TRANSCRIPTIONAL NETWORKS VARIATIONS DURING CELL CYCLE PROGRESSION IN HUMAN EMBRYONIC STEM CELLS

Anna Osnato
Gonville & Caius College

Wellcome Trust - Medical Research Council Cambridge Stem Cell Institute
Department of Surgery
University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy
September 2017
DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

As stipulated by the Degree Committee of Clinical Medicine, this dissertation does not exceed 60,000 words in length (excluding figures, photographs, tables, appendices and bibliography).

Anna Osnato

September 2017

Cambridge
SUMMARY

Differentiation and cell cycle regulation in stem cell have a key function for embryonic development, organ homeostasis and tissue repair. Recent results have shown that these two mechanisms are intrinsically connected. Indeed, cell cycle machinery directly controls maintenance of pluripotency and initiation of differentiation. More precisely, the cell cycle regulator Cyclin D appears to control the transcriptional activity of Activin/Nodal signalling during progression of the cell cycle in human Embryonic Stem Cells (hESCs). As a consequence, hESCs can only differentiate into endoderm in the Early G1 phase when Cyclin Ds are expressed at low levels. These results show the mechanisms by which the cell cycle defines differentiation propensity of stem cells. However, these observations also imply the existence of interplays coordinating extracellular signalling pathways with the epigenetic state, chromatin structure and transcriptional networks during cell cycle progression and these mechanisms remain to be fully uncovered.

Here, I have utilised the FUCCI reporter system combined with ATAC-Seq to analyse chromatin dynamics during cell cycle progression in hESCs. Furthermore, I performed ChIP-Seq analyses to define the genomic location of transcriptional regulators during cell cycle progression as well as RNA-Seq to confirm variation in gene expression pattern. Integration of these data shows that the chromatin status in hESCs is highly dynamic and the core pluripotency transcription factors and epigenetic modifiers change genomic location during cell cycle progression. I also showed that hESCs in the Late G1 phase accumulate transcripts that are important for differentiation and development; therefore, indicating this phase represents a unique portion of the cell cycle for cell fate decisions.

Taken together, these results uncover that transcriptional networks are unexpectedly dynamic during the progression of cell cycle in stem cells. I hypothesise that these modifications are necessary to prime hESCs for different cell fate choices allowing a diversity of differentiation that is otherwise impossible. Overall these mechanisms underline the need to study transcriptional and epigenetic mechanisms in the dynamic context of the cell cycle and have major implications for adult tissue homeostasis and disease.
ACKNOWLEDGEMENTS

Learning how to look at the world around me with new eyes and open mind is what I constantly challenge myself to do, in order to become the person I want to be. It is hard, but it is the most rewarding exercise I have ever engaged in. We are the people we meet, the journeys we undertake, the knowledge we let in. The more we let the world in, the more we are gifted with fulfilment in return. But none of these can be done alone, because ‘No man is an island’.

A place like Cambridge gave me the possibility to infinitely expand my horizons, and pushed me to reach highs that I did not consider within my means before. I will be forever grateful to my mentor Ludovic, it is hard to express with words the gratitude for the unconditioned support that I felt as a part of his lab family. The inspiring science we endlessly discussed during our meetings taught me how to challenge my views, how to deal with failures and come out of them even stronger. This work exists because you believe in me.

The unique place where I pursued this work, the LRM, has given me infinite strength, passion for science, and, most importantly, beautiful friendships. To Will and Daniel, a bit of this work is also yours, the way you supported me during these 3 years, and kept me one piece during thesis writing, has been truly wonderful. Another special mention to Steph, an unlimited source of knowledge and support, you have silently built up confidence in myself and taught me how stand up for my ideas and my work. Another massive thank you to the LRM crew, the most beautiful colleagues I could have ever asked for, there is no bad day when you have people like these with you. Ultimately, I could not have made it without the daily support of my friends Loukia and Sarah: we are so different but compensate each other as I did not know it was possible with such different characters, I have learnt so much from you, and I will be forever grateful for that.

I want to heartily thank who made this possible: the Wellcome Trust for the funding, but most importantly Prof. Austin Smith, and the PhD committee, who gave me the chance to start this PhD journey, even before I fully knew what it meant. I want to thank especially Brian, who supported me during some rough days of my rotation year, and made me feel I always had a door to knock on in case of need. With them, I want to thank Prof. Viola Calabrò from the University of Napoli, as the very first to have believed in me, my first mentor and guide.

When you are a wandering entity like I am, friends are the only way to call a place home. And Cambridge has felt like home from day one, and for this, I have to
thank Gonville & Caius College. For giving me a home, smiling porters, safe shelter, and the most amazing people I have met in my entire life. To Lily, Helen, and all the sirens, I am lucky to call you friends, our friendship is worth a million blades, and it will forever be. To my friend Ben, you have been truly amazing in the past few weeks, and will always have a special place in my heart. To Lala, Dan, Jens, Tom, Roeland, Itxas0, Julia and all my closest friends who have made my time in Cambridge unique, I never felt alone, and I will be forever grateful to you for this.

Never forget those five back home, Silvia, Gaia, Federica, Chiara, Camilla. You have taught me what friendship is, and with you on my side I feel invincible. When you learn how to love, no distance matters and the strength that this has given me through this journey is truly invaluable. I have to thank you for this. The list of old friends is endless, and include the Sabbesi crazy bunch, my anchor during the academic years in Italy. Together with old friends, I would not be here writing this without Valeria, the first person I met in Cambridge and the one I am never going to let go. A wise guide, you helped me to navigate the mysterious Cambridge world from day one. You have taken the worst of me, and smiled back, I will be never grateful enough to whoever put you on my path, you have been my family when family was too far to reach even with thoughts. Thank you, my sister.

This bring us to the end, to the most important people of my life. Cristina, my sister, my opposite and necessary complement, thank you for bearing with me through all my downfalls. Mamma, your unconditioned love gives me the power to climb the highest peaks, effortlessly. Papà, I will be never grateful enough for your intellectual input, and moral stand. I hope to have a life as fulfilling as yours, you amazing family: you taught me the importance of being kind, honest, hard-working, and to always believe in myself. With you on my side, I know I can become whatever I want.

The last few words to Cambridge, to Paris, to Amsterdam, to Napoli, to Europe: I was lucky enough to call all these places home, to meet so many truly amazing people along the way, and if I could I would thank them all one by one. Each word in this thesis is yours. You are my new eyes and my open mind, and you made me able to learn what true happiness is. Grazie.

And as always, ad maiora.
Contents

1 INTRODUCTION ......................................................................................................................... 1
  1.1 THE CELL CYCLE IN STEM CELL AND EARLY DEVELOPMENT ............................... 2
    1.1.1 Cell cycle .................................................................................................................. 2
    1.1.2 Cell cycle in different cell types ................................................................. 8
    1.1.3 Cell fate decisions and cell cycle ............................................................ 13
    1.1.4 Tools to study cell cycle progression ................................................. 16
  1.2 HUMAN EMBRYONIC STEM CELLS (hESCs) ............................................. 19
    1.2.1 PSCs and hESCs ................................................................................................. 19
    1.2.2 Pluripotency transcriptional network ................................................. 22
    1.2.3 Signalling pathways regulating pluripotency .................................. 24
    1.2.4 Chromatin epigenetic control of pluripotency .................................. 27
  1.3 CELL CYCLE REGULATION OF PLURIPOTENCY .................................... 32
    1.3.1 Chromatin organisation during cell cycle ......................................... 32
    1.3.2 Cell cycle regulation of histone marks and methylation .................. 34
    1.3.3 Transcriptional network changes during cell cycle progression ....... 34
  1.4 OPEN QUESTIONS AND PROJECT AIMS ......................................................... 36

2 MATERIAL AND METHODS ................................................................................................. 37
  2.1 CELL CULTURE .............................................................................................................. 38
    2.1.1 Germ layers specification ................................................................................. 38
  2.2 MOLECULAR BIOLOGY ............................................................................................... 39
    2.2.1 Immunofluorescence (IF) ................................................................................ 39
    2.2.2 Western Blotting (WB) ..................................................................................... 39
    2.2.3 Quantitative real-time PCR (qPCR) .............................................................. 40
    2.2.4 Flow cytometry .................................................................................................. 41
    2.2.5 Cell sorting ......................................................................................................... 41
    2.2.6 Cell synchronisation .......................................................................................... 42
  2.3 CHROMATIN IMMUNOPRECIPITATION (ChIP) .................................................. 42
  2.4 ASSAY FOR TRANSPOSASE-ACCESSIBLE CHROMATIN (ATAC) ............... 43
  2.5 GENETIC MANIPULATIONS ....................................................................................... 44
    2.5.1 Transfection ....................................................................................................... 44
    2.5.2 Genotyping ........................................................................................................ 45
  2.6 BIOINFORMATICS ANALYSES AND PIPELINES ....................................... 46
    2.6.1 ChIP-seq ............................................................................................................... 46
    2.6.2 ATAC-seq ........................................................................................................... 47
LIST OF TABLES

TABLE 1. KEY CELL CYCLE FEATURES IN PLURIPOTENT CELLS ........................................ 9
TABLE 2. CELL CULTURE MEDIA COMPOSITION .................................................................. 38
TABLE 3. ANTIBODIES ....................................................................................................... 39
TABLE 4. qPCR PRIMERS .................................................................................................... 41
TABLE 5. CHIP-qPCR PRIMERS ........................................................................................ 43
TABLE 6. GENOTYPING PRIMERS ....................................................................................... 45
LIST OF FIGURES

FIGURE 1.1 CELL CYCLE SCHEMATIC IN EUKARYOTES ............................................. 2
FIGURE 1.2 EXPRESSION OF CDKs, CYCLINS AND CDKI THROUGH THE CELL CYCLE .... 3
FIGURE 1.3 ACTIVATION/INACTIVATION OF CYCLIN/CDK COMPLEXES ....................... 4
FIGURE 1.4 PRB REGULATION ................................................................................... 5
FIGURE 1.5 MITOSIS AND CYTOKINESIS .................................................................... 7
FIGURE 1.6 PSCs AND SOMATIC CELL CYCLE ............................................................ 11
FIGURE 1.7 CELL CYCLE REGULATION OF SMAD2/3 ............................................. 14
FIGURE 1.8 CELL FATE PROPENSITY DURING CELL CYCLE PROGRESSION .............. 15
FIGURE 1.9 FUCCI hESCs SYSTEM ............................................................................. 17
FIGURE 1.10 HUMAN EMBRYONIC STEM CELLS ..................................................... 19
FIGURE 1.11 DEVELOPMENTAL POTENTIAL OF STEM CELLS .................................. 20
FIGURE 1.12 CORE PLURIPOTENCY TRANSCRIPTIONAL NETWORK ...................... 22
FIGURE 1.13 SIGNALLING PATHWAYS REGULATING PLURIPOTENCY IN hESCs ........ 26
FIGURE 1.14 3D GENOME ORGANISATION ............................................................... 27
FIGURE 3.1 SCHEMATIC FOR FUCCI FACS SORTING ............................................ 52
FIGURE 3.2 CHIP OPTIMISATION SUMMARY .................................................................. 53
FIGURE 4.1 ATAC-SEQ SCHEMATIC ....................................................................... 56
FIGURE 4.2 ATAC-SEQ FRAGMENT SIZE PLOT ......................................................... 57
FIGURE 4.3 ATAC-SEQ IRD PLOT ............................................................................. 57
FIGURE 4.4 ATAC-SEQ METAPLOT ....................................................................... 58
FIGURE 4.5 ATAC-SEQ VENN OVERLAP ................................................................. 58
FIGURE 4.6 ATAC-SEQ FEATURES DISTRIBUTION PLOTS ..................................... 59
FIGURE 4.7 ATAC-SEQ GENE ONTOLOGY AND PATHWAY ANALYSES .................. 60
FIGURE 4.8 ATAC-SEQ DIFFERENTIALLY OPEN REGIONS ..................................... 61
FIGURE 4.9 DIFFERENTIALLY OPEN REGIONS METAPLOT AND HEATMAP ......... 62
FIGURE 4.10 DIFFERENTIALLY OPEN REGIONS FEATURES DISTRIBUTION PLOTS .... 63
FIGURE 4.11 DIFFERENTIALLY OPEN REGIONS GENE ONTOLOGY ......................... 64
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>2i</td>
<td>dual inhibition</td>
</tr>
<tr>
<td>5C</td>
<td>Chromosome Conformation Capture Carbon Copy</td>
</tr>
<tr>
<td>AAVS1</td>
<td>AAVS1 locus: PPR1R12C gene on human chromosome 19</td>
</tr>
<tr>
<td>ALK</td>
<td>Activin receptor-Like Kinases</td>
</tr>
<tr>
<td>APC</td>
<td>Anaphase- Promoting Complex, also known as cyclosome</td>
</tr>
<tr>
<td>ATAC</td>
<td>Assay for transposase-accessible chromatin</td>
</tr>
<tr>
<td>AURK</td>
<td>Aurora Kinase</td>
</tr>
<tr>
<td>BCL-3</td>
<td>B Cell Lymphoma 3 protein</td>
</tr>
<tr>
<td>BMP</td>
<td>Bone Morphogenetic Protein</td>
</tr>
<tr>
<td>bp</td>
<td>base-pairs</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>CBX</td>
<td>Cromobox protein</td>
</tr>
<tr>
<td>CDB</td>
<td>Cell Dissociation Buffer</td>
</tr>
<tr>
<td>CDC25</td>
<td>Cell Division Cycle 25</td>
</tr>
<tr>
<td>CDC6</td>
<td>Cell Division Cycle 6</td>
</tr>
<tr>
<td>CDKIs</td>
<td>Cyclin Dependent Kinase Inhibitors</td>
</tr>
<tr>
<td>CDKs</td>
<td>Cyclin Dependent Kinases</td>
</tr>
<tr>
<td>CDM</td>
<td>Chemically Defined Medium</td>
</tr>
<tr>
<td>CDT1</td>
<td>Chromatin Licensing/DNA Replication Factor 1</td>
</tr>
<tr>
<td>CHEK2</td>
<td>CHEckpoint Kinase 2</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin Immunoprecipitation</td>
</tr>
<tr>
<td>co-SMAD</td>
<td>common-mediator SMAD</td>
</tr>
<tr>
<td>CpG</td>
<td>Cytosine-phosphate-Guanine dinucleotide</td>
</tr>
<tr>
<td>cPRC1</td>
<td>Canonical PRC1</td>
</tr>
<tr>
<td>CTCF</td>
<td>CCCTC-binding factor</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Cyc</td>
<td>cyclins</td>
</tr>
<tr>
<td>DAPI</td>
<td>4',6-diamidino-2-phenylindole</td>
</tr>
<tr>
<td>dEN</td>
<td>Definitive Endoderm</td>
</tr>
<tr>
<td>dNTPs</td>
<td>deoxynucleotides</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>ECCs</td>
<td>Embryonal Carcinoma Cells</td>
</tr>
<tr>
<td>EG1</td>
<td>Early G1 phase</td>
</tr>
<tr>
<td>EpiSCs</td>
<td>Epiblast stem cells</td>
</tr>
<tr>
<td>ERK</td>
<td>Extracelluar signal-Regulated Kinase</td>
</tr>
<tr>
<td>EZH2</td>
<td>Enhancer of Zeste Homolog 2</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence Activated Cell Sorting</td>
</tr>
<tr>
<td>FBS</td>
<td>foetal bovine serum</td>
</tr>
<tr>
<td>FGF</td>
<td>Fibroblast Growth Factors</td>
</tr>
<tr>
<td>FGFRs</td>
<td>FGF receptors</td>
</tr>
<tr>
<td>FLR</td>
<td>Footprint Log-likelihood Ratio</td>
</tr>
<tr>
<td>FPKM</td>
<td>Fragments Per Kilobase of transcript per Million mapped reads</td>
</tr>
<tr>
<td>FS</td>
<td>Forward Scatter</td>
</tr>
<tr>
<td>FUCCI</td>
<td>Fluorescent, Ubiquitination-based Cell Cycle indicator</td>
</tr>
<tr>
<td>G1/S</td>
<td>G1/S transition</td>
</tr>
<tr>
<td>GREAT</td>
<td>Genomic Regions Enrichment of Annotations Tool</td>
</tr>
<tr>
<td>H2AK119ub</td>
<td>Histone 2A Lysine 119 ubiquitination</td>
</tr>
<tr>
<td>H3K27ac</td>
<td>Histone 3 Lysine 27 acetylation</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>Histone 3 Lysine 27 trimethylation</td>
</tr>
<tr>
<td>H3K4me1</td>
<td>Histone 3 Lysine 4 mono-methylation</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>Histone 3 Lysine 4 trimethylation</td>
</tr>
<tr>
<td>H3K9ac</td>
<td>Histone 3 Lysine 9 acetylation</td>
</tr>
<tr>
<td>H3K9me3</td>
<td>Histone 3 Lysine 9 trimethylation</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>hESCs</td>
<td>human embryonic stem cells</td>
</tr>
<tr>
<td>HRP</td>
<td>Horseradish Peroxidase</td>
</tr>
<tr>
<td>I-SMAD</td>
<td>Inhibitory-SMAD</td>
</tr>
<tr>
<td>ICM</td>
<td>inner cell mass</td>
</tr>
<tr>
<td>IDR</td>
<td>Irreproducibility Discovery Rate</td>
</tr>
<tr>
<td>IF</td>
<td>Immunofluorescence</td>
</tr>
<tr>
<td>IGFs</td>
<td>Insulin-like Growth Factors</td>
</tr>
<tr>
<td>IGV</td>
<td>Integrative Genomics Viewer</td>
</tr>
<tr>
<td>iKD</td>
<td>Inducible KnockDown</td>
</tr>
<tr>
<td>iPSC</td>
<td>inducible pluripotent stem cell</td>
</tr>
<tr>
<td>LADs</td>
<td>Lamina-Associated Domains</td>
</tr>
<tr>
<td>LG1</td>
<td>Late G1 phase</td>
</tr>
<tr>
<td>LIF</td>
<td>Leukemia Inhibitory Factor</td>
</tr>
<tr>
<td>MAPK</td>
<td>Mitogen-Activated Protein Kinase</td>
</tr>
<tr>
<td>MBT</td>
<td>mid-blastula transition</td>
</tr>
<tr>
<td>MCM</td>
<td>Minichromosome maintenance protein complex</td>
</tr>
<tr>
<td>MEF</td>
<td>Mitotically-inactivated mouse Embryonic Fibroblasts</td>
</tr>
<tr>
<td>mESCs</td>
<td>mouse embryonic stem cells</td>
</tr>
<tr>
<td>MLL</td>
<td>Mixed Lineage Leukaemia</td>
</tr>
<tr>
<td>ncPRC1</td>
<td>Non-canonical PRC1</td>
</tr>
<tr>
<td>NPAT</td>
<td>Nuclear Protein Ataxia-Telangiectasia locus</td>
</tr>
<tr>
<td>ORC</td>
<td>Origin Recognition Complex</td>
</tr>
<tr>
<td>OSN</td>
<td>OCT4, SOX2, NANOG</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate-Buffered Saline</td>
</tr>
<tr>
<td>PcG</td>
<td>Polycomb Group</td>
</tr>
<tr>
<td>PCGF</td>
<td>Polycomb group ring-finger domain proteins</td>
</tr>
<tr>
<td>Pen/Strep</td>
<td>Penicillin-Streptomycin</td>
</tr>
</tbody>
</table>
PFA  Paraformaldehyde
PHC  Polyhomeotic homologous protein
PI3K  Phosphatidylinositol 3-Kinase
PKB  Protein Kinase B
PLK1  Polo-like kinase 1
PMSF  Phenylmethane sulfonyl fluoride
PP1  Protein Phosphatase 1
pRB  tumour suppressor protein Retinoblastoma
PRC  Polycomb Repressive Complex
PrE  Primitive Endoderm
pre-RC  pre-Replicative Complex
PSCs  Pluripotent Stem Cells
PTMs  Post-Translational Modifications
PVA  Polyvinyl alcohol
qPCR  Quantitative real-time PCR
R-point  restriction point
R-SMAD  Receptor-regulated SMAD
RNAPII  RNA polymerase II
RYBP  YY1-binding protein
S/G2/M  S/G2/M phase
SAM  Sterile alpha motif
SKP2  ubiquitin ligase S-phase Kinase-associated Protein 2
SMAD  Sma- and Mad-related protein
SS  Side Scatter
STAT  Signal Transducer and Activator of Transcription
TADs  Topological-Associated Domains
TE  trophectoderm
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFs</td>
<td>Transcription Factors</td>
</tr>
<tr>
<td>TGFβ</td>
<td>Transforming Growth Factor β</td>
</tr>
<tr>
<td>trxG</td>
<td>Trithorax Group</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription Start Sites</td>
</tr>
<tr>
<td>WB</td>
<td>Western Blotting</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-Type</td>
</tr>
</tbody>
</table>
RELIST OF APPENDICES

BIOINFORMATICS OUTPUT .................................................................................................................. 132
OCT4 SOPTIKD PLASMID MAP ........................................................................................................ 135
SCR SOPTIKD PLASMID MAP ........................................................................................................ 136
PTP6 OCT4 ERT2 PLASMID MAP ..................................................................................................... 137
1 INTRODUCTION

Cell cycle regulation has a key function in development, organ homeostasis and disease. Recently new interest has emerged in the stem cell field, arising from the demonstration that the cell cycle directly controls maintenance of pluripotency and initiation of differentiation. Nonetheless, the precise molecular mechanisms, transcriptional networks and epigenetic states have not been investigated comprehensively yet. This chapter starts with an overview of the main events taking place during cell cycle and their regulation in stem cells and early development, followed by a summary of recent finding relating these mechanisms to cell fate propensity, and the tools available to do so. As human embryonic stem cells (hESCs) are proven to be the best model to investigate these questions, a subchapter describing their transcriptional and epigenetic states during cell cycle progression follows. The introduction concludes summarising the main questions that this PhD thesis aims to address.
1.1 The cell cycle in stem cell and early development

The cell cycle is the sequence of events leading to cellular division and duplication of DNA in order to produce two daughter cells. This subchapter summarises the current knowledge about cell cycle structure and its regulation in stem cell and early development, and the available tools to study cell fate propensity during cell cycle progression.

1.1.1 Cell cycle

In eukaryotes, the cell cycle is divided in two major phases: interphase and the M phase. During interphase, the cell duplicates its DNA preparing for mitosis. It is subdivided into three distinct phases: G1, the period following cell division and preceding DNA replication when the cell prepares for DNA duplication, S phase, the period during which DNA and histones synthesis occurs, and the G2 phase, which follows replication and precedes mitosis, during which cells produce proteins necessary for mitosis. The M phase includes the processes of mitosis, in which duplicated chromosomes segregate, and cytokinesis, in which the cell is physically divided into two daughter cells (Figure 1.1).

![Cell cycle schematic in eukaryotes](image)

Figure 1.1 Cell cycle schematic in eukaryotes.

During interphase, in the G1 phase the cell prepares for DNA duplication (S phase), and produces proteins necessary for mitosis (M phase). During mitosis, chromosomes segregate, and the cell physically divides into two daughter cells (cytokinesis).

Control of cell cycle progression

Checkpoints ensure correct chromosome segregation, cell size and DNA integrity after replication, allowing progression of the cell cycle. Absence of proliferation
signals drives cells arrest in G1 and quiescence, apoptosis or differentiation. Different layers of regulation ensure correct cell cycle progression, and different combinations of the following mechanisms determine whether the cells divide or enter a quiescent (G0) state. Firstly, checkpoints passages are regulated by different sets of cyclins (Cyc) and Cyclin Dependent Kinases (CDKs) which control the activity of DNA replication and the cell division enzymes (Figure 1.2). Usually CDKs are constitutively expressed, whereas the cyclins are quickly synthetized and degraded, allowing tight regulation of their respective CDKs.

![Figure 1.2 Expression of CDKs, cyclins and CDKI through the cell cycle.](image)

CDK activity during the early G1 is very low, but it increases in the late G1, when CDK4 and CDK6 associate with the D-type Cyclins (D1, D2, D3). In the late G1 phase, in order to enter the S phase and pass the restriction point (R-point), CDK2 is activated by CycE, and the complex further phosphorylates pRB, enhancing E2F mediated transcription. In the S phase, CycA/CDK2 drive the synthesis of proteins essential for DNA replication. At the end of S phase, Cyclin A is associated with CDK1 and CDK2. During the G2 phase, Cyclin A is degraded, while Cyclin B is synthesized and CycB/CDK1 complexes form to enter mitosis. In addition, CDKIs directly bind and inhibit Cyclins-CDKs complexes, arresting cell cycle progression. The INK4 family of CDKIs (p16^{INK4A}, p15^{INKB}, p18^{INK4C}, p19^{INK4D}) inhibit the CycD-CDK4/6 complex; CIP/KIP, instead, inhibits a wide range of CDKs (p21^{CIP1}, p27^{KIP1}, p57).

To ensure this rapid regulation, cyclins are firstly ubiquitinilated by the Anaphase-Promoting Complex (APC, also known as cyclosome) and then quickly degraded, triggering inactivation of the respective associated CDK (Meyer and Rape, 2011). Secondly, in order to trigger the abrupt activation/inactivation of such complexes,
Cyclin/CDK complexes contain inhibitory phosphates, and therefore require de-phosphorylation to become active (Figure 1.3).

![Diagram of Cyclin/CDK complexes activation/inactivation](image)

Figure 1.3 Activation/inactivation of Cyclin/CDK complexes.

Cyc/CDKs contain inhibitory phosphates and require de-phosphorylation to become active. This is regulated by WEE1 kinase and CDC25 phosphatase: CDC25 phosphatase removes inhibitory phosphates, WEE1 kinase phosphorylates and inhibits CDK activity.

In addition to these canonical cell cycle functions, the D-type cyclins and the kinases CDK4 and CDK6 have been shown to perform several non-canonical functions, such as transcription, differentiation, and metabolism (Hydbring et al., 2016). Lastly, additional control on CDKs activity is provided by Cyclin Dependent Kinase Inhibitors (CDKIs), which directly bind and inhibit Cyclins-CDKs complexes, resulting in arrest of cell cycle progression. The INK4 family of CDKIs (p16^{INK4A}, p15^{INKB}, p18^{INK4C}, p19^{INK4D}) are known to inhibit the CycD-CKD4/6 complex. CIP/KIP, instead, inhibits a wide range of CDKs: p21^{CIP1}, p27^{KIP1}, and p57 (Sherr and Roberts, 1999). Many of them are involved in surveillance mechanisms to monitor the status of cell cycle events such as DNA replication and chromosome compaction, to determine whether or not the cell cycle progresses (Figure 1.2).

Regulation of G1 phase

G1 phase is an important decision point for the cell and an intense period of metabolic activity and cell growth. Based on intracellular and extracellular cues, the cell cycle machinery can either hold the cell in G1 (or in a continued non-proliferative state G0) or commit to complete an entire cell cycle. The length of this phase is indeed very variable in the different cell types. CDK activity during early G1 is very low, which promotes the formation of pre-replication complexes at the origin of DNA replication. By late G1, CDK activity is evident due to the association of CDK4 and CDK6 with the D-type cyclins (D1, D2, D3). CDK4 and CDK6 are highly tissue-specific homologous serine/threonine kinases, known to
phosphorylate a mostly overlapping set of target proteins, confirmed by gene knock-out experiments in vivo (Anders et al., 2011; Malumbres and Barbacid, 2009). CycD-CDK4/6 promote cell cycle progression by two major mechanisms (Sherr and Roberts, 2004). First, they sequester p21\textsuperscript{CIP1} and p27\textsuperscript{KIP1}, two CDKIs that bind to and prevent activation of CycE-CDK2 kinase. Second, they control the G1 phase transition through the regulation of the tumour suppressor protein Retinoblastoma (RB, encoded by the RB1 gene) and the related proteins p107 (RBL1) and p130 (RBL2). When hypo-phosphorylated, RB binds to the E2F transcription factor, preventing its binding to genes necessary for the S phase transition. In response to numerous signals during the G1 phase, pRB becomes hyper-phosphorylated and inactivated by CDK4/6, leading to the release of E2F, and subsequent activation of genes promoting G1/S transition. In the late G1 phase, in order to enter the S phase and pass the restriction point (R-point), CDK2 is activated by CycE, and the complex further phosphorylates pRB, enhancing E2F mediated transcription (Figure 1.4). For the rest of the cell cycle, pRB remains hyper-phosphorylated, until it is dephosphorylated by the serine/threonine Protein Phosphatase 1 (PP1) during mitotic exit.

Figure 1.4 pRB regulation.
During the G1 phase, pRB is hyper-phosphorylated and inactivated by CDK4/6, E2F is released and can activate genes promoting G1/S transition. In the late G1 phase, CDK2 is activated by CycE and the complex further phosphorylates pRB, enhancing E2F mediated transcription in order to enter the S phase and pass the restriction point (R-point).

Regulation of S phase
The S phase is mostly dedicated to DNA synthesis, in order to ensure that each chromosome has two sister chromatids. During this phase, the cells contain double the amount of DNA, and the CycA/CDK2 complex drives the synthesis of proteins essential for DNA replication. Exception made for histones, transcription and
protein synthesis are very low during this phase (Cameron and Greulich, 1963; Nelson et al., 2002; Wu and Bonner, 1981). During the early G1 phase, several proteins including the Origin Recognition Complex (ORC) are recruited on specific sites on the DNA (origin of replication sequences), in order to assemble the pre-Replicative Complex (pre-RC). In the first step of replication initiation, ORC recruits Cell Division Cycle 6 (CDC6) and Chromatin Licensing/DNA Replication Factor 1 (CDT1), and together they load a pair of DNA helicases to form the pre-RC. This is composed of six ORC proteins (ORC1-6), CDC6, CDT1, and a heterohexamer of the six Minichromosome maintenance protein complex (MCM2-7). At the start of the S phase, the MCM heterohexamer is phosphorylated by CDC7 and CDK2, which displace CDC6 and recruit MCM10. MCM10 cooperates with MCM2-7 in the recruitment of CDC45. CDC45 then recruits key components of the replisome; the replicative DNA polymerase α and its primase forming the replication forks (Takisawa et al., 2000). Phosphorylation of CDC6 ensures that DNA replication cannot be initiated before the next cell cycle.

Regulation of G2/M phase
During the G2 phase, CycB/CDK1 complex is accumulated and immediately phosphorylated by the inhibitory kinases WEE1, in order to be maintained in its inactive form. At the G2/M transition, Polo-like kinase 1 (PLK1) and Aurora kinases (Aurora A and Aurora B) phosphorylate and activates Cell Division Cycle 25 (CDC25), a phosphatase that dephosphorylates and activates the CycB/CDK1 complex, promoting mitosis by activating genes required for chromosome segregation. CycB/CDK1 is able to self-activate through a positive feedback loop that activates more CDC25 and inactivates WEE1 (Arroyo and Raychaudhuri, 1992; Kato et al., 1993; Pardee, 1974; Rubin et al., 2001). When the cell enters the M phase, complexes of Condensins are phosphorylated by CycB/CDK1, and their active form drive chromosomal condensation (Kimura et al., 1998).

Regulation of mitosis and cell division
Mitosis is the process during cell division in which the replicated DNA molecules of each chromosome are segregated into the two nuclei, and ensures that they receive a complete and equivalent complement of genetic material. It is divided into prophase, prometaphase, metaphase, anaphase, and telophase, and followed by cytokinesis (Figure 1.5). During prophase, condensed chromosomes are organised in the typical mitotic rod-shaped structure, composed by two sister chromatids attached together at the centromere by cohesin complexes. In this
phase, the cytoskeleton is disassembled, and the mitotic spindle, the microtubule machinery required for chromosomes segregation, is assembled (Walczak and Heald, 2008). During prometaphase, the nuclear envelope abruptly breaks and chromosomes attach to the mitotic spindle via their kinetochores. At metaphase chromosomes are aligned at the centre of the spindle, between the cell poles. After centromeres split, during anaphase sister chromatids are synchronously separated from one another into opposite regions of the dividing cell though microtubule shortening. Once segregated, during telophase a new nuclear envelope is formed and a complex mainly composed by actin and myosin filaments, known as the contractile ring, is assembled to allow cytokinesis. Finally, the cytoplasm segregation occurs with the physical division of the two daughter cells and segregated chromosomes return in their dispersed interphase condition (Glotzer, 2005).

**Figure 1.5 Mitosis and cytokinesis.**

Mitosis is divided into prophase, prometaphase, metaphase, anaphase, telophase, and cytokinesis. During prophase, chromosomes are condensed (black bars), the cytoskeleton is disassembled, and the mitotic spindle (blue) is assembled. During prometaphase, the nuclear envelope is disrupted (dotted line) and chromosomes attach to the mitotic spindle. At metaphase, chromosomes align at the centre of the spindle, and centromeres split. During anaphase, sister chromatids are synchronously separated. Once segregated, during telophase, a new nuclear envelope is formed (dotted line) and the contractile ring (yellow) is assembled. During cytokinesis, the cytoplasm segregates and the daughter cells physically divides.

**Additional cell cycle checkpoints**

The cell cycle control organisation uses several mechanisms to block cell cycle progression, or complete withdrawal from the cell cycle, in case of DNA damage. A well-known mechanism has been described at the G1/S transition and is activated in response to DNA damage. DNA damage increases the concentration and activity of p53, a transcription factor that in turns activates the inhibitor p21, which binds to CDK4/6 and CDK2 to prevent S phase entry and DNA replication (Matlashewski et al., 1984). This cell cycle arrest allow the damage to be repaired or, in case it is too severe, can induce apoptosis (Yonish-Rouach et al., 1991). This factor has been reported to be mutated or absent in 50% of all human
cancers. In addition, DNA damage is detected by dedicated proteins and triggers cell cycle arrest via CHEckpoint Kinase 2 (CHEK2) and p53 in G1 phase or via CHEK1 in S or G2 phase (Flaggs et al., 1997; Sanchez et al., 1997). In case of incorrect DNA replication, another mechanism of cell cycle arrest has been described at the G2/M transition. CDC25 is able to inhibit itself and prevent the removal of the CycB/CDK1 inhibitory phosphate, necessary for cell to progress into mitosis, leaving the complex inactive and M phase delayed until damage is repaired and replication completes (Strausfeld et al., 1991).

Additional diversions from canonical cell cycle progression involve delay or total withdraw for prolonged periods or permanently. In many terminally differentiated cells, such as red cells, the cell cycle machinery is completely shut down and cell stop dividing. In some other cell types, such as liver cells, in the presence of damage or proper signal, cells can exit the G0 phase and proliferate.

1.1.2 Cell cycle in different cell types
The features described so far represent the underlying mechanisms that regulate the cell cycle in all eukaryotes. The cell cycle structure and the single phases are characteristics of all the cell types; however, their length and regulation are highly variable. In addition to this, they lack strict checkpoints, and CDK inhibitors follow different expression patterns to allow alternative regulations. More specifically, pluripotent cells are characterised by a shorter cell cycle compared to adult stem cells and fully differentiated cells (Becker et al., 2006; Savatier et al., 1994). In addition to this, pRB activity is absent or strongly reduced (Conklin et al., 2012; Dannenberg et al., 2000), and Cyclins/CDKs are highly active, whereas they are cell cycle regulated in somatic cells (Stead et al., 2002). The following paragraphs illustrate the main differences between the pluripotent cell cycle in early development, mouse Embryonic Stem Cells (mESCs), and human Embryonic Stem Cells (hESCs), also summarised in Table 1.
### Table 1. Key cell cycle features in pluripotent cells

<table>
<thead>
<tr>
<th></th>
<th>mESCs</th>
<th>hESCs</th>
<th>Somatic cells</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell cycle length</strong></td>
<td>~12h</td>
<td>~18h</td>
<td>~24h</td>
</tr>
<tr>
<td><strong>Cell cycle distribution</strong></td>
<td>mostly S-phase (&gt;50%)</td>
<td>mostly S-phase (&gt;50%)</td>
<td>mostly G1-phase (&gt;50%)</td>
</tr>
<tr>
<td><strong>pRB</strong></td>
<td>Hyper-phosphorylated</td>
<td>Hyper/hypo-phosphorylated</td>
<td>Hypo-phosphorylated</td>
</tr>
<tr>
<td><strong>Cyc/CDK activity</strong></td>
<td>Constitutively expressed (exc. CDK1, CycD)</td>
<td>Mostly cell cycle regulated</td>
<td>Cell cycle regulated</td>
</tr>
<tr>
<td><strong>CDKI</strong></td>
<td>Absent</td>
<td>Expressed at low levels</td>
<td>Highly expressed</td>
</tr>
<tr>
<td><strong>Checkpoints</strong></td>
<td>Inactive</td>
<td>Partially active</td>
<td>Active</td>
</tr>
</tbody>
</table>

**Cell cycle in early development**

In order to sustain the fast growth of the mammalian embryo, the cell cycle machinery must be coupled to specific embryonic events to ensure that growth and differentiation are synchronised. For instance, during mouse gastrulation, if a critical embryonic mass is not generated, formation of the three germ layers is delayed (Power and Tam, 1993; Rands, 1986). The rapid cell cycles associated with early embryonic development has been well described in lower vertebrates such as *Xenopus* (Murray and Kirschner, 1989), *Drosophila* (Edgar and Lehner, 1996) and *Danio rerio* (Yarden and Geiger, 1996). In these model organisms, the cell cycle effectively only alternates DNA synthesis (S phase) and chromosomes segregation (M phase), lacking fully formed G1/G2 phases. In the mouse, from the second cell division, a short G1 phase (1–2 hours) is distinguishable but the G2 phase at this time is remarkably long (12–16 hours) and importantly, marks the beginning of zygotic genome activation (Flach et al., 1982; Sikora-Polaczek et al., 2006). In the subsequent divisions, the duration of the G2 phase is shorter and comparable to that of G1 and cell division is mostly driven by *de novo* RNA and protein synthesis (Mac Auley et al., 1993). During epiblast development, pluripotent cells maintain a short G1 phase and cycle quickly compared to their extra-embryonic counterpart (Lawson et al., 1991; Snow and Bennett, 1978). The late pre-implantation and early post-implantation cells from the mouse embryo proliferate at remarkably fast rates: between 4.5 and 6.0 days post coitum (dpc) the epiblast has an average generation time of 10h (Solter et al., 1971). However, at the gastrulation stage instead, between 6.5 and 7 dpc, cell cycle accelerates even further, generating more than 4,000 cells with a mean generation time of 4.5h (Power and Tam, 1993). This means that the switch is not triggered by implantation itself, but rather at gastrulation by the formation of the primitive...
streak. This fast proliferation allows rapid induction of differentiation in the embryo, initiated by the interplay of different signalling pathways.

Many knock-out studies have proven that during development most Cyclins and CDKs are functionally redundant (Fantl et al., 1995; Geng et al., 2003; Ortega et al., 2003; Parisi et al., 2003), as well as E2F and RB/RB-like p107 and p130 (Clarke et al., 1992; Lee et al., 1992; Lee et al., 1996). The only exceptions are CycA2, CycB1 and CDK1: CycA2 and CycB1 homozygous null mutants survive only until 5.5 dpc (Brandeis et al., 1998; Murphy et al., 1997), whereas in absence of CDK1 embryos fail to develop to the morula and blastocyst stages (Santamaria et al., 2007). Whether the rapid division reported in these studies is just a useful mechanism to increase the cell number before gastrulation or something that is inherent to the machineries associated with the pluripotent state is challenging to investigate in vivo, but from studies using Pluripotent Stem Cells (PSCs) it seems that rapid rates of cell proliferation are not an absolute prerequisite for maintenance of pluripotency (Li et al., 2012b; Singh and Dalton, 2009). In addition to this, considering the relatively low cell number available at this stage of development, it is difficult to investigate further mechanistic insights, and only PSCs or model system can be used for this purpose. However, recent work in *Xenopus* embryos has looked into how the cell cycle lengthens during the mid-blastula transition (MBT) and found that changes in the nuclear/cytoplasmic ratio impact concentrations of factors required for DNA replication (Collart et al., 2013). The onset of the S phase is delayed by the dilution effect resulting in an extended G1 phase. These changes have long been thought to affect the cell cycle length in early development, but it is still unknown whether this mechanism is conserved in mammalian embryos. In addition to this, most of these studies have been performed using mouse embryos, which might not be the best model for human cell cycle regulation, considering their shorter developmental timeframe.

**Cell cycle in pluripotent stem cell**

The analyses described so far do not reveal specific roles for cell cycle regulators during early development. This is due to the limited number of pluripotent cells in the embryo and the consequent difficulties in generating enough material to perform biochemical analyses of CDK activity in vivo. This issue was overcome by the derivation of embryonic stem cell lines, firstly from mouse (mESCs) and then from human embryos (hESCs) (Evans and Kaufman, 1981; Martin, 1981; Thomson et al., 1998), together with the development of the inducible pluripotent stem cell
(iPS C) technology (Takahashi et al., 2007; Takahashi and Yamanaka, 2006). These cells have the remarkable capacity to divide indefinitely in vitro while maintaining their ability to differentiate into the three primary germ layers. Interestingly, they lack the capacity to undergo quiescence or senescence and their growth is not subject to contact inhibition. However, it is important to remember that their embryonic equivalent does not remain a stable stem cell pool, but rapidly differentiates at the gastrulation stage. For this reason, this opens a unique window into the study of early human development and its links with cell cycle progression. As previously described, the cell cycle machinery is very well conserved in different cell types and among species. However, the length of the cell cycle and the molecular mechanisms regulating its progression are highly specific. PSCs have a quicker cell cycle compared to their fully differentiated counterparts (Figure 1.6). In the rodent epiblast, undifferentiated cells lack of a fully formed G1 and G2 phases, and about 60% of their cell cycle time is dedicated to DNA synthesis (S phase) (Stead et al., 2002). Similar findings have been reported in mESCs (Savatier et al., 1994), hESCs (Becker et al., 2006), murine embryonal carcinoma (Kranenburg et al., 1995; Mummery et al., 1987), embryonal germ cells (Resnick et al., 1992), and other primates (Fluckiger et al., 2006).

![Figure 1.6 PSCs and somatic cell cycle.](image)

Schematic comparing pluripotent and somatic cell cycle: pluripotent stem cells have a quicker cell cycle compared to their differentiated counterpart, characterised by a shorter G1 phase and a high proliferative state with most time spent in the S phase.
In mESCs, Cyclins E/A and CDK2 are constitutively expressed and highly active during the cell cycle, whereas they are cell cycle regulated in the majority of somatic cell types (Stead et al., 2002). However, not all of the Cyclins/CDKs follow this pattern: CycB1/CDK1, for example, has been reported to be cell cycle regulated (White et al., 2005). Cyclin Ds have different patterns: D1 is expressed at low levels, D2 is completely absent, whereas D3 is expressed at robust levels and forms active complexes with CDK6 (Faast et al., 2004; Stead et al., 2002). The constitutive activation of CDK2 throughout the cell cycle enhances pRB phosphorylation, holding it in a constitutive hyper-phosphorylated state, therefore inactive. The absence of pRB activity allows quick progression of the cell cycle in the S phase, completely overcoming the R-point and disconnecting E2F target genes from cell cycle progression. Indeed, knockout of all three RB-like proteins (pRB, p107, p130) has no effect on proliferation in mESCs, but strongly impaired their differentiation capacities (Dannenberg et al., 2000; Sage et al., 2000), suggesting that a tight cell cycle regulation is indispensable to drive differentiation. Moreover, pRB/p107/p130 triple knockout speeds up the cell cycle in mouse embryonic fibroblasts (MEFs), enhancing their proliferation rate (Dannenberg et al., 2000; Sage et al., 2000). The mechanisms controlling the unrestrained activity of CDKs in mESCs have not been investigated thoroughly, but their uncontrolled cycle is partially due to the lack of CDKI, and it is conserved from the embryo at the peri-implantation stage (Faast et al., 2004). Additional layers of control are achieved by micro-RNAs, such as the miR-290 family, that have been shown to also accelerate the G1/S transition phase by pRB-dependent and independent mechanisms (Wang et al., 2008; Wang et al., 2013).

hESCs have also been reported to be very proliferative cells, with a similar cell cycle structure to the mESCs with a truncated G1 phase that is lengthened upon differentiation (Becker et al., 2006; Calder et al., 2013). Their pluripotent state is associated with mostly hyper-phosphorylated pRB, with a small fraction still showing hypo-phosphorylation (Conklin et al., 2012), also reported in primates (Fluckiger et al., 2006). The pRB pathway appears to be active in G1 phase (Filipczyk et al., 2007; Sela et al., 2012), with different mechanisms of regulation proposed, relying on Cyclin D and A (Becker et al., 2010; Filipczyk et al., 2007). The pluripotent status of hESCs relies on CDK2 activity (Neganova et al., 2009), and lack of G1 checkpoint regulation (Coronado et al., 2013). However, in this case most Cyclins and CDKs are cell cycle regulated, including CycA/CDK2: CDK6, CDK2, CycD1/D3 showed the highest expression in G1; CDK4 in G1 and S;
Chapter 1: Introduction

and CycE, CDK1, CycA, CycB1, and CycD2 in G2, CDC25A in S phase (Neganova et al., 2009). CDKI are not completely absent, but maintained at low levels through independent regulations: p21 is regulated by the epigenetic modifier JMJD5 (Zhu et al., 2014) and the miR-302 family (Dolezalova et al., 2012) whereas the ubiquitin ligase S-phase Kinase-associated Protein 2 (SKP2) instead targets p27 for degradation (Egozi et al., 2007). The tight modulation of these two inhibitors is likely resulting in elevated CDKs activity and a rapid cell cycle.

1.1.3 Cell fate decisions and cell cycle

As described in the paragraphs above, even though its structure is highly conserved, the cell cycle machinery is highly dynamic in both early development and during differentiation in vivo and in vitro. Already at gastrulation, for example, when lineage commitment occurs, the G phases lengthen, triggering an increased cell division time (Snow and Bennett, 1978). Moreover, most of the cell cycle changes are determined by modification of the G1 phase, suggesting a pivotal role in cell identity determination for this phase. The length of this phase indeed increases in parallel with a reduction in pluripotency markers expression and an elevation in expression of early differentiation markers (Calder et al., 2013; Faast et al., 2004; Stead et al., 2002). Therefore, the connection between rapid proliferation and broad differentiation capacities hints towards a mechanistic link between cell cycle progression, especially for the G1 phase, and cell fate commitment. The earliest described events during differentiation involve decreased CDKs activity (Stead et al., 2002) and full activation of the pRB pathway (Sela et al., 2012), resulting in G1 lengthening. Indeed, in mESCs elevated CDKs levels are required to maintain self-renewal: knockdown of CycE induces spontaneous differentiation delaying G1/S transition, whereas its overexpression makes the cells less sensitive to Leukemia Inhibitory Factor (LIF) withdrawn (Coronado et al., 2013). The intriguing idea that cell fate decisions are taken in the G1 phase has been thoroughly investigated, but only recently mechanistically proven. Early studies in Embryonal Carcinoma Cells (ECCs) showed that colony forming activity was more responsive in G1 phase (Pierce et al., 1984; Wells, 1982), but the lack of valuable tools to explore cell cycle dynamics made this route hard to pursue. Only recently, studies have mechanistically connected cell cycle regulation with cell fate propensity, showing that the differentiation capacity in hESCs varies during the progression of cell cycle, and that differentiation is only initiated in the G1 phase (Calder et al., 2013; Coronado et al., 2013). Cyclins and
CDKs have been shown to affect stem cell fate, for example forced G1 lengthening by chemical inhibition of CDK4 has been shown to increase differentiation (Roccio et al., 2013). Most interestingly, our lab has shown that hESCs in the early G1 phase can only initiate differentiation into endoderm, whereas the late G1 phase is only permissive for neuroectoderm specification (Pauklin and Vallier, 2013) (Figure 1.7). This is regulated by the levels of Cyclin D, which have low abundance in the early G1 phase and allow the Activin/Nodal effector Sma- and Mad-related protein 2 and 3 (SMAD2/3) to enter the nucleus and bind endoderm genes. In contrast, high levels of Cyclin D in the late G1 activate CDK4/6 that leads to the phosphorylation of SMAD2/3 linker region, thereby preventing its entry into the nucleus, thus inhibiting the up-regulation of endodermal genes and allowing neuroectoderm specification (Pauklin and Vallier, 2013). Consistent with this, in self-renewing conditions, developmental genes have been reported to be transitorily transcribed in the G1 phase at very low levels, to then become fully activated in response to differentiation cues (Calder et al., 2013; Singh et al., 2013).

![Figure 1.7 Cell cycle regulation of SMAD2/3.](image-url)

During the EG1 phase, SMAD2/3 enters the nucleus, it binds to endoderm genes, allowing endoderm specification. During the LG1 phase, high levels of CycD activate CDK4/6. This phosphorylates SMAD2/3, preventing its entry into the nucleus and inhibiting the up-regulation of endodermal genes, allowing neuroectoderm differentiation instead. Adapted from Pauklin and Vallier, 2013.

Lastly, it has been shown that factors controlling the G2/M phase decrease differentiation propensity in this phase, and they are absent in the G1 phase when differentiation is allowed (Gonzales et al., 2015). This is mediated by the ATM/ATR-CHEK2-p53 and Cyclin B1 pathways, active during the S and G2
phases inhibiting pluripotent state dissolution. Interestingly, in support of these findings, it has been shown that mESCs in G2/M reprogram more efficiently in B lymphocytes and fibroblasts than of those in G1 (Tsubouchi et al., 2013). This result is particularly interesting because it shows that what has been described so far is valid not only in differentiation but also in the opposite direction, during reprogramming. During reprogramming, iPSCs show a similar cell cycle structure than PSCs, with a short G1 phase (Mikkelsen et al., 2008; Sridharan et al., 2009). However, alteration of the cell cycle occurs early during reprogramming, and confirming the idea that it is important for establishment of pluripotency rather than just maintenance. Interestingly, the ectopic expression of CycD/CDK4 or downregulation of pRB increases reprogramming efficiency up to 10-fold, whereas knockdown of CycD/E or ectopic expression of CDK1 severely impairs it (Ruiz et al., 2011). This work suggests that promoting S phase entry increases reprogramming efficiency, whereas lengthening the G1 phase has the opposite effect.

All these observations taken together suggest that the G1 phase represents a so-called “window of opportunity”, when PSCs are able to respond to differentiation signals and are therefore competent for differentiation during this window of time (Figure 1.8). Importantly, these mechanisms have partially revealed how the cell cycle can direct cell fate choice, and has implied that manipulating cell cycle regulators could improve differentiation of stem cells. However, these results imply that other factors, such as chromatin structure and epigenetic status, may also be involved in controlling the transcriptional activity of signalling effectors in early/late G1 phase.

![Cell fate propensity during cell cycle progression](image)

**Figure 1.8 Cell fate propensity during cell cycle progression.**

Schematic representation of cell fate propensity during cell cycle progression: in the G1 phase cells are responsive to differentiation signal, whereas they sustain pluripotency during the S/G2/M phase.
1.1.4 Tools to study cell cycle progression

The study of the cell cycle and the related mechanisms have been challenged by the technical problems impairing analyses of their regulations in vivo and also by the lack of suitable in vitro model systems. Model organisms are effective to uncover the basics of cell cycle control, but more challenging to investigate human specific features, especially in early development. The techniques that can be used spans from conventional light microscopy and live imaging to more complex genetically modified cell lines, and can be obtained by either physical or chemical fractionation.

Early physical methods, such as counterflow centrifugal elutriation or density gradient centrifugation, rely on cell fractionation based on size and sedimentation density (Guidozzi, 1997; Lindahl, 1948, 1956). However, these techniques are relatively slow and the fractions quite often substantially overlap, increasing the heterogeneity of the samples. Alternatively, cells can be separated on the selection of mitotic cells that detach from flasks during culturing, with the so called mitotic shake-off (Terasima and Tolmach, 1963). An old approach that is still used is based on flow cytometry and cell sorting according to cell size and DNA content (Arndt-Jovin and Jovin, 1977; Juan et al., 2002). DNA is stained with the appropriate dye, such as Propidium iodide (Phenanthridinium compound), Hoeschst 33342 (Bisbenzimidazole), or DAPI (4′-6-diamidino-2-phenylindole), that quantitatively binds the DNA.

Chemical methods are more frequently reported, and mostly consist of a chemical arrest followed by a release period to allow a synchronous cell cycle progression. Different cell cycle inhibitors can be used to block cells at precise points during cell cycle progression, based on the biochemical process they are inhibiting. They include inhibitors of DNA replication (aphidicolin, hydroxyurea, mimosine and thymidine) or inhibitors of mitotic spindle formation (nocodazole, colchicine) (Deysson, 1968; Matsui et al., 2012; Zieve et al., 1980). Alternatively, cells can be blocked by addition of CDK inhibitors such as roscovitine. However, roscovitine has been reported to arrest cell cycle progression not only at the G1/S transition (Mgbonyebi et al., 1999), but also in G2/M checkpoints (De Azevedo et al., 1997). However, these traditional techniques used for synchronization introduce cell cycle perturbations and, quite often, cytotoxicity, or lead to differentiation in PSCs.
More recently, a new powerful tool has been introduced which allows unperturbed cells to be isolated based on their cell cycle position using fluorescence activated cell sorting. The FUCCI (Fluorescent, Ubiquitination-based Cell Cycle indicator) reporter system (Sakaue-Sawano et al., 2008) adapted to hESCs (Pauklin and Vallier, 2013) overcame these issues, allowing the analysis of cell cycle-specific events in hESCs (Figure 1.9).

The cell cycle is not only regulated at the transcriptional and post-translational levels, but also by ubiquitin mediated proteolysis (Ang and Harper, 2004; Nakayama and Nakayama, 2006). Two E3 ligases complexes, APC<sup>Cdh1</sup> and SCF<sup>Skp2</sup> ubiquitinate several proteins in a cell cycle dependent manner. SCF<sup>Skp2</sup> complex is active in the S and G2 phases and is a direct substrate of the APC<sup>Cdh1</sup> complex, which is active in the late M and G1 phases (Vodermaier, 2004). Therefore, these two ligase activities oscillate reciprocally during the cell cycle. The FUCCI system uses two direct substrates of the APC<sup>Cdh1</sup> and SCF<sup>Skp2</sup> complexes, GEMININ and CDT1, fused to two fluorescent proteins. CDT1, a DNA replication factor which is accumulated during the late G1 fused to mKO2 (Kusabira Orange), and GEMININ, a DNA replication inhibitor which is accumulated during the S/G2/M phase fused to mAG (Azami Green). This reporter system enables the real-time following of cell cycle progression, but also permits the sorting of cells by Fluorescence Activated Cell Sorting (FACS) according to their cell cycle phase: cells in early G1 do not express any fluorescent protein,
cells in late G1 express CDT1-mKO2 red, and cells in S/G2/M express GEMININ-
mAG green. The establishment of this system provides tremendous benefits to researchers interested in the role of the cell cycle in development. Several variations of the FUCCI system have been published so far, including a reporter for G0 arrest (Oki et al., 2014) and FUCCI4, which reports simultaneously on all four cell-cycle phases in live cells (Bajar et al., 2016).
1.2 Human Embryonic Stem Cells (hESCs)

As described in the previous section, the relatively low cell number available and the ethical implication of using embryos for human studies make the use of PSCs the only system available to investigate early stage of development. Model organisms can be used to elucidate early biological process, but there are fundamental differences in how development and, for our purpose, cell cycle is regulated in mice and humans. hESCs are one of the most powerful tool we have to try to recapitulate human developmental biology and validate mechanisms described in other model organisms. This section introduces hESCs, the transcriptional networks and signalling pathways regulating pluripotency, and concludes with an overview on their epigenetic state.

1.2.1 PSCs and hESCs

Human embryonic stem cells (hESCs) are pluripotent cells derived from the inner cell mass (ICM) at the pre-implantation blastocyst stage of a human embryo. They can proliferate indefinitely in vitro while maintaining the ability to differentiate into the three primary germ layers, endoderm, mesoderm and neuroectoderm (Figure 1.10) (Thomson et al., 1998).

![Figure 1.10 Human Embryonic Stem Cells.](image)

Human Embryonic Stem Cells (hESCs) are derived from the inner cell mass (ICM) of a pre-implantation human blastocyst. They can indefinitely self-renew in vitro while maintaining the ability to differentiate into the three primary germ layers, endoderm, mesoderm and neuroectoderm.
hESCs can have a normal karyotype over prolonged period of time, maintain high telomerase activity, and exhibit remarkable proliferative potential for long term in culture. Their derivation followed the previous generation of mESCs from the ICM of the mouse blastocyst (Evans and Kaufman, 1981; Martin, 1981), the first karyotypically normal embryonic stem cell line. Since these pioneering experiments, PSCs have been obtained from several other sources such as non-human primate (Thomson et al., 1995), various developmental stages, including the post-implantation epiblast (Brons et al., 2007; Tesar et al., 2007), the germ line (Boroviak et al., 2014; Kanatsu-Shinohara et al., 2004; Ko et al., 2009; Matsui et al., 1992), and by cell reprogramming (Takahashi et al., 2007; Takahashi and Yamanaka, 2006). These different cell types exhibit distinctive properties depending on their developmental stage of derivation and the maintenance conditions (Figure 1.11). By convention, PSCs derived from the pre-implantation embryo are known as ES cells or naïve, whereas lines generated from slightly later stages are called EpiSCs (Brons et al., 2007; Tesar et al., 2007) or primed (Nichols and Smith, 2009). The hallmark of naïve pluripotency is the formation of chimaeras, but for humans this is comprehensibly restricted on ethical grounds. Non-human primate ES cells fail to generate chimaeras with pre-implantation embryos (Tachibana et al., 2012), and conventional hESCs have been reported to inefficiently do so (James et al., 2006). Therefore, conventional hESCs, even if derived at similar pre-implantation stage to naïve mESCs, exhibit many features that makes them more similar to primed EpiSCs.

![Developmental potential of stem cells](image)

**Figure 1.11 Developmental potential of stem cells.**

Totipotency is defined by the capacity to develop an entire organism, including extra-embryonic tissues. Naïve PSCs form teratoma and a chimaeric animal when injected into pre-implantation embryos, whereas primed PSCs form teratomas, but form chimaeras less efficiently. Tissue-specific stem cells generate cell types related to a specific tissue, but do not form teratomas or chimaeras (adapted from De Los Angeles, 2015).
This can be explained by a possible adaptation into a primed state of the ICM-derived hESCs: it has been shown that PSCs lines derived from different mouse strains from the ICM can adapt in vitro into a primed state (Hanna et al., 2009). This is because of the genetic background of the donor, that seems to have an impact on the “naïve requirements” conditions, therefore inducing a primed state if these requirements are not fulfilled (Hanna et al., 2009). Other characteristics that makes hESCs more similar to primed pluripotency are the use of the OCT4 proximal enhancer, higher level of DNA methylation compared to mESCs, and a propensity for X inactivation (Hanna et al., 2010), while relying on Activin/Nodal/ the Transforming Growth Factor-β1 (TGFβ) (Vallier et al., 2005). Human pre-implantation embryos are hypomethylated, whereas ICM outgrowths undertake genomic re-methylation: hESCs maintain such methylation, similar to EpiSCs (Guo et al., 2014; Smith et al., 2014). Some pluripotency factors, such as KLF4, are expressed in mESCs and human pre-implantation epiblast, but repressed in hESCs and EpiSCs (Yan et al., 2013). However, some key transcription factors detected in mouse such as KLF2 seem to be absent in the human pre-implantation epiblast, and vice versa for KLF17, revealing some species-specific mechanisms (Blakeley et al., 2015). In general, hESCs can be considered to be “less primed” than EpiSCs: they do not upregulate Fibroblast Growth Factors 5 (FGF5) or N-cadherin, but express high levels of E-cadherin, NANOG and PRDM14 (Chia et al., 2010; Gafni et al., 2013). Recently, their methylation state has been described to be more similar to FBS/LIF condition rather than EpiSCs (Hackett et al., 2013; Shipony et al., 2014). Ultimately, they are able to generate germ cells, unlike EpiSCs (Gkountela et al., 2013; Kee et al., 2009).

Nevertheless, several human naïve PSCs have been reported (Gafni et al., 2013; Guo et al., 2016; Takashima et al., 2014; Theunissen et al., 2014). However, some of the signalling on which they rely on still differ from mESCs, suggesting that there might be some species-specific requirements (Weinberger et al., 2016) or that they represent different pluripotency states (Smith, 2017), suggesting that further investigation into the very early transition in development is required. It is also important to note that naïve cells must first transition to the primed state in order to further differentiate into the three germ layers. It has been hypothesized that naïve hPSCs have a broader and less biased differentiation potential due to lack of certain epigenetic marks (Takashima et al., 2014), but no extensive investigation confirmed this hypothesis yet. In addition to this, naïve cells show loss of imprinting (Pastor et al., 2016), and the epigenetic differences appear to be
less influential than genetic variability (Kyttala et al., 2016). Taken all this together, while naïve cells represent an important tool to study the transition from totipotency to pluripotency and early development in human, conventional hESCs have more potential for translational application.

1.2.2 Pluripotency transcriptional network

The pluripotency state is ruled by an extensively interconnected pluripotency genes regulatory network, mostly conserved among the different pluripotent stages. This regulatory network is able to sustain self-renewal, while maintaining differentiation abilities, and keeping critical developmental genes in an inactive but poised state. The core pluripotency factors are OCT4 (OCT3/4, POU5F1), SOX2 and NANOG (Figure 1.12).

![Core Pluripotency Transcriptional network](image)

In mESCs, OCT4, SOX2 and NANOG promote self-renewal while blocking extraembryonic differentiation. In hESCs all the factors promote self-renewal but each factor controls specific fates repressing embryonic lineage specification.

OCT4 known is uniquely expressed in PSCs and primordial germ cells (PGCs), and is essential for both *in vivo* and *in vitro* pluripotency (Nichols et al., 1998; Niwa et al., 2000; Scholer et al., 1989). During mouse development, *Oct4* is expressed from the oocyte to the post-implantation epiblast and the emerging mesendoderm, and it is downregulated gradually from the anterior to the posterior part of the embryo (Palmieri et al., 1994; Rosner et al., 1990; Scholer et al., 1990). It is required for PSCs derivation and ICM specification, inhibiting trophoderm (TE) differentiation (Nichols et al., 1998), and accordingly its loss induces TE
differentiation in mESCs (Niwa et al., 2000). Oct4 overexpression, on the contrary, triggers mesendoderm differentiation (Ivanova et al., 2006). In hESCs there has been discrepancy in results reported by different groups, but it seems that the levels of Bone Morphogenetic Protein 4 (BMP4) pathways regulates different developmental fates: loss of OCT4 drives ectoderm differentiation in the absence of BMP4 but specifies extraembryonic lineages in high BMP4 conditions; whereas its overexpression has little impact on both self-renewal and differentiation (Wang et al., 2012).

SOX2 is required for the epiblast formation as its knockout in the embryo is lethal at the post-implantation stage, and it is known to work as a heterodimer with OCT4, regulating its expression (Avilion et al., 2003; Masui et al., 2007). It is interestingly also expressed in several lineages at later developmental stages: is it highly expressed in the anterior portion of the post-implantation epiblast (the region that will give rise to neuroectoderm), to then being expressed in the brain, the neural tube, and the gut endoderm (Avilion et al., 2003; Li et al., 1998; Zappone et al., 2000). In mESCs, loss of Sox2 promotes trophectoderm differentiation, whereas its overexpression leads to neuroectoderm differentiation (Thomson et al., 2011). In hESCs instead, loss of SOX2 has a little impact on self-renewal and in long term, but if combined with SOX3 knockdown, it triggers the expression of mesendoderm markers (Wang et al., 2012). SOX2 overexpression does not have any impact on self-renewal, but enhances neuroectoderm differentiation (Wang et al., 2012).

NANOG is required for the epiblast specification by blocking Primitive Endoderm (PrE) differentiation (Chazaud et al., 2006). During mouse development, Nanog is expressed in the pre-implantation epiblast, but post-implantation is restricted to the primitive streak, to then gradually being dowregulated (Chambers et al., 2003; Mitsui et al., 2003). Nanog conditional knockout is compatible with self-renewal, but enhances the propensity to differentiate into TE and PrE (Ivanova et al., 2006). However, mESCs lacking Nanog cannot be derived (Chambers et al., 2003), but they can self-renew when established by genetic deletion (Chambers et al., 2007). On the contrary, when overexpressed, NANOG enables LIF-independent self-renewal in mESCs (Chambers et al., 2003). Similarly, NANOG overexpression in hESCs enables Activin/Nodal-independent self-renewal (Vallier et al., 2009a), however, its knockdown induces neuroectoderm and impairs mesendoderm differentiation (Teo et al., 2011; Vallier et al., 2009a).
Genome-wide studies have shown co-localisation of these three factors on the chromatin, co-occupying a substantial portion of their target genes and their own regulatory regions (Boyer et al., 2005). In this way, they provide a platform for recruiting additional factors to integrate regulatory inputs from signalling pathways and epigenetic mechanisms (Hackett and Surani, 2014; Young, 2011).

Albeit the core pluripotency factors are shared between the naïve and primed state (Nichols and Smith, 2009), there are some notable differences. For instance, in naïve mESCs the core pluripotency factors OCT4, SOX2, and NANOG (OSN) are irreplaceable, promoting the expression of pluripotency markers by mostly repressing extra-embryonic differentiation (Boyer et al., 2005). They are surrounded by additional pluripotency factors (Klf2, Klf4, Sall4, Essrb, and Tbx3) that can be replaced, and most importantly are downregulated in primed hESCs (Dunn et al., 2014; Ivanova et al., 2006). Another important difference is that in hESCs, OSN maintain pluripotency by blocking lineage specification, but also have important roles in driving differentiation (Wang et al., 2012).

### 1.2.3 Signalling pathways regulating pluripotency

The core pluripotency factors uphold the central network of signalling pathways necessary for self-renewal. The first PSCs line was derived and maintained on a Mitotically-inactivated mouse Embryonic Fibroblasts layer of cells (MEF) in the presence of a conditioned media obtained by embryonal carcinoma cells and serum (Martin, 1981), which turned out to be a source of LIF, WNT, BMP4 and other factors. It is important to remember that the embryonic equivalent of PSCs is not a stable stem cell pool, but it is regulated in a way that rapidly induce differentiation at the gastrulation stage. The core pluripotency factors indeed promote differentiation through the activation of Fgf4, an autocrine factor that controls the FGF4-Extracellular signal-Regulated Kinase (ERK) (Kunath et al., 2007; Yuan et al., 1995). ERK activation regulates RNA polymerase II (RNAPII) and the Polycomb Repressive Complex 2 (PRC2) occupancy at developmental genes, to maintain them in a poised state. Accordingly, ERK or FGF4 inhibition in PSCs blocks all differentiation potential (Tee et al., 2014), and this can be achieved in culture with LIF and BMP. LIF signals through the Signal Transducer and Activator of Transcription 3 (STAT3), to activate Klf4 and Tfcp2l1 (Niwa et al., 1998) and maintain pluripotency, primarily through blocking mesendoderm differentiation. Interestingly, overexpression of both Klf4 and Tfcp2l1 makes LIF dispensable for maintenance of self-renewal, via upregulation of Nanog (Niwa et
An additional downstream effector of LIF is the B Cell Lymphoma 3 protein (BCL-3), that seems to interact with Oct4 and β-catenin (Chen et al., 2015). BMP instead blocks neuroectoderm differentiation through the SMAD transcription factors (Ying et al., 2003). For this reason, conventional PSCs cultures relies on a combination of foetal bovine serum (FBS) and LIF (Smith et al., 1988). However, it has been demonstrated that these conditions result in heterogeneous populations, and only a subset of cells represent the truly naïve pluripotent state (Niakan et al., 2010; Singh et al., 2007; Toyooka et al., 2008). In order to enhance the naïve state, it has been shown that mESCs can be maintained without extrinsic LIF and BMP by dual inhibition (2i) of the MAPK/ERK kinase (MEK)–ERK pathway and glycogen synthase kinase 3 (GSK3) using the two small molecules inhibitors PD0325901 and CHIR99021 (Ying et al., 2008). Even if not strictly necessary, LIF in combination with 2i (2i/LIF) is used to enhance self-renewal.

Since their derivation, it has been evident that hESCs were obviously different from their murine counterpart in their phenotype and growth factors requirements (Figure 1.13). Their growth is LIF/STAT independent (Humphrey et al., 2004), and whilst BMP4 is not required (Xu et al., 2002), its inhibition promotes self-renewal (Levine and Brivanlou, 2006). They instead rely on TGFβ1/Activin A/Nodal and FGF2 pathways to maintain pluripotency and control cell fate decisions (Vallier et al., 2005; Vallier et al., 2004a). Initially thought to be differences caused by species divergences, it is now believed that they represent two different developmental stages. Indeed, hESCs share these features with EpiSCs (Brons et al., 2007), even if they can be considered to be slightly “less primed” than EpiSCs (see 1.2.1). Inhibition of Activin/Nodal signalling in hESCs induces neuroectoderm differentiation (Brown et al., 2011; Vallier et al., 2009a; Xu et al., 2008), whereas inhibition of FGF receptors can be rescued by and increased concentration of Activin (Vallier et al., 2005). These features are shared with hiPSC (Vallier et al., 2009b), therefore the most used hPSCs culture condition are feeder-free, in chemical defined media in the presence of Activin/Nodal/TGFβ and FGF2. Activin/Nodal/TGFβ pathway signals through several heteromeric transmembrane receptor complexes (two type I and two type II serine/threonine kinase receptors) (Wrana et al., 1992), recruiting and activating by phosphorylation type I Activin receptor-Like Kinases (ALK) (Reissmann et al., 2001). These receptors then phosphorylate two important transcription factors, SMAD2 and SMAD3 (Nakao et al., 1997), part of the Receptor-regulated Sma- and Mad-related proteins (R-
SMAD) family (Graff et al., 1996; Zhang et al., 1996). Consequently, SMAD2/3 form homo- and heteromeric complexes that bind to the common-mediator SMAD (co-SMAD) SMAD4 (Lagna et al., 1996), allowing nuclear translocation and gene activation. SMAD2/3 is inactivated by de-phosphorylation by PP1MA (Lin et al., 2006), or alternatively its activation can be self-inhibited by Inhibitory-SMAD7 (I-SMAD) in a negative feedback loop (Hayashi et al., 1997). Receptor binding is antagonised by the Nodal inhibitors LEFTY1/2 and CER1 (Bouwmeester et al., 1996; Meno et al., 1996) and the Activin inhibitor Follistatin (Harrison et al., 2005).

FGF2 activates the FGF receptors (FGFRs) stimulating the Mitogen-Activated Protein Kinase (MAPK) pathway (Dvorak et al., 2005; Li et al., 2007), and subsequently regulating NANO activity (Xu et al., 2008), which leads to suppression of both cell death and apoptosis genes (Eiselleova et al., 2009).

To sustain pluripotency, several other pathways can be activated which are not necessary to maintain self-renewal. The Insulin-like Growth Factors (IGFs) sustain pluripotency by reinforcing the activation of the Phosphatidylinositol 3-Kinase (PI3K)/Protein Kinase B (PKB/AKT) pathway (Wang et al., 2007), which is already activated by the FGF signalling (Eiselleova et al., 2009), resulting in increasing proliferation (Campbell et al., 2012) whilst blocking mesendoderm specification (McLean et al., 2007). Low level of WNT/β-catenin have been reported to promote the primed pluripotent state. WNT is known to mark the primitive streak in the embryo and is required for gastrulation (Liu et al., 1999; Mohamed et al., 2004), and high levels promote mesendoderm differentiation (Singh et al., 2012).

![Diagram](image)

**Figure 1.13 Signalling pathways regulating pluripotency in hESCs.**

The pluripotency state relies on the Activin A/Nodal and Fibroblast Growth Factors 2 (FGF2) pathways, the insulin-like growth factor (IGF) pathway, and low level of WNT/β-catenin pathway. Contrary to mESCs, Bone Morphogenetic Protein 4 (BMP4) is not required and its inhibition promotes self-renewal.
1.2.4 Chromatin epigenetic control of pluripotency

Chromatin structure and its epigenetic regulation are tightly linked to the transcriptional network and the signalling pathways regulating pluripotency. It is known that compared to their differentiated counterpart, PSCs have a generally de-condensed chromatin, which resembles the mouse ICM in vivo (Ahmed et al., 2010). Its conformation rearranges extensively during fate acquisition, and it is formed in a hierarchical manner (Efroni et al., 2008). Briefly, nucleosomes form clutches that engage in dynamic longer distance loops. Some of these loops give rise to gene domains within Topological-Associated Domains (TADs) that are established or stabilised by protein complexes (CTCF and cohesin) and regulatory components (transcription factors or Polycomb). Interaction among TADs of the same epigenomic type forms compartments (A and B) and combination of compartments in the same chromosome forms chromosome territories (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012) (Figure 1.14).

Figure 1.14 3D genome organisation.

Inside the nucleus, chromatin is divided in A/B compartments, and lamina-associated domains (LADs). Each compartment contains several 3D structures called topological-associated domains (TADs). In A compartments (green), chromatin is open and mostly nucleosome free, therefore permissive for transcription. This allows high occupancy of the RNA polymerase II (RNAPII, dark green) complex and transcription factors (TF). In this conformation, TADs boundaries are demarcated by CCCTC-binding factor (CTCF, blue), cohesins (pink ring) and mediators (grey), a conformation that allow the formation of promoter–enhancer loops and activation of the transcriptional machinery. In B compartments (red), chromatin is condensed and inactive. A and B compartments are characterised by a prominent epigenomic signature, correlating with their activation state. LADs are composed of condensed heterochromatin enriched at the nuclear periphery and physically bound to lamin proteins (Adapted from Rajarajan, 2016).
3D organisation of the genome
In any given nucleus, the DNA is wrapped around nucleosomes, which are made of 146/147bp of DNA coiled around a histone octamer (two H2A-H2B dimers, and a H3-H4 tetramer) (Luger et al., 1997). Nucleosomes are linked by the linker histone H1 and a short sequence of linker DNA (Harshman et al., 2013). They are assembled in heterogeneous groups of varying sizes, called clutches, which are interspersed with nucleosome-depleted regions (Ricci et al., 2015). These characterise two different chromatin compartments. A-compartments are mostly nucleosome free regions, characterised by an open chromatin state permissive for transcription, highly occupied by RNAPII and Transcription Factors (TFs). B-compartments are characterised by condensed chromatin, nucleosome enriched, with lower TFs and RNAPII occupancy (Lieberman-Aiden et al., 2009). Recently, it has been shown that these two main compartments can be further subdivided into six different sub-compartments (two active A compartment and four inactive B compartment), each bearing a distinctive pattern of epigenetic features (Rao et al., 2014; Wang et al., 2016; Wijchers et al., 2016). Each compartment contains several megabases of DNA compacted in 3D structures called TADs (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). They have an average size of ~185 kb (range 40kb-3Mb), and are characterised by a prominent epigenomic signature, correlating with their compartment super-structure. Interestingly, constituent loci within the same TAD interact with each other much more frequently than with loci located in outside these domains (Rao et al., 2014). Stable chromatin loops are observed at the borders of ~40% of the domains, and these boundaries are also enriched for insulator proteins such as CCCTC-binding factor CTCF (~80%), active transcription marks, and housekeeping genes (~35%) (Dixon et al., 2012). This spatial partition of the genome into TADs correlates with many linear genomic features such as coordinated gene expression and histone modifications, and especially enhancer–promoter interactions seem to be mostly constrained within a TAD (Shen et al., 2012). These features are indeed known to mostly control cell-type specific gene expression (Buecker and Wysocka, 2012; Ong and Corces, 2011). Enhancers are occupied by TFs that in turn recruit RNAPII via co-activators to target genes triggering their activation (Banerji et al., 1981; Benoist and Chambon, 1981; Gruss et al., 1981; Tjian and Maniatis, 1994). These regulatory regions can be located far upstream or downstream to the promoter of the gene, and their regulatory function is achieved by looping to the promoters of the target gene (Ernst et al., 2011; Hnisz et al., 2013). Essential for chromatin looping and
3D genome organisation are the mediator complex (Kagey et al., 2010), cohesin (Hadjur et al., 2009) and CTCF (Splinter et al., 2006). The mediator complex is found at promoters and enhancers of active genes and it enables RNAPII elongation (Kagey et al., 2010). It interacts with cohesin, another protein also involved in sister chromatid cohesion (Nasmyth and Haering, 2009), that interacts with CTCF. The formation of an enhancer-promoter loop starts from the binding of a TF to the promoter and enhancer, and the recruitment of the co-activator complex mediator, which in turn recruits the cohesin complex. The cohesin complex acts as a ring that creates or stabilise the loop. CTCF operates in parallel or synergistically to Mediator (Phillips-Cremins et al., 2013), allowing the activation and modulation of the RNAPII core transcription machinery.

The 3D genome includes also Lamina-Associated Domains (LADs) composed of condensed heterochromatin and physically bound to Lamin proteins at the inner nuclear lamina (Guelen et al., 2008). Finally, at even larger scales, chromatin is organised into individual chromosome territories (one for each chromosome), which rarely interact (Lichter et al., 1988; Lieberman-Aiden et al., 2009; Pinkel et al., 1988).

Epigenetic marks
In order to maintain remodelling potential, the 3D pluripotent structure of the genome must be highly dynamic. All its compaction levels are characterised by different ways of imposing an epigenetic control over these features. This is regulated by several epigenetic modifications, that cooperate to regulate local chromatin accessibility and allow TFs binding. Among them, histone modifications (Kouzarides, 2007) and DNA methylation (Smith and Meissner, 2013) are the best studied.

Histone tails can be marked by several covalent changes; at least eleven types of modifications have been reported including methylation, acetylation, phosphorylation, ubiquitination and sumoylation. They represent one of the most dynamic feature in the genome, and have a substantial impact on transcriptional regulation and chromatin remodelling via three main mechanisms (Kouzarides, 2007). First, histone modification can directly alter chromatin packaging modulating nucleosomes interactions (Ruthenburg et al., 2007). Second, they can recruit specific binding proteins to directly regulate gene expression (Wysocka et al., 2005). Third, they can inhibit the binding of transcriptional regulators to chromatin, blocking gene expression (Margueron et al., 2005). Histone modifications are
involved in gene activation (for example, Histone 3 Lysine 4 trimethylation [H3K4me3], Histone 3 Lysine 9 acetylation [H3K9ac], etc.) or in gene suppression (for example, Histone 3 Lysine 9 trimethylation [H3K9me3], Histone 3 Lysine 27 trimethylation [H3K27me3], etc). Histones in PSCs are marked by an increased amount of activating marks and a decrease in suppressive marks (Efroni et al., 2008; Meshorer et al., 2006). These are regulated by two main complexes of chromatin-modifying factors: the Trithorax Group (trxG) and Polycomb Group (PcG) proteins (Schuettengruber et al., 2007). The trxG proteins were discovered as activators of HOX genes in Drosophila, and a subset of these has been found to catalyse the trimethylation of H3K4. In mammals, this is catalysed by multi-subunit complexes that include SET1A, SET1B, and the Mixed Lineage Leukaemia (MLL) proteins 1-4 (Shilatifard, 2012). On the contrary, the PcG proteins were identified as silencers of HOX genes (Schuettengruber et al., 2007). PcG proteins form the multi-subunit Polycomb-repressive complexes (PRCs) 1 and 2 (Margueron and Reinberg, 2010; Simon and Kingston, 2013). PRC2 catalyses the deposition of the repressive H3K27me3 mark, whereas PRC1 is recruited by this mark to repress gene expression inducing RNAPII pausing via ubiquitination of Histone 2A Lysine 119 (H2AK119ub) (Margueron et al., 2009; Min et al., 2011; Morey et al., 2012). This is catalysed by RING1B, the main component of PRC1 complex, and the deposition of this mark in turn induces PRC2 recruitment to reinforce transcriptional repression (Cao et al., 2005; Cooper et al., 2014; Wang et al., 2004). RNAPII has been shown to be bound at silent bivalent genes in ESCs, and it is maintained by PRC1-mediated H2AK119ub. Removal of RING1A/B results indeed in changes in the RNAPII complexes composition, and subsequently, in transcriptional activation (Stock et al., 2007). In addition, a number of RNAPII variants have been reported at PRC target sites. For example, developmental regulators are bound by unproductive RNAPII, whereas other genes are bound by elongating RNAPII. Indeed, RING1B depletion shows that PRC1 and unproductive RNAPII bind to the same chromatin and maintain developmental regulators (Brookes et al., 2012). Other compositions of the PRC1 complex and have been described and they each appear to directly compact chromatin (Gao et al., 2012; Tavares et al., 2012). Together with RING1B, all PRC1 complexes are alternatively constituted by one of the six Polycomb group ring-finger domain proteins (PCGF1–PCGF6). Canonical PRC1 (cPRC1) complexes include PCGF2/4 and are defined by one chromobox protein (CBX2/4/6–8) that binds H3K27me3 and a Polyhomeotic homologous protein (PHC1–PHC3), which contains a sterile alpha motif (SAM)
domain controlling repression (Robinson et al., 2012). Non-canonical PRC1 (ncPRC1) have instead the zinc-finger domain and YY1-binding protein (RYBP), which associates with PCGF1 (ncPRC1.1), PCGF3/5 (ncPRC1.3/PRC1.5), or PCGF6 (ncPRC1.6) (Reviewed in (Schuettengruber et al., 2017). The activity of each complex is characterised by the different PCGF protein, dependent on the presence of RYBP (Rose et al., 2016). Apart from the canonical recruitment via H3K27me3, PRC1 in its different configurations has been shown to be recruited independently to chromatin marks in certain genome contexts (Kahn et al., 2016), however a strong correlation with hypomethylated CGIs has been reported (Boyer et al., 2006). Therefore, it is possible that PRC1 is recruited to target genes via multiple interactions with diverse TFs according to the different configurations, similar to induction of transcription via interactions with multiple enhancer-bound TFs. In addition to this repressive function, PRC1 has been described having regulatory roles in chromatin loop formation but also higher order interactions, in order to form Polycomb-containing TADs (reviewed in (Entrevan et al., 2016)). Importantly, PRC1 plays a critical role in the establishing Polycomb long-range interactions via oligomeration mediated by the SAM domain of the Ph/PHC1 subunit of cPRC1 (Kundu et al., 2017).

Genome-wide mapping in PSCs have shown how histone marks – or combination of them – are distinctive of precise genomic regions: Histone 3 Lysine 4 monomethylation (H3K4me1) and Histone 3 Lysine 27 acetylation (H3K27ac) define active enhancers, H3K4me3 active promoters, and H3K27me3 repressed promoters (Calo and Wysocka, 2013; Zhou et al., 2011). Interestingly, the presence of a large group of developmental gene promoters containing both activating H3K4me3 and repressing H3K27me3 marks have been described (Azuara et al., 2006; Bernstein et al., 2006). These marks, termed bivalent, are prevalent on developmental genes that show little or no expression in the pluripotent state, but are thought to be in a poised state enabling rapid activation upon differentiation cues. Bivalent domains are mostly found on Cytosine-phosphate-Guanine dinucleotide (CpG) islands, which are CpG-dense regions that are largely resistant to DNA methylation, prevalent at the Transcription Start Sites (TSS) of housekeeping and developmental genes (Bernstein et al., 2006).
1.3 Cell cycle regulation of pluripotency

Cell division and the transition from the M phase to the G1 is associated with dramatic changes in nuclear architecture, including restructuring of the nuclear envelope, chromosome de-condensation and extensive chromosome repositioning in 3D space (see 1.1). In the presence of differentiation signals, G1 phase would potentially establish a favourable epigenetic and structural environment that allows developmental programs to be activated, defining this phase as a window of opportunity for differentiation. This favourable environment is a combination of permissive chromatin state, highly dynamism of cell cycle regulated proteins, and a unique combination of epigenetic marks.

1.3.1 Chromatin organisation during cell cycle

The 3D organisation of the genome is highly dynamic and requires a tight regulation in order to be established, maintained, and reset each cell cycle, whilst extensively changing during fate acquisition. In PSCs, chromatin is open and much less compact when compared to somatic cells, implying that the overall structure is less condensed (Efroni et al., 2008). Its organisation and accessibility are predictive of gene expression, and DNA replication timing is correlated with an accessible chromatin structure (Orkin and Hochedlinger, 2011). Chromatin modifications and conformational changes during the cell cycle occur through two sequential events. First, during the S phase, newly synthesised DNA is wrapped around histones and Post-Translational Modifications (PTMs) are ready to be re-established. Then, during mitosis, many chromatin remodelling complexes and transcription factors detach from chromatin and the nuclear structure breaks down. As described in Chapter 1.1.1, during the M phase chromosomes change dramatically their conformation, suggesting that many organisational aspects of the interphase chromosome, such as TADs, are lost, and transcription is shut down. DNA staining shows that mitotic chromosomes are characterised by a mostly cell type-invariant banding pattern (Craig and Bickmore, 1993). Interestingly, immunofluorescence analyses have demonstrated that metaphase chromosomes are still composed of different domains characterised by a diverse chromatin status (Terrenoire et al., 2010), which have been speculated to correspond to A/B compartments in interphase. Recently, using Chromosome Conformation Capture Carbon Copy (5C), followed by high-throughput sequencing (Hi-C), it has been shown that loop enrichment and TADs insulation emerge very early in the G1
phase, decrease during the S phase as soon as replication commences, and it is ultimately completely absent in mitotic cells. However, A/B compartments appear to have a delayed trend: they are undetectable in G1 phase, to only reappear during the S phase, plateauing in G2, and disappearing during mitosis (Nagano et al., 2017; Naumova et al., 2013). Interestingly, these studies confirm the concept that chromosomal structure is not inherited through mitosis but it is established de novo during the early G1 phase (Thomson et al., 2004). However, mitotic chromosomes are thought to hold an epigenetic memory of the corresponding cell type, in order to transfer the information about the gene set that has to be switched on in the next cell cycle (Zaidi et al., 2010). Considering that TADs have been described as the most invariant units of chromatin structure, it is thought that their borders can be “bookmarked”, to allow quick rebuilding of the 3D structure after cell division, and restore the locations of previously active genes and regulatory elements. It appears that several key factors remain associated with chromatin through mitosis (Follmer et al., 2012; Kadauke et al., 2012; Yan et al., 2013a), or that some chromatin regions simply have to remain accessible so that specific factors can re-bind in the next G1 (Martinez-Balbas et al., 1995), and this is sufficient to drive self-assembly of higher 3D structures. Examples are FOXA1 in adult liver (Caravaca et al., 2013), GATA1 in haematopoietic cells (Kadauke et al., 2012), Oct4, Sox2, Klf4, and H3K27ac in mESCs (Deluz et al., 2016; Liu et al., 2017; Teves et al., 2016).

To summarise, the model based on current literature is the following: during prophase, most TFs dissociate from chromatin and this may cause a loss of higher 3D structures such as compartments. Then, the chromosomes stochastically fold, whilst TADs boundaries and enhancers remain marked by bookmarking proteins or remain in a nucleosome-free status. This bookmarking does not affect the stochastic folding of the genome, but ensure the rapid re-establishment of TADs and re-binding of TFs after division to maintain cell identity. After TADs are established, A/B compartments are reconstructed, but only later in the S phase.

This can interestingly correlate with the previously described propensity for cells to differentiate in the G1 phase: absence of compartmentalisation might represent a favourable environment for fate change, whilst maximum insulation ensures quick regulation and promoter-enhancer looping at the required loci.
1.3.2 Cell cycle regulation of histone marks and methylation
As hESCs transition through G1, not only the chromosomal architecture remodels, but also epigenetic modifications influence this permissive state. The globally open chromatin structure is characterized by high levels of epigenetic marks indicative of active transcription, such as H3K4me3 and H3/H4 acetylation (Azuara et al., 2006). Histones biogenesis itself is cell cycle regulated; they are actively synthesised in the S phase, regulating DNA replication speed (Groth et al., 2007). Entry into S-phase is triggered by CycE/CDK2, which also phosphorylates Nuclear Protein Ataxia-Telangiectasia locus (NPAT), to initiate transcription of the histone genes (Ma et al., 2000). Interestingly, it has been shown that increased biosynthesis of histones in S phase correlates with nucleosome-depleted hypersensitive regions, and with the cell cycle regulated recruitment of RNAPII and NPAT to the H3 and H4 locus, triggering their transcription (Medina et al., 2012), indicating cell cycle regulated binding patterns of TFs. Important for the inheritance of the modifications on the newly synthesised histone is the cell cycle regulated activity of the PRC2 complex. To allow for H3K27me3 to be deposited on new histones in the S phase, the PRC2 complex is recruited to mature histones already marked by the H3K27me3 modification during the G1 phase (Hansen et al., 2008). Additional examples of the tight interplay between the cell cycle machinery and PRC2 mediated gene silencing is the evidence that CDK1 and CDK2 can phosphorylate Enhancer of Zeste Homolog 2 (EZH2), the catalytic subunit responsible for the methylation activity of PRC2, to reinforce the silencing of differentiation genes through deposition of H3K27me3 (Chen et al., 2010). However, especially at bivalent domains, other histone modifications have been shown to fluctuate in self-renewing hESCs, with a transient increase in H3K4me3 at developmental genes during the G1 phase (Singh et al., 2015). Consistent with this, developmental genes are transiently transcribed at low levels in G1, to then become fully express in the presence of differentiation cues (Singh et al., 2013). This also coincides with higher level of 5-hydroxymethylation, known to promote demethylation and activation of target genes (Guo et al., 2011).

1.3.3 Transcriptional network changes during cell cycle progression
Such open chromatin state also implements the transcriptional circuitry controlling pluripotency, mainly driven by OCT4, NANOG and SOX2, and it has been proposed that this state allows transcriptional networks to efficiently switch upon induction of differentiation (Meshorer and Misteli, 2006). Genome-wide studies
have shown co-localisation of these three factors on the chromatin, co-occupying a substantial portion of their target genes, and are known to be involved in both gene activation and repression in vivo, in order to maintain hESC pluripotency (Boyer et al., 2005). These pluripotency factors have also been demonstrated to control cell cycle progression. NANOG, for example, binds and regulates CDK6 and CDC25A, genes required for G1/S transition, and its overexpression accelerates S phase entry (Zhang et al., 2009). Another study in ECCs showed analogous features: NANOG overexpression upregulates several G1 regulators including CDK6, Cyclin E, and Cyclin Ds (Choi et al., 2012). A similar role has been proposed for OCT4 in mESCs, where it has been shown to be required for transition through the G1-phase: downregulation of Oct4 leads to G1 arrest, mediated by p21 upregulation (Lee et al., 2010). OCT4 has also been shown to form a complex with CDK1 to enhance its binding on the trophectoderm marker Cdx2 and repress trophectoderm differentiation (Li et al., 2012a). This interaction has been further characterised and appears to have other important roles in controlling cell cycle progression: Oct4 has been shown to inhibit Cdk1 activation to counteract high CycB concentration, to ensure normal G2/M progression and prevent premature mitotic entry (Zhao et al., 2014). In addition to this, it has been shown in hESCs that OCT4 and SOX2 bind to a conserved promoter region of miR-302, a cluster of eight micro-RNAs expressed specifically in hESCs able to induce reprogramming without any requirement for exogenous TFs through the activation of OCT4 (Anokye-Danso et al., 2011; Suh et al., 2004). MiR-302a represses the translation of CycD1 and CDK4, and its downregulation causes differentiation and G1 block (Card et al., 2008). OCT4 has been shown to also drive the expression of inhibitors of the pRB-phosphatase PP1 to maintain the hyper-phosphorylation status of pRB in pluripotency, and its downregulation during differentiation is driven by another miRNA, miR-335, which blocks Oct4 expression and pRB phosphorylation (Schoeftner et al., 2013). Conversely, the cell cycle has also been shown to regulate the core pluripotency network. For instance, it has been shown that SOX2 is phosphorylated by CDK2 and this is required for the establishment of pluripotency during reprogramming (Ouyang et al., 2015). Mechanistic analyses of the link between cell cycle progression and the transcriptional network have only recently become an area of investigation for the field. In mESCs, it has been shown that OCT4 binding to the chromatin is regulated by phosphorylation during mitosis: Aurora Kinase B (AURKB) phosphorylates OCT4 during G2/M phase to trigger its dissociation from the chromatin. PP1 dephosphorylates phospho-OCT4.
during the M/G1 transition, which prompts OCT4 to reset pluripotency transcription on re-entry into the following G1 phase (Shin et al., 2016).

1.4 Open questions and project aims
From the literature, it is clear that during cell cycle progression, evident chromatin remodelling events take place to ensure correct chromosomal segregation and cell division. This has been evident since the first observation of heterochromatin in 1930 (Muller and Altenburg, 1930). However, it has only recently become appreciated that this organisation is tightly linked to control of transcriptional landscape. Whilst there is an abundance of evidence demonstrating how cell cycle affects pluripotency and fate specification, the link between chromatin changes and the transcriptional network during cell cycle progression remains still unknown. Furthermore, the observation that the activity of the Activin/Nodal-SMAD2/3 signalling pathway is cell cycle dependent in hESCs, suggested that the transcriptional regulatory network itself could be cell cycle regulated. This could drive or be driven by both changes in chromatin structure and epigenetic modifications during cell cycle progression. Considering the importance of these regulations, not only for cell fate specification but also, more broadly, for developmental biology, we decided to investigate the interplays between chromatin modifications and transcriptional networks during cell cycle progression in hESCs.

The specific aims of this PhD projects are:
- To investigate the dynamic changes in chromatin accessibility during cell cycle progression.
- To study the regulation of transcription factors binding during cell cycle.
- To correlate changes in the binding pattern with gene transcription during the cell cycle.
- To understand how these mechanisms could control maintenance of pluripotency and induction of differentiation.
2 MATERIAL AND METHODS
2.1 Cell culture

FUCCI-Human Embryonic Stem Cells (hESCs) were previously derived from H9 WiCell (Madison, WI) and used for all the experiments. H9 FUCCI-hESCs were grown in Chemically Defined Medium (CDM), supplemented with Bovine Serum Albumin (BSA), recombinant Activin A (10 ng/ml) and Fibroblast growth factor 2 (FGF2, 12 ng/ml), both from Dr Marko Hyvonen (Department of Biochemistry, University of Cambridge) on 0.1% Gelatin and MEF-medium coated plates, and passaged weekly using Collagenase IV as previously described (Brons et al., 2007; Pauklin and Vallier, 2013). All the FUCCI-hESCs lines used were kept under selection with Puromycin (1mM) and Geneticin (G-418, 0.2 mg/ml), or alternatively G-418 (0.2 mg/ml) and Blasticidin (1mM). Media composition is summarised in Table 2.

Table 2. Cell Culture media composition

<table>
<thead>
<tr>
<th>Medium</th>
<th>Components</th>
<th>Concentration</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CDM BSA/PVA</strong></td>
<td>IMDM:F12 (1:1)</td>
<td>-</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td>Chemically defined lipids</td>
<td>1%</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td>Insulin</td>
<td>7μg/ml</td>
<td>Roche</td>
</tr>
<tr>
<td></td>
<td>Transferrin</td>
<td>15μg/ml</td>
<td>Roche</td>
</tr>
<tr>
<td></td>
<td>Penicillin-Streptomycin</td>
<td>1%</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td>1-thioglycerol</td>
<td>450mM</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td></td>
<td>Bovine Serum Albumin (BSA)</td>
<td>5mg/ml</td>
<td>Europa Bio</td>
</tr>
<tr>
<td></td>
<td>OR Polyvinyl alcohol (PVA)</td>
<td>1mg/ml</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td><strong>MEF</strong></td>
<td>Advanced DMEM F12</td>
<td>-</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td>Foetal bovine serum (FBS)</td>
<td>10%</td>
<td>Biosera</td>
</tr>
<tr>
<td></td>
<td>L-glutamine</td>
<td>1%</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td>β-mercaptoethanol</td>
<td>100μM</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td></td>
<td>Penicillin-Streptomycin</td>
<td>1%</td>
<td>Gibco</td>
</tr>
<tr>
<td><strong>Gelatin</strong></td>
<td>Embryo Transfer Water</td>
<td>-</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td></td>
<td>Gelatine from porcine skin</td>
<td>0.1%</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td><strong>Collagenase</strong></td>
<td>Advanced DMEM F12</td>
<td>-</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td>Knockout serum replacer (KSR)</td>
<td>20%</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td>L-glutamine</td>
<td>1%</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td>Collagenase IV</td>
<td>1mg/ml</td>
<td>Gibco</td>
</tr>
</tbody>
</table>

2.1.1 Germ layers specification

Definitive Endoderm (dEN) was induced for 3 days in CDM-Polyvinyl alcohol (PVA) (-Insulin) with 20ng/ml FGF2, 10μM LY294002, 100ng/ml Activin A, and 10ng/ml
BMP4, changing media daily, as previously described (D'Amour et al., 2005; Touboul et al., 2010).

Neuroectoderm was induced for 6 days in CDM-BSA with 12ng/ml FGF2 and 10μM SB431542 (Activin/Nodal/TGFβ inhibitor; Tocris), as previously described (Vallier et al., 2009c).

2.2 Molecular biology

2.2.1 Immunofluorescence (IF)
FUCCI hESCs were fixed for 20 min at 4°C in 4% paraformaldehyde (PFA)/Phosphate-Buffered Saline (PBS, Gibco) and washed three times with PBS. Non-specific bindings were blocked incubating cells with 10% Donkey Serum (Biorad) and 0.1% Triton X-100 (Sigma-Aldrich) PBS for 30 min. Cells were then incubated with the appropriate antibody (1:100, Table 3) in 1% Donkey Serum and 0.1% Triton X-100 PBS overnight at 4°C. Cells were then washed three times with PBS and incubate with the appropriate secondary antibodies (1:1,500; AlexaFluor donkey anti-goat/rabbit/mouse IgG, Invitrogen) in 1% Donkey Serum and 0.1% Triton X-100 PBS 1h at room temperature protected from light. Cells were then washed three times with PBS, for nuclei staining Hoechst (1:10,000, Sigma-Aldrich) was added to the first wash. Images were acquired using Zeiss LSM 700 Confocal Microscope, and Zen 2009 imaging software.

Table 3. Antibodies

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Catalogue number</th>
<th>IF</th>
<th>WB</th>
<th>Flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCT4 - ChIP Grade</td>
<td>ab19857 - abcam</td>
<td>1:200</td>
<td>1:1,000</td>
<td>1:200</td>
</tr>
<tr>
<td>OCT-3/4 (C-10)</td>
<td>sc-5279 - Santa Cruz</td>
<td>1:100</td>
<td>1:1,000</td>
<td>1:100</td>
</tr>
<tr>
<td>OCT-3/4 (N-19)</td>
<td>sc-8628- Santa Cruz</td>
<td>1:100</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NANOG</td>
<td>af1997 - R&amp;D Systems</td>
<td>1:100</td>
<td>1:500</td>
<td>1:100</td>
</tr>
<tr>
<td>SOX2</td>
<td>af2018 - R&amp;D Systems</td>
<td>1:100</td>
<td>1:1,000</td>
<td>1:100</td>
</tr>
<tr>
<td>SOX17</td>
<td>af1924 - R&amp;D Systems</td>
<td>1:100</td>
<td>-</td>
<td>1:100</td>
</tr>
<tr>
<td>CTCF</td>
<td>07-729 - Merk Millipore</td>
<td>1:100</td>
<td>1:1,000</td>
<td>-</td>
</tr>
<tr>
<td>RING1B</td>
<td>39663 - Active Motif</td>
<td>-</td>
<td>1:1,000</td>
<td>-</td>
</tr>
<tr>
<td>TRA-1-60</td>
<td>sc-21705 - Santa Cruz</td>
<td>1:50</td>
<td>-</td>
<td>1:100</td>
</tr>
</tbody>
</table>

2.2.2 Western Blotting (WB)
Cells were harvested by scraping in phosphate-buffered saline, centrifuged at 250g for 5 min at 4°C, and re-suspended in ice-cold CellLytic M buffer (Sigma-
Aldrich) supplemented with cOmplete Protease Inhibitor (Roche). After an incubation for 30 min on ice, lysates were clarified by centrifugation at 16,000g for 5 min at 4°C. The supernatants were transferred to new tubes, and proteins in cellular extracts were quantified using Protein Quantification Kit-Rapid (Sigma) according to manufacturer’s instructions. 1% β-mercaptoethanol (Sigma-Aldrich) was added to LDS NuPage Sample Buffer (Invitrogen NP0007) and protein denatured at 95°C for 5 min. Same amount of protein per sample (1μg to 50μg depending on the relative abundance) was loaded on a precast NuPAGE 4-12% Bis-Tris protein gel (Invitrogen) and run for 45min, 200V, 130 mA. Proteins were transferred from the gels to Immuno-Blot® PVDF Membranes (Bio-Rad) using the Bio-Rad mini-Trans Blot system for 30 min at room temperature, and blotted with the appropriate primary antibody (1:1,000, Table 2) overnight at 4°C. Afterwards, membranes were washed three times for 5 min with 0.05% Tween-20 PBS, incubated for 1h at room temperature with the appropriate secondary antibody (1:10,000, anti-mouse/rabbit/goat) conjugated to Horseradish Peroxidase (HRP, Sigma-Aldrich), and developed using the Pierce ECL detection kit (Thermo Scientific) according to manufacturer’s instructions.

2.2.3 Quantitative real-time PCR (qPCR)
RNA was extracted using the GenElute Mammalian Total RNA Miniprep Kit and the On-Column DNase I Digestion Set (both from Sigma-Aldrich) following manufacturer’s instructions. cDNA was synthesised using 500ng of RNA in a reaction with 250ng random primers, 0.5mM deoxynucleotides (dNTPs), 20U RNaseOUT, and 25U of SuperScript II (all Invitrogen) in a total of 20μl. cDNA was diluted 30-fold and 5μl were used for qPCR using SensiMix SYBR low-ROX (Bioline) with 150nM Forward and Reverse primers (Sigma-Aldrich; see Table 4 for primer sequences). Samples were run in technical duplicates on 96-well plates on a Stratagene Mx-3005P (Agilent) or 384-well plate on a QuantStudio Flex Real-Time PCR System, and results were analysed using the delta-delta cycle threshold (ΔΔCt) approach using RPLP0/PBGD as housekeeping genes. Primers were designed using PrimerBlast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/), and tested to produce a single PCR product, with an efficiency between 80% and 120%.
Table 4. qPCR primers

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>POUF51</td>
<td>FW AGTGAGAGGCAACCTGGAGA</td>
</tr>
<tr>
<td></td>
<td>REV ACACTGGGACACATCCCTTC</td>
</tr>
<tr>
<td>NANOG</td>
<td>FW CATGAGTGGATCCAGCTTG</td>
</tr>
<tr>
<td></td>
<td>REV CCTGAATAAGCAGATCCATGG</td>
</tr>
<tr>
<td>SOX2</td>
<td>FW TGGACAGTTACGCACAT</td>
</tr>
<tr>
<td></td>
<td>REV CGAAGACATGCTGTATGG</td>
</tr>
<tr>
<td>SOX17</td>
<td>FW CGCAGGGAATTGAAGTGA</td>
</tr>
<tr>
<td></td>
<td>REV GGATCAGGGACCTGTACAC</td>
</tr>
<tr>
<td>EOMES</td>
<td>FW ATCATTACGAAACAGGGCAGG</td>
</tr>
<tr>
<td></td>
<td>REV CCGGTTTGGATTTTGTAAGG</td>
</tr>
<tr>
<td>MIXL1</td>
<td>FW GGTACCAGACATCCTTGG</td>
</tr>
<tr>
<td></td>
<td>REV TAATCTCCGCGCTAGCCAAA</td>
</tr>
<tr>
<td>T</td>
<td>FW TGCTTCCCAGAGACCCAGTT</td>
</tr>
<tr>
<td></td>
<td>REV GATCACTTCTTTTCTTTGATCAAG</td>
</tr>
<tr>
<td>SOX1</td>
<td>FW+REV QUANTITECT primers (QIAGEN): QT00215299</td>
</tr>
<tr>
<td>PAX6</td>
<td>FW CTTTGCTGGAAATCAGG</td>
</tr>
<tr>
<td></td>
<td>REV AGCCAAGTGGTGGAAGACTC</td>
</tr>
<tr>
<td>PBGD</td>
<td>FW GGAGCCATGTCGTGTAACGG</td>
</tr>
<tr>
<td></td>
<td>REV CCACGGAATCCTCTCATCT</td>
</tr>
</tbody>
</table>

2.2.4 Flow cytometry

Single cell suspensions were generated by incubating cells in Cell Dissociation Buffer (CDB, Gibco) for 8 min at 37°C followed by repetitive pipetting. Cells were washed twice with PBS and fixed for 20 min at 4°C with PBS 4% PFA. After three washes with PBS, cells were first permeabilised for 20 min at RT with PBS 0.1% Triton X-100, then blocked for 30 min at RT with PBS 10% Donkey Serum. Cells were incubated with primary antibody (Table 2) for 1h at RT in PBS 1% Donkey Serum 0.1% Triton X-100, followed by secondary antibody in the same manner. Cells were then washed three times with the same buffer after each incubation. Flow cytometry was performed using a Cyan ADP flow-cytometer, and at least 20,000 events were recorded. Flow cytometry analysis was performed using FlowJo software.

2.2.5 Cell sorting

FUCCI-hESCs were washed with PBS and collected by incubating with CDB for 10 min at 37°C. When needed, crosslinking of proteins to DNA was performed prior to sorting, according to the specific experiment (see Chromatin Immunoprecipitation section). Cells were washed with 1% PBS/BSA and incubated in 1% PBS/BSA with
anti-TRA-1-60-antibody mouse monoclonal IgM (1:100, Santa Cruz Biotechnology, Inc.) for 20 min on ice, together with the secondary antibody, Alexa Fluor® 647 goat anti-mouse IgM (1:1,500, Thermo Fisher). Cells were washed with PBS, re-suspended in 6ml of PBS and sorted into the different cell cycle phases using BD FACS AriaTM III, BD Fusion or BD Influx sorters by the Cambridge NIHR BRC Cell Phenotyping Hub.

2.2.6 Cell synchronisation
FUCCI-hESCs were grown for 2 days after plating in standard conditions (see 2.1). Synchronisation was achieved incubating cells with 100 ng/ml of Nocodazole (Sigma-Aldrich M1404) for 16-20h in the required media, usually CDM-BSA supplemented with Activin and FGF2 (see 2.1). Cells were then washed once with warm media and downstream experiments performed.

2.3 Chromatin Immunoprecipitation (ChIP)
Crosslinking of protein to DNA was performed prior to sorting in 1% PBS/Formaldehyde, rocking for 10 min at room temperature. All the experiments were performed with single crosslinking and neutralised with Glycine (0.125M), for 5 min shaking at room temperature. ChIP was then carried out as described before (Brown et al., 2011). In sum, after sorting, cells were washed in PBS and sonicated for 20 times, 30 sec ON / 45 sec OFF, in pre-chilled sonication tubes (Bioruptor® Pico, Diagenode). A pre-clearing step was performed using the appropriate IgG (100 ng), for 1h rotating at 4°C, followed by Protein G Agarose beads (1:1 Protein G:PBS, Roche) incubation, in order to remove non-specific interactions. The supernatants were split in the appropriate number of tubes (1 per antibody/IgG, see Table 2), incubated overnight at 4°C, with 5μg of antibody. The protein/DNA complexes were immunoprecipitated with Protein G Agarose beads. Rnase A (1μg) was used to degrades the RNA and reverse the crosslinking effects of the formaldehyde, 5 hours at 67°C. An additional incubation with Dithiothreitol (DTT, 100mM, Sigma-Aldrich), 37°C for 30 min was performed in case of double crosslinking. Proteinase K (Sigma-Aldrich) was added overnight at 45°C in order to degrade all the protein. The DNA was precipitated with phenol/chloroform and ethanol (Sigma). All the steps were performed in presence of protease inhibitors (Leupeptin, Sodium Butyrate, and Phenylmethane sulfonyl fluoride [PMSF], all from Sigma-Aldrich).
Alternatively, after optimisations (see Chapter 3.2), ChIP was performed using LowCell# ChIP kit (C0101007, Diagenode) and DNA recovered using iPure kit (C03010012, Diagenode) according to manufacturer’s instructions, with the following modifications. 5μg of antibody were used per IP, and the quantity of magnetic beads adjusted accordingly, based on their binding capacity (10μl of beads bind ~3μg). Elution was made in Buffer C from the iPure kit, compatible with qPCR and next generation sequencing. For all the experiments, the same number of cells was used per IP after sorting, normalised for the less abundant cell cycle phase, usually EG1 ~1M cells.

Table 5. ChIP-qPCR primers

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>POUF51</td>
<td>FW ACATGCTTCGGAACAGGAGG</td>
</tr>
<tr>
<td></td>
<td>REV GGTGTGGAGATTCCAGCCAA</td>
</tr>
<tr>
<td>NANOG</td>
<td>FW GGTCTGTGGCTCGGTTTTC</td>
</tr>
<tr>
<td></td>
<td>REV ATGAGGAACCAGCTCAGTC</td>
</tr>
<tr>
<td>SOX2</td>
<td>FW CTAAGCAAGGCATTTTCCCCC</td>
</tr>
<tr>
<td></td>
<td>REV TGTGAGCAAGAACTGGCGAA</td>
</tr>
<tr>
<td>POU5F1</td>
<td>FW ACCCATTCCCTGTTCACTGAC</td>
</tr>
<tr>
<td>NEGATIVE</td>
<td>REV GGCCAGGGTCTCTTTTCTG</td>
</tr>
</tbody>
</table>

2.4 Assay for transposase-accessible chromatin (ATAC)
TRA-1-60* hESCs were sorted in the different cell cycle phases (EG1, LG1, G1/S, S/G2/M). Samples were generated from two independent experiments, 100,000 cells per sample, as previously described in (Kumasaka et al., 2016) with some modifications. In brief, cells were washed with ice-cold PBS and incubated with 500μl of ice-cold sucrose buffer (10mM Tris-Cl pH 7.5, 3mM CaCl$_2$, 2mM MgCl$_2$ and 0.32M sucrose) for 12 min on ice. To collect the nuclei, Triton X-100 was added to a final concentration of 0.5% and cells incubated for additional 6 min on ice after mixing. Samples were then resuspended in 1ml of sucrose buffer and transferred to a fresh 1.5 ml Eppendorf tube. The nuclei were centrifuged at 1500g for 3 min at 4°C and the supernatant discarded. Nuclei pellets resuspended in 50μl of Nextera DNA Sample Preparation Kit (Illumina FC-121-1030) for tagmentation, comprising 25μl 2x Tagment DNA buffer, 20μl nuclease-free water and 5μl Tagment DNA Enzyme 1. The tagmentation reaction mixture was transferred in a 1.5ml low-binding tube and incubated at 37°C for 30 min and stopped by the
addition of 250μl Buffer PB (Qiagen). The tagmented chromatin was purified using the MinElute PCR purification kit (Qiagen 28004), according to the manufacturer’s instructions, eluting in 11.5μl of buffer EB (Qiagen). For PCR amplification, 10μl of the tagmented chromatin was mixed with 2.5μl of Nextera PCR primer cocktail and 7.5μl of Nextera PCR mastermix (Illumina FC-121-1030) in low-binding PCR tube. Primers were from the Nextera Index kit (Illumina FC-121-1011), using 2.5μl of an i5 primer and 2.5μl of an i7 primer per PCR, totalling 25μl. PCR amplification was performed as follow: 72°C for 3 min and 98°C for 30 sec, followed by 12 cycles at 98°C for 10 sec 63°C for 30 sec and 72°C for 3 min. The sample volumes were then increased to 50μl by adding Qiagen EB buffer. The PCR primers were removed with 1x 0.9:1 SPRI beads (Beckman Coulter, Cat no. A63880) according to manufacturer’s instruction and samples eluted in 20μl. Following amplification, excess of primers was removed running the samples of 50ml 1% agarose TAE gel at 90V for 25 min, and size selection obtained by cutting from 120bp to 1kb. MinElute Gel Extraction kit (Qiagen 28604) was used to extract DNA, eluted in 20μl of buffer EB.

2.5 Genetic manipulations

The generation of the inducible lines was performed using lipofection, both for random integration (ERT2) and genomic targeting (iKD). Genomic targeting was followed by genotyping to screen for off-target events.

2.5.1 Transfection

Lipofection was performed as previously described (Vallier et al., 2004b). In brief, FUCCI-hESCs cells were seeded into 6 well plates using medium without Penicillin-Streptomycin (Pen/Strep, Gibco), in low density. After 48-72h (depending on colony size), cells were transfected using Lipofectamine® 2000 Transfection Reagent (Thermo Fisher), according to manufacturer’s instruction. In sum, medium was aspirate and replaced with 1mL/well of Opti-MEM® Reduced Serum Media (Thermo Fisher). 4μg of (each) plasmid was resuspended in 0.25ml of Opti-MEM medium/well. 10μl of Lipofectamine 2000 were dissolved in 0.25ml of Opti-MEM medium/well, and incubated for 5 min at room temperature. Lipofectamine 2000 and the plasmid were mixed together and incubated for 20 min at room temperature. 0.5ml of the DNA/Lipofectamine mix was added drop by drop to the well and incubate for 18-20h. Opti-MEM was then replaced with regular media without antibiotics. Antibiotic selection was started 4-5 days after the transfection
(Puromycin 1 mg/ml, G-418 50μg/ml). After 6 days, clones were picked into 24 wells plate. Each clone was passaged into 12 well plates, for analysis of transgene expression by qPCR, immunostaining or western blot.

2.5.2 Genotyping
hESCs clones from targeting experiments (Inducible KnockDown, iKD) were screened by genomic PCR to verify site-specific targeting, determine the number of alleles targeted, and exclude off-target effects. Genomic DNA was isolated using the Mammalian Genomic DNA Miniprep Kit (Sigma Aldrich) according to the manufacturer’s instructions. PCRs were performed using 100ng of genomic DNA as template in a 10μl reaction using LongAmp Taq DNA Polymerase (NEB) according to manufacturer’s instructions in 2% DMSO. The cycling parameters were as follows:

- **WT AAVS1 locus**
  - Denaturation at 94°C for 5 min, 94°C for 15 sec, primer annealing at 68°C for 30 min, extension at 65°C for 2 min, 65°C for 5 min.

- **5’ INT PCR**
  - Denaturation at 94°C for 5 min, 94°C for 15 sec, primer annealing at 68°C for 30 sec, extension at 65°C for 1 min 30 sec, 65°C for 5.

- **3’ INT PCR**
  - Denaturation at 94°C for 5 min, 94°C for 15 sec, primer annealing at 60°C for 30 sec, extension at 65°C for 1 min 30 sec, 65°C for 5 min.

- **5’ BB PCR** and **3 min BB PCR**
  - Denaturation at 94°C for 5 min, 94°C for 15 sec, primer annealing at 60°C for 30 sec, extension at 65°C for 2 min, 65°C for 5 min.

Primers sequences are listed in Table 6.

<table>
<thead>
<tr>
<th>Table 6. Genotyping primers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PCR Type</strong></td>
</tr>
<tr>
<td>WT</td>
</tr>
<tr>
<td>5’ END</td>
</tr>
<tr>
<td>3’ END</td>
</tr>
<tr>
<td>5’ BB</td>
</tr>
<tr>
<td>3’ BB</td>
</tr>
</tbody>
</table>
2.6 Bioinformatics analyses and pipelines

The following data processing pipelines have been used for all the bioinformatics analyses according to ENCODE (Landt et al., 2012). All the analyses have been run on the Sanger Farm Computer Cluster. All the output files can be found in the annex CD-ROM, listed in the appendix.

2.6.1 ChIP-seq

TRA-1-60* hESCs were sorted in the different cell cycle phases (EG1, LG1, S/G2/M). From two independent experiments, chromatin immunoprecipitation was performed as described in 2.3. Library preparation and sequencing were performed at the Wellcome Trust Sanger Institute next-generation sequencing facility. Size selection was applied, and fragments between 100bp and 400bp (average length of 200bp) were used to prepare barcoded sequencing libraries using NEBNext Sample Prep Kit1 (NEB) following manufacturer’s instructions. 10ng of input material (as a pool of the different phases) for each experiment was processed in parallel as control. To test whether pooled inputs were an acceptable control, one experiment was run with one input per cell cycle phase (see Chapter 4). No differences were reported with > 90% peaks shared with the pooled input (data not shown). Equimolar amounts of each library were pooled, and 10 samples/lane were multiplexed on Illumina HiSeq 2000, 2 X 75bp paired-end reads. On average, 400M reads per lane were generated, mapped to human genome GRCh38.p15 reference assembly and stored as cram files. Cram were converted in bam keeping only reads with mapping score above 10 (q > 10) using Samtools view (Li et al., 2009). From two independent replicates, consistency between replicates was measured using the Irreproducibility Discovery Rate (IDR) pipeline (https://sites.google.com/site/anshulkundaje/projects/idr). This approach is based on an independent peak-calling for the two replicates with a relaxed cut-off (-p1e-1, --to-large) using MACS (Zhang et al., 2008). This allows to sample enough noise in the experiment for efficient statistical comparison of replicates in subsequent steps, meaning that many false positives are present in this peak set and they are not considered as final peaks for the separate replicates. Peaks are then ranked for p-value, and the same number is compared based on reproducibility (top 100,000 peaks). This is used to score rankings between peaks in each replicate to determine an optimal cut-off for significance. The final peak calls (bed and bigBed format) are the set of peak calls that pass IDR at a threshold of 5-10% max. This is the conservative peak set and it is used for further analyses.
For visualisation purposes, biological replicates were combined. Normalised bedGraph format files were generated for each sample using BEDTools v2.24.0 (Quinlan and Hall, 2010). The reads mapped at both DNA strands from 5' to 3' direction were extended to a length of 300bp, and the read-enrichment was normalized by million mapped reads and size of the library. bedGraph files were converted to bigWig using UCSC tool bedGraphToBigWig, and visualised on the Biodalliance genome viewer (Down et al., 2011) or the Integrative Genomics Viewer (IGV) (Robinson et al., 2011).

Differential binding was assessed using the G-test implemented in diffReps (Shen et al., 2013) in collaboration with Dr Pedro Madrigal. This tool scans the whole genome using a sliding window, performing millions of statistical tests and report the significant hits. diffReps takes into account the biological variations within a group of samples and uses that information to enhance the statistical power. For this, we used a sliding window of 300bp, \( p < 1E^{-5} \). Regions with differential binding that did not overlap with previously detected peaks through the IDR pipeline were removed. Genomic regions were ranked by their fold change (FC), and reported as differential only if both the \( |FC| > 1.5 \) and corrected p-value \( padj < 1E^{-6} \).

For the annotation of genomic intervals, Heatmaps and Metaplot generation, the R packages Genomation and ChIPpeakAnno were used (Akalin et al., 2015; Zhu et al., 2010). GREAT (Genomic Regions Enrichment of Annotations Tool) was used to associate genes with regulatory regions (McLean et al., 2010), combined with UCSC liftOver tool to convert bed coordinates from Homo Sapiens GRCh38/hg38 assembly to GRCh37/hg37 assembly (https://genome.ucsc.edu/cgi-bin/hgLiftOver).

2.6.2 ATAC-seq
TRA-1-60\(^+\) hESCs were sorted in the different cell cycle phases (EG1, LG1, G1/S, S/G2/M). Samples generated from two independent experiments, 100,000 were used per sample, as described in 2.4. Library preparation and sequencing were performed at the Wellcome Trust Sanger Institute next-generation sequencing facility. 8 ATAC-seq libraries were prepared with one of i5 and i7 Nextera tags combination (see 2.4), and pooled equimolarly. Sequencing was performed on Illumina HiSeq 2000, 2 X 75bp paired-end reads obtaining more than 60M mapped reads per library. Reads were mapped to human genome GRCh38.p15 reference assembly and stored as cram files. Duplicates were removed with Samtools rmdup (Li et al., 2009). Cram were converted in bam keeping only reads with mapping
score above 5 (q > 5) and from chromosome 1-22 and X (H9 is a female line) to remove mitochondrial contamination using Samtools view (Li et al., 2009). From two independent replicates, consistency between replicates was measured using the Irreproducibility Discovery Rate (IDR) pipeline, as described above in 2.6.1. Conservative peak set was used for further analyses. Diagnostic plots of fragment size distribution were generated using the R package ATACseqQC (Ou J, 2017). For visualisation purposes, biological replicates were combined as described above (2.6.1) to generate normalised bigWig.

Differential accessibility was assessed using the G-test implemented in diffReps (Shen et al., 2013) in collaboration with Dr Pedro Madrigal, as described above (2.6.1), with minor changes. Specifically, a sliding window of 600bp was used, p < 1E-6.

Footprinting analyses were done in collaboration with Dr Pedro Madrigal combining two methods: Footprint Log-likelihood Ratio (FLR) (Yardimci et al., 2014) and Wellington (Piper et al., 2013), adjusted for ATAC-seq. To get high quality reproducible footprints, motif matches with FLR ≥10.0 in each biological replicate were classified as “bound”. TF footprints that were present in both replicates, for a motif with at least 50 footprints, were considered as final predicted TF binding sites. FLR mean of the replicates was calculated, only based on unique genomic regions (for those palindromic sequences with a footprint reported in each strand, we obtained the mean FLR). The quality of footprint was assessed calculating the Protection Score, removing scores below 0 (Gusmao et al., 2016). Wellington instead uses a “non-motif centric” approach, which considers the imbalance in this strand-specific information to efficiently identify DNA footprints. For this, a relaxed cut-off has been used (-p 32 -A -fdr 0.1 --FDR_limit -4 --pv_cutoffs -4 -fp 4,30,1 -fdriter 500 —one_dimension) on merged replicates (to increase sequencing depth), in order to achieve a significant overlap with FLR, and a final list generated. A conservative list of footprints was generated by overlap between FLR and Wellington, removing not expressed genes according to RNA-seq data (Fragments Per Kilobase of transcript per Million mapped reads, FPKM > 1.0).

In order to generate a set of differential footprints, an extension of Wellington has been used, Wellington-bootstrap (Piper et al., 2015), adapted to ATAC-seq. Footprints with score > 20 were selected in order to select the most confident differential footprint (-p 16 -A -fdr 0.05 --FDR_limit -10 -fp 4,30,1 -fdriter 100). In
order to obtain a final list, selected footprints are filtered using the conservative list generated using FLR and Wellington combined.

For visualisation of footprints, bigWigs of ATAC-seq data containing Tn5 insertion frequencies have been generated using NucleoATAC (Schep et al., 2015) from the bam files (--bam $bamFile --out $outputBasename --cores 16).

2.6.3 RNA-seq

TRA-1-60⁺ hESCs were sorted in the different cell cycle phases (EG1, LG1, S/G2/M). From three independent experiments, RNA was extracted using GenEluteTM Mammalian Total RNA Miniprep Kit and On-Column DNase I Digestion Set (both Sigma-Aldrich) and a total of 9 samples fulfilling the submission requirements (0.5-1.0μg total RNA in nuclease free dH₂O with A260/A280 and A60/A230 ratios above 1.8) were sent for library preparation and next-generation sequencing at CGS (Cambridge Genomic Services). PolyA purified library preparation kit was used, and samples were multiplexed in one HIGH Output NextSeq run Cycle, 2 X 75bp, paired-end reads. A total of 462,569,268 reads evenly distributed across the lanes were obtained (50,547,439 on average per sample). Reads were trimmed using Sickle (Joshi, 2011), an application for trimming FASTQ files based on quality values (Sanger), with a q=20 (quality threshold) and l=30 (length threshold). For the alignment, the transcriptome was built with TopHat2 (Kim et al., 2013) (v2.1.0), based on Bowtie2 (Langmead and Salzberg, 2012) (v2.2.6) using as reference genome the human GRCh38.p6 and GTF annotation from Ensembl (http://ftp.ensembl.org/pub/release-83/gtf/homo_sapiens/). All analyses were performed using this genome assembly. Alignment was performed using TopHat2 with standard parameters. Using Samtools view (Li et al., 2009), reads with mapping quality > 10 were kept for further analyses. Data were quantitated and differential expression analysis was performed using SeqMonk (Andrews, 2007). In sum, data were quantitated using the RNA-seq pipeline, and differential expression analysis was performed using the Intensity Difference Test and the R package DESeq2 (for this, uncorrected raw read counts were used, see below), generating a list of transcripts passing the filter, with a p-value cutoff < 0.05. Hierarchial clustering plots (r > 0.7). Gene Ontology analyses were performed using the Enrichment analysis tool by the Gene Ontology Consortium (Panther) (Thomas et al., 2003), and validated using Enrich (Chen et al., 2013).
The Intensity Difference Test and DESeq2

The intensity difference test performs a statistical test using the general distribution of the data to test whether any individual point is likely to be an outlier from the overall distribution. It is a pairwise test, but can be extended to multiple samples by doing a series of pairwise tests and combining the results.

DESeq2 performs an internal normalisation where geometric mean is calculated for each gene across all samples – that is why raw reads are needed. Then, the counts for each gene in each sample is divided by this mean. The median of these ratios in a sample is the size factor for that sample. This procedure corrects for library size and RNA composition bias, which can arise for example when only a small number of genes are very highly expressed in one experiment condition but not in the other.
3 EXPERIMENTAL APPROACH

The experimental setting for most of the work in this dissertation is based on a very straightforward but powerful approach. Using a FUCCI-hESCs line previously validated (Pauklin and Vallier, 2013) and described in 1.1.5, we were able to live-follow cell cycle progression and sort cells according to their cell cycle phase, and collect them for downstream analyses.

3.1 Cell sorting
Prior to sorting, cells were stained for TRA-1-60, a pluripotency marker, in order to remove background differentiated cells that occasionally arise in culture. Cells were then size selected for Forward (FS) and Side Scatter (SS), and doublets were removed. TRA-1-60+ cells were then sorted in the four different cell cycle phases according to the FUCCI reporter: early G1 (EG1) for double negative (mKO2- / mAG+), late G1 (LG1) for red-mKO2+, G1/S transition for double positive (yellow-mKO2+ / mAG+), and S/G2/M for green-mAG+. Sorted cells were then used for ATAC-seq, ChIP-seq, and RNA-seq (Figure 3.1). Whereas the green-mAG+ population is clearly defined, the transition from early to late G1 is not distinct, since the red-mKO2 is gradually accumulated during G1 progression. Therefore, the gating for the EG1 was set according to the S/G2/M population, which is mKO2-, and the remaining cells, opportunely spaced, are considered mKO2+. TRA-1-60+ only have been used for unsorted control when needed, in order to have a pool of cells that maintain the original heterogeneity but have gone through the same sorting process as the cell phase specific samples. As described before, in hESCs most cells are usually in the S phase, therefore the partition obtained by
sorting was typically the following: about 10% in EG1, 20% LG1, <5% in G1/S transition, and about 65% in S/G2/M. For this reason, all the experiments have been performed normalising for cell number, usually to the lowest sample, depending on the downstream application. ChIP-seq, especially for transcription factors, does require a significant number of cells. For this reason, the transition phase G1/S has been excluded from this analysis, whereas it was included in the ATAC-seq, considering the relatively low number of cells required for this assay (Buenrostro et al., 2013). On average, about 1 million cells per cell cycle phase have been used for ChIP-seq and RNA-seq, whereas 100,000 cells per phase have been used for ATAC-seq.

Figure 3.1 Schematic for FUCCI FACS sorting.

(A) Representative colony of FUCCI-hESCs showing cells in EG1 (no fluorescence), LG1 (red), G1/S transition (yellow), and S/G2/M (green). Scale bar represents 100μm. (B) Schematic for downstream analyses after sorting (ATAC-seq, ChIP-seq, and RNA-seq). (C) Gating strategy: 1, cells are size selected for forward scatter area (FSC-A) and side scatter area (SSC-A); 2, doublets are removed (FSC-A, FSC-W, width); 3 TRA-1-60+ cells are selected; and 4, sorted in the four different cell cycle phases according to the FUCCI reporter: EG1 for double negative (mKO2-/mAG-), LG1 for red-mKO2+, G1/S transition for double positive (yellow-mKO2+/mAG+), and S/G2/M for green-mAG+. For all events, average and standard deviation are reported with n=10.
3.2 Chromatin Immunoprecipitation optimisation

As briefly mentioned, ChIP experiments require a significant number of cells in order to efficiently immunoprecipitate chromatin. However, considering our experimental settings following cell sorting, we firstly needed to optimise our ChIP protocol, and be able to scale it down to a number of cells that could be consistently obtained after sorting. ChIP relies on antibodies able to specifically immunoprecipitate the proteins of interest, but also on their robust cross-linking to chromatin. Therefore, our approach was focused on three main points: antibody screenings, adaptation of the ChIP protocol to low number of cells, and sonication tests. The process is summarised and the results reported for OCT4 ChIP (Figure 3.2).

![ChIP optimisation summary](image)

**Figure 3.2 ChIP optimisation summary.**

(A) OCT4 antibodies screening with standard ChIP protocol (10M cells), comparing different commercially available ChIP-grade antibodies: N-19 from Santa Cruz (sc1), C-10 from Santa Cruz (sc2), and ab19857 from abcam (ab1). (B) Cell number tests using Diagenode ChIP kit for Low Cell Number, in combination with the iPure kit from Diagenode, comparing efficiency using 500k, 250k, 100k cells. ChIP analyses reported as % of input, comparing IP vs IgG control, n=3. (C) Sonication tests comparing from 4 to 19 cycles, increasing by 3. Arrow indicates ideal fragments between 100-500bp.
The first optimisation step compared different commercially available ChIP-grade antibodies: in this case, N-19 and C-10 from Santa Cruz, and ab19857 from abcam (see Material and Methods for further details). The Anti-OCT4 antibody from abcam was proven to be the more efficient, showing a better enrichment for two well-known OCT4 binding sites (OCT4, NANOG). The next step was to develop a ChIP protocol compatible with low cell numbers, testing different commercially available kits and modifying the related protocols in order to make them compatible our experimental settings. After several tests, the most efficient and reproducible protocol was the one developed using the Diagenode ChIP kit for Low Cell Number, in combination with the iPure kit from Diagenode to recover the precipitated chromatin (see Materials and Methods for more details). This combination allowed us to use down to 250k cells per IP in a very robust and reproducible way (Figure 3.2), with a 100k cells lower limit. These kits use a combination of magnetic beads and buffer able to maximise the chromatin precipitation and the DNA recovery, and most importantly with no unspecific binding and very reproducible results. In parallel, sonication tests were performed for the different cell numbers, to avoid over sheared samples, incompatible with downstream sequencing. Fragments between 100-500 bp are considered ideal for ChIP-seq, as a good compromise between read mappability and enrichment resolution/specificity. This was achieved sonicating 1M cells for 4-5 cycles (Figure 3.2). These optimisation steps, here summarised, generated a robust and reproducible protocol for all the downstream ChIP-seq application following cell sorting described in the thesis.
Chapter 4: Cell cycle regulation of chromatin structure

4 CELL CYCLE REGULATION OF CHROMATIN STRUCTURE

During cell cycle progression, chromatin alternate between an invariant mitotic conformation and a cell-type specific interphase organisation (Naumova et al., 2013). However, the chromatin dynamics during interphase have not been investigated until very recently, both at conformational and accessibility levels. It has been shown that the genome is not stably folded at any point in interphase, with loop enrichment and TADs insulation emerging very early in the G1 phase, but A/B compartments reappearing only during the S phase (Nagano et al., 2017). To investigate this into greater details, ATAC-see, a technique that combines ATAC-seq and direct imaging, has revealed chromatin openness dynamics. This showed a global chromatin decondensation that peaks in LG1, to then refold during S/G2/M. Intriguingly, the majority of open loci in EG1 were enhancers, but they were mostly promoters in LG1 (Chen et al., 2016). These results raised the possibility that some enhancers become accessible first in EG1, and many promoters become only accessible in LG1.

However, none of these studies uses a precise system to isolate the different cell cycle phases in order to investigate chromatin openness, nor has looked into mechanistic regulation for such changes. Moreover, most of them were not performed in hESCs, where differentiation propensity has been shown to be controlled by cell cycle (Pauklin and Vallier, 2013). Therefore, investigating chromatin dynamic in these cells could reveal new regulation for cell fate decision. In order to address these questions, we performed ATAC-seq on sorted FUCCI-
hESCs, and identified a highly dynamic chromatin pattern during cell cycle progression. In addition to this, footprinting analyses revealed a strong association with the transcription factor CTCF, here proposed to regulate such changes.

4.1 Changes in chromatin structure during cell cycle progression

In order to investigate chromatin dynamics during the cell cycle, we sorted FUCCI-hESCs in the different cell cycle phases (EG1, LG1, G1/S, S/G2/M) and performed ATAC-seq to study chromatin accessibility (Buenrostro et al., 2013). ATAC-seq is a technique used for mapping chromatin accessibility genome-wide based on a hyperactive Tn5 transposase, which inserts sequencing adapters into accessible chromatin. Fragments can be sequenced and used to map regions of high accessibility and nucleosome position (Figure 4.1).

![Figure 4.1 ATAC-seq schematic.](image)

ATAC-seq (Assay for transposase-accessible chromatin sequencing) is an in vitro technique based on the hyperactive Tn5 transposase that inserts sequencing adaptors into open chromatin followed by deep sequencing. This provides base-pair resolution of nucleosome-free regions of the genome.

In order to assess the quality of the ATAC libraries, we looked at fragment size density (Figure 4.2). The insert size distribution of sequenced fragments showed a clear periodicity of approximately 200 bp, with most reads in nucleosome-free regions, suggesting a general state of open chromatin in all the samples. The remaining fragments are protected by increasing multiples of nucleosomes, showing a sharp periodicity equal to the helical pitch of DNA, confirming an efficient DNA cutting. Using the IDR pipeline on two replicates (see Material and Methods, 2.6.2) we found 88,479 reproducible peaks in EG1, 117,283 in LG1, 99,109 in G1/S, 80,398 in S/G2/M, showing higher degrees of general accessibility in the LG1 phase (Figure 4.3).
Figure 4.2 ATAC-seq fragment size density plot.

ATAC-seq fragment size density for the 4 sequenced libraries, EG1, LG1, G1/S, and S/G2/M sorted cells. Inset histograms (log-transformed) show periodicity to four nucleosomes.

Figure 4.3 ATAC-seq IDR plot.

(A) IDR (Irreproducible Discovery Rate) plots reporting reproducible ATAC-seq peaks between two replicates in EG1, LG1, G1/S, S/G2/M. IDR threshold < 5% (B) Total number of reproducible ATAC-seq peaks in the different cell cycle phases EG1, LG1, G1/S, S/G2/M.
Interestingly, a global open structure was confirmed by comparing all undifferentiated conditions with definitive Endoderm (dEN) (Madrigal et al., unpublished) (Figure 4.4). From a total of 153,007 unique open regions, overlapping showed a highly dynamic chromatin state during cell cycle progression: only ~33% of the peaks are present in all the phases, whereas ~67% of the peaks are present in either a single phase (~34%) or more than one, but not all (~33%) (Figure 4.5).

**Figure 4.4 ATAC-seq Metaplot.**

(A) Metaplot of ATAC-seq signal +/- 1kb around nucleosome free regions in EG1, LG1, G1/S, S/G2/M, compared to definitive Endoderm (dEN, data from Madrigal et al., unpublished). (B) Browser tracks for ATAC-seq in hESCs compared to dEN over the POU5F1 locus.

**Figure 4.5 ATAC-seq Venn overlap.**

Venn diagram showing peaks overlap for the ATAC-seq in EG1, LG1, G1/S, S/G2/M, and table summary reporting number and percentage of peaks found in a single phase, more than one phase (2-3), or all phases.
However, features distribution did not reveal specific changes, with most of the peaks falling into intergenic or intronic regions in all the phases (Figure 4.6). More specifically, after associating the resulting genomic regions with putative genes using GREAT (McLean et al., 2010), the majority was found at +/- 500 to 50 kb from the transcription starting site (TSS), most likely in distal regulatory regions such as enhancers (Figure 4.6).

**Figure 4.6 ATAC-seq features distribution plots.**

(A) Feature distribution for the ATAC-seq peaks in EG1, LG1, G1/S, S/G2/M. (B) Distribution of region-gene associations around transcription starting site (TSS), binned by distance in kb (0-5, 5-50, 50-500, >/<= 500).

Enrichment analyses for biological processes showed involvement in positive and negative regulation of transcription from RNAPII promoter in all phases, consistent with the idea of chromatin openness as a way to control initiation of transcription (Orkin and Hochedlinger, 2011) (Figure 4.7). Interestingly, all phases were also enriched for positive regulation of transcription by TFs localisation, suggesting that, rather than directly correlating with regulation of gene expression, different chromatin openness during the cell cycle might represent a favourable environment for TFs binding to allow differentiation or maintenance of pluripotency. Indeed, this GO term includes any process that activates or increases the frequency, rate or extent of DNA-dependent transcription using a mechanism that involves the localization of a transcription factor (GO:0061586). Consistent with cell cycle progression, EG1 showed enrichment for mitosis exit genes, and, very interestingly, LG1 was enriched for neuron differentiation genes. This is in agreement with previous findings in our lab, where the LG1 has been described as the preferential phase for neuroectoderm specification (Pauklin and Vallier, 2013).
Other general enriched terms were metabolism, developmental biology, signal transduction, gene expression and cell cycle related pathways in human. Protein-protein interaction (PPI) hub, a literature based protein-protein interaction network, where each transcription factor protein is a term in the library and the proteins that interact with it are the set elements, was enriched for cell cycle genes (CDK1, CDK2, Ubiquitin) and SMAD2, that interestingly ranked in the top 10 in the EG1 phase, and lower in LG1 and G1/S (Figure 4.7). This also agrees with the previously described cell cycle regulation of SMAD2/3, present in the nucleus mostly in the EG1 phase (Pauklin and Vallier, 2013).

Figure 4.7 ATAC-seq Gene Ontology and Pathway analyses.

(A) Gene ontology for Biological Processes in the different cell cycle phases EG1, LG1, G1/S, S/G2/M for ATAC-seq gene associated regions. Combined score is calculated by multiplying the unadjusted p-values with the z-scores. (B) Additional enrichment analyses for Pathways (Reactome) and Protein-Protein interaction (PPI Hub), a literature based protein-protein interaction network, reported with ranking based on combined score.
Taken together, ATAC-seq in sorted H9-FUCCI revealed a highly dynamic chromatin structure that opens after cell division in EG1, peaks in LG1, and re-condense in preparation of mitosis in S/G2/M.

4.2 Differential chromatin openness during cell cycle progression

The simple overlap of peaks in the different cell cycle phases does not give strong insights on the dynamics of these changes. Therefore, in order to describe in greater details these dynamics, we used diffReps, a tool developed to detect differential sites using a sliding window to scan the genome and identifies regions that show read count differences (Shen et al., 2013). This tool is independent of any peak calling program and allows multiple comparisons. We performed pairwise comparison (EG1 versus LG1, LG1 versus G1/S, G1/S versus S/G2/M, and S/G2/M versus EG1) to detect differentially open sites. They were then filtered for the conservative list of reproducible peaks described in 4.1, in order to remove background changes. Using this tool, about 1,000 statistically significant (|FC| >1.5 and corrected p-value p-adj < 1E-6) highly dynamic regions were identified, resulting mostly from changes throughout the EG1 phase: 715 regions from EG1 to LG1, 195 from S/G2/M to EG1 (Figure 4.8).

![Figure 4.8](image.png)

**Figure 4.8 ATAC-seq differentially open regions.**

Number of dynamic regions during cell cycle progression obtained by pair-wise comparisons using diffReps (FC>1.5/FC<-1.5 and corrected p-value p-adj<1E-6). In pink, regions closing between phase A and B, in green number of regions opening between A and B.

Interestingly, most regions were opening during the S/G2/M to EG1 transition (179 peaks) and closing during the EG1 to LG1 transition (433 peaks), suggesting that the most dynamic phase during the cell cycle is indeed the EG1, when chromatin rapidly opens and reorganises after division in order to maintain cell identity or respond to differentiation stimuli. Interestingly, when plotting all the dynamic
regions, EG1 was indeed the most open phase, followed by a gradual chromatin compaction that culminates in the S/G2/M (Figure 4.9).

![Figure 4.9 Differentially open regions Metaplot and Heatmap.](image)

**(A)** Metaplot of ATAC-seq signal for differentially open regions (diffReps) +/- 1kb around nucleosome free regions in EG1, LG1, G1/S, S/G2/M, compared to definitive Endoderm (dEn, data from Madrigal et al., unpublished). **(B)** Heat-maps display normalised ATAC-seq signal centred on the merged peaks, +/- 1kb, for all the replicates.

Very little changes were detected between the G1 phase and the S/G2/M, suggesting that after the highly active G1 phase, most activities in the cells are uniquely devoted to DNA replication and preparation to cell division. Features distribution confirmed most peaks in intronic or intergenic regions, with an interesting exception (Figure 4.10). Indeed, region closing during the EG1 to LG1 transition are the only changing promoters and exons, suggesting that whereas distal regulatory regions are the most open and dynamic, there is a subset of regions that might directly correlate with gene expression within the G1 phase. Characterising these peaks in more details associating them to genomic regions, confirmed that changes mostly happened in distal regulatory regions (+/- 500 to 50 kb from TSS), with the exception of regions closing during EG1 to LG1 transition, as already shown (Figure 4.10).
Enrichment analyses for biological processes showed, as previously described, involvement in positive and negative regulation of transcription from RNAPII promoter and cell cycle in most phases, and some interesting cell phase specific features (Figure 4.11). Most regions opening in the EG1 phase correlated with regulation of transcription and chromatin organisation regulating it (EG1vsLG1 down and S/G2/MvsEG1 up). Cell cycle terms are instead highly enriched during the S/G2/M (S/G2/MvsEG1 down), when the DNA synthesis and cell division machinery is more active. Interestingly, several MAPK/ERK and phosphatidylinositol 3-kinase (PI3K) signalling activation terms were enriched in the regions opening at the EG1 to LG1 transition. Both ERK1/2 and PI3K/AKT cascades are downstream targets of the FGF pathway in hESCs and have been shown to cooperate for maintaining the pluripotent state (Li et al., 2007). ERK1/2 has been described as master regulator of G1/S progression (Meloche and Pouyssegur, 2007). In addition, PI3K antagonises the ability of hESCs to differentiate in response to Activin/Nodal, and its inhibition efficiently promotes differentiation of hESCs into mesendoderm and then definitive endoderm (McLean et al., 2007). The activation or enhancing of these pathways only in the LG1 phase supports the previous findings showing that hESCs can only differentiate into mesendoderm in the EG1 phase (Pauklin and Vallier, 2013), when PI3K activity could be lower. To support this, several mesendoderm genes were closing during the EG1 to LG1 transition (FOXH1, HNF1A, PBX3, PAX7) and in contrast the LG1
phase is enriched for genes regulating neural development (EPHA3, Neuregulins, GABRB1, MYO16) (Figure 4.11). Other cell cycle genes are included, supporting normal cell cycle progression in the different phases: regions closing closes between EG1 and LG1 for example include CDK11A, highly expressed CDK during G2/M phase; or CDKN1C, inhibitor of several G1 cyclins.

![Diagram]

**Figure 4.11 Differentially open regions Gene Ontology.**

(A) Gene Ontology for Biological Processes for ATAC-seq differentially open gene-associated regions (diffReps). Combined score is calculated by multiplying the unadjusted p-values with the z-scores. (B) Browser tracks for ATAC-seq in the different cell cycle phases (EG1, LG1, G1/S, S/G2/M) and compared to dEN over the FOXH1 and GABRB1 loci.

Taken together, differential opening of chromatin during cell cycle progression revealed phase specific patterns correlating with cell fate propensity. Importantly, distinct chromatin structure is shown to regulate transcription, but rather than directly, they might represent alternative chromatin states for differential TFs binding to allow differentiation or maintenance of pluripotency.

### 4.3 Footprinting analysis

In order to investigate whether TFs occupancy changes during the cell cycle, we decided to perform footprinting analyses in the different cell cycle phases. This is based on the idea that the DNA corresponding to a binding motif is selectively resistant to digestion by nonspecific DNA nucleases or hyperactive transposases, therefore leaving a “footprint” when a TF is binding a specific site in the genome. Combining FLR (Yardimci et al., 2014) and Wellington (Piper et al., 2013)
pipelines, a conservative list of footprints has been generated. The list has been then filtered for expressed genes (RNA-seq, see Chapter 5), in order to remove transcription factors predicted but not expressed in hESCs. This resulted in a total of 414,437 footprints in the different phases (103,168 in EG1, 90,302 in LG1, 121,006 in G1/S, 99,961 in S/G2/M). Of these, 56,806 were covering unique regions, 7% (4,009 footprints) were shared by all, whereas ~56% were cell phase specific (31,636 total footprints, 9,134 footprints only in EG1, 5,338 only in LG1, 10,399 only in G1/S, and 6,765 only in S/G2/M), suggesting a dynamic occupancy of TFs (Figure 4.12). Of all the footprints, we decided to look which of them were overlapping with dynamic genomic regions, in order to identify possible factors regulating chromatin changes. Among the total of 1,807 overlapping footprints, 74% (1,330 footprints) were found in the EG1, 8% (147) in LG1, 11% (203) in G1/S, and 7% (127) in S/G2/M (Figure 4.12).

![Figure 4.12 ATAC-seq footprints analysis.](image)

(A) Total number of footprints generated from ATAC-seq in the different cell cycle phases EG1, LG1, G1/S, S/G2/M. (B) Total number of unique footprints in the different cell cycle phases. (C) Number of predicted footprints overlapping with dynamic regions (ATAC-diffReps) in the different cell cycle phases. Final footprints were obtained combining FLR (Yardimici et al., 2014) and Wellington (Piper et al., 2013) pipelines (combining list with Footprint Likelihood Ratio (FLR) ≥10.0, Protection Score > 0).

In all the phases, apart from the G1/S transition, the top-ranking footprint corresponded to the TFs CTCF (Figure 4.13). More specifically, CTCF footprint represented ~50% of all the footprints in the EG1 phase, dropping to ~10% in the remaining phases. On the contrary, in the other phases there was a wider distribution of factors’ footprints, and noticeably a gradual increase of FOX-protein footprints (FOXA2, FOXC1, FOXO3, FOXO4), previously described as pioneering factors able to bind condensed chromatin (Zaret and Carroll, 2011). When comparing EG1 phase with the others, CTCF footprints was remarkably more
frequent in all the comparisons, whereas the other phases showed a wider distribution with more factors accumulating (Figure 4.14).

Figure 4.13 Transcription factors footprints.

(A) Percentage of footprints related to single transcription factors generated from ATAC-seq in the different cell cycle phases EG1, LG1, G1/S, S/G2/M. (B) Pairwise comparisons of all cell cycle phases, reporting percentages of footprints for specific transcription factors, log scale axis.

Figure 4.14 CTCF footprint in differentially open regions.

Aggregate ATAC-seq footprint for CTCF in the different cell cycle phases plus definitive endoderm (dEn, Madrigal et al., unpublished) generated over differentially open regions within the genome.
This suggests an important role for CTCF in regulating chromatin structure at the start of the cell cycle, whereas a plethora of different TFs follow to regulate the specific transcriptional networks. In addition, these analyses predict a novel specific binding of CTCF in the EG1 phase.

### 4.4 CTCF ChIP-seq

CTCF was originally characterized as an insulator protein, restricting enhancer–promoter interactions both *in vitro* and *in vivo* (Kurukuti et al., 2006; Xie et al., 2007). It is ubiquitously expressed and it is essential for embryonic development (Soshnikova et al., 2010; Wan et al., 2008). It is enriched at TAD boundaries (Dixon et al., 2012), but the majority of its binding sites are distributed within TADs and are thought to be also involved in intra-TAD interactions (Handoko et al., 2011). However, the cell cycle dynamics of such protein have not been investigated yet, despite growing evidences linking its activity with genome folding. According to our footprinting analyses, we expected to see a stronger chromatin interaction of CTCF during the EG1 phase, which seems to be the most dynamic phase of the cell cycle in terms of chromatin organisation. In order to validate this hypothesis, we performed ChIP-seq in sorted cells for CTCF after ChIP optimisation (see Chapter 3.2). From two replicates, using the IDR pipeline (see Material and Methods, 2.6.2) we found a total of 6,177 reproducible peaks, of which 4,878 peaks were found in EG1, 506 peaks in LG1, and 793 peaks in S/G2/M (Figure 4.15). Genome wide mapping showed again stronger binding in the EG1 phase (Figure 4.15), confirming footprinting analyses prediction.

![Figure 4.15 CTCF ChIP-seq.](image)

(A) Total number of peaks in the different cell cycle phases for CTCF ChIP-seq. (B) Metaplot of ChIP-seq signal for CTCF +/- 1kb around CTCF binding sites in EG1, LG1, S/G2/M, compared to input. (C) Heat-maps display normalised CTCF ChIP-seq signal centred on the merged peaks for all the replicates +/- 1kb, in the different cell cycle phases.
Of these, 5,606 peaks were unique, with 91% (5,125) present in a single phase, and only 9% in two or more phases. Of note, 4,401 of the unique peaks were found in EG1 (Figure 4.16). Feature distribution did not reveal a cell cycle specific location, with most peaks present in intergenic regions in all the phases (Figure 4.17), and after associating peaks with genomic regions using GREAT, we confirmed that most features were found in distal regulatory regions (+/- 500 to 50 kb and +/- 50 to 5 kb).

![CTCF ChIP-seq Venn overlap](image)

**Figure 4.16 CTCF ChIP-seq Venn overlap.**

Venn diagram showing peaks overlap for the CTCF ChIP-seq in EG1, LG1, S/G2/M, and table summary reporting number and percentage of peaks found in a single phase, more than one phase (2-3), or all phases.

![CTCF ChIP-seq features distribution](image)

**Figure 4.17 CTCF ChIP-seq features distribution.**

(A) Feature distribution for the CTCF ChIP-seq peaks in EG1, LG1, S/G2/M. (B) Distribution of region-genes associated regions around transcription starting site (TSS), binned by distance in kb (0-5, 5-50, 50-500, >/<= 500).
Enrichment analyses for biological processes mostly overlapped with ontologies from all ATAC-seq peaks, confirming that CTCF binding might be indeed regulating such processes. In more details, in all phases CTCF binding seemed to control positively and negatively transcription, and interestingly, MAP/ERK terms in LG1, as previously found (Figure 4.18). Additional ontologies showed interesting features: in EG1 for example, Transcription Factor Protein-Protein Interactions (PPIs) database showed enrichment for many components of the SMAD family (SMAD2, SMAD3). This means that the genes associated with CTCF peaks mostly encode for proteins described as SMADs interactors, in agreement with the previously described activation of the SMAD2/3 pathway in EG1.

Figure 4.18 CTCF ChIP-seq Gene Ontology.

(A) Gene Ontology for Biological Processes in the different cell cycle phases EG1, LG1, S/G2/M for CTCF ChIP-seq gene associated regions. (B) Additional enrichment analyses in EG1 for transcription factors Protein-Protein Interaction (TFs_PPIs), a literature based protein-protein interaction network, reported with ranking. Combined score is calculated by multiplying the unadjusted p-values with the z-scores.

Ultimately, to validate footprinting prediction we overlapped CTCF footprints in all the phases with CTCF ChIP-seq peaks. Of the 708 CTCF footprints, ~60% (430 peaks) overlapped with CTCF peaks called with ChIP-seq, confirming the validity of the prediction (Figure 4.19). However, the vast majority of the CTCF peaks have not been predicted by the footprinting analyses, underlying some of the limitation of these technique in predicting de novo binding sites.
The overlapping of features and ontologies between ATAC-seq peaks and CTCF binding sites, together with the footprinting predictions, were strongly supporting our hypothesis of chromatin conformation control by CTCF. In order to validate this hypothesis, we overlapped CTCF peaks with ATAC peaks and found that ~77% (4,762 peaks) of 6,177 CTCF peaks overlapped with ATAC peaks, confirming strong correlation between chromatin openness and CTCF occupancy (Figure 4.20). However, most ATAC peaks do not correlate with CTCF occupancy, suggesting that other factors might be contributing to chromatin openness during the cell cycle together with CTCF. An interesting pattern arises when comparing cell cycle specific peaks: ~82% of CTCF peaks overlap with ATAC in EG1, but this overlap dropped to 60% in LG1 and 51% in S/G2/M (Figure 4.20). This could mean that CTCF binding highly correlates with chromatin openness in EG1 to re-arrange chromatin after cell division, whereas in the other phases it might have additional functions.

Figure 4.19 CTCF ChIP-seq / CTCF footprint overlap.
Venn diagram reporting the overlap of CTCF footprints predicted by ATAC-seq and CTCF ChIP-seq peaks. 60% represent the percentage of footprints overlapping with ChIP-seq peaks.

Figure 4.20 CTCF ChIP-seq / ATAC-seq overlap.
(A) Venn diagram reporting the overlap of CTCF ChIP-seq peaks and ATAC-seq peaks. 77% represent the percentage of CTCF ChIP-seq peaks with ATAC-seq peaks. (B) Cell phase specific overlap of CTCF ChIP-seq peaks and ATAC-seq peaks in the EG1, LG1 and S/G2/M phase.
Taken together, footprinting predictions, ontologies and ATAC peaks overlaps supported the hypothesis that CTCF is among the most important factors controlling chromatin conformation after cell division. Its stronger binding during the EG1 phase, just after mitotic exit, support the idea that CTCF might be the drive for restructuring chromatin to its structure prior to chromosomes condensation, in order to preserve genome integrity.

4.5 Discussion

The results reported in this chapter represent a first description of chromatin dynamics during cell cycle progression in hESCs, and suggest a mechanism regulating these changes. Firstly, we showed that, despite an overall open structure compared to its differentiated counterpart (Efroni et al., 2008), chromatin is highly dynamic during cell cycle progression in hESCs, showing higher degrees of accessibility in the LG1 phase. This is in agreement with recent reports that described similar dynamics, with a global chromatin decondensation that peaks in LG1, to then refold during S/G2/M (Chen et al., 2016). However, most rearrangements were found to happen from the S/G2/M phase to EG1, and from EG1 to LG1, indicating the EG1 phase as the most dynamic in terms of restructuring. This sequence of modifications is not surprising, considering that during mitosis chromosomes stochastically fold, and chromatin structure has to be established de novo at the beginning of each cell cycle (Efroni et al., 2008). However, it is important to note that our sorting strategy does not discern S phase to G2/M, and that the majority of cells are likely to be in the S phase, given the high proliferation rate of hESCs. This means that we are certainly underestimating the compaction state of mitotic cells, and only have a snapshot of early compacting regions, and stronger chromatin interactions. Alternative strategies, such as nocodazole synchronisation (Deysson, 1968; Matsui et al., 2012; Zieve et al., 1980), or the FUCCI4 (Bajar et al., 2016) could help to analyse the last phase of the cell cycle and investigate in greater details the changes around the M phase. However, we were mostly interested in changes happening during interphase, and methods to study molecular mechanisms in M phase are not always reliable. For instance, it has been shown that TFs exclusion from mitotic chromosomes could be largely a fixation artefact (Teves et al., 2016). Our main questions were whether putative conformation changes were driving or establishing cell identity, and whether these changes were regulated by cell cycle events. Therefore, a system that allowed us to dissect in great details the early cell cycle, when cells are
susceptible to alternative differentiation cues (Pauklin and Vallier, 2013), would suffice. Therefore, after we established that chromatin is indeed very dynamic, we described that most changes happened in distal regulatory regions, in agreement with what previously described using ATAC-see (Chen et al., 2016), suggesting that a hierarchical reforming of higher structures is the first step upon mitotic exit. This is in agreement with a recent publication reporting that the genome is not stably folded at any point in interphase, with loop enrichment and TADs insulation emerging very early in the G1 phase, but A/B compartments reappearing only during the S phase (Nagano et al., 2017). Interestingly, this could correlate with the propensity for hESCs to differentiate in the G1 phase: absence of compartmentalisation might represent a favourable environment for fate change, whilst maximum insulation ensures quick regulation and promoter-enhancer looping at the required loci, probably mediated by specific TFs binding. In agreement with these, our data showed that differential opening of chromatin during cell cycle progression revealed phase specific patterns, partially correlating with cell fate propensity: several mesendoderm genes were closing during the EG1 to LG1 transition (FOXH1, HNF1A, PBX3, PAX7) and in contrast the LG1 phase is enriched for genes regulating neural development (EPHA3, Neuregulins, GABRB1, MYO16). These distinct chromatin structures might represent alternative chromatin states for differential TFs binding to allow differentiation or maintenance of pluripotency, such as SMAD2/3 driving mesendoderm in the EG1 phase (Pauklin and Vallier, 2013). In order to identify factors regulating this chromatin rearrangements, we performed footprinting analyses with our ATAC-seq dataset. These analyses predicted an important role for CTCF in regulating chromatin structure in the early cell cycle, whereas a plethora of different TFs follow to regulate the specific transcriptional network. Indeed, ChIP-seq for CTCF confirmed its cell cycle regulation, and its prominent binding in the EG1 phase. Its binding has been shown to correlate with chromatin openness, but this correlation decreases during cell cycle progression. This suggests a tighter link in the early cell cycle, when CTCF could be necessary for re-establishing chromatin structure, but facultative for its maintenance after DNA replication, when cell identity is established. Other factors might be involved and responsible for maintaining such established structure, confirmed by the fact that compartmentalisation is CTCF independent (Nora et al., 2017). Recent reports using live imaging confirmed that 80% of CTCF protein binds chromatin transiently, and that this binding decreases from G1 to S phase (Agarwal et al., 2017). Moreover, separate studies reported its
eviction from chromatin during mitosis, regulated by mitotic phosphorylation that affects its DNA binding capacities (Sekiya et al., 2017). Post-translational modifications, and especially phosphorylation, are a classical way to quickly regulate cell cycle events. Indeed, another report showed how the cell cycle regulated activity of the Casein Kinase 2 (CK2), known to be active in the G2/M phase to maintains CycB/CDK1 activity, regulates CTCF mitotic inactivation (El-Kady and Klenova, 2005). This led us to hypothesise that CTCF is required to rearrange chromatin structure after cell division, when chromatin compacts in a mostly stochastic way and in a cell-type independent manner, but its binding is dispensable in the S/G2/M phase, or restricted to more classical TF activities rather than having a structural role. A recent report showed that indeed CTCF is required, in a dose-dependent manner, to instruct chromatin loops and TADs insulation (Nora et al., 2017). In addition to this, they also show that higher chromatin structures such as compartments are not affected by CTCF depletion. Interestingly, compartments have been shown to reappear only in the S phase (Nagano et al., 2017), and the finding that in our system CTCF binding is much lower in the S phase supports this. Surprisingly, this paper also reports that, despite CTCF depletion dramatically slows proliferation, cells are not blocked in any specific phase of the cell cycle (Nora et al., 2017). However, looking closely at their supplementary data, there is an evident enrichment of the G1/S fraction to the detriment of the S/G2/M phase, supporting the reported slower proliferation and an accumulation of cells at the G1 checkpoint. This might reflect an impairment for the cell to reconstruct the required chromatin conformation necessary to maintain cell identity and ensure correct proliferation. Interestingly, CTCF has been previously reported as a regulator of cell cycle arrest (Qi et al., 2003), and its overexpression results in S phase block (Agarwal et al., 2017). This might suggest a role for CTCF in rearranging chromatin in the EG1 phase, when cell identity needs to be re-established, but also a role in regulating correct DNA replication. Its depletion induces G1 accumulation when chromatin cannot fold properly, and its upregulation prevents its eviction from chromatin during the S phase, impairing DNA synthesis, where it is thought to compete with the replication machinery for binding sites (Agarwal et al., 2017).

The model we propose, summarising the literature cited above, is the following: during mitosis, most TFs dissociate from chromatin and this might cause loss of higher 3D structures such as compartments. Then the chromosomes stochastically fold, but TADs boundaries and enhancers remains marked by bookmarking
proteins or remain in a nucleosome-free status. This bookmarking does not affect the stochastic folding of the genome, but ensure the rapid re-establishment of TADs and re-binding of TFs in the EG1 phase to maintain cell identity. We propose that this is regulated by CTCF, that shows strong binding in EG1, indicating that it might be acting as a chromatin remodelling factor only at the beginning of the cell cycle, with only a TF role during the rest of the cell cycle. After TADs are established, A/B compartments are reconstructed, and chromatin compacts, but only later in the S phase.

This model described an important structural function for CTCF after mitosis, but did not reveal insights on how the transcriptional network is re-established. Moreover, ATAC-seq showed that chromatin dynamically changes at important developmental genes, but the link with their expression in yet unknown. Ultimately, CTCF ChIP-seq have shown for the first time a cell cycle regulated binding for a transcription factor, suggesting that not only gene expression but also ChIP-seq dataset should be reconsidered in terms of cell phase heterogeneity. Furthermore, it will be fundamental to establish whether gene expression correlates with chromatin dynamics, and most importantly whether CTCF preferential occupancy in the EG1 phase is shared among all transcription factors or it is instead a unique feature with a functional role.
In the previous section, we showed that chromatin conformation changes during the cell cycle. In addition to this, chromatin openness is known to be a generally favourable environment for different TFs to bind and regulate gene expression. Accordingly, developmental genes have been reported to be transitorily transcribed in the G1 phase at very low levels, to then become fully activated in response to differentiation cues (Singh et al., 2013). This characterise the G1 phase as a time window when cells are responsive to differentiation signals. In support of this, cell fate propensity also changes during cell cycle progression, in agreement with gene expression patterns (Pauklin and Vallier, 2013). However, the link between chromatin openness and gene expression during cell cycle has not been investigated yet.

We have shown that a dynamic chromatin state during the cell cycle might be responsible for maintaining hESCs plasticity, but a direct link is still missing. In order to address this question, we performed RNA-seq on sorted FUCCI-hESCs and identified cell cycle specific expression patterns that recapitulates differentiation propensity in the G1 phase. Interestingly, activation of transcription mostly overlaps with an open chromatin state, confirming a link between chromatin openness and gene expression. However, the overlap characterises mostly constantly open regions, whereas differentially expressed genes do not correspond to dynamic regions. This suggests that additional mechanisms are required to activate alternative fates and expression patterns during the cell cycle, and that
local chromatin rearrangements do not control gene expression in hESCs. Alternatively, these changes could be needed to create a favourable environment in hESCs in order to receive and respond to differentiation signals. Differential binding of TFs or epigenetic modifiers could be playing an important role for such regulation.

5.1 Gene expression changes during cell cycle progression

In order to investigate gene expression changes during the cell cycle, we sorted FUCCI-hESCs in the different cell cycle phases (EG1, LG1, S/G2/M) and performed RNA-seq. From triplicates, QC analyses confirmed good quality of the library, and consistency across the replicates (Figure 5.1). Nearly 100% of reads fell into genes/exons, no ribosomal RNA (rRNA) or mitochondrial (MT) contamination, and confirms the opposing-strand specific library prep. After normalisation of reads in log_{2}RPM, the reads distribution showed a good overlap on a Cumulative Count Distribution Plot, and DataStore Tree plot showed expected clustering of the samples.

Figure 5.1 RNA-seq QC.
(A) RNA-Seq QC plot reports percentages of reads falling into genes, exons, rRNA, mitochondrial and sense-strand (indicating an opposing strand specific library) for the different samples. (B) Cumulative Distribution Plot, reporting data normalisation. This plot orders the probe values from lowest to highest, and then samples this set of values at common percentiles through the distribution. The x-axis reports the distribution, and the y-axis the value that the probe in that position has in the data store. (C) Datastore Tree reports clustering of different samples/replicates. This uses a Pearson correlation to calculate a distance matrix between all of the data stores, and it constructs a neighbour joining tree from the matrix.
Differential expression analyses showed 394 differentially expressed genes (Figure 5.2), consistent with previous reports (Singh et al., 2013). Gene Ontology analyses showed an enrichment for cell cycle terms as top hits for biological processes, with most of the genes known to be regulators of G2/M transition and cell division, such as *CDK1*, *CDC25C* and *AURKA* (Figure 5.2). In agreement with this, pathway analyses also showed enrichment for cell cycle related terms. Hierarchical clustering grouped differentially regulated genes in 3 clusters, namely Cluster 5, 7 and 12 (Figure 5.3). The most populated Cluster 5 was recapitulating the general trend, with genes mostly activated during the S/G2/M phase and gradually downregulated during the G1, to show the lowest expression in the LG1. We found in this group important G2/M regulators such as *CDK1* and CycB1 (*CCNB1*) and *CDC25C*, or S/G2 transition such as CycA2 (*CCNA2*). Gene ontology confirmed cell cycle terms enrichment, for both biological processes and pathways. Cluster 12 followed in number grouping a list of genes mildly downregulated during cell cycle progression from EG1 to S/G2/M. However, ontology enrichment for biological processes was quite low, mostly made by metabolic terms, suggesting that during cell cycle metabolic activity might decrease during cell cycle progression. Interestingly, pathway analysis showed enrichment for chromatin rearrangement terms, such as promoter opening and histones acetylation/deacetylation, in agreement with our finding of highly dynamic chromatin around the EG1 phase. Ultimately, Cluster 7 showed as well extremely interesting features: among upregulated transcripts in LG1, top hit for biological processes ontologies was endoderm formation and general developmental processes, including key endoderm genes such as *EOMES, MIXL1, SOX17, FOXA2, GATA4* and *GATA6*. These results have been also independently validated by qPCR (Figure 5.4).
Figure 5.2 Differentially expressed genes during cell cycle.

(A) Hierarchical Profile Plot reporting 394 differentially expressed genes in the different cell cycle phases, each column represents one replicate (p-value cut-off < 0.05, per-probe normalised). (B) Examples of the differentially expressed cell cycle regulator CDK1, CDC25C, AURKA, all up-regulated in the S/G2/M phase. (C) Gene ontology for Biological Processes and Reactome Pathway analysis for the 394 differentially expressed genes. Combined score is calculated by multiplying the unadjusted p-values with the z-scores.

Figure 5.3 Differentially expressed genes Gene Ontology.

Hierarchical Profile Plot reporting 394 differentially expressed genes in the different cell cycle phases, divided in 3 main clusters, each column represents one replicate (p-value cutoff < 0.05), and relative gene ontology for Biological Processes. Combined score is calculated by multiplying the unadjusted p-values with the z-scores.
Taken together, these results reveal an important cell phase specific transcription pattern, and confirmed that the mechanisms occurring in G1 “prime” hESCs for differentiation through the expression of transcripts that are important for lineage specification and development.

5.2 Chromatin regulated transcriptome changes

In order to assess whether gene expression correlates with chromatin openness, and most importantly whether cell cycle regulated transcripts are activated by changes in chromatin conformation, we looked at the overlap of features between the two datasets. We firstly simply overlapped all of the 13,478 expressed transcripts (log$_2$RPM > 1) with all the associated features of the ATAC-seq peaks: ~85% of total transcripts overlap with ATAC-seq peaks (Figure 5.5), meaning that most expressed genes are indeed in regions of open chromatin. To note, the number of expressed genes is similar in all the phases, with an increase of ~200 genes during the LG1 phase, which interestingly correlates with the maximum of ATAC peaks. However, when looking at phase specific overlaps, we found that ~83% of expressed transcripts overlaps with ATAC-peaks in EG1, ~70% in LG1, and ~80% in S/G2/M (Figure 5.5). This suggested that, even though the overlap is quite high in all the phases, the differentially expressed genes cannot be explained only by chromatin openness changes since the percentage decreases in LG1. Specifically, considering that the number of total transcripts increased in LG1, this highlighted that these transcriptional changes might not be regulated by chromatin rearrangements. In order to test this, we overlapped the list of differentially expressed genes with the list of dynamic ATAC-regions described in 4.2.

**Figure 5.4 qPCR validation of RNA-seq differentially expressed developmental genes.**

RNA-seq log$_2$RPM values for the genes SOX17, EOMES, and MIXL1, validated by q-PCR (values relative to housekeeping gene PBGD).
Strikingly, out of this comparison only ~3% (13 genes) of differentially expressed genes overlapped with dynamic regions, whereas the majority was part of constantly open regions (Figure 5.6). This observation implies that dynamic changes in chromatin structure are not involved in transcriptional regulations, and that additional mechanisms, likely to involve transcription factors regulations, are required to activate transcription at specific stages of the cell cycle.

![Figure 5.5 ATAC-seq / RNA-seq overlap.](image)

(A) Venn diagram reporting the overlap of expressed genes (RNA-seq) and chromatin open regions (ATAC-seq). 77% represents the percentage of expressed genes (log2RPM > 1) present in open chromatin regions, as overlapping with ATAC-seq peaks. (B) same as (A), but in the different cell cycle phases EG1, LG1, and S/G2/M. (C) Number of expressed genes (log2RPM > 1) in the different cell cycle phases.

![Figure 5.6 Differentially expressed genes / differentially open regions overlap.](image)

Venn diagram reporting the overlap of differentially expressed genes (RNA-seq diff) and ATAC dynamic regions (ATAC-seq diff). 3.3% represents the percentage of differentially expressed genes to which correspond changes in chromatin structure.
Taken together, these results suggested that overall gene expression correlates with chromatin openness. However, small changes in the expression patterns such as variation in the expression of cell cycle regulated genes do not correlate with chromatin changes. This suggests that their activation or repression relies on alternative mechanisms, such as differential binding of TFs or epigenetic modifiers activity during the cell cycle.

5.3 Discussion

In the previous chapter, we have described chromatin dynamics during cell cycle progression, but how this links to hESCs transcriptome is yet unknown. Here, in order to address this question, we performed RNA-seq on sorted FUCCI-hESCs and identified a subset of genes showing cell phase specific fluctuations. Most of these genes, such as CDK1, CDC25C and AURKA, are known to be cell cycle regulators, necessary for G2/M transition and cell division. Thus, these variations are related to cell cycle regulation/progression and not directly to cell fate choice. However, an interesting group of developmental regulators, mostly involved in endoderm differentiation, was found to be upregulated in the LG1 phase. This is in agreement with previous reports (Singh et al., 2013), and includes important genes such as SOX17, EOMES, GSC, MIXL1, FOXA2, GATA6 and GATA4, among many others. These developmental regulators are transiently expressed by a cell cycle-dependent mechanism in the LG1 phase, that represents a narrow window of time in which hESCs are indeed responsive to differentiation signals. This means that in pluripotency condition, differentiation signals are not sufficient to disrupt self-renewal and induce the expression of developmental regulators to a threshold that could induce lineage specification. However, low expression of these genes is transiently activated in G1, and this “primed” state allows a quick up-regulation in case of differentiation signals.

Interestingly, activation of transcription largely overlaps with an open chromatin state, with an average of 80% of expressed genes showing an open chromatin conformation, suggesting a link between chromatin openness and gene expression. However, the overlap characterises constantly open regions, whereas at differentially expressed genes do not correspond changes in chromatin structure. Indeed, only 3% of the differentially expressed genes correspond to chromatin conformation changes, suggesting that additional mechanisms are required to activate alternative fates and expression patterns during the cell cycle. We therefore hypothesised that differential binding of TFs, chromatin marks or
Epigenetic modifiers could be playing an important role for such regulation. Recent reports have investigated this and they shown that, especially at bivalent domains, H3K4me3 changes in self-renewing hESCs, corresponding to a transient increase at developmental genes during the LG1 phase, regulated by CDK2-dependent recruitment of MLL2 (Singh et al., 2015). However, it is still unknown what makes the G1 phase unique in terms of cell fate decision. Here, we describe a model where an open chromatin conformation correlates with steady-state gene expression, but its changes cannot explain cell cycle regulation of some transcripts. However, conformational changes might create different structural environments for the binding of TFs or epigenetic modifiers.
Chapter 6: Pluripotency transcriptional network during cell cycle progression

6 PLURIPOTENCY TRANSCRIPTIONAL NETWORK DURING CELL CYCLE PROGRESSION

While chromatin remodelling dynamics revealed a tight link with general transcription, they do not address changes in cell cycle regulated transcription and do not fully explain difference in fate propensity during cell cycle. For this reason, we then hypothesised that alternative mechanisms regulate these processes, perhaps using open chromatin as a template for differential binding of TFs or epigenetic modifiers. In addition to this, GO analyses for ATAC-seq peaks revealed, together with general regulation of transcription, terms involved in positive regulation of transcription by TFs localisation. This links to ATAC peaks several processes that activate transcription using a mechanism that involves the localization of a TFs. Moreover, as we have already shown for CTCF, differential binding of TFs during the cell cycle could be an interesting way to quickly control induction of differentiation. Previous work in the lab already reported cell cycle regulated binding of SMAD2/3, that in turns affect cell fate propensity (Pauklin and Vallier, 2013). In this work, we showed how differentiation in early lineages is only allowed in the G1, with hESCs differentiation into mesendoderm in EG1 and into neuroectoderm in LG1. However, little is known on how the pluripotency network counterbalances this differentiation propensity and maintains cell identity, and cell
cycle regulation of alternative factors might be an interesting mechanism to investigate. Likely candidates appeared to be OCT4, SOX2 and NANOG, well-known master regulators of pluripotency and differentiation in hESCs. In the introduction, we have described how these factors have been shown to regulate and being regulated by the cell cycle, and interestingly, they also share binding patterns with SMAD2/3. Indeed, NANOG interacts directly with SMAD2/3 to regulate their transcriptional activity in the nucleus, and ChIP analyses showed that they also bind to the same genomic region in several target genes (Vallier et al., 2009a). In addition, SMAD2/3 can bind endoderm genes without inducing their expression, suggesting the presence of other factors blocking the expression of such genes. Indeed, down regulation of NANOG upon differentiation releases SMAD2/3, which can then interact with additional partners (Brown et al., 2011).

During differentiation, OCT4 has also been described as co-effector of SMAD2/3. It has been shown that during Primitive Streak (PS) induction, OCT4 binds several regulatory regions in proximity of β-catenin (main effector of the Wnt pathway) and SMAD2/3, and this interaction is required for gene activation (Funa et al., 2015). Therefore, being OCT4, SOX2 and NANOG tightly linked to the cell cycle regulated factor SMAD2/3, and themselves connected with cell cycle regulation, we decided to investigate their binding patterns during cell cycle progression.

Interestingly, we showed that core pluripotency factors bind chromatin mostly in the S/G2/M phase, when cells are not responsive to differentiation cues. We also confirmed that core pluripotency factors work as a complex to maintain the pluripotency network and block differentiation in all the cell cycle phases, but to different degrees: the transcriptional network they regulate is weaker in the G1 phase when hESCs are responsive to differentiation signals, whereas is stronger and tightly interconnected during the S/G2/M phase, making cells unresponsive to differentiation stimuli.

6.1 Cell cycle regulation of OCT4

Interesting hints were given already by additional terms enriched in different databases with ATAC-seq peaks and RNA-seq expressed genes. As already mentioned, GO analyses for ATAC-seq peaks revealed terms involved in positive regulation of transcription by TFs localisation in all the phases. In addition to this, TFs loss-of-function (LOF) databases from Gene Expression Omnibus (GEO), a database that links gene expression signatures to TFs perturbations, highlighted in all phases but EG1 an enrichment for OCT4, suggesting that its activity might be
indispensable only in the later cell cycle. Moreover, when looking at dynamic regions opening during EG1 to LG1 transition, we found enrichment for several POU family members, according to the TRANSFAC and JASPAR PWMs database, which scans for binding motifs at the promoters of the input genes. Finally, also differentially expressed genes in the LG1 phase showed enrichment for OCT4 in TF-LOF and motif scans identified binding of SMAD3, PRC components and OCT4 (Figure 6.1).

<table>
<thead>
<tr>
<th>TFs LOF</th>
<th>LG1 ATAC-seq</th>
<th>#1</th>
<th>#9</th>
</tr>
</thead>
<tbody>
<tr>
<td>POU5F1_hESCs_up</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POU5F1_hESCs_down</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRANSFAC PWMs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POU1F1</td>
<td></td>
<td>#1</td>
<td></td>
</tr>
<tr>
<td>POU2F2</td>
<td></td>
<td>#6</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TFs LOF</th>
<th>LG1 RNA-seq</th>
<th>#1</th>
<th>#10</th>
</tr>
</thead>
<tbody>
<tr>
<td>POU5F1_hESCs_up</td>
<td></td>
<td></td>
<td>#1</td>
</tr>
<tr>
<td>NANO2_mESCs_up</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ChEA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMAD3_hESCs</td>
<td></td>
<td>#1</td>
<td></td>
</tr>
<tr>
<td>POU5F1_hESCs</td>
<td></td>
<td>#4</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6.1 Additional Ontologies for ATAC-seq and RNA-seq.

Additional enrichment analyses for transcription factors loss-of-function (TFs LOF) database (links gene expression signatures to TFs perturbations), TRANSFAC and JASPAR PWMs database (scans for binding motifs at the promoters of the input genes), and ChEA (a library of transcription factors featured in papers in pubmed), in ATAC-seq and region and upregulated genes in the LG1 phase.

All these, taken together, support the hypothesis that OCT4 downstream activity might be required only in the last part of the cell cycle to sustain the pluripotency network, whereas being dispensable in the EG1. In order to investigate OCT4 binding dynamics during the cell cycle, we sorted FUCCI-hESCs in the different cell cycle phases (EG1, LG1, S/G2/M) and performed ChIP-seq to study genome occupancy. From two replicates, using the IDR pipeline (see Material and Methods, 2.6.2) we found a total of 4,843 reproducible peaks in the different phases, 611 peaks in EG1, 1,378 in LG1, and 2,854 in S/G2/M, suggesting that OCT4 binding increased during cell cycle progression. Genome wide mapping showed again weakest binding in the EG1, increasing in the LG1, to reach higher level of binding in the S/G2/M phase (Figure 6.2), thereby confirming our original hypotheses. Of these, 3,391 peaks were unique, with the majority (68%) present in a single phase, 21% in two phases, and only 11% during all cell cycle. Of note, 1,787 of the unique peaks were found in S/G2/M only, 53% of all unique peaks (Figure 6.3). Feature distribution did not reveal a cell cycle specific location, with most peaks present in introns and intergenic regions in all the phases (Figure 6.4), and after associating peaks with genomic regions using GREAT, we confirmed that most features were found in distal regulatory regions (+/- 500 to 50 kb and +/- 50 to 5 kb).
Figure 6.2 OCT4 ChIP-seq.

(A) Metaplot of ChIP-seq signal for OCT4 +/- 1kb around binding sites in EG1, LG1, S/G2/M, compared to input. (C) Heat-maps display normalised OCT4 ChIP-seq signal centred on the merged peaks, +/- 1kb, for all the replicates in the different cell cycle phases.

Figure 6.3 OCT4 ChIP-seq Venn overlap.

(A) Total number of peaks in the different cell cycle phases for OCT4 ChIP-seq. (B) Venn diagram showing peaks overlap for the OCT4 ChIP-seq in EG1, LG1, S/G2/M, and table summary reporting number and percentage of peaks found in a single phase, in two phases, or all phases.
Chapter 6: Pluripotency transcriptional network during cell cycle progression

Pathways analyses showed enrichment in all the phases for pluripotency networks terms, as well as differentiation (Figure 6.5). Interestingly, the top hit in the S/G2/M phase was indeed pluripotency network, whereas in LG1 was ectoderm differentiation.

**Figure 6.4 OCT4 ChIP-seq features distribution.**

(A) Feature distribution for the OCT4 ChIP-seq peaks in EG1, LG1, S/G2/M. (B) Distribution of region-genes associated regions around transcription starting site (TSS), binned by distance in kb (0-5, 5-50, 50-500, >/< 500).

**Figure 6.5 OCT4 ChIP-seq Gene Ontology.**

Gene ontology for Biological Processes in the different cell cycle phases EG1, LG1, S/G2/M for OCT4 ChIP-seq gene associated regions. Combined score is calculated by multiplying the unadjusted p-values with the z-scores.
Taken together, these results revealed a novel mechanisms of cell cycle regulated binding of the core pluripotency factor OCT4. Its increased binding in the S/G2/M phase could preserve a pluripotency identity during such phase, whereas lower occupancy in G1 could permit differentiation, in a permissive chromatin status. This is in agreement with a recent study that shows how the cell cycle states dominantly block pluripotent state dissolution (PSD) in hESCs (Gonzales et al., 2015). Specifically, the S/G2 phase is shown to attenuate PSD because of intrinsic propensity towards the pluripotent state, independent of G1 phase, confirming our findings.

6.2 Cell cycle regulation of SOX2 and NANOG

This novel observation that one of the core pluripotency factors did not have a consistent binding pattern during cell cycle progression, led us to hypothesise that other members of the core pluripotency network such as SOX2 and NANOG could behave similarly. In order to validate this hypothesis, we sorted FUCCI-hESCs in the different cell cycle phases (EG1, LG1, S/G2/M) and performed additional ChIP-seq to study genome occupancy for SOX2 and NANOG. From two replicates each, using the IDR pipeline (see Material and Methods, 2.6.2) we found a total of 5,190 reproducible peaks for SOX2 and 3,321 reproducible peaks for NANOG, distributed in the different cell cycle phases. As previously found, the majority of peaks were found in both cases in the S/G2/M, with 750 peaks in EG1, 101 in LG1, and 4,339 in S/G2/M for SOX2, and 84 in EG1, 215 in LG1, and 3,022 in S/G2/M for NANOG (Figure 6.6), confirming an increased number of binding sites during cell cycle progression.

**Figure 6.6 SOX2 and NANOG ChIP-seq.**

Total number of peaks in the different cell cycle phases for SOX2 and NANOG ChIP-seq.
Genome wide mapping showed that both SOX2 and NANOG bind a lower number of genomic regions in the EG1 and LG1, while reaching a maximal number of binding sites in the S/G2/M phase (Figure 6.7). For SOX2, 4,370 unique peaks were identified, with the majority (83%) present in a single phase, 15% in two phases, and only 2% shared between all the phases. Similarly to OCT4, the majority of peaks (3,600) were only present in the S/G2/M phase (Figure 6.8). NANOG showed a similar pattern: of 3,076 unique peaks, 93% (2,896) were present in a single phase, 170 in two of the phases and only 37 shared between the phases. As for OCT4 and SOX2, the majority of unique NANOG peaks (2,816) were found only in the S/G2/M phase (Figure 6.8).

Figure 6.7 SOX2 and NANOG Heatmaps and Metaplots.

(A) Metaplot of ChIP-seq signal for SOX2 and NANOG +/- 1kb around respective binding sites in EG1, LG1, S/G2/M, compared to input. (C) Heat-maps display normalised SOX2 and NANOG ChIP-seq signal centred on the merged peaks, +/- 1kb, for all the replicates in the different cell cycle phases.
For both SOX2 and NANOG, feature distribution did not reveal a cell cycle specific location, with most peaks present in introns and intergenic regions in all the phases (Figure 6.9), and after associating peaks with genomic regions using GREAT, we confirmed that most features were found in distal regulatory regions (+/- 500 to 50 kb and +/- 50 to 5 kb). Pathways analyses showed terms involved in both pluripotency maintenance and differentiation. For SOX2, LG1 phase top rank enrichment was ectoderm differentiation. Instead, NANOG top hit in EG1 was endoderm differentiation, whereas LG1 were ectoderm terms (Figure 6.10), reinforcing previously described cell fate propensity (Pauklin and Vallier, 2013). However, it is important to note that all the phases ranked as top hits both pluripotency and differentiation terms, with small differences in p-values. Thus, a tight balance could exist between their binding patterns, thereby regulating differentiation propensity, and that this cannot be ascribed to a single factor binding across the genome.

Figure 6.8 SOX2 and NANOG Venn overlap.
Venn diagrams showing peaks overlap for SOX2 and NANOG ChIP-seq in EG1, LG1, S/G2/M, and table summary reporting number and percentage of peaks found in a single phase, in two phases, or all phases.
Chapter 6: Pluripotency transcriptional network during cell cycle progression

Gene ontology for Biological Processes in the different cell cycle phases EG1, LG1, S/G2/M for SOX2 and NANOG ChIP-seq gene associated regions. Combined score is calculated by multiplying the unadjusted p-values with the z-scores.

Figure 6.9 SOX2 and NANOG features distributions.

(A) Feature distribution for SOX2 and NANOG ChIP-seq peaks in EG1, LG1, S/G2/M. (B) Distribution of region-genes associated regions around transcription starting site (TSS), binned by distance in kb (0-5, 5-50, 50-500, >/500).

Figure 6.10 SOX2 and NANOG Gene Ontologies.
6.3 Combinatorial binding of core pluripotency factors

The core pluripotency factors are known to function together to positively regulate their own promoters, forming an interconnected auto-regulatory loop (Boyer et al., 2005). Considering how single factors have been shown to dynamically bind during cell cycle progression, we decided to investigate the effects on regulation of the OCT4/SOX2/NANOG (OSN) network. We firstly overlapped peaks in the three datasets, and showed that of the 5,904 unique peaks, 30% (1,767) were shared, with the majority involved in signalling pathways regulating pluripotency (Figure 6.11).

![Venn diagram showing peaks overlap for OCT4, SOX2 and NANOG ChIP-seq](image)

**Figure 6.11 OCT4, SOX2, NANOG Venn overlap.**

Venn diagrams showing peaks overlap for OCT4, SOX2 and NANOG ChIP-seq, and table summary reporting gene ontology for biological processes for shared peaks (OSN), and OCT4/NANOG/SOX2 only peaks. Combined scores are reported, calculated by multiplying the unadjusted p-values with the z-scores.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Term</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSN</td>
<td>Signalling pathways regulating pluripotency</td>
<td>42.45</td>
</tr>
<tr>
<td>OCT4</td>
<td>Histone deubiquitination</td>
<td>25.88</td>
</tr>
<tr>
<td>NANOG</td>
<td>Insulin receptor signalling pathway</td>
<td>24.44</td>
</tr>
<tr>
<td>SOX2</td>
<td>Positive regulation of transcription involved in neuron differentiation</td>
<td>28.97</td>
</tr>
</tbody>
</table>

Single factor binding revealed specific features: OCT4 only peaks were enriched for deubiquitination terms, NANOG only for insulin receptor signalling pathway, and SOX2 regulation of transcription involved in neuron differentiation. Cell phase specific analyses showed similar trends to the single TFs (Figure 6.12). Interestingly, whereas OCT4 seemed to gradually bind more sites during cell cycle progression, SOX2 and NANOG occupancy are mutually exclusive during the G1 phase: in the EG1, SOX2 was binding more sites, whereas in the LG1 NANOG was more prominent, suggesting that they might regulate alternative cell fates in during the G1. The EG1 phase was enriched for terms related to pluripotency, however, OCT4 and SOX2 only bound-regions were mostly ectoderm genes (*PAX6, OLIG3, TFAP2C, CTNNB2*), whereas very few genes were only bound by NANOG, some of them related to endoderm differentiation (*FOXA1, DKK1*). In the LG1 we found enriched terms for pluripotency and differentiation (top hit ectoderm
differentiation), with OCT4 and NANOG only bound-regions mostly ectoderm genes again, and very few only-SOX2 bound genes, with some genes related to mesendoderm commitment. The S/G2/M was enriched again for pluripotency terms for all the TFs, with a combination of pluripotency and differentiation terms in the single-factor bound genes (Ectoderm for SOX2 and Endoderm for NANOG) (Figure 6.12).

Figure 6.12 OCT4, SOX2, NANOG Gene ontology.
(A) Gene ontology for biological processes in OSN dataset in the different cell cycle phases. Combined scores are reported, calculated by multiplying the unadjusted p-values with the z-scores. (B) Venn overlap for OSN in the different cell cycle phases, EG1, LG1, S/G2/M, and table summary reporting gene ontology for biological processes in OSN overlaps for genes only bound by, respectively, OCT4, SOX2 and NANOG in the different cell cycle phases. When marked with *, terms are not significant.

Taken together, these results suggested that the core pluripotency factors sustain the pluripotency network in all the cell cycle phases acting as a complex, but also reveal that single factor could specifically regulate differentiation mostly in the G1 phase.
6.4 Transcriptional regulation and binding patterns of pluripotency factors

Core transcription factors have been reported as co-occupying and activating expression of genes necessary to maintain the pluripotent state, while contributing to repression of genes encoding for lineage-specific genes (Boyer et al., 2005). In order to distinguish these two complementary roles and better characterise the transcriptional network regulated by the core pluripotency factors OCT4, SOX2 and NANOG, we overlapped TFs binding with RNA-seq data. OCT4 binding alone overlapped with 20% of all the expressed genes, and similar occupancy was found for NANOG (16%) and SOX2 (21%). When combined, the three factors together were found to occupy regulatory regions of 30% of the expressed genes, confirming their key role in governing the pluripotency transcriptional network (Figure 6.13). As expected, the overlapping regions included the OSN auto-regulatory loop (*POU5F1, SOX2, NANOG*), and top hit for pathway analysis was indeed pluripotency network. However, 32% (1,924) of bound genes were not expressed, confirming a potential repressive role. Of these genes, GO analyses showed enrichment for mostly neuronal targets, confirming that, as previously shown, most of the genes bound but not activated by OSN are involved in neuroectoderm development (Figure 6.14).

![Figure 6.13 OCT4, SOX2, NANOG ChIP-seq / RNA-seq overlap.](image)

(A) Percentage of expressed genes (log2RPM > 1) bound by OCT4, SOX2, and NANOG. (B) Venn diagram reporting the overlap of expressed genes (RNA-seq) and OSN ChIP-seq. 30% represents the percentage of expressed genes (log2RPM > 1) bound by OCT4, SOX2 or NANOG, as overlapping with OSN peaks.
Interestingly, phase specific binding confirmed what previously found, but revealed additional mechanistic dynamics: 32% (508) of the genes were bound but not expressed in EG1, 31% (620) in LG1, and 32% (1,823) in S/G2/M. In all phases, top hit for expressed genes was PluriNetWork, whereas not expressed-bound genes — therefore repressed — were involved in neural development (Figure 6.15).

Figure 6.14 OCT4, SOX2, NANOG repressed genes GO.

Gene ontology for Biological Processes of OSN repressed genes (bound by ChIP-seq but not expressed by RNA-seq). Combined score is calculated by multiplying the unadjusted p-values with the z-scores.

Figure 6.15 OCT4, SOX2, NANOG ChIP-seq / RNA-seq overlap.

Venn diagrams reporting overlap for OSN ChIP-seq and expressed genes (RNA-seq) in EG1, LG1, and S/G2/M. Percentages represent the number of repressed genes (bound by ChIP-seq but not expressed by RNA-seq). Relative gene ontologies for Biological Processes of OSN repressed genes. Combined score is calculated by multiplying the unadjusted p-values with the z-scores.
While the ratio of active/inactive genes remained constant during the cell cycle, it was evident that the absolute number of ON/OFF bound genes increased during cell cycle progression. To speculate, this could suggest a “looser” OSN network in the G1 phase, with less pluripotency/active and differentiation/repressed genes than the S/G2/M phase. This could allow hESCs to be more responsive to differentiation cues in the G1 phase, whereas sustain a stronger pluripotency network in the S/G2/M phase. It is important to notice that, both at the transcriptional and protein levels, OSN showed no differences in the different cell cycle phases (Figure 6.16). This suggested that their differential binding could be regulated by post-translational modification, and rather than their expression, is their binding to the chromatin that regulates the transcriptional network.

![Figure 6.16 OCT4, SOX2, NANOG RNA-seq, qPCR and WB during cell cycle.](image)

(A) qPCR for OCT4, NANOG and SOX2 in EG1, LG1, and S/G2/M, relative to housekeeping gene PBGD (two-ways ANOVA, n=3). (B) RNA-seq data for OCT4, SOX2, and NANOG, in EG1, LG1, and S/G2/M, in log2RPM. (C) Western-blot and relative quantification for OCT4, SOX and NANOG in EG1, LG1, and S/G2/M, plus unsorted (UNS) control (α-tubulin as loading control, n=1).

Taken together, these results confirmed that the core pluripotency factors work as a complex to maintain the pluripotency network and block differentiation in all the cell cycle phases, but the transcriptional network they regulate is weaker in the G1 phase.
phase when hESCs are responsive to differentiation signals, whereas is stronger and tightly interconnected during the S/G2/M phase.

6.5 Cell cycle regulation of RING1B

The novel finding of a cell cycle regulated OSN transcriptional network raised the question whether other TFs or epigenetic modifiers could follow the same binding pattern, or whether different types of regulation happen during the cell cycle. Among several OSN interactors, we decided to investigate the PRC1 complex, and more specifically one of its main subunits RING1B. Indeed, in mESCs RING1B has been shown to be functionally linked to the pluripotency network, since its engagement at target genes is OCT4-dependent (Endoh et al., 2008). However, PRC1 has been described having several roles, from chromatin compaction to positive and negative regulation of expression (Aranda et al., 2015). PRC1 is an E3 ubiquitin (Ub) ligase that, through its canonical activity, transfers the mono-Ub mark to the C-terminal tail of Histone H2A at K119 inducing RNAPII pausing (Cao et al., 2005; Wang et al., 2004). Nonetheless, PRC1 is also capable of chromatin compaction, a histone-independent function, and this activity appears also important in gene silencing (Francis et al., 2004). H2A ubiquitination is dispensable for chromatin compaction but it is essential for maintaining developmental genes in a repressed state during pluripotency (Endoh et al., 2012). Ultimately, it also has been described as activators of gene expression during development and in non-canonical PRC conformations (Frangini et al., 2013; Gao et al., 2014; Kondo et al., 2014). Interestingly, RING1B deficiency has been shown to cause gastrulation arrest and cell cycle inhibition (Voncken et al., 2003), emphasising its important role in cell cycle progression.

For these reasons, we decided to investigate RING1B genome occupancy during cell cycle progression, performing ChIP-seq in FUCCI-hESCs sorted cells as previously described. Interestingly, RING1B showed a unique binding pattern, with high occupancy during the EG1 phase, sharply decreasing during the LG1 phase, to then partially increase during the S/G2/M. From two replicates, using the IDR pipeline (see Material and Methods, 2.6.2) we found a total of 22,535 reproducible peaks, of which 11,932 peaks were found in EG1, 3,068 peaks in LG1, and 7,535 peaks in S/G2/M (Figure 6.17). Genome wide mapping confirmed peak calling reporting the strongest binding in the EG1 phase, which decreased in the LG1, and slightly increased in the S/G2/M (Figure 6.17). Of these, 13,357 peaks were unique, with 50% present in a single phase (6,680), 33% in two phases (4,367),
and 17% present in all phases (2,310). Of note, 5,257 of the unique peaks were found in EG1 (Figure 6.18). Feature distribution did not reveal a cell cycle specific location, with most peaks covering genes, from 5UTR to gene body, including promoters, introns and exons, in all the phases (Figure 6.19), and after associating peaks with genomic regions using GREAT, we confirmed that most features were around the TSS (+/- 0 to 5 kb), preferentially upstream.

Figure 6.17 RING1B ChIP-seq.
(A) Total number of peaks in the different cell cycle phases for RING1B ChIP-seq. (B) Metaplot of ChIP-seq signal for RING1B +/- 1kb around its binding sites in EG1, LG1, S/G2/M, compared to INPUT. (C) Heat-maps display normalised RING1B ChIP-seq signal centred on the merged peaks, +/- 1kb, for all the replicates in the different cell cycle phases.

Figure 6.18 RING1B Venn overlap.
Venn diagram showing peaks overlap for RING1B ChIP-seq in EG1, LG1, S/G2/M, and table summary reporting number and percentage of peaks found in a single phase, two phases, or all phases.
Chapter 6: Pluripotency transcriptional network during cell cycle progression

Enrichment analyses revealed in all phases that RING1B binding controlled positively and negatively transcription from RNAPII promoters (Figure 6.20). Additional ontologies for cellular components showed interesting features regulating nucleosomes, DNA packaging, and PRC1 binding in all phases, with no specific function in any of the cell phases. Interestingly, EG1 only bound regions were enriched for CTCF binding sites, together with SMC3, core subunit of the cohesin complex (Wendt et al., 2008), whereas S/G2/M only bound regions were enriched for SUZ12 and EZH2, PRC2 components.

Figure 6.19 RING1B ChIP-seq features distribution.
(A) Feature distribution for RING1B ChIP-seq peaks in EG1, LG1, S/G2/M. (B) Distribution of region-genes associated regions around transcription starting site (TSS), binned by distance in kb (0-5, 5-50, 50-500, >/< 500).

Enrichment analyses revealed in all phases that RING1B binding controlled positively and negatively transcription from RNAPII promoters (Figure 6.20). Additional ontologies for cellular components showed interesting features regulating nucleosomes, DNA packaging, and PRC1 binding in all phases, with no specific function in any of the cell phases. Interestingly, EG1 only bound regions were enriched for CTCF binding sites, together with SMC3, core subunit of the cohesin complex (Wendt et al., 2008), whereas S/G2/M only bound regions were enriched for SUZ12 and EZH2, PRC2 components.

Figure 6.20 RING1B Gene Ontology.
Gene ontology for Biological Processes in the different cell cycle phases EG1, LG1, S/G2/M for RING1B ChIP-seq gene associated regions. Additional enrichment analyses in EG1 and S/G2/M for Encode Consensus TFs, an ENCODE annotated transcription factors library, listed by ranking.

99
Finding enriched terms for both positive and negative regulation of transcription, together with the fact that RING1B was found on regions co-bound by different factors during the cell cycle progression, prompted the hypothesis that RING1B binding could have indeed different roles in the different cell cycle phases. To support this, RING1B showed a binding pattern that only partially correlates with OCT4 activity, previously described as mostly bound to chromatin in the S/G2/M phase, suggesting that it might have an OCT4-dependent role only in the late cell cycle, whereas having an alternative role, perhaps structural, in the early cell cycle. In order to test our hypothesis, we firstly overlapped RING1B occupancy with CTCF in the different cell cycle phases. Strikingly, almost 40% of the RING1B bound regions were also occupied by CTCF in the EG1, whereas this percentage drops to 6% and 9% respectively in LG1 and S/G2/M (Figure 6.21), suggesting that RING1B might have divergent roles in the different phases. Specifically, we propose a structural role in the EG1 phase, where it collaborates with CTCF in reconstituting chromatin structure after mitosis, whereas it might have a canonical repressor function in the S/G2/M, possibly OCT4-dependent, when the pluripotency network is reconstituted.

![Venn diagrams reporting overlap for RING1B ChIP-seq and CTCF ChIP-seq in EG1, LG1, and S/G2/M. Percentages represent the number of genes bound by both RING1B and CTCF.](image)

**Figure 6.21 RING1B / CTCF ChIP-seq overlaps.**

Intrigued by its dual role, we investigated whether RING1B binding was only correlating with transcriptional repression. Strikingly, overlapping RNA-seq data with RING1B ChIP-seq, we found that only ~18% of the genes bound by RING1B are not expressed, and this is consistent in all the phases (Figure 6.22), meaning that the majority of genes occupied by RING1B are expressed in hESCs. However, in order to exclude transcriptionally-leaky genes bound by RING1B, we ranked the genes based on their expression, and selected the top 20%, to compare RING1B occupancy on the highest expressed genes with their activation.
We then overlapped this list with RING1B peaks, and found that ~50% of the genes bound by RING1B are highly expressed (Figure 6.22). On the other hand, looking at cell phase specific transcripts, in the EG1 only 35% of the expressed genes are not bound by RING1B, whereas in the LG1 78% are RING1B free, and 52% in S/G2/M (Figure 6.23).

Figure 6.22 RING1B ChIP-seq / RNA-seq overlap.

(A) Venn diagrams reporting overlap for RING1B ChIP-seq and RNA-seq in EG1, LG1, and S/G2/M. Percentages represent genes repressed by RING1B compared to the total of genes bound in the different cell cycle phases (RING1B bound but not expressed by RNA-seq). (B) Overlap of RING1B bound genes with highly expressed genes by RNA-seq top 20% (genes are ranked based on their expression, and the top 20% is selected). Percentage represents the number of genes bound and not expressed.

We then overlapped this list with RING1B peaks, and found that ~50% of the genes bound by RING1B are highly expressed (Figure 6.22). On the other hand, looking at cell phase specific transcripts, in the EG1 only 35% of the expressed genes are not bound by RING1B, whereas in the LG1 78% are RING1B free, and 52% in S/G2/M (Figure 6.23).

Figure 6.23 RNA-seq / RING1B ChIP-seq Venn overlap.

Venn diagrams reporting overlap for RING1B ChIP-seq and RNA-seq in EG1, LG1, and S/G2/M. Percentages represent the number of genes expressed and not bound by RING1B.
This reinforced the hypothesis of RING1B repressive role in LG1 and S/G2/M, where the majority or the expressed genes are RING1B free. Interestingly, the lowest RING1B occupancy was found to be in the LG1 phase, which also correspond to the phase with the highest percentage of RING1B-free expressed genes and, as described in the previous chapter, primed expression of developmental genes. Specifically, we hypothesised that RING1B lower occupancy during the LG1 phase could correlate with the priming of developmental genes in such phase. In order to test our hypothesis, we overlapped RING1B occupancy with the list of genes that showed increased expression in the LG1 phase (Cluster 7). Interestingly, 81% of the genes upregulated in the LG1 phase do not show RING1B occupancy. Moreover, the residue 19% showed a substantial reduction in RING1B occupancy, suggesting that decreased occupancy could anyway induce expression priming (Figure 6.24).

Figure 6.24 RING1B browser tracks.
(A) Browser tracks for RING1B ChIP-seq in the different cell cycle phases (EG1, LG1, S/G2/M) over the SOX17, MIXL1 and GSC loci. (B) Venn diagram reporting overlap for upregulated genes (RNA-seq) and RING1B ChIP-seq in LG1. Percentages represent the genes upregulated in LG1 and bound by RING1B.
All these results, taken together, showed that RING1B is cell cycle regulated, minimising its genome occupancy in the LG1 phase and correlating with increased expression of developmental genes. In addition to this, its higher occupancy in the EG1 phase strongly correlated with CTCF binding, suggesting a parallel chromatin remodelling role. To summarise, we hypothesise that, together with its canonical function as a transcriptional repressor, during the cell cycle RING1B has different roles. Its specific binding in the EG1 helps to re-structure chromatin after mitosis, together with CTCF. During the LG1, its eviction from chromatin allows priming of developmental genes and cells are responsive to differentiation stimuli. Finally, during the S/G2/M phase, it re-binds in an OCT4-dependent manner, driven by the re-establishment of the pluripotency network.

6.6 Discussion
In this chapter, we described for the first time the binding patterns of the core pluripotency factors during the cell cycle in hESCs. Interestingly, they have been shown to bind the chromatin mostly in the S/G2/M phase, whereas little binding has been found in the early cell cycle. Their higher binding during the S/G2/M phase could represent a pluripotency gatekeeper during such phase, whereas lower occupancy in G1 could permit differentiation, in a context of permissive chromatin. This agrees with the fact that the S/G2 phase has been shown to block the pluripotent state dissolution in hESCs (Gonzales et al., 2015). Moreover, these results suggest that the core pluripotency factors sustain the pluripotency network in all the cell cycle phases acting as a complex, but reveal that single factors could regulate differentiation in the G1 phase, mostly blocking neuroectoderm. Importantly, TFs are known to dissociate from chromatin during mitosis, and thus could strongly re-bind only later in the cell cycle, when chromatin structure have been defined, and cells are committed to complete another pluripotency cell cycle. This is in agreement with previous findings, where OCT4 has been shown to regulate mitotic entry and dissociate from chromatin during G2/M phase (Zhao et al., 2014). Other transcription factors have been shown to regulate cell cycle progression: NANOG, for instance, have been described regulating S-phase entry via CDK6 and CDC25A (Zhang et al., 2009). Interestingly, this paper showed that overexpression of NANOG in hESCs shortens the time needed for S-phase entry, and causes a significant increase in ESC proliferation while enhancing their pluripotent phenotype. We could speculate that NANOG overexpression increases its genome binding in all the phases of the cell cycle, forcing a rapid formation of
the pluripotency network, normally fully reformed only in the S/G2/M phase. However, it is important to note that RNA-seq and qPCR validations showed no difference in gene expression for the core pluripotency factor during cell cycle progression, and western blot confirmed it at the protein level. Therefore, alternative mechanisms should regulate this differential binding. Considering that many cell cycle related events rely on post-translational modification of proteins, to ensure quick activation/inhibition, it is reasonable to hypothesise that similar mechanisms regulate pluripotency. It is known, for example, that OCT4 can be phosphorylated at Ser236 in hESCs (Swaney et al., 2009), and this might influence the DNA binding and its transcriptional activity, since this residue is contained within the POU-homeodomain. In general, all the pluripotency factors have been described to be regulated by post-translational modifications, shown to be important to regulate their levels and activity and to balance between pluripotency and differentiation (Cai et al., 2012). Therefore, it is reasonable to hypothesise that a similar mechanism could be regulating their differential binding. Further work must be done in this direction in order to identify mechanisms of regulation of these factors.

The novel finding of a cell cycle regulated transcriptional network prompted the idea that other TFs or epigenetic modifiers could follow the same binding pattern. In order to investigate whether different types of regulations happen during the cell cycle among different epigenetic modifiers, we focused on RING1B, one of the main subunits of the PRC1 complex. RING1B has been described to have roles in chromatin compaction, but also in regulation of expression. It also has been shown to be functionally linked to the pluripotency network, since its engagement at target genes is dependent on OCT4 (Endoh et al., 2008). Through its canonical activity, PRC1 transfers the mono-Ub mark to the C-terminal tail of Histone H2A at K119 inducing RNAPII pausing (Cao et al., 2005; Wang et al., 2004). On the other hand, PRC1 exerts a histone-independent function to regulate chromatin compaction, and this activity appears also important in gene silencing (Francis et al., 2004). Polycomb, in general, have been shown to prevent premature expression of primed genes and thus to stably maintain a pluripotent ES cell identity in culture (Azuara et al., 2006). Indeed, it has been shown that a null mutation in RING1B leads to a reduction of PRC1 proteins and a loss of H2AK119ub1 in ES cells, inducing de-repression of lineage markers (Leeb and Wutz, 2007). Intrigued by this possibility, we investigated whether RING1B activity was cell cycle regulated, and whether we could dissect its multiple roles in a cell cycle regulated manner. The
first important observation was the fact that most RING1B bound genes were expressed in hESCs, with only ~20% of these genes transcriptionally repressed. Even the overlap with the top 20% of the expressed transcripts revealed that ~50% of the RING1B bound genes are still expressed, suggesting that its binding does not directly correlate with transcriptional repression, but instead was also linked to active transcription. Together with this, we also found RING1B to co-occupy the genome with CTCF in the EG1, with almost 40% of its binding sites also bound by CTCF. On the contrary, CTCF was found on only 6% and 9% of RING1B bound regions in LG1 and S/G2/M, respectively. This lead us to speculate on a structural role for RING1B in the EG1 phase, where it works with CTCF in re-structuring chromatin after cell division, whereas it could have a canonical repressor function in the S/G2/M, OCT4-dependent, when the pluripotency network is reconstituted. Interestingly, we also found that its genomic binding anti-correlates with priming of developmental genes in the LG1 phase, suggesting that its genome occupancy could correlate with repression of such genes. In more details, we found that in more than 80% of the genes upregulated in LG1 are not bound by RING1B, and the remainder 20% present a strong reduction in it occupancy. Moreover, it has been shown that RING1B blocks p16^{INK4A}, a CDK inhibitor that regulates CDK4/6 activity (Voncken et al., 2003). Interestingly, our RNA-seq show that p16^{INK4A} is indeed upregulated in the LG1 phase, when RING1B is evicted from its locus. This implies also an important developmental role for such regulation, since RING1B deficiency has been shown to cause gastrulation arrest and cell cycle inhibition due to deregulation of p16^{INK4A} (Voncken et al., 2003). To further support this, RING1B depletion have also been shown to increase H3K4me3 levels at its target genes (Endoh et al., 2008), and previous reports have shown how during the LG1 phase this mark is increased through a CDK2-dependent phosphorylation, supporting transcriptional leakiness of this phase we previously described (Singh et al., 2015). Here, the transcriptional repressor RING1B is indeed shown to anti-correlate with the upregulation of developmental genes in the LG1 phase, suggesting the idea that it may regulate their expression. It would be interesting to investigate mechanismistically its alternative roles, whether it is only acting on local transcriptional repression or it is also regulating higher chromatin structures, perhaps in a histone-independent manner. Intriguingly, it has been shown that PRC1-repressed loci form isolated self-interacting domains of compacted chromatin independently of H2AK119Ub. These domains are different from TADs, and are thought to repress developmental regulators during pluripotency.
Moreover, PRC1 loss upon differentiation causes de-compaction of developmentally regulated loci and induction of cell fate specification (Kundu et al., 2017). This mechanism could be shared with the one we described in this chapter, combining a histone-independent role for RING1B with direct transcriptional repression. Indeed, in some cases, PRC1 domains overlap the boundaries of previously characterised TAD domains (Kundu et al., 2017), known to be enriched for CTCF (Dixon et al., 2012).

To summarise, here we showed cell cycle regulation of the core pluripotency factors OSN. They have been shown to strongly bind chromatin during the S/G2/M, displaying a lower binding in the early cell cycle. The relevance of this process might be in the fact that the chromatin is “freer” from these factors during the G1 phase, therefore able to quickly respond to differentiation cues. OSN in this case might act as factors maintaining the pluripotent state, and their clearing from chromatin in the G1 phase confirms this as a “window of opportunity” for differentiation. However, this phase is not completely free of regulation: CTCF and RING1B might collaborate or have complementary roles in restructuring chromatin after division. However, whereas CTCF is mostly a chromatin remodelling factor, its binding then decreases during cell cycle, whereas RING1B might switch function turning itself into a canonical transcriptional repressor, instructed by the re-formed pluripotency network in the S/G2/M phase. The RING1B activity switch correlates with its lower binding in the intermediate LG1 phase, where indeed developmental genes are temporarily de-repressed.

Most importantly, these findings highlighted an important conceptual point related to transcriptional regulation. That is to say, in order to reveal specific functions for TFs, or binding patterns that will be otherwise masked by a heterogeneous population, it will be important to study transcriptional regulation in a cell cycle context. In a heterogeneous sample of hESCs, for instance, the majority of cells (∼60%) will be representative of the S phase stage, whereas a very little proportion will be representing the early cell cycle, phases that we have shown to be of most interest in terms of chromatin remodelling and transcriptional rearrangements.
Chapter 7: Cell cycle specific perturbation of pluripotency

7 CELL CYCLE SPECIFIC PERTURBATION OF PLURIPOTENCY

In the previous chapters, we have described the dynamic changes that occur during cell cycle progression in chromatin organisation, gene expression, and core pluripotency network. These observations led to a model that explains different cell cycle dependent events previously unlinked. Just after mitosis, chromatin needs to be re-organised in order to maintain cell identity, and CTCF plays a key structural role in this. The plasticity that derives from these rearrangements allows the low expression of several developmental genes in the early cell cycle during the G1 phase, accompanied by a decrease in RING1B occupancy. This makes the cells responsive to differentiation stimuli, defining the G1 phase as a window of opportunity for differentiation. This is reinforced by the absence in the G1 phase of a strong pluripotency network regulated by OSN binding, which have been shown to mostly bind their targets during the S/G2/M phase.

To test our model, we decided to perturb the system in a cell phase specific manner. Our aim was to see whether forcing the binding or withdrawing one of the controlling factors in a specific cell cycle phase was enough to induce loss of pluripotency and, more interestingly, was able to impair differentiation capacities of hESCs. In order to do this, we generated two inducible lines in a FUCCI reporter background, to overexpress and knock-down the TF OCT4. We focussed on this factor since it is generally considered the master regulator of pluripotency, and its
levels have to be tightly regulated to ensure correct differentiation (Nichols et al., 1998; Niwa et al., 2000; Scholer et al., 1989).

Here, we confirmed that overexpression and knock-down of OCT4 have a strong effect on maintenance of pluripotency. Interestingly, we show that a short induction of OCT4 in the early cell cycle is already enough to impair differentiation, and that its withdrawn during the S/G2/M reduces pluripotency. This confirms the role of OCT4 in maintaining the pluripotency network active in the S/G2/M phase and that forcing its presence in the G1 phase can alter the differentiation capacities of such phase.

7.1 Generation and characterisation of iKD and ERT2 OCT4 hESCs FUCCI lines

In order to test our hypotheses, two inducible lines have been generated: a FUCCI H9 OCT4 inducible Knock-Down (OCT4 iKD) and a FUCCI H9 OCT4 inducible ERT2 (OCT4-ERT2).

The iKD system has been previously developed in the lab, and it is based on genetic engineering of a human genomic safe harbours combined with an improved tetracycline-inducible method (Bertero et al., 2016). In this system, the OCT4 shRNA is under the control of a tetracycline inducible promoter (TET-ON system), and the TET-OCT4 shRNA cassette, together with the tetracycline-responsive element, is integrated into the AAVS1 locus, a constantly open chromatin region. Upon tetracycline addiction, the OCT4 shRNA is expressed and thereby leads to the knockdown of OCT4 mRNA (Figure 7.1).

Figure 7.1 OCT4 iKD.

Single step optimised inducible knock-down. AAVS1 locus targeting, with OCT4 shRNA under the control of a tetracycline inducible promoter (Bertero et al., 2016). ZFN: zinc-finger nucleases; 5'-HAR/3'-HAR: upstream/downstream homology arm; H1-TO: Tetracycline-inducible H1 Pol III promoter carrying one tet operon after the TATA box; CAAG: CMV early enhancer, chicken β-actin and rabbit β-globin hybrid promoter; TetR: Tetracycline-sensitive repressor protein. (Adapted from Bertero et al., 2016).
Chapter 7: Cell cycle specific perturbation of pluripotency

H9-FUCCI cells were transfected with this vector, or a scramble control, and the resulting colonies were picked for derivation (3 scramble controls, 8 OCT4). All the clones were genotyped, in order to identify the correct integration events and the absence of off-target events by PCR (see Material and Methods). Two OCT4 iKD clones showed correct integration and absence of random integration events (Figure 7.2).

<table>
<thead>
<tr>
<th></th>
<th>correct integration</th>
<th>off-target</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WT 5’ END</td>
<td>3’ END</td>
</tr>
<tr>
<td>scr1</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>OCT4</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

One of the positive clones (#11) and one scramble (#2) were used to further characterise the system. Upon addition of TET, the OCT4 knock-down clearly induced differentiation, as visible by changes in morphology 72h after induction (Figure 7.3).

Figure 7.2 OCT4 iKD Genotyping.
OCT4 iKD genotyping screening after targeting. Green colour indicates expected results, red indicates unwanted targeting, summarised in bottom the table.

One of the positive clones (#11) and one scramble (#2) were used to further characterise the system. Upon addition of TET, the OCT4 knock-down clearly induced differentiation, as visible by changes in morphology 72h after induction (Figure 7.3).

Figure 7.3 OCT4 iKD morphology.
OCT4 iKD phenotype. FUCCI-H9 morphology after 72h +/- TET in OCT4 iKD.
The efficiency of the knockdown was tested by a qPCR time course, showing a significant decrease in the OCT4 transcript already after 12h (Figure 7.4). At the protein level, OCT4 was very low already at 24h, to then disappear completely at 48h (Figure 7.5). In order to characterise the effect of the knockdown, we performed qPCR analyses on differentiation markers (SOX17 and EOMES for endoderm, T and MESP1 for mesoderm, PAX6 and SOX1 for ectoderm). Interestingly, reduction of OCT4 levels induced downregulation of NANOG and upregulation of the endoderm markers SOX17 and EOMES, together with a reduction of T background levels, and a mild increase, only at day 3 for PAX6 (Figure 7.6). This suggested that knock-down of OCT4 induces mostly endoderm differentiation, in agreement with previous reports (Bertero et al., 2016).

Figure 7.4 OCT4 iKD time course.
OCT4 knock-down qPCR at 0h, 6h, 12h, 24h, 48h, 72h after TET induction (+TET), compared to -TET control, values relative to PBGD (two-ways ANOVA, n=3).

Figure 7.5 OCT4 iKD WB time-course.
OCT4 knock-down WB at 0h, 6h, 12h, 24h, 48h, 72h after TET induction (+TET), compared to -TET control. α-tubulin as loading control.
Figure 7.6 OCT4 iKD induces endoderm differentiation.

(A) OCT4 knock-down qPCR at 0h, 6h, 12h, 24h, 48h, 72h after TET induction (+TET), compared to -TET control, values relative to PBGD for SOX17 and EOMES. (C) T and MESP1. (D) SOX1 and PAX6 (two-ways ANOVA, n=3). (B) Flow-cytometry analyses for in OCT4 iKD for SOX17 at 48h and 72h after induction (SOX17+, red) compared to -TET (SOX17-, blue).
In parallel, another system has been developed in order to generate and inducible-OCT4 overexpressing line (Figure 7.7). The ERT2 system relies on the fusion of OCT4 with a modified fragment of the Estrogen receptor (OCT4-ERT2), which sequesters OCT4 in the cytoplasm, where it is prevented to bind the DNA. In presence of an Estrogen receptor antagonist, such as 4-Hydroxytamoxifen (4OHT), OCT4 rapidly relocates into the nucleus where it can bind the DNA. H9-FUCCI cells have been transfected with a pTP6 vector containing the OCT4-ERT2 cassette, or with a pTP6 empty vector control. Single colonies were picked after selection (3 controls, 9 OCT4-ERT2).

Western blot showed the correct generation of the OCT4-ERT2, with the presence of WT band of 39 kDa and a fusion-protein band of 74 kDa (Figure 7.8). After 72h of 4OHT induction, OCT4-ERT2 hESCs showed severe morphological changes, suggesting differentiation (Figure 7.9). However, it was hard to determine the nature of the differentiated cells, since no clear markers of differentiation was detected by qPCR, apart from downregulation of background T (Figure 7.10). To further characterise the efficiency of the translocation, we looked at the localisation of the OCT4 protein by immunofluorescence in undifferentiated cells upon 4OHT induction (Figure 7.11). Already after 4h induction, OCT4-ERT2 seemed to have relocated in the nucleus. However, considering the high levels of endogenous

Figure 7.7 OCT4-ERT2 system.

The modified fragment of the Estrogen receptor (ERT2) is fused to OCT4, which is retained in the cytoplasm. In presence of an Estrogen receptor antagonist, such as 4-Hydroxytamoxifen (4OHT), the fusion protein binds to it and relocates into the nucleus.
OCT4, the relocation was not so obvious. To circumvent this, we differentiated the cells into neuroectoderm for 6 days, to remove endogenous OCT4 protein, and then added 4OHT for 4h. In these conditions, OCT4 staining was spread over the cytoplasm and, only upon 4OHT induction, was strongly relocated into the nucleus (Figure 7.11).

![Image of Western Blot](image)

**Figure 7.8 OCT4-ERT2 WB.**

OCT4 WB for 2 control lines (lane 1-2) and 4 OCT4 ERT2 lines (lanes 3-6), wild-type OCT4 band at 39 kDa and the OCT4 ERT2 fusion-protein band at 74 kDa. α-tubulin as loading control.

![Image of FUCCI-H9 morphology](image)

**Figure 7.9 OCT4-ERT2 morphology.**

FUCCI-H9 morphology after 72h +/- 4OHT in OCT4-ERT2 line.
Figure 7.10 OCT4-ERT2 qPCR time course.

OCT4-ERT2 qPCR at 0h, 6h, 12h, 24h, 48h, 72h after 4OHT induction (+4OHT), compared to -4OHT control, values relative to PBGD for SOX17, T and PAX6 (two-ways ANOVA, n=3).

Figure 7.11 OCT4 ERT2 translocation.

(A) OCT4 ERT2 line in pluripotency conditions, +4OHT induction shows relocation of cytoplasmic OCT4 ERT fusion protein into the nucleus. (B) Control line (Ctrl ERT2 line) vs OCT4 ERT after 6 days in SB to induce neuroectoderm differentiation: in the control line, endogenous OCT4 is absent compared to pluripotency condition (AF), in OCT4 ERT2 lines cytoplasmic OCT4 is visible upon SB treatment after 6 days, and upon +4OHT addiction relocates into the nucleus. Scale bars 50 μm.
Chapter 7: Cell cycle specific perturbation of pluripotency

Taken together, these results showed the successful generation of two inducible lines to alternatively knockdown or overexpress OCT4. Moreover, we showed that in our culture conditions knock-down of OCT4 induced rapid upregulation of endoderm markers, whereas its overexpression resulted in growth impairment suggesting differentiation.

7.2 Cell phase specific binding of OCT4 regulates pluripotency and endoderm differentiation in hESCs

In order to test whether perturbing OCT4 binding was able alone to block differentiation or alternatively impair self-renewal, we set up two parallel experiments. We decided to investigate OCT4 withdrawal during pluripotency in the S/G2/M phase, when its binding was shown to be maximum, and its overexpression while inducing endoderm differentiation specifically during the G1 phase, when its binding was minimum. Both systems, for simplicity, relied on the use of cell cycle synchronisation by nocodazole, and on the FUCCI colours as a quick read-out. Nocodazole is a small molecule that reversibly interferes with the polymerization of microtubules, arresting the cell in the G2 and M phases of the cell cycle: cells do enter mitosis but cannot form metaphase spindles (Deysson, 1968; Matsui et al., 2012; Zieve et al., 1980). However, after nocodazole withdrawn, cells can re-enter the cell cycle as a synchronous population without any effect on pluripotency or differentiation capacities (Yiangou, 2017 in preparation).

In order to test for OCT4 knock-down, cells were incubated for 20h in presence of nocodazole, to block them in the G2/M phase, and in presence of tetracycline, to knockdown OCT4. According to previous time course, this timing should be enough to reduce OCT4 levels specifically in the S/G2/M phase, since most changes at the protein level happened between 12h and 24h, and by 12h most cells will be already in the S/G2/M phase. Following nocodazole removal, cells have been grown for additional 24h in pluripotency conditions. Interestingly, compared to the asynchronous population, nocodazole treated cells after 24h showed lower levels of OCT4, suggesting that lack of OCT4 in the S/G2/M phase is more detrimental to maintenance of pluripotency (Figure 7.12).
Figure 7.12 OCT4 knock-down impairs pluripotency in the S/G2/M phase.

(A) Schematic of the experiment: cells are incubated for 20h with nocodazole (G2/M phase block) and with tetracycline (+TET), to down-regulate OCT4. After nocodazole and TET removal, cells have been grown for additional 24h in pluripotency conditions. (B) OCT4 iKD qPCR upon nocodazole/TET addiction (-20h), at removal of NOC and TET (0h), after 24h in pluripotency conditions (two-ways ANOVA, n=3).
On the other hand, we wanted to test whether translocation of OCT4 into the nucleus during the G1 phase was detrimental for differentiation. In order to do this, we synchronised the cells for 16h using nocodazole, and then released them in endoderm inducing condition for 24h, in presence or absence of a short pulse (4h) of 4OHT, to induce OCT4 translocation specifically in the G1 phase. Interestingly, despite both conditions show upregulation of endoderm genes, markers such as T and EOMES were significantly lower when OCT4 was induced during the G1 phase (Figure 7.13), suggesting that in this condition differentiation was less efficient.

![Diagram](image)

Figure 7.13 OCT4 overexpression impairs differentiation in the G1 phase.

(A) Schematic of the experiment: cells are incubated for 16h with nocodazole (G2/M phase block). After release, cells are incubated in endoderm inducing conditions (FLyAB) for a short pulse of 4h with 4OHT to induce OCT4 ERT2 translocation specifically in the G1 phase. Cells are collected 24h after endoderm induction. (B) OCT4 ERT2 qPCR for OCT4, T, MIXL1, and EOMES 24h after OCT4 ERT2 induction (+4OHT), values relative to housekeeping gene PBGD (two-ways ANOVA, n=3).

Taken together, these results suggest that binding of OCT4 in specific phases of the cell cycle has important roles in maintaining pluripotency and blocking differentiation. Its withdrawal in the S/G2/M phase reduce pluripotency, whereas its induction in the G1 phase impairs differentiation.
7.3 Discussion

In this chapter, we investigated basic but important mechanistic insights on the effect of cell cycle regulation of pluripotency factors. Here, we confirmed that both overexpression and knock-down of OCT4 have a strong effect on pluripotency in an asynchronous population, as previously reported. However, results in literature are rather controversial, due to a great discrepancy reported by different groups, therefore this shed some light on the effect of perturbing the pluripotency network in our system. Previously published studies are hard to compare mostly because of the differences in the culture conditions: many of them rely on mouse embryonic fibroblast (MEF) monolayers or MEF-conditioned media, together with KSR. Upon OCT4 KD, early reports show upregulation of endoderm markers such as GATA4 and GATA6, together with the induction of CDX2 and other trophoblast-associated markers (Babaie et al., 2007; Hay et al., 2004). However, in some of these reports the induction of trophoblast markers was found to be only induced in condition sub-optimal for self-renewal. Recent reports show instead that low levels of OCT4 induce neuroectoderm differentiation in the absence of BMP4 and specify primitive endoderm in the presence of BMP4 (Wang et al., 2012). Similar discrepancies have been reported for OCT4 overexpression: early reports show induction of differentiation (Rodriguez et al., 2007), whereas recent publications show no effect on self-renewal (Wang et al., 2012). Most of the differences can be addressed to the discrepancies in culture conditions, or in the levels of BMP in the culture (Wang et al., 2012). Specifically, they it has been showed that high levels of OCT4 permit hESC self-renewal in the absence of BMP4 and promote primitive streak (PS) differentiation in the presence of BMP4 (Wang et al., 2012).

Interestingly, we found that OCT knock-down mostly led to the up-regulation of endoderm markers SOX17 and EOMES, and only a mild increase at day 3 for the neuroectoderm marker PAX6. Having shown that OCT4 is mostly repressing neuroectoderm genes during pluripotency, we were expecting a stronger neuroectoderm phenotype. However, in agreement with what previously shown in our lab (Bertero et al., 2016), endoderm induction was stronger, and we hypothesised that longer depletion might increase the levels of neuroectoderm markers as well, considering that previous studies looked at the expression after 5 days of knock-down. On the other hand, OCT4 overexpression also led to an interesting phenotype, characterised by impairment for self-renewal, but no clear differentiation phenotype. It would be interesting to investigate whether OCT4
overexpression might lead to any lineage in the long term, or whether proliferation is impaired due to a cell cycle block perhaps.

Having a system that allows us to knock-down and overexpress OCT4 in a FUCCI background, we decided to test whether perturbing OCT4 binding in a precise cell cycle phase was able to block differentiation or alternatively impair self-renewal. Using this system, we showed that a short induction of OCT4 in the early cell cycle was enough to impair differentiation, and that its withdrawal during the S/G2/M reduced pluripotency. These results were very interesting since they directly linked OCT4 binding with cell cycle regulation. In the first experiment we showed that, after G2/M synchronisation and release, inducing OCT4 nuclear translocation in the early cell cycle just for a short pulse of 4h significantly impaired differentiation. Early endoderm markers such as EOMES and T were significantly lower than the control, indicating that whereas differentiation is not fully blocked, it is severely impaired in the early stages. This is in agreement with our model that describe a chromatin landscape in the G1 phase with a lower OCT4 occupancy, that in turns makes the cells responsive to differentiation stimuli. Inducing OCT4 translocation specifically in this phase can in fact reduce the ability of the cells to respond to differentiation, enforcing a S/G2/M-like transcriptional network, when pluripotency is enforced. This is also confirmed by the fact that, as we showed in a reverse experiment, withdrawing of OCT4 in the S/G2/M results in lower levels of OCT4 compared to the unsynchronised population, suggesting that its knock-down might have stronger effects in the S/G2/M rather than in a heterogeneous population. In this case, a fraction of the cells will be withdrawn of OCT4 in the G1 phase without having a major effect on pluripotency, normally resetting the pluripotency network when they progress through the cell cycle.

However, these experiments are quite preliminary and only represent a proof of principle, as both approaches present some pitfalls. The ERT2 system represent a quick and efficient way to induce nuclear translocation of the fusion protein, however, its binding and the resulting functional effect needs to be validated with additional experiments. In order to do this, a ChIP experiment where OCT4 ERT2 is induced in the EG1 phase should show increased binding, possibly to levels similar to those in the S/G2/M phase. This will be important to link OCT4 binding and differentiation capacities directly, to not base conclusion only on correlations. On the other hand, although the iKD system has been proven to be very efficient (Bertero et al., 2016), it does not have the tight time resolution that the
experiments proposed here should have. With this system, it takes about 12h to significantly reduce OCT4 transcript, and 12h to 24h to remove the protein. For our purpose, a combination of G2/M block with nocodazole and knock-down induction was enough to check for effects in this phase, but to investigate other phases a different approach should be used. A conditional degradation strategy such as the auxin-inducible degron (AID) system (Nishimura et al., 2009) combined with FUCCI sorting, could be a better approach to quickly and reversibly induce degradation of OCT4.

Taken together, these experiments represent a first mechanistic hint into the role of OCT4 binding during cell cycle progression. We describe a model where OCT4 maintains the pluripotency network active in the S/G2/M phase, and its lower occupancy in the G1 is more permissive for differentiation. Indeed, forcing its binding in the G1 phase can alter the differentiation capacities of hESCs, and its depletion in the S/G2/M can impair pluripotency quicker than an unsorted population.
8 CONCLUSIONS AND FUTURE DIRECTIONS

In this dissertation, we have described for the first time the dynamic changes that occur in chromatin organisation, gene expression, and core pluripotency network during cell cycle progression in hESCs. We then further validated these findings with functional experiments perturbing the system in a cell cycle specific manner. Most importantly, we highlighted for the first time the importance of studying transcription factors binding in a cell cycle context. Indeed, in an asynchronous population, the majority of cells show characteristics of their longer phase of the cell cycle, which in the case of pluripotent cells is the S phase. However, our analyses suggest that the S phase is relatively inactive for transcriptional reorganisation while the G1 phase is much more dynamic. Surprisingly, we found that even the core pluripotency factors OCT4, NANOG and SOX2 showed unexpected binding dynamics, establishing a tight and fully formed transcriptional network only in the S/G2/M phase. Moreover, we showed that studying cell phase specific binding could help predicting or dissecting divergent roles for transcription factors and epigenetic modifiers such as RING1B. Thus, we would like to propose a model that complements what is already known in terms of cell cycle dynamics, linking it with cell fate acquisition and maintenance of self-renewal.
8.1 CTCF and chromatin structure during cell cycle progression

In the first part of this dissertation, we have described for the first time chromatin accessibility during cell cycle progression in hESCs, and revealed that extensive rearrangements happen at the beginning of a new cell cycle. It is known that during mitosis, most transcription factors dissociate from chromatin and this might cause loss of higher 3D structures such as compartments. Then the chromosomes stochastically fold, in order to allow proper chromatin compaction, but it is thought that TADs boundaries and enhancers remain marked by bookmarking proteins or persist in a nucleosome-free status. This bookmarking does not affect the stochastic folding of the genome, but ensure the rapid re-establishment of TADs and re-binding of some TFs in the EG1 phase to maintain cell identity (Nagano et al., 2017). We propose that this is regulated by CTCF, that shows strong binding in EG1, indicating that it might be acting as a chromatin remodelling factor only at the beginning of the cell cycle, with a canonical transcriptional role during the rest of the cell cycle. Additional epigenetic modifiers could collaborate, perhaps driving the re-establishment of repressive domains (Kundu et al., 2017) after mitosis. After TADs are established in the EG1, they are functionally locked only in the S phase, when A/B compartments are reconstructed, and the core pluripotency factors can then fully re-establish their binding patterns. In the meanwhile, the tight window of the G1 phase, when the core pluripotency factors are less present on the DNA, represent the optimal time for PSCs to respond to differentiation cues. This results as a combination of different factors: a) TADs are not fully re-established yet in EG1 and changes in inter-TADs interactions that follows differentiation are allowed; b) this permissive chromatin state is also free of repressive binding of the core pluripotency factors OCT4, NANOG and SOX2, showed to fully re-bind their target genes only in the S/G2/M phase; c) transcriptional leakiness for developmental regulators, regulated by higher levels of H3K4me3 mark and low RING1B occupancy, represent an additional favourable environment to respond to differentiation stimuli.

It is important to mention that the model we proposed requires further validations in order to be fully established, and we hope to confirm the proposed functional divergences with further mechanistically experiments. The last chapter represents indeed a starting point to address these questions. The finding that inducing or withdrawing the binding of a single factor in a specific cell cycle phase can de-regulate such system is encouraging, but further experiments should be performed
in that directions. Firstly, it will be important to validate that the function of CTCF in re-arranging chromatin is mostly restricted to the EG1 phase, when its chromatin binding is stronger, and whether this factor is essential for chromatin remodelling. Such architectural role for CTCF has been thoroughly described, and lately, using an auxin-inducible degron system, it has been shown that it instructs chromatin loops and TAD insulation genome wide, without impacting higher order compartmentalization (Nora et al., 2017). This is in agreement with our findings, supporting a model where CTCF, promptly bound to chromatin in the EG1 phase, “scans” the DNA searching for marks to re-establish TADs and insulation. Higher order compartmentalisation is not affected since it has been shown to be CTCF independent, and indeed only happens later in the cell cycle when CTCF genome occupancy is massively reduced. For instance, we expect that knock-down of CTCF in a G1 synchronous population would have a stronger effect compared to an asynchronous population. It has been shown that acute depletion of CTCF can be tolerated for 2-3 days without obvious cell death or differentiation, but depleting for longer slowed cell proliferation dramatically (Nora et al., 2017). The delay in a phenotypical response could be explained by the fact that in an asynchronous population of PSCs, most cells will be in the S phase, when CTCF might be not needed, therefore they have to complete a full cycle and wait until the next division to see an effect. Moreover, additional mechanisms of chromatin remodelling might compensate for CTCF depletion, such as PRC1-associated domains, further delaying its phenotypical effect. It is also important to notice that higher levels of compartmentalisation are not affected by CTCF depletion, therefore not the entire chromatin structure is de-stabilised and PSCs might be able to pass an additional cell cycle before stopping to proliferate. It would be also interesting to investigate the differentiation capacities of CTCF depleted cells: being the drive for chromatin looping its depletion might alter differentiation capacities of PSCs in the G1 phase, even before having an effect on pluripotency, since promoter-enhancer looping necessary for lineage induction might be impossible, or even make cells competent for differentiation in the S phase.

8.2 Epigenetic modifiers integration during cell cycle progression

The interesting observation that CTCF binding strongly overlaps with the epigenetic modifiers RING1B occupancy, but only in the EG1 phase, confirmed the importance of studying TFs binding in the context of the cell cycle, and led us to
the hypothesis that RING1B might have a structural role in the early cell cycle and have a canonical role only after the pluripotency network is re-established. Indeed, after validating the structural role for CTCF, it might be interesting to test how it connects with RING1B occupancy, and whether they collaborate to re-structure chromatin after division. In the last few years, it has been established a clear link between PcG proteins and 3D genome organisation, but these findings almost always describe correlations. Therefore, it will be important to define whether the architectural functions are required, or whether they are a consequence of gene regulatory processes. Our findings suggest that the different roles are temporally distinct during the cell cycle, therefore could be functionally uncoupled. In this scenario, characterising a cell phase specific interactome would be extremely interesting in order to define whether at different functions correspond different canonical/non-canonical PRC1 complexes. This possibility could explain the differential binding patterns, since for example, already among the canonical PRC1 complexes, different CBXs dictate different chromatin recruiting. Moreover, the different composition of PRC1 complexes could explain also the apparently divergent roles: we found that RING1B is not only able to carry out structural functions but also could be used dynamically to change gene regulatory states, both repressing, and surprisingly, activating expression. Several reports have already suggested the possibility for PcG to be involved with gene activation (Frangini et al., 2013; Gao et al., 2014; Kondo et al., 2014; van den Boom et al., 2016), but none of them reports such temporal dissection of roles during cell cycle progression. Having a complete interactome in the different phases could then facilitate the identification of potential cell cycle regulated factors that might drive RING1B differential occupancy and biological function. Ultimately, in order to characterise the impact of CTCF and RING1B on 3D chromatin structure, it would be interesting to perform Hi-C or ChIA-PET on sorted FUCCI hESC, and finally link higher order chromatin with local rearrangements described in this work. This will allow to furtherly characterise the chromatin landscape that makes the G1 the right environment to respond to differentiation cues: the dynamic but local rearrangements that we described in this work are likely to be accompanied by higher-order chromatin reorganisation, as described in (Nagano et al., 2017).

8.3 Core pluripotency factors and cell cycle

The finding that the core pluripotency factors weakly bind the G1 phase of the pluripotent cell cycle reinforces the leading notion in the field that the G1 phase
represent a window of opportunity for differentiation, whereas the S/G2/M phase preserves pluripotency. This led us to hypothesise that differential transcription factors occupancy will play an essential role in determining cell fate propensity. A chromatin landscape free of OCT4, NANOG and SOX2 in the early cell cycle represents a favourable environment for mild de-repression of lineage markers, as we shown by RNA-seq, but most likely allow additional factors, such as SMAD2/3, to freely bind and induce differentiation in presence of the appropriate stimuli, affecting cell fate propensity (Pauklin and Vallier, 2013). An additional but equally interesting consideration could be made in the context of DNA replication. We could speculate that during the S phase, genetic stability must be safeguarded and correct replication of DNA ensured. OCT4, SOX2 and NANOG could be blocking differentiation in this phase in order to preserve genetic stability, avoiding the establishment of epigenetic marks that could compromise cell identity. This could mean that they not only control fate choice but also genomic integrity, a dual role previously proposed for OCT4 (Zhao et al., 2014). In addition to this, whereas OCT4 seemed to increase its binding during cell cycle progression, SOX2 and NANOG occupancy are mutually exclusive during the G1 phase, suggesting that they might block/induce divergent cell fates. Further functional experiments are required in order to validate this hypothesis, where factor could be overexpressed or withdrawn in a cell phase specific manner. Of note, we already reported the feasibility of such experiment by integrating a FUCCI-hESCs line with an inducible system, together with chemical synchronisation by nocodazole. Using this system, we preliminary showed that a short induction of OCT4 in the G1 phase was enough to impair differentiation, and that its knockdown during the S/G2/M reduced pluripotency. These results directly linked OCT4 binding with cell cycle regulation, suggesting that forcing its presence on the chromatin in the G1 could re-create an S/G2/M-like network, when the pluripotent state is harder to dissolve. In agreement with this, withdrawing of OCT4 in the S/G2/M phase causes reduction of pluripotency, indicating that a decrease in OCT4 levels in this phase of the cell cycle makes cells more susceptible to self-renewal disruption. It would be interesting to further investigate this hypothesis by inducing differentiation specifically in the S/G2/M after OCT4 knockdown. Indeed, OCT4 depleted cells should be able to differentiate in any phase of the cell cycle, while wild type cells can only do so in the G1 phase. Ultimately, it will be fundamental to validate this findings in additional cell lines, and we are planning to partially replicate these results in iPSC-FUCCI lines. In the long term, it will also be of high interest to test
whether the transcriptional regulations uncovered here might be also relevant in other cellular contexts such as adult stem cells.

Overall, these mechanisms underline the need to study transcriptional and epigenetic mechanisms in the dynamic context of the cell cycle, and highlight major potential implications for stem cell biology, embryonic development and, ultimately, adult tissue homeostasis.

**Figure 8.1 Schematic summary of the proposed working model.**

In the EG1 phase, immediately after cell division, chromatin is highly dynamic and CTCF binding is proposed to regulate changes in accessibility. This is possibly complemented by a non-canonical role for RING1B, described as a factor able to remodel chromatin. Core pluripotency factors OCT4, SOX2 and NANOG (OSN) show low occupancy in the G1 phase, and this allows hESCs to respond to differentiation cues. In the S/G2/M phase, core pluripotency factors fully re-bind chromatin re-establishing the pluripotency network, sustaining self-renewal and blocking differentiation. In this phase, canonical RING1B/PRC1 complexes ensure correct repression of developmental genes.

### 8.4 Future directions

In this dissertation, we demonstrate that binding of key transcription factors changes during cell cycle progression, and this possibly affects chromatin accessibility and differentiation capacities. This work shows for the first time the importance of studying transcription factors binding in a cell cycle context, but opens a lot of questions that need to be addressed in order to understand the importance of such mechanisms.

The first important point would be to validate the connection between CTCF occupancy and chromatin accessibility. To prove that CTCF binding does not just correlate with chromatin openness, it would be interesting to test the effect of a CTCF knock-down on both gene expression and chromatin organisation. More specifically, we would like to show that the structural function of CTCF is mostly
restricted to the EG1 phase, when it mostly binds chromatin. We expect that knock-down of CTCF in a G1 synchronous population would have a stronger effect compared to an asynchronous population. A quick and effective way to achieve this can be the generation of a FUCCI line combined with an auxin-inducible degron system for CTCF (Nishimura et al., 2009), in order to be able to sort cells and induce degradation in a cell phase specific manner. Alternatively, a more laborious approach could be generating a line to induce the expression of CTCF in a cell cycle specific manner, using CDT1 or GEMININ promoters in a CTCF null background. Following this, it would be important to check whether CTCF depletion corresponds to changes in chromatin accessibility performing ATAC-seq, and possibly Hi-C to look at higher chromatin structures. In parallel, it would be important to investigate the differentiation capacities of CTCF depleted cells: considered the drive for chromatin looping, its depletion might alter differentiation capacities of hESCs.

After validating the proposed structural role for CTCF, it will be interesting to test how it connects with RING1B occupancy, and whether they collaborate to restructure chromatin after division. There is growing evidence that PcG proteins and 3D genome organisation are tightly linked, and here we indeed propose divergent roles for RING1B in both regulating chromatin structure and act as a canonical repressor. This is supported by our RNA-seq data showing that a significant proportion of expressed genes are indeed RING1B bound. However, both these hypotheses need to be validated. The first approach could be to analyse the RING1B interactome in sorted cells: we expect to see enrichment for chromatin remodelers in EG1 and canonical PRC interactors in the S/G2/M. Starting from these results, a cell cycle specific depletion as described above might help dissecting these roles, as S/G2/M reduction should de-repress key developmental regulators, while EG1 should perturb chromatin structures. Insights might be gained also by further analysing the available datasets, in order to better characterise the nature of RING1B bound but expressed genes with bound but repressed ones. Together with this, it would be interesting to see whether the overlap with OCT4 ChIP-seq dataset only occurs in the S/G2/M phase, as proposed, and then further characterise the link between OCT4 and RING1B binding. Indeed, using the system proposed above, we can investigate whether OCT4 depletion specifically in S/G2/M reduces RING1B binding, and more in general further develop the experiments reported in Chapter 7. Indeed, using the degron system described above, it will be possible to have better time resolution
and a more specific depletion of OCT4, NANOG and SOX2 in the different cell cycle phases. This will represent a good validation for the experiments described in Chapter 7, and can be combined with further analyses such as ATAC-seq in order to assess chromatin accessibility after depletion of specific factors. Then we would like to analyse the effects on all the three germ layers, in order to check whether depletion any of the factors impairs or directs the cells specifically towards one germ layers only in EG1 or LG1. Finally, it would be of special interest to repeat some of these experiments in human naïve cells, to understand whether cell cycle dynamics are conserved between the two states and, most importantly, their pluripotent state relies on such dynamical binding.
8.5 Conclusions

According to the University of Cambridge guidelines for the award of a PhD Degree, the candidate must demonstrate that his or her dissertation “represents a significant contribution to learning, for example through the discovery of new knowledge, the connection of previously unrelated facts, the development of new theory, or the revision of older views”. To this end, this dissertation has accomplished such criteria through the following:

- Develop an experimental procedure to test cell phase specific events in hESCs, describing a platform applicable to different downstream analyses.
- Reveal local chromatin dynamics during cell cycle progression and proposing regulatory factors, such as CTCF.
- Uncover a dynamic transcriptional network during cell cycle progression that regulates pluripotency and differentiation capacities in hESCs.
- Demonstrate the existence of a diverse fate propensity in the different cell cycle phases, with a G1 phase able to respond to differentiation cues and an S/G2/M phase keeper of pluripotency.
- Establish cell cycle specific function of the master regulator of pluripotency OCT4.
- Highlight the importance of studying transcription factors binding in a cell cycle context.

Collectively, the results presented in this dissertation advanced both the knowledge and the methodologies available to the field, and will significantly contribute to several exciting future developments in the years to come.
Chapter 9: Appendices

9 APPENDICES
All the tables and output files listed below are included in the annex CD-ROM.

<table>
<thead>
<tr>
<th>File name</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATAC_CTCF_overlap</td>
<td>bed</td>
</tr>
<tr>
<td>ATAC_EG1</td>
<td>bed</td>
</tr>
<tr>
<td>ATAC_EG1_genes</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_EG1vsLG1_diff_down_final_bed37_enrich</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_EG1vsLG1_diff_final</td>
<td>bed</td>
</tr>
<tr>
<td>ATAC_EG1vsLG1_diff_up_final_bed37_enrich</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_EG1vsLG1_up_TRANSFAC_and_JASPAR_PWMs_table</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_G1S</td>
<td>bed</td>
</tr>
<tr>
<td>ATAC_G1S_genes</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_G1SvsSG2M_diff_final</td>
<td>bed</td>
</tr>
<tr>
<td>ATAC_GO_Biological_Process_2017b_EG1</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_GO_Biological_Process_2017b_G1S</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_GO_Biological_Process_2017b_LG1</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_GO_Biological_Process_2017b_SG2M</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_KEGG_2016_EG1</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_LG1</td>
<td>bed</td>
</tr>
<tr>
<td>ATAC_LG1_genes</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_LG1vsG1S_diff_final</td>
<td>bed</td>
</tr>
<tr>
<td>ATAC_PPI_Hub_Proteins_EG1</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_PPI_Hub_Proteins_G1S</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_PPI_Hub_Proteins_LG1</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_PPI_Hub_Proteins_SG2M</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_Reactome_2016_EG1</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_Reactome_2016_G1S</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_Reactome_2016_LG1</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_Reactome_2016_SG2M</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_SG2M</td>
<td>bed</td>
</tr>
<tr>
<td>ATAC_SG2M_genes</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_SG2MvsEG1_diff_down_final_bed37_enrich</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_SG2MvsEG1_diff_final</td>
<td>bed</td>
</tr>
<tr>
<td>ATAC_SG2MvsEG1_diff_up_final_bed37</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_SG2MvsEG1_diff_up_final_bed37_enrich</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_TF-LOF_Expression_from_GEO_G1S</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_TF-LOF_Expression_from_GEO_LG1</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_TF-LOF_Expression_from_GEO_SG2M</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_WELLINGTON_FOOTPRINTS_FINAL_EG1</td>
<td>bed</td>
</tr>
<tr>
<td>ATAC_WELLINGTON_FOOTPRINTS_FINAL_G1S</td>
<td>bed</td>
</tr>
<tr>
<td>ATAC_WELLINGTON_FOOTPRINTS_FINAL_LG1</td>
<td>bed</td>
</tr>
<tr>
<td>ATAC_WELLINGTON_FOOTPRINTS_FINAL_SG2M</td>
<td>bed</td>
</tr>
</tbody>
</table>
Chapter 9: Appendices

<table>
<thead>
<tr>
<th>Dataset Name</th>
<th>File Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATAC_bootstrap_EvL_EG1_only</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_bootstrap_EvL_LG1_only</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_bootstrap_LvS_G1S_only</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_bootstrap_LvS_LG1_only</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_bootstrap_MvE_EG1_only</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_bootstrap_MvE_SG2M_only</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_bootstrap_SvM_G1S_only</td>
<td>txt</td>
</tr>
<tr>
<td>ATAC_bootstrap_SvM_SG2M_only</td>
<td>txt</td>
</tr>
<tr>
<td>CTCF_EG1</td>
<td>bed</td>
</tr>
<tr>
<td>CTCF_EG1_GO_Biological_Process_2017b_CTCF</td>
<td>txt</td>
</tr>
<tr>
<td>CTCF_EG1_TranscriptionFactor_PPIs_CTCF</td>
<td>txt</td>
</tr>
<tr>
<td>CTCF_LG1</td>
<td>bed</td>
</tr>
<tr>
<td>CTCF_LG1_Biological_Process_2017b_CTCF</td>
<td>txt</td>
</tr>
<tr>
<td>CTCF_LG1_ENCODE_and_ChEA_Consensus_TFs_from_ChIP-</td>
<td>txt</td>
</tr>
<tr>
<td>X_CTCF</td>
<td></td>
</tr>
<tr>
<td>CTCF_SG2M</td>
<td>bed</td>
</tr>
<tr>
<td>CTCF_SG2M_Biological_Process_2017b_CTCF</td>
<td>txt</td>
</tr>
<tr>
<td>CTCF_SG2M_ENCODE_TF_ChIP_seq_2015_CTCF</td>
<td>txt</td>
</tr>
<tr>
<td>NANOG_EG1</td>
<td>bed</td>
</tr>
<tr>
<td>NANOG_EG1_WikiPathways_2016_table</td>
<td>txt</td>
</tr>
<tr>
<td>NANOG_LG1</td>
<td>bed</td>
</tr>
<tr>
<td>NANOG_LG1_WikiPathways_2016_table</td>
<td>txt</td>
</tr>
<tr>
<td>NANOG_SG2M</td>
<td>bed</td>
</tr>
<tr>
<td>NANOG_SG2M_WikiPathways_2016_table</td>
<td>txt</td>
</tr>
<tr>
<td>OCT4_EG1</td>
<td>bed</td>
</tr>
<tr>
<td>OCT4_EG1_WikiPathways_2016_table</td>
<td>txt</td>
</tr>
<tr>
<td>OCT4_LG1</td>
<td>bed</td>
</tr>
<tr>
<td>OCT4_LG1_WikiPathways_2016_table</td>
<td>txt</td>
</tr>
<tr>
<td>OCT4_SG2M</td>
<td>bed</td>
</tr>
<tr>
<td>OCT4_SG2M_WikiPathways_2016_table</td>
<td>txt</td>
</tr>
<tr>
<td>OSN_EG1</td>
<td>bed</td>
</tr>
<tr>
<td>OSN_EG1_not_expressed_GO_Biological_Process_2017b_table</td>
<td>txt</td>
</tr>
<tr>
<td>OSN_LG1</td>
<td>bed</td>
</tr>
<tr>
<td>OSN_LG1_expressed_WikiPathways_2016_table</td>
<td>txt</td>
</tr>
<tr>
<td>OSN_LG1_not_expressed_GO_Biological_Process_2017b_table</td>
<td>txt</td>
</tr>
<tr>
<td>OSN_N_only_GO_Biological_Process_2017b_table</td>
<td>txt</td>
</tr>
<tr>
<td>OSN_O_only_GO_Biological_Process_2017b_table</td>
<td>txt</td>
</tr>
<tr>
<td>OSN_LG1_all_expressed_WikiPathways_2016_table</td>
<td>txt</td>
</tr>
<tr>
<td>OSN_N_only_GO_Biological_Process_2017b_table</td>
<td>txt</td>
</tr>
<tr>
<td>OSN_SG2M</td>
<td>bed</td>
</tr>
<tr>
<td>OSN_SG2M_expressed_WikiPathways_2016_table</td>
<td>txt</td>
</tr>
<tr>
<td>OSN_SG2M_not_expressed_GO_Biological_Process_2017b_table</td>
<td>txt</td>
</tr>
<tr>
<td>OSN_all_expressed_WikiPathways_2016_table</td>
<td>txt</td>
</tr>
<tr>
<td>OSN_all_not_expressed_GO_Biological_Process_2017b_table</td>
<td>txt</td>
</tr>
<tr>
<td>OSN_shared_KEGG_2016_table</td>
<td>txt</td>
</tr>
</tbody>
</table>
OCT4 sOPTiKD PLASMID MAP
Chapter 9: Appendices

PTP6 OCT4 ERT2 PLASMID MAP
10 REFERENCES


Chia, N.Y., Chan, Y.S., Feng, B., Lu, X., Orlov, Y.L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M.S., Huss, M., Soh, B.S., Kraus, P., Li, P., Lufkin, T., Lim, B.,


receptor-mediated signalling through Smad2, Smad3 and Smad4. EMBO J 16, 5353-5362.


