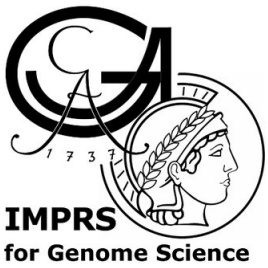
Bioinformatics analysis of multi-omics data elucidates the primary function of Oct4 in gene regulation and pluripotency maintenance

Dissertation

For the award of the degree
"Doctor of Philosophy" (Ph.D.)

Division of Mathematics and Natural Sciences
of the Georg-August-Universität Göttingen

Within the doctoral program



International Max Planck Research School for Genome Science
of the Göttingen Graduate School for Neurosciences, Biophysics, and Molecular Biosciences
(GGNB)

submitted by

Le Xiong

from Wuhan, China

Göttingen 2021

Thesis Advisory Committee

Prof. Dr. Patrick Cramer

Max Planck Institute for Biophysical Chemistry, Department of Molecular Biology, Göttingen, Germany

Prof. Dr. Bernd Wollnik

University Medical Center Göttingen, Institute of Human Genetics, Göttingen, Germany

Prof. Dr. Stefan Klumpp

Georg August University Göttingen, Institute for Nonlinear Dynamics, Göttingen, Germany

Members of the Examination Board

Prof. Dr. Patrick Cramer (1st reviewer)

Max Planck Institute for Biophysical Chemistry, Department of Molecular Biology, Göttingen, Germany

Prof. Dr. Bernd Wollnik (2nd reviewer)

University Medical Center Göttingen, Institute of Human Genetics, Göttingen, Germany

Further members of the Examination Board

Prof. Dr. Stefan Klumpp

Georg August University Göttingen, Institute for Nonlinear Dynamics, Göttingen, Germany

Prof. Dr. Argyris Papantonis

University Medical Center Göttingen, Institute of Pathology, Göttingen, Germany

Dr. Ufuk Günesdogan

Göttingen Center for Molecular Biology, Department of Developmental Biology, Göttingen, Germany

Dr. A. Marieke Oudelaar

Max Planck Institute for Biophysical Chemistry, Research Group Genome Organization and Regulation, Göttingen, Germany

Date of oral examination: December 17, 2021

Acknowledgements

I would like to first express my deepest gratitude to my supervisor Prof. Dr. Patrick Cramer for giving me the opportunity to work in such an outstanding and fantastic scientific environment. I am extremely grateful that Patrick gave me enough freedom to explore and develop my own interests and ideas during my PhD. Patrick's continuous guidance and support have inspired me and helped me overcome all challenges and difficulties throughout my PhD. It has been a fruitful and incredible time to work under his supervision over the last four years.

I would also like to thank Prof. Dr. Hans R. Schöler for the opportunity to collaborate with his laboratory and for his valuable comments and suggestions on my PhD work.

A special thanks to my thesis committee members, Prof. Dr. Bernd Wollnik and Prof. Dr. Stefan Klumpp, for keeping track of my PhD progress and providing valuable comments and suggestions along the way.

I am grateful to Dr. Michael Lidschreiber for his help, discussion and supervision throughout my PhD work. I have learned a lot on how to analyze, interpret and summarize the data in a proper, rigorous and unbiased way from him. A huge thanks to Dr. Michael Lidschreiber for reviewing this thesis. I also want to thank Dr. Björn Schwalb for sharing his code and teaching me on how to analyze TT-seq data at the very beginning of my PhD. It was also a great pleasure to work with Dr. Michael Lidschreiber and Dr. Björn Schwalb for the R and NGS course tutorial.

A huge thanks to Dr. Livia Caizzi, who helped me and performed ChIP-seq experiments for my PhD project. Her data improved the quality of the work and is essential for the completeness of the story. I also want to thank her for her encouragement and it was a great pleasure to work with her. I owe a big thanks to Dr. Jinmi Choi who mentored me when I got started for my PhD. Her guidance and supervision helped me overcome all obstacles in the very beginning of my PhD.

I am also grateful to my collaborator Dr. Erik Tolen for experiments and discussions on data analysis and the results, which are essential and instrumental for the development of the project.

I would also like to thank all the bioinformaticians: Dr. Sara Martins, Eusra Mohammad, Gabriel Villamil and Arjun Devadas for sharing office, mutual help and discussions on data analysis. In addition, I also want to say thanks to system biology lab mates Dr. Kristina Zumer, Kseniia Lysakovskaia, Aiturgan Zheenbekova, Taras Velychko for discussions, support and feedback during my PhD. A special thanks to Dr. Kristina Zumer for reviewing part of this thesis. A further huge thanks to Petra Rus and Kerstin Maier for maintaining of the sequencing facility and Janine Blümel and Almuth Burgdorf for lab running.

I always enjoy to be part of the International Max Planck Research School for Genome Science (IMPRS-GS). A special thanks to the current program coordinators Dr. Henriette Irmer and Frauke Bergmann and the former coordinator Dr. Katja Lidschreiber. I would also like to thank all the friends from this program including Dr. Xin Wang, Dr. Wanwan Ge, Ruoshi Zhang, Shu Zhang and Xiaojuan Li for their friendship.

I would also like to thank Dr. Haibo Wang, who has shared four years of close friendship with me since I joined Cramer lab. It was always a great fun to travel together with him to different places during my PhD. Also, I would like to thank him for sharing with me his wisdom on science and all other topics. Another special thanks to Dr. Peng Liu for friendship and reviewing part of my thesis.

Last but not least, I want to thank my girlfriend Wei Zhu, who constantly supports me with encouragement and love. I am also deeply grateful to my brother, my parents and all other family members for their unconditional love, continuous support and encouragement.

Summary

The transcription factor (TF) Oct4 is essential for maintaining pluripotency *in vitro* as well as *in vivo*. The crucial function of Oct4 in control of pluripotency has been extensively studied during reprogramming of somatic cells to induced pluripotency cells (iPSCs). However, how Oct4 maintains pluripotency in native embryonic stem cells (ESCs) remains far from being fully understood. Moreover, Oct4 cooperates with the TF Sox2 at thousands of genomic sites in ESCs to fulfill its functional role, but the interplay between Oct4 and Sox2 in pluripotency control has not been completely elucidated.

In this study, we aim to investigate the primary function of Oct4 in the control of pluripotency in mouse ESCs. To achieve this goal, we used a doxycycline inducible Oct4 loss-of-function mouse ES cell line. We then conducted a high-resolution time course study to monitor changes in the transcriptome by transient transcriptome sequencing (TT-seq), changes in chromatin accessibility by assay for transposase accessible chromatin with sequencing (ATAC-seq), and changes in Oct4 and Sox2 occupancy by chromatin immunoprecipitation and sequencing (ChIP-seq). We first used the TT-seq and ATAC-seq data to segment the genome into transcription units (TUs) and non-transcribed regions. We then classified TUs into different types including mRNAs and enhancer RNAs (eRNAs). Differential gene expression analysis revealed a rapid downregulation of the components of the pluripotency transcriptional network with super enhancer (SE)-controlled genes being immediately and strongly affected upon Oct4 depletion. We then derived a refined annotation of putative enhancers and identified Oct4-bound transcribed enhancers in mouse ESCs. We found that Oct4 depletion led to immediate loss of Oct4 binding and a rapid loss of eRNA synthesis at Oct4-bound transcribed enhancers, which correlated well with a decrease in mRNA synthesis from nearby putative target genes. In contrast, chromatin accessibility at Oct4-bound enhancers was generally decreased only later or remained unchanged at a subset of enhancers, arguing that Oct4 primarily functions in transcription activation, not chromatin opening. Furthermore, we found that binding of Sox2 at Oct4-bound enhancers remained unchanged during the first 6 hours of Oct4 depletion and it correlated well with the delayed chromatin accessibility changes.

Our data provide new insights into the direct target genes of Oct4. We identified 446 early down-regulated genes upon Oct4 depletion and these genes are strongly enriched for stem cell population maintenance, suggesting that a key pluripotency gene network centered around SEs

is the primary/direct downstream targets of Oct4. Importantly, our results suggest that the key role of Oct4 in pluripotency maintenance in ESCs is to act as a transcriptional activator that stimulates transcription of pluripotency enhancers and their target genes, whereas Sox2 may act as a factor that renders enhancers accessible. In summary, our work illustrates the primary function of Oct4 and its interplay with Sox2 in the controlling the enhancer landscape to regulate pluripotency gene expression in mouse ESCs.

Publications

Part of this work has been published or is in the process of publication:

Oct4 primarily controls enhancer activity rather than accessibility

Le Xiong[*], Erik A. Tolen[*], Jinmi Choi, Livia Caizzi, Kenjiro Adachi, Michael Lidschreiber, Patrick Cramer[#], Hans R. Schöler[#]

([*]) joint first authorship, ([#]) corresponding author

Manuscript in revision, doi: https://doi.org/10.1101/2021.06.28.450119

Contribution: The initial design of the study was by EAT, KA and HRS, with input from JC and PC. LX carried out bioinformatics analyses. EAT carried out experiments except that LC performed Oct4 ChIP-seq and JC performed TT-seq, with input from EAT. ML and JC provided assistance with bioinformatics analyses. PC and HRS supervised research. LX, EAT and PC wrote the manuscript with input from LC, ML and HRS.

Structures and implications of TBP–nucleosome complexes

Haibo Wang, **Le Xiong**, and Patrick Cramer

PNAS July 27, 2021 118 (30) e2108859118, doi: https://doi.org/10.1073/pnas.2108859118

Contribution: HW designed and conducted experiments and data analysis except for bioinformatic analysis, which was performed by LX. PC supervised research. HW and PC wrote the manuscript.

Table of Contents

I. Introduction

1. Transcription factor Oct4

1.1 Oct4 is a member of POU family proteins

Oct4 is a critical developmental regulator and encoded by *Pou5f1* gene on chromosome 17 in the mouse genome (Jerabek et al., 2014). Oct4 is a member of the POU family TFs (Tantin, 2013; Zhao, 2013). The acronym POU is derived from the names of three TFs: Pit1, Oct1/ Oct2, Unc86 (Tantin, 2013; Zhao, 2013). The POU proteins contain a well-conserved DNA-binding domain referred to as the POU domain (Tantin, 2013; Zhao, 2013). Based on sequence similarity of the POU domain, POU proteins are divided into six classes (POU I to POU VI) (Zhao, 2013). Oct4 is a member of the class POU V. Three classes (POU II, III and V) display high affinity for the standard octamer binding motif and thereby are also termed Oct proteins (Zhao, 2013). The mouse *Pou5f1* gene shares over 80% sequence similarity with the human *POU5F1* gene, which is located on chromosome 6 of the human genome (Jerabek et al., 2014).

The POU domain of Oct4 is comprised of a ~80 amino acids N-terminal POU-specific domain (POU_S) and a ~50 amino acids C-terminal POU homeodomain (POU_{HD}) (**Figure 1**) (Tan and Takada, 2020). The two domains are connected by a flexible linker region, which varies in sequences and length among different POU family members (Zhao, 2013). The two domains use classical helix-turn-helix class DNA-binding domain to specifically recognize 4 base pairs (bps) DNA sequence ATGC and AAAT, respectively (Tan and Takada, 2020) (**Figure 1**). Together, Oct4 can bind the consensus DNA sequence 'ATGCAAAT', which is known as the octamer binding motif (**Figure 1**). Besides the POU domain, Oct4 contains two additional transactivation domains in the N and C terminal that exhibit little sequence conservation (Zhao, 2013).

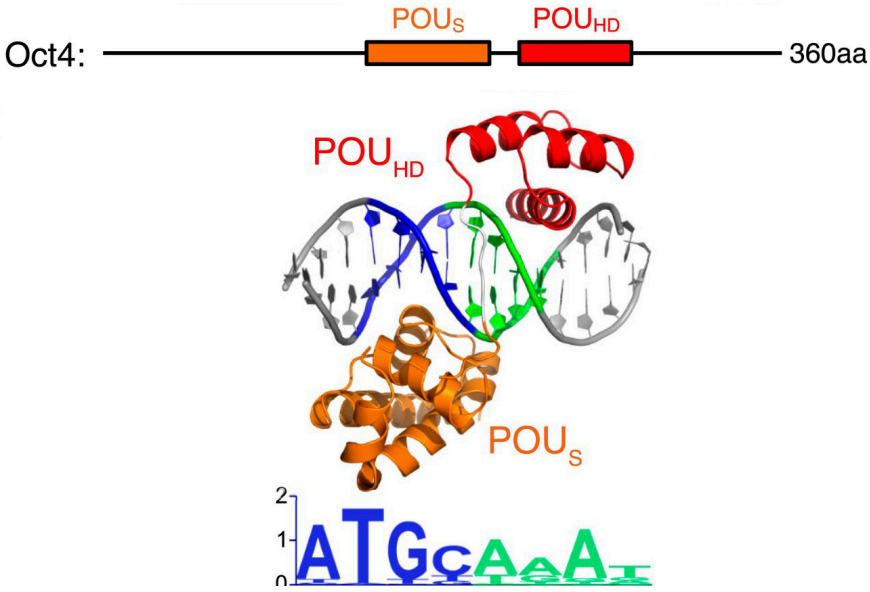


Figure 1. The POU domain of Oct4.

The POU domain of Oct4 contains POU_S (orange) and POU_{HD} (red) domains connected by a flexible linker region (white). The structure of Oct4 POU domain binding to its octamer binding motif (lower). Figure was adapted from (Tan and Takada, 2020).

1.2 Oct4 is essential for both maintenance and induction of pluripotency

The expression of Oct4 during mouse embryo development is summarized in (**Figure 2**) (Jerabek et al., 2014). Oct4 is a maternal factor and similar levels of mRNA have been detected in the zygote and unfertilized oocyte (Jerabek et al., 2014). After fertilization, Oct4 is reduced transiently at the 2 and 4-cell stages (Jerabek et al., 2014). Oct4 expression is increased at the morula stage at 2.5 days post coitum (dpc). At the 3.5-dpc blastocyst, high level of Oct4 expression is maintained in the cells of inner cell mass (ICM), whereas down-regulation of Oct4 is observed in the cells of trophectoderm (Jerabek et al., 2014). After embryo differentiation, the expression of Oct4 is gradually turned off and it is only detected in primordial germ cells in the adult mouse eventually (**Figure 2**).

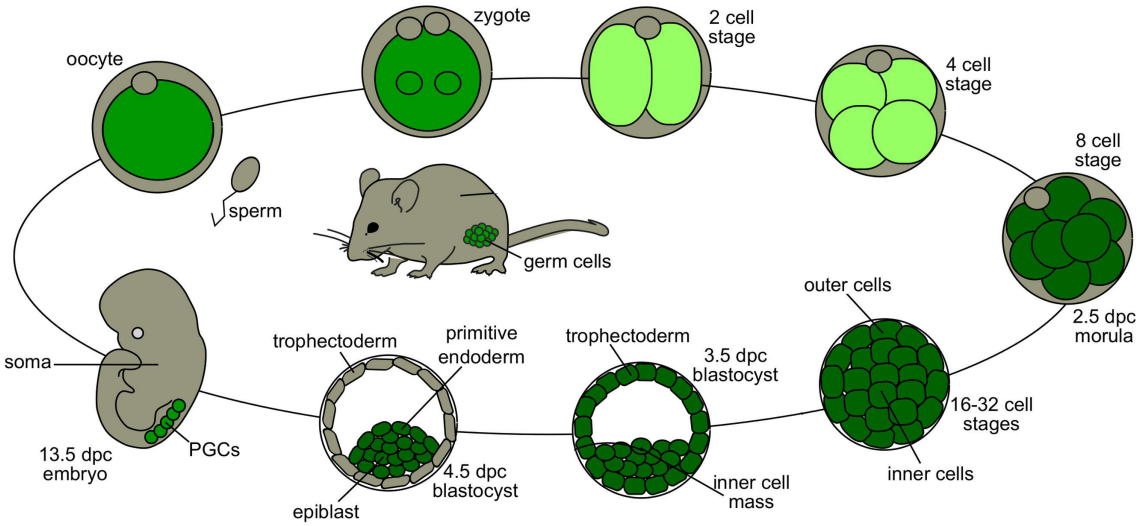


Figure 2. The expression level of Oct4 during mouse life cycle.

The expression of Oct4 during mouse life cycle is summarized by the intensity of green color. Figure was adapted from (Jerabek et al., 2014).

Oct4 is a key factor that controls pluripotency both *in vivo* and *in vitro* (Jerabek et al., 2014). *In vivo*, deletion of the DNA binding and transactivation domains of Oct4 leads to impairment of early mouse embryos development (Nichols et al., 1998). *In vitro,* the functions of Oct4 have been extensively studied in embryonic stem cells (ESCs). ESCs are the cells cloned from the ICM of embryo and are capable of self-renewal and generating all the three germ layers including endoderm, mesoderm and ectoderm. Using a doxycycline inducible Oct4 expression system, early study showed that up-regulation of Oct4 led to mouse ESCs differentiating into primitive mesodermal and endodermal lineage, whereas mouse ESCs were differentiated into trophectoderm lineage when Oct4 was dramatically reduced (Niwa et al., 2000). This suggests that the expression of Oct4 is tightly controlled and dysregulation of Oct4 leads to loss of ESCs pluripotency (Niwa et al., 2000).

The crucial function of Oct4 for pluripotency has also been demonstrated by the discovery that somatic cells can be reprogrammed into induced pluripotent cells (iPSCs) using a combination of TFs including Oct4, Sox2, Klf4 and c-Myc (Yamanaka factors) (Takahashi and Yamanaka, 2006). After Yamanaka's pioneer work, extensive studies have been focused on understanding the individual role of the four factors by replacing them with other factors (Feng et al., 2009; Kim et al., 2009a; Kim et al., 2009b; Nakagawa et al., 2008; Sterneckert et al., 2012; Yu et al., 2007). In comparison to Sox2 and Klf4, which can be replaced by other members of their families during iPSCs reprogramming (Nakagawa et al., 2008), Oct4 is the only factor that

could not be replaced by any other members of its family. Thus, Oct4 has been considered to be unique during iPSCs reprogramming.

2. Molecular mechanisms by which Oct4 controls pluripotency

2.1 Oct4 is part of the core ESCs transcriptional regulatory network

The essential function of Oct4 for pluripotency maintenance in ESCs has been attributed to its pivotal role as part of the core of pluripotency transcriptional regulatory network (Boyer et al., 2005; Chen et al., 2008; Kim et al., 2008; Loh et al., 2006; Young, 2011). Oct4 recognizes enhancers (**see 3.1**), which are transcriptional regulatory element in the genome that controls cell-type specific gene expression (Banerji et al., 1981). Enhancers often contain multiple TF binding sites and are recognized by a cluster of TFs (Spitz and Furlong, 2012). Initial study using high throughput chromatin immunoprecipitation with DNA microarrays (ChIP-Chip) has shown that ~90% genes occupied by Oct4 are co-targeted by TFs Sox2 and Nanog, suggesting a combinatorial regulation of gene expression in ESCs (Boyer et al., 2005). Strikingly, it has been found that Oct4, Sox2 and Nanog regulate themselves and other well-known pluripotency genes (Boyer et al., 2005; Loh et al., 2006). This suggests that some of the key pluripotency regulators cooperate together to control their own expression, indicating the existence of an interconnected transcriptional regulatory loop/network in ESCs (Young, 2011).

A later study combining ChIP with massively parallel DNA sequencing (ChIP-seq) to map thirteen sequence-specific TFs found that Smad1 and Stat3 are also co-localized with Oct4, Sox2 and Nanog at a majority of their occupied locus (Chen et al., 2008). Smad1 is a key factor associated with bone morphogenic protein (BMP) signaling pathways and Stat3 is a master regulator involved in leukemia inhibitory factor (LIF) pathways (Do et al., 2013; Hall et al., 2009; Niwa et al., 2009; Ying et al., 2003). This suggests that the core transcriptional network in ESCs maintained by Oct4, Sox2 and Nanog is further integrated with external signaling pathways (Chen et al., 2008). ESCs can differentiate into a variety of cell types based on distinct differentiation-inducing signals. Thus, the co-occupancy of Smad1 and Stat3 with Oct4, Sox2 and Nanog provides the molecular mechanisms by which different signaling pathways are incorporated into the core transcriptional regulatory network to control pluripotency state of ESCs (Chen et al., 2008) (**Figure 3**).

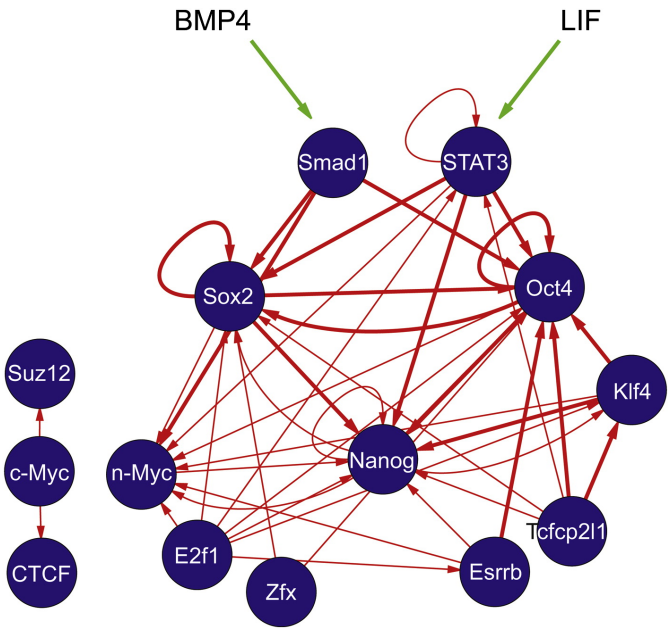


Figure 3. Transcriptional regulatory network in mouse ESCs.

The figure was adapted from (Chen et al., 2008), where they constructed a transcriptional regulatory network by combining thirteen TFs ChIP-seq binding profiles with gene expression data during mouse differentiation.

Last but not least, some of microRNAs that are highly expressed in ESCs were found to be co-occupied by Oct4, Sox2 and Nanog (Marson et al., 2008; Xu et al., 2009). These microRNAs are down-regulated upon ESCs differentiation, suggesting an important role of non-coding RNAs in controlling pluripotency (Marson et al., 2008). Thus, the regulatory non-coding RNAs have also been integrated into the pluripotency transcriptional network. Transcriptional regulatory network such as feed-forward loop has been shown to provide fast and rapid response upon internal or external signal changes or perturbations (Alon, 2007). Taken together, these evidence lead to the idea that existence of transcriptional regulatory network controlled by key pluripotency regulators, signaling factors and some of the regulatory non-coding RNAs is the basis of the ESCs state control and Oct4 is part of the core of this network (Young, 2011).

2.2 Oct4 controls super enhancers to maintain cell identity of ESCs

Oct4 function in maintaining of ESCs pluripotency has also been attributed to its role in the establishment of super-enhancers (SEs), which show high occupancy with TFs and coactivators (Whyte et al., 2013). The genes located close to Oct4 occupied SEs have been implicated in control of pluripotent state of ESCs (Whyte et al., 2013). Depletion of Oct4 by short hairpin

RNA at 3, 4 and 5 days leads to down-regulation of the genes that are preferentially regulated by SEs (Whyte et al., 2013).

Whether SEs represent a new type of regulatory entity remains controversial (Blobel et al., 2021; Pott and Lieb, 2015). A liquid-liquid phase separation (LLPS) model has been proposed to underlie the formation of SEs (Hnisz et al., 2017). Transcription related proteins such as TFs and coactivators often contain intrinsically disordered regions (IDRs) and can form multivalent interaction with each other (Boehning et al., 2018; Cho et al., 2018; Chong et al., 2018; Sabari et al., 2018). *In vitro*, mixture of these proteins with high concentration leads to the formation of phase-separated droplets (Boehning et al., 2018; Cho et al., 2018; Chong et al., 2018; Martin and Mittag, 2018; Sabari et al., 2018).

The phase separation capability of Oct4 has been investigated recently and Oct4 alone does not form droplets *in vitro* (Boija et al., 2018). However, when Oct4 is mixed with IDR of Med1, a subunit of Mediator coactivator, droplets are formed to a near-micrometer sized and show fast recovery after photobleaching (FRAP) (Boija et al., 2018). Specifically, the acidic residues in the two transactivation domains of Oct4 have been identified to be crucial for the formation of Oct4-Med1-IDR droplets *in vitro* (Boija et al., 2018). Mutation of these acidic residues *in vivo* leads to attenuated transcriptional activation function of Oct4 (Boija et al., 2018). In summary, these results support the notion that Oct4 maintains ESCs pluripotency by controlling SEs and a phase separation process may underline the formation of SEs.

2.3 Oct4 can act as a pioneer factor

In eukaryotes, DNA is compacted and wrapped by histone octamers to form nucleosomes, which consist of the basic structural unit of chromosome (Luger et al., 1997). DNA within nucleosome is generally inaccessible to most TFs. However, a group of TFs called pioneer factors are capable of directly interacting with nucleosomal DNA and thereby create an open chromatin environment to initiate subsequent cell fate transition (Zaret, 2020).

Several lines of evidence support that Oct4 can act as a pioneer factor during somatic cell reprogramming to iPSCs (Chen et al., 2016; Chronis et al., 2017; Li et al., 2017; Soufi et al., 2012; Soufi et al., 2015; Velychko et al., 2019). Except for c-Myc, all other three Yamanaka factors Oct4, Sox2 and Klf4 have been found to be capable of interacting with nucleosome *in*

vivo and *in vitro* (Soufi et al., 2012; Soufi et al., 2015). Binding of c-Myc with nucleosome is enabled by Oct4, Sox2 and Klf4. Mechanistically, Oct4, Sox2 and Klf4 bind nucleosome by recognizing partial binding motifs (Soufi et al., 2015). For Oct4, 42% of Oct4 targeted nucleosome-enriched sites contain ATGC motif that is recognized by POUS domain and 28% of them contain AAAT motif that is recognized by POUHD domain (Soufi et al., 2015).

The binding preference of Oct4 on the nucleosome has been assessed recently by combing EMSA with next generation sequencing (Michael et al., 2020). It has been shown that Oct4 prefers to interact with the entry-exits sites of nucleosome, while Sox2 appears to be less specific and prefers to target the entry-exits sites and the sites near the nucleosome dyad (Michael et al., 2020). High resolution structure of Oct4-Sox2-nucleosome complex has been solved (Michael et al., 2020). Oct4 and Sox2 interact with the nucleosome at super helix location (SHL) -6 and releases the entry-exit DNA away from the histone core (**Figure 4**) (Michael et al., 2020). Oct4 recognizes a partial DNA motif using its POUS domain (**Figure 4**). In summary, Oct4 uses one of its DNA binding domains to target a partial DNA motif on the nucleosome and thereby acts as a pioneer factor during reprogramming.

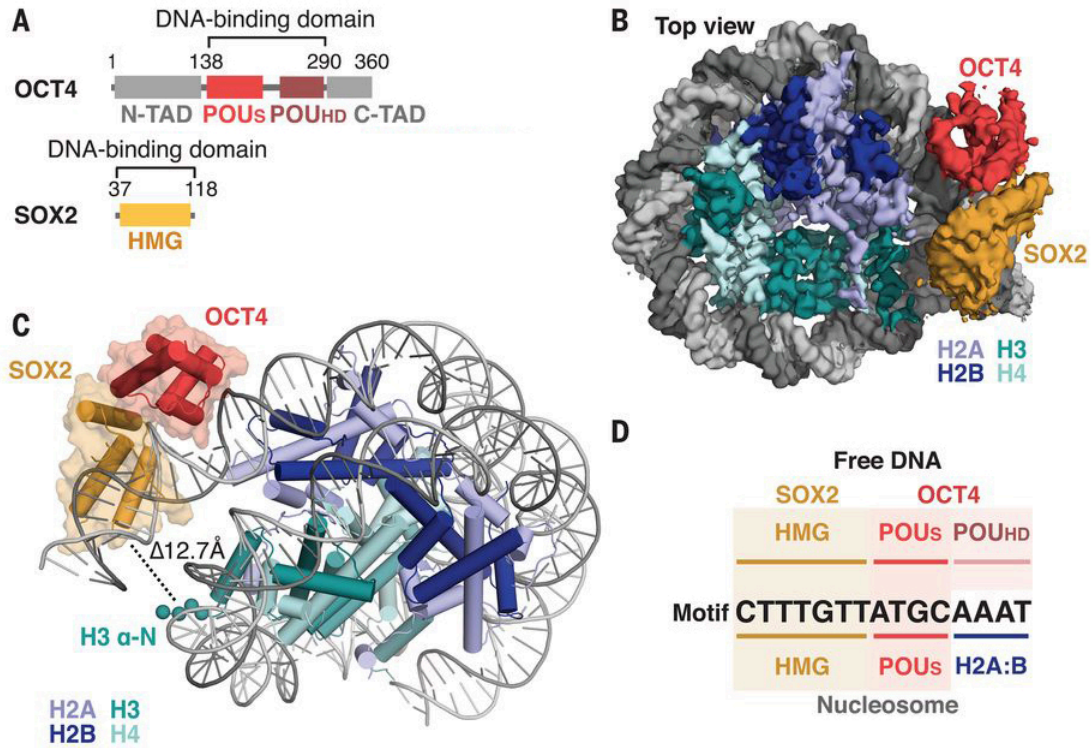


Figure 4. Cryo-EM structure of Oct4-Sox2-nucleosome (SHL-6) complex.

(A) Domain representation of Oct4 and Sox2. (B) and (C) Structures of Oct4 and Sox2 interacting with nucleosome at SHL-6. (D) Oct4 uses its POUS domain to recognize a partial DNA motif on the nucleosome. Figure was adapted from (Michael et al., 2020).

Whether Oct4 also acts as a pioneer factor or not in native ESCs environment, where the chromatin is already opened and accessible, has been studied recently (King and Klose, 2017). In mouse ESCs, depletion of Oct4 for 24 hours leads to a decrease of chromatin accessibility at majority of its occupied sites, accompanied by loss of Sox2 and Nanog binding (King and Klose, 2017). This suggests that Oct4 may act as a pioneer factor to support other factors binding in ESCs. Moreover, by analyzing public ChIP-seq data of the well-known chromatin remodeling complexes, Brg1 (also named Smarca4) has been identified to be correlated with Oct4 dependent accessibility (King and Klose, 2017). Brg1 belongs to SWI/SNF family chromatin remodeler which contains a catalytic ATPase subunit and is involved in chromatin remodeling (de la Serna et al., 2006). Taken together, Oct4 requires the chromatin remodeler Brg1 to facilitate its function as a putative pioneer factor in mouse ESCs (King and Klose, 2017).

2.4 Oct4 and Sox2 cooperation

The proper function of Oct4 in pluripotency maintenance requires its direct cooperation with Sox2, which uses its high mobility domain (HMG) to recognize a 7-bp DNA with a consensus sequence CTTTGTT (Ambrosetti et al., 1997; Yuan et al., 1995). Oct4 co-occupies with Sox2 at thousands of genomic sites in mouse ESCs (Chen et al., 2008; Loh et al., 2006). Oct4 and Sox2 bind enhancers adjacently to a well-defined composite DNA motif (CATTGTTATGCAAAT), which is consisted of a Sox2 binding motif, followed immediately by an Oct4 octamer motif (Chen et al., 2008; Loh et al., 2006). This motif was found to be enriched in well-known enhancers of pluripotency genes such as Klf4 and Nanog (Aksoy et al., 2013; Rodda et al., 2005). Structurally, Oct4 can interact with Sox2 directly by protein-protein interaction (Merino et al., 2014). Mutation of this interface leads to impairment of stem cell pluripotency, demonstrating that the partnership between Oct4 and Sox2 is essential for the maintenance of pluripotency (Merino et al., 2014). Biophysical study reveals that the interaction between Oct4 and Sox2 is mediated by DNA, suggesting that the interaction does not exist prior to DNA binding (Lam et al., 2012).

The cooperativity between Oct4 and Sox2 has been assessed *in vitro* (Chang et al., 2017). Strikingly, among all tested spacer distances, only the adjacent composite motif (0 spacer) conveys a high cooperativity for Oct4 and Sox2 (Chang et al., 2017). The cooperativity between Oct4 and Sox2 has also been assessed *in vivo* in live mouse ESCs by single molecular imaging (Chen et al., 2014). Both Oct4 and Sox2 recognize their target DNA in a trial and error sampling

fashion consisting of short time 3D diffusion and nonspecific collisions, followed by specific DNA interaction with much longer time duration (Chen et al., 2014). Their study supports a model that Oct4/Sox2/DNA ternary complex is assembled in a hierarchical manner *in vivo* with Sox2 first engaging the DNA and assisting Oct4 binding (Chen et al., 2014).

The cooperativity between Oct4 and Sox2 has also been studied during reprogramming (Li et al., 2017; Malik et al., 2019). Although Oct4 and Sox2 alone can interact with nucleosome, neither of them can create open chromatin in the same way as both factors together (Li et al., 2017). Individually, Sox2 alone can open almost twice as many regions as Oct4 alone, and most of the regions opened by Oct4 could also be opened by Sox2 (Li et al., 2017). Moreover, the regions targeted by Sox2 are more strongly enriched with nucleosomes in comparison to Oct4 (Soufi et al., 2015). Thus, compared to Oct4, Sox2 acts as a much stronger pioneer factor during reprogramming (Li et al., 2017; Malik et al., 2019).

3. Enhancers in transcription regulation

3.1 Transcriptional enhancers and its identification

Enhancers were initially described as short DNA fragments that can positively regulate gene expression in a distance and orientation independent manner (Banerji et al., 1981; Benoist and Chambon, 1981). Enhancers facilitate gene transcription by recruiting lineage-specific TFs and general coactivators including Mediator to stabilize pre-initiation complex (PIC) assembly at promoters (Cramer, 2019; Spitz and Furlong, 2012). In genomic era, enhancers can be identified in a systematical way by making use of the well-known hallmarks of enhancers, which include the detection of DNase hypersensitivity or chromatin opening region, a high level of histone H3 lysine 4 monomethylation (H3K4me1) and histone H3 lysine 27 acetylation (H3K27ac), the binding of multiple TFs, the occupancy of coactivators and acetyltransferases such as p300 (Shlyueva et al., 2014; Visel et al., 2009). This leads to the identification of ~1 million putative enhancers in human genome in the encyclopedia of DNA elements (ENCODE) project (2012; Rivera and Ren, 2013).

Enhancers can be transcribed and give rise to non-coding RNAs (ncRNAs) called enhancer RNAs (eRNAs) (De Santa et al., 2010; Kim et al., 2010; Tuan et al., 1992). Consequently, eRNAs can be used to identify enhancers (Andersson et al., 2014a; De Santa et al., 2010;

Schwalb et al., 2016). Enhancers with clear eRNA transcription exhibit higher occupancy of TFs, co-activators and more enriched for active histone modifications (Hah et al., 2013; Kim et al., 2010; Melgar et al., 2011; Zhu et al., 2013). Importantly, by using large-scale reporter assays, it has been shown that transcribed enhancers often show several folds stronger reporter activity than non-transcribed enhancers that are identified purely by histone modifications (Andersson et al., 2014a). Thus, compared with the aforementioned chromatin marks, eRNA is a better and more reliable feature for identification of active enhancers (Li et al., 2016). Consequently, identification of active enhancers by mapping eRNAs have been conducted in various cell types during distinct biologically processes (Andersson et al., 2014a; Choi et al., 2021; De Santa et al., 2010; Hah et al., 2013; Henriques et al., 2018; Kim et al., 2010; Lidschreiber et al., 2021; Melgar et al., 2011; Michel et al., 2017).

3.2 Regulation of enhancer transcription

The regulation of enhancer transcription has been extensively studied (Andersson et al., 2014b; Fischer et al., 1985; Hah et al., 2013; Koch et al., 2011; Li et al., 2016). Several lines of evidence support that transcription initiation at enhancers occurs in a way similar to promoters (Andersson and Sandelin, 2020). Recruitment of TATA-box binding protein (TBP) and general TFs happens similarly at both enhancers and promoters (Koch et al., 2011). eRNAs are capped and transcription of enhancers gives rise to bidirectional transcripts, which is also observed at promoters (Andersson et al., 2014a; Andersson et al., 2014b; Core et al., 2014; Melgar et al., 2011). Enhancer and promoter transcription differs after initiation. Different from mRNAs, eRNAs are generally un-spliced and this may due to the lack of U1 splice sites at enhancer regions (Andersson et al., 2014a; Andersson et al., 2014b; Core et al., 2014). Transcription elongation of RNA polymerase II (Pol II) at coding regions is often correlated with phosphorylation of the C terminal domain (CTD) of Pol II at serine 2 (S2) and an enrichment of histone H3 lysine 36 trimethylation (H3K36me3) at gene bodies, which rarely happens during elongation of enhancer transcription (Bonn et al., 2012; Koch et al., 2011). Different from mRNAs that are relatively stable, eRNAs are unstable and generally lower expressed, which may due to its fast decay and degradation by exosome during transcription elongation (Lubas et al., 2015; Pefanis et al., 2015).

3.3 The function of enhancer transcription

The function of enhancer transcription remains poorly understood (Lewis et al., 2019; Li et al., 2016; Sartorelli and Lauberth, 2020). Transcription of enhancer may simply represent transcriptional noise from Pol II random searching of highly accessible regions (Struhl, 2007). However, evidence suggests that the process of enhancer transcription may play a role in gene regulation (Ho et al., 2006; Kaikkonen et al., 2013; Ling et al., 2004). It comes from the fact that the CTD of Pol II can recruit more than one hundred proteins with various functions during transcription elongation (Bentley, 2014). These proteins can influence chromatin organization when travelling together with transcribing Pol II. For example, it has been shown that in microphage, inhibition of enhancer transcription leads to a decrease of recruitment of histone methyltransferases Mlls and the reduction of H3K4me1 deposition at enhancers, and this effect is independent of the corresponding eRNAs (Kaikkonen et al., 2013). This suggests a role of enhancer transcription in chromatin remodeling.

Alternatively, eRNAs may play a direct role in gene regulation (Bose et al., 2017; Li et al., 2013; Mousavi et al., 2013; Schaukowitch et al., 2014; Sigova et al., 2015). In mouse muscle cells, depletion of the eRNAs originated from two enhancer regions targeted by TFs MyoD and MyoG results in a decrease of chromatin accessibility at Myod1 promoter (Mousavi et al., 2013). This suggests a role of eRNAs in control of local chromatin organization (Mousavi et al., 2013). eRNAs have also been found to be involved in control of histone acetylation during transcription (Bose et al., 2017). A study has shown that the histone acetylation enzyme CBP/p300 can directly interact with eRNA *in vitro* and *in vivo* using its histone acetyltransferase (HAT) domain (Bose et al., 2017). Binding of eRNA with HAT domain leads to increased activity of CBP/p300 and the subsequent deposition of H3K27ac at enhancers (Bose et al., 2017). These studies support an active role of eRNAs in control of chromatin remodeling and local chromatin organization to regulate gene expression (Sartorelli and Lauberth, 2020).

eRNAs may be directly involved in enhancer-promoter interactions (Hsieh et al., 2014; Li et al., 2013; Pnueli et al., 2015). Depletion of specific eRNAs in human cancer cells leads to decreased enhancer-promoter interactions (Li et al., 2013; Pnueli et al., 2015). The eRNAs were found to interact with potential looping factors such as cohesin and Mediator (Hsieh et al., 2014; Li et al., 2013). However, studies in other cellular systems have shown no profound changes of

enhancer-promoter interactions after depletion of specific eRNAs (Hah et al., 2013; Schaukowitch et al., 2014). This suggests that the function of eRNAs relies on its specific biological context. eRNAs can also interact with TF directly like Yin Yang 1(YY1) to increase its occupancy at enhancers (Sigova et al., 2015). This may also contribute to the stabilization of enhancer-promoter looping because YY1 is found to be involved in bridging enhancer-promoter loop (Weintraub et al., 2017). In addition to transcription initiation regulation, eRNAs can also regulate Pol II pause release by interacting with the negative elongation factor (NELF) (Schaukowitch et al., 2014). In most cases, eRNAs are thought to play a role in *cis* because eRNAs are short-lived and unstable and mostly located in chromatin-bound fractions. However, some studies have demonstrated that eRNAs can play a role in *trans* and can even relocate to other chromatin regions (Hsieh et al., 2014; Mousavi et al., 2013; Ørom et al., 2010). In the latter situation, the eRNAs are found to be polyadenylated, which may lead to the increased eRNA stability and thereby acts in *trans* (Hsieh et al., 2014; Mousavi et al., 2013; Ørom et al., 2010). This suggests that the function of eRNAs is also related to its post-transcriptional regulation.

Bases on their functions, eRNAs can be generally divided into three classes (Li et al., 2016) (**Figure 5**). For the first class, the transcribed eRNAs are mostly non-functional and may purely represent the transcriptional noise. For the second class, the transcription process of enhancers plays a role in control of gene transcription. For the last group, the eRNAs per se contribute to gene transcription through interaction with different regulators.

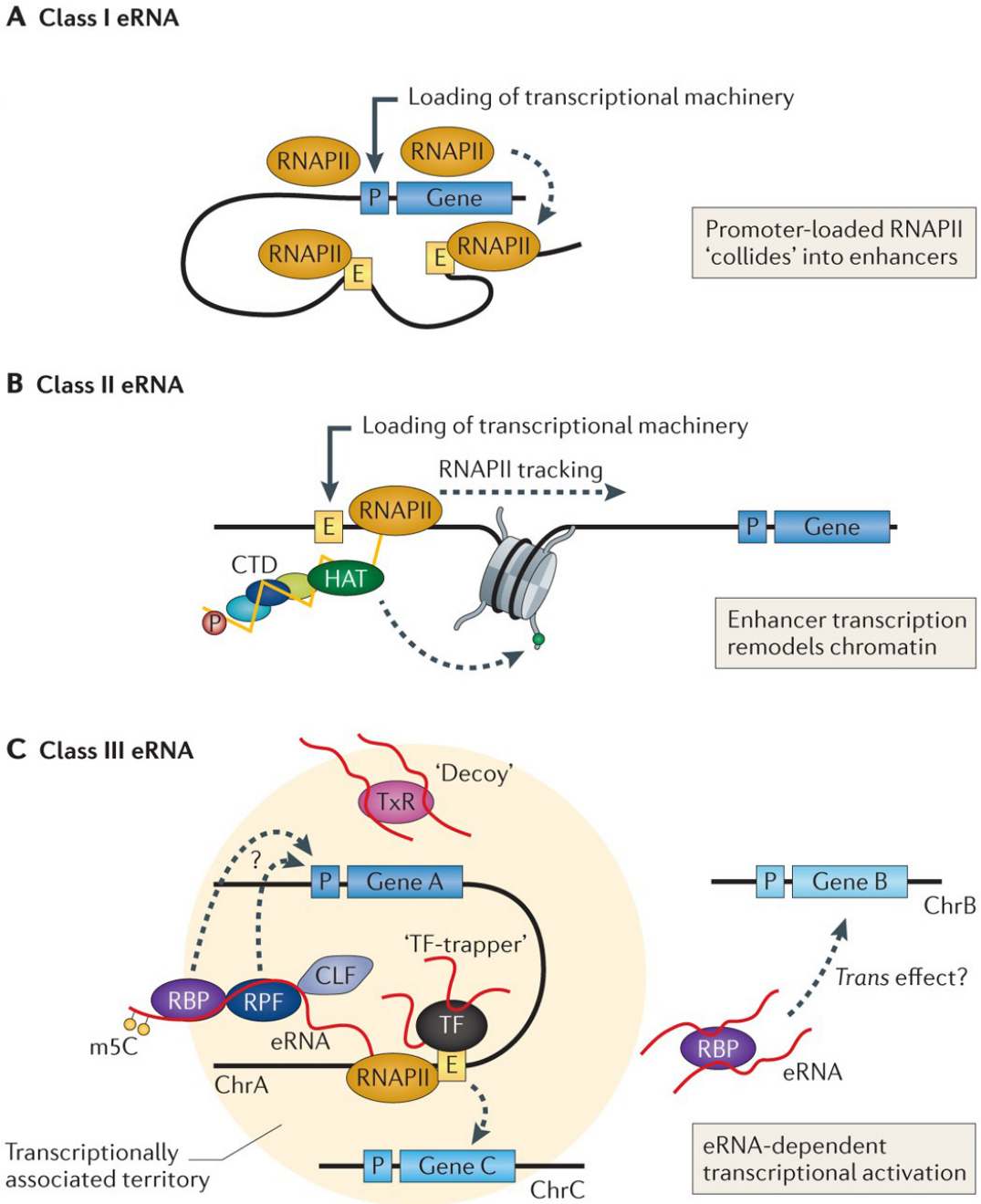


Figure 5. Three functional classes of eRNAs.

(A) Class I eRNA purely represents transcriptional noise of enhancers because of its proximity to high concentration of active transcriptional machinery. (B) For Class II eRNA, the transcription process of enhancers plays a role in gene regulation. Specifically, the CTD of transcribing Pol II can interact with different proteins to remodel local chromatin structure such as histone acetylation. (C) For Class III eRNA, the eRNA per se contributes to gene regulation by interacting with various regulators including chromosome loop factors (CLF), TF, transcriptional repressors (TxRs), RNA processing factors (RPFs) and so on, in a *cis* or *trans* manner, on the basis of its eRNA stability. Figure was adapted from (Li et al., 2016).

3.4 Enhancer-promoter interactions and chromatin organization

Enhancer-promoter communication is thought to happen within topological associating domains (TADs), which are self-interacting genomic regions that are identified by chromosome conformation capture assays including Hi-C (Andrey et al., 2013; Furlong and Levine, 2018; Hnisz et al., 2016; Oudelaar and Higgs, 2021; Smallwood and Ren, 2013). The DNA sequences within TADs show much more frequent interaction than its interaction with DNA sequences outside TADs (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). The boundaries of TADs are enriched for CCCTC-binding factor (CTCF), cohesin and actively transcribed genes including housekeeping genes and transfer RNA genes (Dixon et al., 2012). A loop extrusion model that is facilitated by CTCF and cohesin has been proposed to underline the formation of TADs (Fudenberg et al., 2016; Sanborn et al., 2015). This includes loading of cohesin complexes that can form a ring around chromatin and extrude DNA actively through ATP hydrolysis until it encounters a barrier element, which is thought to be a convergent CTCF proteins at the anchors of the loop (Fudenberg et al., 2016; Sanborn et al., 2015). The cohesin loop extrusion model has been strongly supported by recent studies showing that human cohesin complexes can drive formation of DNA loops *in vitro* (Davidson et al., 2019; Golfier et al., 2020; Kim et al., 2019). The loop extrusion model is further supported by depletion of cohesin *in vivo*, which leads to complete loss of TADs (Rao et al., 2017; Wutz et al., 2017). In contrast, depletion of cohesin release factor Wapl results in longer time dwell of cohesin and thereby larger TADs (Haarhuis et al., 2017).

Currently, whether TADs and chromatin organization indeed play an instructive role in gene expression or not remains controversial. Locus-specific experiments have demonstrated that disruption of TADs boundaries leads to inappropriate enhancer-promoter interactions and mis-regulation of gene expression. For instance, depletion of CTCF binding sites at TAD boundaries of sonic hedgehog gene and a-globin gene loci leads to impairment of enhancer-promoter interactions, accompanied by changes in gene expression (Hanssen et al., 2017; Paliou et al., 2019). Moreover, mutations that disrupt boundaries of TADs have been reported to cause certain cancers as well as developmental disorders including limb malformations (Corces and Corces, 2016; Spielmann et al., 2018). These studies provide strong evidence that TADs and chromatin organization play an instructive role in gene regulation by controlling proper enhancer-promoter interactions. However, recent genome-wide studies have demonstrated that depletion of cohesin only leads to modest effect on gene transcription, although TADs are completed eliminated (Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017).

Enhancers are thought to form a loop with promoters through directly interaction (Kagey et al., 2010). However, it remains unclear what kinds of force underlines the formation of enhancer-promoter loop (Schoenfelder and Fraser, 2019). Some studies support that TFs play a direct role in enhancer-promoter looping. Lineage specific TFs, such as the pluripotency factor Klf4, hematopoietic TFs Ldb1, Klf1 and GATA1, have been reported to be directly involved in enhancer-promoter loops (Deng et al., 2012; Di Giammartino et al., 2019; Song et al., 2007; Vakoc et al., 2005). Whether enhancer-promoter looping is the cause of gene expression has also been studied. For example, artificial tethering of TF Ldb1 to beta-globin promoter by engineered zinc-finger TF leads to forced enhancer-promoter looping and activation of beta-globin genes in immature murine erythroid cells (Deng et al., 2014). This suggests that TFs function in enhancer-promoter looping is the cause of gene activation (Deng et al., 2014). However, it has also been reported that establishment of enhancer-promoter interactions could precede activation of the corresponding gene transcription (Ghavi-Helm et al., 2014). In summary, these results delineate a general picture that enhancer-promoter looping and the activation of transcription could happen simultaneously or in a dynamic hierarchical manner.

In addition to TFs, the coactivator Mediator has also been suggested to play a direct role in enhancer-promoter interactions (Kagey et al., 2010). However, recent two studies have shown that depletion of Mediator leads to modest influences on enhancer-promoter interactions, suggesting that Mediator is not essential for enhancer-promoter interactions (El Khattabi et al., 2019; Jaeger et al., 2020). Whether loop extrusion by cohesin also plays a direct role for enhancer-promoter interactions or not still remains unclear. A recent study has demonstrated that depletion of cohesin leads to only modest reduction of long-range enhancer-promoter interactions, indicating that cohesin associated loop extrusion is not essential for the maintenance of enhancer-promoter interactions (Aljahani et al., 2021).

4. Motivation and aim of this thesis

The crucial function of Oct4 in control of pluripotency has been extensively studied during iPSC reprogramming (Chen et al., 2016; Chronis et al., 2017; Li et al., 2017; Soufi et al., 2012; Soufi et al., 2015; Velychko et al., 2019). However, how Oct4 maintains pluripotency in native ESCs remains far from being fully understood. Moreover, Oct4 cooperates with the TF Sox2 at thousands of genomic sites in ESCs to fulfill its functional role, but the interplay between Oct4 and Sox2 in pluripotency control has not been completely elucidated.

In this thesis, we aim to study the primary function of Oct4 in gene regulation and maintenance of pluripotency. To achieve this goal, a rapid and earlier depletion of Oct4 is necessary. To understand how Oct4 controls enhancer landscape to maintain pluripotency, enhancer activity and accessibility need to be monitored simultaneously. To monitor enhancer activity, transcriptional enhancers need to be identified and their activity by eRNA synthesis should be examined. To understand how changes in enhancer activity and accessibility are related to gene transcription, changes in mRNA synthesis also need to be monitored simultaneously. Taken together, a rapid depletion of Oct4 and simultaneous monitoring of changes in enhancer activity by eRNA, chromatin accessibility and downstream mRNA synthesis changes will allow a better understanding of the primary function of Oct4 in gene regulation. We have developed TT-seq, a method that captures newly synthesized RNA with high sensitivity (Schwalb et al., 2016). TT-seq combines a short pulse of 4-thiouridine (4sU) labeling with RNA fragmentation and monitors transcription changes at both enhancers and promoters genome-wide. TT-seq can quantify changes in enhancer transcription during cellular processes such as T-cell stimulation, the heat shock response and transdifferentiation, and is thus ideally to monitor immediate transcriptome changes after perturbation (Choi et al., 2021; Gressel et al., 2019; Michel et al., 2017).

In this work, to investigate the primary function of Oct4 in the control of pluripotency, we used a doxycycline (DOX) inducible Oct4 loss-of-function mouse ESCs that allows for rapid depletion of Oct4 (Niwa et al., 2000). This system was previously used to study the effect of Oct4 depletion after 24 hours of DOX treatment (King and Klose, 2017). In order to investigate the direct, primary role of Oct4, we conducted a time course experiment collecting samples after 0, 3, 6, 9, 12 and 15 hours of DOX treatment. We then monitor changes in the transcriptome by TT-seq, changes in chromatin accessibility by ATAC-seq, and changes in

Oct4 and Sox2 occupancy by ChIP-seq. Combing these data allows us to dissect the order of events that occur at enhancers and promoters upon rapid Oct4 depletion and draws a conclusion on the primary function of Oct4 in control of gene expression and pluripotency.

II. Methods and Materials

The methods presented in this chapter have been published:
Le Xiong*, Erik A. Tolen* et al. Oct4 primarily controls enhancer activity rather than accessibility (* joint first authorship)
Manuscript in revision, doi: https://doi.org/10.1101/2021.06.28.450119

A detailed list of items from the published manuscript can be found in Appendix (page 104).

1. Experimental methods

The following experiments were conducted by Dr. Erik Tolen (Department of Cell and Developmental Biology, MPI for Molecular Biomedicine).

1.1 Employed cell lines and doxycycline treatment

Mouse embryonic stem cells (mESCs) harbouring a doxycycline repressible Pou5f1 transgene (ZHBTc4 cells, **Figure 6**) (Niwa et al., 2000) were propagated on gelatin-coated plates in equal parts DMEM-F12 (Life Technology, 21331-020) and Neuralbasal (Life Technology, 21103-049) Medium supplemented with 2% Fetal Bovin Serum (Sigma Aldrich, G1393-100 ml), 2% Knockout Serum Replacement Medium (Gibco, 10828-028), 0.04 µg/ml Leukemia Inhibitory Factor, Pennicilin/Streptomycin (Sigma Aldrich, P4333-100ml), 0.1 mM-Mercaptoethanol (Gibco, 31350-010), 0.5x B27 supplement (Life Technology, 12587-010), 0.5x N2 supplement (Gibco, AM9759, 3 µM CHIR99021 (Cayman Chemicals, 13122), 1 µM PD0325901 (Biomol, 103034-25). Cells were passaged using Accutase (Sigma Aldrich, 6964-100ml), Pou5f1 expression was abolished by treatment with 1µg/ml doxycycline (Sigma Aldrich D9891-1G) for 3 hours, 6 hours, 9 hours, 12 hours, 15 hours (verified by western blot) and 24 hours (verified by immunofluoresence and western blot).

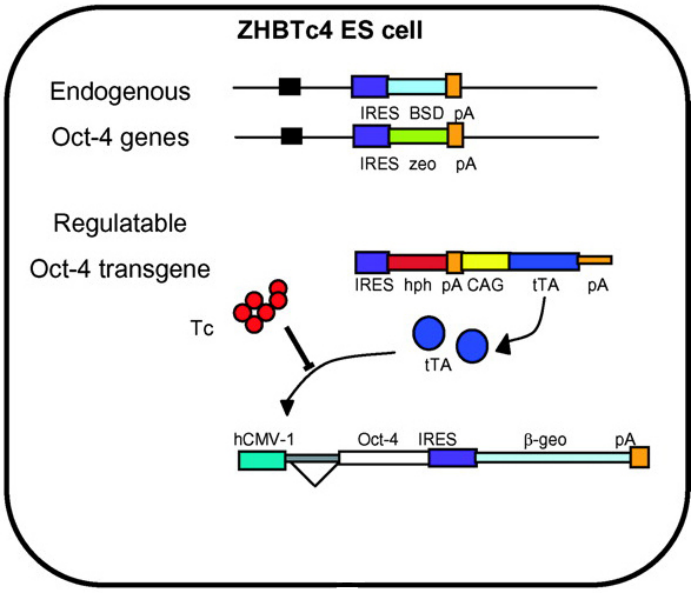


Figure 6. ZHBTc4 mouse ES cell used in this study.

Endogenous *Pou5f1* alleles were truncated and depleted by IRES-BSD-pA and IRES-zeo-pA. A regulatable *Pou5f1* transgene was inserted into mouse genome and addition of tetracycline leads to silence of its expression. Figure was adapted from (van den Berg et al., 2008).

1.2 Sample preparation and Western blotting

ZHBTc4 cells were washed with PBS and harvested using Accutase (Sigma Aldrich, 6964-100 ml) at the given time points of loss-of-Oct4. Cells were centrifuged for 5 minutes at 1400 rpm, supernatant was aspirated and cell pellet re-suspended as single cell suspension in cell culture medium. For whole cell lysate, cell pellets were weighed. Cell pellets were re-suspended in 4x LDS buffer based on weight. Equal volume for each sample was loaded on to SDS-PAGE gels. For chromatin samples, cell number was determined using counting chamber. $2x10^7$ cells were cross-linked with final concentration of 1% formaldehyde AppliChem, A0877,0250) for 8 minutes and quenched for 5 minutes with 125 mM Glycine Sigma Aldrich, G8898-1KG). Cross-linked cells were centrifuged for 5 minutes at 1350 x g and washed twice with 1 ml of cold PBS. Cells were either stored at -80°C or directly processed for chromatin extraction. For whole cell lysate samples equal volume for each sample was loaded on to SDS-PAGE gels. Chromatin samples were loaded equally based on DNA concentrations. Blots were probed for Oct4 (Santa Cruz, sc-5279), Sox2 (Santa Cruz, sc17320) Nanog (Bethyl Laboratories, A300-397A), Histone 3 (H3) for chromatin samples (Abcam ab1791), Tubulin for whole cell lysate samples (Sigma Aldrich, T6199) overnight at 4°C. Next day, blots were washed and probed with secondary antibodies anti-mouse-HRP (Jackson Labs, 115-035-044), anti-goat-HRP

(R&D systems, HAF019), anti-rabbit-HRP (GE healthcare, NA934) at room temperature for 2 hours. Blots were exposed to film using ECL (GE healthcare, RPN2232).

1.3 Immunofluorescence

ZHBTc4 cells were treated for 0 and 24 hours. Next, cells were cross-linked with 4% paraformaldehyde (Sigma Aldrich, D6148-500G) for 30 minutes. Formaldehyde was quenched with 50 mM Glycine (Sigma Aldrich, G8898-1KG) for 15 minutes. Cells were stained for Oct4 (Santa Cruz sc-5279) or Sox2 (Santa Cruz, sc17320) overnight at 4°C. The following day samples were incubated with anti-goat-alexa-488 (Thermoscientific, A11078) or anti-goat-alexa-568 (Thermoscientific, A11061) and Hoechst (Sigma Aldrich, H6024).

1.4 ATAC-seq

ATAC-seq was performed as described (Buenrostro et al., 2013) with a few alterations. ZHBTc4 cells were harvested using Accutase (Sigma Aldrich, 6964-100 ml) at 0 hours, 3 hours, 6 hours, 9 hours, 12 hours, 15 hours. Nuclear isolation of 5x10^4 cells was followed by treatment with Nextera Tn5 enzyme (Illumina, 20034198) for 45 minutes at 37°C. PCR amplification of the samples was performed using Nextera primers 1 and 2 and NEBNext High fidelity master mix (NEB, M0541S) for 12 cycles as determined by KAPA Real-Time Library Amplification Kit (Peqlab, KK2701). Libraries were purified over Machnery-Nagel PCR spin column (Macherey Nagel, 740609.50S) and AMPure XP beads (Beckman Coulter, A63881) in a 1:1.8 ratio. Sequencing of libraries was performed on a Illumina NEXTseq 550.

1.5 Sox2 ChIP-seq

Cross-linked pellets (as described in sample preparation for western blot) were thawed on ice. Protease inhibitor (Roche, 4693124001) was added to all buffers. Pellets were resuspended in lysis buffer 1 (50 mM Hepes-KOH pH7.5, 140 mM NaCl, 1 mM EDTA pH8, 10% glycerol, 0.5% IGEPAL CA630, 0.25% Triton-X100) and lysed for 30 minutes on ice. Samples were pelleted and washed with lysis buffer 2 (10 mM Tris-HCl pH8, 200 mM NaCl, 1 mM EDTA, 0.5M EGTA) for 10 minutes on roller bank at 4 C. Samples were pelleted and re-suspended in SDS sonication buffer (10mM Tris-HCl, 1mM EDTA, 0.5% SDS), incubated on ice for 10

minutes and transferred to TPX sonication tubes (Diagnod, C30010009). Chromatin was sonicated in Diagnod Bioruptor 4 x 15 minutes at 30 seconds ON and 30 seconds OFF, high setting in a cooled water bath. Sheared chromatin was centrifuged for 10 minutes 15000 rpm at 4°C. 25 µl of sample was de-crosslinked overnight at 65°C and distribution of size was checked on 1.4% agarose gel. 3µg of Sox2 antibody (Neuromics, GT15098) was coupled to Dynabeads protein G (Thermo Fisher Scientific, 10009D) for 2 hours at 4°C for each sample. 50µg of chromatin was used for each immunoprecipitation (IP). Chromatin was diluted using ChIP dilution buffer (10mM Tris-HCl pH8, 125mM NaCl, 0.125% Sodium deoxycholate, 1.25% Triton-X100). Antibody-chromatin mix was incubated overnight at 4°C rotating end-over-end. Samples were washed with low salt buffer (20mM Tris-HCl pH8, 150mM NaCL, 2mM EDTA, 0.1% SDS, 1% Triton-X100), twice using high salt buffer (20mM Tris-HCl pH8, 500mM NaCl, 0.1% SDS, 1% Triton-X100), twice using RIPA washing buffer (50mM HEPES-KOH pH7.6, 250mM LiCl, 1mM EDTA, 1% IGEPAL CA630, 0.7% Sodium deoxycholate) and once with TE buffer containing 50mM NaCl. Bound chromatin was eluted using 105µl pre-warmed elution buffer (10mM Tris-HCl pH8, 5mM EDTA, 300mM NaCl, 0.5% SDS) for 15 minutes at 65°C. RNase A (Invitrogen, 1004D was added and the samples were incubated overnight at 65°C. Next day, samples were treated with Proteinase K (AppliChem, A4392,0010) for 2 hours at 55°C. Samples were purified using Machnerey-Nagel PCR spin column (Macherey Nagel, 740609.50S). DNA quantity was done using Qubit 3.0 (Life Technology, Q33126). 25ng of DNA was used to prepare sequencing libraries using NEBNext® Ultra™ DNA Library Prep Kit (NEB, E7370L) according to manufactures manual. Purity and size distribution was analyzed on Fragment Analyzer. Libraries were sequenced on a HiSeq 1500 (Illumina).

The following TT-seq experiment was conducted by Dr. Jinmi Choi (Department of molecular Biology, MPI for Biophysical Chemistry).

1.6 TT-seq

TT-seq was performed as described (Schwalb et al., 2016) with minor alterations. In brief, two biological replicates at the aforementioned timepoints were produced for TT-seq. 1*10^8 cells were labeled with 500 µM 4-thiouridine (4sU, Carbosynth, 13957-31-8) for 5min. Cells were harvested and lysed using TRIzol (Ambion, 1559018) and stored at -80°C. Prior to RNA isolation, RNA spike-ins were added at 5 ng per 1*10^8 cells. Details regarding the used spike-in sequences and generation of the spike-in mix can be found in (Wachutka et al., 2019). Total RNA was isolated using TRIzol (Ambion, 1559018) according to manufactures instructions, and subsequentially fragmented to 1500-5000 bp using Covaris S220 Ultrasonicator. Nascent RNA was purified as described (Schwalb et al., 2016) with minor modifications. Following purification using streptavidin pulldown, the collected RNA was purified using RNeasy micro Kit (Qiagen, 74004), as well as DNase treatment (Qiagen, 79254). Sequencing libraries were produced using NuGen Ovation Universal RNA-seq System (Nugen, 0343). Size selected libraries were analyzed on a Fragment Analyzer before sequencing on a Illumina NEXTseq 550.

The following Oct4 ChIP-seq experiment was conducted by Dr. Livia Caizzi (Department of molecular Biology, MPI for Biophysical Chemistry).

1.7 Oct4 ChIP-seq

Cross-linked pellets (as described in sample preparation for western blot) were thawed on ice. Protease inhibitor (Roche, 4693124001) was added to all buffers. A pellet coming from 3 x 10^7 cells was resuspended in Farnham Lysis buffer (5 mM Pipes pH 8, 85 mM KCl, 0.5 % NP-40) and lysed for 10 minutes on ice. Samples were pelleted for 5 min at 1,700 g at 4 °C. Samples were washed with PBS and pelleted for 5 min at 1,700 g at 4 °C. Samples were re-suspended in 1 mL of SDS sonication buffer (10 mM Tris-HCl 7.5 pH, 1 mM EDTA, 0.4 % SDS), incubated on ice for 10 minutes and transfer to AFA milliTube. Sonication was performed with a S220 Focused-ultrasonicator (Covaris) with the following parameters: duty cycle 5 %, peak incident power 140 W, cycle per burst 200, processing time 840 sec, degassing mode continuous, water run level 8. 25 µl of sample was de-crosslinked overnight at 65°C and distribution of size was checked on 1.4% agarose gel. 40 µg of Oct4 antibody (R&D, AF1759) was coupled to Dynabeads protein G (Thermo Fisher Scientific, 10009D) for 2 hours at room temperature for each sample. 100 µg of chromatin was used for each IP. 100 ng of *Drosophila S2* sheared crosslinked-chromatin (Covaris S200 parameters: duty cycle 5 %, peak incident power 140 W, cycle per burst 200, processing time 1,800 sec, degassing mode continuous, water run level 8) were added to 100 g of chromatin as spike-ins control. Chromatin was diluted in IP buffer (50 mM Hepes pH 7.9, 150 mM NaCl, 1 mM EDTA, 1 % Triton X-100, 0.1 % Sodium-deoxycholate) to obtain a 0.1 % final concentration of SDS. 1 % of diluted chromatin was kept as input at 4 °C. Antibody-chromatin mix was incubated overnight at 4°C rotating end-over-end. Samples were washed 5 times with IP wash buffer (100 mM Tris HCl pH 7.5, 500 mM LiCl, 1 % NP-40, 1 % Sodium-deoxycholate) and one time with TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA). Immuno-bound chromatin was eluted at 70 °C for 10 min with elution buffer (0.1M $NaHCO_3$, 1 %SDS) and de-crosslinked overnight at 65 °C. After RNAse A treatment at 37 °C for 1.5 h and proteinase K treatment at 45 °C for 2 h, DNA was extracted with one volume phenol:chloroform:isoamyl alcohol 25:24:1 (Sigma-Aldrich, P2069) and precipitated for 30 min at -80 °C with 200 mM NaCl and 100 % ethanol. Pellet was washed with 70 % ethanol and resuspended in TE buffer. DNA quality and size distribution were checked on Fragment Analyzer. 3 ng of DNA was used for library preparation according to NEBNext® Ultra™ II DNA Library Prep Kit (NEB, E7645S). Purity and size distribution was

analyzed on Fragment Analyzer. Size-selected libraries were sequenced on Illumina NEXTseq 550.

2. Computational methods and data statistic

I carried out all bioinformatics analyses.

2.1 TT-seq data preprocessing

Paired-end 42 bp reads were mapped to the mouse genome assembly mm10 using STAR 2.5.3 (Dobin et al., 2013) with the following parameters: outFilterMismatchNmax 2, outFilterMultimapScoreRange 0 and alignIntronMax 500,000. SAMtools (Li et al., 2009) was then used to remove alignments with MAPQ smaller than 7 (-q 7) and only proper pairs (-f 2) were selected. HTSeq-count (Anders et al., 2015) was used to calculate fragment counts for different features. Further data processing was carried out using the R/Bioconductor environment.

2.2 Antisense bias correction

In TT-seq, we used spike-ins to infer the antisense bias ratio during library preparation. Reads mapped to spike-ins were counted by HTSeq-count for both sense and antisense strands. Antisense bias ratios c_j were evaluated for each sample j by the following equation:

$$c_j = median_i \left(\frac{k_{ij}^{antisense}}{k_{ij}^{sense}} \right)$$

k_{ij} denotes mapped reads for each of the spike-ins *i* in each sample *j*.

We then used the estimated ratio c_j to correct observed read counts for each strands using the following formula:

$$k_{ij}{}_{\text{corrected}} = \frac{k_{ij}{}^{sense} - c_j * k_{ij}{}^{antisense}}{1 - c_j{}^2}$$

Here we assume the observed sense counts ${k_{ij}}^{sense}$ come from the sum of real sense counts ${k_{ij}}_{\text{corrected}}$ and a small percentage of antisense counts introduced by antisense bias of library preparation.

2.3 Transcription unit annotation and classification

Annotation of transcription unit (TU) was performed as described (Schwalb et al., 2016) with minor modifications. Briefly, the whole genome was segmented into 200-bp consecutive bins and the midpoint of TT-seq fragments was then used to calculate the coverage for each bin for each sample. A pseudo-count was added to each bin to avoid noisy signals. In order to create a unified annotation independent of a specific time point, all TT-seq samples were combined. The R/Bioconductor package GenoSTAN (Zacher et al., 2017) was then used to learn a two-state hidden Markov model with a PoissonLog-Normal emission distribution in order to segment the genome into "transcribed" and "untranscribed" states. Transcribed regions overlapping at least 20% of their length with GENCODE annotated protein-coding gene or lincRNA and overlapping with an annotated exon were classified as mRNA/lincRNA and the rest was defined as ncRNA. Transcribed regions mapping to exons of the same protein-coding gene or lincRNA were combined to create a consecutive transcription unit. In order to avoid spurious predictions, ATAC-seq data was used to call open chromatin regions (see below) and TUs for which their promoter (+/- 1kb of transcription start site, TSS) did not overlap with an opening chromatin region were removed. A minimal expression threshold was optimized based on expression difference between TUs for which their promoter did or did not overlap with an opening chromatin region. This resulted in 26822 TUs originating from an open chromatin region with a minimal RPK of 26.5. In order to overcome low expression or mappability issues, ncRNAs that were only 200bp (1bin) apart were merged. Subsequently, TU start and end sites were refined to single nucleotide precision by finding borders of abrupt coverage increase or decrease between two consecutive segments in the four 200-bp bins located around the initially assigned start and stop sites via fitting a piecewise constant curve to the TT-seq coverage profiles for both replicates using the segmentation method from the R/Bioconductor package (Huber et al., 2006).

ncRNAs were then classified into the following four categories according to their respective location relative to protein-coding genes: upstream antisense RNA (uaRNA), convergent RNA (conRNA), antisense RNA (asRNA) and intergenic RNA (incRNA) (**Figure 7C**). ncRNAs

located on the opposite strand of an mRNA were classified as asRNA if the TSS was located > 1 kbp downstream of the sense TSS, as uaRNA if the TSS was located < 1 kbp upstream of the sense TSS, and as conRNA if the TSS was located < 1 kbp downstream of the sense TSS. The remaining ncRNAs were classified as incRNA.

To annotate putative eRNAs we selected asRNAs and incRNAs. Since highly synthesized mRNAs can give rise to spurious and un-continuous downstream transcription signal, we restricted the analysis to a subset of asRNAs and incRNAs that were located at least 1kb far away from promoter related RNAs including mRNAs, uaRNAs, conRNAs and defined them as putative eRNAs. We then merged them if the putative eRNAs fell within 1kb of each other. eRNAs within 1kb of Oct4-occupied chromatin regions were defined as Oct4-regulated eRNAs and the corresponding Oct4 peaks were classified as Oct4-bound transcribed enhancers. Genome browser views showing coverages and annotations were plotted by software pyGenomeTracks (Lopez-Delisle et al., 2021).

2.4 Differential gene expression analysis

R/Bioconductor package DESeq2 (Love et al., 2014) was used to call differentially expressed mRNAs and eRNAs applying DESeq2's default size factor normalization. For both cases DESeq2 size factors were calcualted using counts for protein-coding genes. An adjusted P-value of 0.01 was used to identify significantly changed mRNAs or eRNAs by comparing each time point to the 0 hours measurement.

In DESeq2, the default size factor is evaluated by the following equation:

$$s_j = median_i \frac{k_{ij}}{\prod_{j=1}^{m} {k_{ij}}^{1/m}}$$

Size factor is calculated for each sample j. In the formula, the denominator can be seen as a pseudo reference sample evaluated by calculating the geometric mean for each gene i across all samples from 1 to m. k_{ij} denotes the raw TT-seq count for gene i in sample j. Size factor s_j is then evaluated by the median of the ratios of the j-th sample's TT-seq counts to the pseudo-reference sample.

2.5 Principal component analysis

For each replicate, size factor normalized feature counts were obtained and DESeq2's default variance stabilizing transformation was applied. The DESeq2 plotPCA function was used to generate PCA plots with the default parameters. By default, the top 500 genes with highest row variance are selected and used for principal components analysis.

2.6 K-means clustering

Size factor normalized feature counts were aggregated for two biological replicates for each time point and then the data matrix was subjected to Z-score transformation before clustering.

$$Z = \frac{x - \mu}{\sigma}$$

Z = standard score, x = observed count, μ = mean of the sample, σ = standard deviation of the sample.

K-means clustering was performed using the kmeans function in R. We first performed 2000 rounds clustering with 2 groups to create a cluster center matrix. This matrix was used as an input for an additional round clustering to generate the cluster center, which was used as the cluster centers of the final K-means clustering.

2.7 ATAC-seq data processing

Paired-end 76 bp reads were obtained for each of the samples and Nextera Transposase adapter sequence was removed using (Martin, 2011). Bowtie2 (Langmead and Salzberg, 2012) was used to align paired-end reads to the mouse genome assembly mm10 with the "--local" and "--no-discordant" options. SAMtools (Li et al., 2009) was then used to remove alignments with MAPQ smaller than 7 (-q 7) and only proper pairs (-f 2) were selected. Reads mapped to custom blacklist regions and mitochondria were removed. Two replicates were pooled and MACS2 (Zhang et al., 2008) was used to call open chromatin peaks with options: -f BAMPE -g mm --broad --broad-cutoff 0.05. Peaks for all time point were then merged to create non-overlapping unified peaks. Further data processing was carried out using the R/Bioconductor environment.

For quantitative comparison, HTSeq-count (Anders et al., 2015) was used to calculate fragment counts for the non-overlapping unified peaks and DESeq2 was used to call regions with significantly changed chromatin accessibility. For normalization, ATAC-seq peaks overlapping with promoters of protein-coding genes with unchanged mRNA expression were used to calculate DESeq2 size factors (**see 2.4**). An adjusted P-value of 0.01 was used to identify significantly changed open chromatin regions by comparing each time point to the 0 hour measurement.

2.8 ChIP-seq data processing

Paired-end ChIP-seq data processing was done as described for ATAC-seq data. For single-end ChIP-seq data, Bowtie2 (Langmead and Salzberg, 2012) was used for mapping with "--local" option. SAMtools (Li et al., 2009) was then used to remove alignments with MAPQ smaller than 7 (-q 7). All published ChIP-seq data were fully processed by ourselves. Detailed information for all ChIP-seq samples can be found in (**Table 3-5, section 2.13 Data statistic**). For peak calling of published paired-end Oct4, Sox2 and Nanog ChIP-seq data, three replicates were pooled and MACS2 (Zhang et al., 2008) was used to call peaks with options: -f BAMPE -g mm. For our paired-end Oct4, Sox2 ChIP-seq data, peaks were called by MACS2 with the same options for each individual replicate.

For our Oct4 ChIP-seq data, data was normalized using added *D. melanogaster* spike-ins. Normalization factors were obtained by dividing the total *D. melanogaster* spike-ins read counts for each sample by the total spike-ins read counts of the sample with the lowest spike-ins read counts. For our Sox2 ChIP-seq data, data was normalized using total number of uniquely mapped reads. ChIP-seq coverages were divided by the respective normalization factors. HTSeq-count (Anders et al., 2015) was used to calculate fragment counts for peak features. Reads counts were divided by the respective normalization factors.

To avoid noisy signal, we only used Sox2 peaks that were detected by both replicates (consensus peaks) at 0 hours samples for Sox2 occupancy analysis. We overlapped Sox2 consensus peaks with Oct4-sensitive, -insensitive and -independent enhancers and kept the overlapped peaks for the analysis in (**Figure 19A-B, Figure 20A-B, Figure 23A-B**). For Oct4-bound transcribed enhancers this resulted in 154, 67 and 111 Sox2 consensus peaks that overlapped with Oct4-sensitive, -insensitive and -independent enhancers in (**Figure 19A-B,**

Figure 20A-B). For Oct4-bound non-transcribed enhancers this resulted in 1010 and 478 Sox2 consensus peaks that overlapped with Oct4-sensitive and -insensitive enhancers in (**Figure 23A-B**). For Oct4 occupancy analysis in (**Figure 22F**), we used the same strategy and this resulted in 844 and 187 Oct4 consensus peaks for Oct4-sensitive and -insensitive enhancers used in (**Figure 22F**).

2.9 GO enrichment analysis

The gene ontology enrichment analysis for differentially expressed mRNAs was performed by DAVID Bioinformatics Resources (Huang da et al., 2009). Genomic regions enrichment analysis was performed by GREAT (McLean et al., 2010).

2.10 Transcription factor binding motif analysis

DNA sequences +/-500bp around Oct4 ChIP-seq peak summit were extracted and FIMO (Grant et al., 2011) was used to find individual motif occurrences. DNA motifs of Oct4, Sox2, Sox2-Oct4 composite, Klf4 and Esrrb were download from JASPAR database (Fornes et al., 2020). Percent of motif occurrence was calculated by counting how many sequences contained the queried motifs in the total of subjected sequences.

2.11 Super enhancer annotation in mESC

Previously, 231 large enhancer domains were identified to be super-enhancers from 8794 sites that are co-occupied by Oct4, Sox2 and Nanog in mESC (Whyte et al., 2013). We downloaded the annotation from the original publication and used UCSC LiftOver (Hinrichs et al., 2006) to convert the coordinates from mouse mm9 genome assembly to mm10. To extract the nearest genes of super enhancers (to use as putative target genes), we first defined the promoters of TT-seq annotated protein-coding genes by taking the +/- 1kb regions around the transcription start sites. We then used the nearest function from the R package GenomicRanges to find the nearest genes by using the super enhancers annotation as query object and the promoter annotation as subject object. This resulted in 150 putative target genes of annotated super enhancers and was used in (**Figure 11**).

2.12 Public RNA-seq data analysis

For the RNA-seq data analyzed in (**Figure 28**), the normalized RNA-seq counts table for genes for all samples was downloaded from original paper (Bates et al., 2021).

2.13 Data statistic

Table 1. Sequencing statistics of TT-seq samples generated in this study.

All samples were sequenced on a NEXTseq 550 sequencing platform in 42bp paired-end mode.

No.	Hours of DOX treatment	Replicate no.	Sequenced reads	Mapped reads	Duplicates（%）
1	0h	1	139,767,320	115,716,286	42.6
2		2	137,196,527	115,935,348	31.8
3	3h	1	134,759,731	113,952,490	31.3
4		2	140,542,668	119,816,299	35.4
5	6h	1	139,202,151	117,349,804	38.5
6		2	139,904,081	118,716,899	33.7
7	9h	1	143,258,532	120,721,286	42.5
8		2	131,906,425	110,880,848	35.2
9	12h	1	136,024,166	115,730,201	38.3
10		2	136,854,827	116,897,968	33.4
11	15h	1	141,129,470	120,097,702	40.9
12		2	132,219,606	111,909,955	31.6

Table 2. Sequencing statistics of ATAC-seq samples generated in this study.

All samples were sequenced on a NEXTseq 550 sequencing platform in 76bp paired-end mode.

No.	Hours of DOX treatment	Replicate no.	Sequenced reads	Mapped reads	Duplicates （%）
1	0h	1	75,879,782	60,778,817	7.0
2		2	84,285,910	66,234,761	7.0
3	3h	1	99,108,163	76,192,137	7.0
4		2	95,058,225	75,287,164	8.0
5	6h	1	100,888,477	80,437,745	7.0
6		2	96,527,924	77,644,843	7.0
7	9h	1	98,029,091	76,592,320	6.0
8		2	96,395,215	76,365,304	7.0
9	12h	1	104,560,309	82,912,008	7.0
10		2	91,177,600	72,588,902	7.0
11	15h	1	93,561,692	73,702,559	8.0
12		2	82,330,900	65,056,516	6.0

Table 3. Sequencing statistics of Oct4 ChIP-seq samples generated in this study.

All samples were sequenced on a NEXTseq 550 sequencing platform in 43bp paired-end mode.

No.	Hours of DOX treatment	Replicate no.	Sequenced reads	Mapped reads	Duplicates（%）
1	0h	1	45,489,989	37,572,722	8.0
2		2	44,844,975	35,827,918	9.0
3	3h	1	45,200,637	37,381,092	7.0
4		2	40,712,696	33,473,643	7.0
5	6h	1	35,615,780	28,101,180	11.0
6		2	44,105,296	34,059,983	12.0
7	9h	1	39,889,543	32,484,669	8.0
8		2	33,968,970	27,705,901	8.0
9	12h	1	49,893,843	40,471,411	9.0
10		2	40,182,623	31,715,336	11.0
11	15h	1	54,230,702	43,871,081	8.0
12		2	53,405,989	38,050,210	10.0

Table 4. Sequencing statistics of Sox2 ChIP-seq samples generated in this study.

All samples were sequenced on a HiSeq 1500 (Illumina) sequencing platform in 76bp paired-end mode.

No.	Hours of DOX treatment	Replicate no.	Sequenced reads	Mapped reads	Duplicates （%）
1	0h	1	31,157,808	26,708,870	8.0
2		2	30,358,437	26,105,044	8.0
3	3h	1	29,224,107	25,024,914	8.0
4		2	23,143,269	19,813,671	8.0
5	6h	1	32,774,215	28,167,501	8.0
6		2	26,661,227	22,987,161	7.0
7	9h	1	29,314,133	25,170,019	9.0
8		2	25,428,443	21,723,313	8.0
9	12h	1	26,165,258	22,383,270	9.0
10		2	24,258,711	20,855,734	8.0
11	15h	1	26,728,664	22,835,623	8.0
12		2	24,926,639	21,369,856	9.0

Table 5. List of previously published ChIP-seq datasets used in this study.

No.	Read mode	Data	Hours of DOX treatment	Cell line	Available at
1	Paired-end	Oct4	0h	ZHBTc4	NCBI Gene Expression Omnibus (GSE87822)
			24h		
2	Paired-end	Sox2	0h	ZHBTc4	NCBI Gene Expression Omnibus (GSE87822)
			24h		
3	Paired-end	Nanog	0h	ZHBTc4	NCBI Gene Expression Omnibus (GSE87822)
			24h		
4	Single-end	Klf4		C57BL/6J	NCBI Gene Expression Omnibus (GSE 90895)
5	Single-end	Esrrb		C57BL/6J	NCBI Gene Expression Omnibus (GSE 90895)
6	Single-end	H3K4me1		C57BL/6J	NCBI Gene Expression Omnibus (GSE 90895)
7	Single-end	H3K4me3		C57BL/6J	NCBI Gene Expression Omnibus (GSE 90895)
8	Single-end	H3K27ac		C57BL/6J	NCBI Gene Expression Omnibus (GSE 90895)

2.14 Data availability

The sequencing data sets generated in this study have been deposited in the Gene Expression Omnibus (GEO) database under accession code GSE174774.

3. Software

Table 6. Software used in this study.

Software	Source	Identifier
STAR (2.5.3)	Dobin et al., 2013	https://github.com/alexdobin/STAR, RRID:SCR_015899
Samtools (1.6)	Li et al., 2009	https://github.com/samtools/samtools, RRID: SCR_002105
HTSeq (0.9.1)	Anders et al.,2015	https://htseq.readthedocs.io/en/master/, RRID:SCR_005514
GenoSTAN (2.18.0)	Zacher et al., 2017	https://www.bioconductor.org/packages/release/bioc/html/STAN.html
tilingArray (1.68.0)	Huber et al., 2006	https://www.bioconductor.org/packages/release/bioc/html/tilingArray.html
pyGenomeTracks (3.5.1)	Lopez-Delisle et al., 2021	https://github.com/deeptools/pyGenomeTracks, RRID:SCR_016366
Cutadapt (1.16)	Martin, 2011	https://cutadapt.readthedocs.io/en/stable/guide.html, RRID:SCR_011841
MACS2 (2.1.1.20160309)	Zhang et al., 2008	https://pypi.org/project/MACS2/, RRID:SCR_013291
Bowtie 2 (2.3.4.1)	Langmead and Salzberg., 2012	http://bowtie-bio.sourceforge.net/bowtie2/index, RRID:SCR_005476
DESeq2 (1.26.0)	Love et al, 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html, RRID:SCR_015687
DAVID Bioinformatcis Resources (6.8)	Huang da et al., 2009	https://david.ncifcrf.gov, RRID:SCR_001881
GREAT (4.04)	McLean et al., 2009	http://great.stanford.edu/public/html/, RRID:SCR_005807
FIMO (5.0.4)	Grant et al., 2011	https://meme-suite.org/meme/doc/fimo.html, RRID:SCR_001783
UCSD LiftOver	Hinriches et al., 2006	https://genome.ucsc.edu/cgi-bin/hgLiftOver, RRID:SCR_018160

III. Results

The results presented in this chapter have been published:
Le Xiong*, Erik A. Tolen* et al. Oct4 primarily controls enhancer activity rather than accessibility (* joint first authorship)
Manuscript in revision, doi: https://doi.org/10.1101/2021.06.28.450119

A detailed list of items from the manuscript can be found in Appendix (page 104).

Author contributions

The initial design of the study was by EAT, KA and HRS, with input from JC and PC. LX carried out bioinformatics analyses. EAT carried out experiments except that LC performed Oct4 ChIP-seq and JC performed TT-seq, with input from EAT. ML and JC provided assistance with bioinformatics analyses. PC and HRS supervised research. LX, EAT and PC wrote the manuscript with input from LC, ML and HRS.

Chapter summary

The transcription factor Oct4 is essential for maintaining stem cell pluripotency and for efficient cell reprogramming, but its functional roles are far from being understood. Here, we investigate the functions of Oct4 by rapidly depleting Oct4 from mouse embryonic stem cells and conducting a time-resolved multi-omics analysis. Oct4 depletion leads to an immediate loss of its binding to putative enhancers that are accessible in chromatin. Loss of Oct4 is accompanied by a concomitant decrease in mRNA synthesis from putative target genes that are part of the transcriptional network that maintains pluripotency. Oct4 binding to enhancers does not correlate with chromatin accessibility, whereas Sox2 may retain accessibility after Oct4 depletion and even in the absence of eRNA synthesis. These results suggest that Oct4 primarily controls enhancer activity rather than accessibility to maintain pluripotency gene network in ESCs.

1. Rapid Oct4 depletion and transcription unit annotation

To investigate the role of Oct4 in maintaining pluripotency, we used a doxycycline (DOX) inducible Oct4 loss-of-function mouse embryonic stem cell line (mESC) ZHBTc4 that allows for rapid depletion of Oct4 (**Section II, 1.1, Figure 6**) (Niwa et al., 2000). This system was

previously used to study the effect of Oct4 depletion after 24 hours of DOX treatment (Friman et al., 2019; King and Klose, 2017). In order to investigate the direct, primary role of Oct4, we conducted a time course experiment collecting samples after 0, 3, 6, 9, 12 and 15 hours of DOX treatment. We found that Oct4 protein levels were already reduced after 6 hours of DOX treatment and decreased strongly after 9 hours (**Figure 7A, whole cell lysate**). Oct4 protein depletion was nearly complete at 24 hours of treatment, while Sox2 and Nanog protein levels remained essentially unchanged for extended times before decreasing (**Figure 7A, Figure 8A**). Chromatin binding of Oct4 was impaired already after 3 hours of DOX treatment, whereas Sox2 binding decreased after 9 hours and Nanog binding remained unchanged over the entire time course (**Figure 7A, chromatin**).

To monitor the effect of a rapid Oct4 depletion on transcription, we conducted TT-seq (Schwalb et al., 2016) at 0, 3, 6, 9, 12 and 15 hours after DOX treatment (**Figure 7B**). RNA labelling with 4-thiouridine (4sU) was carried out for 5 minutes (min) and two independent biological replicates were generated for each time point (**Section II, 2.13, Table 1**). To study the role of Oct4 in maintaining chromatin accessibility, we performed ATAC-seq (Buenrostro et al., 2013) over the same time course (**Figure 7B, Section II, 2.13, Table 2**). TT-seq and ATAC-seq data were highly reproducible (**Figure 8B-C**).

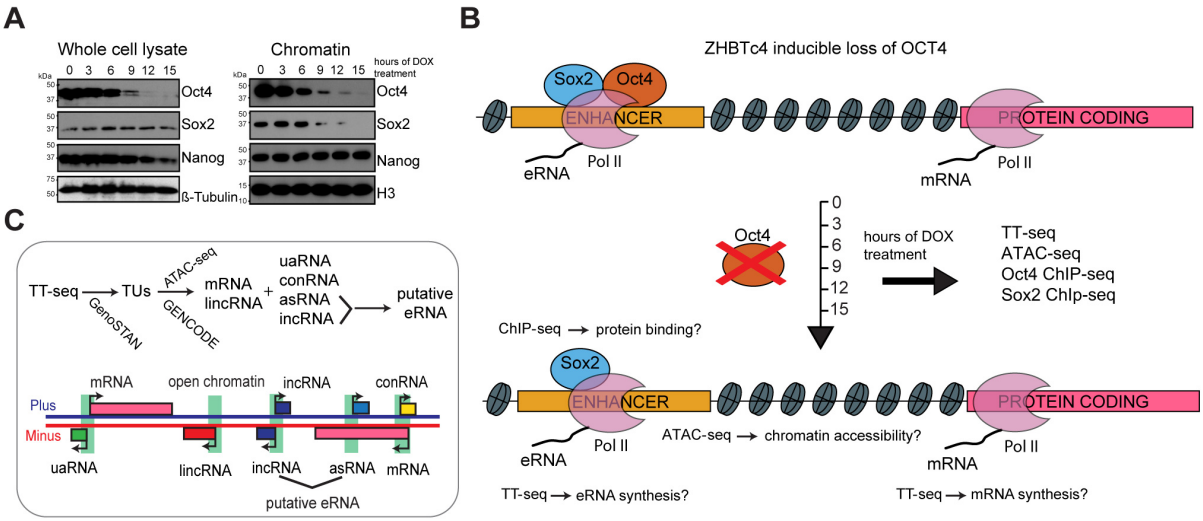


Figure 7. Rapid Oct4 depletion in ZHBTc4 mouse ESCs.

(A) Western blot analysis of whole cell lysate and chromatin samples over time course of DOX treatment using Oct4, Sox2, Nanog, ß-Tubulin (control) and H3 (control) antibodies. (B) Schematic of methodology and samples collected. TT-seq, ATAC-seq, Oct4 and Sox2 ChIP-seq experiments were performed after 0, 3, 6, 9, 12 and 15 hours of DOX treatment in ZHBTc4 mouse ESCs. (C) Transcription unit annotation. Genome segmentation with GenoSTAN was used to annotate transcription units (TUs) from TT-seq data. ATAC-seq data and mouse GENCODE annotation were then combined to classify TUs.

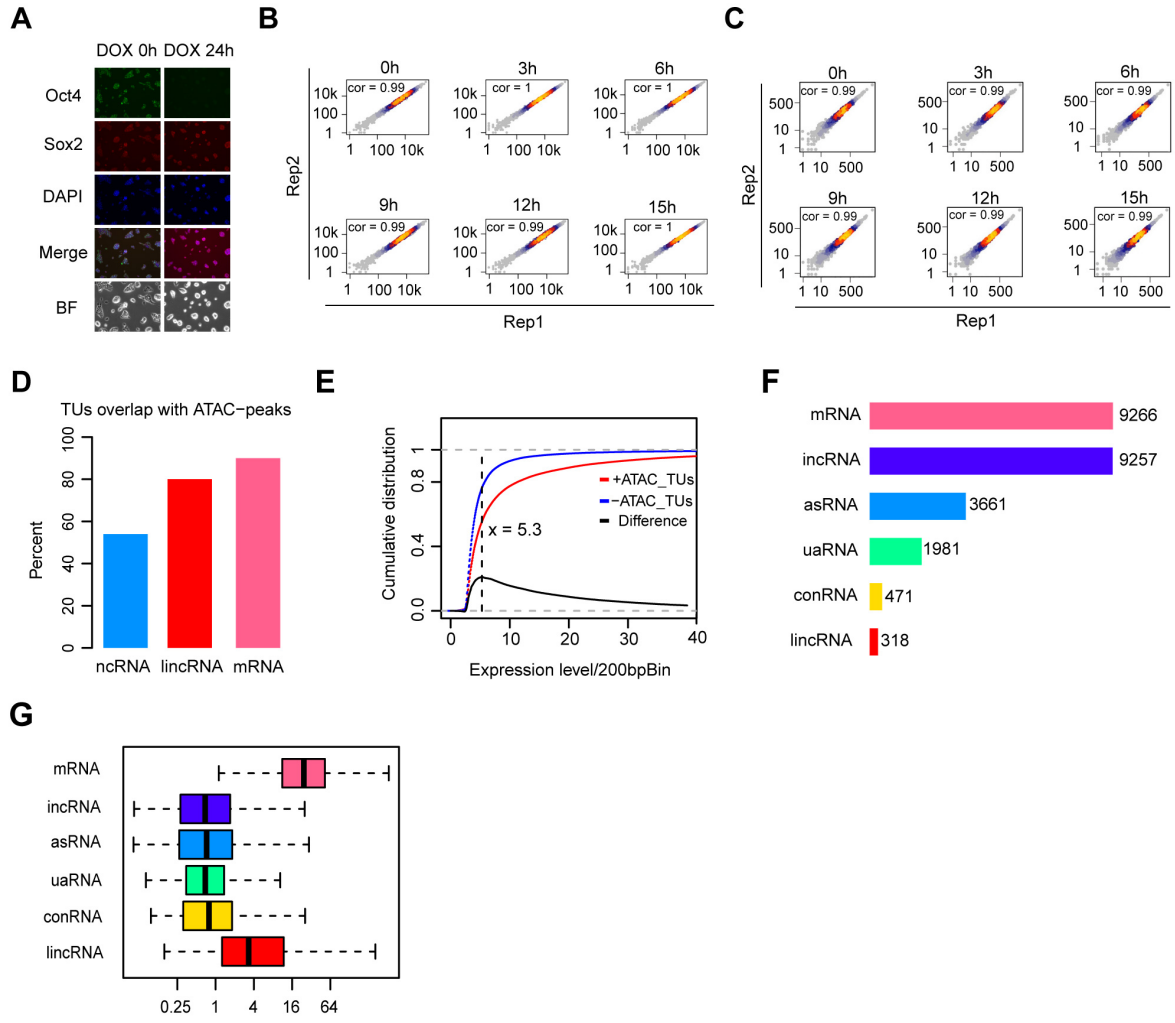


Figure 8. Transcription unit annotation in ZHBTc4 mouse ES cell line.

(A) Immunofluorescence of ZHBTc4 cells after 0 and 24 hours of 1µg/ml DOX treatment. Green: Oct4, red: Sox2, blue: DAPI, BF: Brightfield. (B) Scatter plots showing correlation between two biological replicates of TT-seq data at each time point. Normalized TT-seq counts of annotated mRNA (n=9,266) were plotted. Spearman's rank correlation coefficient was calculated and shown in each plot. (C) Scatter plots showing correlation between two biological replicates of ATAC-seq data at each time point. Normalized ATAC-seq counts for Oct4-bound transcribed enhancers (n=2,223) were plotted. Spearman's rank correlation coefficient was calculated and shown in each plot. (D) Bar graph illustrating percentage of overlap of GenoSTAN annotated ncRNAs, lincRNAs and mRNAs with chromatin open regions defined by ATAC-seq peaks. (E) Cumulative distribution of expression level of TUs with (red) or without (blue) ATAC-seq peaks overlapping. Black line indicates the difference between distribution of expression level for TUs with or without ATAC-seq peaks overlapping. Dash black line indicates the maximum point for the difference. (F) Numbers of different types of annotated TUs. (G) Length distribution of different types of annotated TUs.

We then used the TT-seq data to segment the genome into transcription units (TUs) and non-transcribed regions using GenoSTAN (Zacher et al., 2017) (**Figure 7C and Section II, 2.3 TU annotation methods**). To avoid spurious predictions, TUs detected by TT-seq had to exceed a

minimal expression cutoff of RPK > 26.5 and had to originate from an open chromatin region identified by ATAC-seq (**Figure 8D-E**). We sorted TUs into protein-coding RNAs (mRNAs) and long intergenic non-coding RNAs (lincRNAs) based on GENCODE annotation (Frankish et al., 2019). The remaining non-coding TUs were classified as upstream antisense RNA (uaRNA), convergent RNA (conRNA), antisense RNA (asRNA) and intergenic RNA (incRNA) units according to the location relative to mRNA (**Figure 7C and Section II, 2.3 TU annotation methods**). This resulted in 9,266 mRNAs, 9,257 incRNAs, 3,661 asRNAs, 1,981 uaRNAs, 471 conRNAs and 318 lincRNAs (**Figure 8F**). The length distribution of the detected RNA units (**Figure 8G**) agreed with previous estimations (Michel et al., 2017; Schwalb et al., 2016).

2. Oct4 maintains the transcriptional network governing pluripotency

We first investigated changes in mRNA synthesis during Oct4 depletion. Changes could already be observed after 3 hours of DOX treatment (**Figure 9A**), in agreement with chromatin fractionation results (**Figure 7A**, **chromatin**). Differential gene expression analysis (Love et al., 2014) detected 769 significantly down-regulated and 829 up-regulated genes (adjusted P-value = 0.01) after 15 hours of DOX treatment (**Figure 9B-D**). To dissect the kinetics of mRNA synthesis changes of differentially expressed genes, we performed k-means clustering and classified early and late down- and up-regulated genes. Early down-regulated genes (446 genes) showed the strongest decrease in mRNA synthesis after 6-9 hours of DOX treatment (**Figure 9E-F, left**), whereas late down-regulated genes (323 genes) decreased most strongly after 12-15 hours (**Figure 9E-F, right**). Early and late up-regulated genes behaved similarly (**Figure 10A-C**). Gene ontology (GO) analysis (Huang da et al., 2009) showed that early down-regulated genes were enriched for stem cell population maintenance (**Figure 9G**), whereas late down-regulated genes were enriched for DNA replication and cell cycle (**Figure 10D**). Early up-regulated genes showed enrichment for carbohydrate metabolic process (**Figure 10E**), whereas late up-regulated genes were slightly enriched for *in utero* embryonic development (**Figure 10F**). These findings reflect the known transition of ZHBTc4 cells towards the trophectoderm upon loss of Oct4 (Niwa et al., 2000).

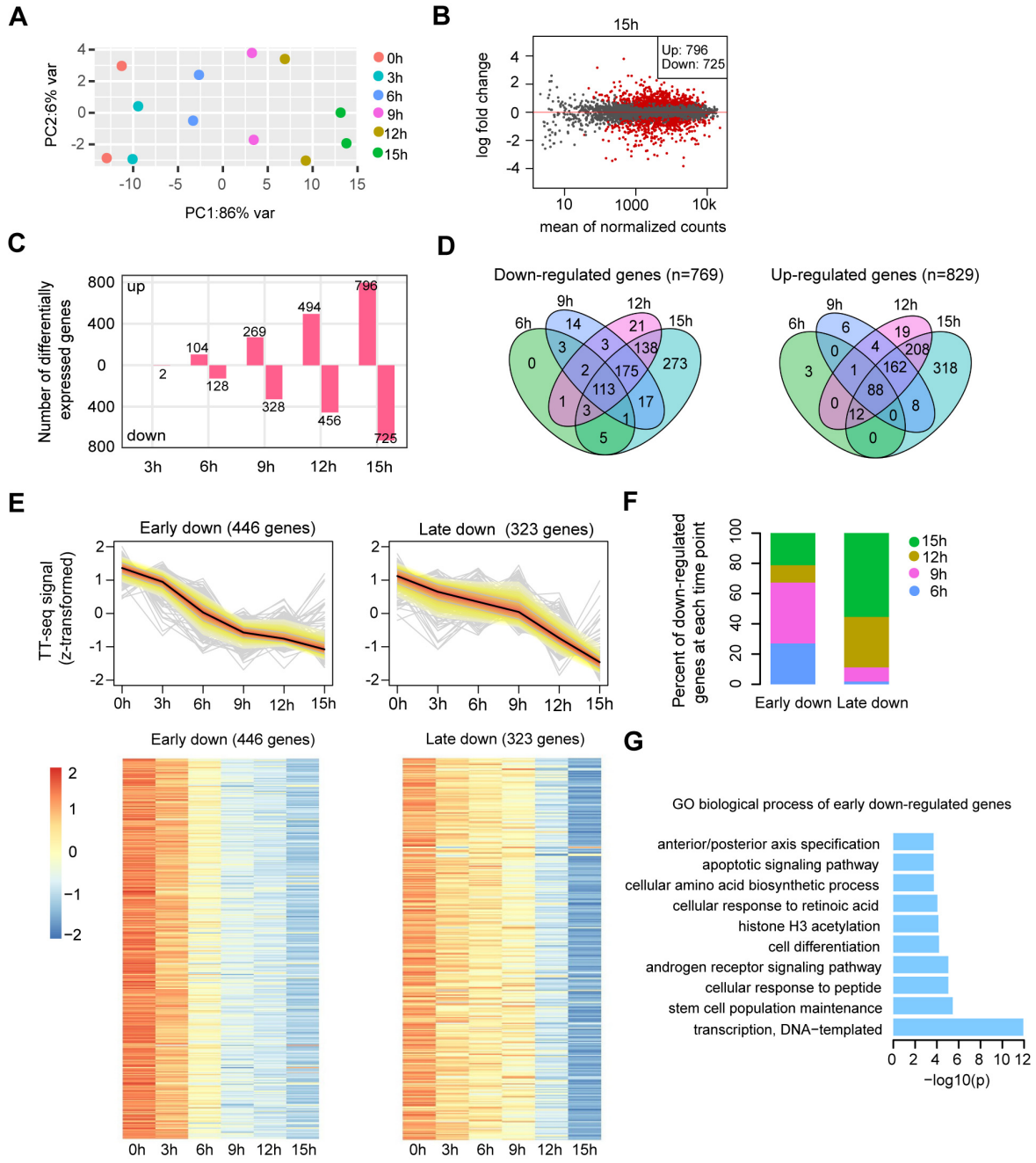


Figure 9. Oct4 maintains the transcriptional network governing pluripotency.

(A) Principle component analysis of TT-seq annotated mRNA (n=9,226) over time course of DOX treatment (0, 3, 6, 9, 12, 15 hours) for two biological replicates. (B) MA plot showing differentially regulated mRNAs detected by DESeq2 after 15 hours of DOX treatment. Statistically significantly regulated mRNAs are depicted in red. (C) Number of differentially expressed mRNAs detected by DESeq2 after each time point of DOX treatment. (D) Venn diagram showing the overlap for differentially expressed mRNAs detected by DESeq2 at each time point of DOX treatment. (E) K-means clustering of 769 down-regulated mRNAs into early (left) and late (right) down-regulated groups. Y axis indicates z-score transformed TT-seq counts. Bottom: heatmap indicating the kinetics of early and late down-regulated genes. (F) Differentially regulated time point composition for early (left) and late (right) down-regulated gene groups. Y axis indicates percentage. (G) GO biological process enrichment of early down-regulated mRNAs.

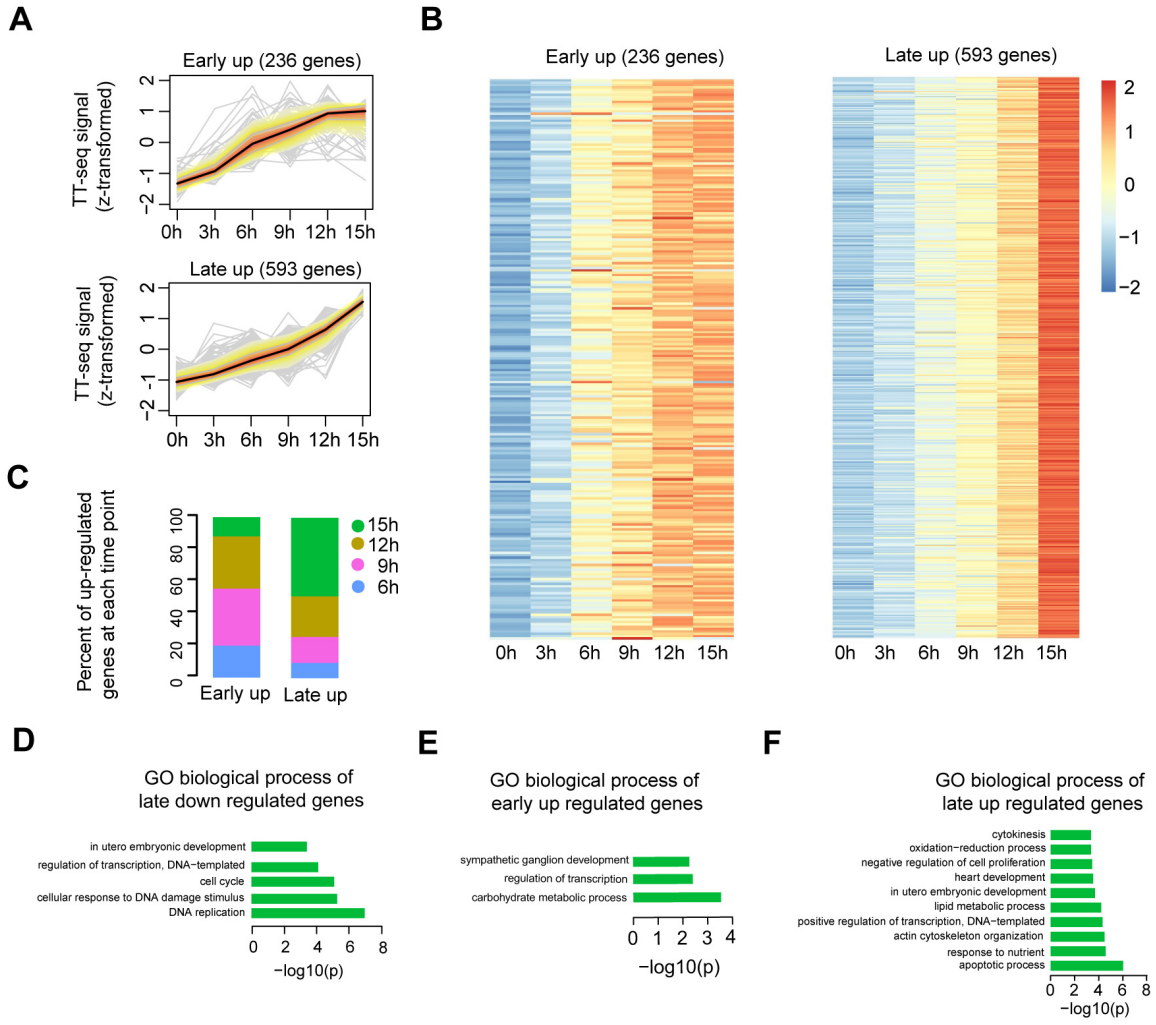


Figure 10. Transcription dynamics of early- and late-up regulated genes and their GO enrichment.

(A) Heatmap indicating the kinetics of early and late down-regulated genes. (B) K-means clustering of up-regulated genes divided into early (top) and late (bottom) up-regulated gene groups. Y axis indicates z-score transformed TT-seq counts. (C) Differentially regulated time point composition for early (left) and late (right) up-regulated gene groups. Y axis indicates percentage. (D-F) GO biological process enrichment for late down regulated genes (D), early up-regulated genes (E) and late up-regulated genes (F).

We then assessed if there is an enrichment of putative SE-controlled genes among the early down-regulated genes. Indeed, of the 150 transcribed genes that are nearest to SEs (Whyte et al., 2013), we found 60 genes to be significantly down-regulated, of which 45 were early down-regulated (**Figure 11A, P-value=2.6e-24, Fisher's exact test**). We then compared the kinetics of mRNA synthesis changes of the putatively SE-controlled down-regulated genes (60) to other down-regulated genes (709). This showed that mRNA synthesis of SE-controlled down-regulated genes was particularly sensitive to Oct4 depletion (**Figure 11B**). Among the

60 SE genes that were down-regulated, we found many pluripotency genes at early time points (**Figure 11C**). At 6 hours of Oct4 depletion we found a significant decrease of synthesis for *Esrrb*, *Klf2*, *Klf4*, *Utf1* and *Tbx3*. At 9 hours of depletion, *Sox2*, *Nanog* and *Prdm14* were significantly down-regulated, and *Nr5a2* and *Fgf4* after 12 hours. Taken together, our analysis of early mRNA synthesis changes upon Oct4 depletion revealed a rapid downregulation of the components of pluripotency transcriptional network with SE-controlled genes being immediately and strongly affected. Thus, consistent with previous findings (Whyte et al., 2013), Oct4 is strictly required to maintain the transcriptional network underlying pluripotency.

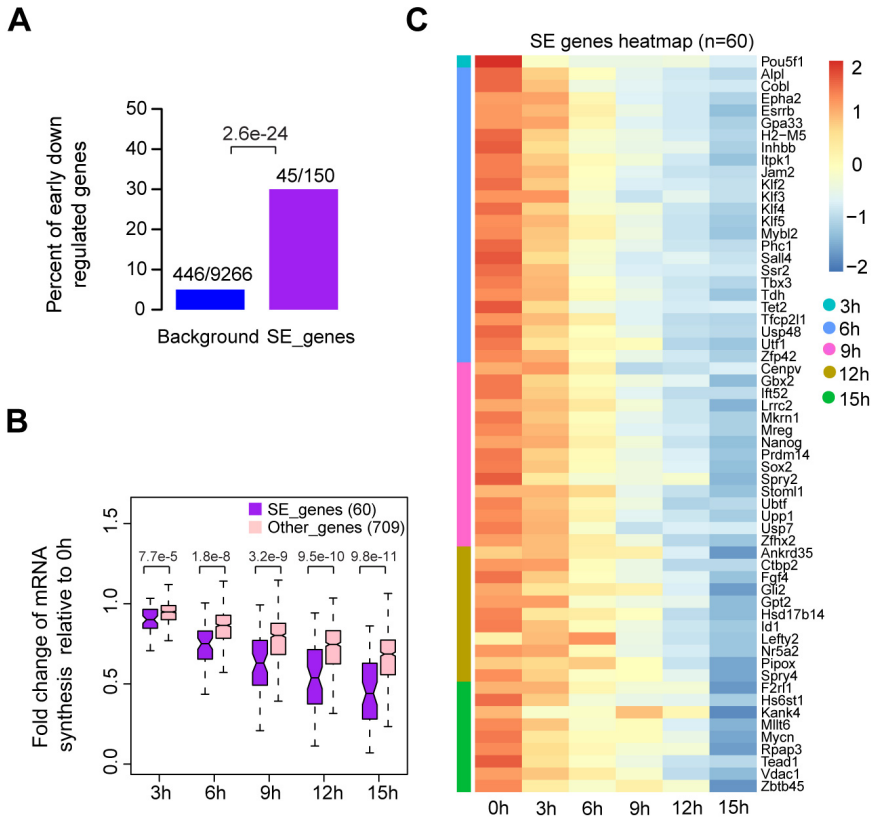


Figure 11. Down-regulation of SE gene networks upon Oct4 depletion.

(A) Bar graph depicting the percentage of early down-regulated mRNAs in all annotated mRNAs (blue, as background) and SE-controlled mRNAs (purple). SE annotation was obtained from (Whyte et al., 2013). P-value was calculated by Fisher's exact test. (B) Boxplot indicating the changes in mRNA synthesis for putative SE-controlled down-regulated genes (n=60, purple) versus other down-regulated genes (n=709, pink). Y axis indicates fold change of mRNA synthesis relative to 0 hours. P-values were calculated by Wilcoxon rank sum test. Black bars represent the median values for each group. Lower and upper boxes are the first and third quartiles, respectively. The ends of the whiskers extend the box by 1.5 times the interquartile range. Outliers are not drawn. (C) Heatmap indicating the kinetics of SE-controlled down-regulated genes (n=60). Genes were ordered by the corresponding time of significant down regulation.

3. Oct4-bound transcribed enhancers produce high levels of eRNA

To understand how loss of Oct4 leads to rapid destabilization of the pluripotency gene network, we combined our TT-seq and ATAC-seq data with published Oct4 ChIP-seq data (King and Klose, 2017) to derive a refined annotation of putative enhancers in mESCs (**Figure 12A**). First, we defined transcribed enhancers by annotating putative eRNAs (**Figure 12A**). We selected asRNAs and incRNAs that originated over 1 kb away from promoter-related RNAs (mRNA, conRNA and uaRNA) and merged those located less than 1 kb apart (**Figure 7C**). This resulted in 8,727 putative eRNAs, consisting of 2,468 asRNAs and 6,259 incRNAs, with a median length ~700 bp (**Figure 12B-C**). To annotate putative Oct4-regulated eRNAs, we used available Oct4 ChIP-seq data (King and Klose, 2017) (**Section II, 2.13, Table 5**). Most Oct4 ChIP-seq peaks (91%) overlapped with open chromatin regions identified by ATAC-seq (**Figure 12D**). Out of the 8,727 putative eRNAs, 2,221 overlapped with 2,231 Oct4-bound sites (Klf4 SE shown as an example in **Figure 12E**). We refer to these Oct4-bound sites as Oct4-bound transcribed enhancers (**Figure 12A**). The majority of Oct4-bound transcribed enhancers (90%) were marked by histone H3 lysine 4 mono-methylation (H3K4me1) (**Figure 12F**). Oct4-bound transcribed enhancers were strongly enriched for biological processes related to stem cell population maintenance (**Figure 12G**). In contrast, the Oct4-unbound transcribed enhancers (**Figure 12A**) were enriched for other biological processes (**Figure 12H**). Moreover, Oct4-regulated eRNAs originating from Oct4-bound enhancers were significantly longer and showed higher levels than other putative eRNAs (**Figure 12F**, P-value < 2.2e-16, Wilcoxon rank sum test).

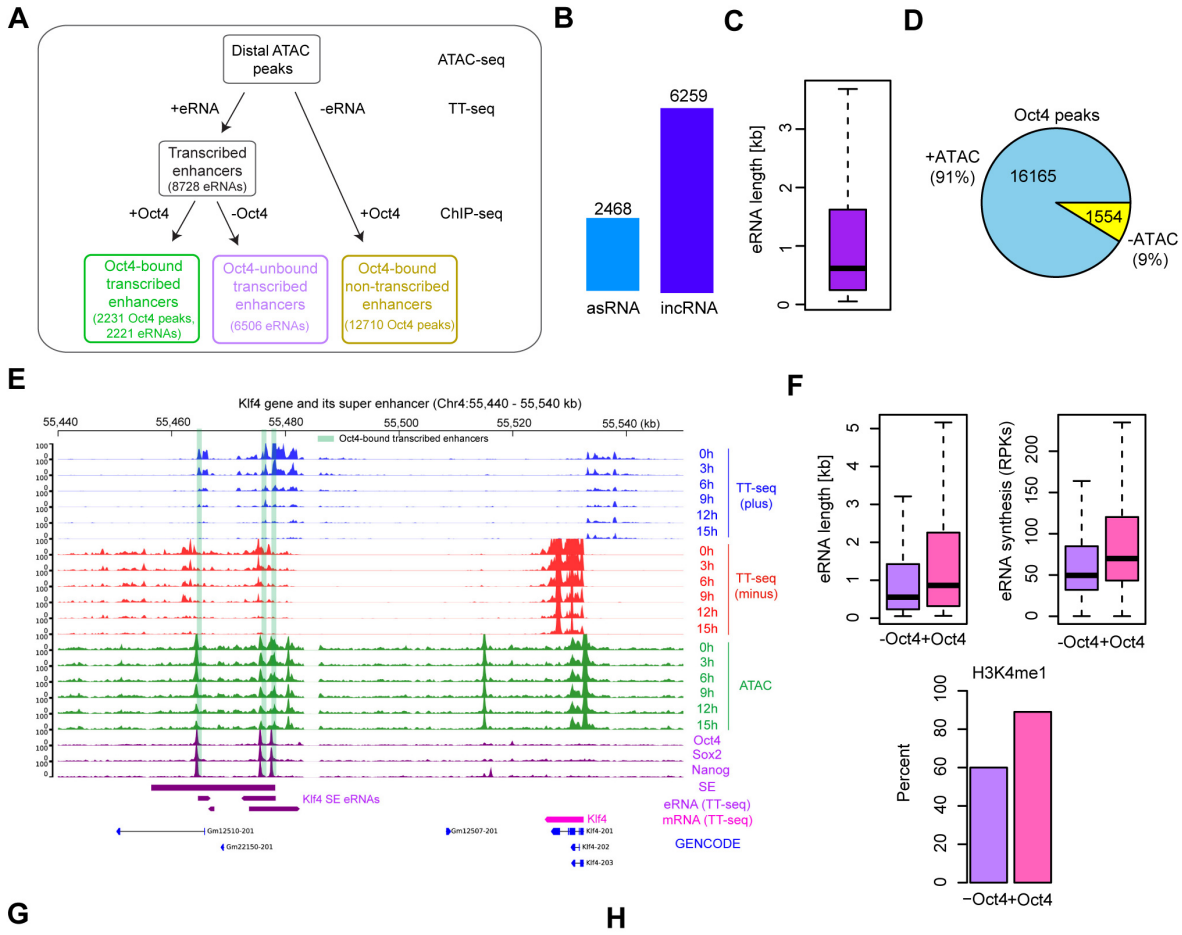


Figure 12. Annotation of putative Oct4-bound/regulated enhancer classes in mESCs.

(A) Diagram illustrating classification of Oct4 binding sites by combing ATAC-seq, TT-seq annotated eRNAs, and Oct4 ChIP-seq peaks. Oct4 ChIP-seq data at 0 hours DOX treatment was obtained from (King and Klose, 2017). (B) Total number of asRNAs and incRNAs classified as putative eRNAs. (C) Length distribution of putative eRNAs. (D) Percentage of Oct4 binding sites overlapped with chromatin open regions detected by our ATAC-seq data. Oct4 ChIP-seq data was obtained from (King and Klose, 2017). (E) Genome browser view of Oct4-bound transcribed enhancers at *Kfl4* SE. Tracks from top to bottom: TT-seq coverages of plus strand (blue), minus strand (red) and ATAC-seq coverages (green) at 0, 3, 6, 9, 12 and 15 hours; ChIP-seq coverages of Oct4, Sox2 and Nanog (purple) from ZHBTc4 mouse ES cell at 0 hours (King and Klose, 2017); SE annotation (Whyte et al., 2013); *Klf4* SE eRNAs annotated by TT-seq (purple arrow); *Klf4* gene (magenta arrow) annotated using TT-seq data; mouse GENCODE annotation for *Klf4* gene (blue). Biological replicates were merged for visualization. (F) Length, synthesis and H3K4me1 occupancy of Oct4-regulated eRNAs (n=2,221) versus other eRNAs (n=6,506). ChIP-seq data of H3K4me1 was obtained from (Chronis et al., 2017). (G-H) GO biological process enrichment of Oct4-regulated eRNAs (G) versus other eRNAs regions (H).

Of the remaining Oct4-bound accessible sites, 1,098 produced mRNAs, and 12,710 produced no detectable RNAs and were referred to as Oct4-bound non-transcribed enhancers (**Figure 12A, 13A**). We then performed metagene analysis for TF enrichment at Oct4-bound transcribed and non-transcribed enhancers using published data (Chronis et al., 2017; King and Klose, 2017) (**Section II, 2.13, Table 5**). Whereas both groups of enhancers showed similar H3K4me1 levels, Oct4-bound transcribed enhancers showed an enrichment with active histone marks H3K27ac and H3K4me3, higher chromatin accessibility, and higher occupancies with Oct4, Sox2, Nanog, Klf4 and Esrrb (**Figure 13B**). Moreover, Oct4-bound transcribed enhancers were located closer to their nearest active putative target genes (median distance of 37 kb) as compared to non-transcribed enhancers (median distance 89 kb) (**Figure 13C**). Finally, we investigated eRNA synthesis at SEs. Half of Oct4-bound sites within SEs produced eRNAs (**Figure 13D**). The eRNAs obtained from SEs were generally longer and had higher synthesis rates compared to eRNAs from typical enhancers (TEs), and SEs showed higher occupancy with Oct4, Sox2, Nanog, Klf4 and Esrrb (**Figure 13E-F**). These efforts led to an enhancer annotation in mESCs and suggested that Oct4-bound enhancers are transcriptionally more active than other enhancers.

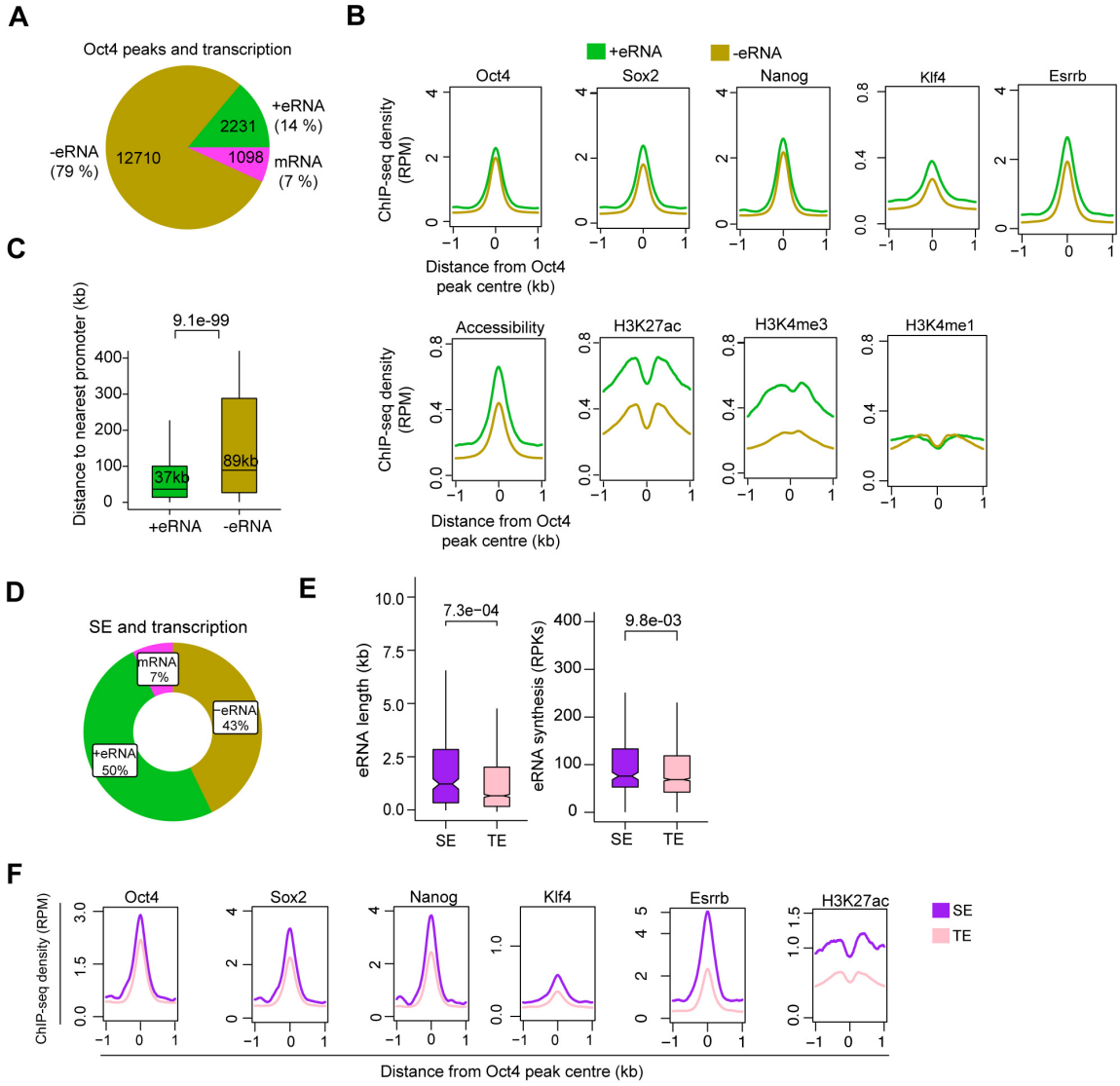


Figure 13. Characterization of Oct4 bound and unbound transcribed enhancers.
(A) Pie chart indicating the overlap of Oct4 binding sites with regions of active transcription (eRNA or mRNA) or no transcription annotated by TT-seq. (B) Metagene plot showing the occupancy for transcription factors Oct4, Sox2, Nanog, Klf4, Esrrb, chromatin accessibility, and H3K27ac and H3K4me3 histone modifications at Oct4-bound transcribed enhancers (n=2,231) and Oct4-bound non-transcribed enhancers (n=12,710). Y axis depicts ChIP-seq coverage density in reads per million (RPM). ChIP-seq data of Oct4, Sox2 and Nanog were obtained from (King and Klose, 2017), Klf4, Esrrb, H3K27ac and H3K4me3 histone modifications data were obtained from (Chronis et al., 2017). (C) Genomic distance distributions of Oct4-bound transcribed enhancers to their nearest gens versus Oct4-bound non-transcribed enhancers to their nearest genes. P-values were calculated by Wilcoxon rank sum test. (D) Pie chart depicting the overlap of Oct4-bound sites (n=514) at SE that show eRNA transcription (n=256, 50%), mRNA transcription (n=38, 7%) or no transcription (n=220, 43%) by TT-seq. (E) Length and synthesis (reads per kilobases, RPKs) distributions of SE eRNAs (n=243) versus TE eRNAs (n=1,978). P-value were calculated by Wilcoxon rank sum test. (F) Metagene analysis of Oct4, Sox2, Nanog, Klf4, Esrrb occupancy and H3K27ac histone modification at Oct4 occupied TE versus SE. ChIP-seq data of Oct4, Sox2 and Nanog were obtained from (King and Klose, 2017)and Klf4, Esrrb and H3K27ac data from (Chronis et al., 2017). Y axis depicts ChIP-seq coverage density in reads per million (RPM).

4. Oct4 is often required for enhancer transcription

We next analyzed Oct4-bound transcribed enhancers (**Figure 15A**) with respect to changes in their eRNA synthesis upon Oct4 depletion. Synthesis of eRNAs was highly reproducible between the two biological replicates (**Figure 14A**). PCA analysis revealed that the changes of Oct4-regulated eRNA synthesis followed a similar trajectory to that seen for mRNAs (**Figure 14B, Figure 9A**). Differential expression analysis of eRNAs (Love et al., 2014) detected significant down-regulation of 782 Oct4-regulated eRNAs after 15 hours of DOX treatment (**Figure 14C, 15A, adjusted P-value = 0.01**).

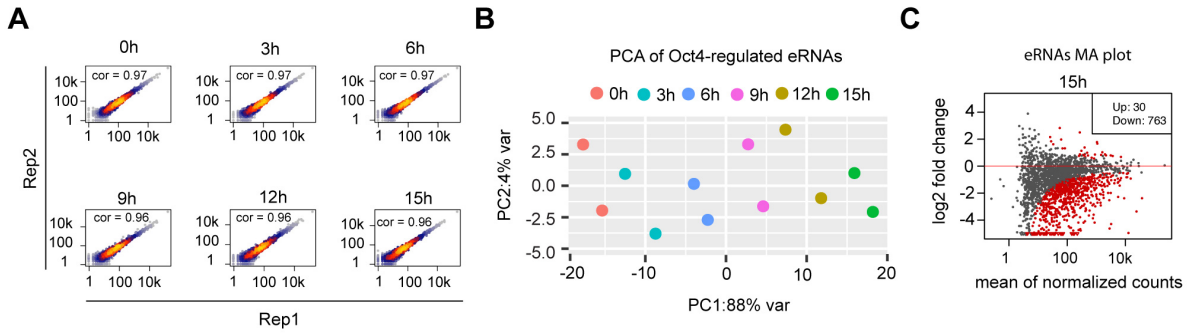


Figure 14. eRNA synthesis changes at Oct4-bound transcribed enhancers.

(A) Scatter plots showing correlation between two biological replicates of TT-seq data for Oct4-regulated eRNAs (n=2,221) for each time point of DOX treatment. Spearman's rank correlation coefficient was calculated and shown in each plot. (B) Principle component analysis of Oct4-regulated eRNA synthesis changes over time-course of DOX treatment (0, 3, 6, 9, 12, 15 hours) for two biological replicates. (C) MA plots showing differentially regulated Oct4-regulated eRNAs detected by DESeq2 after 15 hours of DOX treatment. Statistically significantly regulated eRNAs are depicted in red.

The kinetics analysis showed that for the down-regulated eRNAs synthesis decreased already after 3 hours (**Figure 15B**). Moreover, SE eRNAs were more strongly down-regulated compared to TE eRNAs (**Figure 15C-D**). Taken together, these results suggest that Oct4 is required for eRNA synthesis at about one third of putative Oct4-bound transcribed enhancers including SEs.

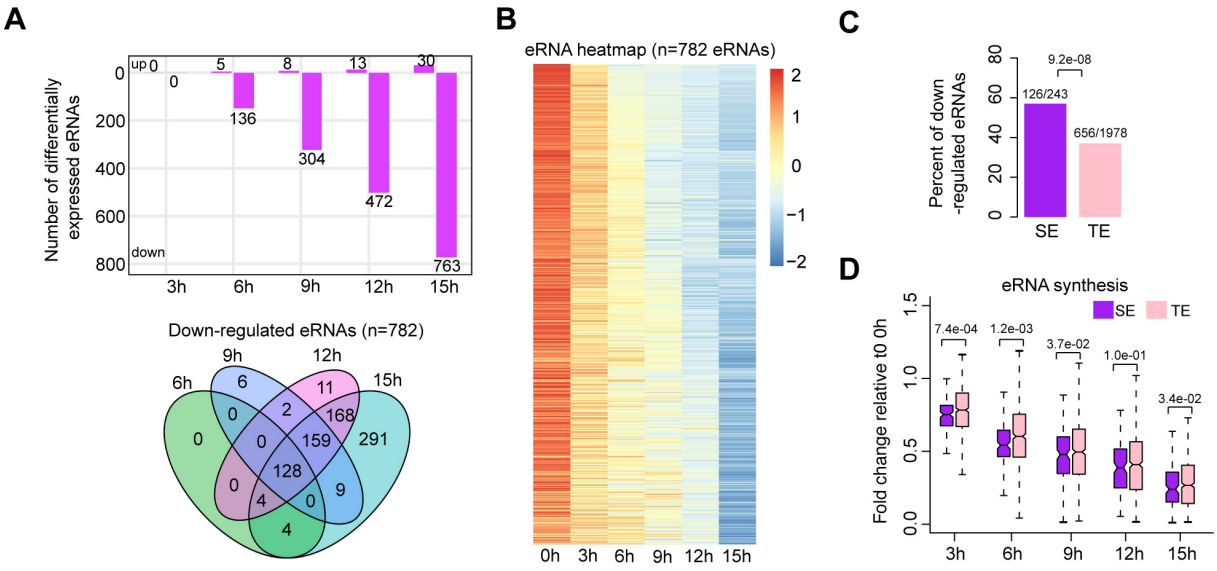


Figure 15. Oct4 is required for enhancer transcription.

(A) Number of differentially expressed Oct4-regulated eRNAs detected by DESeq2 (top) and Venn diagram showing overlapping of differentially expressed Oct4-regulated eRNA (bottom) at each time point of DOX treatment. (B) Heatmap visualizing the kinetics of Oct4-regulated down-regulated eRNAs (n=782). (C) Bar chart indicating the percentage of down-regulated eRNAs at SE and TE. P-value was calculated by Fisher's exact test. (D) Boxplot indicating synthesis changes of down-regulated eRNAs at SE (n=126) versus TE (n=656). Y axis indicates fold change of eRNAs synthesis relative to 0 hours. P-values were calculated by Wilcoxon rank sum test. Black bars represent the median values for each group. Lower and upper boxes are the first and third quartiles, respectively. The ends of the whiskers extend the box by 1.5 times the interquartile range. Outliers are not drawn.

5. Oct4 binds enhancers to activate putative target genes

To investigate whether Oct4 depletion leads to a loss of Oct4 binding to enhancers, we performed ChIP-seq of Oct4 after 0, 3, 6, 9, 12 and 15 hours of DOX treatment (**Section II, 2.13, Table 3**). In agreement with chromatin fractionation results (**Figure 7A**, **chromatin**), Oct4 occupancy decreased after 3 hours and strongly dropped after 9 hours of DOX treatment (**Figure 16A**). This is consistent with the observed decrease in eRNA synthesis (**Figure 15B**). In accordance with our previous results (**Figure 15D**), occupancy of Oct4 decreased more strongly at SEs compared to TEs (**Figure 16B**). These results show that Oct4 binding is required for eRNA synthesis at a subset of Oct4-bound transcribed enhancers and particularly at SEs.

To investigate whether the observed decrease in Oct4 occupancy and eRNA synthesis at Oct4-bound transcribed enhancers coincided with a decrease of target mRNA synthesis, we then paired transcriptionally down-regulated SEs with their nearest transcribed genes and kept those pairs with down-regulated genes. This resulted in 62 enhancer-gene pairs. For them, we

found that eRNA and mRNA synthesis decreased already after 3 hours of DOX treatment and occurred with similar trajectories over the entire time course, coinciding with a decrease in Oct4 occupancy (**Figure 16C**). This is further illustrated for *Klf4* and *Sox2* genes. Synthesis of both genes decreased at 3 hours and changes over time correlated well with changes of Oct4 occupancy and eRNA synthesis at their well-studied SEs (**Figure 16D-E**). These results are consistent with a function of Oct4 in activating both enhancer transcription and mRNA synthesis from putative target genes.

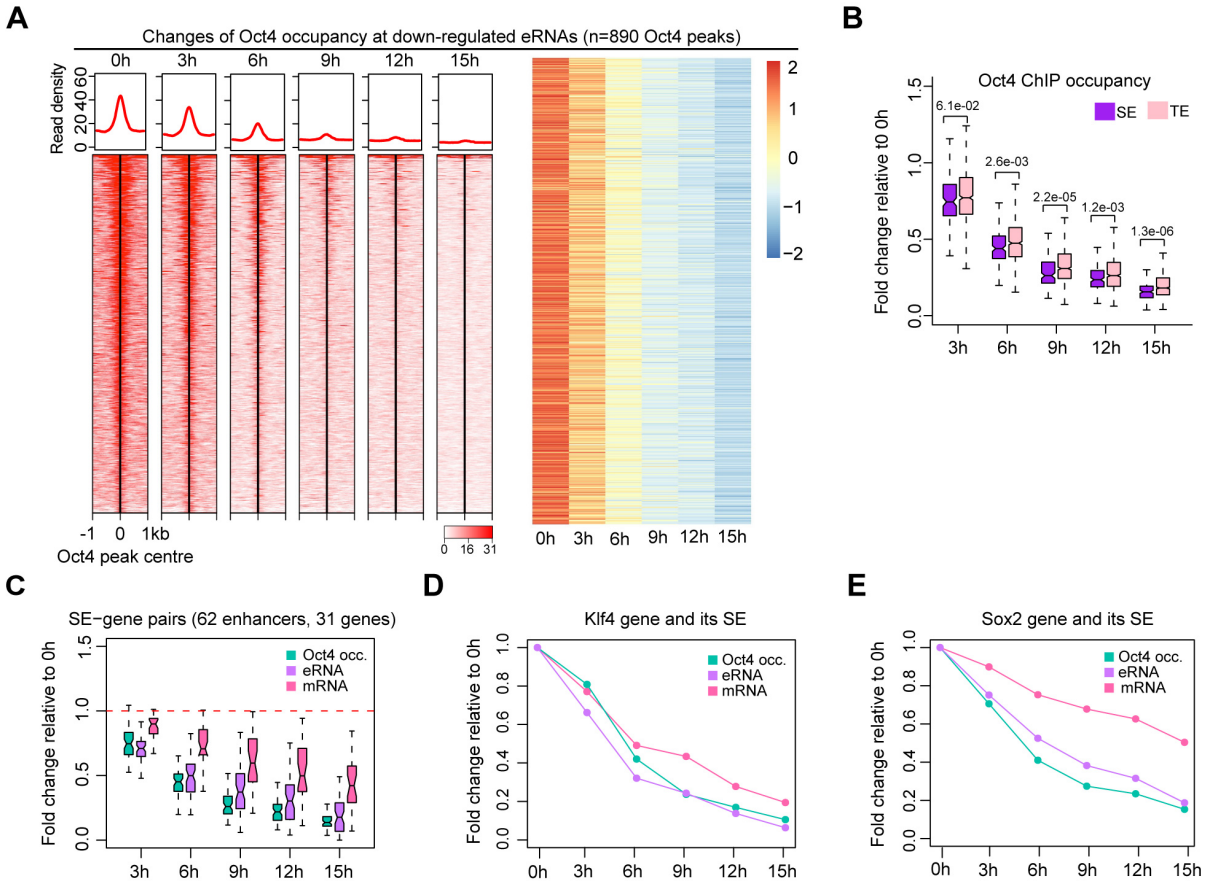


Figure 16. Oct4 is required for gene transcription.
(A) Changes of Oct4 occupancy at down-regulated eRNAs, illustrated by Oct4 ChIP-seq coverage (left) and count heatmaps (right). 782 down-regulated eRNAs were originated from 890 Oct4-bound transcribed enhancers (peaks). Normalized read densities were presented, in left, peaks were ranked by Oct4 read densities. (B) Boxplot indicating the corresponding Oct4 occupancy changes at down-regulated SE versus TE eRNAs defined in Figure 17 (D). Y axis indicates fold change of Oct4 occupancy relative to 0 hours. P-values were calculated by Wilcoxon rank sum test. (C) Boxplot indicating changes of Oct4 occupancy, eRNA and mRNA synthesis for 62 SE-gene pairs. Y axis indicates fold change relative to 0h. Transcriptionally down-regulated SEs were paired with their nearest transcribed genes and pairs with down-regulated genes were kept. Enhancers number were counted by individual Oct4 peaks within SEs. (D) Fold changes of Oct4 occupancy, eRNA and mRNA synthesis at *Klf4* gene and its associated SE. Fold changes of Oct4 occupancy and eRNA synthesis at SE were represented by average of the individual enhancers within the SE (illustrated as IGV track in **Figure 18C**). (E) Fold changes of Oct4 occupancy, eRNA and mRNA synthesis at *Sox2* gene and its associated

SE. Fold changes of Oct4 occupancy and eRNA synthesis at SE were represented by average of the individual enhancers within the SE (illustrated as IGV track in **Figure 18D**).

6. Oct4 binding does not directly correlate with enhancer accessibility

We next investigated changes in chromatin accessibility at Oct4-bound transcribed enhancers. Principal component analysis indicated that accessibility changes started to occur after 6 hours of DOX treatment, with substantial changes until 12 hours and followed by a subtle change at 15 hours (**Figure 17A**). To call significantly changed accessible chromatin regions we used DESeq2 (Love et al., 2014). In contrast to mRNAs and eRNAs (**Figure 9C, Figure 15A**), only few enhancers were detected to have significantly altered chromatin accessibility at 6 hours (**Figure 17B-C**). 15 hours of DOX treatment resulted in a significant decrease of chromatin accessibility at 726 enhancers (adjusted P-value = 0.01) (**Figure 17B, D**). The kinetic analysis showed that for these enhancer, chromatin accessibility remained unchanged at 3 hours (**Figure 17D**). These results show that decreased Oct4 binding does not directly correlate with chromatin accessibility changes, which occur later than Oct4 depletion. This delayed effect on chromatin accessibility argues that the primary function of Oct4 is not to render or keep chromatin accessible.

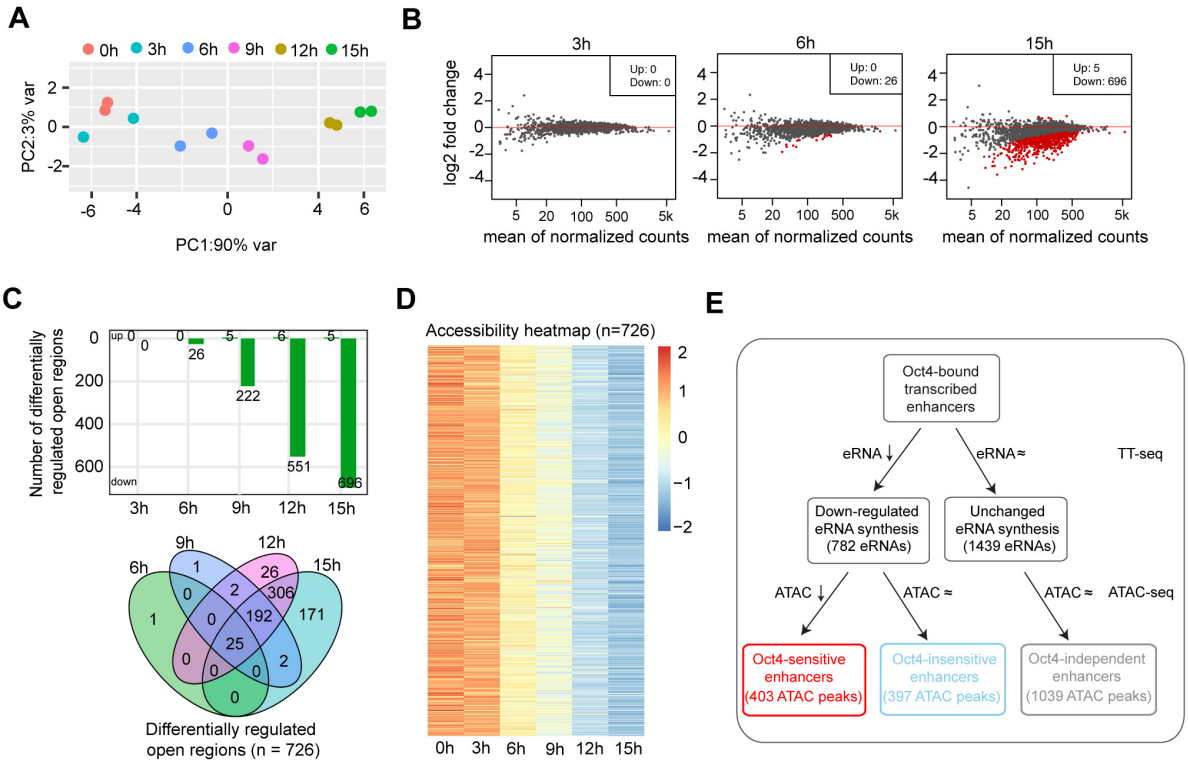


Figure 17. Change of enhancer accessibility is delayed upon Oct4 depletion.

(A) Principle component analysis of chromatin accessibility changes at Oct4-bound transcribed enhancers (n=2,223) over time course of DOX treatment (0, 3, 6, 9, 12, 15 hours) for two biological replicates. (B) MA plots showing differentially regulated chromatin open regions detected by DESeq2 at 3, 6 and 15 hours DOX treatment for Oct4-bound transcribed enhancers. Statistically significant features are depicted in red. (C) Number of differentially regulated chromatin open regions detected by DESeq2 for Oct4-bound transcribed enhancers (top) and Venn diagram showing overlapping of detected differentially regulated open regions (bottom) for each time point of DOX treatment. (D) Heatmap visualizing the kinetics of chromatin accessibility changes at Oct4-occupied accessibility decreased sites (n=726). (E) Diagram indicating classification of Oct4-sensititive, -insensitive and -independent enhancers defined by change of eRNA synthesis and chromatin accessibility at Oct4-bound transcribed enhancers.

To further investigate whether Oct4 may have roles in altering chromatin, we classified the down-regulated Oct4-bound transcribed enhancers (**Figure 15B)** based on their respective changes in chromatin accessibility. We defined two groups of down-regulated transcribed enhancers, showing either decreased or unchanged chromatin accessibility (**Figure 17E**). In both groups, Oct4 binding to chromatin decreased over the time course and was associated with a decrease in eRNA synthesis (**Figure 18A-B**). For the first group, depletion of Oct4 led to a decrease in eRNA synthesis and a reduction of chromatin accessibility, with down regulation of eRNA synthesis preceding the decrease in accessibility (**Figure 18A**). This is illustrated for the SE near *Klf4* gene, which contains three enhancers of this group (**Figure 18C)**. Synthesis of *Klf4* gene decreased at 3 hours and occurred simultaneously with decrease of Oct4 occupancy and eRNA synthesis at its SE **(Figure 16D)**, whereas the corresponding chromatin accessibility started to decrease at 9 hours (**Figure 18C)**. For the second group, depletion of Oct4 led to a decrease in eRNA synthesis without changes in chromatin accessibility (**Figure 18B**). This is illustrated by the *Sox2* gene SE and *Mir290* SE (Whyte et al., 2013)(**Figure 18D-E**). The remaining Oct4-bound transcribed enhancers showed no changes in eRNA synthesis and chromatin accessibility upon Oct4 depletion (**Figure 18F**). We hereafter refer the three groups of Oct4-bound transcribed enhancers as Oct4-sensitive, Oct4-insensitive and Oct4-independent enhancers (**Figure 17E**). In conclusion, these results indicate that Oct4 binding to putative enhancers does not directly correlate with chromatin accessibility.

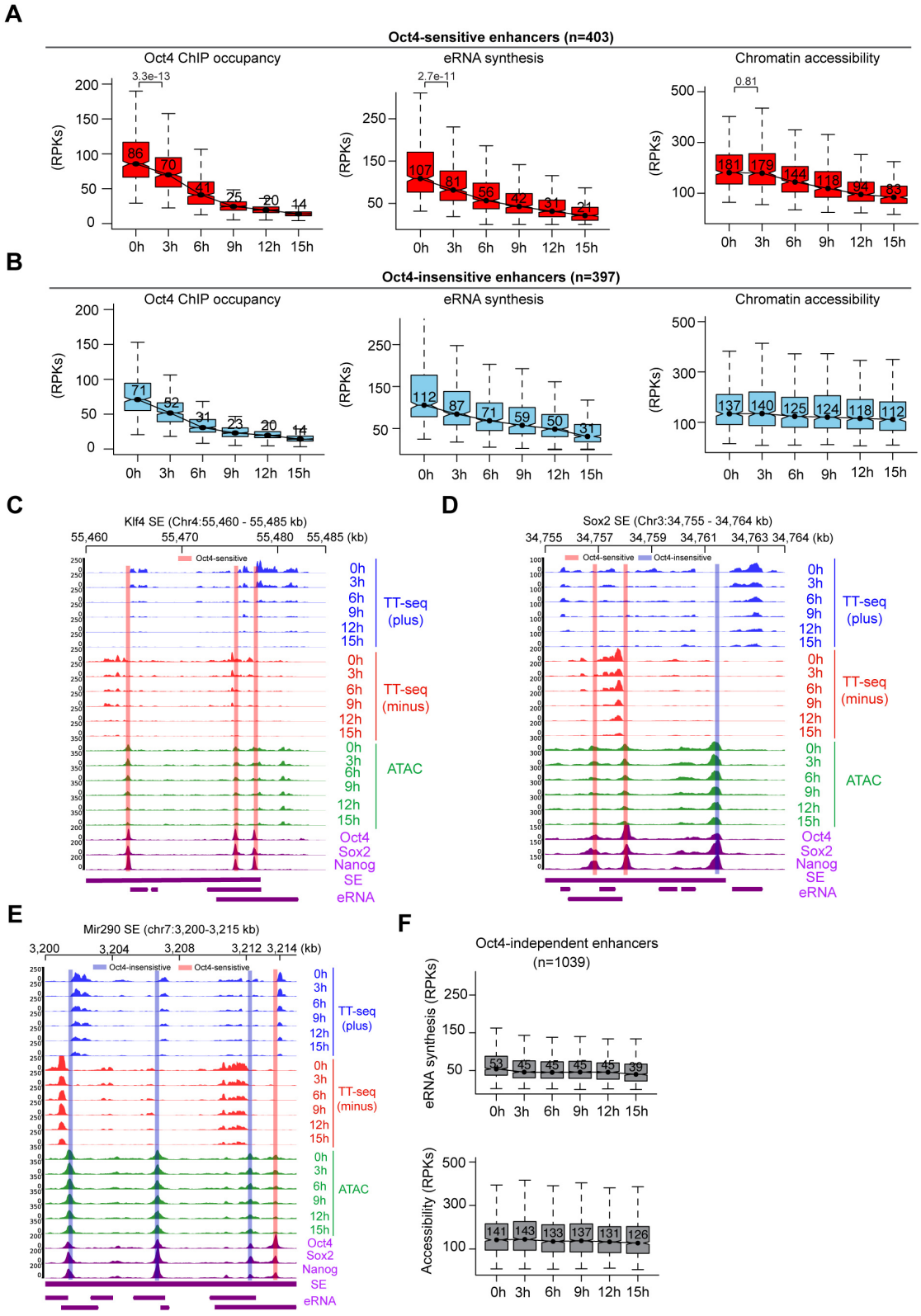


Figure 18. Oct4 binding does not correlate with enhancer accessibility changes.

(A) Boxplots indicating the changes of Oct4 occupancy, eRNA synthesis and chromatin accessibility at Oct4-sensitive enhancers (n=403). P-values were calculated by Wilcoxon rank sum test. Y axis represents read counts per kilobases (RPKs). Black bars represent the median values for each group. Lower and upper boxes are the first and third quartiles, respectively. The ends of the whiskers extend the box by 1.5 times the interquartile range. Outliers are omitted. (B) Boxplots indicating the changes of Oct4 occupancy, eRNA synthesis and chromatin accessibility at Oct4-insensitive enhancers (n=397).

(C) Genome browser view for changes of eRNA synthesis and chromatin accessibility at *Klf4* SE including three Oct4-sensitive enhancers. Tracks from top to bottom: TT-seq coverages of plus strand (blue), minus strand (red) and ATAC-seq coverages (green) at 0, 3, 6, 9, 12 and 15 hours; ChIP-seq coverages of Oct4, Sox2 and Nanog (purple) from ZHBTc4 mouse ES cell at 0 hours (King and Klose, 2017); SE annotation (Whyte et al., 2013); *Klf4* SE eRNAs annotated by TT-seq (purple arrow). Biological replicates were merged for visualization. (D) Genome browser view for changes of eRNA synthesis and chromatin accessibility at *Sox2* SE including two Oct4-sensitive enhancers and one Oct4-insensitive enhancer. Tracks were ordered in the same way as in (C). (E) Genome browser view for changes of eRNA synthesis and chromatin accessibility at *Mir290* SE including three Oct4-insensitive enhancers and one Oct4-sensitive enhancer. Tracks from top to bottom: TT-seq coverages of plus strand (blue), minus strand (red) and ATAC-seq coverages (green) at 0, 3, 6, 9, 12 and 15 hours; ChIP-seq coverages of Oct4, Sox2 and Nanog (purple) from ZHBTc4 mouse ES cell at 0 hours (King and Klose, 2017); SE annotation (Whyte et al., 2013); *Mir290* SE eRNAs annotated by TT-seq (purple arrow). Biological replicates were merged for visualization. (F) Boxplots showing the changes of eRNA synthesis and chromatin accessibility at Oct4-independent enhancers (n=1,039).

7. Sox2 may contribute to retained enhancer accessibility upon Oct4 depletion

We then asked whether Sox2 may contribute to the delayed decrease in chromatin accessibility observed at Oct4-sensitive enhancers (**Figure 17E**). We performed ChIP-seq of Sox2 over the same time course (**Section II, 2.13, Table 4**). At Oct4-sensitive enhancers, Sox2 remained bound from 0 to 6 hours and started to decrease after 9 hours of treatment (**Figure 19A-B**). At Oct4-insensitive and Oct4-independent enhancers, Sox2 occupancy was stable over the entire time course (**Figure 19A-B, 20A-B**). Moreover, we analyzed published Oct4, Sox2 and Nanog ChIP-seq data at 0 and 24 hours after DOX treatment (King and Klose, 2017). Oct4-sensitive enhancers showed a strong loss of all three TFs at 24 hours, whereas Sox2 occupancy was unchanged and Nanog occupancy increased at 24 hours at Oct4-insensitive enhancers and Oct4-independent enhancers (**Figure 20C-E**).

These findings are well illustrated at the exemplary genomic regions comprising putative SEs of *Klf4*, *Sox2* and *Mir290* (**Figure 19C-D, 20F**). Within these regions, we observed a decrease of chromatin accessibility and Sox2 occupancy at the Oct4-sensitive enhancers after 9 hours of DOX treatment (**Figure 19C-D, 20F**), whereas the Oct4-insensitive enhancers remained accessible and occupied by Sox2 (**Figure 19D, 20F**). Taken together, these findings suggest that Sox2 is involved in temporarily keeping chromatin open during the first 6 hours of DOX treatment for Oct4-sensitive enhancers.

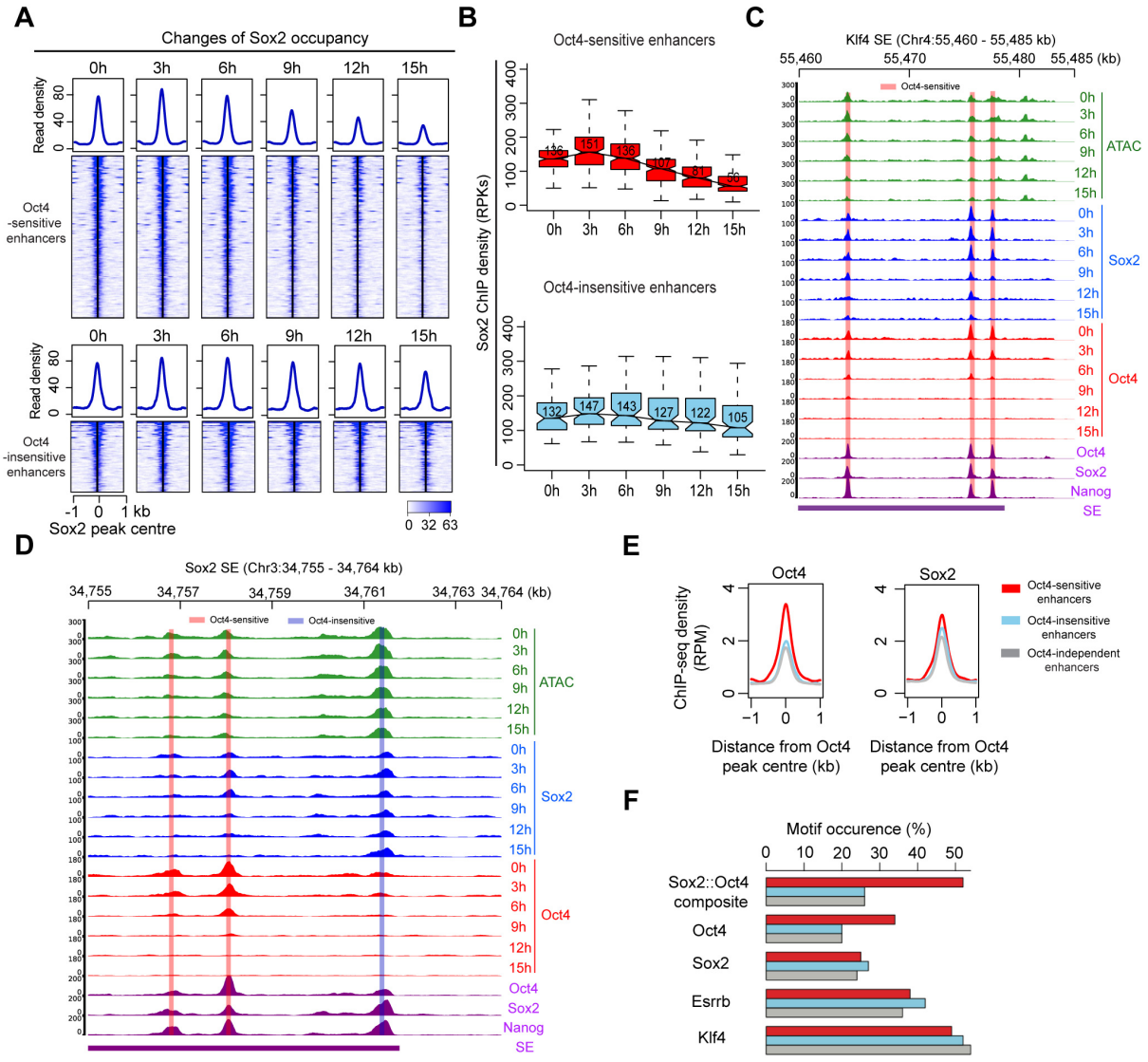


Figure 19. Sox2 may contribute to retained enhancer accessibility upon Oct4 depletion.

(A) Heatmap showing changes of Sox2 occupancy at Oct4-sensitive and Oct4-insensitive enhancers over the entire time course of DOX treatment. Normalized read densities were presented, peaks were ranked by Sox2 read densities. (B) Same as (A), but using boxplots to depict quantification of Sox2 occupancy changes at Oct4-sensitive and Oct4-insensitive enhancers. Y axis represents Sox2 ChIP-seq density in reads per kilobases (RPKs). Black bars represent the median values for each group. Lower and upper boxes are the first and third quartiles, respectively. The ends of the whiskers extend the box by 1.5 times the interquartile range. Outliers are omitted. (C) Genome browser view illustrating changes of chromatin accessibility, Sox2 and Oct4 occupancy at *Klf4* SE. Tracks from top to bottom: ATAC-seq coverages (green), ChIP-seq coverages for Sox2 (blue) and Oct4 (red) at 0, 3, 6, 9, 12 and 15 hours; ChIP-seq coverages for Oct4, Sox2 and Nanog (purple) from ZHBTc4 mouse ES cell at 0 hours (King and Klose, 2017); SE annotation (Whyte et al., 2013). Biological replicates were merged for visualization. (D) Genome browser view illustrating changes of chromatin accessibility, Sox2 and Oct4 occupancy at *Sox2* SE. Tracks were visualized and ordered in the same way as in (C). (E) Metagenes analysis of Oct4 and Sox2 occupancy at Oct4-sensitive, -insensitive and -independent enhancers at 0 hours, data were obtained from (King and Klose, 2017). Y axis depicts ChIP-seq coverage density in

reads per million (RPM). (F) Percentage of motif occurrence at Oct4-sensitive, -insensitive and -independent enhancers for Sox2-Oct4 composite motif and Oct4, Sox2, Esrrb and Klf4 motifs.

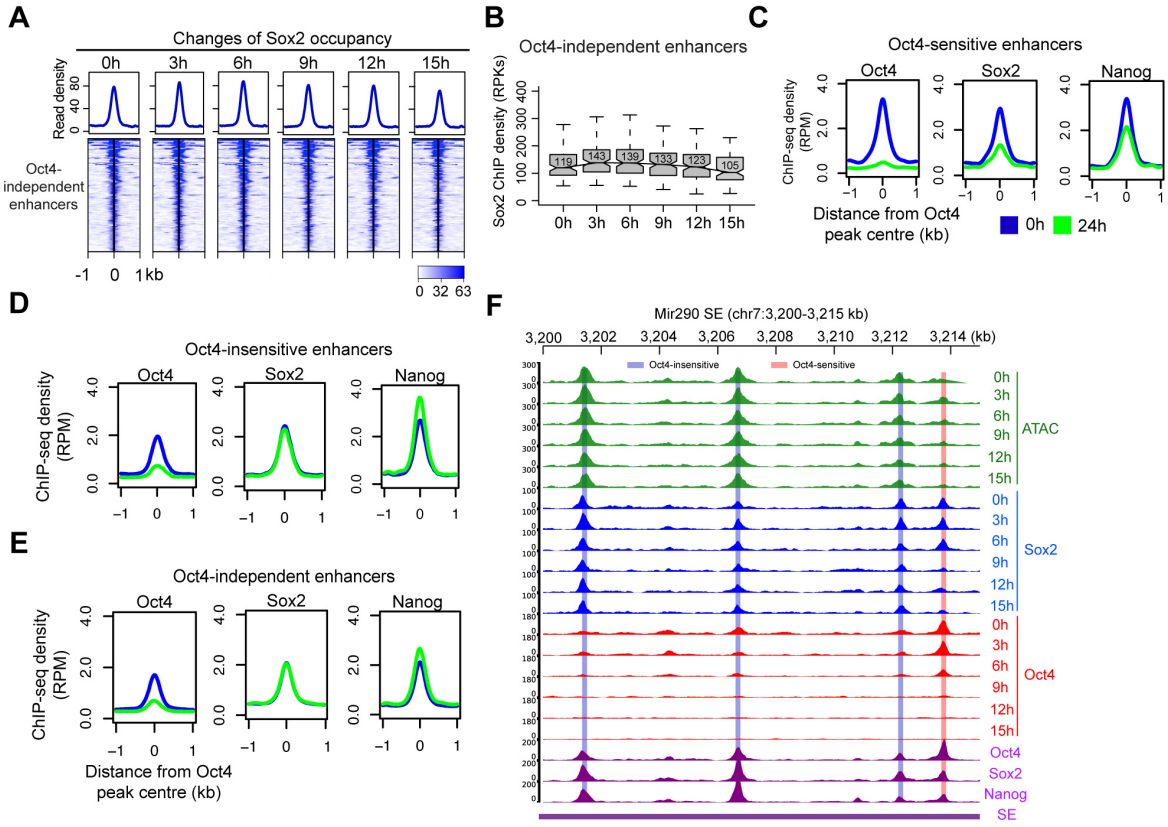


Figure 20. Characterization of Oct4-sensitive, -insensitive and -independent enhancers.

(A) Heatmap showing changes of Sox2 occupancy at Oct4-independent enhancers over the entire time course of DOX treatment. Normalized read densities were presented, peaks were ranked by Sox2 read densities. (B) Same as (A), but using boxplot to depict quantification of Sox2 occupancy changes at Oct4-independent enhancers. Y axis represents Sox2 ChIP-seq density in reads per kilobases (RPKs). (C) Metagenes analysis of Oct4, Sox2 and Nanog occupancy after 0 and 24 hours of DOX treatment at Oct4-sensistive enhancers. ChIP-seq data were obtained from (King and Klose, 2017). Y axis depicts ChIP-seq coverage density in reads per million (RPM). (D) Metagenes analysis of Oct4, Sox2 and Nanog occupancy after 0 and 24 hours of DOX treatment at Oct4-insensistive enhancers. (E) Metagenes analysis of Oct4, Sox2 and Nanog occupancy after 0 and 24 hours of DOX treatment at Oct4-independent enhancers. (F) Genome browser view illustrating changes of chromatin accessibility, Sox2 and Oct4 occupancy at *Mir290* SE. Tracks from top to bottom: ATAC-seq coverages (green), ChIP-seq coverages for Sox2 (blue) and Oct4 (red) at 0, 3, 6, 9, 12 and 15 hours; ChIP-seq coverages for Oct4, Sox2 and Nanog (purple) from ZHBTc4 mouse ES cell at 0 hours (King and Klose, 2017); SE annotation (Whyte et al., 2013). Biological replicates were merged for visualization.

8. Oct4 may cooperate with Sox2 to render enhancers accessible

To further characterize the differences between Oct4-sensitive and Oct4-insensitive enhancers, we analyzed publicly available ChIP-seq data. Oct4 and Sox2 co-localize in both enhancer

groups, with 95% and 86% of Oct4-sensitive and Oct4-insensitive enhancers, respectively, overlapping with Sox2 peaks (**Figure 21A**). In metagene plots, Oct4-sensitive enhancers showed ~1.5-fold enrichment of Oct4 occupancy compared to Oct4-insensitive enhancers (**Figure 19E**), whereas Sox2, Nanog, Klf4 and Esrrb were only slightly enriched if at all (**Figure 19E, 21B**). Oct4-sensitive enhancers also displayed higher levels of H3K27ac, whereas H3K4me1 showed similar levels (**Figure 21B**). Oct4-independent enhancers showed lower signals for pluripotency TFs and histone modifications, in line with the observed low transcriptional activity (**Figure 19E, 21B**). According to genomic region enrichment analysis (McLean et al., 2010), Oct4-sensitive enhancers were enriched for stem cell population maintenance (**Figure 21C**) and Oct4-insensitive enhancers for neural differentiation and development (**Figure 21D**). Enhancers of both types may target the same nearest active gene (**Figure 19D, 21E**).

To investigate whether a specific binding motif may be related to the enrichment of Oct4 occupancy in Oct4-sensitive enhancers, we performed motif analysis. Our results showed a strong enrichment for both Oct4 and the canonical composite DNA motifs at Oct4-sensitive enhancers only (**Figure 19F**), whereas no difference was found for other TFs. These findings suggest that Oct4 influences chromatin accessibility preferentially when the Sox2-Oct4 composite motif is present in DNA.

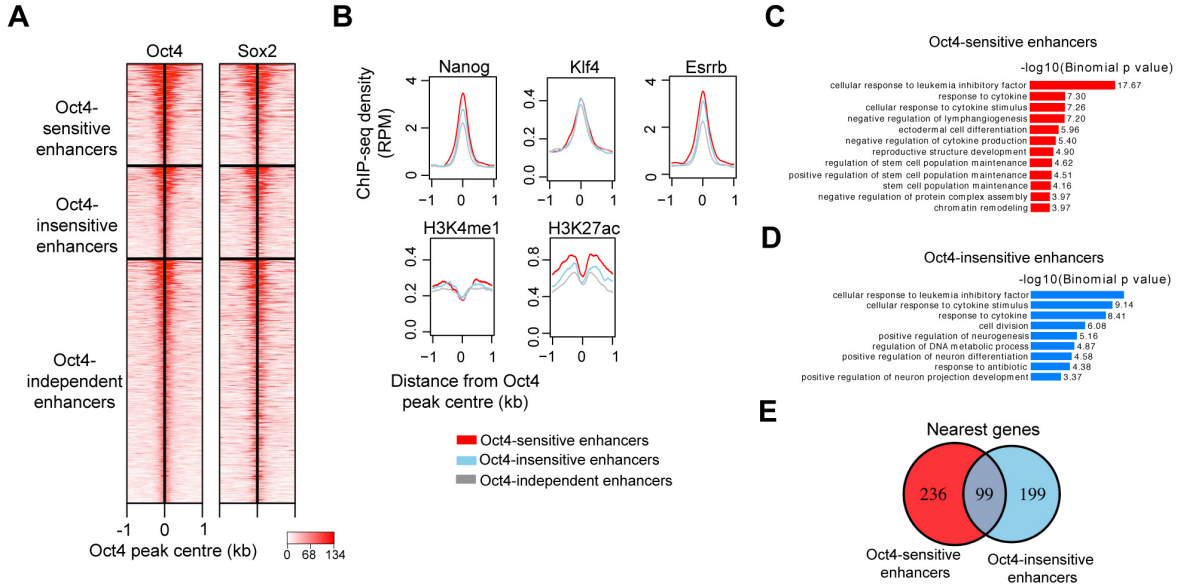


Figure 21. Characterization of Oct4-sensitive, -insensitive and -independent enhancers.
(A) Heatmap showing co-occupancy of Oct4 and Sox2 at Oct4-sensitive, -insensitive and -independent enhancers at 0 hours. ChIP-seq data were obtained from (King and Klose, 2017). (B) Metagenes analysis of Nanog, Klf4, Esrrb occupancy and H3K4me1, H3K27ac histone modifications at Oct4-sensitive, -insensitive and -independent enhancers at 0 hours. Nanog ChIP-seq data was obtained from (King and

Klose, 2017), other ChIP-seq data were obtained from (Chronis et al., 2017). (C)GO biological process enrichment of Oct4-sensitive enhancers. (D)GO biological process enrichment of Oct4-insensitive enhancers. (E) Overlap of nearest target genes of Oct4-sensitive and Oct4-insensitive enhancers.

9. Sox2 maintains chromatin accessibility in the absence of eRNA synthesis

To gain insights into the change of chromatin accessibility at Oct4-bound sites without detected eRNA synthesis, we analyzed the 12,710 Oct4-bound non-transcribed enhancers (**Figure 22A-B**). PCA analysis showed similar transitions for these enhancers over the time course as observed for Oct4-bound transcribed enhancers (**Figure 22A**). After 15 hours of DOX treatment, we detected 4,985 enhancers with significantly reduced chromatin accessibility (**Figure 22B-D, adjusted P-value = 0.01**). We refer to them as Oct4-sensitive enhancers and the remaining ones as Oct4-insensitive enhancers (**Figure 22E**). For both enhancer groups, Oct4 occupancy decreased already after 3 hours of DOX treatment (**Figure 22F**), and for Oct4-sensitive enhancers the Oct4 occupancy decrease preceded the decrease in chromatin accessibility (**Figure 22F, top**).

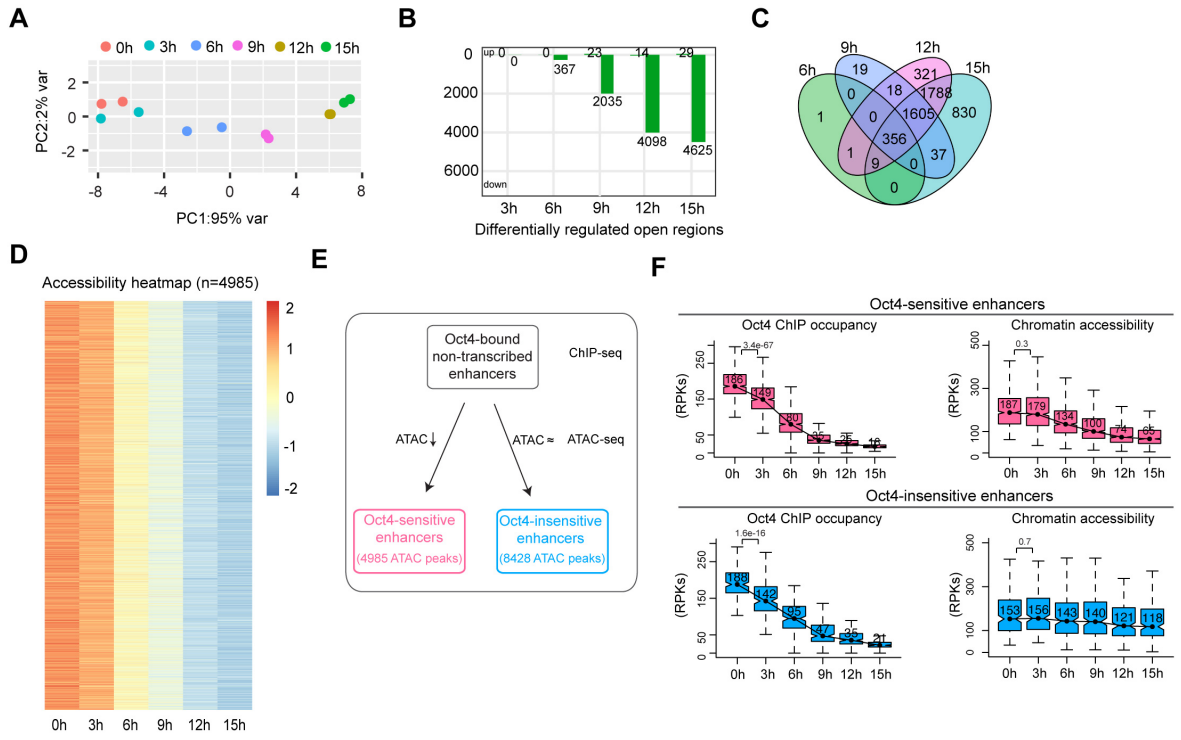


Figure 22. Chromatin accessibility changes at Oct4-bound non-transcribed enhancers.

(A) Principle component analysis of chromatin accessibility changes at Oct4 bound non-transcribed enhancers (n=12,710) over time course of DOX treatment (0, 3, 6, 9, 12, 15 hours) for two biological replicates. (B) Number of differentially regulated chromatin open regions detected by DESeq2 at each time point for Oct4 bound non-transcribed enhancers. (C)Venn diagram showing overlapping of differentially regulated chromatin open regions from (B) at each time points. (D)Heatmap visualizing the kinetics of chromatin accessibility changes at differentially regulated chromatin open regions (n=4,985). (E) Diagram indicating classification of Oct4-sensititive and -insensitive enhancers by changes of chromatin accessibility at Oct4-bound non-transcribed enhancers. (F)Boxplots illustrating changes of Oct4 occupancy and chromatin accessibility at Oct4-sensitive and -insensitive enhancers. P-values were calculated by Wilcoxon rank sum test. Y axis represents read counts per kilobases (RPKs). Black bars represent the median values for each group. Lower and upper boxes are the first and third quartiles, respectively. The ends of the whiskers extend the box by 1.5 times the interquartile range. Outliers are omitted.

We then investigated whether Sox2 may keep chromatin accessible also at Oct4-bound non-transcribed enhancers. Analysis of the Sox2 binding kinetics revealed a decrease of Sox2 binding at Oct4-sensitive enhancers after 9 hours of DOX treatment (**Figure 23A-B).** Oct4-insensitive enhancers showed only a slight decrease of Sox2 binding at 15 hours. Moreover, we analyzed published Oct4, Sox2 and Nanog ChIP-seq data at 0 and 24 hours after DOX treatment (King and Klose, 2017). Oct4-sensitive enhancers showed a strong loss of all three TFs at 24 hours, whereas at Oct4-insensitive enhancers Sox2 and Nanog occupancies were essentially unchanged (**Figure 23C**). This is consistent with our earlier observations at Oct4-bound transcribed enhancers (**Figure 19C-E**). Furthermore, metagene and motif analysis revealed an enrichment of Oct4 occupancy and the composite DNA motif at Oct4-sensitive enhancers (**Figure 23D-E**). Occupancy of other pluripotency factors and associated histone modifications also revealed a similar pattern as observed for Oct4-bound transcribed enhancers (**Figure 23D**). Together, these findings suggest that Sox2 can maintain chromatin accessibility at enhancers in the absence of eRNA synthesis during the first 6 hours of DOX treatment.

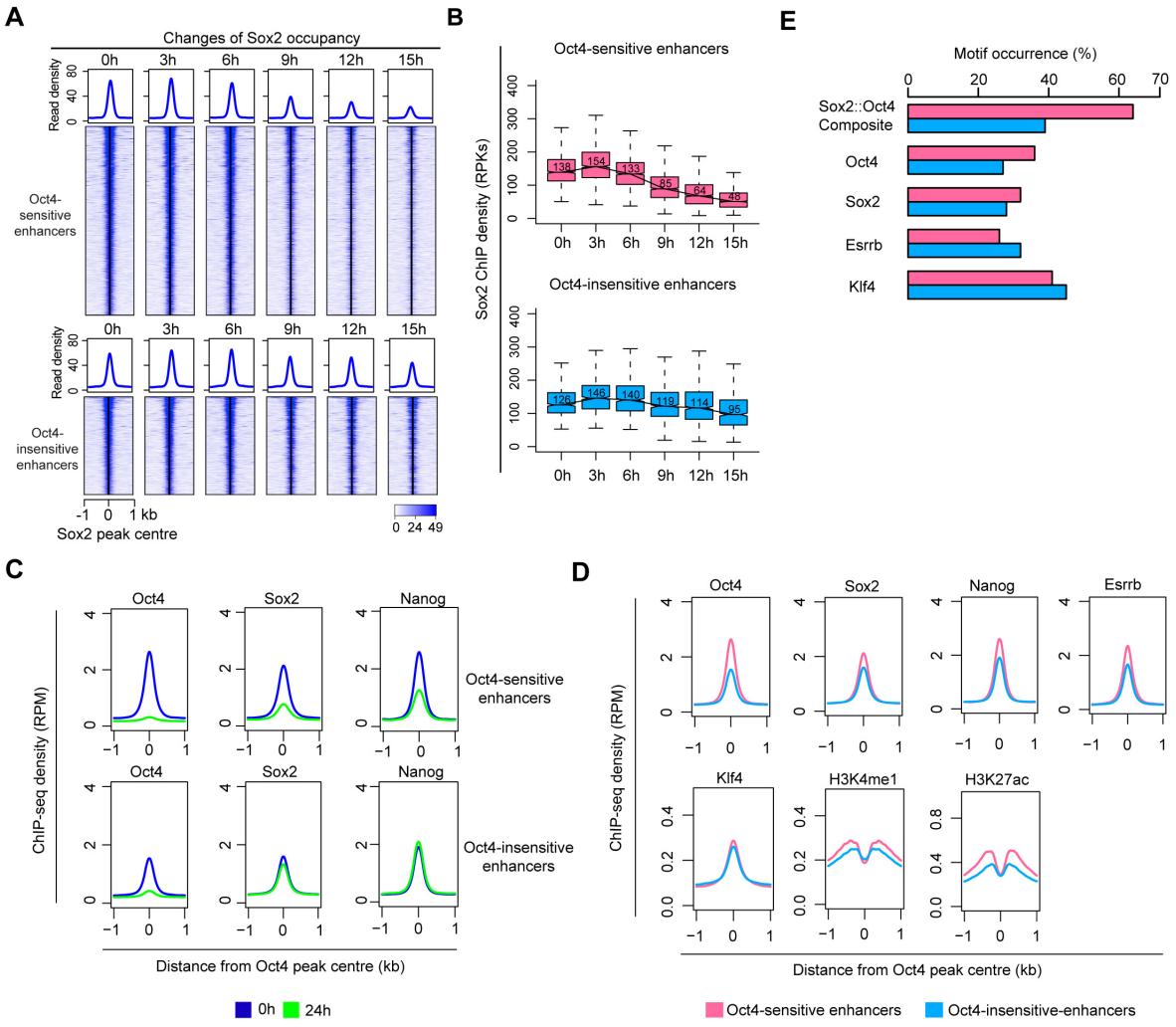


Figure 23. Sox2 maintains chromatin accessibility in the absence of eRNA synthesis.

(A) Heatmap showing changes of Sox2 occupancy at Oct4-sensitive and Oct4-insensitive enhancers over the entire time course of DOX treatment. Normalized read densities were presented, peaks were ranked by Sox2 read densities. (B) Same as (C), but using boxplots to depict quantification of Sox2 occupancy changes at Oct4-sensitive and Oct4-insensitive enhancers. Y axis represents Sox2 ChIP-seq density in reads per kilobases (RPKs). (C) Metagenes analysis of Oct4, Sox2 and Nanog occupancy after 0 and 24 hours of DOX treatment at Oct4-sensitive and -insensitive enhancers. ChIP-seq data of Oct4, Sox2 and Nanog were obtained from (King and Klose, 2017). (D) Metagene analysis of Oct4, Sox2, Nanog, Esrrb and Klf4 occupancy and H3K4me1, H3K27ac histone modifications at Oct4-sensitive and insensitive enhancers. ChIP-seq data of Oct4, Sox2 and Nanog were obtained from (King and Klose, 2017), other ChIP-seq data were obtained from (Chronis et al., 2017). Y axis depicts ChIP-seq coverage density in reads per million (RPM). (E) Percentage of motif occurrence at Oct4-sensitive and -insensitive enhancers for Sox2-Oct4 composite motif and Oct4, Sox2, Esrrb and Klf4 motifs.

IV. Discussion

Part of the discussion presented in this chapter has been published:
Le Xiong*, Erik A. Tolen* et al. Oct4 primarily controls enhancer activity rather than accessibility (* joint first authorship)
Manuscript in revision, doi: https://doi.org/10.1101/2021.06.28.450119

A detailed list of contents from the manuscript can be found in Appendix (page 104).

1. Direct target genes of Oct4

To study the primary function of Oct4 in maintenance of pluripotency, we conducted a high-resolution time course of Oct4 depletion in mouse ESCs using a doxycycline inducible system. To investigate the early effects of rapid Oct4 depletion on gene transcription, we monitored nascent RNA synthesis changes by TT-seq. Hundreds of genes were found to be down-regulated already at 6 hours of DOX treatment. Oct4 was fully depleted between 12-15 hours, accompanied by dysregulation of approximately 1500 genes. A previous study showed that a prolonged depletion of Oct4 for 24 hours led to dysregulation of thousands of genes (King and Klose, 2017). Our data is consistent with their discovery, but suggests that changes in gene transcription happen much faster and earlier than expected previously. Both studies indicate that there is a large and complex downstream gene network controlled by Oct4.

Here, we identified 446 genes that were early down-regulated upon depletion of Oct4 and these genes are strongly enriched for stem cell population maintenance. We found a strong enrichment of SE nearest genes within this group and most of them are well known key pluripotency regulators (e.g. Sox2 and Klf4). This suggests that a key pluripotency gene network centered around SEs is the primary/direct downstream targets of Oct4. Recently, another study reported immediate down-regulation of key pluripotency regulators upon rapid Oct4 protein degradation by an auxin-degron system (Bates et al., 2021). In this system, Oct4 protein levels are largely reduced within 0.5 hours and fully depleted within 2 hours of auxin addition (Bates et al., 2021). In our system, we depleted Oct4 via silencing gene transcription. When we compared our data with their RNA-seq data (**see analysis in section VI, 1.3**), we found similar changes in gene transcription for different groups of genes, suggesting that our observations are independent on the specific ways of Oct4 depletion.

2. TF co-occupancy contributes to enhancer transcription

Interestingly, although Oct4 occupies tens of thousands of distal regions, only a small group of them are actively transcribed and produce eRNAs. In line with previous discoveries in other cellular systems (Hah et al., 2013; Kim et al., 2010; Melgar et al., 2011; Zhu et al., 2013), we found that transcribed enhancers exhibit higher occupancy of pluripotency TFs, active histone modifications and greater chromatin accessibility in comparison to non-transcribed enhancers. This suggests that TF co-occupancy at enhancers contributes to transcription and synthesis of eRNAs at enhancers. We found an enrichment of enhancer transcription at SEs, consistent with their high occupancy of pluripotency factors and co-activators (Hnisz et al., 2013; Whyte et al., 2013). Our data appears to be consistent with several imaging studies showing that TFs (e.g. Sox2), co-activators (e.g. Mediators and Brd4) and Pol II itself are often un-evenly distributed in the nucleus and clustered at specific loci (Boehning et al., 2018; Cho et al., 2018; Liu et al., 2014; Sabari et al., 2018).

3. Oct4 maintains pluripotency gene expression by controlling enhancer activity rather than accessibility

In line with rapid down-regulation of pluripotency genes, eRNAs derived from Oct4 bound transcribed enhancers also decreased immediately upon loss of Oct4 binding. Moreover, the observed eRNA synthesis changes correlated well with mRNA synthesis changes from putative Oct4 target genes. Surprisingly, however, we found that loss of Oct4 binding did not directly correlate with changes in chromatin accessibility changes, which were generally delayed. Specifically, we defined two groups of enhancers named Oct4-sensitive and Oct4-insensitive enhancers. For the Oct4-sensitive enhancers, we observed a decrease of chromatin accessibility but the changes occurred later than Oct4 depletion. Importantly, for this group of enhancers we found that down-regulation of putative target genes occurred earlier than the delayed decrease of chromatin accessibility (e.g. Klf4 gene), suggesting that accessibility changes are not directly required for Oct4 functions in gene expression. For Oct4-insensitive enhancers, depletion of Oct4 led to loss of Oct4 binding and simultaneous decrease of eRNAs synthesis, whereas the chromatin accessibility remained unchanged in the full-time course. These data strongly support that Oct4 maintains pluripotency gene expression by primarily controlling enhancer activity rather than accessibility.

4. Sox2 may maintain enhancers accessible

Using the same doxycycline inducible system, it has been found that depletion of Oct4 for 24 hours led to a loss of chromatin accessibility at the majority of Oct4 occupied sites, accompanied by loss of Sox2 and Nanog binding (King and Klose, 2017). This led to the conclusion that Oct4 also acts as a pioneer factor to maintain an open chromatin environment and stabilizes binding of Sox2 and Nanog in mouse ESCs (King and Klose, 2017). In our high temporal resolution data, Sox2 remained bound at both Oct4 sensitive and -insensitive enhancers during the first 6 hours of Oct4 depletion and the chromatin accessibility remained largely unchanged, although modest effects were observed at 6 hours. Thus, considering the early effects of Oct4 depletion in our data suggests a different model in which Sox2 rather than Oct4 might play a more important role in control of chromatin accessibility. Compared to Oct4-insensitive enhancers, we found an enrichment of Oct4 binding and the Sox2-Oct4 composite motif at Oct4-sensitive enhancers. This suggests that enrichment of Oct4 on chromatin results from its cooperation with Sox2 through the composite motif. Taken together, these results suggest an emerging model that Sox2 assists Oct4 binding at enhancers and Oct4 controls enhancer activity, whereas Sox2 maintains enhancer accessibility.

Consistent with our genome-wide *in vivo* analysis, structural studies found that Sox2 alone can bind a nucleosome and displace part of the nucleosomal DNA from the octamer surface (Dodonova et al., 2020), and that Oct4 and Sox2 can co-occupy a nucleosome, contacting each other (Michael et al., 2020). It was also found that Oct4 and Sox2 bind cooperatively to a composite DNA motif *in vitro* (Chang et al., 2017) and that Sox2 can increase the specificity for Oct4 binding i*n vivo* (Merino et al., 2014; Michael et al., 2020). Also supporting our findings, single molecule imaging showed that Sox2 assists Oct4 binding in live ESCs (Chen et al., 2014), although reanalysis of this data suggests that the effect of Oct4 and Sox2 on one-another's binding may vary across different binding sites (Biddle et al., 2019).

Our model appears to be inconsistent with a previous report showing that Oct4 has a stronger influence on chromatin accessibility than Sox2 in mouse ESCs (Friman et al., 2019). By separately depleting Oct4 for 24 hours and Sox2 for 26 hours using a doxycycline inducible system, they found that loss of Oct4 led to stronger influences on chromatin accessibility than Sox2 (Friman et al., 2019). Specifically, based on changes in accessibility, the authors defined three enhancer groups: Oct4 dependent enhancers (only dependent on Oct4 for accessibility, n

= 8869), Sox2 dependent enhancers (only dependent on Sox2 for accessibility, n = 1834) and co-dependent enhancers (dependent on both Oct4 or Sox2 for accessibility, n = 2973) (Friman et al., 2019). Surprisingly, Oct4 and Sox2 were found to co-occupy all enhancers in all three groups. These data seems to support that Oct4 is more important than Sox2 for control of chromatin accessibility. However, in their study, a residual level Sox2 protein was still detectable even after 40 hours of Sox2 depletion due to the much longer half-live of the Sox2 protein compared to Oct4 (Friman et al., 2019). Thus, the different effects on chromatin accessibility observed in their study may be due to residual levels of Sox2 that remained bound to chromatin (Friman et al., 2019). To investigate this further, it will be important to develop an efficient and rapid Sox2 protein degradation system in the future.

Importantly, in contrast to a previous report (King and Klose, 2017), a recent study using an auxin-degron system found that immediate loss of Oct4 protein leads to a global increase rather than decrease of Nanog occupancy at enhancers in mouse ESCs (Bates et al., 2021). Their observation resembles our finding of Sox2 remaining bound at enhancers upon loss of Oct4 binding. It remains unknown how chromatin accessibility is changed in their study, but the observed increased binding of Nanog suggests that the chromatin accessibility is unlikely to be changed after immediate loss of Oct4 protein (Bates et al., 2021). Thus, both studies support a model in which other pluripotency factors can remain bound after immediate loss of Oct4 and they may play a direct role in rendering chromatin accessible.

Our model that Oct4 controls enhancer activity and Sox2 maintains chromatin accessibility is further supported by published results on the functions of Oct4 and Sox2 during reprogramming of somatic cells to induced pluripotent stem cells (iPSC) (Chen et al., 2016; Chronis et al., 2017; Li et al., 2017; Soufi et al., 2012; Soufi et al., 2015). Several lines of evidence indicate that during reprogramming Sox2 acts as a driver of chromatin opening, whereas Oct4 is an accessory factor (Li et al., 2017; Malik et al., 2019) that is not strictly required (An et al., 2019; Velychko et al., 2019). Thus, a general model emerges that Oct4 controls enhancer activity, whereas Sox2 governs enhancer accessibility, although both factors often cooperate and engage in functional interactions with other factors.

5. Two step enhancer activation and inactivation model

Enhancer activation is thought to happen in a two-step manner: priming of enhancers and activation of enhancers (Li et al., 2016). In the first step, pioneer factors are required to create an open chromatin environment that allows binding of other collaborative TFs and important cofactors including histone methyltransferases that deposit H3K4me1 or H3K4me2 (Li et al., 2016). The second step requires the recruitment of other important cofactors, including the histone acetyltransferases CBP/p300, which deposit histone acetylation marks, and finally, general TFs and Pol II are recruited to enhancers and eRNAs are produced (Li et al., 2016). In contrast, little is known about how inactivation of enhancers occurs. In our study, a group of enhancers was inactivated in two steps upon Oct4 depletion. First enhancer transcription was decreased, followed by a decrease in chromatin accessibility. Thus, based on our results and available literature we therefore propose that the activation and inactivation of enhancers both occur in a two-step process with distinct factors dominating each step but also often cooperating.

6. Implication for SEs

Whether SEs represent a specific class of enhancers is still highly controversial (Blobel et al., 2021; Pott and Lieb, 2015). Our data support that the activity of SEs is highly susceptible to perturbation. We found that loss of Oct4 led to preferential decrease of Oct4 binding and eRNA synthesis at SEs in comparison to other enhancers and the nearest genes to SEs were strongly down-regulated. This suggests a crucial role of Oct4 in the maintenance of SEs to control pluripotency gene expression in mouse ESCs. In addition, our data also provide new insights into the property of SEs. It has been proposed that the individual enhancers with SEs may behavior similar and act as a single entity and the entire SE may collapse as a functional unit after perturbation (Hnisz et al., 2017). However, our data suggest that, at least for some of the SEs, this is not the case. For example, for the Sox2 and mir290 SEs, some of the individual enhancers, its activity by eRNA can be down-regulated, but the chromatin accessibility can remain unchanged after loss of Oct4. Whereas for other individual enhancers within the SEs, loss of Oct4 led to changes in both enhancer activity and accessibility. This suggests that the individual enhancers within SEs can behavior in a very different way in response to Oct4 concentration change, which may provide a means of precise gene regulation.

V. Future Perspectives

1. Oct4 depletion system

In this study, we used a doxycycline inducible Oct4 loss-of-function system in mouse ESCs to deplete Oct4 via the transcription level. After 6 hours of Oct4 depletion, minimal levels of Oct4 occupancy remained at Oct4-bound transcribed enhancers and Oct4 was only fully depleted between 12-15 hours. Currently, it cannot be excluded that a minimal level of Oct4 is sufficient to maintain the chromatin accessibility.

In here, I suggest two experiments to disentangle this question in the future. The first experiment is to recover Oct4 by washout of doxycycline after a full depletion of Oct4 protein between 12-15 hours and to capture RNA synthesis and chromatin accessibility changes at enhancers and promoters during a recovery time course. When Oct4 levels increase again during recovery, collecting data at the early stage of recovery would allow us to answer whether or not minimal level Oct4 levels are sufficient to make chromatin more accessible. If we would observe an immediate increase of eRNA synthesis whereas chromatin accessibility remains unchanged during the early stages of recovery, this would strongly support our conclusion that Oct4 mainly contributes to enhancer activity rather than accessibility. The second experiment is to conduct the same experiments including TT-seq, ATAC-seq and ChIP-seq of Oct4 and Sox2, but use the recently developed Oct4 protein degradation system (Bates et al., 2021). In this system, Oct4 protein can be fully depleted within 2 hours of auxin addition (Bates et al., 2021).

2. Pioneer factor model or TF cooperativity

The pioneer factor model emphasizes that there is a key factor that can first interact with nucleosomes and create an opening chromatin environment to facilitate the binding of other factors and the subsequent activation of enhancers and cell fate transition (Zaret and Carroll, 2011). Oct4, Sox2 and other pluripotency factors (e.g. Klf4) have all been demonstrated to be able to act as a pioneer factor during iPSCs reprogramming (Soufi et al., 2012; Soufi et al., 2015). However, a previous study showed that selection of pluripotency enhancers during early stage iPSC reprogramming requires a tight cooperation among Oct4, Sox2 and Klf4 (Chronis et al., 2017). Expression of another pluripotency factor, Esrrb, together with Oct4, Sox2 and Klf4, led to early accessibility of some pluripotency enhancers which are not accessible without the expression of Esrrb (Chronis et al., 2017). This supports a different model in which cooperative binding of TFs rather than a single pioneer factor orchestrates iPSCs reprogramming (Chronis et al., 2017).

Here, we studied the function of Oct4 in native mouse ESCs where the chromatin is already open. We found that Sox2 can remain bound after loss of Oct4 during the early stage of enhancer inactivation. In agreement with our observation, another study reported increased binding of the pluripotency factor Nanog after immediate loss of Oct4 protein (Bates et al., 2021). Thus, both studies support a model in which binding of different pluripotency factors can be quite independent on whether Oct4 is in present or not.

In our work, we studied the function of Oct4 under the hypothesis that Oct4 contributes to enhancer activity and accessibility in an equal concentration-dependent manner. It remains unknown whether this is the case and our work is the first study to monitor changes in enhancer activity and chromatin accessibility simultaneously with high-sensitive upon depletion of a pioneer-like factor. In the future, it will be important to perform similar experiments for other pluripotency factors such as Sox2, Klf4 and Esrrb to understand the individual role of different factors in control of enhancer activity, accessibility and gene regulation. Upon depletion, if none of them can result in simultaneous changes in enhancer accessibility and eRNA synthesis, this would suggest that the chromatin accessibility is maintained generally by cooperation among different TFs rather than a single factor. Alternatively, it may also suggest that an unequal concentration-dependent mechanism underlies the transcriptional activity and chromatin opening activity of TFs.

3. The function of enhancer transcription and eRNA

In this study, we used eRNA synthesis as a proxy for enhancer activity. However, the functional role of eRNAs in gene regulation during the Oct4-dependent enhancer inactivation remains completely unknown. In general, the function of eRNAs or enhancer transcription remains largely unknown (**see Section I, 3.3**). A previous study suggested that eRNAs can interact with the histone acetylation enzyme CBP/p300 to regulate its histone acetyltransferase activity *in vivo* and *in vitro* (Bose et al., 2017). In addition, enhancer transcription is thought to also play a role in deposition of histone methylation (H3K4me1/2) during microphage enhancer activation (Kaikkonen et al., 2013). In our study, it remains unknown how different histone marks are changed after loss of Oct4 and how changes of eRNA synthesis and chromatin accessibility are temporally coordinated with the chromatin state. Thus, it will be important in the future to also perform ChIP-seq for the aforementioned histone modifications during the same time course used in here. Combing these data with our TT-seq and ATAC-seq data will provide new insights into the order of events during Oct4-dependent enhancer inactivation and shed light on the functional roles of enhancer transcription and eRNAs in control of chromatin remodeling and the epigenome.

4. Enhancer cooperativity

Cell identity genes are often controlled by multiple enhancers (e.g. SEs), however, it still remains largely unknown what the nature of SEs is (Pott and Lieb, 2015).Previous work from our lab using computational modelling of time-course TT-seq data showed that during human leukemia B-cells to macrophages transdifferentiation, multiple enhancers could behave in an additive or synergistic way to regulate gene expression (Choi et al., 2021). SEs are well studied in mouse ESCs and there are plenty of additional functional genomics data accessible in public databases. Thus, it will be interesting to use the same computational pipeline to model our TT-seq data to identify different types of enhancers cooperativity during the Oct4-dependent enhancer inactivation process. Moreover, here we identified early and late-down regulated genes upon Oct4 depletion. In the future it will be interesting to link the modelling results with this information to understand the biological function of the different types of enhancer cooperativity.

5. The role of Oct4 in enhancer-promoter interaction

It remains unknown whether Oct4 could play a role in controlling enhancer-promoter interactions. Previous studies showed that the three dimensional genome structure in pluripotent cells is organized around pluripotency factors, suggesting a role of these factors in chromatin organization (de Wit et al., 2013; Denholtz et al., 2013). Evidence supports that the pluripotency factor Klf4 is involved in enhancer-promoter interactions during reprogramming and in pluripotency cells (Di Giammartino et al., 2019). Oct4 may play an indirect role in control of enhancer-promoter interactions by stabilizing cohesin binding at enhancers (Liu et al., 2021). Prolonged depletion of Oct4 led to a decrease of chromatin accessibility and loss of binding of cohesin at pluripotency enhancers (Liu et al., 2021). Depletion of Sox2, but not Nanog, led to similar effects as observed for Oct4, demonstrating that Oct4 and Sox2 are important to create an open chromatin platform for loading of cohesion (Liu et al., 2021). Interestingly, a recent study showed that Oct4 is involved in reorganization of TADs during reprogramming, which dependents on the phase separation capacity of Oct4 (Wang et al., 2021).

Our work and the system used in here could also be an ideal system to study the direct functional role of Oct4 or other pluripotency factors in control of enhancer-promoter interactions. We observed changes in transcription at both enhancers and promoters between 3-9 hours of Oct4 depletion and the chromatin accessibility was significantly decreased only afterwards. Thus, to investigate the direct role of Oct4 or other pluripotency factors in enhancer-promoter looping, the data needs to be collected before 9 hours of Oct4 depletion. In this way, any secondary effects caused by decreased chromatin accessibility can be excluded. Potential future experiments include HiChIP (Mumbach et al., 2016) against Oct4 or next-generation Capture-C (Davies et al., 2016) for some selected gene locus (e.g. Klf4 or Sox2 genes) to examine the changes in enhancer-promoter interactions. Combing these data with our TT-seq data will provide fundamental new insights into the roles of Oct4 in enhancer-promoter interactions.

VI. Supplementary information

1. Additional analysis for section III

In this section, I performed additional analysis to support the discovery in section III.

1.1 Analysis of enhancers not bound by Oct4

In section III, we mainly focused on Oct4-bound transcribed enhancers and provided analysis regarding their changes in eRNA synthesis, chromatin accessibility and occupancy of Sox2. In here, I also provided the same analysis but for Oct4-unbound transcribed enhancers and used it as a positive control.

Briefly, I performed differential expression analysis of eRNAs at Oct4-unbound transcribed enhancers (**Figure 24A-E**). Based on this, I also generated a barplot to compare the percentage of differentially regulated eRNAs at each time points for Oct4-bound and -unbound transcribed enhancers, respectively, as shown in (**Figure 24F**). We found down-regulation of eRNAs synthesis occurred predominately at Oct4-bound transcribed enhancers in comparison to Oct4-unboulatend transcribed enhancers (**Figure 24F**). Although a total of ~400 eRNAs (7.4%) were down-regulated at Oct4-unbound transcribed enhancers, most of the downregulation occurred between 12-15 hours of Oct4 depletion. A tiny amount of up-regulated eRNAs (<2%) were detected at both Oct4-bound and -unbound transcribed enhancers (**Figure 24F, right**). These data suggest that upon Oct4 depletion, inactivation of enhancers occurred rapidly and predominately at Oct4-bound transcribed sites, although a small portion of Oct4-unbound transcribed sites were also influenced later, most likely representing secondary effects.

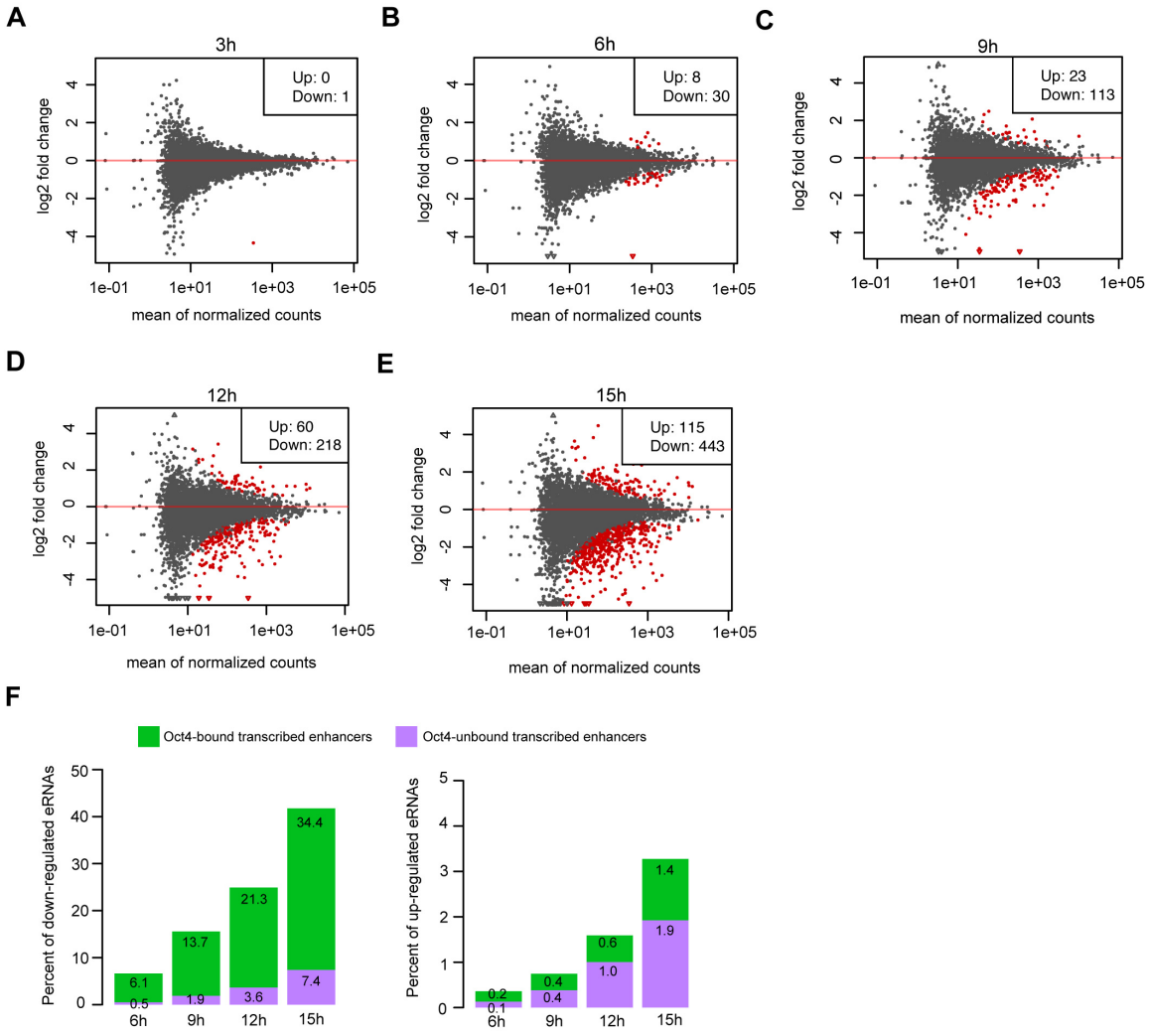


Figure 24. Differential expression analysis of eRNAs at Oct4-unbound transcribed enhancers.

(A-E) MA plots showing differentially regulated eRNAs detected by DESeq2 at 3, 6, 9, 12 and 15 hours of DOX treatment for Oct4-unbound transcribed enhancers. Statistically significant changes are depicted in red. (F) Percent of down-regulated (left) and up-regulated (right) eRNAs at Oct4 bound and -unbound transcribed enhancers at 6, 9, 12 and 15 hours.

I then called significantly changed accessible chromatin regions for Oct4-unbound transcribed enhancers (**Figure 25A-E**). At 15 hours of DOX treatment, only 48 enhancers (0.3%) showed significant change of chromatin accessibility, indicating that the chromatin accessibility remained largely unaltered across the full-time course for Oct4-unbound transcribed enhancers. The accessibility boxplot (**Figure 25F**) also confirmed no change of chromatin accessibility at Oct4-unbound transcribed enhancers.

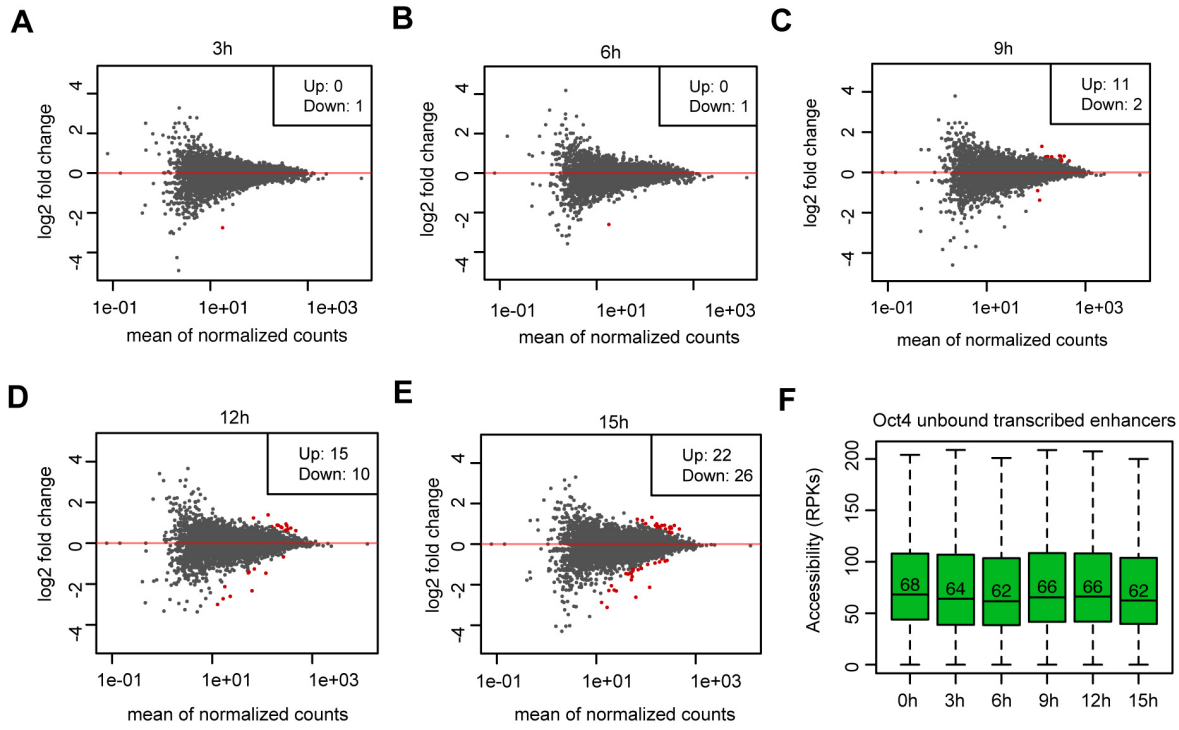


Figure 25. Analysis of differentially regulated chromatin accessible regions for Oct4-unbound transcribed enhancers.

(A-E) MA plots of differentially regulated chromatin accessible regions detected by DESeq2 at 3,6,9,12 and 15 hours of DOX treatment at Oct4-unbound transcribed enhancers. Statistically significant changes are depicted in red. (F) Boxplots showing chromatin accessibility at Oct4-unbound transcribed enhancers.

For Sox2 occupancy, we found a tiny portion (307 enhancers, 4.3%) of Oct4-unbound transcribed enhancers which overlapped with our Sox2 ChIP-seq peaks (within a 1kb window). For this subgroup, we also found no appreciable change of Sox2 occupancy across the full-time course, as shown in (**Figure 26A-B**).

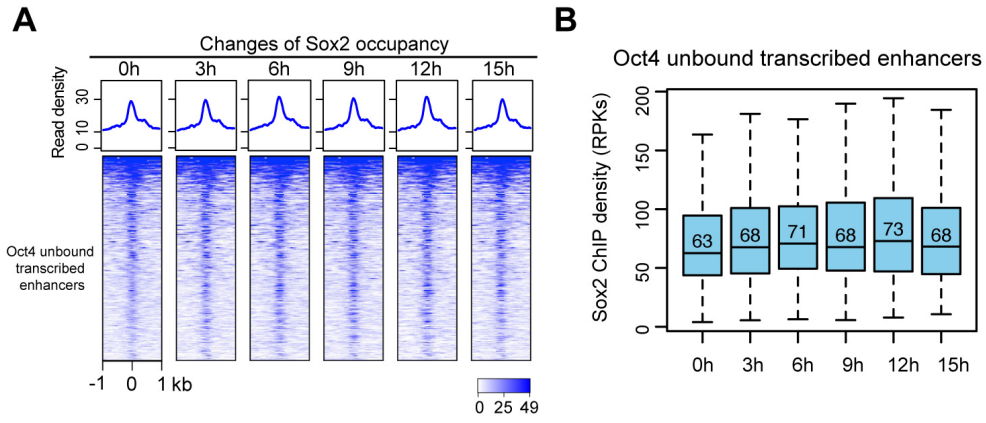


Figure 26. Sox2 occupancy at Oct4 unbound transcribed enhancers.

(A) Heatmap showing changes of Sox2 occupancy at Oct4-unbound transcribed enhancers over the entire time course of DOX treatment. Normalized read densities were presented, peaks were ranked by Sox2 read densities. (B) Same as (A), but using boxplots to depict quantification of Sox2 occupancy changes at Oct4-unbound transcribed enhancers. Y axis represents Sox2 ChIP-seq density in reads per kilobases (RPKs).

1.2 Impact of Oct4 depletion on super enhancers

For around half of the super enhancers bound by Oct4 a decrease in eRNAs synthesis did not lead to a statistically significant decrease of nearest gene expression. To further elaborate this, We first plotted their mRNA fold changes to test the possibility that there is a decrease of mRNA synthesis but not significant enough to call differential expression. We found no appreciable changes in their mRNA synthesis (**Figure 27A**). We then analyzed the occupancy of Oct4 and the fold change of eRNAs synthesis at these SEs. We observed much lower occupancy of Oct4 and a less dramatic decrease of eRNA synthesis at these SEs in comparison to the transcriptionally down-regulated SEs for which the nearest genes were down-regulated (**Figure 27B**). This suggests that the amount of bound Oct4 and the degree of eRNA synthesis changes at decommissioned SEs is likely to play a role in determining gene expression. Currently, it is unclear whether the decommissioned SEs may have regulatory ability beyond the nearest gene.

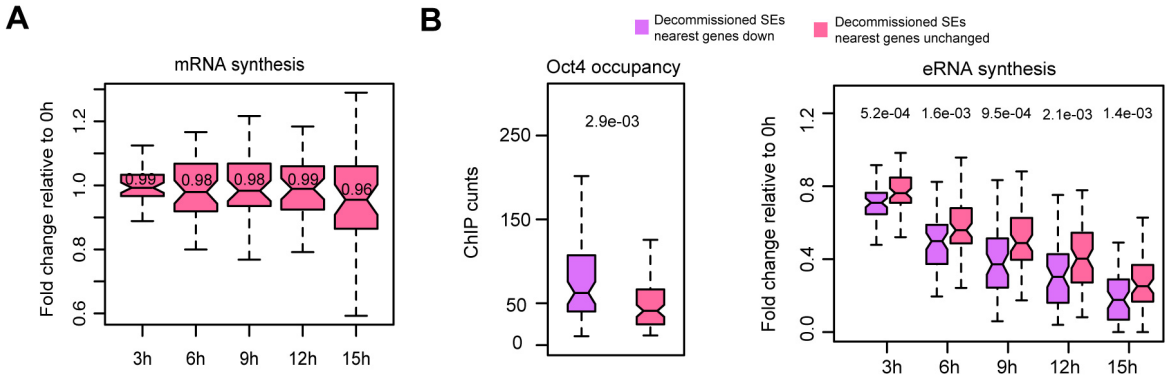


Figure 27. Comparison of eRNA-down regulated SEs whose nearest gene were down-regulated or unchanged upon Oct4 depletion.

(A) mRNA fold changes of the nearest genes of the another half of the super enhancers for which the eRNA synthesis were down-regulated but the nearest genes were unchanged statistically. (B) Occupancy of Oct4 and fold changes in eRNA synthesis at transcriptionally down-regulated for which the nearest genes were down-regulated or unchanged.

1.3 Analysis of public RNA-seq data

A recent work reported immediate mechanisms of Oct4 in mouse ESCs by depletion of Oct4 protein within 2 hours using an auxin-degron-tagged system (Bates et al., 2021). They showed by RNA-seq that a full depletion of Oct4 leads to rapid down-regulation of key pluripotency factors (Bates et al., 2021). Here, in our study, we used a Tet-off Oct4 depletion system that silences Oct4 gene expression and this resulted in a gradual decrease of Oct4 level during most of the time course.

In this section, I analyzed the RNA-seq data from their study and compared it with our TT-seq data. We have classified differentially expressed genes into early and late down- and up-regulated gene groups. Similar effects on gene expression changes were observed in their RNA-seq data for these different gene groups (**Figure 28A-B**). In addition, compared with other down-regulated genes, we also found SE-controlled genes were strongly influenced in their RNA-seq data (**Figure 28C**). Taken together, these results demonstrate that although we used different Oct4 depletion systems, similar mRNA response dynamics can be observed in both system.

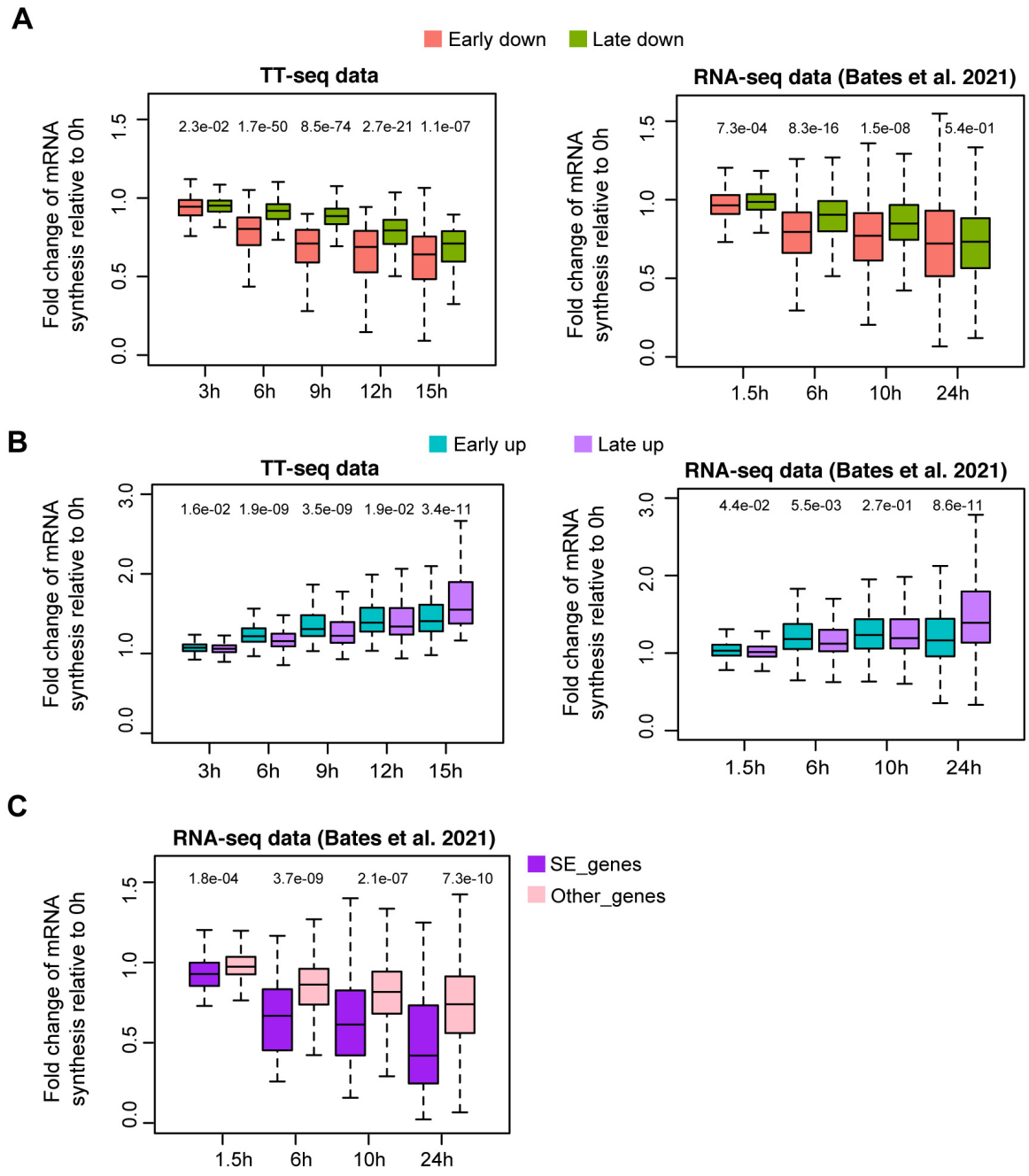


Figure 28. Response dynamics of early-down, late-down, early-up and late-up regulated genes in our TT-seq data and a recent RNA-seq data by Oct4 protein degragation.

(A) Response dynamics of early/late-down regulated genes in our TT-seq and their RNA-seq data. (B) Response dynamics of early/late-up regulated genes in our TT-seq and their RNA-seq data. (C) Response dynamics of SE nearest down-regulated genes and other down regulated genes in our TT-seq and their RNA-seq data. The different groups of genes were classfied by analysis of our TT-seq data (**see section III, part 2, Figure 9-11**) and RNA-seq data taken from (Bates et al., 2021).

2. Transcription units and enhancers annotation

Transcription units, eRNA annotation, Oct4-bound_transcribed_enhancers and other annotations generated in this study can be found in Gene Expression Omnibus database with query number GSE174774.

3. Sequence of the spike-ins.

The following six ERCC spike-ins were used in TT-seq experiments.

Spike-in 2

AATACCTTTACAAATGCTTTAACAAGAGGAAATTGTGTTTTTGCCAATTTAAGAC
CTAATTTAATAGTTAAACCATTAACCTTAGTTGTTCCAAGGCATAATATAGAGAG
TGAGATACAGGATGAGCTATTTCAGGGAGTTATTCAGTATGCAGTTGCCAAGGCA
GTTGCTGATTTAGATTTAGATGAAGATTTAAAGGTTGTTGTCTCTGTTAATGTCCC
AGAGGTTCCAATAACCAATTTAAATAAAAGAAAACTCTTCCAATACTTCTATGCC
TCAGCAAAGTTAGCTATAAACAGAGCTTTAAATGAATATCCTTCAAAAGAGAAG
GTAAAGAAAGAGAAATATAGAGCTTTGCATCCATTAGTTGGATTTAGGGATGTTA
GATTGGAGTATCCTCCATATCTACAAATTGCTTTGGATGTCCCAACTATGGAGAA
TTTGGAATTTTTGTTACAAACAATTCCAAATAGCGACCACATCATCTTAGAGGCT
GGAACACCACTAATTAAAAAGTTTGGTTTAGAGGTTATTGAAATAATGAGAGAAT
ATTTTGATGGCTTTATTGTTGCTGATTTAAAAACCTTAGACACTGGAAGGGTTGA
GGTAAGATTGGCATTTGAAGCAACAGCTAATGCAGTGGCAATAAGTGGAGTAGC
ACCAAAATCAACAATAATTAAAGCTATCCACGAATGTCAAAAATGTGGTTTAATC
AGCTATTTGGATATGATGAACGTCTCTGAACCTCAAAAATTATATGATTCATTAA
AATTAAAGCCAGATGTTGTTATCTTGCATAGAGGGATTGATGAGGAGACATTTGG
AATTAAAAAGGAATGGAAATTTAAGGAAAACTGCTTATTAGCAATTGCTGGAGG
AGTTGGTGTGGAGAATGTTGAAGAGCTTTTAAAAGAATATCAAATATTAATCGTT
GGTAGAGCAATTACAAAATCAAAAGACCCAGGAAGAGTAATTAGGATTTTATAA
ACAAGATGGGTTAAAAAAAAAAAAAAAAAAAAAAA

Spike-in 4

TTTCGACGTTTTGAAGGAGGGTTTTAAGTAATGATCGAGATTGAAAAACCAAAAA
TCGAAACGGTTGAAATCAGCGACGATGCCGAATTTGGTAAGTTTGTCGTAGAGCC
ACTTGAGCGTGGATATGGTACAACTCTGGGTAACTCCTTACGTCGTATCCTCTTAT
CCTCACTCCCTGGTGCCGCTGTAACATCAATCCAGATAGATGGTGTACTGCACGA
ATTCTCGACAATTGAAGGCGTTGTGGAAGATGTTACAACGATTATCTTACACATT
AAAAAGCTTGCATTGAAAATCTACTCTGATGAAGAGAAGACGCTAGAAATTGAT
GTACAGGGTGAAGGAACTGTAACGGCAGCTGATATTACACACGATAGTGATGTA
GAGATCTTAAATCCTGATCTTCATATCGCGACTCTTGGTGAGAATGCGAGTTTCC
GAGTTCGCCTTACTGCTCAAAGAGGACGTGGGTATACGCCTGCTGACGCAAACAA
GAGAGGCGATCAGCCAATCGGCGTGATTCCGATCGATTCTATCTATACGCCAGTT
TCCCGTGTATCTTATCAGGTAGAGAACACTCGTGTAGGCCAAGTTGCAAACTATG
ATAAACTTACACTTGATGTTTGGACTGATGGAAGCACTGGACCGAAAGAAGCAA
TTGCGCTTGGTTCAAAGATTTTAACTGAACACCTTAATATATTCGCTGGTTTAACT
GACGAAGCTCAACATGCTGAAATCATGGTTGAAGAAGAAGAAGATCAAAAAGAG
AAAGTTCTTGAAATGACAATTGAAGAATTGGATCTTTCTGTTCGTTCTTACAACTG
CTTAAAGCGTGCGGGTATTAACACGGTTCAAGAGCTTGCGAACAAGACGGAAGA
AGATATGATGAAAGTTCGAAATCTAGGACGCAAATCACTTGAAGAAGTGAAAGC
GAGACTAGAAGAACTTGGACTCGGACTTCGCAAAGACGATTGACTAGTTTCCCTT
GTGAACTAGGATTTTCCCGGGTACAAAAAAAAAAAAAAAAAAAAA

Spike-in 5

ACTGTCCTTTCATCCATAAGCGGAGAAAGAGGGAATGACATTGTTCTTACACGGC
ACAAGCAGACAAAATCAACATGGTCATTTAGAAATCGGAGGTGTGGATGCTCTCT
ATTTAGCGGAGAAATATGGTACACCTCTTTACGTATATGATGTGGCTTTAATACG
TGAGCGTGCTAAAAGCTTTAAGCAGGCGTTTATTTCTGCAGGGCTGAAAGCACAG
GTGGCATATGCGAGCAAAGCATTCTCATCAGTCGCAATGATTCAGCTCGCTGAGG
AAGAGGGACTTTCTTTAGATGTCGTATCCGGAGGAGAGCTATATACGGCTGTTGC
AGCAGGCTTTCCGGCAGAACGCATCCACTTTCATGGAAACAATAAGAGCAGGGA
AGAACTGCGGATGGCGCTTGAGCACCGCATCGGCTGCATTGTGGTGGATAATTTC
TATGAAATCGCGCTTCTTGAAGACCTATGTAAAGAAACGGGTCACTCCATCGATG
TTCTTCTTCGGATCACGCCCGGAGTAGAAGCGCATACGCATGACTACATTACAAC
GGGCCAGGAAGATTCAAAGTTTGGTTTCGATCTTCATAACGGACAAACTGAACGG
GCCATTGAACAAGTATTACAATCGGAACACATTCAGCTGCTGGGTGTCCATTGCC
ATATCGGCTCGCAAATCTTTGATACGGCCGGTTTTGTGTTAGCAGCGGAAAAAAT
CTTCAAAAAACTAGACGAATGGAGAGATTCATATTCATTTGTATCCAAGGTGCTG
AATCTTGGAGGAGGTTTCGGCATTCGTTATACGGAAGATGATGAACCGCTTCATG
CCACTGAATACGTTGAAAAAATTATCGAAGCTGTGAAAGAAAATGCTTCCCGTTA
CGGTTTTGACATTCCGGAAATTTGGATCGAACCGGGCCGTTCTCTCGTGGGAGAC
GCAGGCACAACTCTTTATACGGTTGGCTCTCAAAAAGAAGTGGATAAGCTGTACA
ATCGTTTCATCATTCGGCGTGCGAATTAAAAAAAAAAAAAAAAAAAAAAAAA

Spike-in 8

AGATGTATATATGATGTCCTTGGACGGGGTGGCGCAGTATTACTGCAAGAGAGCG
GACAGATTAGTGTGTTGGAGCCGACACATCAAAGGTTCGTCCGGGGACCGATCTG
CAGCCTACGGGACATTTATCCGTAAAAGCATGGCGCTGTTTCGTACTTATCGGAG
GCCAGGTATCGTCGCGGCGAGTCTCCCCGACGACGGAGATGGGCGTTACTATCTG
GGCCGTCTCGTACTCTGTTACTTGGCACAGATGCGAGCCCTCGTAATGTGCATCA
GCTAAGGGCGATATTATAATGCGACGTTTGTACGGATTCGTTACTAACGTGTTGG
ACGCTAGTGGAATATGTGTCGTTGGTTAGCCTACCCATGGCTTTCGCGGCGACAC
ATGCTTAGACTCTTTCAAAACTTCGGTGAAGTTCACTCAAGCCGCGGAGCGCCGT
CGTAATTCACTAGGGATGGCGGTACCCGTGCCCGTCCGATTCGTAGCAACCTGCA
TCACGATTTTGTCTTCGGGCGACTTATCAGATACGGTAATGTAAATACCTGGCATT
TGGGCACTTCTTGCGTTTAAGCGGGAAAGATCGCGAGGGCCCGCTATTTGCGATA
CTTCCCATGTCGGTGCCGTCGCCTCTATGTACTCGGAGACGTTAATGCAGAGGCT
AAGGACAATTTACCATGACTCGGTAATCCGTTCGTCAAGCAGGTAGCTCGAGTCT
CCCCACGGACACGTAGTGGGTTTGTAACGATCGATACCGAGTCTTTTTGTCTAGT
AGAACCAACCAACCATTAAGGAGTTCACTAGCACATCTTTGCGACCCGATCGTCC
GTGTGTCGCGTAATACTTTTGTTATGACGAGACATACGCTCAAGCCCTGGGTAGC
TAGTCGCGGAGGCACGTTACCGCGCACAACCCCTATTCGTTTACATGTACATCGC
ATCTGAGGTAGTACACTTCCGGCGTACGTGAGTATTTGCGCGTAATAAGCGCGTG
TTTAGCTGATCCCCTCTCGTATCGAGGTTAAGGCAGATTAGTGCCCAGTAATTGC
GTTTTTTTGTCGTTGTCGCAGAACGCGATTTGCTCCGAAAGCTTTAAGCCGTGGAA
AAAAAAAAAAAAAAAAAAAAA

Spike-in 9

TCCAGATTACTTCCATTTCCGCCCAAGCTGCTCACAGTATACGGGCGTCGGCATC
CAGACCGTCGGCTGATCGTGGTTTTACTAGGCTAGACTAGCGTACGAGCACTATG
GTCAGTAATTCCTGGAGGAATAGGTACCAAGAAAAAAACGAACCTTTGGGTTCC
AGAGCTGTACGGTCGCACTGAACTCGGATAGGTCTCAGAAAAACGAAATATAGG
CTTACGGTAGGTCCGAATGGCACAAAGCTTGTTCCGTTAGCTGGCATAAGATTCC
ATGCCTAGATGTGATACACGTTTCTGGAAACTGCCTCGTCATGCGACTGTTCCCC
GGGGTCAGGGCCGCTGGTATTTGCTGTAAAGAGGGGCGTTGAGTCCGTCCGACTT
CACTGCCCCCTTTCAGCCTTTTGGGTCCTGTATCCCAATTCTCAGAGGTCCCGCCG
TACGCTGAGGACCACCTGAAACGGGCATCGTCGCTCTTCGTTGTTCGTCGACTTCT
AGTGTGGAGACGAATTGCCAGAATTATTAACTGCGCAGTTAGGGCAGCGTCTGA
GGAAGTTTGCTGCGGTTTCGCCTTGACCGCGGGAAGGAGACATAACGATAGCGA
CTCTGTCTCAGGGGATCTGCATATGTTTGCAGCATACTTTAGGTGGGCCTTGGCTT
CCTTCCGCAGTCAAAACCGCGCAATTATCCCCGTCCTGATTTACTGGACTCGCAA
CGTGGGTCCATCAGTTGTCCGTATACCAAGACGTCTAAGGGCGGTGTACACCCTT
TTGAGCAATGATTGCACAACCTGCGATCACCTTATACAGAATTATCAATCAAGCT
CCCCGAGGAGCGGACTTGTAAGGACCGCCGCTTTCGCTCGGGTCTGCGGGTTATA
GCTTTTCAGTCTCGACGGGCTAGCACACATCTGGTTGACTAGGCGCATAGTCGCC
ATTCACAGATTTGCTCGGCAATCAGTACTGGTAGGCGTTAGACCCCGTGACTCGT
GGCTGAACGGCCGTACAACTCGACAGCCGGTGCTTGCGTTTTACCCTTAAAAAAA
AAAAAAAAAAAAAAAAA

Spike-in 12

TATTGGTGGAGGGGCACAAGTTGCTGAAGTTGCGAGAGGGGCGATAAGTGAGGC
AGACAGGCATAATATAAGAGGGGAGAGAATTAGCGTAGATACTCTTCCAATAGT
TGGTGAAGAAAATTTATATGAGGCTGTTAAAGCTGTAGCAACTCTTCCACGAGTA
GGAATTTTAGTTTTAGCTGGCTCTTTAATGGGAGGGAAGATAACTGAAGCAGTTA
AAGAATTAAAGGAAAAGACTGGCATTCCCGTGATAAGCTTAAAGATGTTTGGCTC
TGTTCCTAAGGTTGCTGATTTGGTTGTTGGAGACCCATTGCAGGCAGGGGTTTTA
GCTGTTATGGCTATTGCTGAAACAGCAAAATTTGATATAAATAAGGTTAAAGGTA
GGGTGCTATAAAGATAATTTAATAATTTTTGATGAAACCGAAGCGTTAGCTTTGG
GTTATGAAACTCCATGATTTTCATTTAATTTTTTCCTATTAATTTTCTCCTAAAAAG
TTTCTTTAACATAAATAAGGTTAAAGGGAGAGCTCTATGATTGTCTTCAAAAATA
CAAAGATTATTGATGTATATACTGGAGAGGTTGTTAAAGGAAATGTTGCAGTTGA
GAGGGATAAAATATCCTTTGTGGATTTAAATGATGAAATTGATAAGATAATTGAA
AAAATAAAGGAGGATGTTAAAGTTATTGACTTAAAAGGAAAATATTTATCTCCAA
CATTTATAGATGGGCATATACATATAGAATCTTCCCATCTCATCCCATCAGAGTTT
GAGAAATTTGTATTAAAAAGCGGAGTTAGCAAAGTAGTTATAGACCCGCATGAA
ATAGCAAATATTGCTGGAAAAGAAGGAATTTTGTTTATGTTGAATGATGCCAAAA
TTTTAGATGTCTATGTTATGCTTCCTTCCTGTGTTCCAGCTACAAACTTAGAAACA
AGTGGAGCTGAGATTACAGCAGAGAATATTGAAGAACTCATTCTTTAGATAATGT
CTTAGGTTAAAAAAAAAAAAAAAAAAAAAA

4. Resource table

Table 7. Resource table.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Western Blot: Mouse monoclonal against Oct-3/4	Santa Cruz	Cat#sc-5279, RRID:AB_628051
Western Blot: Goat polyclonal against Sox2	Santa Cruz	Cat#sc17320, RRID:AB_2286684
Western Blot: Rabbit polyclonal against Nanog	Bethyl Laboratories	Cat#A300-397A RRID:AB_386108
Western Blot: Rabbit polyclonal against Histone 3	Abcam	Cat#ab1791, RRID:AB_302613
Western Blot: Mouse monoclonal against anti-alpha-Tubulin	Sigma Aldrich	Cat#T6199 RRID:AB_477583
Immunofluorescence: Mouse monoclonal against Oct-3/4	Santa Cruz	Cat#sc-5279 RRID:AB_628051
Immunofluorescence: Goat polyclonal against Sox2	Santa Cruz	Cat#sc17320, RRID:AB_2286684
ChIP: Goat polyclonal against Sox2	Neuromics	Cat#GT15098 RRID:AB_2195800
ChIP: Goat polyclonal against Oct-3/4	R&D	Cat#AF1759 RRID:AB_354975
Chemicals, peptides, and recombinant proteins		
Doxycycline	Sigma Aldrich	Cat#D9891-1G
4-thiouridine	Carbosynth	Cat#NT06186
Nextera Tn5 enzyme	Illumina	Cat#20034198
Formaldehyde 16% concentrate stock methanol-free	Thermo Fisher Scientific	Cat#28908
Critical commercial assays		
Ovation Universal RNA-seq System	NuGEN	Cat#0343-32
NEBNext® Ultra™ II DNA Library Prep Kit	NEB	Cat# E7370L
Illumina Tagment DNA Enzyme and Buffer large kit	Illumina	Cat# 20034198
Human LIF	Prepared in house	N/A
CHIR99021	Cayman Chemicals	13122
PD0325901	Biomol	103034-25
Experimental models: Cell lines		
Mouse: ESC_ZHBTc4: Pou5f1$^{BSD/Zeo}$; Tg(CAG-tTA); Tg(TetO-Pou5f1)	Niwa et al 2000	N/A
Saccharomyces cerevisiae strain: BY4741	Euroscarf	ACC-Y00000

VII. Other contribution

1. Analysis of ChIP-exo occupancy data for TBP and TFIIA

In this chapter, I present other contribution from collaboration that has already been published and the contents presented in here have been taken from:

Structures and implications of TBP–nucleosome complexes

Haibo Wang, Le Xiong, and Patrick Cramer
PNAS July 27, 2021 118 (30) e2108859118, doi: https://doi.org/10.1073/pnas.2108859118

A detailed list of items from the published manuscript can be found in Appendix (page 106).

Contribution: HW designed and conducted experiments and data analysis except for bioinformatic analysis, which was performed by LX. PC supervised research. HW and PC wrote the manuscript.

Material and methods

Published TBP (Spt15) and TFIIA (Toa1) ChIP-exo data, and yeast (*S. cerevisiae*) genome features, including four groups of protein-coding genes (RP, STM, TFO, UNB) and their +1 nucleosome dyad annotations were taken from (Rossi et al., 2021). TBP annotated peak file was downloaded from https://github.com/CEGRcode/2021-Rossi_Nature/blob/master/04_ChExMix_Peaks/Spt15_CX.bed, and TFIIA annotated peak file from https://github.com/CEGRcode/2021-Rossi_Nature/blob/master/04_ChExMix_Peaks/Toa1_CX.bed. The corresponding TBP and TFIIA filtered .bam files were downloaded from http://www.yeastepigenome.org with sample numbers 8,599 and 14,838, respectively. To identify genes that are targeted by TBP or TFIIA, we overlapped the annotated peaks (1,872 peaks for TBP, 1,587 peaks for TFIIA) with gene promoter regions (TSS +/− 200 bp). This resulted in 103 RP, 284 STM, 483 TFO, 291 UNB genes targeted by TBP and 103 RP, 283 STM, 342 TFO, 203 UNB genes targeted by TFIIA. For TBP and TFIIA coverage calculation, only Read_1 reads were used and the reads were extended to 8-bp length from the 5′ end (the exonuclease stop site). The strand-specific pile-up coverages were then calculated and combined. All genes were aligned at the corresponding +1 nucleosome dyads to generate the metagene plot. For heatmap, genes were sorted by the

distance between annotated peaks summit and +1 nucleosome dyads in a decrease manner. All processing was carried out using the R/Bioconductor environment.

Results

To examine the relative location of TBP and nucleosomes in vivo, we reanalyzed available genome-wide chromatin immunoprecipitation–exonuclease digestion (ChIP-exo) occupancy data for yeast TBP and TFIIA (Toa1 subunit), and genome-wide nucleosome positioning data. In metagene plots, TBP and TFIIA occupancy showed peaks ∼80 to 90 bp upstream of the +1 nucleosome dyad, near the edge of the nucleosome (**Figure 29**). A subset of inducible genes with TATA sequences in their promoters, however, showed high occupancy for TBP and TFIIA within the nucleosome near its upstream edge (**Figure 29A-B**), maybe reflecting TBP binding to a TATA box at SHL –6. However, genome-wide occupancy data were obtained from a cell population, and it is likely that most of the ChIP signal for TBP reflects PICs in a context when the nucleosome is depleted. Therefore, it remains to be studied further how TBP and chromatin remodelers cooperate to enable PIC assembly in vivo.

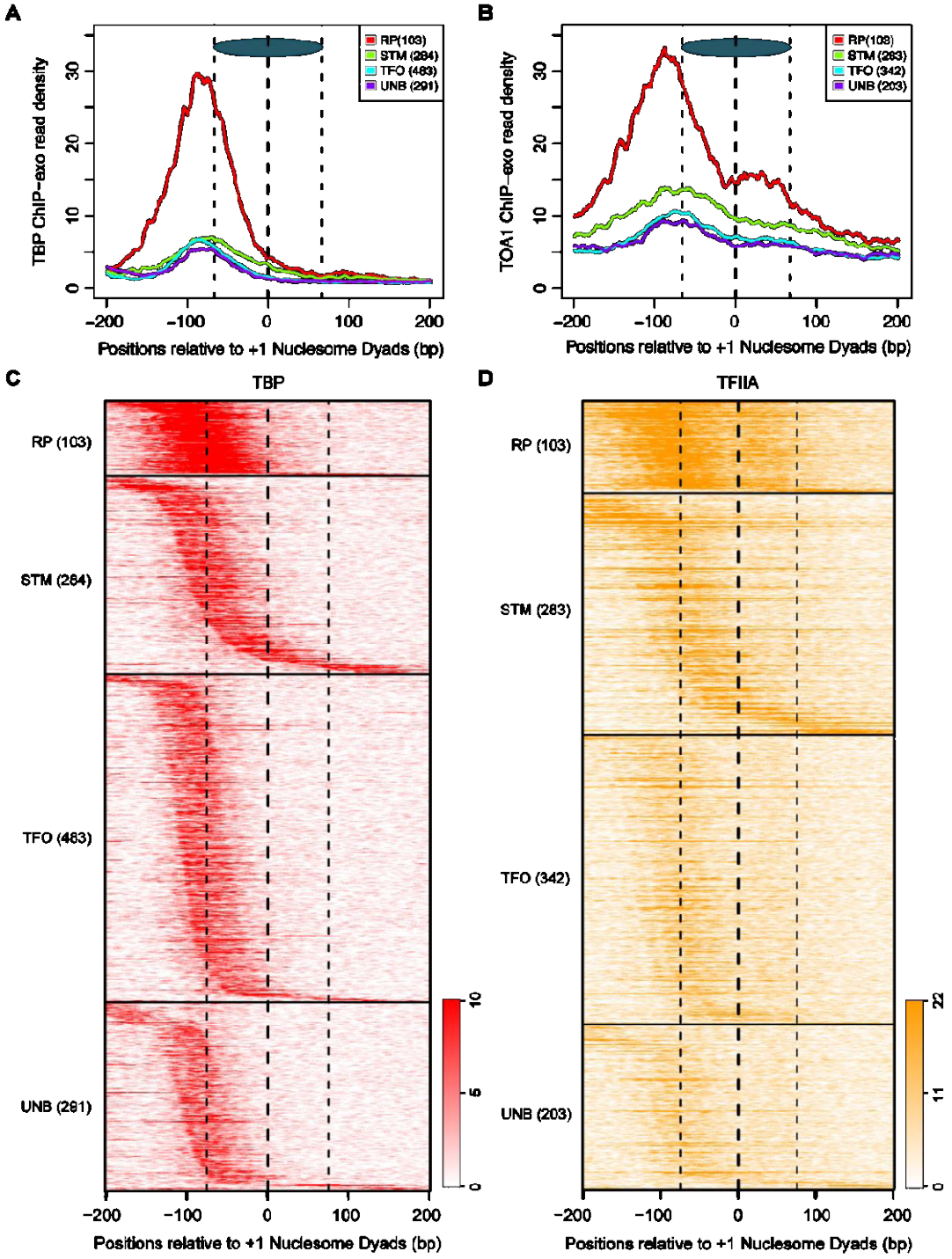


Figure 29. Distribution of TBP and TFIIA ChIP-exo genome-wide occupancy with respect to the location of the +1 nucleosome in yeast cells.

The blue oval denotes the nucleosome with its boundaries shown as vertical dashed lines. (A-B) Metagene plots of TBP (panel A) and TFIIA (panel B) occupancy around the +1 nucleosome for four subgroups of genes as follows: RP: genes encoding ribosomal proteins; STM: genes with promoters associated with inducibility, characteristically bound by sequence-specific transcription factors and major cofactor meta-assemblages SAGA, TUP and/or Mediator/SWI–SNF; TFO: genes with promoters that lacked STM cofactors but typically contained the insulator Abf1 or Reb1; UNB: genes with promoters that only showed bound PIC. (C-D) Heatmaps of TBP and TFIIA occupancy around the +1 nucleosome, sorted by the distance between the two for four subgroups of genes as in panels A and B.

References

(2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Aksoy, I., Jauch, R., Chen, J., Dyla, M., Divakar, U., Bogu, G.K., Teo, R., Leng Ng, C.K., Herath, W., Lili, S.*, et al.* (2013). Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. The EMBO journal *32*, 938-953.

Aljahani, A., Hua, P., Karpinska, M.A., Quililan, K., Davies, J.O.J., and Oudelaar, A.M. (2021). Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF. 2021.2008.2010.455796.

Alon, U. (2007). Network motifs: theory and experimental approaches. Nature reviews Genetics *8*, 450-461.

Ambrosetti, D.C., Basilico, C., and Dailey, L. (1997). Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. Molecular and cellular biology *17*, 6321-6329.

An, Z., Liu, P., Zheng, J., Si, C., Li, T., Chen, Y., Ma, T., Zhang, M.Q., Zhou, Q., and Ding, S. (2019). Sox2 and Klf4 as the Functional Core in Pluripotency Induction without Exogenous Oct4. Cell reports *29*, 1986-2000.e1988.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T.*, et al.* (2014a). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455-461.

Andersson, R., Refsing Andersen, P., Valen, E., Core, L.J., Bornholdt, J., Boyd, M., Heick Jensen, T., and Sandelin, A. (2014b). Nuclear stability and transcriptional directionality separate functionally distinct RNA species. Nature communications *5*, 5336.

Andersson, R., and Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. Nature reviews Genetics *21*, 71-87.

Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., Trono, D., Spitz, F., and Duboule, D. (2013). A switch between topological domains underlies HoxD genes collinearity in mouse limbs. Science (New York, NY) *340*, 1234167.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299-308.

Bates, L.E., Alves, M.R.P., and Silva, J.C.R. (2021). Auxin-degron system identifies immediate mechanisms of OCT4. Stem cell reports *16*, 1818-1831.

Benoist, C., and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promotor region. Nature *290*, 304-310.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nature reviews Genetics *15*, 163-175.

Biddle, J.W., Nguyen, M., and Gunawardena, J. (2019). Negative reciprocity, not ordered assembly, underlies the interaction of Sox2 and Oct4 on DNA. eLife *8*.

Blobel, G.A., Higgs, D.R., Mitchell, J.A., Notani, D., and Young, R.A. (2021). Testing the super-enhancer concept. Nature reviews Genetics.

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T., Marie-Nelly, H., McSwiggen, D.T., Kokic, G., Dailey, G.M., Cramer, P.*, et al.* (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. Nature structural & molecular biology *25*, 833-840.

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M.*, et al.* (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842-1855.e1816.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., and Furlong, E.E. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nature genetics *44*, 148-156.

Bose, D.A., Donahue, G., Reinberg, D., Shiekhattar, R., Bonasio, R., and Berger, S.L. (2017). RNA Binding to CBP Stimulates Histone Acetylation and Transcription. Cell *168*, 135-149.e122.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G.*, et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

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

Chang, Y.K., Srivastava, Y., Hu, C., Joyce, A., Yang, X., Zuo, Z., Havranek, J.J., Stormo, G.D., and Jauch, R. (2017). Quantitative profiling of selective Sox/POU pairing on hundreds of sequences in parallel by Coop-seq. Nucleic acids research *45*, 832-845.

Chen, J., Chen, X., Li, M., Liu, X., Gao, Y., Kou, X., Zhao, Y., Zheng, W., Zhang, X., Huo, Y.*, et al.* (2016). Hierarchical Oct4 Binding in Concert with Primed Epigenetic Rearrangements during Somatic Cell Reprogramming. Cell reports *14*, 1540-1554.

Chen, J., Zhang, Z., Li, L., Chen, B.C., Revyakin, A., Hajj, B., Legant, W., Dahan, M., Lionnet, T., Betzig, E.*, et al.* (2014). Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. Cell *156*, 1274-1285.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J.*, et al.* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106-1117.

Cho, W.K., Spille, J.H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, II (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science (New York, NY) *361*, 412-415.

Choi, J., Lysakovskaia, K., Stik, G., Demel, C., Söding, J., Tian, T.V., Graf, T., and Cramer, P. (2021). Evidence for additive and synergistic action of mammalian enhancers during cell fate determination. eLife *10*.

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X.*, et al.* (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science (New York, NY) *361*.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell *168*, 442-459.e420.

Corces, M.R., and Corces, V.G. (2016). The three-dimensional cancer genome. Current opinion in genetics & development *36*, 1-7.

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nature genetics *46*, 1311-1320.

Cramer, P. (2019). Organization and regulation of gene transcription. Nature *573*, 45-54.

Davidson, I.F., Bauer, B., Goetz, D., Tang, W., Wutz, G., and Peters, J.M. (2019). DNA loop extrusion by human cohesin. Science (New York, NY) *366*, 1338-1345.

Davies, J.O., Telenius, J.M., McGowan, S.J., Roberts, N.A., Taylor, S., Higgs, D.R., and Hughes, J.R. (2016). Multiplexed analysis of chromosome conformation at vastly improved sensitivity. Nature methods *13*, 74-80.

de la Serna, I.L., Ohkawa, Y., and Imbalzano, A.N. (2006). Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. Nature reviews Genetics *7*, 461-473.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS biology *8*, e1000384.

de Wit, E., Bouwman, B.A., Zhu, Y., Klous, P., Splinter, E., Verstegen, M.J., Krijger, P.H., Festuccia, N., Nora, E.P., Welling, M.*, et al.* (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature *501*, 227-231.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell *149*, 1233-1244.

Deng, W., Rupon, J.W., Krivega, I., Breda, L., Motta, I., Jahn, K.S., Reik, A., Gregory, P.D., Rivella, S., Dean, A.*, et al.* (2014). Reactivation of developmentally silenced globin genes by forced chromatin looping. Cell *158*, 849-860.

Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M., and Plath, K. (2013). Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. Cell stem cell *13*, 602-616.

Di Giammartino, D.C., Kloetgen, A., Polyzos, A., Liu, Y., Kim, D., Murphy, D., Abuhashem, A., Cavaliere, P., Aronson, B., Shah, V.*, et al.* (2019). KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. Nature cell biology *21*, 1179-1190.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Do, D.V., Ueda, J., Messerschmidt, D.M., Lorthongpanich, C., Zhou, Y., Feng, B., Guo, G., Lin, P.J., Hossain, M.Z., Zhang, W.*, et al.* (2013). A genetic and developmental pathway from STAT3 to the OCT4-NANOG circuit is essential for maintenance of ICM lineages in vivo. Genes & development *27*, 1378-1390.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Dodonova, S.O., Zhu, F., Dienemann, C., Taipale, J., and Cramer, P. (2020). Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. Nature *580*, 669-672.

El Khattabi, L., Zhao, H., Kalchschmidt, J., Young, N., Jung, S., Van Blerkom, P., Kieffer-Kwon, P., Kieffer-Kwon, K.R., Park, S., Wang, X.*, et al.* (2019). A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. Cell *178*, 1145-1158.e1120.

Feng, B., Jiang, J., Kraus, P., Ng, J.H., Heng, J.C., Chan, Y.S., Yaw, L.P., Zhang, W., Loh, Y.H., Han, J.*, et al.* (2009). Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. Nature cell biology *11*, 197-203.

Fischer, A., Virelizier, J.L., Arenzana-Seisdedos, F., Perez, N., Nezelof, C., and Griscelli, C. (1985). Treatment of four patients with erythrophagocytic lymphohistiocytosis by a combination of epipodophyllotoxin, steroids, intrathecal methotrexate, and cranial irradiation. Pediatrics *76*, 263-268.

Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D.*, et al.* (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic acids research *48*, D87-d92.

Frankish, A., Diekhans, M., Ferreira, A.M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J.*, et al.* (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic acids research *47*, D766-d773.

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. eLife *8*.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell reports *15*, 2038-2049.

Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. Science (New York, NY) *361*, 1341-1345.

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature *512*, 96-100.

Golfier, S., Quail, T., Kimura, H., and Brugués, J. (2020). Cohesin and condensin extrude DNA loops in a cell cycle-dependent manner. eLife *9*.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics (Oxford, England) *27*, 1017-1018.

Gressel, S., Schwalb, B., and Cramer, P. (2019). The pause-initiation limit restricts transcription activation in human cells. Nature communications *10*, 3603.

Haarhuis, J.H.I., van der Weide, R.H., Blomen, V.A., Yáñez-Cuna, J.O., Amendola, M., van Ruiten, M.S., Krijger, P.H.L., Teunissen, H., Medema, R.H., van Steensel, B.*, et al.* (2017).

The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell *169*, 693-707.e614.

Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. Genome research *23*, 1210-1223.

Hall, J., Guo, G., Wray, J., Eyres, I., Nichols, J., Grotewold, L., Morfopoulou, S., Humphreys, P., Mansfield, W., Walker, R.*, et al.* (2009). Oct4 and LIF/Stat3 additively induce Krüppel factors to sustain embryonic stem cell self-renewal. Cell stem cell *5*, 597-609.

Hanssen, L.L.P., Kassouf, M.T., Oudelaar, A.M., Biggs, D., Preece, C., Downes, D.J., Gosden, M., Sharpe, J.A., Sloane-Stanley, J.A., Hughes, J.R.*, et al.* (2017). Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. Nature cell biology *19*, 952-961.

Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. Genes & development *32*, 26-41.

Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F.*, et al.* (2006). The UCSC Genome Browser Database: update 2006. Nucleic acids research *34*, D590-598.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934-947.

Hnisz, D., Day, D.S., and Young, R.A. (2016). Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. Cell *167*, 1188-1200.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13-23.

Ho, Y., Elefant, F., Liebhaber, S.A., and Cooke, N.E. (2006). Locus control region transcription plays an active role in long-range gene activation. Molecular cell *23*, 365-375.

Hsieh, C.L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C.J., Lee, G.S., Chen, S.*, et al.* (2014). Enhancer RNAs participate in androgen receptor-driven looping that

selectively enhances gene activation. Proceedings of the National Academy of Sciences of the United States of America *111*, 7319-7324.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research *37*, 1-13.

Huber, W., Toedling, J., and Steinmetz, L.M. (2006). Transcript mapping with high-density oligonucleotide tiling arrays. Bioinformatics (Oxford, England) *22*, 1963-1970.

Jaeger, M.G., Schwalb, B., Mackowiak, S.D., Velychko, T., Hanzl, A., Imrichova, H., Brand, M., Agerer, B., Chorn, S., Nabet, B.*, et al.* (2020). Selective Mediator dependence of cell-type-specifying transcription. Nature genetics *52*, 719-727.

Jerabek, S., Merino, F., Schöler, H.R., and Cojocaru, V. (2014). OCT4: dynamic DNA binding pioneers stem cell pluripotency. Biochimica et biophysica acta *1839*, 138-154.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C.*, et al.* (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Molecular cell *51*, 310-325.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049-1061.

Kim, J.B., Greber, B., Araúzo-Bravo, M.J., Meyer, J., Park, K.I., Zaehres, H., and Schöler, H.R. (2009a). Direct reprogramming of human neural stem cells by OCT4. Nature *461*, 649-643.

Kim, J.B., Sebastiano, V., Wu, G., Araúzo-Bravo, M.J., Sasse, P., Gentile, L., Ko, K., Ruau, D., Ehrich, M., van den Boom, D.*, et al.* (2009b). Oct4-induced pluripotency in adult neural stem cells. Cell *136*, 411-419.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S.*, et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182-187.

Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J., and Yu, H. (2019). Human cohesin compacts DNA by loop extrusion. Science (New York, NY) *366*, 1345-1349.

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

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C.*, et al.* (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. Nature structural & molecular biology *18*, 956-963.

Lam, C.S., Mistri, T.K., Foo, Y.H., Sudhaharan, T., Gan, H.T., Rodda, D., Lim, L.H., Chou, C., Robson, P., Wohland, T.*, et al.* (2012). DNA-dependent Oct4-Sox2 interaction and diffusion properties characteristic of the pluripotent cell state revealed by fluorescence spectroscopy. The Biochemical journal *448*, 21-33.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods *9*, 357-359.

Lewis, M.W., Li, S., and Franco, H.L. (2019). Transcriptional control by enhancers and enhancer RNAs. Transcription *10*, 171-186.

Li, D., Liu, J., Yang, X., Zhou, C., Guo, J., Wu, C., Qin, Y., Guo, L., He, J., Yu, S.*, et al.* (2017). Chromatin Accessibility Dynamics during iPSC Reprogramming. Cell stem cell *21*, 819-833.e816.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X.*, et al.* (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature *498*, 516-520.

Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nature reviews Genetics *17*, 207-223.

Lidschreiber, K., Jung, L.A., von der Emde, H., Dave, K., Taipale, J., Cramer, P., and Lidschreiber, M. (2021). Transcriptionally active enhancers in human cancer cells. Molecular systems biology *17*, e9873.

Ling, J., Ainol, L., Zhang, L., Yu, X., Pi, W., and Tuan, D. (2004). HS2 enhancer function is blocked by a transcriptional terminator inserted between the enhancer and the promoter. The Journal of biological chemistry *279*, 51704-51713.

Liu, N.Q., Maresca, M., van den Brand, T., Braccioli, L., Schijns, M., Teunissen, H., Bruneau, B.G., Nora, E.P., and de Wit, E. (2021). WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. Nature genetics *53*, 100-109.

Liu, Z., Legant, W.R., Chen, B.C., Li, L., Grimm, J.B., Lavis, L.D., Betzig, E., and Tjian, R. (2014). 3D imaging of Sox2 enhancer clusters in embryonic stem cells. eLife *3*, e04236.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J.*, et al.* (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nature genetics *38*, 431-440.

Lopez-Delisle, L., Rabbani, L., Wolff, J., Bhardwaj, V., Backofen, R., Grüning, B., Ramírez, F., and Manke, T. (2021). pyGenomeTracks: reproducible plots for multivariate genomic datasets. Bioinformatics (Oxford, England) *37*, 422-423.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology *15*, 550.

Lubas, M., Andersen, P.R., Schein, A., Dziembowski, A., Kudla, G., and Jensen, T.H. (2015). The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. Cell reports *10*, 178-192.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Malik, V., Glaser, L.V., Zimmer, D., Velychko, S., Weng, M., Holzner, M., Arend, M., Chen, Y., Srivastava, Y., Veerapandian, V.*, et al.* (2019). Pluripotency reprogramming by

competent and incompetent POU factors uncovers temporal dependency for Oct4 and Sox2. Nature communications *10*, 3477.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J.*, et al.* (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521-533.

Martin, E.W., and Mittag, T. (2018). Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. Biochemistry *57*, 2478-2487.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 *17*, 3.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nature biotechnology *28*, 495-501.

Melgar, M.F., Collins, F.S., and Sethupathy, P. (2011). Discovery of active enhancers through bidirectional expression of short transcripts. Genome biology *12*, R113.

Merino, F., Ng, C.K.L., Veerapandian, V., Schöler, H.R., Jauch, R., and Cojocaru, V. (2014). Structural basis for the SOX-dependent genomic redistribution of OCT4 in stem cell differentiation. Structure (London, England : 1993) *22*, 1274-1286.

Michael, A.K., Grand, R.S., Isbel, L., Cavadini, S., Kozicka, Z., Kempf, G., Bunker, R.D., Schenk, A.D., Graff-Meyer, A., Pathare, G.R.*, et al.* (2020). Mechanisms of OCT4-SOX2 motif readout on nucleosomes. Science (New York, NY) *368*, 1460-1465.

Michel, M., Demel, C., Zacher, B., Schwalb, B., Krebs, S., Blum, H., Gagneur, J., and Cramer, P. (2017). TT-seq captures enhancer landscapes immediately after T-cell stimulation. Molecular systems biology *13*, 920.

Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Molecular cell *51*, 606-617.

Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nature methods *13*, 919-922.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nature biotechnology *26*, 101-106.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379-391.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nature genetics *24*, 372-376.

Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature *460*, 118-122.

Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930-944.e922.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J.*, et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381-385.

Ørom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q.*, et al.* (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46-58.

Oudelaar, A.M., and Higgs, D.R. (2021). The relationship between genome structure and function. Nature reviews Genetics *22*, 154-168.

Paliou, C., Guckelberger, P., Schöpflin, R., Heinrich, V., Esposito, A., Chiariello, A.M., Bianco, S., Annunziatella, C., Helmuth, J., Haas, S.*, et al.* (2019). Preformed chromatin

topology assists transcriptional robustness of Shh during limb development. Proceedings of the National Academy of Sciences of the United States of America *116*, 12390-12399.

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., Federation, A., Chao, J., Elliott, O., Liu, Z.P.*, et al.* (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. Cell *161*, 774-789.

Pnueli, L., Rudnizky, S., Yosefzon, Y., and Melamed, P. (2015). RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin α-subunit gene. Proceedings of the National Academy of Sciences of the United States of America *112*, 4369-4374.

Pott, S., and Lieb, J.D. (2015). What are super-enhancers? Nature genetics *47*, 8-12.

Rao, S.S.P., Huang, S.C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D.*, et al.* (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305-320.e324.

Rivera, C.M., and Ren, B. (2013). Mapping human epigenomes. Cell *155*, 39-55.

Rodda, D.J., Chew, J.L., Lim, L.H., Loh, Y.H., Wang, B., Ng, H.H., and Robson, P. (2005). Transcriptional regulation of nanog by OCT4 and SOX2. The Journal of biological chemistry *280*, 24731-24737.

Rossi, M.J., Kuntala, P.K., Lai, W.K.M., Yamada, N., Badjatia, N., Mittal, C., Kuzu, G., Bocklund, K., Farrell, N.P., Blanda, T.R.*, et al.* (2021). A high-resolution protein architecture of the budding yeast genome. Nature *592*, 309-314.

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C.*, et al.* (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science (New York, NY) *361*.

Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J.*, et al.* (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proceedings of the National Academy of Sciences of the United States of America *112*, E6456-6465.

Sartorelli, V., and Lauberth, S.M. (2020). Enhancer RNAs are an important regulatory layer of the epigenome. Nature structural & molecular biology *27*, 521-528.

Schaukowitch, K., Joo, J.Y., Liu, X., Watts, J.K., Martinez, C., and Kim, T.K. (2014). Enhancer RNA facilitates NELF release from immediate early genes. Molecular cell *56*, 29-42.

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. Nature reviews Genetics *20*, 437-455.

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science (New York, NY) *352*, 1225-1228.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C.H., Mirny, L.*, et al.* (2017). Two independent modes of chromatin organization revealed by cohesin removal. Nature *551*, 51-56.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458-472.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nature reviews Genetics *15*, 272-286.

Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science (New York, NY) *350*, 978-981.

Smallwood, A., and Ren, B. (2013). Genome organization and long-range regulation of gene expression by enhancers. Current opinion in cell biology *25*, 387-394.

Song, S.H., Hou, C., and Dean, A. (2007). A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. Molecular cell *28*, 810-822.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994-1004.

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell *161*, 555-568.

Spielmann, M., Lupiáñez, D.G., and Mundlos, S. (2018). Structural variation in the 3D genome. Nature reviews Genetics *19*, 453-467.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nature reviews Genetics *13*, 613-626.

Sterneckert, J., Höing, S., and Schöler, H.R. (2012). Concise review: Oct4 and more: the reprogramming expressway. Stem cells (Dayton, Ohio) *30*, 15-21.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nature structural & molecular biology *14*, 103-105.

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

Tan, C., and Takada, S. (2020). Nucleosome allostery in pioneer transcription factor binding. Proceedings of the National Academy of Sciences of the United States of America *117*, 20586-20596.

Tantin, D. (2013). Oct transcription factors in development and stem cells: insights and mechanisms. Development (Cambridge, England) *140*, 2857-2866.

Tuan, D., Kong, S., and Hu, K. (1992). Transcription of the hypersensitive site HS2 enhancer in erythroid cells. Proceedings of the National Academy of Sciences of the United States of America *89*, 11219-11223.

Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., and Blobel, G.A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Molecular cell *17*, 453-462.

van den Berg, D.L., Zhang, W., Yates, A., Engelen, E., Takacs, K., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2008). Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression. Molecular and cellular biology *28*, 5986-5995.

Velychko, S., Adachi, K., Kim, K.P., Hou, Y., MacCarthy, C.M., Wu, G., and Schöler, H.R. (2019). Excluding Oct4 from Yamanaka Cocktail Unleashes the Developmental Potential of iPSCs. Cell stem cell *25*, 737-753.e734.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F.*, et al.* (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854-858.

Wachutka, L., Caizzi, L., Gagneur, J., and Cramer, P. (2019). Global donor and acceptor splicing site kinetics in human cells. eLife *8*.

Wang, J., Yu, H., Ma, Q., Zeng, P., Wu, D., Hou, Y., Liu, X., Jia, L., Sun, J., Chen, Y.*, et al.* (2021). Phase separation of OCT4 controls TAD reorganization to promote cell fate transitions. Cell stem cell *28*, 1868-1883.e1811.

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L.*, et al.* (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573-1588.e1528.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307-319.

Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D.A., Stocsits, R.R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M.J.*, et al.* (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. The EMBO journal *36*, 3573-3599.

Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J.A., and Kosik, K.S. (2009). MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. Cell *137*, 647-658.

Ying, Q.L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281-292.

Young, R.A. (2011). Control of the embryonic stem cell state. Cell *144*, 940-954.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R.*, et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science (New York, NY) *318*, 1917-1920.

Yuan, H., Corbi, N., Basilico, C., and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes & development *9*, 2635-2645.

Zacher, B., Michel, M., Schwalb, B., Cramer, P., Tresch, A., and Gagneur, J. (2017). Accurate Promoter and Enhancer Identification in 127 ENCODE and Roadmap Epigenomics Cell Types and Tissues by GenoSTAN. PloS one *12*, e0169249.

Zaret, K.S. (2020). Pioneer Transcription Factors Initiating Gene Network Changes. Annual review of genetics *54*, 367-385.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes & development *25*, 2227-2241.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.

Zhao, F.Q. (2013). Octamer-binding transcription factors: genomics and functions. Frontiers in bioscience (Landmark edition) *18*, 1051-1071.

Zhu, Y., Sun, L., Chen, Z., Whitaker, J.W., Wang, T., and Wang, W. (2013). Predicting enhancer transcription and activity from chromatin modifications. Nucleic acids research *41*, 10032-10043.

Appendix

Listed Contents

List of figures

List of tables

List of items from the manuscript

The following are listed figures, tables and chapters that are taken from the published manuscript:

Le Xiong*, Erik A. Tolen* et al. Oct4 primarily controls enhancer activity rather than accessibility

(* joint first authorship)

Manuscript in revision, doi: https://doi.org/10.1101/2021.06.28.450119

A detailed author contributions can be found in **Section III. Results (page 37)**.

Figures

Figure 7 corresponds to Figure 1

Figure 8 corresponds to Figure 1-figure supplement 1

Figure 9A-B correspond to Figure 2-figure supplement 1A-B

Figure 9C-D correspond to Figure 2A-B

Figure 9E corresponds to Figure 2C and Figure 2-figure supplement 1C

Figure 9F corresponds to Figure 2D

Figure 9G corresponds to Figure 2E

Figure 10A-F correspond to Figure 2-figure supplement 1D-I

Figure 11A-C correspond to Figure 2F-H

Figure 12A corresponds to Figure 3A

Figure 12B-H correspond to Figure 3-figure supplement 1A-G

Figure 13A corresponds to Figure 3B

Figure 13B corresponds to Figure 3C and Figure 3-figure supplement 1H

Figure 13C corresponds to Figure 3-figure supplement 1I

Figure 13D corresponds to Figure 3D

Figure 13E-F correspond to Figure 3-figure supplement 1J-K

Figure 14 corresponds to Figure 4-figure supplement 1

Figure 15A-D correspond to Figure 4A-D

Figure 16A-E correspond to Figure 4E-I

Figure 17A-B correspond to Figure 5-figure supplement 1A-B

Figure 17C-E correspond to Figure 5A-C

Figure 19A-D correspond to Figure 5D-G

Figure 18E-F correspond to Figure 5-figure supplement 1C-D

Figure 19 corresponds to Figure 6

Figure 20A-F correspond to Figure 6-figure supplement 1A-F

Figure 21A-E correspond to Figure 6-figure supplement 1G-K

Figure 22A-D correspond to Figure 7-figure supplement 1A-D

Figure 22E-F correspond to Figure 7A-B

Figure 23A-B correspond to Figure 7C-D

Figure 23C-E correspond to Figure 7-figure supplement 1E-G

Tables

Table1 corresponds to Table S1

Table2 corresponds to Table S2

Table3 corresponds to Table S5

Table4 corresponds to Table S6

Table5 corresponds to Table S4

Chapters

The following sections in this thesis were taken from the referred manuscript:

Part of Summary

Chapter II. Methods and Materials (sections 1.1-1.7, 2.1, 2.3-2.11, 2.13)

Chapter III. Results (sections 1-9)

Chapter IV. Discussion (part of sections 3-5)

The following are listed figures and sections that are taken from the published manuscript:

Structures and implications of TBP–nucleosome complexes
Haibo Wang, Le Xiong, and Patrick Cramer
PNAS July 27, 2021 118 (30) e2108859118, doi: https://doi.org/10.1073/pnas.2108859118

A detailed author contributions can be found in **section VII. Other contribution (page 81)**.

Figures

Figure 29 corresponds to Figure S7

Sections

Material and Methods, Results were taken from the referred manuscript.

Abbreviations

4sU	4-thiouridine
asRNA	antisense RNA
ATAC-seq	assay for transposase accessible chromatin with sequencing
bps	base pairs
BMP	bone morphogenic protein
Brg1	brahma-related gene 1
ChIP-Chip	chromatin immunoprecipitation with DNA microarrays
ChIP-exo	chromatin immunoprecipitation-exonuclease digestion
ChIP-seq	chromatin immunoprecipitation and sequencing
conRNA	convergent RNA
CTCF	CCCTC-binding factor
CTD	C terminal domain
DNA	deoxyribonucleic acid
DOX	doxycycline
dpc	days post coitum
EMSA	electrophoretic mobility shift assays
ERCC	External RNA Controls Consortium
eRNA	enhancer RNA
ESCs	embryonic stem cells
Esrrb	estrogen-related receptor beta
FRAP	fast recovery after photobleaching
GO	gene ontology
H3K27ac	histone H3 lysine residue 27 acetylation
H3K4me1	histone 3 lysine 4 monomethylation
HAT	histone acetyltransferase
HMG	high mobility domain
ICM	inner cell mass
IDR	intrinsically disordered region
incRNA	intergenic RNA
iPSCs	induced pluripotent cells
Klf4	Krüppel-like factor 4
LCR	locus control regions
LIF	leukemia inhibitory factor

LLPS	liquid-liquid phase separation
Med1	mediator subunit 1
mESCs	mouse embryonic stem cells
min	minutes
ncRNA	non-coding RNA
NELF	negative elongation factor
Oct4	octamer-binding transcription factor 4
PCA	principal component analysis
Pol II	DNA-dependent RNA polymerase II
POU_{HD}	POU homeodomain
POU_S	POU specific domain
RNA	ribonucleic acid
SE	super enhancer
SHL	super helix location
Smad1	SMAD Family Member 1
Sox2	SRY (sex determining region Y)-box 2
Stat3	signal transducer and activator of transcription 3
SWI/SNF	SWItch/Sucrose Non-Fermentable
TADs	topological associating domains
TBP	TATA-box binding protein
TE	typical enhancer
TF	transcription factor
TSS	transcription start site
TT-seq	transient transcriptome sequencing
TU	transcription unit
uaRNA	upstream antisense RNA
YY1	Yin Yang 1