Department of Gene Regulation, Stem Cells and Cancer
Centre for Genomic Regulation

Doctoral Thesis 2019

Universitat Pompeu Fabra

IDENTIFICATION AND CHARACTERIZATION OF THE MOLECULAR PATHWAYS REGULATING THE CELL CYCLE-LINKED PLURIPOTENCY EXIT

Dissertation presented by

Anna Alcaine Colet

For the degree of Doctor in Biomedical Research

Work carried out under the supervision of Dr. Sergi Aranda and Dr. Luciano Di Croce in the Epigenetic Events in Cancer Group in the Gene Regulation, Stem Cells and Cancer Program, at the Centre for Genomic Regulation (CRG)

Abstract

The self-renewing nature of embryonic stem cells (ESCs) is a consequence of their ability to proliferate while maintaining pluripotency, by means the capacity to differentiate into all adult cell types. Despite the remarkable scientific advances in the knowledge regarding the mechanism(s) controlling pluripotency, little is known about how cell cycle is coordinated with self-renewal and differentiation.

Here we have shown a segregation of pluripotency during cell cycle progression, being cells in S-G2/M more primed for cardiac lineage commitment. By characterizing the proteome dynamics during cell cycle, we have identified the DNA demethylation enzyme thymine DNA glycosylase (TDG) as the molecular determinant for cell cycle pluripotent dissociation. Our TDG ChIP-seq data suggests a functional link between TDG and TP53 by which TP53 transcriptional programs would prime S-G2/M cells for cardiomyocyte differentiation.

With this study, we have enlightened the mechanisms underlying mouse ESC fate decisions control during cell cycle.

Resum

La naturalesa d'autorenovar-se de les cèl·lules mare és una conseqüència de la seva capacitat de proliferar mantenint la seva pluripotència, essent aquesta la capacitat de diferenciar-se en tots els tipus cel·lulars trobats en l'organisme adult. Tot i els remarcables avenços científics en el coneixement dels mecanismes que controlen la pluripotència cel·lular, es desconeix com el cicle cel·lular està coordinat amb l'autorenovació i la diferenciació de les cèl·lules mare.

Aquí hem pogut mostrar que la pluripotència cel·lular és segregada durant el cicle cel·lular, essent les cèl·lules en fase S-G2/M més llestes per l'inici de la diferenciació cap al llinatge cel·lular cardiac. Caracteritzant el dinamisme proteomic Durant el cicle cel·lular, hem identificat la proteïna DNA demethylation enzyme thymine DNA glycosylase (TDG) com el determinant molecular per la dissociació de la pluripotència durant en el cicle cel·lular. La información obtinguda per ChIP-seq de TDG ens suggereix una enllaç functional entre TDG I TP53, a través del qual els programes de transcripció de TP53 prepararíen les cèl·lules en fase S-G2/M per la diferenciació cap a cèl·lules cardíaques.

Amb aquest estudi hem aportat nous coneixements sobre els mecanismes que controlen les decisions sobre el destí cel·lular de les cèl·lules mare de ratolí durant el cicle cel·lular.

Table of contents

INTRODUCTION

1. Mouse embryonic stem cells (mESCs)

1.1 Development and mESCs origin

The concept of mouse pluripotent cells with embryonic origin first appeared with the discovery and characterization of embryonal carcinoma cells (EC cells), originated from teratocarcinomas. Teratocarcinomas are malign tumours comprising differentiated cells from the three germ layers as well as undifferentiated EC cells. EC cells were shown to proliferate indefinitely while being able to form all cells found in the original teratocarcinoma, after being injected into mice (Kleinsmith and Pierce, 1964). Technical advances enabled to stablish EC cell lines *in vitro* for further characterization and the specific teratocarcinoma stem cell identification (Martin and Evans, 1974). These EC stem cells were described to have the power to form embryoid bodies (Martin and Evans, 1975), a three-dimension cell aggregates containing different cell types resembling early embryos.

Soon after, the first attempts to derive pluripotent cells directly from mouse embryos arrived. The first mouse embryonic stem cell (mESC) lines were derived in 1981 from *in vitro* cultured blastocysts embryos on fibroblasts feeder layers (Evans and Kaufman, 1981; Martin, 1981).

Blastocyst embryos are reached at E3.5 stage (3.5 days post conception) of mouse development (**Figure I.1**). After fertilization occurs, several rounds of cell division preceded the morula generation at 16-cell stage and its following cellular compaction. Later, at E3.0 stage, cavitation starts and with it the formation of the blastocoel, a fluid-filled cavity. This process finishes with the origin of a 32 cell embryo stage called blastocyst. In the blastocyst, two different structures are found: the inner cell mass (ICM) and the outer cell layer (**Figure I.1**). At this stage, the first cell fate decision is taken during development: cells in the ICM, where unspecified pluripotent cells can be found, and trophectoderm (TE) cells which will arise from the external layer of the early blastocyst embryo. No much time has to pass until the second lineage specification event occurs, as at late blastocysts stage ICM cells will decide between becoming epiblast (EPI) cells or primitive endoderm (PrE) cells. While TE and PrE cells will result in extra-embryonic tissues (placenta, and yolk sack, allantois and amnion respectively), EPI cells (from where ESCs can be derived from (Evans and Kaufman, 1981; Martin, 1981)) will give rise to all the cells of the future embryo (Brook and Gardner, 1997). Each of the three different cell types shows specific transcription programs. For instance, TE cells are characterized by the expression of the transcription factor caudal type homeobox 2 (CDX-2) (Beck et al., 1995), PrE cells by the expression of Gata-binding proteins 6 and 4 (GATA-6 and GATA-4) (Chazaud et al., 2006; Rossant et al., 2003) and

EPI cells by NANOG expression (Chambers et al., 2003) (**Figure I.1**).

At around E6.5 stage, after the embryo is implanted into the uterus, gastrulation begins. During this process, pluripotent EPI cells give rise to the primitive streak and subsequent cell differentiation occurs. First, mesendoderm cells are formed, which are the precursors for mesodermal and endodermal cells. Later, ectodermal lineage cells are established. Thus, at the end of the gastrulation process, cells from the three germ layers endoderm, mesoderm and ectoderm are found in the embryo. These cells will derive to all cell types in the organism. Moreover, primordial germ cells are also produced during gastrulation. Primordial germ cells will eventually start the migration to the gonads and will give rise to germ cells (oocytes and spermatozoa).

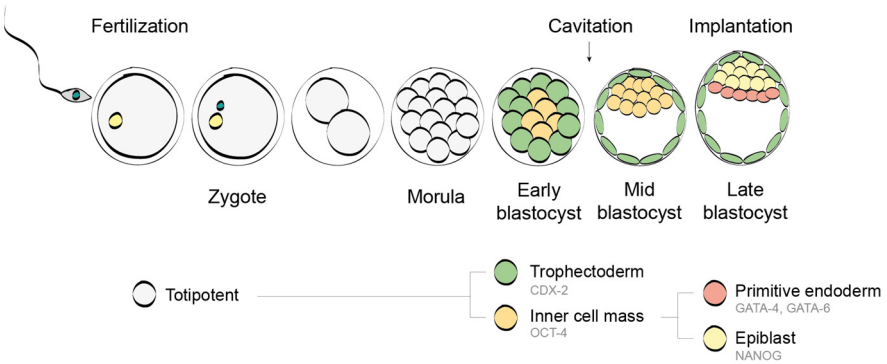


Figure I.1. Early mouse developmental stages. Fertilization precedes zygote formation. After several rounds of cell division, the morula is formed. Later, early blastocyst formation anticipates the cavitation process and the blastocoel formation, where the inner cell mass (ICM) is found. Then, cells from the ICM will develop either to primitive endoderm or epiblast cells and implantation will happen. Cells follow the lineage color-code indicated at the bottom. Specific lineage markers are indicated in grey.

1.2 Characteristics of mESCs and potential applications

mES cell lines are stablished *in vitro* by culturing dissociated morulae (Eistetter, 1989), intact E3.5 blastocyst embryos (Evans and Kaufman, 1981) or dissected ICM (Martin, 1981) on mitotically inactivated fibroblasts feeder layers. Alternatively, mESCs can also be derived from EPI cells from E5.5-E7.5 post-implantation embryos (Brons et al., 2007; Tesar et al., 2007). ESCs are defined by two main characteristics: the pluripotency and the self-renewal capacities (**Figure I.2**). While the pluripotency is the ability of ESCs to differentiate to all cell types found in the proper embryo (including the three germ layers endoderm, mesoderm and ectoderm), self-renewal capacity is defined by the ability to grow indefinitely while maintaining their undifferentiated state.

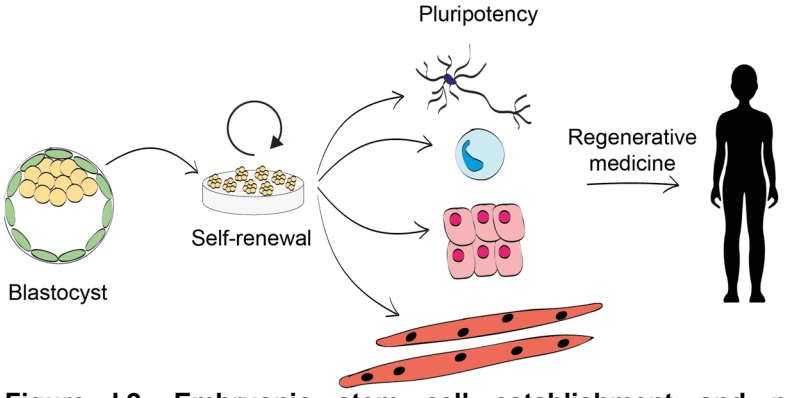


Figure I.2. Embryonic stem cell establishment and potential applications. ESCs can be derived from blastocyst embryos. They possess two main characteristics: self-renewal and pluripotency. Differentiation of pluripotent cells *in vitro* to virtually all cell types make ESCs an attractive model for regeneration medicine purposes.

Prove of concept *in vivo* experiments for pluripotency include teratoma and chimera formation.

Teratomas are benign tumours containing differentiated cells from the three germ layers. Their formation can be promoted by ESCs transplantation into immunodeficient mice. ESCs pluripotency is then assessed by analyzing the cell contribution to endoderm, mesoderm and ectoderm lineages.

Similarly, chimera formation assay allows the study of the pluripotency capacity of ESCs by injecting ESCs into the ICM of a blastocyst (or a developmentally earlier embryo) (Tam and Rossant, 2003). The formation of a full organism with original ESCs contribution is expected to be obtained. A proper pluripotent ESCs injected into a blastocyst not only has to give rise to a chimera embryo, but also to contribute to the germline.

Alternatively, pluripotency can be assessed *in vitro* by embryoid body (EB) formation, where endodermal, mesodermal and ectodermal cells are generated from ESCs. EBs are three-dimension aggregates of pluripotent cells grown in suspension induced to differentiate spontaneously by removing pluripotency-maintaining factors from the culture medium.

Concerning self-renewal, in order to avoid genetic anomalies due to their rapid proliferation rate, ESCs possess a high telomerase activity, which has been described to be linked with pluripotent transcriptomic programs. As an example, telomerase reverse transcriptase (*Tert*) expression, which prevents shortening of telomeres, has been described to be

regulated by the pluripotency transcription factor KLF4 (Wong et al., 2010). This telomerase activity enhancement is not only true for *in vitro* cultured mESCs, but also for *in vivo* (Bekaert et al., 2004).

Pluripotent cells need to maintain their genomic stability, as they will give rise to all cells in the adult organisms. Therefore, mutations in pluripotent cells would be inherited in differentiated cells, potentially leading to malignant transformations. Mutation rates have been shown to be 100 fold less frequent in ESCs compared to somatic cells (Hong et al., 2007). A complementary mechanism for ESCs to avoid the transmission of mutations is through the action of TP53. As mESCs lack a proper cell cycle checkpoint, in the presence of genomic instability, TP53 acts promoting differentiation. By doing so, cells will thus enter apoptosis as differentiated cells restore an intact checkpoint and thus genomic instability will not be transmitted to the cellular progeny (Lin et al., 2005).

ESCs can be kept in culture endlessly as well as to differentiate them into virtually any cell type. The special nature and features of ESCs make them an attractive model not only to study early developmental events, but also to be used as a tool to study gene function via transgenesis by chimera generation. Additionally, ESCs allow for *in vitro* differentiation studies and therefore are an interesting system for regenerative medicine and gene therapy purposes as well (**Figure I.2**).

Up to date, the only stem cell-based therapies applied to humans are matched donor treatments. The most established and commonly used one is bone marrow transplantation, which consist on replacing the patient's bone marrow by a bone marrow from a healthy donor, including hematopoietic stem cells. However, there are several studies that have used hESCs with the aim to treat diseases (see data at www.clinicaltrials.gov), such as cardiovascular diseases, diabetes, retina or liver injuries (Mahla, 2016; Volarevic et al., 2018; Shufaro and Reubinoff, 2004; Song et al., 2015). Thus, hESCs are differentiated *in vitro* and then differentiated cells are transplanted to non-human organisms (Shiba et al., 2012; Fernandes et al., 2015; Tolosa et al., 2015). Some of the ongoing studies are in clinical trials, involving differentiated cells transplantation into patients (Schwartz et al., 2012). However, several drawbacks on the use of hESCs to treat diseases limit their application. For example, *in vitro* protocols for cell differentiation produce a mix population of cells, lacking thus the purity needed for transplantation therapies. On the other side, the direct use of hESCs is controversial for ethical issues as early human embryos are their source of obtention. Moreover, clinical applications of hESCs entails safety issues, as non-desired differentiated cell types and tissues may be produced in patients (Volarevic et al., 2018). Improved *in vitro* differentiation protocols, further characterization on the effect transplantations have on animal disease models and control on hESCs growth are some of the aspects that should be

addressed in the future for the use of hESCs for therapy applications (Reubinoff, 2004).

1.3 Gene expression program governing mESCs nature

Intrinsic and extrinsic signals contribute on keeping the undifferentiated state of mESCs. Intrinsic elements include master transcription pluripotency factors; while extrinsic ones, signaling pathways.

1.3.1 ESCs transcription factors

Fundamental transcription factors govern mESCs identity, keeping its pluripotency capacity. OCT-4 (also known as POU5F1) and NANOG were two of the first identified intrinsic transcriptional determinants in mESCs and early development in the ICM cells (Chambers et al., 2003; Chambers and Smith, 2004; Mitsui et al., 2003; Nichols et al., 1998). Consistently, *Oct-4* or *Nanog* disruption in mouse embryos fail to form a proper ICM, as no pluripotent cells are formed (Nichols et al., 1998; Mitsui et al., 2003). Instead, TE cell differentiation occurs in *Oct-4* mutated embryos (Nichols et al., 1998), while extraembryonic endoderm is formed in *Nanog* disrupted embryos (Mitsui et al., 2003).

Although constitutively *Nanog* expression in mESCs can keep their self-renewal capacity without any other external stimulation (Chambers et al., 2003), *Nanog* deletion does not cause direct mESCs differentiation (Chambers et al., 2007). This suggest that NANOG has a central role in keeping the pluripotent network activated and, therefore, the pluripotent state. However, its loss is not sufficient to induce differentiation.

Sex Determining Region Y-Box 2 (SOX-2) is another master regulator of pluripotency in mESCs, and its disruption causes embryonic defects due to improper ICM formation as well (Avilion et al., 2003). SOX-2 indirectly regulates *Oct-4* expression (through other transcription factors), acting as a pluripotency stabilizer (Masui et al., 2007). In fact, upon *Sox2* deletion, mESCs loose pluripotency and undergo TE-like cell differentiation (Masui et al., 2007).
OCT-4 and SOX-2 directly interact as well, forming heterodimers that work activating gene transcription (Loh et al., 2006). The formation of the heterodimer enhance the transcription factor activity, acting with more affinity and specificity in binding their target genes as heterodimers than as single homodimers (Loh et al., 2006).

In mESCs, chromatin immunoprecipitation followed by ultra-sequencing (ChIP-seq) analysis revealed OCT-4, NANOG and SOX-2 to co-occupy common target genes (Chen et al., 2008;

Loh et al., 2006). OCT-4, NANOG and SOX-2 work together activating the expression of other pluripotent regulators and silencing differentiation genes (Chen et al., 2008; Loh et al., 2006). Not only this, but they also bind to their own promoters, regulating then their own expression. (**Figure I.3**). These findings showed how the coordinated action of the core pluripotency regulators maintains the identity of mESC.

In fact, the pluripotent circuitry in mESCs comprises many other transcription factors, such as Estrogen Related Receptor Beta (ESRRB), Kruppel Like Factor 4 (KLF-4), Kruppel Like Factor 2 (KLF-2) or Transcription Factor CP2 Like 1 (TFCP2L1) (Li et al., 2017, 2018). This network includes direct interaction between transcription factors, which working in cooperation with the core factors ensures the cell identity of mESCs.

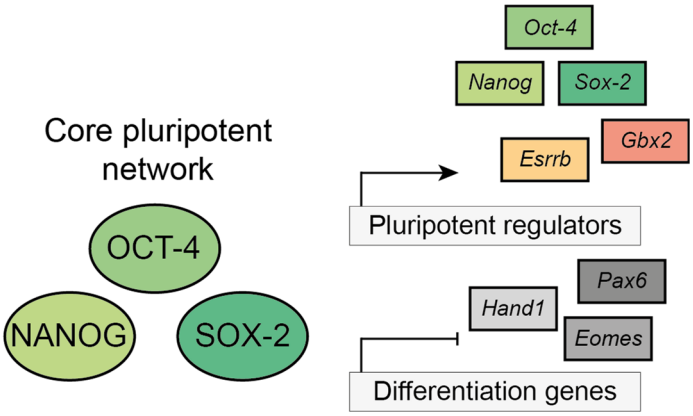


Figure I.3. The pluripotency network maintains mESCs identity. OCT-4, NANOG and SOX-2 form the core pluripotent network, which works regulating gene transcription, activating other pluripotency factors (such as *Esrrb* or *Gbx2*) as well as silencing differentiation genes (such as *Pax6* or

Eomes) (Loh et al., 2006; Jaenisch and Young, 2008). Thus, this pluripotent network works maintaining mESCs identity. Ovals represent transcription factors, while transcripts are represented by rectangles.

1.3.2 Signaling pathways ruling pluripotency in mESCs

LIF and JAK/STAT3 signaling pathway

To ensure the maintenance of pluripotency *in vitro*, mESCs were initially co-cultured with mitotically inactivated fibroblast feeders (Evans and Kaufman, 1981). Later, the cytokine Leukemia Inhibitory Factor (LIF) secreted by feeder cells was discovered and this permitted to grow mESCs on feeder free conditions (Smith et al., 1988; Williams et al., 1988). *In vitro* cultured mESCs are then kept in an undifferentiated state by growing them in serum medium supplemented with LIF in the absence of fibroblast feeders. LIF withdrawal follows ESCs differentiation (Niwa et al., 1998).

LIF is a member of the IL-6 cytokines family and binds to its receptor LIFR and the IL-6 family co-receptor subunit glycoprotein gp130. This binding leads to the LIFR-gp130 heterodimer formation and subsequent activation of the JAK (Janus Kinase) signaling (Lütticken et al., 1994; Stahl et al., 1994), which precedes signal transducer and activator of transcription 3 (STAT3) recruitment and following activation trough phosphorylation (Boeuf et al., 1997; Zhong et al., 1994; Niwa et al., 1998). Finally, STAT3 activation initiates its homodimerization and the successive nucleus translocation

and transcriptional activation of target genes (Sasse et al., 1997; Huang et al., 2015), some of which are well known pluripotency factors such as *Klf-4* or *Tcfcp2l1* (Hall et al., 2009; Martello et al., 2013; Ye et al., 2013) (**Figure I.4**, left). While KLF-4 is one of the well-known Yamanaka factors able to reprogram somatic to pluripotent cells (Takahashi and Yamanaka, 2006), TCFCP2L1 is found being part of the pluripotency network and its knockdown is the only one that fails to keep the undifferentiated state of mESCs (Martello et al., 2013).

This opens a link between external signals (JAK/STAT3 signaling through LIF) and internal signals (pluripotency factors), pointing out to a coordinated activity between both to keep mESCs identity.

Although the STAT3 pathway is not the only LIF-induced signaling pathway (Burdon et al., 1999), LIF-STA3 pathway is the responsible for sustaining ESC self-renewal and pluripotency. In agreement, constitutive JAK-STAT activation in the absence of LIF is sufficient to maintain the self-renewal capacity of ESCs (Matsuda et al., 1999; Niwa et al., 1998). In line with this, STAT3F mutation, which causes STAT3 dimerization and nuclear translocation abrogation, blocks self-renewal and pluripotency in ESCs even in the presence of LIF and, therefore leads to ESCs differentiation (Niwa et al., 1998). STAT3 activation output seems to be dose dependent, as poor STAT3 follows ESC differentiation, and excessive activation

also leads to lineage specification, in this case specifically towards TE (Tai et al., 2014).

Wnt/β-catenin signaling pathway

Another pathway regulating pluripotency and development is the Wnt/β-catenin signaling pathway. The canonical Wnt signaling pathway, which is the most studied Wnt pathway, relies its activity on the amount of Cadherin-associated protein beta 1 (β-catenin) present in the cytoplasm. β-catenin is a transcriptional activator and regulates the expression of lineage/developmental genes. In the absence of the WNT ligand, β-catenin is phosphorylated by the Axin degradation complex, which is formed by casein kinase 1 (CK1) and glycogen synthase kinase 3 (GSK3) among other proteins. Once phosphorylated, β-catenin is recognized and subsequently degraded by the proteasome system. T cell factor/Lymphoid enhancer factor (TCF/LEF) family of transcription factors suppresses the Wnt signaling through association with transcriptional co-repressors. Therefore, when WNT is not present, β-catenin levels are kept down and the pathway is shut-off.

However, when a WNT ligand is present and binds to its receptor FRIZZLED (Fz) and its co-activator low-density lipoprotein receptor related protein 6 (LRP6), the Axin complex is recruited to the receptors. Consequently, the degradation complex cannot target β-catenin and it is then stabilized, leading to the activation of the Wnt pathway and therefore to

β-catenin gene activation in association with TCF/LEF (**Figure I.4**, middle).

Wnt pathway is needed for proper mouse embryo development as *β-catenin* knockout causes gastrulation defects (Haegel et al., 1995). Furthermore, *in vitro* cultured mESC depleted of *β-catenin* revealed a cell adhesion defect shown on the cell colony formation capacity (Haegel et al., 1995). Yet, mESCs devoided of β-catenin are able to differentiate towards endoderm, mesoderm and ectoderm tissues when injected into a blastocysts embryo (Huelsken et al., 2000). Thus showing β-catenin is not totally required for pluripotency and self-renewal, although its lack causes some abnormalities in mESCs.

FGF/ERK signaling pathway

Pluripotent self-renewal is also affected by the FGF/ERK signaling pathway. FGF signaling is activated by the FGF receptor FGFR-ligand interaction, which results on the FGFR auto-phosphorylation and subsequent activation of the pathway (together with other pathways, such as PI3K-AKT).

FGF signaling controls early embryo development, by regulating the initial differentiation events. Mouse embryos depleted of *Fgf4*, an important activator of the pathway produced by pluripotent cells, do not show any defect until implantation occurs, causing the embryo's death (Feldman et al., 1995). Supporting this, *in vitro* cultured mESCs, do not show any abnormality when *Fgf4* is disrupted (Wilder et al.,

1997). Yet, defects start to appear when differentiation is triggered (Wilder et al., 1997). mESCs depleted from *Fgf4* fail on differentiating towards mesoderm and ectoderm lineages (Kunath et al., 2007; Stavridis et al., 2007). Therefore, FGF signaling pathway it is not required for pluripotency and self-renewal capacities maintenance but it is key for proper differentiation. When ESCs pluripotency is lost, cells differentiate irreparably. And, as mentioned, ERK1/2 signaling activation through FGF4 leads to cell differentiation. Hence, inhibition of the ERK1/2 pathway creates a differentiation blockage and thus maintains the pluripotency state (**Figure I.4**, right).

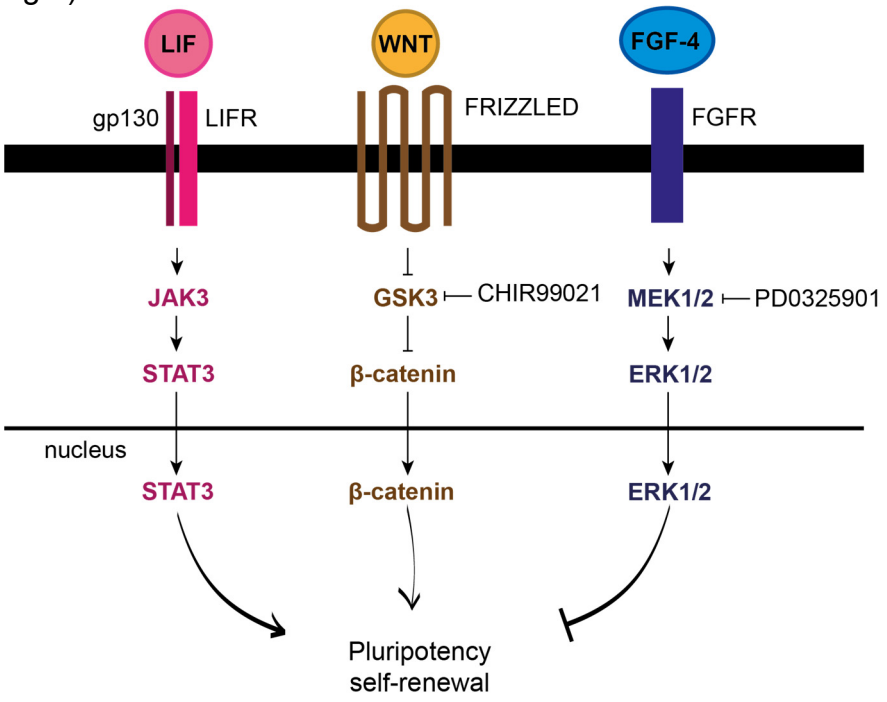


Figure I.4. Signaling pathways regulating pluripotency. LIF/JAK/STAT3, WNT/β-catenin and FGF/ERK are three of the signaling networks that keep the pluripotent state in ESCs. Their coordinated activity regulate gene expression, maintaining the pluripotency factors

transcription. PD and CHIR refer to PD0325901 (an ERK pathway inhibitor), and CHIR99021 (a GSK3 inhibitor), respectively.

1.4 The pluripotent states in mESCs

Once the signaling pathways governing the pluripotent state were defined, improving the *in vitro* cell culture conditions became a target. The use of feeder cells or serum supplemented mediums make the pluripotent self-renewal of mESCs a complex multi-factorial dependent state.

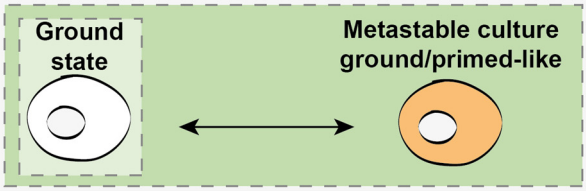
Therefore, in order to establish a serum-free medium, Ying and collaborators make use of two inhibitors for two of the signaling pathways supporting differentiation in ESCs: PD0325901, an inhibitor for the ERK pathway, and CHIR99021, a GSK3 (from the Wnt signaling pathway) inhibitor (Ying et al., 2008) (**Figure I.4**, middle and right). By blocking two differentiation-promoting elements, these two inhibitors act supporting self-renewal and pluripotency. Moreover, CHIR99021 alone enhance cell survival and therefore compensates for the apoptotic effects seen by blocking the ERK pathway alone (Ying et al., 2008). Even after LIF withdrawal, mESCs were able to be grown keeping its pluripotent self-renewal in the presence of PD0325901 and CHIR99021 (Ying et al., 2008). PD0325901 and CHIR99021 inhibitors were added to the serum-free N2B27 medium, constituting the so-called 2i/LIF (Ying et al., 2008).

Two different pluripotency states exist *in vivo*: i) the naïve state, which corresponds to the pluripotent cells from pre-implantation blastocyst embryos and, ii) the primed state, which corresponds to the pluripotent cells found in post-implantation embryos (Nichols and Smith, 2009). Naïve state cells can be expanded *in vitro* in two different culture conditions: 2i/LIF and serum/LIF. However, serum/LIF grown cells show a metastable culture, transiting from a naïve state to a prime-like state (Abranches et al., 2013; Toyooka et al., 2008).

Cells grown in 2i/LIF medium show different morphological, transcriptional, epigenetic and functional features compared to serum/LIF-grown cultured cells (**Figure I.5**). Morphologically, naïve cells grown in 2iLIF conditions show more homogeneous dome-shaped colonies, compared to ESCs grown in serum/LIF, which show more flat colonies. At the level of the pluripotency factors expression such as Stella, Pecam1, NANOG and SSEA1, metastable cells show more fluctuations, with mix population of positive and negative cells for such markers (Chambers et al., 2007; Hayashi et al., 2008). On the contrary, mESCs grown in 2iLIF express pluripotency factors, such as *Nanog*, more uniformly (Chambers et al., 2007; Hayashi et al., 2008; Wray et al., 2011; Ying et al., 2008). Transcriptomic comparative analysis of 2i/LIF and serum/LIF cultured mESCs revealed developmental genes as the serum/LIF most enriched category (Marks et al., 2012). Moreover, novel state specific genes were identified *in vitro*

and *in vivo* by transcriptome analysis, such as *Spic* for ICM and 2i/LIF grown cells, and *Fst* for EpiSCs and serum/LIF grown cells (Bernardo et al., 2018). Cells grown in 2i/LIF are characterized by a global genomic hypomethylation (Leitch et al., 2013) and reduced tri-methylation of lysine 27 on histone H3 (H3K27me3) marked bivalent promoters, shown by chromatin immunoprecipitation and deep sequencing (ChIP-seq), compared to serum/LIF grown cells (Marks et al., 2012). However, recent proteomic and ChIP-seq analysis revealed an increase on the total H3K27me3 abundance in 2i/LIF compared to serum/LIF cultured cells, being the 2i/LIF-specific peaks randomly allocated in the genome (Mierlo et al., 2019). H3K27me3 as well as tri-methylation of lysine 4 on histone H3 (H3K4me3) correspond to generally repressive and active post-transcriptional histone modifications, respectively. Promoters co-occupied by these two marks are called bivalent and are thought to act by priming developmental genes in ESCs for rapid and proper later activation during differentiation (Bernstein et al., 2006).

These described changes in transcription and epigenetics between 2i/LIF and serum/LIF grown cells are also accompanied by functional differences, including the proliferation speed rate and the cell cycle control (mentioned below in chapter 3).



Culture conditions	2i/LIF	serum/LIF
Colony morphology	Dome-shaped	More flat
Pluripotent markers	Homogeneous	Heterogeneous
Differentiation markers	Low	High
DNA hypomethylation	Yes	No
H3K27me3 levels	Low	High

Figure I.5. Pluripotent cell properties for naïve and metastable ground-primed mESCs. ESCs cultured in 2i/LIF or in serum supplemented with LIF (serum/LIF) show different biological features, which are depicted in the first column. The second and the third column correspond to ground and the metastable cell culture comprising ground and primed-like states respectively. H3K27me3 levels referrer to the levels of the mark found on bivalent promoters.

Considering the above differences between 2i/LIF and serum/LIF cultured cells, it is therefore essential to take into account the cell culture conditions in which mESCs are maintained *in vitro* when taking conclusions about their pluripotent identity.

1.5 Mouse and human embryonic stem cells: similarities and differences

Human ESCs (hESCs) were first isolated from the ICM of blastocysts embryos in 1998 by Thomson and co-workers (Thomson et al., 1998). Soon it became clear that hESCs had different characteristics compared to mESCs. Starting from its

cell culture *in vitro*, hESCs show different requirements to mESCs, as they do not need LIF to be kept in their undifferentiated state and do need for FGF2 and Activin A signaling activation instead (Thomson et al., 1998; Bertero et al., 2015; Brons et al., 2007; Humphrey et al., 2004; Daheron et al., 2004). So hESCs keep their pluripotent state through different signaling pathways compared to mESCs. In fact, not only hESCs require FGF2 and Activin A signaling, but they are also necessary for pluripotency maintenance in mouse EpiSCs (Brons et al., 2007). Not only this, but hESCs also share their transcriptomic profile and the transcriptional networks driving pluripotency with mouse EpiSCs. For instance, hESCs OCT-4 targets show a higher overlap with EpiSCs than with mESC (Tesar et al., 2007). Thus, cultured hESCs resembles more the primed EpiSCs rather than the naïve ESCs found in the mouse ICM, suggesting a similar pluripotent state between hESCs and mEpiSCs.

2. Chromatin organization in mESCs

2.1 Global chromatin structure in mESCs

Eukaryotic genome is packed into the cell nucleus in a complex named chromatin composed of DNA, proteins and RNA. The nucleosome is the basic chromatin unit, being formed by 147 DNA base pairs wrapped around a histone octamer formed by the four core histones H3, H4, H2A and H2B. Chromatin is key not only for genome packaging, but also to regulate gene transcription, DNA replication and DNA repair processes.

Chromatin can be divided into two states: heterochromatin and euchromatin. While heterochromatin refers to a highly condensed chromatin, euchromatin is attributed to a more open chromatin. This openness goes hand in hand with high chromatin accessibility and transcriptional activation permissibility. In ESCs, chromatin is known to have unique characteristics compared to differentiated cells, including an open chromatin status (Gaspar-Maia et al., 2011; Meshorer and Misteli, 2006; Mattout and Meshorer, 2010) (**Figure I.6**).

This is not a specific feature of naïve, but is also common in primed pluripotent cells (Schlesinger and Meshorer, 2019). The increased openness of ESC chromatin compared to somatic cells is supported by direct electron microscopy observations (Park et al., 2004) or by the more diffused

staining of the heterochromatin HP1 protein (Meshorer and Misteli, 2006).

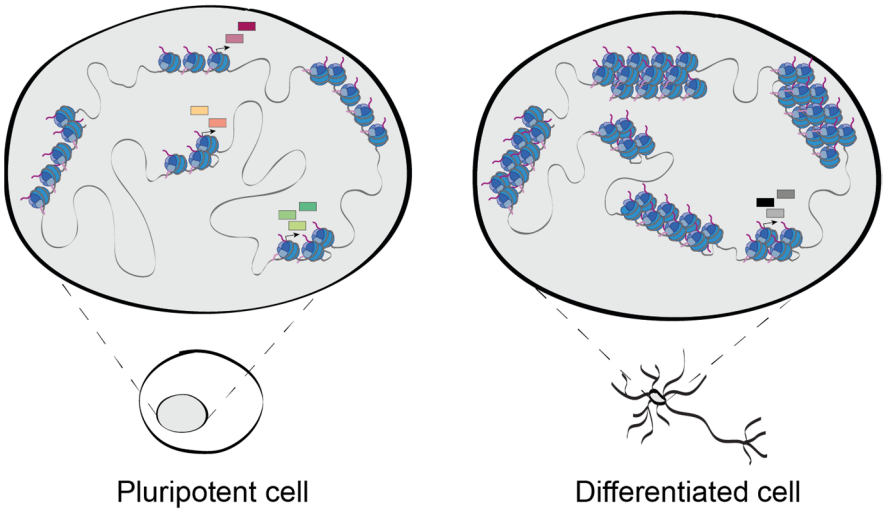


Figure I.6. Chromatin structure in pluripotent versus differentiated cells. Chromatin structure is different between uncommitted or pluripotent and differentiated cells. While chromatin in pluripotent cells (left) is open and permissive for transcription (colored rectangles represent transcripts), it is more condensed in differentiated cells (right) and less gene transcription occurs.

Overall, pluripotent cells regardless of their specific state possess an open chromatin structure, proposed to be necessary for its genome plasticity to differentiate to any cell type.

2.2 Epigenetics in mESCs

The transcriptional program of a cell defines its identity. The epigenome helps to maintain the cell specific transcriptional networks, such as the ones driving pluripotency in mESCs, by

transmitting the epigenetic information through cell divisions. DNA modifications, post-translational modifications (PTMs) of histones, histone variants incorporation or non-coding RNA (ncRNA) mediated processes are different signals for epigenetic regulation, which help controlling the characteristic chromatin landscape of ESCs and, therefore, gene transcription.

2.2.1 DNA methylation and demethylation

DNA methylation constitute an important layer for epigenetic control and, therefore, gene expression regulation. It is catalyzed by two types of DNA methyltransferases depending on their purpose: the *de novo* methyltransferases, such as DNMT3a and DNMT3b, and the maintenance methyltransferases, such as DNMT1 (**Figure I.8**). The latest ones are responsible for transferring the methylation pattern to daughter cells during replication.

Generally, DNA methylation is associated with gene repression and takes place on CpG islands, although methylation outside CpG regions can be seen in ESCs (Lister et al., 2009), being mostly found in major satellite repeats (**Figure I.9**). This unique distribution of 5-methylcytosine (5-mC) in ESCs suggests different mechanisms for DNA methylation in these cells.

DNA methylation can be reverted by passive dilution after serial cell cycle divisions (if the maintenance DNA methyltransferases are disrupted) or by active DNA demethylation performed by Ten-eleven-translocation (TET) dioxygenases, comprising TET1, TET2 and TET3. This family of proteins were described to act catalyzing the conversion from 5-mC to 5-hydroxymethylcytosine (5-hmC) (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009).

During mammalian development, two waves of DNA demethylation drive and precede key developmental processes (**Figure I.7**). The first one takes place right after fertilization occurs. These demethylation events are achieved through passive dilution after several rounds of cell division in the maternal genome, and through active DNA demethylation processes in the paternal one (Santos et al., 2002).
In later stages of development, the second demethylation wave happens guiding Primordial Germ Cells (PGCs) migration towards genital ridges, where they will give rise to the gametes, being activated their specific genes. Thus, DNA methylation/demethylation constitute an important mechanism to set the patterns needed for proper embryo development *in vivo*.

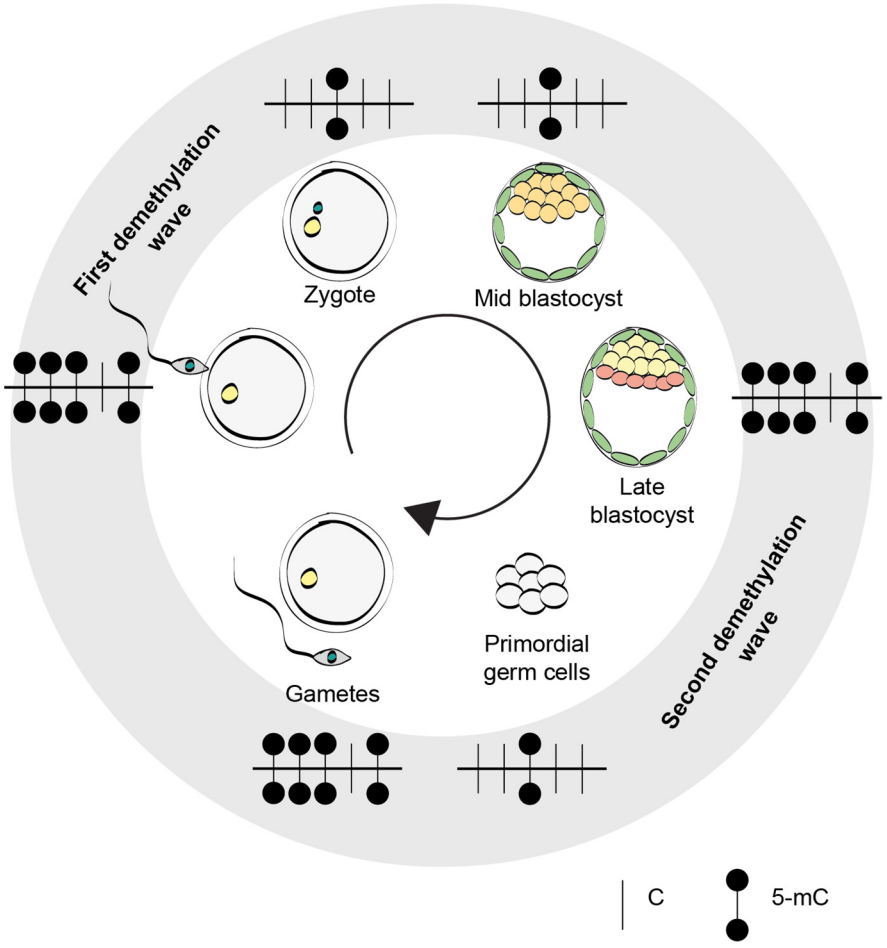


Figure I.7. DNA methylation dynamism during mouse development. Two waves of DNA demethylation occur during development: after fertilization and during primordial germ cells migration/maturation. Blastocyst methylation patterns refers to ICM cells. Black circles represent 5-mC.

The 2i/LIF cell culture conditions capture mESCs to the ground pluripotent state. Under these conditions genome methylation in mESCs is almost abolished, leading to a hypomethylated state, compared to serum/LIF cultured cells (being 20% on CpG regions in 2i/LIF in front of 80% in serum/LIF grown cells) (Stadler et al., 2011; Leitch et al., 2013; Habibi et al., 2013;

Ficz et al., 2013). This state is mainly accomplished by the downregulation of UHRF1, the co-factor of DNMT1, as well as the downregulation of DNMT3A and DNMT3B DNA methyltransferases and their co-factor DNMT3L (von Meyenn et al., 2016; Leitch et al., 2013). The hypomethylated state of 2i/LIF cultured mESCs is another feature which makes them more similar to the *in vivo* ICM cells from E3.5 embryo. Alternatively, the serum/LIF cultured cells show a methylation state more similar to E6.5 post-implantation embryo (Habibi et al., 2013).

DNA methylation is dispensable for pluripotency maintenance of mESCs (Tsumura et al., 2006). However, when pushed for differentiation, mESCs demand the DNA methylation in order to silence pluripotency factors, such as *Nanog* and *Oct-4* (Schmidt et al., 2012).

DNA methylation through the 5-mC mark is not the only methylation form that exerts a regulatory gene expression mechanism. In fact, 5-hmC is found to be distributed in CpG islands at gene bodies on active genes, promoters of active and paused genes and active enhancers (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009; Pastor et al., 2013; Stroud et al., 2011; Szulwach et al., 2011) (**Figure I.8**). TET1, TET2 and TET3 proteins were later described to act further oxidizing 5-hmC to 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-CaC) (Ito et al., 2011; He et al., 2011) (**Figure I.8**). Complete DNA demethylation to an unmodified cytosine (C) is achieved either from 5-fC or 5-CaC by the activity of thymine DNA

glycosylase (TDG) coupled to base excision repair (BER) (He et al., 2011; Maiti and Drohat, 2011). 5-fC and 5-CaC are relatively low abundant in the genome compared to 5-mC, being 20 and 3 parts per million (ppm) of the total cytosine respectively (Ito et al., 2011).This suggest that TDG removes 5-fC and 5-CaC actively and efficiently and that these two marks exist temporarily.

While 5-mC genome-wide occupancy can be determined by bisulfite conversion-based or affinity-based methods (Frommer et al., 1992; Booth et al., 2012), 5-fC and 5-CaC DNA modifications cannot be distinguished using this methodology (Song et al., 2012; Raiber et al., 2017). Immunoprecipitation methods based on the use of specific antibodies coupled to high throughput sequencing (MeDIP) were then developed and allowed the detection of 5-hmC, 5-fC and 5-CaC genome-wide (Ficz et al., 2011; Koh et al., 2011). As an alternative to antibody-based methods and to circumvent bias toward high density regions, chemical labeling methods using aldehyde reactive probe were established (Pastor et al., 2011). However, this methods may also label abasic sites, product of DNA repair processes (Ide et al., 1993; Nakamura et al., 1998). Other methods have been developed as well to characterize genomic occupancy of DNA modifications marks at single base resolution (Xia et al., 2015).

5-fC has been found to occupy CpG islands of gene bodies and promoters corresponding to active genes, satellite repeats and active/poised enhancers as well as other regulatory

regions (Raiber et al., 2012; Shen et al., 2013; Song et al., 2013; Neri et al., 2015) (**Figure I.8**). This distribution overlaps with 5-hmC, being common 72% of the occupied regions by the two marks (Raiber et al., 2012; Xia et al., 2015). TDG was shown to be active in these sites, as its knock-down leads to an increase of 5-fC (Raiber et al., 2012; Xia et al., 2015).

On the other hand, 5-CaC has been determined to be located at CpG islands of active promoters and active/poised enhancers and other regulatory regions (Neri et al., 2015; Shen et al., 2013) (**Figure I.8**).

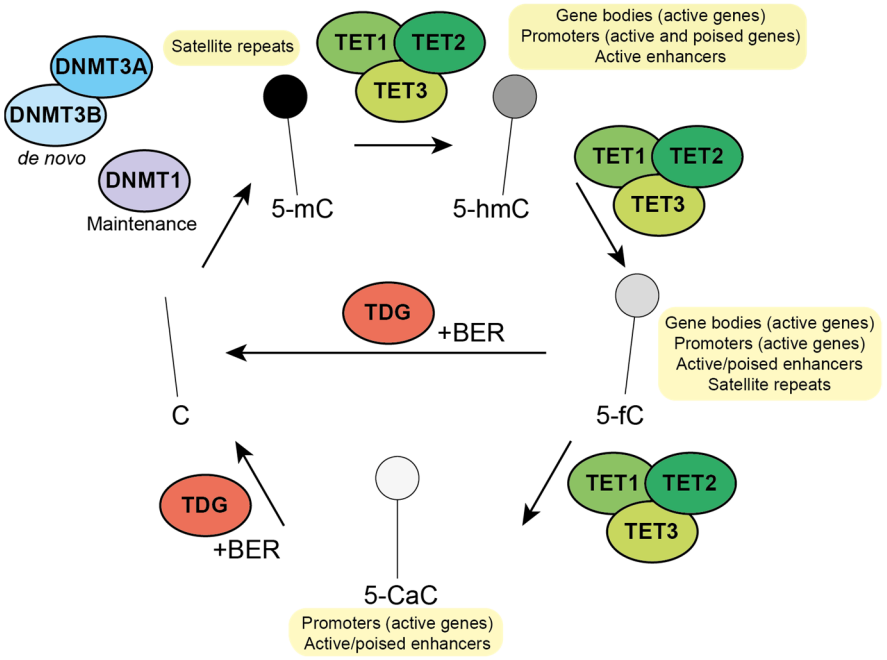


Figure I.8. Active DNA demethylation pathway. Unmodified cytosine is methylated by DNMTs resulting in 5-mC. TET proteins then oxidize 5-mC to 5-hmC, 5-fC and 5-CaC. Finally, 5-fC and 5-CaC are excised by TDG coupled with BER mechanisms, returning to the initial unmodified state. Genome-wide distribution of 5-mC, 5-hmC, 5-fC and 5-CaC is indicated.

The oxidized 5-mC forms of the DNA demethylation cycle have been proposed to be not only intermediates but also different layers of epigenetic regulation, as their genomic occupancy is associated with different degrees of gene expression. 5-fC enriched regions have been found to be co-occupied by the transcriptional co-activator p300 (Song et al., 2013), suggesting a role of this mark in recruiting chromatin factors. However, we are still far from having a complete picture and further studies will be needed to clarify the biological function of DNA modifications.

2.2.2 Histone post-translational modifications

Not only DNA, but histones can be chemically modified as well, leading to another layer of epigenetic mechanism for chromatin accessibility and transcriptional regulation. More specifically, at least nine different histone post-translational modifications (such as methylation, acetylation or phosphorylation) have been described (Kouzarides, 2007),. While promoters of actively transcribed genes are generally covered with H3K4me3 and tri-methylation of lysine 36 on histone H3 (H3K36me3), promoters and gene bodies of repressed genes are commonly decorated with H3K27me3 (Kouzarides, 2007; Bernstein et al., 2006; Benayoun et al., 2014; Young et al., 2011) (**Figure I.9**).

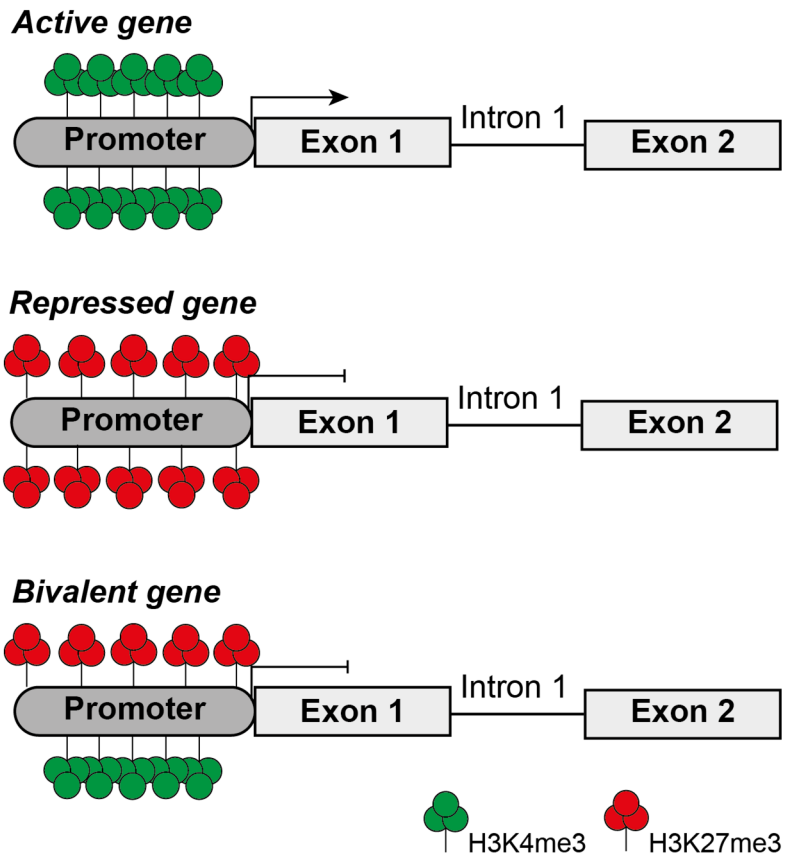


Figure I.9. H3K4me3 and H3K27me3 are located at promoter of active and repressed genes respectively. While promoters of active and repressed genes are decorated with H3K4me3 and H3K27me3 respectively, bivalent genes promoters are decorated with both active and repressive marks. Bivalency is thought to poise developmental genes for rapid differentiation when specification stimuli are present in the media.

In 2006, Berstein and colleagues described the co-existence of two opposing histone modification in gene promoters of mESCs (Bernstein et al., 2006) (**Figure I.9**). Thus, the active H3K4me3 and the repressive H3K27me3 marks are placed together on non-expressed developmental genes. This co-occupancy is called bivalency and is thought to repress

differentiation genes while keeping them poised for later activation. By having this chromatin configuration, mESCs would rapidly activate developmental genes upon differentiation stimuli by removing H3K27me3.

3. Cell cycle in mESCs

ESCs are characterized by its fast proliferative rate, as opposed to somatic or differentiated cells. This high proliferation of pluripotent cells is accomplished by having a short G1 phase (Becker et al., 2006; Singh and Dalton, 2009). When ESCs undergo differentiation upon specification cues exposure, proliferation rates decrease and cell cycle is re-structured (Savatier et al., 1996). More concretely, their G1 phase is elongated (Fluckiger et al., 2006; Stead et al., 2002), going from less than 2h G1 phase in mESCs, to around 10h for a differentiated mouse cell (Stead et al., 2002) (**Figure I.10**). Strikingly, when differentiated cells fate is changed back to pluripotent through the reprogramming process, the G1 phase is shortened again (Ghule et al., 2011). Both the shortened and enlarged G1 phase in pluripotent and differentiated cells respectively are not specific traits of cultured cells, but it is also true for *in vivo* cells (Snow, 1977).

Not only in ESCs, but other stem cell populations have been described to present a short G1 cell cycle phase that is elongated when differentiation occurs. Neural stem cells are one example, as during the process of neurogenesis the G1 phase length is increased by four fold, being shorter in the more stem population and longer in the differentiated one (Takahashi et al., 1995). In order to test if G1 cell expansion was the cause or the consequence of neural differentiation,

Lange and co-workers shortened G1 phase by overexpressing Cdk4-CyclinD1 in cortical progenitors. In fact, by shortening the G1 phase they could inhibit neurogenesis and promote the proliferation of neural progenitors (Lange et al., 2009). Not only in neurogenesis, but the cell cycle-differentiation link has also been proposed to occur in hematopoietic stem cells (Orford and Scadden, 2008). However, more studies will have to be performed to clarify it.

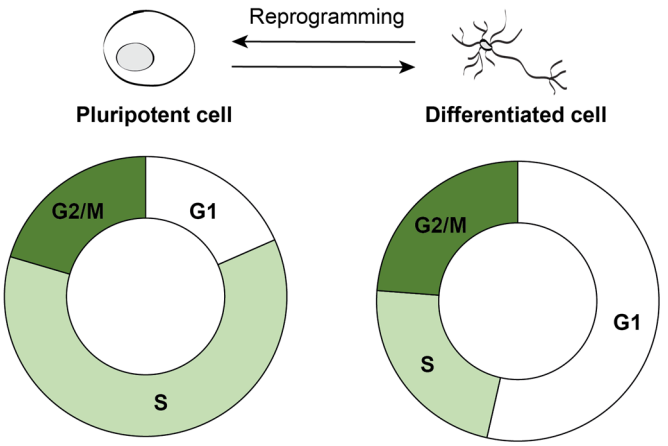


Figure I.10. Cell cycle structure in pluripotent and differentiated cells. Cell cycle structure is represented in proportion of cells in each cell cycle phase (G1, S, G2/M) in pluripotent mESCs (left) and differentiated cells (right). Although the time spend in S phase is similar between both cell states, G1 is elongated when cells differentiate.

This association between cell cycle and stemness is not specific for stem cell populations, but it is also happening in other species. Fly, frog, fish, monkey or human uncommitted cells also show high proliferation rates accompanied by a short G1 phase (Edgar and Lehner, 1996; Heasman, 2006; Murray and Kirschner, 1989; Yarden and Geiger, 1996; Fluckiger et al., 2006; Becker et al., 2006).

All together suggests a connection between cell cycle and pluripotency or differentiation. However, the biological implications of this abbreviated cell cycle as well as the mechanisms behind it are still under debate.

3.1 Pluripotent versus somatic cell cycle

To have a better idea of how and why pluripotent cells possess a truncated G1 phase and the implications of it, it is important to focus the attention on the cell cycle differences between pluripotent and somatic cells.

In somatic cells, the cell cycle is driven and tightly regulated by the cyclin-dependent protein kinases (CDKs), which form complexes with their regulatory subunits called cyclins. Specific CDK-cyclin complexes are formed in concrete steps of the cell cycle to regulate the progression to the following cell cycle phase. Thus, CDKs show a phase-specific activity (Stead et al., 2002). Transition from G1 to S phase is regulated by the CDK4/6-cyclin D and CDK2-cyclin E complexes. Such complexes phosphorylate and inactivate retinoblastoma (RB) proteins. This inhibition lead to the activation of the E2F transcription factor target genes, which are in charge of the transcription of the elements/factors needed for next cell cycle phases. This step, which is called restriction point (R point), is critical for the cell, as it will decide whether the cell progresses

through the cell cycle and enters S phase or exit cell cycle and enters a G0 quiescent state (Pardee, 1974).

While this is true for somatic cells, mESCs exhibit some differences.

First, they lack a CDK-cyclin cell cycle regulation, as CDKs and cyclins are constitutively active during the entire cell cycle (Stead et al., 2002). This sustained CDKs activation (especially for CDK2) leads to a permanent biochemical RB inactivation and subsequent cell cycle progression. This would explain in part the quick G1-S phase transition present in mESCs. Consistently, triple knockout (KO) for all RB family members has no major effect on mESC proliferation, highlighting their dispensability (Dannenberg et al., 2000). An exception for this constitutive activation of CDKs is the case of CDK1 and its corresponding subunit cyclin B1, which are cell cycle-regulated (Stead et al., 2002).

Second, CDK inhibitory molecules (CDKIs) are not present in mESCs (Stead et al., 2002; Savatier et al., 1994), contributing to the sustained activity of CDKs throughout the cell cycle.

Finally, the existence of microRNAs controlling the cell cycle of mESCs add another layer of cell cycle regulation. Micro-RNAs are non-coding small RNAs that bind mRNAs and regulate their transcription. These micro-RNAs target CDKIs, further supporting the CDKs constant activation (Wang et al., 2008).

All these make mESCs unique in terms of cell cycle regulation and suggest other mechanisms regulating its progression that are still to be identified.

However, all the described characteristics are true for mESCs cultured in serum/LIF, while some differences apply when mESCs are grown in 2i/LIF. mESCs grown in the presence of the two inhibitors show a longer G1 phase (ter Huurne et al., 2017; Jaime-Soguero et al., 2017). CHIR99021, one of the two inhibitors, activates the Wnt pathway by inhibiting its inhibitor GSK3. The Wnt pathway has been described to reduce mESCs proliferation, thus explaining the longer G1 phase in 2i/LIF cultured cells (Jaime-Soguero et al., 2017). In contrast to serum/LIF cells, CDKIs such as P16, P21 or P27 are cell cycle-regulated in 2i/LIF mESCs. The presence of these CDKIs are responsible for the RB hypo-phosphorylation, which in turn is mediated by the ERK signaling pathway inhibition by PD0325901 (ter Huurne et al., 2017). In parallel, P53 pathway, which regulates the G1 check-point, is active in 2i/LIF mESCs in contrast to serum/LIF cells (Huurne et al., 2019), arguing in favor of a G1 check-point restoration in 2i/LIF. Moreover, P53 activates Rb1, contributing thus to G1 elongation as well (Huurne et al., 2019).

Not only mESCs cultured in 2i/LIF conditions but also hESCs do not share all the mentioned cell cycle regulatory mechanisms of mESCs cultured in serum/LIF, although they do in having a shortened G1 phase. Thus, hESCs show tight cell cycle regulation of CDKs and cyclins as well as high levels of CDKIs, both being cell cycle regulatory mechanisms shared with somatic cells (Neganova et al., 2009). Furthermore, in hESCs, CDK2 inhibition downregulates the expression of

pluripotency factors, such as OCT-4, indicating a direct link between cell cycle machinery disruption and pluripotency/differentiation (Filipczyk et al., 2007; Neganova et al., 2009). However, no such effect is seen in mESCs (Li et al., 2012).

These differences may account for intra-species differences and suggest different or alternative explanations for the cell cycle-linked pluripotency exit between these two species.

3.2 The cell cycle-linked pluripotency exit

The direct cell cycle pluripotency exit link was first described in 1987 by Mummery and colleagues, when by using EC cells they showed that G1 cells have a higher propensity to respond to differentiation cues upon retinoic acid treatment (Mummery et al., 1987a; b) (**Figure I.11**). Later studies wanted to extrapolate this to ESCs, and used chemical or genetic approaches to evaluate if G1 cells were also more susceptible for differentiation (Filipczyk et al., 2007; Neganova et al., 2009). However, although similar conclusions were raised, the cell cycle arrest itself due to synchronization drugs effects was not excluded to be the cause for G1 differentiation bias.

To bypass this limitation, Sela and colleagues (2012) made used of the differential cell sizes shown in different phases of the cell cycle to separate cells from G1 to M phases by centrifugal elutriation (Sela et al., 2012). Similar to what was

shown before, the authors described a lower ability to form colonies and a consequent more propensity to differentiate by G1 cells (Sela et al., 2012). Nonetheless, apoptotic effects were not ruled out from this study.

More recent studies used a fluorescent ubiquitin cell cycle indicator approach named Fluorescent Ubiquitination-based Cell Cycle Indicator (FUCCI) to overcome previous methods' detriments (Sakaue-Sawano et al., 2008). This new methodology allows to monitor cell cycle in time and space without the cell cycle perturbations or cytotoxic effects produced by synchronization drugs. It also allows to separate cells in different cell cycle phases by fluorescent activated cell sorting (FACS). Using the FUCCI approach coupled to FACS sorting, Coronado and colleagues described G1 as a cell cycle phase in which mESCs are more permissive for differentiation (they give rise to more differentiated colonies) after retinoic acid induction (Coronado et al., 2013). Similarly, using the FUCCI system, Calder and co-workers showed G1 as the cell cycle phase less capable of producing pluripotent colonies in hESCs, characterized by the lack of positive staining for alkaline phosphatase (Calder et al., 2013). Complementary to this was the fact that S and G2 phases of the cell cycle were shown to restrict the pluripotent state dissolution in hESCs, while not in G1. This makes G1 a permissive phase of the cell cycle for differentiation initiation (Gonzales et al., 2015).

All these studies pointed out G1 as the cell cycle phase from which ESCs differentiate, and therefore the window of opportunity for cell differentiation (**Figure I.11**).

3.3 Molecular mechanisms for cell cycle-linked pluripotency exit

The question about why G1 represents a time opportunity for cells to differentiate soon emerged. And with the question, the attempts to find a mechanism that could explain this link between cell cycle and differentiation. Mainly two models have been proposed up to date. While one supports cell cycle machinery as the direct cause, the other points towards epigenetics and gene regulation.

3.3.1 D cyclins drives the cell cycle-linked pluripotency exit

Using the FUCCI technology, Pauklin and co-workers focused on studying the molecular mechanism behind the cell cycle and cell fate decisions in hESCs (Pauklin and Vallier, 2013). In their experiments, cell cycle was shown to play a role in determining cell fate, as early G1 cells are restricted to endodermal differentiation, while late G1 to ectodermal (Pauklin and Vallier, 2013). Endoderm specification is achieved *via* activation of the Activin/Nodal signaling. This signaling pathway was described

to be controlled by cyclin D proteins, which activate CDK4/6 in the late G1 phase of the cell cycle. The CDK4/6 activation leads then to the phosphorylation of SMAD2 and SMAD3, which prevents their nuclear entrance and consequent activity. Thus, in late G1 phase, when cyclin D-CDK4/6 complex is more abundant, SMAD2/3 phosphorylation results in the Activin/Nodal signaling inhibition, which blocks endodermal differentiation and allow for ectodermal specification (Pauklin and Vallier, 2013). Therefore, cyclin D, a cell cycle player, orchestrate hESCs differentiation either to the endoderm lineage in early G1 or to ectodermal lineage in late G1.

Later on, the same authors described cyclin D1 as the cyclin D protein that participates in the developmental-cell cycle bias in hESCs, through an independent mechanism of the Activin/Nodal/Smad2/3 pathway. Cyclin D1 was shown to be bound to chromatin, sitting in late G1 together with coactivators on neuroectodermal genes and with corepressors on endodermal genes, therefore regulating transcription of developmental genes (Pauklin et al., 2016). Thus, cyclin D1 alone, together with its activity through the Activin/Nodal/Smad2/3 pathway, would act coordinating the hESCs cell fate decisions during cell cycle (**Figure I.11**).

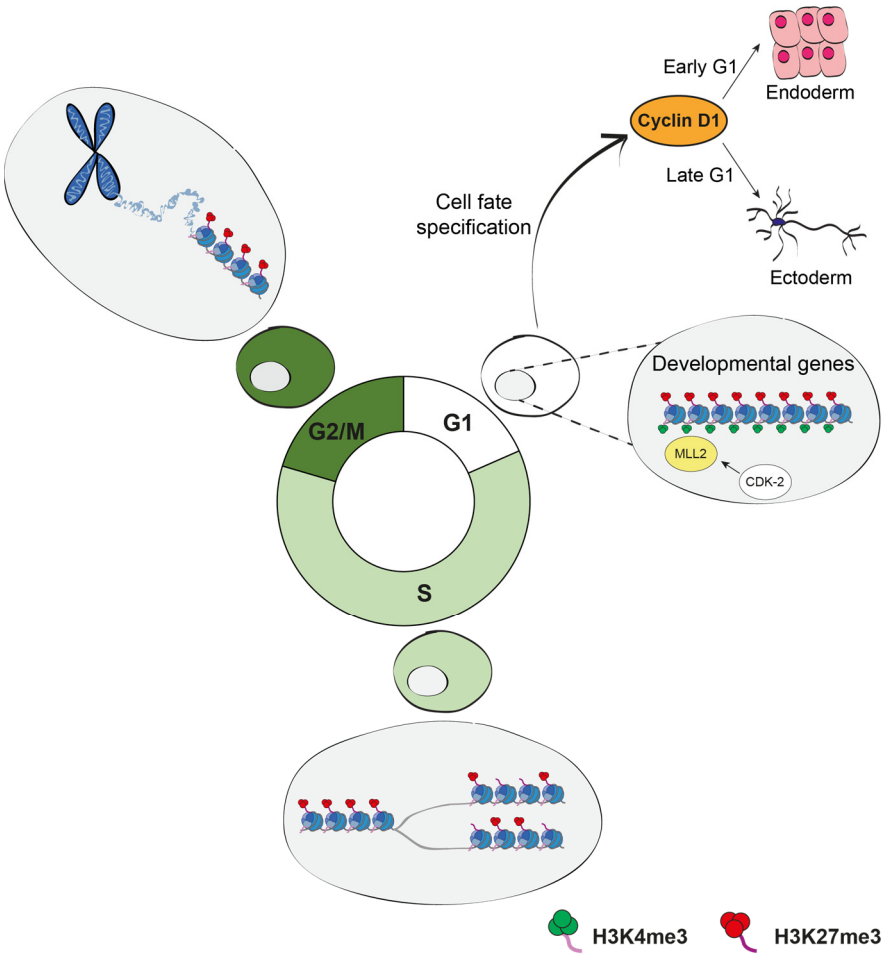


Figure I.11. G1 phase as a window opportunity for ESCs differentiation. ESC differentiation starts from the G1 phase of the cell cycle. Two models have been proposed to explain the mechanism for this. The first one consist on a cell-cycle regulated differentiation bias through the action of cyclin D1. The second mechanism involves epigenetics, pointing to G1 established bivalency though MLL2 action as the explanation for G1 cells being ready to respond to differentiation cues, according to Singh and co-workers (Singh et al., 2015). However, other authors described M phase as the cell cycle phase where bivalency is stablished, preparing the cells for differentiation in G1 upon differentiation cues (not depicted) (Grandy et al., 2015).

3.3.2 Bivalency as the mechanism for the cell cycle-linked pluripotency exit

In parallel with the described studies, and with the same aim of understand the molecular mechanisms behind the G1 cells differentiation priming, RNA-sequencing analysis (RNA-seq) in hESCs from FACS sorted FUCCI cells was performed. In line with previous observations, the analysis revealed a dynamic regulation of the cell transcriptome during cell cycle, being upregulated developmental genes (including markers of three germ layers) in G1 (Singh et al., 2013). In association with this upregulation, 5-hmC, an epigenetic mark associated with transcription activation, was found enriched in G1 cell population (Singh et al., 2013).

The occupancy of the two epigenetic marks that account for bivalency in ESCs H3K4me3 and H3K27me3 was also assessed by ChIP-seq in hESCs by Singh and co-workers (Singh et al., 2015). Bivalency was shown to be cell-cycle regulated and proposed to be crucial for the differentiation priming in G1 (Singh et al., 2015) (**Figure I.11**). Specifically, the active mark H3K4me3 was more enriched in G1 cells in approximately half of the ~2000 defined as bivalent genes. These genes were developmental genes, in agreement with the upregulation shown in transcription before (Singh et al., 2013). However, non-cell-cycle regulated bivalent genes were not related to development or differentiation (Singh et al., 2015). The molecular mechanism behind this H3K4me3

dynamics involves MLL2 phosphorylation by CDK2 and, thus, its cell cycle regulation. MLL2 is a histone methyltransferase that mediates the H3K4me3 mark deposition. Chromosome architecture is also altered during cell cycle. More specifically, promoter-enhancer loops are described to occur on the developmental genes *Gata-6* and Sex Determining Region Y-Box 17 (*Sox-17*) and to be key for later proper differentiation, as the deletion of their enhancers leads to an inability to upregulate these genes when differentiation is triggered. Despite the described observed changes in H3K4me3 during the cell cycle, no changes in H3K27me3 were observed (Singh, 2015) (**Figure I.11**).

Similarly, using chemically synchronized hESCs, H3K27me3 was also shown to be constant across cell cycle by Grandy and co-workers (Grandy et al., 2015). However, H3K4me3 was described to be enriched in the M phase of the cell cycle, being G1 the cell cycle phase with a lower enrichment. These results were accompanied by MLL enrichment during mitosis (and a reduction during G1). Nonetheless, when cells were differentiated, the cell cycle H3K4me regulation was abolished. Thus, H3K4me3 was suggested to be responsible for pluripotent cell preparation during mitosis for later developmental gene activation in G1 phase, in case differentiation stimuli are present (Grandy et al., 2015).

As a developing field, results still show inconsistencies on how bivalency is regulated cell cycle wise and how it is linked to cell fate biases seen in early/late G1 phase of the cell cycle in

hESCs (Singh et al., 2013). Further studies will be needed to address this matter and see if these mechanisms are also kept in other pluripotent cells like mESCs, where cyclin Ds are expressed at low levels and are not cell cycle-regulated. Chromatin configuration and proper quantification of material by spike-in from different cell cycle phases should be taken into account for further reports. Spike-in normalization consist in RNA with known sequence and quantity that is used to correct for the differences in total RNA amount in the different samples.

More generally, the question on why G1 phase is the window of time when cells start their lineage commitment is still open in the field and further analysis on the mechanism underlying the cell cycle-pluripotency link may be key to address it.

AIM

The aim of this thesis was to identify and characterize the molecular pathways regulating the cell cycle-linked pluripotency exit. To achieve the main aim, the thesis project was sub-divided into two work packages, which in turn enclose different objectives:

- Work package I: Identification of the molecular pathways regulating the cell cycle-linked pluripotency exit.

  The work package I is divided into:

  - Functional characterization, which includes the following objectives:

    I. Interrogate the impact of cell cycle progression on the self-renewal capacity of mESCs.

    II. Interrogate the impact of cell cycle progression on the pluripotency capacity of mESCs *in vitro*.

    III. Interrogate the impact of cell cycle progression on the pluripotency capacity of mESCs *in vivo*.

  - Molecular characterization, comprising the following objectives:

    IV. Investigate the transcriptome dynamics during cell cycle in mESCs.

V. Inspect the proteome dynamics during cell cycle in mESCs.

From this first work package, a candidate protein was expected to be obtained, which would be the main focus for the work package II:

- Work package II: Characterization of the molecular pathways regulating the cell cycle-linked pluripotency exit.
  The work package II includes the following objectives:

  VI. Characterize functionally the impact of the candidate protein on cell cycle-linked pluripotency exit.
  VII. Characterize at the molecular level the role of the candidate protein on cell cycle-linked pluripotency exit.

RESULTS

1. Gem-mESCs as a model to study cell cycle

In order to accomplish the aim of this thesis, that is to identify and characterize the molecular pathways regulating the cell cycle-linked pluripotency exit, Gem-mESCs (Aranda et al., 2014) were chosen as a cell model based on the FUCCI system. The FUCCI system, described by Sakaue-Sawano and co-workers, consist on the expression of two different fluorescent proteins fused to GEMININ and CDT1, two proteins that oscillate during cell cycle (Sakaue-Sawano et al., 2008). The FUCCI system confers some advantages compared to more traditional methods for studying cell cycle, such as the use of synchronizing drugs or centrifugal elutriation. For instance, it overcomes the problems of cytotoxicity induced by conventional synchronizing drugs (Boward et al., 2016). Moreover, synchronizing drugs have not been probed to not affect pluripotency of mESCs. While centrifugal elutriation allows the separation of cells in different cell cycle phases by size differences without perturbing cell cycle (McEwen et al., 1968; Banfalvi, 2011), it requires specific equipment and set up conditions. Moreover, the FUCCI tool allows for cell cycle progression tracking by imaging, as well as FACS cell sorting in different phases of the cell cycle, obtaining thus pure population of G1, S and G2/M phases.

The Gem-mESCs line based on the FUCCI system was generated by stably transfecting R1 mESCs to constitutively express the fusion protein composed of the human cell cycle

indicator GEMININ (fragment from the aminoacid 1 to 110) and the green fluorescent protein monomeric Azami-Green1 (mAG1) (Aranda et al., 2014) (**Figure R.1**, FUCCI reporter mAG1-Geminin). This fusion protein starts to be synthesized in the S phase of the cell cycle and it is progressively accumulated in the cell until the end of the mitosis, when it is targeted by the proteasome system and ultimately degraded. This cell cycle-dependent expression allows to follow cell cycle dynamics, as G1 cells are not fluorescent, S phase cells are fluorescent with a moderate intensity and G2/M cells, high fluorescent (**Figure R.1**, color scheme).

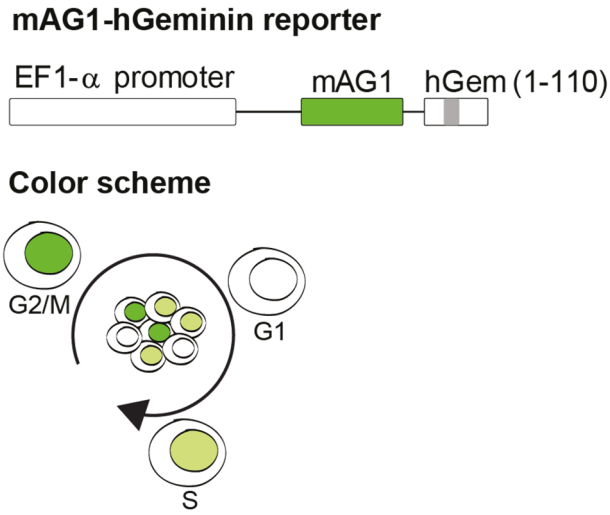


Figure R.1. mAG1-hGeminin reporter as a cell cycle indicator. (Top) Scheme of the mAG1-hGeminin reporter used, including EF1-α promoter, mAG1 and hGeminin (1-110 aminoacids). (Bottom) Color code that follow expressing Gem-mESCs. G1 cells do not express the fusion protein (being therefore not fluorescent), S cells start to expressing it (mid fluorescent) and G2/M cells are the cells where the fusion protein is more accumulated and show higher fluorescence intensity. Adapted from Aranda et al., 2014.

Gem-mESCs were cultured in 2i/LIF medium, supporting the ground pluripotent state of mESCs and resembling more to the cells found in the blastocyst *in vivo* embryo (Kolodziejczyk et al., 2015). All the experiments promoting mESCs pluripotency presented in this thesis were performed by culturing the cells in 2i/LIF conditions.

In order to validate Gem-mESCs as a good model to study the cell cycle-linked pluripotency exit, cells were checked and compared with its parental ES cell line R1 for cell morphology by conventional light microscopy (**Figure R.2**) and gene expression by RNA-seq (**Figure R.3**). No changes in cell morphology were observed, as Gem-mESCs displays the typical dome-shaped colony formation (**Figure R.2**).

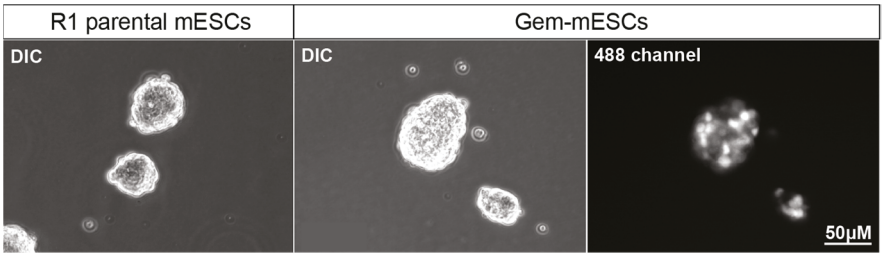


Figure R.2. Cell morphology is equivalent between parental R1 and Gem-mESCs. Cell morphology was compared between parental R1 and Gem-mESCs by Differential Interference Contrast (DIC) microscopy (left and middle) and green fluorescence was checked in Gem-mESCs (right). Note that green-fluorescent signal is observed for some of the colony cells, indicating cells in S and G2/M phases.

RNA-seq analysis showed 125 genes up-regulated in Gem-mESCs and 524 down-regulated compared to R1 (**Figure R.3**, left). Gene Ontology (GO) analysis of the differentially expressed genes showed extracellular matrix as the most enriched category, being downregulated in Gem-mESCs,

possibly due to specific clone characteristics. However, pluripotent and differentiation related categories were not significantly enriched (**Figure R.3**, right). These indicates that Gem-mESCs are equivalent to their parental cell line R1.

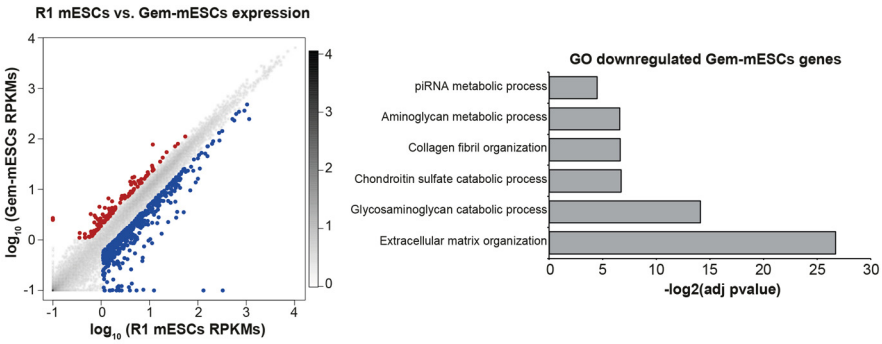


Figure R.3. Gene expression analysis comparing parental R1 and Gem-mESCs. (Left) Scatter plot showing the RPKMs of R1 and FUCCI mESCs of expressed genes (equal or higher than 1 RPKM in one of the two conditions). The average of gene expression values per each gene of 2 independent replicates is provided. Red and blue dots indicate differentially expressed genes by 2 or more folds, upregulated in Gem-mESCs or R1 parental mESCs respectively. The legend on the right represents dot density. Darker zones represent higher dot density. (Right) GO analysis of the differentially expressed genes. Note that the most enriched categories account for extracellular matrix. Pluripotent or differentiation-related categories are not significantly different between both cell lines.

One of the advantages of the use of the Gem-mESCs is that they can be sorted according to their green fluorescence into G1 and S-G2/M population of cells by FACS (**Figure R.4**). Purity of sorted populations was evaluated by propidium iodide (PI) incorporation, as it intercalates the DNA and allows for the DNA content assessment. Cell sorting purity was shown to be around 90% in both negative and positive green fluorescent sorted cells. By means of this, 90% of sorted cells were actually

G1 cells and S-G2/M cells (for negative and positive sorted cells respectively) (**Figure R.5A**).

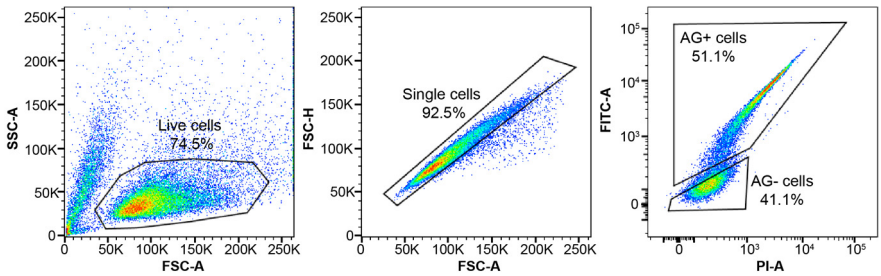


Figure R.4. Gating strategy for Gem-mESCs sorting. Gates used to FACS sort Gem-mESCs into green positive (AG+) and green negative (AG-) cells. First, live cells were determined based on SSC-A and FSC-A. Then doublets of cells were discarded and finally positive and negative green cells were selected.

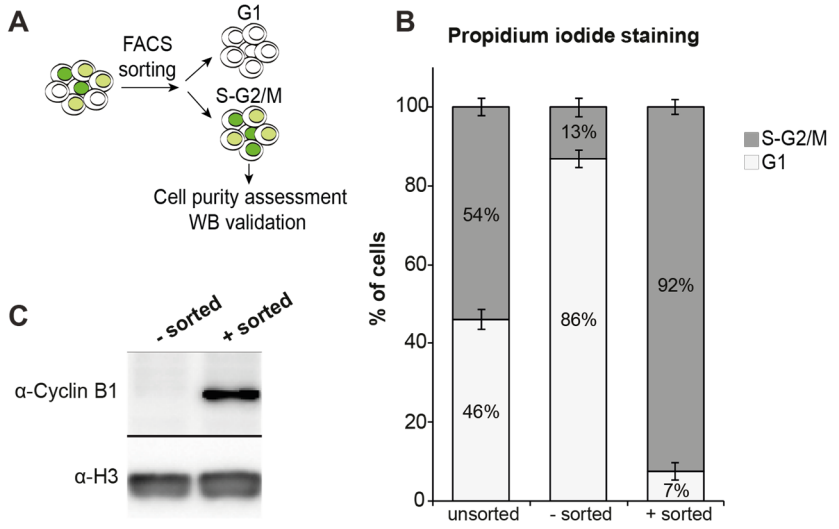


Figure R.5. Validation of negative and positive sorted populations of FUCCI mESCs as G1 and S-G2/M cells. (A) Scheme of the FACS strategy to sort cells in G1 and S-G2/M and posterior analysis. **(B)** Negative (-) and positive (+) sorted FUCCI mESCs were stained with PI and DNA content was assessed. Unsorted initial cell population was included as well. Percentage of cells in G1 and S-G2/M phases of the cell cycle is indicated for each sample condition. The average of 4 independent replicates is

provided. Error bars indicate +/- SEM. **(C)** Cell sorting was validated by Western Blot using specific antibodies.

Cell sorting efficacy was also validated extracting total protein from positive and negative sorted cells and checking protein amounts of Cyclin B1, a specific S-G2/M cyclin, by Western Blot (WB) (**Figure R.5C**). These results show a good cell sorting efficiency and thus, green negative and positive cell sorted cells to be G1 and S-G2/M respectively.

To check when the cell cycle is restored in G1 and S-G2/M sorted cells, they were cultured after cell sorting was performed and DNA content by PI was assessed at different time points (0, 8, 24 and 48 hours) (**Figure R.6A**). From PI analysis we concluded that the cell cycle structure is re-stablished to the original proportion of G1 and S-G2/M cells within 24 hours (**Figure R.6B**).

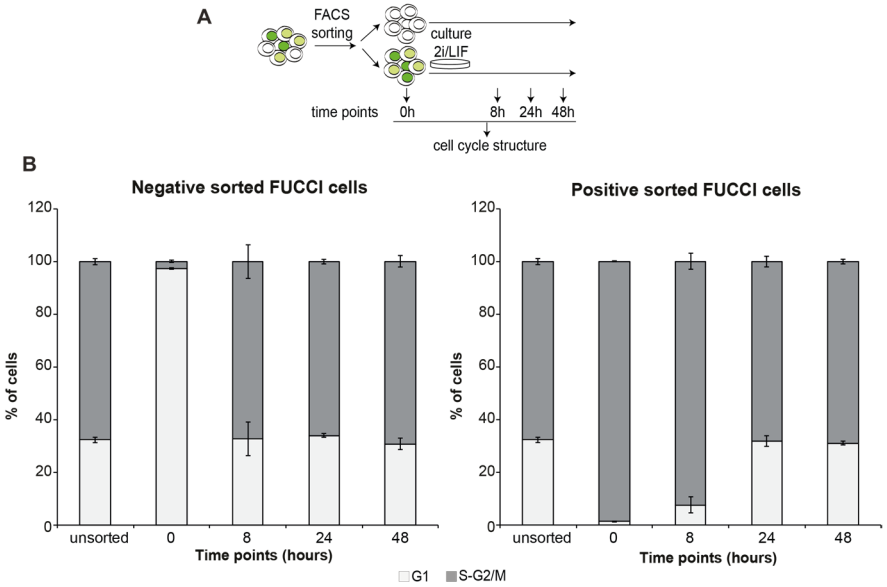


Figure R.6. Cell cycle structure restoration after cell sorting in 2i/LIF cultured cells. (A) Scheme of the FACS strategy to sort cells in G1 and S-G2/M, posterior cell culture in 2i/LIF and the different time points where cell

cycle structure was analyzed. **(B)** DNA content was assessed by PI incorporation in negative (left) and positive (right) sorted FUCCI mESCs right after cell sorting and 8, 24 and 48 hours after the sorting was performed. Unsorted initial cell population is included as well. Percentage of cells in G1 and S-G2/M phases of the cell cycle is indicated for each sample. The average of 2 independent replicates is provided. Error bars indicate +/- SEM.

All these experiments place Gem-mESCs as a suitable cell model to study cell cycle impact on self-renewal and pluripotency.

2. Self-renewal capacity is not altered during cell cycle

To interrogate the impact cell cycle progression has in the self-renewal capacity of mESCs, Gem-mESCs were sorted according to its cell cycle phase in equal number of G1 and S-G2/M cells (**Figure R.7.A**). After the sorting, G1 and S-G2/M cells were cultured in 2i/LIF separately. After two days in culture, cells were dissociated, counted and equal number of cells were cultured for 3 additional days. At this time point, the number of cells was assessed again for both conditions. RNA from G1 and S-G2/M cells was also collected after 2 days in culture and global transcriptome was analyzed by RNA-seq (**Figure R.7A**). After 2 days in culture, S-G2/M cells produced almost two times more cells compared to G1. However, cell number after 5 days in culture (including one cell splitting, platting equal cell numbers) revealed no changes between G1 and S-G2/M cell number (**Figure R.7B**). Thus, indicating equivalent self-renewal potential during cell cycle.

We speculate that the reason for S-G2/M cells to have higher number of cells at day 2 is related to the fact that they are more ahead of the cell cycle and ready to divide. In fact, when the two culture conditions reach the same cell cycle asynchrony (at 24h) (**Figure R6**), they yield the number of cells supporting an equivalent self-renewal capacity.

Supporting this, RNA-seq analysis at 48h, did not show major changes in expression (N=2; RPKM>1; 2 fold-change

difference) between the different cell cycle phases (**Figure R.7C**). Taken together, these data indicate that self-renewal is equivalent in G1 and S-G2/M cells.

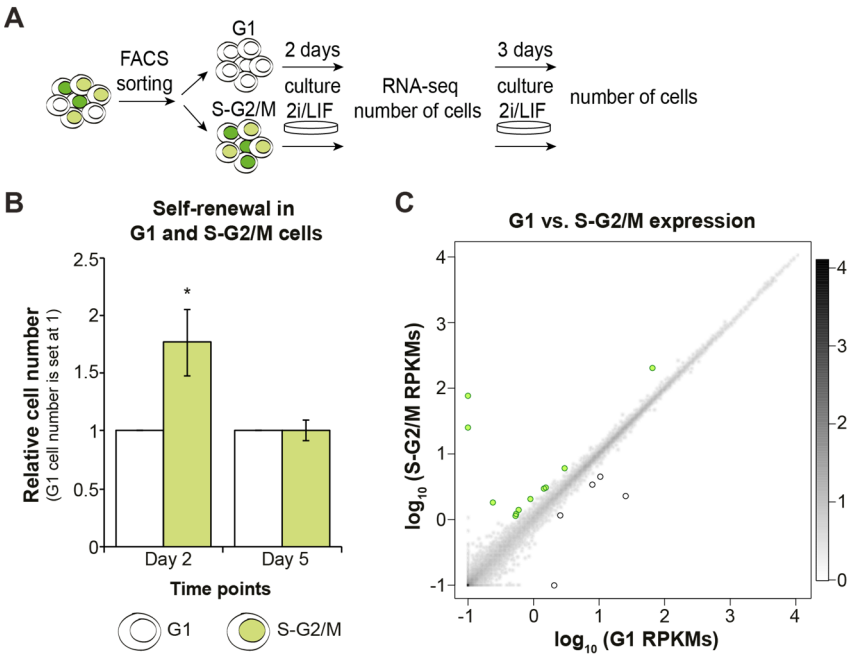


Figure R.7. Self-renewal capacity is not affected by cell cycle progression. (A) Scheme of the strategy used to assess self-renewal capacity during cell cycle. Cells were sorted in G1 and S-G2/M phases, and were put in culture to count cell number after 2 and 5 days, and analyze the global transcriptome after 2 days. **(B)** Cell number obtained for initially sorted G1 and S-G2/M cells 2 and 5 days after sorting. The averaged relative number of cells from 4 independent replicates is indicated. A student T-test was performed to test statistical significance. "*" means p-value<0.05; "**" p-value<0.01 and "***" p value<0.001. **(C)** Scatter plot showing differentially expressed genes by 2 or more folds in G1 and S-G2/M cells. The average of gene expression values per each gene of 2 independent replicates is provided. White and green-colored dots represent significantly upregulated genes in G1 and S-G2/M cells respectively. The legend on the right represents dot density. Darker zones represent higher dot density.

3. Pluripotency capacity is altered during cell cycle

In order to evaluate the impact of cell cycle progression on the pluripotency capacity of mESCs, we performed embryoid body (EB) differentiation from G1 and S-G2/M sorted cells (**Figure R.8A**). EB differentiation constitute the gold standard method to assess pluripotency *in vitro*, as if initial cells are pluripotent, they will give rise to cells from the three germ layers (endoderm, mesoderm and ectoderm).

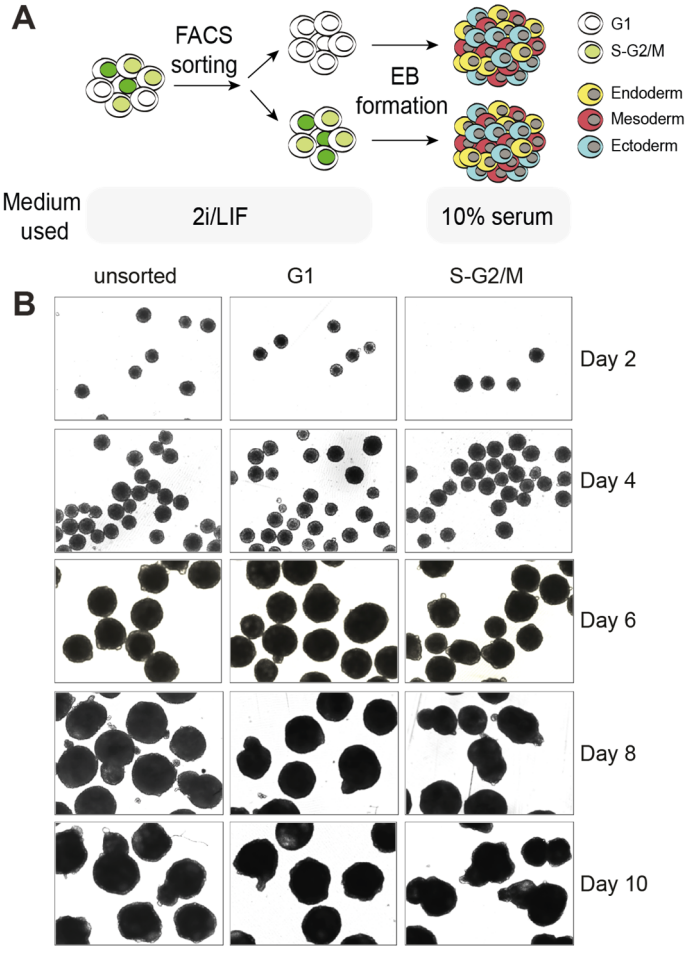


Figure R.8. G1 and S-G2/M cells are able to form EBs. (A) Schematic representation of the cell sorting performed to obtain G1 and S-G2/M cells, and following EB differentiation process, where endoderm, mesoderm and ectodermal cells are obtained. EBs are obtained by culturing the cells in 10% serum medium without LIF. **(B)** EBs at different time points of the differentiation process (from day 2 to day 10). Note that no major changes in cell size nor morphology can be observed.

Interestingly, virtually pure population of G1 and S-G2/M cells were both able to form EBs without major changes in morphology nor size (**Figure R.8B**).

To check when the cell cycle is restored in G1 and S-G2/M sorted cells in differentiating conditions, they were cultured after cell sorting in 10% serum, and DNA content was assessed by PI staining at different time points (0, 8, 24 and 48 hours) (**Figure R.9A**). From this analysis we concluded that the cell cycle structure is re-stablished to the original proportion of G1 and S-G2/M cells within 24 hours (**Figure R.9B**), equal to what was shown in 2i/LIF cultured cells (**Figure R.6B**). Thus, proportion of G1 and S-G2/M cells is only different for 24 hours of differentiation, when the differentiation stimuli are sensed by mESCs. After those 24 hours, cell asynchrony is equivalent between conditions.

To discard possible differences in EB size in G1 and S-G2/M conditions that could affect the differentiation process of mESCs (Hwang et al., 2009), the number of cells per EB was calculated at day 4 of differentiation. Number of cells per EB were not significantly different between EBs obtained from G1 and S-G2/M cells (**Figure R.9C**).

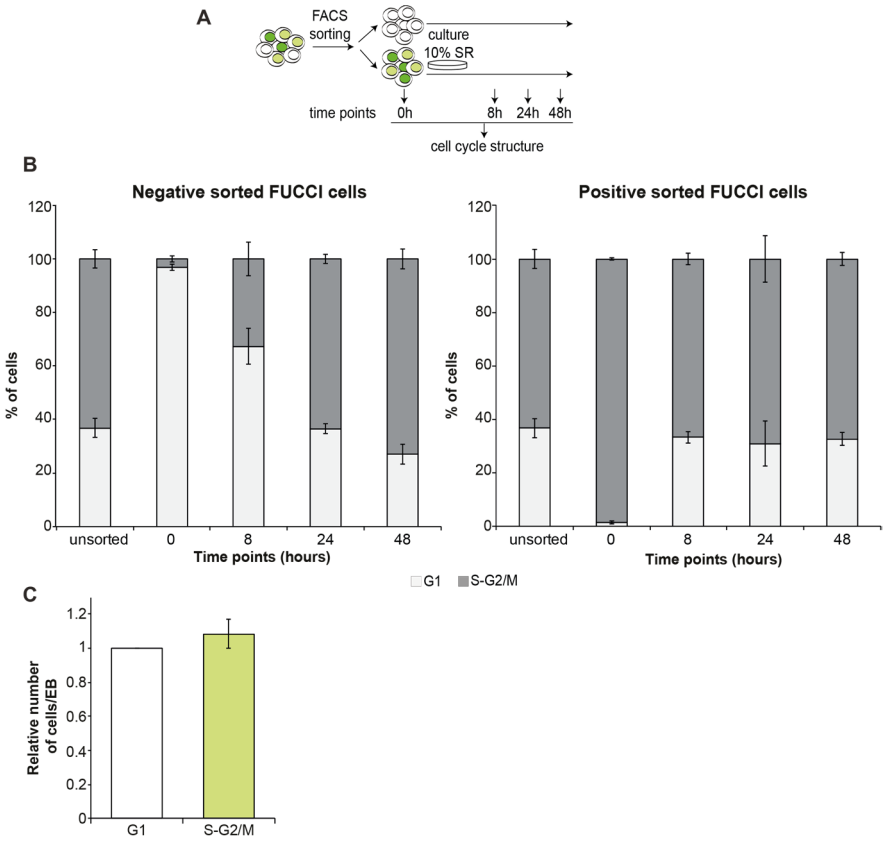


Figure R.9. Cell cycle structure restoration after cell sorting in serum cultured cells. (A) Scheme of the FACS strategy to sort cells in G1 and S-G2/M, posterior cell culture in 10% serum medium (10% SR) and the different time points where cell cycle structure was analyzed. **(B)** DNA content was assessed by PI incorporation in negative (left) and positive (right) sorted FUCCI mESCs right after cell sorting and 8, 24 and 48 hours after the sorting was performed. Unsorted initial cell population is included as control. Percentage of cells in G1 and S-G2/M phases of the cell cycle is indicated for each sample. The average of 2 independent replicates is provided. Error bars indicate +/- SEM. **(C)** Relative number of cells per EB at day 4 of differentiation. The average of 3 independent replicates is provided. Error bars indicate +/- SEM.

To further characterize the pluripotency capacity during the cell cycle, single cell RNA-seq (scRNA-seq) (in collaboration with Atefeh Lafzy and Holger Heyn, from CNAG-CRG) was performed. scRNA-seq allow to determine cell heterogeneity in

a given population. Moreover, by performing scRNA-seq on EBs, the number of cells from each germ layer based on their transcriptome will be assessed and thus pluripotency in G1 and S-G2/M conditions.

For this purpose, EBs obtanined from G1 and S-G2/M were disaggregated at different time points (from day 0 to 10) and single-cells were sorted for scRNA-seq. (**Figure R.10A**).

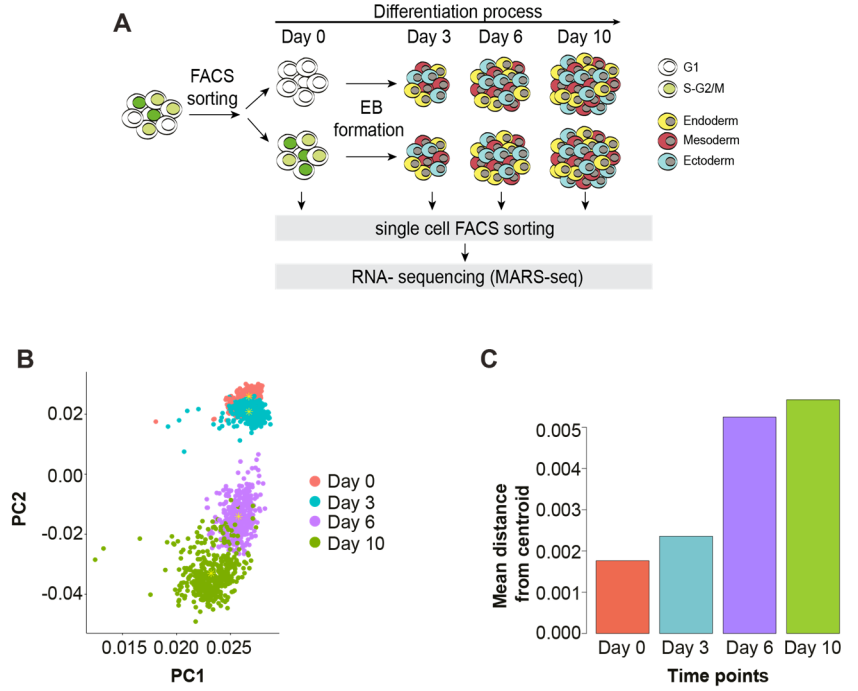


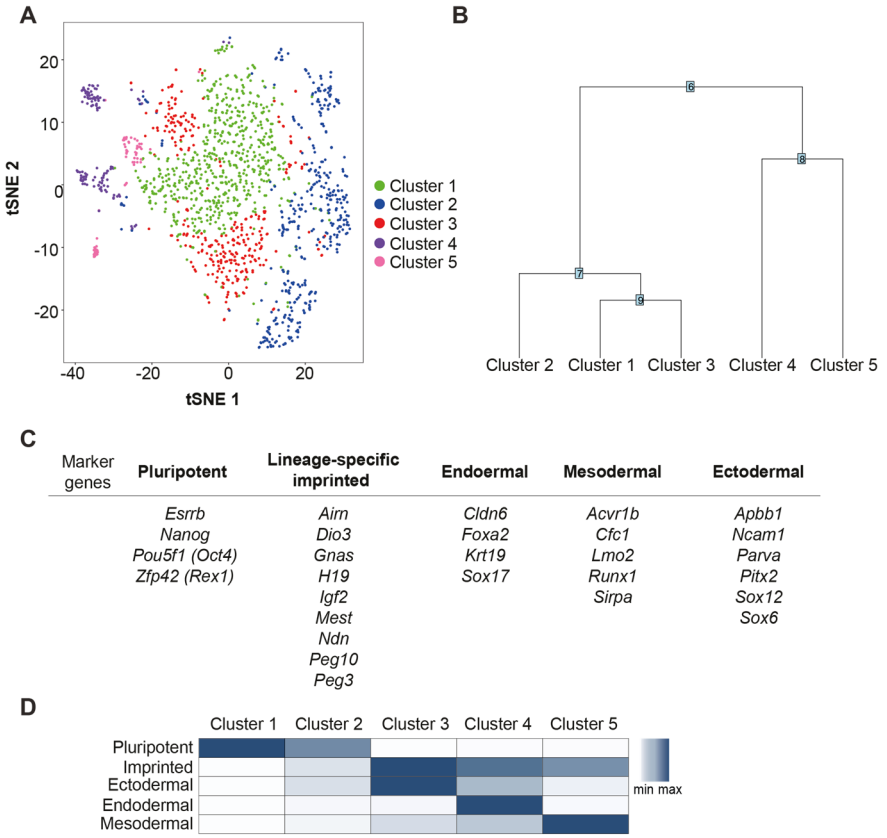
Figure R.10. Single cell RNA-sequencing enable to track the differentiation process of EBs. (A) Scheme of the experimental procedures followed. First, FUCCI mESCs were FACS sorted according to cell cycle. Then, an EB formation assay was performed, where samples were taken at different time points to then be single cell FACS sorted for scRNA-seq. **(B)** PCA based on single cell transcriptome of day 0-day 10 single cells. Day 0 and day 3 occupy a more reduced space compared to day 6 and day 10, corresponding to more transcriptionally homogeneous cells. Cells are colored according to the corresponding time point (see color legend). **(C)** Cell dispersion for each time points. Dispersion is given as the mean distance from the centroid of each cluster.

We sequenced a total of 1,919 cells. An initial filtering analysis discarded cells with a poor RNA-seq quality data (highly reduced number of reads). This resulted in a total of 1,531 cells analyzed from 0 to 10 days of EB differentiated from initially G1 and S-G2/M sorted cells. At transcriptional level, in order to avoid confounding factors to identify cellular identities, highly expressed and variable genes such as histones or cell cycle genes were discarded. After this filtering, a total of 14,904 genes across the 1,531 cells were used for the analysis.

The transcriptomes of single cells were used for principal-component analysis (PCA), which allow the visualization of the variation present in our dataset. PCA analysis resulted in the formation of 4 groups of single cells spatially separated corresponding to day 0, day 3, day 6 and day 10 (**Figure R.10B**). Thus, the differentiation time point is what make the single cells more different between each other. PCA visually identified groups of day 6 and day 10 single cells seemed to be more dispersed than groups of day 0 and day 3 single cells. In order to quantify this dispersion, the mean distance from the centroid was calculated per each cluster. An increased dispersion of day 6 and day 10 clusters was then confirmed (**Figure R.10C**). This reflect the increased heterogeneity that is expected to occur along the differentiation process, going from more uniform pluripotent cells to a variety of committed and differentiated cells.

Unsupervised clustering was performed (Lafzi et al., 2018) by pooling all time points and the two conditions (G1 and S-G2/M)

to increase statistical power. Clusters were then represented in a low dimensional space produced by t-distributed stochastic neighbor embedding (tSNE). tSNE allow to place neighbor single cells closer, and thus it allows for a better visualization compared to PCA. Clustering and subsequent tSNE representation of the data allowed the identification of 5 different clusters (**Figure R.11A**), based on their transcriptome. Cluster 1 and cluster 2 resulted to be the more similar between each other (**Figure R.11B**). In order to assign cell identities to the clusters, we used previously annotated maker genes (**Figure R.11C**), and checked their expression in the different clusters (**Figure R.11D**). We found that cluster 1 express predominantly pluripotent genes, while cluster 2 pluripotent and lineage-specific imprinted genes. On the other hand, cluster 3, 4 and 5 mainly express lineage-specific imprinted genes together with ectodermal, endodermal and mesodermal genes respectively (**Figure R.11D**). Moreover, we found that most of the cells from cluster 1 correspond to day 0 cells, cells from cluster 2 to day 3 cells and cells from clusters 3-5 correspond to cells from day 6 and 10. Therefore, considering the expression score according to gene markers and the temporal origin of the cells forming the cluster, we assigned cell identities for clusters 1-5, being pluripotent mESCs, early uncommitted progenitors, ectodermal, endodermal and mesodermal cells respectively (**Figure R.11D**).



Marker genes	Pluripotent	Lineage-specific imprinted	Endoermal	Mesodermal	Ectodermal
	Esrrb	*Airn*	*Cldn6*	*Acvr1b*	*Apbb1*
	Nanog	*Dio3*	*Foxa2*	*Cfc1*	*Ncam1*
	Pou5f1 (Oct4)	*Gnas*	*Krt19*	*Lmo2*	*Parva*
	Zfp42 (Rex1)	*H19*	*Sox17*	*Runx1*	*Pitx2*
		Igf2		*Sirpa*	*Sox12*
		Mest			*Sox6*
		Ndn			
		Peg10			
		Peg3			

Figure R.11. Subpopulation identification of single cell EBs cells. (A) tSNE plot representation of the supervised cluster formation including all data sets. **(B)** Cluster analysis tree, showing distances between clusters according to their transcriptome similarities/differences. **(C)** Gene list used to assign cluster identities. Pluripotency, lineage-specific imprinted, endodermal, mesodermal and ectodermal differentiation marker genes are shown. **(D)** Heat map showing gene markers expression levels on the different clusters.

To analyze the identified clusters during the differentiation process in G1 and S-G2/M conditions, we assessed the number of cells per cluster in all the time points from day 0 to 10 for G1 and S-G2/M (**Figure R.12A**). Importantly, at day 3 of differentiation, there are significantly less cells from the pluripotent cluster in G1 compared to S-G2/M condition.

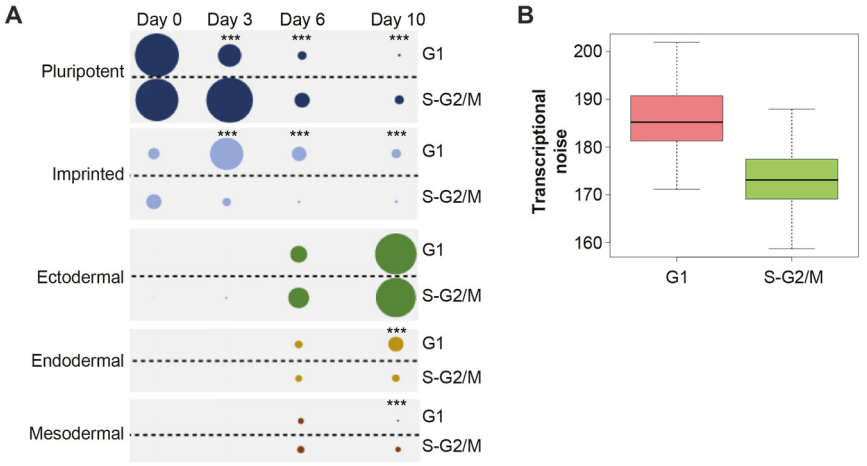


Figure R.12. G1 cells show an accelerated differentiation and pluripotency is segregated during cell cycle. (A) Representation of the number of cells per each cluster during EB formation (day 0-day 10) in G1 and S-G2/M conditions. Circle size represents number of cells. For statistical analysis, a Fisher's exact test was used. "*" means p-value<0.05; "**" p-value<0.01 and "***" p value<0.001. **(B)** Transcriptional noise in G1 and S-G2/M cells at day 0 (mESCs prior to EB differentiation). Transcriptional noise is calculated as the average distance between each cell to rest of the other cells in each time point for G1 and SG2-M cells separately. The distribution of these average distances is plotted in the boxplot.

Additionally, at the same time point, more cells are present in the lineage-specific imprinted cluster in G1 compared to S-G2/M. This indicates that pluripotency maker genes are shutted-down earlier when cells differentiate from G1 compared to S-G2/M, pointing towards a faster differentiation from this cell cycle phase. This was in agreement with previous reports, where G1 was described to be a window of time for mESCs differentiation (Sela et al., 2012; Coronado et al., 2013). Interestingly, transcriptional noise was found to be higher in G1 mESCs compared to S-G2/M (**Figure R.12B**).

Transcriptional noise has been shown to precede cell fate decisions or differentiation (Mohammed et al., 2017; Eldar and Elowitz, 2010). Thus, increased transcriptional noise in G1 cells could explain the rapid pluripotency exit and differentiation.

Surprisingly, at day 10 of differentiation we detected a significantly increased number of cells in endodermal cluster and in mesodermal cluster for EBs originally derived from G1 and S-G2/M cells respectively (**Figure R.12A**). Thus, EBs from G1 cells gave rise to more endodermal cells compared to S-G2/M, and EBs from S-G2/M cells generated more mesodermal cells.

These results show that, as previously shown, cells in G1 are primed for an accelerated differentiation. Notably, we also found that the cell cycle state of mESCs influences the ultimate lineage-commitment in the presence of a similar differentiation context (i.e. cell media). This, together with the rapid recovery of asynchrony on sorted cells (**Figure R.9B**) and the equivalent number of cells per EB after LIF withdrawal (**Figure R.9C**), point to the existence of intrinsic cell fate determinant with differential cell cycle activity.

4. S-G2/M cells are primed for cardiomyocyte differentiation

The previous scRNAseq data was in line with a serendipitous observation of beating phenomenon from forming EBs. The beating phenomenon consist in a constant and synchronous pulsing contraction and relaxion of formed EBs, as a result of the presence of cardiomyocytes, with mesodermal origin, present in the EBs. A manual quantification of number of beating EBs formed from initially G1 and S-G2/M sorted cells (**Figure R.13A**) showed a striking difference with more % of beating EBs derived from S-G2/M cells (**Figure R.13B**).

Consistently, genome wide transcriptome analysis by bulk RNA-seq revealed cardiac/heart categories as the most enriched in S-G2/M coming EBs at day 8 of the differentiation process (**Figure R.13C**). The increased expression of selected cardiomyocyte genes was further validated by real time quantitative PCR (qPCR) (**Figure R.13D**).These increased gene expression could be mainly due to two facts: either cardiac genes are more expressed in EBs from S-G2/M cells or S-G2/M cells give rise to more cardiomyocytes. To discriminate between these two possibilities, ACTC1, a cardiomyocyte marker, staining was performed and analyzed by FACS at single cell level (**Figure R.13E**). FACS data showed an increase presence of cardiac cells in S-G2/M EBs, arguing in favor of the second hypothesis (**Figure R.13.E**).

Overall, these results support the mesodermal differentiation preference observed in S-G2/M cells by scRNA-seq and, in particular, towards cardiomyocyte fate.

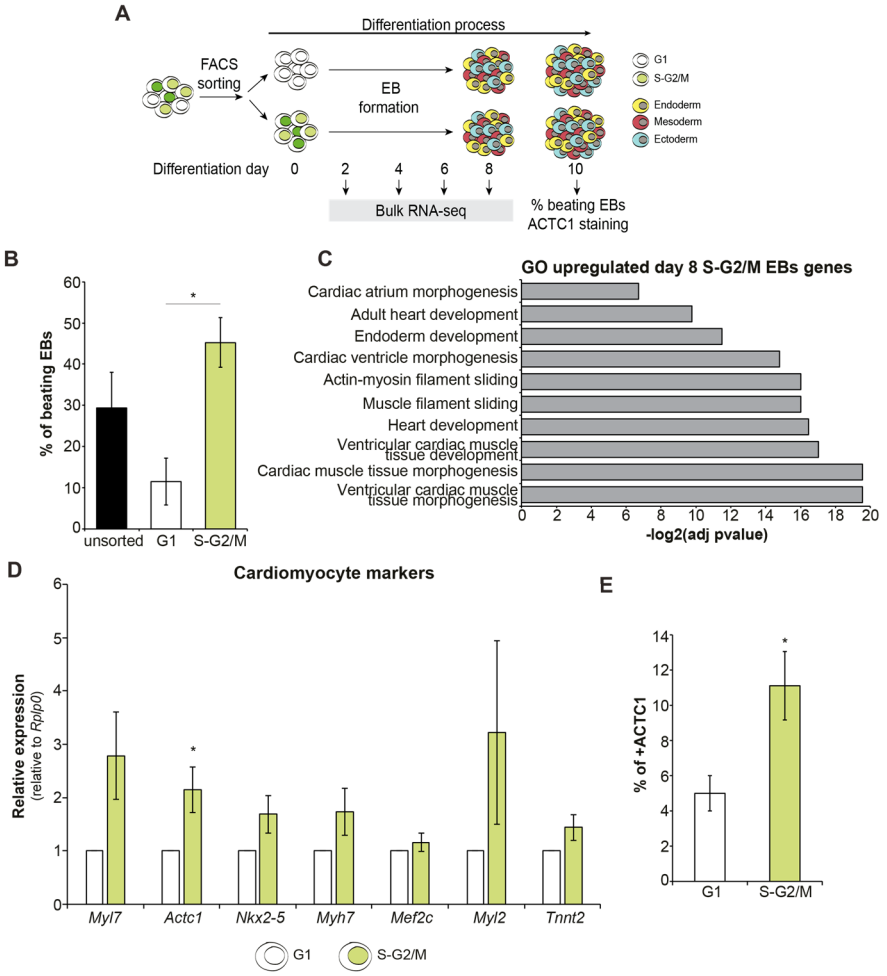


Figure R.13. Embryoid bodies from S-G2/M mESCs show more cardiomyocyte differentiation. (A) Scheme of EB differentiation including the time points when samples were taken for bulk RNA-seq, beating EBs assessment and ACTC1 staining. **(B)** Percentage of beating EB for initial G1 and S-G2/M sorted conditions. An unsorted condition is included as a control. An EB was counted as "beating EB" if any beating area was observed. The average of 3 independent replicates is provided. Error bars indicate +/- SEM. A student T-test was performed to test statistical significance. "*" means p-value<0.05; "**" p-value<0.01 and "***" p value<0.001. **(C)** GO analysis of the up-regulated genes at day 8 of the EB

differentiation for S-G2/M condition. Note cardiac/heart enriched categories. **(D)** qPCR analysis of day 8 EBs from G1 and S-G2/M cells for cardiomyocyte markers. The average of relative expression against the housekeeping gene *Rplp0* of 5 independent replicates is provided. Error bars indicate +/- SEM. A student T-test was performed to test statistical significance. "*" means p-value<0.05; "**" p-value<0.01 and "***" p value<0.001. **(E)** Percentage of positive cells for ACTC1, a cardiomyocyte marker. An average of 3 independent replicates is provided. Error bars represent +/- SEM. A student T-test was performed to test statistical significance. "*" means p-value<0.05; "**" p-value<0.01 and "***" p value<0.001.

Additionally, EBs from G1 sorted mESCs present higher expression of differentiation genes at day 4 of the EBs formation, compared to S-G2/M. This was shown by GO analysis, which revealed categories related to developmental processes as top enriched (**Figure R.14A**). Specifically, early mesodermal differentiation categories were among the most enriched. Some of these genes were further validated by qPCR (**Figure R.14B**). Taken together, these results also support the increased propensity for G1 mESCs to differentiate once differentiation is triggered.

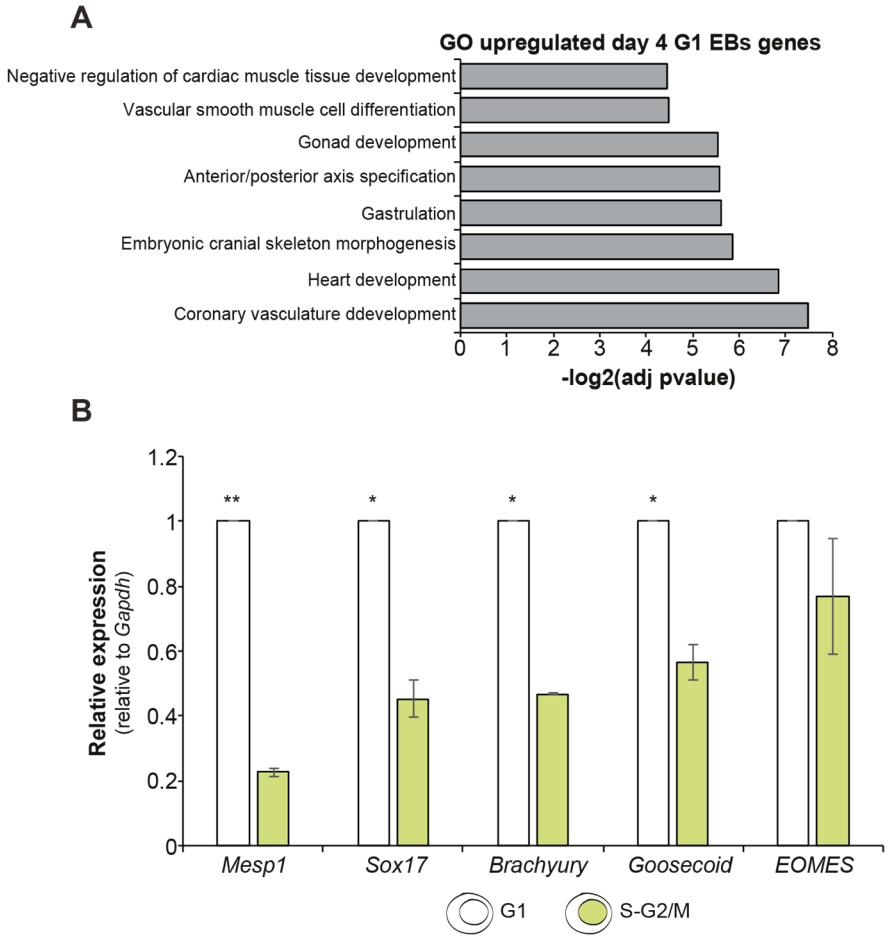


Figure R.14. G1 mESCs exit the pluripotent state earlier compared to S-G2/M. (A) GO analysis of the upregulated genes at day 4 of EB differentiation from G1 condition. Note developmental-related categories. **(B)** qPCR analysis of day 4 EBs from G1 cells for early differentiation markers. The average of relative expression against the housekeeping gene *Gapdh* of 2 independent replicates is provided. Error bars indicate +/- SEM. A student T-test was performed to test statistical significance. "*" means p-value<0.05; "**" p-value<0.01 and "***" p value<0.001.

5. Cell cycle is determinant for *in vivo* cell specification

In order to test whether cell cycle had an impact on pluripotency *in vivo* as well, two different approaches were followed in collaboration with Laura Batlle and Marta Vila from the Tissue Engineering Unit at the Centre for Genomic Regulation (CRG). The first one being a teratoma formation assay and the second one, a chimera formation assay, both of them aimed to evaluate pluripotency *in vivo*.

The teratoma formation assay consist in the injection of ESCs into immunodeficient mice. These cells, if pluripotent, will form teratomas, which are benign tumors containing cells from the three germ layers. Thus, for the teratoma formation assay, 300,000 G1 or S-G2/M cells were injected into each flank of 3 immunodeficient mice per condition (**Figure R.15A**). A total of 6 and 3 teratomas were extracted from mice injected with G1 and S-G2/M mESCs respectively (as one S-G2/M condition mouse had to be sacrificed because of illness, probably due to its immunodeficiency nature). Teratomas were then paraffin embedded for being posteriorly sectioned using a microtome. Sections were then stained with hematoxylin-eosin (H&E) and the percentage of cells from each germ layer (endoderm, mesoderm and ectoderm) per teratoma was estimated by visual inspection of H&E staining and morphology (**Figure R.15B**). No significant differences could be seen in germ layer contribution in G1 and S-G2/M injected cells (**Figure R.15C**).

Lack of correlation with the *in vitro* pluripotency assessment experiments may be due to technical difficulties when estimating the percentage of each germ layer per teratoma, as it was performed by visual inspection based on cell morphology. Cell heterogeneity could also be masking cardiomyocyte cells among all mesodermal cells.

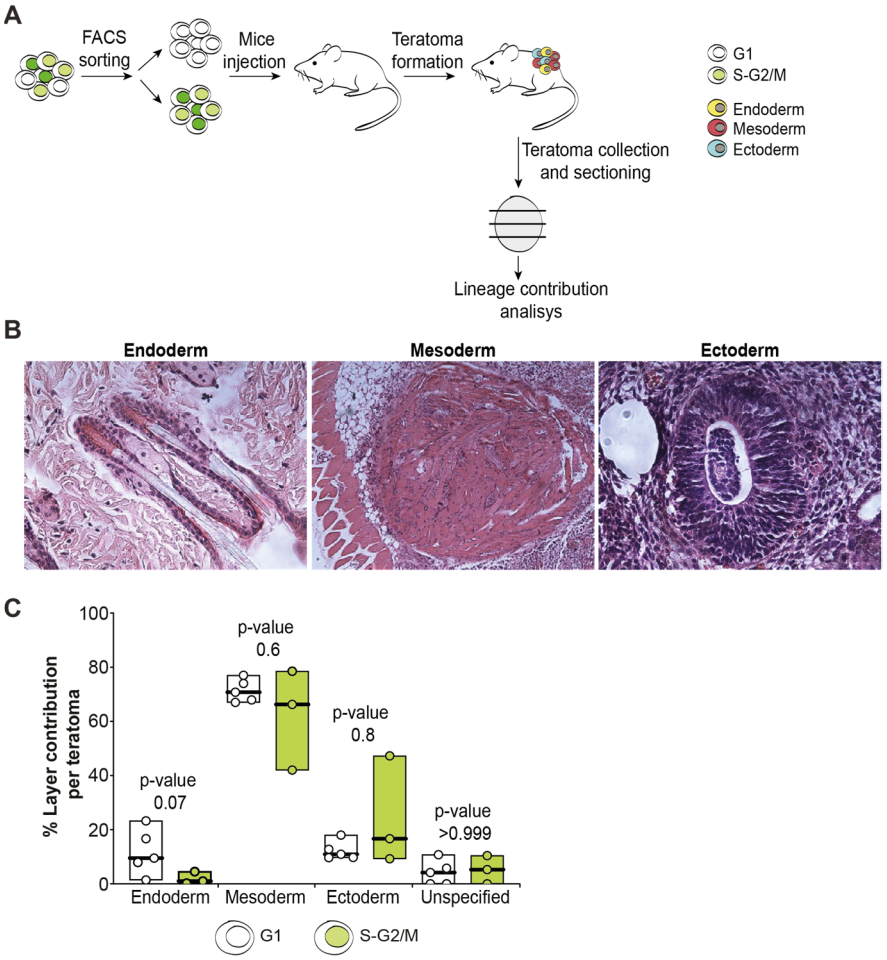


Figure R.15. Teratoma analysis did not show any pluripotency segregation. (A) Scheme showing the experimental steps from FACS sorting to the analysis of lineage contribution per teratoma. **(B)** Histological sections derived from teratomas. (Left) Gut cells from the endodermal lineage, together with mesenchymal cells (from mesoderm). (Middle) Muscle cells, mesenchymal cells and adipose tissue, being the three of

them from mesodermal origin. (Right) Neuroectodermal cells from ectodermal lineage. **(C)** Box plots showing the percentage of endoderm, mesoderm, ectoderm and unspecified contribution per teratoma from G1 and S-G2/M injected cells. 6 and 3 teratomas were analyzed for G1 and S-G2/M conditions respectively. Mann-Whitney test was used for statistical analysis. P-value is indicated in each case.

On the other hand, chimera formation is accomplished by the injection of ESCs into early embryos (blastocyst or earlier embryos). The pluripotent injected cells will contribute to the formation of tissues in the animal, from the three germ layers, forming thus a chimeric organism containing cells genetically different. Cells injected in earlier developmental stages, will contribute more to the adult animal.

Thus, for the chimera contribution assay, in order to be able to track the injected cells and assess its contribution to the chimeric organism, Gem-mESCs were genetically modified using the PiggyBac system to incorporate dsRed, a fluorescent reporter gene expressed constitutively (**Figure R.16**).

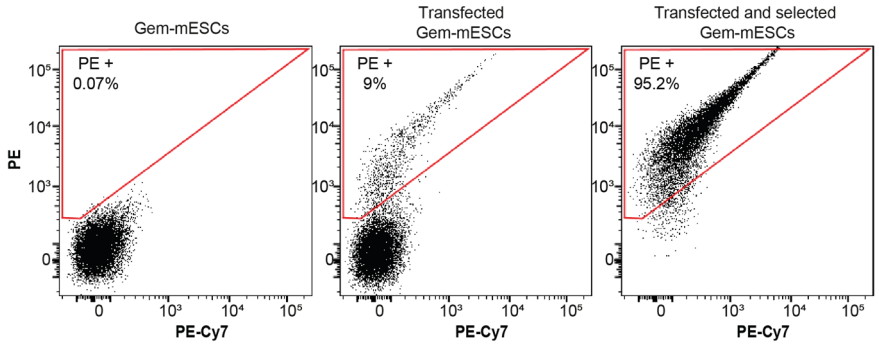


Figure R.16. dsRed-PiggyBac Gem-mESCs establishment. Gem-mESCs (left) were transfected with dsRed PiggyBac (middle). After selection was performed, 95% of the Gem-mESCs were positive for the dsRed reporter gene (right).

PiggyBac systems includes two elements: the PiggyBac transposon and the PiggyBac transposase. The PiggyBac system turns to be a genetic mobile element capable of efficiently transposing via a “copy and paste mechanism” your sequence of interest, being it dsRed in this case. PiggyBac has already been used in mouse, resulting in stable reporter gene expression (Leeb et al., 2012). The dsRed-PiggyBac Gem-mESC cell line was stablished by stable transfection of Gem-mESCs. Cells were then selected, obtaining almost a pure population of dsRed expressing cells (**Figure R.16**).

Once the dsRed-PiggyBac Gem-mESC cell line was established, cells were sorted in dsRed positive and G1 and S-G2/M cells and mouse chimeras were generated to analyze the chimera contribution (**Figure R.17A**). To achieve that, 10-20 mESCs were injected into 15-16 E3.0-3.5 blastocysts, which were then transferred into a pseudopregnant mouse. Embryos were collected at stage E18.5 and organs representing the three germ layers were extracted for posterior FACS analysis in order to assess the chimera contribution of G1 and S-G2/M cells. While placenta and yolk sack were collected as extra-embryonic tissues (tissues mainly formed by cells with primitive endoderm and trophectoderm embryonic origin respectively), heart, liver and brain were collected as tissues mainly formed by cells with mesodermal (*e.g.* cardiomyocytes), endodermal (*e.g.* hepatocytes) and ectodermal (*e.g.* neurons) embryonic origin respectively. The

mentioned organs were extracted from E18.5 embryos and disaggregated mechanically and enzymatically incubating them overnight at 4ºC to obtain single cell suspensions. Contribution to chimeric embryos was assessed by checking the percentage of positive dsRes cells per organ by FACS. No significant differences were observed in the number of viable and chimeric embryos obtained from G1 and S-G2/M injected embryos (**Figure R.17B)**. However, strikingly, G1 injected cells contributed significantly more to yolk sack than S-G2/M (**Figure R.17C**). This suggests that S-G2/M mESCs display a restriction towards primitive endodermal cell commitment, showing the implication of cell cycle in cell fate decisions *in vivo*. Brain was another organ where cells contributed differentially, as S-G2/M cells were more committed towards this mainly ectodermal tissue.

However, we found that cells from G1 and S-G2/M cell cycle phases contribute equally to heart. Thus, the *in vivo* chimera assays seemgly did not recapitulate our *in vitro* EB assays.

Considering that mESCs were injected at the E3.0-3.5 blastocyst stage, we speculate that at the time when the epiblast engages the initial differentiation *in vivo* (E6.5) our initially synchronized cells in G1 or SG2M might display a similar asynchrony. Thus, at the onset of lineage-specification, the two experimental conditions would be equivalent cell cycle-wise. At the light of the results and our interpretation, we would need to inject Gem-mESCs sorted cells at E6.5. However, this

approach entails technical difficulties as E6.5 embryos have already undergone implantation into the mother.

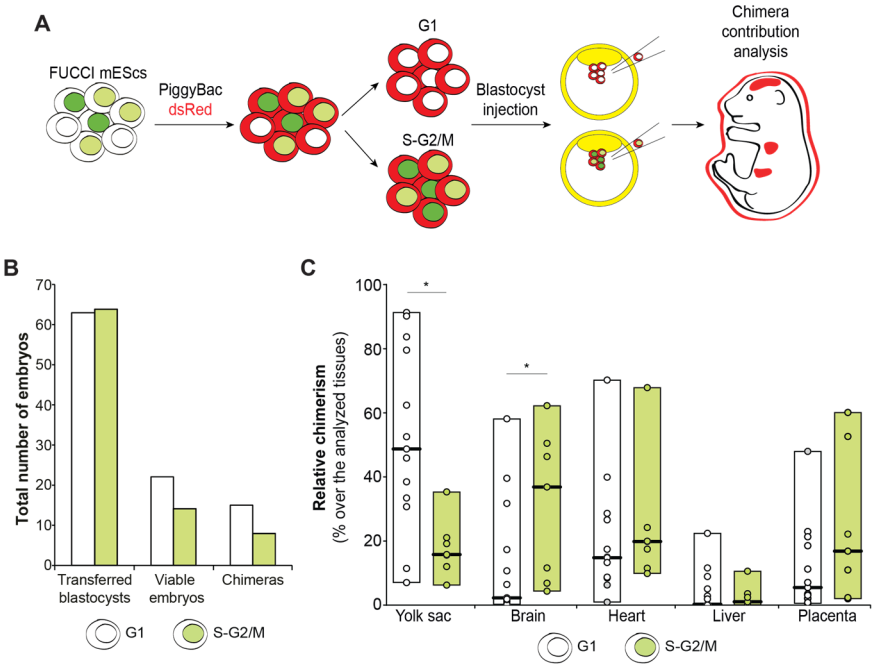


Figure R.17. Pluripotency is segregated during cell cycle *in vivo*. (A) Schematic representation of the followed experimental procedure, from PiggyBac dsRed Gem-mESC cell line establishment to cell injection and chimera contribution analysis. **(B)** Sum of transferrer blastocyst and viable and chimeric E18.5 embryos from G1 and S-G2/M conditions. **(C)** Percentage of contribution over the total analyzed tissues for yolk sac, brain, heart, liver and placenta. Data from 3 independent replicates is shown. Each dot represents data from one organ in one embryo. To test statistical significance, a Mann-Whitney test was performed. "*" means p-value<0.05; "**" p-value<0.01 and "***" p value<0.001.

6. Transcriptomic differences across cell cycle progression

We then focused our attention on characterizing the molecular mechanism that could explain the *in vitro* S-G2/M bias towards cardiomyocyte differentiation. For this purpose, equal number of Gem-mESCs were sorted to G1 and S-G2/M and total RNA was extracted in order to perform full transcriptome analysis by RNA-seq (**Figure R.18A**). As S-G2/M cells show an approximately 30% increase in whole transcriptome compared to G1 due to replication (**Figure R.18B**), spike-in normalization was included in the RNA-seq analysis to correct for the differences in total RNA amount in G1 and S-G2/M samples. Bioinformatics analysis were performed by Enrique Blanco and Mar González from the laboratory.

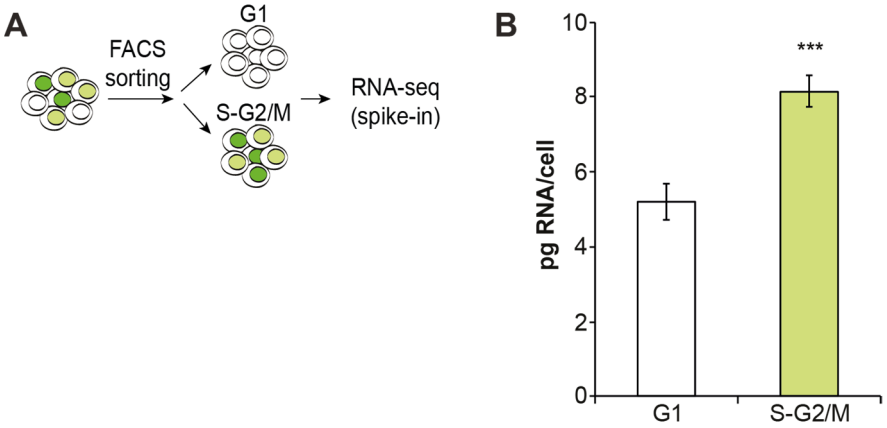


Figure R.18. Bulk RNAs increase in S-G2/M mESCs. (A) Schematic representation of the strategy followed to perform RNA-seq, including spike-in. **(B)** Amount of RNA (in pg) found per mESCs in G1 and S-G2/M cell cycle phases. The average of 2 independent replicates is provided. Error bars indicate +/- SEM. A student T-test was performed to test

statistical significance. “*” means p-value<0.05; “**” p-value<0.01 and “***” p value<0.001.

After spike-in correction, the average of the two independent performed replicates was calculated and differentially expressed genes were identified considering at least 2 fold change difference between replicates and 1 RPKM expression in at least one of the conditions. Thus, analysis from RNA-seq showed a global increase of transcripts in S-G2/M cells compared to G1 (**Figure R.19A**). In particular, 336 were the

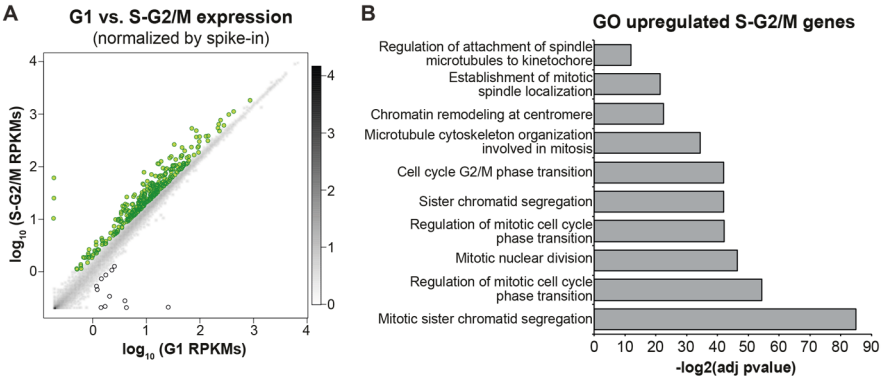


Figure R19. S-G2/M mESCs show a global increased gene transcription levels. (A) Scatter plot showing the differentially expressed genes during the cell cycle (full genome). The average of 2 independent replicates is provided. White and green-colored dots represent significantly upregulated genes in G1 and S-G2/M cells respectively. The legend on the right represents dot density. Darker zones represent higher dot density. **(B)** GO analysis of differentially expressed genes, upregulated in S-G2/M compared to G1. Note typical cell cycle-related GO categories for these cell cycle phases.

genes upregulated in S-G2/M mESCs in front of 12 genes that were upregulated in G1 mESCs. While GO analysis did not show any significantly enriched category in G1 cells due to very low gene number, expected cell cycle-related categories were

the most enriched in S-G2/M cells (**Figure R.19B**). Some of the upregulated genes include *Cdk1*, *Cyclin B1*, *Aurka* and *Cenpa*, which are well-known regulators of cell cycle progression from S to M phases (**Figure R.20**). Moreover, some pluripotency genes, although most of them were not significant, showed differential gene expression, being more expressed in S-G2/M cells. Unexpectedly, also some developmental genes were up-regulated in these cell cycle phases. This upregulation was specific of this subset of genes as they were upregulated with respect to total transcriptome.

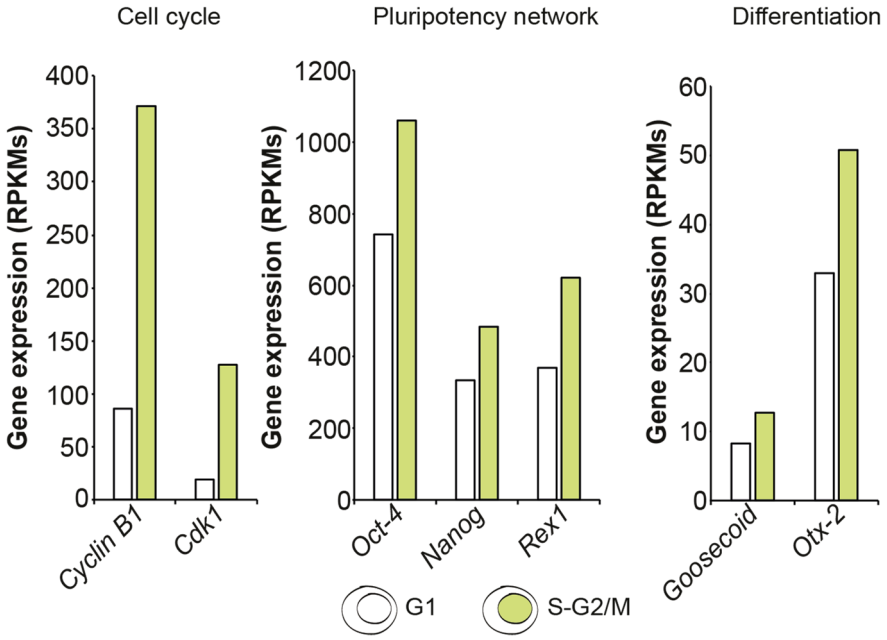


Figure R.20. mESCs show a dynamic gene regulation during cell cycle. Gene expression from RNA-seq experiment showing RPKM values for cell cycle, pluripotency and differentiation markers.

Then, we wanted to validate the observed differential expression during the cell cycle by qPCR. To correct for the

global transcriptomic increased in S-G2/M cells compared to G1 (**Figure R.18B**), we designed a strategy for introducing spike-in correction in qPCR experiments. The strategy take advantage of mESCs and Kc167 *Drosophila melanogaster* cells (**Figure R.21A**).

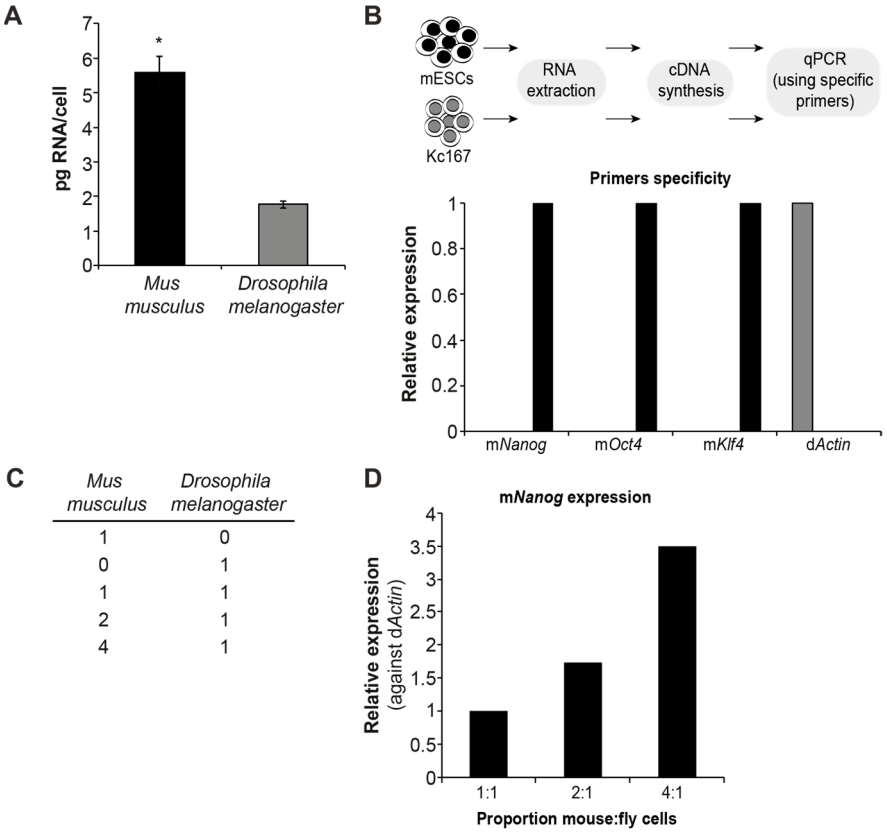


Figure R.21. *Drosophila melanogaster* cells use to perform qPCR experiments correcting by spike-in. (A) Amount of total RNA per cell in mouse and fly cells (*Mus musculus* and *Drosophila melanogaster*). The average of 2 independent replicates is provided. Error bars indicate +/- SEM. A student T-test was performed to test statistical significance. "*" means p-value<0.05; "**" p-value<0.01 and "***" p value<0.001. **(B)** Primers specificity for *Mus musculus* and *Drosophila melanogaster* was tested by checking specific gene expression in mouse and fly cells. **(C)** Proportion of mouse:fly cells used for the analysis. **(D)** *Nanog*, as a mouse specific gene, expression was checked by qPCR in pooled samples of mouse and fly cells. 1:1, 2:1 and 4:1 refers to mouse:fly proportion of cells. The relative expression against d*Actin* is provided.

We first evaluated the amount of RNA per cell in Kc167 *Drosophila* cells and mESCs. mESCs resulted to have 3 times more RNA per cell than *Drosophila* cells (**Figure R.21A**). Then, in order to evaluate the specificity of the primers used, RNA was converted to cDNA and qPCR was performed using specific primers for mouse and for *Drosophila* cells (**Figure R.21B**). We validated that primers used for *Nanog*, *Oct4* and *Klf4* were specific for mouse genes, while d*Actin* primer was specific for the fly gene. Once the primer specificity was tested, we mixed mESCs and Kc167 *Drosophila* cells in different proportions (**Figure R.21C**) to evaluate the accuracy of our approach on quantifying the abundance of a given transcript (*e.g.* mouse *Nanog*) normalizing against the specific *Drosophila Actin* gene. The relative expression of mouse *Nanog* in mESCs normalized against *Drosophila Actin* was directly proportional to the initial amount of mESCs in the mix (**Figure R.21D**). Thus, we verified our experimental strategy for qPCR analysis including *Drosophila* cells spike-in.

Then, to validate the observed differentially expressed genes during cell cycle (**Figure R.20**) by qPCR, equal number of sorted G1 and S-G2/M cells were pooled with Kc167 *Drosophila* cells (in 4:1 proportion of mESCs:*Drosophila* cells (**Figure R.22A**). RNA was then extracted from the pooled cells and gene expression was checked by qPCR. Quantification of RNA expression was normalized by the *Drosophila Actin* gene expression values. Thus, genes changing their expression

during cell cycle progression were validated by using the explained strategy (**Figure R.22B**).

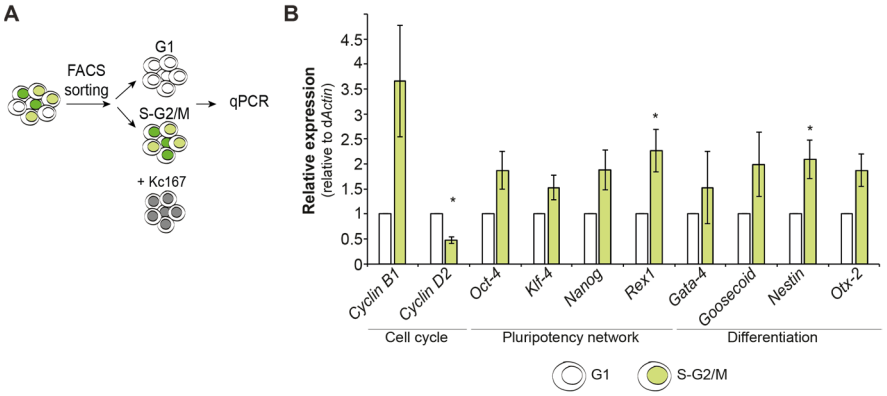


Figure R.22. Validation of the mESCs dynamic gene regulation during cell cycle. (A) Schematic representation of the followed procedure: Gem-mESCs were sorted to G1 and S-G2/M cells and mixed with *Drosophila* cells. Then, RNA was extracted from the pooled samples and gene expression was checked by qPCR. **(B)** qPCR analysis of cell cycle, pluripotency and differentiation marker genes. The average of relative expression against the *Drosophila Actin* of 3 independent replicates is provided. Error bars indicate +/- SEM. A student T-test was performed to test statistical significance. "*" means p-value<0.05; "**" p-value<0.01 and "***" p value<0.001.

However, none of the differentially expressed genes could explain the cardiomyocyte enhanced phenotype observed *in vitro*. Thus, transcriptome analysis did not shed light into the molecular characterization of the shown pluripotency segregation of mESCs during cell cycle.

7. Identification of differentially expressed proteins during the cell cycle

As transcriptome analysis could not identify the molecular mechanism for the S-G2/M preference for cardiomyocyte differentiation, we focused on interrogating the proteome dynamism during cell cycle with the same objective. For this purpose, the same amount of Gem-mESCs were sorted to G1 and S-G2/M, from which total amount of protein was extracted and processed for shotgun proteomics (**Figure R.23A**). Proteomics and the posterior analysis were done in collaboration with Eduard Sabido and Eva Borras from the Universitat Pompeu Fabra (UPF) and CRG Proteomics Unit. From shotgun proteomic analysis, protein fold-changes were quantified between samples considering the three most abundant peptides per protein. A total of 47,522 unique peptides were identified, corresponding to 6,453 proteins. 3,318 proteins resulted to be significantly differentially expressed during cell cycle by using an ANOVA model (FDR<0.05). Similarly to what was observed transcriptomically (**Figure R.18B**), the global proteome of S-G2/M cells showed higher expression values compared to G1 cells (**Figure R.23B, C**). This goes in line with the global increase in cell volume, RNA and DNA content observed in S-G2/M cells compared to G1. Thus, indicating that in ESCs, similarly to other cell types (Padovan-Merhar et al., 2015), there is a tight connection between the cell volume and the mount of major molecular cell

components. Specifically, 2,916 proteins were significantly more expressed in S-G2/M cells, against 402 proteins in G1 (**Figure R.23C**).

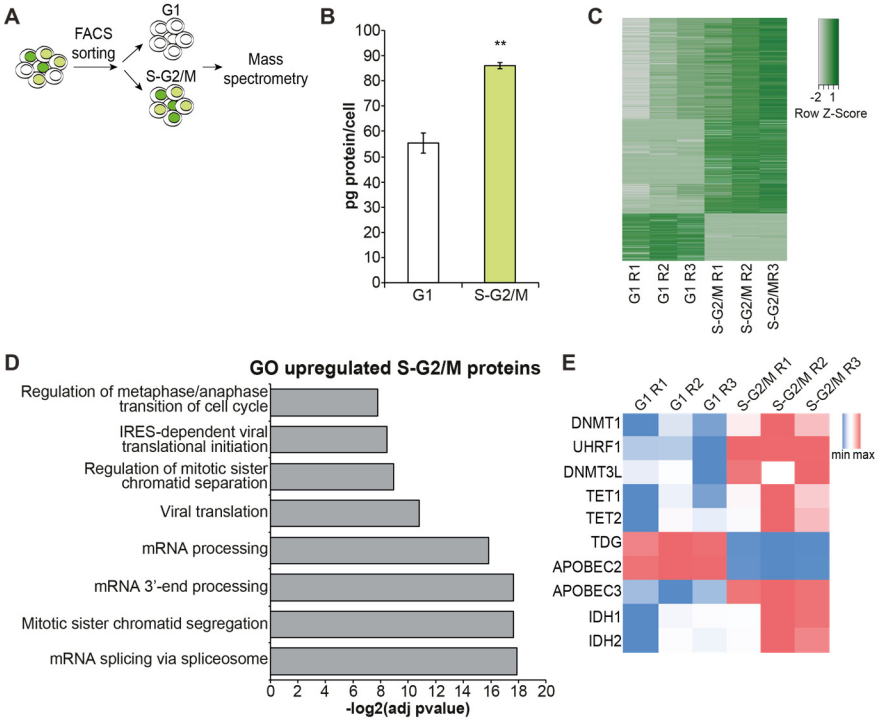


Figure R.23. Proteome is dynamically regulated during cell cycle. (A) Schematic representation of the experimental strategy followed to analyze protein expression during cell cycle. **(B)** Protein abundance per G1 and S-G2/M cells. The average of 4 independent replicates is provided. Error bars indicate +/- SEM. A student T-test was performed to test statistical significance. "*" means p-value<0.05; "**" p-value<0.01 and "***" p value<0.001. **(C)** Heat map showing the differentially expressed proteins during the cell cycle (FDR<0.05). 3 independent replicates are shown as R1, R2 and R3 for each condition. **(D)** GO Biological Process analysis of upregulated S-G2/M compared to G1 mESCs. **(E)** DNA methylation/demethylation proteins expression in G1 and S-G2/M cells. 3 independent replicates are provided. The color-code goes from blue to red, indicating minimal and maximal protein abundance respectively.

The fact that we could observe more proteins enriched in G1 cells compared to expressed genes may be due to post-translational protein regulation.

GO term enrichment analysis of differentially expressed proteins during cell cycle indicated a significant enrichment in functional categories related with cell cycle progression and RNA processing during S-G2/M, as expected. (**Figure R.23D**). However, we did not find any significative increased in any functional GO term for the proteins displaying an increased abundance in G1 cells. However, manual analysis indicate that DNA methylation/demethylation involved proteins were dynamically expressed during the cell cycle progression (**Figure R.23E**). This includes DNMTs, such as the *de novo* methyltransferase DNMT1 and its recruiter protein UHRF1, TET proteins and TDG. Additionally, APOBEC proteins (involved in the demethylation process though deamination) and IDH proteins were also found to be cell cycle regulated (**Figure R.23E**).

TDG was one of the enriched G1 proteins most significantly changing its expression during cell cycle not only among DNA methylation proteins but also among all proteins (being in the 115 top enriched proteins in G1). The TDG upregulation in G1 has also been described in human HeLa cells, as a mechanism to functionally separate DNA glycosylases activity during cell cycle (Hardeland et al., 2007). TDG has also been described to be important during development, as its depletion causes embryonic death at around E11.5 embryonic stage due

to internal hemorrhage (Cortázar et al., 2011; Cortellino et al., 2011). Among other abnormalities, dead embryos were described to present defects on the developing heart (Cortellino et al., 2011). As we observed a cardiac differentiation preference during cell cycle progression and as TDG is significantly more expressed in G1 cells, we speculated that TDG could be involved in the molecular mechanism explaining this cardiomyocyte differentiation priming. Thus, we focused on studying TDG for its possible involvement in the cardiac differentiation priming observed in S-G2/M mESCs.

To confirm the dynamics of TDG expression during cell cycle in mESCs, we performed Western Blot analysis with G1 and S-G2/M sorted samples using a specific antibody against TDG. *Tdg* encodes for three different isoforms, being isoform number 2 the one encoding for a longest variant. Isoform 1 contains a shorter N-terminus and isoform 3 is one aminoacid shorter compared to isoform 2. All the isoforms are equally expressed in mESCs and display the same dynamics being more expressed in G1 than in S-G2/M cells (**Figure R.24A** top). Moreover, TDG overexpression in G1 cells was validated in sorted Gem-mESCs by immunofluorescence (**Figure R.24B**). Not only in Gem-mESCs (being R1 its parental cell line), but TDG overexpression in G1 phase was also validated in a different ES cell line (E14Tg2A) by Western Blot (**Figure R.24A** bottom). In this case, E14Tg2A cells were fixed using paraformaldehyde and their DNA was stained by DRAQ5. Then, cells were sorted according to their DNA content in G1

and S-G2/M and total protein was extracted for posterior Western Blot analysis.

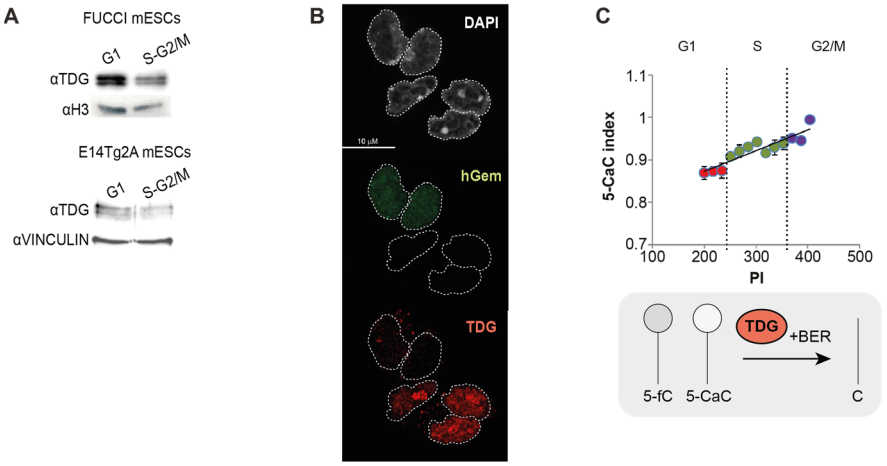


Figure R.24. TDG is overexpressed in G1 cell cycle phase. (A) WB validation of TDG G1 overexpression in FUCCI mESCs (top) and in E14Tg2A (bottom). H3 and VINCULIN were used as loading controls. Protein extraction was performed from equal number of sorted cells. **(B)** TDG immunostaining in FUCCI mESCs. In white, DAPI stained nuclei are shown (top). In green, mAG1-hGeminin can be observed (middle) and in red, TDG expression. Note that negative cells for green fluorescence (cells in G1) express more TDG protein. **(C)** (Top) 5-CaC DNA modification abundance analysis during cell cycle. A 5-CaC specific antibody was used for staining and DNA quantity was assessed by PI staining. Signal intensity was measured by FACS. 5-CaC index refers to 5-CaC intensity measured by FACS divided by DNA content. The average of 3 independent replicates is provided. Error bars indicate +/- SEM. (Bottom) Scheme representing TDG substrates and product.

In order to check if the increased in TDG found in G1 cells is also accompanied by an increased TDG activity, 5-CaC levels, as a TDG substrate, were measured. For this purpose, mESCs were disaggregated, fixed and stained for 5-CaC, using a specific antibody, and for DNA content using PI. In line with the dynamic expression of TDG during cell cycle, 5-CaC levels were shown to be increased in S-G2/M cells (**Figure R.24C**).

This result indicates that TDG activity is differentially distributed during cell cycle, being higher in G1 phase.

For all of these, we considered TDG as our candidate for the pluripotent segregation observed during cell cycle in mESCs *in vitro*.

8. TDG determines cell cycle priming for cardiomyocyte differentiation

In order to experimentally simulate the transitory depletion of TDG in S-G2/M cells and to assess if TDG is the molecular determinant for the observed pluripotency segregation during cell cycle, we transfected asynchronous mESCs with small interferent RNA (siRNA) against *Tdg* (**Figure R25A**). siRNA transfection results in a transient depletion of TDG from 48 to 96 hours, being the protein fully re-expressed at 120 hours (**Figure R25B**). Once mESCs have been depleted for TDG for 48 hours (from 48 to 96 hours post transfection), EB formation protocol was started (**Figure R.25A**, **B**). At day 10 of differentiation, the percentage of beating EBs as well as the percentage of positive cells for ACTC1 staining were measured to monitor cardiomyocyte differentiation (**Figure R.25A**).

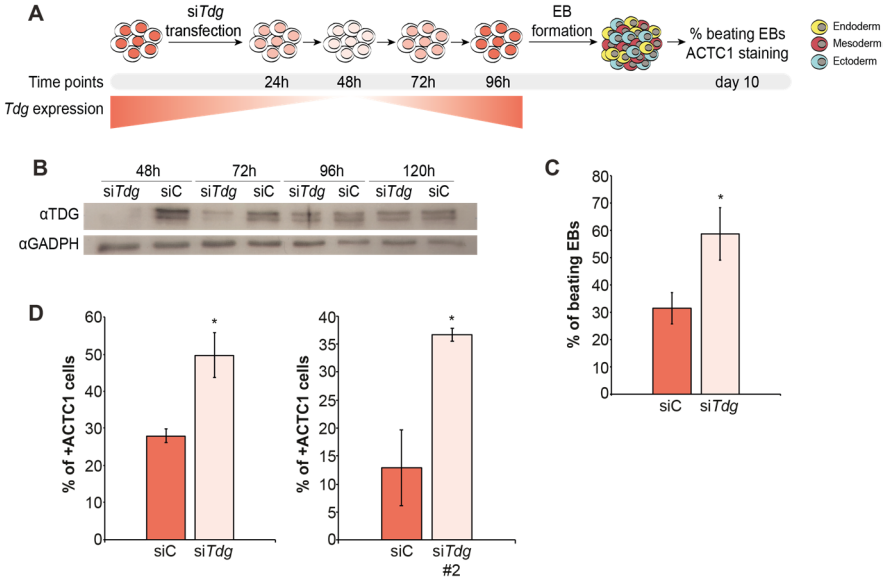


Figure R.25. Transiently depleted *Tdg* mESCs are primed for cardiac differentiation. (A) Schematic representation of the experimental steps. mESCs were transfected with an siRNA against *Tdg*. At 96 hours post-transfection, when mESCs have sensed for longer time the KD effect, EB differentiation was initiated. 10 days later, the percentage of beating EBs and ACTC1 positive cells were assessed. **(B)** TDG expression in siC and si*Tdg* cells at different time points (48h, 76h, 96h and 120h) after siRNA transfection. GAPDH is included as loading control. **(C)** Percentage of beating EBs for si*Tdg* and siC transfected cells. The average of 6 independent replicates is provided. Error bars indicate +/- SEM. A student T-test was performed to test statistical significance. "*" means p-value<0.05; "**" p-value<0.01 and "***" p value<0.001. **(D)** Percentage of positive cells for ACTC1 cardiomyocyte marker in siC and si*Tdg* transfected cells. Data produced using a second siRNA against *Tdg* (*siTdg #2*) is shown as well. An average of 3 independent replicates is provided for both si*Tdg*. Error bars represent +/- SEM. A student T-test was performed to test statistical significance. "*" means p-value<0.05; "**" p-value<0.01 and "***" p value<0.001.

Transiently *Tdg* depleted mESCs produced an increased percentage of beating EBs (**Figure R.25C**) and ACTC1 positive cells (**Figure R.25D**). These results indicate that *Tdg* transiently depleted cells are more primed for cardiac differentiation compared to the control. As an independent validation, we used a secondary siRNA against *Tdg* with a different seed sequence. We found a similar increase in the number of positive ACTC1 cells in EBs at day 10 of differentiation (**Figure R.25D**, right), supporting the specific action of TDG on cardiomyocyte specification.

Additionally, RNA-seq performed on EBs from si*Tdg* and control cells showed an increased expression of some key cardiomyocyte gene markers (**Figure R.26A**). The analysis of the increased expression on cardiac lineage markers was extended and further verified by qPCR (**Figure R.26B**). Thus, si*Tdg* transiently transfected mESCs mimics S-G2/M, being

both more susceptible to differentiate towards the cardiac lineage.

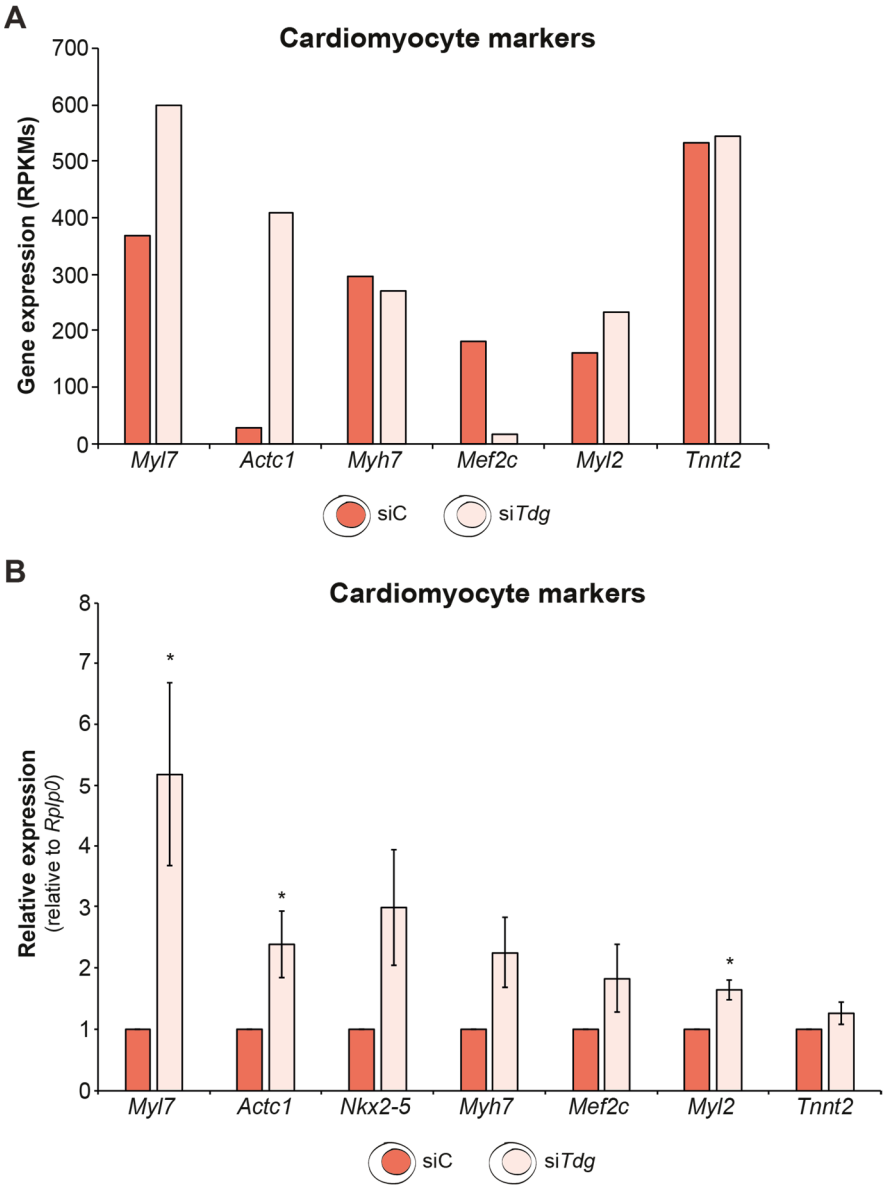


Figure R.26. EBs from transiently depleted *Tdg* mESCs showed increased expression of cardiomyocyte markers. (A) Gene expression from RNA-seq experiment showing RPKM values for key cardiac markers. **(B)** qPCR analysis of cardiomyocyte markers. The average relative expression against the housekeeping gene *Rplp0* of 4 independent replicates is provided. Error bars indicate +/- SEM. A student T-test was

performed to test statistical significance. “*” means p-value<0.05; “**” p-value<0.01 and “***” p value<0.001.

However, TDG transient depletion does not affect mESCs nature. This was shown as transiently *Tdg* depleted mESCs had no morphological (**Figure R.27A**), cell cycle structure (**Figure R.27B**) or transcriptomic (**Figure R.27C, D**) differences compared to control cells. Cell morphology was checked using conventional microscopy and the characteristic dome-shaped colonies were observed in transfected mESCs for both siControl and si*Tdg* (**Figure R.27A**). Cell cycle structure was assessed by fixing the cells at 48 and 96 hours post-transfection and staining them with PI, and no differences were observed between conditions (**Figure R.27B**). Finally, in order to analyze a potential impact of TDG on transcription, we analyzed the whole transcriptome by RNA-seq on 48 and 96 hours post-transfection samples. Comparative analysis of siControl and *siTdg* transcriptome showed that few genes (54 at 48h and 26 at 96h) were differentially expressed by more than 2 fold-change. This indicates that TDG has minor impact on gene expression in mESCs.

Moreover, in order to discard that the effect on beating EBs is due to a difference on EB cellularity, the number of cells per EB was measured in siControl and *siTdg* transfected cells. No differences in cell number were observed between EBs from siC and si*Tdg* transfected cells (**Figure R.27E**), indicating that

the increased cardiomyocyte differentiation was independent of the cell number present per EB.

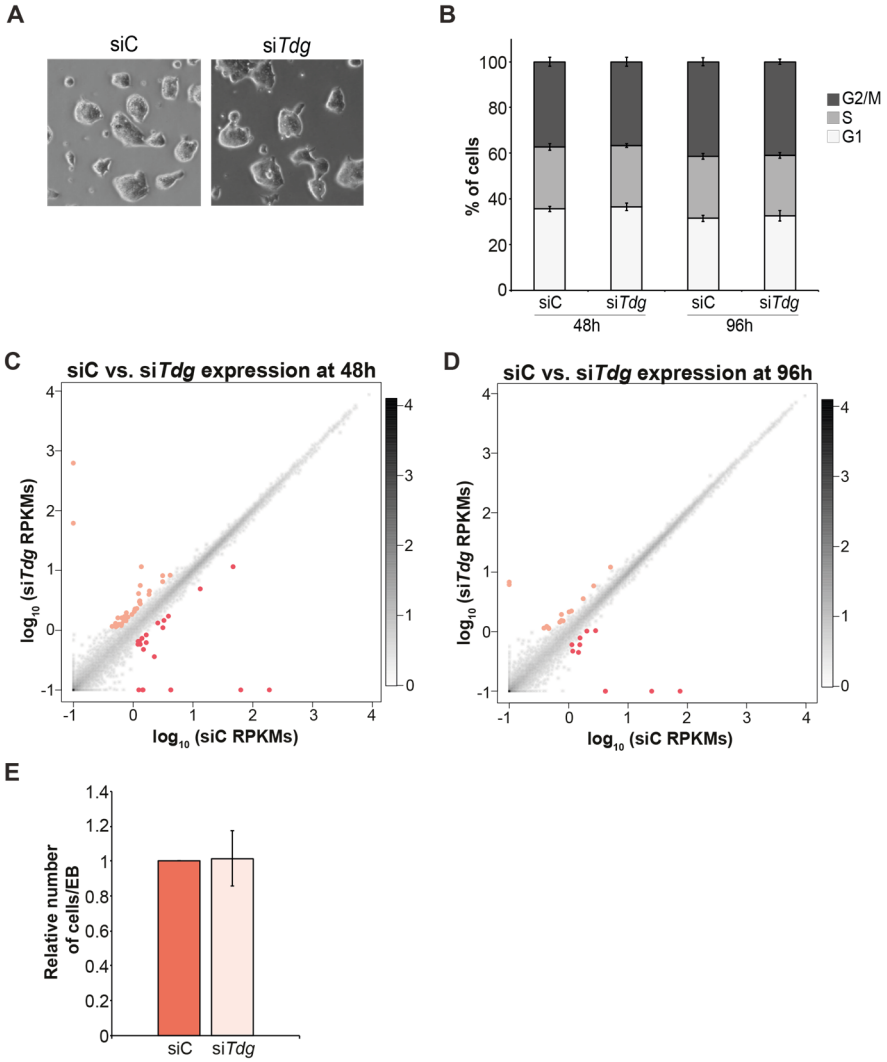


Figure R.27. Cardiomyocyte priming is independent of *Tdg* depletion effect in mESCs. (A) mESC morphology in siC and si*Tdg* transfected cells, 48h after transfection. **(B)** Percentage of cells in G1, S or G2/M from siC and si*Tdg* transfected cells at 48h and 96h post-transfection. Cell cycle structure was determined by PI staining. The average of 4 independent replicates is provided and error bars indicate +/- SEM. **(C)** Scatter plot showing the differentially expressed genes in siC si*Tdg* 48h post-transcription (full genome). The average of 2 independent replicates is provided. Dark and light pink-colored dots represent significantly upregulated genes in siC and si*Tdg* cells respectively. The legend on the

right represents dot density. Darker zones represent higher dot density. **(D)** Scatter plot showing the differentially expressed genes in siC si*Tdg* 96h post-transcription (full genome). The average of 2 independent replicates is provided. Dark and light pink-colored dots represent significantly upregulated genes in siC and si*Tdg* cells respectively. The legend on the right represents dot density. **(E)** Relative number of cells per EB at day 4 of differentiation. 4 independent replicates are shown. Error bars indicate +/- SEM.

Taken together, these results suggest TDG, a cell cycle-dynamically expressed protein, as the molecular determinant for pluripotency segregation, priming in its absence mESCs for cardiomyocyte differentiation upon differentiation cues.

9. TDG together with its partner TP53 act as molecular determinants for cell cycle pluripotency segregation

To deepen into the molecular mechanisms by which TDG regulates lineage commitment in mESCs, TDG genome-wide occupancy was determined by ChIP-seq. ChIP-seq allow the identification of genome-wide protein binding sites by using specific antibodies. Therefore, by performing ChIP-seq for TDG we would get insights into its role in gene regulation. To identify regions significantly occupied by TDG, we performed peak analysis using MACS2. We identified a total of 1822 regions occupied by TDG (called peaks). We found a significant occupancy on regulatory regions of genes (distal, proximal, 5' and 3' UTR) and inside gene bodies (CDS) (**Figure R.28A, B**). In order to identify potential TDG regulated genes, we assigned the 1,822 regions to 1,607 closest genes.

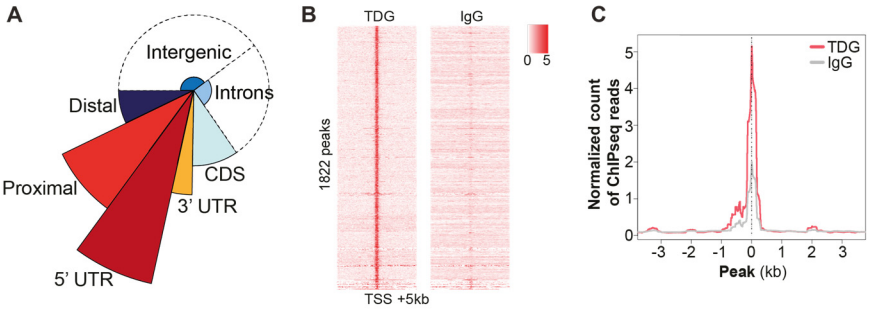


Figure R.28. TDG peaks are located around TSSs. (A) Superimposed pie chart representing the TDG peaks distribution in intergenic, introns, CDS, 3' UTR, 5'UTR, proximal and distal regions. The genome-wide distribution of each genomic category has been considered. **(B)** Heat map from TDG peaks localization centered to the peak summit (± 5kb). IgG has

been included as a negative control. **(C)** Metagene plot showing TDG peaks ±3Kb around the TSS. TDG is indicated in pink and IgG control, in grey.

GO enrichment analysis of TDG peaks was performed in order to identified associated categories. Using the ChEA library (including data from published TF ChIP-seq data) revealed a close relationship between TDG and the transcription factor TP53 target genes (**Figure R.29A**). Moreover, motif analysis using MEME-ChIP revealed TP53 as the most significant motif for TDG peaks (**Figure R.29B**). Furthermore, co-occupancy analysis with other 200 different factors in ESCs positioned again TP53 as the most significant sharing TDG target genes (**Figure R.29C**).

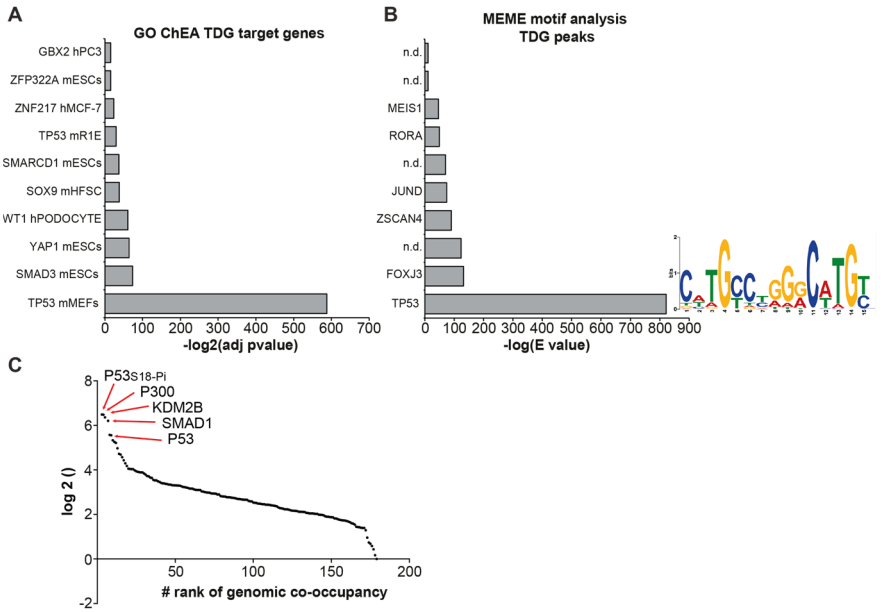


Figure R.29. TDG and TP53 co-occupy target sites genome-wide. (A) GO analysis using ChEA library, containing ChIP-seq data from TF already published. **(B)** Motif enrichment analysis for TDG peaks. Note TP53 DNA motif shows as the most significantly enriched. DNA motif is provided next to TP53 category. **(C)** Percentage of TDG co-occupancy with chromatin factors regulating ESCs identity. The coordinates of TDG peaks were

compared against available data of the genomics coordinates enriched for 179 factors in mESCs, using the BinDB software (Livyatan et al., 2015).

These data indicate a co-occupancy between TDG and TP53. In fact, TDG was previously described to bind TP53 on human lung cancer cell line H1299, modulating its activity by being a transcriptional co-activator (Kim and Um, 2008).

In line with this, target gene expression of si*Tdg* and control cells point TDG as a gene activator, as 341 of its target genes get downregulated upon its downregulation at 96h by gene set enrichment analysis (GSEA) (**Figure R.30.A**). The difference between 341 downregulated target genes and the reduced number of prior identified differentially expressed genes (26 in total) between siC and si*Tdg s*amples 96h post-transfection (**Figure R.27C,D**), is due to the criteria used to determine differentially expressed genes. For differentially expressed genes analysis, 2 fold-change of difference between samples and at least 1 RPKM of expression in one of them was applied. However, GSEA does not use any cut-off, but ranks the genes according to their difference in expression between samples and defines if a prior given subset of genes is significantly different.

GO analysis of the downregulated 341 TDG target genes after *Tdg* KD, showed developmental-related categories among the most enriched ones, such as axogenesis or dendritic spine morphogenesis (**Figure R.30B**). Thus, TDG together with TP53 may act activating transcriptional programs that would

facilitate early differentiation priming. They might also interfere on regulatory networks for later blockade of cardiac lineage specification observed in G1 cells.

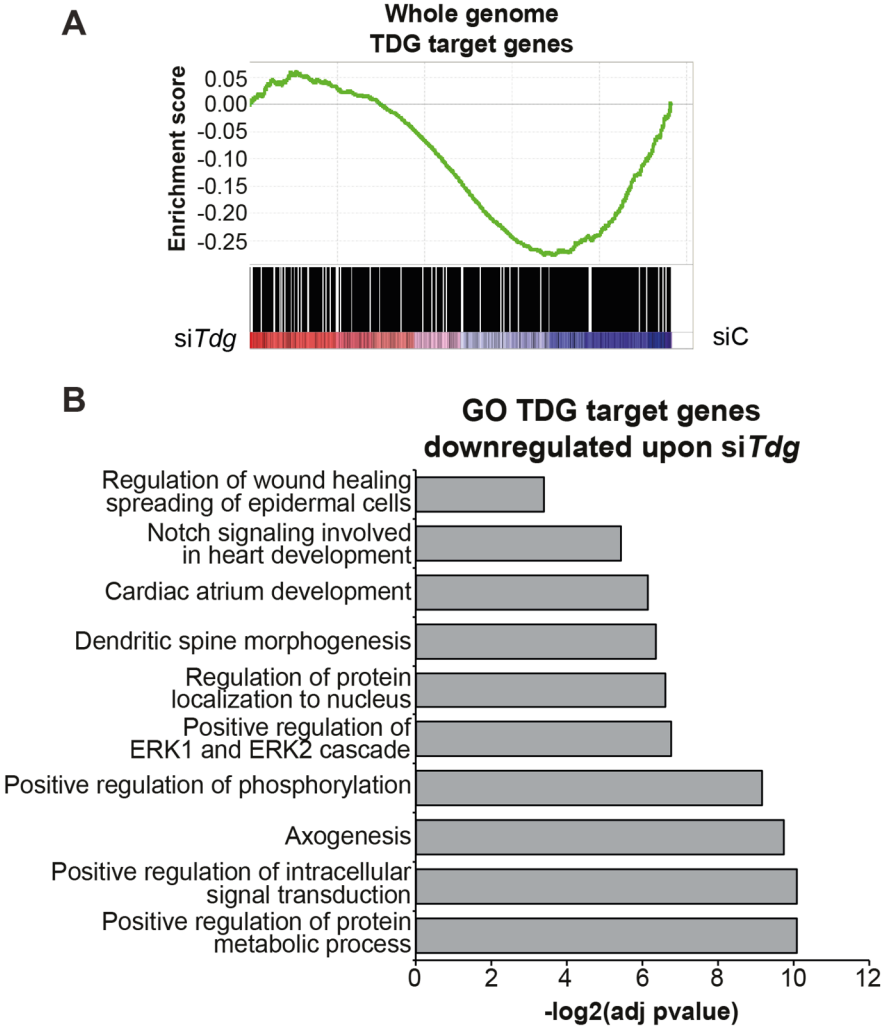


Figure R.30. TDG is a gene activator in mESCs. (A) Gene set enrichment analysis (GSEA) plot for RNA-seq data from siC and si*Tdg* transfected cells at 96h post-transfection. TDG target genes were chosen as gene set and expression data of siC and siTdg cells was ranked according to gene expression. si*Tdg* condition is positioned to the left, while siC, to the right. **(B)** GO analysis of the TDG target 341 genes that are more affected by *Tdg* transient depletion.

DISCUSSION

More than 30 years have passed since cell cycle and pluripotency were linked together. Back then, EC cells were proposed to initiate cell differentiation from G1 cell cycle phase (Mummery et al., 1987a; b). However, only recently new implemented tools have allowed to address the question of how cell cycle controls pluripotency. The FUCCI system has provided a great advance in the field as it allows to monitor cell cycle progression in living ESCs as well as to separate them efficiently by cell cycle phases using FACS sorting. Importantly, this neither generate perturbations in the cell cycle machinery nor produce cytotoxic effects. Thus, the FUCCI tool has permitted to direct the firsts studies about the molecular mechanisms of the cell cycle pluripotency-exit link in hESCs (Pauklin and Vallier, 2013; Pauklin et al., 2016; Singh et al., 2013, 2015). However, these studies did not reach the same conclusion regarding the mechanism behind cell cycle and lineage commitment, although they could be complementary. Ones pointed directly the cell cycle machinery (precisely Cyclin D1) as the determinant for G1 pluripotency exit and described how pluripotency is segregated through the cell cycle via Cyclin D1 (Pauklin and Vallier, 2013; Pauklin et al., 2016). Yet, others described epigenetic mechanisms as the cause for the G1 pluripotency exit. Specifically, bivalent domains (Singh et al., 2015; Grandy et al., 2015) and 5-hmC DNA modification (Singh et al., 2013) establishment during the cell cycle with the corresponding changes in gene expression were proposed to be the cause. Nonetheless, these studies showed

discrepancies on how bivalency is dynamically regulated during cell cycle progression (Singh et al., 2015; Grandy et al., 2015) and could not explain the G1 pluripotency segregation previously observed (Pauklin and Vallier, 2013). Thus, further studies will need to be performed to shed light on the molecular mechanism involved in this area of research. Moreover, all these studies were performed using hESCs as a cell model. Thus, an extra level of complexity is added when considering other species such as mESCs, which differ from hESCs in many aspects cell cycle-wise. Importantly, as metastable mESCs cultured in serum/LIF do not show cell cycle phase specificity on the CDK activity (Stead et al., 2002), the Cyclin D1 model is automatically not valid for mESCs growing in serum/LIF. An important aspect to be considered is the fact that different features exist in the chromatin composition and transcriptional programs depending on the culture medium used. While 2i/LIF cultured mESCs resemble naïve ICM cells of E3.5 blastocyst embryos and serum/LIF cultured mESCs are in metastable state resembling both naïve ICM cells and primed EpiSCs, hESCs mirror primed EpiSCs from post-implantational embryos. Thus, comparing mESCs and hESCs is not yet accurate as they do not share the same *in vivo* counterparts and the cell cycle mechanisms for lineage specification initiation may also differ between naïve and primed cells. To compare such mechanisms between both species, more experiments will be needed to identify a

consensus cell culture medium supporting the naïve state of hESCs.
All these questions, and others, are still under debate in the field and will have to be addressed in future studies.

In this thesis we have focused on the investigation of the cell cycle pluripotency exit in naïve 2i/LIF cultured mESCs and its molecular characterization. For this purpose, we have explored the segregation of the two main characteristics of ESCs (self-renewal and pluripotency capacity) during cell cycle progression. We have shown that while cell cycle does not impact on self-renewal, it does on pluripotency, being S-G2/M cells primed for cardiomyocyte differentiation. In order to examine the molecular mechanisms behind this segregation, we have characterized the transcriptome and the proteome of mESCs during the cell cycle. From this analysis we have placed TDG, which is more expressed in G1 phase, as our candidate determinant for the observed segregation of pluripotency. We have shown how, by transiently depleting *Tdg*, we are able to recapitulate the S-G2/M cardiomyocyte differentiation priming. ChIP-seq experiment revealed TP53 as a possible partner for TDG, which would act together shaping lineage commitment transcriptional networks.

Gem-mESCs as a model to study cell cycle

The FUCCI system, published in 2008 by Sakaue-Sawano and co-workers (Sakaue-Sawano et al., 2008) has supposed a revolution on the molecular study of cell cycle progression. It has allowed to bypass the use of chemical drugs, which interfere with cell cycle machinery and produces cytotoxic effects, often leading to spontaneous cell differentiation. Moreover, the FUCCI system permits cell sorting thanks to its cell cycle phase-specific fluorescence. It also overcomes the inconveniences of other techniques, such as the set up needed for the use of the centrifugal elutriation technique, for which possible apoptotic effects has not been discarded. Overall, the FUCCI tool represents the most attractive way for studying the molecular aspects of cell cycle. For all these reasons we have chosen our model system based on FUCCI. We have used an already published mESCs cell line (Aranda et al., 2014), which we called Gem-mESCs, constitutively expressing the cell cycle regulated GEMININ protein fused to the fluorescent protein AzamiGreen. Thus, cells in G1 cell cycle phase do not show any green fluorescence, while S-G2/M cells do. Therefore, allowing the physical separation of G1 and S-G2/M by FACS sorting. We have cultured Gem-mESCs in 2i/LIF conditions as cells cultured in the presence of these two inhibitors and LIF have been proved to resemble more to the *in vivo* ICM cells from blastocyst embryos (Kolodziejczyk et al., 2015). For further characterization of the cell line, transcriptomic analysis

were performed, comparing Gem-mESCs to the parental mESCs cell line R1. Moreover, validations were also performed for sorted cells to ensure correct cell cycle phases purification.

All together verified Gem-mESCs as a suitable cell model to study cell cycle.

Our model is missing one of the two FUCCI probes, specifically the one marking G1 phase (CDT1 fused to the fluorescent protein KO2) (Sakaue-Sawano et al., 2008). A disadvantage for Gem-mESCs in front of the original FUCCI system, is that it cannot identify early and late G1 cells. However, for this study we wanted to have a global picture of the cell cycle role in pluripotency exit comparing G1 and S-G2/M cells.

Another disadvantage of our system would be that we cannot distinguish between negative green fluorescent cells being G1 cells or cells that have lost the AG-Geminin construct. To overcome this, we culture the Gem-mESCs in the presence of geneticin (the selection antibiotic) and we routinely check DNA content of sorted cells.

As a way to strength the results of our functional assays and to discard potential technical bias, such as clonal specific effect on Gem-mESCs or possible differential effects of the FACS sorting process, alternative methods are being implemented in the laboratory. These include the usage of centrifugal elutriation and nocodazole treatment to obtain G1 and S-G2/M cells. Centrifugal elutriation relies on the physical separation of cells based on cell cycle size differences. Instead, nocodazole

acts by chemically inhibiting microtubules formation, therefore arresting cells in G2/M phases of the cell cycle. Nocodazole has been very recently described to be the most efficient small molecule to synchronize cells in G2/M cell cycle phases while not affecting them transcriptomically (Yiangou et al., 2019). In both cases, after either having a G1 and S-G2/M separated cells or G2/M arrested cells, an EB differentiation protocol will be performed to assess cardiomyocyte differentiation. We expect to obtain similar results to FACS sorted G1 and S-G2/M Gem-mESCs, being S-G2/M cells more primed for cardiac lineage commitment.

Self-renewal capacity is not altered during cell cycle

Once our cell model was validated for having minor transcriptomic changes compared to the R1 parental cell line, and for being efficient in separating cells in different cell cycle phases by FACS sorting, we started to characterize functionally mESCs during the cell cycle. First, we focused on the self-renewal capacity, one of the two major characteristics of ESC. Self-renewal is often misused in published reports as a synonym for pluripotency, the other main ESC characteristic. However, these are two different concepts that we clearly separate during this thesis. While pluripotency is the ability to differentiate to all cell types, self-renewal is the capacity to

divide indefinitely while maintaining the undifferentiated or pluripotent state.

We have shown that the same number of cells is produced regardless of the initial cell cycle phase in which cells were cultured. Moreover, the transcriptomic profile was not affected between cultured cells having been sorted in G1 and S-G2/M. From these results we therefore concluded that self-renewal capacity is not altered during the cell cycle. Supporting this, both G1 and S-G2/M sorted cells only need 24 hours to restore their cell cycle profile, meaning the original percentages of cells in G1, S and G2/M phases of the cell cycle.

Pluripotency capacity is altered during cell cycle

Following the objective of characterizing the two properties of ESCs, pluripotency was assessed *in vitro* by scRNA-seq from EBs. scRNA-seq allows the identification of cell heterogeneity in a given population. Thus, by performing this experiment we expected to determine the different cell types produced differently (if so) during the EB differentiation protocol starting from G1 and S-G2/M cells. From this analysis we were able to show that while G1 cells give rise to more endodermal cells, S-G2/M cells have a higher propensity to differentiate towards mesodermal cells compared to S-G2/M and G1 respectively.

Pluripotency segregation was already described before in hESCs, although it was restricted to early and late G1 phase,

being early G1 and late G1 cells more primed for endodermal and ectodermal differentiation respectively (Pauklin and Vallier, 2013; Pauklin et al., 2016). This differentiation bias was described to be driven by Cyclin D proteins, which oscillate during cell cycle (Pauklin and Vallier, 2013; Pauklin et al., 2016). However, alternative mechanisms must exist in mESCs, as Cyclin Ds are not regulated during cell cycle (Stead et al., 2002).

Not only this but scRNA-seq allowed us to confirm the already described existence of a G1 window of time permissive for lineage commitment upon differentiation cues (Sela et al., 2012; Calder et al., 2013; Gonzales et al., 2015; Coronado et al., 2013). We speculate that the increased transcriptional noise found in G1 mESCs could be the reason for the higher differentiation sensitivity observed in these cells compared to S-G2/M. In fact, increased transcriptional noise preceding lineage specification has been proposed before as a mechanism for cell fate-decision making process during early mouse development (Mohammed et al., 2017).

S-G2/M cells are primed for cardiomyocyte differentiation

The observed pluripotency segregation of S-G2/M cells towards mesodermal lineage was further characterized and delineated to cardiomyocyte specification. For this, we

performed EB differentiation and assessed the percentage of beating EBs. Moreover, we monitored the differentially expressed genes at late time points of differentiation and ACTC1 cardiomyocyte marker positive cells from G1 and S-G2/M originally formed EBs. Analysis showed an increased percentage of beating EBs from S-G2/M conditions, together with heart related categories when GO analysis were performed from day 8 bulk RNA-seq from EBs. ACTC1 cardiac marker staining revealed an increased number of differentiated cardiac cells from original S-G2/M mESCs. Thus, we concluded S-G2/M cells are more primed for cardiomyocyte differentiation compared to G1 cells. *In vivo* pre-cardiac mesodermal cells are known to undergo high levels of DNA synthesis (Pasumarthi and Field, 2002). Considering this, our hypothesis is that pre-cardiac mesodermal cells start cardiac differentiation from S phase cells.

Moreover, bulk RNA-seq also validated the faster differentiation initiation from G1 compared to S-G2/M.

However, no endodermal priming was observed from G1 mESCs from bulk RNA-seq data. We hypothesize that endodermal differentiation could be masked by the EB cell heterogeneity, being bulk RNA-seq not sensitive enough to capture this difference in lineage specification priming. Specific protocols for endoderm differentiation would be required to validate this specific segregation of pluripotency from G1 mESCs.

Cell cycle is determinant for *in vivo* cell specification

With the aim to test the relevance of the cell cycle-pluripotency dissociation *in vivo*, we performed for the first time chimera contribution analysis of G1 and S-G2/M cells. From this, we have concluded that the cell cycle impact on pluripotency is not specific of *in vitro* systems, but it also occurs *in vivo*. Specifically, G1 cells injected mESCs into blastocist embryos contributed significally more to extraembryonic yolk sac compared to S-G2/M cells.

Nonetheless, the *in vitro* observed differentiation priming for G1 and S-G2/M cells for endodermal and mesodermal lineages was not observed from the *in vivo* chimera formation assay. Yet, we believe that the developmental timing that represent cultured mESCs is not equivalent to the timing when mESCs were injected. mESCs were injected into E3-E3.5 early blastocyst stage embryos, prior to primitive endoderm and epiblast cells cell fate decision is taken. Thus, suggesting cell cycle plays a role on cell fate decision at different developmental stages.

In fact, during early mouse development, pluripotent cells are found as an asynchronous cell population (Mohammed et al., 2017). This is not only true for mouse but also for human early embryos, proven after analyzing single cell RNA-seq data from Petropoulos and co-workers report (data not shown)

(Petropoulos et al., 2016). Asynchrony in different stages of development could be thus crucial for cell fate decisions.
G1 and S-G2/M mESCs injections at earlier (before trophectoderm and ICM decision is taken) or later (before epiblast specification) time points would be key to validate this hypothesis.

Transcriptomic and proteomic differences across cell cycle progression

To characterize at the molecular level the functional differences between G1 and S-G2/M cells, an unbias approach was followed and the transcriptome and proteome were evaluated on these cells.

Transcriptomic analysis did not show changes that could help understanding the observed phenotype in G1 and S-G2/M cells. Transcriptome analysis during the cell cycle did not go in line with what was already published in hESCs, as developmental genes were not found to be upregulated in G1 cells (Singh et al., 2013) but on the contrary, in S-G2/M cells. We are currently re-analyzing the data considering lowly expressed genes (RPKMs lower than 1) and checking if developmental genes appear to be differentially enriched in G1 cells compared to S-G2/M in this case.
Nonetheless, these initial contradictory results could be explained by intra-species differences, as they show different

cell cycle regulation. Moreover, mESCs and hESCs *in vivo* counterparts exhibit different pluripotent states, being naïve and primed ESCs respectively. Primed ESCs are molecularly readier to differentiate and this could account for the observed discrepancies on developmental genes expression.

On the other hand, proteomic analysis showed changes in expression during cell cycle. We found DNA methylation/demethylation proteins as highly dynamic during progression of cell cycle. More precisely, we identified TDG as one of the most enriched proteins in G1. *Tdg* depletion was described to cause embryonic lethality (Cortellino et al., 2011; Cortázar et al., 2011) and heart abnormalities were shown to be one of the phenotypes in mice (Cortellino et al., 2011). Then, we postulated TDG as our candidate molecular determinant for the *in vitro* observed pluripotency segregation. TDG upregulation in G1 cells has been described in human HeLa cells before (Hardeland et al., 2007). TDG was shown to be degraded by the proteasome system when cells enter the S phase of the cell cycle until G2 (Hardeland et al., 2007). A proteasome degradation mechanism of TDG would go in line with the fact that we did not observe significant changes in expression for the *Tdg* gene.

Of note, G1 enriched proteins were a minority compared to S-G2/M. And among them, we identified several mitochondrial

proteins. In fact, mitochondrial proteins have been described to be differentially expressed in differentiating cells compared to pluripotent cells (Enomoto et al., 2015). In line with this, pluripotent cells rely on the glycolysis metabolism and differentiated cells, on mitochondrial respiration (Xu et al., 2013). Thus, a metabolic change is important to be happening when cells initiate differentiation. Being G1 a window of time for differentiation to occur, we speculate that differential expression of mitochondrial proteins may prepare the cell for lineage commitment in response to differentiation cues.

TDG determines cell cycle priming for cardiomyocyte differentiation

With the purpose of validating our candidate protein, we characterize TDG functionally. For this purpose, *Tdg* was knocked-down transiently and EB differentiation towards cardiomycoytes was performed. These results showed that *Tdg* transient downregulation resulted in an increased percentage of beating EBs and positive ACTC1 cells, sharing the same phenotype as S-G2/M cells. To further verify TDG implication on cardiac lineage segregation during the cell cycle, a transiently overexpressing *Tdg-3HA* system is being generated in the laboratory. After transfecting the cells, EB differentiation will be performed and the percentage of beating EBS together with the percentage of positive ACTC1 cells will

be determined. We expect to obtain a reduced cardiomyocyte differentiation priming in *Tdg-3HA* overexpressing cells, when compared to control cells.

Moreover, as EB differentiation turns to be spontaneous and a lot of variability could be seen between replicates, more directed cardiac lineage protocols should be established and performed for further characterization. An example for such variability can be observed comparing the EB differentiation experiments performed with in G1 and S-G2/M cells and siC and si*Tdg* transfected cells. While between 5 and 11% of the cells were positively stained by ACTC1 in cell cycle experiments (**Figure R.13E**), 30-50% of cells were positive for ACTC1 in transfected cells (**Figure R.25.D**).

Although we have tried some cardiomyocyte differentiation protocols (Morey et al., 2015; Craft et al., 2013; Kattman et al., 2011; Kokkinopoulos et al., 2016), we have faced several technical problems not obtaining differentiated cardiomyocytes up to date. However, by transiently knocking down *Tdg* we have obtained comparable results to published protocols, being cardiomyocyte cells around 50% of the total cell population in our data and 60% in published reports (measured in both cases using cardiac markers) (Kattman et al., 2011; Kokkinopoulos et al., 2016). Thus, by simply transiently depleting *Tdg* we could substantially improve cardiac differentiation using a very simple protocol, going from around 30% to 50% of differentiated cells.

Taken together, these results suggest TDG as the molecular determinant for the observed cardiomyocyte priming differentiation, acting by blocking this lineage specification in G1 cells. Importantly, we have linked TDG with pluripotency segregation for the first time.

TDG localizes genome-wide on TP53 target genes

To expand the characterization of the molecular pathways regulating cell cycle-linked pluripotency segregation, TDG ChIP-seq was performed. By ChIP-seq we identified TDG to be bound to regulatory regions and gene bodies. GO analysis, motif enrichment analysis and genomic co-occupancy with other factors pointed towards a TDG-TP53 association genome-wide. The downregulation of target genes upon *Tdg* knockdown suggest TDG to be a gene expression activator. Moreover, we showed that the downregulated TDG target genes upon *Tdg* KD are developmental related genes, suggesting a possible link between TDG and the phenotypic effect we observed during cell cycle.

TDG and TP53 were shown previously to physically interact and regulate each other expression in human cell lines (Kim and Um, 2008; Da Costa et al., 2012). However, they were never linked to play a role on pluripotency segregation before. Further experiments, such as IP and the doble *Tdg-Tp53* knockdown, are planned to be performed to verify TDG-TP53

interaction and to test their functional association. Moreover, mutagenesis on TDG TP53-interaction domain is being performed to determine whether the TDG-associated phenotype during cell cycle is dependent on TP53 interaction. In particular, the TDG E321Q mutation, which has been described to affect TP53 interaction (abrogating SUMO-1 binding) (Mohan et al., 2007). After TDG E321Q mutation has been achieved, cells depleted from wild-type TDG will be transfected with TDG E321Q and EB differentiation will be performed to assess cardiomyocyte differentiation. By losing TDG-TP53 interaction, we expect to lose the effect on cardiac lineage specification as well.

From all the data generated in this thesis, we can mainly conclude that: (1) G1 cells downregulate pluripotency genes and, subsequently, upregulate early differentiation genes more rapidly than S-G2/M mESCs upon differentiation cues. (2) S-G2/M mESCs are more primed to cardiac lineage specification compared to G1. (3) TDG is overexpressed in G1 cells compared to S-G2/M cells. (4) *Tdg* transient downregulation prime mESCs for cardiomyocyte differentiation. (5) TDG occupies TP53 binding sites. (6) TDG is a transcriptional activator. From this, our proposed model includes a cell cycle-regulated TDG association with TP53 in G1, which leads to gene activation. This role of TDG binding to TP53 and potentiating transcriptional activation has already been

suggested to happen in the human cancer cell line H1299 (Kim and Um, 2008). Thus, TP53 transcriptional programs would be activated. TP53 has been shown to act promoting ESCs differentiation by downregulating pluripotency genes (Lin et al., 2005; Jain and Barton, 2018). Thus, TP53 enhanced activity by the TDG activator role would lead to early differentiation from G1 cells (**Figure D.1**). However, TP53 transcriptional programs have a dual role in regulating ESCs nature. While some cell differentiation programs are induced, other are inhibited (Aylon Oren M, 2019; Jain and Barton, 2018). Thus, the TP53 regulatory network may also act blocking cardiac differentiation specifically in G1 cells, being potentiated by TDG (**Figure D.1**). Alternatively, cardiac differentiation blockade may be the result of a lineage exhaustion, after early mesodermal genes having been expressed prematurely in G1 mESCs. On the contrary, in S-G2/M cells, where TDG is less present, TDG-TP53 interaction would not occur and cells will be primed for cardiomyocyte differentiation (**Figure D.1**).

By exploiting the acquired knowledge from this thesis on the role of TDG in pluripotency segregation, improvements on *in vitro* cardiac lineage-directed protocols could be implemented.

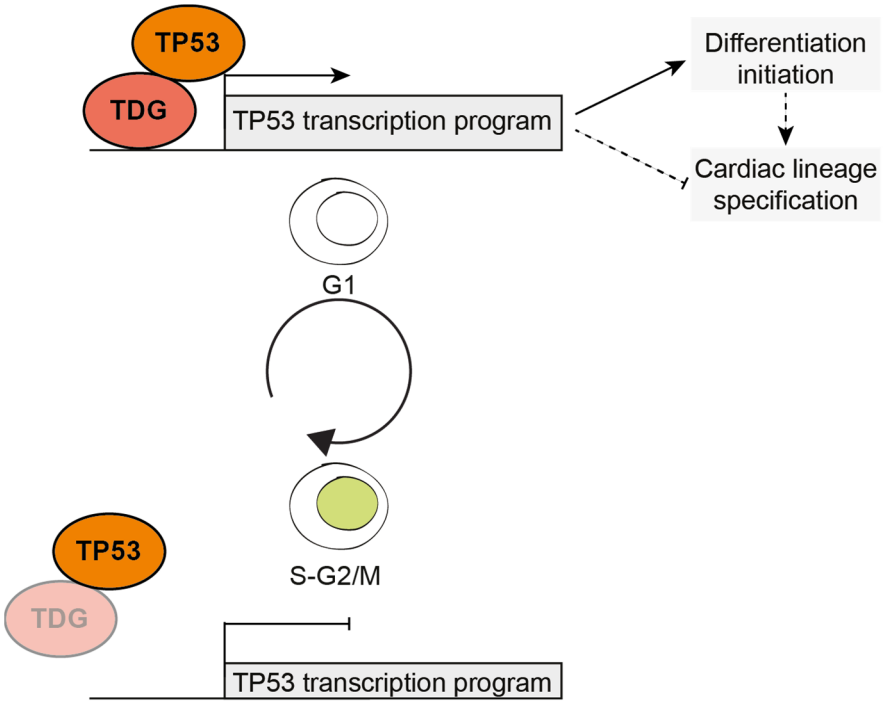


Figure D.1. Model for TDG-TP53 mediated G1 pluripotency-exit and cell cycle-linked segregation. TDG, being expressed in G1 cell cycle phase, binds to TP53 regulating its activity. This results in TP53 transcriptional program activation, which would lead to differentiation initiation and potentially to cardiac lineage specification blockage. This could be also achieved through the activation of early differentiation programs, by exhausting the cardiac lineage particular lineage program.

This would be of high importance for regenerative medicine purposes and for drug discovery screenings.

On the other hand, this could also be applied in other research areas such as cancer. In fact, *Tdg* has been seen to be de-regulated in different cancer types (data not shown, from the "Oncomine" and "The human protein atlas" databases) (Mancuso et al.). First-generation TDG inhibitors has been isolated and its anti-cancer activity has already been tested (Mancuso et al., 2019). Thus, we believe cell cycle-regulation

of TDG may have potential medical applications, not only in the field of regenerative medicine but also in cancer.

However, there are still many basic outstanding questions that remain. Why is differentiation initiated from G1 cell cycle phase? Why and when is cell cycle restructured when differentiation occurs? Are the mechanisms behind this link universal or specific for different species or cell types? By answering them, the knowledge on the field will be increased and this would be important for later applications for medical purposes.

CONCLUSIONS

From the work presented in this thesis we can draw the following main conclusions:

1. Gem-mESCs represents an adequate and useful model to study cell cycle molecularly.

2. Cell cycle does not impact on the self-renewal capacity of mESCs.

3. G1 cell cycle phase represents a window of time for cell fate commitment, being cells in this cell cycle fate more sensitive upon differentiation cues.

4. Pluripotency is segregated during cell cycle progression *in vitro*, being S-G2/M cells more primed for cardiomyocyte differentiation.

5. Cell cycle segregates pluripotency *in vivo* as well, contributing G1 cells more towards primitive endodermal tissues, after being injected in early E3.0-E3.5 blastocysts.

5. Transcriptomic changes observed during cell cycle could not explain S-G2/M cardiomyocyte priming.

5. Proteome analysis during cell cycle progression identified TDG as one of the more expressed G1 proteins compared to S-G2/M cell cycle phases.

6. Transient depletion of *Tdg* results in an increased cell propensity to differentiate towards the cardiac lineage.

7. TDG genome-wide distribution overlaps with that found in TP53.

8. TDG binds to TSSs and acts as an activator of gene expression.

9. TDG controls the activity of developmental genes.

MATERIALS AND METHODS

Cell culture

Gem-mESCs (Aranda et al., 2014), R1 mESCs (male mESCs of 129S1/SvImJ and 129X1/SvJ strains; from A. Nagy, Toronto, Canada) and E14TG2a mESCs (male mESCs of 129/Ola strain; Sigma-Aldrich) were grown feeder-free on tissue culture plates coated with 0.1% gelatin (Millipore). Coating was achieved by incubating plates with gelatin at 37ºC for at least 15 minutes.

mESCs were cultured in 2i/LIF medium to support naïve culture conditions: DMEM/F12:Neurobasal 1:1, supplemented with 0.5x N2 supplement, 1x B27 supplement, 0.5mM β-mercaptoethanol, glutamax, MEM non essential amino acids , penicillin-streptomycin, BSA V 0,033% (all from Gibco), 1 μM PD0325901 and 3μM CHIR99021 (Selleckchem), and Leukemia Inhibitory Factor (LIF). Alternatively, cells were grown in metastable culture conditions in serum-free medium (Knockout Dulbecco's modified Eagle's medium (DMEM)), supplemented with 20% Knockout Serum Replacement (KSR), 0.5mM β-mercaptoethanol, 2mM glutamax, MEM non-essential amino acids, penicillin-streptomycin (all from Gibco) supplemented with LIF.

FACS sorting

FACS sorting was performed at the CRG Flow Cytometry Unit. Either FACSAria II SORP or BD Influx sorters (BD Biosciences) were used for cell sorting.

Gem-mESCs

For Gem-mESCs, cells were collected using Trypsin (Gibco) to obtain a single cell suspension. Cells were then filtered to get rid of cell aggregates and sorting was performed.

E14TG2a

E14TG2a cells were collected using Trypsin (Gibco) to obtain a single cell suspension and were then fixed with 4% paraformaldehyde for 10 minutes at room temperature in rotation. Fixation was stopped by addition of 0.125M glycine (pH=7.4) for 5 minutes at room temperature in rotation. 3 PBS washes were then performed (with centrifugation steps of 5 minutes at 500g between washes).

1.5xE6 cells were incubated on ice for 10 minutes with DRAQ5 (abcam) at 5**µM** concentration to stain DNA content. Cell sorting was then performed.

Cell transfection

mESCs were reverse transfected with lipofectamine (Lipofectamine 2000; Invitrogen) with 30nM siRNA (from Sigma-Aldrich) directed to *Tdg* or with control siRNA (siRNA universal negative control #1; Sigma-Aldrich). The sequence of the siRNAs used against *Tdg* are listed in the **Table MM.1** below.

siRNA	Forward	Reverse
Tdg #1	GAAAUCCGGCAAGUCUA CA[dT]	UGUAGACUUGCCGGAUUU C[dT]
Tdg #2	GAAACUCUGUGCUACGU CA[dT]	UGACGUAGCACAGAGUUU C[dT]

Table MM.1. siRNA sequences against *Tdg*. Forward and reverse sequences for si*Tdg* #1 and si*Tdg* #2 are provided. All the sequences are shown in a 5' to 3' direction.

Embryoid body formation

For EB formation, mESCs were plated in hanging-drops at 1000-2000 cells per 20µl concentration in Glasglow Minimum Essential Medium (Sigma-Aldrich) supplemented with 10% Fetal Bovine Serum (FBS), 0.5mM β-mercaptoethanol, 2mM glutamax, MEM non-essential amino acids, penicillin-streptomycin (Gibco). After 2 days in culture, EBs were collected and cultured in non-treated plates. The medium was changed every 2-3 days. EBs were collected at different timepoints and processed for RNA extraction.

For Gem-mESCs, cells were initially FACS sorted and then the EB differentiation protocol was started.

Percentage of beating EBs assessment

To assess the percentage of beating EBs, EBs were plated on gelatinized 96 well plates individually (1 EB per well) at day 5 of differentiation, so that they get attached on the plate. Generally, 48 EBs were plated per condition. The medium was changed every 2-3 days and the percentage of beating EBs was assessed by counting the number of EBs showing any beating area over the total number of EBs.

Number of cells per EB assessment

To count the number of cells per EB, same number of EBs were collected per condition and were disaggregated using TrypLE Express (Invitrogen) to obtain a single cell suspension.

Number of cells per ml was then assessed using a Neubauer chamber.

Protein extraction and Western Blotting

Whole cell extracts were prepared in lysis buffer (whose composition is indicated below). Samples were then boiled for 10 minutes at 96ºC for protein denaturation. Protein extracts were quantified by BCA protein assay kit (Thermo Scientific), following instructions from manufacturers. 4x Laemli Buffer was then added to 10-20ug of protein and samples were boiled for 5 more minutes at 96ºC. Samples were then resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) in running buffer (detailed below). Protein samples were then transferred to a nitrocellulose membrane (Amersham) by running them in transfer buffer (detailed composition indicated below) for 1 hour at 4ºC. Transferred membranes were blocked in 10% milk in TBS-Tween (TBST) (detailed composition below) for at least 30 minutes. Membranes were then incubated with primary antibodies in 5% milk in TBST overnight at 4ºC (see **Table MM.2** for primary antibodies list). The following day, membranes were washed twice for 15 minutes in TBST and incubated with secondary antibodies conjugated with horseradish (Dako) in 5% milk in TBST at 1:2000 concentration for 45 minutes at room temperature (see **Table MM.3** for secondary antibodies list). Two more TBST washes were followed and finally proteins were detected using an enhanced chemiluminescence reagent

(Amersham ECL Prime Western Blotting detection reagent, Life Sciences). Chemiluminescence was read by using a Fuji LAS-3000 analyzer.

Lysis buffer
25mM Tris-HCl, pH=7.5
1% SDS
1mM EDTA
20mM β-glycerolphosphate
2mM sodium orto-vanadate
30mM PPi-Na
Cocktail protease inhibitors

Running buffer	Transfer buffer	TBST
25mM Tris-base	25mM Tris-HCl, pH= 8.3	10mM Tris-HCl, pH=7.5
200mM glycine	200nM glycine	100mM NaCl
0.1% SDS	20% methanol	0.1% Tween 20

PI staining

ESCs were ethanol fixed and then resuspended in propidium iodide (Sigma-Aldrich) at 20ug/ml with RNase A (Thermo Scientific) for 1 hour at 37ºC to stain the DNA. Cells were then placed on ice protected from light and analyzed using the CellQuest software (Becton Dickinson) and analyzed with the FlowJo software (version 10).

Immunostaining

Gem-mESCs

Gem-mESCs were plated on gelatinized chamber slides (Lab-Tek chamber slide, Thermo Scientific) in 2i/LIF conditions. 1-2 days later, cells were fixed in 4% paraformaldehyde at room temperature for 10 minutes. Then, 3 PBS washes were performed after which cells were incubated with blocking solution (see composition below) for at least 30 minutes at room temperature. Then, primary antibody was incubated for 2 hours at room temperature in blocking solution with 5% FBS (instead of 10%) (see **Table MM.2** for primary antibodies list). Cells were then washed 3 times in PBS and secondary antibody was incubated for 1 hour at room temperature in the dark (see **Table MM.3** for secondary antibodies list). After 3 more washes in PBS, mounting medium with DAPI (Biotium) was applied and slides for confocal microscopy visualization were prepared.

For 5-CaC staining, same protocol was followed in cell suspension. After secondary antibody incubation, PI staining was performed. Cells were then placed on ice in the dark and analyzed using the CellQuest software (Becton Dickinson) and analyzed with the FlowJo software (version 10).

EBs

EBs were disaggregated using TrypLE Express (Invitrogen) to obtain a single cell suspension. Cells were then fixed in 4% paraformaldehyde in rotation at room temperature for 10

minutes, after which cells were washed 3 times in PBS, centrifuging the cells for 5 minutes at 500g between washes. After having been filtered to discard cell aggregates, cells were incubated with blocking solution (indicated composition below) for at least 30 minutes at room temperature. Then, incubation with primary antibody against ACTC1 for 2 hours at room temperature was performed in blocking solution with 5% FBS (instead of 10%) (see **Table MM.2** for primary antibodies list). Cells were then washed 3 times in PBS and incubated with the secondary antibody for 1 hour at room temperature in the dark (see **Table MM.3** for secondary antibodies list). After 3 more washes, cells were analyzed using LSR II (Becton Dickinson). Data was collected using the CellQuest software (Becton Dickinson) and analyzed with FlowJo (version 10).

Blocking solution
1% Bovine serum albumin (BSA)
10% FBS
0.1% Triton X-100
PBS

Teratoma formation assay

The teratoma formation assays were performed in collaboration with CRG Tissue Engineering Unit.

Gem-mESCs were FACS sorted according to green fluorescence intensity to obtain G1 and S-G2/M. 3xE5 cells and 1:15 matrigel were injected in 100ul of M2 medium (Millipore) per flank in 3 immunodeficient SCID BEIGE mice per each condition. After 3 weeks, mice were sacrificed and

teratomas were isolated and washed with PBS. Then teratomas were fixed in 4% PFA and paraffin embedded. Finally, sections were obtained from 3 different regions using a microtome per each teratoma and hematoxylin-eosin staining was performed. By visual inspection, percentage of endodermal, mesodermal and ectodermal cells were assessed for each section and the averaged percentage of lineage contribution per teratoma was then calculated.

All experimental protocols were performed in accordance with recommendations for the proper care and use of laboratory animals [local (law 32/2007); European (EU directive n° 86/609, EU decree 2001- 486) regulations, and the Standards for Use of Laboratory Animals n° A5388-01 (NIH)] and were approved by the local ethical committee (CEEA-PRBB).

Chimera contribution assay

PiggyBac cell line establishment

Gem-mESCs were co-transfected using Lipofectamine 2000 (Invitrogen) with a PiggyBac vector carrying a CAG promoter-driven dsRed reporter gene, together with a PiggyBac transposase-encoding plasmid (1:10 transposon-transposase relation). Cells were then selected with Hygromycine at 1:5000 concentration (Sigma-Aldrich) for 5 days.

FACS sorting, blastocyst injection, chimera formation and analysis

The chimera contribution assays were performed in collaboration with CRG Tissue Engineering Unit.

Gem-PiggyBac-dsRed mESCs were FACS sorted according to red and green fluorescence intensity (positive cells for dsRed reporter gene and both negative (G1) and positive (S-G2/M) for AzamiGreen). 8-14 cells either G1 or S-G2/M were then injected to 14-16 E3.0-E3.5 mouse blastocysts of CD1 mice (white coat colour), which were then transferred to a pseudopregnant mother 2.5 days post coitum (from CD1 mouse strain) for chimaera generation. 3 independent injections were performed per condition. Embryos were collected at E18.5 stage. The heart, liver, brain, placenta and yolk sac were dissected from each embryo and disaggregated by using 1ml 0.05% Trypsin-EDTA with 100ug/ml DNase I, incubating the samples overnight at 4ºC. The following day, samples were filtered to get rid of cellular aggregates and lysis buffer (RBC) was used to eliminate erythrocyte cells from the samples, a source of auto fluorescence. Single cell suspensions were analyzed by FACS (Fortessa, Becton Dickinson) in order to quantify the percentage of positive dsRed cells (chimeric cells). Data was collected using the CellQuest software (Becton Dickinson) and analyzed with FlowJo (version 10).

All experimental protocols were performed in accordance with recommendations for the proper care and use of laboratory animals [local (law 32/2007); European (EU directive n° 86/609, EU decree 2001- 486) regulations, and the Standards for Use of Laboratory Animals n° A5388-01 (NIH)] and were approved by the local ethical committee (CEEA-PRBB).

RNA extraction and cDNA synthesis

RNA was extracted using the RNeasy mini kit (Qiagen) following manufacturer's instructions. Total RNA was quantified with NanoDrop, and equal amounts of RNA were used to generate cDNA with qScript cDNA synthesis kit (Quanta Biosciences). SYBR Green I PCR Master Mix (Roche) and Roche LightCycler 480 were used to perform qPCR reactions. All the primers used are listed in the **Table MM.4**. Expression values were normalized by the housekeeping gene *Rplp0* or *Gapdh.* For spike-in qPCR, Kc167 *Drosophila* cells were added to mESCs in a 1:4 proportion, RNA was extracted from pooled cells and cDNA was synthetized and finally retrotranscribed. Expression values were normalized by *dActin*.

RNA-sequencing

Library preparation and sequencing

RNA-seq (library preparation and sequencing run) was performed at the CRG Genomics Unit. RNA samples were quantified and quality control was performed using a Bioanalyzer. Ribosomal RNA was depleted prior to library preparation, for which TruSeq Stranded Total RNA Library Prep (Illumina) was used. Single read (50bp) mRNA-polyA selection sequencing was then performed from at least 1ug of RNA using the HiSeq2500 sequencer (Illumina).

Bioinformatics analysis (performed by Enrique Blanco and Mar González-Ramírez)

The RNA-seq samples were mapped against the mm9 mouse genome assembly using TopHat (Trapnell et al., 2009) with the option –g 1 to discard those reads that could not be uniquely mapped in just one region. Cufflinks and Cuffdiff were run to quantify the expression in FPKMs of each annotated transcript in RefSeq and to identify the list of differentially expressed genes for each case (Trapnell et al., 2013).

RNA-seq samples normalized by spike-in were mapped against a synthetic genome constituted by the mouse chromosomes (mm9) and the ERCC92 spike-in mix of sequences using Bowtie with the option -m 1 to discard reads that did not map uniquely to one region. The expression values of the mouse genes were corrected by performing local regression (LOESS) using the expression of the ERCC sequences in the same conditions as a guide, as previously described (Lovén et al., 2013).

RNA-seq samples normalized by spike-in were mapped against a synthetic genome constituted by the mouse and the fruit fly chromosomes (mm9 + dm3) using Bowtie with the option -m 1 to discard reads that did not map uniquely to one region. The expression values of the mouse genes were corrected by performing local regression (LOESS) using the expression of fly genes in the same conditions as a guide, as previously described (Taruttis et al., 2017).

DESeq2 (Love et al., 2014) was run to quantify the expression of every annotated transcript using the RefSeq catalog of

exons and to identify each set of differentially expressed genes.

GSEA of the pre-ranked lists of genes by DESeq2 stat value was performed with the GSEA software (Subramanian et al., 2005). Mouse genes were ranked by the ratio between si*Tdg* and siC RNA-seq expression or SG2M and G1 RNA-seq expression.

Single-cell RNA-sequencing (performed in collaboration with Atefeh Lafzy and Holger Heyn, from CNAG-CRG)

Gem-mESCs were single-cell FACS sorted according to AzamiGreen fluorescence (non-fluorescent (G1) and fluorescent cells (S-G2/M)) in 384 well plates, containing lysis buffer and reverse-transcription (RT) primers. Cells from both populations were also FACS sorted to form EB. EBs were disaggregated with 0.05% Trypsin-EDTA (Gibco) to obtain a single cell suspension at different time points of the EB differentiation process: day 3, 6 and 10. Cells were then single-cell FACS sorted in 384 well plates.

Library preparation (MARSeq)

To construct single cell libraries from polyA-tailed RNA, we applied massively parallel single-cell RNA sequencing (MARSseq) (Paul et al. 2015; Jaitin et al. 2014). Briefly, single cells were FACS-sorted into 384-well plates, containing lysis buffer (0.2% Triton (Sigma-Aldrich); RNase inhibitor (Invitrogen)) and reverse-transcription (RT) primers. The RT

primers contain the single cell barcodes and unique molecular identifiers (UMIs) for subsequent de-multiplexing and correction for amplification biases, respectively. Single cell lysates were denatured and immediately placed on ice. The RT reaction mix, containing SuperScript III reverse transcriptase (Invitrogen) was added to each sample. In the RT reaction, spike-in artificial transcripts (ERCC, Ambion) were included at a dilution of 1:16xE6 per cell. After RT, the cDNA was pooled using an automated pipeline (epMotion, Eppendorf). Unbound primers were eliminated by incubating the cDNA with exonuclease I (NEB). A second pooling was performed through cleanup with SPRI magnetic beads (Beckman Coulter). Subsequently, pooled cDNAs were converted into double-stranded DNA with the Second Strand Synthesis enzyme (NEB), followed by clean up and linear amplification by T7 *in vitro* transcription overnight. Afterwards, the DNA template was removed by Turbo DNase I (Ambion) and the RNA was purified with SPRI beads. Amplified RNA was chemically fragmented with Zn2+ (Ambion), then purified with SPRI beads. The fragmented RNA was ligated with ligation primers containing a pool barcode and partial Illumina Read1 sequencing adapter using T4 RNA ligase I (NEB). Ligated products were reversed transcribed using the Affinity Script RT enzyme (Agilent Technologies) and a primer complementary to the ligated adapter, partial Read1. The cDNA was purified with SPRI beads. Libraries were completed through a PCR step using the KAPA Hifi Hotstart ReadyMix (Kapa

Biosystems); and a forward primer that contains Illumina P5-Read1 sequence and the reverse primer containing the P7-Read2 sequence. The final library was purified with SPRI beads to remove excess primers. Library concentration and molecular size was determined with High Sensitivity DNA Chip (Agilent Technologies). The libraries consist of 192 single cell pools. Multiplexed pools (2) were run in one Illumina HiSeq 2500 Rapid two lane flow cell following the manufacturer's protocol. Primary data analysis was carried out with the standard Illumina pipeline.

Data preprocessing

Sequencing was carried out as paired-end reads, wherein the first read contains the transcript sequence and the second read the cell barcode and UMIs. Quality check of the generated reads was performed with the FastQC quality control suite. Samples that reached the quality standards were then processed to deconvolute the reads to single-cell level by de-multiplexing according to the cell and pool barcodes. Reads were filtered to remove polyT sequences. Sequencing reads were mapped to the mouse reference genome (Gencode release M12, assembly GRCm38) with the RNA pipeline of the GEMTools 1.7.0 suite using default parameters (6% of mismatches, minimum of 80% matched bases, and minimum quality threshold of 26). Cells with less than 60% of reads mapping to the reference genome or more than 2×106 total reads were discarded. Gene quantification was performed using UMI corrected transcript information to correct for

amplification biases, collapsing read counts for reads mapping on a gene with the same UMI (allowing an edit distance up to two nucleotides in UMI comparisons). Only unambiguously mapped reads were considered. Thresholds were set to reduce technical noise, but to conserve the sensitivity to identify low frequency outlier cell populations and to capture differences between fresh and cryopreserved cells.

Data Analysis

Population dispersion

In order to calculate the dispersion of cell populations in each time point in 2D PCA space, we performed kmeans on the first two principle components with K=4 (corresponding to the 4 time points) and calculated the centroids of each time points (Fig X.a). Later, for each subpopulation we calculated the average distance of all cells in the same population from their corresponding centroid. We represented these average distances in a barplot (Fig X.b). This would give an idea of how dispersed cells are in each of the distinct time points.

Clustering

For clustering, we filtered the expression matrix by removing genes that are expressed in less than 10 cells. Also low-quality cells were filtered out based on the distribution of the number of non-zero count genes per cell (minimum number of genes detected), to remove cells with less than two median absolute deviations (MAD) with respect to the median. Gene expression levels for each cell were normalized by the total expression, multiplied by a scale factor (10,000), and log-trans- formed.

Batches were then regressed out, and scaled Z scored residuals of the model were used as normalized expression values. We defined the 10% most variable genes based on their average expression and dispersion as highly variable genes (HVG). We reduced the dimensionality of the data by performing principle component analysis (PCA) on HVG. To find subpopulations, clustering was performed using the R package Seurat (Stuart, Butler et al., bioRxiv, 2018) on the first 8 PCAs based on the amount of variance explained by them. To cluster cells, a K-nearest neighbor (KNN) graph constructed on a Euclidean distance matrix in PCA space was calculated and then converted to a shared nearest neighbor (SNN) graph, in order to find highly interconnected communities of cells (Xu and Su, 2015). Cells were then clustered using the Louvain method to maximize modularity (Waltman and Van Eck, 2013). To display data, the t-distributed stochas- tic neighbor embedding (t-SNE) was applied to cell loadings of selected PCs, and the cluster assignments from the graph-based clustering were used (Fig X.c,d).

Transcriptional Noise

To compare transcriptional noise within the G1 and SG2-M population of cells in each time point, we calculated the average distance between each cell to rest of the other cells in each time point for G1 and SG2-M cells separately. We plotted the distribution of these average distances in a boxplot with the idea of having a higher distribution means the cells are more

far from each other which can be interpreted as higher transcriptional noise.

All plates were sequenced with massively parallel single-cell RNA sequencing (MARS-Seq) method. A total of 1531 cells were sequenced, which were virtually pooled to then be able to identify cell clusters based on gene expression. Finally, cluster composition was de-convoluted to be able to identify the amount of G1 and S/G2/M cells forming each cluster.

ChIP-sequencing

ChIP sample preparation

Two 15cm mESCs plates were crosslinked in 1% parafolmaldehid (PFA) for 10 minutes at room temperature in rotation. Fixation was stopped by addition of 0.125M glycine (pH=7.4) for 5 minutes at room temperature in rotation. Crosslinked cells were then washed 3 times in PBS and collected in ice-cold PBS with protease inhibitors (Roche) using cell scrapers. Cell pellets obtained by 5 minutes centrifugation at 1250g at 4ºC and were then resuspended in 30x volume of ice-cold IP buffer (composition is detailed below). Samples were sonicated for 60 cycles using a Bioruptor (Diagenode) at maximum output (30 seconds on/30 seconds off). After sonicaton, cells were centrifuged for 15 minutes at maximum speed at 4ºC and the supernatant was collected. To check fragmented chromatin size, 1% of de material (input) was reversed-crosslinked in high salt buffer (composition is listed below) for 3 hours at 65ºC at 1000rpm.

DNA was then purified using the PCR purification kit (Qiagen) and 0.5-0.8µg was analyzed in a 1% agarose gel to corroborate chromatin was sheared in 200-500bp fragments. 60µg of DNA were immunoprecipitated with 5µg of antibody (against you protein of interest and rabbit IgG as control; see **Table MM.2** for antibodies) in 650ul of total volume overnight at 4ºC in rotation. The following day, 42µl of protein A agarose beads blocked with 0.05% BSA in IP buffer (Diagenode) were added to ChIP samples and incubated for 2 hours at 4ºC. 3 and 1 washed of low salt and high salt buffer were performed respectively, centrifuging the cells at 3000rpm for 3 minutes between washes. ChIP samples were eluted in 200ul of freshly-prepared elution buffer (see below buffer components). Reverse crosslinking was then performed by overnight incubation at 65ºC at 1000rpm. Finally, DNA was purified using the PCR purification kit (Qiagen) the following day.

SDS buffer	Triton dilution buffer	IP buffer
100mM NaCl	100mM NaCl	1 volume SDS buffer
50mM Tris-HCl, pH=8.1	100mM Tris-HCl, pH=8.6	0.5 volume Triton dilution buffer
5mM EDTA, pH=8.0	5mM EDTA, pH=8.0	Proteinase inhibitors
0.5% SDS	5% Triton X-100	
Water	Water	

ChIP low salt buffer	ChIP high salt buffer	Elution buffer
50mM HEPES, pH=7.5	50mM HEPES, pH=7.5	0.1M $NaHCO^3$
140mM NaCl	500mM NaCl	1% SDS
1% Triton X-100	1% Triton X-100	Water
Water	Water	

Library preparation and sequencing

ChIP-seq was performed at the CRG Genomics Unit (from library preparation to sequencing run). ChIP samples were quantified and quality control was performed using Qubit. ChiP-seq library preparation was performed by using 2-10µg of DNA using the NEBNext Ultra DNA library Prep Kiy for Illumina (New England Biolabs). Prepared libraries were then sequences using a HiSeq2500 sequencer (Illumina).

Bioinformatics analysis (performed in collaboration with Enrique Blanco and Mar González-Ramírez)

ChIPseq samples were mapped against the mm9 mouse genome assembly using Bowtie with the option –m 1 to discard those reads that could not be uniquely mapped to just one region (Langmead et al., 2009).

MACS was run with the default parameters but with the shift-size adjusted to 100 bp to perform the peak calling against the corresponding control sample (Zhang et al., 2008).

The genome distribution of each set of peaks was calculated by counting the number of peaks fitted on each class of region according to RefSeq annotations (Leary et al., 2016). Distal region is the region within 2.5 Kbp and 0.5 Kbp upstream of the transcription start site (TSS). Proximal region is the region within 0.5 Kbp upstream of the TSS. UTR, untranslated region; CDS, protein coding sequence; intronic regions, introns; and the rest of the genome, intergenic. Peaks that overlapped with more than one genomic feature were proportionally counted the same number of times.

Spie charts were generated by calculating the genome distribution of all features in the full genome, and the R caroline package was used to combine the piechart of each set of peaks with the full genome distribution (Feitelson).
Each set of target genes was retrieved by matching the ChIPseq peaks in the region 2.5 Kbp upstream of the TSS until the end of the transcripts as annotated in RefSeq.
Reports of functional enrichments of GO and other genomic libraries were generated using the EnrichR tool (Kuleshov et al., 2016).

The aggregated plots showing the average distribution of ChIPseq reads around the TSS of each target gene were generated by counting the number of reads for each region according to RefSeq and then averaging the values for the total number of mapped reads of each sample and the total number of genes in the particular gene set.
The aggregated plots showing the average distribution of ChIPseq reads of a collection of peaks were generated by counting the number of reads around the summit of each peak and normalizing for the total number of peaks in the set.
Motif analysis of the sequences within the ChIPseq peaks of one experiment was performed with the MEME-ChIP tool, adjusting the MEME motif width between 5 and 15 bps (Machanick and Bailey, 2011).

Proteomics (in collaboration with Eduard Sabido and Eva Borras)

Sample processing

Cell lysates were reduced, alkylated and digested to peptide mixes according to the filter-aided sample preparation (Wisniewski et al., 2009) method using LysC 1:10 ratio(w:w; enzyme:substrate) at 37º C overnight followed by trypsin 1:10 ratio (w:w; enzyme:substrate) at 37 º C for 8 hours. Tryptic peptide mixtures were desalted using a C18 UltraMicroSpin column (Rappsilber et al., 2007).

LCMS Analysis

Samples were analyzed in an Orbitrap hybrid Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) coupled to a Proxeon nano-LC equipped with a C18 reversed-phase chromatography column using a 120 min gradient (water/acetonitrile). The instrument was operated in DDA mode and full MS scans with 1 micro scans at resolution of 120,000 were used over a mass range of m/z 300-2,000 with detection in the Orbitrap. Following each survey scan the top twenty most intense ions were selected for fragmentation at normalized collision energy of 35%. Fragment ion spectra produced via collision-induced dissociation (CID) were acquired in the linear ion trap. All data were acquired with Xcalibur software. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al., 2016) partner repository with the dataset identifier PXD008335.

Data Analysis

Acquired data were analyzed using the Proteome Discoverer software suite (v1.4, Thermo Fisher Scientific) and data were searched with Mascot search engine against SwissProt Mouse where the most common contaminants were added (Bunkenborg et al., 2010). A precursor ion mass tolerance of 7 ppm at the MS1 level was used, and up to three missed cleavages for trypsin were allowed. The fragment ion mass tolerance was set to 0.5 Da. Oxidation of methionine and protein acetylation at the N-terminal were defined as variable modification. Carbamidomethylation on cysteines was set as a fix modification. The identified peptides were filtered using a FDR < 1 % using Percolator.

For quantification, peptides observed in one single biological replicate were not considered and values completely missing in one condition were imputed with a minimal value extracted +/- error. Areas of the three most abundant peptides per protein were log-transformed and used to calculate protein fold-changes and p-values using an ANOVA model. P-values were adjusted for multiple hypothesis testing using BH.

Primary antibodies list

Primary antibody	Host	Application	Dilution	Source	Catalog N
GAPDH	mouse	WB	1:5000	Santa Cruz Biotechnology	
Cyclin B1		WB	1:1000	Santa Cruz Biotechnology	
H3	rabbit	WB	1:10000	Abcam	ab1791
VINCULIN	mouse	WB	1:1000	Sigma-Aldrich	V9131
TDG	rabbit	WB	1:1000	Active Motif	61437, 61438
5-Carboxylcytosine	rabbit	Flow Cyt.	2ul per 2ug	Active Motif	61225, 61226
TDG	rabbit	ChIP	5ug	Thermo Scientific	PA5-29140
IgG	rabbit	ChIP	5ug	Abcam	ab172730
ACTININ-α	mouse	Flow Cyt.	1:800	Sigma-Aldrich	A7811

Table MM.2. List of primary used antibodies. The targeted protein, host, application, the dilution used, the source and the catalog number are provided for each antibody. WB means western blot; Flow Cyt., flow cytometry.

Secondary antibodies list

Secondary antibody	Host	Application	Dilution	Source	Catalog N
anti-mouse - horseradish	rabbit	WB	1:2000	Dako	P0260
anti-rabbit-horseradish	goat	WB	1:2000	Dako	P0448
anti-rabbit-650	goat	IF	1:500	Bethyl	A120-101D5
anti-mouse-647	donkey	Flow Cyt.	1:500	Invitrogen	A31571

Table MM.3. List of secondary used antibodies. Secondary antibody, host, application, the dilution used, the source and the catalog number are provided for each antibody. IF means immunofluorescence.

Primer list

Gene	Forward	Reverse
Nanog	TTCTTGCTTACAAGGGTCTGC	AGAGGAAGGGCGAGGAGA
Oct4	GTTGGAGAAGGTGGAACCAA	CTCCTTCTGCAGGGCTTT
Klf4	CGGGAAGGGAGAAGA	GAGTTCCTCACGCCA
Zfp42	AGTGTGCAGTGCAGCCAG	TGCTTTCTTCTGTGTGCAGG
Nestin	CTGCAGGCCACTGAAAAGTT	TCTGACTCTGTAGACCCTGCTTC
Gata4	TTCGCTGTTTCTCCCTCAAG	CAATGTTAACGGGTTGTGGA
Otx2	GACTGCAGGGCAGAGACG	GGTAGATTTGGAGTGACGGAAC
Cyclin B1	GCGCTGAAAATTCTTGACAAC	TTCTTAGCCAGGTGCTGCAT
Cyclin D2	CCCGACTCCTAAGACCCATC	TTGGAAGCTAGGAACATGCAC
Tdg	TTGTGGCATTGCTTCAAATG	CTGCCCATTCGGAACATC
Myl7	CCCATCAACTTCACCGTCTT	AACATGCGGAAGGCACTC
Actc1	CCGATCGTATGCAAAAGGAA	CTGGAAGGTGGACAGAGAGG
Nkx2-5	GACGTAGCCTGGTGTCTCG	GTGTGGAATCCGTCGAAAGT
Myh7	CTCAGAGCTCAAGCGGGATA	CCAGCCATCTCCTCTGTCA
Mef2c	TGATCAGCAGGCAAAGATTG	GGATGGTAACTGGCATCTCAA
Myl2	CAACGGCTGCATCAACTATG	GGCCAGGAAAGACTACCACA
Tnnt2	CCTGCTGAGGCTGAACAGAT	CAGACATGCTCTCGGCTCTC
Goosecoid	AAAGCCTCGCCGGAGAA	AGCTGTCCGAGTCCAAATCG
Sox17	CTGTGGAGGTGAGGGACTG	AGACAGTCTCCCCATGTAGCTC
Brachyury	TGCTTCCCTGAGACCCAGTT	GATCACTTCTTTCCTTTGCATCAAG
Mesp1	CCTTCGGAGGGAGTAGATC	AAAGCTTGTGCCTGCTTCA
EOMES	GGCAAAGCGGACAATAACAT	AGCCTCGGTTGGTATTTGTG
GAPDH	GTATGACTCCACTCACGGCAAA	TTCCCATTCTCGGCCTTG
Rplp0	TTCATTGTGGGAGCAGAC	CAGCAGTTTCTCCAGAGC
d*Actin*	GCGTCGGTCAATTCAATCTT	AAGCTGCAACCTCTTCGTCA

Table MM.4. List of primers used for qPCR experiments. Forward and reverse sequences are provided. All the sequences are shown in a 5' to 3' direction.

REFERENCES

Abranches, E., E. Bekman, and D. Henrique. 2013. Generation and Characterization of a Novel Mouse Embryonic Stem Cell Line with a Dynamic Reporter of Nanog Expression. *PLoS One*. 8:1–13. doi:10.1371/journal.pone.0059928.

Aranda, S., D. Rutishauser, and P. Ernfors. 2014. Identification of a large protein network involved in epigenetic transmission in replicating DNA of embryonic stem cells. *Nucleic Acids Res.* 42:6972–6986. doi:10.1093/nar/gku374.

Avilion, A.A., S.K. Nicolis, L.H. Pevny, L. Perez, N. Vivian, and R. Lovell-Badge. 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* 17:126–140. doi:10.1101/gad.224503.

Aylon Oren M, Y. 2019. The Paradox of p53: What, How and Why? *Cold Spring Harb. Perspect. Med.*

Banfalvi, G. 2011. Synchronization of Mammalian Cells and Nuclei by Centrifugal Elutriation. *Methods Mol. Biol.* 25–45. doi:10.1007/978-1-61779-182-6.

Beck, F., T. Erler, A. Russell, and R. James. 1995. Expression of Cdx-2 in the mouse embryo and placenta: Possible role in patterning of the extra-embryonic membranes. *Dev. Dyn.* 204:219–227. doi:10.1002/aja.1002040302.

Becker, K. a., P.N. Ghule, J. a. Therrien, J.B. Lian, J.L. Stein, A.J. Van Wijnen, and G.S. Stein. 2006. Self-renewal of human embryonic stem cells is supported by a shortened

G1 cell cycle phase. *J. Cell. Physiol.* 209:883–893. doi:10.1002/jcp.20776.

Bekaert, S., H. Derradji, and S. Baatout. 2004. Telomere biology in mammalian germ cells and during development. *Dev. Biol.* 274:15–30. doi:10.1016/j.ydbio.2004.06.023.

Benayoun, B.A., E.A. Pollina, D. Uçar, S. Mahmoudi, K. Karra, E.D. Wong, K. Devarajan, A.C. Daugherty, B. Anshul, E. Mancini, B.C. Hitz, R. Gupta, T.A. Rando, C. Julie, M.P. Snyder, J.M. Cherry, and A. Brunet. 2014. H3K4me3 breadth is linked to cell identity and transcriptional consistency. 158:673–688. doi:10.1016/j.cell.2014.06.027.H3K4me3.

Bernardo, A.S., A. Jouneau, H. Marks, P. Kensche, J. Kobolak, K. Freude, V. Hall, A. Feher, Z. Polgar, C. Sartori, I. Bock, C. Louet, T. Faial, H.H.D. Kerstens, C. Bouissou, G. Parsonage, K. Mashayekhi, J.C. Smith, G. Lazzari, P. Hyttel, H.G. Stunnenberg, M. Huynen, R.A. Pedersen, and A. Dinnyes. 2018. Mammalian embryo comparison identifies novel pluripotency genes associated with the naïve or primed state. *Biol. Open*. 7:bio033282. doi:10.1242/bio.033282.

Bernstein, B.E., T.S. Mikkelsen, X. Xie, M. Kamal, D.J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S.L. Schreiber, and E.S. Lander. 2006. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem

Cells. *Cell*. 125:315–326. doi:10.1016/j.cell.2006.02.041.

Bertero, A., P. Madrigal, A. Galli, N.C. Hubner, I. Moreno, D. Burks, S. Brown, R. a. Pedersen, D. Gaffney, S. Mendjan, S. Pauklin, and L. Vallier. 2015. Activin/Nodal signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark. *Genes Dev.* 29:702–717. doi:10.1101/gad.255984.114.

Boeuf, H., C. Hauss, F. De Graeve, N. Baran, and C. Kedinger. 1997. Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. *J. Cell Biol.* 138:1207–1217. doi:10.1083/jcb.138.6.1207.

Booth, M.J., M.R. Branco, G. Ficz, D. Oxley, W. Reik, and S. Balasubramanian. 2012. Reports Quantitative Sequencing of 5-Methyl- cytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Sci. Express reports*. 3–6.

Boward, B., T. Wu, and S. Dalton. 2016. Control of cell fate through cell cycle and pluripotency networks. 34:1427–1436. doi:10.1097/SLA.0000000000001177.Complications.

Brons, I.G.M., L.E. Smithers, M.W.B. Trotter, P. Rugg-Gunn, B. Sun, S.M. Chuva De Sousa Lopes, S.K. Howlett, A. Clarkson, L. Ahrlund-Richter, R. a. Pedersen, and L. Vallier. 2007. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*. 448:191–195. doi:10.1038/nature05950.

Brook, F.A., and R.L. Gardner. 1997. The origin and efficient derivation of embryonic stem cells in the mouse. *Proc. Natl. Acad. Sci.* 94:5709–5712. doi:10.1073/pnas.94.11.5709.

Bunkenborg, J., G. Espadas García, M.I. Peña Paz, J.S. Andersen, and H. Molina. 2010. The minotaur proteome : Avoiding cross-species identifications deriving from bovine serum in cell culture models. 3040–3044. doi:10.1002/pmic.201000103.

Burdon, T., C. Stracey, I. Chambers, J. Nichols, and A. Smith. 1999. Signaling mechanisms regulating self-renewal and differentiation of pluripotent embryonic stem cells. 43:30–43.

Calder, A., I. Roth-Albin, S. Bhatia, C. Pilquil, J.H. Lee, M. Bhatia, M. Levadoux-Martin, J. McNicol, J. Russell, T. Collins, and J.S. Draper. 2013. Lengthened G1 Phase Indicates Differentiation Status in Human Embryonic Stem Cells. *Stem Cells Dev.* 22:279–295. doi:10.1089/scd.2012.0168.

Chambers, I., J. Silva, D. Colby, J. Nichols, B. Nijmeijer, M. Robertson, J. Vrana, K. Jones, L. Grotewold, and A. Smith. 2007. Nanog safeguards pluripotency and mediates germline development. *Nature*. 450:1230–4. doi:10.1038/nature06403.

Chambers, I., and A. Smith. 2004. Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene*. 23:7150–7160. doi:10.1038/sj.onc.1207930.

Chambers, I., S. Tweedie, J. Nichols, S. Lee, D. Colby, M. Robertson, and A. Smith. 2003. Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. *Cell*. 113:643–655. doi:10.1016/s0092-8674(03)00392-1.

Chazaud, C., Y. Yamanaka, T. Pawson, and J. Rossant. 2006. Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway. *Dev. Cell*. 10:615–624. doi:10.1016/j.devcel.2006.02.020.

Chen, X., H. Xu, P. Yuan, F. Fang, M. Huss, V.B. Vega, E. Wong, Y.L. Orlov, W. Zhang, J. Jiang, Y.H. Loh, H.C. Yeo, Z.X. Yeo, V. Narang, K.R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W.K. Sung, N.D. Clarke, C.L. Wei, and H.H. Ng. 2008. Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell*. 133:1106–1117. doi:10.1016/j.cell.2008.04.043.

Coronado, D., M. Godet, P.Y. Bourillot, Y. Tapponnier, A. Bernat, M. Petit, M. Afanassieff, S. Markossian, A. Malashicheva, R. Iacone, K. Anastassiadis, and P. Savatier. 2013. A short G1 phase is an intrinsic determinant of naïve embryonic stem cell pluripotency. *Stem Cell Res.* 10:118–131. doi:10.1016/j.scr.2012.10.004.

Cortázar, D., C. Kunz, J. Selfridge, T. Lettieri, Y. Saito, E. MacDougall, A. Wirz, D. Schuermann, A.L. Jacobs, F.

Siegrist, R. Steinacher, J. Jiricny, A. Bird, and P. Schär. 2011. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature*. 470:419–423. doi:10.1038/nature09672.

Cortellino, S., J. Xu, M. Sannai, R. Moore, E. Caretti, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, K. Abramowitz, M.S. Bartolomei, F. Rambow, M.R. Bassi, M. Fanciulli, C. Renner, A.J. Klein-szanto, D. Kobi, I. Davidson, C. Alberti, and L. Larue. 2011. Demethylation by Linked Deamination-Base Excision Repair. *Cell*. 146:67–79. doi:10.1016/j.cell.2011.06.020.Thymine.

Da Costa, N.M., A. Hautefeuille, M.P. Cros, M.E. Melendez, T. Waters, P. Swann, P. Hainaut, and L.F.R. Pinto. 2012. Transcriptional regulation of thymine DNA glycosylase (TDG) by the tumor suppressor protein p53. *Cell Cycle*. 11:4570–4578. doi:10.4161/cc.22843.

Craft, A.M., N. Ahmed, J.S. Rockel, G.S. Baht, B.A. Alman, R.A. Kandel, A.E. Grigoriadis, and G.M. Keller. 2013. Specification of chondrocytes and cartilage tissues from embryonic stem cells. *Development*. 140:2597–2610. doi:10.1242/dev.087890.

Daheron, L., S.L. Opitz, H. Zaehres, W.M. Lensch, P.W. Andrews, J. Itskovitz-Eldor, and G.Q. Daley. 2004. LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. *Stem Cells*. 770–778.

Dannenberg, J.-H., A. van Rossum, L. Schuijff, and H. te Riele. 2000. Ablation of the Retinoblastoma gene family

deregulates G1 control causing immortalization and increased cell turnover under growth-restricting conditions. *Mol. Biol. Cell*. 3051–3064. doi:10.1101/gad.847700.function.

Edgar, B.A., and C.F. Lehner. 1996. Developmental control of cell cycle regulators: A fly's perspective. *Science (80-.).* 274:1646–1652. doi:10.1126/science.274.5293.1646.

Eistetter, H.R. 1989. Pluripotent embryonal stem cell lines can be Established from disggregated mouse moruale. *Development*. 31:275–282.

Eldar, A., and M.B. Elowitz. 2010. Functional roles for noise in genetic circuits. *Nature*. 467:167–173. doi:10.1038/nature09326.

Enomoto, K., K. Watanabe-Susaki, M. Kowno, H. Takada, A. Intoh, Y. Yamanaka, H. Hirano, H. Sugino, M. Asashima, and A. Kurisaki. 2015. Identification of novel proteins differentially expressed in pluripotent embryonic stem cells and differentiated cells. *J. Med. Investig.* 62:130–136. doi:10.2152/jmi.62.130.

Evans, M.J., and M.H. Kaufman. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature*. 292:154–156.

Feitelson, D.G. Comparing Partitions With Spie Charts. 1–7.

Feldman, B., W. Poueymirou, V.E. Papaioannou, T.M. Dechiara, and M. Goldfarb. 1995. Requirement of FGF-4 for Postimplantation Mouse Development. 267:246–249.

Fernandes, S., J.J.H. Chong, S.L. Paige, M. Iwata, B. Torok-

Storb, G. Keller, H. Reinecke, and C.E. Murry. 2015. Comparison of human embryonic stem cell-derived cardiomyocytes, cardiovascular progenitors, and bone marrow mononuclear cells for cardiac repair. *Stem Cell Reports*. 5:753–762. doi:10.1016/j.stemcr.2015.09.011.

Ficz, G., M.R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. a. Hore, C.J. Marques, S. Andrews, and W. Reik. 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*. 473:398–404. doi:10.1038/nature10008.

Ficz, G., T.A. Hore, F. Santos, H.J. Lee, W. Dean, J. Arand, F. Krueger, D. Oxley, Y.-L. Paul, J. Walter, S.J. Cook, S. Andrews, M.R. Branco, and W. Reik. 2013. FGF Signaling Inhibition in ESCs Drives Rapid Genome-wide Demethylation to the Epigenetic Ground State of Pluripotency. 351–359. doi:10.1016/j.stem.2013.06.004.

Filipczyk, A.A., A.L. Laslett, C. Mummery, and M.F. Pera. 2007. Differentiation is coupled to changes in the cell cycle regulatory apparatus of human embryonic stem cells. 45–60. doi:10.1016/j.scr.2007.09.002.

Fluckiger, A.-C., G. Marcy, M. Marchand, D. Négre, F.-L. Cosset, S. Mitalipov, D. Wolf, P. Savatier, and C. Dehay. 2006. Cell Cycle Features of Primate Embryonic Stem Cells. *Stem Cells*. 24:547–556. doi:10.1634/stemcells.2005-0194.

Frommer, M., L.E. Mcdonald, D.S. Millar, C.M. Collist, F.

Wattt, G.W. Griggt, P.L. Molloyt, and C.L. Paul. 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. 89:1827–1831.

Gaspar-Maia, A., A. Alajem, E. Meshorer, and M. Ramalho-Santos. 2011. Open chromatin in pluripotency and reprogramming. *Nat. Rev. Mol. Cell Biol.* 12:36–47. doi:10.1038/nrm3036.

Ghule, P.N., R. Medina, C.J. Lengner, M. Mandeville, M. Qiao, Z. Dominski, J.B. Lian, J.L. Stein, A.J. Van Wijnen, and G.S. Stein. 2011. Reprogramming the pluripotent cell cycle: Restoration of an abbreviated G1 phase in human induced pluripotent stem (iPS) cells. *J. Cell. Physiol.* 226:1149–1156. doi:10.1002/jcp.22440.

Gonzales, K.A.U., H. Liang, Y.S. Lim, Y.S. Chan, J.C. Yeo, C.P. Tan, B. Gao, B. Le, Z.Y. Tan, K.Y. Low, Y.C. Liou, F. Bard, and H.H. Ng. 2015. Deterministic Restriction on Pluripotent State Dissolution by Cell-Cycle Pathways. *Cell*. 162:564–579. doi:10.1016/j.cell.2015.07.001.

Grandy, R. a., T.W. Whitfield, H. Wu, M.P. Fitzgerald, J.J. VanOudenhove, S.K. Zaidi, M. a. Montecino, J.B. Lian, A.J. van Wijnen, J.L. Stein, and G.S. Stein. 2015. Genome-Wide Studies Reveal that H3K4me3 Modification in Bivalent Genes Is Dynamically Regulated during the Pluripotent Cell Cycle and Stabilized upon Differentiation. *Mol. Cell. Biol.* 36:615–627. doi:10.1128/mcb.00877-15.

Habibi, E., A.B. Brinkman, J. Arand, L.I. Kroeze, H.H.D. Kerstens, F. Matarese, K. Lepikhov, M. Gut, I. Brun-heath, N.C. Hubner, R. Benedetti, L. Altucci, I.G. Gut, H. Marks, H.G. Stunnenberg, and J.H. Jansen. 2013. Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells. 1–10. doi:10.1016/j.stem.2013.06.002.

Haegel, H., L. Larue, M. Ohsugi, L. Federov, K. Herrenknecht, and R. Kemler. 1995. Lack of β-catenin affects mouse development at gastrulation. *Development*. 121:3529–3537.

Hall, J., G. Guo, J. Wray, I. Eyres, J. Nichols, L. Grotewold, S. Morfopoulou, P. Humphreys, W. Mansfield, R. Walker, S. Tomlinson, and A. Smith. 2009. Oct4 and LIF/Stat3 Additively Induce Krüppel Factors to Sustain Embryonic Stem Cell Self-Renewal. *Cell Stem Cell*. 5:597–609. doi:10.1016/j.stem.2009.11.003.

Hardeland, U., C. Kunz, F. Focke, M. Szadkowski, and P. Schär. 2007. Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2. *Nucleic Acids Res.* 35:3859–3867. doi:10.1093/nar/gkm337.

Hayashi, K., S.M.C. de S. Lopes, F. Tang, and M.A. Surani. 2008. Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States. *Cell Stem Cell*. 3:391–401. doi:10.1016/j.stem.2008.07.027.

He, Y., B. Li, Z. Li, P. Liu, Y. Wang, and Q. Tang. 2011. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. 333:1303–1307. doi:10.1126/science.1210944.Tet-Mediated.

Heasman, J. 2006. Patterning the early Xenopus embryo. *Development*. 133:1205–1217. doi:10.1242/dev.02304.

Hong, Y., R.B. Cervantes, E. Tichy, J.A. Tischfield, and P.J. Stambrook. 2007. Protecting genomic integrity in somatic cells and embryonic stem cells. 614:48–55. doi:10.1016/j.mrfmmm.2006.06.006.

Huang, G., S. Ye, X. Zhou, D. Liu, and Q.L. Ying. 2015. Molecular basis of embryonic stem cell self-renewal: From signaling pathways to pluripotency network. *Cell. Mol. Life Sci.* 72:1741–1757. doi:10.1007/s00018-015-1833-2.

Huelsken, J., R. Vogel, V. Brinkmann, B. Erdmann, C. Birchmeier, and W. Birchmeier. 2000. Requirement for beta-catenin in anterior-posterior axis formation in mice. 148:567–578.

Humphrey, R.K., G.M. Beattie, A.D. Lopez, N. Bucay, C.C. King, M.T. Firpo, S. Rose-John, and A. Hayek. 2004. Maintenance of pluripotency in human embryonic stem cells is STAT3 independent. *Stem Cells*. 22:522–530.

Huurne, M., T. Peng, G. Yi, G. Van Mierlo, H. Marks, and G. Hendrik. 2019. Critical role for P53 in regulating the cell cycle of ground state embryonic stem cells. *bioRxiv*.

ter Huurne, M., J. Chappell, S. Dalton, and H.G.

Stunnenberg. 2017. Distinct Cell-Cycle Control in Two Different States of Mouse Pluripotency. *Cell Stem Cell*. 21:449-455.e4. doi:10.1016/j.stem.2017.09.004.

Hwang, Y.-S., B.G. Chung, D. Ortmann, N. Hattori, H.-C. Moeller, and A. Khademhosseini. 2009. Microwell-mediated control of embryoid body size regulates embryonic stem cell fate via differential expression of WNT5a and WNT11. *Proc. Natl. Acad. Sci.* 106:16978–16983. doi:10.1073/pnas.0905550106.

Ide, H., K. Akamatsu, Y. Kimura, K. Michiue, K. Makino, A. Asaeda, Y. Takamori, and K. Kubo. 1993. Synthesis and Damage Specificity of a Novel Probe for the Detection of Abasic Sites in DNA. *Biochemistry*. 32:8276–8283. doi:10.1021/bi00083a031.

Ito, S., L. Shen, Q. Dai, S.C. Wu, L.B. Collins, J. a Swenberg, C. He, and Y. Zhang. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*. 333:1300–1303. doi:10.1126/science.1210597.

Jaenisch, R., and R. Young. 2008. Stem Cells, the Molecular Circuitry of Pluripotency and Nuclear Reprogramming. *Cell*. 132:567–582. doi:10.1016/j.cell.2008.01.015.

Jaime-Soguero, A. De, F. Aulicino, G. Ertaylan, A. Griego, A. Cerrato, A. Tallam, A. Sol, M.P. Cosma, and F. Lluis. 2017. Wnt / Tcf1 pathway restricts embryonic stem cell cycle through activation of the Ink4 / Arf locus. *PLoS Genet.*

Jain, A.K., and M.C. Barton. 2018. P53: Emerging Roles in Stem Cells, Development and Beyond. *Development*. 145:dev158360. doi:10.1242/dev.158360.

Kattman, S.J., A.D. Witty, M. Gagliardi, N.C. Dubois, M. Niapour, A. Hotta, J. Ellis, and G. Keller. 2011. Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. *Cell Stem Cell*. 8:228–240. doi:10.1016/j.stem.2010.12.008.

Kim, E.J., and S.J. Um. 2008. Thymine-DNA glycosylase interacts with and functions as a coactivator of p53 family proteins. *Biochem. Biophys. Res. Commun.* 377:838–842. doi:10.1016/j.bbrc.2008.10.058.

Kleinsmith, L.J., and G.B. Pierce. 1964. Multipotentiality of Single Embryonal Carcinoma Cells. *Cancer Res.* 24:1544–1551.

Koh, K.P., A. Yabuuchi, S. Rao, Y. Huang, K. Cunniff, J. Nardone, A. Laiho, M. Tahiliani, C.A. Sommer, G. Mostoslavsky, R. Lahesmaa, S.H. Orkin, S.J. Rodig, G.Q. Daley, and A. Rao. 2011. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*. 8:200–213. doi:10.1016/j.stem.2011.01.008.

Kokkinopoulos, I., H. Ishida, R. Saba, S. Coppen, K. Suzuki, and K. Yashiro. 2016. Cardiomyocyte differentiation from mouse embryonic stem cells using a simple and defined protocol. *Dev. Dyn.* 245:157–165.

doi:10.1002/dvdy.24366.

Kolodziejczyk, A.A., J.K. Kim, J.C.H. Tsang, T. Ilicic, J. Henriksson, K.N. Natarajan, A.C. Tuck, X. Gao, M. Bühler, P. Liu, J.C. Marioni, and S.A. Teichmann. 2015. Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell*. 17:471–485. doi:10.1016/j.stem.2015.09.011.

Kouzarides, T. 2007. Review Chromatin Modifications and Their Function. 693–705. doi:10.1016/j.cell.2007.02.005.

Kriaucionis, S., and N. Heintz. 2009. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science (80-.).* 324:929–930. doi:10.1126/science.1169786.

Kuleshov, M. V, M.R. Jones, A.D. Rouillard, N.F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S.L. Jenkins, K.M. Jagodnik, A. Lachmann, M.G. Mcdermott, C.D. Monteiro, W. Gundersen, and A. Ma. 2016. Enrichr : a comprehensive gene set enrichment analysis web server 2016 update. 44:90–97. doi:10.1093/nar/gkw377.

Kunath, T., M.K. Saba-El-Leil, M. Almousailleakh, J. Wray, S. Meloche, and A. Smith. 2007. FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development*. 134:2895–2902. doi:10.1242/dev.02880.

Lafzi, A., C. Moutinho, S. Picelli, and H. Heyn. 2018. Tutorial: guidelines for the experimental design of single-cell RNA

sequencing studies. *Nat. Protoc.* 13:2742–2757.

Lange, C., W.B. Huttner, and F. Calegari. 2009. Cdk4/CyclinD1 Overexpression in Neural Stem Cells Shortens G1, Delays Neurogenesis, and Promotes the Generation and Expansion of Basal Progenitors. *Cell Stem Cell*. 5:320–331. doi:10.1016/j.stem.2009.05.026.

Langmead, B., C. Trapnell, M. Pop, and S.L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. 10. doi:10.1186/gb-2009-10-3-r25.

Leary, N.A.O., M.W. Wright, J.R. Brister, S. Ciufo, D. Haddad, R. Mcveigh, B. Rajput, B. Robbertse, B. Smith-white, D. Ako-adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C.M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, S. Joardar, V.K. Kodali, W. Li, D. Maglott, P. Masterson, M. Mcgarvey, M.R. Murphy, K.O. Neill, S. Pujar, S.H. Rangwala, D. Rausch, L.D. Riddick, C. Schoch, A. Shkeda, S.S. Storz, H. Sun, F. Thibaud-nissen, I. Tolstoy, R.E. Tully, R. Vatsan, C. Wallin, D. Webb, W. Wu, M.J. Landrum, A. Kimchi, T. Tatusova, M. Dicuccio, P. Kitts, T.D. Murphy, and K.D. Pruitt. 2016. Reference sequence (RefSeq) database at NCBI : current status , taxonomic expansion , and functional annotation. 44:733–745. doi:10.1093/nar/gkv1189.

Leeb, M., R. Walker, B. Mansfield, J. Nichols, A. Smith, and A. Wutz. 2012. Germline potential of parthenogenetic

haploid mouse embryonic stem cells. *Development*. 139:3301–3305. doi:10.1242/dev.083675.

Leitch, H.G., K.R. Mcewen, A. Turp, V. Encheva, T. Carroll, N. Grabole, W. Mansfield, B. Nashun, J.G. Knezovich, A. Smith, M.A. Surani, and P. Hajkova. 2013. Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* 20:311–316. doi:10.1038/nsmb.2510.

Li, M., J. Carlos, and I. Belmonte. 2017. Ground rules of the pluripotency gene regulatory network. *Nat. Rev. Genet.* 1–12. doi:10.1038/nrg.2016.156.

Li, M., J. Carlos, and I. Belmonte. 2018. Deconstructing the puripotency gene regulatory network. *Nat. Cell Biol.* 20:382–392. doi:10.1038/s41556-018-0067-6.

Li, V.C., A. Ballabeni, and M.W. Kirschner. 2012. Gap 1 phase length and mouse embryonic stem cell self-renewal. *Proc. Natl. Acad. Sci.* 109:12550–12555. doi:10.1073/pnas.1206740109.

Lin, T., C. Chao, S. Saito, S.J. Mazur, M.E. Murphy, E. Appella, and Y. Xu. 2005. p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. *Nat. Cell Biol.* 7:165–171. doi:10.1038/ncb1211.

Lister, R., M. Pelizzola, R.H. Dowen, R.D. Hawkins, G. Hon, J. Tonti-Filippini, J.R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A.H. Millar, J.A. Thomson, B. Ren, and J.R. Ecker. 2009.

Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 315–322. doi:10.1007/BF02932019.

Livyatan, I., Y. Aaronson, D. Gokhman, R. Ashkenazi, and E. Meshorer. 2015. BindDB : An Integrated Database and Webtool Platform for ‘“ Reverse-ChIP ”’ Epigenomic Analysis. 17:647–648. doi:10.1016/j.stem.2015.11.015.

Loh, Y.H., Q. Wu, J.L. Chew, V.B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K.Y. Wong, K.W. Sung, C.W.H. Lee, X.D. Zhao, K.P. Chiu, L. Lipovich, V.A. Kuznetsov, P. Robson, L.W. Stanton, C.L. Wei, Y. Ruan, B. Lim, and H.H. Ng. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* 38:431–440. doi:10.1038/ng1760.

Love, M.I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. 1–21. doi:10.1186/s13059-014-0550-8.

Lovén, J., D.A. Orlando, A.A. Sigova, C.Y. Lin, P.B. Rahl, C.B. Burge, D.L. Levens, T.I. Lee, and R.A. Young. 2013. Revisiting Global Gene Expression Analysis. 151:476–482. doi:10.1016/j.cell.2012.10.012.Revisiting.

Lütticken, C., U.M. Wegenka, J. Yuan, J. Buschmann, C. Schindler, A. Ziemiecki, A.G. Harpur, A.F. Wilks, K. Yasukawa, T. Taga, T. Kishimoto, G. Barbieri, S. Pellegrini, M. Sendtner, P.C. Heinrich, and F. Horn.

1994. Association of transcription factor APRF and protein kinase Jak1 with the interleukin-6 signal transducer gp130. *Science (80-.).* 263:89–92. doi:10.1126/science.8272872.

Machanick, P., and T.L. Bailey. 2011. MEME-ChIP : motif analysis of large DNA datasets. 27:1696–1697. doi:10.1093/bioinformatics/btr189.

Mahla, R.S. 2016. Stem cells applications in regenerative medicine and disease therapeutics. *Int. J. Cell Biol.* 2016. doi:10.1155/2016/6940283.

Maiti, A., and A.C. Drohat. 2011. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. *J. Biol. Chem.* 286:35334–35338. doi:10.1074/jbc.C111.284620.

Mancuso, P., R. Tricarico, V. Bhattacharjee, L. Cosentino, Y. Kadariya, J. Jelinek, E. Nicolas, M. Einarson, N. Beeharry, K. Devarajan, R. a. Katz, D.G. Dorjsuren, H. Sun, A. Simeonov, A. Giordano, J.R. Testa, G. Davidson, I. Davidson, L. Larue, R.W. Sobol, T.J. Yen, and A. Bellacosa. Thymine DNA glycosylase as a novel target for melanoma. *Oncogene*. doi:10.1038/s41388-018-0640-2.

Mancuso, P., R. Tricarico, V. Bhattacharjee, L. Cosentino, Y. Kadariya, J. Jelinek, E. Nicolas, M. Einarson, N. Beeharry, K. Devarajan, R.A. Katz, D.G. Dorjsuren, H. Sun, A. Simeonov, A. Giordano, J.R. Testa, G. Davidson,

I. Davidson, L. Larue, R.W. Sobol, T.J. Yen, and A. Bellacosa. 2019. Thymine DNA glycosylase as a novel target for melanoma. *Oncogene*. 38:3710–3728. doi:10.1038/s41388-018-0640-2.

Marks, H., T. Kalkan, R. Menafra, S. Denissov, K. Jones, H. Hofemeister, J. Nichols, A. Kranz, A. Francis Stewart, A. Smith, and H.G. Stunnenberg. 2012. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*. 149:590–604. doi:10.1016/j.cell.2012.03.026.

Martello, G., P. Bertone, and A. Smith. 2013. Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *EMBO J.* 32:2561–2574. doi:10.1038/emboj.2013.177.

Martin, G.F., and M.J. Evans. 1974. Copyright@ 1974 by MIT. 2. 163–172 pp.

Martin, G.R. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 78:7634–8.

Martin, G.R., and M.J. Evans. 1975. Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. *Proc. Natl. Acad. Sci.* 72:1441–1445. doi:10.1073/pnas.72.4.1441.

Masui, S., Y. Nakatake, Y. Toyooka, D. Shimosato, R. Yagi, K. Takahashi, H. Okochi, A. Okuda, R. Matoba, A.A. Sharov, M.S.H. Ko, and H. Niwa. 2007. Pluripotency

governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* 9:625–635. doi:10.1038/ncb1589.

Matsuda, T., T. Nakamura, K. Nakao, T. Arai, M. Katsuki, T. Heike, and T. Yokota. 1999. STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO J.* 18:4261–4269. doi:10.1093/emboj/18.15.4261.

Mattout, A., and E. Meshorer. 2010. Chromatin plasticity and genome organization in pluripotent embryonic stem cells. *Curr. Opin. Cell Biol.* 22:334–341. doi:10.1016/j.ceb.2010.02.001.

McEwen, C.R., R.W. Stallard, and E.T. Juhos. 1968. Separation of Biological Particles by Centrifugal Elutriation. *Anal. Biochem.* 23:369–377.

Meshorer, E., and T. Misteli. 2006. Chromatin in pluripotent embryonic stem cells and differentiation. *Nat. Rev. Mol. Cell Biol.* 7:540–546. doi:10.1038/nrm1938.

von Meyenn, F., M. Iurlaro, E. Habibi, N.Q. Liu, A. Salehzadeh-Yazdi, F. Santos, E. Petrini, I. Milagre, M. Yu, Z. Xie, L.I. Kroeze, T.B. Nesterova, J.H. Jansen, H. Xie, C. He, W. Reik, and H.G. Stunnenberg. 2016. Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells. *Mol. Cell*. 62:848–861. doi:10.1016/j.molcel.2016.04.025.

Mierlo, G. Van, L. De Clerck, A.M. Dirks, L. De Clerck, A.B.

Brinkman, M. Huth, and S.L. Kloet. 2019. Integrative Proteomic Profiling Reveals PRC2- Dependent Epigenetic Crosstalk Maintains Ground- State Pluripotency. 24:1–15. doi:10.1016/j.stem.2018.10.017.

Mitsui, K., Y. Tokuzawa, H. Itoh, K. Segawa, M. Murakami, K. Takahashi, M. Maruyama, M. Maeda, and S. Yamanaka. 2003. The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. 113:1–12.

Mohammed, H., I. Hernando-herraez, A. Savino, J. Nichols, J.C. Marioni, W. Reik, H. Mohammed, I. Hernando-herraez, A. Savino, A. Scialdone, and I. Macaulay. 2017. Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. *CellReports*. 20:1215–1228. doi:10.1016/j.celrep.2017.07.009.

Mohan, R.D., A. Rao, J. Gagliardi, M. Tini, and M.O.L.C.E.L.L.B. Iol. 2007. SUMO-1-Dependent Allosteric Regulation of Thymine DNA Glycosylase Alters Subnuclear Localization and CBP / p300 Recruitment □ †. 27:229–243. doi:10.1128/MCB.00323-06.

Morey, L., A. Santanach, E. Blanco, L. Aloia, E.P. Nora, B.G. Bruneau, and L. Di croce. 2015. Polycomb Regulates Mesoderm Cell Fate-Specification in Embryonic Stem Cells through Activation and Repression Mechanisms. *Cell Stem Cell*. 17:300–315. doi:10.1016/j.stem.2015.08.009.

Mummery, C.L., C.E. van den Brink, and S.W. de Laat. 1987a. Commitment to differentiation induced by retinoic acid in P19 embryonal carcinoma cells is cell cycle dependent. *Dev. Biol.* 121:10–19. doi:10.1016/0012-1606(87)90133-3.

Mummery, C.L., M. a van Rooijen, S.E. van den Brink, and S.W. de Laat. 1987b. Cell cycle analysis during retinoic acid induced differentiation of a human embryonal carcinoma-derived cell line. *Cell Differ.* 20:153–160.

Murray, A.W., and M.W. Kirschner. 1989. Cyclin synthesis drives the early embryonic cell cycle. *Nature*. 342:189–92. doi:10.1038/340301a0.

Nakamura, J., V.E. Walker, P.B. Upton, S. Chiang, Y.W. Kow, and J. a Swenberg. 1998. Chemically Induced Depurination under Physiological Conditions. *Sites J. 20Th Century Contemp. French Stud.* 58:222–225.

Neganova, I., X. Zhang, S. Atkinson, and M. Lako. 2009. Expression and functional analysis of G1 to S regulatory components reveals an important role for CDK2 in cell cycle regulation in human embryonic stem cells. *Oncogene*. 28:20–30. doi:10.1038/onc.2008.358.

Neri, F., D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. DalBello, and S. Oliviero. 2015. Single-Base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter DNA Methylation Dynamics. *Cell Rep.* 10:674–683. doi:10.1016/j.celrep.2015.01.008.

Nichols, J., and A. Smith. 2009. Naive and Primed Pluripotent States. *Cell Stem Cell*. 4:487–492. doi:10.1016/j.stem.2009.05.015.

Nichols, J., B. Zevnik, K. Anastassiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Schöler, and A. Smith. 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*. 95:379–391. doi:10.1016/S0092-8674(00)81769-9.

Niwa, H., T. Burdon, I. Chambers, and A. Smith. 1998. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.* 12:2048–2060. doi:10.1101/gad.12.13.2048.

Orford, K.W., and D.T. Scadden. 2008. Deconstructing stem cell self-renewal: Genetic insights into cell-cycle regulation. *Nat. Rev. Genet.* 9:115–128. doi:10.1038/nrg2269.

Padovan-Merhar, O., G.P. Nair, A. Biaesch, A. Mayer, S.W. Foley, A.R. Wu, L.S. Churchman, A. Singh, and A. Raj. 2015. Single mammalian cells compensate for difference in cellular volume and DNA copy number through independent global transcriptional mechanisms. 58:339–352. doi:10.1016/j.molcel.2015.03.005.Single.

Pardee, A.B. 1974. A restriction point for control of normal animal cell proliferation. *Proc. Natl. Acad. Sci. U. S. A.* 71:1286–90.

Park, S.H., S.H. Park, M.C. Kook, E.Y. Kim, S. Park, and J.H. Lim. 2004. Ultrastructure of human embryonic stem cells

and spontaneous and retinoic acid-induced differentiating cells. *Ultrastruct. Pathol.* 28:229–238. doi:10.1080/01913120490515595.

Pastor, W.A., L. Aravind, and A. Rao. 2013. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* 341–356. doi:10.1016/j.biotechadv.2011.08.021.Secreted.

Pastor, W.A., U.J. Pape, Y. Huang, H.R. Henderson, R. Lister, M. Ko, E.M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G.Q. Daley, X.S. Liu, J.R. Ecker, P.M. Milos, S. Agarwal, and A. Rao. 2011. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*. 473:394–397. doi:10.1038/nature10102.

Pasumarthi, K.B.S., and L.J. Field. 2002. Cardiomyocyte Cell Cycle Regulation. 1044–1054. doi:10.1161/01.RES.0000020201.44772.67.

Pauklin, S., P. Madrigal, A. Bertero, and L. Vallier. 2016. Initiation of stem cell differentiation involves cell cycle-dependent regulation of developmental genes by Cyclin D. 421–433. doi:10.1101/gad.271452.115.

Pauklin, S., and L. Vallier. 2013. The cell-cycle state of stem cells determines cell fate propensity. *Cell*. 155:135. doi:10.1016/j.cell.2013.08.031.

Petropoulos, S., D. Edsga, B. Reinius, and S. Linnarsson. 2016. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation

Resource Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. 1012–1026. doi:10.1016/j.cell.2016.03.023.

Raiber, E., D. Beraldi, G. Ficz, H.E. Burgess, M.R. Branco, P. Murat, D. Oxley, M.J. Booth, W. Reik, and S. Balasubramanian. 2012. Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol.* 13:R69. doi:10.1186/gb-2012-13-8-r69.

Raiber, E., R. Hardisty, P. Van Delft, and S. Balasubramanian. 2017. Mapping and elucidating the function of modified bases in DNA. *Nat. Rev. Chem.* 1. doi:10.1038/s41570-017-0069.

Rappsilber, J., M. Mann, and Y. Ishihama. 2007. Protocol for micro-purification , enrichment , pre-fractionation and storage of peptides for proteomics using StageTips. doi:10.1038/nprot.2007.261.

Reubinoff, B. 2004. Human embryonic stem cells—potential applications for regenerative medicine. *Int. Congr. Ser.* 1266:45–53. doi:10.1016/j.ics.2004.02.002.

Rossant, J., C. Chazaud, Y. Yamanaka, M. Jones, E.J. Robertson, A. Smith, and R.G. Edwards. 2003. Lineage allocation and asymmetries in the early mouse embryo. *Philos. Trans. R. Soc. B Biol. Sci.* 358:1341–1349. doi:10.1098/rstb.2003.1329.

Sakaue-Sawano, A., H. Kurokawa, T. Morimura, A. Hanyu, H. Hama, H. Osawa, S. Kashiwagi, K. Fukami, T. Miyata, H.

Miyoshi, T. Imamura, M. Ogawa, H. Masai, and A. Miyawaki. 2008. Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell*. 132:487–498. doi:10.1016/j.cell.2007.12.033.

Santos, F., B. Hendrich, W. Reik, and W. Dean. 2002. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev. Biol.* 241:172–182. doi:10.1006/dbio.2001.0501.

Sasse, J., U. Hemmann, C. Schwartz, U. Schniertshauer, B. Heesel, C. Landgraf, J. Schneider-Mergener, P.C. Heinrich, and F. Horn. 1997. Mutational analysis of acute-phase response factor/Stat3 activation and dimerization. *Mol. Cell. Biol.* 17:4677–4686. doi:10.1128/mcb.17.8.4677.

Savatier, P., S. Huang, L. Szekely, K. Wiman, and J. Samarut. 1994. Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. *Oncogene*. 9:809–18.

Savatier, P., H. Lapillonne, L.A. van Grunsven, B.B. Rudkin, and J. Samarut. 1996. Withdrawal of differentiation inhibitory activity/leukemia inhibitory factor up-regulates D-type cyclins and cyclin-dependent kinase inhibitors in mouse embryonic stem cells. *Oncogene*. 12:309–22.

Schlesinger, S., and E. Meshorer. 2019. Open Chromatin, Epigenetic Plasticity, and Nuclear Organization in Pluripotency. *Dev. Cell*. 48:135–150. doi:10.1016/j.devcel.2019.01.003.

Schmidt, C.S., S. Bultmann, D. Meilinger, B. Zacher, A. Tresch, K.C. Maier, C. Peter, D.E. Martin, H. Leonhardt, and F. Spada. 2012. Global DNA Hypomethylation Prevents Consolidation of Differentiation Programs and Allows Reversion to the Embryonic Stem Cell State. 7. doi:10.1371/journal.pone.0052629.

Schwartz, S.D., J.P. Hubschman, G. Heilwell, V. Franco-Cardenas, C.K. Pan, R.M. Ostrick, E. Mickunas, R. Gay, I. Klimanskaya, and R. Lanza. 2012. Embryonic stem cell trials for macular degeneration: A preliminary report. *Lancet*. 379:713–720. doi:10.1016/S0140-6736(12)60028-2.

Sela, Y., N. Molotski, S. Golan, J. Itskovitz-Eldor, and Y. Soen. 2012. Human embryonic stem cells exhibit increased propensity to differentiate during the G1 phase prior to phosphorylation of retinoblastoma protein. *Stem Cells*. 30:1097–1108. doi:10.1002/stem.1078.

Shen, L., H. Wu, D. Diep, S. Yamaguchi, A.C. D'Alessio, H.-L. Fung, K. Zhang, and Y. Zhang. 2013. Genome-wide analysis reveals TET-and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell*. 153:692–706.

Shiba, Y., S. Fernandes, W.Z. Zhu, D. Filice, V. Muskheli, J. Kim, N.J. Palpant, J. Gantz, K.W. Moyes, H. Reinecke, B. Van Biber, T. Dardas, J.L. Mignone, A. Izawa, R. Hanna, M. Viswanathan, J.D. Gold, M.I. Kotlikoff, N. Sarvazyan, M.W. Kay, C.E. Murry, and M.A. Laflamme. 2012. Human ES-cell-derived cardiomyocytes electrically

couple and suppress arrhythmias in injured hearts. *Nature*. 489:322–325. doi:10.1038/nature11317.

Shufaro, Y., and B.E. Reubinoff. 2004. Therapeutic applications of embryonic stem cells. 18:909–927. doi:10.1016/j.bpobgyn.2004.07.002.

Singh, A.M. 2015. Cell cycle-driven heterogeneity: On the road to demystifying the transitions between “poised” and “restricted” pluripotent cell states. *Stem Cells Int.* 2015. doi:10.1155/2015/219514.

Singh, A.M., J. Chappell, R. Trost, L. Lin, T. Wang, J. Tang, H. Wu, S. Zhao, P. Jin, and S. Dalton. 2013. Cell-cycle control of developmentally regulated transcription factors accounts for heterogeneity in human pluripotent cells. *Stem Cell Reports*. 1:532–544. doi:10.1016/j.stemcr.2013.10.009.

Singh, A.M., and S. Dalton. 2009. The Cell Cycle and Myc Intersect with Mechanisms that Regulate Pluripotency and Reprogramming. *Cell Stem Cell*. 5:141–149. doi:10.1016/j.stem.2009.07.003.

Singh, A.M., Y. Sun, L. Li, W. Zhang, T. Wu, S. Zhao, Z. Qin, and S. Dalton. 2015. Cell-Cycle Control of Bivalent Epigenetic Domains Regulates the Exit from Pluripotency. *Stem Cell Reports*. 5:323–336. doi:10.1016/j.stemcr.2015.07.005.

Smith, A.G., J.K. Heath, D.D. Donaldson, G.G. Wong, J. Moreau, M. Stahl, and D. Rogers. 1988. Inhibition of pluripotential embryonic stem cell differentiation by

purified polypeptides. *Nature*. 336:688–690.

Snow, M.H.L. 1977. Gastrulation in the mouse: Growth and regionalization of the epiblast. *Development*. 42:293–303.

Song, C.-X., C. Yi, and C. He. 2012. Mapping New Nucleotide Variants in the Genome and Transcriptome. *Nat Biotechnol.* 30:1107–1116. doi:10.1038/nbt.2398.Mapping.

Song, C., K.E. Szulwach, Q. Dai, Y. Fu, S. Mao, L. Lin, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, and L. Li. 2013. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. 153:678–691. doi:10.1016/j.cell.2013.04.001.Genome-wide.

Song, W.K., K.M. Park, H.J. Kim, J.H. Lee, J. Choi, S.Y. Chong, S.H. Shim, L. V. Del Priore, and R. Lanza. 2015. Treatment of macular degeneration using embryonic stem cell-derived retinal pigment epithelium: Preliminary results in Asian patients. *Stem Cell Reports*. 4:860–872. doi:10.1016/j.stemcr.2015.04.005.

Stadler, M.B., R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Scholer, E. Van Nimwegen, C. Wirbelauer, E.J. Oakeley, D. Gaidatzis, V.K. Tiwari, and D. Schubeler. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. 480:490–495. doi:10.1038/nature10716.

Stahl, N., T.G. Boulton, T. Farruggella, N.Y. Ip, S. Davis, B. A, F.W. Quelle, O. Silvennoinen, G. Barbieri, S. Pellegrini,

J.N. Ihle, G.D. Yancopoulos, S. Science, N. Series, N. Jan, and B.A. Witthuhn. 1994. Association and Activation of Jak-Tyk Kinases by I Receptor Components CNTF-LIF-OSM-IL-6. *Adv. Sci.* 263:92–95.

Stavridis, M.P., J.S. Lunn, B.J. Collins, and K.G. Storey. 2007. A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. *Development*. 134:2889–2894. doi:10.1242/dev.02858.

Stead, E., J. White, R. Faast, S. Conn, S. Goldstone, J. Rathjen, U. Dhingra, P. Rathjen, D. Walker, and S. Dalton. 2002. Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. *Oncogene*. 21:8320–8333. doi:10.1038/sj.onc.1206015.

Stroud, H., S. Feng, S. Morey Kinney, S. Pradhan, and S.E. Jacobsen. 2011. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.* 12:R54. doi:10.1186/gb-2011-12-6-r54.

Subramanian, A., P. Tamayo, V.K. Mootha, S. Mukherjee, and B.L. Ebert. 2005. Gene set enrichment analysis : A knowledge-based approach for interpreting genome-wide.

Szulwach, K.E., X. Li, Y. Li, C.X. Song, J.W. Han, S.S. Kim, S. Namburi, K. Hermetz, J.J. Kim, M.K. Rudd, Y.S. Yoon, B. Ren, C. He, and P. Jin. 2011. Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.* 7.

doi:10.1371/journal.pgen.1002154.

Tahiliani, M., K.P. Koh, Y. Shen, W. a Pastor, Y. Brudno, S. Agarwal, L.M. Iyer, R. David, L. Aravind, and A. Rao. 2009. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. 324:930–935. doi:10.1126/science.1170116.Conversion.

Tai, C.-I., E.N. Schulze, and Q.-L. Ying. 2014. Stat3 signaling regulates embryonic stem cell fate in a dose-dependent manner. *Biol. Open*. 3:958–965. doi:10.1242/bio.20149514.

Takahashi, K., and S. Yamanaka. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 126:663–76. doi:10.1016/j.cell.2006.07.024.

Takahashi, T., R. Nowakowski, and V. Caviness. 1995. The Cell Cycle of the Pseudostratified Embryonic Murine Cerebral Wall. *J. Neurosci.* 15:6046–6057.

Tam, P.P.L., and J. Rossant. 2003. Mouse embryonic chimeras: Tools for studying mammalian development. *Development*. 130:6155–6163. doi:10.1242/dev.00893.

Taruttis, F., M. Feist, P. Schwarzfischer, W. Gronwald, D. Kube, R. Spang, and J.C. Engelmann. 2017. External calibration with Drosophila whole-cell spike-ins delivers absolute mRNA fold changes from human RNA-Seq and qPCR data. doi:10.2144/000114514.

Tesar, P.J., J.G. Chenoweth, F.A. Brook, T.J. Davies, E.P.

Evans, D.L. Mack, R.L. Gardner, and R.D.G. Mckay. 2007. New cell lines from mouse epiblast share defining features with human embryonic stem cells. 448:196–202. doi:10.1038/nature05972.

Thomson, J.A., J. Itskovitz-Eldor, S.S. Shapiro, M.A. Waknitz, J.J. Swiergiel, V.S. Marshall, and J.M. Jones. 1998. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science (80-.).* 282:1145–1147. doi:10.1126/science.282.5391.1145.

Tolosa, L., J. Caron, Z. Hannoun, M. Antoni, S. López, D. Burks, J.V. Castell, A. Weber, M.J. Gomez-Lechon, and A. Dubart-Kupperschmitt. 2015. Transplantation of hESC-derived hepatocytes protects mice from liver injury. *Stem Cell Res. Ther.* 6:1–17. doi:10.1186/s13287-015-0227-6.

Toyooka, Y., D. Shimosato, K. Murakami, K. Takahashi, and H. Niwa. 2008. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development*. 135:909–918. doi:10.1242/dev.017400.

Trapnell, C., L. Pachter, and S.L. Salzberg. 2009. TopHat : discovering splice junctions with RNA-Seq. 25:1105–1111. doi:10.1093/bioinformatics/btp120.

Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, H. Pimentel, S.L. Salzberg, J.L. Rinn, and L. Pachter. 2013. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. 7:562–578.

doi:10.1038/nprot.2012.016.Differential.

Tsumura, A., T. Hayakawa, Y. Kumaki, and S. Takebayashi. 2006. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1 , Dnmt3a and Dnmt3b. 805–814. doi:10.1111/j.1365-2443.2006.00984.x.

Vizcaíno, J.A., A. Csordas, N. Del-Toro, J.A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-riverol, F. Reisinger, T. Ternent, Q. Xu, R. Wang, and H. Hermjakob. 2016. 2016 update of the PRIDE database and its related tools. 44:447–456. doi:10.1093/nar/gkv1145.

Volarevic, V., B.S. Markovic, M. Gazdic, A. Volarevic, N. Jovicic, N. Arsenijevic, L. Armstrong, V. Djonov, M. Lako, and M. Stojkovic. 2018. Ethical and safety issues of stem cell-based therapy. *Int. J. Med. Sci.* 15:36–45. doi:10.7150/ijms.21666.

Wang, Y., S. Baskerville, A. Shenoy, J.E. Babiarz, L. Baehner, and R. Blelloch. 2008. Embryonic stem cell–specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat. Genet.* 40:1478–1483. doi:10.1038/ng.250.Embryonic.

Wilder, P.J., D. Kelly, K. Brigman, C.L. Peterson, T. Nowling, Q. Gao, R.D. Mccomb, M.R. Capecchi, and A. Rizzino. 1997. Inactivation of the FGF-4 Gene in Embryonic Stem Cells Alters the Growth and/or the Survival of Their Early Differentiated Progeny. 629:614–629.

Williams, R.L., D.J. Hilton, S. Pease, T.A. Willson, C.L.

Stewart, D.P. Gearing, E.F. Wagner, D. Metcalf, N.A. Nicola, and N.M. Gough. 1988. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature*. 336:684–687.

Wisniewski, J.R., A. Zougman, N. Nagaraj, and M. Mann. 2009. Universal sample preparation method for proteome analysis. 6:3–7. doi:10.1038/NMETH.1322.

Wong, C.W., P.S. Hou, S.F. Tseng, C.L. Chien, K.J. Wu, H.F. Chen, H.N. Ho, S. Kyo, and S.C. Teng. 2010. Krüppel-like transcription factor 4 contributes to maintenance of telomerase activity in stem cells. *Stem Cells*. 28:1510–1517. doi:10.1002/stem.477.

Wray, J., T. Kalkan, S. Gomez-Lopez, D. Eckardt, A. Cook, R. Kemler, and A. Smith. 2011. Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat. Cell Biol.* 13:838–845. doi:10.1038/ncb2267.

Xia, B., D. Han, X. Lu, Z. Sun, A. Zhou, Q. Yin, H. Zeng, M. Liu, X. Jiang, W. Xie, C. He, and C. Yi. 2015. Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. *Nat. Methods*. 12:1047–1050. doi:10.1038/nmeth.3569.

Xu, X., S. Duan, F. Yi, A. Ocampo, G.H. Liu, and J.C. Izpisua Belmonte. 2013. Mitochondrial regulation in pluripotent stem cells. *Cell Metab.* 18:325–332. doi:10.1016/j.cmet.2013.06.005.

Yarden, A., and B. Geiger. 1996. Zebrafish cyclin E regulation during early embryogenesis. 11.

Ye, S., P. Li, C. Tong, and Q.L. Ying. 2013. Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. *EMBO J.* 32:2548–2560. doi:10.1038/emboj.2013.175.

Yiangou, L., R. a. Grandy, C.M. Morell, R. a. Tomaz, A. Osnato, J. Kadiwala, D. Muraro, J. Garcia-Bernardo, S. Nakanoh, W.G. Bernard, D. Ortmann, D.J. McCarthy, I. Simonic, S. Sinha, and L. Vallier. 2019. Method to Synchronize Cell Cycle of Human Pluripotent Stem Cells without Affecting Their Fundamental Characteristics. *Stem Cell Reports.* 12:165–179. doi:10.1016/j.stemcr.2018.11.020.

Ying, Q.L., J. Wray, J. Nichols, L. Batlle-Morera, B. Doble, J. Woodgett, P. Cohen, and A. Smith. 2008. The ground state of embryonic stem cell self-renewal. *Nature*. 453:519–523. doi:10.1038/nature06968.

Young, M.D., T.A. Willson, M.J. Wakefield, E. Trounson, D.J. Hilton, M.E. Blewitt, A. Oshlack, and I.J. Majewski. 2011. ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Res.* 39:7415–7427. doi:10.1093/nar/gkr416.

Zhang, Y., T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, W. Li, and X.S. Liu. 2008. Open Access Model-based Analysis of ChIP-Seq (MACS). doi:10.1186/gb-2008-9-

9-r137.

Zhong, Z., Z. Wen, and J.E.D. Jr. 1994. Stat3: A STAT family member activated by tyrosine phosphorylation in response to epidermal growth factor and interleukin-6. 264:95–98.

RESEARCH ARTICLES

A manuscript from this thesis' results is under preparation.

From a collaboration in the lab the following paper was published:

Aranda, S., A. Alcaine-Colet, E. Blanco, E. Borràs, C. Caillot, E. Sabidó, and L. Di Croce. 2019. Chromatin caputure links the metabolic enzyme AHCY to stem cell proliferation. *Sci. Adv*. 5. doi:10.1126/sciadv.aav2448

Moreover, the following review article has been submitted:

Blanco, E., M. González-Ramírez, A. Alcaine-Colet, S. Aranda and L. Di Croce. 2019. The bivalent genome: characterization, structure and regulation. *Trends Genet.*

ACKNOWLEDGEMENTS

I thank all the membres from the present and past of the Di Croce family for the unconditional suport during my PhD, especially to Sergi Aranda and Luciano Di Croce for the close supervision. I will miss all of you and hope the best for your futures.
I thank our collaborators Holger Heyn and Atefeh Lafzi from CNAG for the single cell RNA-sequencing experiment performance and analysis and discussions; and Eduard Sabido and Eva Borràs from the Prometomics Unit at CRG and UPF for the proteomic analysis performed and discussions.
I thank Laura Batlle and Marta Vila from the Tissue Engineering Unit at CRG for carrying out the teratoma and chimera assays and their analysis; Òscar Fornas, Alexandre Bote, Erika Ramírez, Eva Julià from the Flow Cytometry Unit at CRG and UPF for the technical assistance and advise for FACS analysis and FACS sorting; the Genomics Unit from the CRG for assistance with sequencing.

This work received the support of the Secretary for Universities and Research of the Ministry of Economy and Knowledge of the Government of Catalonia. This work in the the laboratory was supported by grants from the Spanish “Ministerio de Educación y Ciencia” (SAF2013-48926-P), La Marato TV3, the Fundación Vencer El Cáncer (VEC) and from the European Commission's 7th Framework Program 4DCellFate grant number 277899.