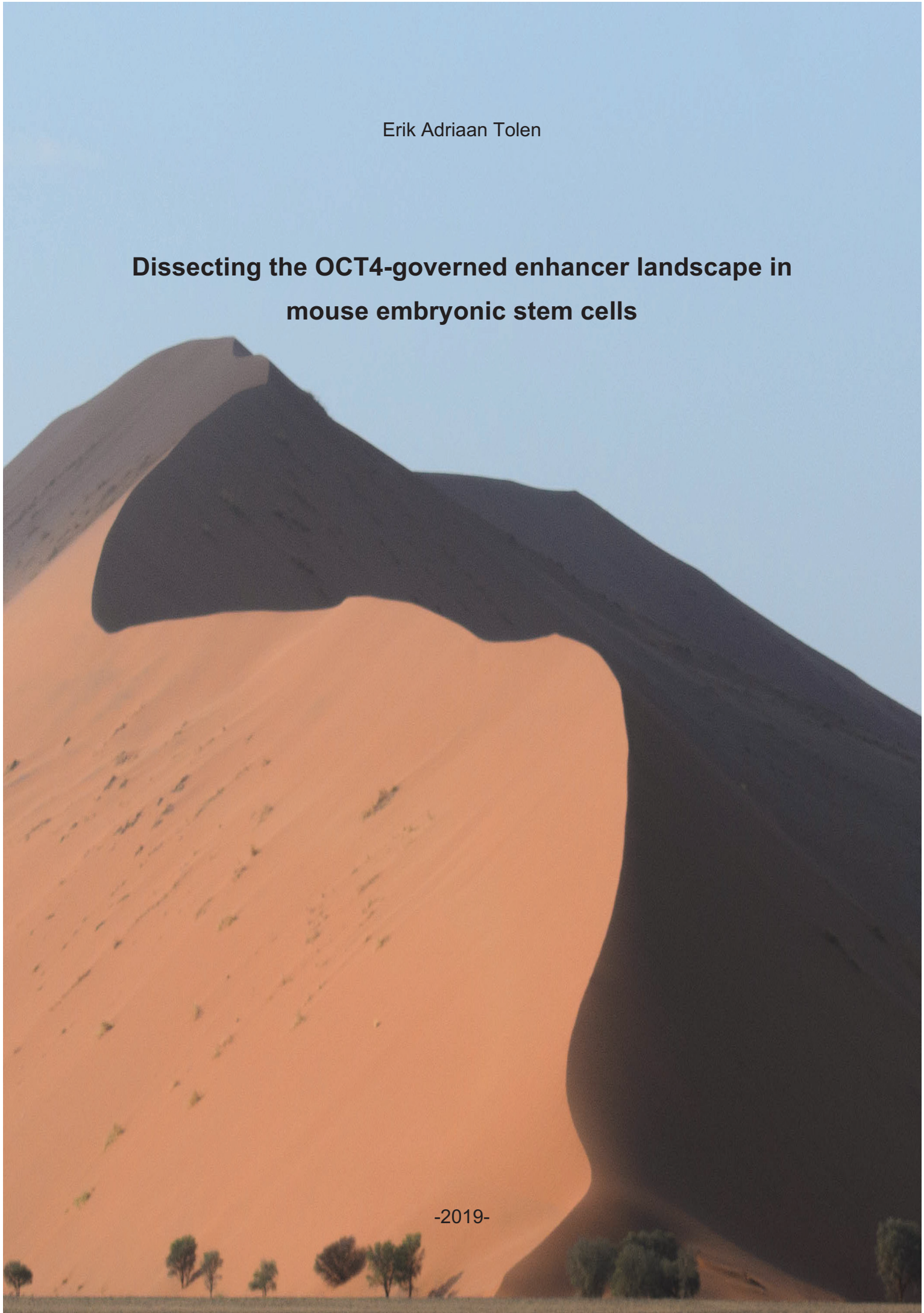Erik Adriaan Tolen

Dissecting the OCT4-governed enhancer landscape in mouse embryonic stem cells

-2019-

Biologie

Dissecting the OCT4-governed enhancer landscape in mouse embryonic stem cells

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften im Fachbereich Biologie

der Mathematisch-Naturwissenschaftlichen Fakultät

der Westfälischen Wilhelms-Universität Münster

Max-Planck-Institut für Molekulare Biomedizin

Münster, Deutschland

Vorgelegt von

Erik Adriaan Tolen

aus Amstelveen, Niederlande

- 2019-

Dekanin: Prof. Dr. Susanne Fetzner

Erster Gutachter/Prüfer: Prof. Dr. Hans R. Schöler

Zweiter Gutachter/Prüfer: Prof. Dr. Michael Hippler

Tag der mündlichen Prüfung: 20th März 2019

Tag der Promotion: 12 April 2019

Curriculum vitae

Acknowledgements

I am sincerely grateful to Prof. Dr. Hans R Schöler for providing me with the opportunity to work in his laboratory, for his continuous support and for letting me explore my own ideas.

I would like to thank Prof. Dr. Patrick Cramer for the opportunity to collaborate with his laboratory and utilize the TT-seq method as well as the resources to explore the genomics data obtained this project.

My deepest appreciation to Dr. Kenjiro Adachi for the countless hours of discussing the results and experimental setups and for his support to overcome all obstacles along the way.

I would like to thank Dr. Jinmi Choi for her guidance during the TT-seq experiments and all help with the analysis of the genomics data. Her input has been instrumental for the data analysis.

I would like to thank Le Xiong for data analysis and fun discussions regarding the results, which contributed to improve the project.

A big thank you to all the colleagues in the Department for Cell and Developmental Biology at the Max-Planck-Institute for Molecular Biomedicine for the support and discussions over the years. And for the fun we had in and outside of the lab.

I am thankful to my parents and sister for their continuous belief in me. Mama, papa, zus, dankjulliewel!

I am thankful to my parents-in-law for all of their support.

I am forever grateful to my wife and my daughter who are my inspiration and whose love has helped me to overcome any difficulty during my PhD. Without them I would not have managed to finish the tough journey and survived the long days in the lab.

Table of contents

Abstract

The transcriptional network of mouse embryonic stem cells (mESC) is governed by the three transcription factors (TF) OCT4, SOX2 and NANOG. Besides governing the mESC identity these TF are also involved in maintaining the accessibility of the genome. Alterations in the OCT4 expression levels are known to precede the differentiation of mESC into primitive endoderm or trophectoderm. OCT4 regulates gene expression by binding to distal cis-regulatory elements called enhancers. Enhancers are defined by specific histone modification patterns, TF occupancy and chromatin accessibility. Recently, it was shown that active enhancers synthesize non-coding RNAs (ncRNAs) called enhancer RNAs (eRNAs). The detailed molecular mechanisms underlying the regulation of transcription by OCT4 through enhancer activity have not been elucidated so far.

Most studies investigating OCT4 function in transcriptional regulation have focused on the use of genome wide binding data from chromatin immunoprecipitation followed by next generation sequencing (ChIP-seq) experiments. A large number of OCT4-occupied genomic locations have been mapped, but it remains to be determined which of these binding sites fulfils a functional role.

We utilized an OCT4 loss-of-function model in conjunction with 'transient transcriptome sequencing' (TT-seq) to increase our understanding of OCT4 and its regulatory role in the pluripotency network. The TT-seq method allows for the analysis of RNA synthesis and detection of instable RNA transcripts such as eRNA in a kinetic manner. In addition, we performed the 'assay for transposase-accessible chromatin' (ATAC-seq) on the OCT4 loss-of-function model to monitor changes to the chromatin accessibility. We incorporated publicly available ChIP-seq data into our TT-seq and ATAC-seq data sets. This experimental setup allowed us to study changes to newly synthesized RNA as well as alterations to chromatin accessibility and to integrate OCT4 binding sites to identify OCT4 regulated regions. In addition, incorporation of SOX2 and NANOG ChIP-seq data enabled analysis of TF cooperativity between OCT4, SOX2 and/or NANOG.

At the transcriptional level we found many differentially synthesized eRNA. We also observed significant changes to chromatin accessibility. Strikingly, the OCTSOX composite motif was enriched in regions that show a decrease in chromatin accessibility upon loss of OCT4. TF cooperativity was found to be highest at OCT4-

bound sites with eRNA synthesis. Curiously, only 17% of OCT4 binding events overlapped with transcriptional activity.

In-depth analysis of OCT4 bound sites that overlap with eRNA synthesis unveiled four distinct clusters displaying specific transcription dynamics and chromatin accessibility kinetics. The clusters with the strongest loss in eRNA synthesis and chromatin accessibility also show an enrichment for the OCTSOX composite motif. There was a greater distance between OCT4 and SOX2 binding sites at loci that do not show striking changes in transcriptional activity and chromatin accessibility upon OCT4 loss.

Taken together, this study reports novel insights into OCT4-mediated molecular mechanisms at the transcriptional and chromatin accessibility levels in the pluripotent stem cell regulatory network.

Abbreviations

°C	degree centigrade
1D	unidirectional
2D	bidirectional
4sU	4-thiouridine
4sU-seq	4-thiouridine sequencing
5-caC	5-carboxylcytosine
5-fC	5-fluorocystosine
5-hmC	5-hydroxymethylcytosine
5-mC	5-methylcytosine
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
BAF	Brg1 associated factor complex (protein, mouse, according to international protein nomenclature guidelines)
bdHMM	bi-directional Hidden Markov Models
BMP	Bone morphogenic protein (protein, mouse, according to international protein nomenclature guidelines)
BRE	B recognition element
BRG1	Brahma-related gene 1 (protein, mouse, according to international protein nomenclature guidelines)
BSA	Bovine serum albumin
CAGE-seq	Capped analysis of gene expression using sequencing
CBP/p300	CREB-binding protein – E1A binding protein p300 (Protein)
CDK8	Cyclin dependent kinase 8 (protein, mouse, according to international protein nomenclature guidelines)
cDNA	Copy DNA
ce eRNA	core enhancer eRNA
CHD	Chromodomain helicase DNA binding
ChIP-qPCR	Chromatin immunoprecipitation followed by quantitative polymerase chain reaction
cm	Centimeter
conRNA	Convergent RNA
CR2/3	*Pou5f1* proximal enhancer
CR4	*Pou5f1* distal enhancer
CT	Chromosome territories

CTCF	11-zinc finger protein or CCCTC-binding factor (protein, mouse, according to international protein nomenclature guidelines)
Da	Dalton
DAM-id	DNA adenine methyltransferase identification
DDR eRNA	DDR enhancer eRNA
DEG	Differentially expressed genes
DIP-seq	DNA immunoprecipitation sequencing
DNA	Deoxyribonucleic acid
DNMT	DNA methyl transferase
dNTP	deoxy nucleotide triphosphate
DOX	Doxycycline
Dpc	Days post coitum
ECC	Embryonic carcinoma cell
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
eExons	Enhancer exons
EGTA	Ethyleneglycol-bis(β-aminoethylesther)-N,N,N',N'-tetraacetic acid
EMSA	Electrophoretic mobility shift assay
EPI	Epiblast, primitive ectoderm
ERK	Extracellular signal-regulated kinase
eRNA	Enhancer RNA
ESC	Embryonic stem cell
Esrrb	Estrogen-related receptor beta (Gene symbol)
ESRRB	Estrogen-related receptor beta (protein, mouse, according to international protein nomenclature guidelines)
FBS	Fetal bovine serum
g	Gravity
Gapdh	Glyceraldehyde 3-phosphate dehydrogenase (protein mouse, according to international protein nomenclature guidelines)
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase (protein mouse, according to international protein nomenclature guidelines)
GO	Gene ontology
GRO-seq	Global run-on followed by sequencing
GSK3	Glycogen synthase kinase 3

GTF	General transcription factors
H3	Histone 3
H3K4me1	Histone 3 lysine 4 monomethylation
H3K4me2	Histone 3 lysine 4 dimethylation
H3K4me3	Histone 3 lysine 4 trimethylation
H3K27ac	Histone 3 lysine 27 acetylation
HCl	Hydrogen chloride
HCP	CpG island promoter
HDAC	Histone deacetylases
HMG	High mobility group
HRP	Horseradish peroxidase
ICM	Inner cell mass
Id	Inhibitor of differentiation (Gene symbol)
IF	Immunofluorescence
IFNβ	Interferon beta (Protein, mouse, according to international protein nomenclature guidelines)
JAK-STAT	Janus Kinase Signal Transducer and Activator of Transcription proteins (protein, mouse, according to international protein nomenclature guidelines)
kb	Kilobase
Klf4	Krüppel-like factor 4 (Gene symbol)
KLF4	Krüppel-like factor 4 (protein, mouse, according to international protein nomenclature guidelines)
KOH	Potassium hydroxide
LAD	Lamina-associated domains
LCP	No CpG island promoter
LCR	Locus control region
LIF	Leukaemia Inhibitory Factor
lincRNA	Long intergenic non coding RNA
lincROR	Long intergenic non coding RNA, regulator of reprogramming
LINE	Long interspersed elements
lncRNA	Long non coding RNA
mA	Milli ampere
MAD	Mothers against decapentaplegic

MAPK	Mitogen-activated protein kinase (protein, mouse, according to international protein nomenclature guidelines)
MED1	Mediator subunit 1 (protein, mouse, according to international protein nomenclature guidelines)
mESC	Mouse embryonic stem cells
mg	milligram
miRNA	microRNA
mM	mill molar
MMLV	Molone murine leukemia virus
mNET-seq	Mammalian native elongation transcript sequencing
mRNA	Messenger RNA
NaCl	Sodium chloride
ncRNA	Non-coding RNA
NDR	Nucleosomal depleted region
NFR	Nucleosomal free region
nM	Nanomolar
NTU	Non-transcriptional unit
OCT4	Octamer-binding transcription factor 4 (protein, mouse, according to international protein nomenclature guidelines)
OSKM	OCT4, SOX2, KLF4, cMyc
OSN	OCT4, SOX2, NANOG
PAD	Pericentromere-associated domains
PAF1	Polymerae associated factor 1 (protein, mouse, according to international protein nomenclature guidelines)
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCA	Principle component analysis
PCR	Polymerase chain reaction
PI(3)K	Phosphoinositide 3-kinase (protein, mouse, according to international protein nomenclature guidelines)
PIC	Pre-initiation complex
Pou5f1	Pou domain class 5 transcription factor (Gene symbol)
POU$_{hd}$	POU-homeodomain
POU$_s$	POU-specific domain
PP	Polypropylene

PRC2	Polycomb repressive complex 2
PrE	Primitive endoderm
PRO-seq	Precise run-on sequencing
RNA	Ribonucleic acid
RNA-seq	Ribonucleic acid sequencing
RNAP	RNA polymerase
RPK	Reads per kilobase
Rplp0	60s acidic ribosomal protein p0 (Gene symbol)
RPM	Reads per million mapped reads
RT-qPCR	Reverse transcriptase quantitative polymerase chain reaction
SDs	Super-enhancer domains
SDS	Sodium dodecyl sulfate
SE	Super-enhancer
SELEX-seq	Systematic evolution of ligans by exponential enrichment sequencing
SMA	Small worm phenotype
Sox2	SRY (sex determining region Y)-box 2 (Gene symbol)
SOX2	SRY (sex determining region Y)-box 2 (protein, mouse, according to international protein nomenclature guidelines)
SRR2	SOX2 regulatory region 2
SV40	Simian vacuolating virus 40
TAD	Transactivation domain
TADs	Topological associated domains
TBP	TATA binding protein (protein, mouse, according to international protein nomenclature guidelines)
TE	Throphoectoderm
TEs	Typical enhancers
TET	Ten-eleven translocation protein
TF	Transcription factor
TFAP2γ	Transcription factor AP2 gamma (protein, mouse, according to international protein nomenclature guidelines)
TFBS	Transcription factor binding site
TIP60	Histone acetyl transferase 5, also known as KAT5 (protein, mouse, according to international protein nomenclature guidelines)

TRIM28	Tripartite motif-containing 28 (protein, mouse, according to international protein nomenclature guidelines)
TSS	Transcriptional start site
TT-seq	Transient transcriptome sequencing
TUs	Transcriptional units
uaRNA	Up-stream antisense RNA
Utf1	Undifferentiated embryonic cell transcription factor 1 (Gene symbol)
UTF1	Undifferentiated embryonic cell transcription factor 1 (protein mouse, according to international protein nomenclature guidelines)
V	Volt
YY-1	Yin yang 1 (protein, mouse, according to international protein nomenclature guidelines)
µg	Microgram
µl	Microliter
µM	Micro molar

1. Introduction

1.1 Mouse embryonic development

The early development of the mouse embryo is comprised of well-controlled events starting at fertilization of the oocyte and ending in gastrulation of the embryo. The fertilized oocyte gradually loses totipotency, its developmental potential to contribute to the entire organism and extraembryonic tissues (Nicholas & Hall 1942; Hillman et al. 1972). At the late 8-cell stage, embryo compaction leads to cellular polarization (Hillman et al. 1972; Stephenson et al. 2010). At the morula stage, the apical-basal polarity triggers the first cell lineage segregation via two rounds of asymmetrical cell division (Tarkowski & Wróblewska 1967). The polar outer cell layers of the morula will differentiate into trophoectoderm (TE) via suppression of Hippo signaling. The apolar inner cells of the morula will become the inner cell mass (ICM) in the blastocyst. Suppression of Hippo signaling by the outside cells triggers activation of genes involved in the TE lineage differentiation (Nishioka et al. 2009). The segregation of the inner and outer cell layers becomes distinguishable when a cavity forms on the inside of the compacted morula, instigating the formation of the early blastocyst. The TE on the outside of the blastocyst is essential for the formation of the placenta, whereas the ICM contributes to the three germ layers to form the fetus and some extraembryonic tissues: ectoderm, mesoderm, and endoderm. The difference between TE and ICM is demarcated by the expression of different transcription factors (TF). TE cells express high levels of caudal type homeobox 2 (CDX2) and transcription factor AP2 gamma (TFAP2γ) or AP2γ, whereas ICM cells express Octamer transcription factor 4 (OCT4), Sex determining region Y-box 2 (SOX2), and NANOG (Chambers et al. 2003; Niwa et al. 2005; Dietrich & Hiiragi 2007; Ralston & Rossant 2008; Guo et al. 2010).

The developmental potential to contribute to the three germ layers is called pluripotency. The pluripotent ICM cells can be cultured *in vitro* to generate embryonic stem cells (ESC) (Amabile & Meissner 2009). *In vitro*, ESC can divide indefinitely while maintaining their developmental potential (Evans & Kaufman 1981).

Cells within the ICM of the early blastocyst have a similar transcriptional landscape as ESC (Ohnishi et al. 2014). As the blastocyst grows, the Primitive Endoderm (PrE) is being formed. The late blastocyst is made up of the PrE, TE, and epiblast (EPI, also called primitive ectoderm).

After implantation the blastocyst undergoes gastrulation, allowing the PrE to give rise to the extraembryonic parietal and visceral endoderm whereas the EPI contributes towards the three germ layers and germ cells (Gardner & Rossant 1979; Gardner 1985; Lawson et al. 1991; Tam & Zhou 1996). Despite the differences between the ICM and EPI in terms of transcriptional profile and their dependency on different growth signals is the EPI also pluripotent because the EPI can also contribution to the tissues of all three germ layers (for review see: Silva & Smith 2008). The concept of pluripotency is very intriguing because of the two main features, firstly, the ability of becoming all cells of the adult organism while, secondly, maintaining an undifferentiated state *in vitro* indefinitely. This makes pluripotent cells a powerful tool to study the processes underlying cell fate decisions and the transcription regulation guiding these decisions.

1.1.1 *In vitro* pluripotency

Pluripotent stem cells hold a tremendous potential for research and therefore many people have tried to capture these cells *in vitro*. In the 1950s, the first experiments were performed with cells derived from tetracarcinomas. These tumors are comprised of undifferentiated embryonal carcinoma cells (ECC) and differentiated cells ranging across the three germ layers (Stevens & Little 1954). In the 1970s, the first mouse and human ECC lines were established *in vitro* to study cellular differentiation (Kahan & Ephrussi 1970; Hogan et al. 1977). However, these cells exhibited extensive chromosomal abnormalities rendering them unsuitable for research aimed at more clinical applications (Atkin et al. 1974; Brinster 1974; Stewart & Mintz 1982; Pleasure & Lee 1993). Further advances in pluripotency research came in the early 1980s when Evans and colleagues isolated ICM cells from mouse and managed to culture these mouse embryonic stem cells (mESC) *in vitro*. It was possible to maintain ICM cells on a layer of mouse embryonic fibroblasts (MEF), termed "feeder cells", in cell culture medium that contained fetal bovine serum (FBS) (Evans & Kaufman 1981; Martin 1981). Improvements of the cell culture media led to the omission of the "feeder cells", via the addition of the cytokine Leukemia Inhibitory Factor (LIF) (Smith et al. 1988; R L Williams et al. 1988). At the molecular level, LIF acts via the activation of the Janus tyrosine kinase and signal transducer and activator of the transcription proteins (JAK-STAT) pathway (Niwa et al. 1998; Matsuda et al. 1999; Niwa et al. 2009). LIF receptor activation prompts the activation

of JAK and subsequent phosphorylation of TF STAT3. STAT3 is involved in the activation of Kruppel-Like Factor 4 (KLF4), which in turn positively regulates the TFs OCT4, SOX2, and NANOG. LIF also activates the mitogen-activated protein kinase (MAPK) pathway initiating extracellular signaling regulated kinase (ERK) phosphorylation which is a stimulus for differentiation of mESC (Burdon et al. 1999). A third pathway that is activated by LIF is the phosphatidylinositol-4,5-bisphosphate 3-kinase (PI(3)K) pathway counteracting the activation of ERK by LIF (Paling et al. 2004). The addition of serum to the culture medium is important because FBS contains various growth factors. One of those factors belongs to the family of bone morphogenic proteins (BMP). Specifically, BMP4 is important for the activation of TF "small" worm phenotype (SMA) and "Mothers Against Decapentaplegic" (MAD) together called SMAD. More specifically, BMP4 activates SMAD1 which in turn triggers the expression of inhibitor of differentiation (ID) genes which specifically suppress the expression of neuroectoderm genes (Ying et al. 2003). Moreover, serum stimulates the Wingless-type (WNT) pathway via the binding to the WNT receptor Frizzled. Binding of this receptor leads to the suppression of the glycogen synthase kinase 3 (GSK3). Suppression of GSK3 rescues the level of beta-catenin that is important for the activation of pluripotency-related TFs such as OCT4 and NANOG (Sato et al. 2004).

Our understanding of the mechanisms underlying pluripotency has enabled the development of a defined medium to culture pluripotent stem cells. This was necessary because serum-containing medium led to spontaneous differentiation. The defined medium, termed 2i/LIF, contains specific supplements: LIF and two small molecule inhibitors (Ying et al. 2008). The first inhibitor, PD0325901, blocks the kinase activity of ERK1 and ERK2, whereas the second inhibitor, CHIR99021, blocks the GSK3β kinase, rendering the use of serum obsolete. It is worth noting that prolonged exposure of mESC to this 2i/LIF medium alters the cells' ability of differentiation due to chromosomal abnormalities that occur over time (Chen et al. 2015).

1.1.2 Obtaining pluripotency in fully differentiated cells

Somatic cell plasticity was demonstrated by the fact that various terminally differentiated cell populations could be reprogrammed into other cell types. First, in somatic cell nuclear transfer (SNCT) nuclei of differentiated cells were transferred into enucleated oocytes, where they reset to a pluripotent state (Briggs & King 1952; King & Briggs 1955; Gurdon 1962). Later it was shown that fibroblasts can be transdifferentiated to muscle cells via overexpression of MyoD (Davis et al. 1987), B-cells into macrophages via C/EBPα overexpression (Xie et al. 2004) and fibroblasts to neuronal cells via forced expression of ASCL1, BRN2, and MYTL1 (Wapinski et al. 2013). Remarkably, adult mouse and human fibroblasts can be reprogrammed to an induced pluripotent state via the overexpression of the four transcription factors OCT4, SOX2, KLF4, and c-MYC (OSKM) (Takahashi & Yamanaka 2006; Takahashi et al. 2007).

The establishment of pluripotency in somatic cells is a slow and imprecise process that goes through a stochastic initial phase followed by a deterministic latter phase (Buganim et al. 2013; Smith et al. 2016). During the initial stages of reprogramming the fibroblast undergoes mesenchymal-to-epithelial transition which is marked by silencing of fibroblast-specific genes and upregulation of epithelial markers as well as genes involved in proliferation (Li et al. 2010; Samavarchi-Tehrani et al. 2010; Polo et al. 2012). Next, the pluripotency factors are stably upregulated (Polo et al. 2012). Approximately two weeks after induction of OSKM the transcriptional profile of the induced pluripotent cells was similar to ESC. However, it takes additional time to also establish the epigenetic landscape that resembles ESC (Polo et al. 2012; Koche et al. 2011; Lee et al. 2014). To establish the pluripotent transcriptional landscape in somatic cells OSKM need to access parts of the genome that are silenced and seemingly inaccessible (For review see: Zaret & Carroll 2011). It was shown that OCT4, SOX2 and KLF4, but not c-MYC, can bind to regions in the genome that seemingly lack histone modifications and DNA methylation and are resistant to DNaseI, as early as 48 hours post induction (Soufi et al. 2012; Soufi et al. 2014). These observations lead to the conclusion that OCT4, SOX2 and KLF4 act as pioneer factors that can open up the chromatin to allow for activation of regulatory elements essential for the inception of pluripotency. More recently this view has been questioned and it remains to be determined whether or not OCT4, SOX2 and KLF4 have pioneering activity (K. Chen et al. 2016; Chronis et al. 2017).

1.2 OCT4 and the *Pou5f1* locus

OCT4 belongs to a bigger family of TFs called the PIT-OCT-UNC (POU) TFs (Sturm & Herr 1988). All TFs in the POU family show high conservation within their DNA-binding domains and they are classified into six sub-families. OCT4 belongs to sub-family V. TFs in the other sub-families are PIT1(I), OCT1 (II), OCT6 (III), BRN3 (IV), BRN5 (VI) among others.

The TF OCT4 is encoded by the *Pou5f1* gene on chromosome 17 of the mouse genome. *Pou5f1* comprises five exons and 4 introns making a 38,216 Dalton (Da) protein (Schöler et al. 1990; Krishnan et al. 1995). *Pou5f1* expression is tightly controlled by the regulatory elements upstream of the locus. *Pou5f1* control region 1 contains the proximal promoter with the transcriptional start site (TSS) (Pikarsky et al. 1994; Schoorlemmer et al. 1994; Sylvester & Schöler 1994; Ben-Shushan et al. 1995; Minucci et al. 1996) In order to control the precise levels of expression the *Pou5f1* locus contains two enhancer elements, the proximal enhancer (CR2/3) and the distal enhancer (CR4). Both the distal and proximal enhancer allow for a hierarchal spatial-temporal regulation of expression during development. The proximal enhancer is essential for the regulation of OCT4 levels in the epiblast, whereas the distal enhancer is important for the expression levels in mESC and in germ cells (Yeom et al. 1996; Tesar et al. 2007).

The OCT4 protein consists of a centrally located DNA-binding domain called the homeodomain. The Oct4 homeodomain consists of three units, the POU specific domain (POU_{sp}), the POU homeo domain (POU_{hd}) and a linker region bridging the POU_{sp} and POU_{hd} (van Leeuwen et al. 1997). The DNA-binding module is flanked by transactivation domains on the amino-terminal and transactivation domains on the carboxy-terminal (Brehm et al. 1997; Ambrosetti et al. 2000; Niwa et al. 2002). The DNA-binding domain of OCT4, like other POU TF, recognizes a consensus sequence ATTTGCAT or the palindromic ATGCAAAT (Aurora & Herr 1992; Verrijzer et al. 1992; Klemm & Pabo 1996). OCT4 binds its target sites as a homodimer or as a heterodimer. Oct4 shows a strong preference for heterodimerization (Jacobson et al. 1997; Tomilin et al. 2000; Ambrosetti et al. 1997; Nishimoto et al. 1999). In mESC OCT4 binding is most commonly accompanied by binding of the high mobility group (HMG)-box TF SOX2. OCT4 and SOX2 dimerize at the SOX-OCT motif or composite motif (CTTTCTGATTTGCAT) (Xi Chen et al. 2008; Lam et al. 2012a).

1.2.1 The developmental role of OCT4

Throughout the early developing embryo maternally provided OCT4 protein and mRNA are present, with a strong upregulation of *Pou5f1* transcriptional activity upon zygotic genome activation (Rosner et al. 1990; Schöler, Ruppert, et al. 1990). At day 4.5 post coitum (dpc), OCT4 expression is limited to the EPI and the PrE, whereas no OCT4 expression can be found in the TE (Rosner et al. 1990). At the onset of gastrulation, OCT4 expression is limited to the germ cell lineage (Rosner et al. 1990; Schöler et al. 1990).

OCT4 deletion impairs development at the transition of implantation demonstrating its important role in development. Totipotency was not impaired when maternal OCT4 was removed. However, knockout of zygotic *Pou5f1* was detrimental for pluripotency of the ICM (Wu et al. 2013). OCT4 also plays an important role in the formation of the PrE, where OCT4 initiates the expression of FGF4 helping to induce cell autonomy of the PrE (Frum et al. 2013).

A well-controlled balance between OCT4 and SOX2 is important for correct formation of the germ cell layers. OCT4 contributes to the formation of mesendodermal lineages while inhibiting gene expression important for neural ectodermal differentiation. SOX2 on the other hand has a positive effect on the differentiation towards the neural ectoderm while suppressing mesendoderm formation (Thomson et al. 2011). BMP4 plays an important role in OCT4-mediated differentiation, as BMP4 is regulated by OCT4 to influence the differentiation outcome (Wang et al. 2012). Unlike SOX2, which plays a role in the transcriptional regulation of neural stem cells, OCT4 is restricted to the germ cell lineage and is not involved in the transcription regulation of any adult stem cell population (Lengner et al. 2007).

1.2.2 The role of OCT4 in mouse embryonic stem cells

OCT4 plays a key role in mESC, where it is at the core of the transcriptional regulatory network. This was demonstrated by the fact that small changes (roughly 2-fold) in the expression levels of OCT4 led to differentiation of mESC *in vitro*. Increased expression by as little as 1.5-fold led to differentiation of mESC to PrE and mesoderm lineages, whereas the downregulation of OCT4 below 60% led to differentiation towards the TE (Niwa et al. 2000). Although OCT4 is at the core of the regulatory network, the two TFs SOX2 and NANOG play essential roles as well.

OCT4, SOX2, and NANOG (OSN) are supported by a larger number of TFs that are non-essential individually, but together support the pluripotency network. Among those TFs are KLF4, ESRRB, cMYC, and DPPA3.

OCT4 and SOX2 work in close collaboration due to the presence of the SOX-OCT binding motif. Genome-wide binding analysis showed that OCT4 and SOX2 regulate their own gene expression by binding to their respective enhancers. Binding of OCT4 and SOX2 is also found at the NANOG enhancer as well as at the enhancers of many of the supportive TFs (Nishimoto et al. 1999; Ambrosetti et al. 2000; Chew et al. 2005; Catena et al. 2004; Kuroda et al. 2005; Okumura-Nakanishi et al. 2005; Rodda et al. 2005). These observations indicate that ESC have a complex regulatory network in which the core factors and the supportive factors are linked to each other via feedback loops. Interestingly, genes that are regulated via the LIF and BMP pathways show an enrichment of SOX-OCT and NANOG motifs, indicating that external regulation is integrated into the regulatory system (Xi Chen et al. 2008). In addition, roughly 200 proteins have been described to interact with OCT4 in mESC (Pardo et al. 2010; van den Berg et al. 2010; Ding et al. 2012; Esch et al. 2013). Among the interactors of OCT4 are TFs, various co-factors (e.g. histone modifying enzymes), chromatin remodelers (e.g. CHD1, BRG1 and INO80), basal transcription machinery (e.g. Mediator) and parts of the Polycomb Repressive Complex 1 (e.g. RNF2 and Ring 1).

1.3 Process of gene transcription

All uni- and multi-cellular organisms carry their essential information in the form of a genome. In order to extract the information necessary for the cells survival a process called transcription has evolved. Gene transcription is a highly regulated process that involves a broad range of proteins which will be exemplified in the sections below.

The process of gene transcription is complex and does not only dependent on the proteins described below. Regulatory DNA elements also play an important role in transcription, these regulatory elements are addressed in section 1.5.

1.3.1 Gene transcription in mammalian cells

Mammalian cells carry their vital information in the form of deoxyribonucleic acid (DNA) molecules. Information stored in the DNA is first transferred to ribonucleic acid (RNA) molecules via a series of highly regulated steps. In a further step, RNA molecules can be translated into proteins (Allison et al. 1985; Huet et al. 1983). In general, the regulatory steps of DNA-transcription to RNA are highly conserved throughout evolution. Transcription is carried out by a group of enzymes called RNA polymerases (RNAP) which are found in prokaryotes, archaea, and eukaryotes. In mammalian cells, three RNA polymerases (RNAP) can be found: RNAPI, RNAPII and RNAPIII. RNAPI is involved in the production of the ribosomal RNAs such as 18S and 28S rRNA, RNAPII transcribes messenger RNA (mRNA) and non-coding RNAs (ncRNA) such as long intergenic non-coding RNA and enhancer RNAs (eRNA) and RNAPIII generates short RNAs such as 5S rDNA and tRNAs (Cramer et al. 2008).

1.3.2 RNAPII-mediated transcription

RNAPII-mediated transcription of protein-coding genes is regulated via multiple steps starting at the formation of a pre-initiation complex (PIC), which ultimately guides RNAPII to the promoter region of the target gene (He et al. 2013). Activation of a gene commonly starts with the binding of specific gene activators to regulatory elements called enhancers, which can be located up- or downstream of the gene promotor (He et al. 2013). TFs recruit the Mediator complex for the enhancer. The enhancer is positioned in proximity of the promoter via conformational changes to the chromatin landscape. This allows the Mediator complex to recruit general transcription factors (GTFs) (such as TFIIA, TFIID, and others) for the promotor (Thakur et al. 2008; Vojnic et al. 2011). Recruitment of the GTFs and formation of the PIC happens in a stepwise manner. Firstly, TATA-binding protein (TBP), as part of the GTF TFIID, binds close to the transcription start site (TSS) (Baek et al. 2002). Next, TFIIA, TFIIB, and RNAPII bind together with TFIIF. Lastly, TFIIE and TFIIH complete the formation of the PIC (Esnault et al. 2008; Sakurai et al. 1996). After completion of the PIC, the CTD of RNAPII is phosphorylated at specific serine residues initiating transcription. RNAPII pauses at some, but not all genes, where RNAPII stalls approximately 50bp downstream of the TSS. This is a result of the

presence of NELF and DSIF complex (Zhou et al. 2012; Larochelle et al. 2012). Release of paused RNAPII is regulated by phosphorylation steps of the RNAPII CTD allowing for RNAPII elongation into the gene body. Transcription is terminated upon reaching the transcription terminator at the 3' end of the gene, releasing RNAPII from the gene and the mRNA from the polymerase (Larochelle et al. 2012).

1.4 The epigenome

DNA is wrapped around protein octamers called histones to enable the storage of vast amounts of DNA in the small space of the nucleus (Kornberg et al. 1974; Olins et al. 1974). The DNA-Histone complex is referred to as chromatin; DNA wrapped around one histone is one nucleosome. Chromatin allows for the storage of the DNA in higher order structures. The individual nucleosome can influence the accessibility of the chromatin e.g. by post-translational modification. There are various types of post-translational histone modifications both at the unstructured amino-terminal domains and at the highly structured globular domains. So far the following post-translational modifications have been described: methylation, acetylation, ubiquitination, sumoylation, proline-isomerization, ADP-ribosylation, and phosphorylation. Methylation of the histone tails can be found in mono-, di-, or tri-methylation. It is important to note that the histone modifications are highly dynamic and regulated by various enzymes (for review see: Lawrence et al. 2016).

Another form of genomic regulation occurs via the modification of DNA itself in the form of methylation of cytosine residues in a CpG context called 5-methylcytosine (5-mC) (Bird 1986). In contrast to histone modifications, DNA methylation is regulated by only a few enzymes, DNA-methyltransferase (DNMT) 1, 3a, 3b, and the catalytically inactive 3l. DNMT1 is considered to be maintenance DNMT, whereas DNMT3a, b, and l are de novo methyltransferases (Jurkowska et al. 2011). Removal of 5-mC occurs either passively by a gradual loss of 5-mC over the course of cell divisions. Another mechanism is the active removal via the ten-eleven-translocation methylcytosine dioxygenases (TET) 1, 2, and 3 creating intermediate products 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC) which can be removed through the TDG pathway (Bhutani et al. 2011).

Taken together, histone modifications and DNA methylation serve to dynamically regulate gene expression and to mark active and inactive regions in the genome. These modifications can also be used to distinguish between the various regulatory elements.

1.5 Cis-regulatory elements: promoters versus enhancers

Gene transcription requires the formation of the multi-factor PIC around the promoter element (see section 1.3.2).

A promoter generally contains various elements, among them core promoter elements like CpG islands, TATA-box, B-recognition element (BRE), a TSS, and a binding site for RNAPII (Smale & Kadonaga 2003; Gershenzon & Ioshikhes 2005; Lagrange et al. 1998). The promoter can be enriched in tri-methylation of the histone 3 tail at lysine 4 (H3K4me3) (Barski et al. 2007; Guenther et al. 2007; Santos-Rosa et al. 2002). In addition, it often contains a nucleosome-free region (NFR), which is surrounded by well-positioned -1 and +1 nucleosomes and allows for the recruitment of all the factors necessary for gene transcription (Radman-Livaja & Rando 2010). Other histone modifications such as H3K27ac and H3K36me can also be found around promoters. Inactive promoters can be marked with DNA-methylation of their respective CpG islands and/or presence of high levels of H3K27me3 (Schuettengruber et al. 2009; Ru Cao et al. 2002).

Enhancer elements can be identified by similar characteristics: binding of TFs, acetylation of various lysine residues of the histone 3 amino terminal tail (H3K27ac, H3K9ac, H3K18ac and H3K27ac), H3K4 methylation (mono-, di-, and tri), H3K27me3, DNA methylation, RNAPII presence, and/or active transcription of eRNA (Brownell et al. 1996; Ghisletti et al. 2010; Heintzman et al. 2009; Visel et al. 2009; Goldberg et al. 2010; Jin et al. 2009; Spitz & Furlong 2012; Natoli & Andrau 2012). The sensitivity to DNaseI indicates the presence of open chromatin, whereas Mediator components such as Med1 and/or histone acetyl transferases such as CBP/p300 can be indicative of the presence of an enhancer (Gross 1988; Hsieh et al. 2014; Whyte et al. 2013; Hnisz et al. 2013). Generally, enhancers can be divided into three categories: active, poised, and repressed enhancers.

Commonly, active enhancers show histone modifications H3K27ac and H3K4me2 (Zhang et al. 2015). Some enhancers are marked by H3K122ac instead of H3K27ac (Pradeepa et al. 2016). Moreover, binding of TFs, RNAPII, CBP/p300, and Med1 is a

prerequisite for active enhancers. (Whyte et al. 2013; Hnisz et al. 2013)**.** Recently, it was shown that active enhancers produce eRNAs. Transcriptional activity of enhancers is thought to give a direct read-out for the strength of the enhancer (Mikhaylichenko et al. 2018).

Inactive or repressed enhancers are demarcated by an abundance in H3K27me3, binding of repressive transcription factors as well as absence of RNAPII and eRNA (Zentner et al. 2011; Rada-Iglesias et al. 2011). These enhancers also show reduced levels of H3K4me1, while some early developmental enhancers exhibit 5mC to stably repress enhancer activity (Rada-Iglesias et al. 2011). Generally, those enhancers do not show active transcription of eRNAs.

Poised enhancers allow for rapid activation. Those poised enhancers show various histone modifications, H3K4me1, low levels of H3K27ac, and low levels of eRNAs. Upon activation, those characteristics are changed/increased to the levels of active enhancers (Creyghton et al. 2010).

It is important to note that the identification of promoters and enhancers has become increasingly difficult due to the overlapping characteristics of the two cis-regulatory elements (Andersson 2015).

1.5.1 Transcriptional enhancers

The mouse genome harbors roughly 24,000 protein-coding genes, which make up a total of 1-1.5% of the total genomic content. The remainder of the genome consists of various transposable elements, retroviral elements, non-protein-coding DNA elements, and regulatory regions. In order to regulate the differences between the various cell types, evolution has developed a way to tidily control their respective gene expression programs via enhancer elements (Andersson et al. 2014; Long et al. 2016).

Enhancers were initially discovered in the simian virus 40 (SV40), in which a genomic element of 2 x 72bp led to increased expression of its target gene (Banerji et al. 1981). Subsequently, the first mammalian enhancer was described: The gene encoding the immunoglobulin heavy chain contains an enhancer within the first intron of the gene (Gillies et al. 1983; Banerji et al. 1983; Mercola et al. 1983). Depending on the spatial arrangement of the locus the enhancer was active or inactive allowing for temporal control of the locus.

1.5.2 Enhancer emergence and evolution

Enhancers play important roles in the regulation of gene expression, and their evolution has been subject of extensive research. Some theories concerning the emergence of enhancers are listed below.

The duplication of the genome is a very important evolutionary event. In addition to the inception of new genomic loci, genome duplication also allowed for repurposing of certain regions into regulatory elements due to the lack of selective pressure to maintain both gene copies. It has been shown that duplication of non-coding DNA and previously existing enhancers led to the formation of newly acquired regulatory elements (Taylor & Raes 2004; Lan & Pritchard 2016; Allan et al. 1995).

Ancestral regulatory elements that reside in the genome have been shown to develop into enhancers due to the lower evolutionary energy barrier and the presence of transcription factor binding sites (TFBS) (Rebeiz et al. 2011). Although it remains a controversial issue, there have been cases in the mouse genome in which exons develop enhancer activity due to their epigenetic make up. Those enhancer exons (eExons) showed up in DNaseI hypersensitivity data of about 80 different tissues samples, indicating a dual activity for these exons (Vierstra et al. 2014; Frankel et al. 2011; Cretekos et al. 2008).

Drosophila studies suggest that neutral DNA elements as short as 15-mer, can evolve into regulatory regions by random mutations (Smith, Riesenfeld, et al. 2013). Theoretically it would take 0.5 to 10 million years to establish novel enhancers for the anterior-posterior patterning in Drosophila (Duque & Sinha 2015).

Transposable elements are another possible source of enhancer elements because they are prevalent in the genome and vulnerable to mutations (Consortium 2001; Feschotte 2008). For example, the human Alu elements have been identified as a potential source of enhancer evolution. About 1×10^{6} Alu elements were found in the human genome, comprising about 10% of the total human genome. So far, the function of Alu elements is not well understood. It has been shown that Alu elements resemble their enhancer counterparts and can act as transcriptional enhancers *in vitro* (Su et al. 2014).

Additionally, DNA methylation is prone to mutagenesis due to the error-prone repair of the deaminated methylated CpG from C to T (Jones 2012; Duncan & Miller 1980; Sved & Bird 1990). p53 enhancers are scattered throughout the genome and it is

believed that C to T mutations due to spontaneous cytosine deamination have led to the accumulation of these enhancers (Zemojtel et al. 2009).

Rearrangements of the chromosome context has also been shown to drive the formation of enhancers by bringing a different genomic locus into proximity of an enhancer that was previously out of reach (Cande et al. 2009).

1.5.3 Enhancer characteristics

Enhancers are genomic loci of 50 to 1500bp in size and have evolved to control the complex transcriptional profiles of various mammalian cell types (Blackwood & Kadonaga 1998; Pennacchio et al. 2013). Thousands of enhancers can be found in the mouse genome. Transcriptional enhancers are located close to the target gene, sometimes (Blackwood & Kadonaga 1998; Pennacchio et al. 2013) even within an intron, or multiple 100 kbp away from their target gene. One of their key features is the ability to influence gene transcription despite their distant genomic position from the target gene (for review see: Kim & Shiekhattar 2015). The enhancers' ability to regulate gene expression over a distance and the promoters' ability to interact with multiple enhancers allows for a vast complexity of transcriptional regulation, which is at the heart of diversity of biological life (for review see: Long et al. 2016).

Enhancers are highly enriched for TFBS allowing TFs to regulate the spatial-temporal activity. In general, binding of multiple TFs at the same time is important as well as the binding of cell-type specific TFs (For review see: Buecker & Wysocka 2012).

1.5.4 Enhancer spectrum: from the enhanceosome model to the 'flexible billboard model'

In general, there is a broad spectrum of enhancer functions, ranging from the enhanceosome model to the flexible billboard model. On the one hand the enhanceosome model is based on a key characteristic, namely a firm organization of the TFBS. This means that the TFBS are in a specific orientation and that the spacing between the TFBS is precise (for review see: Long et al. 2016). Although descriptions are rather hard to find, computer simulations modeling TF binding indicated that this type of enhancer should be more common in the mammalian genome (Guturu et al. 2013). A prime example of an enhancer that fits the enhanceosome model is the viral-inducible interferon-β (IFNβ) enhancer. This

enhancer requires the combinatorial binding of eight TFs to allow the formation of a complex surface for DNA binding (Thanos & Maniatis 1995).

On the other hand, the flexible billboard model relies on TF cooperativity based on the precise and accurate presence of TFBS. Moreover, the binding of some TFs might be more important than the binding of other TFs (Long et al. 2016). Billboard enhancers rely on the indirect cooperativity between TFs, this means that TF interaction on the DNA is relevant (Arnosti & Kulkarni 2005). The flexibility of the billboard enhancers is also shown by the fact that TFBS are not always the consensus sequence; this helps to increase the specificity of the enhancers (Crocker et al. 2015; Farley et al. 2015). An example of a billboard enhancer is the Otx-a enhancer, which is important for the formation of the neural plate. The Otx-a enhancer contains imperfect TFBS for the GATA and ETS TFs. Optimization of the binding sites leads to an increased enhancer activity, which is likely due to the higher affinity of the TF to their binding site. This indicates that suboptimal TFBS help to fine tune the gene expression by influencing the TF binding specificity (Farley et al. 2015). The conservation of enhancers across different species as well as the synthetic analysis of enhancer activity support the flexible billboard model (Taher et al. 2011; Smith, Taher, et al. 2013).

1.5.5 Enhancer activity: eRNA expression

Active enhancers have been shown to produce ncRNAs named eRNAs. Based on their size (>200bp) a substantial fraction of eRNAs is part of the long non-coding RNA (lncRNA) class (Ørom & Shiekhattar 2013; Kim et al. 2010; de Santa et al. 2010; Pulakanti et al. 2013). Most eRNAs can be distinguished from lncRNAs because eRNAs are predominately unspliced, more precisely 30% of the lncRNAs and only 5% of the eRNAs are spliced (Li et al. 2016). In addition, lncRNAs have substantial levels of H3K4me3 at their respective TSS, whereas eRNA promoters lack this characteristic. Another feature is the directionality. lncRNAs are generally transcribed in an uni-directional (1D) manner, eRNAs on the other hand can be found in uni- and bidirectionality (2D) (Kim et al. 2010; Kim et al. 2015). 1D eRNAs are generally capped and poly-adenylated, whereas bidirectionally transcribed eRNAs lack the poly-adenylation (Djebali et al. 2012; Koch et al. 2011; Lam et al. 2013; Natoli & Andrau 2012). eRNAs are also particularly vulnerable to exosome activity, which is involved in degrading RNA with its 3'-to-5' exoribonuclease activity (Pefanis

et al. 2015). This could explain why their half-life is shorter despite similar transcriptional rates as mRNA and lncRNA (Lam et al. 2013). eRNAs also show a cell-type specific expression pattern which is in line with the observation that their enhancers also tend to be cell-type specific (such as super-enhancers, see section 1.6.12) (Djebali et al. 2012; Whyte et al. 2013; Hnisz et al. 2013).

Besides providing an additional parameter for determining the activity of an enhancer, eRNAs have also been implicated in biological processes. Therefore, they are not merely a side product of an active RNAPII present at the enhancer (Lam et al. 2014; Kim et al. 2015; Rothschild & Basu 2017).

eRNAs have been implicated in TF trapping, e.g. by recruiting the TF Yin-Yang 1 (YY-1) to promoter-enhancer contact points (Sigova et al. 2015; Weintraub et al. 2017). Knockout of the exosome leads to accumulation of eRNA, however it also resulted in reduced YY-1 recruitment. This was likely due to improper processing of eRNA abolishing their recruitment capacity (Pefanis et al. 2015; Sigova et al. 2015; Weintraub et al. 2017). More recently, it was also shown that YY-1 recruits the BAF complex to promoter-enhancer sites to control transcription (Wang et al. 2018). Moreover, eRNAs have been found to be involved in YY-1- and CTCF-mediated chromatin looping (Beagan et al. 2017), and they are connected to Mediator and cohesin complex reinforcement at promoter-enhancer contact (Li et al. 2013; Lai et al. 2013; Feng et al. 2006; Hsieh et al. 2014). It was also suggested that eRNAs can recruit histone-modifying enzymes such as CBP/p300 to enhancer elements to reinforce their active status by increasing H3K27ac marks (Bose et al. 2017). In addition, eRNAs have been implicated in sequestering transcriptional co-factor NELF from the paused RNAPII, allowing for productive elongation (Schaukowitch et al. 2014). In mouse myoblast cells, two eRNAs were shown to be transcribed from the *MyoD* locus, the core enhancer eRNA (^{ce}eRNA) and the DDR enhancer eRNA (^{DDR}eRNA). The ^{ce}eRNA contributes to the transcriptional regulation of the *MyoD* gene in cis. The ^{DDR}eRNA is expressed in myoblasts upon differentiation into myotubes and functions in trans. The ^{DDR}eRNA is thought to recruit the cohesin complex to *Myogenin* to form the correct spatial chromatin organization necessary for muscle differentiation (Tsai et al. 2018).

1.6 Cellular identity and the pluripotent transcription regulatory networks

1.6.1 Cellular identity

Each cell of a metazoan carries identical genetic material. Therefore the diversity must arise from differences in the RNA levels and the quantity of proteins (Barrero et al. 2010). Differences between cell types are largely adapted via the integration of developmental and environmental stimuli. Developmental plasticity diminishes throughout the development of the organism and is finally restricted to small populations of multipotent progenitor cells, such as hematopoietic stem cells (Fisher 2002). The hematopoietic stem cells maintain the capability of giving rise to every cell type of the hematopoietic system while retaining the capacity of self-renewal (Nakauchi et al. 2001). In order to maintain the cells' phenotype there generally are large regulatory networks in place allowing to sustain a transcriptional network (for review see: Holmberg & Perlmann 2012).

1.6.2 Transcription factors and transcriptional regulation

One of the first steps in transcriptional regulation is mediated by so-called TFs, which can initiate the activation or repression of their target genes by directly binding to the DNA (for review see: Lee and Young 2013). Hence, TFs are the 'DNA-interpreters' and play essential roles in gene regulation and maintenance of cellular identity. Approximately 10% of all protein-coding genes encode TFs, making them the largest portion of protein-coding genes in the genome (Levine & Tjian 2003; Vaquerizas et al. 2009). In the mouse genome, there are roughly 3,200 putative TFs 1,200 of which have been described in scientific publications (Fulton et al. 2009). These ~3,200 TFs can be divided into various subclasses based on their DNA-binding domains: Basic helix-loop-helix domains (Littlewood & Evan 1995), Zinc-coordinated DNA-binding domains (Laity et al. 2001), Helix-turn-Helix (Wintjens & Rooman 1996), beta-Scaffold Factors with Minor Groove Contacts and others (Johnson & McKnight 1989; Lambert et al. 2018). TFs often dimerize into homodimers or heterodimers to exert their function (Botquin et al. 1998; Jacobson et al. 1997; Tomilin et al. 2000; Reményi et al. 2001). This dimerization is dictated via the DNA sequence, depending on the motif it allows for homodimerization or heterodimerization (Tomilin et al. 2000; Lam et al. 2012b).

24

Most TFs recognize DNA sequences that range from 6 to 12 base pairs (bp) (for review see: Spitz & Furlong 2012). The emergence of next-generation sequencing techniques, such as ChIP-seq, SELEX-seq, and DamID-seq, has allowed for the identification of TF consensus sequences and discovered the ability of TFs to bind variations of those consensus sequences (Lambert et al. 2018)**.**

TFs exert their activation and repressive capacity by binding to the aforementioned enhancer elements or promotor regions of their target genes. Most TFs do not influence gene activation or repression on their own, rather they recruit co-activators or co-repressors. TFs can activate gene expression either via recruitment of the transcription machinery and/or release of paused RNAPII (Fuda et al. 2009) or recruitment of epigenetic machinery, such as histone modifiers and chromatin remodelers (Li et al. 2007). However, due to their cell type-specific expression TFs are at the core of the transcriptional regulatory network (Mullen et al. 2011; Gertz et al. 2013; Choudhury & Ramsey 2016).

The presence of TF binding sites throughout the genome, sometimes in motifs that are imperfect but close to TFBS for a different TF, allows for increased complexity of TF regulation (Spivakov et al. 2012; Lambert et al. 2018). This feature enables the cell to modulate gene expression in a hierarchical-temporal manner. Transcriptional regulatory networks are essential during development to initiate cellular differentiation as well as maintenance of cellular identity. In each given cell type, there are a few TFs that function as core factors in its specific regulatory network.

Pluripotent stem cells harbour a great developmental potential as their progeny give rise to the entire organism (for review see: Amabile & Meissner 2009). To regulate the spatial-temporal gene expression of mESC, a complicated network of many players governs gene expression to ensure the maintenance of the pluripotent cellular identity (Barrero et al. 2010). The TFs OCT4, SOX2, and NANOG are at the heart of this pluripotent regulatory network (Chambers & Smith 2004; Niwa 2007; Silva & Smith 2008). The distinctive expression pattern of OCT4 and NANOG together with genetic experiments pinpointed their role as pivotal regulators (Niwa et al. 2000; Chambers & Smith 2004; Chambers et al. 2003; Nichols et al. 1998; Mitsui et al. 2003; Tapia et al. 2015). SOX2, the third TF in the core of the pluripotency network, cooperates with OCT4 as a heterodimer (Ambrosetti et al. 2000; Avilion et al. 2003; Masui et al. 2007).

1.6.3 TF cooperativity

OCT4 is known to be a cooperative transcription factor, which means that it prefers to bind genomic loci together with other transcription factors in a context-dependent manner. In mESC, OCT4 prefers to bind to the DNA together with SOX2 at the OCTSOX motif, also referred to as the composite motif (Reményi et al. 2003; Niwa 2007). Moreover, NANOG, the other key pluripotency factor, often binds in close proximity to the OCT4-SOX2 heterodimer (Boyer et al. 2005; Loh et al. 2006; X Chen et al. 2008; Kim et al. 2008; Whyte et al. 2013). Among the POU TFs OCT4 is unique in its hetero dimerization as most other POU TF prefer to bind to their target sites as homodimers (Rhee et al. 1998). Another SOX factor, SOX17, shows little cooperativity with OCT4 on the OCTSOX composite motif and is incapable of inducing pluripotency when it replaces SOX2 in the reprogramming cocktail (OCT4, SOX2, KLF4, cMYC). On an alternative motif, the compressed motif, OCT4 and SOX17 actually cooperate to initiate the formation of the primitive endoderm (Jauch et al. 2011; Aksoy et al. 2013). Changing a single amino acid in SOX17 E57K, which is part of the OCT4 interface of SOX17, allows SOX17 to form iPSC (Jauch et al. 2011). These observations underline that cooperativity among the pluripotency TF strongly impacts their binding capacity and therefore their role regarding lineage specification.

Coming back to OCT4-SOX2 cooperativity, the strong dimerization at the OCTSOX motif was underlined by a study that performed electronic mobility shift assay (EMSA) on a variety of composite motif configurations. The conclusion of this study is that the OCTSOX motif without spacing in between the OCT4 and SOX2 TFBS showed the strongest dimerization (Jauch et al. 2011). In addition, including 1 or 2 nucleotides in between the OCT and SOX motif almost completely abolished the OCT4-SOX2 dimerization, whereas 3-10 nucleotides of spacing only reduced the cooperativity. Another study showed that on the FGF4 enhancers, which have naturally 3 nucleotides in between OCT4 and SOX2 binding sites, increase in spacing between OCT4 and SOX2 motif leads to decrease in enhancer activity. The OCT4-SOX2 complex was not formed on spacing mutant despite SOX2 still binding (Ambrosetti et al. 1997). In the context of reprogramming of mouse embryonic fibroblasts, it is was shown that distinct configurations of the OCTSOX motif are more important for the induction and maintenance of pluripotency than other OCTSOX motifs (Tapia et al. 2015).

1.6.4 The regulatory network in mESC

There are two important principles for maintenance of the pluripotency networks. First, the core factors regulate each other's expression via positive feedback loops. Second, the core factors positively influence the expression of a larger number of support TFs necessary to maintain the pluripotent identity while suppressing the expression of differentiation genes (Xi Chen et al. 2008; Chew et al. 2005; Matoba et al. 2006; Bilodeau et al. 2009; L. a Boyer et al. 2005; Boyer et al. 2006; Lee et al. 2006; Loh et al. 2006; Marson, Levine, et al. 2008; Pasini et al. 2008; Pasini et al. 2004). In order for these principles to work the core factors drive an active transcriptional landscape while maintaining an open chromatin (Efroni et al. 2008; Meshorer et al. 2006). mESC express high levels of specific chromatin remodeler complex members in order to sustain the open chromatin landscape that is necessary for the hyperactive transcription to take place (Efroni et al. 2008; Ho et al. 2009; Gaspar-Maia et al. 2009).

1.6.5 Genome organization

The genome contains four distinct levels of organization to allow for sufficient accessibility of roughly 2 meters of DNA into each nucleus. Chromosomes are placed at defined spaces in the nucleus, so-called chromosome territories (CT) (Cremer & Cremer 2010). The compartmentalization within each chromosome places specific regions within a chromosome towards the nuclear lamina (B compartments) or rather towards the inside of the nucleus (A compartments) (Boettiger et al. 2016; Lieberman-Aiden et al. 2009; Rao et al. 2014; Wang et al. 2016; Vieux-Rochas et al. 2015; Lin et al. 2012). Smaller sub-structures, so-called topological associated domains (TAD), are formed, which create gene neighborhoods and allow for restricted interactions (Nora et al. 2012; Dixon et al. 2012; Sexton et al. 2012). Long-range chromatin loops create contact points between enhancers and promoters (Ptashne 1986; Schleif 1992; Tolhuis et al. 2002; Palstra et al. 2003).

Genome organization in pluripotent stem cells harbours all of the aforementioned characteristics; nonetheless, some of the structures stand out in comparison to differentiated cells (Meshorer et al. 2006; Jaenisch & Young 2008; Wiblin 2005; Denholtz & Plath 2012). In comparison to differentiated cells, pluripotent cells show a distinctive feature in the higher order chromatin. Inactive portions of the pluripotent

genome show a lack of long-range contact in their B compartments indicative of a relaxed organisation of these regions (de Wit et al. 2013). In somatic cells pericentromere-associated domains (PADs) overlap with lamina-associated domains (LADs), which are part of the B compartments and considered to be inactive. Interestingly, the pluripotent genome PADs showed higher transcriptional activity and significantly less overlap with LADs (Wijchers et al. 2015).

OCT4, NANOG, and KLF4 are instrumental in the correct formation of the 3D genome in pluripotent stem cells. Loss of one of these factors causes reduction in the long-range interaction loops (Apostolou et al. 2013; de Wit et al. 2013; Wei et al. 2013). There is a strong propensity for the promoters of the pluripotency genes to form long-range interactions with each other in both cis and trans (Apostolou et al. 2013; de Wit et al. 2013; Kieffer-Kwon et al. 2013). Genomic loci that harbour multiple TFBS for OSN show a strong tendency for long-range interactions, which disappear upon differentiation (Phillips-Cremins et al. 2013; Apostolou et al. 2013; de Wit et al. 2013; Wei et al. 2013; Denholtz et al. 2013). Interestingly, the binding of OSN coincides at some loci with the presence of Polycomb Repressive Complex 2 (PRC2). Perturbation of the sub-component EED of PRC2 leads to loss of the interaction points. Curiously, PRC2 is often associated with repressive chromatin; however, in this context PRC2 enrichment sites and OSN-bound sites both reside in A compartments (Denholtz et al. 2013). In addition to their role in long-range loop formation, the pluripotency factors also extensively contribute to so-called sub-TAD interactions. Sub-TAD interactions take place within a TAD. These sub-TAD structures are unique to pluripotent cells (Kagey et al. 2010; Phillips-Cremins et al. 2013; Nitzsche et al. 2011). In mESC, regulation of the sub-TAD contacts of the *Nanog* locus seems to be partially controlled by OCT4 (Levasseur et al. 2008). Overall, roughly 76% of enhancers in mESC interact with genes beyond the adjacent active gene, with over 40% of those interactions being interchromosomal (Zhang et al. 2013).

1.6.6 TF binding in the genome

The cooperativity of OCT4 and SOX2 is essential to the maintenance of ESC identity and the induction of pluripotency in somatic cells. Alterations of the protein-protein interactions by mutations or post-translational modifications impact the transcriptional profiles of ESC (for review see: Li & Belmonte 2017). Interestingly, mutations

affecting the OCT4-SOX2 interface severely impact the reprogramming capacity; however, changes to the DNA-binding domains seem less critical (Tapia et al. 2015; Jerabek et al. 2017). OCT4, unlike the other POU factors, prefers to bind as a heterodimer at the OCT-SOX motif (Jerabek et al. 2017; Mistri et al. 2015). Another feature of OCT4 is that it is claimed to be a pioneer TF, which refers to TFs capable of binding closed chromatin with the purpose of opening it up (Soufi et al. 2014; Soufi et al. 2012; For review see: Zaret & Carroll 2011).

The molecular mechanisms by which OCT4 functions as a pioneer factor remain to be fully elucidated. It was suggested that this might work via Brahma-related gene 1 (BRG1) which is a subunit of the Brahma-associated factor (BAF) complex (King & Klose 2017). A decrease in chromatin accessibility around OCT4 TFBS was observed upon loss of OCT4 in mESC due to a decrease of BRG1 recruitment at these loci. This finding indicates that recruitment of BRG1 by OCT4 helps OCT4 to stably bind to its target sites allowing the co-binding of SOX2 and NANOG. In support of these findings it was shown that the BAF complex is a facilitator of reprogramming and that over-expression of BAF components leads to increased reprogramming efficiency (Singhal et al. 2010). The proposition of OCT4 pioneer activity has been questioned by observations made in DNaseI sequencing data, which revealed that OCT4 acts as a "settler" TF rather than a pioneer TF (Sherwood et al. 2014). Moreover, in the context of somatic cell reprogramming to iPSC OCT4 together with SOX2 and KLF4 are primarily involved in silencing somatic enhancers. It was also shown that binding of OCT4 alone did not explain the opening of the chromatin (Chronis et al. 2017). Furthermore, it was observed that the binding of SOX2 might be required for the binding of OCT4 (Chen et al. 2014). In relation to the repression of developmental genes it was suggested that the pluripotency factors pre-mark cell type-specific enhancers already in mESC (Kim et al. 2018). Macrophage enhancers were pre-bound by one or two pluripotency factors keeping an open chromatin profile and maintaining demethylated DNA by recruitment of Tet1 to these sites.

The main binding events of the pluripotency factors are their own enhancers as well as the enhancers of the support factors such as ESRRB (Ivanova et al. 2006; X. Zhang et al. 2008), TCF3 (Cole et al. 2008; Marson et al. 2009), SALL4 (Zhang et al. 2006; Wu et al. 2006), TBX3 (Niwa et al. 2009; Ivanova et al. 2006), KLF4 (Jiang et al. 2008; Niwa et al. 2009), PRDM14 (Chia et al. 2010), and CDH1 (Gaspar-Maia et al. 2009) among many others.

1.6.7 Transcriptional co-factors

TFs bind to the DNA and are used as docking sites for activating/repressing protein complexes that do not exhibit any DNA-binding capacity (co-activators/co-repressors) (Young 2011; Li & Belmonte 2017). Most of the co-factors such as p300, cohesin, and mediators are ubiquitously expressed in nearly all cell types. Unlike these other cell types mESC are particular vulnerable when the levels of these co-factors are reduced (Kagey et al. 2010; Fazzio & Panning 2010). These co-factors can recruit and control the transcriptional apparatus (Conaway et al. 2005; for review see: Malik & Roeder 2005; Roeder 1998; Taatjes 2010). Knockdown experiments indicated that loss of 60kDa Tat-interactive protein (TIP60)-p400 complex, RNA polymerase associated factor 1 (PAF1) or tripartite motif-containing protein 28 (TRIM28) are detrimental to mESC (Fazzio et al. 2008; Ding et al. 2015; Hu et al. 2009). Depletion of co-factor p300 has little effect on mESC maintenance, even though it is found near most active promoters, but its loss interferes with the mESC differentiation potential (Xi Chen et al. 2008; Zhong & Jin 2009).

Mediator is a multi-subunit complex with kinase activity via the cyclin dependent kinase 8 (CDK8) module. CDK8 can phosphorylate SMAD1/5 or SMAD2/3 in their respective linker region which either allows for dimerization and subsequent activation or triggers proteasome-mediated degradation and repression (Alarcón et al. 2009; Fryer et al. 2004; Gao et al. 2009; Taatjes 2010). In the context of pluripotency Mediator and cohesin are important for the chromatin architecture. This is demonstrated by the fact that reduction in their levels results in dissolution of the pluripotent regulatory network (Gorkin et al. 2014). Mediator plays a role in the recruitment of Nipbl, a factor important in the loading of cohesin (Ciosk et al. 2000; Kagey et al. 2010). Loading of cohesin onto Mediator-bound DNA sites allows for formation of loops to structure the chromatin in a cell type-specific manner. Mediator has been found to co-occupy enhancers that are bound by OCT4, SOX2 and NANOG to link these enhancers to active promoters in mESC (Xi Chen et al. 2008; Zhong & Jin 2009). This co-occupancy seems to be especially strong at so-called super enhancers, which control the expression of cell identity genes (more details on super enhancers in section 1.6.12) (Whyte et al. 2013; Hnisz et al. 2013).

1.6.8 Integration of external signalling pathways

The core pluripotency network needs to be supported by external factors *in vitro* to maintain the cellular identity (see section 1.6.1). Especially the suppression of differentiation factors is key to the maintenance of pluripotency *in vitro* (for review see: Silva & Smith 2008; Pera & Tam 2010). Some of the downstream effectors of the various signalling pathways (LIF, BMP, WNT) are TFs that bind to enhancers bound by OSN and others (Xi Chen et al. 2008; Zhang et al. 2006; Wu et al. 2006; Cole et al. 2008; Tam et al. 2008; X Chen et al. 2008). This indicates that the external signalling pathways are hardwired into the core regulatory network of ESC identity. Interestingly, loss of OCT4-binding at its target enhancers also initiates the loss of binding of these signalling TFs. The notion that signalling is well integrated into the core regulatory network is further supported by the fact that WNT-stimulation during somatic cell reprogramming increases the formation of iPSC (Lluis et al. 2008; Marson, Foreman, et al. 2008).

Maintenance of pluripotency *in vitro* is achieved by the use of 2i/LIF medium, which is a well-defined medium containing LIF, GSK3β-inhibitor, and MEK-inhibitor (Section 1.1.1). Cells cultured in 2i/LIF medium are considered to be in a pluripotent ground state, also known as naïve pluripotency. These cells closely resemble their *in vivo* counterparts (unlike mESC cultured in Serum/LIF conditions) (Ying et al. 2008; Marks et al. 2012; Boroviak et al. 2014). Inhibition of ERK via the MEK-inhibitor has multiple effects on the pluripotency network. The FGF4-ERK pathway destabilizes the expression of KLF2, a TF imperative for pluripotency (Yeo et al. 2014). The MEK-inhibitor and the GSK3β-inhibitor (via stimulation of the Wnt-β-catenin pathway) stabilize the KLF2-regulated circuitry and promote ground state pluripotency (Qiu et al. 2015). ERK also phosphorylates NANOG and thereby destabilizes NANOG and reduces the transactivation capacity (Kim et al. 2014).

Remarkably, genetic perturbation of ERK1 and ERK2 has a detrimental effect on pluripotency: telomere length is no longer maintained, the genome becomes unstable and vulnerable to damage, and pluripotency support factors are no longer properly expressed, ultimately leading to programmed cell death (Chen et al. 2015). In the same study, data suggests that MEK inhibition might work via ERK-dependent and ERK-independent signalling, explaining why MEK inhibition helps pluripotency but ERK-knockout is detrimental.

GSK3β-inhibition stabilizes beta-catenin to help self-renewal and resemble the activation of canonical WNT which is thought to preserve the ground pluripotent state (Lyashenko et al. 2011; Wray et al. 2011; Berge et al. 2011; Yi et al. 2011). GSK3β inhibition-mediated stabilization of beta-catenin leads to increased occupancy of TCF3 at beta-catenin binding sites, which coincides with OCT4- and SOX2-binding in ESC. Curiously, TCF3 is a transcriptional repressor and its activity antagonistic for the OCT4-SOX2 function. Beta-Catenin alleviates transcriptional repression by TCF3, therefore securing the ground state pluripotency (Faunes et al. 2013; Wray et al. 2011). GSK3β-inhibition can be circumvented by stimulation of ESRRB expression, whereas deletion of ESRRB leads to loss of naïve pluripotency (Martello et al. 2012).

In summary, 2i/LIF media secures naïve pluripotency by stimulating the expression of the core pluripotency network as well as the support factors (Li & Belmonte 2017; Hackett & Surani 2014).

1.6.9 Epigenomic landscape of mESC

mESC have a comparatively loose and open chromatin landscape (Meshorer et al. 2006). The core transcription factors maintain this open chromatin landscape via the activities of various histone- and DNA-modifying enzymes (Fouse et al. 2008; Hawkins et al. 2010; Meissner et al. 2008; Mikkelsen et al. 2007). The genes regulated by the core pluripotency TFs can be largely divided into two groups, the genes containing CpG islands in their promoters (HCP) and the genes that do not (LCP).

Among the HCP genes there are regulators with a capacity for self-renewal as well as house-keeping genes. In mESC these HCP genes are largely depleted of 5mC and contain high levels of H3K4me3 indicating transcriptional activity and protection against DNA methylation. HCP genes can also display a combination of histone modifications such as H3K4me3 and H3K27me3. The so-called bivalent domains mark genes that can be rapidly upregulated or quickly silenced, e.g. upon differentiation (Mikkelsen et al. 2007). Among these bivalently marked genes there are master regulators of differentiation.

The LCP genes display DNA methylation, lack H3K4 or H3K27 histone methylation marks and show low transcriptional activity (Meissner et al. 2008). These genes are largely tissue-specific genes that become active upon differentiation, when DNA

methylation is lost and H3K4me3 is gained. Some LCP genes are repressed in a more stringent manner during differentiation, shown by a gain of H3K27me, H3K9me3 and DNA methylation (Hawkins et al. 2010).

1.6.10 ncRNA in mESC

The vast majority of the mammalian genome does not encode for protein-coding genes, but rather expresses ncRNA. A growing number of ncRNAs have been implicated in various biological processes such as regulation of gene expression, X chromosome inactivation, genomic imprinting, silencing of repeats, and dosage compensation (Lee 2009; Wilusz et al. 2009; Zaratiegui et al. 2007; Surface et al. 2010). Both classes of ncRNAs, microRNAs (miRNAs) and lncRNAs are also part of the regulatory network in mESC. miRNAs function mostly through the fine tuning of gene expression by manipulating mRNA stability and translation rates, whereas lncRNAs have been shown to recruit chromatin regulatory complexes to the DNA (Bracken & Helin 2009; Guenther & Young 2010; Surface et al. 2010; Zhao et al. 2010).

OSN seem to regulate miRNAs in two distinct manners (Marson et al. 2008). On the one hand, OSN activate miRNA clusters that help to modulate the expression of essential ESC genes. Upon differentiation these miRNA immediately degrade the pluripotency factors. On the other hand, OSN recruit activating and repressive complexes to loci of miRNA to prepare these loci for rapid activation/repression upon differentiation.

More specifically, OSN regulates the expression of the miR290-and miR-302 clusters via direct binding of these loci. miR-290 and miR-302 have been insinuated to promote cell cycle progression and prevent differentiation, thus supporting self-renewal capacity (Gangaraju & Lin 2009; Houbaviy et al. 2003). miR clusters 375, 124, 9, and 296 are actively silenced by OSN, as they are important for differentiation towards pancreas, neural fate, and existing pluripotency (Marson et al. 2008).

Long intergenic non-coding RNA (lincRNA) regulator of reprogramming (linc-ROR) was characterized as a regulator in somatic cell reprogramming filtering out miRNAs that negatively impact the expression of pluripotency factors (Sheik Mohamed et al. 2010; Loewer et al. 2010; Wang et al. 2013). In human ESC, lncRNAs transcribed from endogenous retroviral elements such as *HERV-H* are involved in the

recruitment of OCT4 or co-factors CBP/p300, MED6, and MED12 to enhancer sites (Lu et al. 2014).

Long interspersed nuclear elements (LINEs) are amply represented in the genome. LINE-1 is a retroviral element thought of as detrimental to the cells; it is found to be involved in various diseases, neurological disorders, diverse cancers and hematopoetic diseases (Burns 2017). Interestingly, LINE-1 elements are highly expressed during development (in mESC and germ cells); however, there are seemingly low frequencies of retrotranspositioning (Kano et al. 2009; Richardson et al. 2017; Newkirk et al. 2017). In mESC, LINE-1 has various functions. It can function as a scaffold for the binding of nucleolin to ribosomal DNA (rDNA) genes and it can recruit KAP1/TRIM28, which assist in the transcription of rDNA genes, ultimately contributing to mESC self-renewal (Percharde et al. 2018). On the contrary, LINE-1 transcripts are important in silencing of Dux, the master regulator of the 2-Cell stage. Expression of LINE-1 has been shown to coincide with the exiting of the 2-Cell stage, and knockdown of LINE-1 in mESC leading to reprogramming to the 2-Cell stage (Percharde et al. 2018).

1.6.11 Chromatin remodelers

TFs and histone modifications can exert their effects on the chromatin via the recruitment of protein complexes called ATP-dependent nucleosome remodelers. These ATP-dependent nucleosome remodelers can enzymatically influence the structure of the chromatin (for review see: Clapier & Cairns 2009; Ho & Crabtree 2010). Nucleosome remodelers alter the DNA-accessibility which in turn can affect the transcriptional activity and binding capabilities of TFs. Various members of nucleosome remodelling complexes have been shown to be essential for pluripotency or proper differentiation of ESC (Ho & Crabtree 2010; Gaspar-Maia et al. 2009; Schnetz et al. 2010; Bilodeau et al. 2009; Klochendler-Yeivin et al. 2000). Two members of the chromodomain helicase DNA-binding (CHD) family have been implicated in different roles. CHD1 can be found at active promoters that are directly regulated by the core pluripotency factors OCT4 and SOX2. Depletion of CHD1 leads to improper formation of the PrE (Gaspar-Maia et al. 2009). A second member of this family, CHD7, is found at active enhancers that are also bound by OSN, where it helps to modulate the gene expression of the enhancer's target genes (Schnetz et al. 2010).

1.6.12 Transcriptional control of cell identity genes

Genes essential for cellular identity are being controlled by specific regions in the genome called locus control regions (LCR) (Fraser & Grosveld 1998). Control of the *human β-globin* gene expression relies on a LCR that regardless of the orientation and position is able to maintain physiological levels of β-globin (Grosveld et al. 1987). Other genes were also identified to have LCR: T-cell specific *CD2*, *CD4* and *TCRα/δ* (Diaz et al. 1994; Ortiz et al. 1997; Hong et al. 1997), B-cell specific MHC class II Ea (Carson & Wiles 1993), the neuron-specific gene *S100β* and others (Friend et al. 1992). These regions were sensitive to DNaseI and contained high numbers of TFBS (Fraser & Grosveld 1998). More recently, super-enhancers were described as regions in mESC that are marked by high occupancy of TFs such as OSN, transcription machinery and co-factors regulating gene expression levels vital to the cellular identity (Whyte et al. 2013; Hnisz et al. 2013; Lovén et al. 2013). Super-enhancers are computationally defined by three criteria: The TFs OSN must be bound, enhancers within a 12.5kb range were added together (so-called stitched enhancers) and stitched enhancers needed to show a defined amount of Med1-binding (Whyte et al. 2013). In addition to the above-mentioned computational determination of super-enhancers there are some biological determinants, which are listed below. The size of a super-enhancer in mESC is about ~8700bp (as opposed to a 'regular' mESC enhancer with about ~700bp) (Whyte et al. 2013; Lovén et al. 2013; Hnisz et al. 2013). Other co-factors, such as CBP/p300, cohesin, and CTCF are also bound at super-enhancers (Hnisz et al. 2013). Various histone modifications (such as H3K4me2, H3K4me1, and H3K27ac) are reported to be abundantly present at super-enhancers. Genes encoding core pluripotency factors tend to have nearby super-enhancers, which are essential for the high levels of transcription found at these loci (Whyte et al. 2013). Another key feature of super-enhancers is that they exhibit high eRNA expression signals (Pulakanti et al. 2013; Hnisz et al. 2013). Super-enhancers are not only found in pluripotent stem cells, but also in various other cell types at lineage-specific genes. Interestingly, super-enhancers overlap with the previously described LCR (Hnisz et al. 2013). Besides super-enhancers stretched enhancers can also be found. Stretched enhancers are genomic regions that are defined by enrichment for several histone modifications over a larger region (>3kb). In comparison to super-enhancers stretch enhancers are more prevalent by an order of magnitude (Parker et al. 2013).

Many pluripotency-related super-enhancers and their corresponding protein-coding genes are situated in CTCF- and cohesin-structured chromatin loops termed super-enhancer domains (SDs) (Dowen et al. 2014). Deletion of these CTCF sites led to disruption of chromatin loops and distortion of transcriptional output, highlighting the importance of SDs insulation. .Interestingly, differentiation of pluripotent cells did not seem to alter the structure of the cohesin- and CTCF-mediated loops, which is supported by the notion that cohesion and CTCF are ubiquitous in function (Phillips-Cremins et al. 2013; Dowen et al. 2014). In contrast, it was observed that OCT4, SOX2 and NANOG distal super enhancers create enhancer-promoter contact loops that are highly tissue-specific and disappear upon differentiation of pluripotent cells. This suggests that chromatin structure is cell-type specific and that lineage-specific TF could play an important role in shaping the genomic landscape (Kagey et al. 2010; Phillips-Cremins et al. 2013; Gaspar-Maia et al. 2011; Li et al. 2014; Levasseur et al. 2008). In summary, it remains to be shown what the exact role OCT4 in the context of enhancer regulation and gene expression really is.

1.7 <u>Aim of this thesis</u>

Despite the major advancements in next-generation sequencing and genetic approaches there remains a gap in our understanding of the pluripotent regulatory network. Most of the acquired data stems from utilization of individual genomic methods to address TF binding or transcriptomic analysis.

In this study, we aim to address the gaps that remain in our understanding of the pluripotent regulatory network by combining various genomic approaches.

We set out to dissect the OCT4 governed mouse Embryonic Stem Cell regulatory network. We utilized a loss of function model developed by Niwa et al in 2000 to address the effects of acute loss of OCT4 on the pluripotent regulatory network. Combining this loss of function model with the novel transient transcriptome followed by deep sequencing enables us to observe changes in the transcriptome at the nascent RNA level and to address alterations of transient RNA species such as eRNAs. Integration of chromatin immunoprecipitation followed by high-throughput sequencing will allow us to correlate the changes in the transcriptome to the changes in the binding of transcription factors OCT4 and SOX2 upon loss of OCT4. Lastly, overlaying analysis for transposase accessible chromatin sequencing will be applied to examine the changes of the chromatin accessibility upon Oct4 withdrawal.

In summary, our goal is to define the direct target regulatory elements of OCT4, to improve our understanding of the cooperativity of OCT4 and SOX2, and to enhance our knowledge of the protein-coding genes that are under direct OCT4 control. The data in this thesis are derived from the combined efforts of the laboratories of Prof. Dr. Hans R Schöler and of Prof. Dr. Patrick Cramer.

2. Materials & Methods

2.1 Materials

2.1.1 Chemical substances

1,4-Dithiothreitol (DTT, Sigma Aldrich, DTT-RO-Roche)

4-thiouridine (Carbosynth, 13957-31-8)

6x DNA loading dye (ThermoFisher Scientific, R0611)

Accutase® solution (Sigma Aldrich, 6964-100ml)

Acryl amid solution (Serva Electrophorsis, 10688)

Agarose Universal (Bio Buget Technologies GmbH, 10-35-1020)

Agencourt RNAClean XP beads (NuGene, S0152)

Ammonium persulfate (Serva Electrophorsis, 13376.02

B27 supplement (Life Technology, 12587-010)

Beta mercaptoethanol (Gibco, 31350-010)

Biotin(EZ-link HPDP-) (ThermoFisher Scientific, 21341)

Bovine Serum Albumin (Sigma Aldrich, A3059-100G)

CHIR99021 (Cayman Chemicals, 13122)

Chloroform (Sigma Aldrich, C2432)

Chloroform (Merck, 107024)

cOmplete tablets, mini protease inhibitor (Roche, 4693124001)

DAPI dihydrochloride (PARTEC, #05-7202)

DMEM/F12 (Life Technology, 21331-020)

DNA marker (ThermoFisher Scientific, SM0331)

DNase Turbo (Invitrogen, AM2239)

Doxycycline hyclate (Sigma Aldrich, D9891-1G)

ECL Prime Western blotting solution (GE Healthcare, RPN2232)

Ethanol (Roth, 9065-3)

Ethanol (Merck, 100983)

Ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid (EGTA) (Sigma Aldrich E4378)

Ethylenediaminetetraacetic acid (EDTA) (Gibco, 15575020)

Ethylenediaminetetraacetic acid (EDTA) (AppliChem, A3553,1000)

Fetal Bovine Serum (FBS) (Sigma Aldrich, G1393-100ml)

Formaldehyde (AppliChem, A0877,0250)

Gelatin (Sigma Aldrich, G13393)

GelRed Nucleic Acid Stain (Biotrend 41002-1)

Glycerol (Roth, 3783.1)

Glycin (Sigma Aldrich, G8898-1KG)

H_2O (Gibco, 10977035)

HEPES (Roth, 6763.2)

Hoechst 33258 dye (Sigma Aldrich, H6024)

Hydrochloric acid (Sigma Aldrich, 30721)

Igepal CA-630 (Sigma Aldrich, I8896)

Isopropanol (Roth, 9866-1)

Isopropanol (Merck, 109634)

Knockout Serum Replacement (Gibco, 10828-028)

L-Glutamin (Sigma Aldrich, G7513-100ml)

Leukemia Inhibitory Factor (Home made)

Lithium Chloride (Roth, P007.1)

Magnesium Chloride (AppliChem, A3618,0500)

Methanol (Roth, 4627.6)

N,N-Dimethylformamide (Sigma Aldrich, D4551)

N2 supplement (Gibco, AM9759)

Neuralbasal (Life technology, 21103-049)

NP-40 (Sigma, 18896-100ml)

Paraformaldehyde (Sigma Aldrich, D6148-500G)

PD0325901 (Biomol, 103034-25)

Penicillin/streptomycin (Sigma Aldrich, P4333-100ml)

Phosphate Buffered Saline (Sigma Aldrich, D8537-500ml)

Potassium Hydroxde (AppliChem, A1575,1000)

Precision Plus Protein Kaleidoscope Protein Standard (Bio-Rad, 161-0375)

Protein G dynabeads (Invitrogen, 1004D)

Proteinase K (AppliChem, A4392,0010)

RNaseA PureLink (Invitrogen, 12091-021)

Sodium Chloride (AppliChem, 1.316.591.211)

Sodium Chloride (Gibco, AM9759)

Sodium dodecyl sulfate (Roth, CN30.3)

Sodium Hydorxide (AppliChem, A1551.1000)

Sodimdeoxycholate (Sigma Aldrich, 30970-100G)

SYBRgreen mix (Bio-Rad, 172-5124)

SYBRGreen mix (ThermoFisher Scientific, 4309155)

Tetramethylethylenediamine (TEMED) (Sigma Aldrich, 77-86-1)

Tris (Roth, 5429.3)

Tris (Sigma Aldrich, 77-86-1)

Triton-X100 (Promega, H5141)

TRIzol (Ambion, 15590018)

Tween-20 (AppliChem, A4974,0500)

Tween-20 (Sigma Aldrich, P9416)

2.1.2 Commercially Available kits

Bioanalyser ChIP/Gel (Agilent Technology, 5067-4626)

M-MLV Reverse transcription (Promega, M1701/M531A)

miRNase Micro kit (Qiagen, 217084)

Multiplex Oligos for Illumina (Index Primers Set 1) (New England BioLabs, E7335L)

Multiplex Oligos for Illumina (Index Primers Set 2) (New England BioLabs, E7500L)

NEBNext Ultra DNA Library prepration kit (New England BioLabs, E7370L)

NEBNext High Fidility 2x PCR mix (New England BioLabs, M05413)

Nextera DNA library preparation kit (Illumina 15028212)

Nucleospin Gel and PCR Cleanup kit (Machery & Nagel, 740.609.250)

Ovation Universal RNA-seq system (NuGen, 343)

Real-Time Library amplification kit (KAPA Biosciences, 7959028001)

µMACS Steptavidin kit (Miltenyi, #130-074-101/130-042-701)

2.1.3 Laboratory devices

2100 Bioanalyser (Agilent Technology, 2100)

Biometra Trio (Biometra T professional)

Bioruptor Standard sonication device (Diagnode, B01010002)

Cell culture microscope (Olympus, CKX41)

Centrifuge cold (Tabletop) (Eppendorf, 5415R)

Centrifuge (Tabletop) (Eppendorf, 5702)

Centrifuge cold (Tabletop) (Eppendorf, 5415R)

Covaris sonicator (Coavris, S220)

DynaMag2 (Life Technology, 12321D)

Electrophoresis power supply (Amersham Bioscience, EP530)

End-over-end rotor (CMV)

Eppendorf Research pipets (p10, p100, p1000) (Eppendorf)

Fast Real-Time PCR System with 96-Well block module (Applied Biosystems 7900HT)

Fluorescence microscope camera (Leica, DGC350FX)

Fragment Anaylser (Advance Analytical, FSV2CE2F)

Fully Automated Inverted Research Microscope (Leica DMI6000B)

GeneAmp 9700 PCR machine (Applied Biosystems)

Gilson Pipets (p2, p10, p20, p100, p200, p1000)

Highspeed table top centrifuge (Eppendorf, #2-16KL)

HyperCassette (18x24) (Amersham Bioscience)

Liebherr comfort (Liebherr)

Microwave (Samsung)

Mini-Protean 3 Cell/Mini Trans Blot Module (Bio-Rad, 153BR)

Multifuge (Heraus Multifuge 4KR)

Nalgene Cryo 1 Freezing container (ThermoFisher Scientific, 5100-0001)

Nanodrop 2000 (ThermoFisher Scientific, 2000)

Nanodrop 1000 (ThermoFisher Scientific, 1000)

ORICA microsope camera (Hamamatsu, C10600)

Pipet boy (Drummond)

Plate Fuge (Benchmark Scientific, C2000)

qTOWER3 (Analytik Jena AG, TOWER3)

Quantstudio 5 (Applied Biosystems)

Qubit 2.0 (Life Technology)

Qubit 3.0 (Life Technology)

Rollerband (Stuart, SRT9)

Sanyo Ultra low (Sanyo)

Sigma rotor (Sigma Aldrich, #12181)

Tape Station (Agilent Technology, 2200)

Thermo electron cooporation Froma Direct Heat CO_2 incubator (ThermoFisher Scientific)

Thermomixer Comfort (Eppendorf)

Thermocycler (Biometra Tprofessional)

Thermomixer C (Eppendorf, 5832000015)

Tissue Culture hood (HERA SAFE, K5-15)

Vortx bioanalyser chip (IKA Works)

Vortex Genie 2 (Scientific Industries)

VWR Galaxy minister silverline (VWR)

Waterbath (GFZ)

Water cooler (ThermoFisher Scientific, RTE-7)

2.1.4 Disposables

1.5ml tubes (Gkisker, G052BP)

1.5ml Low Protein binding tubes (Eppendorf, 22431081)

1000µl pipet tips

100mm tissue culture plate (Sarstedt, 83.3902)

10µl pipet tips

150mm tissue culture plates (Sarstedt, 83.3903)

15ml falcon tubes (Sarstedt, 62.544.502)

200µl pipet tips

2.0ml Low Protein binding tubes (Eppendorf, 22431002)

384-well qPCR plates (Axygen, 321-22-051)

50ml conical tubes (Sarstedt, 62.54.254)

6-well tissue culture plate (Sarstedt, 83.3920)

60mm tissue culuture plates (Sarstedt, 83.3901)

96-well PCR plate (VWR, 732-2387)

Biosphere Filter tips 1250µl PP plastic (Sarstedt, 70.1186.210)

CryoTube™ (ThermoFisher Scientific, 3777224)

DNA LoBind Tube 2.0ml (Eppendorf, 0030 108.078)

Filter tips 1000µl (TipONE, 1122-1830)

Filter tips 100µl (TipONE, S1120-1840)

Filter tips 10µl (TipONE, S1120-3810)

Filter tips 200µl (TipONE, 112-8810)

Filter tips 20µl (TipONE, S1120-1810)

Gloves (MicroFlex, YN705715749)

MicroAmp Fast Optical 96-well qPCR reaction plate (Aplied Biosystems, N8010560)

microTUBEs with AFA fiber (Covaris, #520045)

Optical Clear Adhesive Seal Sheets (ThermoFisher Scientific, AB-1170)

Phase Lock Gel Heavy (2ml) (5 Prime, 2302830)

Transfer Membranes Immobilon-P PVDF (Merck Millipore, IPVH00010)

Whatman™ Filterpaper 3MM CHR (GE Healthcare Life Sciences, 3030-917)

2.2 Methods

2.2.1 Cell culture

Mouse embryonic stem cells carrying a doxycycline repressible *Pou5f1* transgene (ZHBTc4 cells (Niwa et al. 2000)) were cultured at 37°C on gelatin-covered plates in 50% Dulbecco's Modified Medium/F12 and 50% Neural Basal Medium containing 2% Fetal Bovine Serum, 2% Knockout Serum Replacement Medium, 0.5x N2 supplement, 0.5x B27 supplement, 1x penicillin/streptomycin, 2mM L-Glutamin, 0.1mM β-Mercaptoethanol, 1μM PD0325901, 3μM CHIR99021, and 0.04μg/ml Leukemia Inhibitory Factor. 1μg/ml doxycycline was added to the culture medium at 0 hours, 3 hours, 6 hours, 9 hours, 12 hours, 15 hours, 18 hours, and 24 hours to obliterate *Pou5f1* expression.

2.2.2 Freezing and thawing cells

Cells at ~70% density in a 10cm cell culture dish were washed with PBS and dissociated using Accutase® solution. After a 2-minute incubation at 37°C the cells were harvested with their appropriate cell culture medium and collected into 15ml falcon tubes. Cell were pelleted at 1400 rpm for 5 minutes. Supernatant was aspirated and cells were re-suspended in 2.5ml of 'freezing medium' containing 50% cell culture medium and 50% FBS/DMSO in an 80% to 20% ratio. 500μl (1/5 of a 10cm plate) was transferred to labeled CyroTube™ vials and frozen at -80°C in a Nalgene™ Cryo 1 Freezing Container to achieve a -1°C/min cooling rate for optimal cryopreservation of the cells. Two days after this freezing process the vials were transferred into liquid nitrogen storage for maximum preservation. Cells were thawed in a 37°C water bath. The thawed cells were diluted into appropriately pre-warmed cell culture medium in a 1:10 ratio. Cells were centrifuged for 5 minutes at 1400 rpm, the supernatant was aspirated and the cell pellet re-suspended using 1ml of pre-warmed cell culture medium. Cells were transferred to a labeled 6cm or 10cm cell culture dish containing cell culture medium.

2.2.3 Immunofluorescence

Loss of *Pou5f1* gene expression was tested via immunofluorescence. Cells were washed four times with phosphate buffered saline (PBS) and cross-linked using 4% paraformaldehyde for 30 minutes at room temperature. The cross-link reaction was quenched with 50mM Glycine for 15 minutes at room temperature. After quenching the cells were washed four times with PBS. Cell membranes were permeated with 0.1% Triton-X100 and blocked with 5% BSA and 1% FBS in PBS for > 2 hours rocking at 4C. Primary antibody against OCT4 and/or SOX2 was added (see table 1) in blocking solution (5% BSA/1% FBS in PBS) and incubated for > 2 hours rocking at room temperature/4°C overnight. Cells were washed four times with PBS. Cells were incubated with secondary antibody (see table 2) and Hoechst in PBS for 20 minutes at room temperature in the dark. Cells were washed four times with PBS.

Table 1. Primary antibodies used for immunofluorescence

Antibody	Origin	Dilution	Company	ID
OCT4	Goat	1:1000	Santa Cruz	sc5279
SOX2	Goat	1:500	Santa Cruz	sc17320

Table 2. Secondary antibodies used for immunofluorescence

Antibody	Origin	Dilution	Company	ID
Anti-goat-alexa-488	Rabbit	1:1000	ThermoFisher Scientific	A11078
Anti-goat-alexa-568	Rabbit	1:1000	ThermoFisher Scientific	A11061

2.2.4 RNA isolation (TRIzol)

60-80% confluent wells of 6-well plate were washed one time with PBS. Next, 1ml of TRIzol® was added and the cell/TRIzol® mix was homogenized with a pipette and incubated for 5 minutes at room temperature. TRIzol® samples were either stored at -80°C or directly processed for RNA extraction. 1/5 volume of chloroform was added to the TRIzol®, mixed vigorously and incubated at room temperature for 5 minutes. Samples were centrifuged for 15 minutes at 12.000 x g at 4°C. The upper aqueous phase was transferred to a new Eppendorf tube. 1 volume of 2-propanol was added to the aqueous phase and mixed by inverting. Samples were incubated at room

temperature for 10 minutes before centrifugation for 30 minutes at 12000 x g at 4°C. The supernatant was discarded and RNA pellets were washed 2 times with 75% freshly prepared ethanol and centrifuged for 5 minutes at 12.000 x g at 4°C. All ethanol was removed carefully and the RNA pellets were air-dried for 5-10 minutes. RNA pellets were re-suspended in RNase-free water. The RNA concentration was measured using Nanodrop1000.

2.2.5 Reverse transcription (cDNA)

RNA was reverse-transcribed into cDNA using the MMLV Reverse transcription kit from Promega. For each sample 1μg RNA was added to a total volume of 18.9μl H_2O to which 6.1μl master mix was added. This master mix contained 5μl of 5X MMLV Master mix, 0.5μl of 100mM dNTPs, 0.5μl Oligo(dT), 0.1μl MMLV reverse transcriptase enzyme (200units/μl). The mixture was incubated according to protocol specified in table 3. Next, the mixture was diluted with 175μl H2O to 5ng/μl for quantitative polymerase chain reaction (qPCR). cDNA samples were stored at -20°C.

Table 3. PCR protocol used for RT-PCR

Step	Temperature	Time
Annealing	42°C	60:00
Extention	72°C	10:00
Soak	4°C	∞

2.2.6 Reverse Transcribed-quantitative Polymerase Chain Reaction

RT-qPCR in 384-well plates in triplicate format was performed as described below. followed, in a 96-well plate 4μl cDNA (5ng/μl) was added to 31μl master mix (17.5μl 2x Bio-Rad SYBR® Green, 13.2μl H_2O, 0.15μl 100mM forward primer and 0.15μl 100mM reverse primer). cDNA/Mastermix was mixed well with a multi-channel pipette by pipetting up and down and washing the well walls. Each well of a 96-well plate was used to pipette 10μl into a well of a 384-well plate making up a triplicate. The 384-well plate was sealed off with a optically clear adhesive seal sheet and centrifuged for 1 minute at 600 x g. It was placed in a Applied Biosystems Quantstudio™ 5; thermal cycling conditions were used as described in table 4,

including melt curve analysis. Ct value data was exported and analyzed in Microsoft Excel using the

$^{-2\Delta\Delta}$Ct method and normalized against housekeeping genes *Gapdh* or *Rplp0*. Primer sequences are listed in table 5.

Table 4. PCR protocol used for RT-qPCR

Step	Temperature	Time	Number of cycles
Polymerase activation	95°C	0:30	1
Denaturing	95°C	0:15	40
Annealing/extension	60°C	0:45	
Melt curve	95°C	0:15	1
	60°C	1:00	
	95°C	0:15	

Table 5. Oligonucleotide sequences of RT-qPCR primers

Name	Sequence 5' > 3'
*Gapdh*_fwd	CCAATGTGTCCGTCGTGGAT
*Gapdh*_rev	TGCCTGCTTCACCACCTTCT
*Rplp0*_fwd	CAAAGCTGAAGCAAAGGAAGAG
*Rplp0*_rev	AATTAAGCAGGCTGACTTGGTTG
*Pou5f1*_fwd	CACGAGTGGAAAGCAACTCA
*Pou5f1*_rev	AGATGGTGGTCTGGCTGAAC
*Sox2*_fwd	TTCGAGGAAAGGGTTCTTGCTG
*Sox2*_rev	TCCTTCCTTGTTTGTAACGGTCCT
*Klf4*_fwd	TGTGTCGGAGGAAGAGGAAGC
*Klf4*_rev	ACGACTCACCAAGCACCATCA
*Esrrb*_fwd	AGGCTCTCATTTGGGCCTAGC
*Esrrb*_rev	ATCCTTGCCTGCCACCTGTT
*Nanog*_fwd	GAACGGCCAGCCTTGGAAT
*Nanog*_rev	GCAACTGTACGTAAGGCTGCAGAA
*Utf1*_fwd	ACGTGGAGCATCTACCAGGT
*Utf1*_rev	TAGACTGGGGGTCGTTTCTG

2.2.7 Cross-linking of ZHBTc4

ZHBTc4 cells were washed with PBS and harvested using Accutase® at the given time points of loss of OCT4. Cells were centrifuged for 5 minutes at 1400rpm, supernatant was aspirated and cell pellet re-suspended as single cell suspension in cell culture medium. Cell number was determined using counting chamber. $2x10^7$ cells were diluted in a total of 3.6ml of cell culture medium in a 15ml falcon tube. 100µl 37% formaldehyde was added, mixed with the cell suspension and incubated at room temperature for 8 minutes while inverted every 30 seconds. 370µl of 1.375M Glycine was added to quench the formaldehyde and incubated for 5 minutes. Cross-linked cells were centrifuged for 5 minutes at 1350 x g and washed twice with 1 ml of cold PBS, between washes centrifuged for 5 minutes at 1350 x g. Cells were either stored at -80°C or directly processed for chromatin extraction.

2.2.8 Chromatin isolation

The cross-linked cell pellets were thawed and loosened by flicking the 15ml falcon tube. Loose cell pellets were re-suspended in 1 ml of 'lysis buffer 1' containing 50mM Hepes-KOH pH7.5, 140mM NaCl, 1mM EDTA pH8.0, 10% Glycerol, 0.5% Igepal CA630, 0.25% Triton-X100, 1x protease inhibitor. Cell suspension was incubated on a roller bank for 30 minutes at 4°C. Next, the cell suspension was centrifuged for 5 minutes at 1350 x g at 4°C. Supernatant was removed and the remaining pellet was re-suspended in 1ml of 'lysis buffer 2' containing 10mM Tris-HCl pH8, 200mM NaCl, 1mM EDTA, 0.5mM EGTA. Cell suspension was incubated for 10 minutes on a roller bank at 4°C and centrifuged for 5 minutes at 1350xg at 4°C. Supernatant was removed and remaining cell pellet was re-suspended in 2ml SDS sonication buffer containing 10mM Tris-HCl pH8, 10mM EDTA pH8, 0.5% SDS, 1x protease inhibitor ($1x10^7$ cells/ml) and incubated for 10 minutes on ice. 1ml of SDS sonication buffer containing cell pellet was transferred to TPX sonication tubes for sonication. $1x10^7$ cells were sonicated in a Diagnod Bioruptor® for four rounds of 15 minutes, pulse 30 seconds ON/30 seconds OFF, high intensity. Supernatant was collected into 2ml Eppendorf Protein LoBind tubes and centrifuged for 10 minutes at 15000 rpm at 4°C. Supernatant was pooled and stored in 2ml Eppendorf Protein LoBind tubes at -80°C until chromatin immunoprecipitation. 25µl of chromatin was set aside to assess the

shearing efficiency. 175µl H$_2$O, 10µl 5M NaCl, and 1µl of 20mg/ml RNaseA were added to the 25µl of chromatin and incubated overnight at 65°C in a hybridization oven. Next, 1µl of 20mg/ml Proteinase K was added and incubated at 55°C for 2 hours at 600rpm in a thermal mixer. Chromatin was purified using Machnery-Nagel PCR spin column purification kit and eluted with 15µl pre-warmed (65°C) elution buffer. DNA concentration was measured on a Nanodrop 1000 and a total of 1µg DNA was run on 1.2% agarose gel for 45 minutes at 90V/400mA. Size distribution was determined based on 1kb marker.

2.2.9 Westernblot analysis on chromatin material

Equal amounts of chromatin were diluted using 4x lami sample buffer. Samples were boiled at 98°C for 45 minutes and centrifuged at 16100 x g for 1 minute. 0.75mm SDS-Page gels were prepared using 12% running gel (12% acrylamide/Bis-acrylamide, 0.75M Tris-HCl (pH 8.8), 0.1% SDS, 0.1% ammonium persulfate, 0.33% TEMED) and 4% stacking gel (4% Acrylamide/Bis-acrylamide, 0.18M Tris-HCl (pH6.8), 0.1% SDS, 0.3% Ammonium persulfate, 0.96% TEMED). Gels were assembled in a Bio-Rad running chamber containing 1x running buffer covering the entire gel. 1µg of chromatin of each sample was loaded per well (for H3 blots 1/10 of volume was loaded). Gels were run at 10V per gel /400mA until the samples reached the running gel, the voltage was elevated to 40V until the samples reach the bottom of the SDS-Page gel. Next, the protein samples were transferred to PVDF membrane using Bio-Rad transfer system. Transfer was performed in 1x transfer buffer containing 15% methanol for 1,5 hours at 300V, 400mA at 4°C on a magnetic stirrer. PVDF membranes were blocked for 1 hour in 5% skim milk in PBS containing 0.05% Tween-20 (PBS-T) at room temperature followed by incubation with primary antibody (see table 6) in PBS-T containing 5% skim milk overnight at 4°C on a rollerbank. Next, the membranes were washed extensively with PBS-T (4-6x for 5-10 minutes) and incubated with secondary HRP-coupled antibody (see table 7) in PBS-T with 5% skim milk for 1 hour at room temperature. Then the membranes were washed extensively with PBS-T and a final wash of PBS. Membranes were exposed to ECL to activate the HRP on the secondary antibody. Protein expression was visualized using film and developer.

Table 6. Primary antibodies used for Western blotting

Antibody	Origin	Dilution	Company	ID
OCT4	Mouse	1:2500	Santa Cruz	sc5279
SOX2	Goat	1:1000	Santa Cruz	sc17320
NANOG	Rabbit	1:5000	Bethyl Laboratories (Biomol)	A300-397A
ESRRB	Mouse	1:5000	R&D systems	pp-h6705-00
H3	Rabbit	1:20000	Abcam	ab1791

Table 7. Secondary antibodies used for Western blotting

Antibody	Origin	Dilution	Company	ID
Anti-mouse-HRP	Goat	1:20000	Jackson Labs	115-035-044
Anti-goat-HRP	Chicken	1:20000	R&D systems	HAF019
Anti-rabbit-HRP	Donkey	1:20000	GE Healthcare	NA934

2.2.10 Chromatin immunoprecipitation

25µl protein G Dyna Beads® per 25µg of chromatin was separated in 1.5ml protein low binding tubes using a DynaMag2™ magnetic stand, and storage buffer was removed. 500µl cold PBS was added and the beads were gently shaken for 30 seconds. After a brief spin the tubes were again placed on the magnetic stand. PBS was removed and the beads were washed two more times with 500µl PBS containing 0.02% Tween-20 and 0.1% BSA (PBS-T-B). After removal of the last PBS-T-B wash the beads were re-suspended in 250µl PBS-T-B and 2µg of antibody was added to conjugate antibody to the magnetic beads (list of antibodies in table 9). Beads/PBS-T-B/antibody mixture was incubated for 2 hours at 4°C on an end-over-end rotator. Next, the beads were washed as described before using PBS containing 0.02% Tween-20 (PBS-T). After removal of the last wash the beads were re-suspended in 4 times the volume of the chromatin samples with the lowest concentration using ChIP dilution buffer containing 10mM Tris-HCl pH8, 125mM NaCl, 0.125% Sodiumdeoxycholate, 1.25% Triton X-100, 1x protease inhibitor. The other samples were adjusted by adding SDS sonication buffer into the ChIP reaction. Next, 25µg chromatin was added and incubated with the antibody-coupled beads overnight at 4°C on an end-over-end rotator. 10% chromatin was set aside for input and stored at

4°C overnight. The following day all liquid was collected with a brief spin, tubes were placed on magnetic stand and the supernatant was removed without disturbing the beads. Beads were washed with low salt buffer (20mM Tris-HCl pH8, 150mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100), 2 times high salt buffer (20mM Tris-HCl pH8, 500mM NaCl, 0.1% SDS, 1% Triton X-100), 2 times RIPA washing buffer (50mM Hepes-KOH pH7.6, 250mM LiCl, 1mM EDTA, 1% Igepal CA630, 0.7% Sodium Deoxycholate), and 1 time Tris/EDTA (TE)-containing 50mM NaCl as described below:

500µl of the wash buffer was added to the beads and the tubes were rotated on the removable plastic rack of the DynaMag™ 2 for 30 seconds. Next, the liquid was collected with a brief spin and the tubes were placed on the magnetic rack. After 30 seconds the liquid was removed and the next washing step was performed. After the final wash the tubes were centrifuged briefly to collect all remaining TE-containing 50mM NaCl and placed back onto the magnetic stand. All liquid was removed. Next, the beads were re-suspended in 105µl ChIP Elution Buffer (10mM Tris-HCl pH8, 5mM EDTA, 300mM NaCl, 0.5% SDS). Tubes were incubated in a thermal mixer at 65°C 1400rpm for 15 minutes. Next, the tubes were centrifuged for 1 minute at 16.000 x g and placed onto the magnetic stand. Supernatant was transferred to a new 1.5ml protein low binding Eppendorf tube and 1µl 20mg/ml RNaseA was added to each sample. Samples were incubated overnight at 65°C. For input samples, ChIP elution buffer was added up to a final volume of 100µl and 1µl 20mg/ml RNase A was added. Input samples were also incubated overnight at 65°C. The next day, tubes were removed from 65°C and briefly centrifuged to collect all condensation. 1µl 20mg/ml Proteinase K was added and samples were incubated in a thermal mixer at 55°C for 2 hours at 600rpm. Last, the samples were purified over Machnery-Nagel PCR spin columns and eluted with 100µl pre-warmed (65°C) elution buffer. At this step, parallel reactions were pooled onto the column. For ChIP-seq samples a total elution volume of 50µl. Samples were stored at -20°C.

Table 8. Antibodies used for ChIP

Antibody	Origin	Amount per	Company	ID
OCT4	Goat	20µl/25µg chromatin	R&D	AF1759
OCT4	Goat	10µl/25µg chromatin	Santa Cruz	SC8628
SOX2	Goat	1.5µg/25µg chromatin	Neuromics	GT15098
NANOG	Rabbit	2µl/25µg chromatin	Bethyl Laboratories	A300-397A
ESRRB	Mouse	3µl/25µg chromatin	R&D	pp-h6705-00
KLF4	Goat	10µl/25µg chromatin	R&D	AF3158

2.2.11 ChIP-qPCR

Input samples at 10% were serial-diluted to 2%, 0.4%, 0.08%, and 0.016%. For ChIP-sequencing samples the input was serial-diluted to 1%, 0.2%, 0.04%, and 0.008%. ChIP-qPCR samples were processed the same way as RT-qPCR samples with a slight difference in the PCR protocol (see table 9). qPCR on ChIP-seq samples was done as follows; 2µl ChIP-seq sample was diluted 4 to 8-fold (normal elution volume for 25µg is 100µl, for Oct4 ChIP 100µg chromatin was used, for Sox2 ChIP 50µg). qPCR in a 384-well plate in duplicate format was performed as described in RT-qPCR section 2.2.6 with the following differences, in a 96-well plate 2.67µl ChIP material was added to 20.67µl master mix (11.67µl 2x Bio-Rad SYBR® Green, 8.8µl H_2O, 0.1µl 100mM forward primer, 0.1µl 100mM reverse primer). Analysis of the data was performed using the Growth function in excel to superimpose the ChIP samples onto a standard curve of the input. Samples were normalized against background regions such as 28s rDNA, the complete list of ChIP-qPCR primers can be found in table 10.

Table 9. PCR protocol used for ChIP-qPCR

Step	Temperature	Time	Number of cycles
Polymerase activation	95°C	0:30	1
Denaturing	95°C	0:15	40
Annealing/extension	57°C	0:45	
Melt curve	95°C	0:15	1
	60°C	1:00	
	95°C	0:15	

Table 10. Oligonucleotide sequences of ChIP-qPCR primers

Name	Sequence 5' > 3'
grDNA-28S_F	CTGGGTATAGGGGCGAAAGAC
grDNA-28s_R	GGCCCCAAGACCTCTAATCAT
grDNA-IGS1_F	GGCCAGTTCCTCCTGCCTTCTGTT
grDNA-IGS1_R	ACTGTGGATGGAGCGTGCATGTGT
gPou5f1-a(cr4)_2F	TGGGCAGACGGCAGATGCATAACA
gPou5f1-a(cr4)_2R	GGGACCCCTCCCCAACCATCTTCT
gSox2-b(srr2)_F	AGTCCAAGCTAGGCAGGTTCCCCT
gSox2-b(srr2)_R	TGCCCGAGCCCGGGAAATTCTTTT
gNanog-a(pe)_F	GATGCCCCCTAAGCTTTCCCTCCC
gNanog-a(pe)_R	TAATCCCACCTGCAGGGTCCACCA
gUtf1-a_F	GGGGAGGGCTTAGGTGCAGGTAGA
gUtf1-a_R	GGCCGGATGGGCCCAGAATTTGTA
gKlf4_b_F	CCAAGTTGAAATTGATGAGTGTGT
gKlf4_b_R	ACACATTGAAATTCACCCACTTT
gKlf4_e_F	CCTGAGCAACCTTCCTGGCAAAGG
gKlf4_e_R	CCCACACTTTGTTGGTCAGCTGCTT

2.2.12 ChIP-seq library construction

Library construction was performed using NEBNext® Ultra™ DNA Library Prep Kit from Illumina® (NEB E7370) according to the manufacturers protocol (version 5.0 5/17) with alterations at the Adaptor amplification step. In brief, ChIP samples were end-repaired to allow for adaptor ligation using NEBNext® End Prep. Samples were incubated for 30 minutes at 20°C followed by 30 minutes at 65°C. Next, NEBNext® Adaptor for Illumina® was diluted 10-fold to 1.5µM using 10mM Tris-HCl pH7.4-containing 10mM NaCl. Adaptors were ligated for 15 minutes at 20°C followed by opening of the adaptor with the USER™ enzyme for 15 minutes at 37°C. Libraries were cleaned up without size selection using AMPure XP beads at a 1:1 ratio. PCR enrichment of Adaptor-ligated DNA was performed using KAPA Biosystems KAPA Real-time Library Amplification Kit (KK2702). In a 96-well format, 15µl library was mixed with 25µl 2x KAPA HiFi HotStart Real-time PCR Master Mix, 5µl Index primer (see table 11) and 5µl universal primer. In addition, 4 standards were run in duplicate to determine the optimal stopping point for library amplification within the linear phase of the amplification. PCR protocol used as described in table 12. Next, the libraries were purified using two rounds of AMPure XP beads in a 1:0.9 ratio. Libraries were

quantitatively and qualitatively analyzed using Qubit® 3.0 and Agilent Bioanalyzer. Libraries were sequenced on an Illumina NextSeq 550. Remaining libraries were stored at -20°C.

Table 11. Oligonucleotide sequences of ChIP-seq library indexing primers

Product	Index primer sequence	Index primer sequence read
NEBNext Index 1 Primer	5´- CAAGCAGAAGACGGCATACGAGATCGTGATGTG ACTGGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	ATCACG
NEBNext Index 2 Primer	5´-CAAGCAGAAGACGGCATACGAGATACATCGGTG ACTGGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	CGATGT
NEBNext Index 3 Primer	5´-CAAGCAGAAGACGGCATACGAGATGCCTAAGTGA CTGGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	TTAGGC
NEBNext Index 4 Primer	5´-CAAGCAGAAGACGGCATACGAGATTGGTCAGTGA CTGGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	TGACCA
NEBNext Index 5 Primer	5´-CAAGCAGAAGACGGCATACGAGATCACTGTGTGA CTGGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	ACAGTG
NEBNext Index 6 Primer	5´-CAAGCAGAAGACGGCATACGAGATATTGGCGTGA CTGGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	GCCAAT
NEBNext Index 7 Primer	5´-CAAGCAGAAGACGGCATACGAGATGATCTGGTGA CTG GAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	CAGATC
NEBNext Index 8 Primer	5´-CAAGCAGAAGACGGCATACGAGATTCAAGTGTGA CTG GAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	ACTTGA
NEBNext Index 9 Primer	5´-CAAGCAGAAGACGGCATACGAGATCTGATCGTGA CTG GAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	GATCAG
NEBNext Index 10 Primer	5´-CAAGCAGAAGACGGCATACGAGATAAGCTAGTGA CTG GAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	TAGCTT
NEBNext Index 11 Primer	5´-CAAGCAGAAGACGGCATACGAGATGTAGCCGTGA CTG GAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	GGCTAC
NEBNext Index 12 Primer	5´-CAAGCAGAAGACGGCATACGAGATTACAAGGTGA CTG GAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	CTTGTA
NEBNext Index 13 Primer	5´-CAAGCAGAAGACGGCATACGAGATTGTTGACTGTGA CT GGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	AGTCAA
NEBNext Index 14 Primer	5´-CAAGCAGAAGACGGCATACGAGATACGGAACTGTG ACTGGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	AGTTCC

Product	Index primer sequence	Index primer sequence read
NEBNext Index 15 Primer	5´-CAAGCAGAAGACGGCATACGAGATTCTGACATGTG ACTGGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	ATGTCA
NEBNext Index 16 Primer	5´-CAAGCAGAAGACGGCATACGAGATGCGGACGGGT GACTGGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	CCGTCC
NEBNext Index 18 Primer	5´-CAAGCAGAAGACGGCATACGAGATGTGCGGACGT GACTGGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	GTCCGC
NEBNext Index 19 Primer	5´-CAAGCAGAAGACGGCATACGAGATCGTTTCACGT GACT GGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	GTGAAA
NEBNext Index 20 Primer	5´-CAAGCAGAAGACGGCATACGAGATAAGGCCACGT GACTGGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	GTGGCC
NEBNext Index 21 Primer	5´-CAAGCAGAAGACGGCATACGAGATTCCGAAACGT GAC TGGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	GTTTCG
NEBNext Index 22 Primer	5´-CAAGCAGAAGACGGCATACGAGATTACGTACGGT GACT GGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	CGTACG
NEBNext Index 23 Primer	5´-CAAGCAGAAGACGGCATACGAGATATCCACTCGT GACT GGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	GAGTGG
NEBNext Index 25 Primer	5´-CAAGCAGAAGACGGCATACGAGATATATCAGTGT GACT GGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	ACTGAT
NEBNext Index 27 Primer	5´-CAAGCAGAAGACGGCATACGAGATAAAGGAATGT GACT GGAGTTCAGACGTGTGCTCTTCCGATC-s-T-3´	ATTCCT
NEBNext Adaptor	5´-/5Phos/GATCGGAAGAGCACACGTCTGAACTCCAGT CdUACACTCTTTCCCTACACGACGCTCTTCCGATC-s-T-3´	N/A
NEBNext Universal PCR Primer	5´-AATGATACGGCGACCACCGAGATCTACACTCTTTCC CTACACGACGCTCTTCCGATC-s-T-3´	N/A

Table 12. PCR protocol used for amplification of ChIP-seq library

Step	Temperature	Time	Number of cycles
Polymerase activation	98°C	0:45	1
Denaturing	98°C	0:15	Based on standards
Annealing	60°C	0:30	
Extension	72°C	0:30	

2.2.13 Transient transcriptome sequencing (TT-seq)

At given time points of OCT4 knockout in ZHBTc4 cells the cell culture medium was supplemented with 500µM 4-thiouridine for 5 minutes (Schwalb et al. 2016). Medium was aspirated and TRIzol® was added and incubated for 5 minutes. All TRIzol® samples were collected, multiple plates were pooled and samples were stored at -80°C. After thawing of the TRIzol® samples 5µl ERCC spike-in RNA mix (6ng/µl and 1µg/µl each, see table 13) per 1×10^8 cells was added and mixed. 1/5 volume of chloroform was added and samples were mixed thoroughly by shaking by hand for 15 seconds and incubated for 5 minutes at room temperature. Next, samples were centrifuged at 12000 x g for 15 minutes at 4°C. The upper aqueous phases were transferred to new tubes and equal volumes of 100% isopropanol were added, mixed by inversion and incubated at room temperature for 10 minutes. Samples were centrifuged at 12000 x g for 30 minutes at 4°C, supernatant was discarded and RNA pellet washed with 10ml of freshly prepared 75% ethanol. Samples were centrifuged 12000 x g for 10 minutes at 4°C, supernatant was removed and 600µl 75% ethanol was added to the RNA pellet to be transferred to new 2ml Eppendorf tube. This process was repeated three times to collect the entire RNA pellet. Samples were centrifuged 12000 x g for 10 minutes at 4°C, supernatant was removed and the RNA pellet was air-dried for 10 minutes at room temperature. RNA pellet was re-suspended in 135µl RNase-free H_2O and incubated 10 minutes at 60°C. RNA was quantified with Nanodrop 2000. 300µg of RNA was adjusted to a total volume of 130µl and prepared for sonication using Covaris® S220 with the following settings: intensity = 100W, duty cycle = 1%, cyc/burst = 200, duration = 10 seconds, temperature 4-8°C. Sonicated RNA was transferred to 2ml PP plastic tubes (Sarstedt). 1µl of fragmented RNA plus 9µl H_2O was set aside as total RNA and stored at -80°C. For biotinylation the RNA was denatured at 60°C for 10 minutes and placed on ice for 2 minutes, condensation was collected with a brief spin and water was added to a final volume of 1400µl and divided into two 2ml PP plastic tubes (150µg RNA per tube). 100µl biotinylation buffer (100mM Tris-HCl (pH7.5), 10mM EDTA (pH8.0), 200µl 1mg/ml HPDP-Biotin) was added to each tube and incubated for 1.5 hours at room temperature at 800 rpm in the dark. Next, 2ml phase lock tubes were prepared by centrifugation at 15000 rpm for 1 minute at room temperature. Biotinylated RNA was added to the phase lock tubes with 700µl chloroform and

mixed vigorously. Tubes were centrifuged at 15000 rpm for 5 minutes at room temperature. The upper aqueous phases were transferred to new 2ml eppendorf tubes; 100μl 5M NaCl and 1ml isopropanol was added and mixed by inverting the tubes. Samples were centrifuged at 15000 rpm for 30 minutes at 4°C, supernatant was removed and RNA pellet washed with freshly prepared 75% ethanol and centrifuged at 15000 rpm for 10 minutes at 4°C. Ethanol was completely removed and RNA pellets of 1 tube were re-suspended in 100μl Rnase-free H_2O, incubated for 5 minutes at room temperature, then for 5 minutes at 60°C and this was then used to resuspend the second of the two tubes, incubated 5 minutes at room temperature and incubated 5 minutes at 60°C. RNA was quantified using Nanodrop 2000. Labeled RNA was separated using streptavidin μBeads. Biotinylated RNA was denatured at 65°C for 10 minutes and placed on ice for 5 minutes. Next, 100μl streptavidin μBeads were added to each sample and incubated for 15 minutes at 24°C 400rpm in the dark. MACSμ columns were washed with 900μl room temperature washing buffer (100mM Tris-HCl (pH7.5), 10mM EDTA (pH8.0), 1M NaCl, 0.1% Tween-20) to pre-run and equilibrate the column. The beads/RNA mixture was added to the columns and the flow-through was collected and re-loaded onto the column three times. Next, the columns were washed three times with pre-warmed (65°C) washing buffer and three times with 900μl room temperature washing buffer. RNA was eluted off the column in two rounds using 100mM DTT for each elution, gently applying pressure on the top of the column for a complete elution. Isolated labeled RNA was purified using Qiagen miRNeasy kit with a few alterations to the manufacturers protocol. At this point, total RNA was thawed and processed as the enriched labeled RNA. 300μl 100% ethanol was added to the 200μl enriched labeled RNA samples, for total RNA 200μl H_2O, 20μl sodium acetate and 300μl 100% ethanol was added. Samples were loaded onto the column and the flow-through was reloaded one more time for optimal yield. Columns were washed with RWT buffer supplemented with isopropanol instead of ethanol. Columns were DNaseI treated for 15 minutes at room temperature. Next, columns were washed with RWT with isopropanol, flow-through was re-loaded one time, washed with RPE buffer and then with freshly prepared 80% ethanol. RNA was eluted off the columns using 15μl RNase-free H2O; the flow-through was reloaded. RNA concentrations were quantified using Nanodrop2000 and enrichment for labeled RNA was determined using RT-qPCR for the ERCC spike-in standards. RNA was stored at -80°C, or directly used for RNA-seq library preparation.

2.2.14 RNA-seq Library construction

RNA-seq libraries were constructed using NuGEN Ovation® Universal RNA-seq System according to the manufacturers user guide. In brief, DNaseI-treated RNA ranging from 10 to 100ng was used for First Strand cDNA synthesis and Second Strand cDNA synthesis. cDNA was then sonicated in a Covaris® S220 with parameters set at: intensity = 175W, duty cycle = 10%, cyc/burst = 200, duration = 180 seconds, temperature 4-8°C, to reach 200bp fragments. cDNA was concentrated using Agencourt RNAClean XP beads in a 1:1.8 ratio. After end repair, adaptor ligation and strand selection, the cDNA was purified with Agencourt RNAClean XP beads in a 1:0.8 ratio. Next, the second strand selection was performed and 5µl of the library was amplified using KAPA Real-time Library Amplification Kit to determine the optimal cycle number for the actual library amplification. The libraries were amplified based on the results of the KAPA Real-time Library Amplification Kit and purified twice using Agencourt RNAClean XP beads in a 1:1 ratio. cDNA libraries were quantified using Qubit® 2.0, Tape station, and fragment analyzer. Libraries were processed on Illumina NextSeq550, and the remainder of the library was stored at -80°C.

2.2.15 Assay for Transposase Accessible Chromatin sequencing

ATAC-seq was performed using the Illumina Nextera DNA library preparation kit as described by Buenrostro et al. 2015 with a few alterations. At indicated time points ZHBTc4 cells were harvested using Accutase®. After centrifugation at 500 x g for 5 minutes the cell pellets were washed in cold (4°C) PBS and counted using a counting chamber. 50000 cells were then re-suspended in 50µl cold PBS and 50µl 2x nuclei isolation buffer (20mM Tris-HCl Ph7.4, 20mM NaCl, 6mM MgCl$_2$, 0.2% Igepal GA630, 0.5% Triton X-100). Cell suspension was incubated for 10 minutes on ice and centrifuged at 500 x g for 10 minutes at 4°C. Supernatant was aspirated and the remaining nuclei were re-suspended in 45µl RNase-free H$_2$O, 50µl TD1 buffer and 5µl TDE1 Tn5 enzyme. Samples were then incubated in thermal mixer for 45 minutes at 37°C at 600rpm. Next, samples were purified over Machnery-Nagel PCR spin column, elution was performed using 12µl. Samples were then PCR-amplified: 10µl transposed DNA, 25µl 2x NEB HiFi PCR enzyme mix, 0.625µl Ad1_nomx (see table

14), 0.625µl Reverse primers with unique barcodes (see table 13) and 13,75 RNase-free H$_2$O. PCR program was run as listed in table 15. After 5 cycles of PCR, 5µl library was used in KAPA Real-time Library Amplification Kit with fluorescent standards to determine the optimal cycle number for library amplification. Next, the remaining 45µl of library was amplified according to the results of the KAPA Real-time Library Amplification Kit. PCR-amplified libraries were purified over Machnery-Nagel PCR spin column and purified again using AMPure XP beads in a 1:1.8 ratio. Libraries were qualitatively and quantitatively assessed using Qubit® 3.0 and Bioanalyzer. Libraries were processed on Illumina NextSeq550, and remaining libraries were stored at -80°C.

Table 13. ERCC spike-in RNA sequences used for TT-seq

ERCC ID	SEQUENCE	LABELED / UNLABELED
ERCC-00004 DQ516752 AC034598 74_A1 AC034599 70_A1	TCTTGCTTCAACAATAACGTCTCTTTCAGAAGGCATTGGTATCTTT TCCCCACTTCCAAGCATTTTTTCAACTAATCTTATGTTATTAACCAT TTCCTTAAATTCTTCTGGGTCTGCTGACAAAGCATGATCAGGACC TTCCATATTTTTATCTAAGGTAAAGTGCTTCTCAATAACATCCGCT CCTAAGGCAACAGAAACTACTGGGGCGAGTATTCCCAATGTATG GTCAGAATATCCCACAGGGATATTGAATATACTTTTCAAGGTTTTA ATAGCGTTTAAATTGACATCTTCATAAGGGGTTGGGTAAGATGAA ATACAATGCAATAAAATAATATCCCTGCATCCATTATTTTCTAAAAC TTTAACTGCTTCCCAAATTTCCCCAATATCAGACATTCCTGTAGAT AAAATCACCGGCTTGCCTGTTTTTGCCACTTTTTCTAATAAGGGAT AAAAGGTTAAATCACCAGAGGCAATTTTAAATCAGGCACATAAAAA AAAAAAAAAAAAAAAAAAAA	Unlabeled
ERCC-00012 DQ883670 AC034598 77_A1 AC034599 73_A1	CGAGAGATGTTTGTAGGTGCGGAATGTGTGCGGTCTACCTTAGCT GTAGTGTGCGATGAACCTACACACAACGTGGTATAGTGGCCGAT CTTAGAGTGATCCTATCACTCCTTACGCACCAGAAGGGATCTGCA TACCAGGCGGAGAACTTGGAAGGCGGCTAGATCACTGAATTGCG GGAATCGGCATTTCGCATTCTTAGGATCTAAACCTTAGACCTCCG CGTGCGATTGCACCTGCTTGGTACAGAGTTACAAGCCCCCCGCA CTTTCTTTGCGGTCGTTAAGAGGGAAATCGCCCAATTAGCAGAGT GTCAGGTGTTACGCGCGATTGAGCCGTCAGAAGAATCGATAGAG CCGCGTCGGGACCTTGATGGTATCTCTGCCTCAGCTAACCTGCTA GGTCCGTCCCCTGGGGATGATCAGGACTGCGGATAGTAAATTGC GGGTTTGAAGCCGGACTTGCCGCCTAGGCAAAGCACAAAAACAT CGGACATGTAGAAGTCTCATCGAACTCCTTTCCCGTTCATGCAGA TACTTCAACTGTGACTAGTGGGGTTCGGGAGCACCCGCACTACTT CATTCTTGGCGGTGGGCCACTTTATGTGACTGTACATGGGACTTC TACTCATACCAATGTAAAGTATAGTTAACGCCCTGTCCACTCTACT CAGGCGTAATCATCGCGGAAGGCTATCCACAGCCCATCAGCGGT CTACATGTCCCAGCAGATTCACCTGTCCTGCGGGTCCGCGTCAC AGCCTATTCTGAGGCTCTAAAGACTATGCGAACCAGGTGTCCCAG TCGATCAGACGACGAAGTCGGGAAGGAAGCATGGATACCAAAAA GGCTTTATATACTGGGTTATCCTAGGGGATGTTTTTACCGGACTG GTCAGCCTCGGTGCGCTCGGCCTAGGCGCTTACTGCATGGGGG CTGTGGGCAATTTGGTATTTCTCAGGACTATGGACAAAAAAAAA AAAAAAAAAAAAA	Labeled

Table 14. Oligonucleotides used for indexing ATAC-seq libraries

Original name	Sequence 5' > 3'	Seq reads
Ad1_nomx	AATGATACGGCGACCACCGAGATCTACACTCGT CGGCAGCGTCAGATGTG	
Ad2.1	CAAGCAGAAGACGGCATACGAGATTCGCCTT AGTCTCGTGGGCTCGGAGATGT	TAAGGCGA
Ad2.2	CAAGCAGAAGACGGCATACGAGATCTAGTACG GTCTCGTGGGCTCGGAGATGT	CGTACTAG
Ad2.3	CAAGCAGAAGACGGCATACGAGATTTCTGCCTG TCTCGTGGGCTCGGAGATGT	AGGCAGAA
Ad2.4	CAAGCAGAAGACGGCATACGAGATGCTCAGGAG TCTCGTGGGCTCGGAGATGT	TCCTGAGC
Ad2.5	CAAGCAGAAGACGGCATACGAGATAGGAGTCC GTCTCGTGGGCTCGGAGATGT	GGACTCCT
Ad2.6	CAAGCAGAAGACGGCATACGAGATCATGCCTAG TCTCGTGGGCTCGGAGATGT	TAGGCATG
Ad2.7	CAAGCAGAAGACGGCATACGAGATGTAGAGAGG TCTCGTGGGCTCGGAGATGT	CTCTCTAC
Ad2.8	CAAGCAGAAGACGGCATACGAGATCCTCTCTGG TCTCGTGGGCTCGGAGATGT	CAGAGAGG
Ad2.9	CAAGCAGAAGACGGCATACGAGATAGCGTAGC GTCTCGTGGGCTCGGAGATGT	GCTACGCT
Ad2.10	CAAGCAGAAGACGGCATACGAGATCAGCCTCG GTCTCGTGGGCTCGGAGATGT	CGAGGCTG
Ad2.11	CAAGCAGAAGACGGCATACGAGATTGCCTCT TGTCTCGTGGGCTCGGAGATGT	AAGAGGCA
Ad2.12	CAAGCAGAAGACGGCATACGAGATTCCTCTAC GTCTCGTGGGCTCGGAGATGT	GTAGAGGA
Ad2.13	CAAGCAGAAGACGGCATACGAGATATCACGAC GTCTCGTGGGCTCGGAGATGT	GTCGTGAT
Ad2.14	CAAGCAGAAGACGGCATACGAGATACAGTGG TGTCTCGTGGGCTCGGAGATGT	ACCACTGT
Ad2.15	CAAGCAGAAGACGGCATACGAGATCAGATC CAGTCTCGTGGGCTCGGAGATGT	TGGATCTG
Ad2.16	CAAGCAGAAGACGGCATACGAGATACAAAC GGGTCTCGTGGGCTCGGAGATGT	CCGTTTGT
Ad2.17	CAAGCAGAAGACGGCATACGAGATACCCAG CAGTCTCGTGGGCTCGGAGATGT	TGCTGGGT
Ad2.18	CAAGCAGAAGACGGCATACGAGATAACCCC TCGTCTCGTGGGCTCGGAGATGT	GAGGGGTT
Ad2.19	CAAGCAGAAGACGGCATACGAGATCCCAAC CTGTCTCGTGGGCTCGGAGATGT	AGGTTGGG
Ad2.20	CAAGCAGAAGACGGCATACGAGATCACCAC ACGTCTCGTGGGCTCGGAGATGT	GTGTGGTG
Ad2.21	CAAGCAGAAGACGGCATACGAGATGAAACC CAGTCTCGTGGGCTCGGAGATGT	TGGGTTTC
Ad2.22	CAAGCAGAAGACGGCATACGAGATTGTGAC CAGTCTCGTGGGCTCGGAGATGT	TGGTCACA

Table 15. PCR protocol used for ATAC-seq library amplification

Step	Temperature	Time	Number of cycles
Polymerase activation	72°C	5:00	1
Denaturing	98°C	0:30	1
Denaturing	98°C	0:10	
Annealing	63°C	0:30	12
Extension	72°C	1:00	

2.2.16 TT-seq data analysis

2.2.16.1 Sequencing data processing

50 bp paired-end reads were obtained for each of the TT-seq samples. Deep sequencing of the TT-seq libraries resulted in 180-300 million read pairs per sample which were mapped to the mm10 assembly version of the mouse genome using STAR 2.5.3 (Dobin & Gingeras 2015). Samtools (Li et al. 2009) was used for quality control and filtering the SAM files. Alignments with an MAPQ smaller than 7 (-q 7) were skipped. Further data analysis was performed using R/Bioconductor environment.

2.2.16.2 Transcription unit annotation and classification

The whole genome was divided into consecutive 200bp bins; the center of the paired-end sequenced fragments was taken to count the sequencing reads into each bin. The replicates for each time point in the time course experiment were merged before annotation. An antisense bias ratio was calculated using the spike-in and this ratio was then used to correct antisense bias for counts. A pseudo-count was added to each bin to overcome noisy signals. The R/Bioconductor package GenoSTAN (Zacher et al. 2017) was used to segment the genome into 'transcribed' and 'untranscribed' states by using a two-state hidden Markov model with a PoissonLog-Normal emission distribution.

2.2.16.3 Defining protein-coding and lincRNA TUs

A minimal of 20% overlap of transcriptional units (TUs) length with a protein-coding gene or a lincRNA annotated in GENCODE as well as an overlap with an exon of the annotated feature, were classified as protein-coding gene and lincRNA, respectively. TUs overlapping with exons of the same protein-coding gene or lincRNA were combined. The remaining TUs were regarded as ncRNA. In order to filter spurious TUs, a minimal expression threshold for TUs was evaluated through Jaccard index based on similarity between TUs and all annotated genes in GENCODE. Next, the precise start and end sites of TUs were refined to nucleotide resolution by fitting a piecewise constant curve to the coverage profiles around initially divided bins using segmentation method from R/Bioconductor package tilingArray (Huber et al. 2006).

2.2.16.4 Defining other TUs

TUs located on the opposite strand within 1kb upstream and 1kb downstream of transcription start sites of protein-coding genes were assigned as upstream antisense RNA (uaRNA) and convergent RNA (conRNA), respectively. TUs located on the opposite strand of annotated protein-coding genes were assigned as antisense_intragenic RNAs. TUs located outside of GENCODE-annotated protein-coding genes and lincRNAs were classified as intergenic RNAs. ncRNAs from antisense_intragenic RNAs and intergenic RNAs were defined as eRNAs as long as there was an overlap with ATAC-Seq peaks (maxgap = 1kb) at one of the time points. All ncRNAs located within 1kb upstream or 1kb downstream of protein-coding genes in the sense direction were discarded in eRNA annotation. Next, eRNAs identified within a 1 kb bin were merged.

2.2.16.5 Differential gene expression analysis

HTSeq-count (Anders et al. 2015) was used to calculate count tables for all TT-seq time points and R/Bioconductor package DESeq2 (Love et al. 2014) was used for differential gene expression analysis. Size factor was evaluated on our set of protein-coding genes. An adjusted P-value of 0.05 was used to identify differentially expressed genes by comparing each time point to the 0 hours of OCT4 loss sample.

2.2.17 ChIP-seq data analysis

2.2.17.1 Data resources

Publicly available paired-end OCT4, SOX2 and NANOG ChIP-seq data at time points 0 hours and 24 hours from the same cell type (ZHBTc4) were downloaded from NCBI Gene Expression Omnibus under accession number GSE87822 (King & Klose 2017). Single-end Mediator (Med1) ChIP-seq data from the same cell type (ZHBTc4) were downloaded under accession number GSM1038259 (Whyte et al. 2013). Single-end ESRRB, KLF4, MYC, p300, H3K4me1, H3K4me3 and H3K27ac ChIP-seq data from ES cell C57BL/6J were obtained from GEO under accession number GSE90895 (Chronis et al. 2017).

2.2.17.2 Mapping and peak calling

Paired/Single-end reads were aligned to the mouse genome assembly version mm10 using Bowtie2.3.4 (Langmead & Salzberg 2012) with the '-local' and '-no-discordant' options. For paired-end reads, only proper pairs (-f99, -f147, -f83 and -f163) with MAPQ bigger than 7 were selected. Peaks were called using the 'callpeak' function of MACS2 (Y. Zhang et al. 2008). Non-uniquely mapping reads and reads mapping to a custom 'blacklist' regions of genome were removed. Metagene plots were generated by R/Bioconductor package metagene (Beauparlant et al. 2018).

2.2.18 ATAC-seq data analysis

2.2.18.1 Mapping and peak calling

Paired-end reads were aligned to mouse genome assembly version mm10 without chromosome M using Bowtie 2.3.4 (Langmead & Salzberg 2012). Peaks were called using the 'callpeak' function from MACS2 with options of '-broad'. Non-uniquely mapping reads and reads mapping to a custom 'blacklist' regions of genome were removed. Peaks at any time point were merged as the final peaks.

2.2.19 Differential binding and k-means cluster analysis

HTSeq-count was used to get count tables for ATAC-seq peaks and DESeq2 was used to call differential accessibility from time point 0 hours, 3 hours, 6 hours, 9 hours, 12 hours and 15 hours. Differential accessibility sites were further classified into early- and late-response peaks through k-means cluster. Briefly, original count tables for two biological replicates were normalized by size factors and then merged by mean. Merged counts for all time points were further transformed by z score and classified into four clusters by k-means function in R.

2.2.20 De novo motif enrichment analysis

DNA sequences +/-500bp around OCT4 ChIP-seq peak summits were extracted through R/Bioconductor package Rsamtools (Li et al. 2009). BaMMmotif (Kiesel et al. 2018) was used for de novo motif enrichment analysis. For the OCTSOX composite motif analysis, OCT4 or SOX2 motif seeds were manually selected and extended around 8 base pairs. In order to evaluate the distance between OCT4 motif and SOX2 motif, FIMO in MEME suite (Grant et al., 2009) was used to scan OCT4 and SOX2 motifs separately. The position information for OCT4 and SOX2 motifs in each query sequence was then extracted and used for distance calculation.

2.2.21 Enhancer-promoter pairing

Enhancers were paired with their target gene based on proximity as well as based on trend of eRNA synthesis and mRNA synthesis. Intragenic eRNAs were paired with the gene they overlapped with. Remaining eRNAs are paired with closest upstream and/or downstream mRNAs.

3. Results

3.1 Characterization of inducible OCT4 loss-of-function cells

The role of OCT4 in embryo development has been studied extensively. OCT4 function has also been investigated in the context of binding to the genome in ESC and iPSC (Marson et al. 2009; Soufi et al. 2012; Chronis et al. 2017; Zhou et al. 2007; Xi Chen et al. 2008; Kim et al. 2008). Precise OCT4 levels are critical for the maintenance of the pluripotency network. Up- and down-regulation leads to loss of pluripotency and differentiation into primitive endoderm and trophectoderm, respectively (Niwa et al. 2000).

The aim of this thesis is to elucidate the role of OCT4 in governing the enhancer landscape in mouse embryonic stem cells (mESC) and to investigate the direct target genes of OCT4. Until now OCT4 activity has been determined largely based on studying its binding to the genome, more specifically to enhancer elements. However, binding per se does not infer activity. More recently it was shown that highly active enhancers show transcriptional activity in the form of enhancer RNAs (eRNAs).

A transgenic mESC line was utilized, in which the *Pou5f1* loci was truncated by insertion of a Hygromycin B/Zeocin resistance cassettes (ZHBTc4, Niwa et al. 2000) (**Figure 1**). Maintenance of pluripotency was established by insertion of a cDNA copy of *Pou5f1* under control of a tetracycline repressible element (**Figure 1**).
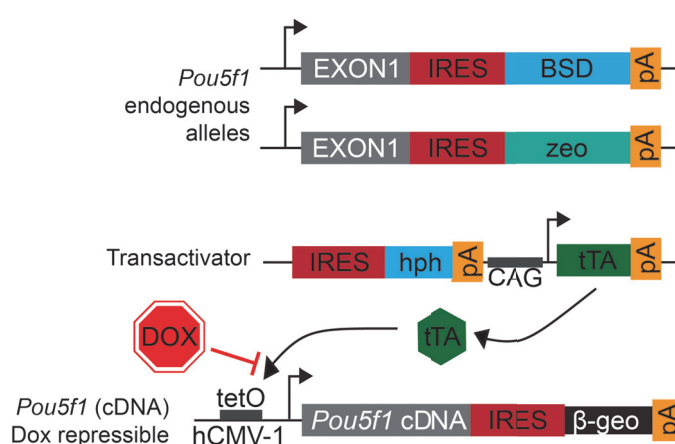

Figure 1 Schematic overview of the genetic alterations of mouse embryonic stem cells.

Pou5f1 alleles were truncated by insertion of IRES-BSD-pA and IRES-Zeo-pA behind exon 1. An OCT4 transgene, which is activated in the absence of tetracycline, was inserted (Image adapted from Niwa et al. 2000).

Addition of doxycycline (DOX) rapidly shuts off *Pou5f1* expression resulting in loss of pluripotency and differentiation to trophectoderm (Niwa et al. 2000).

OCT4 and SOX2 protein levels at 0 hours and 24 hours of DOX treatment were assessed using immunofluorescence (IF) (**Figure 2a**). OCT4 protein was no longer detectable at 24 hours of DOX treatment, whereas SOX2 protein levels appeared unaffected. ZHBTc4 cells were rounding up and started to slightly detach from the tissue culture plate after 24 hours of DOX treatment (**Figure 2a, bright field (BF)**).
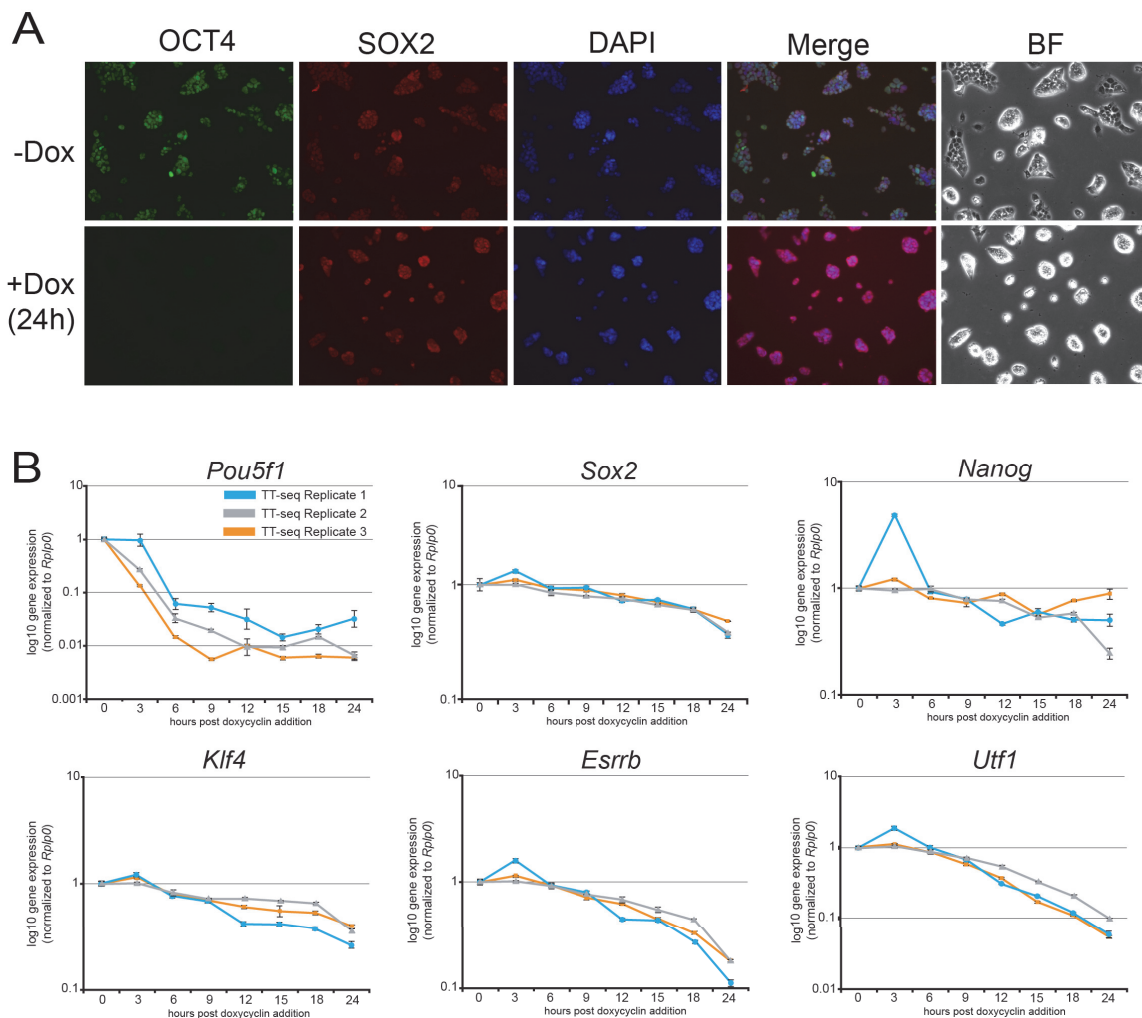

Figure 2 Assessment of protein and RNA expression in ZHBTc4 cells after doxycycline (DOX) treatment.

A: Immunofluorescence of ZHBTc4 cells without DOX and after 24-hour treatment with 1µg/ml DOX. Green: OCT4, red: SOX2, blue: DAPI, BF: Brightfield. **B**: RT-qPCR analysis of gene expression of the pluripotency transcription factors *Pou5f1, Sox2, Nanog, Esrrb, Klf4, Utf1* relative to housekeeping gene *Rplp0* as indicated. n=3 as indicated by blue, orange and grey lines with StDev indicated by technical triplicate.

DOX treatment was performed in a kinetic manner for 0 hours, 3 hours, 6 hours, 9 hours, 12 hours, 15 hours, 18 hours, and 24 hours. At each time point RNA and chromatin were collected to analyze transcriptional changes via RT-qPCR and to address OCT4 occupancy via Western and ChIP-qPCR, respectively. *Pou5f1* expression showed a 10-fold down-regulation after 3 hours of DOX treatment (**Figure 2b**), and at 6 hours *Pou5f1* expression was down-regulated to approximately 100-fold, similar to 24 hours of DOX treatment. *Sox2* and *Nanog* expression seemed to not be affected at early time points of DOX treatment and appeared to be slightly down-regulated after 24 hours of DOX treatment (**Figure 2b**). Support factors such as *Klf4*, *Esrrb* and *Utf1* were affected earlier than *Sox2* and *Nanog*, starting to be down-regulated after 9 to 12 hours of DOX treatment with *Klf4* being less affected than *Esrrb* and *Utf1* (**Figure 2b**).

Global binding of OCT4, NANOG and ESRRB was addressed by separating chromatin material on SDS-Page followed by Western blotting. OCT4 binding was largely lost at 9 hours and completely lost after 12 hours of DOX treatment (**Figure 3a**). NANOG and ESRRB binding was impacted after 9 hours of DOX treatment. However, global binding for ESRRB remained detectable until 18 hours of DOX treatment, whereas NANOG global binding was still detectable at 24 hours of DOX treatment (**Figure 3a**).

Next, Chromatin-Immunoprecipitation followed by quantitative real-time PCR (ChIP-qPCR) was performed for OCT4 and SOX2 to determine the occupancy at specific genomic loci (**Figure 3b**). OCT4 and SOX2 binding was impaired at four enhancer elements (*Pou5f1 CR4, Sox2 SRR2, Nanog* enhancer, *Utf1* enhancer) as early as 9 hours of DOX treatment. OCT4 and SOX2 binding reached background levels at 12 to 15 hours of DOX treatment.

We combined this loss-of-function model with the Transient Transcriptome Sequencing (TT-seq) method. To study global RNA synthesis levels and capture eRNA and other elusive RNA species TT-seq was employed (Schwalb et al. 2016). Integration of other genomic approaches such as ChIP-seq and ATAC-seq enabled us to investigate the OCT4-governed enhancer landscape and assess the direct target genes of OCT4.
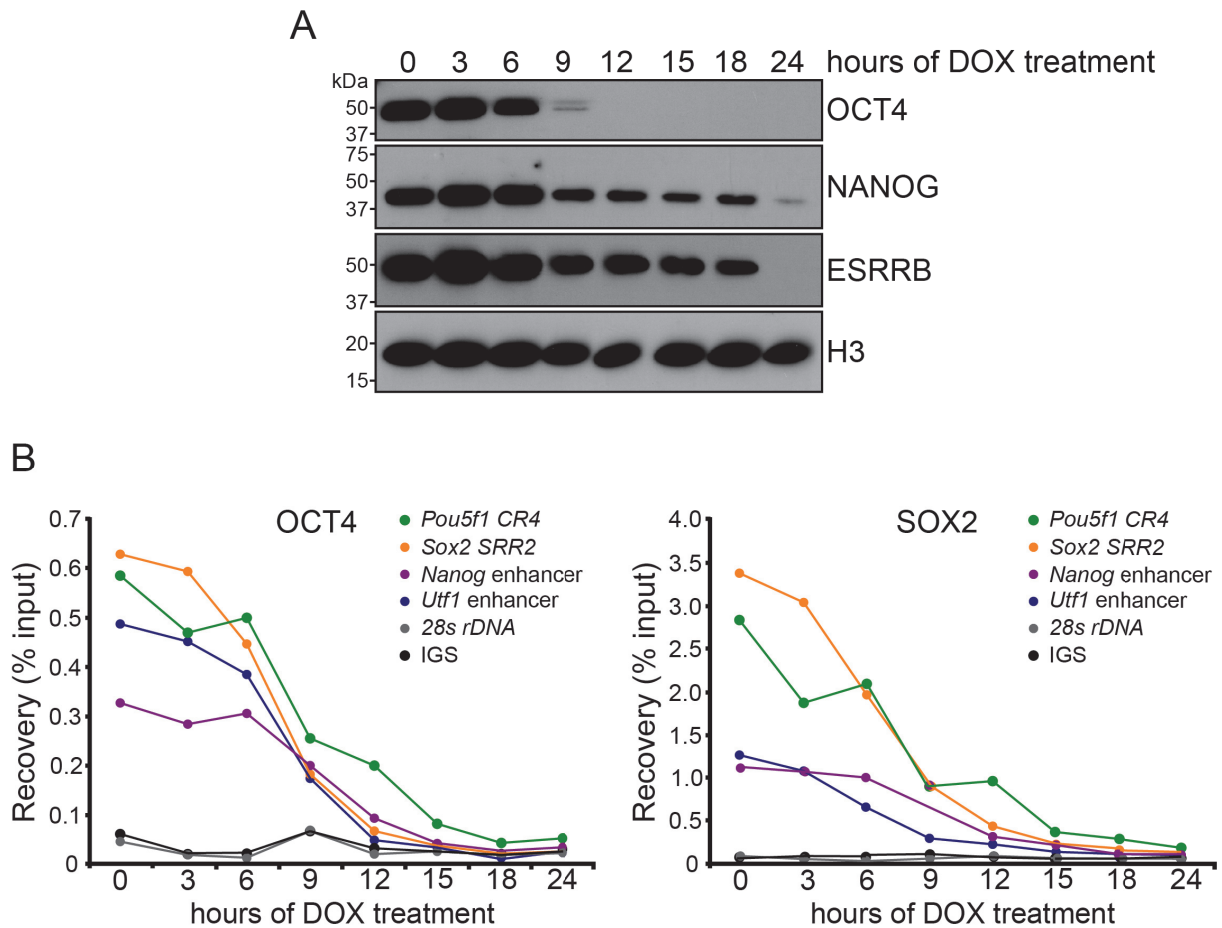
Figure 3 Assessment of TF binding in ZHBTc4 cells in response to DOX treatment.

A: Western blotting analysis of chromatin samples from ZHBTc4 cells over time course of DOX treatment using OCT4, NANOG, ESRRB, H3 (control) antibodies **B**: Occupancy of OCT4 and SOX2 decreased at *Pou5f1 CR4, Sox2 SRR2, Nanog enhancer, Utf1 enhancer, 28s rDNA* (control) and *IGS* (control) over the time course of DOX treatment. Representative sample shown.

3.2 Transient transcriptome sequencing (TT-seq) analysis of ZHBTc4 cells

TT-seq is a method similar to RNA-seq coupled with metabolic labeling with 4-thiouridine (4sU, 4sU-seq) (**Figure 4**). Unlike 4sU-seq, the TT-seq protocol contains a fragmentation step to by-pass the 5' bias of the 4sU-seq. The 5' bias refers to the already transcribed RNA molecules prior to 4sU exposure which dominate the data (Schwalb et al. 2016). Thus, the fragmentation step in TT-seq allows for effective isolation of newly synthesized RNA. ZHBTc4 cells were treated with DOX for 0 hours, 3 hours, 6 hours, 9 hours, 12 hours and 24 hours and were exposed to 4sU at each time point for 5 minutes.
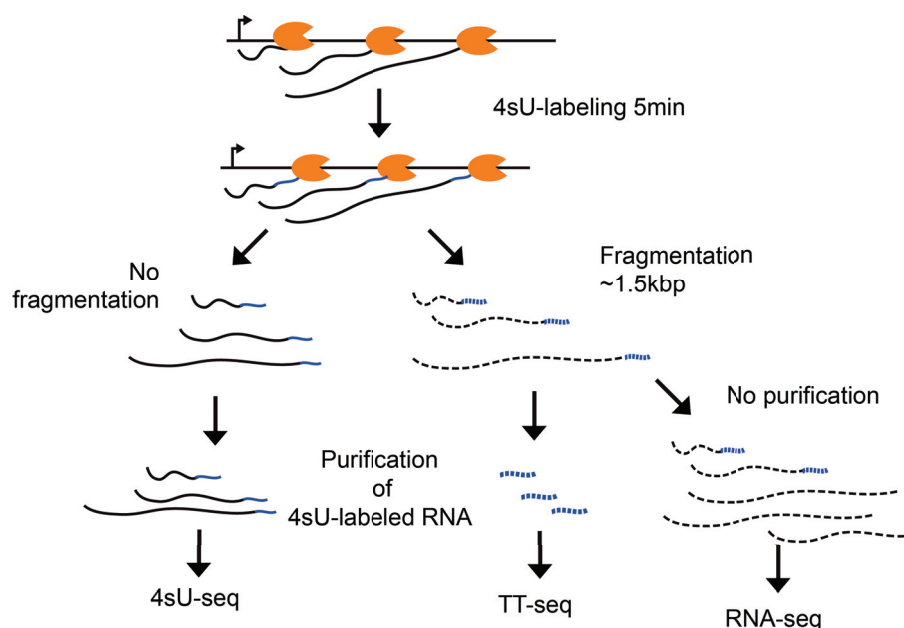
Figure 4 Schematic representation of the 4sU-seq, TT-seq and RNA-seq methods.

TRIzol was used to secure an immediate stop of labeling and preservation of all biomolecules. After RNA isolation and fragmentation, RNA-seq was performed on a small amount of total RNA (**Figure 4**). The remainder of the RNA was used for enrichment of newly synthesized RNA as detected by TT-seq. After library preparation, the samples were pooled and sequenced at low sequencing depth of ~20x10^6 reads per sample (**Table 16**).

A

	TT-seq					
	Replicate 1		Replicate 2		Replicate 3	
	Mapped reads	Duplicate level	Mapped reads	Duplicate level	Mapped reads	Duplicate level
0 hours	21.978.897	14%	24.431.485	42%	24.677.005	35%
3 hours	24.232.519	51%	28.729.212	30%	24.395.194	37%
6 hours	36.386.344	29%	36.767.718	16%	24.238.704	29%
9 hours	24.541.826	40%	36.591.722	21%	25.260.329	22%
12 hours	27.300.130	30%	23.619.214	18%	21.559.283	36%
24 hours	x	x	x	x	24.366.044	25%

B

	RNA-seq			
	Replicate 1		Replicate 2	
	Mapped reads	Duplicate level	Mapped reads	Duplicate level
0 hours	24.752.477	39%	23.817.007	36%
3 hours	25.550.577	38%	21.383.090	52%
6 hours	28.003.937	42%	14.887.283	26%
9 hours	23.321.127	32%	22.401.586	31%
12 hours	25.187.166	32%	22.894.091	35%
24 hours	25.556.903	50%	21.680.791	37%

Table 16 Assessment of quality of shallow sequencing TT-seq and RNA-seq libraries.

A: Number of reads per individual sample of TT-seq, 3 replicates **B**: Percentage of duplicate reads per sample. L: for RNA-seq, 2 replicates.

This shallow sequencing was important for the quality assessment of the sequencing libraries. Duplication levels are an important parameter in RNA-seq, and particularly in TT-seq, as they indicate the complexity of the samples (**Table 16a and b**). Unfortunately, some samples of the three replicates displayed high duplication levels (**Table 16a**), which prevented deep sequencing (~150x10^6 reads per sample) of all three replicates. However, the shallow sequencing data allowed for some preliminary analysis and provided the necessary insight to continue with the experimental setup. At each time point three sequencing data sets were obtained for analysis, except the 24 hours time point where only 1 data set was obtained. Therefore, most of the results only show 0 hours, 3 hours, 6 hours, 9 hours, and 12 hours of treatment. Some analyses include the 1 sample for 24 hours. The TT-seq samples (labeled RNA) correlated well with each other, and the same accounted for the RNA-seq samples (total RNA) (**Figure 5**).
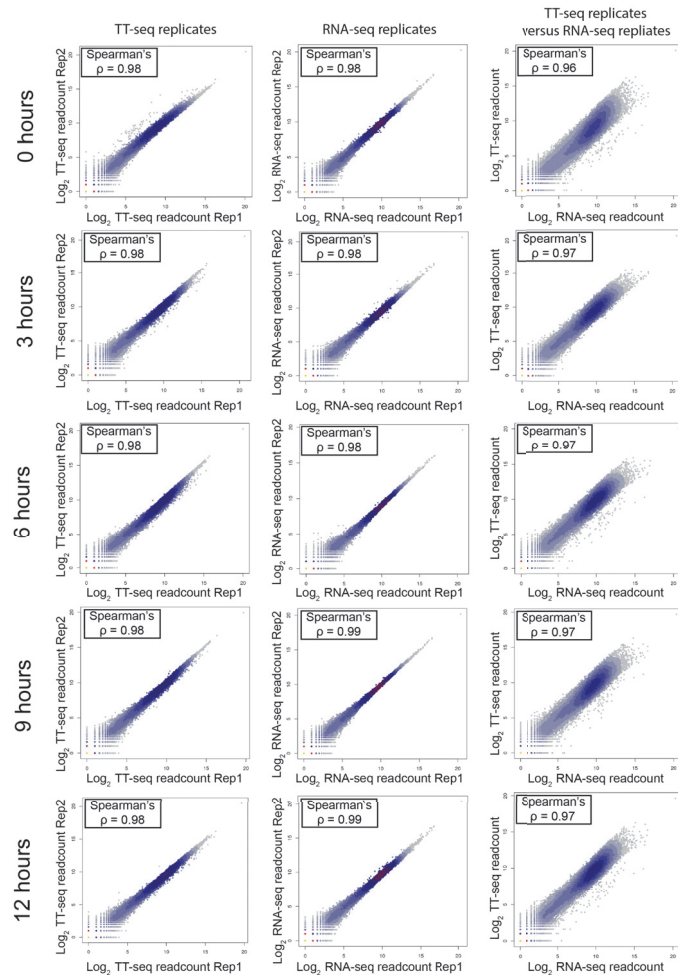

Figure 5 Quality assessment of shallow sequencing TT-seq and RNA-seq libraries.

Scatter plots depicting the Pearson's correlation coefficient for replicates 1 and 2 of TT-seq/RNA-seq and cross-sample analysis for the six time points.

As expected, total RNA and labeled RNA samples correlated less (**Figure 5**) indicating that levels of newly synthesized RNA (monitored by TT-seq) were different from total RNA levels (identified by RNA-seq).

After quality control of the sequencing data the mapped reads were divided into different groups of transcriptional units (TUs) depending on annotation to the reference genome. Transcripts were grouped into protein-coding, long intergenic non-coding RNA (lincRNA), non-coding RNA (ncRNA), antisense RNA, and upstream-antisense RNA (uaRNA). For each transcript class the length was determined (**Figure 6a**). Analysis of the transcript length showed that protein-coding transcripts were generally longer than the other transcript classes. ncRNA, which contains putative eRNA, showed a much smaller size with a wide variability. The total number of TUs within each class over the time course was assessed (**Figure 6b**). There seemed to be little changes to the total number of TUs in the protein-coding class during loss of OCT4. In contrast, ncRNA, antisense RNAs, and uaRNAs showed a more substantial loss of TUs after 24 hours of DOX treatment. Next, protein-coding gene synthesis and lincRNA/ncRNA synthesis were analyzed separately to investigate differential synthesis of the two groups.
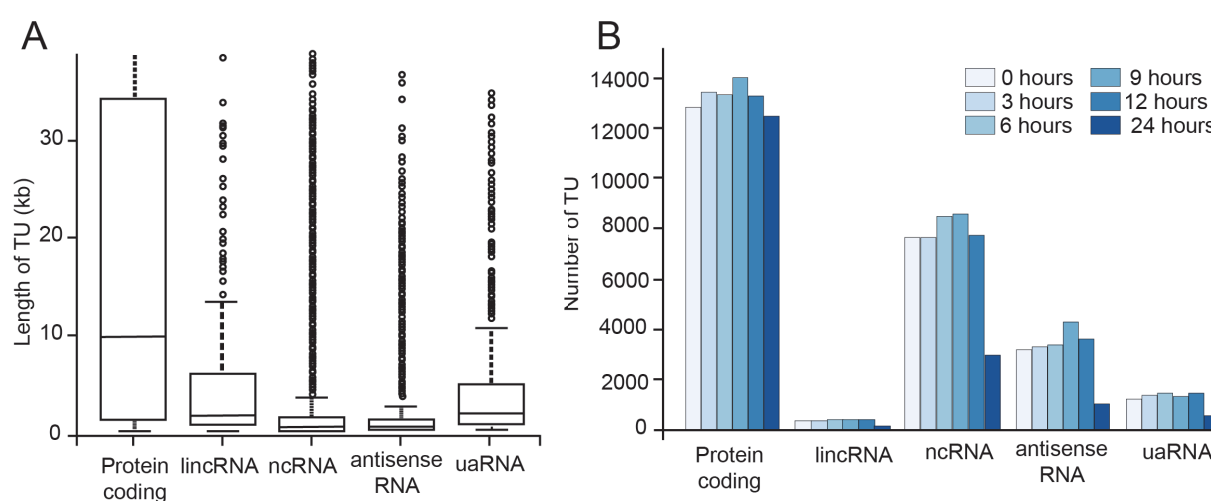

Figure 6 Transcriptional unit (TU) assessment of TT-seq data.

A: Box plot showing median length of TU in the respective TU classes **B**: Total number of TU in each transcript class in ZHBTc4 cells treated with DOX (0, 3, 6, 9, 12, 24 hours).

Differentially expressed genes (DEG) of the protein-coding class were divided into up- and down-regulated genes for both TT-seq and total RNA seq (**Figure 7a**). At early time points of OCT4 loss more down-regulated genes were detected in the labeled RNA samples.

A similar number of up-regulated genes were detected in TT-seq and total RNA-seq data. Principle component analysis of the TT-seq data for 0, 3, 6, 9, and 12 hours of DOX treatment showed a clear separation in PC1, which represents the time course (**Figure 7b**). PC2 of the PCA plot represents sample variability (**Figure 7b**).
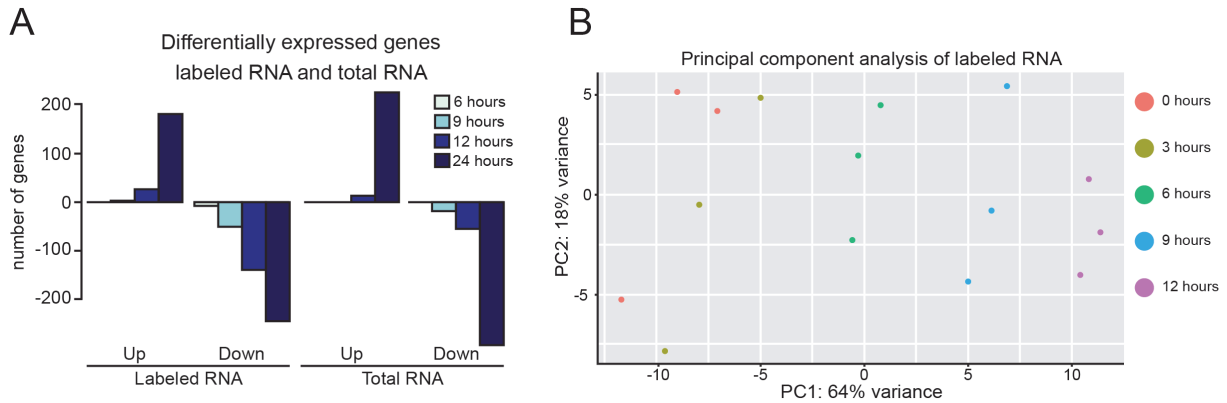

Figure 7 Analysis of differential protein-coding gene expression.

A: Bar chart depicting number of differentially expressed protein-coding genes (up and down) for labeled RNA and total RNA over 6, 9 ,12 and 24 hours of DOX treatment in ZHBTc4 cells (labeled RNA, total RNA) **B**: Principle component analysis of labeled RNA results (0, 3, 6, 9, 12 hours) PC1: Time PC2: Sample variability.

Analysis of differentially synthesized lincRNA/ncRNAs, from now on termed ncRNAs, indicated loss of transcription as early as 3 hours of DOX treatment (**Figure 8a**). Two distinct phases were found when ncRNAs synthesis was lost (**Figure 8b**). Some ncRNAs showed a rapid loss as early as 6 hours, whereas others showed a loss between 12 and 24 hours. Next, publicly available OCT4 ChIP-seq data (Marson et al. 2008) was integrated to assess if OCT4 binds close to genomic loci with differential ncRNAs synthesis. This integrative analysis revealed that most ncRNAs (71%) have their origin close to sites occupied by OCT4 (+/- 1kb) (**Figure 8c**). Next, these differentially synthesized ncRNAs were linked to the closest differentially expressed protein-coding gene (+/- 10kb). We detected 11 up-regulated ncRNAs in proximity to up-regulated protein-coding genes, as well as 19 down-regulated ncRNAs in proximity to down-regulated protein-coding genes (**Figure 8d**). These ncRNAs were termed putative eRNAs. Among the down-regulated putative eRNAs and linked protein-coding genes were *Klf5* and *Etv1*, which are important for maintenance of pluripotency (**Figure 8d**). Some down-regulated putative eRNAs (*Tmem132d, Runxt11, Dlgkb, Dnah8, Dio3* and *Rnf125*) showed a more rapid loss of synthesis than the others.
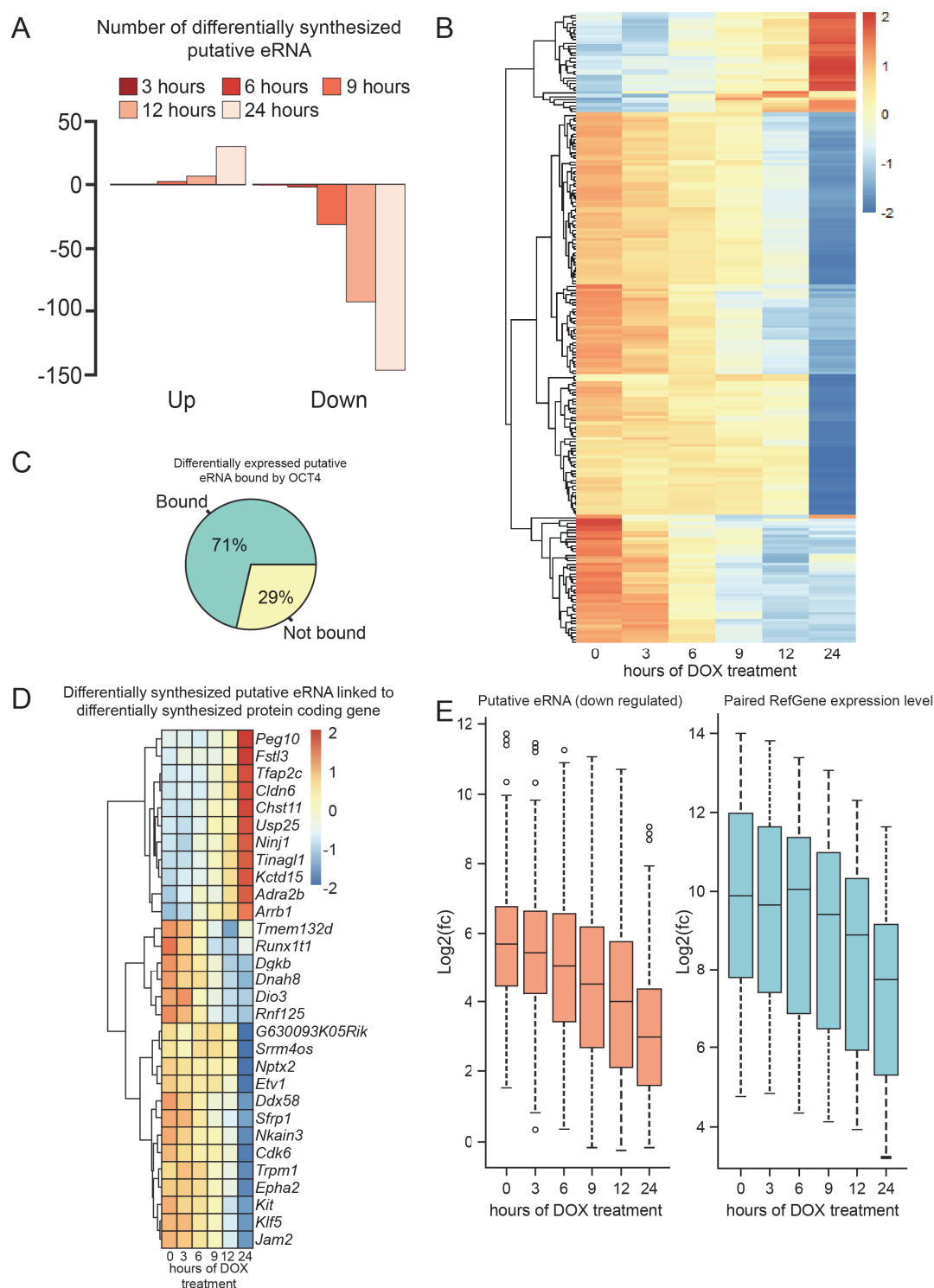
Figure 8 Analysis of differential expression of ncRNA.

A: Bar chart showing differentially expressed ncRNA (up and down) after 6, 9, 12, and 24 hours of DOX treatment of ZHBTc4 cells **B**: Heatmap indicating the global changes in putative eRNA expression **C**: Pie chart depicting the fraction of differentially expressed putative eRNA originating close to sites bound by OCT4 (71%) or not bound by OCT4 (29%) **D**: Heatmap illustrating differentially expressed putative eRNA with close OCT4 binding and their paired differentially expressed protein-coding genes. **E**: Box plots of down regulated putative eRNA bound by OCT4 log2(fc) Salmon: eRNA expression Blue: paired protein-coding gene expression levels after DOX treatment.

Among the up-regulated putative eRNAs and paired protein-coding genes are *Tfap2c* and *Tinagl1*, which are expressed in extra embryonic tissues (**Figure 8d**).

The analysis of the kinetics of down-regulated putative eRNAs and their paired differentially expressed protein-coding genes showed that down-regulation of putative eRNAs precedes the expression changes of the paired protein-coding genes (**Figure 8e**). Putative eRNAs showed decreased expression at 6 to 9 hours, whereas their paired protein-coding genes show a similar decrease at 9 to 12 hours.

Taken together, this analysis documented that the ZHBTc4 cell model paired with TT-seq provides the means to investigate the role of OCT4 in governing the enhancer landscape of mESC and identify direct OCT4 target genes. However, the data quality needed to be improved in order to perform deep sequencing of TT-seq samples. Deep sequencing would likely enable the analysis of lowly synthesized eRNAs/mRNAs occurring upon loss of OCT4.

3.3 Second attempt of TT-seq

We repeated the time course experiment as described above, but added two additional time points. Due to the marked differences between 12 and 24 hours, we also retrieved samples after 15 and 18 hours of DOX treatment (**Figure 6b, 8d-e).**

Quality assessment again was performed utilizing shallow sequencing. Importantly, we found less duplication levels indicating that good quality samples were obtained, even though replicate 1 showed higher duplication levels than replicate 2 (**Table 17a**). Duplication levels of the total RNA-seq data were similar across the various time points and replicates (**Table 17b**).

A

	TT-seq			
	Replicate 1		Replicate 2	
	Mapped reads	Duplicate level	Mapped reads	Duplicate level
0 hours	18.262.931	30%	19.793.637	11%
3 hours	21.110.726	42%	17.046.163	12%
6 hours	16.154.325	47%	21.922.193	17%
9 hours	18.975.631	65%	20.022.925	18%
12 hours	15.617.084	43%	18.649.205	27%
15 hours	19.467.698	49%	17.020.089	13%
18 hours	24.798.739	48%	21.482.764	14%
24 hours	21.286.768	35%	22.099.459	12%

B

	RNA-seq			
	Replicate 1		Replicate 2	
	Mapped reads	Duplicate level	Mapped reads	Duplicate level
0 hours	68.448.897	55%	19.273.593	23%
3 hours	15.153.172	24%	17.131.770	22%
6 hours	15.376.629	38%	18.831.409	24%
9 hours	15.450.102	25%	21.464.008	29%
12 hours	18.355.876	28%	18.542.845	32%
15 hours	12.689.214	46%	20.135.458	28%
18 hours	13.509.973	25%	24.241.660	32%
24 hours	17.234.696	25%	23.810.360	37%

Table 17 Evaluation of the second round of shallow sequencing.

A: Representation of number of mapped sequencing reads per sample and replicate. Percentage of duplicate reads per sample and replicate for shallow sequencing of TT-seq **B**: Representation of number of mapped sequencing reads per sample and replicate. Percentage of duplicate reads per sample and replicate for shallow sequencing of total RNA seq.

We found good correlation between the TT-seq replicates (**Figure 9a**) and between the total RNA-seq samples (**Figure 9b**). Other quality assessments include intron enrichment and enrichment of 4sU incorporated RNA transcripts. TT-seq showed an intron enrichment of about 0.7 whereas RNA-seq only show an intron enrichment of about 0.2 (**Figure 9c**) illustrating that TT-seq and RNA-seq contain pre-mature and mature RNA species, respectively. A clear enrichment of 4sU labeled transcripts compared to unlabeled transcripts was observed in the TT-seq data (**Figure 9d**), but not in the total RNA-seq data (**Figure 9d**) indicating good enrichment of newly synthesized transcripts in the TT-seq approach.
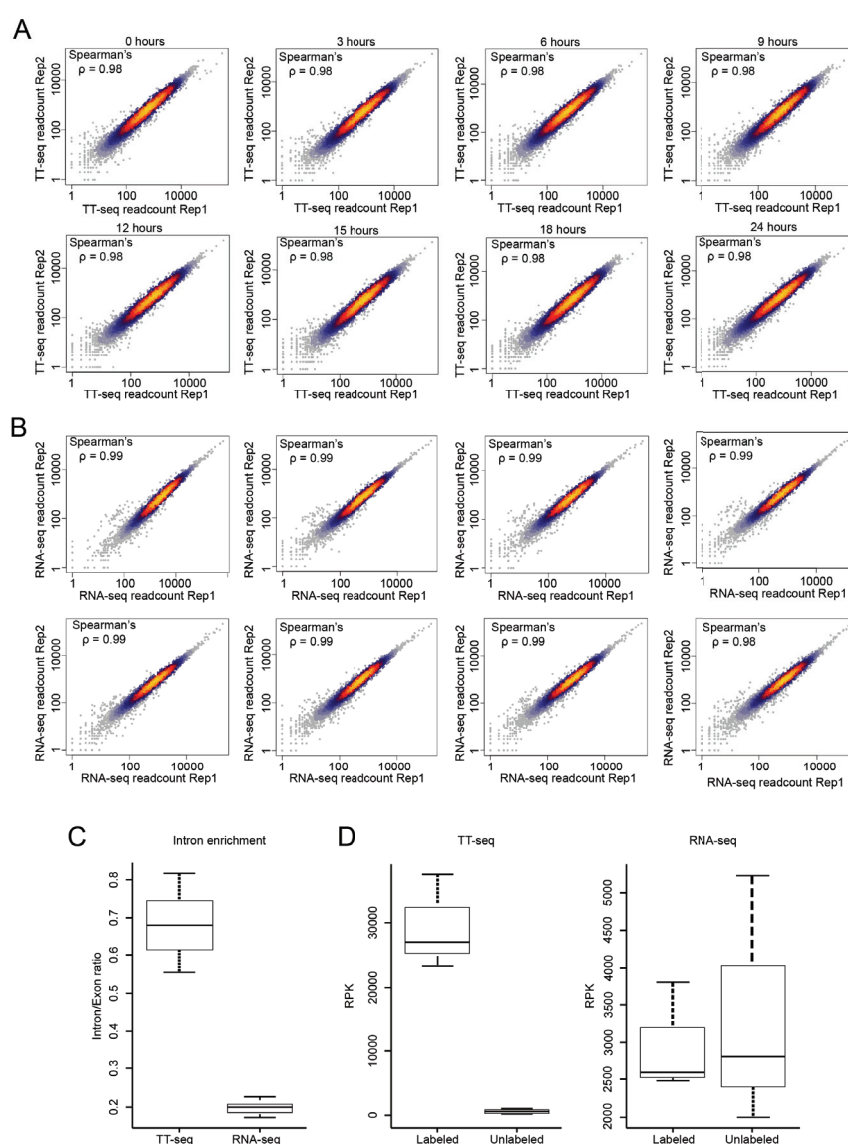

Figure 9 Evaluation of the second round of shallow sequencing.

A: Scatter plot depicting the Pearson's correlation coefficient for replicate 1 and 2 of the second TT-seq shallow sequencing experiment **B**: Scatter plot depicting the Pearson's correlation coefficient for replicate 1 and 2 of the second RNA-seq shallow sequencing experiment **C**: Enrichment of intronic reads in TT-seq versus total RNA-seq represented as intron/exon ratio **D**: Enrichment of labeled RNA

in TT-seq represented as reads per kilobase (RPK) **E**: Enrichment of labeled RNA in total RNA-seq represented as RPK.

The protein-coding genes were analyzed for differential expression. Principle component analysis found a clear separation of TT-seq and RNA-seq samples (**Figure 10a**). The PCA plot also showed clear separation across the various time points and minor variability between the samples of the replicates. Importantly, TT-seq and RNA-seq show the same overall trend regarding the kinetics of OCT4 loss. DEG are found as early as 6 hours of DOX treatment of the ZHBTc4 cells for both up- and down-regulated genes in the TT-seq experiment (**Figure 10b).** Most DEG overlap between time points, with the biggest number of DEG at 24 hours of DOX treatment (**Figure 10c**).
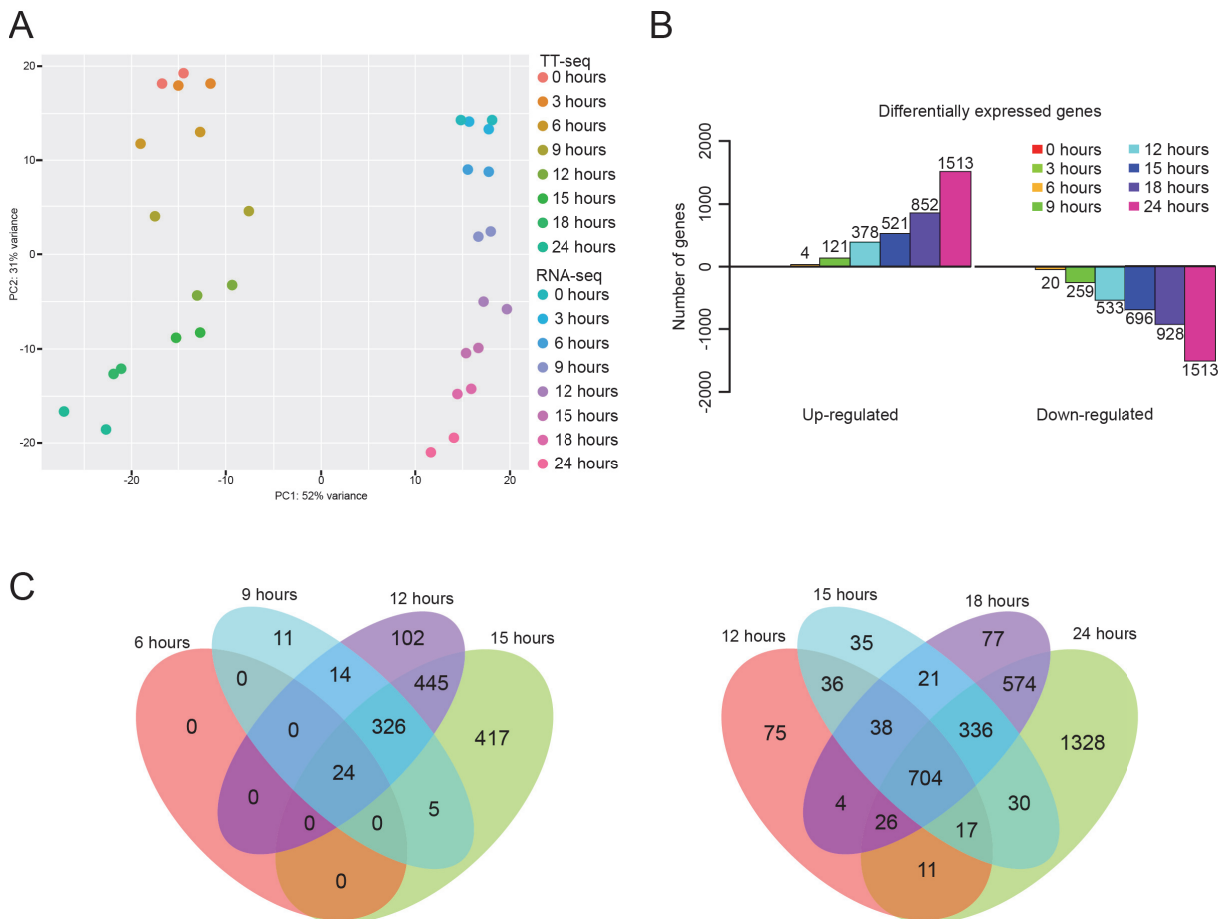

Figure 10 Differential gene expression of protein-coding genes in shallow sequencing data.

A: Principle component analysis of TT-seq and RNA-seq samples. PC1: Sample variability, PC2: Time **B**: Bar chart representing up- and down regulated DEG in TT-seq data across the time course of DOX treatment in ZHBTc4 cells **C**: Venn diagram depicting overlap of up- and downregulated DEG in TT-seq data as defined in (**B**); 6 hours through 15 hours (left) and 12 hours through 24 hours (right).
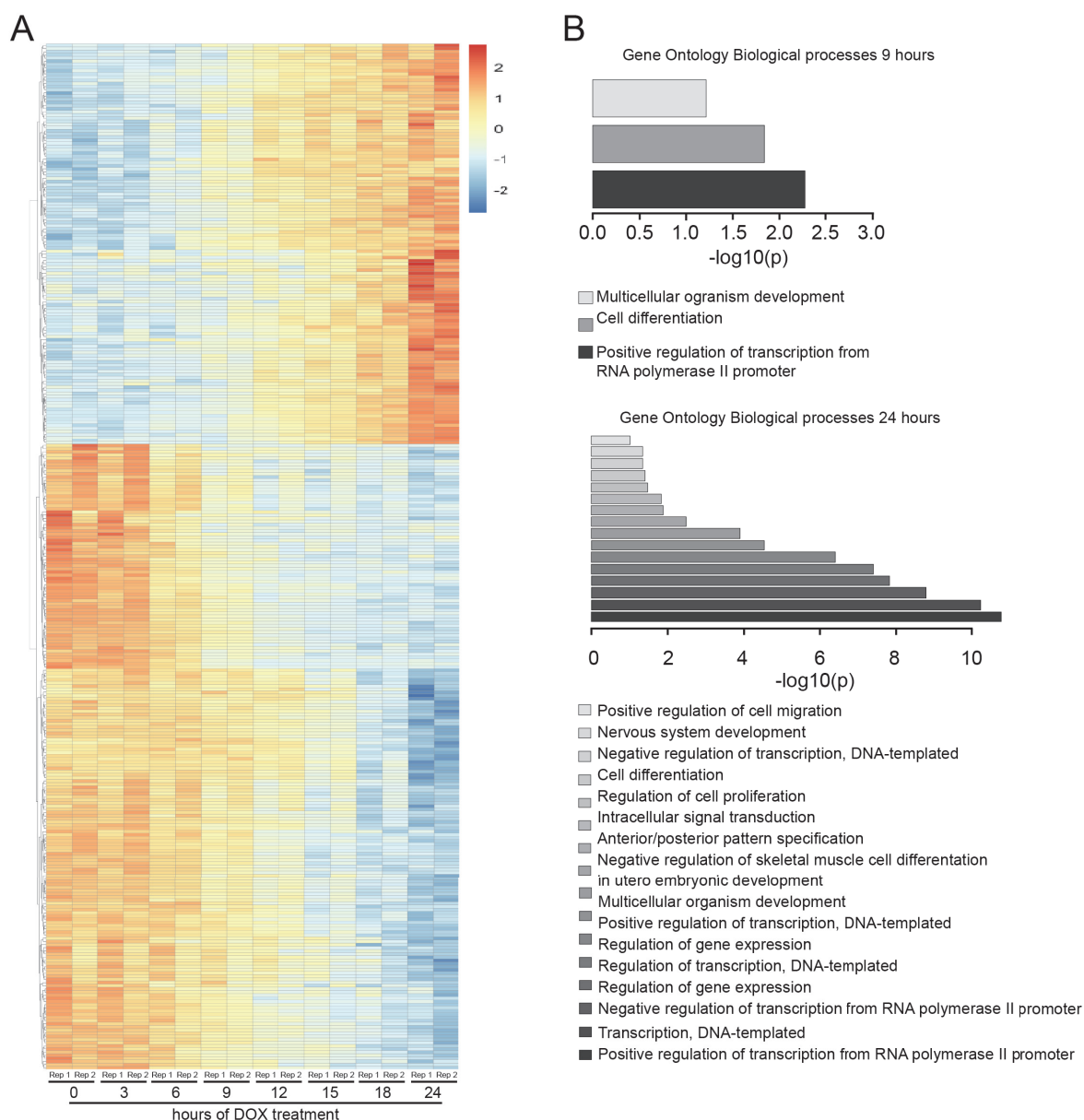
A

B

Gene Ontology Biological processes 9 hours

-log10(p)

☐ Multicellular ogranism development
☐ Cell differentiation
■ Positive regulation of transcription from RNA polymerase II promoter

Gene Ontology Biological processes 24 hours

-log10(p)

☐ Positive regulation of cell migration
☐ Nervous system development
☐ Negative regulation of transcription, DNA-templated
☐ Cell differentiation
☐ Regulation of cell proliferation
☐ Intracellular signal transduction
☐ Anterior/posterior pattern specification
☐ Negative regulation of skeletal muscle cell differentiation
☐ in utero embryonic development
☐ Multicellular organism development
☐ Positive regulation of transcription, DNA-templated
☐ Regulation of gene expression
☐ Regulation of transcription, DNA-templated
☐ Regulation of gene expression
☐ Negative regulation of transcription from RNA polymerase II promoter
☐ Transcription, DNA-templated
■ Positive regulation of transcription from RNA polymerase II promoter

hours of DOX treatment

Figure 11 Heatmap and Gene Ontology analysis of differentially expressed protein-coding genes in shallow TT-seq data.

A: Heatmap indicating the kinetics of up- and down-regulated DEG. Replicates separately shown **B**: Gene Ontology analysis of biological processes for down-regulated protein-coding genes at 9 hours and 24 hours of DOX treatment.

Depiction of DEG in a heatmap illustrated a relatively homogenous cluster for the up-regulated genes whereas the down-regulated genes were divided into two clusters, fast and slow responding genes (**Figure 11a**). Among the up-regulated genes were *Klf6, Gata3, Igf2, Tinagl1*, which are all involved in differentiation of mESC to trophectoderm. Among the down-regulated genes were *Utf1, Klf4, Klf2, Klf5, Esrrb, c-Myc, Foxd3, Tfcp2l1, Sall3, Etv1*, which are all involved in maintaining pluripotency. Gene ontology analysis of the down-regulated DEG showed biological processes

79

involved in differentiation, positive regulation of RNAPII transcription and multicellular organism development at 9 hours of DOX treatment (**Figure 11b**). At 24 hours of DOX treatment most of the biological processes indicated transcription-related processes and embryo development (**Figure 11c**). Taken together these findings indicated that the second TT-seq and RNA-seq data sets were sufficient in quality to proceed with deep sequencing of the TT-seq libraries. As OCT4 binding reached background levels around 12 hours (**Figure 3c**), the libraries from the following time points were utilized for deep sequencing: 0 hours, 3 hours, 6 hours, 9 hours, 12 hours, and 15 hours. The 15-hour time point was included in order to capture the transcriptional changes after OCT4 binding is no longer detectable.

3.4 Analysis of Deep sequenced TT-seq samples

Deep sequencing of the TT-seq libraries was performed at a sequencing depth of about 150×10^6 reads, which allowed for the analysis of sparsely synthesized or rapidly degraded transcripts such as eRNA.

Quality assessment of the deep sequencing data revealed that the first replicate had high duplication levels, whereas the second replicate had acceptable duplication levels (**Table 18**). Despite the high levels of duplication in replicate 1 the two replicates correlated well (**Figure 12**).

	TT-seq			
	Replicate 1		Replicate 2	
	Mapped reads	Duplicate level	Mapped reads	Duplicate level
0 hours	147.578.348	79%	144.764.664	34%
3 hours	156.943.580	89%	155.579.194	73%
6 hours	154.474.743	89%	149.958.010	50%
9 hours	166.600.978	85%	153.486.986	45%
12 hours	167.809.848	90%	142.027.970	55%
15 hours	153.958.756	93%	150.290.914	59%

Table 18 Summary of mapped reads and duplication levels in deep sequenced TT-seq libraries. Representation of number of mapped sequencing reads per sample and replicate. Percentage of duplicate reads per sample and replicate for deep-sequenced TT-seq samples.
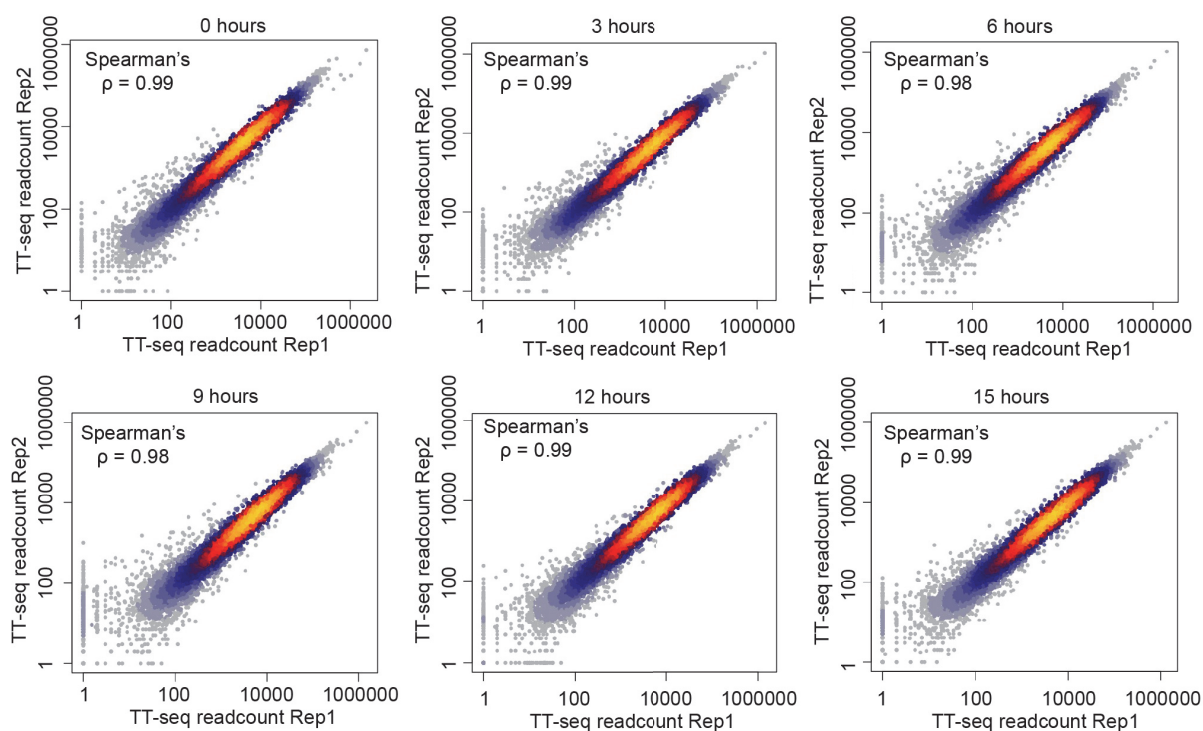
Figure 12 Correlation of TT-seq replicates.

Scatter plot depicting the Pearson's correlation coefficient for replicate 1 and 2 of TT-seq deep sequencing data at the six time points (0 hours, 3 hours, 6 hours, 9 hours, 12 hours and 15 hours).

TUs were annotated using bi-directional Hidden Markov Models (bdHMM) and filtered by Jaccard-index. Jaccard-index filtering was utilized to remove any false positive TUs, which is important for reliable downstream analysis of the TT-seq deep sequencing data (**Figure 13**).
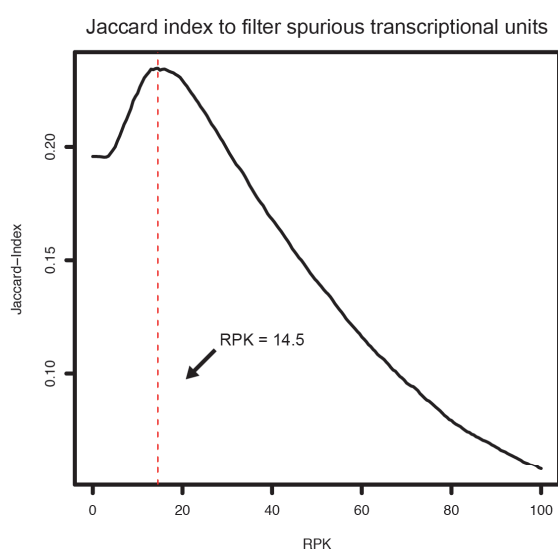
Figure 13 Filtering of TT-seq data to remove spurious TUs.

Jaccard index (y-axis) compared to annotation to GENCODE reference genome for varying RPK (x-axis).
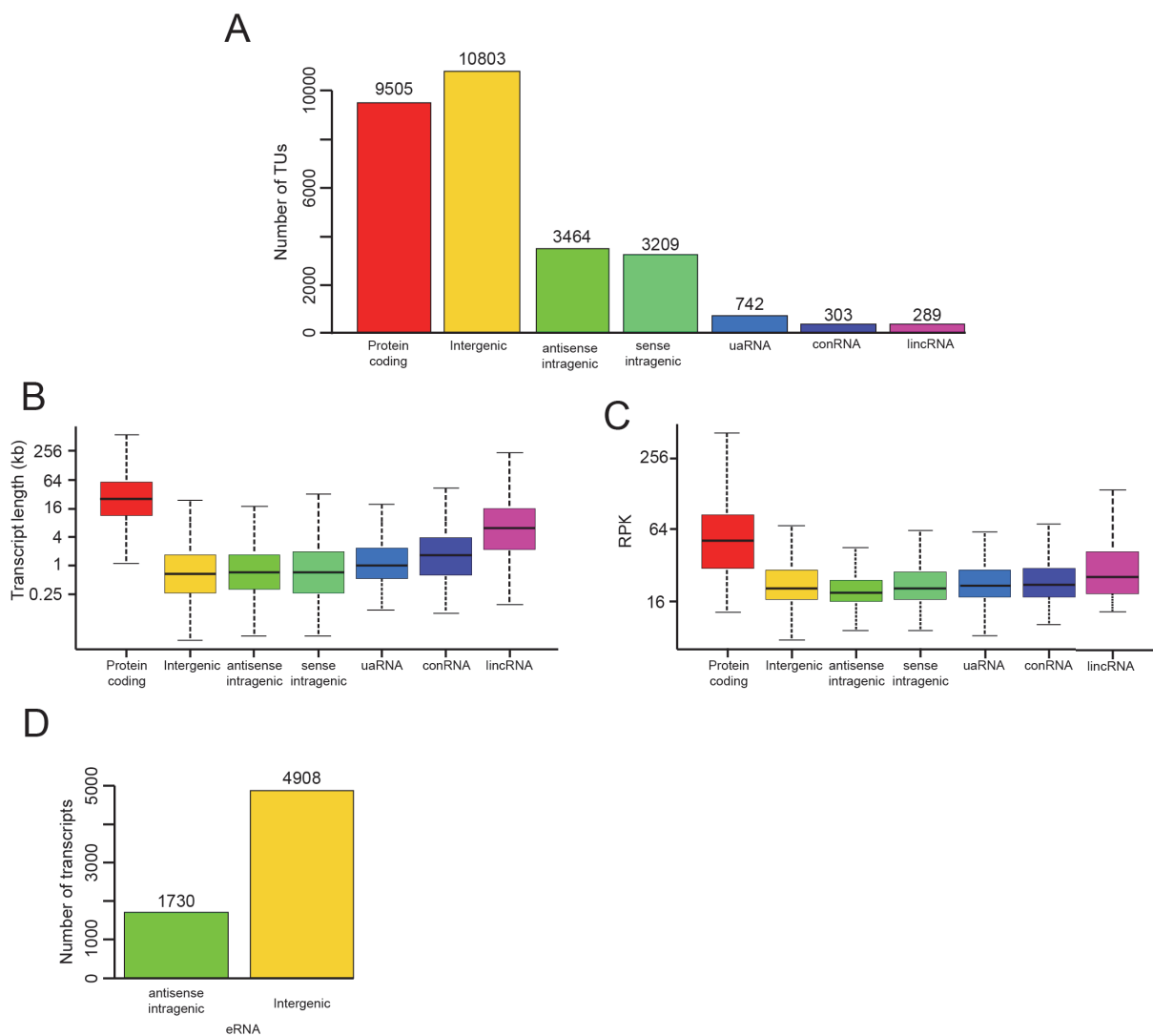
A

B

C

D

Figure 14 Classification of TUs into separate classes based on annotation to GENCODE reference genome.

A: Bar chart depicting number of TUs per TU class **B**: Boxplots depicting length distribution for the different TU classes **C**: Boxplots depicting expression levels for the various TU classes **D**: Bar chart indicating number of eRNA found in antisense, intragenic and intergenic TU classes.

The spurious TUs were removed and the remaining TUs were segregated into various TU classes, namely intergenic, protein-coding, antisense intragenic, antisense intergenic, uaRNA, convergent RNA (conRNA), and lincRNA. The number of TUs for each class was determined (**Figure 14a**). Intergenic and protein-coding TUs made up the majority of the sequencing data with 10803 and 9505 TUs, respectively. We found 3464 antisense intragenic TUs, 3209 sense intragenic TUs, 742 uaRNA, 303 conRNA and 289 lincRNA. The median length for each class was determined, showing the expected trend that protein-coding TUs have a larger median length and lincRNA were generally longer than most of the other ncRNA TU classes (**Figure 14b**). Protein-coding TUs displayed the highest expression levels;

the various ncRNA classes showed similar expression/synthesis levels with lincRNA demonstrating slightly higher expression levels (**Figure 14c**). Putative eRNA were annotated in the antisense intragenic cluster (1730 TU) and the intergenic cluster (4908 TU) (**Figure 14d**). The pluripotency genes *Nanog, Pou5f1, Esrrb, Klf4, Sox2* displayed some of the highest expression levels at 0 hours (**Figure 15a**). Their synthesis rates were affected after 15 hours of DOX treatment (**Figure 15b**).
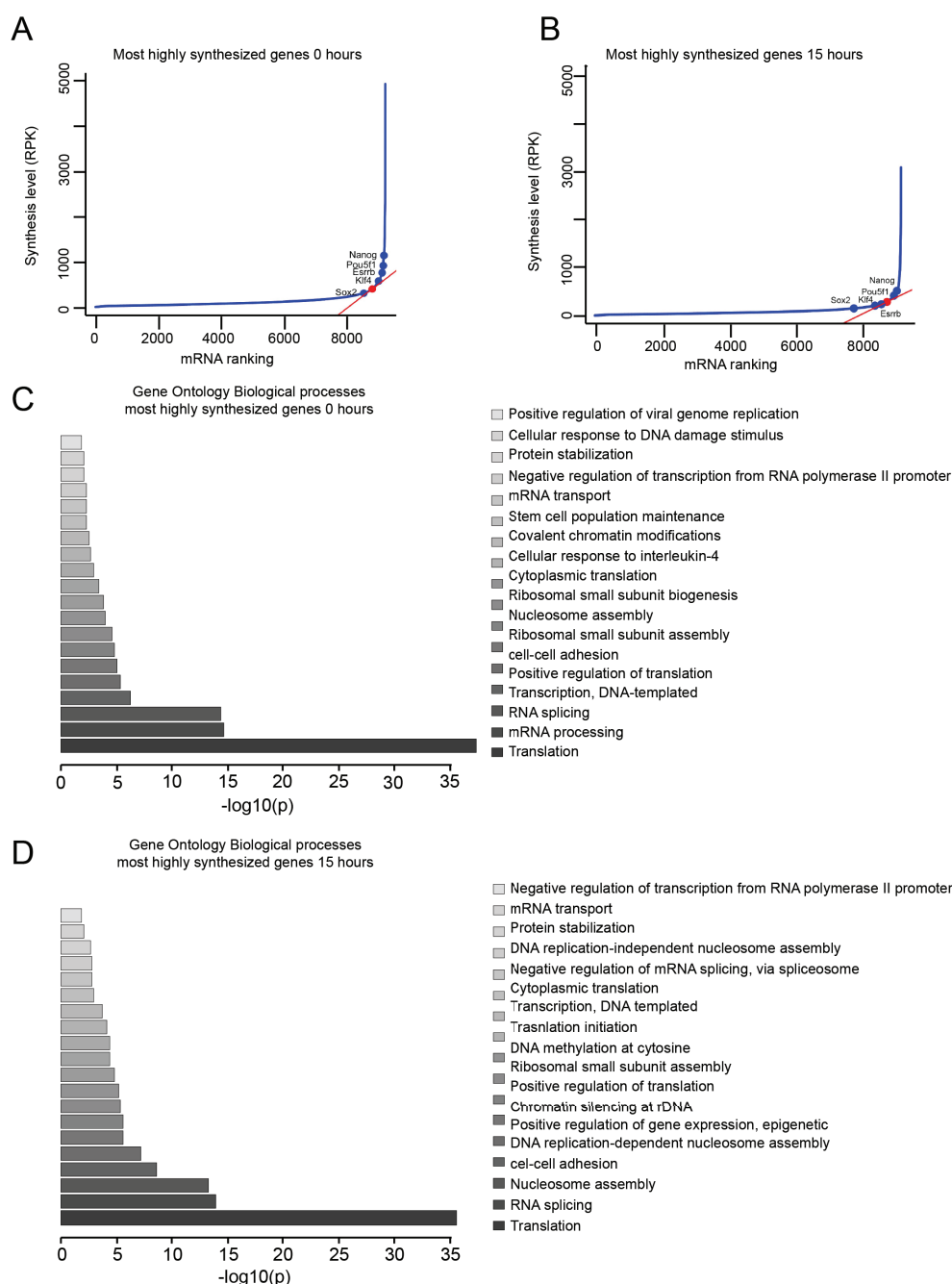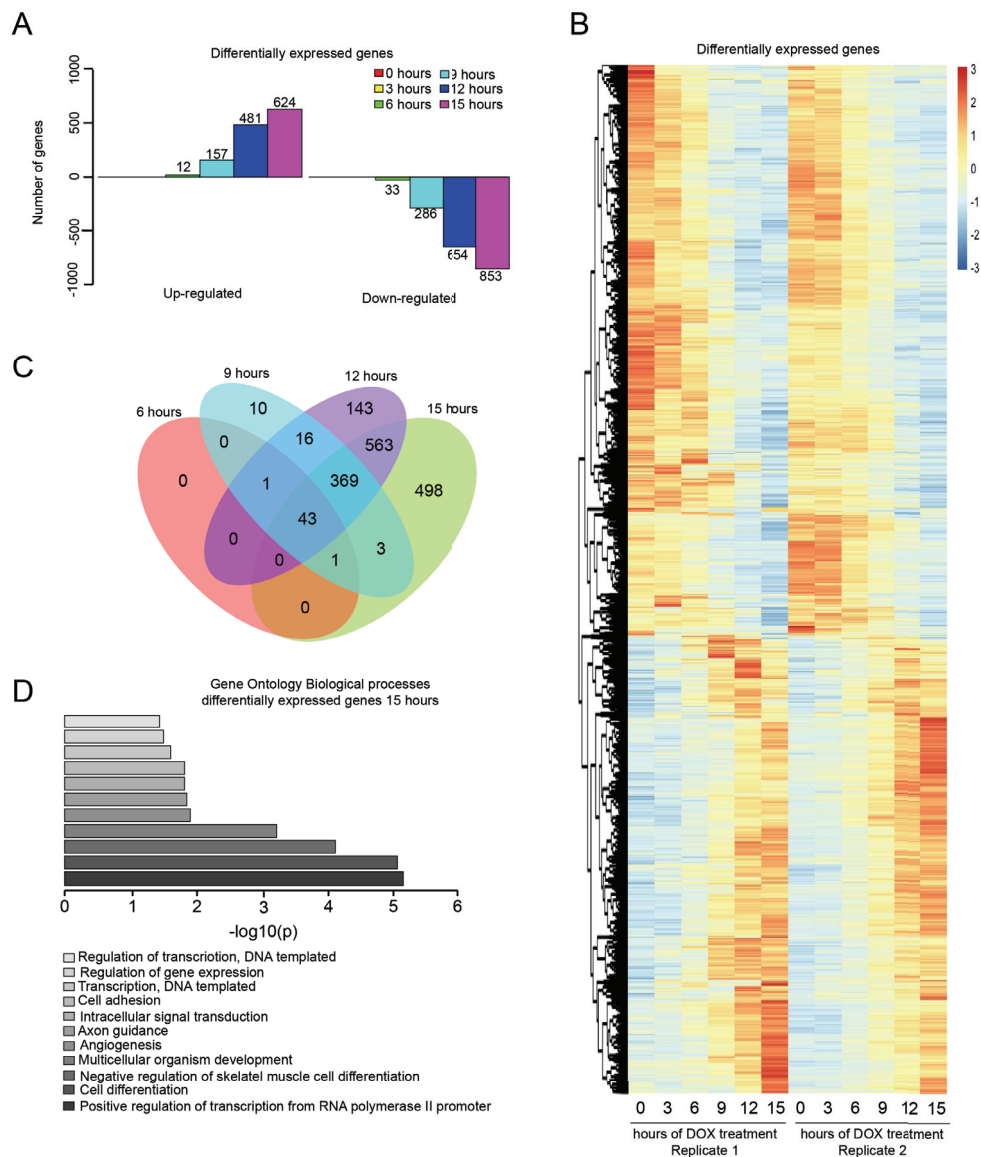

Figure 15 Analysis of protein-coding genes with the highest synthesis levels at 0 and 15 hours of DOX treatment.

A: Ranking of most highly synthesized protein-coding genes at 0 hours of DOX treatment, pluripotency genes are indicated **B**: Ranking of most highly synthesized protein-coding genes at 15 hours of DOX

treatment, pluripotency genes are indicated **C**: GO analysis of biological processes for most highly synthesized protein-coding genes at 0 hours of DOX treatment **D**: GO analysis of biological processes for most highly synthesized protein-coding genes at 15 hours of DOX treatment.

GO analysis of the highest transcribed genes at 0 and 15 hours of DOX treatment showed enrichment of similar biological processes (**Figure 15c and d**). The GO term 'stem cell population maintenance' was enriched at 0, but not at 15 hours. The GO term 'DNA methylation of cytosine' was enriched at 15 hours of DOX treatment, but not at 0 hours. Next, the protein-coding TUs were examined more closely. DEG analysis documented 853 down- and 624 up-regulated genes after 15 hours of DOX treatment (**Figure 16a**).

(Figure legend on next page)

Figure 16 Differential expression analysis of protein-coding genes and corresponding GO terms.

A: Bar chart depicting the up- and down-regulated protein-coding genes **B**: Heatmap indicating the kinetics of up- and down-regulated protein-coding genes. Columns indicate replicates at different time points. Lines display DEG. **C**: Venn diagram indicating the overlap of up- and down-regulated protein-coding DEG as defined in (**A**) **D**: GO analysis of biological processes for down-regulated protein-coding genes at 15 hours of DOX treatment.
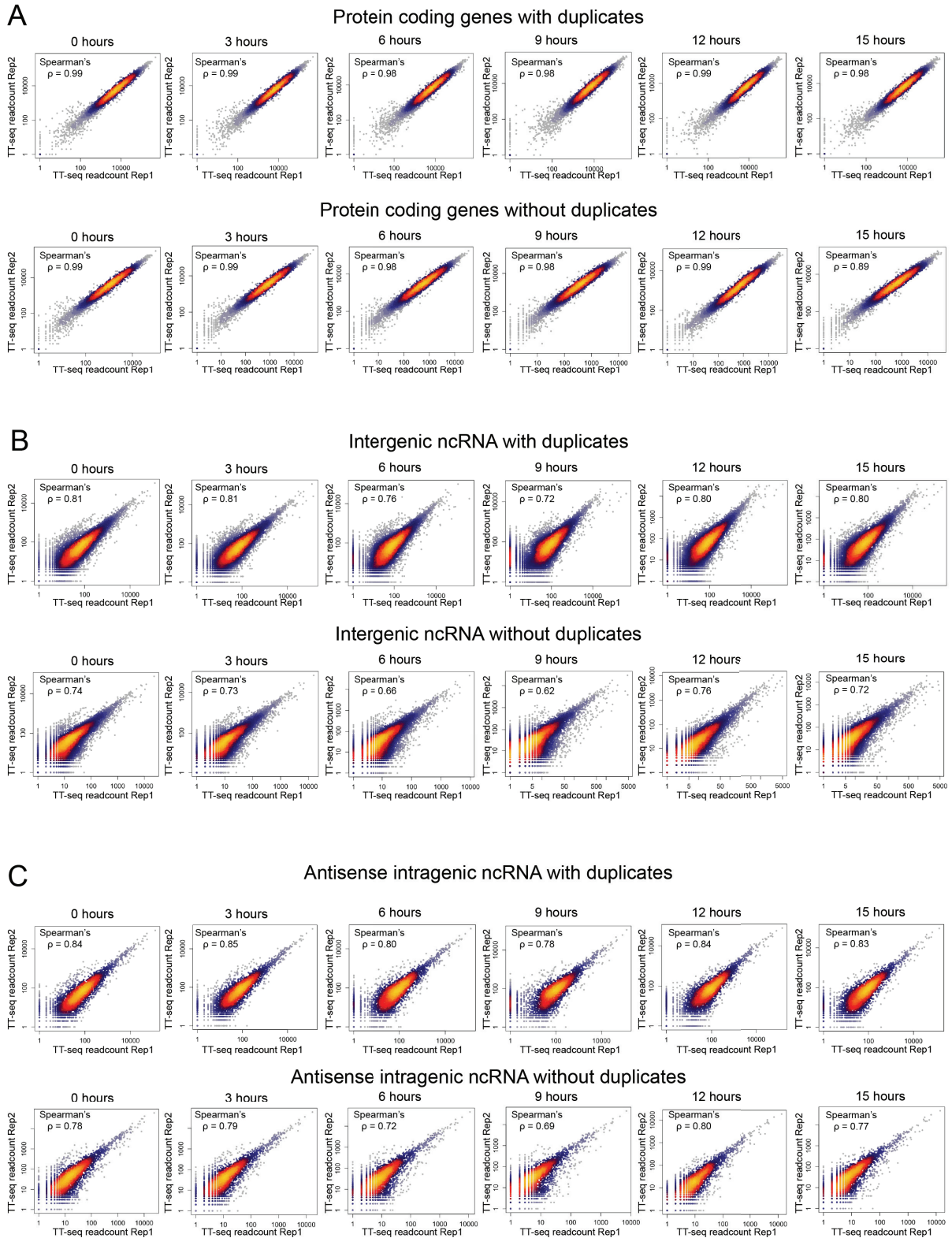
The kinetics of up- and down-regulated genes showed little variation between the replicates (**Figure 16b**). The down-regulated genes included *Utf1, Esrrb, Klf2, Klf4, Klf5, Foxd3, Tbx3,* and *Tfcp2l1*. Genes such as *Tinagl1, Igf2, Klf6, Gata3, Tead4* were included in the up-regulated genes (**Figure 16b**). The strongest up- and down-regulation was documented between 6 hours and 9 hours of DOX treatment. The Venn-diagram indicates the overlap of the TUs between the time points that showed differential gene expression (6 hours, 9 hours, 12 hours, 15 hours) (**Figure 16c**). The differentially expressed protein-coding genes at 15 hours of DOX treatment were analyzed for gene ontology of biological processes. Enrichment for processes such as transcriptional regulation, covalent chromatin modifications, multicellular organism development, cell differentiation, and regulation from RNAPII promoter was identified (**Figure 16d**).

3.5 TT-seq data analysis: the duplicate level challenge

Duplicate levels can interfere with proper analysis of the sequencing data and it needed to be determined whether the duplicates had to be removed. Correlation between the replicates was determined for the different TU classes with and without removal of the duplicates. The protein-coding TUs did not show a loss of correlation when the duplicates are removed (**Figure 17a**). Intergenic ncRNAs, antisense intragenic ncRNAs and putative eRNAs displayed a loss of correlation between the replicates after removal of the duplicates (**Figure 17b, c and e**). Correlation between samples was less affected for the sense intragenic ncRNAs (**Figure 17d**). Taken together these results prompted us to not remove the duplicates for most of the analysis. However, for metagene analysis and annotation of various TU classes the duplicates were removed as they cause reliability issues. Acquiring other genomic data sets such as OCT4 and SOX2 ChIP-seq as well as ATAC-seq will help to reliably analyze the TT-seq data with or without the duplicates. ChIP-seq will help to

determine whether an eRNA originates close to an OCT4 and/or SOX2 occupied region and has therefore relevance to our study. ATAC-seq provides an insight into the accessibility to the chromatin which adds further confidence to eRNA analysis as active enhancers generally display chromatin accessibility.
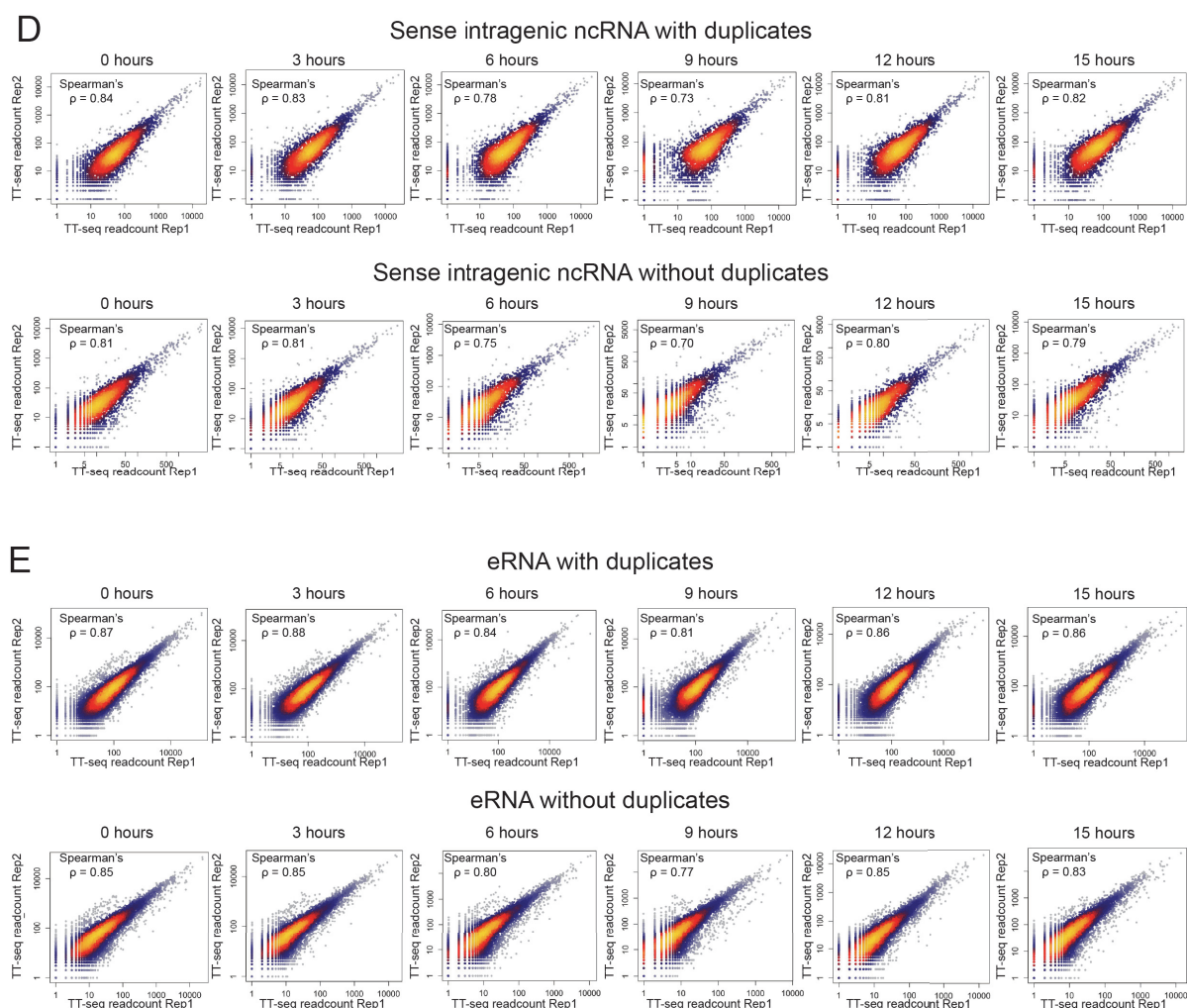
D Sense intragenic ncRNA with duplicates

Sense intragenic ncRNA without duplicates

E eRNA with duplicates

eRNA without duplicates

Figure 17 Correlation assessment of TU classes with and without duplicate reads.

Scatter plots depicting the Pearson's correlation coefficient for TT-seq deep sequencing replicate 1 and 2 in different TU classes: (**A**) protein-coding genes with duplicates (**B**) protein-coding genes without duplicates (**C**) intergenic ncRNA with duplicates (**D**) intergenic ncRNA without duplicates (**E**) antisense intragenic ncRNA with duplicates (**F**) antisense intragenic ncRNA without duplicates (**G**) sense intragenic ncRNA with duplicates (**H**) sense intragenic ncRNA without duplicates (**I**) eRNA with duplicates (**J**) eRNA without duplicates.

3.6 OCT4 & SOX2 ChIP-seq

Next, we employed OCT4- and SOX2-ChIP-seq to determine whether the DEG were direct OCT4 target genes. First, the sonication time for chromatin fragmentation was optimized as e.g. over-fragmentation can lead to increased background levels. Sufficient enrichment for DNA fragments ranging from 200-400bp was achieved after 40 minutes of sonication (**Figure 18a**). Next, ChIP-qPCR was performed to determine the levels of OCT4 binding at the *Pou5f1* distal enhancer (CR4) and *Sox2* SRR2 enhancer; *28s rDNA* served as a control region. Good absolute recovery

represented as percentage of input was obtained after 45 minutes of sonication (**Figure 18b**). Relative enrichment over the *28s rDNA* region also indicated consistent OCT4 binding after 45 minutes of sonication (**Figure 18c**). However, due to the optimal fragment sizes as well as decent absolute recovery and relative enrichment we preceded with 75 minutes of sonication.
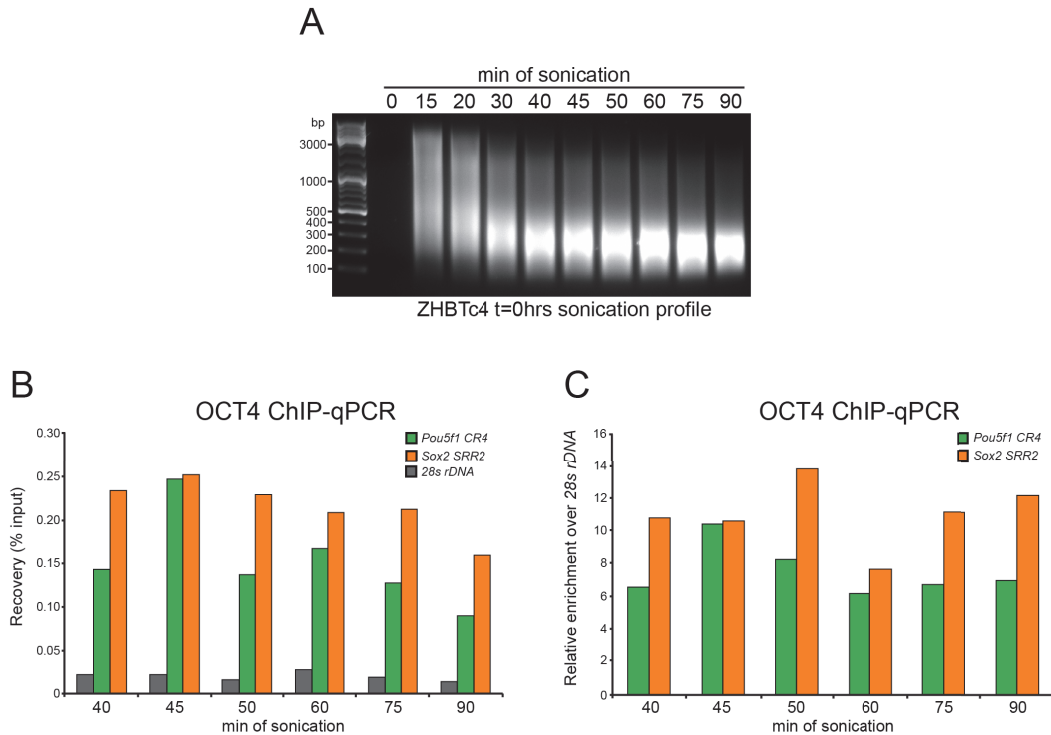

Figure 18 Optimization of sonication for ChIP-seq sample preparation.

A: Size distribution of 1µg of purified chromatin analysed by agarose gel (1.2%) separation at different sonication times **B**: Bar chart representing OCT4 ChIP-qPCR analysis at the *Pou5f1 CR4* and *Sox2 SRR2* loci after various sonication times. Data represented as percentage of input **C**: Bar chart representing OCT4 ChIP-qPCR analysis at the *Pou5f1 CR4* and *Sox2 SRR2* loci after different sonication times. Data represented as relative enrichment over control background loci 28s rDNA.

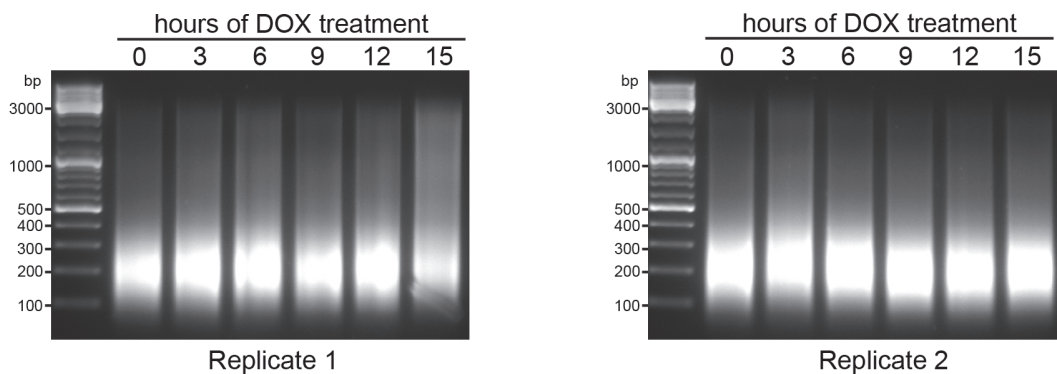
Figure 19 Sonication profile of ChIP-seq samples replicates.

Size distribution after 75 minutes of sonication of 1µg of purified chromatin analysed by agarose gel (1.2%) separation for replicate 1 (left) and replicate 2 (right).

Chromatin was isolated from $2x10^7$ cells ZHBTc4 cells after DOX treatment for 0 hours, 3 hours, 6 hours, 9 hours, 12 hours, and 15 hours and displayed good fragmentation after 75 minutes of sonication (**Figure 19**).

OCT4- and SOX2-ChIP was performed followed by ChIP-qPCR analysis of the *Pou5f1 CR4, Sox2 SRR2*, and *28s rDNA* regions. Absolute recovery was determined for the different samples and time points (**Figure 20**). Replicate 1 showed an unexpected OCT4 binding profile at the *Pou5f1 CR4* and an expected OCT4 binding profile at the *Sox2 SRR2* locus (**Figure 20a**). Replicate 2 showed expected OCT4 binding profiles for both genomic loci (**Figure 20b**). Both replicates showed the expected SOX2 binding profiles at the tested genomic loci (**Figure 20c and d**). Despite the observed inconsistencies we decided to proceed with the ChIP-seq library preparation and sequencing.
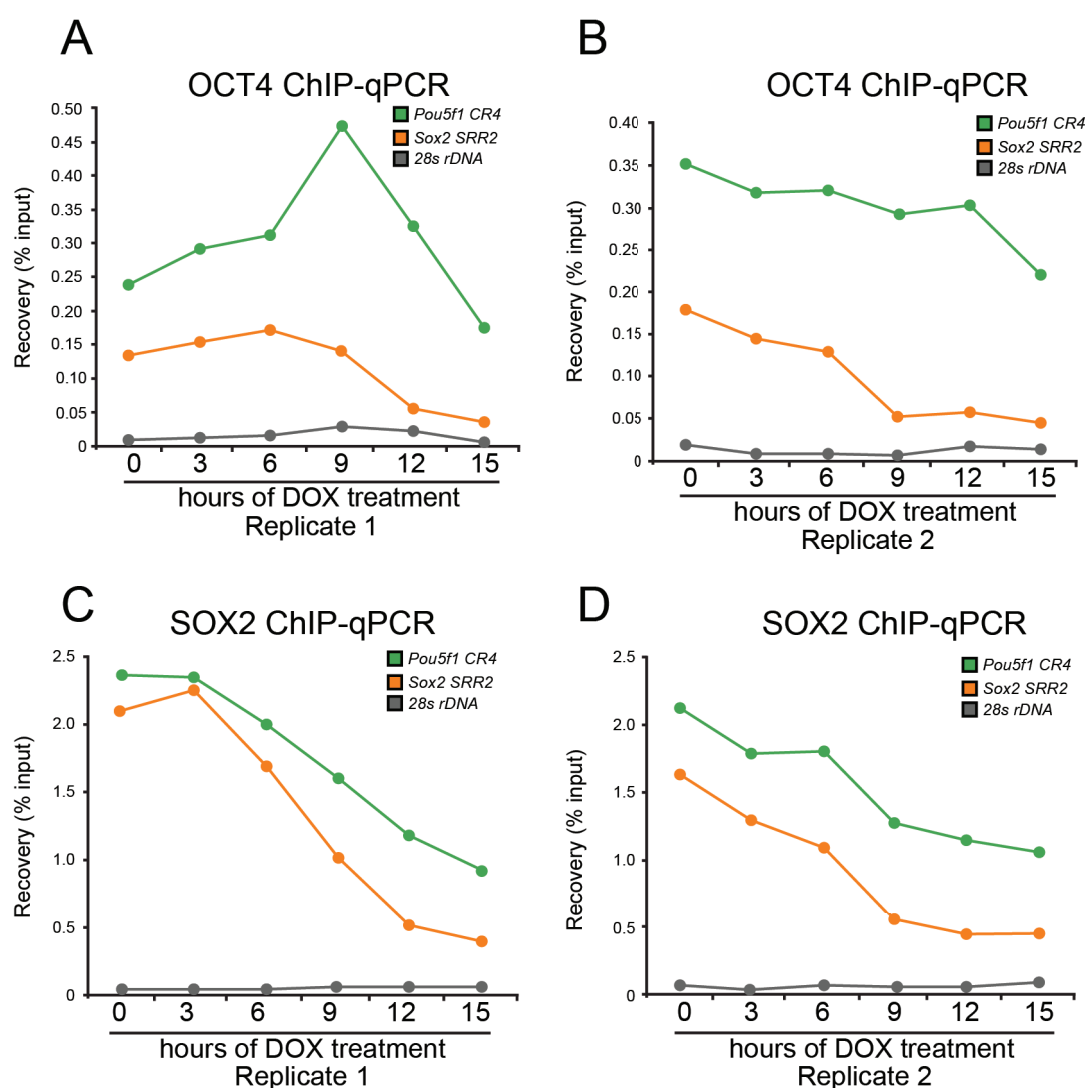

Figure 20 Quality control of OCT4- and SOX2-ChIP samples by ChIP-qPCR at *Pou5f1 CR4* and *Sox2 SRR2* loci.

A: Line graphs indicating OCT4 ChIP-qPCR analysis over DOX treatment time course. Data

presented as percentage of input, *28s rDNA* as control region **B**: Line graphs showing SOX2 ChIP-qPCR analysis over DOX treatment time course. Data presented as percentage of input, *28s rDNA* as control region **C**: Line graphs illustrating OCT4 ChIP-qPCR analysis over DOX treatment time course. Data presented as percentage of input, *28s rDNA* as control region **D**: Line graphs representing SOX2 ChIP-qPCR analysis over DOX treatment time course. Data presented as percentage of input, *28s rDNA* as control region.

After peak calling it became obvious that the data was insufficient (**Table 19**). We observed a very low number of OCT4 peaks and a low number of SOX2 peaks across all time points compared to publicly available data sets (King & Klose 2017). However, about 80% of our SOX2 peaks overlap with the published SOX2 ChIP-seq data in the same cell line (King & Klose 2017). We integrated the publicly available OCT4, SOX2 and NANOG ChIP-seq data sets in ZHBTc4 cells at 0 hours and 24 hours of DOX treatment with our TT-seq data.

	OCT4 peaks	SOX2 peaks
0 hours	1618	3856
3 hours	429	4649
6 hours	326	5158
9 hours	245	2724
12 hours	230	2216
15 hours	202	1226

Table 19 Summary of called peaks in OCT4 and SOX2 ChIP-seq data.

Summary of OCT4 and SOX2 peaks called in ChIP-seq data over the time course of DOX treatment.

3.7 Identification of OCT4-bound enhancers and super-enhancers

Integration of both publicly available OCT4 ChIP-seq and ATAC-seq data sets (King & Klose 2017) with our TT-seq data allowed for the analysis of mRNA and eRNA synthesis at OCT4 binding sites and characterization of OCT4-regulated promoters, enhancers and super-enhancers. Principle component analysis showed a separation of the samples based on the replicates (PC1) and time (PC2) for putative eRNAs (**Figure 21a**). PCA plots for mRNA synthesis displayed a clear separation based on time (PC1), while replicates (PC2) showed some variability (**Figure 21b**). Sample distance analysis showed a similar distribution with clear separation of replicate 1 and 2, as well as inter sample distribution of the individual time points (**Figure 21c and d**).
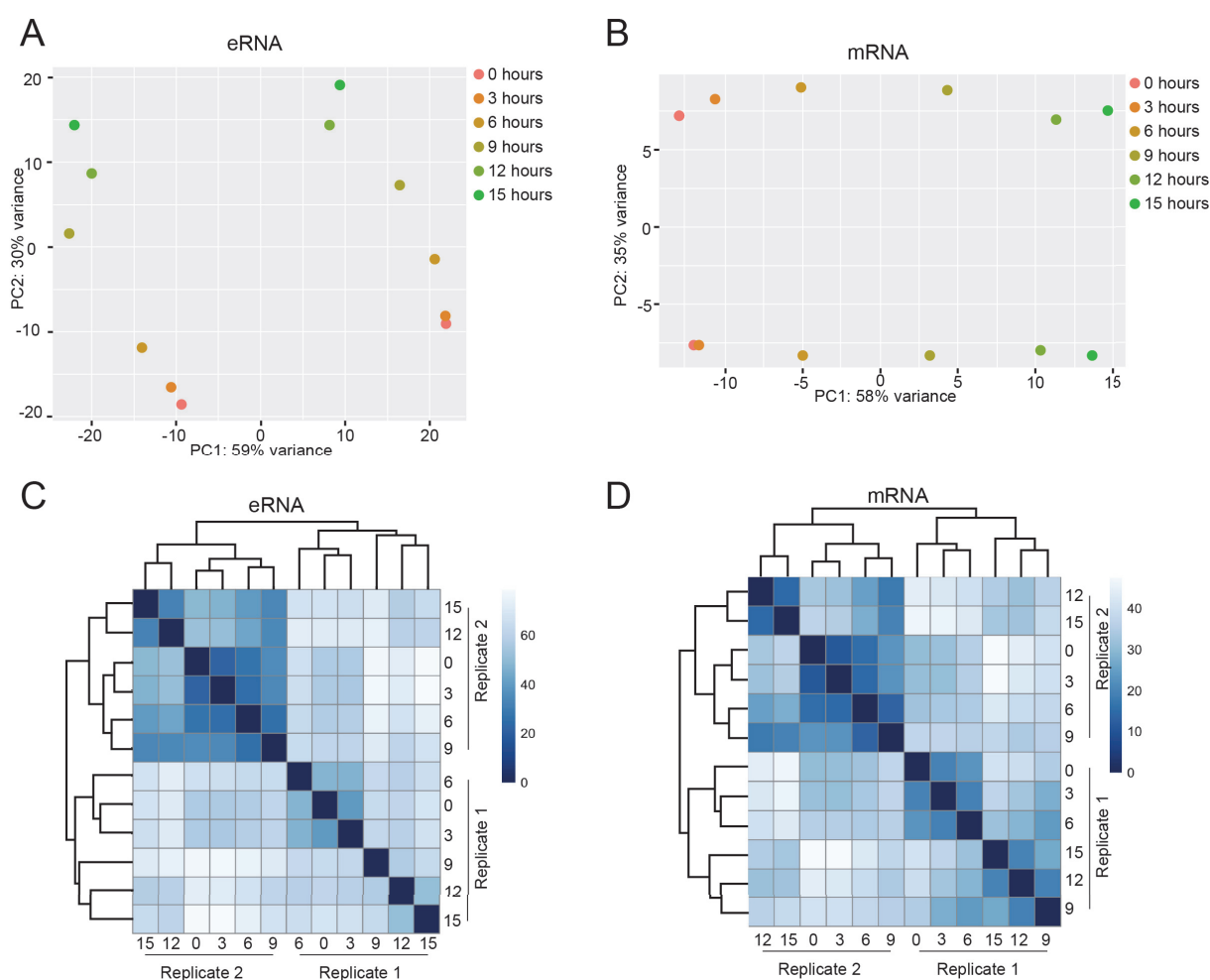
Figure 21 Assessment of sample distance for eRNA and paired protein-coding transcriptional units.

A: Principle component analysis of TT-seq replicates for eRNA TU. PC1 indicating sample variability and PC2 indicating time course **B**: Principle component analysis of TT-seq replicates for mRNA TUs. PC1 depicting sample variability and PC2 indicating time course **C**: Heatmap of sample distance for TT-seq replicates for eRNA TUs **D**: Heatmap of sample distance for TT-seq replicates for mRNA TUs.

Next, ncRNAs in the intergenic, sense intragenic and anti-sense intragenic classes were assessed for eRNA synthesis based on their ATAC-seq and/or OCT4 signal. These loci were further classified into typical enhancers/super-enhancers. In total 7469 eRNA were found, of which 2241 overlapped with sites bound by OCT4, and of which 7455 overlapped with accessible chromatin (**Figure 22a**). 2227 eRNAs showed an overlap with both sites bound by OCT4 and accessible chromatin whereas 5242 showed an overlap with only ATAC-seq peaks (accessible chromatin) (**Figure 22b**). More specifically, 467 eRNAs of the antisense intragenic overlapped with sites bound by OCT4 and accessible chromatin and 1133 eRNAs displayed accessible chromatin only (**Figure 22c**). 1554 intergenic eRNAs overlapped with sites bound by OCT4 and accessible chromatin, while 3214 eRNAs of this class

overlapped with the ATAC-seq signals only. 206 sense intragenic eRNAs showed OCT4 binding and accessible chromatin and 895 displayed accessible chromatin only.
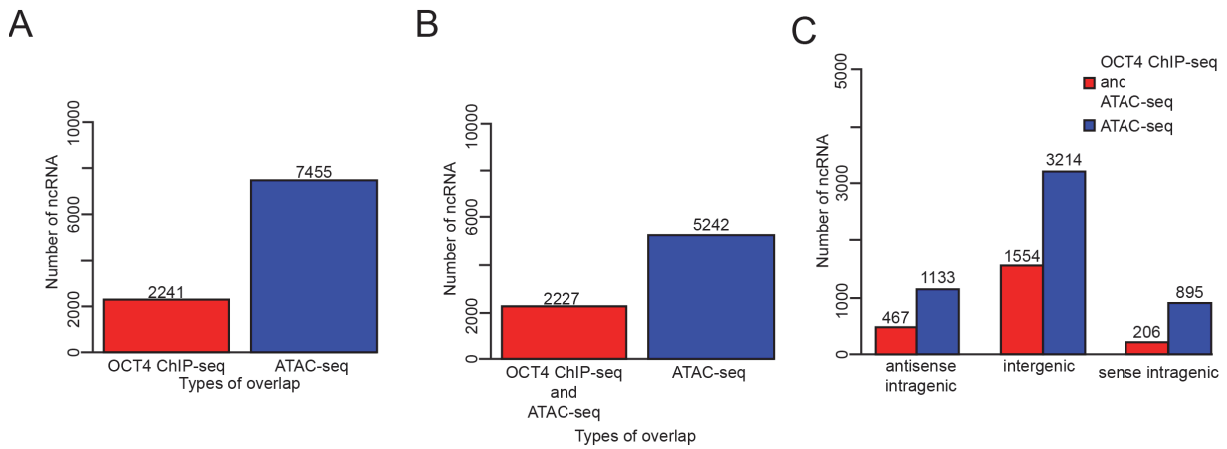
A

B

C

Figure 22 Overlap of accessible chromatin and/or OCT4 binding events with eRNA synthesizing sites.

A: Bar chart depicting overlap of eRNA TUs with OCT4 ChIP-seq peaks and ATAC-seq peaks (King & Klose 2017) **B**: Bar chart indicating the overlap of eRNA TUs with both OCT4 ChIP-seq and ATAC-seq or only ATAC-seq data **C**: Bar chart representing overlap of eRNA TUs in antisense intragenic, intergenic, or sense intergenic classes with both OCT4 ChIP-seq and ATAC-seq or only ATAC-seq data.

Based on OCT4 occupancy 8794 enhancers and 231 super-enhancers were identified. Overall, 810 eRNAs overlapped with 635 typical enhancers and 255 eRNAs were found in 117 super-enhancer regions ($\pm$50% of super-enhancers found in OCT4 ChIP-seq data based on (Whyte et al. 2013)).

Next, the differential synthesis of putative eRNAs (all intergenic, antisense and intragenic ncRNAs) and eRNAs (as defined by overlap of putative eRNAs and OCT4 ChIP-seq) was addressed. The differential expression analysis of putative eRNAs indicated a gain of synthesis for 59 putative eRNAs at 15 hours of DOX treatment. Moreover, 3, 41, 164, and 284 putative eRNAs were lost at 6, 9 ,12, and 15 hours of DOX treatment, respectively (**Figure 23a**). For the differentially synthesized eRNAs we found 43 up-regulated eRNAs, of which 8 eRNAs were up-regulated as early as 9 hours of DOX treatment. A total of 227 eRNAs were down-regulated at 15 hours of DOX treatment, with 3, 35, and 121 eRNAs at time points 6, 9, and 12 hours, respectively (**Figure 23b**). We found 172 differentially synthesized eRNAs within loci bound by OCT4, and 123 differentially expressed eRNAs overlapped with ATAC-seq peaks, i.e. accessible chromatin (**Figure 23c**). More specifically, of the 172 eRNAs

that originate around OCT4-bound sites 76 were not annotated (NA) before, 18 overlapped with super-enhancers and 78 were found in regions of typical enhancers (**Figure 23d**). 118 eRNAs within regions of accessible chromatin were NA, 2 originated from regions within super-enhancers and 3 within typical enhancers (**Figure 23d**).
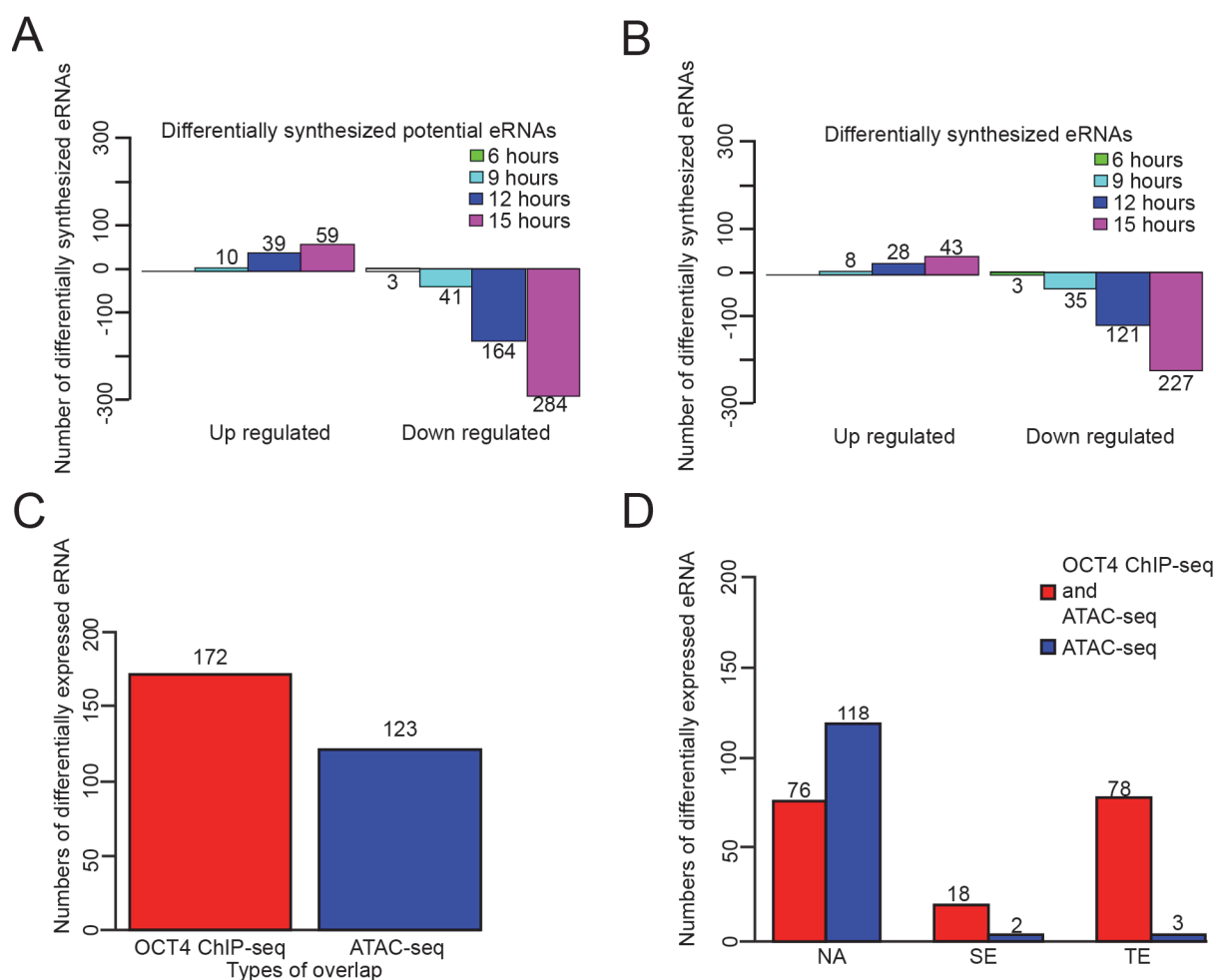

Figure 23 Analysis of differentially synthesized eRNAs and potential eRNAs.

A: Bar chart depicting up- and down-regulated ncRNA (overlap with ATAC-seq) **B**: Bar chart showing up- and down-regulated potential eRNA (overlap with ATAC-seq) **C**: Bar chart illustrating the overlap of differentially synthesized eRNA with either OCT4 ChIP-seq peaks or ATAC-seq peaks **D**: Overlap of differentially synthesized eRNA with either OCT4 ChIP-seq peaks or ATAC-seq peaks that have not been annotated (NA), Super-enhancer (SE), or typical enhancer (TE).

Next, we paired differentially synthesized eRNAs with protein-coding genes based on proximity and similar synthesis behavior and performed gene ontology analysis. Gene ontology analysis revealed biological processes related to transcriptional regulation and chromatin remodeling (**Figure 24**).
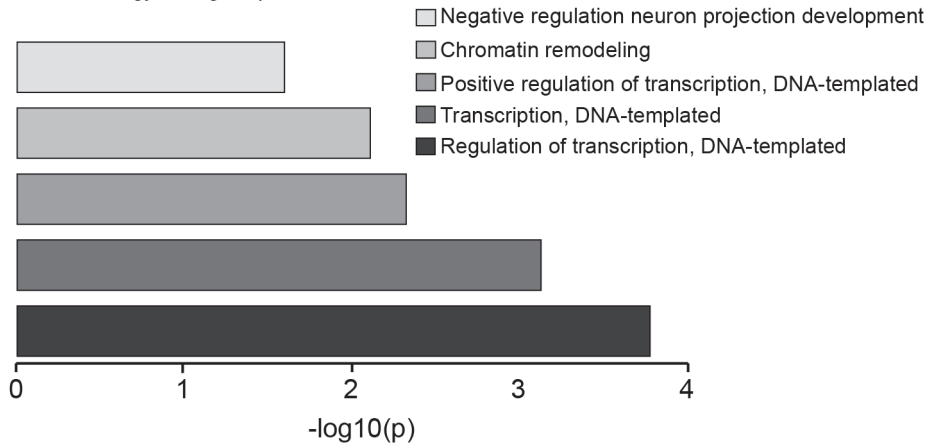
Gene Ontology Biological processes

Figure 24 Gene Ontology analysis of biological processes for protein-coding genes paired with differentially synthesized eRNA.

GO analysis of biological processes for differentially synthesized eRNA that overlap with OCT4 ChIP-seq peaks.

Analyzing the kinetics of synthesis of differentially synthesized eRNAs shows that down-regulated eRNAs can be considered lowly synthesized at 9 hours. This is also the time point where up-regulated eRNAs can be considered synthesized (**Figure 25a**). However, the paired protein-coding genes showed different kinetics and did not always correlate with the kinetics of the paired eRNAs (**Figure 25b**).
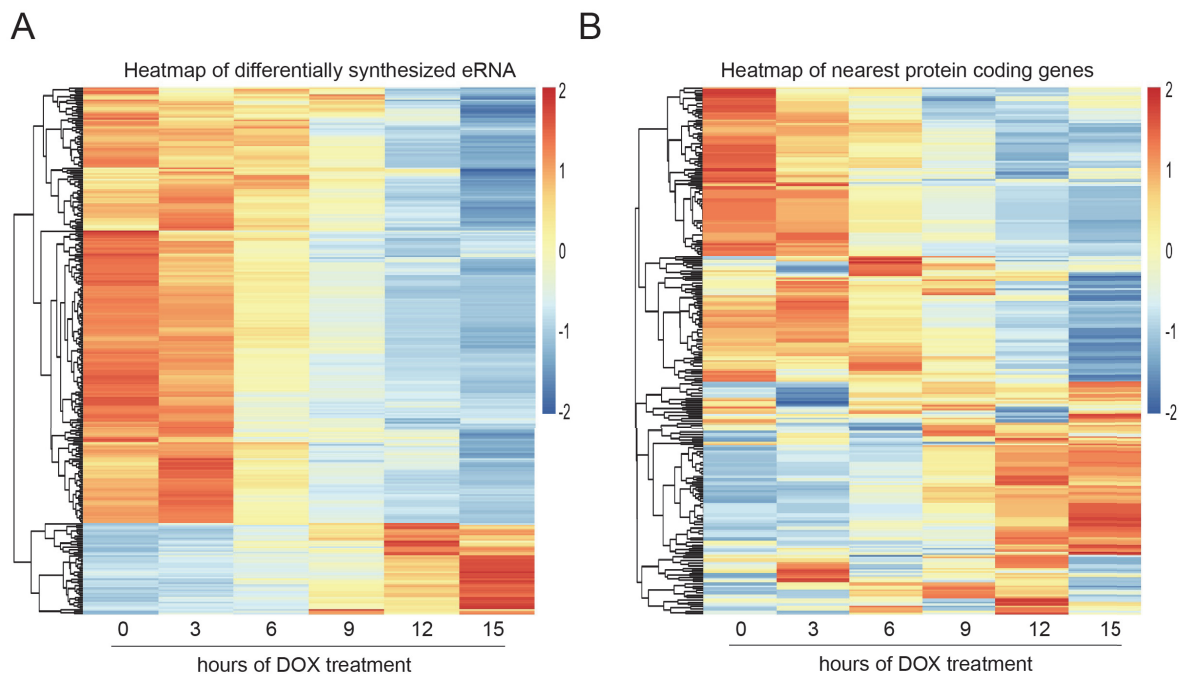

Figure 25 Dynamics of differentially synthesized eRNAs and paired protein-coding genes.

A: Heatmap showing the kinetics of differentially synthesized eRNAs (replicates 1 and 2 were combined) **B**: Heatmap depicting the kinetics of paired protein-coding genes to the differentially synthesized eRNAs shown in (**A**) (replicates 1 and 2 were combined).

Most eRNAs at OCT4-bound sites were down-regulated in a similar manner over the time course of DOX treatment/OCT4 loss, a minority of eRNAs were up-regulated (**Figure 26a**). The analysis of the paired protein-coding genes displayed de- or increased synthesis of these genes, which did not follow an apparent pattern (**Figure 26b**).
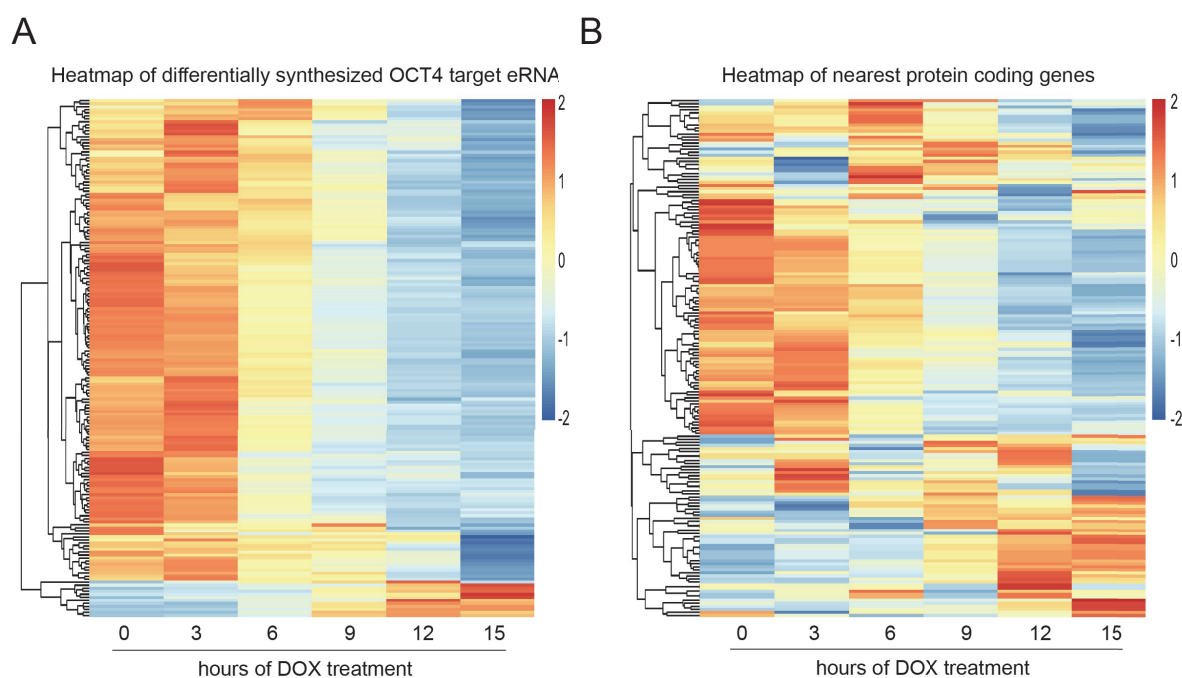

A: Heatmap of differentially synthesized OCT4 target eRNA

B: Heatmap of nearest protein coding genes

Figure 26 Dynamics of differentially synthesized eRNAs at OCT4-bound sites and paired protein-coding genes.

A: Heatmap representing the kinetics of differentially synthesized eRNAs that are targeted by OCT4 (replicates 1 and 2 were combined) **B**: Heatmap illustrating the kinetics of paired protein-coding genes to the differentially synthesized eRNAs shown in (**A**) (replicates 1 and 2 were combined).

Next, we evaluated the kinetics of the protein-coding genes of all differentially synthesized eRNAs (**Figure 27a**). Most protein-coding genes were down-regulated including *Klf4, Klf5, Etv1,* and *Chd1*. The analysis of up-regulated eRNAs displayed two clusters, one cluster being up-regulated at 12 hours and the second cluster at 15 hours of DOX treatment. *Igf2r* and *Tinagl1* were linked to up-regulated eRNA. The list becomes shorter when only the linked differentially expressed protein-coding genes are considered (**Figure 27b**). The kinetics of loss and gain of eRNA synthesis remain very similar to **Figure 27a**. DEG linked to differentially synthesized eRNAs are *Klf5, Zfp42, Etv1* and *Utf1* for the down-regulated genes. *Tinagl1* and *Peg10* are among the up-regulated genes linked to differentially expressed eRNAs (**Figure 27b**).

A

Heatmap of differentially synthesized eRNA

B

Heatmap of differentially synthesized eRNA whose nearest gene is also differentially synthesized

hours of DOX treatment
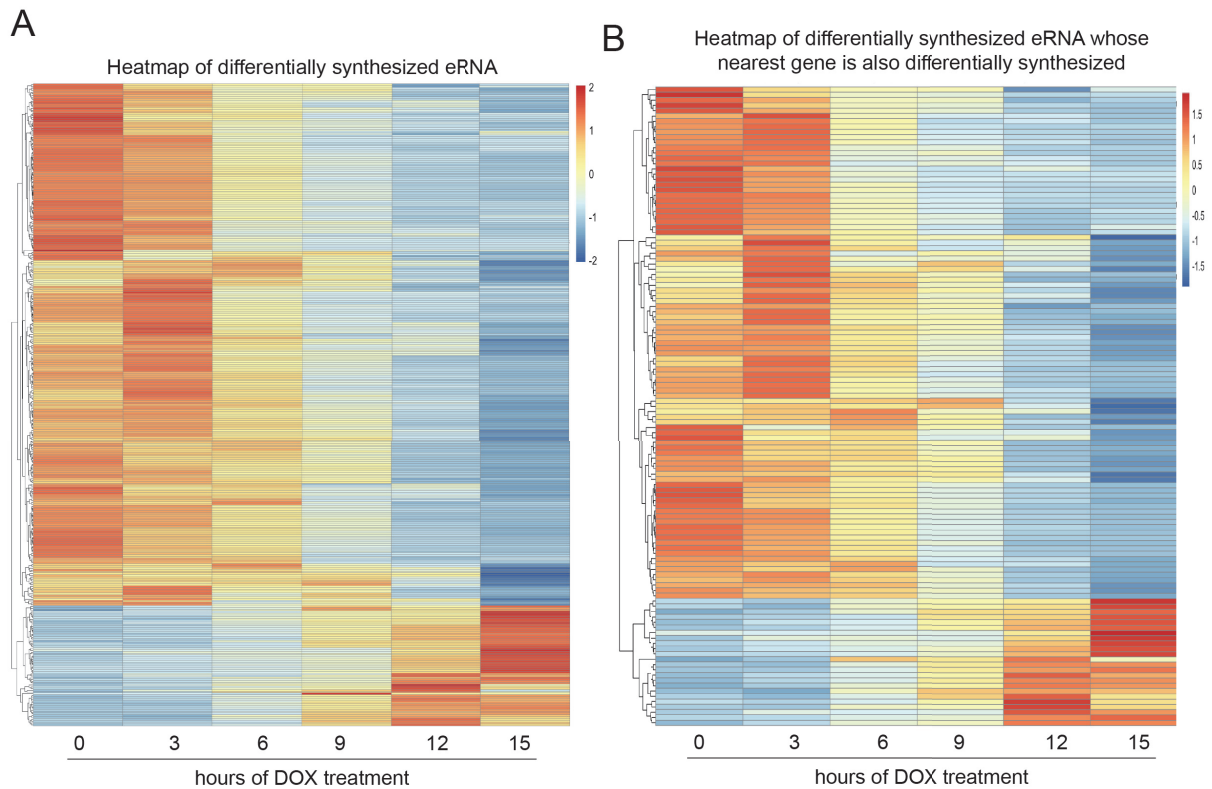
Figure 27 Dynamics of differentially synthesized eRNAs and paired differentially expressed protein-coding genes.

A: Heatmap depicting the kinetics of differentially synthesized eRNAs (replicates 1 and 2 were combined) **B**: Heatmap representing the kinetics of differentially synthesized paired protein-coding genes to the differentially synthesized eRNAs shown in (**A**) (replicates 1 and 2 were combined).

Next, the kinetics of the paired protein-coding genes were addressed (**Figure 28a**). Differentially expressed protein-coding genes were down- and up-regulated in a similar manner (**Figure 28b**). At 9 hours of DOX treatment down-regulated protein-coding genes show weak synthesis levels and up-regulated protein-coding genes show first signs of additional synthesis (**Figure 28b**).

Next, differentially expressed genes that are targeted by super-enhancers were analysed (**Figure 29**). Among the down-regulated genes were *Esrrb, Sall4, Nr5a2, Utf1, Tet1* and *Nanog*.

A: Heatmap of nearest gene of differentially synthesized eRNA

B: Heatmap of nearest differentially synthesized genes of differentially synthesized eRNA
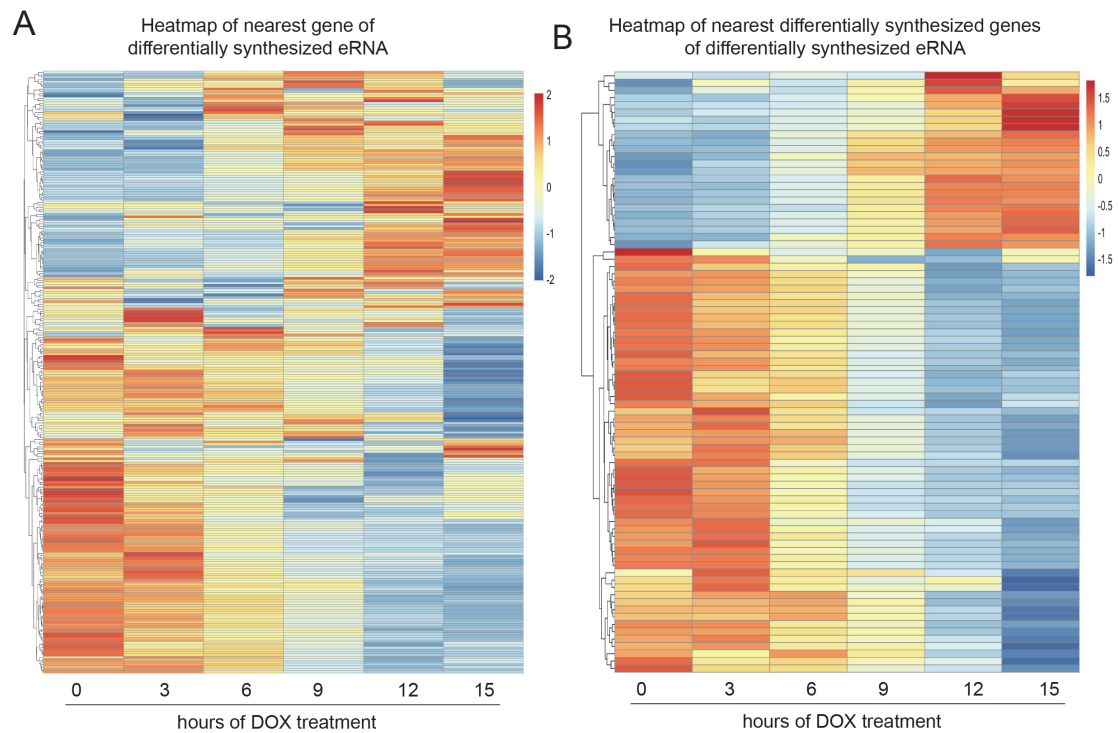
Figure 28 Dynamics of protein-coding genes paired to differentially synthesized eRNAs.

A: Heatmap showing the kinetics of all nearest protein-coding genes to differentially synthesized eRNAs (replicates 1 and 2 were combined) **B**: Heatmap illustrating the kinetics of differentially expressed paired protein-coding genes to differentially synthesized eRNAs (replicates 1 and 2 were combined).
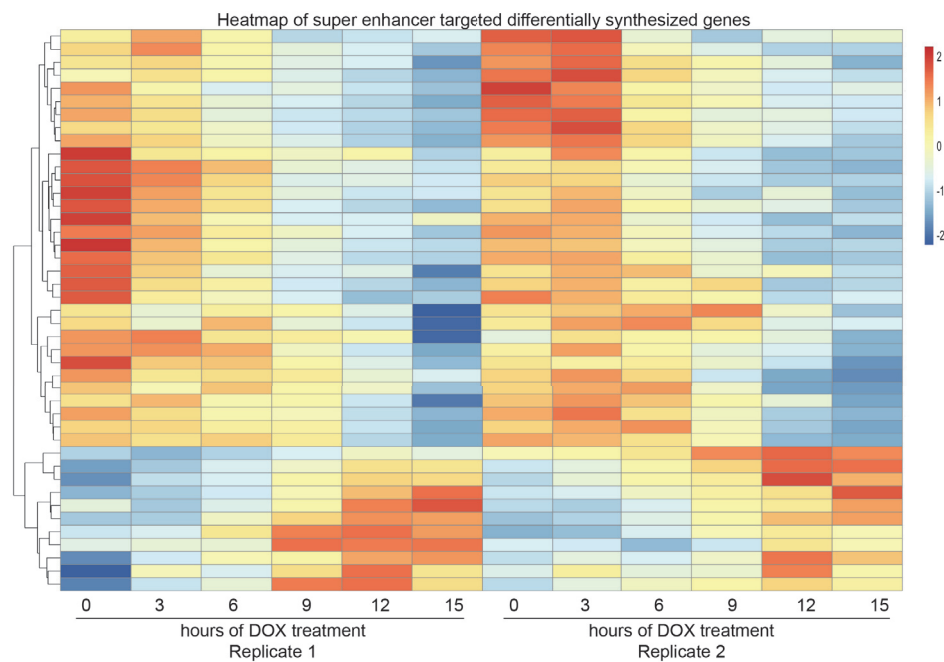
Figure 29 Dynamics of super-enhancer targeted differentially expressed protein-coding genes.

Heatmap indicating differentially expressed protein-coding genes targeted by super-enhancers (replicates 1 and 2 separately shown).

Kinetics of eRNA synthesis at super-enhancers was linked to their corresponding protein-coding genes (**Figure 30a and b**). A total of 200 super-enhancer originating eRNAs were assessed and linked to their corresponding protein-coding genes. We observed a striking difference between replicate 1 and 2.
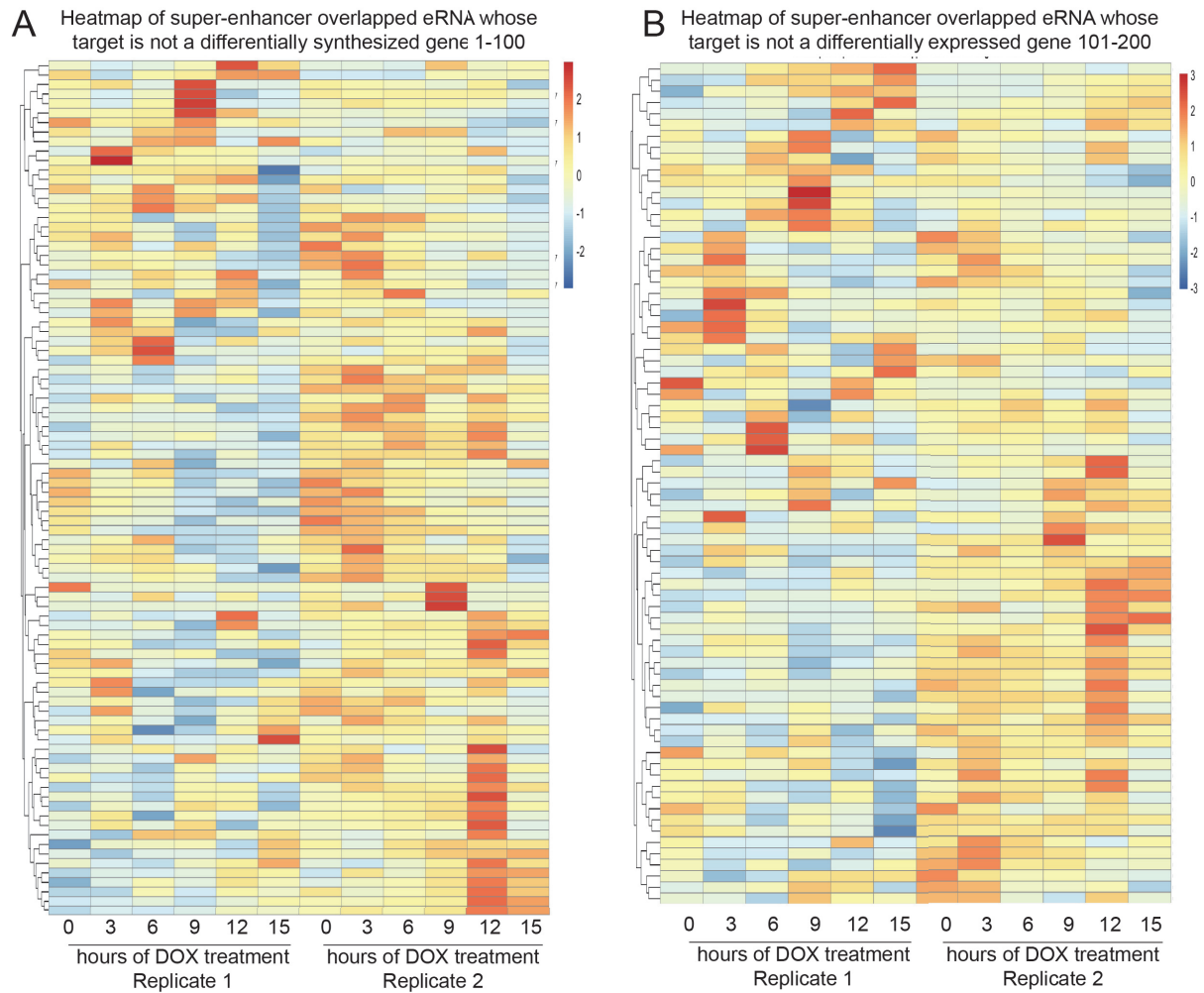
Figure 30 Dynamics of differentially synthesized eRNAs from 200 super-enhancers whose target genes are not differentially expressed.

A: First 100 super-enhancers that are differentially synthesized without their target being differentially expressed (replicates 1 and 2 shown separately) **B**: Second 100 super-enhancers that are differentially synthesized without their target being differentially expressed (replicates 1 and 2 shown separately).

24 differentially expressed super-enhancers were identified (**Figure 31a**). *Klf4*, *Klf5* and *Utf1* were linked to down-regulated super-enhancers. In addition, the kinetics of the differentially expressed super-enhancer target genes were addressed (**Figure 31b**). The eRNAs and protein-coding transcripts for *Klf5, Utf1*, and *Klf4* were down-regulated with similar kinetics.

As the ATAC-seq data from the King & Klose 2017 only contained 0 and 24 hours of DOX treatment we performed our own ATAC-seq experiment to determine the kinetics of chromatin accessibility changes upon loss of OCT4.
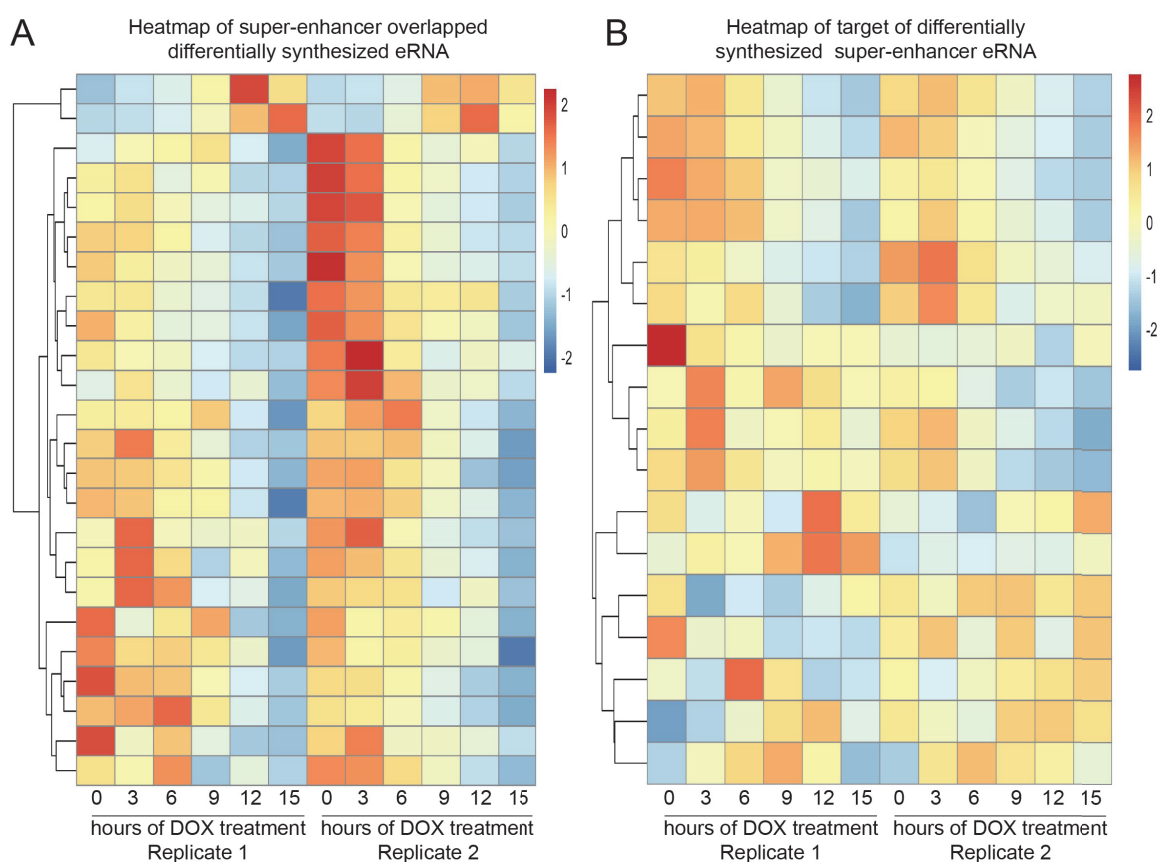
Figure 31 Dynamics of differentially synthesized super-enhancer eRNAs and their differentially expressed paired protein-coding genes.

A: Heatmap indicating differentially synthesized eRNAs (replicates 1 and 2 shown separately) **B**: Heatmap depicting differentially expressed protein-coding genes that are targeted by differentially synthesized eRNAs (replicates 1 and 2 shown separately).

3.8 Chromatin dynamics during loss of OCT4

ATAC-seq allows for analysis of chromatin accessibility and provides insights into chromatin dynamics (Buenrostro et al. 2015). ATAC-seq was performed at the aforementioned time points to assess chromatin dynamics. Bioanalyzer analysis was performed to determine the optimal fragmentation by the Tn5 transposase and documented sub-nucleosome peaks, single nucleosomes and di-nucleosome peaks (**Figure 32**). The ATAC-seq libraries were sequenced and a total number of ~100.000 peaks were called across all time points (**Figure 33a**).

Representative bioanalyzer profile of ATAC-seq samples
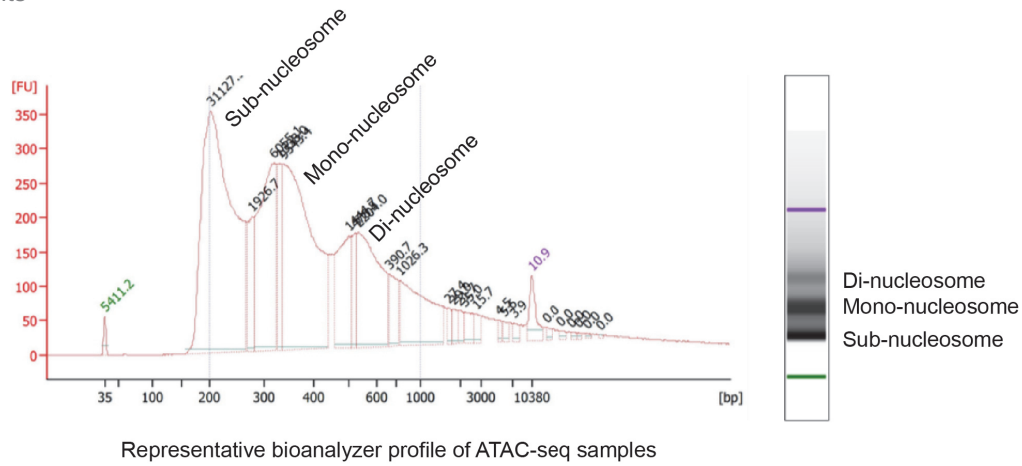
Figure 32 Quality control of ATAC-seq sample preparation.

Bioanalyzer track of representative ATAC-seq sample depicting fragmentation of genome after Tn5 transposase treatment.
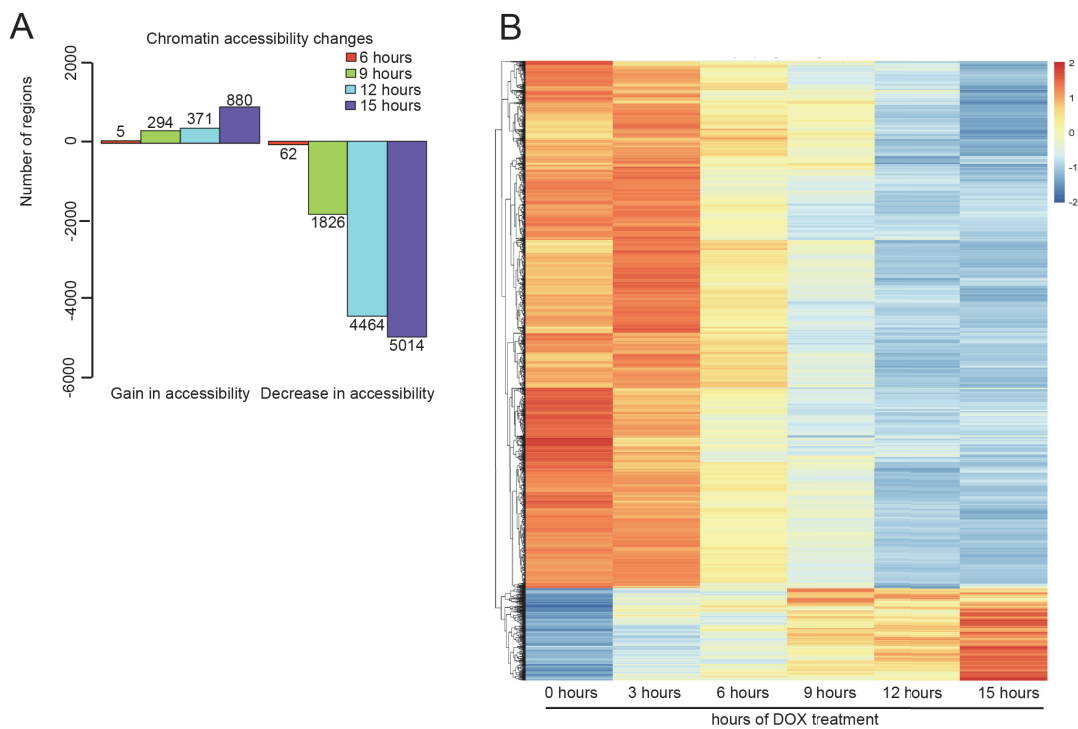
Figure 33 Analysis of alterations to chromatin accessibility.

A: Bar chart representing increased and decreased chromatin accessibility over time course of OCT4 loss **B**: Heatmap depicting kinetics of increased and decreased chromatin accessibility over time course of OCT4 loss as defined in (**A**).

Over the time course we found 880 regions that gained accessibility after loss of OCT4, whereas 5014 regions lost accessibility (**Figure 33a**). The decrease in accessibility was most striking between 6 and 9 hours and between 9 and 12 hours. The global changes occur in a similar manner for both loss and gain of accessibility (**Figure 33b**).

100

A more thorough investigation of the chromatin dynamics revealed three different clusters of loss of accessibility after loss of OCT4 (**Figure 34a**), and two clusters of gain of accessibility (**Figure 34b**). The dynamics of these clusters for individual regions were depicted in heatmaps for loss (**Figure 34c**) and gain (**Figure 34d**) of accessibility. In cluster 1 the changes occurred between 3 and 12 hours of DOX treatment (**Figure 34a and c, left**). Cluster 2 showed a steady loss of accessibility upon loss of OCT4 (**Figure 34a and c, middle**) and cluster 3 displayed a continuous loss of accessibility starting between 3 and 6 hours of DOX treatment (**Figure 34a and c, right**). The first cluster displaying a gain of accessibility reached a plateau around 9 hours of DOX treatment, whereas cluster 2 showed a steady, albeit slower, gain in accessibility (**Figure 34b and d**).
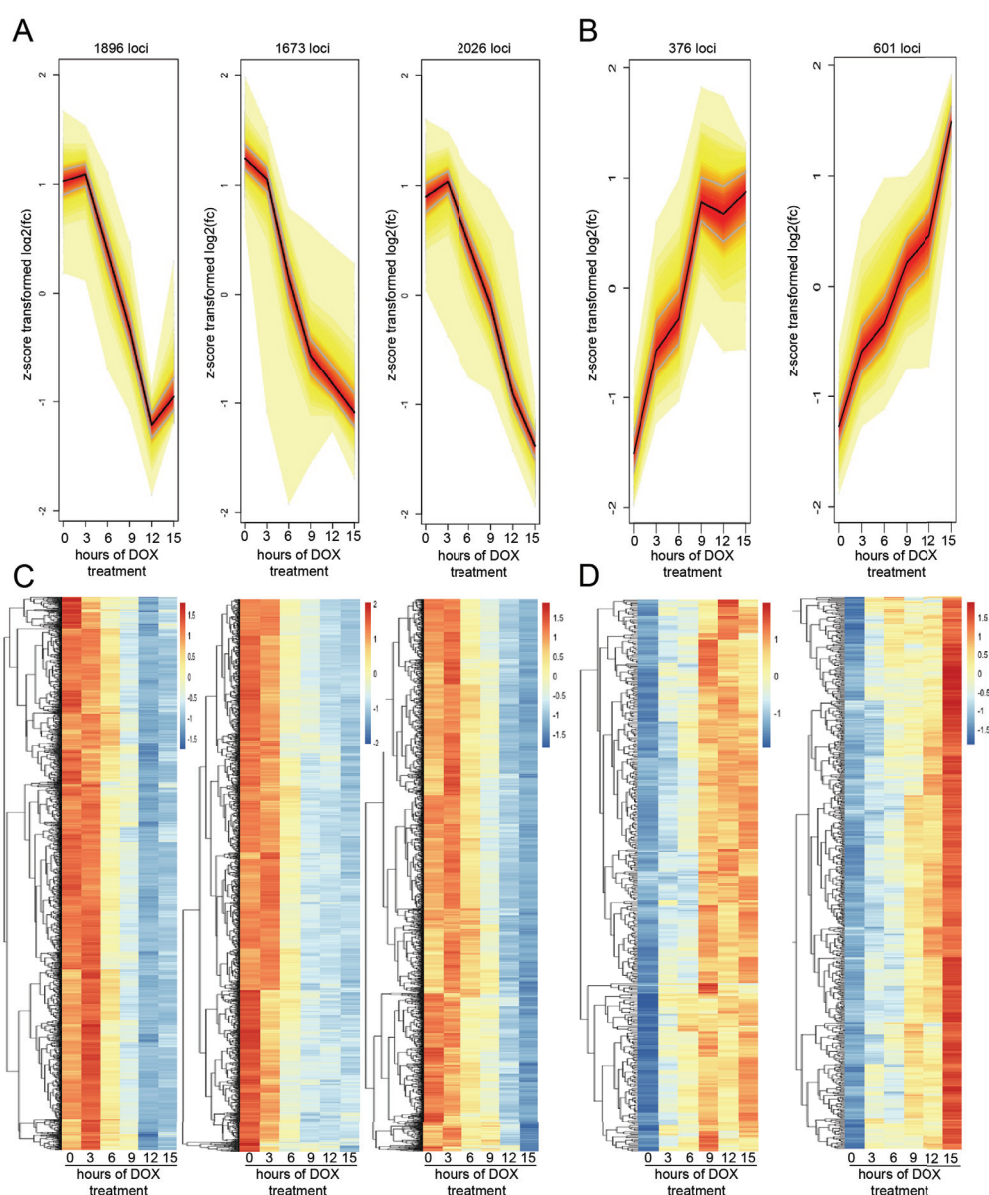
(**Figure legend on next page**)

Figure 34 Analysis of global changes in chromatin accessibility over time course of DOX treatment.

A: Heat signature line graphs showing three clusters with decreased chromatin accessibility across time course of DOX treatment. Cluster I including 1896 genomic loci, cluster II 1673 genomic loci, cluster III 2026 genomic loci (y-axis showing median count) **B**: Heat signature line graphs showing two clusters with increased chromatin accessibility across time course of DOX treatment. Cluster I including 376 genomic loci, cluster II 601 genomic loci (y-axis showing median count) **C**: Heatmaps depicting the kinetics of decreased accessibility in three clusters as defined in (**A**) **D**: Heatmaps depicting the kinetics of increased accessibility in two clusters as defined in (**B**).

Integration of the OCT4 ChIP-seq data revealed that the majority of regions that lose accessibility were bound by OCT4 (**Figure 35**). More specifically, 4615 of 5595 sites losing accessibility were bound by OCT4. In contrast, only 72 of 977 sites gaining accessibility were bound by OCT4 (**Figure 35**).
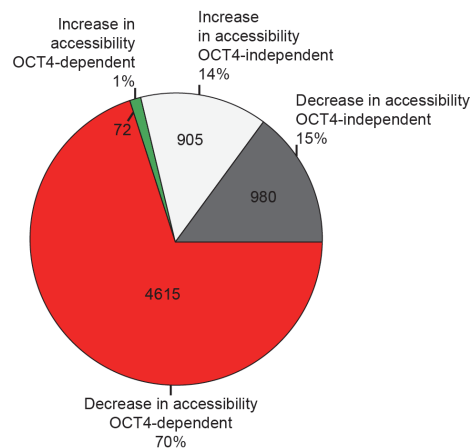
Figure 35 Overlap of chromatin accessibility changes and OCT4 binding events.

Pie chart indicating the overlap of OCT4 binding sites with increased (1%, 72 loci) and decreased (70%, 4615 loci) chromatin accessibility.

Analysis of the ATAC-seq data regarding the various TU classes in the TT-seq data (**see Figure 15a**) displayed that the vast majority of protein-coding, conRNA and uaRNA overlapped with open chromatin (**Figure 36**). 81% of all lincRNA displayed open chromatin, whereas antisense intragenic, intergenic and sense intragenic TUs showed open chromatin in only 53%, 46%, and 37%, respectively (**Figure 36**).

ATAC-seq signal occupation in TUs
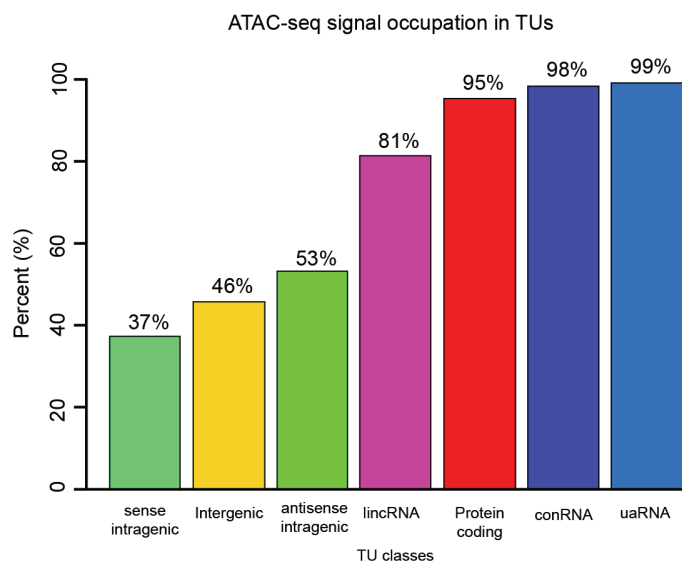
Figure 36 Overlap of transcriptional unit classes with accessible chromatin.

Bar chart illustrating the overlap of ATAC-seq signal (open chromatin) with TU classes as defined in (**Figure 14a**).

3.9 Analysis of global OCT4, SOX2 and NANOG binding at differentially accessible regions

Next, ChIP-seq data sets (OCT4, SOX2, NANOG in ZHBTC4 at 0 and 24 hours (King & Klose 2017)) were integrated in our TT-seq and ATAC-seq data. Global analysis of TF occupancy at genomic OCT4-bound regions was performed. The metagene plots show that OCT4 binding is lost dramatically at both clusters, with and without altered chromatin accessibility, after 24 hours of DOX treatment (**Figure 37a and b**).
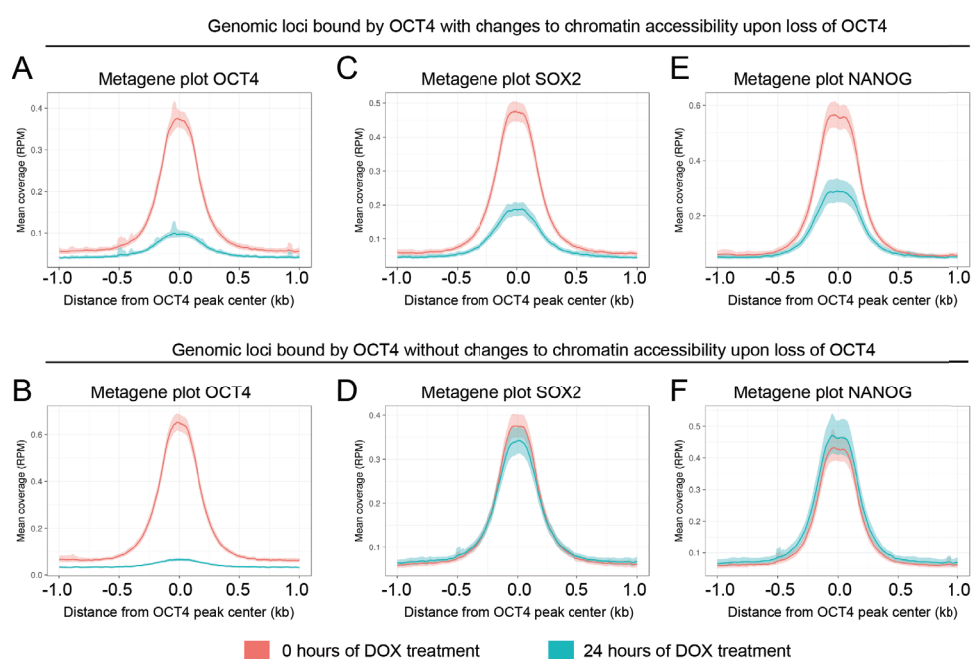
(**Figure legend on next page**)

Figure 37 Analysis of OCT4, SOX2, NANOG global binding dynamics at 0 hours and 24 hours of DOX treatment overlapped with or without changes to chromatin accessibility.

Metagene plot depicting OCT4 binding intensity centered to OCT4-bound sites at time point 0 hours (Red) and 24 hours (Turquoise) of DOX treatment at sites that show a decrease of chromatin accessibility (**A**) or that do not show a decrease of chromatin accessibility (**B**). Metagene plot showing SOX2 binding intensity centered to OCT4-bound sites at time point 0 hours (Red) and 24 hours (Turquoise) of DOX treatment at sites that show decreased chromatin accessibility (**C**) or that do not show decreased chromatin accessibility. (**D**) Metagene plot indicating NANOG binding intensity centered to OCT4-bound sites at time point 0 hours (Red) and 24 hours (Turquoise) of DOX treatment at sites that show a decrease of chromatin accessibility (**E**) or that do not show a decrease of chromatin accessibility (**F**).

OCT4-binding at regions that showed altered chromatin accessibility was slightly more contained. SOX2- and NANOG-binding was partially lost at OCT4-bound sites with altered chromatin accessibility (**Figure 37c and e**). SOX2-binding was mostly contained at OCT4-bound sites without altered chromatin accessibility following loss of OCT4, whereas NANOG displayed a slight increase in binding at these sites (**Figure 37e and f**). Taken together, these results showed two global clusters of OCT4-regulation: (1) OCT4-bound sites that were dependent on the presence of OCT4 to maintain chromatin accessibility. Loss of OCT4 at these sites also showed a decrease in SOX2 and/or NANOG occupancy and a sub sequential decrease in chromatin accessibility. (2) OCT4-bound sites that did not depend on the presence of OCT4 to preserve chromatin accessibility. At these sites SOX2 and/or NANOG were still present and the chromatin remained accessible. *De novo* motif discovery in the regions with altered chromatin accessibility identified the OCTSOX composite motif (**Figure 38a**).

(**Figure legend on next page**)

However, *de novo* motif discovery in regions that did not show alterations to the chromatin accessibility only identified OCT4 binding motifs (**Figure 38b)**. These results indicate that OCT4 loss could be compensated by SOX2- and NANOG-binding at regions where OCT4 and SOX2 or NANOG do not bind closely together. More interestingly, at sites where OCT4 and SOX2 bind together the loss of OCT4 also leads to a loss of SOX2- and NANOG-occupancy and reduction in chromatin accessibility.

3.10 Clustering of OCT4-bound sites based on transcriptional activity

Next, we integrated the TT-seq data into the ChIP-seq and ATAC-seq data to address whether the cooperative TF binding influenced transcriptional activity and to assess the TF binding and chromatin dynamics at sites that show or do not show transcriptional activity. Overall, we found 17719 OCT4-bound sites of which the vast majority (83%) did not overlap with TUs (non-transcriptional units, NTU) (**Figure 39a**).

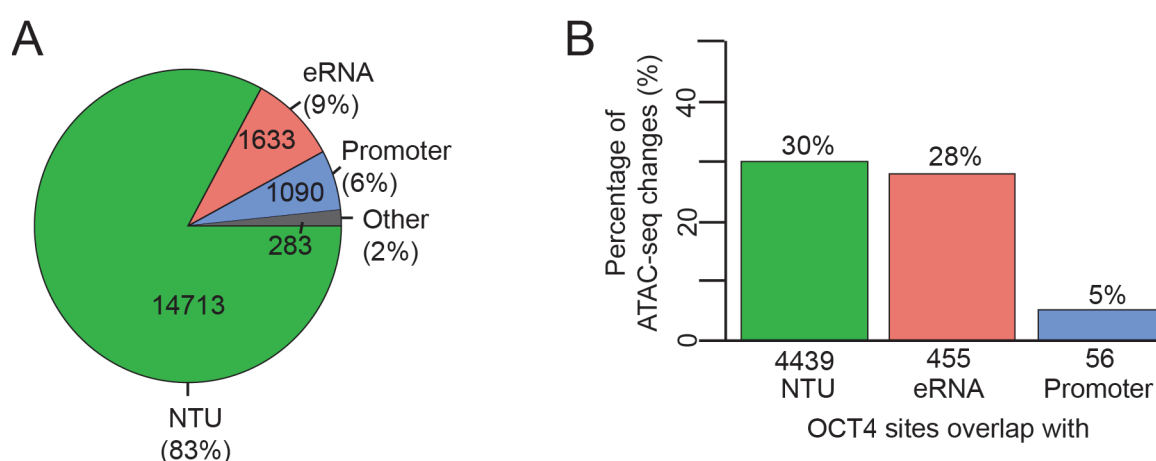**Figure 39 Overlap of OCT4-bound sites with transcriptional units and changes to chromatin accessibility.**

A: Pie chart depicting overlap of OCT4-bound sites that overlap with no transcriptional units (NTU),eRNA, promoter, other. **B**: Bar chart indicating the percentage of chromatin accessibility changes within three groups of OCT4-bound sites that show NTU, eRNA, Promoter as defined in (**A**).

The remaining OCT4-bound sites (17%) overlapped with eRNA TUs (9%), Promoter TUs (6%), and other TUs (2%) (**Figure 39a**). 30% of OCT4-bound NTU sites showed a change in chromatin accessibility (4439 sites), 28% of OCT4-bound sites with eRNA synthesis overlapped with altered chromatin accessibility (28%) and 5% of OCT4-bound sites close to promoter TU displayed changes to chromatin accessibility (56 sites) (**Figure 39b**).

We performed meta gene analysis for OCT4 binding at the NTU, eRNA TUs, and Promoter TUs. We found that OCT4 binding was strongest at sites with eRNA synthesis, followed by OCT4-bound without transcriptional activity (NTU) and by sites close to promoters (**Figure 40a**). In addition, we addressed the occupancy of other TFs (SOX2, NANOG, ESRRB, KLF4, and c-Myc), histone modifications (H3K4me1, H3K4me2, H3K4me3, H3K27ac), chromatin modifier and co-factor (p300 and Med1), and chromatin accessibility (Med1 in ZHBTc4 0 hours (Whyte et al. 2013), ESRRB, KLF4, cMyc, p300, H3K4me1, H3K4me2, H3K4me3, H3K27ac in mESC v6.5 (Chronis et al. 2017).

Binding for SOX2, NANOG, ESRRB and KLF4 but not c-Myc showed the highest enrichment at OCT4-bound sites with eRNA synthesis (**Figure 40b-f**) with the strongest enrichment for ESRRB. SOX2, NANOG and ESRRB also displayed a strong enrichment at OCT4-bound sites without eRNA synthesis. KLF4 did not show a difference in binding at NTU OCT4-bound sites or OCT4-bound promoter TUs. c-Myc displayed a preference for OCT4-bound sites that synthesize promoter TUs followed by eRNA TUs and lastly OCT4-bound NTU sites. Binding of the Mediator subunit Med1 showed a preference for OCT4-bound sites that synthesize eRNA TUs whereas OCT4-bound NTU sites displayed less Med1-binding (**Figure 40g**). Histone modifier p300 also showed a binding preference at OCT4-bound eRNA TUs sites (**Figure 40h**). Distribution of H3K27ac around the OCT4-bound sites with different TUs was different at eRNA TUs and promoter TUs sites (**Figure 40i**). H3K27ac overlapped with OCT4-bound sites at promoter TUs whereas eRNA TUs OCT4-bound sites showed a bi-modal distribution of H3K27ac, which at a lower level was also the case at NTU OCT4-bound sites (**Figure 40i)**. Other histone modifications displayed cluster-specific distributions. We found high H3K4me1 levels and lower H3K4me3 levels at OCT4-bound sites with eRNA synthesis. As expected, we did not observe the same results in promoter regions (Banerji et al. 1983; Calo & Wysocka 2013). H3K4me1 was identified at slightly higher levels at OCT4 NTU sites and appeared to be more bimodal than eRNA TUs OCT4-bound sites (**Figure 40j**).
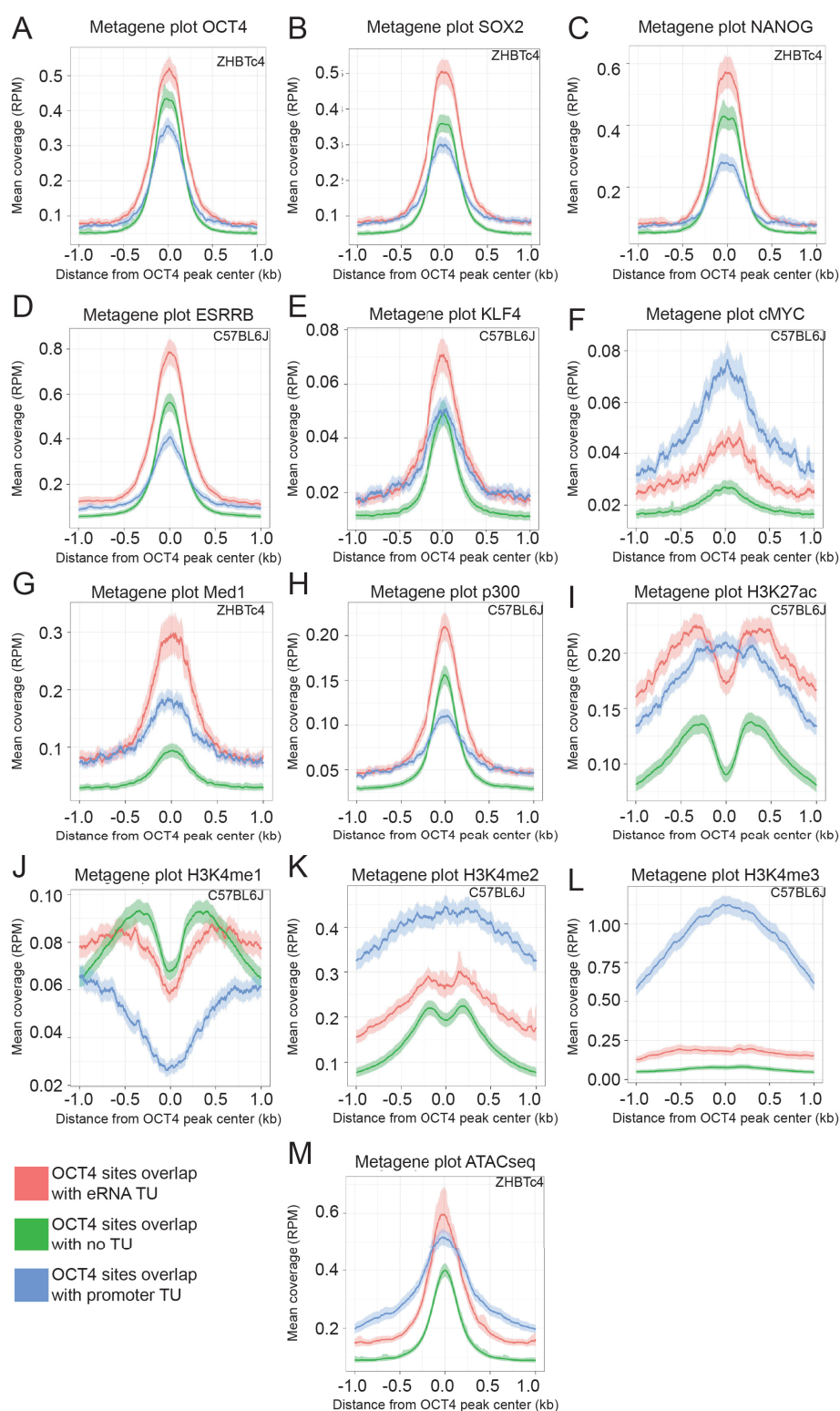
Figure 40 Global analysis of overlap of transcription factor binding with OCT4-bound sites and transcriptional activity.

Metagene analysis of TF binding intensity/histone modification enrichment centered to OCT4-bound sites in three

clusters as defined in (**Figure 39a**) OCT4 (**A**), SOX2 (**B**), NANOG (**C**), ESRRB (**D**), KLF4 (**E**), cMyc (**F**), Med1 (**G**), p300 (**H**), H3K27ac (**I**), H3K4me1 (**J**), H3K4me2 (**K**), H3K4me3 (**L**), Chromatin accessibility (**M**).

Promoter TUs OCT4-bound sites were devoid of H3K4me1. H3K4me2 showed a stronger enrichment for Promoter TUs OCT4-bound sites, whereas a similar distribution was documented at eRNA TUs and NTU OCT4-bound sites seem to be relatively similar, with a slightly higher enrichment at eRNA TUs sites (**Figure 40k**). Promoter TUs OCT4-bound sites were enriched exclusively for H3K4me3 (**Figure 40l**). Analysis of the chromatin accessibility around the three different TU clusters revealed that chromatin was most accessible at eRNA TUs and Promoter TUs OCT4-bound sites with a wider distribution at Promoter TUs OCT4-bound sites (**Figure 40m**). NTU OCT4-bound sites were less accessible in comparison to the other clusters. NTU OCT4-bound sites displayed a narrower distribution of chromatin accessibility (**Figure 40m**). Our results indicate that not all the OCT4-bound sites show transcriptional activity and binding of multiple TFs appears to be important for transcriptional productivity.

3.11 Dynamics of RNA synthesis and chromatin accessibility at OCT4 target sites

To gain further insight into the role of OCT4 at enhancers and promoters, the dynamics of RNA synthesis and chromatin accessibility were assessed. OCT4-bound sites with eRNA synthesis (**Figure 39a**) were analyzed with regard to these kinetics. Four distinct clusters of ATAC-seq and TT-seq dynamics were identified (**Figure 41a**). Cluster I, comprised of 816 sites bound by OCT4, showed a slight decrease in chromatin accessibility and a steady decrease in eRNA synthesis. Cluster II included 554 OCT4-bound sites with no change in chromatin accessibility but an increase in eRNA synthesis. Cluster III containing 219 OCT4-bound sites displayed a decrease in chromatin accessibility and a strongly reduced RNA synthesis. In cluster IV (256 sites with OCT4 binding) the chromatin accessibility was rapidly diminished with a steadier drop in eRNA synthesis. Our results show a maximum decrease in chromatin accessibility at 12 hours of DOX treatment. The individual heatmaps (**Figure 41b and c**) depict the kinetics for chromatin accessibility and eRNA synthesis on an individual gene level (one line represents one gene). Cluster I showed three separate sub-clusters for the ATAC-seq data but similar synthesis kinetics for the paired eRNA. Cluster II displayed some variability within the ATAC-seq signals as well as varying kinetics for the eRNA synthesis at the corresponding genomic loci. Cluster III and IV showed homogenous kinetics for both the chromatin

accessibility and eRNA synthesis changes (**Figure 41b and c**). Taken together we identified four distinct manners of chromatin and transcription dynamics occurring at OCT4-bound sites that synthesize eRNA. In general, chromatin accessibility changes and transcriptional activities were not always coupled (cluster I and II), indicating an independent regulation.
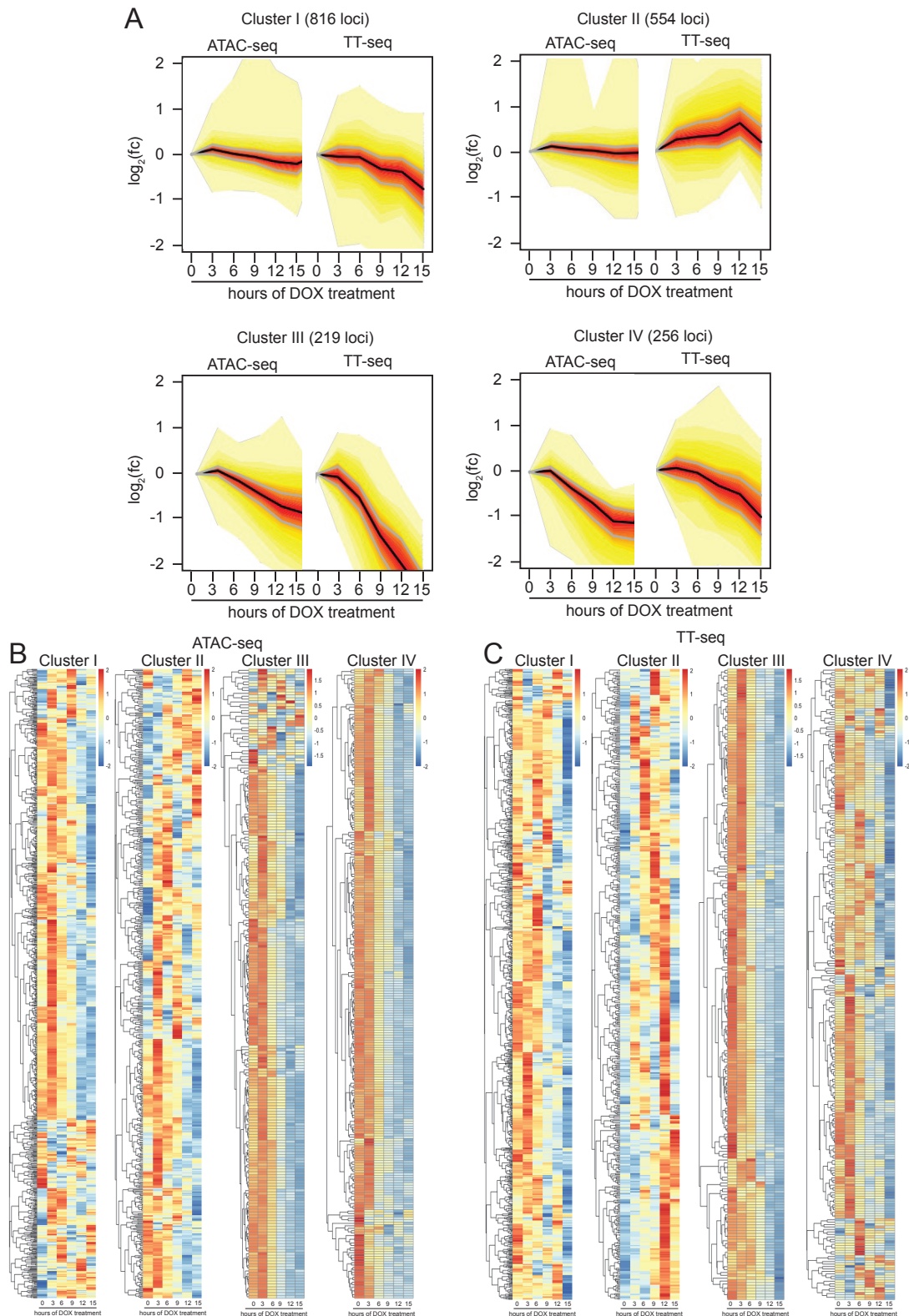
(**Figure legend on next page**)

Figure 41 Clustering of paired ATAC-seq and TT-seq dynamics at OCT4-bound sites that show eRNA synthesis.

A: Heat signature line plot depicting four clusters of ATAC-seq and TT-seq data based on OCT4-bound sites with eRNA synthesis as defined in (**Figure 40a**). Cluster I 816 loci, cluster II 554 loci, cluster III 219 loci, cluster IV 256 loci. Left side ATAC-seq, right side: TT-seq. Y-axis: Log2 fold change, X-axis: time course. **B**: Heatmaps depicting kinetics of chromatin accessibility in individual loci in the four clusters as defined in (**A**) **C**: Heatmaps showing kinetics of transcriptional activity in individual loci in the four clusters as defined in (**A**).

Next, binding of OCT4, SOX2, NANOG, ESRRB, KLF4, p300 and Med1 as well as the occupancy of histone modifications H3K4me1, H3K4me3, and H3K27ac were measured in these four clusters. OCT4, SOX2 and NANOG displayed strong binding to the genomic loci represented in clusters III and IV (**Figure 42a**).
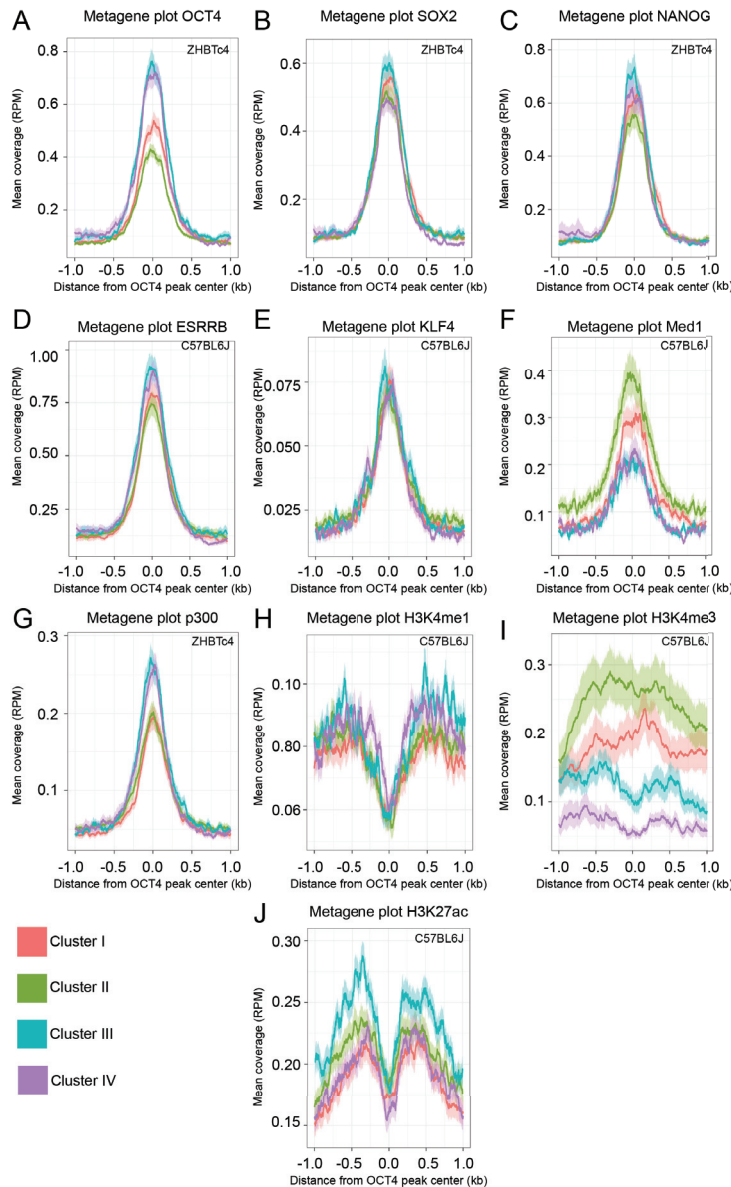
(**Figure legend on next page**)

Figure 42 Global binding analysis of TF occupancy at OCT4-bound sites based on clusters of chromatin accessibility and transcriptional activity dynamics.

Metagene analysis of TF binding intensity/histone modification enrichment centered to OCT4-bound sites in four

clusters as defined in (**Figure 41a**) OCT4 (**A**), SOX2 (**B**), NANOG (**C**), ESRRB (**D**), KLF4 (**E**), Med1 (**F**), p300 (**G**), H3K4me1 (**H**), H3K4me3 (**I**), H3K27ac (**J**).

SOX2 showed the strongest binding to genomic regions in cluster III, followed by cluster I (**Figure 42b**).

NANOG binding showed the strongest enrichment in cluster III and bound equally in clusters I and IV (**Figure 42c**). ESRRB binding showed the strongest enrichment in clusters III and IV, whereas KLF4-binding was similar in the four clusters (**Figure 42d and e**). Mediator subunit Med1 displayed a higher binding enrichment in clusters I and II than in clusters III and IV (**Figure 42f**). Co-factor p300-binding was enriched in clusters III and IV (**Figure 42g**). The histone modification H3K4me1 showed a similar enrichment pattern in the four different clusters (**Figure 42h**). H3K4me3 displayed higher levels of enrichment in clusters I and II (**Figure 42i**). H3K27ac showed a higher enrichment in cluster III, especially up-stream of the OCT4-bound site (**Figure 42j**).

Next, we analyzed the paired protein-coding genes of these clusters for their gene ontology of biological processes. We identified enrichment for 'regulation of transcription' in all four clusters (**Figure 43a-d**). In clusters I, III, and IV the biological processes 'stem cell population maintenance' and 'chromatin remodeling' were enriched.

Transcription factor binding motif discovery for OCT4, ESRRB, and KLF4 in these four clusters revealed enrichment for different motifs (**Figure 44**). The OCT4 motif showed enrichment for the POU_{hd} binding motif in clusters I and II, whereas the complete OCT4 motif was enriched in clusters III and IV (**Figure 44a-d**). The SOX2 motif was identified in clusters I, II and III (**Figure 44e-g**). Similar to OCT4, ESRRB displayed enrichment for the short motif in clusters I and II and for the extended motif in cluster III (**Figure 44h-j**). The KLF4 motif was detected similarly in all four clusters (**Figure 44k-n**). The OCTSOX motif was enriched in clusters III and IV, the SOX motif was enriched in clusters I and II with a low probability for the $OCT4_{hd}$ motif (**Figure 44o-r**).

A Gene Ontology Biological processes
Cluster I

- Glomerular basement membrane development
- Glomerular filtration
- Regulation of chromosome segregation
- Positive regulation of fibroblast proliferation
- Cell cycle
- Mitotic nuclear division
- Lens induction in camera-type eye
- Negative regulation of DNA replication
- Stem cell differentiation
- Negative regulation of transcription from RNA polymerase II promoter
- Positive regulation of transcription from RNA polymerase II promoter
- Cell division
- Stem celll population maintenance
- Transcription, DNA-templated
- Regulation of transcription, DNA-templated

B Gene Ontology Biological processes
Cluster II

- Negative regulation of transcription, DNA-templated
- Sabeceous development
- Cellular protein complex localization
- Negative regulation of gene expression, epigenetic
- Regulation of transcription from RNA polymerase II promoter
- Lung aveolus development
- Multicellular organism development
- Covalent chromatin modification
- Skin development
- Cell-cell adhesion
- Cellular response to insulin stimulus
- Positive regulation of transcription, DNA templated
- Negative regulation of transcription from RNA polymerase II promoter
- Regulation of transcription, DNA-templated
- Transcription, DNA-templated

C Gene Ontology Biological processes
Cluster III

- Response to organic substance
- Nucleosome positioning
- Regulation of cell growth
- Chromatin silencing at rDNA
- Nucleosome assembly
- Covalent chromatin modification
- DNA-templated transcription initiation
- Cellular response to insulin stimulus
- Positive regulation of transcription from RNA polymerase II promoter
- Negative regulation of mTOR signaling
- Rhytmic process
- Positive regulation of transcription, DNA-templated
- Regulation of transcription, DNA-templated
- Transcription, DNA-templated
- Chromatin remodeling

D Gene Ontology Biological processes
Cluster IV

- Wnt signaling pathway
- Blood vessel morphogenesis
- Positive regulation of epithelial to mesenchymal transition
- Osteoclast differentiation
- Apoptotic process
- Rhythmic proces
- Negative regulation of transcription from RNA polymerase II promoter
- Cellular response to transforming growth factor beta stimulus
- Movement of cell of subcellular component
- Positive regulation of bone resorption
- Covalent chromatin modification
- Response to growth factor
- in utero embryonic development
- Stem cell population maintenance
- Osteoblast differentiation

-log10(p)

(Figure legend on next page)

Figure 43 Gene Ontology analysis of genes linked to clusters of chromatin accessibility and transcriptional activity dynamics.

GO analysis of biological processes terms found in clusters I (**A**), II (**B**), III (**C**) and IV (**D**) as defined in (**Figure 41a**).

Figure 44 Transcription factor binding motif analysis in 4 clusters defined by ATAC-seq and TT-seq dynamics on OCT4 peaks that show eRNA synthesis.

De novo motif analysis in the four clusters as defined in (**Figure 41a**) **A**: Profile of OCT4 **B**: Profile of SOX2 **C**: Profile of OCTSOX composite motif **D**: Profile of ESRRB **E**: Profile of KLF4.

Next, binding intensities of OSN at the four clusters was assessed at 0 and 24 hours of DOX treatment and displayed in Metagene plots (**Figure 45**). OCT4 binding was lost in clusters III and IV, while binding was slightly maintained in clusters I and II (**Figure 45a-d**). SOX2 binding levels were also decreased in clusters III and IV, whereas the SOX2-binding levels only showed a slight decrease in cluster I.

In cluster II on the other hand we documented a minor increase in binding levels for SOX2 (**Figure 45e-h**). NANOG binding levels were decreased in clusters III and IV, but slightly increased in cluster I. Binding intensities of NANOG were higher in genomic loci represented in cluster II (**Figure 45i-l**). In addition, we noted a slight shift away from the OCT4-binding peak center in clusters I and II for NANOG (**Figure 45i-j**).
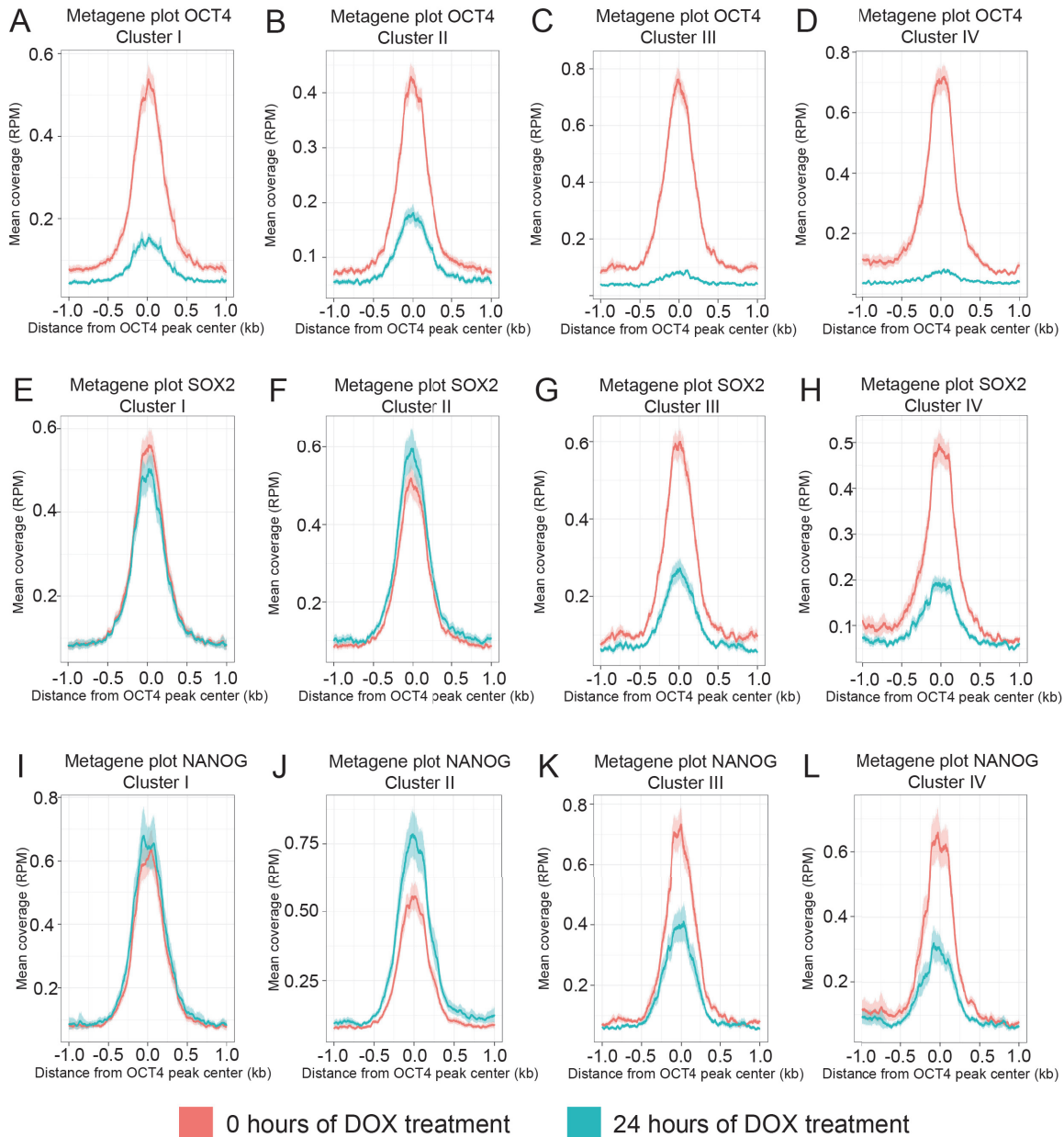
Figure 45 Global analysis of OCT4, SOX2, NANOG binding dynamics upon loss of OCT4 in clusters defined by ATAC-seq and TT-seq dynamics.

Metagene plot depicting TF binding intensity centered to OCT4-bound sites at time point 0 hours (Red) and 24 hours (Turquois) of DOX treatment in four clusters as defined in (**Figure 41a**). OCT4 (**A-D**), SOX2 (**E-H**), NANOG (**I-L**).

Finally, we performed clustering analysis of the RNA synthesis kinetics at OCT4-bound sites that overlap with promoter coding TUs (**as defined in Figure 39a**).
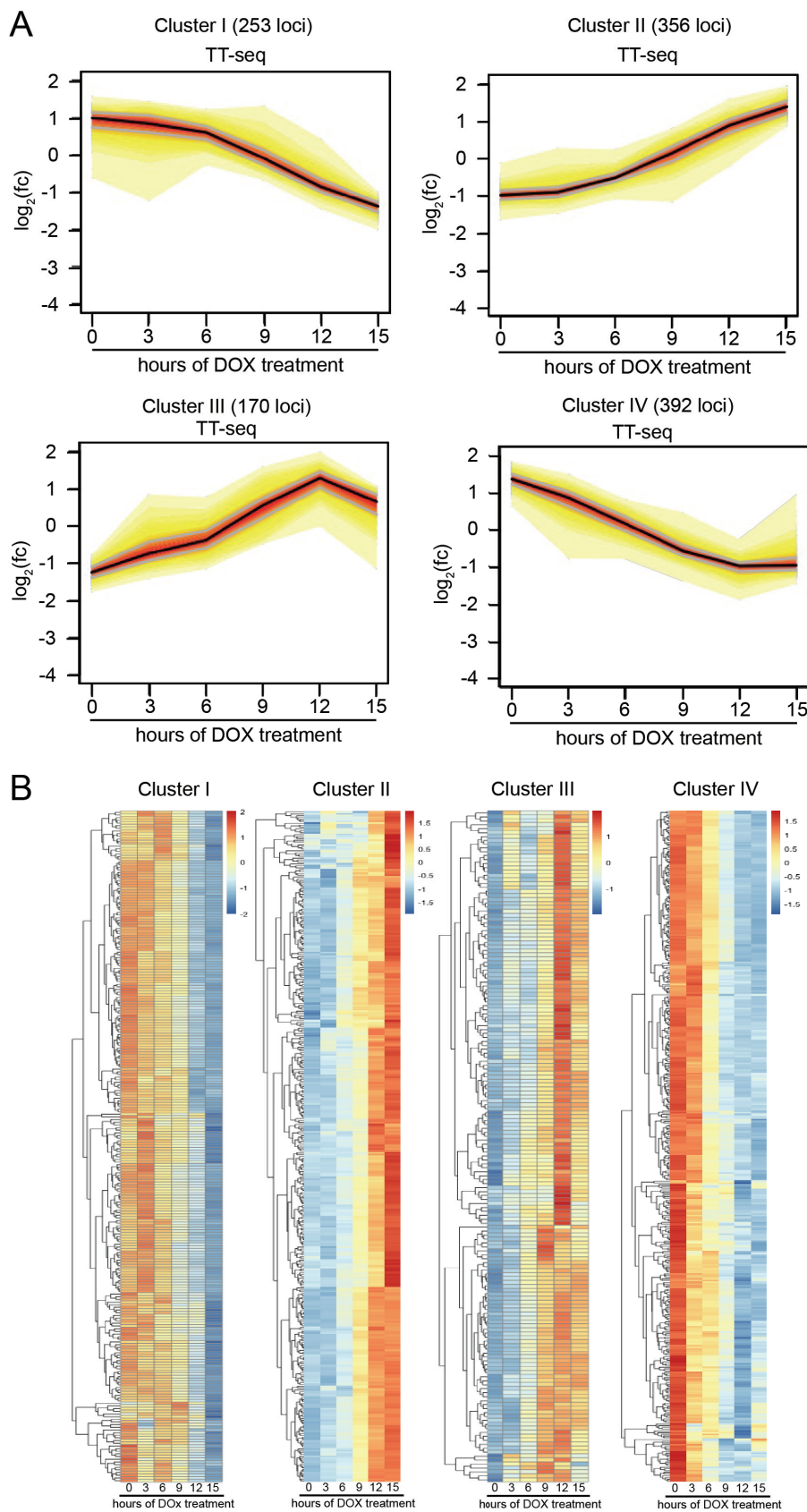
Figure 46 Clustering of differentially expressed protein-coding genes at OCT4-bound sites.

A: Heat signature line plots clustering OCT4-bound sites that show Promoter TU synthesis into 4 clusters based on transcriptional activity. Cluster I 253 loci, cluster II 365 loci, cluster III 170 loci cluster IV 392 loci. Left side: ATAC-seq, right side: TT-seq. Y-axis: Log2 fold change, X-axis: time course **B**: Heatmaps depicting kinetics of transcriptional activity of individual loci in the four clusters as defined in (**A**).

Our results suggest four different responses to loss of OCT4 (**Figure 46a**). Clusters I and IV showed down-regulated mRNA with cluster I including late response- and cluster IV containing immediate response genes (with 253, 356, 170, and 392 genomic loci in the different clusters, respectively) (**Figure 46a**). These kinetics changes were also depicted in the heatmap analysis with cluster IV showing a down-regulation as early as 6 hours of DOX treatment (**Figure 46b**). Cluster II and III displayed increased mRNA synthesis with cluster II containing late response- and cluster III including early response genes (**Figure 46a**). For cluster II we found a steady up-regulation, for cluster III an earlier up-regulation reaching a peak at 12 hours of DOX treatment. The heatmap analysis illustrated the difference in kinetics between clusters II and III in mRNA synthesis after loss of OCT4. Overall, we identified four distinct clusters of differential expression at OCT4-bound sites that show promoter TUs.

3.12 ChIP-qPCR and Western

Unfortunately, the publicly available ChIP-seq data sets did not allow for dynamic analysis of TF binding upon loss of OCT4 due to missing time points. Preliminary ChIP-qPCR analyses were performed over the kinetic time course of loss of OCT4. OCT4 Western blot analysis on chromatin and whole cell lysate was performed to determine if the loss of OCT4-binding is due to protein degradation or lost binding. This analysis showed that OCT4 protein lost global binding between 9 and 12 hours and the total protein levels are decreased (**Figure 47**). SOX2 also showed loss of global binding, but the total protein level was not altered. NANOG-binding and its total protein level were not affected by loss of OCT4.
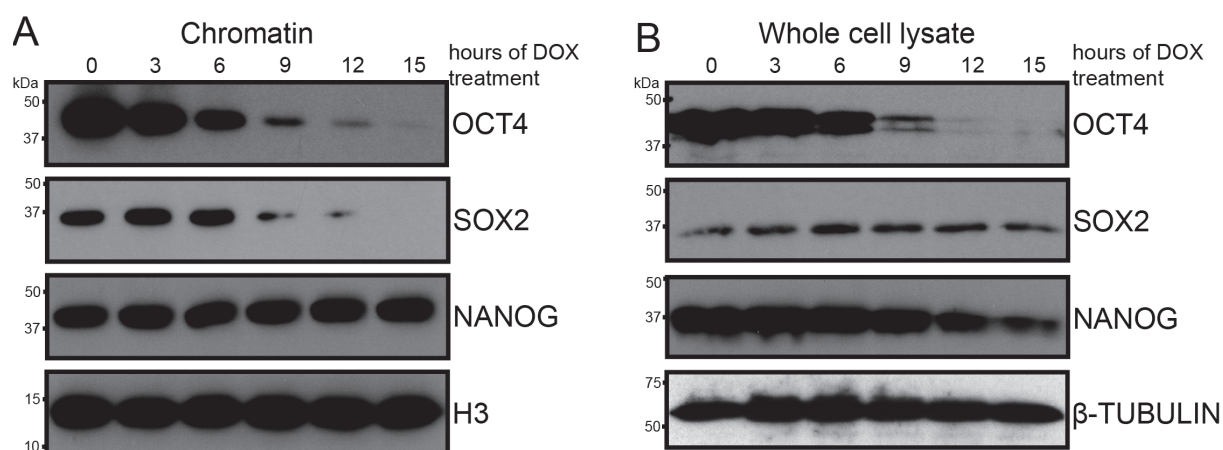
Figure 47 OCT4, SOX2 and NANOG protein expression and binding analysis.

Western blot analysis of chromatin (**A**) and whole cell lysate (**B**) samples of ZHBTc4 cells at time course of DOX treatment using OCT4, SOX2, NANOG, H3 (control) antibodies.

Next, OCT4, SOX2, NANOG, ESRRB and KLF4 binding was assessed at four different enhancer elements (*Pou5f1 CR4, Sox2 SRR2, and two Klf4* enhancers b and e) and two control regions *(28s rDNA and IGS)*. OCT4, SOX2 and NANOG occupancy was decreased at the four enhancers regions (**Figure 48a-c**). ESRRB-binding was decreased at the KLF4 enhancers and did not bind the OCT4 and SOX2 enhancer elements (**Figure 48d**). KLF4 occupancy at the Klf4 b and e element was decreased and was only minimally affected at *Pou5f1 CR4* and *SOX2 SRR2* enhancers (**Figure 48e**).

Taken together, these pieces of preliminary data suggested that OCT4 is important for the binding of other transcription factors as the loss of OCT4-binding coincided with loss of binding of other factors such as SOX2, NANOG and ESRRB.

Figure 48 Kinetics of transcription factor occupancy upon loss of OCT4.

Line graphs representing TF ChIP-qPCR analysis over the time course of DOX treatment. OCT4 (**A**), SOX2 (**B**), NANOG (**C**), ESRRB (**D**), KLF4 (**E**). Data represented as comparison to percentage input.
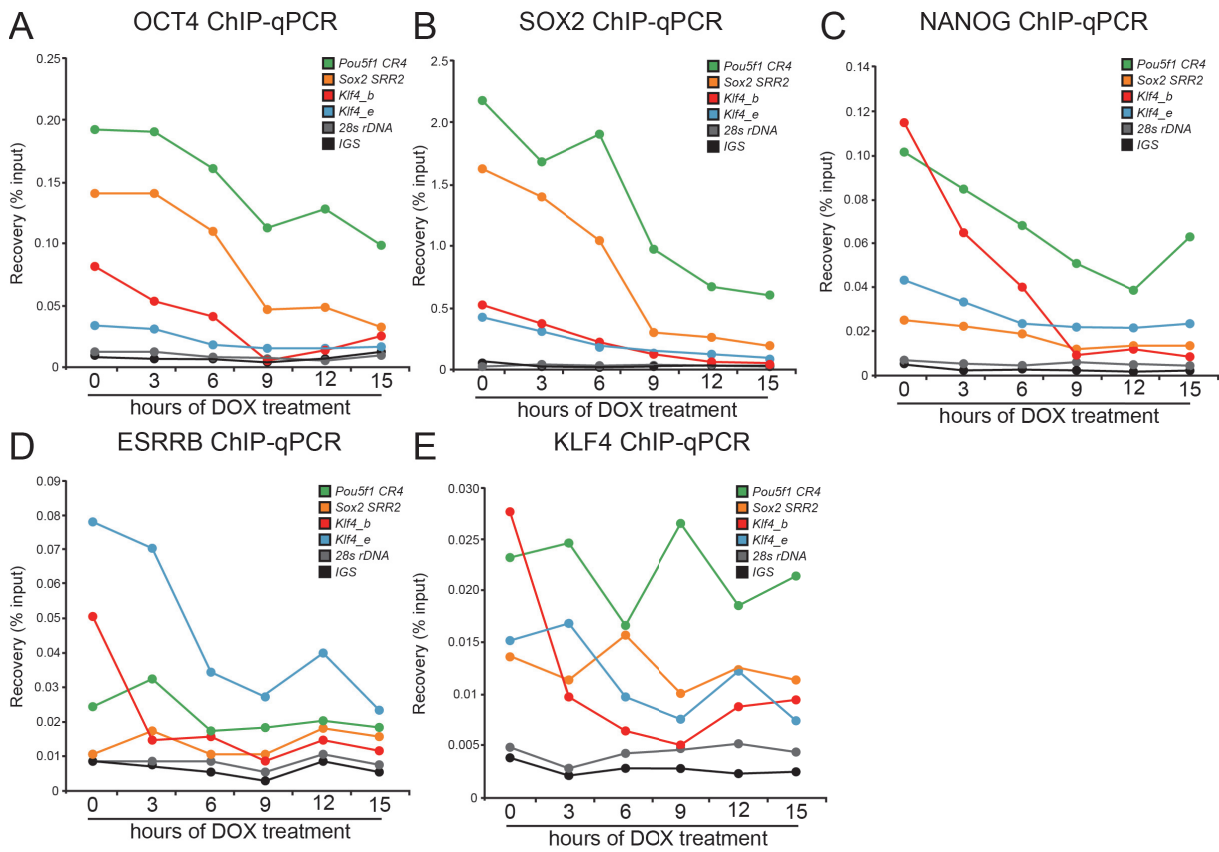
4. Discussion

4.1 Study rationale

OCT4 has been a focal point in stem cell research for a few decades and is known to be vital for mammalian development. It was demonstrated that slight alterations to OCT4 expression lead to rapid differentiation of mESC to primitive endoderm or throphectoderm (Niwa et al. 2000). There have also been studies trying to elucidate the direct target genes based on genome wide binding studies such as ChIP-seq (Kim et al. 2008; Kim et al. 2018; Xi Chen et al. 2008). Sharov et al. attempted to identify the direct OCT4 targets by application of the ZHBTc4 cells (the same cell model used in this thesis) over a time course of DOX treatment (Sharov et al. 2008). In their study microarray analysis was used to analyze transcriptional changes and publicly available ChIP-seq data was integrated to overlap these changes with OCT4 binding events. This study used a gene-based approach and displayed a lack of overlap between the binding events identified in the ChIP-seq data and their differentially expressed genes. Recent findings suggest that microarray analysis is not the right approach to detect relevant enhancer activity, i.e. most active enhancers transcribing eRNA. Therefore, this study provides a novel approach by integrating TT-seq and ChIP-seq data to elucidate the direct target enhancers of OCT4 and their contribution to target gene regulation and maintenance of pluripotency.

Recent studies have aimed to ascribe pioneer activity to OCT4 (Soufi et al. 2012; Soufi et al. 2014), whereas other studies have been arguing against OCT4 being a pioneer TF (Chronis et al. 2017; Li et al. 2017). Our study does not address the potential role for OCT4 as pioneer factor. However, integration of ATAC-seq data into the TT-seq/ChIP-seq data sets allowed us to gain insights into the role of OCT4 in maintenance of chromatin accessibility.

OCT4 has been shown to closely cooperate with SOX2 and NANOG in mESC (Reményi et al. 2003; Niwa 2007; Boyer et al. 2005; Loh et al. 2006; X Chen et al. 2008; Kim et al. 2008; Whyte et al. 2013). It is unclear whether OCT4 or SOX2 is leading this interaction. Integrative analysis of the genome-wide data sets allowed us to elucidate the role OCT4 plays in the OCT4-SOX2 cooperativity, as well as cooperativity with other TFs such as NANOG, KLF4 and ESRRB.

This study is part of an extensive collaboration between the laboratories of Prof. Dr Hans R. Schöler and Prof. Dr Patrick Cramer to address the aforementioned conflicts in our understanding of OCT4 and its role in maintenance of pluripotency in mESC.

The unpublished work presented in this thesis indicates that OCT4 maintains pluripotency in different ways, which appear to be modulated via TFBS and TF cooperativity. Genome-wide analysis of chromatin accessibility upon loss of OCT4 identified two different effects, OCT4 loss of binding accompanied by loss of chromatin accessibility, or OCT4 loss of binding that does not lead to loss of chromatin accessibility. It also appears that these effects depend on TF cooperativity. The analysis of OCT4-bound sites with eRNA synthesis for changes to chromatin accessibility and eRNA synthesis dynamics determined four distinct clusters. The integration of TF binding events with these four clusters revealed cluster-dependent cooperativity changes and variability in the TF motifs.

Taken together, this data provides novel insights into the maintenance of pluripotency by OCT4 and the cooperativity of OCT4 with other TFs. To the best of my comprehension this is the first study combining the novel TT-seq method with ATAC-seq and published ChIP-seq data in an OCT4 loss of function model.

I will discuss my results in the context of previously published findings and examine the key findings in light of present dogmas regarding OCT4 and its role in pluripotency.

4.2 Rationale for use of TT-seq

TF are modulators of transcriptional activity via their DNA-binding properties and recruitment capacity of co-factors that initiate further activation or suppression of corresponding genes. RNA-seq is the most widely used method to study fluctuations in TF activity and to address transcriptional changes. However, RNA-seq relies on the presence of previously transcribed transcripts and does not take the half life and abundance of RNA molecules into account. More specifically, RNA-seq might miss lowly expressed and rapidly degraded RNA transcripts. It has been documented that eRNAs are among the transcripts that are generally missed in RNA-seq analysis due to their rapid turn-over (Djebali et al. 2012; Lam et al. 2013). eRNAs are proposed to reflect the activity of the corresponding enhancer (Mikhaylichenko et al. 2018; Khan et al. 2018; Kim et al. 2015) TF are known to bind to enhancers and therefore a read-out of potential TF activity at a corresponding binding site could be the active

transcription of eRNAs. A variety of techniques are available for the analysis of eRNA synthesis levels, such as GRO-seq, PRO-seq, mNET-seq, CAGE-seq, 4-sU-seq, and TT-seq. GRO-, PRO-, and mNET-seq are complicated techniques that rely on nuclear isolation and substitution of nucleotides with labelled nucleotides to detect nascent RNA. The CAGE-seq method relies on the capture of capped RNA molecules. We opted to use the TT-seq method, which - like 4sU-seq - is based on a brief exposure of live cell culture to 4-thiouridine(4sU) which is incorporated into actively synthesized RNA molecules (**Figure 4**). TT-seq, unlike 4sU-seq, includes a sonication step before nascent RNA enrichment to remove RNA fragments that were already transcribed prior to application of 4sU.

The comparison of the RNA-seq and TT-seq methods shows that TT-seq is more sensitive than RNA-seq (**Figure 7a**). More up-regulated genes were found at earlier time points in the TT-seq data set. Strikingly, the TT-seq method documented down-regulated genes as early as 6 hours and identified more down-regulated genes at 9 and 12 hours of DOX treatment (**Figure 7a**). Expression levels of mRNA are a result of added, newly synthesized mRNA and subtracted, degraded mRNA from pre-existing mRNA levels. TT-seq is advantageous regarding the analysis of down-regulated genes as it omits pre-existing mRNA, which can cover up the changes in mRNA synthesis. Another measure to illustrate the strength of TT-seq is the enrichment for intronic reads (**Figure 9c**). More intronic reads were found in TT-seq data, which supports the discovery of sense- and anti-sense intragenic ncRNAs such as eRNA.

In summary, we chose TT-seq to investigate RNA synthesis. More specifically, TT-seq allowed for the detection of eRNAs and therefore the assessment of enhancer activity. Combining TT-seq analysis with an OCT4 loss-of-function model allowed for the precise analysis of eRNA kinetics upon loss of OCT4 and thereby the investigation of OCT4-controlled enhancers.

4.3 Failed ChIP-seq and rationale for using publicly available data

TF activity can also be analyzed by genome-wide association studies utilizing techniques such as ChIP-seq, DamID-seq, SELEX-seq and DIP-seq. ChIP-seq provides insights into genomic TF binding events in a given cell type under specific conditions. ChIP-seq experiments detect a very large number of binding events, but it

remains to be determined, which binding events and corresponding sites are relevant in which cell type.

Many ChIP-seq studies have been published to identify binding events of pluripotency factors in mESC. For example, it has been shown that the *Pou5f1 CR4* is bound by almost all pluripotency TFs but also some negative regulators such as HDAC2 and LSD1 (Boyer et al. 2005; X Chen et al. 2008; Loh et al. 2006; Whyte et al. 2012). As *Pou5f1 CR4* is one of the strongest enhancers in mESC these findings suggest that binding per se does not infer activity of the bound factor.

Unfortunately, in this study the ChIP-seq experiments in ZHBTc4 cells failed. This might be due to 'over-sonication' of the chromatin and potential loss of recognizable epitopes for the antibodies. The SOX2 ChIP-seq data sets showed a higher number of called peaks than the OCT4 ChIP-seq data sets (**Table 19**). Comparison to a published SOX2 ChIP-seq data set (King & Klose 2017) revealed an overlap of 80% of the peaks so that optimization of the current protocol is likely to improve sample quality. Nevertheless, incorporation of publicly available OCT4, SOX2 and NANOG ChIP-seq data in the same cell line at 0 and 24 hours of Dox treatment allowed for the analysis of OCT4 binding and overlap with transcriptional activity. Strikingly, only 17% of the 17,719 OCT4 binding events coincided with nearby differential transcriptional activity (**Figure 39a**). These findings emphasize that binding per se does not infer activity.

We further integrated ChIP-seq data for Med1 in ZHBTc4 cells (Whyte et al. 2013), as well as ChIP-seq data sets for ESRRB, KLF4, cMYC, p300, H3K27ac, H3K4me1, H3K4me2 and H3K4me3 in C57BL/6J cells (Chronis et al. 2017).

Integration of TT-seq and ATAC-seq dynamics and overlap with publicly available ChIP-seq data sets provided us with the relevant genome-wide data to investigate changes to the pluripotency regulatory network upon loss of OCT4.

4.4 TF cooperativity

Cooperativity of TF is classically defined by the binding of one TF that can affect the binding of another TF (Carey 1998; Kim & Maniatis 1997; Thanos & Maniatis 1995; Tjian & Maniatis 1994). In relation to enhancers, TF cooperativity can be accommodated via different mechanisms. For example, TFs can bend the DNA in order to expose other TFBS (Falvo et al. 1995), recruit other TF via protein-protein interactions (Johnson et al. 1979), or 'co-recruit' other co-factor complexes such as

Mediator (Merika et al. 1998). Our results suggest that TF cooperativity happens in a context-dependent manner. OCT4-bound sites that overlap with eRNA synthesis, promoter TUs or NTU (**Figure 39a**) showed different levels of TF occupancy. OCT4-bound sites with eRNA synthesis displayed the highest levels of TF cooperativity (**Figure 40a-h**). In contrast, OCT4-bound sites without eRNA synthesis had lower levels of TF occupancy.

OCT4 is known to be a cooperative transcription factor, which means that it prefers to bind genomic loci together with other transcription factors in a context-dependent manner. In mESC, OCT4 prefers to bind to the DNA together with SOX2 at the OCTSOX motif, also referred to as the composite motif (Reményi et al. 2003; Niwa 2007). Moreover, NANOG, the third key pluripotency factor, often binds in close proximity to the OCT4-SOX2 heterodimer (Boyer et al. 2005; Loh et al. 2006; X Chen et al. 2008; Kim et al. 2008; Whyte et al. 2013). These observations underline that cooperativity among the pluripotency TF strongly impacts their binding capacity and therefore their role regarding lineage specification. In the context of reprogramming of mouse embryonic fibroblasts, it is was shown that distinct configurations of the OCTSOX motif are more important for the induction and maintenance of pluripotency then other OCTSOX motifs (Tapia et al. 2015).

Clustering of OCT4 sites based on eRNA synthesis kinetics and changes to chromatin accessibility followed by *de novo* motif discovery was performed and 4 distinct clusters were identified (**Figure 41 and 44**).

Clusters I and II indicated the presence of a weak OCT4 consensus sequence with a higher probability for the POU$_{hd}$ half site. (**Figure 44a**). In contrast, clusters III and IV depicted an even probability for the presence of the entire OCT4 motif.

Since OCT4 and SOX2 cooperate, an extended *de novo* motif discovery was performed (**Figure 44c**). Clusters I and II showed the presence of the SOX2 motif, which was surrounded by 'T-rich' regions. These 'T-rich' regions are known to be preferred by the POU$_{hd}$ half site of OCT4. Interestingly, clusters III and IV demonstrated the presence of the OCTSOX composite motif. (**Figure 44d**). Binding of SOX2 in clusters III and IV was largely affected by the loss of OCT4, which could mean that OCT4 is important in the maintenance of SOX2 binding to the OCTSOX composite motif.

Overall, we found interesting differences among the discovered motifs in the four clusters. These findings might help to explain why loci listed in clusters I and II are less sensitive to loss of OCT4 than loci represented in clusters III and IV.

Next, we analyzed OCT4, SOX2 and NANOG occupancy in the four clusters at 0 hours and 24 hours of DOX treatment (**Figure 41** and **Figure 45**). SOX2 and NANOG binding was not affected at OCT4-bound loci that did not show altered chromatin accessibility (**Figure 37**). However, we found differences in SOX2 and NANOG binding after loss of OCT4 in the different clusters based on changes to chromatin accessibility and transcriptional activity (**Figure 45**). In cluster I loss of OCT4 binding did not greatly impact the binding intensity of SOX2 and NANOG (**Figure 45a, e, i**). Our results might suggest that NANOG binds downstream of the OCT4 sites in the presence of OCT4 and switches to more upstream binding sites after loss of OCT4 (**Figure 45i**). SOX2 and NANOG binding intensities were increased in cluster II, but strongly decreased in clusters III and IV (**Figure 45**).

The analysis of the binding intensities and the motif analysis suggest that TF cooperativity between OCT4, SOX2 and NANOG is highest at OCT4-bound sites with the OCTSOX composite motif (clusters III and IV, **Figure 44** and **45**). These findings are in line with previously published data (Tapia et al. 2015; Jauch et al. 2011). Interestingly, our findings might also suggest that OCT4 binding is important for the maintenance of SOX2 binding to the composite motif, which is in contrast to published data (Chen et al. 2014). This does not mean that SOX2 is not important for the recruitment of OCT4. The importance of SOX2 for OCT4-recruitment has been highlighted in many studies. Molecular dynamic simulations have suggested that SOX2 can change the DNA binding preference of OCT4 (Merino et al. 2014). The binding domain of OCT4 consists of two domains of which the POU_{hd} is important for the scanning of the genome and searching for its binding partners. When OCT4 encounters SOX2 at the OCTSOX motif, the binding preferences of OCT4 switch from the POU_{hd} to the POU_{sp} domain (Merino et al. 2014). In support of this notion, OCT1 shows a similar phenomenon (Takayama & Clore 2012a; Takayama & Clore 2012b; Doucleff & Clore 2008). In addition, an EMSA-based study suggested that SOX2 is important for OCT4 recruitment to the composite motif (Aksoy et al. 2013). However and despite the leading role for SOX2 in recruitment of OCT4 to OCTSOX sites our data indicates that OCT4 is the leading TF in the maintenance of OCT4-SOX2-NANOG cooperativity (**Figure 44 and 45**). In contrast, it was shown that SOX2 is expendable for composite motif activation. SOX2 is important for the regulation of TF that maintain sufficient OCT4 levels for pluripotency (Masui et al. 2007).

Overall, these pieces of data show that TF cooperativity is maintained by the binding of OCT4 to the OCTSOX motif. Perturbations to OCT4 lead to a substantial loss of SOX2 and NANOG at the respective loci in clusters III and IV. In contrast, loss of OCT4 binding at sites only bound by OCT4 in clusters I and II did not lead to loss of SOX2 and NANOG binding.

4.5 Enhancer activity and TF occupancy

Classically, enhancers are defined as cis-regulatory elements that provide temporal-spatial control of gene expression. Enhancers are defined by presence of TFBS, specific histone modifications such as H3K27ac and H3K4me1 and presence of co-factors such as Med1 and p300. It was also suggested that enhancers that are active show RNA synthesis in the form of eRNA (Ørom & Shiekhattar 2013; Kim et al. 2010; de Santa et al. 2010; Pulakanti et al. 2013). Moreover, the level of eRNA synthesis and the directionality of transcription is thought to reflect the level of activity of the enhancer (Mikhaylichenko et al. 2018).

Super-enhancers were computationally defined by enhancer elements that are in close proximity to each other and display high levels of occupancy of co-factor Med1 and histone modifications such as H3K27ac (Whyte et al. 2013). In the context of mESC, super-enhancers show high levels of OSN occupancy (Whyte et al. 2013). Metagene analysis at OCT4-bound sites defined in **Figure 39a**, showed distinct patterns regarding H3K27ac, p300 and Med1 occupancy. OCT4-bound sites with eRNA synthesis displayed higher occupancy of Med1 than OCT4-bound sites without eRNA synthesis (**Figure 40g**). These sites also displayed slightly higher levels of p300, a histone acetyl transferase that can acetylate Histone 3 lysine 27 (H3K27) and much higher levels of H3K27ac (**Figure 40h and i**). Levels of H3Km4e1 are also thought to indicate enhancer activity. We found that OCT4-bound sites without transcriptional activity showed a clear bi-modular distribution of H3K4me1, whereas transcriptionally active OCT4-bound sites display a different distribution which appears less bi-modular (**Figure 40j**). These observations support the notion that transcriptionally active enhancers have higher levels of TF occupying their TFBS, higher occupancy of Med1 and p300, and higher levels of histone modifications such as H3K27ac (Whyte et al. 2013).

Next, we analyzed transcriptionally active OCT4-bound sites by integration of the TT-seq data into the OCT4 ChIP-seq data. Our Metagene analysis revealed that OCT4,

SOX2, NANOG, ESRRB, KLF4, Med1 and p300 all prefer to bind OCT4-bound sites that overlap with eRNA synthesis (**Figure 39a**). These findings argue for a stronger TF cooperativity at these loci in comparison to sites without transcriptional activity.

Clustering analysis of the transcriptionally active, OCT4-bound sites was performed based on eRNA synthesis kinetics and chromatin dynamics (**Figure 41**). Cluster III displayed a simultaneous decrease in chromatin accessibility and eRNA synthesis, with the synthesis loss being stronger. Cluster IV depicted a stronger decrease in the accessibility of the chromatin, which is followed by a decrease in eRNA synthesis. *De novo* motif discovery for OCT4 and SOX2 in these four clusters supported that the aforementioned notion that OCT4-SOX2 cooperativity is highly dependent on the DNA context. *As discussed in section 4.5*, this is substantiated by the enrichment of the OCTSOX composite motif being highly enriched in clusters III and IV (**Figure 44**). Our findings support the observations that SOX2 is dispensable for the activation of OCTSOX composite motif enhancers (Masui et al. 2007).

Preliminary analysis of the binding kinetics at four enhancer elements by ChIP-qPCR complies with our genome-wide data sets (**Figure 48**). OCT4 binding was lost at *Pou5f1 CR4, Sox2 SRR2* and *Klf4_e* enhancer elements (**Figure 48a**). SOX2 followed OCT4 loss of binding with similar kinetics, feeding into the premises that OCT4 binding is essential for SOX2 occupancy (**Figure 48b**). Moreover, NANOG binding at the *Pou5f1 CR4* and *Klf4_b* enhancers was lost slightly earlier than OCT4 and SOX2 (**Figure 48c**). Binding of ESRRB at the *Klf4_b* and *Klf4_e* enhancer elements decreased rapidly after loss of OCT4 (**Figure 48d**). Interestingly, KLF4 occupancy at the *Pou5f1 CR4* and *SOX2 SRR2* was stable across the time course of DOX treatment (**Figure 48e**). Together these preliminary ChIP-qPCR results documented that loss of TF occupancy takes place with similar kinetics. However, these results need to be validated.

Taken together these results indicate that transcriptionally active enhancers show higher occupancy of the pluripotency transcription factors. This could potentially translate into higher cooperativity among the pluripotency TF. In addition, our results suggest that transcriptionally active enhancers with the OCTSOX composite motif are more sensitive to loss of OCT4 than enhancers without this composite motif.

4.6 Billboard model versus enhanceosome model

Recently, many studies have attempted to elucidate the molecular mechanisms underlying enhancer activity. It is still unclear to what extent the underlying DNA sequence plays a role in the enhancer activity. There are different enhancer models with the 'flexible billboard model' at one and the enhanceosome model at the other end of the spectrum (Thanos & Maniatis 1995; Long et al. 2016; Ambrosetti et al. 2000; Arnosti & Kulkarni 2005). The billboard model relies on imperfect TFBS and accumulation of different TFs performing redundant roles. The enhanceosome model relies on the presence of perfect TFBS; slight alterations to one TF or the TFBS lead to disruption of the enhanceosome (Arnosti & Kulkarni 2005).

In mESC, it was shown that binding of SOX2 followed by recruitment of OCT4 leads to the establishment of the enhanceosome (Chen et al. 2014).

Analysis of the four clusters of transcriptionally active OCT4-bound sites based on eRNA kinetics and alterations to chromatin accessibility (**Figure 41**) provided some potential insights into the enhancer models. Our data supports both the enhanceosome and the 'flexible billboard' model in a context-dependent manner, as described below.

Analysis of the four clusters with regard to their TF binding intensities of OCT4, SOX2 and NANOG at 0 and 24 hours of DOX treatment displayed two separate effects. Clusters I and II showed no loss of SOX2 and NANOG; clusters III and IV demonstrated a substantial loss of SOX2 and NANOG binding (**Figure 45**). Based on the fact that clusters I and II lacked the perfect composite motif (**Figure 44**), one could argue that the enhancers found in clusters I and II belong to the billboard model of enhancers. The perfect composite motif combined with the loss of SOX2 and NANOG binding upon loss of OCT4, as found in clusters III and IV, could indicate that these enhancers belong to the enhanceosome model. This is also supported by the fact that loss of one TF, in this case OCT4, leads to collapse of the entire enhancer architecture.

In conclusion, loss of OCT4 only led to loss of enhancer activity in the context of the composite OCTSOX motif whereas sites that showed separate binding for OCT4 and SOX2 were not affected by loss of OCT4. These observations support the idea that enhancers in both the billboard and enhanceosome model are targeted by OCT4.

4.7 Chromatin dynamics and TF occupancy upon loss of OCT4

A groundbreaking discovery was published in 2006 demonstrating that overexpression the four TF OCT4, SOX2, KLF4, and cMYC, could revert a somatic cell back into an induced pluripotent stem cell state (Takahashi & Yamanaka 2006; Takahashi et al. 2007). In this original TF cocktail OCT4 is the only TF that cannot be substituted by one of its family members. This is due to the unique preference of OCT4 for heterodimerization (as discussed in section 4.5). However, it was shown that OCT6, when mutated to promote cooperativity with SOX2, can replace OCT4 in the cocktail of ectopically expressed TFs. Thereby, mutated OCT6 is capable of induction of pluripotency, albeit with low efficiency (Jerabek et al. 2017). Interestingly, OCT4 loses its ability to form iPSC when it is mutated to prefer homodimeriziation (as OCT6). These findings indicate that the function of POU factors is mediated via their dimerization preferences which are due to small alterations in their amino acid make-up. In reprogramming, OCT4 binding alone is important for silencing somatic genes at early stages of reprogramming. At later stages OCT4 needs SOX2 to fulfill the establishment of pluripotency (J. Chen et al. 2016; Chronis et al. 2017).

Controversially, our lab has shown that ectopic OCT4 is obsolete in the reprogramming cocktail while endogenous OCT4 remains important for locking in and the maintenance of pluripotency. This was demonstrated by reprogramming of MEF to iPSC with only SOX2, KLF4 and cMYC. The process of reprogramming was much slower in the absence of OCT4, suggesting that OCT4 might be involved in driving proliferation at the early phases of reprogramming. Unexpectedly, the iPSC quality of cells reprogrammed with SKM is better than the quality of OSKM reprogrammed cells (unpublished data from lab).

OSKM need to access closed and silenced chromatin to establish a genomic environment susceptible for the induction of pluripotency. It was postulated that one or more of these TF possess the capacity for pioneer activity. So far, there are different findings supporting pioneer activity for OCT4 (You et al. 2011; Soufi et al. 2012; Soufi et al. 2014). However, there are also reports demonstrating that OCT4 does not possess pioneer activity (Sherwood et al. 2014; Chronis et al. 2017).

OCT4 is capable of generating nucleosomal depleted regions (NDR). Loss- and gain-of-functions studies have demonstrated that loss of OCT4 leads to phasing of the nucleosome into and therefore loss of the NDR and that re-expression of OCT4 leads to re-formation of NDR (You et al. 2011). In addition, OCT4 preferentially targets

DNaseI resistant chromatin during early stages of somatic cell reprogramming (Soufi 2012) and it was documented that OCT4 binds nucleosomal DNA in EMSA (Soufi et al. 2014). On the other hand, OCT4 is not among the TF described to be capable of binding DNaseI insensitive chromatin (Sherwood et al. 2014). Lack of pioneer activity of OCT4 was indicated by genome-wide analysis, e.g. ChIP-seq, at different time points during reprogramming (Chronis et al. 2017). In early reprogramming phases OCT4, SOX2, and KLF4 bind pluripotency enhancers in closed chromatin, but these binding events do not lead to activation of the target regions (Chronis et al. 2017). Rather, OSK bind somatic enhancer to silence these regions by recruiting HDAC1. The early OSK-bound pluripotency enhancers need the presence of another transcription factors such as ESRRB for activation. Another study showed that OCT4 or SOX2 or KLF4 cannot open chromatin alone but need to work in synergy to open chromatin needed for establishing pluripotency program in somatic cells (Li et al. 2017).

In contrast, it was demonstrated that recruitment of BRG1 by OCT4 is essential for co-occupancy of TF at OCT4 sites (King & Klose 2017). Loss-of-function experiments showed that OCT4 is needed for the accessibility and functional regulation of distal regulatory elements in mESC. At these distal regulatory elements OCT4 supports the binding of other pluripotency factors such as SOX2 and NANOG to control gene expression. However, OCT4 alone is insufficient as shown by BRG1 loss-of-function studies which lead to loss of chromatin accessibility and loss of TF occupancy.

Our data indicates that there is a substantial change to chromatin accessibility upon loss of OCT4 (**Figure 22a**). More specifically, we show that decrease and increase of chromatin accessibility occurred around 9 hours after of DOX treatment (**Figure 22b**). The decrease in chromatin accessibility can be divided into three clusters based on its kinetics. Cluster I showed a strong decrease in accessibility between 3 and 12 hours of DOX treatment, cluster II demonstrated a steady loss of chromatin accessibility, and cluster III displayed a steady loss starting at 3 hours of DOX treatment (**Figure 23a**). It is surprising that the alterations occurred so early since binding of OCT4 did not seem to be globally affected at these early time points (**Figure 3a**). In contrast, there are also many sites with increased chromatin accessibility upon loss of OCT4. This increase either plateaus at 9 hours or shows a steady increase until the last time point (**Figure 23b**). Next, we integrated OCT4 ChIP-seq data to determine which changes coincide with OCT4 binding (**Figure 24**). The majority of sites with decreased chromatin accessibility overlapped with OCT4

binding. Only a very small number of sites with increased chromatin accessibility showed an overlap with OCT4 occupancy. These findings support the observations that OCT4 is important for maintaining accessible chromatin, perhaps via recruitment of BRG1 (King & Klose 2017).

Next, in-depth analysis of OCT4-bound sites with and without alterations to chromatin accessibility was performed (**Figure 37**). OCT4-bound sites with a decrease in chromatin accessibility showed a substantial loss of OCT4 binding. Similarly, there is almost a complete loss of OCT4 binding to sites that do not show changes to chromatin accessibility (**Figure 37a and b**). SOX2 and NANOG occupancy was lost at OCT4-bound sites with altered chromatin accessibility (**Figure 37c and e**). In contrast, OCT4-bound sites with unchanged chromatin accessibility showed a minimal loss of SOX2 and a slight increase in NANOG binding (**Figure 37 d and f**). Our data suggests that SOX2 and NANOG can maintain chromatin accessibility in the absence of OCT4 at specific loci. This is in sharp contrast to the observation that OCT4 is the TF recruiting chromatin remodelers to maintain chromatin accessibility (King & Klose 2017).

Motif analysis of these two clusters revealed only the OCT4 motif in OCT4-bound regions with unchanged chromatin accessibility (**Figure 38**). OCT4-bound sites with a substantial decrease of chromatin accessibility showed a strong enrichment for the composite motif (**Figure 38**). This could mean that, in the context of OCT4 binding together with SOX2 at the OCTSOX composite motif, OCT4 is leading in maintaining accessible chromatin.

Next, the changes in chromatin accessibility at OCT4-bound sites overlapping with eRNA synthesis, promoter synthesis or absence of transcriptional activity were analyzed. This analysis revealed that OCT4-bound sites overlapping with promoter TUs are less sensitive to changes in chromatin accessibility (F**igure 39b**) with only 5% of all sites showing changes. Surprisingly, OCT4 sites that show eRNA synthesis or no transcriptional activity seem to be equally affected in terms of changes to chromatin accessibility upon loss of OCT4. Metagene analysis of OCT4-bound sites overlapping with promoter TUs, eRNA TUs or NTU displayed more accessible chromatin at promoter TUs and eRNA TUs than at NTU sites (**Figure 40m**).

We described four distinct clusters based on chromatin dynamics and eRNA synthesis dynamics (**Figure 41**). We found that loci with decreased chromatin accessibility show enrichment for the OCTSOX composite motif, whereas regions

with unaltered chromatin accessibility lacked enrichment for the OCTSOX motif (**Figure 44**).

We confirm earlier findings that SOX2 and NANOG binding is retained in clusters with unchanged chromatin accessibility (**Figure 45 e, f, i, j**). However, in regions with decreased chromatin accessibility SOX2- and NANOG-binding was also lost (**Figure 45 g, h, k, l**). Recently it was shown that NANOG uses multiple different mechanisms in maintenance of pluripotency and self-renewal (Heurtier et al. 2018). These mechanisms include recruitment of other pluripotency TF, such as OCT4, SOX2 and ESRRB to enhancers, thereby fostering chromatin accessibility. Additionally, NANOG is capable of recruiting BRG1 to mediate chromatin accessibility. NANOG can also act as a repressor and actively prevent binding of OCT4 and SOX2.

Together our results confirm published data that OCT4 is important for chromatin accessibility (King & Klose 2017; Heurtier et al. 2018). However, we find that this is only true in a context-dependent manner. In the presence of the OCTSOX composite motif, OCT4 is the main contributor to chromatin accessibility. In contrast, loss of OCT4 binding at the OCT4 motif does not lead to changes in chromatin accessibility, transcriptional activity or occupancy of SOX2 and NANOG. The data suggests an important role for OCT4 in maintenance of chromatin accessibility at sites where OCT4 cooperates with SOX2 on the OCTSOX composite motif.

4.8 mESC Gene regulatory networks

The cellular gene regulatory network is fascinating. In particular, pluripotent cells need to maintain a pluripotent state but also a chromatin landscape that allows for rapid differentiation upon signal cues (Young 2011; Li & Izpisua Belmonte 2018). The pluripotent gene regulatory network has been extensively studied and it has been confirmed that OCT4, SOX2 and NANOG are at the core of the network (Zhou et al. 2007; Chambers & Smith 2004; Niwa 2007; Silva & Smith 2008).

The importance of OCT4 was demonstrated by the fact that subtle alterations to the expression levels lead to loss of pluripotency (Niwa et al. 2000). RNAi experiments of OCT4 have a far more detrimental effect than knockdown of other TFs (Ivanova et al. 2006). Hence, OCT4 appears to be more essential than other TFs.

Microarray data of a OCT4, SOX2, and NANOG loss-of-function models overlapped with public ChIP-seq data sets attempted to identify direct OCT4 target genes (Sharov et al. 2008). They discovered only a small overlap of TF target sites and

differentially gene expression. However, TF often target enhancers and the analysis does not include ncRNA (Sharov et al. 2008). Our data showed that the majority of OCT4 binding events do not overlap with transcriptional activity (**Figure 39a**). There could be four possible explanations for this finding.

It is possible that the majority of OCT4 binding events do not lead to further recruitment of co-factors such as histone acetyltransferases and Mediator. In addition, it is well conceivable that not all of the ~18000 OCT4 binding event found in the ChIP-seq data are functional. In comparison, another study listed with high confidence 1,083 OCT4 binding events (Loh et al. 2006). Therefore, in the event of overlapping ChIP-seq data with other data sets it is important to determine which binding sites are actually functional and what their function is. We found an overlap of OCT4-bound sites and a decrease in chromatin accessibility upon loss of OCT4 (**Figure 24**). OCT4 could be involved in the chromatin structure and architecture at some of these sites.

The small overlap between OCT4-bound sites and transcriptional activity could be due to insufficient sequencing depth, which might result from the increased duplicate levels in replicate 1 (**Table 18**). These duplicate levels complicate the computational analysis of eRNA synthesis. We identified about 50% of the previously described super-enhancers in mESC, which might be due to lower sequencing depth (Whyte et al. 2013).

Previous studies implied that all active enhancers synthesize eRNA (Kim et al. 2010). Our study suggests that perhaps not all active enhancers synthesize eRNA. It is likely that a combination of the above-mentioned explanations applies: the majority of OCT4 binding events are not functional, our analysis misses some differentially synthesized eRNA, and some OCT4 binding events at enhancers are functional but do not lead to synthesis of eRNA.

4.9 Typical enhancers versus super-enhancers

Most protein-coding genes are under the control of enhancers (typical enhancers) that modulate the proper expression levels to maintain cellular homeostasis. Cell identity genes are thought to be under the control of clusters of enhancers that are computationally defined as super-enhancers. Super-enhancers show higher levels of TF, histone modifications and co-factors than typical enhancers. This allows for a stronger influence on the enhancement of their target gene expression (Chapuy et al.

2013; Whyte et al. 2013; Hnisz et al. 2013; Lovén et al. 2013). Super-enhancers are among the active enhancers that show eRNA synthesis (Pefanis et al. 2015; Schmidt et al. 2015).

The analysis of all differentially expressed eRNA revealed no clear trend regarding their kinetics upon loss of OCT4. Not all differentially expressed eRNA had paired protein-coding genes with similar kinetics (**Figure 30a and b**). However, OCT4-targeted enhancers with differentially synthesized eRNA showed a better correlation regarding differential expression of their target genes (**Figure 31a and b**).

In our data we were able to document about 50% of the previously described mESC super-enhancers (Whyte et al. 2013). Interestingly, there is no clear effect regarding loss of eRNA originating from super-enhancers and corresponding loss of paired protein-coding genes (**Figure 35).**

As expected, we found differentially expressed eRNA originating from super-enhancers that are linked to differentially expressed pluripotency genes such as *Klf4, Klf5, Utf1* and *Tbx3* (**Figure 36b**). Surprisingly, there are a number of differentially expressed protein-coding genes that are targeted by super-enhancers not showing differential expression (**Figure 34**). Among those genes are pluripotency factors such as *Sall4, Nr5a2, Nanog,* and *Esrrb*.

Super-enhancers show high levels of TF; they are more sensitive to disruption by loss of TF than typical enhancers and so are their paired protein-coding genes (Whyte et al. 2013). We found that OCT4-bound sites with eRNA synthesis have higher TF occupancy than OCT4-bound enhancers without transcriptional activity (**Figure 40**). Among these OCT4-bound sites we found some differences taking the dynamics of RNA synthesis and chromatin accessibility upon loss of OCT4 into account (**Figure 41**). OCT4 showed the highest binding preference for sites in clusters III and IV, whereas SOX2, NANOG, KLF4, Med1 did not follow this trend. ESRRB was the only other TF showing a binding preference for clusters III and IV albeit with less difference to clusters I and II than OCT4 (**Figure 42a-f**). Our data indicates that SOX2 and NANOG are more sensitive to loss of OCT4 in clusters where the eRNA synthesis is impacted (clusters III and IV, **Figure 45**).

4.10 eRNA synthesis

Enhancer activity is thought to be quantifiable by its level of eRNA expression. The level of eRNA transcription and the directionality of the transcripts could even be a

readout for the strength of an enhancer (Mikhaylichenko et al. 2018). It was postulated that highly active enhancers can act as promoters; with unidirectional eRNA acting as the strongest promoters and strong bidirectional promoters actually functioning as enhancers (Mikhaylichenko et al. 2018). This is further supported by the observation that promoters and enhancers have been repurposed over the course of evolution (Carelli et al. 2018). However, the notion that eRNAs are bi-directionally transcribed has been challenged (Scruggs et al. 2015; Young et al. 2017).

Enhancers with eRNA synthesis also show pausing and elongation control despite RNAPII being less stable at enhancers compared to promoters. Enhancers with high levels of transcriptional activity have higher levels of H3K4me3 than H3K4me1 (Henriques et al. 2018). It remains unclear whether or not eRNAs are functionally relevant for the enhancers. So far only a limited number of papers describe eRNA functions. eRNAs have been proposed to act as decoy in order to sequester negative transcriptional regulators such as NELF to foster transcriptional elongation (Schaukowitch et al. 2014). Additionally, it was demonstrated that CBP/p300 is recruited to sites of eRNA synthesis by eRNA in order to establish acetylated histones (H3K27ac) and stimulation of transcription (Bose et al. 2017).

Interestingly, we found that OCT4-bound sites with active RNA synthesis showed the highest occupancy of p300 (**Figure 40h**). Consequentially, higher levels of H3K27ac were found at OCT4-bound sites with eRNA synthesis (**Figure 40i**). In addition, p300 was mostly enriched at sites without eRNA synthesis and chromatin accessibility upon loss of OCT4 (**Figure 42g**). However, in this context this did not necessarily translate to higher levels of H3K27ac as shown by metagene analysis (**Figure 42j**).

It would be interesting to determine whether there are eRNAs in mESC that are functionally relevant to the maintenance of enhancer activity and thereby the control of the pluripotent stem cell identity. Further analysis of our data sets might enable us to determine whether eRNAs are functionally important for enhancer activity.

4.11 Divergent culturing conditions

mESC can be cultured in different cell culture medium compositions, for example serum/LIF or 2i/LIF (Smith et al. 1988; Williams et al. 1988; Silva et al. 2008). The serum/LIF condition promotes a more primed pluripotent state whereas 2i/LIF promotes naïve pluripotency (Marks et al. 2012; Ying et al. 2008; Marks &

Stunnenberg 2014; Bagci & Fisher 2013; Miyanari & Torres-Padilla 2012; Silva et al. 2009; Leitch et al. 2013). We cultured the ZHBTc4 cells in 2i/LIF medium whereas the publicly available data was acquired from cells propagated in serum/LIF medium (King & Klose 2017; Chronis et al. 2017).

OCT4-dependent and -independent chromatin accessibility changes (**Figure 37 and 38**) were analyzed with the publicly available OCT4, SOX2, and NANOG ChIP-seq data sets (King & Klose 2017) and our own ATAC-seq data set.

In the integrative analysis of the publicly available ChIP-seq data into our TT-seq and ATAC-seq data sets we assessed the overlap of OCT4-bound sites and eRNA synthesis (**Figure 39a**). Therefore, we only analyzed active enhancers bound by OCT4 and filtered out potential differences based on culturing conditions. However, we found that only 9% of all OCT4-bound sites overlap with eRNA synthesis. This small overlap could be an underestimation and due to divergent culturing conditions. Therefore, OCT4, SOX2 and NANOG ChIP-seq data obtained in ZHBTc4 cells cultured in 2i/LIF medium would further elucidate the effects of OCT4 loss in the mESC enhancer landscape.

4.12 OCT4-dependent and -independent control of enhancer activity

Taken together, we propose the following model for OCT4-dependent und –independent control of enhancer activity (**Figure 49**). Loss of OCT4 at OCT4-bound and eRNA-synthesizing sites showed OCT4-dependent and -independent effects (**Figure 39a**, **Figure 41a, Figure 44 and Figure 45**). OCT4-dependent enhancers displayed a loss of SOX2 and NANOG occupancy, a decrease in chromatin accessibility as well as a decrease in eRNA synthesis upon loss of OCT4 (**Figure 49a**). This observation suggests that OCT4 is essential for the maintenance and activity at these enhancers. More specifically, neither SOX2 and/or NANOG occupancy nor eRNA synthesis alone can maintain enhancer activity. This implies that OCT4 binding is necessary for the occupancy of SOX2 and NANOG. It will be important to analyze in a targeted approach, whether or not OCT4 can maintain binding at specific sites upon loss of SOX2 or NANOG. Additionally, it would be interesting to assess whether OCT4 can maintain eRNA synthesis at these sites in the absence of SOX2 or NANOG.
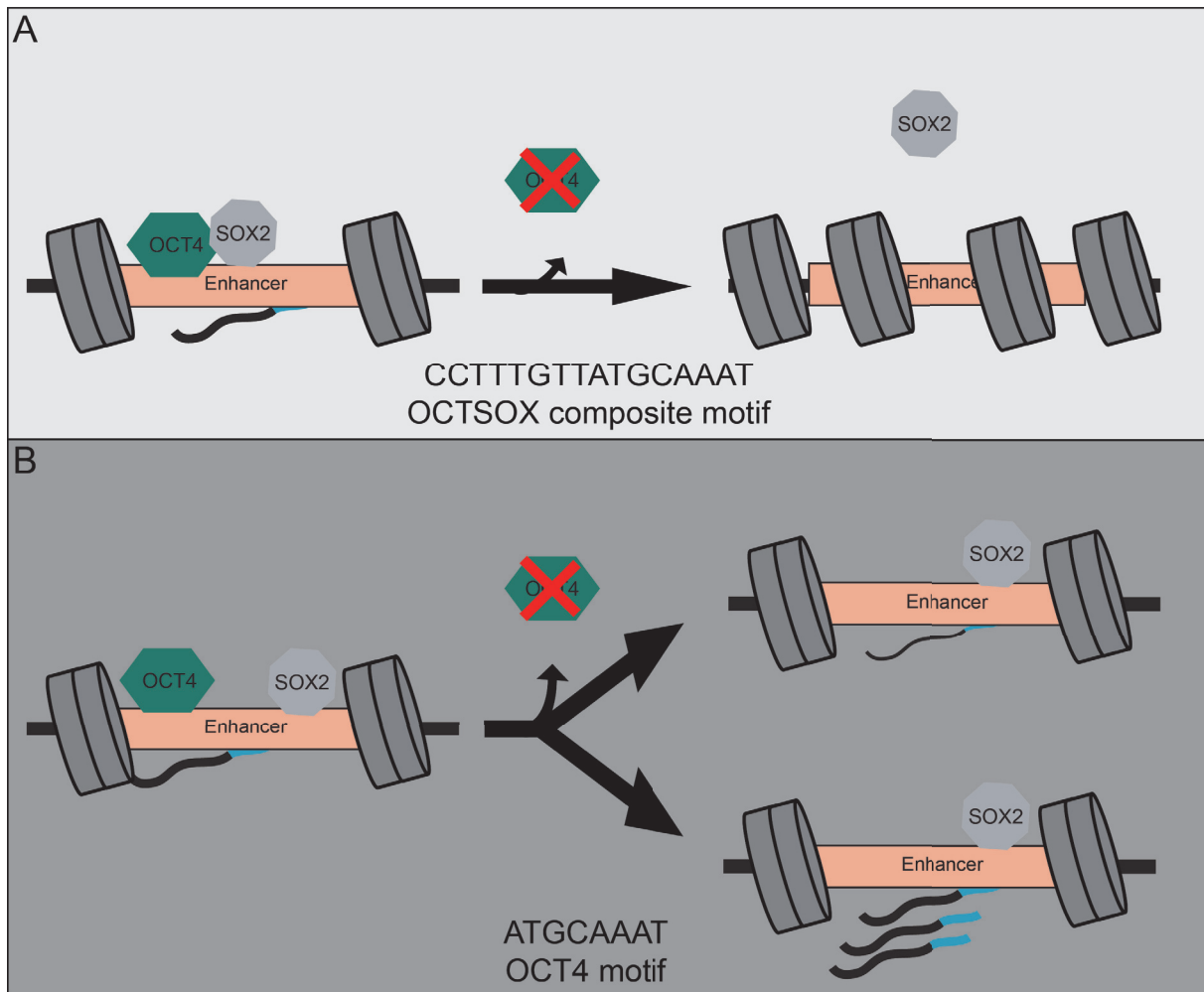
Figure 49 Model of OCT4-dependent and -independent enhancer control.

(**A**) OCT4-dependent enhancer control dependents on TF cooperativity which is mediated by the underlying DNA sequences. Loss of OCT4 at these enhancers leads to a loss of SOX2 and NANOG binding, reduction in the chromatin accessibility and loss of eRNA synthesis. (**B**) OCT4-independent enhancers show two effects upon loss of OCT4. First, SOX2 remains bound and eRNA synthesis is slowly decreasing. Second, SOX2 remains bound and the synthesis rate of eRNA is increased.

OCT4-bound sites not showing a decrease in chromatin accessibility upon loss of OCT4 remained bound by SOX2 and NANOG (**Figure 49b**). Synthesis of eRNAs at these sites either slightly decreased or showed an increase over the time course (**Figure 41a**, **Clusters I and II**). It will be important to analyze some OCT4-independent sites in a SOX2 or NANOG loss-of-function model to determine whether these transcriptional effects are indeed OCT4-independent or whether these sites rely on the presence of multiple TF. If the latter is the case this could mean that the synthesis of eRNA is the driver for enhancer activity and chromatin accessibility.

In addition to all described analyses, we have recently performed in-depth analysis of OCT4-bound sites that show alterations to chromatin accessibility and integration of RNA synthesis. These analyses presented a more detailed insight into the kinetics of eRNA synthesis and alterations to chromatin accessibility upon loss of OCT4. Enhancers regulating expression of pluripotency genes (e.g. *Esrrb*, *Nanog*, and *Klf4*) were more resistant to loss of OCT4 (**data not shown**). However, enhancers driving expression of support factors such as *Tfcp2l1* as well as genes encoding for histone modifying enzymes were among the first to be down-regulated upon loss of OCT4 (**data not shown**). Overall, we found that decrease in eRNA synthesis and chromatin accessibility as well as loss of OCT4 and SOX2 binding followed similar kinetics within each group of enhancers (**data not shown**). More specifically, OCT4, SOX2 and/or NANOG binding, chromatin accessibility, and eRNA synthesis were lost at varying speeds depending on the target gene of the specific enhancer (**data not shown**).

4.13 Summary

The data presented in this thesis elucidate the role of OCT4 in the maintenance of pluripotency. OCT4 is more essential at sites with the OCTSOX composite motif, where it cooperates with SOX2, than at sites, where OCT4 binds separate from SOX2. In addition, OCT4 binding is necessary at the OCTSOX composite motif for SOX2 to remain bound as well as for NANOG to bind.

Loss of OCT4 at the OCTSOX composite motif leads to a decrease in chromatin accessibility (**Figure 49a**). However, there is no effect on chromatin accessibility at loci where OCT4 binds separately from SOX2 and/or NANOG (**Figure 49b**). Curiously, our data shows that at the latter loci, SOX2 and NANOG, can maintain chromatin accessibility and even drive the synthesis of eRNAs. Further analysis of our data and additional target-based assays are needed to dissect the differences between typical and super-enhancers and assessing them at a gene-specific level.

4.14 Future perspective

In this thesis we have collected large amounts of data including TT-seq and ATAC-seq in an OCT4 loss-of-function model. These genome-wide data sets have given us some insights into OCT4's role in governing pluripotency by enhancer regulation.

ChIP-seq analysis of different TF over the OCT4 loss time course would be an important addition to our datasets. Especially OCT4, SOX2, NANOG, ESRRB and KLF4 genome-wide binding data would be valuable in order to determine the kinetics of TF loss upon loss of OCT4. Obtaining this data will allow for a thorough investigation of the order of events such as loss of TF binding, alterations to chromatin accessibility and RNA synthesis. In this thesis we focused on the effects of OCT4 loss. However, it would be instrumental to investigate the effect of SOX2 and NANOG loss to ascertain that OCT4 is the main and responsible TF in the demonstrated events. Improvement of the TT-seq data set, e.g. reduction of duplication levels, could help to discover more eRNA and other ncRNA to complement the existing data. It might allow us to define additional direct OCT4 target genes and regulatory elements and enable us to compile a comprehensive list of elements and genes sensitive to loss of OCT4. These genome-wide data sets will likely help us to establish trends and define target genes to elucidate some pending questions. Target based identification of genome-wide trends will be of major importance to confirm the observed effects. One possible approach is the use of CRISPR/Cas9 to eliminate individual enhancer elements and define which elements within a super-enhancer are more essential for maintenance of pluripotency than others. In addition, the use of RNAi-based approaches could dissect the necessity of specific eRNA.

5. Bibliography

Aksoy, I. et al., 2013. Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO Journal*, 32(7), pp.938–953.

Alarcón, C. et al., 2009. Nuclear CDKs Drive Smad Transcriptional Activation and Turnover in BMP and TGF-β Pathways. *Cell*, 139(4), pp.757–769.

Allan, C.M., Walker, D. & Taylor, J.M., 1995. Evolutionary duplication of a hepatic control region in the human apolipoprotein E gene locus. Identification of a second region that confers high level and liver-specific expression of the human apolipoprotein E gene in transgenic mice. *The Journal of biological chemistry*, 270(44), pp.26278–81.

Allison, L.A. et al., 1985. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell*, 42(2), pp.599–610.

Amabile, G. & Meissner, A., 2009. Induced pluripotent stem cells: current progress and potential for regenerative medicine. *Trends in Molecular Medicine*, 15(2), pp.59–68.

Ambrosetti, D.C. et al., 2000. Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the Fibroblast growth factor-4 enhancer. *Journal of Biological Chemistry*, 275(30), pp.23387–23397.

Ambrosetti, D.C., Basilico, C. & 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(11), pp.6321–6329.

Anders, S., Pyl, P.T. & Huber, W., 2015. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*.

Andersson, R. et al., 2014. An atlas of active enhancers across human cell types and tissues. *Nature*, 507, pp.455–461.

Andersson, R., 2015. Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *BioEssays*, 37(3), pp.314–323.

Apostolou, E. et al., 2013. Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. *Cell stem cell*, 12(6), pp.699–712.

Arnosti, D.N. & Kulkarni, M.M., 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *Journal of Cellular Biochemistry*, 94(5),

pp.890–898.

Atkin, N.B. et al., 1974. Chromosome studies on 14 near-diploid carcinomas of the ovary. *European Journal of Cancer (1965)*, 10(3), pp.143–146.

Aurora, R. & Herr, W., 1992. Segments of the POU domain influence one another's DNA-binding specificity. *Molecular and cellular biology*, 12(2), pp.455–67.

Avilion, A.A. et al., 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes and Development*, 17(1), pp.126–140.

Baek, H.J. et al., 2002. Requirement of TRAP/mediator for both activator-independent and activator-dependent transcription in conjunction with TFIID-associated TAF(II)s. *Molecular and cellular biology*, 22(8), pp.2842–52.

Bagci, H. & Fisher, A.G., 2013. Dna demethylation in pluripotency and reprogramming: The role of Tet proteins and cell division. *Cell Stem Cell*, 13(3), pp.265–269.

Banerji, J., Olson, L. & Schaffner, W., 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell*, 33(3), pp.729–40.

Banerji, J., Rusconi, S. & Schaffner, W., 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27(2 Pt 1), pp.299–308.

Barrero, M.J., Boué, S. & Izpisúa Belmonte, J.C., 2010. Epigenetic Mechanisms that Regulate Cell Identity. *Cell Stem Cell*, 7(5), pp.565–570.

Barski, A. et al., 2007. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, 129(4), pp.823–837.

Beagan, J.A. et al., 2017. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Research*, 27(7), pp.1139–1152.

Beauparlant, C. et al., 2018. metagene: A package to produce metagene plots.

Ben-Shushan, E. et al., 1995. A dynamic balance between ARP-1/COUP-TFII, EAR-3/COUP-TFI, and retinoic acid receptor:retinoid X receptor heterodimers regulates Oct-3/4 expression in embryonal carcinoma cells. *Molecular and cellular biology*, 15(2), pp.1034–1048.

van den Berg, D.L.C. et al., 2010. An Oct4-centered protein interaction network in embryonic stem cells. *Cell stem cell*, 6(4), pp.369–381.

Berge, D. Ten et al., 2011. Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nature Cell Biology*, 13(9), pp.1070–1077.

Bhutani, N., Burns, D.M. & Blau, H.M., 2011. DNA demethylation dynamics. *Cell*, 146(6), pp.866–872.

Bilodeau, S. et al., 2009. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes and Development*, 23(21), pp.2484–2489.

Bird, A.P., 1986. CpG-Rich islands and the function of DNA methylation. *Nature*, 321(6067), pp.209–213.

Blackwood, E.M. & Kadonaga, J.T., 1998. Going the distance: A current view of enhancer action. *Science*, 281(5373), pp.60–63.

Boettiger, A.N. et al., 2016. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, 529(7586), pp.418–422.

Boroviak, T. et al., 2014. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nature Cell Biology*, 16(6), pp.513–525.

Bose, D.A. et al., 2017. RNA Binding to CBP Stimulates Histone Acetylation and Transcription. *Cell*, 168(1–2), p.135–149.e22.

Botquin, V. et al., 1998. New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2. *Genes and Development*, 12(13), pp.2073–2090.

Boyer, L. a et al., 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122(6), pp.947–56.

Boyer, L.A. et al., 2005. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell*, 122, pp.947–956.

Boyer, L.A. et al., 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, 441(7091), pp.349–353.

Bracken, A.P. & Helin, K., 2009. Polycomb group proteins: Navigators of lineage pathways led astray in cancer. *Nature Reviews Cancer*, 9(11), pp.773–784.

Brehm, a, Ohbo, K. & Schöler, H., 1997. The carboxy-terminal transactivation domain of Oct-4 acquires cell specificity through the POU domain. *Molecular and cellular biology*, 17(1), pp.154–162.

Briggs, R. & King, T.J., 1952. Transplantation of living nuclei from blastula cells into enucleated frogs' eggs. *Proceedings of the National Academy of Sciences*, 38(5), pp.455–463.

Brinster, R.L., 1974. The effect of cells transferred into the mouse blastocyst on subsequent development. *The Journal of experimental medicine*, 140(4), pp.1049–56.

Brownell, J.E. et al., 1996. Tetrahymena histone acetyltransferase A: A homolog to

yeast Gcn5p linking histone acetylation to gene activation. *Cell*, 84(6), pp.843–851.

Buecker, C. & Wysocka, J., 2012. Enhancers as information integration hubs in development: lessons from genomics. *Trends in Genetics*, 28(6), pp.276–284.

Buenrostro, J.D. et al., 2015. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, 2015(January), p.21.29.1-21.29.9.

Buganim, Y., Faddah, D. a & Jaenisch, R., 2013. Mechanisms and models of somatic cell reprogramming. *Nature reviews. Genetics*, 14(6), pp.427–39.

Burdon, T. et al., 1999. Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Developmental Biology*, 210(1), pp.30–43.

Burns, K.H., 2017. Transposable elements in cancer. *Nature Reviews Cancer*, 17(7), pp.415–424.

Calo, E. & Wysocka, J., 2013. Modification of enhancer chromatin: what, how, and why? *Molecular cell*, 49(5), pp.825–37.

Cande, J.D., Chopra, V.S. & Levine, M., 2009. Evolving enhancer-promoter interactions within the tinman complex of the flour beetle, Tribolium castaneum. *Development*, 136(18), pp.3153–3160.

Carelli, F.N. et al., 2018. Functional repurposing of regulatory element activity during mammalian evolution. *bioRxiv*.

Carey, M., 1998. The enhanceosome and transcriptional synergy. *Cell*.

Carson, S. & Wiles, M. V, 1993. Far upstream regions of class II MHC Ea are necessary for position-independent, copy-dependent expression of Ea transgene. *Nucleic acids research*, 21(9), pp.2065–72.

Catena, R. et al., 2004. Conserved POU binding DNA sites in the Sox2 upstream enhancer regulate gene expression in embryonic and neural stem cells. *Journal of Biological Chemistry*, 279(40), pp.41846–41857.

Chambers, I. et al., 2003. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, 113(5), pp.643–655.

Chambers, I. & Smith, A., 2004. Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene*, 23(43 REV. ISS. 6), pp.7150–7160.

Chapuy, B. et al., 2013. Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer cell*, 24(6), pp.777–90.

Chen, H. et al., 2015. Erk signaling is indispensable for genomic stability and self-renewal of mouse embryonic stem cells. *Proceedings of the National Academy of*

Sciences of the United States of America, 112(44), pp.E5936–E5943.

Chen, J. et al., 2016. Hierarchical Oct4 Binding in Concert with Primed Epigenetic Rearrangements during Somatic Cell Reprogramming. *Cell Reports*, 14(6), pp.1540–1554.

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

Chen, J., Zhang, Z., Li, L., Chen, B.C., Revyakin, A., Hajj, B., Legant, W., Dahan, M., Lionnet, T., et al., 2014. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell*, 156(6), pp.1274–1285.

Chen, K. et al., 2016. Gadd45a is a heterochromatin relaxer that enhances iPS cell generation. *EMBO reports*, 17(11), pp.1641–1656.

Chen, X. et al., 2008. Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell*, 133, pp.1106–1117.

Chen, X., Vega, V.B. & Ng, H.-H., 2008. Transcriptional regulatory networks in embryonic stem cells. *Cold Spring Harbor symposia on quantitative biology*, 73, pp.203–209.

Chew, J.-L. et al., 2005. Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4/Sox2 Complex in Embryonic Stem Cells. *Molecular and Cellular Biology*, 25(14), pp.6031–6046.

Chia, N.-Y. et al., 2010. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature*, 468(7321), pp.316–320.

Choudhury, M. & Ramsey, S.A., 2016. Identifying cell type-specific transcription factors by integrating chIP-seq and eQTL data-application to monocyte gene regulation. *Gene Regulation and Systems Biology*, 10, pp.105–110.

Chronis, C. et al., 2017. Article Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell*, 168(3), pp.1–18.

Ciosk, R. et al., 2000. Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Molecular Cell*, 5(2), pp.243–254.

Clapier, C.R. & Cairns, B.R., 2009. The biology of chromatin remodeling complexes. *Annual review of biochemistry*, 78, pp.273–304.

Cole, M.F. et al., 2008. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes and Development*, 22(6), pp.746–755.

Conaway, R.C. et al., 2005. The mammalian Mediator complex and its role in transcriptional regulation. *Trends in Biochemical Sciences*, 30(5), pp.250–255.

Consortium, I.H.G.S., 2001. Initial sequencing and analysis of the human genome. *Nature*, 409, pp.860–921.

Cramer, P. et al., 2008. Structure of eukaryotic RNA polymerases. *Annu Rev Biophys*, 37, pp.337–352.

Cremer, T. & Cremer, M., 2010. Chromosome territories. *Cold Spring Harbor perspectives in biology*, 2(3).

Cretekos, C.J. et al., 2008. Regulatory divergence modifies limb length between mammals. *Genes & development*, 22(2), pp.141–51.

Creyghton, M.P. et al., 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*, 107(50), pp.21931–21936.

Crocker, J. et al., 2015. Low affinity binding site clusters confer HOX specificity and regulatory robustness. *Cell*, 160(1–2), pp.191–203.

Davis, R.L., Weintraub, H. & Lassar, A.B., 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, 51(6), pp.987–1000.

Denholtz, M. et al., 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(5), pp.602–616.

Denholtz, M. & Plath, K., 2012. Pluripotency in 3D: Genome organization in pluripotent cells. *Current Opinion in Cell Biology*, 24(6), pp.793–801.

Diaz, P., Cado, D. & Winoto, A., 1994. A locus control region in the T cell receptor alpha/delta locus. *Immunity*, 1(3), pp.207–17.

Dietrich, J.-E. & Hiiragi, T., 2007. Stochastic patterning in the mouse pre-implantation embryo. *Development*, 134(23), pp.4219–4231.

Ding, J. et al., 2012. Oct4 links multiple epigenetic pathways to the pluripotency network. *Cell research*, 22(1), pp.155–67.

Ding, L. et al., 2015. Systems Analyses Reveal Shared and Diverse Attributes of Oct4 Regulation in Pluripotent Cells. *Cell Systems*, 1(2), pp.141–151.

Dixon, J.R. et al., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398), pp.376–380.

Djebali, S. et al., 2012. Landscape of transcription in human cells. *Nature*, 489(7414), pp.101–108.

Dobin, A. & Gingeras, T.R., 2015. Mapping RNA-seq Reads with STAR. *Current protocols in bioinformatics*.

Doucleff, M. & Clore, G.M., 2008. Global jumping and domain-specific intersegment

transfer between DNA cognate sites of the multidomain transcription factor Oct-1. *Proceedings of the National Academy of Sciences*.

Dowen, J.M. et al., 2014. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*, 159(2), pp.374–387.

Duncan, B.K. & Miller, J.H., 1980. Mutagenic deamination of cytosine residues in DNA. *Nature*, 287(5782), pp.560–561.

Duque, T. & Sinha, S., 2015. What Does It Take to Evolve an Enhancer? A Simulation-Based Study of Factors Influencing the Emergence of Combinatorial Regulation. *Genome Biology and Evolution*, 7(6), pp.1415–1431.

Efroni, S. et al., 2008. Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell*, 2(May), pp.437–447.

Esch, D. et al., 2013. A unique Oct4 interface is crucial for reprogramming to pluripotency. *Nat Cell Biol*, 15(3), pp.295–301.

Esnault, C. et al., 2008. Mediator-Dependent Recruitment of TFIIH Modules in Preinitiation Complex. *Molecular Cell*, 31(3), pp.337–346.

Evans, M.J. & Kaufman, M.H., 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature*, 292(5819), pp.154–156.

Falvo, J. V., Thanos, D. & Maniatis, T., 1995. Reversal of intrinsic DNA bends in the IFNβ gene enhancer by transcription factors and the architectural protein HMG I(Y). *Cell*.

Farley, E.K. et al., 2015. Suboptimization of developmental enhancers. *Science*, 350(6258), pp.325–328.

Faunes, F. et al., 2013. A membrane-associated -catenin/Oct4 complex correlates with ground-state pluripotency in mouse embryonic stem cells. *Development*, 140(6), pp.1171–1183.

Fazzio, T.G., Huff, J.T. & Panning, B., 2008. An RNAi Screen of Chromatin Proteins Identifies Tip60-p400 as a Regulator of Embryonic Stem Cell Identity. *Cell*, 134, pp.162–174.

Fazzio, T.G. & Panning, B., 2010. Condensin complexes regulate mitotic progression and interphase chromatin structure in embryonic stem cells. *Journal of Cell Biology*, 188(4), pp.491–503.

Feng, J. et al., 2006. The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes and Development*, 20(11), pp.1470–1484.

Feschotte, C., 2008. Transposable elements and the evolution of regulatory

networks. *Nature Reviews Genetics*, 9(5), pp.397–405.

Fisher, A.G., 2002. Cellular identity and lineage choice. *Nature Reviews Immunology*, 2(12), pp.977–982.

Fouse, S.D. et al., 2008. Promoter CpG Methylation Contributes to ES Cell Gene Regulation in Parallel with Oct4/Nanog, PcG Complex, and Histone H3 K4/K27 Trimethylation. *Cell Stem Cell*, 2(2), pp.160–169.

Frankel, N. et al., 2011. Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature*, 474(7353), pp.598–603.

Fraser, P. & Grosveld, F., 1998. Locus control regions, chromatin activation and transcription. *Current Opinion in Cell Biology*, 10(3), pp.361–365.

Friend, W.C. et al., 1992. Cell-specific expression of high levels of human S100 beta in transgenic mouse brain is dependent on gene dosage. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 12(11), pp.4337–46.

Frum, T. et al., 2013. Oct4 Cell-autonomously promotes primitive endoderm development in the mouse blastocyst. *Developmental Cell*, 25(6), pp.610–622.

Fryer, C.J., White, J.B. & Jones, K.A., 2004. Mastermind recruits CycC:CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover. *Molecular Cell*, 16(4), pp.509–520.

Fuda, N.J., Ardehali, M.B. & Lis, J.T., 2009. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature*, 461(7261), pp.186–192.

Fulton, D.L. et al., 2009. TFCat: The curated catalog of mouse and human transcription factors. *Genome Biology*, 10(3), p.R29.

Gangaraju, V.K. & Lin, H., 2009. MicroRNAs: Key regulators of stem cells. *Nature Reviews Molecular Cell Biology*, 10(2), pp.116–125.

Gao, S. et al., 2009. Ubiquitin Ligase Nedd4L Targets Activated Smad2/3 to Limit TGF-β Signaling. *Molecular Cell*, 36(3), pp.457–468.

Gardner, R.L., 1985. Clonal Analysis of Early Mammalian Development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 312(1153), pp.163–178.

Gardner, R.L. & Rossant, J., 1979. Investigation of the fate of 4-5 day post-coitum mouse inner cell mass cells by blastocyst injection. *Journal of Embryology and Experimental Morphology*, 52, pp.141–152.

Gaspar-Maia, A. et al., 2009. Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature*, 460(7257), pp.863–868.

Gaspar-Maia, A. et al., 2011. Open chromatin in pluripotency and reprogramming. *Nature Reviews Molecular Cell Biology*, 12(1), pp.36–47.

Gershenzon, N.I. & Ioshikhes, I.P., 2005. Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics*, 21(8), pp.1295–1300.

Gertz, J. et al., 2013. Distinct properties of cell-type-specific and shared transcription factor binding sites. *Molecular Cell*, 52(1), pp.25–36.

Ghisletti, S. et al., 2010. Identification and Characterization of Enhancers Controlling the Inflammatory Gene Expression Program in Macrophages. *Immunity*, 32(3), pp.317–328.

Gillies, S.D. et al., 1983. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell*, 33(3), pp.717–28.

Goldberg, A.D. et al., 2010. Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. *Cell*, 140(5), pp.678–691.

Gorkin, D.U., Leung, D. & Ren, B., 2014. The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell*, 14(6), pp.771–775.

Gross, D., 1988. Nuclease Hypersensitive Sites In Chromatin. *Annual Review of Biochemistry*, 57(1), pp.159–197.

Grosveld, F. et al., 1987. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*, 51(6), pp.975–85.

Guenther, M.G. et al., 2007. A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. *Cell*, 130(1), pp.77–88.

Guenther, M.G. & Young, R.A., 2010. Repressive transcription. *Science*, 329(5988), pp.150–151.

Guo, G. et al., 2010. Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote to Blastocyst. *Developmental Cell*, 18(4), pp.675–685.

Gurdon, J.B., 1962. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *Journal of embryology and experimental morphology*, 10(4), pp.622–40.

Guturu, H. et al., 2013. Structure-aided prediction of mammalian transcription factor complexes in conserved non-coding elements. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1632), p.20130029.

Hackett, J.A. & Azim Surani, M., 2014. Regulatory principles of pluripotency: From

the ground state up. *Cell Stem Cell*, 15(4), pp.416–430.

Harris, H., 1959. Turnover of nuclear and cytoplasmic ribonucleic acid in two types of animal cell, with some further observations on the nucleolus. *The Biochemical journal*, 73, pp.362–9.

Hawkins, R.D. et al., 2010. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell*, 6(5), pp.479–491.

He, Y. et al., 2013. Structural visualization of key steps in human transcription initiation. *Nature*, 495(7442), pp.481–486.

Heintzman, N.D. et al., 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459(7243), pp.108–112.

Henriques, T. et al., 2018. Widespread transcriptional pausing and elongation control at enhancers. *Genes and Development*, 32(1), pp.26–41.

Heurtier, V. et al., 2018. The molecular logic of Nanog-induced self-renewal. *bioRxiv*.

Hillman, N., Sherman, M.I. & Graham, C., 1972. The effect of spatial arrangement on cell determination during mouse development. *Journal of embryology and experimental morphology*, 28(2), pp.263–78.

Hnisz, D. et al., 2013. Super-enhancers in the control of cell identity and disease. *Cell*, 155(4), pp.934–47.

Ho, L. et al., 2009. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proceedings of the National Academy of Sciences of the United States of America*, 106(13), pp.5187–91.

Ho, L. & Crabtree, G.R., 2010. Chromatin remodelling during development. *Nature*, 463(7280), pp.474–484.

HOGAN, B. et al., 1977. Isolation of a human teratoma cell line which expresses F9 antigen. *Nature*, 270(5637), pp.515–518.

Holmberg, J. & Perlmann, T., 2012. Maintaining differentiated cellular identity. *Nature Reviews Genetics*, 13(6), pp.429–439.

Hong, N.A. et al., 1997. A targeted mutation at the T-cell receptor alpha/delta locus impairs T-cell development and reveals the presence of the nearby antiapoptosis gene Dad1. *Molecular and cellular biology*, 17(4), pp.2151–7.

Houbaviy, H.B., Murray, M.F. & Sharp, P.A., 2003. Embryonic stem cell-specific microRNAs. *Developmental Cell*, 5(2), pp.351–358.

Hsieh, C.-L. et al., 2014. Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proceedings of the National*

Academy of Sciences, 111(20), pp.7319–7324.

Hu, G. et al., 2009. A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes and Development*, 23(7), pp.837–848.

Huber, W., Toedling, J. & Steinmetz, L.M., 2006. Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics*.

Huet, J. et al., 1983. Archaebacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type. *The EMBO journal*, 2(8), pp.1291–1294.

Ivanova, N. et al., 2006. Dissecting self-renewal in stem cells with RNA interference. *Nature*, 442(7102), pp.533–538.

Jacobson, E.M. et al., 1997. Structure of Pit-1 POU domain bound to DNA as a dimer: Unexpected arrangement and flexibility. *Genes and Development*, 11(2), pp.198–212.

Jaenisch, R. & Young, R., 2008. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell*, 132(4), pp.567–82.

Jauch, R. et al., 2011. Conversion of Sox17 into a pluripotency reprogramming factor by reengineering its association with Oct4 on DNA. *Stem Cells*, 29(6), pp.940–951.

Jerabek, S. et al., 2017. Changing POU dimerization preferences converts Oct6 into a pluripotency inducer. *EMBO reports*, 18(2), pp.319–333.

Jiang, J. et al., 2008. A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nature Cell Biology*, 10(3), pp.353–360.

Jin, C. et al., 2009. H3.3/H2A.Z double variant–containing nucleosomes mark "nucleosome-free regions" of active promoters and other regulatory regions. *Nature Genetics*, 41(8), pp.941–945.

Johnson, A.D., Meyer, B.J. & Ptashne, M., 1979. Interactions between DNA-bound repressors govern regulation by the   phage repressor. *Proceedings of the National Academy of Sciences*.

Johnson, P.F. & McKnight, S.L., 1989. Eukaryotic Transcriptional Regulatory Proteins. *Annual Review of Biochemistry*, 58(1), pp.799–839.

Jones, P.A., 2012. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, 13(7), pp.484–492.

Junion, G. et al., 2012. A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell*, 148(3), pp.473–486.

Jurkowska, R.Z., Jurkowski, T.P. & Jeltsch, A., 2011. Structure and Function of Mammalian DNA Methyltransferases. *ChemBioChem*, 12(2), pp.206–222.

Kagey, M.H. et al., 2010. Mediator and cohesin connect gene expression and

chromatin architecture. *Nature*, 467(7314), pp.430–5.

Kahan, B.W. & Ephrussi, B., 1970. Developmental potentialities of clonal in vitro cultures of mouse testicular teratoma. *Journal of the National Cancer Institute*, 44(5), pp.1015–1036.

Kano, H. et al., 2009. L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes and Development*, 23(11), pp.1303–1312.

Khan, A., Mathelier, A. & Zhang, X., 2018. Super-enhancers are transcriptionally more active and cell-type-specific than stretch enhancers. *bioRxiv*, p.310839.

Kieffer-Kwon, K.R. et al., 2013. Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell*, 155(7), pp.1507–1520.

Kiesel, A. et al., 2018. The BaMM web server for de-novo motif discovery and regulatory sequence analysis. *Nucleic Acids Research*.

Kim, H.S. et al., 2018. Pluripotency factors functionally premark cell-type-restricted enhancers in ES cells. *Nature*.

Kim, J. et al., 2008. An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. *Cell*, 132, pp.1049–1061.

Kim, S.H. et al., 2014. ERK1 phosphorylates Nanog to regulate protein stability and stem cell self-renewal. *Stem Cell Research*, 13(1), pp.1–11.

Kim, T. et al., 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature*, 465(7295), pp.182–187.

Kim, T.K., Hemberg, M. & Gray, J.M., 2015. Enhancer RNAs: A class of long noncoding RNAs synthesized at enhancers. *Cold Spring Harbor Perspectives in Biology*, 7(1), pp.3–5.

Kim, T.K. & Maniatis, T., 1997. The mechanism of transcriptional synergy of an in vitro assembled interferon-β enhanceosome. *Molecular Cell*.

Kim, T.K. & Shiekhattar, R., 2015. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell*, 162(5), pp.948–959.

King, H.W. & 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, pp.1–24.

King, T.J. & Briggs, R., 1955. Changes in the Nuclei of Differentiating Gastrula Cells, As Demonstrated By Nuclear Transplantation. *Proceedings of the National Academy of Sciences of the United States of America*, 41(5), pp.321–5.

Klemm, J.D. & Pabo, C.O., 1996. Oct-1 POU domain-DNA interactions: Cooperative binding of isolated subdomains and effects of covalent linkage. *Genes and*

Development, 10(1), pp.27–36.

Klochendler-Yeivin, A. et al., 2000. The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression. *EMBO Reports*, 1(6), pp.500–506.

Koch, F. et al., 2011. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature Structural and Molecular Biology*, 18(8), pp.956–963.

Koche, R.P. et al., 2011. Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell*, 8(1), pp.96–105.

Kornberg, R.D., Series, N. & May, N., 1974. Chromatin Structure : A Repeating Unit of Histones and DNA Chromatin structure is based on a repeating unit of eight. *Society*, 184(4139), pp.868–871.

Krishnan, B.R., Jamry, I. & Chaplin, D.D., 1995. Feature mapping of the HLA class I region: Localization of the POU5F1 and TCF19 genes. *Genomics*, 30(1), pp.53–58.

Kuroda, T. et al., 2005. Octamer and Sox Elements Are Required for Transcriptional cis Regulation of Nanog Gene Expression. *Molecular and Cellular Biology*, 25(6), pp.2475–2485.

Lagrange, T. et al., 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes & Development*, 12(1), pp.34–44.

Lai, F. et al., 2013. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature*, 494(7438), pp.497–501.

Laity, J.H., Lee, B.M. & Wright, P.E., 2001. Zinc finger proteins: new insights into structural and functional diversity. *Current opinion in structural biology*, 11(1), pp.39–46.

Lam, C.S. et al., 2012a. DNA-dependent Oct4–Sox2 interaction and diffusion properties characteristic of the pluripotent cell state revealed by fluorescence spectroscopy. *Biochemical Journal*, 448(1), pp.21–33.

Lam, C.S. et al., 2012b. DNA-dependent Oct4–Sox2 interaction and diffusion properties characteristic of the pluripotent cell state revealed by fluorescence spectroscopy. *Biochemical Journal*, 448(1), pp.21–33.

Lam, M.T.Y. et al., 2014. Enhancer RNAs and regulated transcriptional programs. *Trends in Biochemical Sciences*, 39(4), pp.170–182.

Lam, M.T.Y. et al., 2013. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature*, 498(7455), pp.511–515.

Lambert, S.A. et al., 2018. The Human Transcription Factors. *Cell*, 172(4), pp.650–665.

Lan, X. & Pritchard, J.K., 2016. Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. *Science*, 352(6288), pp.1009–1013.

Langmead, B. & Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*.

Larochelle, S. et al., 2012. Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nature structural & molecular biology*, 19(11), pp.1108–15. Available at: http://dx.doi.org/10.1038/nsmb.2399.

Lawrence, M., Daujat, S. & Schneider, R., 2016. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends in Genetics*, 32(1), pp.42–56.

Lawson, K.A., Meneses, J.J. & Pedersen, R.A., 1991. Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development*, 113, pp.891–911.

Lee, H.J., Hore, T.A. & Reik, W., 2014. Reprogramming the methylome: Erasing memory and creating diversity. *Cell Stem Cell*, 14(6), pp.710–719.

Lee, J.T., 2009. Lessons from X-chromosome inactivation: Long ncRNA as guides and tethers to the epigenome. *Genes and Development*, 23(16), pp.1831–1842.

Lee, T.I. et al., 2006. Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. *Cell*, 125(2), pp.301–313.

van Leeuwen, H.C. et al., 1997. Linker length and composition influence the flexibility of Oct-1 DNA binding. *The EMBO journal*, 16(8), pp.2043–53.

Leitch, H.G. et al., 2013. Naive pluripotency is associated with global DNA hypomethylation. *Nature Structural and Molecular Biology*.

Lengner, C.J. et al., 2007. Oct4 Expression Is Not Required for Mouse Somatic Stem Cell Self-Renewal. *Cell Stem Cell*, 1(4), pp.403–415.

Levasseur, D.N. et al., 2008. Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. *Genes and Development*, 22(5), pp.575–580.

Levine, M. & Tjian, R., 2003. Transcription regulation and animal diversity. *Nature*, 424(6945), pp.147–151.

Li, B., Carey, M. & Workman, J.L., 2007. The Role of Chromatin during Transcription. *Cell*, 128, pp.707–719.

Li, D. et al., 2017. Chromatin Accessibility Dynamics during iPSC Reprogramming. *Cell Stem Cell*, 21(6), p.819–833.e6.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*.

Li, M. & Belmonte, J.C.I., 2017. Ground rules of the pluripotency gene regulatory network. *Nature Reviews Genetics*, 18(3), pp.180–191.

Li, M. & Izpisua Belmonte, J.C., 2018. Deconstructing the pluripotency gene regulatory network. *Nature Cell Biology*, 20(4), pp.382–392.

Li, R. et al., 2010. A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell stem cell*, 7(1), pp.51–63.

Li, W. et al., 2013. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, 498(7455), pp.516–520.

Li, W., Notani, D. & Rosenfeld, M.G., 2016. Enhancers as non-coding RNA transcription units : recent insights and future perspectives. *Nature Genetics Review*, 17(4), pp.207–223.

Li, Y. et al., 2014. CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. *PLoS ONE*, 9(12).

Lieberman-Aiden, E. et al., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950), pp.289–293.

Lin, Y.C. et al., 2012. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nature Immunology*, 13(12), pp.1196–1204.

Littlewood, T.D. & Evan, G.I., 1995. Transcription factors 2: helix-loop-helix. *Protein profile*, 2(6), pp.621–702.

Lluis, F. et al., 2008. Periodic activation of Wnt/beta-catenin signaling enhances somatic cell reprogramming mediated by cell fusion. *Cell stem cell*, 3(5), pp.493–507.

Loewer, S. et al., 2010. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nature Genetics*, 42(12), pp.1113–1117.

Loh, Y.H. et al., 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genetics*, 38(4), pp.431–440.

Long, H.K., Prescott, S.L. & Wysocka, J., 2016. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell*, 167(5), pp.1170–1187.

Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*.

Lovén, J. et al., 2013. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, 153(2), pp.320–34.

Lu, X. et al., 2014. The retrovirus HERVH is a long noncoding RNA required for

human embryonic stem cell identity. *Nature Structural & Molecular Biology*, 21(4), pp.423–425.

Lyashenko, N. et al., 2011. Differential requirement for the dual functions of β-catenin in embryonic stem cell self-renewal and germ layer formation. *Nature Cell Biology*, 13(7), pp.753–761.

Malik, S. & Roeder, R.G., 2005. Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem Sci*, 30(5), pp.256–263.

Marks, H., Kalkan, T., et al., 2012. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*, 149(Icm), pp.590–604.

Marks, H. & Stunnenberg, H.G., 2014. Transcription regulation and chromatin structure in the pluripotent ground state. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1839(3), pp.129–137.

Marson, A. et al., 2009. Connecting micro-RNA genes to the core transcriptional regulartory circuitry of embryonic stem cells. *NIH Public Access*, 134(3), pp.521–533.

Marson, A., Levine, S.S., et al., 2008. Connecting microRNA Genes to the Core Transcriptional Regulatory Circuitry of Embryonic Stem Cells. *Cell*, 134(3), pp.521–533.

Marson, A., Foreman, R., et al., 2008. Wnt Signaling Promotes Reprogramming of Somatic Cells to Pluripotency. *Cell Stem Cell*, 3(2), pp.132–135.

Martello, G. et al., 2012. Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell*, 11(4), pp.491–504.

Martin, G.R., 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences*, 78(12), pp.7634–7638.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., et al., 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biology*, 9(6), pp.625–635.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. a, et al., 2007. Pluripotency governed by Sox2 via regulation of Oct3 / 4 expression in mouse embryonic stem cells. *Nature Cell Biology*, 9(6), pp.625–635.

Matoba, R. et al., 2006. Dissecting Oct3/4-Regulated Gene Networks in Embryonic Stem Cells by Expression Profiling L. Tora, ed. *PLoS ONE*, 1(1), p.e26.

Matsuda, T. et al., 1999. STAT3 activation is sufficient to maintain an undifferentiated

state of mouse embryonic stem cells. *EMBO Journal*, 18(15), pp.4261–4269.

Meissner, A. et al., 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, 454(7205), pp.766–770.

Mercola, M. et al., 1983. Transcriptional enhancer elements in the mouse immunoglobulin heavy chain locus. *Science (New York, N.Y.)*, 221(4611), pp.663–5.

Merika, M. et al., 1998. Recruitment of CBP/p300 by the IFNβ enhanceosome is required for synergistic activation of transcription. *Molecular Cell*.

Merino, F. et al., 2014. Structural basis for the SOX-dependent genomic redistribution of OCT4 in stem cell differentiation. *Structure*, 22, pp.1274–1286.

Meshorer, E. et al., 2006. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Developmental Cell*, 10(1), pp.105–116.

Mikhaylichenko, O. et al., 2018. The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes and Development*, 32(1), pp.42–57.

Mikkelsen, T.S. et al., 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448(7153), pp.553–560.

Minucci, S. et al., 1996. Retinoic acid-mediated down-regulation of Oct3/4 coincides with the loss of promoter occupancy in vivo. *The EMBO journal*, 15(4), pp.888–99.

Mistri, T.K. et al., 2015. Selective influence of Sox2 on POU transcription factor binding in embryonic and neural stem cells. *EMBO reports*, 16(9), pp.1177–1191.

Mitsui, K. et al., 2003. The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*, 113(5), pp.631–642.

Miyanari, Y. & Torres-Padilla, M.E., 2012. Control of ground-state pluripotency by allelic regulation of Nanog. *Nature*.

Mullen, A.C. et al., 2011. Master transcription factors determine cell-type-specific responses to TGF-β signaling. *Cell*, 147(3), pp.565–576.

Nakauchi, H., Sudo, K. & Ema, H., 2001. Quantitative assessment of the stem cell self-renewal capacity. *Annals of the New York Academy of Sciences*, 938, pp.18-24; discussion 24–5.

Natoli, G. & Andrau, J.-C., 2012. Noncoding Transcription at Enhancers: General Principles and Functional Models. *Annual Review of Genetics*, 46(1), pp.1–19.

Newkirk, S.J. et al., 2017. Intact piRNA pathway prevents L1 mobilization in male meiosis. *Proceedings of the National Academy of Sciences*, 114(28), pp.E5635–E5644.

Nicholas, J.S. & Hall, B. V., 1942. Experiments on developing rats. II. The

development of isolated blastomeres and fused eggs. *Journal of Experimental Zoology*, 90(3), pp.441–459.

Nichols, J. et al., 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*, 95(3), pp.379–391.

Nishimoto, M. et al., 1999. The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Molecular and cellular biology*, 19(8), pp.5453–65.

Nishioka, N. et al., 2009. The Hippo Signaling Pathway Components Lats and Yap Pattern Tead4 Activity to Distinguish Mouse Trophectoderm from Inner Cell Mass. *Developmental Cell*, 16(3), pp.398–410.

Nitzsche, A. et al., 2011. RAD21 cooperates with pluripotency transcription factors in the maintenance of embryonic stem cell identity. *PLoS ONE*, 6(5).

Niwa, H. et al., 2009. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature*, 460(7251), pp.118–22.

Niwa, H., 2007. How is pluripotency determined and maintained? *Development*, 134(4), pp.635–646.

Niwa, H. et al., 2005. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell*, 123(5), pp.917–929.

Niwa, H. et al., 2002. Phenotypic Complementation Establishes Requirements for Specific POU Domain and Generic Transactivation Function of Oct-3/4 in Embryonic Stem Cells. *Molecular and Cellular Biology*, 22(5), pp.1526–1536.

Niwa, H. et al., 1998. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes & development*, 12(13), pp.2048–60.

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

Nora, E.P. et al., 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398), pp.381–385.

Ohnishi, Y. et al., 2014. Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nature Cell Biology*, 16(1), pp.27–37.

Okumura-Nakanishi, S. et al., 2005. Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. *Journal of Biological Chemistry*, 280(7), pp.5307–5317.

Olins, A.L. et al., 1974. Spheroid Chromatin Units (v Bodies) Spheroid Chromatin Units (v Bodies). , 183(4122), pp.330–332.

Ørom, U.A. & Shiekhattar, R., 2013. Long noncoding RNAs usher in a new era in the biology of enhancers. *Cell*, 154(6), pp.1190–3.

Ortiz, B.D. et al., 1997. Adjacent DNA elements dominantly restrict the ubiquitous activity of a novel chromatin-opening region to specific tissues. *The EMBO Journal*, 16(16), pp.5037–5045.

Paling, N.R.D. et al., 2004. Regulation of embryonic stem cell self-renewal by phosphoinositide 3-kinase-dependent signaling. *Journal of Biological Chemistry*, 279(46), pp.48063–48070.

Palstra, R.J. et al., 2003. The β-globin nuclear compartment in development and erythroid differentiation. *Nature Genetics*, 35(2), pp.190–194.

Pardo, M. et al., 2010. An Expanded Oct4 Interaction Network: Implications for Stem Cell Biology, Development, and Disease. *Cell Stem Cell*, 6(4), pp.382–395.

Parker, S.C.J. et al., 2013. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proceedings of the National Academy of Sciences*, 110(44), pp.17921–17926.

Pasini, D. et al., 2008. Regulation of stem cell differentiation by histone methyltransferases and demethylases. *Cold Spring Harbor Symposia on Quantitative Biology*, 73, pp.253–263.

Pasini, D. et al., 2004. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO Journal*, 23(20), pp.4061–4071.

Pefanis, E. et al., 2015. RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell*, 161(4), pp.774–789.

Pennacchio, L.A. et al., 2013. Enhancers: Five essential questions. *Nature Reviews Genetics*, 14(4), pp.288–295.

Pera, M.F. & Tam, P.P.L., 2010. Extrinsic regulation of pluripotent stem cells. *Nature*, 465(7299), pp.713–720.

Percharde, M. et al., 2018. Article A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity Article A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell*, pp.1–15.

Phillips-Cremins, J.E. et al., 2013. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, 153(6), pp.1281–1295.

Pikarsky, E. et al., 1994. Retinoic acid represses Oct-3/4 gene expression through several retinoic acid-responsive elements located in the promoter-enhancer region. *Molecular and cellular biology*, 14(2), pp.1026–1038.

Pleasure, S.J. & Lee, V.M. Y, 1993. NTera 2 Cells: A human cell line which displays

characteristics expected of a human committed neuronal progenitor cell. *Journal of Neuroscience Research*, 35(6), pp.585–602.

Polo, J.M. et al., 2012. A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell*, 151(7), pp.1617–32.

Pradeepa, M.M. et al., 2016. Histone H3 globular domain acetylation identifies a new class of enhancers. *Nature genetics*, (April).

Ptashne, M., 1986. Gene regulation by proteins acting nearby and at a distance. *Nature*, 322(6081), pp.697–701.

Pulakanti, K. et al., 2013. Enhancer transcribed RNAs arise from hypomethylated, Tet-occupied genomic regions. *Epigenetics*, 8(December), pp.1303–1320.

Qiu, D. et al., 2015. Klf2 and Tfcp2l1, Two Wnt/β-Catenin Targets, Act Synergistically to Induce and Maintain Naive Pluripotency. *Stem Cell Reports*, 5(3), pp.314–322.

Rada-Iglesias, A. et al., 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470(7333), pp.279–283.

Radman-Livaja, M. & Rando, O.J., 2010. Nucleosome positioning: How is it established, and why does it matter? *Developmental Biology*, 339(2), pp.258–266.

Ralston, A. & Rossant, J., 2008. Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. *Developmental Biology*, 313(2), pp.614–629.

Rao, S.S.P. et al., 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7), pp.1665–1680.

Rebeiz, M. et al., 2011. Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *Proceedings of the National Academy of Sciences*, 108(25), pp.10036–10043.

Reményi, A. et al., 2003. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes and Development*.

Reményi, A. et al., 2001. Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping. *Molecular Cell*, 8(3), pp.569–580.

Rhee, J.M. et al., 1998. Highly cooperative homodimerization is a conserved property of neural POU proteins. *Journal of Biological Chemistry*.

Richardson, S.R. et al., 2017. Heritable L1 retrotransposition in the mouse primordial germline and early embryo. *Genome Research*, 27(8), pp.1395–1405.

Rodda, D.J. et al., 2005. Transcriptional regulation of Nanog by OCT4 and SOX2. *Journal of Biological Chemistry*, 280(26), pp.24731–24737.

Roeder, R.G., 1998. Role of general and gene-specific cofactors in the regulation of eukaryotic transcription. In *Cold Spring Harbor Symposia on Quantitative Biology*. pp. 201–218.

Rosner, M.H. et al., 1990. A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature*, 345(6277), pp.686–692.

Rothschild, G. & Basu, U., 2017. Lingering Questions about Enhancer RNA and Enhancer Transcription-Coupled Genomic Instability. *Trends in Genetics*, 33(2), pp.143–154.

Ru Cao, L.W. et al., 2002. Role of histone H3 lysine 27 methylation in X-inactivation. *Science*, 298(November), pp.1039–1043.

Sakurai, H. et al., 1996. The yeast GAL11 protein binds to the transcription factor IIE through GAL11 regions essential for its in vivo function. *Proceedings of the National Academy of Sciences of the United States of America*, 93(18), pp.9488–92.

Samavarchi-Tehrani, P. et al., 2010. Functional genomics reveals a BMP-Driven mesenchymal-to-Epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell*, 7(1), pp.64–77.

de Santa, F. et al., 2010. A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biology*, 8(5).

Santos-Rosa, H. et al., 2002. Active genes are tri-methylated at K4 of histone H3. *Nature*, 419(6905), pp.407–411.

Sato, N. et al., 2004. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nature medicine*, 10(1), pp.55–63.

Schaukowitch, K. et al., 2014. Enhancer RNA facilitates NELF release from immediate early genes. *Molecular Cell*, 56(1), pp.29–42.

Schleif, R., 1992. DNA Looping. *Annual Review of Biochemistry*, 61(1), pp.199–223.

Schmidt, S.F. et al., 2015. Acute TNF-induced repression of cell identity genes is mediated by NFκB-directed redistribution of cofactors from super-enhancers. *Genome Research*.

Schnetz, M.P. et al., 2010. CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS Genetics*, 6(7), pp.1–15.

Schöler, H.R., Ruppert, S., et al., 1990. New type of POU domain in germ line-specific protein Oct-4. *Nature*, 344, pp.435–439.

Schöler, H.R., Dressler, G.R., et al., 1990. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *The EMBO journal*, 9(7), pp.2185–95.

Schoorlemmer, J. et al., 1994. Characterization of a negative retinoic acid response element in the murine Oct4 promoter. *Molecular and cellular biology*, 14(2), pp.1122–1136.

Schuettengruber, B. et al., 2009. Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. *PLoS Biology*, 7(1).

Schwalb, B. et al., 2016. TT-seq maps the human transient transcriptome. *Science*, 352(6290), pp.1225–1228.

Scruggs, B.S. et al., 2015. Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Molecular Cell*, 58(6), pp.1101–1112.

Sexton, T. et al., 2012. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, 148(3), pp.458–472.

Sharov, A.A. et al., 2008. Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. *BMC Genomics*, 9, pp.1–19.

Sheik Mohamed, J. et al., 2010. Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. *RNA*, 16(2), pp.324–337.

Sherwood, R.I. et al., 2014. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nature Biotechnology*, 32(2), pp.171–178.

Sigova, A. a. et al., 2015. Transcription factor trapping by RNA in Gene Regulatory Elements. *Science*, 350(6263), pp.978–982.

Silva, J. et al., 2009. Nanog Is the Gateway to the Pluripotent Ground State. *Cell*.

Silva, J. et al., 2008. Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biology*.

Silva, J. & Smith, A., 2008. Capturing Pluripotency. *Cell*, 132(4), pp.532–536.

Singhal, N. et al., 2010. Chromatin-remodeling components of the baf complex facilitate reprogramming. *Cell*, 141, pp.943–955.

Smale, S.T. & Kadonaga, J.T., 2003. The RNA Polymerase II Core Promoter. *Annual Review of Biochemistry*, 72(1), pp.449–479.

Smith, a G. et al., 1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature*, 336, pp.688–690.

Smith, R.P., Riesenfeld, S.J., et al., 2013. A compact, in vivo screen of all 6-mers

reveals drivers of tissue-specific expression and guides synthetic regulatory element design. *Genome Biology*, 14(7).

Smith, R.P., Taher, L., et al., 2013. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nature Genetics*, 45(9), pp.1021–1028.

Smith, Z.D., Sindhu, C. & Meissner, A., 2016. Molecular features of cellular reprogramming and development. *Nature Reviews Molecular Cell Biology*, 17(3), pp.139–154.

Soufi, A. et al., 2014. Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. *Cell*, 161(3), pp.555–568.

Soufi, A., Donahue, G. & Zaret, K.S., 2012. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell*, 151(5), pp.994–1004.

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

Spitz, F. & Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*, 13(9), pp.613–626. Available at: http://dx.doi.org/10.1038/nrg3207.

Spivakov, M. et al., 2012. Analysis of variation at transcription factor binding sites in Drosophila and humans. *Genome Biology*, 13(9).

Stephenson, R.O., Yamanaka, Y. & Rossant, J., 2010. Disorganized epithelial polarity and excess trophectoderm cell fate in preimplantation embryos lacking E-cadherin. *Development*, 137(20), pp.3383–3391.

Stevens, L. & Little, S., 1954. Spontaneous testicular tumors in an inbred strain of mice. *PNAS*, 40, pp.1080–1087.

Stewart, T.A. & Mintz, B., 1982. Recurrent germ-line transmission of the teratocarcinoma genome from the METT-1 culture line to progeny in vivo. *Journal of Experimental Zoology*, 224(3), pp.465–469.

Sturm, R. a & Herr, W., 1988. The POU domain is a bipartite DNA-binding structure. *Nature*, 336(6199), pp.601–604.

Su, M. et al., 2014. Evolution of Alu Elements toward Enhancers. *Cell Reports*, 7(2), pp.376–385.

Surface, L.E., Thornton, S.R. & Boyer, L.A., 2010. Polycomb group proteins set the stage for early lineage commitment. *Cell Stem Cell*, 7(3), pp.288–298.

Sved, J. & Bird, A., 1990. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proceedings of the National Academy of Sciences*, 87(12), pp.4692–4696.

Sylvester, I. & Schöler, H.R., 1994. Regulation of the Oct-4 gene by nuclear receptors. *Nucleic acids research*, 22(6), pp.901–11.

Taatjes, D.J., 2010. The human Mediator complex: A versatile, genome-wide regulator of transcription. *Trends in Biochemical Sciences*, 35(6), pp.315–322.

Taher, L. et al., 2011. Genome-wide identification of conserved regulatory function in diverged sequences. *Genome Research*, 21(7), pp.1139–1149.

Takahashi, K. et al., 2007. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, 131(5), pp.861–872.

Takahashi, K. & Yamanaka, S., 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126, pp.663–676.

Takayama, Y. & Clore, G.M., 2012a. Impact of protein/protein interactions on global intermolecular translocation rates of the transcription factors Sox2 and Oct1 between DNA cognate sites analyzed by z-exchange NMR spectroscopy. *Journal of Biological Chemistry*.

Takayama, Y. & Clore, G.M., 2012b. Interplay between minor and major groove-binding transcription factors Sox2 and Oct1 in translocation on DNA studied by paramagnetic and diamagnetic NMR. *Journal of Biological Chemistry*.

Tam, P.P.L. & Zhou, S.X., 1996. The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Developmental Biology*, 178(1), pp.124–132.

Tam, W.-L. et al., 2008. T-Cell Factor 3 Regulates Embryonic Stem Cell Pluripotency and Self-Renewal by the Transcriptional Control of Multiple Lineage Pathways. *Stem Cells*, 26(8), pp.2019–2031.

Tapia, N. et al., 2015. Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency. *Scientific Reports*, 5.

Tarkowski, A.K. & Wróblewska, J., 1967. Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage. *Journal of Embryology and Experimental Morphology*, 18(1), pp.155–180.

Taylor, J.S. & Raes, J., 2004. Duplication and Divergence: The Evolution of New Genes and Old Ideas. *Annual Review of Genetics*, 38(1), pp.615–643.

Tesar, P.J. et al., 2007. New cell lines from mouse epiblast share defining features

with human embryonic stem cells. *Nature*, 448(7150), pp.196–199.

Thakur, J.K. et al., 2008. A nuclear receptor-like pathway regulating multidrug resistance in fungi. *Nature*, 452(7187), pp.604–609.

Thanos, D. & Maniatis, T., 1995. Virus induction of human IFNβ gene expression requires the assembly of an enhanceosome. *Cell*, 83(7), pp.1091–1100.

Thomson, M. et al., 2011. Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell*, 145(6), pp.875–889.

Tjian, R. & Maniatis, T., 1994. Transcriptional activation: A complex puzzle with few easy pieces. *Cell*.

Tolhuis, B. et al., 2002. Looping and interaction between hypersensitive sites in the active β-globin locus. *Molecular Cell*, 10(6), pp.1453–1465.

Tomilin, A. et al., 2000. Synergism with the coactivator OBF-1 (OCA-B, BOB-1) is mediated by a specific POU dimer configuration. *Cell*, 103(6), pp.853–864.

Tsai, P.F. et al., 2018. A Muscle-Specific Enhancer RNA Mediates Cohesin Recruitment and Regulates Transcription In trans. *Molecular Cell*, 71(1), p.129–141.e8.

Vaquerizas, J.M. et al., 2009. A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics*, 10(4), pp.252–263.

Verrijzer, C.P. et al., 1992. The DNA binding specificity of the bipartite POU domain and its subdomains. *The EMBO journal*, 11(13), pp.4993–5003.

Vierstra, J. et al., 2014. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science*, 346(6212), pp.1007–1012.

Vieux-Rochas, M. et al., 2015. Clustering of mammalian *Hox* genes with other H3K27me3 targets within an active nuclear domain. *Proceedings of the National Academy of Sciences*, 112(15), pp.4672–4677.

Visel, A. et al., 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457(7231), pp.854–858.

Vojnic, E. et al., 2011. Structure and VP16 binding of the Mediator Med25 activator interaction domain. *Nature Structural and Molecular Biology*, 18(4), pp.404–410.

Wang, J. et al., 2018. YY1 Positively Regulates Transcription by Targeting Promoters and Super-Enhancers through the BAF Complex in Embryonic Stem Cells. *Stem Cell Reports*, 10(4), pp.1324–1339.

Wang, S. et al., 2016. Spatial organization of chromatin domains and compartments in single chromosomes. *Science*, 353(6299), pp.598–602.

Wang, Y. et al., 2013. Endogenous miRNA Sponge lincRNA-RoR Regulates Oct4,

Nanog, and Sox2 in Human Embryonic Stem Cell Self-Renewal. *Developmental Cell*, 25(1), pp.69–80.

Wang, Z. et al., 2012. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell*, 10(4), pp.440–454.

Wapinski, O.L. et al., 2013. Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell*, 155(3), pp.621–35.

Wei, Z. et al., 2013. Klf4 organizes long-range chromosomal interactions with the OCT4 locus inreprogramming andpluripotency. *Cell Stem Cell*, 13(1), pp.36–47.

Weintraub, A.S. et al., 2017. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell*, 171(7), p.1573–1588.e28.

Whyte, W.A. et al., 2012. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature*, 482(7384), pp.221–225.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., et al., 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, 153(2), pp.307–19.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., et al., 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, 153(2), pp.307–319.

Wiblin, A.E., 2005. Distinctive nuclear organisation of centromeres and regions involved in pluripotency in human embryonic stem cells. *Journal of Cell Science*, 118(17), pp.3861–3868.

Wijchers, P.J. et al., 2015. Characterization and dynamics of pericentromere-associated domains in mice. *Genome Research*, 25(7), pp.958–969.

Williams, R.L. et al., 1988. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature*, 336(6200), pp.684–687.

Williams, R.L. et al., 1988. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature*.

Wilusz, J.E., Sunwoo, H. & Spector, D.L., 2009. Long noncoding RNAs: Functional surprises from the RNA world. *Genes and Development*, 23(13), pp.1494–1504.

Wintjens, R. & Rooman, M., 1996. Structural Classification of HTH DNA-binding Domains and Protein – DNA Interaction Modes. *Journal of Molecular Biology*, 262(2), pp.294–313.

de Wit, E. et al., 2013. The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature*, 501, pp.227–31.

Wray, J. et al., 2011. Inhibition of glycogen synthase kinase-3 alleviates Tcf3

repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nature Cell Biology*, 13(7), pp.838–45.

Wu, G. et al., 2013. Establishment of totipotency does not depend on Oct4A. *Nature Cell Biology*, 15(9), pp.1089–1097.

Wu, Q. et al., 2006. Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *Journal of Biological Chemistry*, 281(34), pp.24090–24094.

Xie, H. et al., 2004. Stepwise reprogramming of B cells into macrophages. *Cell*, 117(5), pp.663–676.

Yeo, J.C. et al., 2014. Klf2 is an essential factor that sustains ground state pluripotency. *Cell Stem Cell*, 14(6), pp.864–872.

Yeom, Y.I. et al., 1996. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development*, 122(3), pp.881–894.

Yi, F. et al., 2011. Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. *Nature Cell Biology*, 13(7), pp.762–770.

Ying, Q.L. et al., 2003. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell*, 115, pp.281–292.

Ying, Q.L. et al., 2008. The ground state of embryonic stem cell self-renewal. *Nature*, 453(May), pp.519–523.

You, J.S. et al., 2011. OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. *Proceedings of the National Academy of Sciences*, 108(35), pp.14497–14502.

Young, R. a., 2011. Control of the embryonic stem cell state. *Cell*, 144(6), pp.940–954.

Young, R.S. et al., 2017. Bidirectional transcription initiation marks accessible chromatin and is not specific to enhancers. *Genome Biology*, 18(1).

Zacher, B. et al., 2017. Accurate promoter and enhancer identification in 127 ENCODE and roadmap epigenomics cell types and tissues by GenoSTAN. *PLoS ONE*.

Zaratiegui, M., Irvine, D. V. & Martienssen, R.A., 2007. Noncoding RNAs and Gene Silencing. *Cell*, 128(4), pp.763–776.

Zaret, K.S. & Carroll, J.S., 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes & Development*, 25(21), pp.2227–2241.

Zemojtel, T. et al., 2009. Methylation and deamination of CpGs generate p53-binding sites on a genomic scale. *Trends in Genetics*, 25(2), pp.63–66.

Zentner, G.E. et al., 2011. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Research*, 21, pp.1273–1283.

Zhang, J. et al., 2006. Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nature Cell Biology*, 8(10), pp.1114–1123.

Zhang, T., Cooper, S. & Brockdorff, N., 2015. The interplay of histone modifications - writers that read. *EMBO reports*, 16(11), pp.1467–1481. A

Zhang, X. et al., 2008. Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. *Journal of Biological Chemistry*, 283(51), pp.35825–35833.

Zhang, Y. et al., 2013. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature*, 504(7479), pp.306–310.

Zhang, Y. et al., 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biology*.

Zhao, J. et al., 2010. Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. *Molecular Cell*, 40(6), pp.939–953.

Zhong, X. & Jin, Y., 2009. Critical roles of coactivator p300 in mouse embryonic stem cell differentiation and Nanog expression. *Journal of Biological Chemistry*, 284(14), pp.9168–9175.

Zhou, Q. et al., 2007. A gene regulatory network in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences*, 104(42), pp.16438–16443.

Zhou, Q., Li, T. & Price, D.H., 2012. RNA Polymerase II Elongation Control. *Annual Review of Biochemistry*, 81(1), pp.119–143.