin from asynchronous (As) and sorted mitotic (Mit) cells used for downstream ChIP-seq experiments. The first replicate for each condition was analyzed on a Fragment Analyzer and the two remaining replicates on a 1% agarose gel. (G) Western blotting against Sox2 after boiling antibody-bound beads incubated with sonicated extracts from mitotic or asynchronous cells. Antibody-bound beads were used as negative control. LC: light chains; HC: heavy chains.

a



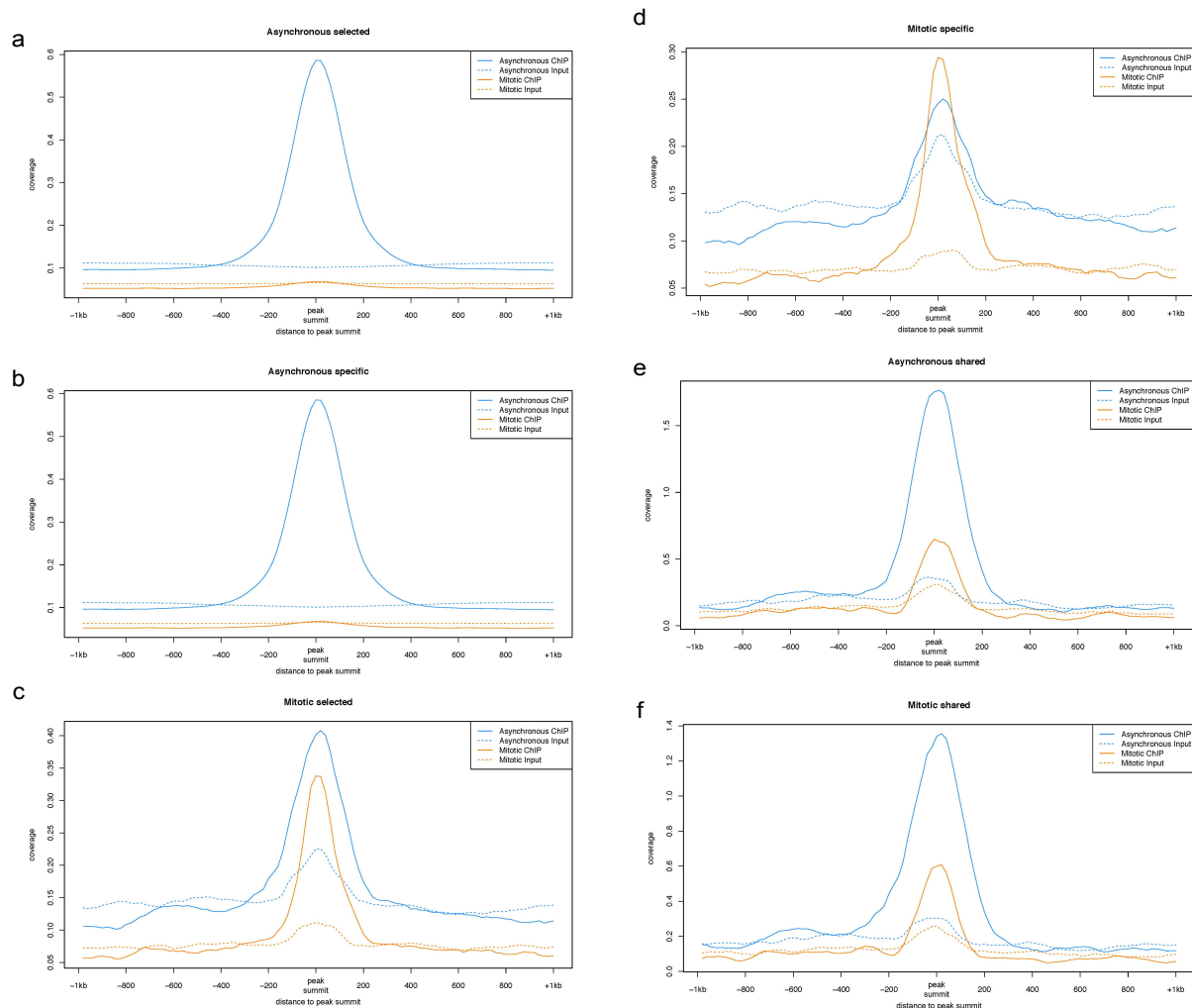
b





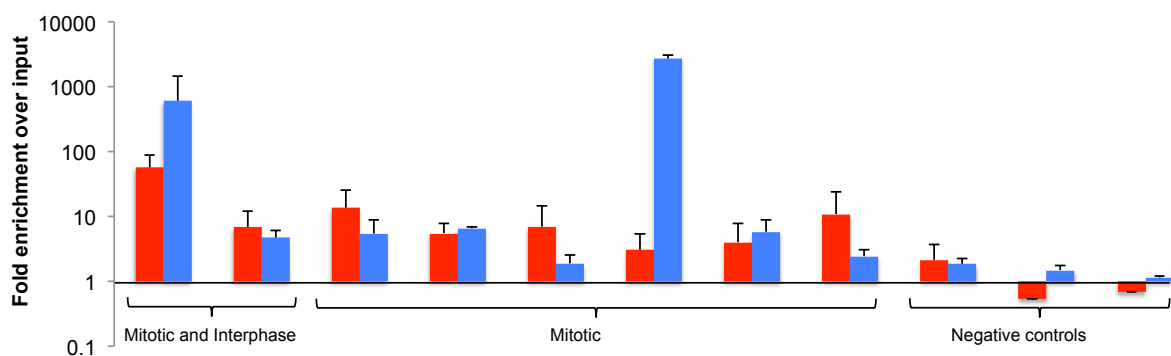
Supplementary Figure 2.6 Peak calling and selection, related to Figure 2.5

(a-b) Peak characteristics of asynchronous (a) and mitotic (b) ChIP-seq samples. Gray: selected peaks; Orange: blacklisted peaks. (c-e) 25th-75th percentile box-plots of peak height (c), peak lengths (d), and peak fold-enrichment (e) of different classes of peaks as indicated on the x-axis. Whiskers: ± 1.5 IQR; black line: median. (f-g) Number of asynchronous and mitotic peaks detected for FoxA1 (Caravaca et al. 2013) (f) and GATA1 (Kadauke et al.) (g) using our analytical pipeline.



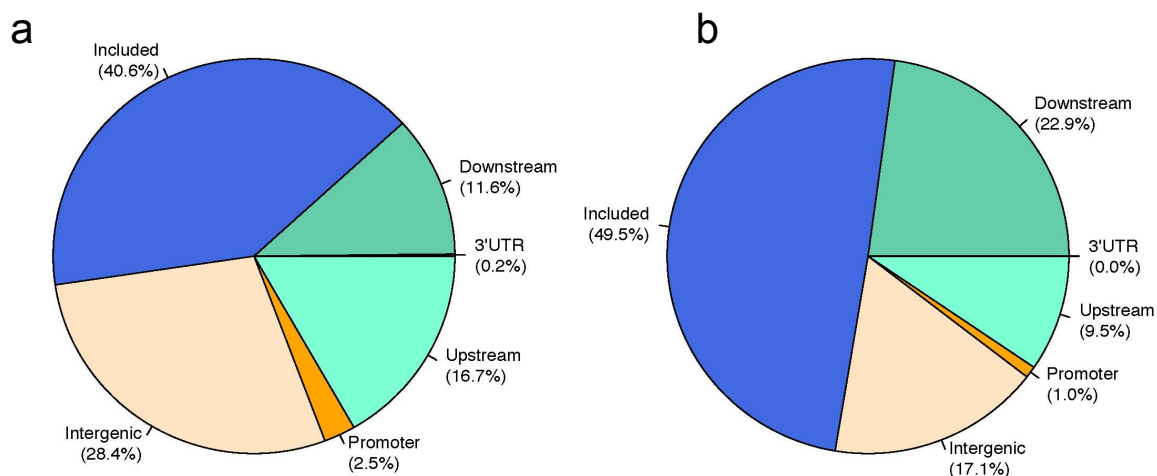
Supplementary Figure 2.7 Relative enrichment of called peaks, related to Figure 2.5.

(a) Asynchronous peaks. (b) Asynchronous peaks not overlapping with mitotic peaks. (c) Mitotic peaks. (d) Mitotic peaks not overlapping with asynchronous peaks. (e) Asynchronous peaks also present in the mitotic samples. (f) Mitotic peaks also present in the asynchronous samples. Blue: Asynchronous samples. Orange: Mitotic samples. Solid lines indicate ChIP samples and dashed lines input samples.



Supplementary Figure 2.8 ChIP-QPCR on selected peaks and negative control regions, related to Figure 2.5.

QPCR primers were designed around the center of peaks called by MACS2, for 2 peaks common to mitotic and asynchronous samples, 6 peaks called only in mitosis, and 3 regions where no peaks were called. Red: Asynchronous ChIP, N=2. Blue: Mitotic ChIP, N=2. Error bars: SE.



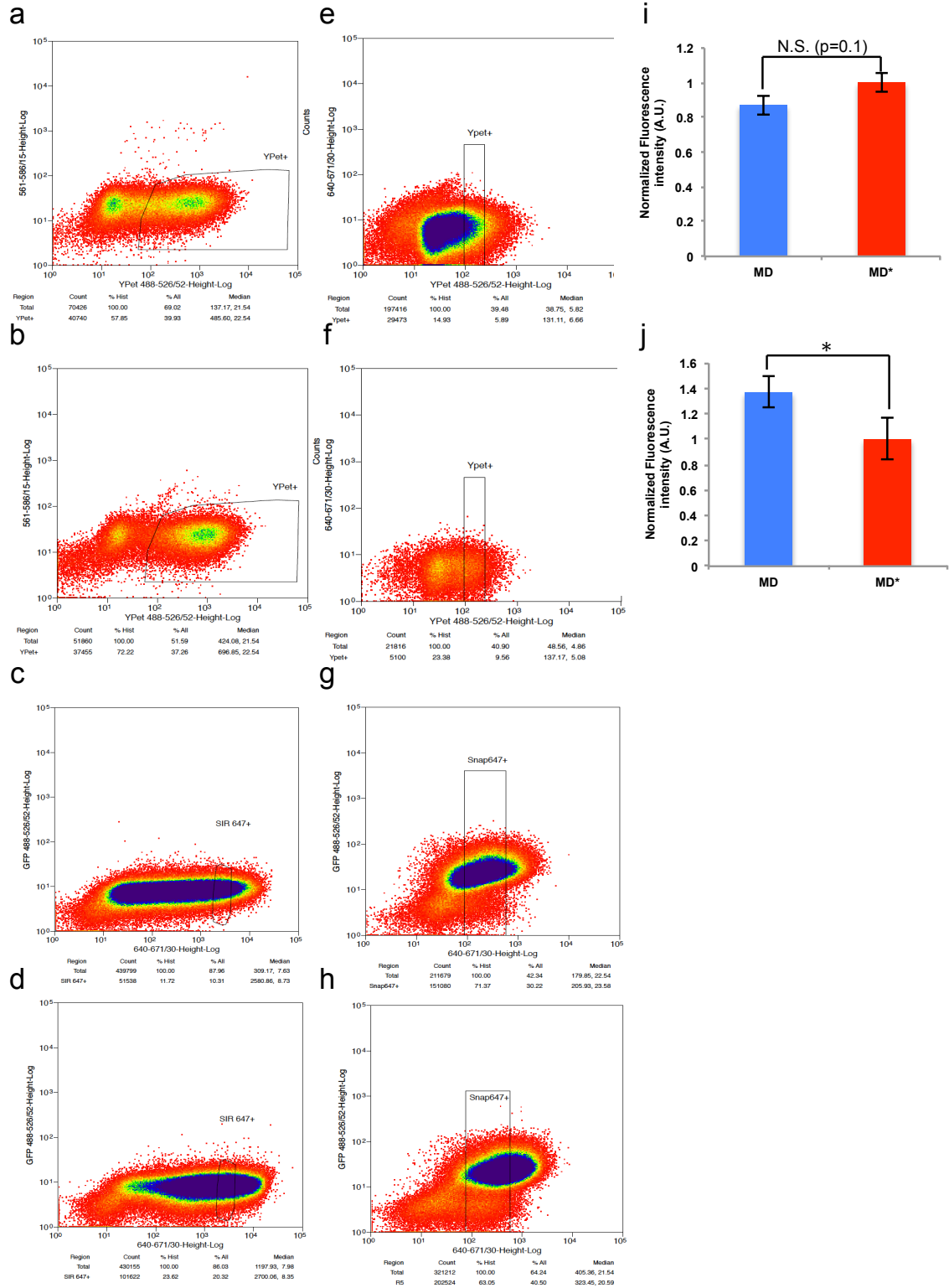
Supplementary Figure 2.9 Genome-wide distribution of ChIP-seq peaks, related to Figure 2.5.

(a) Asynchronous samples. (b) Mitotic samples. Intergenic: regions without genes within a distance of 20kb; Promoter regions: upstream and within 2kb of the gene start; Upstream: between 2kb and 20kb upstream of transcription start sites; Included: within genes; 3'UTR: downstream and within 10% of the distance to the next downstream gene; Downstream: downstream of 3'UTR but at a distance of < 20kb to the closest downstream gene.



Supplementary Figure 2.10 De novo motif identification with MEME in the asynchronous sample, related to Figure 2.5.

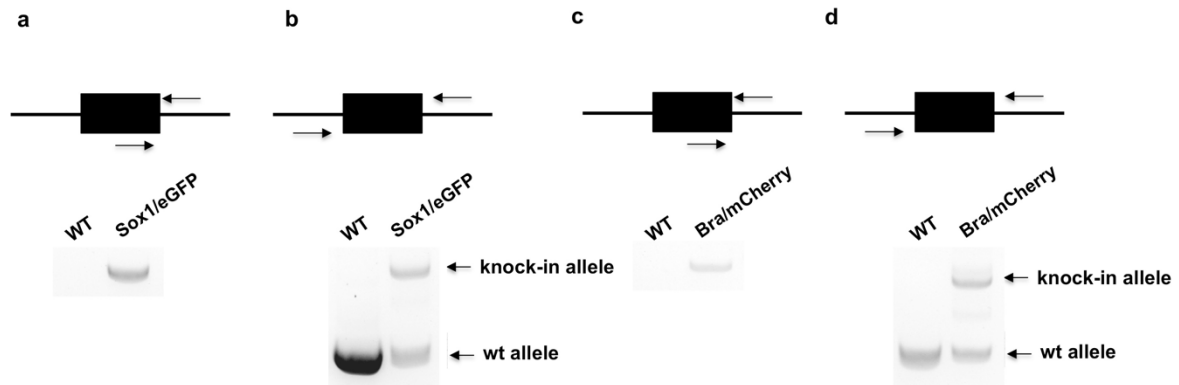
Top-scoring motif in the asynchronous samples, centered on peaks, matching the known Oct4::Sox2 composite binding motif (e-value = 4.1×10^{-343}).



Supplementary Figure 2.11 Sorting strategies for the Sox2 overexpressing cell lines, related to Figure 2.6 and Figure 2.7.

(a-h) Sorting windows of the different cell lines by FACS. (a) TRE3G-Sox2-YPet-MD (b) TRE3G-Sox2-YPet-MD*. (c) TRE3G-SNAP-MD-Sox2. (d) TRE3G-SNAP-MD*-Sox2. (e) PGK-Sox2-YPet-MD. (f) PGK-Sox2-YPet-MD*. (g) PGK-SNAP-MD-Sox2. (h) PGK-SNAP-MD*-Sox2. (i-j): Integrated fluorescence intensity of cells was measured from time-lapse experiments, 5 frames = 25 minutes before the condensation of chromosomes became visible, using cells that were FACS-sorted using the fluorescence windows shown in (a), (b), (c) and (d). Intensities were normalized on the MD* data. Statistical analysis was performed using student's two-tailed t-test with unequal variance. (i) TRE3G-Sox2-YPet-MD (MD; N= 105) and TRE3G-Sox2-

YPet-MD* (MD*; N=109). (j) TRE3G-SNAP-MD-Sox2 (MD; N=100) and TRE3G-SNAP-MD*-Sox2 (MD*; N=102). *:p<0.05



Supplementary Figure 2.12 Genomic analysis of Sox1-P2A-eGFP / Brachyury-P2A-mCherry knock-in ES cells, related to Figure 2.7.

PCR analysis of the genomic insertion sites for the knock-in cassettes of Sox1-P2A-eGFP (Sox1/eGFP) (a-b) and Brachyury-P2A-mCherry (Bra/mCherry) (c-d) in the SBR cell line. The black boxes represent the knock-in cassette with homology arms to the targeted region. The black line represents flanking genomic regions of the expected insertion sites for Sox1 and Bra.

2.7 Tables

Sequence (5' to 3')	Target sequence position
ACCTCCGGGACATGATCA	In Sox2 5' HA (sense)
CGTTTCGCTGCGGAGATT	In Sox2 3' HA (antisense)
CGGTAAGACACTGGGTGTGA	In Firefly Luciferase (sense)
GCATGCTAGCCACAAAGAAA	Downstream of Sox2 3'HA (antisense)
GAGGATGAACACCGGAGTCC	Upstream of Oct4 5' HA (sense)
CTAATTGCTTACACTTGCTCC AGA	Downstream of Oct4 3'HA (antisense)
CTTGTACAGCTCGTCCATGC	In sfGFP (antisense)

Table 2.1: Primers used to verify genomic integration of the Sox2-Fluc and Fluc-Oct4 knock-in cassettes. HA: homology arms.

Sequence (5' to 3')	Target sequence position
AAAAACGTTGTGAGCAAGGGCGAGGAG	eGFP start
GTGCCCCTGACGCACAT	Inside Sox1 5' HA
AGTATGCAAAAGGCAGAACCAC	Downstream of Sox1 3' HA
CATGGATCCATGGTGAGCAAGGGCGAG	mCherry start
GGGCTGCTACAGTCACAGAC	Inside Bra 5' HA
TGCACATTCTCCAAGAAGGA	Downstream of Bra 3' HA

Table 2.2: Primers used to verify genomic integration of the Sox1-P2A-eGFP and Brachyury-P2A-mCherry knock-in cassettes. HA: homology arms.

Sequence (5' to 3')	Target sequence
GCTCCTCAGCATGACTACTACT	Dlx3 cDNA (forward)
CTGCGAGCCCATTGAGATTG	Dlx3 cDNA (reverse)
GAAAACGCGGAATCTGAAACAA	Esx1 cDNA (forward)

TCCTGGTAGTTTGCTTCTTCCA	Esx1 cDNA (reverse)
TTGTGCGCAAAACCTATGTGACC	Rps9 cDNA (forward)
GCCGCCTTACGGATCTTGG	Rps9 cDNA (reverse)

Table 2.3: Primers used for RT-QPCR

Sequence (5' to 3')	Target sequence
CAATACCGCCCATTTCCAGG	Rif1_F
TGGCAGCACATTTCTTAGCG	Rif1_R
GGTGGAGAGTGCTGTCTAGG	Oct4_F
CTGCCCAGAACTCTCAACCT	Oct4_R
GCTGAGTCTAATTGAGGGCC	NF-κB1F
ACACAGTACTGACAGGGCTC	NF-κB1R
GGAGGAGGAGAACAAGGTCC	Nanog_F
TGAGAGAACACAGTCCGCAT	Nanog_R
GGCACCACCCAATCGAAAG	Nrarp_F
ACAATAGACGCTCTTGACGC	Nrarp_R
GAAAGGGGCTGCATGACAAA	Cnpy1_F
CGTGTGAAAGGGGACAAAGG	Cnpy1_R
CAATGAGGTTTCGTGGTCTGC	Ube2r2_F
GTGAACTTCCAGACCTGTGC	Ube2r2_R
AAAGGCAGAACAAAGGAGCC	Mitotic/Interphase peak1_F
CCCGAGGTAGCTTGAGATCA	Mitotic/Interphase peak1_R
GCAAAAGTGTGGCAATCCTAC	Mitotic/Interphase peak2_F
TGGAAACCCAGCCTGTGTG	Mitotic/Interphase peak2_R
GCTAGTCTTGCCAGGTCCTA	Mitotic peak1_F
TAAAGGAGAGTGGGGTCTGG	Mitotic peak1_R
CTTTCATCGGAGCAGCGAAG	Mitotic peak2_F
GTGAGGGAGGACAGACAGAG	Mitotic peak2_F
CTTTCATCTCCCGTCCTGT	Mitotic peak3_F
CAGTGGTCCAGAAGTGACGA	Mitotic peak3_R
GCTCGGCTAAGAAAGGGC	Mitotic peak4_F
CCCGCCCAATCAGCTGTA	Mitotic peak4_R
TAGAACACAGGGCTTGGTGT	Mitotic peak5_F
GCTCACACAGGTTCCCATTC	Mitotic peak5_R
ACCAGTCAGTCCAGCTAGAG	Mitotic peak6_F
CTGACTCTGCTGCCTACCAT	Mitotic peak6_R
GTGGTATAGCTGGGTCCTCA	Negative Region1_F
ACACACCAATGGACGAGTGT	Negative Region1_R
CACATGGACCTTTCAAGCTGT	Negative Region2_F
GACAGGAGATGGACCATGGG	Negative Region2_R
TTGAAATCCGTTGGTCCCAA	Negative Region3_F
GTTCTGTCTCCTCAGCAT	Negative Region3_R

Table 2.4: Primers used for ChIP-QCPR

2.8 Acknowledgements

We thank Andrea Alber for providing the SNAP-tagged Nanog, Esrrb, and Klf4 ES cell lines. We thank Luigi Bozzo, Thierry Laroche, and José Artacho from the Swiss Federal Institute of Technology (EPFL) Bioimaging and Optics Core Facility (EPFL-BIOP); Valérie Glutz, Loïc Tauzin, Miguel Garcia, and André Mozes at the EPFL Flow Cytometry Core Facility (EPFL-FCCF); and Fabien Kuttler at the EPFL Biomolecular Screening Facilities (EPFL-BSF) for assistance in imaging. We thank Marion Leleu from the Bioinformatics and Biostatistics Core Facility (EPFL-BBCF) for ChIP-seq data analysis, Michael Imbeault for advice on ChIP-seq experiments, Massimiliano Caiazzo and Matthias Lütolf for providing Oct4-GFP MEFs and for advice on reprogramming, and Didier Trono and Pierre Gönczy for insightful comments on the manuscript. Work from the laboratory of D.M.S. was supported by the Carigest Foundation, the Swiss National Science Foundation (grant no. PP00P3_144828), and the Novartis Foundation for Biomedical Research (grant no. 15A018). Work from the laboratory of S.M. was supported by the Swiss National Science Foundation (grant no. CR33I2_149850).

Chapter 3 Endogenous fluctuations of OCT4 and SOX2 bias pluripotent cell fate decisions

This chapter represents the following publication which has been formatted to fit with the style of the thesis:

Strebinger, D.*, Deluz, C.*, **Friman, E.T.***, Govindan, S., Alber, A.B. and Suter, D.M. (2019) Endogenous fluctuations of OCT4 and SOX2 bias pluripotent cell fate decisions. *Molecular Systems Biology*, 15(9)

Contribution: In this collaborative study, I expanded the previously generated SBR cell line to harbor fusion tags for SOX2 (SOX2-SNAP, homozygous) and OCT4 (OCT4-HALO, heterozygous) and performed part of the validation of the cell line (called SBROS). The main part of my contribution is the bioinformatics analysis of ATAC-seq data, for which I wrote the corresponding section. I was also heavily involved in editing the manuscript, performing follow-up analysis and experiments for revisions, and responding to reviewers.

Note that most tables and source data have been omitted from this version of the manuscript. These can be found on the publisher's website.

3.1 Abstract

SOX2 and OCT4 are pioneer transcription factors playing a key role in embryonic stem (ES) cell self-renewal and differentiation. How temporal fluctuations in their expression levels bias lineage commitment is unknown. Here we generated knock-in reporter fusion ES cell lines allowing to monitor endogenous SOX2 and OCT4 protein fluctuations in living cells and to determine their impact on mesendodermal and neuroectodermal commitment. We found that small differences in SOX2 and OCT4 levels impact cell fate commitment in G1 but not in S phase. Elevated SOX2 levels modestly increased neuroectodermal commitment and decreased mesendodermal commitment upon directed differentiation. In contrast, elevated OCT4 levels strongly biased ES cells towards both neuroectodermal and mesendodermal fates in undirected differentiation. Using ATAC-seq on ES cells gated for different endogenous SOX2 and OCT4 levels, we found that high OCT4 levels increased chromatin accessibility at differentiation-associated enhancers. This suggests that small endogenous fluctuations of pioneer transcription factors can bias cell fate decisions by concentration-dependent priming of differentiation-associated enhancers.

3.2 Introduction

Embryonic stem (ES) cells can be maintained in a self-renewing state in vitro or be driven to commit to specific fates when exposed to differentiation signals. However, ES cells often exhibit asynchrony and divergences in fate commitment when subjected to the same differentiation cues. This obscures the interpretation of how instructive signals impact cell fate decisions, and limits the generation of pure ES cell-derived cell populations for future regenerative medicine applications (211). Heterogeneity in cell fate commitment points at the coexistence of different cellular states, but these remain largely uncharacterized at the molecular level. Intercellular variability in expression levels of cell fate regulators constitutes a potential source of variable cellular states. One well-studied example is the heterogeneity of NANOG expression in serum + LIF culture conditions, which reflects reversible transitions of ES cells between the naïve and primed states (212). However, protein expression variability of cell fate regulators in ES cells maintained in a more homogeneous, naïve state is poorly explored.

The transcription factors SOX2 and OCT4 (also known as POU5F1) are expressed in ES cells and are strictly required to maintain their pluripotent state (171, 213–218). Recent studies reported that quantitative properties of SOX2 and OCT4 such as expression levels or DNA-binding properties are predictive of cell fate decisions at the four-cell stage (219, 220). Differences in SOX2 concentrations were also shown to change its enhancer binding profile (221), however whether this translates into

differences in cell fate commitment is unknown. SOX2 and OCT4 were also reported to play antagonistic roles in the differentiation of ES cells towards the neuroectodermal (NE) and mesendodermal (ME) fates (170, 171). However, these conclusions were largely based on long-term overexpression/knockdown or indirect correlations from fixed cells. A recent study in human ES cells has shown that OCT4 levels are predictive for the cell fate choice between pluripotent self-renewal and extra-embryonic mesoderm commitment (222). Nevertheless, how endogenous expression levels of SOX2 and OCT4 fluctuate over time in naïve ES cells and whether these fluctuations bias germ layer cell fate commitment remains unknown.

Here we investigated how endogenous variability in SOX2 and OCT4 protein levels impact the ability of ES cells to differentiate towards the NE or ME fates. To do so, we generated knock-in mouse ES cell lines allowing to monitor endogenous fluctuations of SOX2 and OCT4 proteins in live cells. We found that both proteins fluctuate over a 2 to 3-fold range with timescales of approximately one cell cycle in naïve mouse ES cells. Endogenous expression levels of OCT4 and to a lesser extent of SOX2 at the onset of differentiation impact the ability of pluripotent cells to differentiate towards NE and ME upon undirected differentiation. Using ATAC-seq on cells with different endogenous OCT4 levels, we show that OCT4 fluctuations are associated to changes of chromatin accessibility of enhancers involved in cell differentiation.

3.3 Results

3.3.1 Generation of a sox2-SNAP / oct4-halo knock-in ES cell line

We first aimed at generating a cell line allowing monitoring of SOX2 and OCT4 protein levels in single living cells. To do so, we used CRISPR-Cas9 genome editing (210, 223) to knock-in fluorescent tags in fusion to the C-terminus of endogenous SOX2 and OCT4 proteins. We generated a cell line in which both alleles of SOX2 are fused to a C-terminal SNAP tag, and one allele of OCT4 is fused to a C-terminal HALO tag (Figure 3.1A, Supplementary Figure 3.1A-C). Since SNAP and HALO tags allow orthogonal labelling with fluorescent dyes, endogenous SOX2 and OCT4 levels can be measured independently in individual living cells (Figure 3.1B). These knock-ins were generated in a previously established reporter cell line for ME and NE commitment (SBR cell line (224)), thus allowing to monitor SOX2 and OCT4 protein levels in live cells and to track differentiation outcomes. The resulting cell line was named SBROS for Sox1/Brachyury/Reporter/OCT4-HALO/SOX2-SNAP. Because of the heterozygosity of the OCT4-HALO knock-in, we assessed whether OCT4-HALO levels are a good proxy for total OCT4 levels at the single cell level. To do so, we combined immunofluorescence against total OCT4 with direct labelling of OCT4-HALO using the HaloTag TMR dye. We found that total OCT4 and OCT4-HALO levels were well correlated in single cells, suggesting that OCT4-HALO levels allow estimating total OCT4 levels (Figure 3.1C and Supplementary Figure 3.1D). We then verified whether this cell line expresses normal levels of pluripotency markers by quantitative immunofluorescence analysis. In contrast to Western blotting, immunofluorescence is not biased by differences in membrane transfer efficiency of proteins with different molecular weights (such as OCT4 and OCT4-HALO) and allows obtaining distributions of protein levels in the cell population. We found that the SBROS cell line expressed on average 89% of wild type mean OCT4 levels, 111% of wild type mean SOX2 levels, and 129% of wild type mean NANOG levels. The distributions and median expression levels of these proteins were similar to those of wild type E14 ES cells (Figure 3.1D). We also found the half-lives of both OCT4-HALO (7.8 ± 1.3 h) and SOX2-SNAP (8.1 ± 1.0 h) to be close to published half-life values for these proteins (225–227) (Figure 3.1E), and the average cell cycle length of SBROS and wild type E14 ES cells were similar (Figure 3.1F). mRNA levels of pluripotency markers of SBROS cells were mostly unaltered (Supplementary Figure 3.1E). We also verified cloning efficiency and alkaline phosphatase activity after one week of clonal growth, and found it to be comparable between SBROS and its parental SBR cell line (Supplementary Figure 3.1F).

While the normal pluripotent phenotype of SBROS cells implies that the homozygous Sox2-SNAP alleles are functional, this is not necessarily the case for the heterozygous Oct4-Halo allele. To verify the functionality of the OCT4-HALO fusion protein, we drove its expression from the constitutive EF1- α promoter in Zhbtc4 cells, which allow doxycycline (dox)-inducible OCT4 knockout. After 24 hours of dox induction, endogenous OCT4 expression is lost (217) and thus cells rely only on the

OCT4-HALO protein to maintain their pluripotent state. In these conditions, we found that the OCT4-HALO protein was able to fully rescue pluripotency, thus confirming its functionality (Supplementary Figure 3.1G). We also found the interaction of the OCT4-HALO protein with SOX2 to be preserved by co-immunoprecipitation experiments in the SBROS cell line (Supplementary Figure 3.1H), and the ChIP-seq profile of OCT4-HALO to be enriched at WT OCT4 peaks (Supplementary Figure 3.1I).

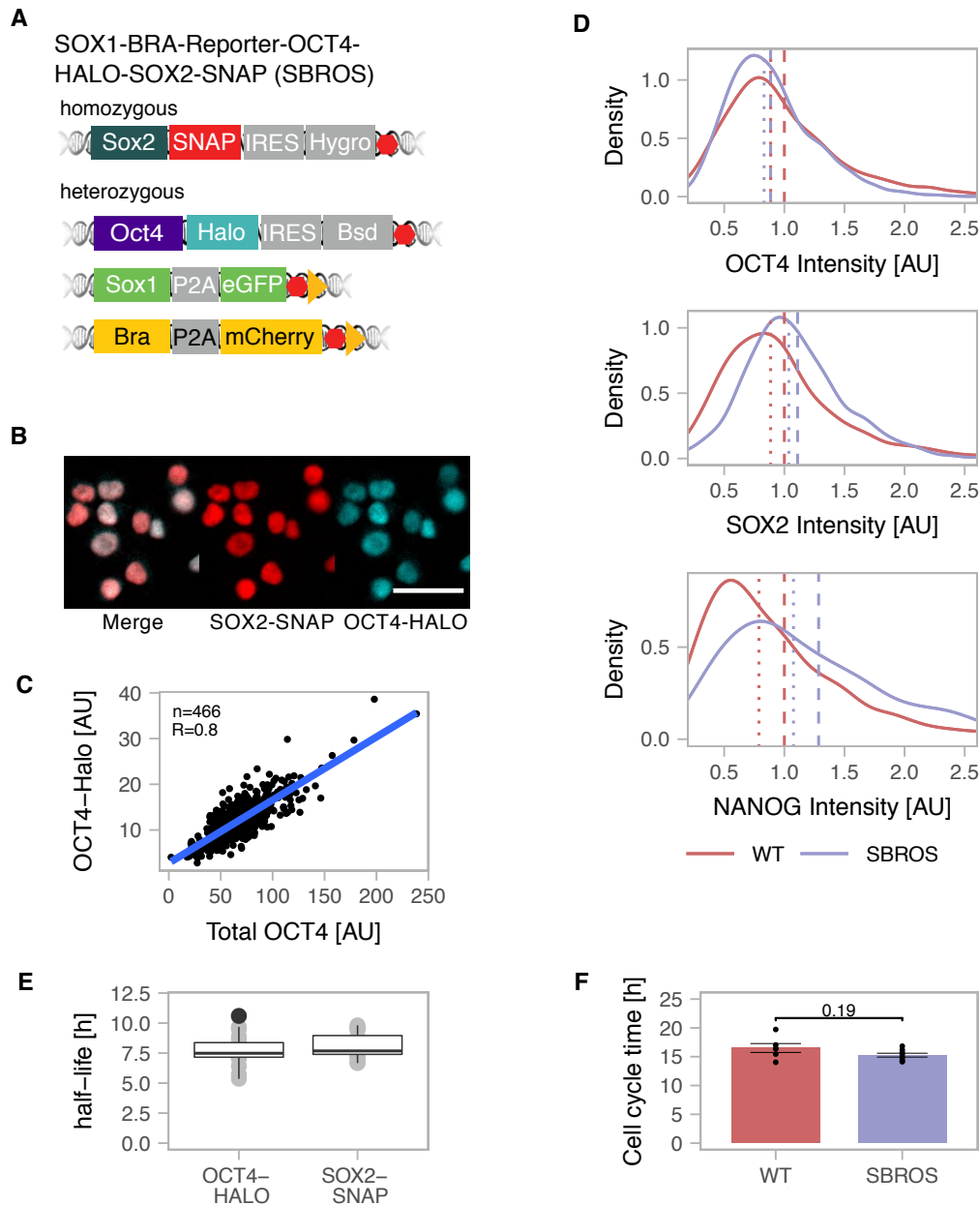


Figure 3.1 Knock-in cell lines.

(A) Scheme of the knock-in alleles of the SOX1-BRA-Reporter-OCT4-HALO-SOX2-SNAP (SBROS) cell line. Red hexagons: STOP codons. Yellow triangles: loxP sites. (B) Example images showing the localization of SOX2-SNAP and OCT4-HALO by fluorescence microscopy of SBROS cells. Scale bar: 50 μ m. (C) Correlation between OCT4-HALO and total OCT4 levels determined by immunofluorescence and HALO labelling. R is Pearson's correlation coefficient. (D) Distributions of OCT4 (n=2682), SOX2 (n=1236) and NANOG (n=2416) levels in WT E14 and SBROS cell lines as determined by quantitative immunofluorescence. Dashed lines: mean protein levels; Dotted lines: median protein levels. (E) Protein half-lives of OCT4-HALO and SOX2-SNAP in SBROS cells. Whiskers: minimum and maximum values; Box: lower and upper quartiles; Solid line: median. Black dots: outliers. (F) Cell cycle duration of E14 WT (n=6) and SBROS (n=8) cells. Error bars: SE. p-value: two-sided t-test with unequal variance.

3.3.2 SOX2 fluctuations regulate OCT4 levels

It has been reported that SOX2 and OCT4 protein levels are positively correlated in single cells (212), however the mechanism underlying this correlation is not understood. While ChIP-seq studies have shown that SOX2 and OCT4 bind to the regulatory sequences of the *Sox2* and *Oct4* genes (213, 228), the functional impact of their binding on *Sox2* and *Oct4* expression is unclear. To determine how SOX2 and OCT4 impact their own and each other's expression levels, we established doxycycline (dox)-inducible ES cell lines allowing precisely timed expression of fusions of SOX2 or OCT4 to the YPet fluorescent protein. The functionality of these fusion proteins was validated by their ability to rescue pluripotency of inducible SOX2 or OCT4 knockout cells (2TS22C (218) or Zhbtc4 cells, respectively) (Supplementary Figure 3.1G and 1J, Tables EV4 and EV5), and by ChIP-seq analysis that showed genome-wide binding profiles similar to WT SOX2 and OCT4, respectively (Supplementary Figure 3.1I).

We then used these constructs to investigate the impact of YPet-OCT4 and YPet-SOX2 expression on endogenous SOX2 and OCT4 protein levels. To do so, YPet-OCT4 and YPet-SOX2 inducible ES cell lines were treated with dox for 0h, 4h or 7h, followed by immunofluorescence staining with anti-OCT4 and anti-SOX2 antibodies. While upon YPet-OCT4 overexpression, SOX2 levels remained stable (Supplementary Figure 3.2A-C), SOX2 overexpression modestly increased OCT4 levels already four hours after induction of SOX2 overexpression (Supplementary Figure 3.2D-F). Interestingly, SOX2 levels slightly decreased as a function of overexpressed YPet-SOX2 over time (Supplementary Figure 3.2G), raising the possibility that SOX2 negatively regulates its own expression. Note that while some studies have suggested that SOX2 positively upregulates *Sox2* expression, these were either based on analysis of SOX2 binding to *Sox2* regulatory regions, which does not allow to infer the functionality of these interactions (229, 230), or using prolonged SOX2 depletion using siRNA (213) that compromises the pluripotent state (1). Furthermore, our findings are in line with previous studies investigating the functional impact of SOX2 on *Sox2* mRNA expression (231, 232).

To confirm negative SOX2 autoregulation, we then monitored endogenous SOX2 levels in live cells upon exogenous SOX2 expression in another knock-in cell line in which both endogenous *Sox2* alleles are fused to *nanoluc*, allowing real-time monitoring of endogenous SOX2 levels by luminescence microscopy. Additionally, a P2A-Firefly Luciferase (FLUC) cassette was also knocked-in in fusion to one allele of *Sox1* to monitor NE commitment (Supplementary Figure 3.2H, Supplementary Figure 3.1A-C). We called this cell line Sox2-Nanoluc-Sox1-Fluc (SNSF). Quantitative immunofluorescence analysis confirmed that the NANOLUC tag did not strongly alter expression level distributions of pluripotency factors, as cells retained 86% of wildtype mean OCT4 levels, 79% of wildtype mean SOX2 levels, 93% of wildtype mean NANOG levels, and distributions and median expression levels of these proteins were similar to wild type E14 cells (Supplementary Figure 3.2I). Growth rates were also unaltered (Supplementary Figure 3.2J). Expression levels of pluripotency markers at the mRNA level (Supplementary Figure 3.1E), cloning efficiency and alkaline phosphatase activity after one week of clonal growth were also similar to WT ES cells (Supplementary Figure 3.1F).

We transduced the SNSF cell line with lentiviral constructs allowing dox-inducible overexpression of SOX2-SNAP or YPet-SOX2-delDBD that lacks the SOX2 DNA binding domain. We then monitored SOX2 protein levels over time by luminescence microscopy after dox induction, as well as SOX2-SNAP or YPet-SOX2-delDBD levels by fluorescence microscopy. While YPet-SOX2-delDBD expression did not impact endogenous SOX2 levels, SOX2-SNAP overexpression reduced SOX2 protein levels to 50% of their initial levels after approximately 9 hours (Figure 3.2A) ($p=1.88 \times 10^{-14}$, Mann-Whitney U test). This time scale is close to the half-life of the SOX2 protein (8.2 ± 1 h (SE) for the half-life of SOX2-SNAP, Figure 3.1E) and thus raises the possibility of a very rapid arrest of SOX2 protein accumulation. To determine whether SOX2 directly alters *Sox2* mRNA levels, we performed RT-QPCR shortly after overexpression of YPet-SOX2 or YPet-SOX2-delDBD. We found that *Sox2* mRNA levels were decreased already 2 hours after dox induction, and were down to 20% of their initial level after 6 hours (Figure 3.2B). This suggests that SOX2 rapidly represses the expression of its own mRNA, in line with earlier studies (231, 232). We next aimed to determine how SOX2 overexpression increased OCT4 levels. Surprisingly, *Oct4* mRNA levels were unaffected after 6 hours of SOX2 overexpression (Figure 3.2B). Since SOX2 and OCT4 form heterodimers, we

reasoned that SOX2 overexpression could increase OCT4 levels by increasing the stability of the OCT4 protein. We thus measured OCT4 half-life by pulse-labelling OCT4-HALO with the Halo-SiR ligand in the SBROS cell line (as described previously in (233)) after 6-8 hours dox induction of SOX2-SNAP expression. We found the half-life of OCT4 to be increased by 50 % (Figure 3.2C) and to be positively correlated to SOX2-SNAP levels in individual cells (Supplementary Figure 3.2K), suggesting that SOX2 increases OCT4 levels by decreasing its degradation rate. In contrast, SOX2 overexpression does not alter its own half-life, as we have shown previously (233). To test whether the increase in OCT4 half-life upon SOX2 overexpression depends on the ability of SOX2 to bind DNA, we generated a cell line allowing dox-inducible overexpression of YPet-SOX2-delDBD. We then treated cells with or without dox for 6-8 hours and pulse-labelled OCT4-HALO to determine its half-life. Unlike SOX2-SNAP, YPet-SOX2-delDBD overexpression did not increase the half-life of OCT4 (Figure 3.2C).

We next aimed to determine how endogenous variations in SOX2 and OCT4 levels affect each other's expression levels. To do so, we labelled SBROS cells with SNAP-SiR647 and HaloTag TMR, and sorted cells for either high or low SOX2-SNAP levels, but with the same, intermediate OCT4-HALO expression levels. In order to minimize the effects of cell cycle progression on differences in SOX2 and OCT4 levels, we sorted cells that were in G1 phase based on DNA content (Figure 3.2D, Supplementary Figure 3.2L). The converse experiment was performed to determine how endogenous OCT4 levels impact SOX2 expression (Figure 3.2D, Supplementary Figure 3.2L). After sorting, cells were kept either in pluripotency maintenance conditions (N2B27+2iLIF) or in differentiation conditions (N2B27). In cells sorted for SOX2-high or SOX2-low levels, OCT4 levels were increased and decreased 8 hours after sorting, respectively (Figure 3.2E). In contrast, in cells sorted for OCT4-high or OCT4-low levels, SOX2 levels remained unchanged 8 hours after sorting (Figure 3.2F). Note that since cells typically require at least 4-6 hours to re-attach to cell culture dishes after plating, a large fraction of cells were still in G1 phase 8 hours after sorting (Supplementary Figure 3.2M). These results suggest that SOX2-high and low cells tend to increase and decrease OCT4 expression levels over time, respectively, and that low amplitude, endogenous variations in SOX2 levels regulate dynamic changes in OCT4 levels.

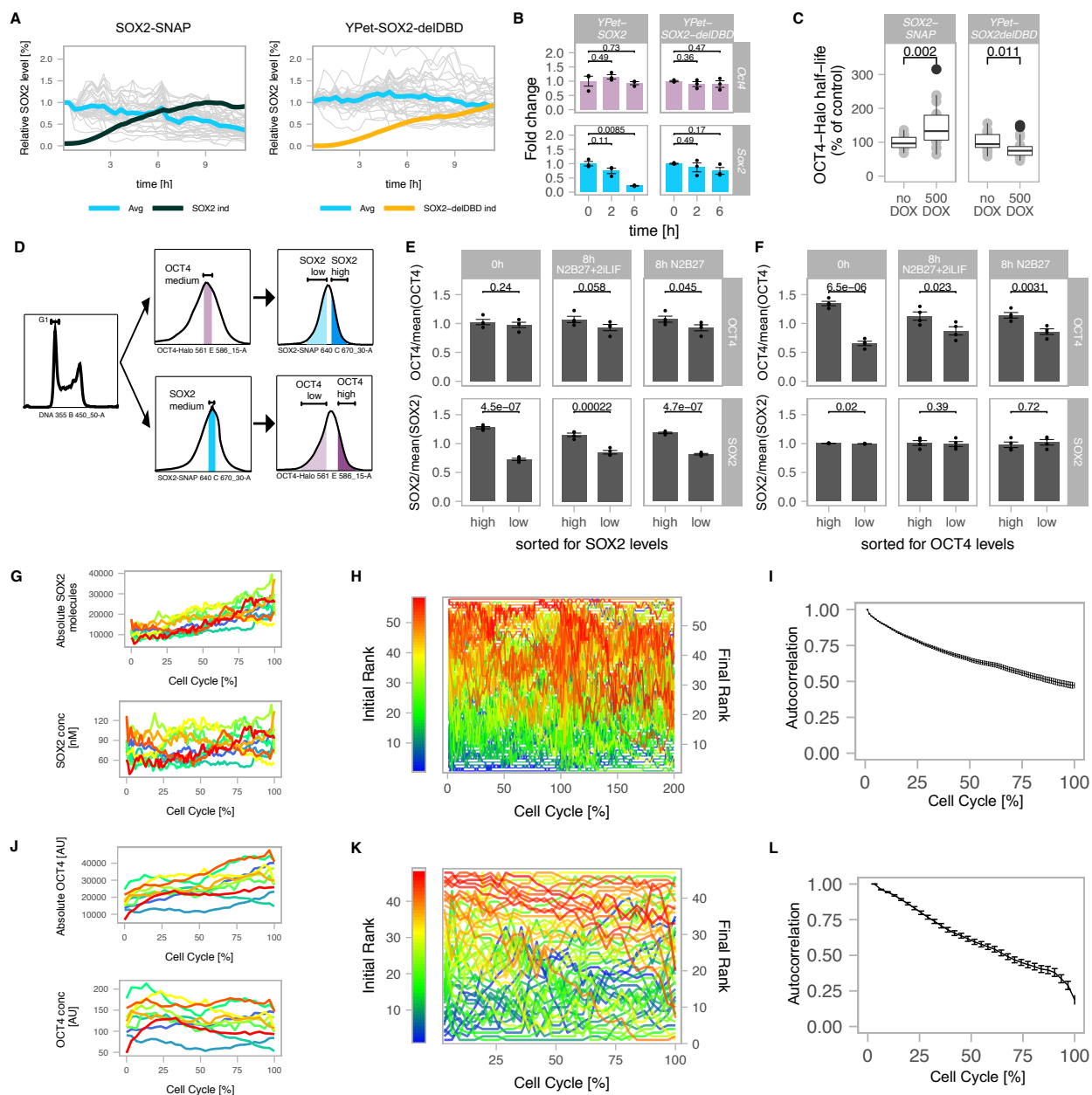


Figure 3.2 Cross-regulation and single-cell fluctuations of SOX2 and OCT4.

(A) Endogenous SOX2 levels in single cells (turquoise) normalized to the value at t=0 in SNSF cells upon induction of SOX2-SNAP (n=37, left) or a truncated SOX2 version missing the DNA binding domain (YPET-SOX2-delDBD, n=47, right). Black line: average SOX2-SNAP levels after induction normalized to the maximum level (n=22); Yellow line: average YPet-SOX2-delDBD levels after induction, normalized to the maximum level (n=20). (B) Sox2 and Oct4 mRNA levels upon overexpression of YPet-SOX2 or YPet-SOX2-delDBD (n=3). p-values: two-sided t-test with unequal variance. (C) Changes of OCT4 half-life upon SOX2-SNAP or YPet-SOX2-delDBD overexpression (n=20 cells each). p-values: Mann-Whitney U test. SBROS cells harbouring inducible SOX2-SNAP or YPet-SOX2-delDBD were treated with 0 or 500 ng/ml dox for 6-8 hours followed by pulse-labelling of OCT4-HALO and live-cell imaging. (D) Sorting strategy to evaluate the impact of endogenous fluctuations of SOX2 and OCT4 on OCT4 and SOX2 protein levels, respectively. (E) Changes of SOX2 and OCT4 levels after sorting for high and low SOX2 levels in cells expressing intermediate OCT4 levels (n=4). Error bars: SE; p-values: one-sided t-test with unequal variance. (F) Changes of SOX2 and OCT4 levels after sorting for high and low OCT4 levels in cells expressing intermediate SOX2 levels (n=4). p-values: one-sided t-test with unequal variance. (G) Representative traces of the absolute number of SOX2-NLUC molecules (top) and inferred concentration in nM (bottom) in in-silico synchronized cells (n=10). (H) Single cells were ranked according to their SOX2 expression at t=0 and assigned a color code for their initial rank from low (blue) to high (red) levels, and changes in ranks over time are shown (n=59). (I) Rank-based autocorrelation function of the SOX2 ranks (n=59). Error bars: SE estimated by bootstrapping. (J) Representative single cell traces of the integrated intensity of OCT4-HALO in

single cells (top) and the corresponding inferred concentration (bottom) in *in-silico* synchronized cells ($n=10$). (K) Representation of the ranks as in (G). (L) Rank-based autocorrelation function as in I ($n=47$). Error bars: SE estimated by bootstrapping.

3.3.3 Characterisation of SOX2 and OCT4 fluctuations

The intercellular variability and the reversion of SOX2 and OCT4 levels towards their mean levels 8 hours after sorting for SOX2-high/low or OCT4-high/low levels (Figure 3.2E and Figure 3.2F, respectively) suggest that these proteins fluctuate over a time scale of hours in individual cells. We thus decided to use our knock-in ES cell lines to directly measure protein expression level fluctuations at the single cell level. We monitored absolute SOX2 levels in the SNSF cell line by luminescence microscopy, using a signal calibration approach we reported previously (234). As expected, SOX2 levels doubled over one cell cycle (Figure 3.2G, Supplementary Figure 3.3A), and SOX2 concentrations calculated after normalization to an inferred nuclear volume (described in (212)) were constant on average (Supplementary Figure 3.3B). In single cells, we found SOX2 concentrations to fluctuate over a 2 to 3-fold range (Figure 3.2G). To measure the time scale of SOX2 concentration fluctuations, individual cells were assigned a rank based on their initial SOX2 level (Figure 3.2H). We then used a rank-based autocorrelation function to determine the time scale of protein level fluctuations. This time scale is referred to as the mixing time (235), and describes how long it takes for a cell to lose its expression rank and thereby to “mix” its expression level with other cells. Using data from either two or a single full cell cycle, we found SOX2 mixing times on the order of one cell cycle (Figure 3.2I and Supplementary Figure 3.3C and D). Since < 2-fold changes in SOX2 expression were reported to compromise pluripotency maintenance (231), our data suggest that rapid readjustment of SOX2 levels may be required to maintain pluripotency despite fluctuation amplitudes of 2-3 fold. We then performed analogous experiments using the SBROS cell line to monitor OCT4-HALO levels over the cell cycle by live fluorescence microscopy, revealing similar fluctuation amplitudes (Figure 3.2J, Supplementary Figure 3.3E and F), and mixing times (Figure 3.2K and L) as compared to SOX2. Thus, both SOX2 and OCT4 display 2 to 3-fold, rapid expression level fluctuations in the pluripotent state.

3.3.4 SOX2 and OCT4 fluctuations impact NE and ME commitment

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 (Figure 3.3A and B). We tracked NLUC and FLUC signals in individual cells over time, and individual cell traces were aligned and normalized for cell cycle duration *in silico*, using linear resampling of the time variable (see Materials and methods). We then grouped cells according to their expression of FLUC, and traced their SOX2 expression levels one cell cycle before, during the cell cycle of FLUC expression onset, and in the subsequent cell cycle. Traces of cells that did not express FLUC were temporally aligned to traces of cells that expressed FLUC, so that both groups of cells were compared at the same average time after which self-renewal signals were withdrawn (Figure 3.3C). Interestingly, 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 correlated with a higher probability of NE differentiation (Figure 3.3C, marked by *). This suggests that SOX2 levels may play a role in NE commitment at the M-G1 transition. This finding is in line with an earlier study from our laboratory, in which we found that the absence of SOX2 at the M-G1 transition suppresses its ability to enhance neuroectodermal fate commitment (224).

We next investigated whether different SOX2 levels at the very onset of differentiation impact NE and ME commitment. To do so, we sorted G1-gated SBROS cells stained with SNAP-SiR647 for low, medium and high SOX2 levels (Figure 3.3D and Supplementary Figure 3.4A). Cells were then released from self-renewal conditions by seeding in N2B27 medium devoid of 2i and LIF, and four days later NE and ME commitment were 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 (Figure 3.3E), suggesting that high SOX2 levels at the time of release from self-renewal enhance NE fate commitment. In contrast, SOX2 levels had only a weak impact on ME commitment in these conditions (Figure 3.3E).

To determine the impact of OCT4 levels at the onset of pluripotency exit on NE and ME commitment, we sorted G1-gated SBROS cells stained with HaloTag TMR in OCT4-low and OCT4-high subpopulations (Figure 3.3D and Supplementary Figure 3.4B) 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 (Figure 3.3F), even though average OCT4 levels differed by < 2-fold between them (Supplementary Figure 3.4B). We next aimed to explore the potential causal relationship between high OCT4 levels and increased NE/ME commitment. To do so, we generated a cell line allowing for inducible expression of SNAP-OCT4 in the SBR background, and treated it with dox for 12h to overexpress OCT4 for a brief period of time, thereby mimicking the timescales of endogenous OCT4 fluctuations (note that we previously demonstrated the functionality of the SNAP-OCT4 fusion protein (224)). Subsequently, cells were sorted for G1 phase and differentiated for four days in N2B27 medium without 2i and LIF (Figure 3.3G and Supplementary Figure 3.4C). Strikingly, this led to a ~ 2-fold increase in the fraction of mCherry+ cells, suggesting that even a brief increase of OCT4 levels prior to self-renewal release enhances ME commitment. However, NE commitment was not enhanced as the fraction of eGFP+ cells remained unchanged (Figure 3.3G, Supplementary Figure 3.4C). We reasoned that this discrepancy could be caused by opposing roles of high OCT4 levels before and after removal of pluripotency signals. To investigate the impact of OCT4 overexpression after release of pluripotency signals, we treated SNAP-OCT4 cells with dox for four days after removal of 2i and LIF in order to overexpress OCT4 during the differentiation process. While this also led to a similar increase in ME differentiation as obtained with a short dox treatment, NE commitment was strongly inhibited (Supplementary Figure 3.4D and E). In summary, our data suggests that high OCT4 levels throughout differentiation inhibit NE commitment (in line with (170)), in contrast to high endogenous OCT4 levels before the onset of differentiation that enhance NE commitment. Finally, providing a short pulse of dox induction in the SNAP-OCT4 inducible SBR cell line before differentiation yields an intermediate NE differentiation outcome, which can be explained by reaching supraphysiological OCT4 levels both before and at the onset of differentiation.

We next measured the combinatorial impact of SOX2 and OCT4 on differentiation by sorting G1-gated SBROS cells into four different subpopulations (Figure 3.3H and Supplementary Figure 3.4F) followed by four days of differentiation (Figure 3.3I, Supplementary Figure 3.4G). As expected, SOX2-high/OCT4-high (SHOH) cells were the most efficient to differentiate towards NE and ME, SOX2-low/OCT4-high (SLOH) cells were less capable to differentiate towards NE as compared to SHOH, and OCT4-low (SHOL and SLOL) populations were strongly impaired in differentiating towards both fates. Overall, OCT4 levels had a larger impact than SOX2 levels on NE and ME differentiation. Furthermore, as OCT4-high cell populations had relatively similar differentiation efficiencies regardless of SOX2 levels, we conclude that the OCT4/SOX2 ratio is not a strong determining factor in differentiation outcome. To ask whether endogenous SOX2 and OCT4 level variability impacts NE and ME commitment during later cell cycle stages, we performed the same experiments in S phase-gated cells. We found the impact of SOX2 on NE commitment and the impact of OCT4 on both NE and ME commitment to be decreased (Figure 3.3J, Supplementary Figure 3.4H). Note that the differentiation efficiency in our conditions was not lower in S phase-sorted as compared to G1 phase sorted cells, in contrast to previous studies (236, 237). In fact, while ME differentiation of S phase-sorted cells was comparable to G1 phase-gated cells, NE differentiation was increased (Supplementary Figure 3.4I). These discrepancies likely result from differences in differentiation protocols and/or endpoint readouts. Regardless, our data suggest that differentiation to NE and ME depends on the levels of OCT4 and SOX2 in G1 but not in S phase, which in the case of SOX2 is in line with its reported function at the Mitosis-G1 transition in NE induction (224).

We next aimed at determining whether SOX2 and OCT4 levels also impact cell fate commitment upon directed differentiation. To do so, we cultured G1-sorted SHOH, SHOL, SLOH and SLOL subpopulations in N2B27 supplemented with 3 μ M CHIR99021 during three days to drive ME differentiation (238). Importantly, only very low percentages of SOX1-positive cells were detected under this regime (Supplementary Figure 3.4J). Surprisingly, we found that OCT4-high and OCT4-low cells displayed a similar potential to differentiate towards ME (Figure 3.3K). In contrast, SOX2-low cells were almost two-fold more efficient than SOX2-high cells to differentiate towards ME (Figure 3.3K). This suggests that the sustained activation of the Wnt pathway overcomes the influence of OCT4 in priming ME differentiation, and that in this context SOX2 expression levels

become determining for ME differentiation efficiency. We next differentiated G1-sorted SHOH, SHOL, SLOH, and SLOL cells in the presence of 1 μ M SB431542 and 25 ng/ml bFGF to direct cells toward neuroectoderm. This increased the proportion of SOX1-positive cells and decreased the proportion of BRA-positive cells to about 0.5% (Supplementary Figure 3.4J). In these conditions, there was no significant difference in the differentiation outcome of SHOH, SHOL, SLOH, and SLOL cells (Figure 3.3L). Taken together, these results suggest that the impact of endogenous OCT4 and SOX2 fluctuations on ES cell differentiation is highly context-dependent.

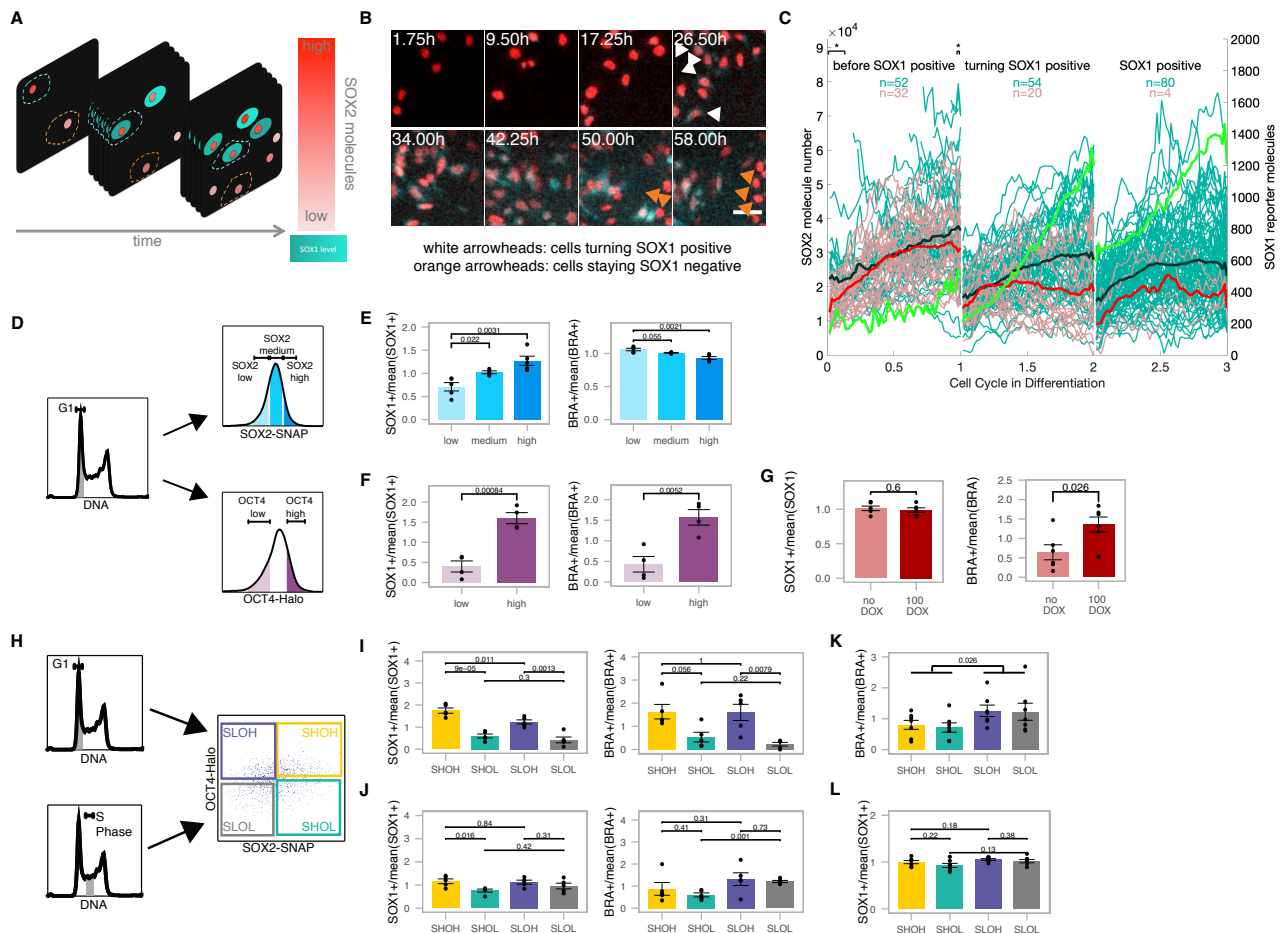


Figure 3.3 Endogenous SOX2 and OCT4 level fluctuations bias differentiation.

(A) Scheme of single-cell tracking for SNSF cells turning SOX1-positive (turquoise cytoplasm) or remaining SOX1-negative throughout the experiment. Turquoise dashed contour: measured region for cells turning SOX1 positive; Red dashed contour: measured region for cells staying SOX1 negative. (B) Snapshot from luminescence movie of SNSF cells in differentiation. White arrowheads: cells starting to express SOX1-P2A-FLUC; Orange arrowheads: cells staying SOX1-negative throughout the experiment. Red: SOX2-NLUC, turquoise: Sox1-P2A-FLUC. Scale bar: 50 μ m. (C) Single cell traces of SOX2 levels in differentiating cells. Turquoise traces: cells becoming SOX1-positive in the middle cell cycle, population average is shown in bold black; Red traces: cells remaining SOX1-negative throughout the experiment, population average is shown in bold; Green line: population average FLUC signal in cells that become SOX1-positive. * $p < 0.05$ determined by two-sided t-test with unequal variance. (D) Strategies to sort SBROS cell populations with different SOX2 and OCT4 levels. (E) Differentiation outcome for cells sorted for different SOX2 levels ($n=5$ for low and high, $n=4$ for medium). p -values: two-sided t-test with unequal variance. (F) Differentiation outcome for cells sorted for different OCT4 levels ($n=4$). p -values: two-sided t-test with unequal variance. (G) Differentiation outcome of uninduced cells or cells induced 12 hours for OCT4-SNAP overexpression before sorting in G1 phase ($n=6$). p -values: two-sided t-test with unequal variance. (H) Sorting strategy for SBROS cells in 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) Differentiation outcome for cells sorted in G1 phase ($n=5$). p -values: two-sided t-test with unequal variance (SOX1) and Mann-Whitney U test (BRA). (J) Differentiation outcome for cells sorted in S phase ($n=5$). p -values: Mann-Whitney U test (SOX1) and two-sided t-test with unequal variance (BRA). (K) Differentiation outcome for cells sorted in G1 phase differentiated in the presence of 3 μ M CHIR ($n=7$). p -values: two-sided

t-test with unequal variance. (L) Differentiation outcome for cells sorted in G1 phase differentiated in the presence of 1 μ M SB431542 and 25 ng/ml bFGF ($n=7$). *p*-values: two-sided *t*-test with unequal variance.

3.3.5 OCT4-high cells open differentiation enhancers

We then aimed to identify the molecular mechanisms by which small and transient endogenous fluctuations of SOX2 and OCT4 can result in major biases in differentiation potential. As SOX2 and OCT4 were shown to regulate chromatin accessibility (239, 240), we reasoned that small changes in their expression level could prime cells for different fates by altering the chromatin accessibility landscape. We thus performed ATAC-seq in G1-sorted SHOH, SHOL, SLOH and SLOL cells. We quantified the fraction of reads in 81'132 open regions identified by peak calling and found no overall differences between conditions (Supplementary Figure 3.5A). We compared the accessibility of each open region in high vs low conditions for SOX2 and OCT4 as well as for SHOH vs SLOL. 3'914 loci (4.8%) were significantly up- or downregulated (FDR < 10%) upon changes in OCT4 alone (538 loci), SOX2 alone (1'259 loci), or SOX2 and OCT4 together, i.e. SHOH vs SLOL (2'117 loci). We grouped these loci into OCT4-regulated, SOX2-regulated and co-regulated loci that were either upregulated (more open in high cells) or downregulated (more open in low cells) (Figure 3.4A and B, Supplementary Figure 3.5B and C). Loci in all groups were close to differentiation-associated genes (Figure 3.4B) but those upregulated by OCT4 alone were the most enriched for genes involved in differentiation processes (Figure 3.4C, Supplementary Figure 3.5D-H). Using a subset of OCT4-upregulated loci that overlap OCT4 peaks and with a more stringent false discovery rate (< 5%) also revealed enrichment for differentiation terms (Supplementary Figure 3.5I), showing that OCT4 binds to regulatory elements near differentiation genes that are more accessible in cells with higher OCT4 levels. In contrast, loci overlapping pluripotency-associated super-enhancers were unaffected by SOX2 and OCT4 levels (Figure 3.4D). Loci in which accessibility was positively correlated to SOX2 and OCT4 levels were enriched for SOX2 and OCT4 binding, while those that were negatively correlated showed less overlap with SOX2 and OCT4 ChIP-seq peaks and were enriched for H3K4me3 (Figure 3.4E), marking active promoters and transcription start sites, and close to promoters (Supplementary Figure 3.5J). Note that while we used SOX2 and OCT4 peaks from two ChIP-seq datasets generated in the same conditions for meaningful comparison, which yielded ~30% of overlap of upregulated loci, including peaks from two additional OCT4 ChIP-seq datasets (with no comparable SOX2 data) increased this number to ~70% (Supplementary Figure 3.5K), indicating that most of these loci are bound by OCT4 and possibly also by SOX2. The OCT motif was most enriched near loci upregulated in OCT4-high cells, and the SOX motif was most enriched near loci upregulated in SOX2-high cells (Supplementary Figure 3.5L). However, the fraction of sites overlapping these canonical motifs was low, suggesting that OCT4 and SOX2 may bind indirectly with other partners. To perform a search for binding partners in these regions, we analysed the overlap with peaks from 3'724 ChIP-seq datasets in ES cells extracted from cistromeDB (241). We trained a random forest model to categorize loci in the different groups (see Materials and methods), including accessible regions that were unaffected, based on the peaks overlapping the regions. While the model performed poorly in predicting the group a region belonged to based on the peaks a region overlapped, it performed better at distinguishing between upregulated, downregulated, and unaffected loci (Supplementary Figure 3.5M). We noticed that many of the top classifiers enriched in downregulated loci included promoter and gene body signatures such as Cdk9, Pol2 and H3K79me2, consistent with these loci being enriched near genes, while upregulated loci were enriched for the enhancer mark H3K4me1 (Appendix table 1). One of the factors with enriched overlap in the upregulated regions was NANOG (Appendix table 1). Accordingly, NANOG binding was strongly enriched at loci whose accessibility was upregulated in OCT4-high or SOX2-high cells compared to regions with downregulated accessibility (Supplementary Figure 3.5N). This suggests that NANOG may act as a binding partner to regulate chromatin accessibility at these loci, in line with a recent study (242). Loci with increased accessibility in OCT4-high cells lose accessibility upon rapid OCT4 depletion (data from (239)), in line with OCT4 directly regulating accessibility at these sites (Supplementary Figure 3.5O) even though we cannot exclude that some loci may be regulated by indirect mechanisms. Taken together, our data suggest that fluctuations of endogenous SOX2 and OCT4 lead to temporal changes in chromatin accessibility, and that high OCT4 levels result in the opening of differentiation-associated enhancers.

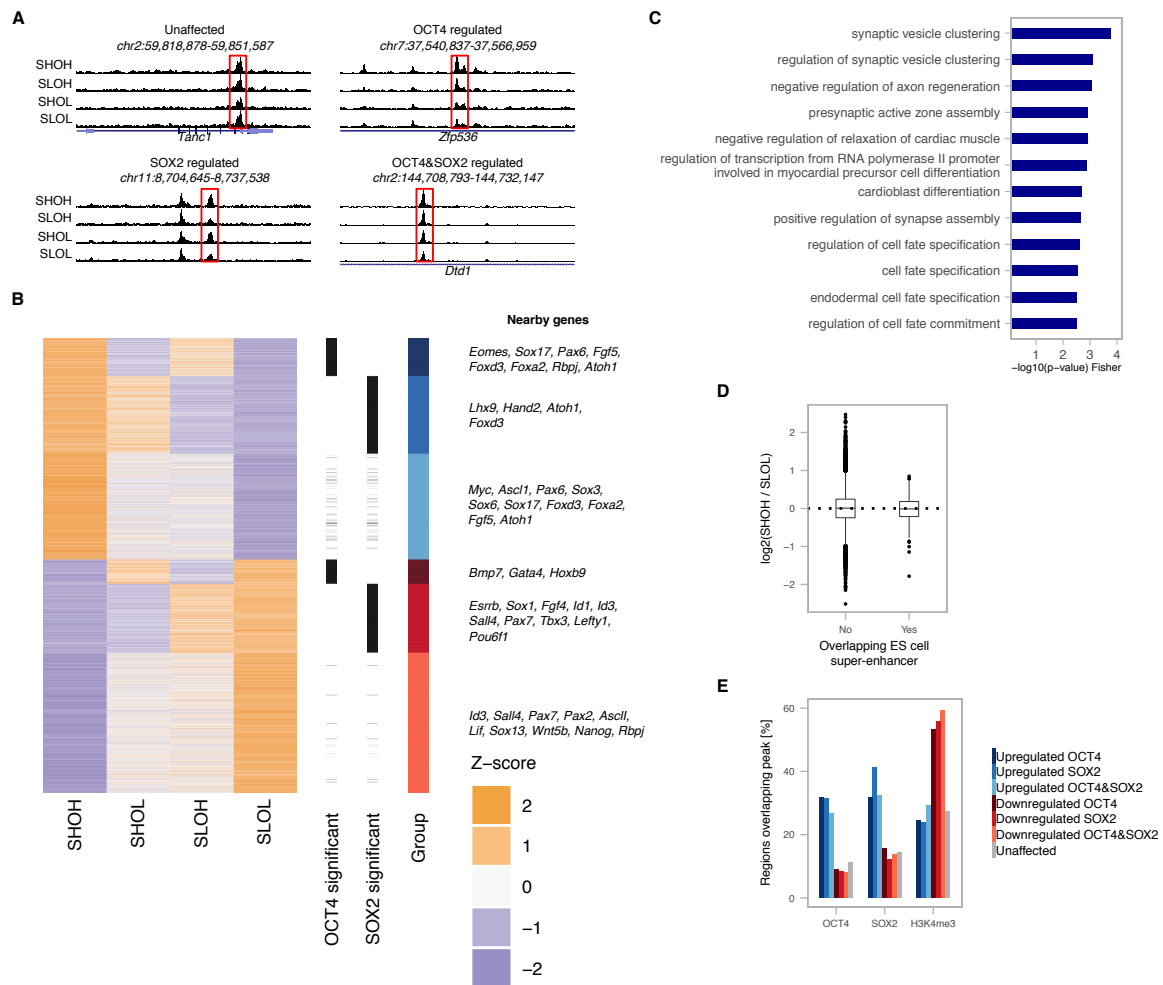


Figure 3.4 SOX2 and OCT4-dependent changes in chromatin accessibility.

(A) Example tracks of the ATAC-seq signal in SHOH, SHOL, SLOH and SLOL populations for loci where chromatin accessibility is unaffected, affected by OCT4 levels, SOX2 levels, or affected by both. (B) Heatmap of relative chromatin accessibility in SHOH/SHOL/SLOH/SLOL samples for all loci with significantly altered accessibility, with examples of genes close to affected regulatory regions. Groups are defined by their significance (OHvsOL/SHvsSL/SHOHvsSLOL) and fold-change direction. (C) Gene Ontology (GO) term enrichment analysis for genes close to regions with higher accessibility in OCT4-high cells (Upregulated OCT4). (D) Chromatin accessibility ratio of SHOH vs SLOL cells in loci overlapping ES cell super-enhancer regions or not. (E) Percentage of regions in each group showing an overlap with OCT4, SOX2 and H3K4me3 ChIP-seq peaks.

3.4 Discussion

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 (243), these are caused by transitions between naïve and primed ES cells (212) 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 bias differentiation potential.

Near-complete OCT4 depletion was shown to lead to upregulation of trophectoderm markers (217). In contrast, less dramatic decreases in WT OCT4 levels lead to enhanced pluripotent self-renewal, as well as compromised differentiation towards ME and NE (244, 245), in line with our findings. Conversely, constant OCT4 overexpression was shown to lead to enhanced differentiation towards the mesoderm lineage (170, 217). This is also in line with our findings that OCT4-high cells are more prone to differentiate towards ME than OCT4-low cells. The enhancement of NE commitment by

elevated SOX2 levels that we described here is also in line with previous studies (170, 171, 246), but here we show that enhanced NE is not only caused by SOX2 overexpression throughout differentiation, but that high endogenous SOX2 levels at the onset of differentiation are sufficient to significantly bias ES cells towards the NE lineage. At first sight, the enhancement of NE commitment by high endogenous OCT4 levels seems to contradict previous work (170). However, this study used either OCT4 overexpression or measured endogenous OCT4 levels at fixed time points, throughout directed NE or ME differentiation. While we confirmed that OCT4 overexpression during differentiation suppresses NE differentiation, we also found that elevated OCT4 levels at the time self-renewal signals are removed increases NE commitment. Taken together, this suggests that while high OCT4 levels consistently increase ME commitment both before and during differentiation, high OCT4 levels enhance NE differentiation only at the onset of self-renewal signals removal. We also found that in the context of directed ME and NE differentiation, the impact of endogenous OCT4 and SOX2 on cell fate commitment becomes much weaker. Therefore, the modulation of signalling pathways can partially or completely buffer the impact of endogenous fluctuations of OCT4 and SOX2 on ES cell differentiation.

The large impact of small amplitude, transient OCT4 fluctuations on ES cell differentiation in the absence of strong signalling cues suggests 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 an explanation for these observations, the reason for differential responses of pluripotency regulatory elements as compared with differentiation-related enhancers is unclear. Using thousands of ChIP-seq datasets in ES cells, we could train a model to predict, with relatively good accuracy, regions that were upregulated or downregulated in accessibility upon high levels of OCT4 or SOX2 (Supplementary Figure 3.5M). We identified NANOG binding as enriched in upregulated regions, consistent with a recent report on the impact of NANOG depletion on OCT4/SOX2 binding and accessibility (242). However, we could not find features allowing to predict whether loci are more dependent on OCT4 vs SOX2. The potential role for cooperativity with other proteins or differential affinity of OCT4/SOX2 binding sites will require further investigation. We observed widespread SOX2-dependent accessibility changes, including at many differentiation enhancer-associated loci. However, loci with increased accessibility in SOX2-high cells were less enriched for differentiation-associated enhancers than in OCT4-high cells (Supplementary Figure 3.5D), which is in line with the weaker effects of SOX2 levels on differentiation. Finally, the fact that cells are more sensitive to SOX2 and OCT4 level variability in G1 phase than S phase raises the possibility that these transcription factors mainly act shortly after mitosis to re-open closed enhancer regions, in line with their reported pioneer transcription factor activity (247) and their essential function in cell fate decisions at the mitosis-G1 transition (224, 248).

3.5 Materials and methods

Cell Culture

The SBR (224) and SBROS cell lines were generated from CGR8 ES cells (Sigma, Cat#07032901-1VL), and the E14 cell line (kindly provided by Didier Trono, EPFL) was used for all ES cell experiments. Cells were routinely cultured on dishes coated with 0.1% gelatin type B (Sigma, Cat#G9391-100G), in GMEM (Sigma, Cat#G5154-500ML) supplemented with 10% ES cell-qualified fetal bovine serum (Gibco, Cat#16141-079), nonessential amino acids (Gibco, Cat#11140-050), 2 mM L-glutamine (Gibco, Cat#25030-024), sodium pyruvate (Sigma, Cat#S8636-100ML), 100 μ M 2-mercaptoethanol (Sigma, Cat#63689-25ML-F), penicillin and streptomycin (BioConcept, Cat#4-01F00-H), homemade leukemia inhibitory factor (LIF), CHIR99021 (Merck, Cat#361559-5MG) at 3 μ M and PD184352 (Sigma PZ0181-25MG) at 0.8 μ M. Cells were passaged by trypsinisation (Sigma, Cat#T4049-100ML) every two to three days at a ratio of 1:10.

For imaging experiments, ES cells were cultured on dishes coated with 5 μ g/ml E-Cadherin, in N2B27 medium supplemented with LIF, CHIR99021 at 3 μ M and PD184352 at 0.8 μ M (N2B27+2iLIF). E-Cadherin coating was performed as previously described (249). Briefly, 5 μ g/ml E-Cadherin (R&D, Cat#8875-EC or Cat#748-EC) in PBS (with Ca²⁺ and Mg²⁺; Amimed, Cat#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 1:1 mixing of DMEM/F12 (Gibco, Cat#11320-033) + N2 supplement (Gibco, Cat#17502-001) medium with Neurobasal (Gibco, Cat#21103-049) + B27 supplement (Gibco, Cat#17504-001) medium, supplemented with penicillin (1000 IU/ml) and streptomycin (1000 mg/ml), 2 mM L-Glutamine and 0.1 mM 2-mercaptoethanol.

HEK293T cells were cultured in DMEM (Gibco, Cat#41966-029) supplemented with 10% fetal bovine serum (Gibco, Cat#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: 8 µg/ml of Blastidicin (Gibco A11139-03), 2 µg/ml of Puromycin (Gibco A11138-03) and 200 µg/ml of Hygromycin B (Invitrogen 10687010). ES cells were transfected using X-tremeGENE 9 transfection reagent (Sigma, Cat#06 365 809 001).

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 3.1A).

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 consists 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 *Oct4*, *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 (210). The guide RNA sequences are listed in Table 3.1.

To generate the SBROS cell line, SBR cells (224) 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 12 nM SNAP-SiR 647, and single cells were sorted for high SNAP expression into 96-well plates and grown out. One clone identified as homozygously targeted by PCR on genomic DNA (Supplementary Figure 3.1B) was further validated by Western Blot (Supplementary Figure 3.1C) 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 Blastidicin 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 (Supplementary Figure 3.1B) was further analysed by Western Blot (Supplementary Figure 3.1C).

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 Blastidicin was initiated two days later. A homozygous Blastidicin-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 (Supplementary Figure 3.1B). This resulting intermediate cell line (Sox2-NLuc) 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 (Supplementary Figure 3.1B). The fusion of NLUC to SOX2 was confirmed by Western blotting (Supplementary Figure 3.1C).

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 *Oct4* allele of the SBROS cell line.

Lentiviral Vector Production and generation of stable cell lines

Lentiviral vectors were produced by Calcium Phosphate co-transfection of HEK293T cells with the envelope (PAX2), packaging (MD2G), and lentiviral construct of interest. The viral vectors were concentrated 120-fold by ultracentrifugation as described previously (207). Stable cell lines were generated by transducing 50,000 cells per well of a 24-well plate with 50 μ l of concentrated lentiviral vector particles. Antibiotic selection was started 48-72 hours later and maintained throughout passaging.

Alkaline phosphatase assays

ES cells were plated at 400 cells per well of a gelatinated 6-well plate in medium with 10% serum, 2i and LIF. The medium was exchanged every two to three days, and after 7 days alkaline phosphatase staining was performed following the manufacturer's instruction (Sigma, Cat#86R-1KT).

Pluripotency rescue colony scoring

Pluripotency rescue colony scoring was performed on the cell colonies right before alkaline phosphatase assays. Dox treatment of Zhbtc4 cells (inducible Oct4 knockout) results in an almost complete loss of colonies, while dox treatment of 2TS22C cells (inducible Sox2 knockout) results in a high number of flat colonies. Therefore, in Zhbtc4-derived cell lines, pluripotency rescue was scored by measuring the ratio of colonies obtained with and without dox treatment. In 2TS22C-derived cell lines, pluripotency rescue was scored by quantifying the ratio of dome-shaped versus flat colonies.

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 (250) and inserted between a BamHI and a XhoI site. Subsequently, 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 (224), in which eGFP was replaced by Fluc using restriction cloning with AclI and SacI. For all pX330 (Addgene Cat#42230) and pX335 (Addgene Cat#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 religated, 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 was 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.

Confirmation PCRs

For all knock-in verification PCRs, genomic DNA was purified using a genomic DNA purification kit (Sigma, Cat#G1N350-1KT) and subsequently used to identify clones with correctly targeted alleles. PCR was done using Phusion High Fidelity DNA Polymerase (ThermoScientific, Cat#F530L). Primers are listed in Table 3.2.

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 (20 mM Tris-HCl pH7.4, 10mM NaCl, 3mM MgCl₂), supplemented with 1 mM PMSF (AppliChem, Cat#A0999.0005) and Protease Inhibitors (Sigma, Cat#P8340-5ML). After 15 minutes of incubation on ice, cells were lysed by the addition of 25 µl of 10% NP-40 (Roche, Cat#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 30 µl RIPA Buffer per million of cells (50 mM Tris-HCl pH7.4, 1% NP-40, 0.5% Sodium Deoxycholate, 0.1% SDS, 2 mM EDTA, 150 mM NaCl), supplemented with PMSF (AppliChem, Cat#A0999.0005) and Protease Inhibitors (Sigma Cat#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 soluble nuclear proteins 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, Cat#23235) and 15 µg of protein were mixed with Laemmli sample buffer (Invitrogen, Cat#NP0007) and loaded on an SDS gel (BioRad, Cat#456-1094) for separation (SDS Running Buffer 25 mM Tris, 190 mM Glycine, 0.1%SDS). Proteins were subsequently transferred (Transfer Buffer 25 mM Tris, 190 mM Glycine, 20% Methanol, 0.1% SDS) from the gel onto a PVDF membrane (Merck, Cat#IPVH07850) using a wet transfer system. The membrane was blocked with 5% milk (Roth, Cat#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, Cat#170-5060) and the signal was detected on a Fusion FX 7 apparatus (Vilber). The antibodies and concentrations used are summarized in Table 3.3.

Quantitative PCR (RT-QPCR)

Total RNA was extracted using the GenElute™ Mammalian Total RNA Miniprep Kit Q-PCR (Sigma, Cat#RTN350), and reverse transcription was performed using an oligoDT primer using superscript II (Thermofisher, Cat#18064014). QPCR was performed on a 7900HT Fast Real-Time PCR System (Thermofischer) with SYBR green reagent (Roche, Cat#04707516001). The *Rps9* cDNA was used for data normalization. Primers used for RT-QPCR are listed in the Table 3.4.

ChIP-seq

The 2TS22C EF1α-YPet-Sox2, Zhbtc4 EF1α-YPet-Oct4 and Zhbtc4 EF1α-Oct4-HALO cells were treated with 1 µg/ml doxycycline for 40-48 hours before fixation to ensure complete depletion of endogenous SOX2 (218) and OCT4 (217). At least 10⁷ cells were fixed in 1% formaldehyde for 10 minutes at room temperature, subsequently quenched with 200 mM Tris-HCl pH 8.0, washed with PBS, spun down, and stored at -80°C. The cell pellet was then resuspended in 1.5 ml LB1 (50 mM HEPES-KOH pH 7.4, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10% Glycerol, 0.5% NP40, 0.25% TritonX-100), incubated 10 min at 4°C, spun down, resuspended in 1.5 ml LB2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA), and incubated 10 min at 4°C. The pellet was spun down and rinsed twice with SDS shearing buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.15% SDS), and resuspended in 0.9 ml SDS shearing buffer. All buffers contained Protease Inhibitor Cocktail in DMSO (Sigma, Cat#P8340) diluted at 1:100. The suspension was transferred to a milliTUBE 1 ml AFA fiber and sonicated on a E220 focused ultrasonicator (Covaris) using the following settings: 20 min, 200 cycles, 5% duty, and 140W. Sonicated chromatin was incubated with 5 µg/10⁷ cells of the αSOX2 Y-17 (Santa Cruz, Cat#sc-17320) or 5µl of antibody/4x10⁶ cells of the αOCT4A C30A3C1 (Cell Signaling Technology, Cat#5677S) at 4°C overnight. 0.5% BSA-blocked Protein G Dynabeads (Thermo Fischer, Cat#10003D) were added to the chromatin and incubated for 3 hours at 4°C. The

chromatin was washed several times at 4°C with 5 min incubation between each wash and 2 min magnetization to collect beads; twice with Low Salt Wash Buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.15% SDS, 1 mM PMSF), once with High Salt Wash Buffer (10 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.15% SDS, 1 mM PMSF), once with LiCl Wash Buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 250 mM LiCl, 1% NP40, 1% sodium deoxycholate, 1 mM PMSF), and with TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). Beads were subsequently resuspended in Elution buffer (TE buffer with 1% SDS and 150 mM NaCl), treated with 400 ng/ml Proteinase K and reverse crosslinked at 65°C overnight. Samples were purified using MinElute PCR purification kit (Qiagen, Cat#28006). For SBR, Zhbtc4 EF1 α -YPet-Oct4 and Zhbtc4 EF1 α -Oct4-HALO, cells were first fixed with 2 mM Disuccinimidyl glutarate (DSG) (Thermo Fisher, Cat#20593) for 50 minutes in PBS at room temperature before proceeding with 1% formaldehyde fixation and chromatin immunoprecipitation as described above. One replicate of each ChIP-seq library was prepared with NEBNext ChIP-seq Library Prep Master Mix Set (NEB, Cat#E6240S) using insert size selection of 250 bp. All libraries were sequenced with 75-nucleotide read length paired-end sequencing on a Illumina NextSeq 500 with 30-50 million reads being sequenced for each sample.

Co-immunoprecipitation

The SBROS and SBR cell lines were plated at a density of about 2 million cells per 10cm dish, and 48 hours later these were treated with 1% formaldehyde for 10 min. Cells were subsequently washed with PBS, and then collected in 1 ml lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1% Triton X-100, inhibitors for proteinase and phosphatase). The lysates were then incubated at 4°C with or without α SOX2 antibody (Santa Cruz, Cat#sc-17320) using 5 μ g/10⁷ cells. The next day, 5% BSA-blocked Protein G Dynabeads (Thermo Fischer, Cat#10003D) were added to the samples, which were then washed four times in 1 ml wash buffer (50 mM Tris pH 7.4, 500 mM NaCl, 1% Triton X-100, inhibitors for proteinase and phosphatase). Samples were subsequently incubated at 98°C for 10 min in 40 μ l of 2 \times Laemmli buffer, centrifuged at 2,000 \times g for 3 min, and frozen at -80°C. Samples were then used for Western Blotting using α -OCT4 C-10 antibody at a dilution of 1:200 (Santa Cruz, Cat#sc-5279).

Immunofluorescence and Image Acquisition

ES cells were fixed for 15 to 30 min with ice-cold 2% PFA (AppliChem, Cat#A0877,0500) in PBS, permeabilised and blocked with chilled PBS-Triton (0.5%, AppliChem, Cat#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, Cat#BP337-500), incubated with 1 μ g/ml DAPI for 15 minutes, washed twice with 0.1% PBS-Tween and once with PBS. The antibody dilutions are listed in Table 3.5.

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

Luminescence Time-lapse microscopy

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, Cat#FD35-100) coated with E-Cadherin in 2 ml of N2B27+2iLIF as described above. For quantitative NANOLUC imaging, knock-in cells were mixed with Calibration cells at a ratio of 1:10 as described previously (234). The medium was supplemented with 1 mM Luciferin (NanoLight Technology, Cat#306A) and 0.5 μ l of RealTime Glo Cell Viability Assay Substrate (Promega, Cat#G9711). For imaging in pluripotency maintenance conditions, images were acquired every 299 s in 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 the NANOLUC channel 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 FLUC 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 FLUC 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 as described above. The next day, the medium was supplemented with 12 nM SNAP-SiR647 dye (NEB, Cat#S9102S) for SOX2-SNAP imaging or 50 nM Halo-SiR ligand (gift from Suliana Manley, EPFL) 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 500 ng/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 Figure 3.3E and F).

Flow Cytometry and Fluorescence-Activated Cell Sorting (FACS)

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

To determine how endogenous SOX2 levels regulate OCT4 expression, we used the following sorting strategy: Cells were gated for G1 based on 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 3.2L). 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 HaloTag TMR and SNAP-SiR647 dyes at a concentration of 100 nM or 12 nM, 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.

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 3.4A) or into ~25% of OCT4-HALO high and low cells (Supplementary Figure 3.4B), 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 3.4F).

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

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 (WPI, Cat#FD35-100) and incubated in N2B27+2iLIF overnight. The next day, the medium was changed to N2B27 supplemented with 1 mM 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 coated with 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.

To direct differentiation towards the mesendoderm, cells were seeded at a density of 60,000 cells/well of a gelatin-coated 12-well plate in N2B27 medium supplemented with 3 μ M of CHIR99021. Three days later, differentiation outcomes were assessed by flow cytometry on a BD Fortessa. To direct differentiation towards the neuroectoderm, cells were seeded at a density of 60,000 cells/well of a gelatin-coated 6-well plate or 30,000 cells/well of a gelatin-coated 12-well plate in N2B27 medium supplemented with 1 μ M SB-431542 and 25 ng/ml of bFGF. Four days later, 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 (251). 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 μ l 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 μ l of transposition reaction mix containing 0.5 μ M of Tn5 transposase (252) (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, Cat#D4003) and eluted in 12 μ l of water. A 65 μ l PCR reaction was setup with 10 μ l of transposed DNA, 0.5 μ M of forward primer Ad1_noMX, 0.5 μ M of multiplexing reverse primer Ad2.x (251), 0.6x SYBR® Green I, and 1x PCR Master Mix (NEB, Cat#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 μ l aliquot was analyzed by qPCR to determine the number of additional cycles needed to avoid amplification saturation as described in (251). The amplified ATAC libraries were purified using a DNA purification kit (Zymo Research, Cat#D4003) and size selected using Agencourt AMPure beads (Beckman Coulter, Cat#63881) (0.55X unbound fraction followed by 1.2X bound fraction). All libraries were sequenced with 75-nucleotide read length paired-end sequencing on an Illumina NextSeq 500 with 30-60 million reads being sequenced for each sample.

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 (253) 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 but devoid of cells) was subtracted. We used a previously reported method to convert the observed light intensity to absolute molecule numbers (234).

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 model for the nuclear size increase during the cell cycle (212) and a reported estimate of the nuclear volume of ES cells (254). To evaluate how cells readjust

their SOX2 levels over time, we used a rank-based autocorrelation (235) 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 (235). As the results using data from 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-sided t-test with unequal variance was performed for the mean SOX2 levels in the cell cycle before cells turn SOX1 positive to evaluate 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 and ChIP-seq analysis

ATAC-seq and ChIP-seq reads were aligned to the mouse reference genome mm10 using STAR 2.5.3a (255) with settings ‘--alignMatesGapMax 2000 --alignIntronMax 1 --alignEndsType EndToEnd --outFilterMultimapNmax 1’. Duplicate reads were removed with Picard (Broad Institute) and reads not mapping to chromosomes 1-19, X, or Y were removed. For each sample, peaks were called with MACS 2.1.1.20160309 (256) with settings ‘-f BAMPE -q 0.01 -g mm’. For ATAC-seq data, peaks from all samples were merged with BEDOPS (257). Peaks overlapping peaks called for ChIP-seq Input data from asynchronous mouse ES cells (GSE89599) were discarded. The HOMER2 (258) function `annotatePeaks.pl` was used with settings ‘-noadj -len 0 -size given’ to count the number of reads for each sample in peaks. TMM-normalization was done with edgeR (259) and analysis of differentially abundant regions with limma (260). 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 SOX2 and OCT4 or OCT4-high/SOX2-high vs OCT4-low/SOX2-low (co-regulated). SOX2 and OCT4 peaks used to determine overlap were taken from GSE87822 and GSE92846 (239, 248). Additional OCT4 ChIP-seq data in Supplementary Figure 3.5K were taken from GSE78073 (130) and GSE56138 (261). H3K4me3 peaks in ES-Bruce4 cells from ENCODE (262) and ES cell super-enhancers from (263) were converted to mm10 using liftOver (264). 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 for both ATAC-seq and ChIP-seq were generated by merging replicate bam files with SAMTools (265) followed by the deepTools 2.4.2 (266) functions `bamCoverage` (with setting ‘--normalizeUsingRPKM’). Average lineplots and ChIP-seq heatmaps were generated using deepTools `computeMatrix` (with setting ‘reference-point’) and `plotHeatmap`. The WT SOX2 ChIP-seq profile in Supplementary Figure 3.1I is based on data from GSE89599 (224). Genome tracks were generated in the UCSC genome browser (267). Motif enrichment was calculated using the HOMER2 function `findMotifs.pl` with setting ‘-size given’. The most frequent known motif corresponding to OCT (OCT4) and SOX (SOX3) family motifs was used. Background enrichment was calculated as the mean of the HOMER-estimated background frequency in all groups. For the random forest model (Supplementary Figure 3.5M and Appendix table 1), we first batch downloaded all peak files from cistromeDB (<http://cistrome.org/db>) (241) in the Mouse_Factor and Mouse_Histone categories. For the groups in Figure 3.4B, we annotated genomic regions overlapping any of the regions in the peak files from cistromeDB and kept only those where Cell_type was “Embryonic Stem Cell”. We generated a data frame in R with the group

belonging of each region and the overlap status (1/0) with the different peak files, consisting 3'725 columns (one for group belonging and 3'724 for the peak identities) and 81'132 rows (one for each region). We used a subset so that each group was represented by the same number of regions, i.e. the lowest number of regions in a group (n=209). These data were split into training (80%, n=1'170) and test (20%, n=293) regions and the training regions were used to train a random forest model. The function `randomForest` was used with the settings 'formula=Group~., mtry=4, ntree=2001', where Group is the column containing the group identity of the region. The model was used on the testing data using the predict function with setting 'type="response"'.

Determination of SOX2 and OCT4 protein half-lives

SBROS cells were seeded at 30,000 cells/cm² on E-cadherin as described above (for SBROS and SBROS overexpressing SOX2-SNAP) or on StemAdhere (Primorigen, Cat#S2071-500UG; for SBROS overexpressing YPet-SOX2-delDBD) in N2B27+2iLIF. Briefly, StemAdhere diluted 1:25 in PBS (with Ca²⁺ and Mg²⁺; Amimed, Cat#3-05K00-I) was added to the culture vessel and incubated for 60 minutes at 37°C. Just before seeding, the StemAdhere solution was removed, the surface of the vessel rinsed once with PBS and filled with the appropriate cell culture medium. After 24 h of cell culture, cells were pulse-labeled with 12 nM of SNAP-SiR 647 or different concentrations of Halo-SiR ligand (see below) (gift from Suliana Manley, EPFL) for 30 min at 37°C. In addition, for OCT4-HALO half-life determination as a function of SOX2-SNAP levels, cells were initially stained with the SNAP-Cell[®] TMR-Star dye (NEB) at a final concentration of 3 µM, and imaged in the Cy3 channel (excitation filter 542/27nm, emission filter 597/45nm). For OCT4-HALO half-life determination upon YPet-SOX2-delDBD overexpression, the YPet channel (excitation filter 513/17nm, emission filter 548/22nm) was initially imaged and only YPet-positive cells were analyzed for Halo-SiR signal decay. Cells were then washed 3x with PBS and incubated in N2B27+2iLIF 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+2iLIF was added. Note that we observed variations in the signal strength of the Halo-SiR that may arise from batch-to-batch differences or loss of signal due to storage. To account for this, we tested several dye concentrations and used the ones that yielded clear exponential decays (too much dye cannot be washed out and the signal does not decay (268)). These concentrations varied between 0.2 nM and 20 nM Halo-SiR. The decay of the fluorescence signal 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, followed by 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 (Figure 3.1E) or the `nls` function in R (Figure 3.2C) (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-SNAP or YPet-SOX2-delDBD overexpression.

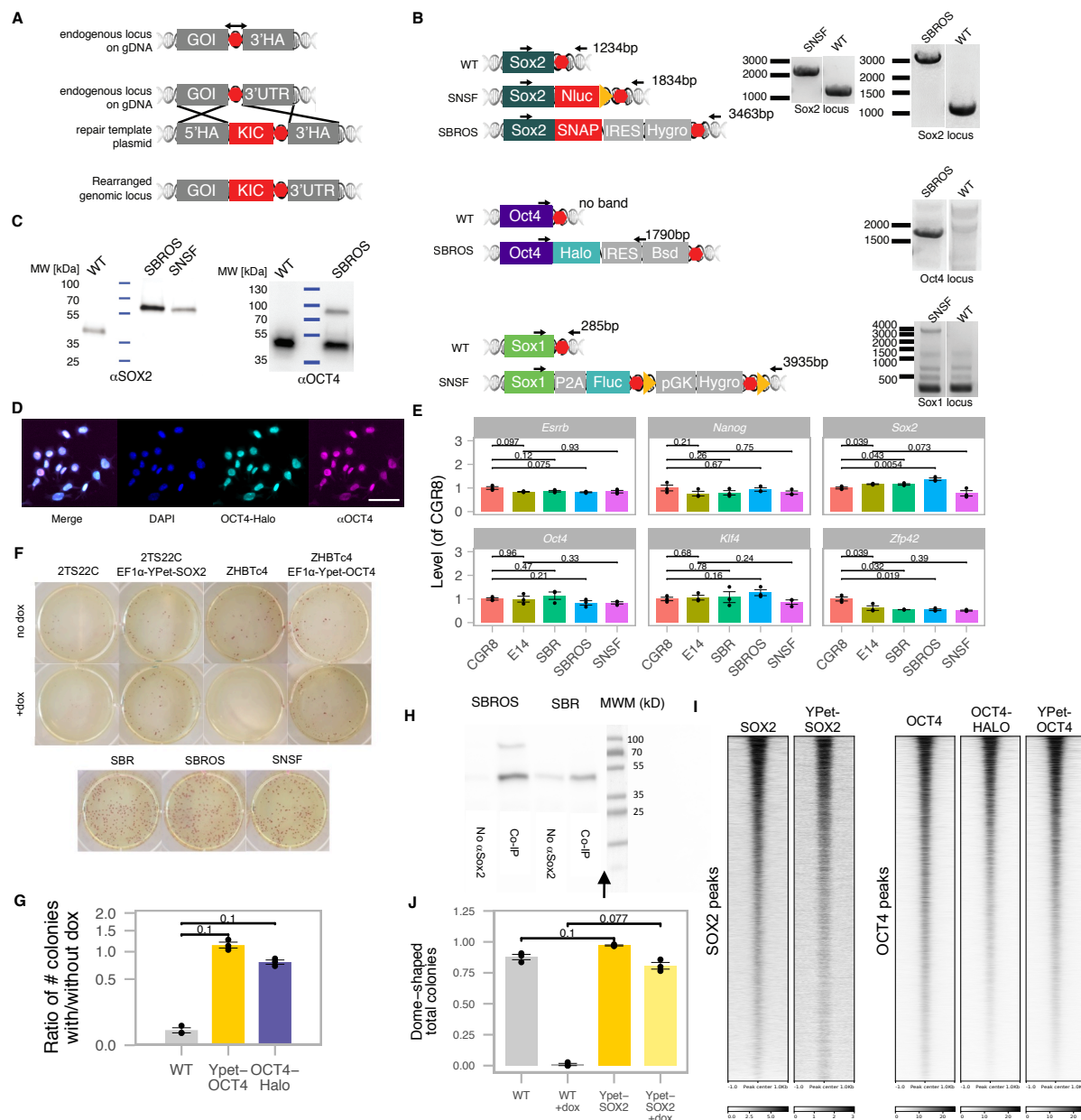
Statistical analysis

Statistical testing was done with either t-tests with unequal variance (Welch) or, when a Shapiro-Wilk's test indicated non-normality (p<0.05), Mann-Whitney U tests, as described in the Figure legends. For Figure 3.3C, two-sided t-tests with unequal variance were performed for each cell cycle progression (time) point. For the autocorrelation functions in Figure 3.2I, L and Supplementary Figure 3.3D the error bars denote the SE estimated by bootstrapping. For Supplementary Figure 3.2K, the p value is based on Pearson correlation.

Data Availability

ATAC-seq and ChIP-seq data from this study have been deposited in GEO (Gene Expression Omnibus) with the accession code GSE126554.

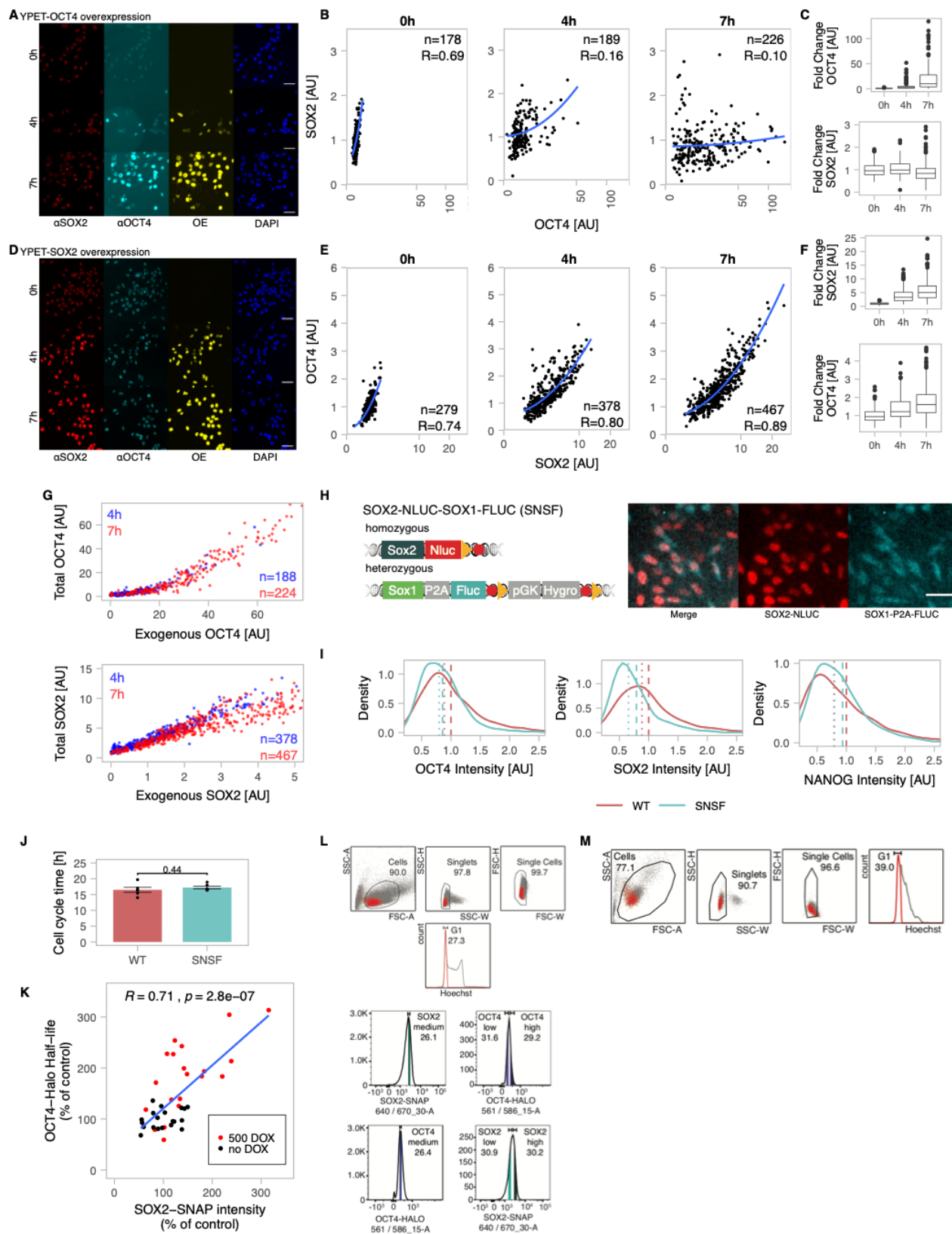
3.6 Supplementary figures



Supplementary Figure 3.1 Validation of knock-in cell lines.

(A) Schematic of the general knock-in strategy used in this study. Red hexagons: STOP codons; KIC: knock-in construct. (B) Location of primers (arrows) and PCR results to verify knocked-in reporters in the SNSF and SBROS cell lines. Red hexagon: STOP codons; Yellow triangles: LoxP sites. (C) Western blot confirming the tagging of endogenous SOX2 and OCT4 in the SNSF and SBROS cell lines. (D) Example of field of view used for the correlation analysis of total OCT4 (αOCT4) and heterozygous OCT4-HALO. Scale bar: 50μm. (E) QPCR analysis of pluripotency marker expression for cell lines used in this study, normalized to Rps9 and to the levels in CGR8 WT mouse ES cells (n=3). p-values: two-sided t-test with unequal variance. (F) Representative images of Alkaline Phosphatase staining in the different cell lines. (G) Ratio of the number of colonies with and without dox, for WT, YPet-OCT4 or OCT4-HALO-expressing Zhbt4 cells (n=3). p-values: Mann-Whitney U test. (H) Co-immunoprecipitation (Co-IP) of OCT4 and OCT4-HALO with SOX2-SNAP (SBROS) and OCT4 with WT SOX2 (SBR). Arrow: Samples and molecular weight marker (MWM) were

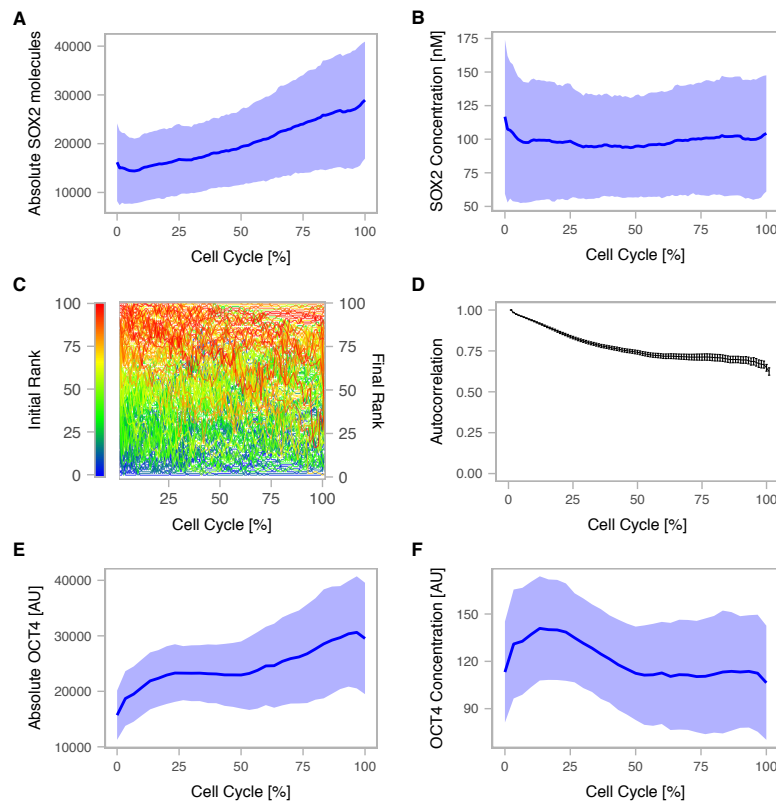
spliced together. (I) Heatmaps for SOX2 (GSE89599) and YPet-SOX2 ChIP-seq signals in SOX2-bound regions and OCT4 (in SBR background), OCT4-HALO and YPet-OCT4 ChIP-seq signals in OCT4-bound regions. (J) Ratio of dome-shaped/flat colonies for WT or YPet-SOX2-expressing 2T522C cells in the absence or presence of dox (n=3). *p*-value: Mann-Whitney U test.



Supplementary Figure 3.2 Cross regulation of SOX2 and OCT4.

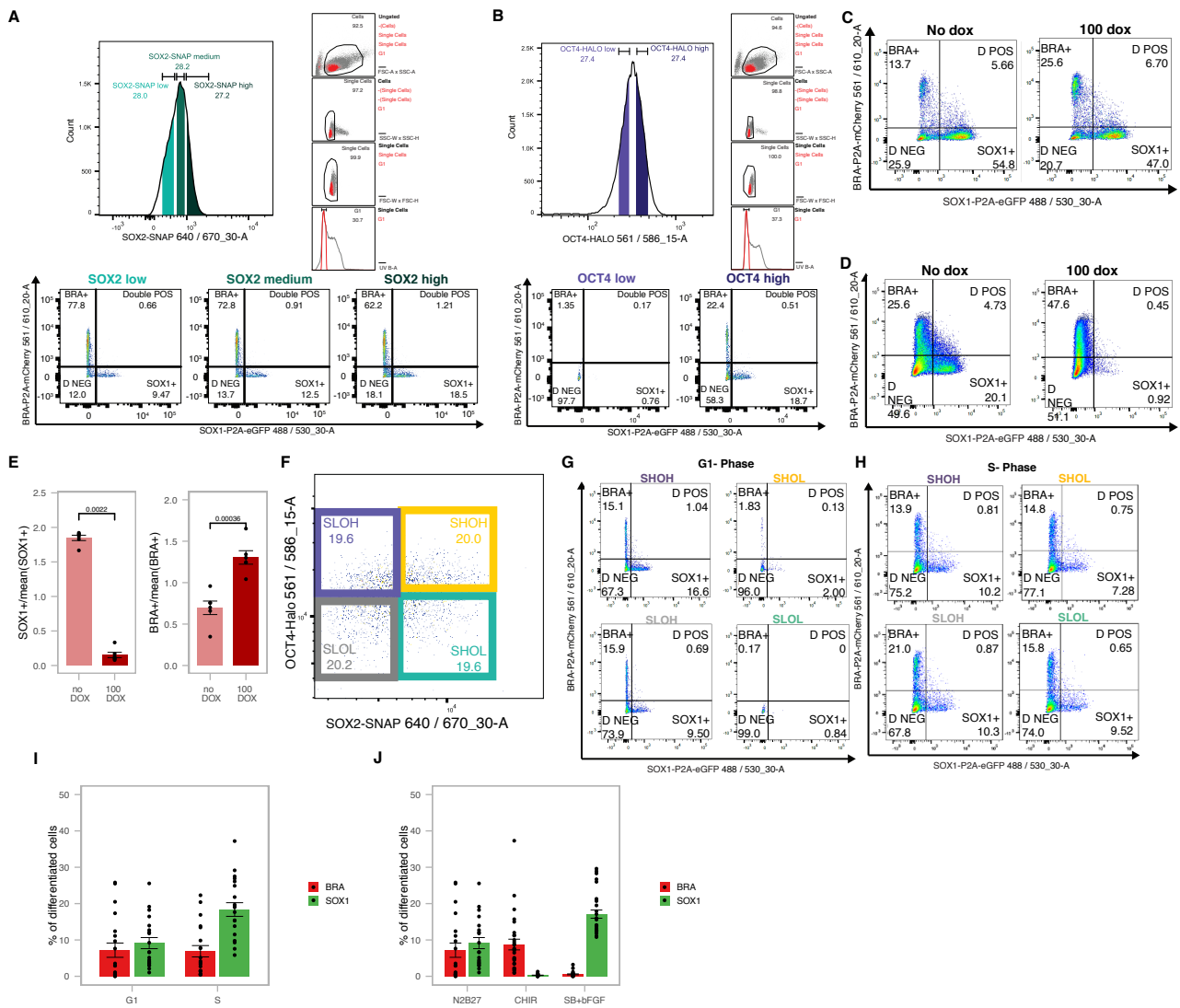
(A-C) YPet-OCT4 overexpression time-course. (A) Fluorescence microscopy images after immunofluorescence staining of SOX2 and OCT4. OE: overexpressed YPet-OCT4. (B-C) Quantifications of

total SOX2 and OCT4 levels in single cells upon YPet-OCT4 overexpression, normalized to the average at 0h. Whiskers: minimum and maximum values; Box: lower and upper quartiles; Solid line: median; Outliers: solid points. (D-F) YPet-SOX2 overexpression time-course. (D) Fluorescence microscopy images after immunofluorescence staining of SOX2 and OCT4. OE: overexpressed YPet-SOX2. (E-F) Quantifications of total SOX2 and OCT4 levels in single cells upon YPet-SOX2 overexpression, normalized to the average at 0h. Whiskers: minimum and maximum values; Box: lower and upper quartiles; Solid line: median; Outliers: solid points. (G) Total OCT4 or SOX2 levels as a function of YPet-OCT4 or YPet-SOX2 overexpression, respectively, 4 hours (blue) and 7 hours (red) after addition of dox. (H) Schematic of the knock-in alleles in the SNSF cell line (Red hexagons: STOP codons; Yellow triangles: LoxP sites) and luminescence microscopy images of differentiating SNSF cells showing the SOX2-NLUC and SOX1-P2A-FLUC signal. (I) Distributions of OCT4 (n=2682), SOX2 (n=1236) and NANOG (n=2416) levels in WT E14 and SNSF cell lines as determined by quantitative immunofluorescence. Dashed lines: mean protein levels; Dotted lines median protein levels. (J) Cell cycle duration of E14 WT (n=6) and SNSF (n=6) cells. Error bars: SE. p-value: two-sided t-test with unequal variance. (K) Correlation between SOX2-SNAP initial fluorescence intensity and OCT4-HALO half-life in single cells induced for SOX2-SNAP overexpression (red) or not (black); values are normalized on the average OCT4-HALO half-life and SOX2-SNAP intensity of the no dox condition. p-value is based on Pearson correlation. (L) Strategy to sort G1 cells for medium endogenous levels of SOX2 and high or low OCT4 levels (top), and conversely (bottom). (M) Raw flow cytometry data of Hoechst profile 8h after sorting as indicated in panel L. Scale-bars: 50 μ m. R is Pearson's correlation coefficient.



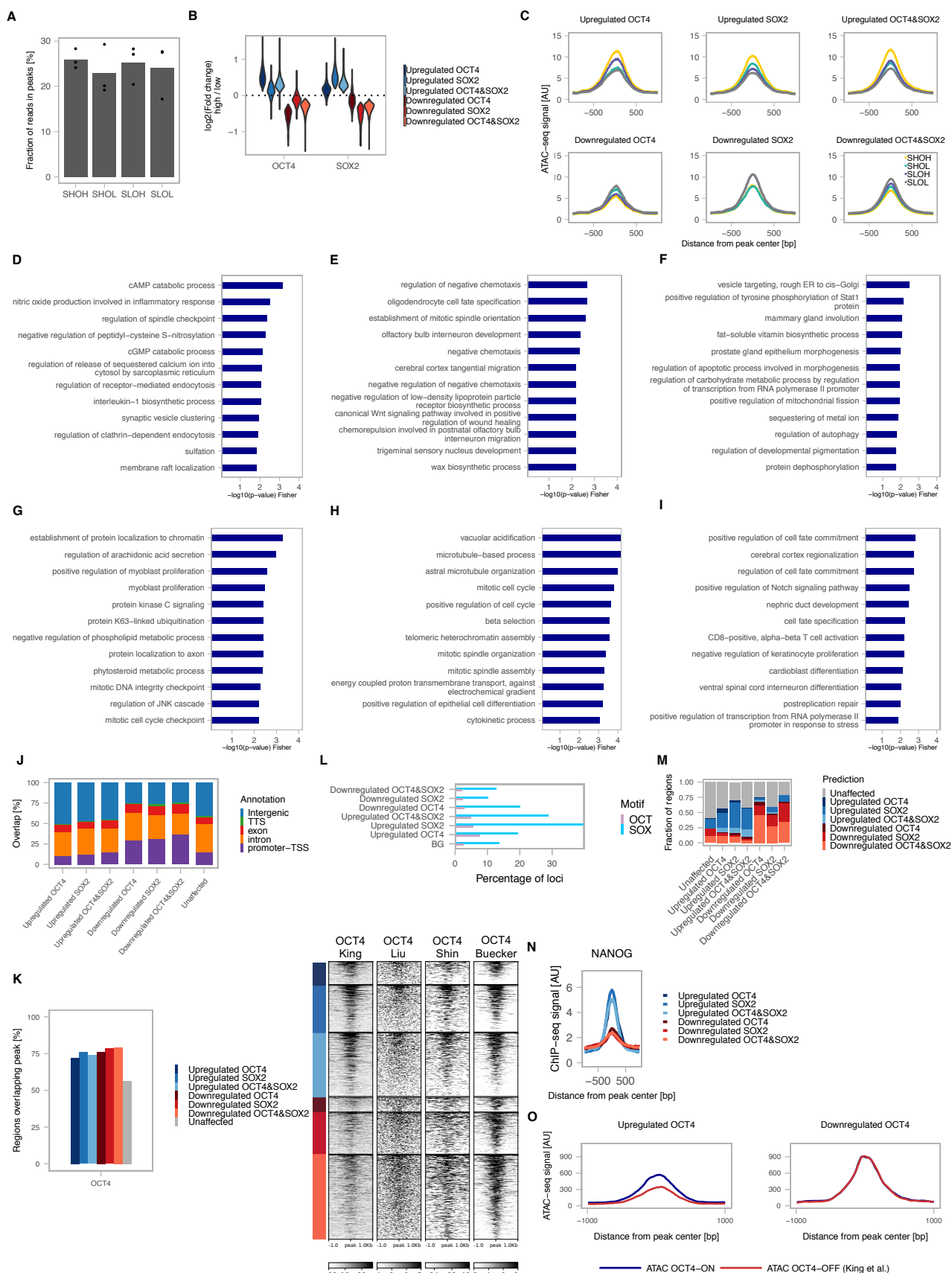
Supplementary Figure 3.3 Fluctuations of SOX2 and OCT4 over the cell cycle.

(A-B) Average SOX2 protein level (A) and concentration (B) in SNSF cells (n=164). Shaded area: SD. (C) Changes of single cell ranks of SOX2 levels over time using data from one cell cycle (n=100). Red: initially high expressing cells; Blue: initially low expressing cells. (D) Rank-based autocorrelation function of the SOX2 ranks. Error bars: SE estimated by bootstrapping. (E-F) Average OCT4 protein level (E) and concentration (F) (n=48) in SBROS cells; shaded area: SD.



Supplementary Figure 3.4 Impact of SOX2 and OCT4 levels on differentiation outcomes.

(A-B) Sorting strategies to evaluate the impact of SOX2 (A) or OCT4 (B) levels on differentiation outcomes and example of flow cytometry data after four days of differentiation. (C) Example of flow cytometry data after four days of differentiation with or without OCT4 overexpression for 12 hours before differentiation. (D) Example of flow cytometry data after four days of differentiation with or without dox throughout differentiation. (E) Differentiation outcome after four days with or without 100 ng/ml dox throughout differentiation. % of SOX1+ and BRA+ cells are normalized to the average values of each experiment (n=6). p-values: Mann-Whitney U test. (F) Sorting strategy for SHOH, SHOL, SLOH and SLOL populations in G1-phase. (G) Example of flow cytometry data of SHOH, SHOL, SLOH and SLOL sorted in G1 phase (see Fig 3I). (H) Example of flow cytometry data of SHOH, SHOL, SLOH and SLOL sorted in S-phase. (I) Percentage of BRA+ and SOX1+ cells after four days of differentiation, when sorted in G1 or S phase (n=5 biological replicates, including SHOH, SHOL, SLOH and SLOL conditions for each replicate both G1 and S). (J) Percentage of BRA+ and SOX1+ cells after four days of differentiation in different conditions, when sorted in G1 phase (n=5 for N2B27, n=7 for CHIR, and n=7 for SB+bFGF biological replicates including SHOH, SHOL, SLOH and SLOL for each replicate).



Supplementary Figure 3.5 ATAC-seq analysis and characterization of differentially accessible regions.

(A) Fraction of reads in ATAC-seq peaks for the different cell populations. (B) Violin plot of the log2 fold-change in OCT4-high vs OCT4-low and SOX2-high vs SOX2-low in different groups of loci. (C) Metaplots of chromatin accessibility for the different cell populations in different groups of loci. (D-H) Top 12 GO terms of

loci in the Upregulated SOX2 (D), Upregulated OCT4&SOX2 (E), Downregulated OCT4 (F) Downregulated SOX2 (G), and Downregulated OCT4&SOX2 (H) groups. (I) Top 12 GO terms for loci in the Upregulated OCT4 group that are bound by OCT4 and with FDR < 5%. (J) Percentage of overlap of regions in the different groups with genomic annotations. (K) Left: Percentage of regions in the different groups overlapping OCT4 peaks from 4 different datasets (GSE87822, GSE92846, GSE78073, and GSE56138). Right: Heatmaps of OCT4 ChIP-seq signals in the different groups from samples in the 4 datasets (OCT4 untreated (King), OCT4 asynchronous (Liu), OCT4 G1 (Shin), and OCT4 2i (Buecker). (L) Fraction of regions in the different groups overlapping OCT and SOX motifs. (M) Prediction of group belonging in the test data based on a trained random forest model (see Materials and Methods). The x-axis shows the real group belonging, and the y-axis shows the percentage of regions predicted to belong to the different regions as colored in the legend. (N) Metaplot of NANOG (GSE87822) binding in the different groups. (O) Changes in accessibility of regions upregulated or downregulated in OCT4-high cells upon OCT4 knock-down (GSE87822).

3.7 Tables

gRNA targets	gRNA sequence (5' to 3')
Oct4	GACTGAGGCACCAGCCCTCCC
Sox2	GCAGCCCTCACATGTGCGACA
Sox1	GACGCACATCTAGCGCCGCG

Table 3.1. Sequences of guide RNAs used to generate knock-in cell lines

Name	Sequence (5' to 3')	Target
Sox1-KI_F2	GTGCCCCTGACGCACAT	In 5' HA
Sox1-KI_R2	CGCTGTGTGCCTCCTCTG	In 3' HA
Seq_Sox2_KI_fw	AGGTGCCGGAGCCCCG	In 5'HA
Sox2_verif_rv_3'	GCATGCTAGCCACAAAGAAA	Downstream of 3' HA
Oct4_KI_verif_fw	GCTCCTCCACCCACCC	Within Oct4
(sf)GFP_verif_fw	CTCGGCATGGACGAGC	In sfGFP
Seqout_IRES_rv	AGACAGGGCCAGGTTTCC	In IRES
COct4_5'_500bp_f	GATCGTCGACTAGCACAAATCCCTTAGCGGT	Upstream of 5' HA

Table 3.2 Primers for knock-in verification

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

Table 3.3 Antibodies used for Western Blotting

Primer name	Primer sequence (5' to 3')
Nanog_qPCR_f	AACCAAAGGATGAAGTGCAAGCGG
Nanog_qPCR_r	TCCAAGTTGGGTTGGTCCAAGTCT
Esrrb_qPCR_f	GCACCTGGGCTCTAGTTGC

Esrrb_qPCR_r	TACAGTCCTCGTAGCTCTTGC
Sox2_qPCR_f	GCACATGAACGGCTGGAGCAACG
Sox2_qPCR_r	TGCTGCGAGTAGGACATGCTGTAGG
Oct4_qPCR_f	GAGGAGTCCCAGGACATGAA
Oct4_qPCR_r	AGATGGTGGTCTGGCTGAAC
Klf4_qPCR_f	CGAACTCACACAGGCGAGAA
Klf4_qPCR_r	CGGAGCGGGCGAATTT
Rps9_qPCR_f	TTGTCGCAAAACCTATGTGACC
Rps9_qPCR_r	GCCGCCTTACGGATCTTGG
Zfp42_qPCR_f	CCCTCGACAGACTGACCCTAA
Zfp42_qPCR_r	TCGGGGCTAATCTCACTTTCAT

Table 3.4 Primers used for QPCR

Target	Species	Dilution	Supplier
Primary Antibodies			
α SOX2	Rabbit	1:200	Invitrogen Cat#48-1400
α OCT4	Mouse	1:50	Cell Signaling Cat#75463
α OCT4	Mouse	1:300	Santa Cruz Cat#sc5279
α NANOG	Rabbit	1:500	Abcam Cat#ab80892
Secondary Antibodies			
α Mouse-IgG-555	Donkey	1:1000	ThermoFisher Cat#A31570
α Mouse-IgG-488	Goat	1:1000	ThermoFisher Cat#A11001
α Mouse-IgG-647	Goat	1:1000	ThermoFisher Cat#A21235
α Rabbit-IgG-647	Chicken	1:1000	ThermoFisher Cat#A21443

Table 3.5 Antibodies used for Immunofluorescence staining

3.8 Acknowledgements

This work was supported by the Swiss National Science Foundation (PP00P3_1144828 and PP00P3_172905) to D.M.S, the Pierre Mercier Foundation, 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. We thank Antonio Meireles-Filho for help with ATAC-seq experiments.

Chapter 4 Dynamic regulation of chromatin accessibility by pluripotency transcription factors across the cell cycle

This chapter represents the following pre-print publication, which is under revision at eLife, formatted to fit with the style of the thesis:

Friman, E.T., Deluz, C., Meireles-Filho, A.C., Govindan, S., Gardeux, V., Deplancke, B. and Suter, D.M. (2019) Dynamic regulation of chromatin accessibility by pluripotency transcription factors across the cell cycle. *bioRxiv*, p.698571

Contribution: In this study that comprises the main contribution from my PhD work, I performed all the computational and bioinformatics analyses and performed most ATAC-seq experiments and part of the ChIP-seq experiments and imaging experiments. Furthermore, I generated the first draft of the manuscript and generated all the figures.

Note that some tables have been omitted from this version of the manuscript. These can be found on the bioRxiv website.

4.1 Abstract

The pioneer activity of transcription factors allows for opening of inaccessible regulatory elements and has been extensively studied in the context of cellular differentiation and reprogramming. In contrast, the function of pioneer activity in self-renewing cell divisions and across the cell cycle is poorly understood. Here we assessed the interplay between OCT4 and SOX2 in controlling chromatin accessibility of mouse embryonic stem cells. We found that OCT4 and SOX2 operate in a largely independent manner even at co-occupied sites, and that their cooperative binding is mostly mediated indirectly through regulation of chromatin accessibility. Controlled protein degradation strategies revealed that the uninterrupted presence of OCT4 is required for post-mitotic re-establishment and interphase maintenance of chromatin accessibility, and that highly OCT4-bound enhancers are particularly vulnerable to transient loss of OCT4 expression. Our study sheds light on the constant pioneer activity required to maintain the dynamic pluripotency regulatory landscape in an accessible state.

4.2 Introduction

Transcription factors (TFs) regulate the expression of genes through interactions with specific DNA sequences located in gene promoters and distal regulatory elements. A minority of TFs display pioneer activity, i.e. they have the ability to bind and induce the opening of nucleosome-occupied chromatin regions, allowing for the subsequent binding of other TFs and co-factors required for transcriptional activation (29, 83, 84). Pioneer TFs thereby play a central role in developmental and reprogramming cell fate decisions, which hinge on large-scale reshaping of the chromatin landscape in tissue-specific regulatory regions (86–88, 91, 269–272). Much less is known about the role of pioneer activity and of its dynamics over the cell cycle in regulating stem cell self-renewal.

The OCT4 (also known as POU5F1) and SOX2 pioneer TFs (87) are absolutely required for the self-renewal of embryonic stem (ES) cells (273, 274). OCT4 and SOX2 can form a heterodimer that binds to a composite motif at thousands of sites in the genome (275–277). A recent study has shown that depletion of OCT4 for 24 hours in ES cells leads to loss of accessibility and co-factor occupancy at a large fraction of its bound enhancers involved in pluripotency maintenance (239). In contrast, the role of SOX2 in the regulation of ES cell chromatin accessibility has not been elucidated. Thus, to which extent the pioneering activities of OCT4 and SOX2 overlap and/or depend on each other to regulate chromatin accessibility in ES cells is unclear.

Self-renewal requires the ability to progress through the cell cycle without losing cell type-specific gene expression. This is not a trivial task since chromatin accessibility of gene regulatory elements

is markedly decreased during S phase and mitosis (51, 72–74). How recovery of chromatin accessibility after DNA replication and mitosis is controlled, and whether it requires pioneer activity is poorly understood. The period of genome reactivation occurring at the mitosis-G1 (M-G1) transition coincides with a particularly favorable context for reprogramming by somatic cell nuclear transfer (mitosis) (63) and increased sensitivity to differentiation signals in human ES cells (G1 phase) (236). Recent evidence also points at cell cycle stage-specific functions of OCT4 and SOX2 in cell fate regulation. OCT4 expression levels in G1 phase affect the propensity of ES cells to differentiate towards neuroectoderm and mesendoderm (278), and depletion of OCT4 at the M-G1 transition impairs pluripotency maintenance of ES cells and leads to a lower reprogramming efficiency upon overexpression in mouse embryonic fibroblasts (248). Depletion of SOX2 at the M-G1 transition impairs both pluripotency maintenance and SOX2-induced neuroectodermal differentiation of ES cells upon release of pluripotency signals (279). Whether the particular sensitivity of M and G1 phases to the action of OCT4 and SOX2 is related to the dynamics of their pioneer activity across the cell cycle is unknown.

Here we studied the interplay of OCT4 and SOX2 in regulating chromatin accessibility of ES cells and dissected the pioneer activity of OCT4 across the cell cycle. We show that most enhancers bound by both TFs depend on only one of them to maintain their open chromatin state, and that cooperative binding of OCT4 and SOX2 is mainly mediated indirectly through changes in chromatin accessibility. Using forms of OCT4 engineered for mitotic or auxin-inducible degradation, we demonstrate the role of OCT4 in re-establishment and continuous maintenance of chromatin accessibility throughout the cell cycle.

4.3 Results

4.3.1 OCT4 and SOX2 regulate chromatin accessibility at mostly distinct loci

OCT4 and SOX2 bind cooperatively to thousands of genomic locations in ES cells both independently and as a heterodimer on a composite OCT4::SOX2 motif. How OCT4 and SOX2 interplay to regulate chromatin accessibility in ES cells is not known. To address this question, we decided to determine genome-wide chromatin accessibility changes upon acute loss of OCT4 or SOX2. To deplete OCT4 and SOX2 from ES cells in an inducible manner, we took advantage of the ZHBTc4 (274) and 2TS22C (273) mouse ES cell lines, in which a Tet-Off promoter regulates the expression of *Oct4* and *Sox2*, respectively (Figure 4.1A). While OCT4 is fully depleted after 24 hours of doxycycline (dox) (274), SOX2 is not, likely due to its longer half-life (273). We determined SOX2 levels by immunofluorescence staining after 26 and 40 hours of dox treatment and found that residual SOX2 expression persisted after 26 hours but not 40 hours of dox treatment (Supplementary Figure 4.1A). Importantly, despite its known role in regulating expression of OCT4 (278, 280), SOX2 depletion for 26 or 40 hours had only a minor impact on OCT4 levels (Supplementary Figure 4.1A-B). In ZHBTc4 cells, as expected 24 hours of dox treatment led to the complete depletion of OCT4 and only mildly affected SOX2 levels (Supplementary Figure 4.1C-D). Therefore, changes in chromatin accessibility upon short-term SOX2 or OCT4 loss are unlikely to be confounded by changes in expression levels of OCT4 and SOX2, respectively.

We performed ATAC-seq in ZHBTc4 cells without dox or with dox for 24 hours, and in 2TS22C cells without dox or with dox for 26 or 40 hours. We first compared chromatin accessibility changes between ZHBTc4 cells +/- dox for 24 hours in our culture conditions (serum + LIF + 2i (S2iL), see Materials and methods) to a previous dataset acquired with ZHBTc4 cells +/- dox for 24 hours but cultured in serum + LIF (SL) (239). The good correlation (Pearson's $R=0.7$) in chromatin accessibility changes at OCT4 binding sites between culture conditions (Supplementary Figure 4.1E) prompted us to take advantage of both datasets for further analysis. We next compared changes in accessibility at SOX2 binding sites in the 2TS22C cell line treated for either 26 or 40 hours with dox, which also displayed a clear correlation (Pearson's $R=0.61$) (Supplementary Figure 4.1F). We reasoned that the 26 hour dox dataset should be less prone to changes in accessibility due to indirect effects of prolonged SOX2 depletion than the 40 hour dox dataset, while the latter should be more sensitive to identify loci that are still accessible at low SOX2 concentrations. We thus called significantly affected loci using limma (281) (false discovery rate (FDR) < 0.05) and selected only those in which the direction of change (decrease or increase in accessibility) was the same for 26

hours vs 40 hours of dox treatment in 2TS22C cells, and likewise for SL vs S2iL in ZHBTc4 cells. In line with previous reports, loss of OCT4 led to decreased accessibility at 20'587 loci, most of which are distal regulatory elements (Figure 4.1B). Loss of SOX2 also led to decreased accessibility mainly at distal elements, but at fewer loci (7'874). We also found that loss of OCT4 led to a gain in accessibility at 20'209 loci, while 1'080 loci gained accessibility upon loss of SOX2 (Figure 4.1B). Loci that lost accessibility were highly enriched for OCT4 and SOX2 ChIP-seq binding while loci that gained accessibility were much less so (Supplementary Figure 4.2A-B).

To compare the loci impacted by OCT4 vs SOX2 depletion, we next focused on all regions that were bound by OCT4 and/or SOX2 as identified from available and newly generated ChIP-seq datasets (see Supplementary Figure 4.2A-B and Methods) and that lost accessibility upon dox treatment. To avoid misrepresenting differences in SOX2 and OCT4 regulation that arise from differences in accessibility due to culture conditions or cell lines, we called significantly different loci (FDR < 0.05) between untreated ZHBTc4 cells cultured in SL vs S2iL conditions as well as between untreated ZHBTc4 cells and 2TS22C cells in S2iL. We then discarded all loci that displayed a large difference (FC > 4) in any of those comparisons. We classified the remaining loci as OCT4-dependent (OD, 8'869 loci), SOX2-dependent (SD, 1'834 loci), and co-dependent (CD, 2'973 loci), as defined by loss of accessibility upon depletion of OCT4 only, SOX2 only, or either of them, respectively (Supplementary Figure 4.3A, Figure 4.1C-E). All three groups were enriched for chromatin marks of enhancers (Supplementary Figure 4.3B). We performed ChIP-seq analysis of the active enhancer mark H3K27ac (282) upon OCT4 or SOX2 loss for 24 hours and 26 hours, respectively. All groups displayed a reduction in H3K27ac, suggesting concordant maintenance of enhancer accessibility and activity by OCT4 and/or SOX2 at these loci (Figure 4.1F-G).

Surprisingly, all groups were enriched for the binding of both OCT4 and SOX2 (Figure 4.2A). 89% of SD sites overlapped with an OCT4 peak and 65% of OD sites overlapped with a SOX2 peak. Therefore, differences in the regulation of chromatin accessibility at these loci cannot simply be explained by differential DNA binding of SOX2 and OCT4. OCT4 has been shown to regulate chromatin accessibility by recruitment of the BAF chromatin remodeling complex, including the BRG1 subunit (239). As SOX2 also interacts with BRG1 in vivo (112), we asked whether SOX2 also regulates chromatin accessibility through BRG1 recruitment. We performed BRG1 ChIP-seq upon SOX2 depletion and reanalyzed ChIP-seq data of BRG1 upon OCT4 depletion (239). We found that loss of accessibility was accompanied by loss of BRG1 in all groups (Figure 4.2B-C). We also reanalyzed ATAC-seq data from cells in which BRG1 has been depleted (239, 283) and found that all groups were dependent on BRG1 to maintain their accessibility (Figure 4.2 – figure supplement 1A). This suggests that OCT4 and SOX2 can regulate chromatin accessibility independently of each other even at sites that are co-occupied and through the recruitment of BRG1.

To understand which features distinguish OD, SD, and CD loci, we performed motif analysis on the underlying sequences. While both OD and CD loci were strongly enriched for the OCT4::SOX2 canonical motif and the OCT motif, SD loci were more enriched for the SOX motif (Figure 4.2D-F and Appendix table 2). SD sites were also enriched for the AP-2 motif (Figure 4.2 – figure supplement 1B). TFAP2C, a member of the AP-2 family, is known to regulate differentiation into trophoblast stem (TS) cells together with SOX2 (117). Interestingly, when reanalyzing data from TS cells (117, 284) we found SD sites to be highly SOX2-bound and accessible compared to OD and CD loci (Figure 4.2 – figure supplement 1C-D). Furthermore, SD loci were enriched near genes that increased in nascent mRNA expression upon loss of OCT4 (239) (Figure 4.2 – figure supplement 1E), which by itself leads to TS cell differentiation (117). In contrast, OD and CD loci were enriched near genes that decreased in nascent mRNA expression upon OCT4 depletion (Figure 4.2 – figure supplement 1E). We next aimed to determine the fraction of pluripotency-associated enhancers falling in the OD, SD, and CD groups. To this end, we checked for enrichment of the nearest gene in three gene ontology (GO) sets associated specifically with pluripotency. We found that OD loci were most enriched in all three GO sets (Figure 4.2G-I). We also analyzed the binding profiles of other pluripotency TFs (ESRRB, NANOG, KLF4, SALL4) (269, 285–287) and found an enrichment in the CD group, although all these TFs bound to some extent to all groups of loci (Figure 4.2 – figure supplement 1F). Notably, all groups were also enriched for the “cell differentiation” GO term (Figure 4.2 – figure supplement 1G), in line with the role of OCT4 and SOX2 in ES cell differentiation. Since SOX2 was shown to require PARP1 to bind to a subset of genomic regions in ES cells (288), we

asked whether PARP1 dependence could explain the differential regulation of chromatin accessibility between these groups. We thus reanalyzed data from wt and PARP1 knockout (KO) ES cells (289, 290), and found a reduction of SOX2 binding in PARP1 KO cells at OD, CD, and SD loci (Figure 4.2 – figure supplement 1H). Thus, PARP1 dependence cannot explain the differential regulation of chromatin accessibility between OD, CD, and SD loci. Overall, these results indicate that OCT4 and SOX2 regulate partially independent sets of pluripotency and differentiation enhancers, with OCT4 having the largest influence on chromatin accessibility of pluripotency-associated regulatory elements.

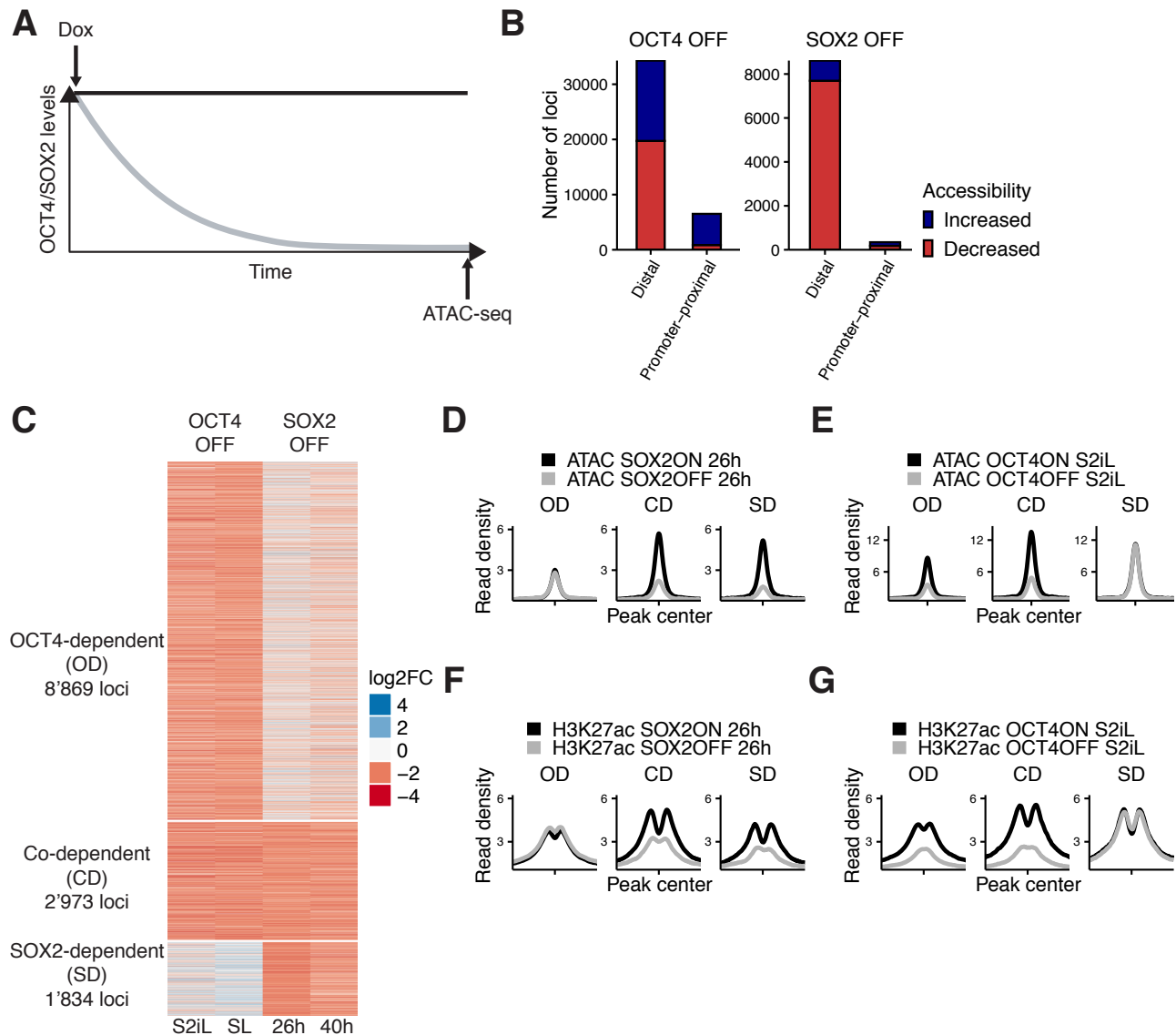


Figure 4.1 Interplay between OCT4 and SOX2 in regulating ES cell chromatin accessibility.

(A) Experimental strategy to compare the effect of OCT4 and SOX2 depletion on chromatin accessibility. (B) Number of regions significantly changed in accessibility upon OCT4 (left) and SOX2 (right) depletion in distal (>1 kb from TSS) and promoter-proximal (<1 kb from TSS) elements. (C) log2 fold-change values of accessibility between dox-treated and untreated cells upon OCT4/SOX2 depletion at OCT4/SOX2 binding sites with decreased accessibility. Loci are grouped into those significantly affected upon OCT4 depletion (OD), SOX2 depletion (SD), or depletion of either factor (CD). Each row corresponds to one individual locus, and each column to a different experimental condition. (D-E) Average RPKM-normalized ATAC-seq signal 2 kb around OD, CD, and SD loci upon SOX2 depletion (D) and OCT4 depletion (E). (F-G) Average RPKM-normalized H3K27ac ChIP-seq signal 2 kb around OD, CD, and SD loci upon SOX2 depletion (F) and OCT4 depletion (G).

4.3.2 Cooperative binding between OCT4 and SOX2 is mainly mediated indirectly through changes in chromatin accessibility

Several lines of evidence suggest that OCT4 and SOX2 exhibit cooperative DNA binding. In vitro electrophoretic mobility shift assays and fluorescence correlation spectroscopy experiments have shown that OCT4 and SOX2 display enhanced binding to the OCT4::SOX2 motif when binding together (109, 291). While in vitro experiments reported OCT4-assisted binding on a purified nucleosomal template (292), single-molecule imaging in live ES cells (293) and ChIP-seq analysis of OCT4 in the presence or absence of SOX2 in fibroblasts (97) have provided evidence that SOX2 assists OCT4 binding in vivo. However, while these experiments suggest that OCT4 and SOX2 can display direct cooperativity, the role this mechanism plays in their colocalization in the complex in vivo chromatin and nuclear environment is unclear. We reasoned that the independent regulation of chromatin accessibility by OCT4 and SOX2 at a large number of loci could result in indirect cooperativity, i.e. each TF could assist the binding of the other through increasing chromatin accessibility. In line with this hypothesis, it was previously shown that upon loss of OCT4, SOX2 binding loss is correlated to the loss in chromatin accessibility (239). However, since most of the in vivo evidence points at a role for SOX2 in mediating cooperative OCT4 DNA-binding rather than vice versa (97, 293), we interrogated genome-wide binding of OCT4 upon loss of SOX2 using ChIP-seq in 2T522C cells treated with dox for 26 hours. We found that changes in OCT4 binding were also highly correlated to changes in chromatin accessibility upon SOX2 loss (Pearson's $R=0.77$) (Figure 4.2 – figure supplement 2A). We next analyzed OCT4 and SOX2 binding in the presence or absence of SOX2 and OCT4, respectively, at OD, CD, and SD loci. We found that OCT4 binding was only slightly decreased at OD sites in the absence of SOX2, while SOX2 binding at SD sites was mildly increased in the absence of OCT4 (Figure 4.2J-K). These findings were also consistent when narrowing down our analysis to sites containing a canonical OCT4::SOX2 motif (Figure 4.2 – figure supplement 2B-E). The slight loss of OCT4 binding at OD sites in the absence of changes in accessibility suggests that other mechanisms such as recruitment by SOX2 may play a role in the binding of OCT4, in line with SOX2 enhancing OCT4 binding.

Upon loss of its partner protein, OCT4 loses binding at 10'264 loci and gains binding at 1'153 loci, while SOX2 loses binding at 7'610 loci and gains binding at 4'423 loci. This indicates that the ability of OCT4 to occupy its specific binding sites is more impacted by the absence of SOX2 than vice-versa, and that SOX2 can get rerouted to new loci in the absence of OCT4. We further noticed that loci gaining accessibility upon loss of OCT4, which are highly enriched for differentiation terms (Supplementary Figure 4.4G), also gained binding by SOX2 (see Supplementary Figure 4.2A columns 6-7 bottom half) and were enriched for the SOX and AP-2 motifs (Appendix table 3). 3'484 loci displayed a significant increase in both accessibility and SOX2 binding. Interestingly, these loci decreased their accessibility upon SOX2 loss (Supplementary Figure 4.5F) and gained BRG1 occupancy concomitantly with OCT4 loss (Supplementary Figure 4.5G), in line with SOX2 recruiting the BAF complex to promote chromatin opening. This may suggest that OCT4 sequesters SOX2 to OCT4-SOX2 sites, and upon OCT4 loss SOX2 is free to bind and increase the accessibility of differentiation-associated regulatory elements. Overall, these results indicate that cooperative binding of OCT4 and SOX2 in ES cells is mainly mediated indirectly through changes in chromatin accessibility. However, while SOX2 enhances OCT4 binding in general, the presence of OCT4 reroutes SOX2 to pluripotency loci.

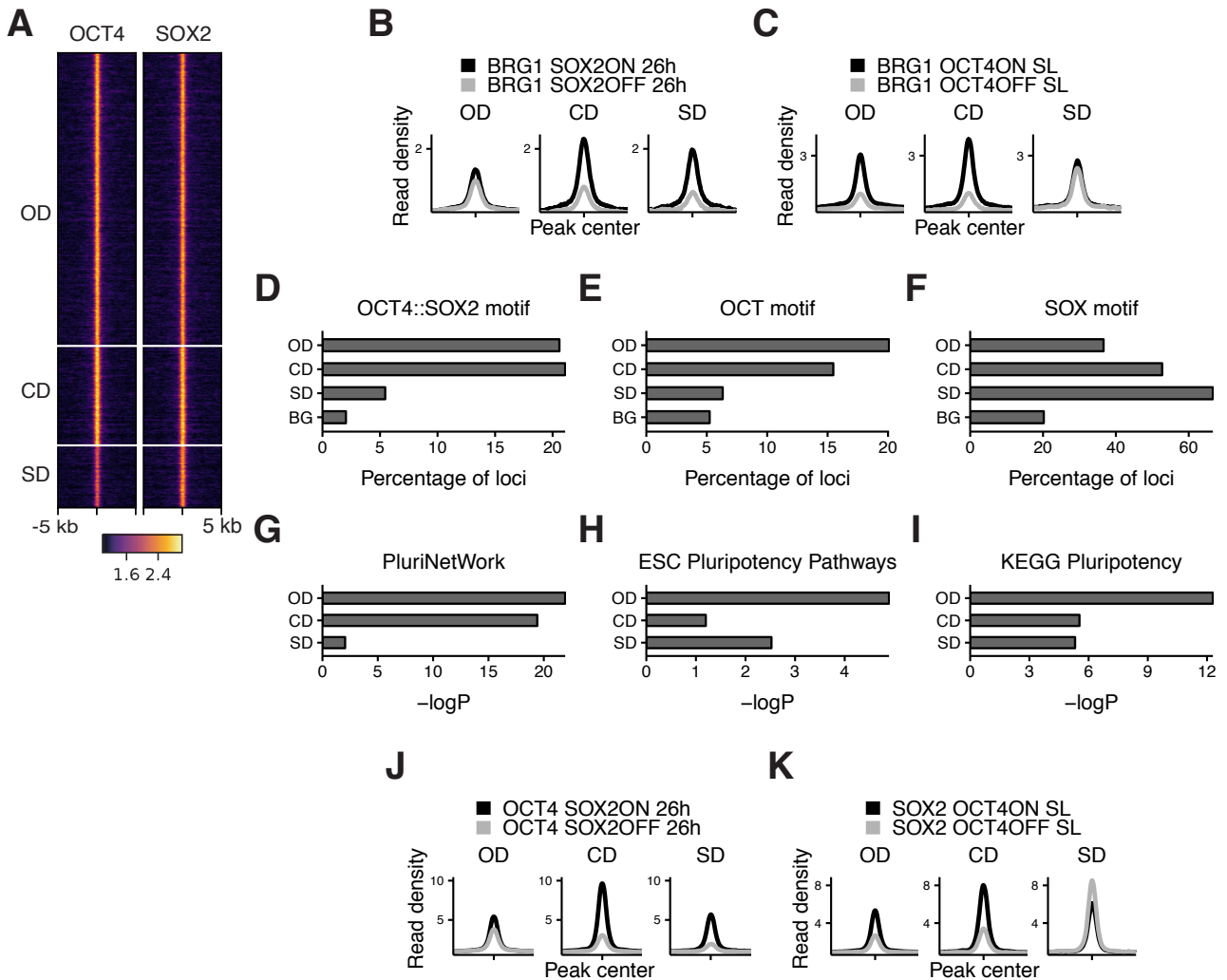


Figure 4.2 Characterization of OCT4/SOX2-dependent loci.

(A) Heatmap of RPKM-normalized OCT4 and SOX2 ChIP-seq binding profiles in untreated ZHBTc4 cells at OD, CD, and SD loci. Each row represents one individual locus. (B-C) Average RPKM-normalized BRG1 ChIP-seq signal 2 kb around OD, CD, and SD loci upon SOX2 depletion (B) and OCT4 depletion (C). (D-F) Frequency of overlap with motifs at OD, CD, and SD loci as well as in background regions (BG) for the canonical OCT4::SOX2 motif (D), the OCT motif (E), and the SOX motif (F). (G-I) Enrichment ($-\log(p)$) values for the closest gene in the OD, CD, and SD groups in the gene ontology sets PluriNetWork (G), ESC Pluripotency Pathways (H), and the KEGG gene set "Signaling pathways regulating pluripotency" (I). (J-K) Average RPKM-normalized OCT4 (J) and SOX2 (K) ChIP-seq signal 2 kb around OD, CD, and SD loci upon SOX2 depletion (J) and OCT4 depletion (K).

4.3.3 OCT4 is required at the M-G1 transition to re-establish enhancer accessibility

Transient depletion of OCT4 or SOX2 at the M-G1 transition has been shown to hinder pluripotency maintenance (248, 279), but the underlying mechanisms are not known. This time window coincides with enhancer reopening upon chromatin decompaction, but whether pioneer factors are involved in this process is not clear. As we found OCT4 to have the broadest influence on accessibility of pluripotency-associated loci, we focused on its role in regulating chromatin accessibility at the M-G1 transition. To allow near-complete loss of OCT4 at the M-G1 transition, we used ZHBTc4 cells constitutively expressing OCT4 fused to a SNAP-tag and a Cyclin B1 degron (mitotic degron; MD) or a mutated version thereof (MD*; Figure 4.3A), which have been described previously (96). Importantly, lower than wildtype levels of OCT4 have been shown to sustain or even enhance pluripotency (119, 120). We thus reasoned that OCT4 levels need to decrease below a certain threshold to impact chromatin accessibility of pluripotency regulatory elements. Furthermore, the MD strategy strongly decreases but does not fully eliminate the target protein (248, 279). We therefore expressed MD-OCT4 and MD*-OCT4 at lower than wildtype levels from the constitutively active but

relatively weak PGK promoter. After lentiviral transduction of the constructs, we stained cells with the SNAP-Cell 647-SiR dye (294) and sorted for a narrow window of SNAP expression to obtain the same average level of OCT4 tagged with MD and MD* across the cell cycle, as described previously (279) (Supplementary Figure 4.6A). We also transduced cells to express YPet-MD in a constitutive manner, which allows for discrimination between cells in early G1 (YPet-negative) and late G1 phase (YPet-positive). In combination with Hoechst staining, this enables sorting cells in early G1 (EG1), late G1 (LG1), S, and late S/G2 (SG2) phase as described previously (96) (Supplementary Figure 4.6B). SNAP-MD-OCT4 levels were correlated with YPet-MD levels in flow cytometry, indicating that OCT4 levels are restored in LG1 in MD-OCT4 cells (Supplementary Figure 4.6C), as shown previously (248). In the absence of dox, these cell lines display no substantial difference in chromatin accessibility at OCT4-regulated loci (Supplementary Figure 4.6D). When grown in the presence of dox, MD*-OCT4 cells maintain a higher fraction of dome-shaped colonies, higher alkaline phosphatase activity, higher expression of pluripotency markers and lower expression of differentiation markers (Figure 4.3- figure supplement 1E-G) than MD-OCT4 cells, as also shown previously (248).

To test whether depletion of OCT4 at the M-G1 transition affects chromatin accessibility, we treated cells with dox for 40 hours to ensure that all cells have gone through at least one cell division expressing only MD or MD*-tagged OCT4. Note that dox-treated cells had a longer G1 phase as compared to wt ES cells, as shown before to be a consequence of lower than wt OCT4 levels (295). However, there was no difference in the proportion of cells in the different cell cycle phases between MD-OCT4 and MD*-OCT4 (Supplementary Figure 4.6H). We sorted cells in EG1, LG1, S, and SG2 phases, performed ATAC-seq, and compared the accessibility between MD-OCT4 and MD*-OCT4 cells at each cell cycle phase. OCT4-regulated loci that increased or decreased in accessibility upon complete OCT4 depletion (see Figure 4.1B) were also affected by transient M-G1 degradation (Figure 4.3B-C, Supplementary Figure 4.6I-J). This shows that OCT4 is required at the M-G1 transition to restore chromatin accessibility and that loci gaining accessibility upon OCT4 loss are also dynamically regulated by OCT4 levels.

To characterize the different dynamic behaviors of chromatin accessibility changes across the cell cycle, we used k-means clustering on the change in accessibility between MD-OCT4 and MD*-OCT4 cells at all accessible loci displaying an OCT4 ChIP-seq peak (Figure 4.3D). Two clusters displayed decreased accessibility in EG1 and recovered their accessibility incompletely (cluster 1) or completely (cluster 2) over the cell cycle. Notably, cluster 2 loci were less affected in EG1 than cluster 1 loci, which likely explains their complete recovery. Cluster 3 loci were characterized by a minor decrease in accessibility but that persisted over the cell cycle, and cluster 4 loci were unaffected by OCT4 loss. In contrast to clusters 1-3, cluster 4 was enriched near TSSs (Supplementary Figure 4.7A) and for the H3K4me3 promoter mark (296) (Supplementary Figure 4.7B), in line with OCT4 generally not affecting accessibility at promoters (Figure 4.1B and (239)). However, cluster 4 also contained many loci enriched for active enhancer marks (H3K4me1 and H3K27ac) (297, 298), similar to clusters 1-3 (Supplementary Figure 4.7B). To test whether active histone marks also acutely change upon rapid loss of OCT4, we performed ChIP-seq for H3K27ac across the cell cycle in cells expressing MD-OCT4 or MD*-OCT4. The difference in H3K27ac across the cell cycle between the SNAP-MD-OCT4 and SNAP-MD*-OCT4 cell lines mimicked the corresponding changes in accessibility, although with smaller amplitude (Supplementary Figure 4.7C-D), suggesting that this modification is also highly dynamic and sensitive to OCT4 levels. We analyzed the fraction of regions in the different clusters overlapping previously annotated ES cell super-enhancers (SEs) and “typical” enhancers (TEs) (299). We found these to be enriched in all clusters compared to non-OCT4 bound regions, with slightly more enrichment in clusters 1 and 3 for both SEs and TEs. This suggests that a large fraction of both SEs and TEs are permanently affected by the transient loss of OCT4 at the M-G1 transition (Supplementary Figure 4.7E).

As mentioned before, pluripotency was shown to be maintained at lower than wild type OCT4 expression levels. To ask whether chromatin accessibility of the observed clusters was OCT4 level-dependent within a higher OCT4 concentration range, we interrogated chromatin accessibility in the context of physiological variations of OCT4 levels. To do so, we took advantage of ATAC-seq data we previously acquired on cells differing in their OCT4 levels by a factor of ~2, due to temporal fluctuations in their endogenous levels (278). We compared chromatin accessibility of the different

clusters we identified for cells expressing high versus low endogenous levels of OCT4 and found virtually no differences in chromatin accessibility between these groups across all clusters (Supplementary Figure 4.7F), consistent with the ability of moderately low OCT4 levels to fully sustain pluripotency.

To understand the reason for the differential impact of transient OCT4 depletion on chromatin accessibility, we performed motif search analysis and compared OCT4 binding at the different clusters. We found a higher enrichment for the canonical OCT4::SOX2 motif (Figure 4.3E) and a higher OCT4 occupancy (Figure 4.3F) at cluster 1 loci. Consistently, cluster 1 contained mostly OD and CD loci identified above (Supplementary Figure 4.7G). We did not find strong differential enrichment for other motifs that could explain the differential regulation of these loci (Appendix table 4). As high OCT4 binding was a signature of the loci most sensitive to transient OCT4 loss, we next aimed to determine the relationship between OCT4 binding and chromatin accessibility. We compared chromatin accessibility in ZHBTc4 cells in the presence or absence of OCT4 in conditions with matched OCT4 ChIP-seq and ATAC-seq data (239). The OCT4 ChIP-seq signal was correlated to loss of accessibility upon OCT4 depletion (Supplementary Figure 4.8A) as shown previously, but also to chromatin accessibility in untreated cells (Supplementary Figure 4.8B), indicating that strong OCT4 binding sites are both highly accessible and sensitive to OCT4 levels. Taken together, these results reveal different classes of OCT4-bound loci that show different cell cycle accessibility dynamics upon OCT4 loss at the M-G1 transition, and that highly bound sites are particularly accessible and sensitive to OCT4 loss for the maintenance of their accessibility and H3K27 acetylation.

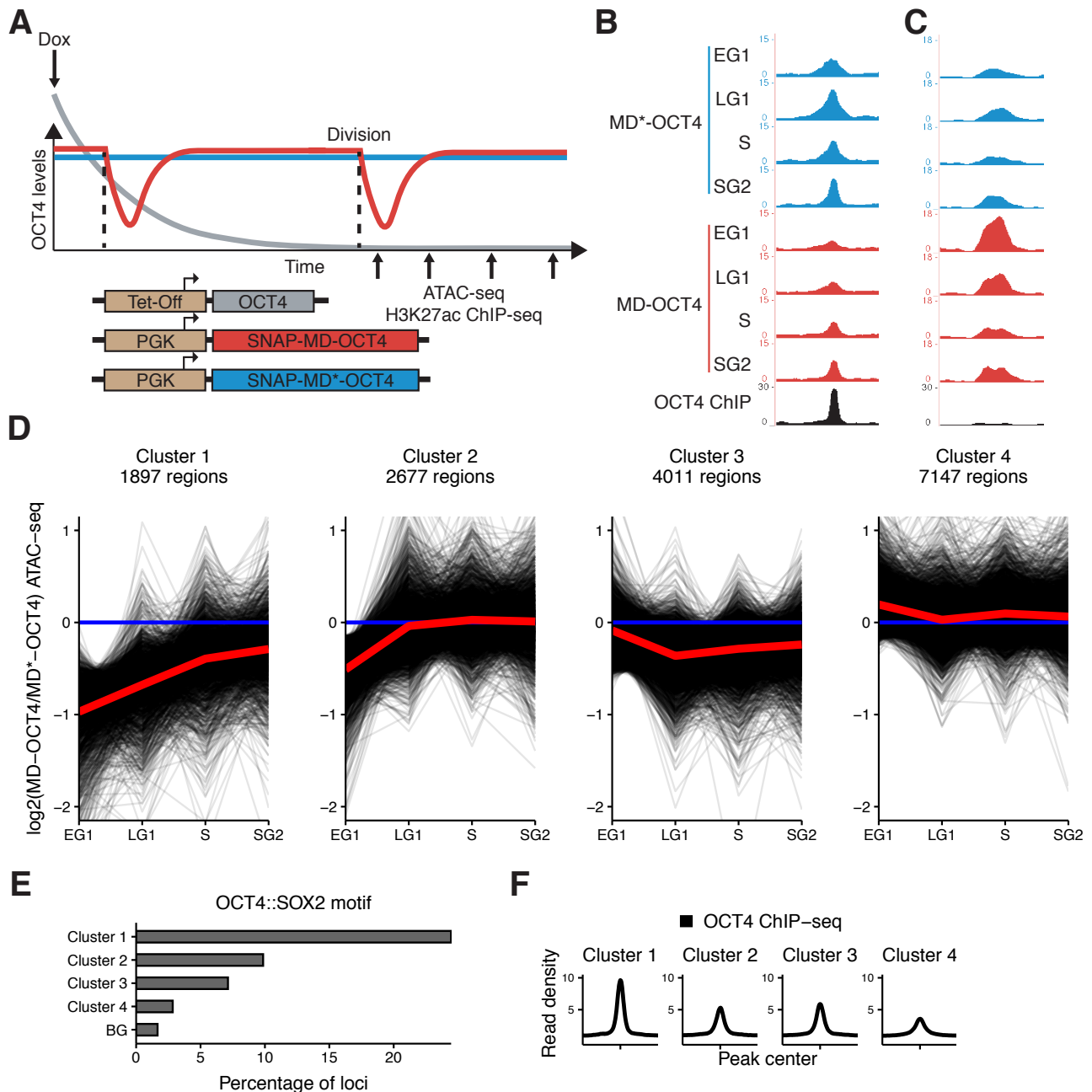


Figure 4.3 Mitotic degradation of OCT4 results in different patterns of accessibility loss.

(A) Experimental strategy used to assess the impact of OCT4 depletion at the M-G1 transition. (B-C) Genome browser tracks of RPKM-normalized accessibility profiles across the cell cycle for one locus that decreases (B) at chr11:6894809-6895533 and one that increases (C) at chr9:41247953-4124841 in accessibility upon transient OCT4 depletion in M-G1. (D) log2 fold-change values of accessibility between MD-OCT4 and MD*-OCT4 (control) cells in different cell cycle phases at all accessible OCT4-bound sites, grouped into four clusters by k-means clustering (see Materials and methods). Each line represents one locus. Red line: mean of each cluster. (E) Frequency of overlap with the canonical OCT4::SOX2 motif in the four clusters as well as in background regions (BG). (F) Average RPKM-normalized OCT4 ChIP-seq signal in untreated ZHBTc4 cells 2 kb around loci in the four clusters.

4.3.4 OCT4 is required throughout the cell cycle to maintain enhancer accessibility

We next asked whether OCT4 also plays a role in maintaining enhancer accessibility in other cell cycle phases. To do so, we generated a cell line allowing drug-inducible degradation of OCT4. Briefly, we used lentiviral vectors to constitutively express the Tir1 ubiquitin ligase (allowing Auxin-inducible ubiquitination and degradation of target proteins (300, 301)) and OCT4 fused to mCherry and an Auxin-inducible degron tag (302, 303) (mCherry-OCT4-AID) in ZHBTc4 cells (Figure 4.4A).

To ensure near-complete OCT4 depletion, we expressed OCT4-AID at low levels using the PGK promoter. We further expressed YPet-MD in this cell line to allow for cell sorting in different cell cycle phases, as described above. Upon addition of indole-3-acetic acid (IAA, also known as Auxin), the AID-tagged OCT4 displayed an exponential degradation profile with a half-life of 0.32 h (Figure 4.4B). After IAA washout, OCT4 recovered to approximately half of the concentration before IAA treatment within ~4.5 hours (Figure 4.4C), in line with the OCT4 protein half-life of ~4 hours (304).

To verify that OCT4 degradation kinetics are similar across the cell cycle, we applied IAA for 0.5 h (partial degradation) and 2 h (full degradation) before analyzing the mCherry signal by flow cytometry. At 2 hours of treatment, mCherry levels were similar to those of mCherry negative cells (Supplementary Figure 4.9A). We observed highly similar changes in the mCherry signal across all cell cycle phases (Supplementary Figure 4.9B-C), consistent with previous reports on the cell cycle-independence of Auxin-mediated protein degradation (143). OCT4 recovery after IAA washout was also very similar across the cell cycle (Supplementary Figure 4.9D). After addition of dox for 24 hours to remove untagged OCT4, we treated cells with IAA or not for 2 hours, sorted for different cell cycle phases, and performed ATAC-seq (Figure 4.4A). The relative magnitude of change in accessibility in the different clusters was consistent with our mitotic degradation experiment (Figure 4.4D). Remarkably, the average loss of accessibility was very similar at all cell cycle phases in clusters 1-3 (Figure 4.4D, Supplementary Figure 4.9E).

Next, we quantified the recovery of chromatin accessibility across the cell cycle. We treated OCT4-AID cells with dox for 24 hours, then with IAA or not for 2.5h, washed out the drug and incubated cells for 4.5 hours, sorted cells in different cell cycle phases and performed ATAC-seq (Figure 4.4A). While both cluster 1 (Supplementary Figure 4.9F) and 2 substantially recovered chromatin accessibility, cluster 3 loci did not (Figure 4.4E), in line with their permanent decrease of accessibility over the cell cycle upon OCT4 degradation at the M-G1 transition (see Figure 4.3D). Overall, these data show that OCT4 is required across the cell cycle to maintain chromatin accessibility at enhancer elements.

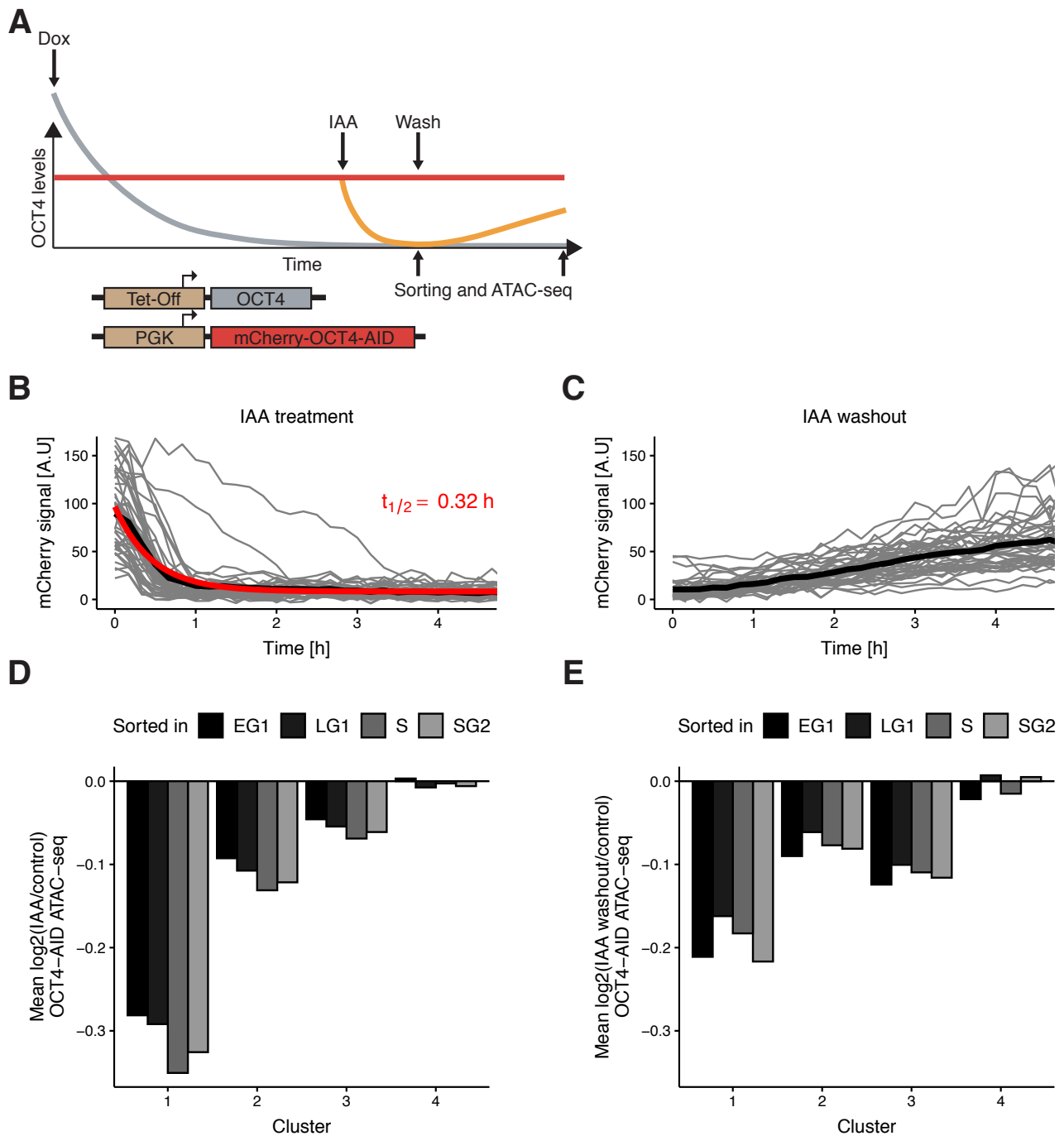


Figure 4.4 Auxin-inducible degradation reveals pioneer activity of OCT4 at different cell cycle phases.

(A) Experimental strategy used to assess the impact of OCT4 depletion and recovery at different cell cycle phases. (B) Red fluorescence (mCherry) signal in mCherry-OCT4-AID cells treated with IAA at $t=0$ as measured by fluorescence microscopy. Gray lines: single cell traces; Black line: population average; Red line: exponential fit. Red text: half-life value derived from the exponential fit. $n=45$ cells from one replicate (C) Red fluorescence (mCherry) signal in mCherry-OCT4-AID treated with IAA for 2.5h and then washed out at $t=0$ as measured by fluorescence microscopy. Gray lines: single cell traces; Black line: population average. $n=45$ cells from one replicate (D) Average \log_2 fold-change values of accessibility between IAA-treated and untreated OCT4-AID cells in the four clusters from Figure 3D at each cell cycle phase. (E) Average \log_2 fold-change values of accessibility between cells first treated with IAA and then washed out, compared to untreated OCT4-AID cells for the four clusters from Figure 3D at each cell cycle phase.

4.3.5 Dynamic relationship between OCT4 concentration and chromatin accessibility

We next aimed to quantify the dynamics of chromatin accessibility changes in response to OCT4 loss. Since residence times of OCT4 on specific DNA sites are in the second-range (293, 305, 279),

we reasoned that if continuous OCT4 re-binding is required to maintain chromatin accessibility, changes in chromatin accessibility and OCT4 concentration should occur in a quasi-synchronized manner. To test this hypothesis, we performed a time-course experiment by treating OCT4-AID cells with IAA for 0.5h, 1h, 2h, 3h, 4h, 6h, and 10h, and performed ATAC-seq at each time point. We took advantage of our different clusters, which showed differential response to OCT4 loss at the M-G1 transition (see Figure 4.3D), and analyzed accessibility loss at these loci over time. At all OCT4-responsive clusters (1-3), accessibility loss was near-complete after 1 hour of IAA treatment (Figure 4.5A-B), in line with accessibility being highly dynamic with OCT4 levels. At 6 and 10 hours of treatment, cluster 4 sites that were insensitive to OCT4 degradation at the M-G1 transition started to lose accessibility, suggesting a broader and potentially indirect impact of OCT4 loss on chromatin accessibility (Figure 4.5A). We thus focused on the first 4 hours of OCT4 removal to estimate the kinetics of accessibility loss. We fitted a single-component exponential function including an offset to account for the residual ATAC-seq signal after OCT4 loss. At clusters 1-3, the half-life of accessibility loss was remarkably close to the half-life of OCT4-AID upon IAA treatment, i.e. around 0.5 hours (Figure 4.5C-E). We were unable to fit an exponential decay to cluster 4, as expected from its OCT4-independent chromatin accessibility regulation (Figure 4.5F). In summary, these data suggest that regulation of enhancer accessibility is extremely dynamic and requires the constant presence of OCT4.

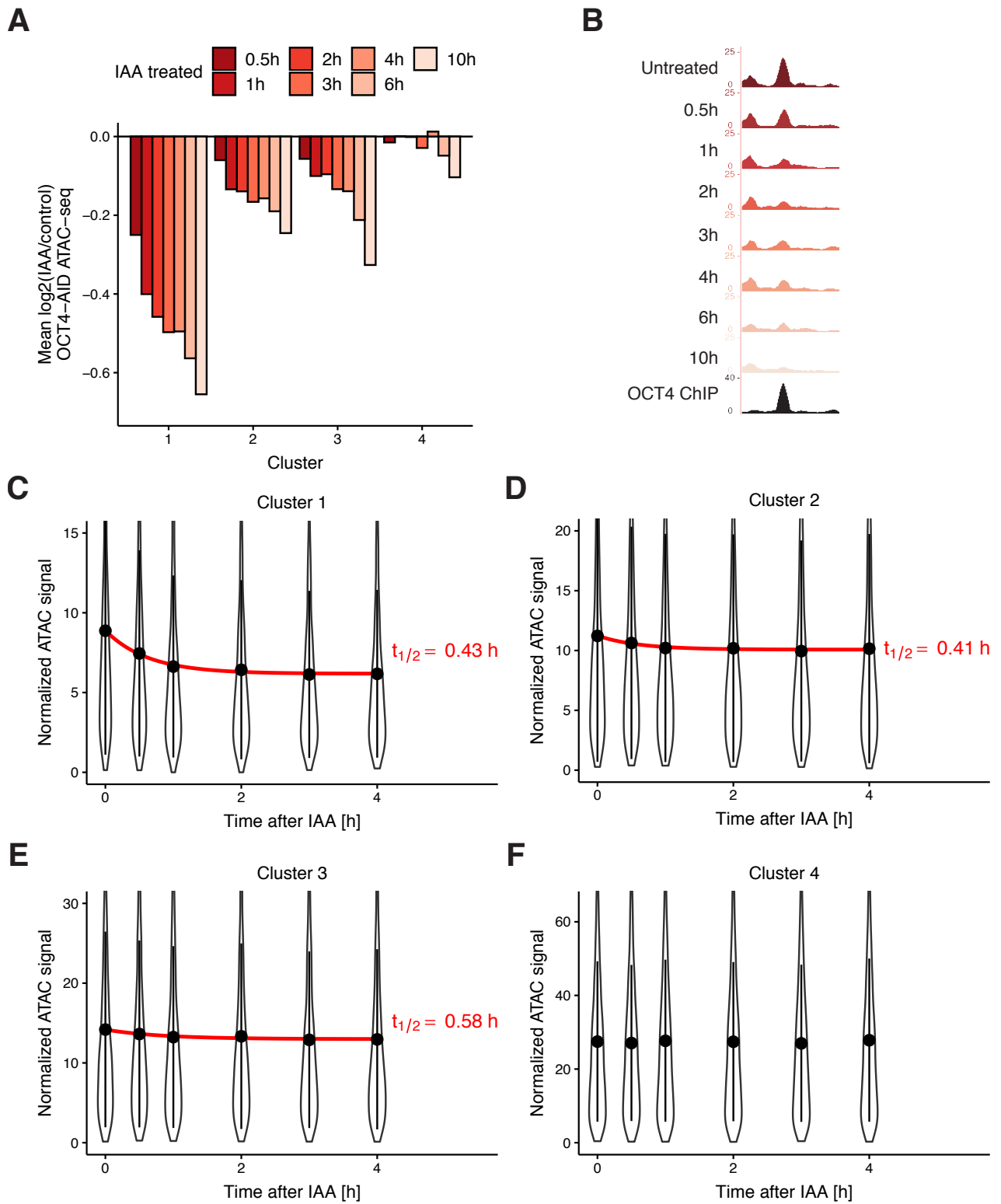


Figure 4.5 Time course analysis of chromatin accessibility changes during OCT4 degradation reveals its highly dynamic pioneer activity.

(A) \log_2 fold-change values of accessibility compared to untreated cells in the four clusters from Figure 3D at different time points of IAA treatment. (B) Genome browser track of accessibility profiles upon treatment with IAA for different durations at a cluster 1 locus at chr3:137779908-137780687. (C-F) Violin plot of normalized ATAC-seq signal across different time points in cluster 1 (C), cluster 2 (D), cluster 3 (E), and cluster 4 (F). Dots: mean; Vertical lines: standard deviation; Red lines in C-E: exponential fit; Red text in C-E: half-life value derived from the exponential fit.

4.4 Discussion

In this study we dissected the roles and interplay of OCT4 and SOX2 in regulating chromatin accessibility in ES cells. To our surprise, we found a large number of enhancers that were bound by both transcription factors but for which chromatin accessibility was regulated by only one of them. In the future it will be interesting to explore whether differences in the topology of OCT4 and SOX2 binding sites on the nucleosome surface or genomic location-dependent DNA residence times could explain these findings. Our results also show that both OCT4 and SOX2 regulate the genomic occupancy of each other mainly via regulation of chromatin accessibility. This is reminiscent of dynamic assisted loading, in which two TFs assist the loading of each other to either the same or nearby DNA binding sites (98, 306).

Surprisingly, upon OCT4 loss chromatin accessibility increased at a large number of genomic sites enriched for proximity to differentiation genes, even when OCT4 was degraded for a brief period of time at the M-G1 transition. The fact that SOX2 occupies these sites and is required to maintain their accessibility suggests that in the absence of OCT4, SOX2 is rerouted to these loci and promotes differentiation together with other partners such as TFAP2C. Therefore, the rapid action of OCT4 in early G1 phase might be required to ensure both the maintenance of chromatin accessibility at pluripotency enhancers and to silence differentiation enhancers. Whether the previously shown association of OCT4 to mitotic chromosomes (248, 279, 305) facilitates its action in early G1 will require further investigation.

We found that OCT4 degradation leads to a rapid decrease in chromatin accessibility at all clusters of OCT4-regulated enhancers across the cell cycle with very similar kinetics, which tightly mirrored changes in OCT4 concentration and thus suggests highly dynamic regulation of chromatin accessibility by OCT4. However, the recovery of chromatin accessibility upon increase of OCT4 concentration displayed locus-dependent behavior. In contrast to clusters 1 and 2, cluster 3 loci did not recover over the time course of several hours either after M-G1 or auxin-induced degradation. While the mechanisms underlying these findings are unclear, permanent loss (cluster 3) or incomplete recovery (cluster 1) of chromatin accessibility may explain why OCT4 loss at the M-G1 transition results in impaired pluripotency maintenance.

Protein depletion by degron systems works by increasing protein degradation rates without affecting their synthesis rate. Therefore, they suffer from an inherent tradeoff in maximizing expression levels when the degron is inactive while minimizing residual expression level when the degron is active. Here we expressed OCT4 at relatively low levels to ensure sufficient depletion, allowing us to show that the pioneering function of OCT4 is required constantly and throughout the cell cycle to maintain enhancer accessibility. However, the relatively low dynamic range of accessibility changes prohibits sensitive detection of specific loci that are quantitatively more or less sensitive to OCT4 loss at different cell cycle phases. Furthermore, whether recurrent, transient loss of OCT4 outside of the M-G1 transition would also lead to pluripotency loss would have to be addressed in future studies.

Here we found that OCT4 is constantly required to maintain chromatin accessibility in self-renewing ES cells. This is reminiscent of a recent study showing that the pioneer factor Zelda is required throughout zygotic genome activation in *Drosophila* for proper gene expression (307). In contrast, in the context of pituitary lineage specification PAX7 requires 72 hours to fully open melanotrope-specific enhancers but is subsequently not required to maintain these (Mayran et al., 2018). It is possible that PAX7 hands over the role of maintaining accessibility to other factors, such as TP1 (308), and is only required at the transition between cell fates. This indicates that pioneering activity can have different manifestations that depend heavily on other regulatory factors and chromatin features. Pluripotent stem cells may be particularly dynamic in this regard, as they need to be able to quickly rewire their gene expression programs upon receiving differentiation signals. In contrast, more differentiated cell types may have mechanisms to avoid precocious changes in gene expression upon subtle alterations in the concentration of TFs. Therefore, the high sensitivity of enhancers to the concentration or activity of pioneer TFs in ES cells could serve as a mechanism to regulate cell fate with precise temporal control.

4.5 Materials and methods

Cell culture

Mouse ES cells were routinely cultured on cell culture-treated dishes coated with 0.1% gelatin (Sigma #G9391-100G) using the following culture medium: GMEM (Sigma #G5154-500ML) containing 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), in-house produced leukemia inhibitory factor (LIF), CHIR99021 (Merck #361559-5MG) at 3µM and PD184352 (Sigma #PZ0181-25MG) at 0.8µM. Cells were passaged by trypsinization (Sigma #T4049-100ML) every two to three days.

Lentiviral vector production

Lentiviral vectors were produced by transfection of HEK 293T cells with the envelope (psPAX2, Addgene #12260), packaging (pMD2.G, Addgene #12259) (309), and lentiviral construct of interest using Calcium Phosphate transfection, as described previously (310). Viral vectors were concentrated 120-fold by ultracentrifugation at 20'000 rpm for 90 minutes at 4°C. 50'000 cells in 1 ml of medium in a 24-well plate were transduced with 50 µl of concentrated lentiviral vector particles to generate stable cell lines.

Cloning of overexpression constructs

pLV-PGK-YPet-MD was derived from pLVTRE3G-Sox2-YPet-MD (279) by amplification of YPet-MD and restriction cloning into pLV-rtTA3G-IRESBsd using AgeI and Sall. pLV-PGK-SNAP-MD-OCT4 and pLV-PGK-SNAP-MD*-OCT4 were derived by amplification of MD or MD* from SNAP-MD-SOX2 (Addgene #115687) and SNAP-MD*-SOX2 (Addgene #115688) (279) and restriction cloning into pLVTRE3G-Oct4-YPet (279) using Sall and XbaI. SNAP-MD-OCT4 and SNAP-MD*-OCT4 were further amplified and cloned by restriction cloning into pLV-rtTA3G-IRESBsd (279) using AgeI and Sall. pLEX-mCherry-OCT4-AID was derived by amplification of OCT4 from pLV-PGK-SNAP-MD-OCT4, AID 71-114 from pEN244 (Addgene #92140) (Nora 2017), and mCherry from pLV-TRE3G-mCherry-PGK-Puro (Suter lab). mCherry and OCT4 were ligated using an XmaI restriction site and mCherry-OCT4 was ligated to AID using a KpnI restriction site. The mCherry-OCT4-AID fragment was cloned into the pLEX_305-C-dTAG backbone (Addgene #91798) (311) using EcoRV and MluI restriction sites. pLV-pCAGGS-Tir1-V5 was derived by amplification of pCAGGS-Tir1-V5 from pEN395 (Addgene #92141) (312) and In-fusion cloning into pLV-PGK-SOX2-SNAP-IRES-Hygro (278) digested using XhoI and XbaI restriction enzymes.

Generation of stable cell lines

To generate MD-OCT4 cell lines, ZHBTc4 cells were transduced with SNAP-MD-OCT4 and SNAP-MD*-OCT4 lentiviral vector particles and sorted to display near-identical average SNAP levels (Supplementary Figure 4.6A), subsequently transduced with PGK-YPet-MD lentiviral vector particles, and sorted to display near-identical average YPet-MD levels. To generate the OCT4-AID cell line, ZHBTc4 cells were transduced with pLV-pCAGGS-Tir1-V5 and PLEX-mCherry-OCT4-AID packaged lentiviral vectors (i.e Tir1-V5 and mCherry-Oct4-AID virus, respectively) and were selected with 2 µg/ml Puromycin (Thermo Fisher Scientific #A1113803) for 10 days. Subsequently, mCherry positive cells were sorted and transduced with PGK-YPet-MD lentiviral particles and sorted for YPet positive cells. Cells that displayed IAA-dependent degradation were selected by sorting a narrow window of mCherry-positive cells, followed by treatment with 500 nM IAA (Sigma #I5148-2G) for 1 hour, and sorting for mCherry-negative cells. All cell lines were maintained in medium without dox (Sigma #D3447-500MG) or IAA prior to experiments.

Treatment conditions

Cells were seeded at a concentration of 9'000-18'000 cells per cm² one day before the start of treatment. ZHBTc4 YPet-MD SNAP-MD-OCT4 and SNAP-MD*-OCT4 were treated with 1 µg/ml dox for 40 hours prior to cell sorting. ZHBTc4 YPet-MD TIR1-V5 mCherry-OCT4-AID cells were treated with 1 µg/ml dox for 24 hours before adding IAA. Dox was maintained throughout the experiment. Cells were treated with 500 nM IAA (or not for control) for 2 hours or treated with 500 nM IAA (or not

for control) for 2.5 hours, washed 5 times with PBS with 2 minutes incubation, and placed back in medium containing 1 μ g/ml dox for 4.5 hours. For the time course experiment, OCT4-AID cells were seeded in different wells of a 24-well plate and treated with dox for 24 hours before adding IAA. Dox was maintained throughout the experiment. IAA was added at different time points (with one well left untreated) prior to cell collection. All wells were collected at the same time and subjected to ATAC-seq as described below.

Cell cycle phase sorting

Cells were trypsinized, resuspended in culture medium with 50 μ M Hoechst33342 (Thermo Fisher Scientific #H3570), and incubated for 15 minutes at 37°C. Cells were then spun down, resuspended in cold PBS with 1% FBS, and sorted according to their YPet-MD and Hoechst profile (Supplementary Figure 4.6B). Cells were sorted at 4°C into 1.5 ml Eppendorf tubes or 15 ml Falcon tubes containing a small amount of PBS with 1% FBS. Sorting for SNAP levels was done on a MoFlo Astrios (Beckman Coulter). All other sorting was done on a FACSaria II or a FACSaria Fusion (BD Biosciences).

ATAC-seq

All ATAC-seq experiments were performed in biological duplicates except for IAA-treated mCherry-OCT4-AID cells sorted in EG1, LG1, and S phase where three replicates were performed. 50'000 cells were collected either directly after trypsinization or after sorting as described above and subjected to ATAC-seq as described previously (147). All centrifugation steps were done at 800g at 4°C. Briefly, cells were centrifuged for 5 minutes and washed with cold PBS, then centrifuged for 5 minutes and resuspended in cold lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% NP-40), and centrifuged for 10 minutes. Subsequently, nuclei were resuspended in a solution of 0.5 μ M Tn5 (in-house preparation according to (313)) in TAPS-DMF buffer (10 mM TAPS-NaOH, 5 mM MgCl₂, 10% DMF) and incubated for 30 minutes at 37°C. DNA was immediately purified using column purification (Zymo #D4004) and eluted in 10 μ l nuclease-free H₂O. Transposed DNA was amplified in a solution containing 1X NEBNextHigh Fidelity PCR Master Mix (NEB #M0541L), 0.5 μ M of Ad1_noMX universal primer, 0.5 μ M of Ad2.x indexing primer and 0.6x SYBR Green I (Thermo Fisher Scientific #S7585) using 72°C for 5 minutes, 98°C for 30 s, and 5 cycles of 98°C for 1s, 63°C for 30s, and 72°C for 60s. 10 μ l of amplified DNA was analyzed by qPCR to determine the total number of cycles to avoid amplification saturation and accordingly amplified with additional 3-7 cycles of 98°C for 10s, 63°C for 30s, and 72°C for 60s. DNA was purified using column purification (Zymo #D4004) and size-selected by taking the unbound fraction of 0.55X AMPure XP beads (Beckman Coulter #A63880) followed by the bound fraction of 1.2X AMPure XP beads. Libraries were sequenced on an Illumina NextSeq 500 using 75 nucleotide read-length paired-end sequencing.

ChIP-seq

All ChIP-seq experiments were performed in biological duplicates. Roughly 10 million cells per sample were collected after trypsinization and fixed with 2 mM disuccinimidyl glutarate (Thermo Fisher Scientific #20593) in PBS for 50 minutes at room temperature, spun down at 600g for 5 minutes and washed once with PBS. Cells were then treated with 1% formaldehyde (Axon Lab #A0877,0500) for 10 minutes at room temperature and quenched with 200mM Tris-HCl pH 8.0 for 10 minutes, washed with PBS and spun down. For ZHBTc4 YPet-MD SNAP-MD-OCT4 and SNAP-MD*-OCT4 cells, cells were subsequently resuspended in cold PBS with 1% FBS and at least 500'000 cells per cell cycle phase were sorted. Fixed cell pellets were kept on ice and resuspended in LB1 (50mM HEPES-KOH pH 7.4, 140mM NaCl, 1 mM EDTA, 0.5mM EGTA, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100), incubated 10 min at 4°C, spun down at 1700g, and resuspended in LB1 a second time, spun down and resuspended in LB2 (10mM Tris-HCl pH 8.0, 200mM NaCl, 1 mM EDTA, 0.5mM EGTA), incubated for 10 min at 4°C, spun down and washed without disturbing the pellet three times with SDS shearing buffer (10mM Tris-HCl pH 8.0, 1mM EDTA, 0.15% SDS) and finally resuspended in SDS shearing buffer. All buffers contained Protease Inhibitor Cocktail

(Sigma #P8340-1ML) at 1:100 dilution. Chromatin was sonicated for 20 min at 5% duty cycle, 140 W, 200 cycles on a Covaris E220 focused ultrasonicator. Sonicated chromatin was equilibrated to 1% Triton X-100 and 150 mM NaCl and incubated with each antibody overnight at 4°C. Antibodies used were anti-BRG1 (Abcam #ab110641) at 5 µg per 10 million cells, anti-OCT4 (Cell Signaling Technology #5677S) at 20 µl per 10 million cells, and anti-H3K27ac (Abcam #ab4729) at 2 µg/25 µg chromatin. Protein G Dynabeads (Thermo Fisher Scientific #10003D) were blocked with 5 mg/ml BSA in PBS, added to chromatin, and incubated at 4°C for 3 hours. Beads were washed twice with Low Salt wash buffer (10mM Tris-HCl pH 8.0, 150mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.15% SDS, 1 mM PMSF), once with High Salt wash buffer (10mM Tris-HCl pH 8.0, 500mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.15% SDS, 1 mM PMSF), once with LiCl wash buffer (10mM Tris-HCl pH 8.0, 1mM EDTA, 0.5mM EGTA, 250mM LiCl, 1% NP40, 1% sodium deoxycholate, 1mM PMSF), and finally with 1X TE (10 mM Tris pH 8.0, 1 mM EDTA) before being resuspended in ChIP Elution buffer (10 mM Tris pH 8.0, 1 mM EDTA, 1% SDS, 150 mM NaCl) with 400 ng/µl Proteinase K (Qiagen #19131) and reverse-crosslinked overnight at 65°C. DNA was purified using a MinElute PCR purification kit (Qiagen #28004) and libraries were prepared with the NEBNext Ultra II DNA Library Prep Kit (NEB #E7645S). Libraries were sequenced on an Illumina NextSeq 500 using 75-nucleotide read length paired-end sequencing.

Pluripotency assays and qRT-PCR

ZHBTc4 YPET-MD cells expressing SNAP-MD-OCT4 or SNAP-MD*-OCT4 were plated at 400 cells per well in a 6-well plate with ES cell medium (see above) with 0 or 1 µg/ml dox and medium was refreshed every other day. At day 8, flat and dome-shaped colonies were scored according to morphology followed by alkaline phosphatase staining (Sigma #86R-1KT). For qRT-PCR experiments, cells were collected at day 8 and RNA was extracted using GenElute Mammalian Total RNA MiniPrep Kit (Sigma #RTN350). cDNA was synthesized using oligodT primers using SuperScript II Reverse Transcriptase (Thermo Fisher Scientific #18064014). qPCR was performed on a 7900HT (Applied Biosystems).

Immunofluorescence microscopy

2TS22C and ZHBTc4 cells were plated in a 96-well plate coated for 1 hour at 37°C with 1:25 diluted StemAdhere (Primorigen Biosciences #S2071-500UG), treated with 1 µg/ml dox for different durations and fixed with 2% formaldehyde for 30 minutes at room temperature, washed with PBS, permeabilized with PBS with 5% FBS and 0.5% Triton X-100 for 30 minutes at room temperature, and incubated with the primary antibody, anti-OCT4 C-10 (Santa Cruz #sc-5279) at 1:500 dilution and anti-SOX2 (ThermoFisher #48-1400) at 1:200 dilution, in PBS with 5% FBS and 0.1% Triton X-100 at 4°C overnight. After washing with PBS, cells were incubated with the secondary antibody, anti-Mouse IgG AF488 (Thermo Fisher Scientific #A28175) and anti-Rabbit IgG AF647 (Thermo Fisher Scientific #A27040) at 1:1000, in PBS with 5% FBS and 0.1% Triton X-100 for 60 minutes at room temperature, with 2 ng/ml DAPI (Thermo Fisher Scientific #62248) for 10 minutes at room temperature and subsequently washed and imaged on an IN Cell Analyzer 2200 (GE Healthcare). Images were background-subtracted using FiJi (314) with a rolling ball radius of 50 pixels and analyzed using CellProfiler (315). Nuclei were identified using the Watershed module on the DAPI channel, objects that were too large or too small were discarded, and the mean intensity in the OCT4 and SOX2 channels was measured within the identified nuclei.

Time-lapse microscopy

ZHBTc4 YPet-MD TIR1-V5 mCherry-OCT4-AID cells were plated in a 96-well plate coated for 1 hour at 37°C with 1:25 diluted StemAdhere (Primorigen Biosciences #S2071-500UG) and imaged on an IN Cell Analyzer 2200 (GE Healthcare) using the TexasRed and Brightfield channels. Cells were treated with IAA or washed just prior to imaging. Images were background-subtracted using FiJi (314) with a rolling ball radius of 50 pixels and nuclei were tracked manually over time using the Manual Tracking plugin in FiJi in the Brightfield channel. The mean signal in 10 pixels around the

tracked spot were measured in the TexasRed channel and the mean background signal at an equivalent sized spot free from cells (background) was subtracted at each time point.

Data analysis for ATAC-seq and ChIP-seq

All sequencing libraries were aligned to the mm10 Mus musculus genome (GRCm38) with STAR 2.6.1c (316) and duplicate reads were removed using Picard (Broad Institute). Reads not mapping to chromosomes 1-19, X, or Y were removed. Peaks were called with MACS 2.1.1.20160309 (317) with settings '-f BAMPE -g mm'. For comparative analysis of ZHBTc4 and 2TS22C cells, all peaks from ZHBTc4 and 2TS22C ATAC-seq experiments were merged with BEDTools (318). For MD-OCT4 and OCT4-AID analyses, peaks from all ATAC-seq experiments of treated cells in the corresponding cell lines were merged. For MD-OCT4 H3K27ac analysis, peak coordinates were expanded by 500 bp on both sides to account for the enrichment profile of H3K27ac. All peaks larger than 5 kb, overlapping peaks called in Input (no immunoprecipitation) samples from ES cells in S2iL (GSE89599) or SL (GSE87822), or overlapping blacklisted peaks (319) were removed. The HOMER2 (320) functions makeTagDirectory and annotatePeaks.pl with settings '-noadj -len 0 -size given' were used for read counting and count tables were loaded into RStudio. TMM Normalization was done with edgeR (321) and analysis of differentially accessible regions was done with limma (281). Contrasts were designed as ~0+Condition+Replicate, where Condition specifies the cell line and treatment and Replicate the date of the experiment, to take into account the paired nature of the experiments. For comparing unpaired experiments, i.e. untreated ZHBTc4 vs 2TS22C cell lines or untreated ZHBTc4 in SL versus S2iL, ~0+Condition was used. For Figure 4.3 – figure supplement 3A-B, the mean of the TMM-normalized reads in the ChIP-seq and ATAC-seq replicates was divided by the nucleotide length of each region. For Figures 5C-F, the mean of the TMM-normalized reads in the replicates was used. The HOMER2 function findMotifsGenome.pl was used with the setting '-size given' for motif searching. The most frequent known motif in target regions of a given class of known motifs (i.e. different versions of SOX and OCT motifs) was used. Background was calculated as the mean of HOMER-estimated background frequency in all groups/clusters. Only motifs with logP < -50 are shown in Tables 1-3. For GO analysis, the closest Entrez gene entry TSS to each region in groups was used and enrichment was calculated using the HOMER2 function findGO.pl with setting 'mouse'. Gene names were converted between assemblies using biomaRt (322). Replicate bam files were merged using SAMTools (323) and converted to bigWig files using the deepTools 3.1.3 (324) function bamCoverage with settings '--normalizeUsingRPKM'. Average lineplots were generated using deepTools computeMatrix (with setting 'reference-point') and custom R code. Heatmaps were generated using the deepTools function plotHeatmap. Genome tracks were made in the UCSC genome browser (325). Plots were generated using ggplot2 (326). Overlap between genomic regions was determined using GenomicRanges (327). The heatmap in Figure 4.1C was generated using ComplexHeatmap (328). Color schemes were taken from colorbrewer2.org and <https://rpubs.com/Koundy/71792>.

Published datasets

Published data were aligned and processed as described above. Processed bigWig files were downloaded from GEO (329) or cistromeDB (241). When necessary, peak files were converted to mm9 using liftOver (330). OCT4 and SOX2 ChIP-seq peaks were derived from newly generated (2TS22C OCT4) and published (ZHBTc4 OCT4 and SOX2) (239) datasets as described above as well as from processed SOX2 ChIP-seq peaks from asynchronous E14 cells (GSE89599) (279) and merged with BEDTools (318). Super-enhancer and typical enhancers were taken from (299) and converted to mm10 using liftOver (330). ChromHMM tracks from mouse ES cells were downloaded from https://github.com/guifengwei/ChromHMM_mESC_mm10 (331). ATAC-seq data from OCT4 high and OCT4 low sorted cells were taken from a previous study (278) and processed as described above, merging SHOH and SLOH samples into OCT4 high and SLOL and SHOL samples into OCT4 low.

K-means clustering

Clusters in Figure 4.3D were generated using the R function `pheatmap` with settings 'clustering_distance_rows = "euclidean", kmeans_k = 4' on a matrix containing the log2 fold-change values in accessibility between MD-OCT4 and MD*-OCT4 at each cell cycle phase (columns) at each OCT4-bound locus (rows). Clusters were ordered according to the lowest mean log2 fold-change in EG1.

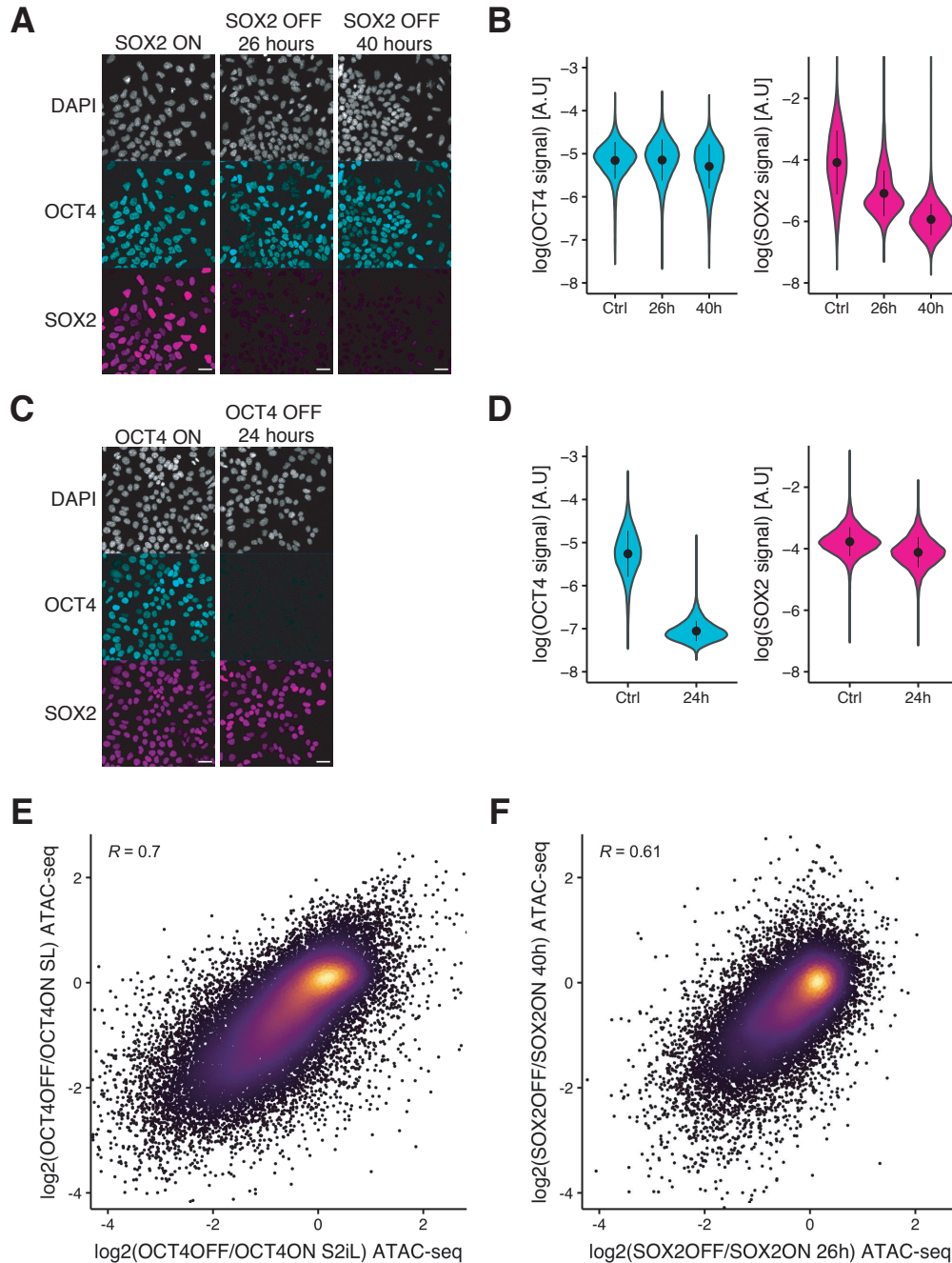
Exponential curve fitting

Exponential decays were fitted using the R function `nls` with the formula $y \sim a * e^{(-b * x)} + c$ where a , b , and c are constants. Half-life values were derived as $\log(2)/b$.

Data availability

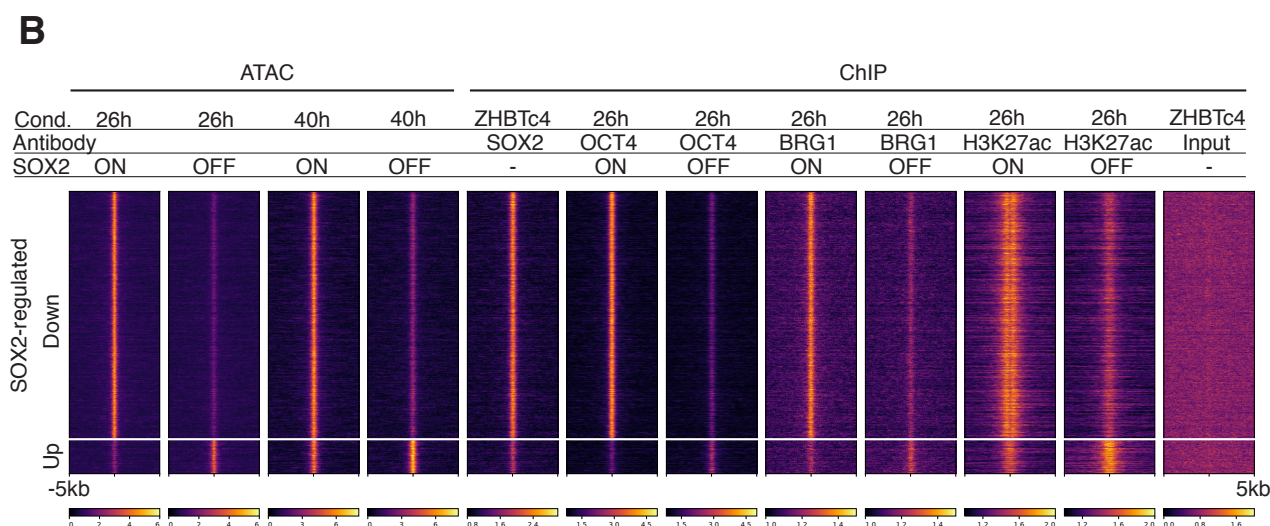
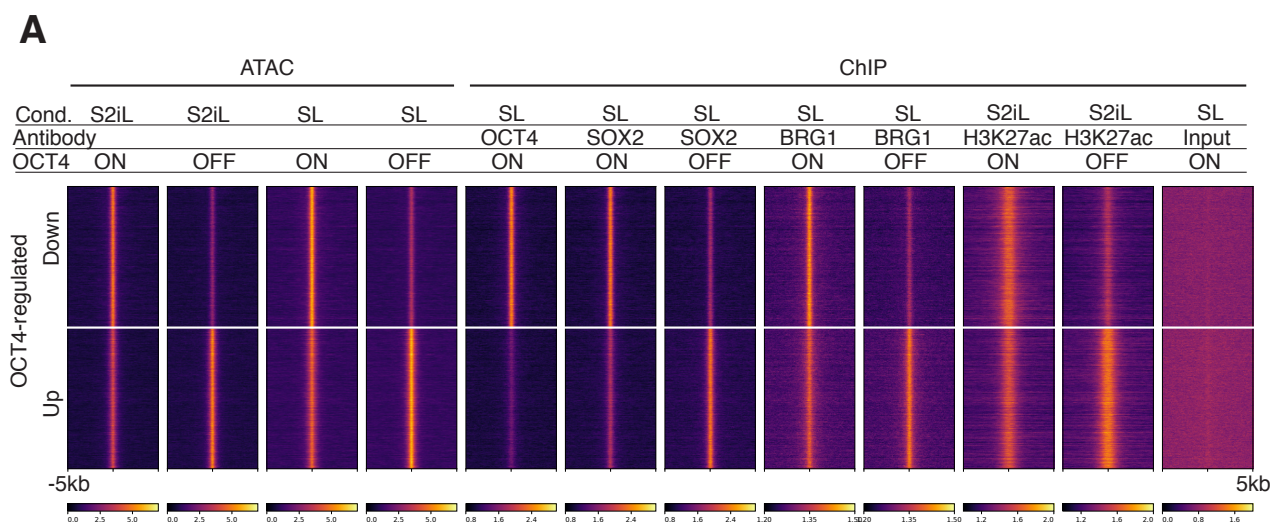
Raw and processed sequencing data have been uploaded to the Gene Expression Omnibus under the accession code GSE134680

4.6 Supplementary figures



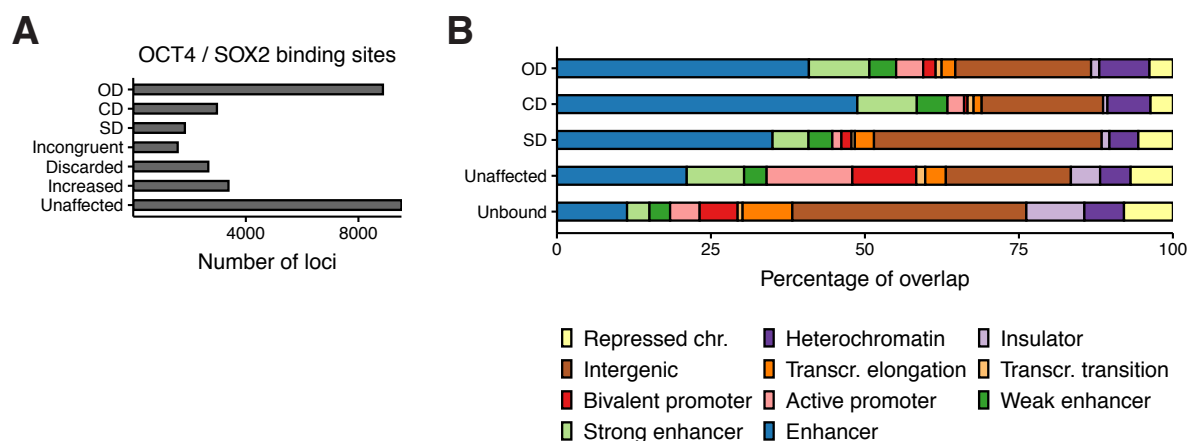
Supplementary Figure 4.1

(A) Immunofluorescence of 2TS22C cells stained for DNA (DAPI), OCT4, and SOX2 without dox treatment (left), and after 26 hours (middle), and 40 hours (right) of dox treatment. (B) Violin plot of background-subtracted log values of immunofluorescence signal in OCT4 (left) and SOX2 (right) channels upon SOX2 depletion. Control: $n=45'601$ cells from 4 biological replicates including 2 technical replicates; 26 hours: $n=42'298$ cells from 3 biological replicates including 2 technical replicates; 40 hours: $n=32'342$ cells from 2 technical replicates. Dots: mean; Vertical lines: standard deviation. (C) Immunofluorescence of ZHBTc4 cells stained for DNA (DAPI), OCT4, and SOX2 without dox treatment (left), and after 24 hours of dox treatment (right). (D) Violin plot of background-subtracted log values of immunofluorescence signal in OCT4 (left) and SOX2 (right) channels upon OCT4 depletion. Control: $n=26'119$ cells from 3 biological replicates. 24 hours: $n=23'157$ cells from 3 biological replicates. Dots: mean; Vertical lines: standard deviation. (E) Correlation between the log2 fold-change values of accessibility upon OCT4 depletion in S2iL (x-axis) and SL (y-axis) at OCT4-bound sites. (F) Correlation between the log2 fold-change values of accessibility upon SOX2 depletion after 26 hours (x-axis) and 40 hours (y-axis) of dox treatment at SOX2 binding sites. R is Pearson correlation coefficient. Scale bars: 30 μ m.



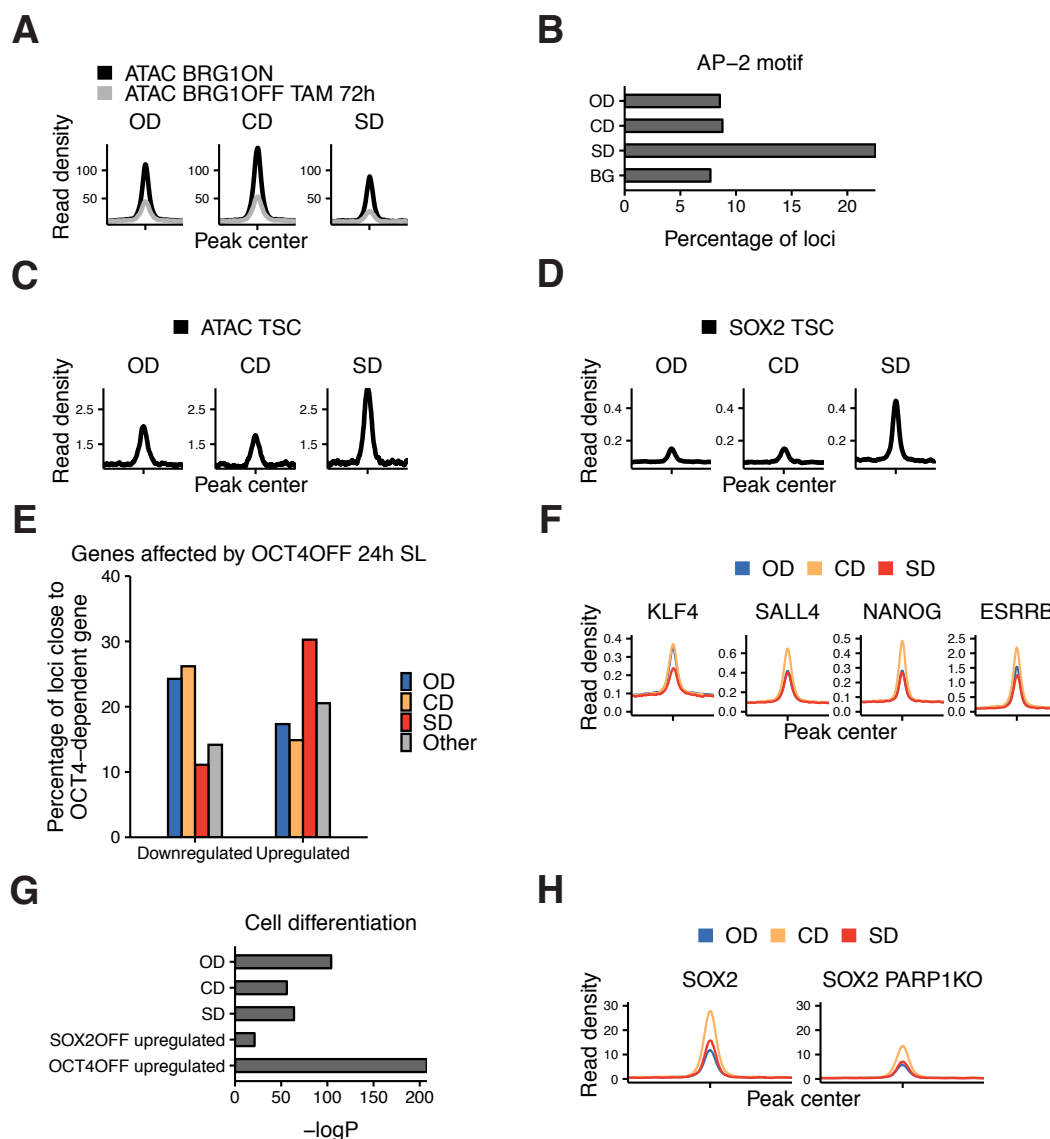
Supplementary Figure 4.2

Heatmaps of RPKM-normalized ATAC-seq and ChIP-seq binding profiles upon OCT4 (A) and SOX2 (B) depletion 5 kb around OCT4-regulated (A) and SOX2-regulated (B) loci. Each row represents one individual locus and each column represents one experimental condition.



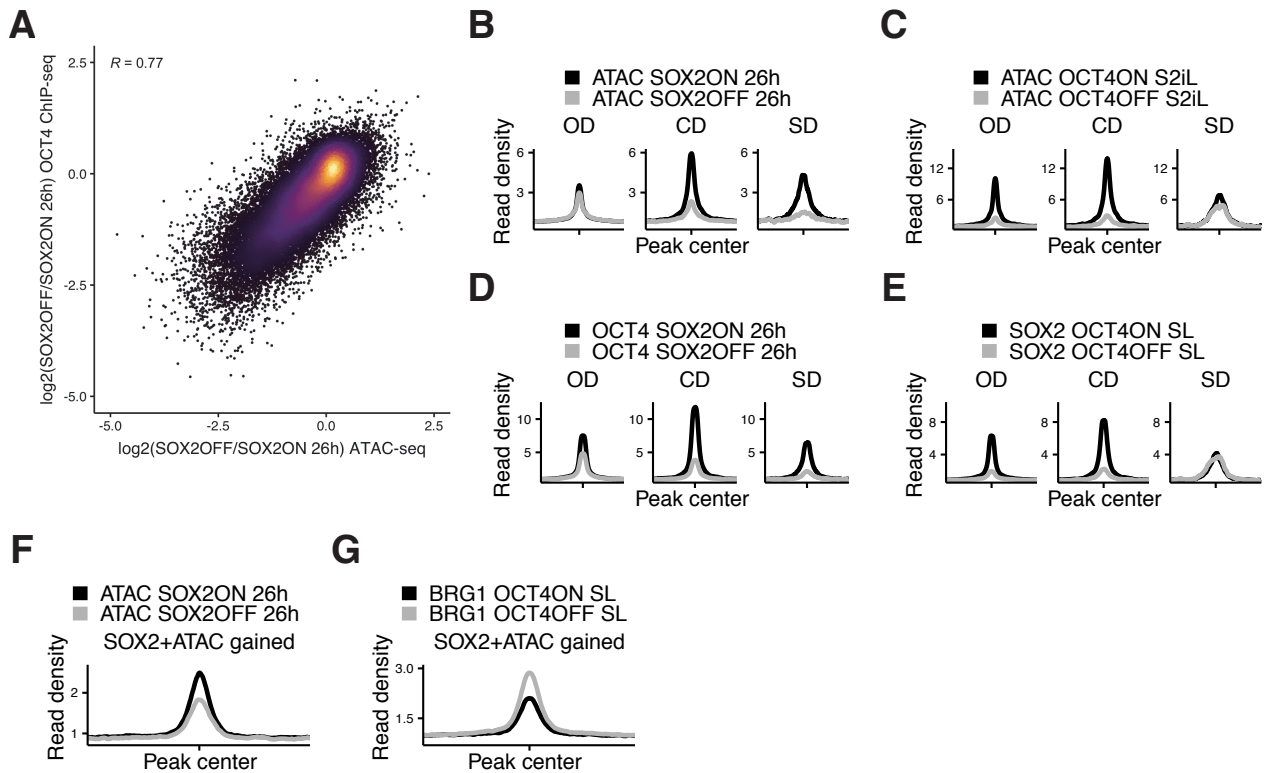
Supplementary Figure 4.3

(A) Classification of all OCT4 and SOX2 binding sites into OD, CD, and SD loci as well as loci that were discarded due to differences in untreated cells between conditions or cell lines (Discarded), due to incongruent effect on accessibility after depletion in different conditions (Incongruent) and those that were unaffected by depletion. (B) ChromHMM signal enrichment at OD, CD, and SD loci as well as loci that were unaffected by depletion or not bound by OCT4 or SOX2.



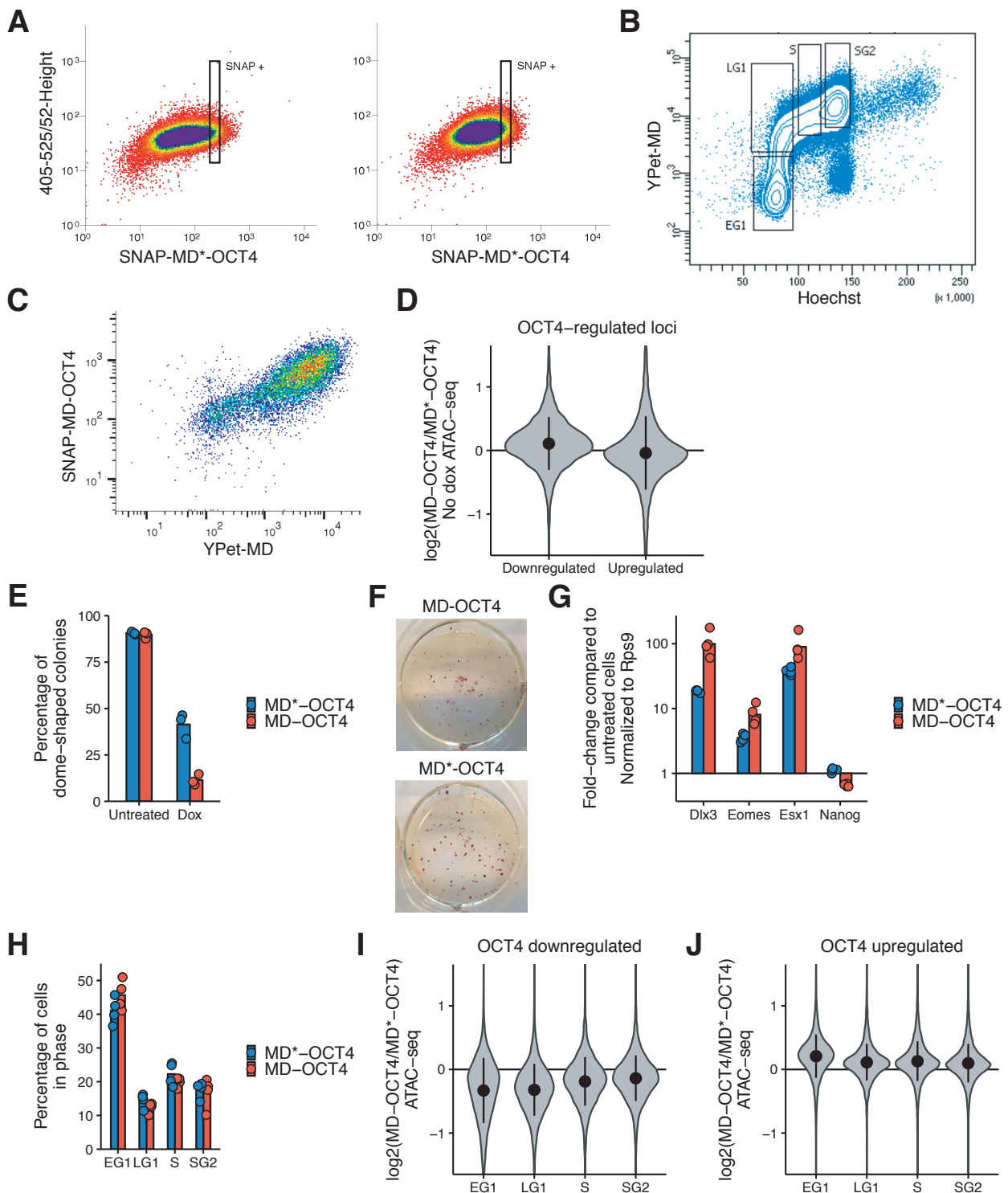
Supplementary Figure 4.4

(A) Average ATAC-seq signal 2 kb around OD, CD, and SD loci in BRG1fl cells that were treated with tamoxifen (TAM) or left untreated. (B) Frequency of the AP-2 motif 2 kb around OD, CD, and SD loci, and in background regions (BG). (C) Average SOX2 ChIP-seq signal in TS cells 2 kb around OD, CD, and SD loci. (D) Average ATAC-seq signal in TS cells 2 kb around OD, CD, and SD loci. (E) Percentage of the closest gene in the OD, CD, and SD groups as well as all other accessible regions (Other) whose nascent RNA levels are downregulated or upregulated upon 24 hours of OCT4 depletion (F) Average ChIP-seq signal of ESRRB, NANOG, KLF4, and SALL4 in ES cells 2 kb around OD, CD, and SD loci. (G) Enrichment (-log(p)) values for the closest gene in the OD, CD, and SD groups in the "Cell differentiation" gene ontology set. (H) SOX2 binding profiles 2 kb around OD, CD, and SD loci in wt and PARP1 KO ES cells.



Supplementary Figure 4.5

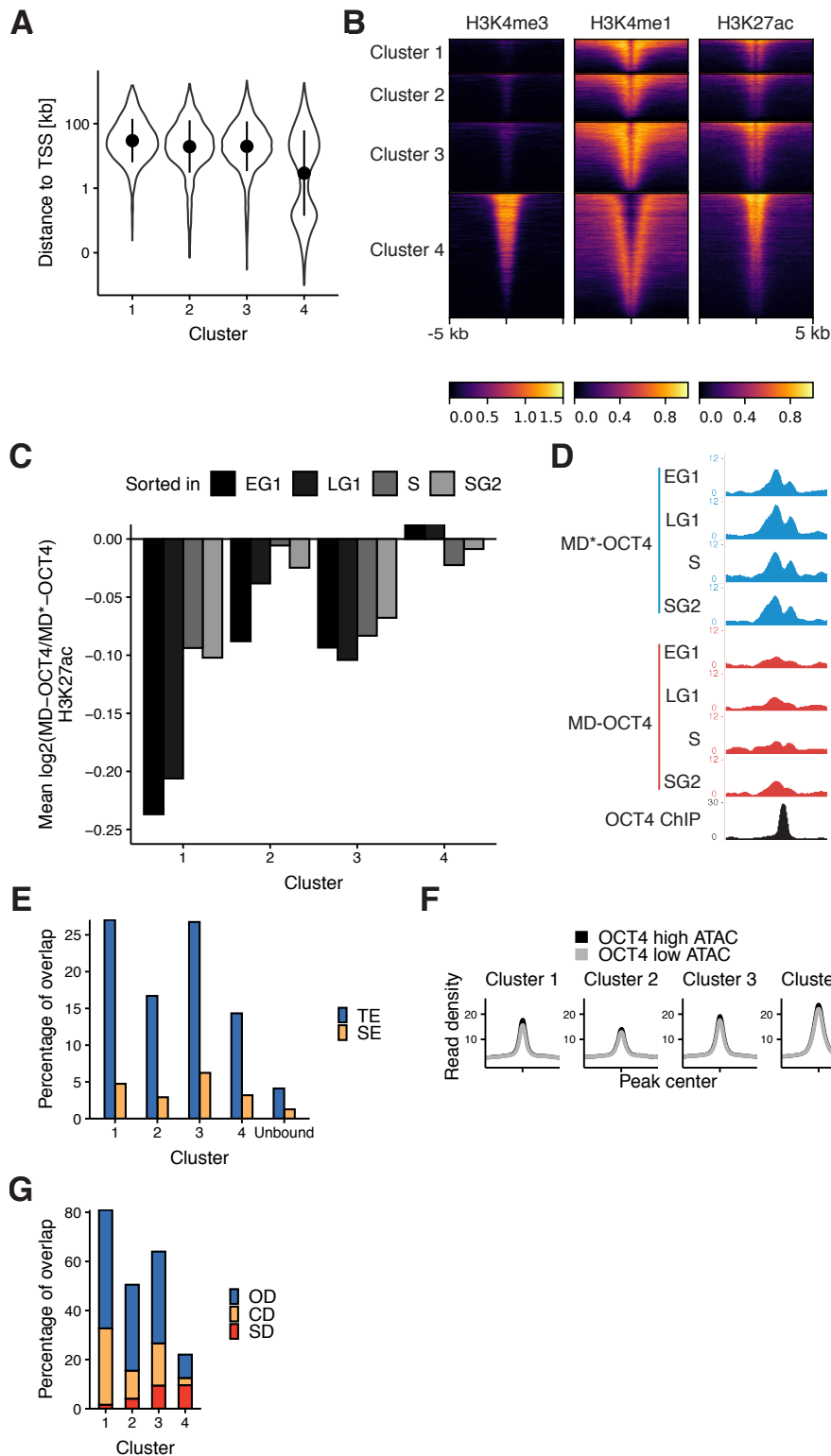
(A) Correlation between \log_2 fold-change values of accessibility (x-axis) and OCT4 binding (y-axis) upon SOX2 depletion in 2TS22C cells with dox treatment for 26 hours. R is Pearson correlation coefficient. (B-C) Average RPKM-normalized ATAC-seq signal 2 kb around OD (n=3'730), CD (n=1'463), and SD (n=273) loci that overlap with a canonical OCT4::SOX2 motif upon SOX2 (B) and OCT4 (C) depletion. (D-E) Average RPKM-normalized OCT4 (D) and SOX2 (E) ChIP-seq signal 2 kb around OD, CD, and SD loci that overlap with a canonical OCT4::SOX2 motif upon SOX2 (D) and OCT4 (E) depletion. (F) Average RPKM-normalized ATAC-seq signal upon SOX2 depletion 2 kb around loci that display a significant increase in accessibility and SOX2 binding upon OCT4 depletion. (G) Average RPKM-normalized BRG1 ChIP-seq signal upon OCT4 depletion 2 kb around loci that display a significant increase in accessibility and SOX2 binding upon OCT4 depletion.



Supplementary Figure 4.6

(A) Gate used to sort SNAP-MD-OCT4 (left) and SNAP-MD*-OCT4 (right) cells for the same average SNAP-Cell 647-SiR signal. Y-axis: Signal amplitude at 405 nm excitation and 526/52 nm emission (negative control). X-axis: Signal amplitude at 640 nm excitation and 671/30 nm emission (SNAP signal). (B) Example of a sorting experiment for different phases of the cell cycle in cells expressing YPet-MD and stained for Hoechst33258. Y-axis: Integrated signal at 488 nm excitation and 525/50 nm emission (YPet). X-axis: Signal amplitude at 355 nm excitation and 450/50 nm emission (Hoechst) (C) Correlation between YPet-MD and SNAP-MD-OCT4 expression in MD-OCT4 cells as measured by flow cytometry. Y-axis: Integrated signal at 640 nm excitation and 670/14 nm emission (SNAP). X-axis: Integrated signal at 488 nm excitation and 525/50 nm emission (YPet). (D) Violin plot of log₂ fold-change values of accessibility between MD-OCT4 and MD*-OCT4 cells in significantly downregulated and upregulated loci (see Figure 4.1B) in unsorted cells in the absence of dox. Dots: mean; Vertical lines: standard deviation. (E) Percentage of dome-shaped colonies as

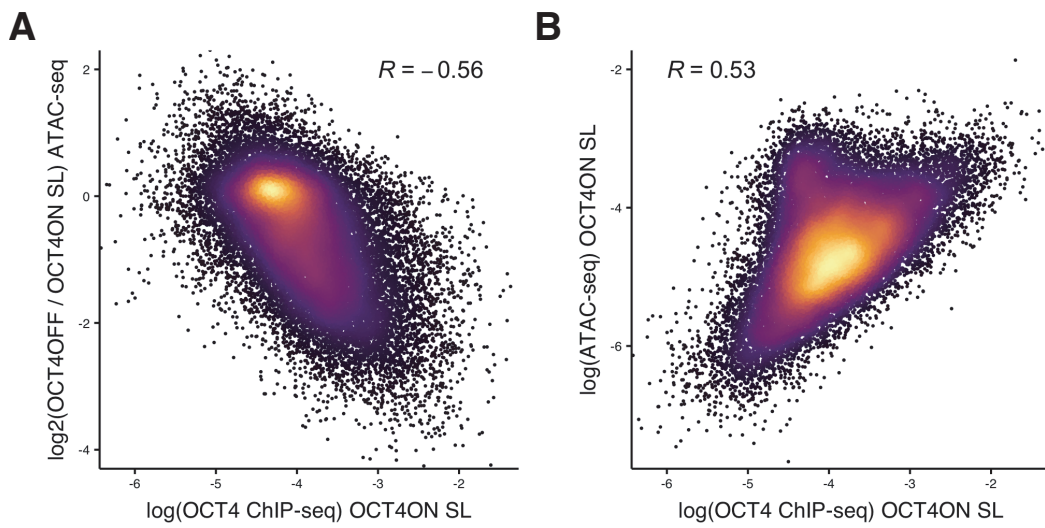
assessed by microscopy in the ZHBTc4 cell line upon dox treatment and with overexpression of SNAP-MD*-OCT4 or SNAP-MD-OCT4. n=3 biological replicates. (F) Representative alkaline phosphatase staining from cells in (E). (G) Fold-change of expression levels of differentiation markers (*Dlx3*, *Eomes* and *Esx1*) and *Nanog*, measured by qRT-PCR in dox-treated versus untreated cells, in MD-OCT4 and MD*-OCT4 cells. Each sample is normalized to the expression of *Rps9*. n=4 biological replicates. (H) Percentage of cells in EG1/LG1/S/SG2 phases as determined by flow cytometry in MD-OCT4 and MD*-OCT4 cells. n=4 biological replicates. (I-J) Violin plot of log2 fold-change values of accessibility between MD-OCT4 and MD*-OCT4 cells in different cell cycle phases at significantly downregulated (I) and upregulated (J) loci (see Figure 4.1B). Dots: mean; Vertical lines: standard deviation.



Supplementary Figure 4.7

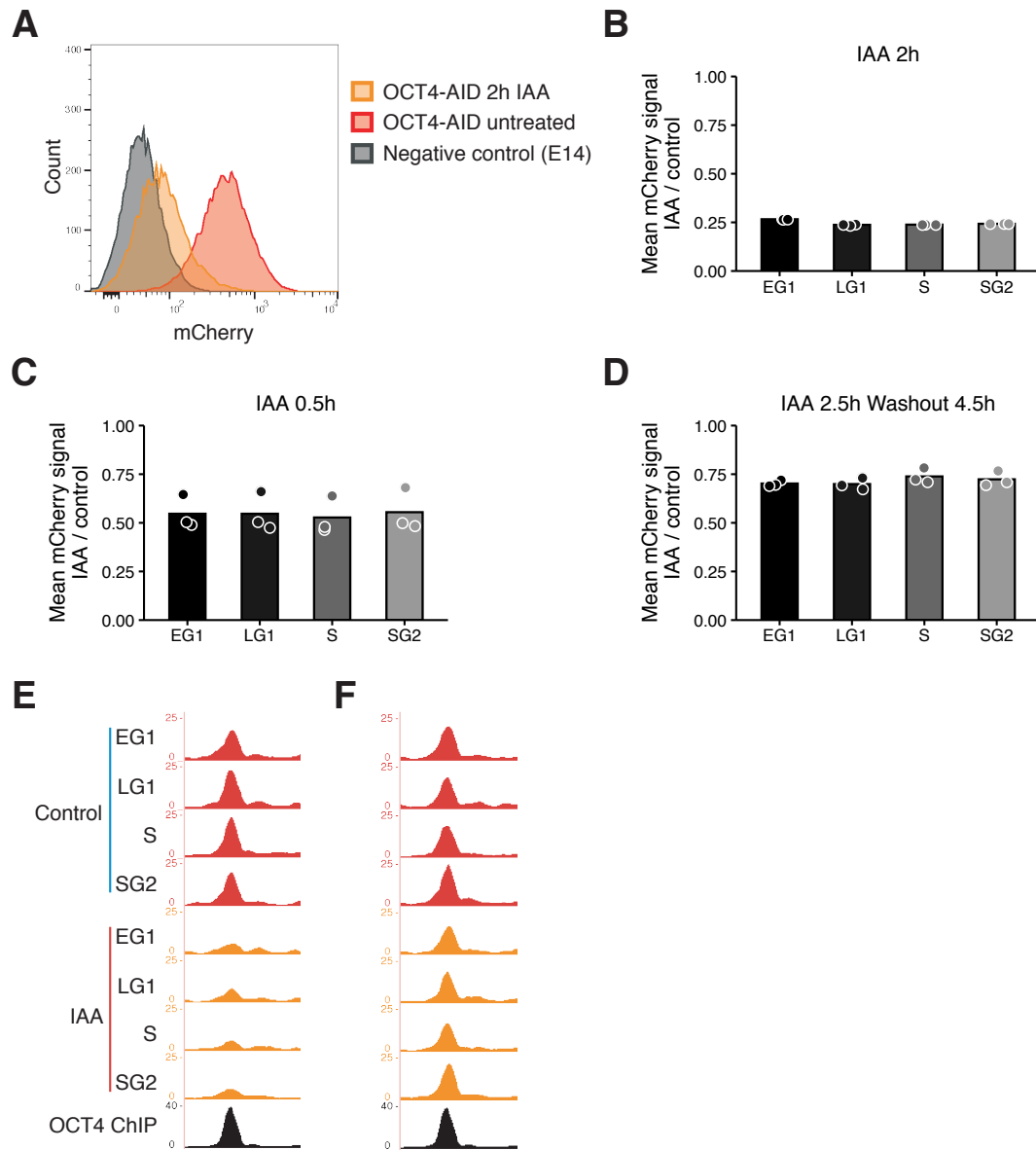
(A) Violin plot of distance to closest TSS in clusters from Figure 3D. Dots: mean; Vertical lines: standard deviation. (B) Heatmap of ChIP-seq signal of H3K4me3, H3K4me1, and H3K27ac in wt ES cells in the different clusters. (C) Average log2 fold-change values of H3K27ac ChIP-seq signal between MD-OCT4 and MD*-OCT4 cells in the different clusters (including 500bp flanking regions at each side) at different cell cycle stages. (D) Genome browser tracks of RPKM-normalized H3K27ac profiles across the cell cycle for a cluster 1 locus (chr11:6894809-6895533) that decreases in accessibility and H3K27ac upon transient OCT4 depletion in M-G1. (E) Percentage of loci in the different clusters and at non-OCT4 bound regions overlapping typical enhancers (TE) and super-enhancers (SE) in mouse ES cells. (F) ATAC-seq signal in

cells sorted for high and low endogenous OCT4 levels in the different clusters. (G) Percentage of loci in the different clusters overlapping OD, CD, and SD loci (see Figure 4.1C).



Supplementary Figure 4.8

(A) Correlation between the log of normalized OCT4 ChIP-seq reads per bp (x-axis) and the log₂ fold-change values of accessibility loss upon OCT4 depletion (y-axis) at all OCT4 binding sites in ZHBTc4 cells. (B) Correlation between the log of normalized OCT4 ChIP-seq reads per bp (x-axis) and the log of normalized ATAC-seq reads per bp (y-axis) at all OCT4 binding sites in ZHBTc4 cells. R is Pearson correlation coefficient.



Supplementary Figure 4.9

(A) Histogram of mCherry signal in untreated mCherry OCT4-AID cells and treated with IAA for 2 hours as well as mCherry negative E14 ES cells as measured by flow cytometry. X-axis: Integrated signal at 561 nm excitation and 610/20 nm emission. Y-axis: Counts. (B-D) Fold-change of red fluorescence (mCherry) signal between treated and untreated mCherry-OCT4-AID cells as determined by flow cytometry in different cell cycle phases upon 2 h IAA treatment (B), 0.5 h IAA treatment (C), and after 2.5 h IAA treatment followed by 4.5 h of washout (D). $n=3$ biological replicates. (E-F) Genome browser tracks of accessibility profiles of a cluster 1 locus at chr10:95455826-95456819 after IAA treatment (E) and after IAA treatment followed by washout (F).

4.7 Acknowledgements

This work was supported by the Swiss National Science Foundation (grants #PP00P3_179068 and PP00P3_17205 to D.M.S.). A.C.A.M.F. was supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme. This work was further supported by AgingX (SystemsX.ch) and SNF (310030_182655). We thank Bastien Mangeat, Elisa Cora, Paolo Ferrari, and Lionel Ponsonnet from the Gene Expression Core Facility for high-throughput sequencing, Miguel Garcia, Loïc Tauzin, Valérie Glutz, and André Mozes from the Flow Cytometry Core Facility for cell sorting, Olivier Burri and Romain Guet from the Bioimaging and Optics Platform for assistance with cell tracking, the staff at Vital-IT and SCITAS for cluster computing, and Armelle Tollenaere for critical reading of the manuscript.

Chapter 5 Conclusion

5.1 Discussion of achieved results

The three studies presented here converge on a few main conclusions that will hopefully serve as a foundation for further work trying to understand the complex actions of transcription factors in regulation of cell fate across the cell cycle. Below I list what are, to my mind, the most important findings and discuss them.

SOX2 and OCT4 can bind to the same genomic regions but regulate accessibility independently of each other

We found that similarly to OCT4, SOX2 relies on the BAF chromatin remodeling complex to promote chromatin accessibility, mostly at enhancer elements. However, upon depletion of either OCT4 or SOX2, the chromatin accessibility regions that are co-occupied by both proteins can be regulated by only one of the two proteins. This result indicates that there must be additional regulatory layers that are not captured by ChIP-seq measurements of binding. We could identify differences in motif frequency between regions of the different groups (dependent on SOX2 or OCT4 or both), suggesting differences in regulation partially occur through the co-binding with other factors, such as SOX2 together with AP-2. However, the motifs alone cannot account for all the observed differences. Due to the ease with which ChIP-seq and similar experiments are performed and compared, many studies infer co-regulation from co-binding to the same genomic regions. Our results suggest caution should be taken when interpreting functionality from binding data and opens up interesting avenues of exploration to understand gene regulation.

SOX2 and OCT4 remain associated to chromatin during mitosis

Live-cell imaging of fluorescently tagged SOX2 and OCT4 revealed clear enrichment of these factors on mitotic chromosomes, which became the starting point for this project. This had likely been overlooked so long due to formaldehyde fixation leading to the apparent eviction of these proteins from mitotic DNA (332). We showed using single-molecule imaging, FRAP, and FLIP that residence times and mobility of bound SOX2 and OCT4 molecules in mitotic cells were similar to interphase cells, but the results suggest that the relative occupancy in mitotic cells was lower. Other work showed that SOX2 binds less (fewer percentage of molecules bound), in agreement with our study, but with roughly half the average residence time compared to interphase cells (333). The origin of this disagreement is unclear but may stem from technical differences in the experimental system (e.g. which protein tag was used, what level of overexpression) or in the details of the microscopic setup. Nevertheless, the values from both studies suggest that some degree of long-lived, and presumably site-specific, binding occurs in mitotic cells. We also tried to address this question using ChIP-seq of SOX2 in mitotic cells. Although we could identify some clear SOX2 peaks, there were three orders of magnitudes fewer peaks than in asynchronous cells. This difference could come from technical problems with ChIP-seq in mitotic cells or from true biological differences. As formaldehyde can lead to eviction of proteins as viewed by fluorescence microscopy, it is not unreasonable to speculate that this might also affect the crosslinking of molecules that are bound on chromatin, as is done in ChIP-seq. A subsequent paper tried to address this by preceding formaldehyde fixation with desuccinimidyl glutarate (DSG) (74). While DSG treatment enabled the retention of mitotically enriched TFs as seen by microscopy, it did not detect substantial mitotic binding of either OCT4 or SOX2 by ChIP-seq. It is noteworthy that another pluripotency TF, ESRRB, can be detected by ChIP-seq using either formaldehyde or DSG and formaldehyde, showing it is possible to detect specific genomic interactions on mitotic chromatin (334). However, it has not been formally excluded that formaldehyde treatment may somehow disrupt the interactions of some TFs, such as OCT4 and SOX2 but not ESRRB, with mitotic chromatin. Furthermore, ChIP-seq on mitotic cells is always performed on Nocodazole-synchronized cell populations enriched in prometaphase. It is also possible that this protocol somehow perturbs the binding of OCT4 and SOX2 or that they may bind more at other phases of mitosis not assessed. Fully answering this question will require native assays devoid of fixation or synchronization. It is also entirely possible that the mitotic association observed in the microscope may be completely or mostly represented by non-specific binding

events. In favor of this scenario is the fact that the few sites occupied by OCT4 and SOX2 (in the order of 10'000-100'000 according to ChIP-seq) is a miniscule fraction of the entire genome, while the proteins appear to coat most of mitotic chromatin. Furthermore, OCT4 was found to be enriched in pericentric heterochromatin where it would most likely not be able to access its binding sites (74). OCT4 is phosphorylated by Aurkb at G2/M phase, which decreases its binding, which would be consistent with at least a decrease in specific binding during mitosis (130). We have shown in a separate study that association to mitotic chromatin as determined by fluorescence microscopy is correlated to several properties of TFs in interphase including binding on-rates, mobility on DNA, and genome occupancy (97). One of the best predictors of mitotic binding in this study was the charge of the DNA binding domain. Manipulating just a few residues to increase the positive charge of TFs increased their association to mitotic chromatin. This suggests that (i) mitotic binding is at least partially representative of non-specific binding events and (ii) non-specific binding helps TFs bind specifically to DNA, presumably by increasing the ability to explore more binding sites in the genome. Finally, I would like to point out that one study showed robust mitotic ChIP-seq data for OCT4 and SOX2 (as well as KLF4) using apparently very similar conditions to those used by us (formaldehyde fixation) and the other group who showed little or no mitotic binding, and thus directly conflicts with ours and others observations (69). At the moment, there is thus no consensus in this matter and the exact role of these proteins during mitosis needs to be further explored.

Depletion of OCT4 and SOX2 at the M-G1 transition affects cell fate

Using the mitotic degron from cyclin B1 fused to SOX2, we could generate cell lines in which SOX2 was degraded at each metaphase-anaphase transition (SOX2-MD). When compared to average protein level-matched control cell lines in which SOX2 was not degraded at this phase, SOX2-MD cells were less capable of sustaining pluripotency. Furthermore, overexpression of SOX2-MD was unable to induce neuroectodermal differentiation, unlike its control. Another group also showed that OCT-MD constructs are deficient in pluripotency maintenance (69) (later confirmed by us), and that when overexpressed together with SOX2, KLF4, and c-MYC is less potent in reprogramming somatic cells to induced pluripotent stem cells. In contrast, we found no difference in reprogramming efficiency when overexpressing SOX2-MD. These data show that the M-G1 transition can be important in cell fate control, but that it is context-dependent. Importantly, we cannot exclude that degradation for similar durations at other points of the cell cycle would not lead to the same phenotypic outcomes. Such experiments are technically challenging because (i) there are not equivalent degron systems for every cell cycle phase and (ii) recovery after degradation must be complete before the end of the cell cycle. However, we believe these results are important because previous studies have merely investigated the impact on transcriptional reactivation (which was often relatively minor) and sometimes using prolonged depletion strategies that have a broader impact across the cell cycle than the ones described here (160, 177, 334). We also tried to assess the impact on transcriptional reactivation after mitosis by blocking SOX2-MD and control cells in mitosis and then releasing them (335). We found only a modest effect on gene reactivation after mitosis for some genes, but this may be hampered by cell cycle synchronization being toxic to ES cells, as we found very poor cell survival. We have not been able to address the function of OCT4 and SOX2 specifically during mitosis, due to lack of tools to interfere with their function only at this short cell cycle phase.

Low-amplitude fluctuations of OCT4 and SOX2 concentrations can affect cell fate outcomes

ES cells can retain their pluripotent state across a large range of OCT4 and SOX2 concentrations. Normally, cells fluctuate over a two to three-fold range of protein levels and the levels of OCT4 and SOX2 are mostly independent. We found that, if challenged to differentiate, the levels of these proteins can bias the cells toward different cell fates. This is an important observation, as it is likely to explain some of the heterogeneity observed when differentiating ES cells. Heterogeneity in differentiation outcomes is hugely problematic for both fundamental research and for potential biomedical applications of stem cells in the future. Understanding the basis for this heterogeneity will thus be important in order to find ways to combat it. While the fluctuating levels of other proteins have been implicated in cell fate decisions (135, 136), these were larger fluctuations happening on longer time scales, and in the case of NANOG signify transitions between different states of

pluripotency. The small differences in protein levels observed in our studies is thus a case of small-scale fluctuations around the mean in phenotypically homogeneous cell populations, which rapidly readjust, that affect the propensity of cells to differentiate to certain lineages. Interestingly, we found that these small fluctuations also affected the chromatin accessibility landscape. High OCT4 levels were correlated to high levels of chromatin accessibility at regions close to differentiation genes, suggesting these may get primed for activation upon differentiation. This hints at a possible mechanism for how fluctuations of TF levels can affect cell fate, although other contributing factors are also likely to be involved and should be explored. It is important to note that the impact OCT4 and SOX2 protein levels had on the differentiation outcome was highly context-dependent. Cells sorted in G1 showed a higher tendency to differentiate towards NE and ME when OCT4 levels were high and NE when SOX2 levels were high, but this effect was almost abolished when sorting cells in S phase. This is consistent with the impact of OCT4 and SOX2 upon depletion in the early phases of the cell cycle (M-G1 transition) observed in our studies and suggests that the early phases of the cell cycle are particularly important to regulate cell fate. Furthermore, when induced to differentiate into specific lineages using pathway agonists or antagonists, the effects were different (SOX2 levels affecting mesendoderm commitment upon directed differentiation using CHIR) or non-existent (upon directed differentiation toward neuroectoderm using a TGF-beta pathway inhibitor and bFGF). These results highlight the importance and context-dependence of TFs in cell fate commitment, and their interaction with other regulatory layers including the cell cycle.

OCT4 regulates highly dynamic chromatin accessibility throughout the cell cycle

Using rapid degradation of OCT4 and sorting cells in different cell cycle phases, we found that the reduction in chromatin accessibility was highly similar across the cell cycle. Furthermore, the different clusters of loci we identified based on the effect of M-G1 degradation behaved similarly across the cell cycle, indicating that highly OCT4-dependent genomic regions are not particularly sensitive to OCT4 loss at the beginning of the cell cycle but require OCT4 all the time. One caveat to this observation is that the magnitude of change in the MD system used for M-G1 degradation cannot be directly compared to the AID system. The control levels of OCT4 are not equal across these cell lines and degradation kinetics are unlikely to be exactly the same. In addition, the length of time OCT4 is depleted is different in the two experiments. We can also not exclude that certain regions are slightly more sensitive to OCT4 loss at certain cell cycle phases, or that more temporally fine-tuned depletion would reveal other results. Nevertheless, there is no evidence from these studies of differential regulation of chromatin accessibility across the cell cycle. This was an unexpected result based on the previous observations that chromatin accessibility is decreased during prometaphase and just after replication. It would have been feasible to imagine that OCT4 depletion during or just after these phases would have had a larger impact on chromatin accessibility than during phases when chromatin is already mostly open. The lack of differences highly suggests that regulation of chromatin accessibility is a very dynamic process, such that OCT4 depletion at any moment has a similar impact. This is substantiated by our data showing that loss of accessibility is highly temporally correlated to the loss of OCT4 protein, and our observation that high endogenous levels of OCT4 and SOX2 (which fluctuate on the time scale of hours) can lead to higher accessibility of certain regulatory elements. It is unclear whether regulation of chromatin accessibility is similarly dynamic in non-dividing cells that do not encounter the obstacles of DNA replication and chromatin compaction, and are also usually terminally differentiated and thus have other mechanisms to prevent cell fate changes. It is interesting to compare the dynamic loss of accessibility to the recovery of accessibility after OCT4 return. While this recovery operated at similar time scales (hours), some loci could not recover accessibility fully (cluster 1) or at all (cluster 3). The mechanisms behind this are currently unclear. It is worth noting that other factors must be involved in the regulation of chromatin accessibility across the cell cycle as OCT4 only regulates parts of the active regions in the genome and it is not present in all cell types. Furthermore, promoters remained almost completely unaffected by even prolonged OCT4 depletion. It remains untested whether these other factors, which include pioneer factors and potentially other categories, display the same dynamic properties as OCT4. Not all pioneers function the same, as evidenced by PAX7 that operates at much longer time scales (hours to days), with regions that can be maintained in an open state even 20 days after activation and subsequent removal of PAX7 (271). However, it is worth noting that after

PAX7 loss these regions are slightly less open than when PAX7 is present, suggesting that PAX7 may also play some quantitative role in their accessibility maintenance even though it is not fully necessary to keep them open. Overall, these results suggest that chromatin accessibility can be highly dynamic and dependent on single TFs.

5.2 Potential future research avenues

The experiments and conclusions presented in this thesis provide answers for specific questions regarding the regulation of ES cells by the TFs OCT4 and SOX2, and hint at possible general mechanisms of gene regulation (e.g. the dynamic regulation of chromatin accessibility). Simultaneously, many new questions have arisen from these findings that may be addressed in future studies. Here I will propose what I consider the most pertinent avenues of exploration to understand these processes.

Determinants of differential regulation of chromatin accessibility by OCT4 and SOX2

We identified multiple different groups and classes of genomic regions whose chromatin accessibility is specifically affected by the levels of OCT4 and SOX2 in different ways. This includes loci dependent on OCT4, SOX2, or both after long-term depletion and clusters of loci differentially responding to mitotic depletion of OCT4. I put a lot of effort into trying to identify determinants of these different regions, mainly looking for sequence motifs and binding by other factors, but these explorations yielded surprisingly little insight. Notably, using machine learning on over 3'500 ChIP-seq datasets in ES cells provided poor predictive power to identify groups of regions that are differentially sensitive to endogenous fluctuations of OCT4 and SOX2. This suggests that (i) binding per se may be insufficient to predict functionality and (ii) the combinatorial complexity operating at the scales of gene regulation may be too large to reliably define all the factors that impact probabilistic functional interactions. Nevertheless, this is an important area of research that should be explored further. I propose the following measures to possibly shed light on this matter: (i) MNase-seq to determine the precise nucleosomal orientation in relation to the sequence motif and how this is altered upon depletion of a factor, (ii) determination of the nature of the binding of TFs to DNA, specifically direct vs indirect binding modes and simultaneous vs sequential co-binding sites, using both computational methods and re-ChIP experiments (e.g. sequential pulldown of SOX2, then OCT4), and (iii) "reverse ChIP" experiments to identify all proteins bound to a region by pulling down specific DNA sequences and performing mass spectrometry, which could identify indirect binding partners without motifs at the region. Combining TF binding data with other sources such as Hi-C to measure chromosome conformation may also give extra predictive power.

The functional role of mitotic chromosome binding by transcription factors

This question has eluded scientists since the discovery of mitotically bound TFs. The difficulty in specifically depleting TFs during mitosis, which is very short, and immediately restoring the function after division is the main factor hindering these investigations. Novel tools must be developed, such as optogenetic disruption strategies combined with precise cell-cycle sorting or synchronization methods. Several hypotheses for the role of TFs in mitosis may be tested with such systems, including but not limited to (i) increased association rates after mitosis, (ii) maintenance of chromatin accessibility during mitosis, and (iii) maintenance of histone marks or other chromatin features during mitosis. It will also be important to use native (non-fixed) cells to study mitotic chromatin binding to definitively determine the relative contribution of specific and non-specific binding. Ideally, this should occur in non-synchronized cells.

Modes of co-factor recruitment by OCT4 and SOX2

OCT4 and SOX2 recognize their binding sites in the genome and can recruit chromatin remodelers and other factors. How this recruitment occurs is not entirely clear. It will be important to understand which residues of the proteins are required for recruitment and how this association is regulated. For example, is OCT4 directly associating with the BAF chromatin remodeling complex to bring it to the region? Recent studies have suggested that transcriptional regulators, including OCT4, can exist

inside phase-separated “condensates” in the nucleus which co-localize with genomic regulatory regions, and it was suggested that the intrinsically disordered domains present in many activation domains of TFs operate via inducing phase separation (336–338). Studying the connection between phase separation and recruitment of co-factors, including chromatin remodelers and other transcriptional regulatory factors, in the regulation of accessibility and transcription will be difficult but important. Some possible experiments to address recruitment include: (i) studying the ability of OCT4 mutants, lacking specific regions of the protein, to associate with various factors (including BRG1 and nucleosomal DNA) in vitro and in vivo, (ii) studying the ability of OCT4 mutants to phase separate in vitro and in vivo, and (iii) single-molecule binding measurements of OCT4 and chromatin remodelers in vitro and in vivo to determine binding parameters in the presence or absence of each other. It will also be interesting to study the interaction between pioneer factors and histone marks and their readers and writers.

The link between chromatin accessibility and transcription

While we have studied the impact of pioneer factors on chromatin accessibility and cell fate, there is a missing link - transcription. It is entirely possible that many of the changes in chromatin accessibility we observe do not have a direct impact on transcription. This is particularly interesting to consider in relation to the observation that enhancers close to differentiation genes get opened up when OCT4 levels are high. These genes are not (or barely) transcribed in ES cells and this may serve as a form of “priming” for later activation. It will be important to understand which enhancers regulate which genes and when. This will require integrative analysis of multiple datasets, including time-resolved measurements of ongoing transcription (for example using run-on methods such as GRO-seq) upon OCT4 depletion. One limitation of our current methods is the relatively low expression level of OCT4 in our overexpression constructs, which yielded clear results but relatively small changes in accessibility and are thus likely to have an even smaller impact on transcription. Novel constructs with a higher dynamic range of protein levels, including possible knock-ins, of OCT4 with inducible depletion would be required to accurately measure such changes.

Exploration of the dynamic regulation of chromatin accessibility

We have hypothesized that the dynamic regulation of chromatin accessibility by OCT4 may be a particular feature of ES cells due to their position in the developmental progression of the embryo, which requires them to rapidly change phenotype in response to signals. It will be interesting to explore whether other pioneer factors share the same function, especially in non-dividing cells. This could be achieved using e.g. cells that can enter quiescence such as fibroblasts when grown at contact-inhibited densities. Similar depletion experiments to those we have done would be appropriate here as well, but would require identifying the major regulatory pioneer factors for the specific cell type in question. It would also be interesting to study the regulation of chromatin accessibility at promoters, which appears more stable across the cell cycle and to TF depletions. Another avenue of exploration would be the even shorter timescales of nucleosome dynamics. For example, how does the dynamic exchange of nucleosomes relate to the short (second-scale) binding times of OCT4. Do chromatin remodelers bind at similar time scales? How long are nucleosomes evicted for? How do other factors interact with the region once nucleosomes are evicted and how does that affect nucleosome occupancy? TFs can recruit factors such as histone modifiers which deposit posttranslational histone modifications which are then recognized by histone readers that recruit other proteins or even the TFs themselves. For example, we observed concomitant loss of H3K27ac with chromatin accessibility upon OCT4/SOX2 depletion. Such feedback loops likely contribute to the action of pioneer factors, but this has not been extensively studied. Indeed, some of the effects on chromatin accessibility we see may be indirect due to loss of binding of other factors. Future research should inquire as to the relative contribution of these different aspects in chromatin regulation. Such investigations would require advanced imaging setups and some way to track nucleosome occupancy in situ. Due to the density of nucleosomes, this may be unfeasible and would perhaps require in vitro setups emulating the regulatory chromatin environment. Specific and temporally sensitive removal or deposition of histone modifications to specific regions in the genome, such as via catalytically inactive Cas9 fused to histone writers, may also shed light on these mechanisms.

References

1. J. B. Gurdon, The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J Embryol Exp Morphol.* **10**, 622–640 (1962).
2. Molecular Biology of the cell - Books - NCBI, (available at <https://www.ncbi.nlm.nih.gov/books/NBK21054/?term=Molecular%20Biology%20of%20the%20cell>).
3. R. M. & R. Philips, » How big are genomes?, (available at <http://book.bionumbers.org/how-big-are-genomes/>).
4. G. Cavalli, E. Heard, Advances in epigenetics link genetics to the environment and disease. *Nature.* **571**, 489–499 (2019).
5. M. Iwafuchi-Doi, K. S. Zaret, Cell fate control by pioneer transcription factors. *Development.* **143**, 1833–1837 (2016).
6. Z. D. Smith, A. Meissner, DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
7. E. Calo, J. Wysocka, Modification of enhancer chromatin: what, how, and why? *Mol. Cell.* **49**, 825–837 (2013).
8. S. Y. Sokol, Maintaining embryonic stem cell pluripotency with Wnt signaling. *Development.* **138**, 4341–4350 (2011).
9. J.-L. Maître, H. Turlier, R. Illukkumbura, B. Eismann, R. Niwayama, F. Nédélec, T. Hiiragi, Asymmetric division of contractile domains couples cell positioning and fate specification. *Nature.* **536**, 344–348 (2016).
10. A. Fiszbein, L. E. Giono, A. Quaglino, B. G. Berardino, L. Sigaut, C. von Bilderling, I. E. Schor, J. H. E. Steinberg, M. Rossi, L. I. Pietrasanta, J. J. Caramelo, A. Srebrow, A. R. Kornblihtt, Alternative Splicing of G9a Regulates Neuronal Differentiation. *Cell Rep.* **14**, 2797–2808 (2016).
11. P. Sampath, D. K. Pritchard, L. Pabon, H. Reinecke, S. M. Schwartz, D. R. Morris, C. E. Murry, A hierarchical network controls protein translation during murine embryonic stem cell self-renewal and differentiation. *Cell Stem Cell.* **2**, 448–460 (2008).
12. C. Furusawa, K. Kaneko, A dynamical-systems view of stem cell biology. *Science.* **338**, 215–217 (2012).
13. S. J. Arnold, E. J. Robertson, Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* **10**, 91–103 (2009).
14. A. Grapin-Botton, in *StemBook* (Harvard Stem Cell Institute, Cambridge (MA), 2008; <http://www.ncbi.nlm.nih.gov/books/NBK27052/>).
15. P. P. L. Tam, D. A. F. Loebel, Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* **8**, 368–381 (2007).
16. P. P. L. Tam, D. A. F. Loebel, S. S. Tanaka, Building the mouse gastrula: signals, asymmetry and lineages. *Curr. Opin. Genet. Dev.* **16**, 419–425 (2006).

17. G. R. Martin, Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7634–7638 (1981).
18. M. J. Evans, M. H. Kaufman, Establishment in culture of pluripotential cells from mouse embryos. *Nature*. **292**, 154–156 (1981).
19. A. G. Smith, J. K. Heath, D. D. Donaldson, G. G. Wong, J. Moreau, M. Stahl, D. Rogers, Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature*. **336**, 688–690 (1988).
20. Q.-L. Ying, J. Wray, J. Nichols, L. Battle-Morera, B. Doble, J. Woodgett, P. Cohen, A. Smith, The ground state of embryonic stem cell self-renewal. *Nature*. **453**, 519–523 (2008).
21. T. Boroviak, R. Loos, P. Bertone, A. Smith, J. Nichols, The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat. Cell Biol.* **16**, 516–528 (2014).
22. A. De Los Angeles, F. Ferrari, R. Xi, Y. Fujiwara, N. Benvenisty, H. Deng, K. Hochedlinger, R. Jaenisch, S. Lee, H. G. Leitch, M. W. Lensch, E. Lujan, D. Pei, J. Rossant, M. Wernig, P. J. Park, G. Q. Daley, Hallmarks of pluripotency. *Nature*. **525**, 469–478 (2015).
23. I. G. M. Brons, L. E. Smithers, M. W. B. Trotter, P. Rugg-Gunn, B. Sun, S. M. Chuva de Sousa Lopes, S. K. Howlett, A. Clarkson, L. Ahrlund-Richter, R. A. Pedersen, L. Vallier, Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*. **448**, 191–195 (2007).
24. P. J. Tesar, J. G. Chenoweth, F. A. Brook, T. J. Davies, E. P. Evans, D. L. Mack, R. L. Gardner, R. D. G. McKay, New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*. **448**, 196–199 (2007).
25. R. D. Kornberg, Eukaryotic transcriptional control. *Trends Cell Biol.* **9**, M46-49 (1999).
26. I. Jonkers, J. T. Lis, Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 167–177 (2015).
27. T. I. Lee, R. A. Young, Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34**, 77–137 (2000).
28. S. Buratowski, Progression through the RNA polymerase II CTD cycle. *Mol. Cell.* **36**, 541–546 (2009).
29. W. Schaffner, Enhancers, enhancers - from their discovery to today's universe of transcription enhancers. *Biol. Chem.* **396**, 311–327 (2015).
30. C. Buecker, J. Wysocka, Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet.* **28**, 276–284 (2012).
31. E. E. M. Furlong, M. Levine, Developmental enhancers and chromosome topology. *Science*. **361**, 1341–1345 (2018).
32. W. de Laat, D. Duboule, Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*. **502**, 499–506 (2013).
33. R. E. Thurman, E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernet, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T. K. Canfield, M. Diegel, D. Dunn, A. K. Ebersol, T. Frum, E. Giste, A. K. Johnson, E. M. Johnson, T. Kutayavin, B. Lajoie, B.-K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E.

- D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, A. Safi, M. E. Sanchez, A. Sanyal, A. Shafer, J. M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M. O. Dorschner, R. S. Hansen, P. A. Navas, G. Stamatoyannopoulos, V. R. Iyer, J. D. Lieb, S. R. Sunyaev, J. M. Akey, P. J. Sabo, R. Kaul, T. S. Furey, J. Dekker, G. E. Crawford, J. A. Stamatoyannopoulos, The accessible chromatin landscape of the human genome. *Nature*. **489**, 75–82 (2012).
34. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. **389**, 251–260 (1997).
 35. S. L. Klemm, Z. Shipony, W. J. Greenleaf, Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
 36. E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. **326**, 289–293 (2009).
 37. J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, B. Ren, Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature*. **485**, 376–380 (2012).
 38. B. van Steensel, E. E. M. Furlong, The role of transcription in shaping the spatial organization of the genome. *Nat. Rev. Mol. Cell Biol.* **20**, 327–337 (2019).
 39. J. S. Becker, D. Nicetto, K. S. Zaret, H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet.* **32**, 29–41 (2016).
 40. M. Shogren-Knaak, H. Ishii, J.-M. Sun, M. J. Pazin, J. R. Davie, C. L. Peterson, Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*. **311**, 844–847 (2006).
 41. S. Bonn, R. P. Zinzen, C. Girardot, E. H. Gustafson, A. Perez-Gonzalez, N. Delhomme, Y. Ghavi-Helm, B. Wilczyński, A. Riddell, E. E. M. Furlong, Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.* **44**, 148–156 (2012).
 42. D. Nicolas, B. Zoller, D. M. Suter, F. Naef, Modulation of transcriptional burst frequency by histone acetylation. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7153–7158 (2018).
 43. T. Wu, Y. F. Kamikawa, M. E. Donohoe, Brd4's Bromodomains Mediate Histone H3 Acetylation and Chromatin Remodeling in Pluripotent Cells through P300 and Brg1. *Cell Rep.* **25**, 1756–1771 (2018).
 44. S. Venkatesh, J. L. Workman, Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 178–189 (2015).
 45. A. J. Bannister, T. Kouzarides, Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).
 46. H. W. King, R. J. Klose, The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *Elife*. **6** (2017), doi:10.7554/eLife.22631.
 47. F. Mueller, T. J. Stasevich, D. Mazza, J. G. McNally, Quantifying transcription factor kinetics: at work or at play? *Crit. Rev. Biochem. Mol. Biol.* **48**, 492–514 (2013).

48. E. Meshorer, D. Yellajoshula, E. George, P. J. Scambler, D. T. Brown, T. Misteli, Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell.* **10**, 105–116 (2006).
49. R. B. Deal, J. G. Henikoff, S. Henikoff, Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science.* **328**, 1161–1164 (2010).
50. A. M. Deaton, M. Gómez-Rodríguez, J. Mieczkowski, M. Y. Tolstorukov, S. Kundu, R. I. Sadreyev, L. E. Jansen, R. E. Kingston, Enhancer regions show high histone H3.3 turnover that changes during differentiation. *Elife.* **5** (2016), doi:10.7554/eLife.15316.
51. K. R. Stewart-Morgan, N. Reverón-Gómez, A. Groth, Transcription Restart Establishes Chromatin Accessibility after DNA Replication. *Molecular Cell* (2019), doi:10.1016/j.molcel.2019.04.033.
52. C. Alabert, T. K. Barth, N. Reverón-Gómez, S. Sidoli, A. Schmidt, O. N. Jensen, A. Imhof, A. Groth, Two distinct modes for propagation of histone PTMs across the cell cycle. *Genes Dev.* **29**, 585–590 (2015).
53. N. Reverón-Gómez, C. González-Aguilera, K. R. Stewart-Morgan, N. Petryk, V. Flury, S. Graziano, J. V. Johansen, J. S. Jakobsen, C. Alabert, A. Groth, Accurate Recycling of Parental Histones Reproduces the Histone Modification Landscape during DNA Replication. *Mol. Cell.* **72**, 239-249.e5 (2018).
54. L. Liu, W. Michowski, A. Kolodziejczyk, P. Sicinski, The cell cycle in stem cell proliferation, pluripotency and differentiation. *Nat Cell Biol.* **21**, 1060–1067 (2019).
55. A. B. Alber, D. M. Suter, Dynamics of protein synthesis and degradation through the cell cycle. *Cell Cycle.* **18**, 784–794 (2019).
56. J. White, S. Dalton, Cell cycle control of embryonic stem cells. *Stem Cell Rev.* **1**, 131–138 (2005).
57. M. Ter Huurne, J. Chappell, S. Dalton, H. G. Stunnenberg, Distinct Cell-Cycle Control in Two Different States of Mouse Pluripotency. *Cell Stem Cell.* **21**, 449-455.e4 (2017).
58. S. Pauklin, L. Vallier, The cell-cycle state of stem cells determines cell fate propensity. *Cell.* **155**, 135–147 (2013).
59. A. M. Singh, S. Dalton, The cell cycle and Myc intersect with mechanisms that regulate pluripotency and reprogramming. *Cell Stem Cell.* **5**, 141–149 (2009).
60. D. Coronado, M. Godet, P.-Y. Bourillot, Y. Tapponnier, A. Bernat, M. Petit, M. Afanassieff, S. Markossian, A. Malashicheva, R. Iacone, K. Anastassiadis, P. Savatier, A short G1 phase is an intrinsic determinant of naïve embryonic stem cell pluripotency. *Stem Cell Res.* **10**, 118–131 (2013).
61. R. Scognamiglio, N. Cabezas-Wallscheid, M. C. Thier, S. Altamura, A. Reyes, Á. M. Prendergast, D. Baumgärtner, L. S. Carnevalli, A. Atzberger, S. Haas, L. von Paleske, T. Boroviak, P. Wörsdörfer, M. A. G. Essers, U. Kloz, R. N. Eisenman, F. Edenhofer, P. Bertone, W. Huber, F. van der Hoeven, A. Smith, A. Trumpp, Myc Depletion Induces a Pluripotent Dormant State Mimicking Diapause. *Cell.* **164**, 668–680 (2016).
62. A. Bulut-Karslioglu, S. Biechele, H. Jin, T. A. Macrae, M. Hejna, M. Gertsenstein, J. S. Song, M. Ramalho-Santos, Inhibition of mTOR induces a paused pluripotent state. *Nature.* **540**, 119–123 (2016).

63. D. Egli, G. Birkhoff, K. Eggen, Mediators of reprogramming: transcription factors and transitions through mitosis. *Nat. Rev. Mol. Cell Biol.* **9**, 505–516 (2008).
64. J. C. Rivera-Mulia, D. M. Gilbert, Replication timing and transcriptional control: beyond cause and effect—part III. *Curr Opin Cell Biol.* **40**, 168–178 (2016).
65. N. Petryk, M. Dalby, A. Wenger, C. B. Stromme, A. Strandsby, R. Andersson, A. Groth, MCM2 promotes symmetric inheritance of modified histones during DNA replication. *Science*. **361**, 1389–1392 (2018).
66. V. Behera, A. J. Stonestrom, N. Hamagami, C. C. Hsiung, C. A. Keller, B. Giardine, S. Sidoli, Z.-F. Yuan, N. V. Bhanu, M. T. Werner, H. Wang, B. A. Garcia, R. C. Hardison, G. A. Blobel, Interrogating Histone Acetylation and BRD4 as Mitotic Bookmarks of Transcription. *Cell Rep.* **27**, 400–415.e5 (2019).
67. E. Javasky, I. Shamir, S. Gandhi, S. Egri, O. Sandler, S. B. Rothbart, N. Kaplan, J. D. Jaffe, A. Goren, I. Simon, Study of mitotic chromatin supports a model of bookmarking by histone modifications and reveals nucleosome deposition patterns. *Genome Res.* **28**, 1455–1466 (2018).
68. P. A. Ginno, L. Burger, J. Seebacher, V. Iesmantavicius, D. Schübeler, Cell cycle-resolved chromatin proteomics reveals the extent of mitotic preservation of the genomic regulatory landscape. *Nat Commun.* **9**, 4048 (2018).
69. Y. Liu, B. Pelham-Webb, D. C. Di Giammartino, J. Li, D. Kim, K. Kita, N. Saiz, V. Garg, A. Doane, P. Giannakakou, A.-K. Hadjantonakis, O. Elemento, E. Apostolou, Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors in Pluripotent Stem Cells. *Cell Rep.* **19**, 1283–1293 (2017).
70. K. C. Palozola, G. Donahue, H. Liu, G. R. Grant, J. S. Becker, A. Cote, H. Yu, A. Raj, K. S. Zaret, Mitotic transcription and waves of gene reactivation during mitotic exit. *Science*. **358**, 119–122 (2017).
71. C. C.-S. Hsiung, C. R. Bartman, P. Huang, P. Ginart, A. J. Stonestrom, C. A. Keller, C. Face, K. S. Jahn, P. Evans, L. Sankaranarayanan, B. Giardine, R. C. Hardison, A. Raj, G. A. Blobel, A hyperactive transcriptional state marks genome reactivation at the mitosis-G1 transition. *Genes Dev.* **30**, 1423–1439 (2016).
72. C. C.-S. Hsiung, C. S. Morrissey, M. Udugama, C. L. Frank, C. A. Keller, S. Baek, B. Giardine, G. E. Crawford, M.-H. Sung, R. C. Hardison, G. A. Blobel, Genome accessibility is widely preserved and locally modulated during mitosis. *Genome Res.* **25**, 213–225 (2015).
73. M. E. Oomen, A. S. Hansen, Y. Liu, X. Darzacq, J. Dekker, CTCF sites display cell cycle–dependent dynamics in factor binding and nucleosome positioning. *Genome Res.* **29**, 236–249 (2019).
74. N. Festuccia, N. Owens, T. Papadopoulou, I. Gonzalez, A. Tachtsidi, S. Vandoermel-Pournin, E. Gallego, N. Gutierrez, A. Dubois, M. Cohen-Tannoudji, P. Navarro, Transcription factor activity and nucleosome organization in mitosis. *Genome Res.* **29**, 250–260 (2019).
75. J. T. Kadonaga, Regulation of RNA Polymerase II Transcription by Sequence-Specific DNA Binding Factors. *Cell*. **116**, 247–257 (2004).
76. P. J. Mitchell, R. Tjian, Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*. **245**, 371–378 (1989).

77. A. Boija, I. A. Klein, B. R. Sabari, A. Dall'Agnesse, E. L. Coffey, A. V. Zamudio, C. H. Li, K. Shrinivas, J. C. Manteiga, N. M. Hannett, B. J. Abraham, L. K. Afeyan, Y. E. Guo, J. K. Rimel, C. B. Fant, J. Schuijers, T. I. Lee, D. J. Taatjes, R. A. Young, Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*. **175**, 1842-1855.e16 (2018).
78. S. Payankulam, L. M. Li, D. N. Arnosti, Transcriptional repression: conserved and evolved features. *Curr Biol*. **20**, R764–R771 (2010).
79. O. Tidin, E. T. Friman, F. Naef, D. M. Suter, Quantitative relationships between SMAD dynamics and target gene activation kinetics in single live cells. *Sci Rep*. **9**, 5372 (2019).
80. S. A. Lambert, A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T. R. Hughes, M. T. Weirauch, The Human Transcription Factors. *Cell*. **172**, 650–665 (2018).
81. M. M. Babu, N. M. Luscombe, L. Aravind, M. Gerstein, S. A. Teichmann, Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol*. **14**, 283–291 (2004).
82. J. T. Kadonaga, The transformation of the DNA template in RNA polymerase II transcription: a historical perspective. *Nat. Struct. Mol. Biol*. **26**, 766–770 (2019).
83. L. A. Cirillo, F. R. Lin, I. Cuesta, D. Friedman, M. Jarnik, K. S. Zaret, Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4. *Molecular Cell*. **9**, 279–289 (2002).
84. K. S. Zaret, J. S. Carroll, Pioneer transcription factors: establishing competence for gene expression. *Genes Dev*. **25**, 2227–2241 (2011).
85. A. Mayran, J. Drouin, Pioneer transcription factors shape the epigenetic landscape. *J. Biol. Chem*. **293**, 13795–13804 (2018).
86. M. Iwafuchi-Doi, K. S. Zaret, Pioneer transcription factors in cell reprogramming. *Genes Dev*. **28**, 2679–2692 (2014).
87. A. Soufi, M. F. Garcia, A. Jaroszewicz, N. Osman, M. Pellegrini, K. S. Zaret, Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell*. **161**, 555–568 (2015).
88. A. Soufi, G. Donahue, K. S. Zaret, Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell*. **151**, 994–1004 (2012).
89. M. Fernandez Garcia, C. D. Moore, K. N. Schulz, O. Alberto, G. Donague, M. M. Harrison, H. Zhu, K. S. Zaret, Structural Features of Transcription Factors Associating with Nucleosome Binding. *Mol. Cell*. **75**, 921-932.e6 (2019).
90. M. Iwafuchi-Doi, G. Donahue, A. Kakumanu, J. A. Watts, S. Mahony, B. F. Pugh, D. Lee, K. H. Kaestner, K. S. Zaret, The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol Cell*. **62**, 79–91 (2016).
91. M. Takaku, S. A. Grimm, T. Shimbo, L. Perera, R. Menafrá, H. G. Stunnenberg, T. K. Archer, S. Machida, H. Kurumizaka, P. A. Wade, GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler. *Genome Biol*. **17**, 36 (2016).

92. D. L. C. van den Berg, T. Snoek, N. P. Mullin, A. Yates, K. Bezstarosti, J. Demmers, I. Chambers, R. A. Poot, An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell*. **6**, 369–381 (2010).
93. M. A. Martínez-Balbás, A. Dey, S. K. Rabindran, K. Ozato, C. Wu, Displacement of sequence-specific transcription factors from mitotic chromatin. *Cell*. **83**, 29–38 (1995).
94. M. Raccaud, D. M. Suter, Transcription factor retention on mitotic chromosomes: regulatory mechanisms and impact on cell fate decisions. *FEBS Lett*. **592**, 878–887 (2018).
95. J. M. Caravaca, G. Donahue, J. S. Becker, X. He, C. Vinson, K. S. Zaret, Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. *Genes Dev*. **27**, 251–260 (2013).
96. S. Kadauke, M. I. Udugama, J. M. Pawlicki, J. C. Achtman, D. P. Jain, Y. Cheng, R. C. Hardison, G. A. Blobel, Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. *Cell*. **150**, 725–737 (2012).
97. M. Raccaud, E. T. Friman, A. B. Alber, H. Agarwal, C. Deluz, T. Kuhn, J. C. M. Gebhardt, D. M. Suter, Mitotic chromosome binding predicts transcription factor properties in interphase. *Nat Commun*. **10**, 487 (2019).
98. E. E. Swinstead, V. Paakinaho, D. M. Presman, G. L. Hager, Pioneer factors and ATP-dependent chromatin remodeling factors interact dynamically: A new perspective: Multiple transcription factors can effect chromatin pioneer functions through dynamic interactions with ATP-dependent chromatin remodeling factors. *Bioessays*. **38**, 1150–1157 (2016).
99. G. Martello, A. Smith, The nature of embryonic stem cells. *Annu. Rev. Cell Dev. Biol*. **30**, 647–675 (2014).
100. S.-J. Dunn, G. Martello, B. Yordanov, S. Emmott, A. G. Smith, Defining an essential transcription factor program for naïve pluripotency. *Science*. **344**, 1156–1160 (2014).
101. I. Chambers, D. Colby, M. Robertson, J. Nichols, S. Lee, S. Tweedie, A. Smith, Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*. **113**, 643–655 (2003).
102. I. Chambers, J. Silva, D. Colby, J. Nichols, B. Nijmeijer, M. Robertson, J. Vrana, K. Jones, L. Grotewold, A. Smith, Nanog safeguards pluripotency and mediates germline development. *Nature*. **450**, 1230–1234 (2007).
103. G. Martello, T. Sugimoto, E. Diamanti, A. Joshi, R. Hannah, S. Ohtsuka, B. Göttgens, H. Niwa, A. Smith, Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell*. **11**, 491–504 (2012).
104. S. Masui, S. Ohtsuka, R. Yagi, K. Takahashi, M. S. Ko, H. Niwa, Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. *BMC Dev Biol*. **8**, 45 (2008).
105. P. Zhang, R. Andrianakos, Y. Yang, C. Liu, W. Lu, Kruppel-like factor 4 (Klf4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression. *J. Biol. Chem*. **285**, 9180–9189 (2010).
106. D. Zeineddine, A. A. Hammoud, M. Mortada, H. Boeuf, The Oct4 protein: more than a magic stemness marker. *Am J Stem Cells*. **3**, 74–82 (2014).
107. G. J. Pan, Z. Y. Chang, H. R. Schöler, D. Pei, Stem cell pluripotency and transcription factor Oct4. *Cell Res*. **12**, 321–329 (2002).

108. L. Hou, Y. Srivastava, R. Jauch, Molecular basis for the genome engagement by Sox proteins. *Semin. Cell Dev. Biol.* **63**, 2–12 (2017).
109. T. K. Mistri, A. G. Devasia, L. T. Chu, W. P. Ng, F. Halbritter, D. Colby, B. Martynoga, S. R. Tomlinson, I. Chambers, P. Robson, T. Wohland, Selective influence of Sox2 on POU transcription factor binding in embryonic and neural stem cells. *EMBO Rep.* **16**, 1177–1191 (2015).
110. A. Gagliardi, N. P. Mullin, Z. Ying Tan, D. Colby, A. I. Kousa, F. Halbritter, J. T. Weiss, A. Felker, K. Bezstarosti, R. Favaro, J. Demmers, S. K. Nicolis, S. R. Tomlinson, R. A. Poot, I. Chambers, A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal. *EMBO J.* **32**, 2231–2247 (2013).
111. J. L. Cox, S. K. Mallanna, X. Luo, A. Rizzino, Sox2 Uses Multiple Domains to Associate with Proteins Present in Sox2-Protein Complexes. *PLOS ONE*. **5**, e15486 (2010).
112. J. Xu, J. Li, Y.-H. E. Loh, T. Zhang, H. Jiang, B. Fritsch, A. Ramakrishnan, L. Shen, P.-X. Xu, Brg1 controls neurosensory cell fate commitment and differentiation in the mammalian inner ear. *bioRxiv*, 434159 (2018).
113. F. Hammachi, G. M. Morrison, A. A. Sharov, A. Livigni, S. Narayan, E. P. Papapetrou, J. O'Malley, K. Kaji, M. S. H. Ko, M. Ptashne, J. M. Brickman, Transcriptional activation by Oct4 is sufficient for the maintenance and induction of pluripotency. *Cell Rep.* **1**, 99–109 (2012).
114. A. Rizzino, E. L. Wuebben, Sox2/Oct4: A delicately balanced partnership in pluripotent stem cells and embryogenesis. *Biochim. Biophys. Acta.* **1859**, 780–791 (2016).
115. H. Niwa, J. Miyazaki, A. G. Smith, Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* **24**, 372–376 (2000).
116. S. Masui, Y. Nakatake, Y. Toyooka, D. Shimosato, R. Yagi, K. Takahashi, H. Okochi, A. Okuda, R. Matoba, A. A. Sharov, M. S. H. Ko, H. Niwa, Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* **9**, 625–635 (2007).
117. K. Adachi, I. Nikaido, H. Ohta, S. Ohtsuka, H. Ura, M. Kadota, T. Wakayama, H. R. Ueda, H. Niwa, Context-Dependent Wiring of Sox2 Regulatory Networks for Self-Renewal of Embryonic and Trophoblast Stem Cells. *Molecular Cell.* **52**, 380–392 (2013).
118. A. Corsinotti, F. C. Wong, T. Tatar, I. Szczerbinska, F. Halbritter, D. Colby, S. Gogolok, R. Pantier, K. Liggat, E. S. Mirfazeli, E. Hall-Ponsole, N. P. Mullin, V. Wilson, I. Chambers, Distinct SoxB1 networks are required for naïve and primed pluripotency. *Elife.* **6** (2017), doi:10.7554/eLife.27746.
119. V. Karwacki-Neisius, J. Göke, R. Osorno, F. Halbritter, J. H. Ng, A. Y. Weiße, F. C. K. Wong, A. Gagliardi, N. P. Mullin, N. Festuccia, D. Colby, S. R. Tomlinson, H.-H. Ng, I. Chambers, 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).
120. A. Radziskeuskaya, G. L. B. Chia, R. L. dos Santos, T. W. Theunissen, L. F. C. Castro, J. Nichols, J. C. R. Silva, A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. *Nat. Cell Biol.* **15**, 579–590 (2013).
121. C. Mulas, G. Chia, K. A. Jones, A. C. Hodgson, G. G. Stirparo, J. Nichols, Oct4 regulates the embryonic axis and coordinates exit from pluripotency and germ layer specification in the mouse embryo. *Development.* **145**, dev159103 (2018).

122. M. Thomson, S. J. Liu, L.-N. Zou, Z. Smith, A. Meissner, S. Ramanathan, Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell*. **145**, 875–889 (2011).
123. F. Koch, M. Scholze, L. Wittler, D. Schifferl, S. Sudheer, P. Grote, B. Timmermann, K. Macura, B. G. Herrmann, Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. *Dev. Cell*. **42**, 514-526.e7 (2017).
124. M. Iwafuchi-Doi, K. Matsuda, K. Murakami, H. Niwa, P. J. Tesar, J. Aruga, I. Matsuo, H. Kondoh, Transcriptional regulatory networks in epiblast cells and during anterior neural plate development as modeled in epiblast stem cells. *Development*. **139**, 3926–3937 (2012).
125. H. B. Wood, V. Episkopou, Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. *Mech. Dev.* **86**, 197–201 (1999).
126. V. Graham, J. Khudyakov, P. Ellis, L. Pevny, SOX2 functions to maintain neural progenitor identity. *Neuron*. **39**, 749–765 (2003).
127. J. Ahlfeld, S. Filser, F. Schmidt, A. K. Wefers, D. J. Merk, R. Glaß, J. Herms, U. Schüller, Neurogenesis from Sox2 expressing cells in the adult cerebellar cortex. *Sci Rep*. **7**, 6137 (2017).
128. C. Guo, L. Liu, Y. Jia, X. Zhao, Q. Zhou, L. Wang, A novel variant of Oct3/4 gene in mouse embryonic stem cells. *Stem Cell Res*. **9**, 69–76 (2012).
129. S. A. Myers, S. Peddada, N. Chatterjee, T. Friedrich, K. Tomoda, G. Krings, S. Thomas, J. Maynard, M. Broeker, M. Thomson, K. Pollard, S. Yamanaka, A. L. Burlingame, B. Panning, SOX2 O-GlcNAcylation alters its protein-protein interactions and genomic occupancy to modulate gene expression in pluripotent cells. *eLife*. **5**, doi:10.7554/eLife.10647.
130. J. Shin, T. W. Kim, H. Kim, H. J. Kim, M. Y. Suh, S. Lee, H.-T. Lee, S. Kwak, S.-E. Lee, J.-H. Lee, H. Jang, E.-J. Cho, H.-D. Youn, Aurkb/PP1-mediated resetting of Oct4 during the cell cycle determines the identity of embryonic stem cells. *Elife*. **5**, e10877 (2016).
131. K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. **126**, 663–676 (2006).
132. A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, S. Tyagi, Stochastic mRNA synthesis in mammalian cells. *PLoS Biol*. **4**, e309 (2006).
133. D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, F. Naef, Mammalian genes are transcribed with widely different bursting kinetics. *Science*. **332**, 472–474 (2011).
134. A. Sigal, R. Milo, A. Cohen, N. Geva-Zatorsky, Y. Klein, Y. Liron, N. Rosenfeld, T. Danon, N. Perzov, U. Alon, Variability and memory of protein levels in human cells. *Nature*. **444**, 643–646 (2006).
135. A. Filipczyk, C. Marr, S. Hastreiter, J. Feigelman, M. Schwarzfischer, P. S. Hoppe, D. Loeffler, K. D. Kokkalis, M. Ende, B. Schauburger, O. Hilsenbeck, S. Skylaki, J. Hasenauer, K. Anastassiadis, F. J. Theis, T. Schroeder, Network plasticity of pluripotency transcription factors in embryonic stem cells. *Nat. Cell Biol*. **17**, 1235–1246 (2015).
136. J. E. Purvis, K. W. Karhohs, C. Mock, E. Batchelor, A. Loewer, G. Lahav, p53 dynamics control cell fate. *Science*. **336**, 1440–1444 (2012).
137. T. W. Mak, J. M. Penninger, P. S. Ohashi, Knockout mice: a paradigm shift in modern immunology. *Nat. Rev. Immunol*. **1**, 11–19 (2001).

138. B. H. Koller, P. Marrack, J. W. Kappler, O. Smithies, Normal development of mice deficient in beta 2M, MHC class I proteins, and CD8+ T cells. *Science*. **248**, 1227–1230 (1990).
139. L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, F. Zhang, Multiplex genome engineering using CRISPR/Cas systems. *Science*. **339**, 819–823 (2013).
140. H. Bouabe, K. Okkenhaug, Gene Targeting in Mice: a Review. *Methods Mol Biol*. **1064**, 315–336 (2013).
141. D. L. Buckley, K. Raina, N. Darricarrere, J. Hines, J. L. Gustafson, I. E. Smith, A. H. Miah, J. D. Harling, C. M. Crews, HaloPROTACS: Use of Small Molecule PROTACs to Induce Degradation of HaloTag Fusion Proteins. *ACS Chem. Biol*. **10**, 1831–1837 (2015).
142. L. A. Banaszynski, L. Chen, L. A. Maynard-Smith, A. G. Lisa Ooi, T. J. Wandless, A Rapid, Reversible, and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules. *Cell*. **126**, 995–1004 (2006).
143. A. J. Holland, D. Fachinetti, J. S. Han, D. W. Cleveland, Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells. *PNAS*. **109**, E3350–E3357 (2012).
144. S. G. Landt, G. K. Marinov, A. Kundaje, P. Kheradpour, F. Pauli, S. Batzoglou, B. E. Bernstein, P. Bickel, J. B. Brown, P. Cayting, Y. Chen, G. DeSalvo, C. Epstein, K. I. Fisher-Aylor, G. Euskirchen, M. Gerstein, J. Gertz, A. J. Hartemink, M. M. Hoffman, V. R. Iyer, Y. L. Jung, S. Karmakar, M. Kellis, P. V. Kharchenko, Q. Li, T. Liu, X. S. Liu, L. Ma, A. Milosavljevic, R. M. Myers, P. J. Park, M. J. Pazin, M. D. Perry, D. Raha, T. E. Reddy, J. Rozowsky, N. Shores, A. Sidow, M. Slattey, J. A. Stamatoyannopoulos, M. Y. Tolstorukov, K. P. White, S. Xi, P. J. Farnham, J. D. Lieb, B. J. Wold, M. Snyder, ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res*. **22**, 1813–1831 (2012).
145. L. P. O'Neill, B. M. Turner, Immunoprecipitation of native chromatin: NChIP. *Methods*. **31**, 76–82 (2003).
146. J. Brind'Amour, S. Liu, M. Hudson, C. Chen, M. M. Karimi, M. C. Lorincz, An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat Commun*. **6**, 6033 (2015).
147. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, 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).
148. J. D. Buenrostro, B. Wu, U. M. Litzenburger, D. Ruff, M. L. Gonzales, M. P. Snyder, H. Y. Chang, W. J. Greenleaf, Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*. **523**, 486–490 (2015).
149. D. M. Prescott, M. A. Bender, Synthesis of RNA and protein during mitosis in mammalian tissue culture cells. *Exp. Cell Res*. **26**, 260–268 (1962).
150. C. A. Spencer, M. J. Kruhlak, H. L. Jenkins, X. Sun, D. P. Bazett-Jones, Mitotic transcription repression in vivo in the absence of nucleosomal chromatin condensation. *J. Cell Biol*. **150**, 13–26 (2000).
151. N. Festuccia, A. Dubois, S. Vandormael-Pournin, E. Gallego Tejeda, A. Mouren, S. Bessonard, F. Mueller, C. Proux, M. Cohen-Tannoudji, P. Navarro, Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. *Nat. Cell Biol*. **18**, 1139–1148 (2016).

152. S. Kadauke, G. A. Blobel, Mitotic bookmarking by transcription factors. *Epigenetics Chromatin*. **6**, 6 (2013).
153. S. K. Zaidi, D. W. Young, M. A. Montecino, J. B. Lian, A. J. van Wijnen, J. L. Stein, G. S. Stein, Mitotic bookmarking of genes: a novel dimension to epigenetic control. *Nat. Rev. Genet.* **11**, 583–589 (2010).
154. S. K. Zaidi, R. A. Grandy, C. Lopez-Camacho, M. Montecino, A. J. van Wijnen, J. B. Lian, J. L. Stein, G. S. Stein, Bookmarking target genes in mitosis: a shared epigenetic trait of phenotypic transcription factors and oncogenes? *Cancer Res.* **74**, 420–425 (2014).
155. D. Chen, C. S. Hinkley, R. W. Henry, S. Huang, TBP dynamics in living human cells: constitutive association of TBP with mitotic chromosomes. *Mol. Biol. Cell.* **13**, 276–284 (2002).
156. A. Grob, C. Colleran, B. McStay, Construction of synthetic nucleoli in human cells reveals how a major functional nuclear domain is formed and propagated through cell division. *Genes Dev.* **28**, 220–230 (2014).
157. C. Lopez-Camacho, A. J. van Wijnen, J. B. Lian, J. L. Stein, G. S. Stein, CBF β and the leukemogenic fusion protein CBF β -SMMHC associate with mitotic chromosomes to epigenetically regulate ribosomal genes. *J. Cell. Biochem.* **115**, 2155–2164 (2014).
158. G. A. Blobel, S. Kadauke, E. Wang, A. W. Lau, J. Zuber, M. M. Chou, C. R. Vakoc, A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. *Mol. Cell.* **36**, 970–983 (2009).
159. S. M. Pockwinse, K. P. Kota, A. J. C. Quaresma, A. N. Imbalzano, J. B. Lian, A. J. van Wijnen, J. L. Stein, G. S. Stein, J. A. Nickerson, Live cell imaging of the cancer-related transcription factor RUNX2 during mitotic progression. *J. Cell. Physiol.* **226**, 1383–1389 (2011).
160. J. M. Caravaca, G. Donahue, J. S. Becker, X. He, C. Vinson, K. S. Zaret, Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. *Genes Dev.* **27**, 251–260 (2013).
161. A. Dey, A. Nishiyama, T. Karpova, J. McNally, K. Ozato, Brd4 marks select genes on mitotic chromatin and directs postmitotic transcription. *Mol. Biol. Cell.* **20**, 4899–4909 (2009).
162. K. S. Zaret, Genome reactivation after the silence in mitosis: recapitulating mechanisms of development? *Dev. Cell.* **29**, 132–134 (2014).
163. S.-J. Dunn, G. Martello, B. Yordanov, S. Emmott, A. G. Smith, Defining an essential transcription factor program for naïve pluripotency. *Science*. **344**, 1156–1160 (2014).
164. A. A. Avilion, Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & Development*. **17**, 126–140 (2003).
165. J. Nichols, B. Zevnik, K. Anastassiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Schöler, A. Smith, Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. *Cell*. **95**, 379–391 (1998).
166. M. F. Cole, R. A. Young, Mapping key features of transcriptional regulatory circuitry in embryonic stem cells. *Cold Spring Harb. Symp. Quant. Biol.* **73**, 183–193 (2008).
167. A. Reményi, K. Lins, L. J. Nissen, R. Reinbold, H. R. Schöler, M. Wilmanns, Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev.* **17**, 2048–2059 (2003).

168. M. Nakagawa, M. Koyanagi, K. Tanabe, K. Takahashi, T. Ichisaka, T. Aoi, K. Okita, Y. Mochizuki, N. Takizawa, S. Yamanaka, Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat. Biotechnol.* **26**, 101–106 (2008).
169. K. Takahashi, S. Yamanaka, Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell.* **126**, 663–676 (2006).
170. M. Thomson, S. J. Liu, L.-N. Zou, Z. Smith, A. Meissner, S. Ramanathan, Pluripotency Factors in Embryonic Stem Cells Regulate Differentiation into Germ Layers. *Cell.* **145**, 875–889 (2011).
171. S. Zhao, J. Nichols, A. G. Smith, M. Li, SoxB transcription factors specify neuroectodermal lineage choice in ES cells. *Molecular and Cellular Neuroscience.* **27**, 332–342 (2004).
172. J. Chen, Z. Zhang, L. Li, B.-C. Chen, A. Revyakin, B. Hajj, W. Legant, M. Dahan, T. Lionnet, E. Betzig, R. Tjian, Z. Liu, Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell.* **156**, 1274–1285 (2014).
173. J. C. M. Gebhardt, D. M. Suter, R. Roy, Z. W. Zhao, A. R. Chapman, S. Basu, T. Maniatis, X. S. Xie, Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat. Methods.* **10**, 421–426 (2013).
174. M. Tokunaga, N. Imamoto, K. Sakata-Sogawa, Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods.* **5**, 159–161 (2008).
175. G. L. Hager, J. G. McNally, T. Misteli, Transcription dynamics. *Mol. Cell.* **35**, 741–753 (2009).
176. A. Soufi, M. F. Garcia, A. Jaroszewicz, N. Osman, M. Pellegrini, K. S. Zaret, Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell.* **161**, 555–568 (2015).
177. S. Kadauke, M. I. Udugama, J. M. Pawlicki, J. C. Achtman, D. P. Jain, Y. Cheng, R. C. Hardison, G. A. Blobel, Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. *Cell.* **150**, 725–737 (2012).
178. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoutte, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
179. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature.* **489**, 57–74 (2012).
180. G. Lukinavičius, K. Umezawa, N. Olivier, A. Honigsmann, G. Yang, T. Plass, V. Mueller, L. Reymond, I. R. Corrêa, Z.-G. Luo, C. Schultz, E. A. Lemke, P. Heppenstall, C. Eggeling, S. Manley, K. Johnsson, A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. *Nat Chem.* **5**, 132–139 (2013).
181. D. Hockemeyer, F. Soldner, E. G. Cook, Q. Gao, M. Mitalipova, R. Jaenisch, A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. *Cell Stem Cell.* **3**, 346–353 (2008).
182. S. Masui, Y. Nakatake, Y. Toyooka, D. Shimosato, R. Yagi, K. Takahashi, H. Okochi, A. Okuda, R. Matoba, A. A. Sharov, M. S. H. Ko, H. Niwa, Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biology.* **9**, 625–635 (2007).

183. C. J. Lengner, F. D. Camargo, K. Hochedlinger, G. G. Welstead, S. Zaidi, S. Gokhale, H. R. Scholer, A. Tomilin, R. Jaenisch, Oct4 expression is not required for mouse somatic stem cell self-renewal. *Cell Stem Cell*. **1**, 403–415 (2007).
184. A. Kubo, K. Shinozaki, J. M. Shannon, V. Kouskoff, M. Kennedy, S. Woo, H. J. Fehling, G. Keller, Development of definitive endoderm from embryonic stem cells in culture. *Development*. **131**, 1651–1662 (2004).
185. Q.-L. Ying, M. Stavridis, D. Griffiths, M. Li, A. Smith, Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nature Biotechnology*. **21**, 183–186 (2003).
186. C. Galonska, Z. D. Smith, A. Meissner, In Vivo and in vitro dynamics of undifferentiated embryonic cell transcription factor 1. *Stem Cell Reports*. **2**, 245–252 (2014).
187. B. P. Mishra, K. I. Ansari, S. S. Mandal, Dynamic association of MLL1, H3K4 trimethylation with chromatin and Hox gene expression during the cell cycle. *FEBS J*. **276**, 1629–1640 (2009).
188. T.-F. Wu, Y.-L. Yao, I.-L. Lai, C.-C. Lai, P.-L. Lin, W.-M. Yang, Loading of PAX3 to Mitotic Chromosomes Is Mediated by Arginine Methylation and Associated with Waardenburg Syndrome. *J. Biol. Chem*. **290**, 20556–20564 (2015).
189. S. Kumar, N. K. Chaturvedi, S. Kumar, R. K. Tyagi, Agonist-mediated docking of androgen receptor onto the mitotic chromatin platform discriminates intrinsic mode of action of prostate cancer drugs. *Biochim. Biophys. Acta*. **1783**, 59–73 (2008).
190. C. Pallier, P. Scaffidi, S. Chopineau-Proust, A. Agresti, P. Nordmann, M. E. Bianchi, V. Marechal, Association of chromatin proteins high mobility group box (HMGB) 1 and HMGB2 with mitotic chromosomes. *Mol. Biol. Cell*. **14**, 3414–3426 (2003).
191. J. Lerner, A. Bagattin, F. Verdeguer, M. P. Makinistoglu, S. Garbay, T. Felix, L. Heidet, M. Pontoglio, Human mutations affect the epigenetic/bookmarking function of HNF1B. *Nucleic Acids Res*. **44**, 8097–8111 (2016).
192. A. Mann, G. Thakur, V. Shukla, M. Ganguli, Peptides in DNA delivery: current insights and future directions. *Drug Discov. Today*. **13**, 152–160 (2008).
193. F. Simeoni, M. C. Morris, F. Heitz, G. Divita, Insight into the mechanism of the peptide-based gene delivery system MPG: implications for delivery of siRNA into mammalian cells. *Nucleic Acids Res*. **31**, 2717–2724 (2003).
194. H. Xing, D. C. Wilkerson, C. N. Mayhew, E. J. Lubert, H. S. Skaggs, M. L. Goodson, Y. Hong, O.-K. Park-Sarge, K. D. Sarge, Mechanism of hsp70i gene bookmarking. *Science*. **307**, 421–423 (2005).
195. C. C.-S. Hsiung, C. R. Bartman, P. Huang, P. Ginart, A. J. Stonestrom, C. A. Keller, C. Face, K. S. Jahn, P. Evans, L. Sankaranarayanan, B. Giardine, R. C. Hardison, A. Raj, G. A. Blobel, A hyperactive transcriptional state marks genome reactivation at the mitosis-G1 transition. *Genes Dev*. **30**, 1423–1439 (2016).
196. S. Pauklin, L. Vallier, The cell-cycle state of stem cells determines cell fate propensity. *Cell*. **155**, 135–147 (2013).
197. S. Pauklin, P. Madrigal, A. Bertero, L. Vallier, Initiation of stem cell differentiation involves cell cycle-dependent regulation of developmental genes by Cyclin D. *Genes Dev*. **30**, 421–433 (2016).

198. D. Coronado, M. Godet, P.-Y. Bourillot, Y. Tapponnier, A. Bernat, M. Petit, M. Afanassieff, S. Markossian, A. Malashicheva, R. Iacone, K. Anastassiadis, P. Savatier, A short G1 phase is an intrinsic determinant of naïve embryonic stem cell pluripotency. *Stem Cell Res.* **10**, 118–131 (2013).
199. R. P. Halley-Stott, J. Jullien, V. Pasque, J. Gurdon, Mitosis gives a brief window of opportunity for a change in gene transcription. *PLoS Biol.* **12**, e1001914 (2014).
200. Y. Sela, N. Molotski, S. Golan, J. Itskovitz-Eldor, Y. Soen, Human embryonic stem cells exhibit increased propensity to differentiate during the G1 phase prior to phosphorylation of retinoblastoma protein. *Stem Cells.* **30**, 1097–1108 (2012).
201. R. J. Lake, P.-F. Tsai, I. Choi, K.-J. Won, H.-Y. Fan, RBPJ, the major transcriptional effector of Notch signaling, remains associated with chromatin throughout mitosis, suggesting a role in mitotic bookmarking. *PLoS Genet.* **10**, e1004204 (2014).
202. M. Nagaoka, U. Koshimizu, S. Yuasa, F. Hattori, H. Chen, T. Tanaka, M. Okabe, K. Fukuda, T. Akaike, E-cadherin-coated plates maintain pluripotent ES cells without colony formation. *PLoS ONE.* **1**, e15 (2006).
203. F. P. A. David, J. Delafontaine, S. Carat, F. J. Ross, G. Lefebvre, Y. Jarosz, L. Sinclair, D. Noordermeer, J. Rougemont, M. Leleu, HTSstation: a web application and open-access libraries for high-throughput sequencing data analysis. *PLoS ONE.* **9**, e85879 (2014).
204. J. Feng, T. Liu, B. Qin, Y. Zhang, X. S. Liu, Identifying ChIP-seq enrichment using MACS. *Nat Protoc.* **7**, 1728–1740 (2012).
205. P. Machanick, T. L. Bailey, MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics.* **27**, 1696–1697 (2011).
206. C. Gubelmann, P. C. Schwalie, S. K. Raghav, E. Röder, T. Delessa, E. Kiehlmann, S. M. Waszak, A. Corsinotti, G. Udin, W. Holcombe, G. Rudofsky, D. Trono, C. Wolfrum, B. Deplancke, Identification of the transcription factor ZEB1 as a central component of the adipogenic gene regulatory network. *Elife.* **3**, e03346 (2014).
207. D. M. Suter, L. Cartier, E. Bettiol, D. Tirefort, M. E. Jaconi, M. Dubois-Dauphin, K.-H. Krause, Rapid generation of stable transgenic embryonic stem cell lines using modular lentivectors. *Stem Cells.* **24**, 615–623 (2006).
208. J. C. M. Gebhardt, D. M. Suter, R. Roy, Z. W. Zhao, A. R. Chapman, S. Basu, T. Maniatis, X. S. Xie, Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat. Methods.* **10**, 421–426 (2013).
209. F. Mueller, T. J. Stasevich, D. Mazza, J. G. McNally, Quantifying transcription factor kinetics: at work or at play? *Crit. Rev. Biochem. Mol. Biol.* **48**, 492–514 (2013).
210. L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, F. Zhang, Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science.* **339**, 819–823 (2013).
211. D. E. Cohen, D. Melton, Turning straw into gold: directing cell fate for regenerative medicine. *Nature reviews. Genetics.* **12**, 243–52 (2011).
212. A. Filipczyk, C. Marr, S. Hastreiter, J. Feigelman, M. Schwarzfischer, P. S. Hoppe, D. Loeffler, K. D. Kokkalis, M. Ende, B. Schauburger, O. Hilsenbeck, S. Skylaki, J. Hasenauer, K. Anastassiadis, F. J. Theis, T. Schroeder, Network plasticity of pluripotency transcription factors in embryonic stem cells. *Nature Cell Biology.* **17**, 1235–1246 (2015).

213. J.-L. Chew, Y.-H. Loh, W. Zhang, X. Chen, W.-L. Tam, L.-S. Yeap, P. Li, Y.-S. Ang, B. Lim, P. Robson, H.-H. Ng, Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4/Sox2 Complex in Embryonic Stem Cells. *Molecular and Cellular Biology*. **25**, 6031–6046 (2005).
214. S. Okumura-Nakanishi, M. Saito, H. Niwa, F. Ishikawa, Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. *J. Biol. Chem.* **280**, 5307–5317 (2005).
215. N. Tapia, C. MacCarthy, D. Esch, A. Gabriele Marthaler, U. Tiemann, M. J. Araúzo-Bravo, R. Jauch, V. Cojocar, H. R. Schöler, Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency. *Scientific Reports*. **5** (2015), doi:10.1038/srep13533.
216. D. L. C. van den Berg, T. Snoek, N. P. Mullin, A. Yates, K. Bezstarosti, J. Demmers, I. Chambers, R. A. Poot, An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell*. **6**, 369–381 (2010).
217. H. Niwa, J. Miyazaki, A. G. Smith, Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* **24**, 372–376 (2000).
218. S. Masui, Y. Nakatake, Y. Toyooka, D. Shimosato, R. Yagi, K. Takahashi, H. Okochi, A. Okuda, R. Matoba, A. A. Sharov, M. S. Ko, H. Niwa, Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature cell biology*. **9**, 625–35 (2007).
219. M. D. White, J. F. Angiolini, Y. D. Alvarez, G. Kaur, Z. W. Zhao, E. Mocskos, L. Bruno, S. Bissiere, V. Levi, N. Plachta, Long-Lived Binding of Sox2 to DNA Predicts Cell Fate in the Four-Cell Mouse Embryo. *Cell*. **165**, 75–87 (2016).
220. M. Goolam, A. Scialdone, S. J. L. Graham, I. C. Macaulay, A. Jedrusik, A. Hupalowska, T. Voet, J. C. Marioni, M. Zernicka-Goetz, Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos. *Cell*. **165**, 61–74 (2016).
221. T. K. Mistri, W. Arindrarto, W. P. Ng, C. Wang, L. H. Lim, L. Sun, I. Chambers, T. Wohland, P. Robson, Dynamic changes in Sox2 spatio-temporal expression promote the second cell fate decision through Fgf4/Fgfr2 signaling in preimplantation mouse embryos. *Biochem. J.* **475**, 1075–1089 (2018).
222. S. C. Wolff, K. M. Kedziora, R. Dumitru, C. D. Dungee, T. M. Zikry, A. S. Beltran, R. A. Haggerty, J. Cheng, M. A. Redick, J. E. Purvis, Inheritance of OCT4 predetermines fate choice in human embryonic stem cells. *Mol. Syst. Biol.* **14**, e8140 (2018).
223. H. Yang, H. Wang, C. S. Shivalila, A. W. Cheng, L. Shi, R. Jaenisch, One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell*. **154**, 1370–1379 (2013).
224. C. Deluz, E. T. Friman, D. Strebing, A. Benke, M. Raccaud, A. Callegari, M. Leleu, S. Manley, D. M. Suter, A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. *Genes & Development*. **30**, 2538–2550 (2016).
225. X. Pan, X. Cang, S. Dan, J. Li, J. Cheng, B. Kang, X. Duan, B. Shen, Y.-J. Wang, Site-specific Disruption of the Oct4/Sox2 Protein Interaction Reveals Coordinated Mesendodermal Differentiation and the Epithelial-Mesenchymal Transition. *J. Biol. Chem.* **291**, 18353–18369 (2016).
226. L. Fang, L. Zhang, W. Wei, X. Jin, P. Wang, Y. Tong, J. Li, J. X. Du, J. Wong, A methylation-phosphorylation switch determines Sox2 stability and function in ESC maintenance or differentiation. *Mol. Cell*. **55**, 537–551 (2014).

227. L. Liu, W. Michowski, H. Inuzuka, K. Shimizu, N. T. Nihira, J. M. Chick, N. Li, Y. Geng, A. Y. Meng, A. Ordureau, A. Kołodziejczyk, K. L. Ligon, R. T. Bronson, K. Polyak, J. W. Harper, S. P. Gygi, W. Wei, P. Sicinski, G1 cyclins link proliferation, pluripotency and differentiation of embryonic stem cells. *Nat. Cell Biol.* **19**, 177–188 (2017).
228. Y.-H. Loh, Q. Wu, J.-L. Chew, V. B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K.-Y. Wong, K. W. Sung, C. W. H. Lee, X.-D. Zhao, K.-P. Chiu, L. Lipovich, V. A. Kuznetsov, P. Robson, L. W. Stanton, C.-L. Wei, Y. Ruan, B. Lim, H.-H. Ng, The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38**, 431–440 (2006).
229. X. Chen, H. Xu, P. Yuan, F. Fang, M. Huss, V. B. Vega, E. Wong, Y. L. Orlov, W. Zhang, J. Jiang, Y.-H. Loh, H. C. Yeo, Z. X. Yeo, V. Narang, K. R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W.-K. Sung, N. D. Clarke, C.-L. Wei, H.-H. Ng, Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell.* **133**, 1106–1117 (2008).
230. L. A. Boyer, T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, R. A. Young, Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell.* **122**, 947–56 (2005).
231. J. L. Kopp, B. D. Ormsbee, M. Desler, A. Rizzino, Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells.* **26**, 903–911 (2008).
232. B. D. Ormsbee Golden, E. L. Wuebben, A. Rizzino, Sox2 expression is regulated by a negative feedback loop in embryonic stem cells that involves AKT signaling and FoxO1. *PLoS ONE.* **8**, e76345 (2013).
233. A. B. Alber, E. R. Paquet, M. Biserni, F. Naef, D. M. Suter, Single Live Cell Monitoring of Protein Turnover Reveals Intercellular Variability and Cell-Cycle Dependence of Degradation Rates. *Mol. Cell.* **71**, 1079-1091.e9 (2018).
234. A. Mandic, D. Strebing, C. Regali, N. E. Phillips, D. M. Suter, A novel method for quantitative measurements of gene expression in single living cells. *Methods.* **120**, 65–75 (2017).
235. A. Sigal, R. Milo, A. Cohen, N. Geva-Zatorsky, Y. Klein, Y. Liron, N. Rosenfeld, T. Danon, N. Perzov, U. Alon, Variability and memory of protein levels in human cells. *Nature.* **444**, 643–6 (2006).
236. S. Pauklin, L. Vallier, The cell-cycle state of stem cells determines cell fate propensity. *Cell.* **155**, 135–47 (2013).
237. D. Coronado, M. Godet, P.-Y. Bourillot, Y. Tapponnier, A. Bernat, M. Petit, M. Afanassieff, S. Markossian, A. Malashicheva, R. Iacone, K. Anastassiadis, P. Savatier, A short G1 phase is an intrinsic determinant of naïve embryonic stem cell pluripotency. *Stem Cell Res.* **10**, 118–131 (2013).
238. D. A. Turner, P. Rué, J. P. Mackenzie, E. Davies, A. Martinez Arias, Brachyury cooperates with Wnt/ β -catenin signalling to elicit primitive-streak-like behaviour in differentiating mouse embryonic stem cells. *BMC Biol.* **12**, 63 (2014).
239. H. W. King, R. J. Klose, The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *Elife.* **6** (2017), doi:10.7554/eLife.22631.

240. M. Raccaud, E. T. Friman, A. B. Alber, H. Agarwal, C. Deluz, T. Kuhn, J. C. M. Gebhardt, D. M. Suter, Mitotic chromosome binding predicts transcription factor properties in interphase. *Nat Commun.* **10**, 487 (2019).
241. S. Mei, Q. Qin, Q. Wu, H. Sun, R. Zheng, C. Zang, M. Zhu, J. Wu, X. Shi, L. Taing, T. Liu, M. Brown, C. A. Meyer, X. S. Liu, Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res.* **45**, D658–D662 (2017).
242. V. Heurtier, N. Owens, I. Gonzalez, F. Mueller, C. Proux, D. Mornico, P. Clerc, A. Dubois, P. Navarro, The molecular logic of Nanog-induced self-renewal in mouse embryonic stem cells. *Nat Commun.* **10**, 1109 (2019).
243. T. Kalmar, C. Lim, P. Hayward, S. Munoz-Descalzo, J. Nichols, J. Garcia-Ojalvo, A. Martinez Arias, Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS biology.* **7**, e1000149 (2009).
244. V. Karwacki-Neisius, J. Göke, R. Osorno, F. Halbritter, J. H. Ng, A. Y. Weiße, F. C. K. Wong, A. Gagliardi, N. P. Mullin, N. Festuccia, D. Colby, S. R. Tomlinson, H.-H. Ng, I. Chambers, 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).
245. A. Radziskeuskaya, G. L. B. Chia, R. L. dos Santos, T. W. Theunissen, L. F. C. Castro, J. Nichols, J. C. R. Silva, A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. *Nat. Cell Biol.* **15**, 579–590 (2013).
246. Z. Wang, E. Oron, B. Nelson, S. Razis, N. Ivanova, Distinct Lineage Specification Roles for NANOG, OCT4, and SOX2 in Human Embryonic Stem Cells. *Cell Stem Cell.* **10**, 440–454 (2012).
247. A. Soufi, M. F. Garcia, A. Jaroszewicz, N. Osman, M. Pellegrini, K. S. Zaret, Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell.* **161**, 555–568 (2015).
248. Y. Liu, B. Pelham-Webb, D. C. Di Giammartino, J. Li, D. Kim, K. Kita, N. Saiz, V. Garg, A. Doane, P. Giannakakou, A.-K. Hadjantonakis, O. Elemento, E. Apostolou, Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors in Pluripotent Stem Cells. *Cell Rep.* **19**, 1283–1293 (2017).
249. M. Nagaoka, U. Koshimizu, S. Yuasa, F. Hattori, H. Chen, T. Tanaka, M. Okabe, K. Fukuda, T. Akaike, E-cadherin-coated plates maintain pluripotent ES cells without colony formation. *PLoS ONE.* **1**, e15 (2006).
250. O. Hobert, PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. *BioTechniques.* **32**, 728–730 (2002).
251. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, 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).
252. W. Chen, V. Gardeux, A. Meireles-Filho, B. Deplancke, Profiling of Single-Cell Transcriptomes. *Curr Protoc Mouse Biol.* **7**, 145–175 (2017).
253. L. Kametsky, T. R. Jones, A. Fraser, M.-A. Bray, D. J. Logan, K. L. Madden, V. Ljosa, C. Rueden, K. W. Eliceiri, A. E. Carpenter, Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics.* **27**, 1179–1180 (2011).

254. K. J. Chalut, M. Höpfler, F. Lautenschläger, L. Boyde, C. J. Chan, A. Ekpenyong, A. Martinez-Arias, J. Guck, Chromatin Decondensation and Nuclear Softening Accompany Nanog Downregulation in Embryonic Stem Cells. *Biophysical Journal*. **103**, 2060–2070 (2012).
255. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).
256. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoutte, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. **9**, R137 (2008).
257. S. Nepf, M. S. Kuehn, A. P. Reynolds, E. Haugen, R. E. Thurman, A. K. Johnson, E. Rynes, M. T. Maurano, J. Vierstra, S. Thomas, R. Sandstrom, R. Humbert, J. A. Stamatoyannopoulos, BEDOPS: high-performance genomic feature operations. *Bioinformatics*. **28**, 1919–1920 (2012).
258. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*. **38**, 576–589 (2010).
259. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**, 139–140 (2010).
260. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. **43**, e47 (2015).
261. C. Buecker, R. Srinivasan, Z. Wu, E. Calo, D. Acampora, T. Faial, A. Simeone, M. Tan, T. Swigut, J. Wysocka, Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell*. **14**, 838–853 (2014).
262. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature*. **489**, 57–74 (2012).
263. W. A. Whyte, D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, R. A. Young, Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. **153**, 307–319 (2013).
264. A. S. Hinrichs, D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T. S. Furey, R. A. Harte, F. Hsu, J. Hillman-Jackson, R. M. Kuhn, J. S. Pedersen, A. Pohl, B. J. Raney, K. R. Rosenbloom, A. Siepel, K. E. Smith, C. W. Sugnet, A. Sultan-Qurraie, D. J. Thomas, H. Trumbower, R. J. Weber, M. Weirauch, A. S. Zweig, D. Haussler, W. J. Kent, The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res*. **34**, D590-598 (2006).
265. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).
266. F. Ramírez, F. Dündar, S. Diehl, B. A. Grüning, T. Manke, deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. **42**, W187-191 (2014).
267. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, D. Haussler, The human genome browser at UCSC. *Genome Res*. **12**, 996–1006 (2002).
268. A. B. Alber, D. M. Suter, Single-Cell Quantification of Protein Degradation Rates by Time-Lapse Fluorescence Microscopy in Adherent Cell Culture. *J Vis Exp* (2018), doi:10.3791/56604.

269. C. Chronis, P. Fiziev, B. Papp, S. Butz, G. Bonora, S. Sabri, J. Ernst, K. Plath, Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell*. **168**, 442-459.e20 (2017).
270. J. Jacobs, M. Atkins, K. Davie, H. Imrichova, L. Romanelli, V. Christiaens, G. Hulselmans, D. Potier, J. Wouters, I. I. Taskiran, G. Paciello, C. B. González-Blas, D. Koldere, S. Aibar, G. Halder, S. Aerts, The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. *Nat. Genet.* **50**, 1011–1020 (2018).
271. A. Mayran, K. Khetchoumian, F. Hariri, T. Pastinen, Y. Gauthier, A. Balsalobre, J. Drouin, Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nat. Genet.* **50**, 259–269 (2018).
272. W. A. Pastor, W. Liu, D. Chen, J. Ho, R. Kim, T. J. Hunt, A. Lukianchikov, X. Liu, J. M. Polo, S. E. Jacobsen, A. T. Clark, TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nat. Cell Biol.* **20**, 553–564 (2018).
273. S. Masui, Y. Nakatake, Y. Toyooka, D. Shimosato, R. Yagi, K. Takahashi, H. Okochi, A. Okuda, R. Matoba, A. A. Sharov, M. S. Ko, H. Niwa, Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature cell biology*. **9**, 625–35 (2007).
274. H. Niwa, J. Miyazaki, A. G. Smith, Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* **24**, 372–376 (2000).
275. L. A. Boyer, T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, R. A. Young, Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell*. **122**, 947–956 (2005).
276. M. Nishimoto, A. Fukushima, A. Okuda, M. Muramatsu, The Gene for the Embryonic Stem Cell Coactivator UTF1 Carries a Regulatory Element Which Selectively Interacts with a Complex Composed of Oct-3/4 and Sox-2. *Molecular and Cellular Biology*. **19**, 5453–5465 (1999).
277. H. Yuan, N. Corbi, C. Basilico, L. Dailey, Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev.* **9**, 2635–2645 (1995).
278. D. Streibinger, E. T. Friman, C. Deluz, S. Govindan, A. B. Alber, D. M. Suter, Endogenous fluctuations of OCT4 and SOX2 bias pluripotent cell fate decisions (2019), doi:10.1101/299073.
279. C. Deluz, E. T. Friman, D. Streibinger, A. Benke, M. Raccaud, A. Callegari, M. Leleu, S. Manley, D. M. Suter, A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. *Genes Dev.* **30**, 2538–2550 (2016).
280. S. J. Dunn, G. Martello, B. Yordanov, S. Emmott, A. G. Smith, Defining an essential transcription factor program for naive pluripotency. *Science*. **344**, 1156–60 (2014).
281. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
282. M. P. Creighton, A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, R. Jaenisch, Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*. **107**, 21931–21936 (2010).

283. L. Ho, E. L. Miller, J. L. Ronan, W. Q. Ho, R. Jothi, G. R. Crabtree, esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nature Cell Biology*. **13**, 903–913 (2011).
284. T. Ishiuchi, H. Ohishi, T. Sato, S. Kamimura, M. Yorino, S. Abe, A. Suzuki, T. Wakayama, M. Suyama, H. Sasaki, Zfp281 Shapes the Transcriptome of Trophoblast Stem Cells and Is Essential for Placental Development. *Cell Rep*. **27**, 1742-1754.e6 (2019).
285. I. Aksoy, V. Giudice, E. Delahaye, F. Wianny, M. Aubry, M. Mure, J. Chen, R. Jauch, G. K. Bogu, T. Nolden, H. Himmelbauer, M. Xavier Doss, A. Sachinidis, H. Schulz, O. Hummel, P. Martinelli, N. Hübner, L. W. Stanton, F. X. Real, P.-Y. Bourillot, P. Savatier, Klf4 and Klf5 differentially inhibit mesoderm and endoderm differentiation in embryonic stem cells. *Nat Commun*. **5**, 3719 (2014).
286. K.-Y. Kim, Y. Tanaka, J. Su, B. Cakir, Y. Xiang, B. Patterson, J. Ding, Y.-W. Jung, J.-H. Kim, E. Hysolli, H. Lee, R. Dajani, J. Kim, M. Zhong, J.-H. Lee, D. Skalnik, J. M. Lim, G. J. Sullivan, J. Wang, I.-H. Park, Uhrf1 regulates active transcriptional marks at bivalent domains in pluripotent stem cells through Setd1a. *Nat Commun*. **9**, 2583 (2018).
287. J. Xiong, Z. Zhang, J. Chen, H. Huang, Y. Xu, X. Ding, Y. Zheng, R. Nishinakamura, G.-L. Xu, H. Wang, S. Chen, S. Gao, B. Zhu, Cooperative Action between SALL4A and TET Proteins in Stepwise Oxidation of 5-Methylcytosine. *Mol. Cell*. **64**, 913–925 (2016).
288. Z. Liu, W. L. Kraus, Catalytic-Independent Functions of PARP-1 Determine Sox2 Pioneer Activity at Intractable Genomic Loci. *Molecular Cell*. **65**, 589-603.e9 (2017).
289. F. Gao, S. W. Kwon, Y. Zhao, Y. Jin, PARP1 Poly(ADP-ribosyl)ates Sox2 to Control Sox2 Protein Levels and FGF4 Expression during Embryonic Stem Cell Differentiation. *J. Biol. Chem*. **284**, 22263–22273 (2009).
290. Y.-G. Yang, U. Cortes, S. Patnaik, M. Jasin, Z.-Q. Wang, Ablation of PARP-1 does not interfere with the repair of DNA double-strand breaks, but compromises the reactivation of stalled replication forks. *Oncogene*. **23**, 3872 (2004).
291. T. K. Mistri, W. Arindrarto, W. P. Ng, C. Wang, L. H. Lim, L. Sun, I. Chambers, T. Wohland, P. Robson, Dynamic changes in Sox2 spatio-temporal expression promote the second cell fate decision through Fgf4/Fgfr2 signaling in preimplantation mouse embryos. *Biochem. J*. **475**, 1075–1089 (2018).
292. S. Li, E. B. Zheng, L. Zhao, S. Liu, Nonreciprocal and Conditional Cooperativity Directs the Pioneer Activity of Pluripotency Transcription Factors. *bioRxiv*, 633826 (2019).
293. J. Chen, Z. Zhang, L. Li, B.-C. Chen, A. Revyakin, B. Hajj, W. Legant, M. Dahan, T. Lionnet, E. Betzig, R. Tjian, Z. Liu, Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells. *Cell*. **156**, 1274–1285 (2014).
294. G. Lukinavičius, K. Umezawa, N. Olivier, A. Honigsmann, G. Yang, T. Plass, V. Mueller, L. Reymond, I. R. Corrêa, Z.-G. Luo, C. Schultz, E. A. Lemke, P. Heppenstall, C. Eggeling, S. Manley, K. Johnsson, A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. *Nat Chem*. **5**, 132–139 (2013).
295. J. Lee, Y. Go, I. Kang, Y.-M. Han, J. Kim, Oct-4 controls cell-cycle progression of embryonic stem cells. *Biochem. J*. **426**, 171–181 (2010).
296. K. Cao, C. K. Collings, M. A. Morgan, S. A. Marshall, E. J. Rendleman, P. A. Ozark, E. R. Smith, A. Shilatifard, An Mll4/COMPASS-Lsd1 epigenetic axis governs enhancer function and pluripotency transition in embryonic stem cells. *Sci Adv*. **4**, eaap8747 (2018).

297. V. Kumar, N. A. Rayan, M. Muratani, S. Lim, B. Elanggovan, L. Xin, T. Lu, H. Makhija, J. Poschmann, T. Lufkin, H. H. Ng, S. Prabhakar, Comprehensive benchmarking reveals H2BK20 acetylation as a distinctive signature of cell-state-specific enhancers and promoters. *Genome Res.* **26**, 612–623 (2016).
298. R. Rickels, H.-M. Herz, C. C. Sze, K. Cao, M. A. Morgan, C. K. Collings, M. Gause, Y.-H. Takahashi, L. Wang, E. J. Rendleman, S. A. Marshall, A. Krueger, E. T. Bartom, A. Piunti, E. R. Smith, N. A. Abshiru, N. L. Kelleher, D. Dorsett, A. Shilatifard, Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nat. Genet.* **49**, 1647–1653 (2017).
299. B. R. Sabari, A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T. I. Lee, I. I. Cisse, R. G. Roeder, P. A. Sharp, A. K. Chakraborty, R. A. Young, Coactivator condensation at super-enhancers links phase separation and gene control. *Science*. **361** (2018), doi:10.1126/science.aar3958.
300. N. Dharmasiri, S. Dharmasiri, M. Estelle, The F-box protein TIR1 is an auxin receptor. *Nature*. **435**, 441 (2005).
301. S. Kepinski, O. Leyser, The Arabidopsis F-box protein TIR1 is an auxin receptor. *Nature*. **435**, 446–451 (2005).
302. M. Morawska, H. D. Ulrich, An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast*. **30**, 341–351 (2013).
303. K. Nishimura, T. Fukagawa, H. Takisawa, T. Kakimoto, M. Kanemaki, An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods*. **6**, 917–922 (2009).
304. A. B. Alber, E. R. Paquet, M. Biserni, F. Naef, D. M. Suter, Single Live Cell Monitoring of Protein Turnover Reveals Intercellular Variability and Cell-Cycle Dependence of Degradation Rates. *Mol. Cell*. **71**, 1079-1091.e9 (2018).
305. S. S. Teves, L. An, A. S. Hansen, L. Xie, X. Darzacq, R. Tjian, A dynamic mode of mitotic bookmarking by transcription factors. *Elife*. **5** (2016), doi:10.7554/eLife.22280.
306. I. Goldstein, V. Paakinaho, S. Baek, M.-H. Sung, G. L. Hager, Synergistic gene expression during the acute phase response is characterized by transcription factor assisted loading. *Nat Commun*. **8**, 1849 (2017).
307. S. L. McDaniel, T. J. Gibson, K. N. Schulz, M. Fernandez Garcia, M. Nevil, S. U. Jain, P. W. Lewis, K. S. Zaret, M. M. Harrison, Continued Activity of the Pioneer Factor Zelda Is Required to Drive Zygotic Genome Activation. *Mol. Cell*. **74**, 185-195.e4 (2019).
308. A. Mayran, K. Sochodolsky, K. Khetchoumian, J. Harris, Y. Gauthier, A. Bemmo, A. Balsalobre, J. Drouin, Pioneer and nonpioneer cooperation drives lineage specific chromatin opening. *bioRxiv*, 472647 (2019).
309. T. Dull, R. Zufferey, M. Kelly, R. J. Mandel, M. Nguyen, D. Trono, L. Naldini, A Third-Generation Lentivirus Vector with a Conditional Packaging System. *Journal of Virology*. **72**, 8463–8471 (1998).
310. D. M. Suter, L. Cartier, E. Bettiol, D. Tirefort, M. E. Jaconi, M. Dubois-Dauphin, K. H. Krause, Rapid generation of stable transgenic embryonic stem cell lines using modular lentivectors. *Stem Cells*. **24**, 615–23 (2006).

311. B. Nabet, J. M. Roberts, D. L. Buckley, J. Paulk, S. Dastjerdi, A. Yang, A. L. Leggett, M. A. Erb, M. A. Lawlor, A. Souza, T. G. Scott, S. Vittori, J. A. Perry, J. Qi, G. E. Winter, K.-K. Wong, N. S. Gray, J. E. Bradner, The dTAG system for immediate and target-specific protein degradation. *Nat. Chem. Biol.* **14**, 431–441 (2018).
312. E. P. Nora, A. Goloborodko, A.-L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, B. G. Bruneau, Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*. **169**, 930-944.e22 (2017).
313. W. Chen, V. Gardeux, A. Meireles-Filho, B. Deplancke, Profiling of Single-Cell Transcriptomes. *Curr Protoc Mouse Biol.* **7**, 145–175 (2017).
314. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nat. Methods*. **9**, 676–682 (2012).
315. A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat, P. Golland, D. M. Sabatini, CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
316. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).
317. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
318. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26**, 841–842 (2010).
319. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature*. **489**, 57–74 (2012).
320. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*. **38**, 576–589 (2010).
321. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**, 139–140 (2010).
322. S. Durinck, Y. Moreau, A. Kasprzyk, S. Davis, B. De Moor, A. Brazma, W. Huber, BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*. **21**, 3439–3440 (2005).
323. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).
324. F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
325. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, D. Haussler, The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

326. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, New York, 2009; <https://www.springer.com/gp/book/9780387981413>), *Use R!*
327. M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, V. J. Carey, Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
328. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* **32**, 2847–2849 (2016).
329. R. Edgar, M. Domrachev, A. E. Lash, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
330. A. S. Hinrichs, D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T. S. Furey, R. A. Harte, F. Hsu, J. Hillman-Jackson, R. M. Kuhn, J. S. Pedersen, A. Pohl, B. J. Raney, K. R. Rosenbloom, A. Siepel, K. E. Smith, C. W. Sugnet, A. Sultan-Qurraie, D. J. Thomas, H. Trumbower, R. J. Weber, M. Weirauch, A. S. Zweig, D. Haussler, W. J. Kent, The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590-598 (2006).
331. G. Pintacuda, G. Wei, C. Roustan, B. A. Kirmizitas, N. Solcan, A. Cerase, A. Castello, S. Mohammed, B. Moindrot, T. B. Nesterova, N. Brockdorff, hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Mol. Cell.* **68**, 955-969.e10 (2017).
332. C. Pallier, P. Scaffidi, S. Chopineau-Proust, A. Agresti, P. Nordmann, M. E. Bianchi, V. Marechal, Association of chromatin proteins high mobility group box (HMGB) 1 and HMGB2 with mitotic chromosomes. *Mol. Biol. Cell.* **14**, 3414–3426 (2003).
333. S. S. Teves, L. An, A. S. Hansen, L. Xie, X. Darzacq, R. Tjian, A dynamic mode of mitotic bookmarking by transcription factors. *Elife.* **5** (2016), doi:10.7554/eLife.22280.
334. N. Festuccia, A. Dubois, S. Vandormael-Pournin, E. Gallego Tejeda, A. Mouren, S. Bessonard, F. Mueller, C. Proux, M. Cohen-Tannoudji, P. Navarro, Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. *Nat. Cell Biol.* **18**, 1139–1148 (2016).
335. C. Deluz, D. Strebing, E. T. Friman, D. M. Suter, The elusive role of mitotic bookmarking in transcriptional regulation: Insights from Sox2. *Cell Cycle.* **16**, 601–606 (2017).
336. A. Boija, I. A. Klein, B. R. Sabari, A. Dall'Agnese, E. L. Coffey, A. V. Zamudio, C. H. Li, K. Shrinivas, J. C. Manteiga, N. M. Hannett, B. J. Abraham, L. K. Afeyan, Y. E. Guo, J. K. Rimel, C. B. Fant, J. Schuijers, T. I. Lee, D. J. Taatjes, R. A. Young, Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell.* **175**, 1842-1855.e16 (2018).
337. S. Chong, C. Dugast-Darzacq, Z. Liu, P. Dong, G. M. Dailey, C. Cattoglio, A. Heckert, S. Banala, L. Lavis, X. Darzacq, R. Tjian, Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science.* **361** (2018), doi:10.1126/science.aar2555.
338. I. Kwon, M. Kato, S. Xiang, L. Wu, P. Theodoropoulos, H. Mirzaei, T. Han, S. Xie, J. L. Corden, S. L. McKnight, Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell.* **155**, 1049–1060 (2013).

Appendix

Importance	Factor	GSMID	Unaffected	Upregulated_OCT4	Upregulated_SOX2	Upregulated_SOX2OCT4	Downregulated_OCT4	Downregulated_SOX2	Downregulated_SOX2OCT4	Sample_Name
0.35736007	POLR2A	GSM284846	0.289349634	0.215804371	0.184131737	0.236263736	0.574162679	0.632825719	0.677713339	Pol2 Nocodazole mES MM8
0.34908722	POLR2A	GSM515678	0.247273952	0.164133739	0.157185629	0.184615385	0.535885167	0.642978003	0.654515327	PolL1_Nterm_Doxycycline24hr_ChIP_Seq
0.32031528	SOX2	GSM2385224	0.121798026	0.279635258	0.414670659	0.30898011	0.081339713	0.091370558	0.097763049	Asynchronous ChIP 1
0.31980247	H3	GSM1000121	0.332318889	0.455927052	0.52994012	0.508791209	0.28708134	0.294416244	0.303231152	LICR_ChIPSeq_ES-E14_H3K4me1_E0
0.31408148	NANOG	GSM1323080	0.217695356	0.200607903	0.158682635	0.194505495	0.33492823	0.404399323	0.385252693	Nanog-sg2-IP
0.2990908	NANOG	GSM1059011	0.282278743	0.468085106	0.579341317	0.514285714	0.368421053	0.390862944	0.416735708	AntiGFP_KD_Nanog
0.28852562	PHC1	GSM1323078	0.36469476	0.398176292	0.350299401	0.391208791	0.492822967	0.631133672	0.634631317	Phc1-sg1-IP
0.28628065	CDK9	GSM1082347	0.303413712	0.252279635	0.233532934	0.285714286	0.502392344	0.615905245	0.632974316	ChIP-Seq for Cdk9 in mES cells
0.28535827	H3K4me1	GSM2588369	0.357287161	0.504559271	0.607784431	0.557142857	0.38277512	0.357021997	0.356255178	MLL2KO_ChIP_H3K4me1
0.28102234	KMT2B	GSM2588428	0.485625113	0.37689696	0.324850299	0.395604396	0.703349282	0.813874788	0.863297432	WT_RA_ChIP_MLL4
0.27922765	NANOG	GSM2417187	0.282460048	0.501519757	0.603293413	0.543956044	0.263157895	0.302876481	0.304888152	ESC_Nanog_ChIP-seq
0.27737146	CTCF	GSM747537	0.348765832	0.164133739	0.193113772	0.220879121	0.358851675	0.448392555	0.454018227	TKO_CTCTF_ChIPseq_rep1
0.27599906	ZIC2	GSM1499117	0.154355202	0.395136778	0.387724551	0.358241758	0.210526316	0.241962775	0.248550124	Zic2_rep2
0.27304203	NANOG	GSM1327154	0.372957083	0.346504559	0.360778443	0.408791209	0.607655502	0.676818951	0.671913836	Nanog-sg3-20ug-IP
0.27185842	H3K4me1	GSM2586537	0.366507809	0.47112462	0.517964072	0.496703297	0.363636364	0.377326565	0.387738194	H3K4me1_ChIP-seq_WT_rep1
0.27164065	H3K27ac	GSM2588293	0.293506695	0.234042553	0.23502994	0.247252747	0.514162679	0.583756345	0.633802817	DKO_RA_ChIP_H3K27Ac_Rep1
0.26991761	ELL3	GSM935891	0.154666011	0.364741641	0.357784431	0.316483516	0.177033493	0.216582064	0.235294118	ChIP-seq of Ell3 in mES
0.26975808	POLR2A	GSM515677	0.228703152	0.175149701	0.175149701	0.198901099	0.488038278	0.566835871	0.602319801	PolL1_Nterm_Doxycycline12hr_ChIP_Seq
0.26936354	NCAPD3	GSM824840	0.217915512	0.164133739	0.121257485	0.163736264	0.459330144	0.534686971	0.566694283	CapD3 WT mES MM8
0.26929262	POLR2A	GSM824845	0.3977990106	0.352583587	0.30988024	0.369736264	0.693736264	0.746192893	0.771338886	Pol2 Asynchronous mES MM8
0.26869978	RAD21	GSM2099829	0.385376467	0.227963526	0.254491018	0.291208791	0.545454545	0.595600677	0.603148302	ChIP_D3_0_E14_Rad21
0.26809429	STAG1	GSM937541	0.329184905	0.179331307	0.19011976	0.218681319	0.33492823	0.428087986	0.437482219	ChIP-Seq for SA1 in mES cells (rep 2)
0.26630816	NANOG	GSM1327155	0.368307907	0.279635258	0.270958084	0.331868132	0.578947368	0.680203046	0.674399337	Nanog-sg3-0.02ug-IP
0.26443885	H3K4me1	GSM970226	0.34038956	0.468085106	0.571856287	0.508791209	0.377990431	0.387478849	0.415900728	ES_Eed_vL_H3K4me1
0.26316021	NELFA	GSM515664	0.193646559	0.139817629	0.148203593	0.169230769	0.488038278	0.502538071	0.561723281	NelfA_ChIP_Seq
0.26149249	POLR2A	GSM2629552	0.322735632	0.258358663	0.232035928	0.294505495	0.598086124	0.62605753	0.671913836	Set1AdSet_ChIP_Pol2_2
0.25966351	H3K79me2	GSM1534641	0.255290217	0.200607903	0.191616766	0.22967033	0.56937799	0.573604061	0.603976802	H3K79me2_Adiarmacin
0.2589673	NRF1	GSM1657384	0.442228496	0.452887538	0.345806383	0.446153846	0.688995215	0.790186125	0.787075394	NRF1_ChIP_Over_1
0.25847017	H3K27ac	GSM2588356	0.321013235	0.294832827	0.267964072	0.271428571	0.607655502	0.617597293	0.654515327	DKO_RA_ChIP_H3K27Ac
0.25664537	CTCF	GSM2870294	0.3494263	0.194528875	0.197607491	0.210989011	0.411483254	0.453468697	0.457332229	ChIP_cCas9_CTCTF
0.25516847	H3K4me1	GSM2588304	0.279805227	0.422492401	0.468562874	0.469230769	0.23923445	0.253807107	0.255178128	TKO_SET1Ash_ChIP_H3K4me1
0.25435101	POLR2A	GSM1566100	0.223212204	0.188449848	0.173652695	0.194505495	0.535885167	0.536379019	0.609776305	v6.5-dC5_PolL1_ChIPSeq
0.25400459	POLR2A	GSM2629675	0.282019736	0.209726444	0.196107784	0.252747253	0.559808612	0.612521151	0.650372825	WT_ChIP_Pol2_1
0.25196346	NRF1	GSM1657385	0.376893988	0.404255319	0.278443114	0.393406593	0.626794258	0.72250423	0.724937862	NRF1_ChIP_Over_2
0.25098056	CTCF	GSM2870298	0.398443368	0.227963526	0.25748503	0.296703297	0.449760766	0.565143824	0.555095278	ChIP_SET_CTCTF
0.25088737	POLR2A	GSM515675	0.167163097	0.112462006	0.125748503	0.152747253	0.488038278	0.497461929	0.517812759	PolL1_Nterm_shcMyc_ChIP_Seq
0.25080219	NRF1	GSM1657379	0.144733093	0.151976584	0.112275449	0.152747253	0.301435407	0.395939086	0.386909693	NRF1_ChIP_to2_1
0.25059464	RAD21	GSM2099833	0.413206765	0.310030395	0.344311377	0.375824176	0.602870813	0.685279188	0.70091135	ChIP_Do_DKO_Rad21
0.24984079	CTCF	GSM1540994	0.217151441	0.072948328	0.05988024	0.096703297	0.200956938	0.27072758	0.257663629	do ChIP rep2
0.24842977	RAD21	GSM2099821	0.444792665	0.340425532	0.341317365	0.389010989	0.636363636	0.673434856	0.690140845	ChIP_D2_0_E14_Rad21
0.24672237	NANOG	GSM1327156	0.24719625	0.197568389	0.160179641	0.225274725	0.454545455	0.485617597	0.485501243	Nanog-sg3-0.2ug-IP
0.24643765	RAD21	GSM2870305	0.300473983	0.164133739	0.214071856	0.321868132	0.440191388	0.458544839	0.487158244	ChIP_cCD_Rad21_rep1
0.24623627	NANOG	GSM1842750	0.255652827	0.440729483	0.528443114	0.471428571	0.339712919	0.338409475	0.345484763	mES_Nanog_3p_2i_1
0.24594042	HDAC1	GSM1335482	0.333872931	0.468085106	0.54491018	0.512087912	0.483253589	0.502538071	0.589892295	HDAc1_ChIPSeq
0.24579329	H3K4me1	GSM1163086	0.438213888	0.589665653	0.628742515	0.612087912	0.387559809	0.439932318	0.490272905	H3K4me1_ESC_Rep_3
0.24574224	H3K4me1	GSM2588375	0.313281787	0.510638298	0.53742515	0.493406593	0.244019139	0.23857868	0.252692626	MLL3KO_ChIP_H3K4me1
0.24527422	RAD21	GSM2870307	0.275648165	0.161094225	0.209580838	0.230574163	0.370558376	0.436548223	0.381938969	ChIP_WT_Rad21_rep2
0.24282737	SMARCA4	GSM2601678	0.292380015	0.32218845	0.29491018	0.337362637	0.574162679	0.573604061	0.61971831	RNAIP2_Smarca4_G784E
0.24254502	H3K4me1	GSM2629645	0.231539278	0.300911854	0.395209581	0.357142857	0.248803828	0.263959391	0.267605634	Set1AdSet_ChIP_H3K4me1_1
0.24247569	CTCF	GSM2259905	0.357144707	0.197568389	0.203592814	0.22967033	0.397129187	0.436548223	0.44076222	ES_2i_CTCTF_ChIPSeq_library
0.2417662	Med1	GSM2520020	0.387383771	0.449848024	0.48502984	0.462637363	0.578947368	0.695431472	0.723280682	IP_Med1_rep2
0.24135392	CTCF	GSM2609187	0.318656272	0.161094225	0.166167665	0.184615385	0.315789474	0.377326565	0.377796189	CTCF_ChIP-seq_CTCTF-AID_shutoff2days_rep1
0.24075059	SOX2	GSM1308179	0.202844909	0.370820669	0.459580838	0.389010989	0.224880383	0.240270728	0.261806131	Sox2
0.24068776	H3K4me1	GSM1359830	0.223004999	0.334346505	0.320359281	0.358241758	0.291866029	0.287648054	0.256835128	Sox2_MES_H3K4me1
0.24063353	H3K4me1	GSM2588371	0.37471833	0.519756839	0.630239521	0.575824176	0.406698565	0.367174281	0.387738194	MLL2KO_RA_ChIP_H3K4me1
0.24014888	H3K4me1	GSM1856424	0.227304514	0.352583587	0.407185629	0.391208791	0.267942584	0.291032149	0.275062138	ChIPseq_H3K4me1_2i
0.23988099	H3K4me1	GSM2588456	0.19775182	0.303951368	0.353293413	0.334065934	0.148325359	0.177664975	0.175642088	WT_nonTsh_RA_ChIP_H3K4me1
0.23954742	ZIC2	GSM1499116	0.125942138	0.32218845	0.324850299	0.308791209	0.153110048	0.194585448	0.186412259	Zic2_rep1
0.23954228	H3K4me1	GSM2630429	0.397005879	0.571428571	0.628742515	0.578021978	0.344497608	0.436548223	0.382767191	MLL4KO_FBS_Lsd1sh_H3K4me1
0.23920437	RAD21	GSM2099825	0.41448885	0.270516717	0.30988024	0.626794258	0.58520643	0.636548223	0.65948633	ChIP_D2_5_E14_Rad21
0.23864571	NANOG	GSM1327155	0.415926338	0.343465046	0.360778443	0.410989011	0.61722488	0.707275804	0.738193869	Nanog-sg3-2ug-IP
0.23833101	RAD21	GSM2870297	0.286733663	0.139817629	0.152694611	0.193406593	0.315789474	0.424703892	0.40347902	ChIP_cCas9_Rad21
0.23792048	GATA1	GSM1582440	0.193361651	0.079027356	0.091317365	0.1	0.425837321	0.443316413	0.469759735	HA BRD2 +GATA1_r1
0.23647117	NANOG	GSM1842753	0.287147582	0.477203647	0.576347305	0.515384615	0.406698565	0.395939086	0.434962717	mES_Nanog_Serum_r2
0.23509651	H3K4me1	GSM2636047	0.236848921	0.358662614	0.419161677	0.354945055	0.205741627	0.226734349	0.225352113	WT_H3K4me1
0.2347985	Med1	GSM1439566	0.293793789	0.255319149	0.300898204	0.488038278	0.514382403	0.514382403	0.557580779	mESC_Med1_Bethyl_ChIPSeq
0.23474584	H3K4me1	GSM2630455	0.4145018	0.601823708	0.661676647	0.608791209	0.354066986	0.402707276	0.390223695	MLL4NT_FBS_H3K4me1
0.23436153	RAD21	GSM824848	0.307531923	0.17325228	0.134730539	0.183516484	0.291866029	0.392554992	0.395194698	Rad21_rep2 WT mES MM8
0.23417671	H3K4me1	GSM2636034	0.288313606	0.352583587	0.275449102	0.358241758	0.296650718	0.399323181	0.378624689	MLL3dSET_H3K4me1_1
0.23401604	KMT2B	GSM2588408	0.419927996	0.373860182	0.402694611	0.416483516	0.660287081	0.736040609	0.77133886	WT_ChIP_MLL4
0.23390326	ASH2L	GSM1258240	0.421611541	0.416413374	0.497005988	0.495604396	0.65502392	0.65820643	0.714995857	E14_Ash2L_GFP
0.23339892	NANOG	GSM1327152	0.260185449	0.285714286	0.263473054	0.314285714	0.454545455	0.507614213	0.517812759	ChIP-seq-s3-transient-IP
0.23331883	POU5F1	GSM1693787	0.082351265	0.246200608	0.254491018	0.201098901	0.133971292	0.082910321	0.096934548	ChIP-seq - mES - Pou5f1 - 12hr KD, 7SK ASO rep2
0.23237737	RAD21	GSM2099845	0.							

Motif.Name	logP_OD	Frequency_OD	logP_CD	Frequency_CD	logP_SD	Frequency_SD
OCT4-SOX2-TCF-NANOG(POU,Homeobox,HMG)/mES-Oct4-ChIP-Seq(GSE11431)/Homer	-2919	20.59%	-891.6	21.09%	-41.82	5.45%
Sox3(HMG)/NPC-Sox3-ChIP-Seq(GSE33059)/Homer	-730.5	36.62%	-740.7	52.71%	-905	66.63%
Sox2(HMG)/mES-Sox2-ChIP-Seq(GSE11431)/Homer	-681	23.16%	-611.2	34.34%	-704.9	44.38%
Sox6(HMG)/Myotubes-Sox6-ChIP-Seq(GSE32627)/Homer	-541.3	31.91%	-465.5	44.40%	-499.7	51.64%
Sox10(HMG)/SciaticNerve-Sox3-ChIP-Seq(GSE35132)/Homer	-337.4	29.12%	-452	43.16%	-636.3	56.16%
Sox17(HMG)/Endoderm-Sox17-ChIP-Seq(GSE61475)/Homer	-521.7	17.81%	-442	26.40%	-488.5	33.10%
Sox15(HMG)/CPA-Sox15-ChIP-Seq(GSE62909)/Homer	-494.2	23.01%	-430.4	33.37%	-584.1	43.95%
KLF5(Zf)/LoVo-KLF5-ChIP-Seq(GSE49402)/Homer	-619.1	24.59%	-213.6	22.47%	-59.68	18.54%
Sox4(HMG)/proB-Sox4-ChIP-Seq(GSE50066)/Homer	-163.1	15.05%	-187.6	21.26%	-424.3	34.57%
Oct4(POU,Homeobox)/mES-Oct4-ChIP-Seq(GSE11431)/Homer	-1278	20.08%	-175.9	15.47%	-5.813	6.32%
Bmi1(POU,Homeobox)/NPC-Bmi1-ChIP-Seq(GSE35496)/Homer	-1173	16.04%	-146.5	11.50%	-4.218	4.31%
KLF6(Zf)/PDAC-KLF6-ChIP-Seq(GSE64557)/Homer	-395	17.69%	-138.9	15.67%	-38.76	13.09%
Sox9(HMG)/Limb-SOX9-ChIP-Seq(GSE73225)/Homer	-105.2	14.31%	-138.3	20.05%	-216.7	27.04%
Sp2(Zf)/HEK293-Sp2-eGFP-ChIP-Seq(Encode)/Homer	-468.8	23.57%	-137.1	19.71%	-54.39	18.81%
EKLf(Zf)/Erythrocyte-Klf1-ChIP-Seq(GSE20478)/Homer	-301.8	6.00%	-122.6	6.12%	-51.21	5.18%
Klf4(Zf)/mES-Klf4-ChIP-Seq(GSE11431)/Homer	-389.9	8.91%	-122.3	8.14%	-49.77	7.14%
Oct6(POU,Homeobox)/NPC-Pou3f1-ChIP-Seq(GSE35496)/Homer	-1038	18.18%	-121.4	13.12%	-3.635	5.73%
Oct11(POU,Homeobox)/NCIH1048-POU2F3-ChIP-Seq(GSE115123)/Homer	-887.2	13.08%	-114.1	9.82%	-6.801	4.25%
KLF3(Zf)/MEF-Klf3-ChIP-Seq(GSE44748)/Homer	-456.8	10.41%	-104.1	8.04%	-41.32	7.25%
TSO1(CPP)/col-TSO1-DAP-Seq(GSE60143)/Homer	-24.11	2.04%	-88.66	5.28%	-3.309	1.85%
Zic(Zf)/Cerebellum-ZIC1.2-ChIP-Seq(GSE60731)/Homer	-387.9	16.51%	-73.14	12.88%	-16.47	9.98%
NFKB-p65(RHD)/GM12787-p65-ChIP-Seq(GSE19485)/Homer	-85.57	7.78%	-68.31	9.49%	-7.064	6.27%
Dlx3(Homeobox)/Keratinocytes-Dlx3-ChIP-Seq(GSE89884)/Homer	-53.66	9.80%	-62.58	14.09%	-2.179	7.63%
KLF14(Zf)/HEK293-KLF14.GFP-ChIP-Seq(GSE58341)/Homer	-225.5	23.99%	-60.36	20.48%	-13.32	18.43%
Oct2(POU,Homeobox)/Bcell-Oct2-ChIP-Seq(GSE21512)/Homer	-597.6	10.63%	-58.75	7.74%	-1.273	3.00%
Zic3(Zf)/mES-Zic3-ChIP-Seq(GSE37889)/Homer	-506.8	12.82%	-57.94	7.70%	-32.87	8.12%
ATHB24(ZFHD)/colamp-ATHB24-DAP-Seq(GSE60143)/Homer	-78.01	11.67%	-52.92	14.33%	-1.115	7.69%
Lhx2(Homeobox)/HfSC-Lhx2-ChIP-Seq(GSE48068)/Homer	-51.45	15.74%	-48.7	19.95%	-0.4345	11.61%
Unknown-ESC-element(?)mES-Nanog-ChIP-Seq(GSE11724)/Homer	-425.8	13.19%	-48.69	8.58%	-20.69	8.12%
ATHB33(ZFHD)/col-ATHB33-DAP-Seq(GSE60143)/Homer	-59.63	18.00%	-45.15	21.70%	-2.916	14.61%
Lhx1(Homeobox)/EmbryoCarcinoma-Lhx1-ChIP-Seq(GSE70957)/Homer	-86.97	18.09%	-43.24	20.92%	-0.1643	12.43%
Sp5(Zf)/mES-Sp5-Flag-ChIP-Seq(GSE72989)/Homer	-203.2	13.44%	-42.04	10.26%	-23.45	11.12%
TCX2(CPP)/colamp-TCX2-DAP-Seq(GSE60143)/Homer	-57.12	19.29%	-41.75	24.32%	-3.702	17.50%
ATHB23(ZFHD)/col-ATHB23-DAP-Seq(GSE60143)/Homer	-68.29	13.17%	-41.72	15.74%	-0.4451	8.62%
AT2G20110(CPP)/colamp-AT2G20110-DAP-Seq(GSE60143)/Homer	-78.19	19.26%	-41.69	23.44%	-3.912	16.52%
SGR5(C2H2)/colamp-SGR5-DAP-Seq(GSE60143)/Homer	-10.33	7.94%	-41.18	12.95%	-54.37	15.38%
SOL1(CPP)/colamp-SOL1-DAP-Seq(GSE60143)/Homer	-56.93	18.91%	-40.47	23.95%	-3.494	17.12%
TEAD4(TEA)/Tropoblast-Tea4-ChIP-Seq(GSE37350)/Homer	-277.9	15.94%	-39.16	13.19%	-101.4	18.59%
TEAD1(TEAD)/HepG2-TEAD1-ChIP-Seq(Encode)/Homer	-248.7	17.71%	-37.11	15.37%	-90.4	19.79%
ERRg(NR)/Kidney-ERRG-ChIP-Seq(GSE104905)/Homer	-255	15.18%	-34.62	11.67%	-57.59	15.59%
TEAD(TEA)/Fibroblast-PU.1-ChIP-Seq(Unpublished)/Homer	-174.4	12.44%	-34.62	11.57%	-74.33	15.16%
Lhx3(Homeobox)/Neuron-Lhx3-ChIP-Seq(GSE31456)/Homer	-54.68	23.38%	-33.2	27.28%	-1.09	19.47%
Nanog(Homeobox)/mES-Nanog-ChIP-Seq(GSE11724)/Homer	-88.82	53.56%	-33.2	54.83%	-7.353	48.85%
LEF1(HMG)/H1-LEF1-ChIP-Seq(GSE64758)/Homer	-136.5	12.66%	-32.1	12.48%	-53.85	15.16%
Klf9(Zf)/GBM-Klf9-ChIP-Seq(GSE62211)/Homer	-147.3	6.49%	-32.08	5.48%	-11.68	5.07%
Nrf2(bZIP)/Lymphoblast-Nrf2-ChIP-Seq(GSE37589)/Homer	-88.99	1.51%	-31.68	1.58%	-16.8	1.36%
TEAD2(TEA)/Py2T-Tea2-ChIP-Seq(GSE55709)/Homer	-189.7	10.59%	-31.01	9.12%	-74.45	12.70%
Foxa2(Forkhead)/Liver-Foxa2-ChIP-Seq(GSE25694)/Homer	-11.26	9.18%	-30.98	12.92%	-52.07	15.49%
ATHB34(ZFHD)/colamp-ATHB34-DAP-Seq(GSE60143)/Homer	-62.54	11.85%	-30.7	13.56%	-0.1669	7.25%
LXh9(Homeobox)/Hct116-LXh9.V5-ChIP-Seq(GSE116822)/Homer	-67.29	20.81%	-29.89	22.97%	-0.8011	16.09%
TEAD3(TEA)/HepG2-TEAD3-ChIP-Seq(Encode)/Homer	-251.5	19.97%	-28.74	16.85%	-78.35	21.54%
Esrnb(NR)/mES-Esrnb-ChIP-Seq(GSE11431)/Homer	-234.6	13.09%	-27.71	9.69%	-48.95	12.81%
SUT1?/SacCer-Promoters/Homer	-77.55	58.77%	-22.77	51.73%	-11.59	54.36%
Maz(Zf)/HepG2-Maz-ChIP-Seq(GSE31477)/Homer	-74.92	15.90%	-22.62	13.22%	-1.644	11.67%
SeqBias: A/T bias	-76.29	94.55%	-22.08	97.54%	-9.329	95.42%
NF-E2(bZIP)/K562-NFE2-ChIP-Seq(GSE31477)/Homer	-73.71	1.52%	-19.57	1.35%	-9.438	1.15%
Bach1(bZIP)/K562-Bach1-ChIP-Seq(GSE31477)/Homer	-100.1	1.68%	-18.42	1.41%	-14.77	1.31%
Etv2(ETS)/ES-ER71-ChIP-Seq(GSE59402)/Homer	-98.83	14.22%	-18.36	12.31%	-3.597	11.18%
Nf5a2(NR)/Pancreas-LRH1-ChIP-Seq(GSE34295)/Homer	-177.5	13.95%	-15.54	10.70%	-5.832	11.01%
ETS1(ETS)/Jurkat-ETS1-ChIP-Seq(GSE17954)/Homer	-119.1	15.93%	-15.52	12.85%	-11.63	13.90%
KLF10(Zf)/HEK293-KLF10.GFP-ChIP-Seq(GSE58341)/Homer	-96.37	9.07%	-15.45	7.74%	-4.991	7.25%
Flt1(ETS)/CD8-FL1-ChIP-Seq(GSE20898)/Homer	-98.83	15.33%	-14.81	12.31%	-4.393	12.27%
ERG(ETS)/VCaP-ERG-ChIP-Seq(GSE14097)/Homer	-118.4	23.45%	-14.69	19.71%	-3.153	18.81%
KAN2(G2like)/colamp-KAN2-DAP-Seq(GSE60143)/Homer	-55.28	17.25%	-13.9	18.47%	-9.814	16.96%
Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer	-93.62	6.60%	-13.3	5.52%	-18.55	6.22%
Bach2(bZIP)/JCILy7-Bach2-ChIP-Seq(GSE44420)/Homer	-68.75	3.15%	-13.11	2.66%	-14.53	2.94%
Zfp281(Zf)/ES-Zfp281-ChIP-Seq(GSE81042)/Homer	-59.28	3.99%	-12.63	3.03%	-4.549	2.94%
Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer	-77.93	8.52%	-12.53	7.70%	-16.88	8.18%
GABPA(ETS)/Jurkat-GABPA-ChIP-Seq(GSE17954)/Homer	-70.05	12.43%	-11.82	10.36%	-24.37	13.47%
AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	-74.1	9.37%	-11.46	8.34%	-13.7	8.72%
AP-2gamma(AP2)/MCF7-TFAP2C-ChIP-Seq(GSE21234)/Homer	-1.118	8.55%	-11.29	8.78%	-188.8	22.52%
Jun-AP1(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	-67.68	3.43%	-11.25	2.89%	-12.09	3.05%
Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer	-58.36	6.92%	-11.02	6.39%	-18.63	7.09%
AP-2alpha(AP2)/Hela-AP2alpha-ChIP-Seq(GSE31477)/Homer	-0.5143	6.09%	-11	6.59%	-219.4	19.85%
COUP-TFII(NR)/K562-NR2F1-ChIP-Seq(Encode)/Homer	-103.8	19.55%	-10.74	15.91%	-25.27	19.25%
Tcf4(HMG)/Hct116-Tcf4-ChIP-Seq(SRA012054)/Homer	-52.31	7.35%	-10.68	7.16%	-32.24	9.81%
SF1(NR)/H295R-Nr5a1-ChIP-Seq(GSE44220)/Homer	-160.4	10.31%	-10.57	7.20%	-19.63	9.81%
SeqBias: CG bias	-60.27	78.19%	-10.48	69.69%	-17.04	75.30%
Elk1(ETS)/Hela-Elk1-ChIP-Seq(GSE31477)/Homer	-61.15	6.75%	-10.09	4.91%	-5.574	5.45%
JunB(bZIP)/DendriticCells-JunB-ChIP-Seq(GSE36099)/Homer	-59.8	6.82%	-10.07	6.12%	-19.91	7.20%
Nf5a2(NR)/mES-Nf5a2-ChIP-Seq(GSE19019)/Homer	-126.5	10.42%	-10	7.77%	-8.971	9.05%
EAR2(NR)/K562-NR2F6-ChIP-Seq(Encode)/Homer	-106.6	18.14%	-9.252	14.26%	-24.16	17.56%
ETV1(ETS)/GIST48-ETV1-ChIP-Seq(GSE22441)/Homer	-96.63	19.79%	-8.701	15.84%	-13.76	18.65%
ETV4(ETS)/HepG2-ETV4-ChIP-Seq(ENCODE)/Homer	-104.6	15.66%	-8.558	11.60%	-5.317	12.60%
Fosl2(bZIP)/3T3L1-Fosl2-ChIP-Seq(GSE56872)/Homer	-61.78	4.41%	-7.867	3.57%	-15.33	4.36%
COUP-TFII(NR)/Aria-Nr2f2-ChIP-Seq(GSE46497)/Homer	-91.84	22.57%	-7.283	18.06%	-22.75	21.86%
BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer	-61.48	8.29%	-7.106	7.23%	-13.25	8.07%
EHF(ETS)/LoVo-EHF-ChIP-Seq(GSE49402)/Homer	-70.96	18.71%	-6.999	16.11%	-9.568	18.48%
EWS-FL1-fusion(ETS)/SK_N_MC-EWS-FL11-ChIP-Seq(SRA014231)/Homer	-53.57	8.26%	-6.597	6.73%	-2.382	6.71%
ELF3(ETS)/PDAC-ELF3-ChIP-Seq(GSE64557)/Homer	-73.56	11.62%	-6.439	9.59%	-17.31	12.21%
Elk4(ETS)/Hela-Elk4-ChIP-Seq(GSE31477)/Homer	-61.26	6.43%	-6.105	4.31%	-3.91	4.96%
GLIS3(Zf)/Thyroid-Glis3.GFP-ChIP-Seq(GSE103297)/Homer	-114.3	16.95%	-5.981	11.44%	-6.016	12.80%
NFE2L2(bZIP)/HepG2-NFE2L2-ChIP-Seq(Encode)/Homer	-60.1	1.53%	-5.673	1.11%	-13.52	1.53%
Eif4(ETS)/BMDM-Eif4-ChIP-Seq(GSE88699)/Homer	-72.84	15.18%	-5.565	11.87%	-16.52	15.16%
MYNN(Zf)/HEK293-MYNN.eGFP-ChIP-Seq(Encode)/Homer	-54.56	4.19%	-4.843	3.46%	-1.71	2.89%
RAR.RXR(NR),DR5/ES-RAR-ChIP-Seq(GSE56893)/Homer	-116.7	4.65%	-4.471	2.76%	-1.399	0.33%
RAR.RXR(NR),DR5/ES-RAR-ChIP-Seq(GSE56893)/Homer	-116.7	4.65%	-4.471	2.76%	-22.45	4.96%
Pit1(Homeobox)/GCrat-Pit1-ChIP-Seq(GSE8009)/Homer	-59.04	14.02%	-2.433	12.28%	0	6.32%
Unknown4/Drosophila-Promoters/Homer	-54.44	2.10%	-1.621	1.35%	-0.05563	0.60%
SPDEF(ETS)/VCaP-SPDEF-ChIP-Seq(SRA014231)/Homer	-72.82	14.88%	-1.262	10.63%	-19.62	15.05%
RARg(NR)/ES-RARg-ChIP-Seq(GSE30538)/Homer	-55.92	2.37%	-0.5941	1.21%	-0.2176	1.20%
RAR.RXR(NR),DR5/ES-RAR-ChIP-Seq(GSE56893)/Homer	-116.7	4.65%	-0.1349	0.13%	-1.399	0.33%
RAR.RXR(NR),DR5/ES-RAR-ChIP-Seq(GSE56893)/Homer	-116.7	4.65%	-0.1349	0.13%	-22.45	4.96%

Appendix table 2

Enrichment (logP) and frequency of known motifs from HOMER in the sequences of OD, CD, and SD loci.

MotifName	logP_OCT4up	Frequency_OCT4up	logP_SOX2up	Frequency_SOX2up
AP-2alpha(AP2)Hela-AP2alpha-ChIP-Seq(GSE31477)Homer	-1850	25.30%	-12.51	12.69%
AP-2gamma(AP2)MCF7-TFAP2C-ChIP-Seq(GSE21234)Homer	-1807	30.09%	-8.492	15.28%
Sox3(HMG)NPC-Sox3-ChIP-Seq(GSE33059)Homer	-1334	30.04%	-5.00E-06	11.30%
Sox3(HMG)mES-Sox3-ChIP-Seq(GSE11431)Homer	-1170	18.16%	-0.002081	5.29%
Sox3(HMG)CPA-Sox3-ChIP-Seq(GSE32969)Homer	-1079	19.31%	-0.006203	7.13%
DMRT6(DM)Testis-DMRT6-ChIP-Seq(GSE60440)Homer	-1041	6.68%	-0.5532	1.39%
Sox6(HMG)Myotubes-Sox6-ChIP-Seq(GSE32627)Homer	-1032	25.14%	-1.60E-05	9.81%
DMRT1(DM)Testis-DMRT1-ChIP-Seq(GSE48829)Homer	-1018	7.02%	-0.4795	1.48%
TEAD4(TEA)Troponin-Tead4-ChIP-Seq(GSE33550)Homer	-940	15.09%	-94.19	20.46%
Sox10(HMG)SkeletalTerve-Sox3-ChIP-Seq(GSE35132)Homer	-934.7	26.13%	-6.00E-06	10.74%
TEAD1(TEA)HepG2-TEAD1-ChIP-Seq(Encode)Homer	-899.2	16.35%	-84.61	21.57%
Sox4(HMG)proB-Sox4-ChIP-Seq(GSE50066)Homer	-859.5	15.67%	-0.007798	5.93%
Sox17(HMG)Endoderm-Sox17-ChIP-Seq(GSE81475)Homer	-854.5	13.86%	-5.40E-05	3.43%
TEAD2(TEA)Troponin-Tead2-ChIP-Seq(Encode)Homer	-811.9	11.97%	-81.51	16.48%
TEAD3(TEA)HepG2-TEAD3-ChIP-Seq(Encode)Homer	-726.8	17.44%	-72.59	22.59%
TEAD2(TEA)PyZT-TEAD2-ChIP-Seq(GSE55709)Homer	-698.4	9.92%	-66.4	13.24%
Sp2(ZY)HEK293-Sp2-eGFP-ChIP-Seq(Encode)Homer	-547.9	29.60%	-58.08	30.37%
KLF5(ZY)Lovo-KLF5-ChIP-Seq(GSE49402)Homer	-456.5	25.62%	-60.78	27.96%
ERRG(NR)Kidney-ERRG-ChIP-Seq(GSE104005)Homer	-390.3	12.96%	-23.89	13.52%
Sp5(ZY)mES-Sp5-Flag-ChIP-Seq(GSE72969)Homer	-386.1	21.92%	-40.85	21.48%
Esrab(NR)mES-Esrab-ChIP-Seq(GSE11431)Homer	-384.6	10.75%	-24.17	11.39%
KLF3(ZY)MEF-Klf3-ChIP-Seq(GSE44748)Homer	-365.8	12.30%	-43.58	12.41%
Klf4(ZY)mES-Klf4-ChIP-Seq(GSE33151)Homer	-301.2	6.16%	-46.85	10.65%
Gata6(ZY)HUG-1N-GATA6-ChIP-Seq(GSE51936)Homer	-299.8	10.94%	-2.75	8.06%
Sox9(HMG)Limb-SOX9-ChIP-Seq(GSE73225)Homer	-296.3	13.13%	-0.001191	6.30%
GATA3(ZY)Treg-Gata3-ChIP-Seq(GSE20888)Homer	-290.6	16.30%	-0.9243	12.13%
Gata3(ZY)K562-GATA3-ChIP-Seq(GSE18259)Homer	-286.6	8.42%	-2.667	6.02%
Gata4(ZY)Heart-Gata4-ChIP-Seq(GSE33151)Homer	-283	11.84%	-3.254	9.35%
COUP-TFII(NR)K562-NR2F1-ChIP-Seq(Encode)Homer	-278.9	18.91%	-17.05	20.19%
KLF14(ZY)HEK293-KLF14-GFP-ChIP-Seq(GSE58341)Homer	-276.2	30.68%	-39.05	32.41%
Gata1(ZY)K562-GATA1-ChIP-Seq(GSE18829)Homer	-272.8	7.74%	-2.998	5.56%
EARG(NR)K562-NR2F6-ChIP-Seq(Encode)Homer	-268.6	17.34%	-10.82	17.41%
KLF2(ZY)PDAC-KLF2-ChIP-Seq(GSE45573)Homer	-239.4	20.48%	-44.75	22.96%
TRPS1(ZY)MCF7-TRPS1-ChIP-Seq(GSE107013)Homer	-239.1	20.46%	-2.065	17.50%
Sp1(ZY)PromoterHomer	-219.5	8.83%	-13.55	5.37%
COUP-TFII(NR)Artis-Nr2Z-ChIP-Seq(GSE46487)Homer	-214.3	21.49%	-11.1	22.41%
Foxd2(ZY)P13-Foxd2-ChIP-Seq(GSE6872)Homer	-211.7	4.64%	-13.82	4.56%
Jun-AP1(bZIP)K562-c-Jun-ChIP-Seq(GSE31477)Homer	-211.3	3.76%	-14.82	3.98%
Isl1(Homeobox)Neuron-Isl1-ChIP-Seq(GSE31456)Homer	-197.3	20.56%	-5.76	20.19%
Klf9(ZY)GBM-Klf9-ChIP-Seq(GSE82211)Homer	-193.7	9.10%	-42.54	11.20%
LEF1(HMG)H1-LEF1-ChIP-Seq(GSE4758)Homer	-191.1	8.71%	-53.74	14.81%
SF1(NR)H29R-Nr5a1-ChIP-Seq(GSE44220)Homer	-187.7	8.43%	-11.05	8.43%
EKLF(ZY)Erythocyte-Klf1-ChIP-Seq(GSE20478)Homer	-184.6	4.12%	-50.43	7.50%
N5a2(NR)Pancreas-LRH1-ChIP-Seq(GSE34295)Homer	-183.5	11.38%	-5.432	10.46%
Fra1(bZIP)BT549-Fra1-ChIP-Seq(GSE46166)Homer	-177.6	6.26%	-8.196	6.39%
N5a2(NR)mES-N5a2-ChIP-Seq(GSE19019)Homer	-172.7	9.04%	-1.175	8.43%
Junb(bZIP)DendriticCulm-Junb-ChIP-Seq(GSE36369)Homer	-165.4	6.38%	-11.88	6.94%
Eras(NR)HepG2-Era-ChIP-Seq(GSE31477)Homer	-161.6	29.05%	-0.6525	25.56%
Fra2(bZIP)Striatum-Fra2-ChIP-Seq(GSE43429)Homer	-158.3	5.71%	-14.22	6.67%
BATF(bZIP)Th17-BATF-ChIP-Seq(GSE38756)Homer	-156.8	7.23%	-6.142	7.04%
AISG(ZY)GBM-ATF3-ChIP-Seq(GSE33812)Homer	-155	7.24%	-6.527	7.78%
ETS1(ETS)Jurkat-ETS1-ChIP-Seq(GSE17554)Homer	-148.9	14.39%	-15.86	16.48%
SeqBias: CG bias	-143.5	87.09%	-8.367	87.59%
AISG05790(MYRelated)Ico1-AISG05790-DAP-Seq(GSE60143)Homer	-137.6	9.13%	-0.7869	6.57%
NLP1(RWPK)Ico1-NLP1-DAP-Seq(GSE60143)Homer	-137.4	21.58%	-2.5	19.63%
EBF1(EBF)Hear-E2f4-ChIP-Seq(GSE21512)Homer	-134.2	14.69%	-1.211	11.85%
Bach2(bZIP)OCILy7-Bach2-ChIP-Seq(GSE44420)Homer	-134.1	3.01%	-11.3	3.52%
AP-1(bZIP)ThioMac-Pu.1-ChIP-Seq(GSE21512)Homer	-133.9	7.98%	-7.689	8.33%
GABPA(ETS)Jurkat-GABPA-ChIP-Seq(GSE17954)Homer	-125.4	13.46%	-13.65	14.54%
Nanog(Homeobox)mES-Nanog-ChIP-Seq(GSE11724)Homer	-124.4	41.65%	-2.827	41.57%
AISG11280(MYRelated)Ico1-AISG11280-DAP-Seq(GSE60143)Homer	-122.2	8.56%	-0.590	6.11%
AISG08520(MYRelated)Ico1-AISG08520-DAP-Seq(GSE60143)Homer	-121.8	10.69%	-0.2923	7.78%
RARa(NR)K562-RARa-ChIP-Seq(Encode)Homer	-121.1	32.40%	-7.795	35.37%
At1g49010(MYRelated)Ico1-At1g49010-DAP-Seq(GSE60143)Homer	-120.5	17.20%	-0.09223	13.24%
THRB(NR)Liver-NR4A1-ChIP-Seq(GSE33813)Homer	-118.6	46.90%	-1.825	45.83%
NF-E2B2(ZY)K562-NFE2B2-ChIP-Seq(GSE31477)Homer	-117.5	1.27%	-11.64	1.57%
Bach1(bZIP)K562-Bach1-ChIP-Seq(GSE31477)Homer	-117.2	1.16%	-8.562	1.30%
Nrf2(bZIP)Lymphoblast-Nrf2-ChIP-Seq(GSE37589)Homer	-116.7	1.10%	-5.905	1.02%
EBF4(ETS)BMDM-E4-ChIP-Seq(GSE36999)Homer	-115.6	14.76%	-11.82	16.57%
Tcf7(HMG)GM12878-Tcf7-ChIP-Seq(Encode)Homer	-113.8	4.46%	-38.55	9.35%
Tcf4(HMG)Hct1-Tcf4-ChIP-Seq(SRA012054)Homer	-112.5	5.88%	-68.36	12.69%
ELF1(ETS)Jurkat-ELF1-ChIP-Seq(SRA014231)Homer	-111	8.77%	-10.1	8.33%
PQM-1(7)cElegans-L3-ChIP-Seq(modEncode)Homer	-109.5	3.68%	-1.997	2.96%
SPDEF(ETS)VCaP-SPDEF-ChIP-Seq(SRA014231)Homer	-106.2	13.59%	-11.79	15.83%
Hoxa9(Homeobox)ChickenEG-Hoxa9-Flag-ChIP-Seq(GSE86085)Homer	-103.4	25.49%	-2.169	25.46%
UnknownsDrosophila-PromotersHomer	-98.8	6.50%	-0.3619	4.81%
ERG(ETS)VCaP-ERG-ChIP-Seq(GSE14097)Homer	-97.39	19.84%	-19.34	24.44%
Nurr7(NR)K562-NR4A1-ChIP-Seq(GSE31363)Homer	-96.56	2.23%	-16.64	3.43%
AISG5990(MYRelated)Ico1-AISG5990-DAP-Seq(GSE60143)Homer	-96.87	12.05%	-1.06	10.00%
Lhx3(Homeobox)Neuron-Lhx3-ChIP-Seq(GSE31456)Homer	-95.55	15.00%	-6.212	16.11%
Foxa2(Forkhead)Liver-Foxa2-ChIP-Seq(GSE25694)Homer	-94.51	7.94%	-8.118	8.70%
Maz(ZY)HepG2-Maz-ChIP-Seq(GSE31477)Homer	-94.34	22.44%	-20.83	23.33%
ELF3(ETS)PDAC-ELF3-ChIP-Seq(GSE46557)Homer	-94.27	8.60%	-5.906	10.19%
ELF5(ETS)T47D-ELF5-ChIP-Seq(GSE303407)Homer	-91.95	10.09%	-6.185	10.52%
TCFL2(HMG)K562-TCFL2-ChIP-Seq(GSE29196)Homer	-91.29	1.44%	-84.71	6.02%
caudal(Homeobox)Drosophila-Embryos-ChIP-Chip(modEncode)Homer	-91.21	7.01%	-3.455	6.57%
Foxa3(Forkhead)Liver-Foxa3-ChIP-Seq(GSE77670)Homer	-91.09	3.61%	-6.29	3.61%
c-Jun-CRE(bZIP)K562-c-Jun-ChIP-Seq(GSE31477)Homer	-88.06	3.24%	-3.489	2.87%
EBM4(ETS)Hela-Eb4-ChIP-Seq(GSE31477)Homer	-86.6	6.97%	-11.67	9.26%
Dlx3(Homeobox)Keratinocytes-Dlx3-ChIP-Seq(GSE89884)Homer	-86.56	5.85%	-1.549	4.81%
LXH9(Homeobox)Hct116-LXH9-V5-ChIP-Seq(GSE116822)Homer	-86.41	12.89%	-3.741	13.52%
AT2038300(G2like)Ico1-AT2038300-DAP-Seq(GSE60143)Homer	-86.29	11.77%	-12.87	15.19%
AR2038301(G2like)Ico1-AR2038301-DAP-Seq(GSE60143)Homer	-83.09	3.58%	-3.623	3.15%
AT240460(G2like)Ico1-AT240460-DAP-Seq(GSE60143)Homer	-82.66	13.93%	-22.66	19.28%
SUT17(SacCer-Promoters)Homer	-82.47	70.72%	-15.55	71.57%
Lhx2(Homeobox)HFSC-Lhx2-ChIP-Seq(GSE48068)Homer	-82.16	9.67%	-3.701	9.81%
Ahrf(bZIP)Zyfl1-Ahrf-ChIP-Seq(GSE56872)Homer	-81.42	4.71%	-3.8	4.35%
E2f2(ETS)ES-E2F1-ChIP-Seq(GSE59402)Homer	-80.84	11.61%	-12.57	14.26%
FoxL2(Forkhead)Ovary-FoxL2-ChIP-Seq(GSE60658)Homer	-79.37	7.74%	-8.862	8.89%
FOXA1(Forkhead)MCF7-FOXA1-ChIP-Seq(GSE26831)Homer	-79.13	9.32%	-8.797	10.46%
ELT-3(Gata)cElegans-L1-ELT3-ChIP-Seq(modEncode)Homer	-78.7	4.03%	-1.483	3.24%
Tcf12(HLH)MG12878-Tcf12-ChIP-Seq(GSE32465)Homer	-77.93	13.10%	-6.653	14.63%
KAN2(G2like)Ico1-KAN2-DAP-Seq(GSE60143)Homer	-77.41	11.27%	-17.56	15.37%
Zfp281(ZY)ES-Zfp281-ChIP-Seq(GSE81042)Homer	-77.1	4.78%	-11.62	6.20%
EBF2(EBF)BrownAdipose-EBF2-ChIP-Seq(GSE97114)Homer	-75.62	11.21%	-8.837	11.67%
EPH2(ZY)Lovo-EHP-ChIP-Seq(GSE49402)Homer	-75.56	16.22%	-6.505	17.87%
PIE1(ETS)CDB-PIE1-ChIP-Seq(GSE20898)Homer	-75.54	15.29%	-19.59	18.24%
SGR5(C2H2)Ico1-SGR5-DAP-Seq(GSE60143)Homer	-72.12	7.16%	-0.1526	4.63%
Tbet(T-box)CD8-Tbet-ChIP-Seq(GSE33802)Homer	-70.76	10.01%	-22.66	14.07%
Stat3(Stat)mES-Stat3-ChIP-Seq(GSE11431)Homer	-70.35	6.46%	-6.006	7.04%
PPAR(NR)DR1/3TL1-Ppar-ChIP-Seq(GSE13511)Homer	-69.14	10.92%	-0.9811	10.19%
CDX4(Homeobox)Zebrafish-Embryos-CDX4-Myc-ChIP-Seq(GSE48254)Homer	-68.97	7.60%	-3.548	7.41%
Thr1(T-box)Cortex-Thr1-ChIP-Seq(GSE71384)Homer	-68.84	13.00%	-8.989	15.46%
FOXA1(Forkhead)LNCAP-FOXA1-ChIP-Seq(GSE27824)Homer	-67.12	11.12%	-7.993	12.04%
Rorin(THAP)ES-Thap11-ChIP-Seq(GSE15522)Homer	-66.2	0.61%	-6.629	0.56%
Nux5.1(Homeobox)Intest-Nux5.1-ChIP-Seq(GSE40975)Homer	-66.12	21.96%	-1.215	21.20%
CTCF(ZY)CD4+CTCF-ChIP-Seq(Baeti_et_al_)Homer	-66.04	2.27%	-26.64	3.89%
Six4(Homeobox)MCF7-SIX4-ChIP-Seq(Encode)Homer	-65.31	0.82%	-19.91	1.85%
Tcf3(HMG)mES-Tcf3-ChIP-Seq(GSE11724)Homer	-65.18	3.26%	-82.81	10.28%
Tcf21(HLH)A1NerySmoothMuscle-Tcf21-ChIP-Seq(GSE81369)Homer	-62.33	12.26%	-4.043	12.50%
MYB101(MY)B101-Ico1-MYB101-DAP-Seq(GSE60143)Homer	-61.59	19.36%	-2.674	19.26%
Tbx21(T-box)GM12878-Tbx21-ChIP-Seq(Encode)Homer	-61.45	9.48%	-17.55	12.96%
ETS(ETS)PromoterHomer	-60.62	5.51%	-7.203	5.37%
RXR(NR)DR1/3TL1-RXR-ChIP-Seq(GSE13511)Homer	-59.53	13.25%	-1.896	13.52%
Aq4(HLH)AML-Thap4-ChIP-Seq(GSE45738)Homer	-58.85	15.35%	-6.941	16.85%
Tbx6(T-box)JES2-Tbx6-ChIP-Seq(GSE25824)Homer	-58.81	10.73%	-13.17	13.80%
Unknown(Homeobox)Limb-p300-ChIP-SeqHomer	-58.14	5.78%	-0.9111	5.19%
HNFA4a(NR)DR1/HepG2-HNF4a-ChIP-Seq(GSE25021)Homer	-57.69	5.38%	-1.369	5.00%
ATHB33(ZFH)Ico1-ATHB33-DAP-Seq(GSE60143)Homer	-56.45	10.61%	-2.949	10.93%
Foxo3(Forkhead)U2OS-Foxo3-ChIP-Seq-MTAB-2701)Homer	-56.11	6.96%	-16.54	9.17%
Zc3h2(ZY)mES-Zc3h2-ChIP-Seq(GSE37899)Homer	-55.68	6.70%	-61.79	16.02%
JunD(bZIP)K562-JunD-ChIP-SeqHomer	-54.41	1.17%	-1.814	0.74%
Ocat-Sox17(POU)Homeobox(HMG)F9-Sox17-ChIP-Seq(GSE44553)Homer	-54.05	1.54%	-0.4392	0.65%
Hoxa1(Homeobox)ChickenMSG-Hoxa11-Flag-ChIP-Seq(GSE86088)Homer	-54.04	19.91%	-1.95	20.56%
Ahrf(bZIP)K562-Ahrf-ChIP-Seq(GSE31477)Homer	-53.77	6.11%	-5.396	6.30%
Lhx1(Homeobox)EmbryoCarcinoma-Lhx1-ChIP-Seq(GSE70957)Homer	-53.37	9.46%	-2.766	10.00%
PPARa(NR)DR1/Liver-Ppara-ChIP-Seq(GSE47954)Homer	-52.99	12.30%	-4.127	13.52%
STAT5(Stat)mCD4+Stat5-ChIP-Seq(GSE12346)Homer	-52.17	3.84%	-4.316	4.54%
Bcl6(ZY)Liver-Bcl6-ChIP-Seq(GSE31578)Homer	-52.16	14.84%	-2.731	16.11%
ETV1(ETS)GIST48-ETV1-ChIP-Seq(GSE22441)Homer	-51.77	18.92%	-11.02	21.57%
Unknown-ESC-element(T)mES-Nanog-ChIP-Seq(GSE11724)Homer	-51.42	7.65%	-70.1	16.57%
FOXX1(Forkhead)HEK293-FOXX1-ChIP-Seq(GSE51673)Homer	-51.21	9.64%	-8.403	11.02%
Cdx2(Homeobox)mES-Cdx2-ChIP-Seq(GSE14586)Homer	-50.82	5.68%	-2.195	5.46%
Bcl11a(ZY)HSPC-BCL11A-ChIP-Seq(GSE104676)Homer	-50.3	8.26%	-1.642	7.69%
Foxo1(Forkhead)RAW-Foxo1-ChIP-Seq(Fan_et_al_)Homer	-50.26	20.49%	-1.777	20.00%

Appendix table 3

Enrichment (logP) and frequency of known motifs from HOMER in the sequences of loci gaining accessibility upon loss of OCT4 (OCT4up) and SOX2 (SOX2up).

Motif.Name	logP_Cluster1	Frequency_Cluster1	logP_Cluster2	Frequency_Cluster2	logP_Cluster3	Frequency_Cluster3	logP_Cluster4	Frequency_Cluster4
OCT4-SOX2-TCF-NANOG(POU.Homeobox.HMG)mES-Oct4-ChIP-Seq(GSE11431)Homer	-798.4	24.46%	-230.7	9.86%	-219.2	7.13%	-63.78	2.84%
KLf6(Zf)LoVo-KLf6-ChIP-Seq(GSE49402)Homer	-259.6	29.26%	-93.63	21.74%	-471.7	31.26%	-232	30.39%
Oct4(POU.Homeobox)mES-Oct4-ChIP-Seq(GSE11431)Homer	-244.7	19.35%	-151.7	13.41%	-92.72	8.65%	-97.85	6.25%
Sox3(HMG)NPC-Sox3-ChIP-Seq(GSE33059)Homer	-238.9	41.80%	-188.6	34.33%	-710.1	44.40%	-562.2	30.71%
Bmi1(POU.Homeobox)NPC-Bmi1-ChIP-Seq(GSE35496)Homer	-212.3	14.92%	-170.4	11.24%	-81.47	6.48%	-98.57	4.91%
KLf6(Zf)PDAC-KLf6-ChIP-Seq(GSE64557)Homer	-204.5	22.19%	-55.09	15.95%	-329.4	23.34%	-131.7	23.84%
Oct6(POU.Homeobox)NPC-Pou3f1-ChIP-Seq(GSE35496)Homer	-182.5	16.76%	-139	12.74%	-58.19	7.55%	-62.54	5.61%
Sox2(HMG)mES-Sox2-ChIP-Seq(GSE11431)Homer	-172.2	24.62%	-155.2	20.36%	-644.2	29.19%	-434	18.37%
Klf4(Zf)mES-Klf4-ChIP-Seq(GSE11431)Homer	-163.7	11.86%	-79.77	8.18%	-340.4	12.94%	-147	9.88%
Oct11(POU.Homeobox)NCIH1048-POU2F3-ChIP-Seq(GSE115123)Homer	-163.4	12.49%	-124.3	9.04%	-92.36	6.08%	-111.3	4.73%
Sox6(HMG)Myotubes-Sox6-ChIP-Seq(GSE32627)Homer	-163	35.32%	-102.4	27.87%	-469.7	36.90%	-314.1	24.08%
Sp2(Zf)HEK293-Sp2-eGFP-ChIP-Seq(Encode)Homer	-159.8	25.36%	-85.42	22.23%	-358.4	30.37%	-258.8	34.20%
EKLF(Zf)Erythrocyte-Klf1-ChIP-Seq(GSE20478)Homer	-156.6	9.07%	-71.76	5.49%	-311.1	9.45%	-133	5.15%
Sox17(HMG)Endoderm-Sox17-ChIP-Seq(GSE1475)Homer	-142.8	19.82%	-107	15.13%	-468.3	22.09%	-303.2	13.46%
Sox10(HMG)SciaticNerve-Sox3-ChIP-Seq(GSE35132)Homer	-137.7	34.05%	-92.32	27.53%	-491.8	38.15%	-329.8	25.37%
Sox15(HMG)CPA-Sox15-ChIP-Seq(GSE62909)Homer	-127.8	24.46%	-116.4	20.66%	-546.8	30.09%	-332.6	18.41%
KLf3(Zf)MEF-Klf3-ChIP-Seq(GSE44748)Homer	-126.5	10.96%	-81.05	9.53%	-355.5	14.76%	-184.6	15.06%
KLF14(Zf)HEK293-KLF14-GFP-ChIP-Seq(GSE58341)Homer	-107.6	26.99%	-49.5	23.94%	-158.7	28.70%	-117.1	35.65%
Oct2(POU.Homeobox)Boel1-Oct2-ChIP-Seq(GSE21512)Homer	-97.63	9.75%	-65.19	6.91%	-63.5	4.86%	-60.6	3.72%
ERRg(NR)Kidney-ESRRG-ChIP-Seq(GSE104905)Homer	-89.66	17.98%	-112	16.17%	-227.3	19.30%	-139	12.65%
Esrrb(NR)mES-Esrrb-ChIP-Seq(GSE11431)Homer	-86.28	15.76%	-113.5	14.23%	-215.5	16.70%	-130.1	10.23%
Sp5(Zf)mES-Sp5-Flag-ChIP-Seq(GSE72989)Homer	-73.93	14.44%	-45.36	14.23%	-131.7	17.15%	-140.6	25.10%
LEF1(HMG)H1-LEF1-ChIP-Seq(GSE64758)Homer	-70.55	16.50%	-36.37	12.22%	-114.6	14.04%	-68.54	8.88%
Sox9(HMG)Limf-SOX9-ChIP-Seq(GSE7325)Homer	-63.08	18.29%	-31.32	13.34%	-147.8	18.32%	-111.7	12.12%
TEAD4(Zf)Tropoblast-Tea4-ChIP-Seq(GSE27350)Homer	-62.03	16.08%	-158.8	18.75%	-143.4	16.21%	-309.2	14.82%
TEAD1(TEAD)HepG2-TEAD1-ChIP-Seq(Encode)Homer	-61.83	18.29%	-133.2	19.76%	-163.3	18.50%	-292.1	16.09%
Nf5a2(NR)Pancreas-LRH1-ChIP-Seq(GSE34295)Homer	-58.67	16.45%	-53.52	13.26%	-139.3	17.33%	-67.3	10.34%
Klf9(Zf)GBM-Klf9-ChIP-Seq(GSE62211)Homer	-57.7	7.91%	-24.55	6.61%	-123.6	9.35%	-69.14	10.56%
SF1(NR)H295R-Nf5a1-ChIP-Seq(GSE44220)Homer	-56.56	12.34%	-43.73	9.60%	-144.3	13.41%	-92.06	8.00%
Sox4(HMG)proB-Sox4-ChIP-Seq(GSE50066)Homer	-53.74	16.61%	-47.45	14.31%	-256	20.67%	-223.9	14.22%
TEAD2(TEA)PyT2-Tea2-ChIP-Seq(GSE55709)Homer	-53.56	11.54%	-121.6	12.92%	-116.3	11.07%	-233.6	9.82%
Tcf4(HMG)Hct116-Tcf4-ChIP-Seq(SRA012054)Homer	-49.13	11.02%	-33.75	8.59%	-59.46	8.75%	-46.56	5.78%
TEAD3(TEA)HepG2-TEAD3-ChIP-Seq(Encode)Homer	-46.24	19.35%	-124.7	21.82%	-129.2	19.70%	-241.3	17.15%
Nf5a2(NR)mES-Nf5a2-ChIP-Seq(GSE19019)Homer	-43.16	12.12%	-43.14	10.35%	-120.1	13.69%	-72.18	8.31%
Maz(Zf)HepG2-Maz-ChIP-Seq(GSE31477)Homer	-39.4	16.66%	-12.49	15.80%	-51.51	18.52%	-42.24	26.51%
COUP-TFII(NR)K562-NR2F1-ChIP-Seq(Encode)Homer	-37.09	21.61%	-40.47	19.57%	-98.04	23.39%	-68.76	16.93%
EAR2(NR)K562-NR2F6-ChIP-Seq(Encode)Homer	-35.99	19.87%	-47.2	18.79%	-107.8	22.11%	-84.81	16.19%
COUP-TFII(NR)Arla-Nr2f2-ChIP-Seq(GSE46487)Homer	-34.69	24.57%	-51.07	23.65%	-84.51	25.95%	-78.43	20.51%
Erra(NR)HepG2-Erra-ChIP-Seq(GSE17954)Homer	-34.53	32.58%	-27.42	29.55%	-52.89	32.81%	-33.47	26.08%
TEAD(TEA)Fibroblast-PU.1-ChIP-Seq(Unpublished)Homer	-33.78	12.07%	-91.65	13.75%	-127.9	13.41%	-198.1	11.08%
RAR-RXR(NR)DR5/ES-RAR-ChIP-Seq(GSE56893)Homer	-31.68	5.38%	-35.7	4.56%	-10.9	0.65%	-78.31	3.41%
RAR-RXR(NR)DR5/ES-RAR-ChIP-Seq(GSE56893)Homer	-31.68	5.38%	-35.7	4.56%	-77.3	5.86%	-1.256	0.24%
RAR-RXR(NR)DR5/ES-RAR-ChIP-Seq(GSE56893)Homer	-31.68	5.38%	-35.7	4.56%	-77.3	5.86%	-78.31	3.41%
RAR-RXR(NR)DR5/ES-RAR-ChIP-Seq(GSE56893)Homer	-31.68	5.38%	-0.03208	0.11%	-10.9	0.65%	-78.31	3.41%
RAR-RXR(NR)DR5/ES-RAR-ChIP-Seq(GSE56893)Homer	-31.68	5.38%	-0.03208	0.11%	-77.3	5.86%	-1.256	0.24%
RAR-RXR(NR)DR5/ES-RAR-ChIP-Seq(GSE56893)Homer	-31.68	5.38%	-0.03208	0.11%	-77.3	5.86%	-78.31	3.41%
KLF10(Zf)HEK293-KLF10-GFP-ChIP-Seq(GSE58341)Homer	-26.31	9.54%	-13.92	8.33%	-71.02	11.47%	-40.38	10.59%
GABPA(ETS)Jurkat-GABPa-ChIP-Seq(GSE17954)Homer	-22.96	12.81%	-16.11	11.62%	-14.91	11.47%	-74.59	14.44%
ETS1(ETS)Jurkat-ETS1-ChIP-Seq(GSE17954)Homer	-21.77	15.34%	-21.31	14.06%	-14.87	13.09%	-92.42	15.71%
Flt1(ETS)CD8-Flt1-ChIP-Seq(GSE20898)Homer	-21.54	14.65%	-17.29	14.46%	-16.33	13.76%	-51.53	17.13%
Unknown-ESC-element(?)mES-Nanog-ChIP-Seq(GSE11724)Homer	-20.23	8.33%	-32.87	8.44%	-97.08	11.59%	-51.04	8.07%
Zic(Zf)Cerebellum-ZIC1.2-ChIP-Seq(GSE60731)Homer	-18.67	10.91%	-41.48	11.99%	-74.92	13.81%	-30.7	11.15%
ELF1(ETS)Jurkat-ELF1-ChIP-Seq(SRA014231)Homer	-18.35	6.27%	-16.38	6.72%	-13.57	6.03%	-57.08	10.00%
Nrf2(bZIP)Lymphoblast-Nrf2-ChIP-Seq(GSE37589)Homer	-18.33	1.53%	-39.79	1.61%	-51.47	1.52%	-46.06	0.98%
SUT17(SacCer-Promoters)Homer	-16.15	54.98%	-38.29	62.01%	-46.33	65.20%	-104.4	76.69%
Zic3(Zf)mES-Zic3-ChIP-Seq(GSE37889)Homer	-15.15	6.38%	-43.87	8.55%	-141.4	11.84%	-51.25	9.78%
Tcfcp2l1(CP2)mES-Tcfcp2l1-ChIP-Seq(GSE11431)Homer	-14.95	2.74%	-3.702	1.61%	-59.99	3.84%	-3.121	1.55%
Isl1(Homeobox)Neuron-Isl1-ChIP-Seq(GSE31456)Homer	-14.92	27.73%	-15.47	25.07%	-22.84	24.91%	-57.43	19.55%
SeqBias: A/T bias	-12.68	96.68%	-18.9	91.26%	-14.09	90.85%	-50.5	71.33%
SeqBias: CG bias	-11.54	75.38%	-38.87	80.65%	-28.38	83.40%	-168.4	90.43%
Eik4(ETS)Hela-Eik4-ChIP-Seq(GSE31477)Homer	-9.898	5.54%	-13.66	6.84%	-8.142	5.83%	-84.08	11.73%
SPDEF(ETS)VCaP-SPDEF-ChIP-Seq(SRA014231)Homer	-8.612	13.18%	-15.2	13.67%	-13.78	13.24%	-54.35	13.49%
AP-2gamma(AP2)MCF7-TFAP2C-ChIP-Seq(GSE21234)Homer	-5.453	9.01%	-58.48	15.76%	-33.39	13.41%	-219.2	23.67%
Foxa2(Forkhead)Liver-Foxa2-ChIP-Seq(GSE25694)Homer	-3.96	9.44%	-12.56	10.01%	-31.27	10.62%	-77.14	8.77%
AP-2alpha(AP2)Hela-AP2alpha-ChIP-Seq(GSE31477)Homer	-3.658	6.54%	-55.95	12.55%	-33.87	10.75%	-237.1	19.53%
Foxa3(Forkhead)Liver-Foxa3-ChIP-Seq(GSE77670)Homer	-2.423	3.74%	-8.468	4.22%	-20.28	4.39%	-83.36	4.37%
FOXA1(Forkhead)LNCAP-FOXA1-ChIP-Seq(GSE27824)Homer	-1.994	14.13%	-12.06	15.39%	-33.35	15.56%	-69.95	12.48%
RAR-RXR(NR)DR5/ES-RAR-ChIP-Seq(GSE56893)Homer	-1.78	0.37%	-35.7	4.56%	-10.9	0.65%	-78.31	3.41%
RAR-RXR(NR)DR5/ES-RAR-ChIP-Seq(GSE56893)Homer	-1.78	0.37%	-35.7	4.56%	-77.3	5.86%	-1.256	0.24%
RAR-RXR(NR)DR5/ES-RAR-ChIP-Seq(GSE56893)Homer	-1.78	0.37%	-35.7	4.56%	-77.3	5.86%	-78.31	3.41%
RAR-RXR(NR)DR5/ES-RAR-ChIP-Seq(GSE56893)Homer	-1.78	0.37%	-0.03208	0.11%	-10.9	0.65%	-78.31	3.41%
RAR-RXR(NR)DR5/ES-RAR-ChIP-Seq(GSE56893)Homer	-1.78	0.37%	-0.03208	0.11%	-77.3	5.86%	-1.256	0.24%
RAR-RXR(NR)DR5/ES-RAR-ChIP-Seq(GSE56893)Homer	-1.78	0.37%	-0.03208	0.11%	-77.3	5.86%	-78.31	3.41%
FOXA1(Forkhead)MCF7-FOXA1-ChIP-Seq(GSE26831)Homer	-1.152	11.23%	-9.909	12.74%	-33.46	13.16%	-85.02	11.04%
BORIS(Zf)K562-CTCF-ChIP-Seq(GSE32465)Homer	-0.9167	1.27%	-68.22	4.97%	-18.14	2.72%	-48.42	4.95%
Ronin(THAP)ES-Thap11-ChIP-Seq(GSE51522)Homer	-0.749	0.11%	-0.03609	0.04%	-0.2719	0.07%	-50.67	0.94%
FoxL2(Forkhead)Ovary-FoxL2-ChIP-Seq(GSE60858)Homer	-0.7155	8.75%	-2.98	9.34%	-9.697	9.03%	-54.1	8.34%
HY5(bZIP)colamp-HY5-DAP-Seq(GSE60143)Homer	-0.6456	3.48%	-5.258	5.45%	-5.773	5.31%	-51.27	8.86%
CTCF(Zf)CD4+-CTCF-ChIP-Seq(Barski_et_al.)Homer	-0.09925	0.74%	-89.7	4.48%	-12.82	1.99%	-92.98	3.54%

Appendix table 4

Enrichment (logP) and frequency of known motifs from HOMER in the sequences of the different clusters.

Curriculum Vitae

Contact information

Name	Torgil Elias Friman
E-mail address	elias.friman@gmail.com
Telephone	+41(0)788 320 442
Home address	Place du Tunnel 21, 1005 Lausanne, Switzerland

Personal information

Date of birth	1989-10-14
Nationality	Swedish
Residence	Switzerland (B permit)
Degrees	BSc, MSc
ORCID	0000-0001-9944-6560
Google Scholar ID	GrXDI28AAAAJ

Employment history

Feb 2015 – Oct 2019	Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland Position: PhD student Supervisor: Prof. David M. Suter Topic: Regulation of embryonic stem cell fate Full-time
Oct 2014 – Jan 2015	Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland Position: Research internship Supervisor: Prof. David M. Suter Topic: Regulation of embryonic stem cell fate Full-time
Feb 2014 – Jul 2014	University of Michigan, Ann Arbor, Michigan, USA Position: Master thesis researcher Supervisor: Prof. James C. Bardwell Topic: Molecular chaperones Full-time
May 2013 – Jul 2013	University of Florida, Gainesville, Florida, USA Position: Research internship Supervisor: Prof. Claudio F. Gonzalez Topic: Hydrolytic enzymes Full-time
Dec 2010 – Jan 2014	Forex Bank AB, Stockholm, Sweden Position: Customer service representative Part-time (irregular)

Oct 2008 – Jun 2009	Gustavsberg Health Clinic, Värmdö, Sweden Position: Internship (Statistical analysis and administration) Part-time
---------------------	--

Education

Feb 2015 – Oct 2019	Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland PhD in Biotechnology & Bioengineering PhD thesis: “The role of OCT4 and SOX2 in regulating chromatin accessibility and cell fate across the cell cycle in embryonic stem cells”
Aug 2012 – Dec 2012	University of Texas at Austin, Austin, Texas, USA Exchange studies as part of master program
Aug 2009 – Aug 2014	KTH Royal Institute of Technology, Stockholm, Sweden 5-year program (BSc+MSc) Bachelor of Science (Biotechnology) Master of Science in Engineering (Medical Biotechnology) Master thesis: “Analysis of periplasmic proteins induced in Escherichia coli strains genetically selected for enhanced protein folding” Official graduation date 23/10/2014
Aug 2005 – Jun 2008	Nacka Gymnasium High School, Nacka, Sweden Natural Sciences program with foreign exchange

Personal qualifications

Languages	Swedish; Native English; Fluent French; Intermediate
Computer skills	Bioinformatics, R, UNIX/Bash, Python, cluster computing, Adobe Illustrator, LaTeX, Office suite
Wet lab skills	CRISPR-Cas9 knock-in/out ChIP-seq, ATAC-seq Mammalian cell culture Live-cell imaging Lentiviral vector production Miscellaneous molecular biology techniques (e.g. cloning, qRT-PCR, western blotting, immunofluorescence, flow cytometry, protein purification)
Teaching	300 hours of teaching assistance Research supervisor for two master and one bachelor student

Elected responsibilities	<p>President of the Association des doctorants en sciences de la vie à l'EPFL (ADSV) (May 2018 – May 2019)</p> <p>PhD student representative EDBB program (Oct 2018 – Oct 2019)</p>
Presentations at international conferences	<p><u>Chromatin Meets South</u> Marseille, France. July 2019. Oral presentation</p> <p><u>International Conference on Systems Biology</u> Lyon, France. October 2018. Oral presentation</p> <p><u>European Society of Gene & Cell Therapy</u> Lausanne, Switzerland. October 2018. Poster presentation</p> <p><u>Gordon Research Conference "Chromatin Structure and Function"</u> Newry, Maine, USA. July 2018. Poster presentation</p>
Non-scientific speaking engagements	<p><u>I, Scientist</u> Berlin, Germany. September 2019. Plenary speaker (topic: Power abuse in academia)</p> <p><u>Max Delbrueck Center</u> Berlin, Germany. September, 2019. Invited speaker (topic: Power abuse in academia)</p> <p><u>Department of Astronomy, Stockholm University</u> Stockholm, Sweden. August, 2019. Invited speaker (topic: Power abuse in academia)</p>

Awards

2018	LS2 Travel Grant recipient for Gordon Research Conference "Chromatin Structure and Function"
2014	KTH General Fund scholarship recipient for academic achievement
2013	<p>CHUST scholarship recipient for research at University of Florida</p> <p>Henrik Göran Sandviken Fund scholarship recipient for academic achievement</p> <p>KTH General Fund scholarship recipient for academic achievement</p>

Publications

*Equal contribution

Friman, E.T., Deluz, C., Meireles-Filho, A.C., Govindan, S., Gardeux, V., Deplancke, B. and Suter, D.M. (2019) Dynamic regulation of chromatin accessibility by pluripotency transcription factors across the cell cycle. *bioRxiv*, p.698571 (in revision at *eLife*)

Strebing, D.*, Deluz, C.*, **Friman, E.T.***, Govindan, S., Alber, A.B. and Suter, D.M. (2019) Endogenous fluctuations of OCT4 and SOX2 bias pluripotent cell fate decisions. *Molecular Systems Biology*, 15(9)

Tidin, O., **Friman, E.T.**, Naef, F. and Suter, D.M. (2019) Quantitative relationships between SMAD dynamics and target gene activation kinetics in single live cells. *Scientific Reports*, 9(1), p.5372

Raccaud, M., **Friman, E.T.***, Alber, A.B.*, Agarwal, H., Deluz, C., Kuhn, T., Gebhardt, J.C.M. and Suter, D.M. (2019) Mitotic chromosome binding predicts transcription factor properties in interphase. *Nature Communications*, 10(1), p.487

Deluz, C., Strebinger, D., **Friman, E.T.** and Suter, D.M. (2017) The elusive role of mitotic bookmarking in transcriptional regulation: insights from Sox2. *Cell Cycle*, 16(7), pp.601-606

Deluz, C.*, **Friman, E.T.***, Strebinger, D.*, Benke, A., Raccaud, M., Callegari, A., Leleu, M., Manley, S. and Suter, D.M. (2016) A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. *Genes & Development*, 30(22), pp.2538-2550

Lennon, C.W., Thamsen, M., **Friman, E.T.**, Cacciaglia, A., Sachsenhauser, V., Sorgenfrei, F.A., Wasik, M.A. and Bardwell, J.C., (2015) Folding optimization in vivo uncovers new chaperones. *Journal of Molecular Biology*, 427(18), pp.2983-2994

Professional memberships

Jun 2016 - Present	Member of Life Sciences Switzerland (LS2) society
--------------------	---

Interests

Scientific interests	Gene regulation, transcription, chromatin, developmental biology, epigenetics
Other academic interests	Power abuse in science, scientific fraud and misconduct, open science, effective altruism