The Cell Cycle and Pluripotency

Menno C. ter Huurne

Radboud Universiteit Nijmegen

The Cell Cycle and Pluripotency

Proefschrift

ter verkrijging van de graad van doctor

aan de Radboud Universiteit Nijmegen

op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,

volgens besluit van het college van decanen

in het openbaar te verdedigen op 14-11-2018

om 12.30 uur precies

door

Menno C. ter Huurne

geboren op 12 oktober 1983

te Enschede

Contents

Chapter 0

General Introduction

Embryonic Development

Almost all forms of life, as most of us are familiar with, are the result of a close collaboration between billions of cells. Together, these cells form the animal body plan with numerous highly specialized organs and tissues. This beautiful and at the same time extremely complex cooperative multicellular entity has been build up from an 1-cell embryo that was born upon the fusion of an oocyte and sperm cell. The subsequent development of a multicellular organism comprehends two tightly coordinated processes that run in parallel: the steady increase in cell numbers through serial cell divisions and the tightly controlled cell specification through (epi)genetic regulation of gene expression. During cell division cells copy their genetic material and divide into two genetically identical daughters cells, a process that is already repeated roughly 40 times during human embryonic development prior to birth. Upon fertilization the oocyte starts to proliferate and at the morula stage, when the embryo consists of 16-32 cells, these cells from a ball-like structure. Characteristic to this initial phase of embryonic development is the rapid expansion of cell numbers while the volume of the embryo remains constant, resulting in compaction of the cells. Albeit all these cells contain the exact same genetic material, the fate of these cells branches for the first time: the outer trophoblast cells will develop into the extra-embryonic tissue whereas the inner embryoblast cells will eventually form the embryo and yolksac. Because the trophoblast cells produce a fluid that is stored within the ball-like structure, a cavity is formed, the blastocoel. Within the so-called blastocyst the embryoblasts clump together on one side of the embryo, forming a group of cells that is called the Inner Cell Mass (ICM). At this stage ICM cells from the early blastocyst will undergo global epigenetic reprogramming that results in global erasure of maternally and paternally inherited epigenetic marks (reviewed by Wu, H. & Zhang[1]). At the same time Fibroblast Growth Factor (FGF)/Extracellular signal-Regulated Kinase (ERK) signaling in the ICM drives differential expression of the key Transcription Factors (TFs) NANOG and GATA6 that will instruct the establishment of the epiblast, that will form the embryo, and the extra-embryonic hypoblast, respectively (reviewed by Frum, T. & Ralston[2]). Around time of implantation of the blastocyst in the uterine wall another cavity (the amniotic cavity) forms and the epiblast reorganizes in a cup-shaped structure, the egg cylinder. This structure is composed of a double cell layer of primed epiblast and primitive endoderm cells. The epiblast cells will form the ectoderm whereas the mesoderm and definitive endoderm will subsequently be formed through invagination and migration of ectodermal cells, processes tightly controlled by TGF-β and Wnt signaling (reviewed by Beddington, R. S. P. & Robertson, E. J.[3]). These three germ layers will eventually give rise to all cell types of the animal body plan.

Embryonic Stem Cells; what's the fuss about?

Prior to germ layer specification cells from the blastocyst reside for a limited period of time in an undifferentiated state and are called Embryonic Stem Cells (ESCs). The establishment of in vitro culture conditions that allowed scientist to capture these cells in their undifferentiated state in the early 80s of the last century opened new exciting possibilities for research[4]. Similar to their in vivo counterparts, in vitro

cultured ESCs are pluripotent, meaning that they have the potential to differentiate in any kind of cell type of the fully developed body plan. The ultimate proof that these in vitro propagated ESCs are truly pluripotent comes from the fact that they will continue to contribute to embryonic development and the eventual body plan upon injection in the ICM. An ESC-specific TF network regulates the expression of genes important for the maintenance of pluripotency and the repression of lineage specifying genes. At the core of this network act the TFs OCT4, NANOG and SOX2, which have overlapping genomic binding sites and coordinately ensure maintenance of pluripotency[5]. Intensive research on the signaling pathways that mediate cell fate specification has resulted in a wide range of protocols that provide efficient differentiation of pluripotent stem cells into cells from all three lineages (reviewed by Bartfeld, S. & Clevers, H.[6]).

Another characteristic of ESCs is there unrestrictive proliferative capacity which facilitates the cultivation of large numbers of cells. Although new methods are being developed to study molecular processes on single cell level, large cell quantities facilitate a variety of biochemical analysis to decipher molecular processes involved in development and disease, like gene expression and transcription factor binding. Moreover, the virtually unlimited number of ESCs that can be obtained is crucial in regenerative medicine, where the ultimate goal is to use stem cells to replace or support lost or damaged tissues[7]. Using well-defined differentiation conditions in combination with 3D culture platforms scientist have been able to grow numerous fully differentiated mini-organs from stem cells (reviewed by Bartfeld, S. & Clevers, H.[6]). These so-called organoids might not only serve to replace lost tissue but also allow scientists to study the role of genes in developmental and disease. In this regard recent advances in genome wide targeted gene editing techniques offer new exciting possibilities to study genes and regulatory elements in development and disease in an unbiased approach[8–10].

The above-mentioned properties and potential clinical applications have driven scientist to unravel both the molecular mechanisms employed by ESCs to remain pluripotency as well as the pathways that regulate ESC proliferation. Interestingly, research over the past decades has shown that members of the cell cycle regulatory network can regulate the pluripotency network and vice versa (reviewed by Kareta, M. S., Sage, J. & Wernig, M.[11]). One of the main goals of this thesis is to comprehensively describe how the ESC cell cycle is regulated. To this end a basic understanding of the mammalian cell cycle is crucial and I will therefor first elaborate on this subject.

The Cell Cycle; the Basics

In proliferating cells the mitotic phase, during which the cell divides, alternates with an interphase. During interphase cells grow, replicate their DNA and prepare for mitosis. The interphase can be subdivided in three phases, gap-phase 1 (G1), the DNA synthesis-phase (S) and gap-phase 2 (G2). In the currently prevailing view progression through the cell cycle is driven by protein complexes that consist of a cyclin and a Cyclin Dependent Kinase (CDK). The cyclins, as their name already implies, are a family of proteins first discovered in sea urchins that undergo cyclic expression during the course of a cell cycle[12]. The cyclins are

highly conserved throughout the animal kingdom and at present over 25 mammalian proteins containing a 'cyclin box' have been identified with a wide range of functions[13]. The primary function of the best studied cyclins is however considered to be the activation of CDKs. Pioneering studies in yeast and frogs identified CDKs as mediators of cell cycle progression, but work over the past three decades has identified CDKs and CDK-like proteins that function in either cell cycle progression or in transcriptional regulation (reviewed by Malumbres, M.[14]). Although the current model designates the activity of different CDK/Cyclin complexes to different phases of the cell cycle it has to be mentioned that there is a significant degree of plasticity in both the role of cyclins as well as CDKs. CDK1 can for example interact with CYCLIN D and CYCLIN E, can functionally replace CDK2-4 and is the only CDK essential for the first cell divisions after fertilization[15]. Similarly, the combined deletion of all Cyclin D and Cyclin E isoforms does not necessarily stop the cell cycle[16].

In their active configuration the pocket proteins RB, P107 and P130 bind and inhibit chromatin-bound E2F TFs. Waves of cell cycle phase-specific cyclins bind to and activate CDKs that in turn activate E2F-mediated transcription through phosphorylation and inactivation of the pocket proteins[17]. To date, nine structurally related proteins have been classified as the E2F protein family. In general these members are considered to either mediate transcriptional activation (E2F1, E2F2, E2F3a & E2F3b) or repression (E2F4, E2F5, E2F6, E2F7 & E2F8) although switching between both functions has been observed indicating that this model is simplified[18]. The E2F transcription factors drive the expression of a suite of proteins needed for the next phase of the cell cycle.

G1-phase

Cells in G1-phase grow and produce proteins necessary for DNA replication. Progression through the G1-phase is dependent on extracellular mitogens that stimulate the ERK signaling pathway. Upon stimulation of FGF membrane receptors consecutive phosphorylation steps culminate in the activation of transcription factors, amongst others MYC, that regulate the expression of cell cycle mediators like CYCLIN D[19]. In the widely accepted model an initial burst of CYCLIN D/CDK4-6 activity results in phosphorylation and inactivation of the pocket proteins, RB, P107 and P130, thereby alleviating repression of the E2F transcription factors[20]. Consequently, E2F activity results in the expression of Cyclin E and Cdc25a, a phosphatase that removes inhibitory phosphate groups from CDK2. Subsequently, CDK2 partners with Cyclin E into a complex that further increases the phosphorylation of the pocket proteins, thereby participating in a feed forward loop that ensures progression into S-phase[21]. It is assumed that this feed forward loop results in a bimodal E2F transcription factor activity[19,22] and marks a specific point during G1-phase at which is switched to either quiescence or growth, the so-called "Restriction point"[23]. If extrinsic mitogenic stimuli result in sufficient CYCLIN D/CDK4-6 activity to induce Cyclin E/CDK2 activity, cells cross this point and will complete a full cell cycle, irrespective of the availability of extracellular stimuli. Interestingly though, malignant cells as well as certain primary cells lines do not have a restriction point and proceed to S-phase in the absence of serum stimulation[19,21,23].

Alternative to progressing through G1-phase, cells can also (ir)reversibly enter a non-proliferative quiescent state. Cellular quiescence can protect cells from stress and toxicity and is of particular importance for long lived cells like adult stem cells and terminally differentiated cells[24]. In case of lack of nutrients or in case of DNA damage, several pathways are able to halt G1-progression to prevent these cells from entering S-phase in order to maintain genomic integrity. P53 plays a key role in halting G1-phase progression. After its discovery in the late 70s / early 80s of the last century it became clear that P53 is an important tumor suppressor gene and mutated in a wide range of tumours[25–27]. Upon DNA damage or stress P53 controls several responses that can lead to either apoptosis or cell cycle quiescence[28,29]. P53 can prevent cell cycle commitment either by counteracting the expression of cyclins[30] or by regulating the expression of several Cyclin-Dependent Kinase inhibitor's (CDKi's). The latter interfere with CDK/cyclin complex formation and thereby abrogate G1-phase progression. P53 is a direct transcriptional activator of Cdkn1A, the first mammalian CDKi described that codes for the protein P21. The P53-P21 axis is now being recognized as a major pathway in cell cycle control. P21 together with P27 and P57 belong to the Cip and Kip family of proteins that prohibit cell cycle progression through a multitude of mechanisms. Although these proteins are able to directly regulate gene transcription they are best known for their ability to prohibit RB phosphorylation by competing with CDK1-2/CYCLIN E complex formation. This process will delay S-phase entry and allow DNA damage repair (reviewed by Abbas, T. & Dutta, A.[31]).

S-phase

Once the quality of DNA is ensured and the cells have passed the restriction point and committed to another cell cycle, the DNA is replicated during Synthesis (S) phase. Replication of the DNA starts at particular DNA sequences, so called origins of replication (Ori's), from where replication forks move in both directions. The sequence with which these Ori's fire is tightly regulated (known as the replication timing program) and highly specific for each cell type[32]. Faithful replication of the DNA can be compromised by various endogenous and exogenous conditions[33]. In concordance with the G1 checkpoint, several surveillance pathways are active during S-phase to either recognize flawed DNA replication or signal when DNA replication forks stall[34].

G2-phase

Prior to cell division the cell increases in size and synthesize proteins needed for division during G2-phase. Progression through this phase is mediated by CYCLIN B/CDK1 complexes and involves a positive feedback mechanism driven by the CDC25 proteins (reviewed by Schmidt, M. *et al.*[35]). Interestingly, recent studies have indicated that decisions on G1-phase progression are already made during G2-phase of the preceding cell cycle. Both mediators of mitogenic signaling as well as stress response pathways in the mother cell are transmitted upon mitosis influencing progression through G1 phase of the newborn daughter cells[36]. ERK signaling during G2 phase results in elevated levels of cyclin E RNA which are transferred during mitosis and abolish the restriction point in the newborn daughter cells[37]. Another very recent study

indicates that not only Cyclin D RNA but also DNA damage-induced P53 protein is transferred from mother to daughter cells and influence G1-progression. Activated P53 results in higher levels of Cdkn1a RNA in the daughter cells which upon translation counteracts CYCLIN D/CDK2 activity[30].

Back to the Embryonic Stem Cell and its cell cycle

During early embryonic development cell proliferation and migration are tightly controlled. Cells in the ICM of the early blastocyst proliferate at a rapid pace with doubling times of around 8 to 10 hours. During the peri-implantation stage these cells start to proliferate even faster, resulting in a doubling time as short as roughly 4 hours. Early studies using ESCs as in vitro counterparts of ICM cells revealed that these cell cycle characteristics can be attributed to the precocious expression of CDKs, in particular CDK2, and their catalytically active subunits (the cyclins), resulting in an absence of hypo-phosphorylated RB protein[38,39]. Furthermore, ESCs lack the expression of the CDKi's and, if expression of the CDK-inhibitor P16 is forced, seem insensitive to CDKi activity. This could also explain why proliferation of ESCs is unaffected neither by serum starvation nor by contact inhibition. As a result of this characteristic cell cycle regulatory network, mouse ESCs reside only for a short period of time in G1 phase, which is reflected by a high number of cells in S-phase when compared to somatic cells[39]. Because several studies have indicated that differentiation is initiated during G1 phase, reflected by changes in the epigenetic landscape and expression of developmental genes, it has been postulated that the characteristic short G1-phase protects ESCs from differentiating[40–42]. On the other hand, however, it was shown that artificial elongation of G1-phase did not affect pluripotency[43] and certain signaling pathways not only ensure maintenance of pluripotency but elongate the G1-phase as well[44,45]. Additionally, a recent study revealed that P53-driven enhancement of TGF-β signaling blocks pluripotent state dissolution specifically during S and G2 phase, which forms an alternative explanation for the fact that ESCs are more prone to differentiate during G1 phase[46]. So, it is still not entirely clear to what extend the short G1-phase contributes to the pluripotency network in ESCs or is merely a consequence of the ESC-specific transcriptional program.

Serum and 2i ESCs

Although the isolation and propagation of ESCs on growth-arrested Mouse Embryonic Fibroblasts (MEFs) was first described almost 50 years ago the elucidation of the exact pathways that ensure maintenance of pluripotency and allow in vitro culturing of ESCs is still ongoing. Seminal research in the late 80s led to the discovery that the addition of LIF potently inhibits ESC differentiation and could replace MEFs, allowing long-term propagation of naive ESCs in the absence of feeders[47,48]. However, the in vitro expansion of pluripotent ESCs remained dependent on the addition of serum components. Disadvantages of the serum include batch-to-batch variation that causes variation in cell culture conditions. Moreover, the addition of this black-box containing a mixture of unknown growth factors and signaling molecules could potentially

hamper the identification of true ESC characteristics. Research over the last couple of decades resulted in the identification of several key signaling pathways involved in stem cell maintenance and differentiation[49,50]. Using this knowledge, a method to propagate ESCs in vitro without the need for either MEFs or serum was developed in 2008. This new culture system allows maintenance of pluripotent ESCs in vitro solely by shielding them from differentiation-inducing cues proving that "extrinsic stimuli are dispensable for the derivation, propagation and pluripotency of ES cells"[51]. The inhibition of FGF/ERK signaling by PD0325901 (hereafter referred to as "PD") protects ESCs from auto-inductive FGF whereas CHIR9021 ("Chiron") inhibits GSK3-α and –β and restores biosynthetic pathways essential for self-renewal. In addition, Chiron prevents neural differentiation presumably through the activation of TGF-β signaling and relieving repression of amongst others Esrrb and Nanog[51,52]. Several lines of research have been set out to compare ESCs cultured in conditions containing either serum or the two small molecule inhibitors (referred to as "naive serum ESCs" and "ground state 2i ESCs", respectively) and revealed some remarkable differences. Although genes of the core pluripotency network are expressed to a similar level, some factors that contribute to maintenance of pluripotency in naive serum ESCs, e.g. MYC and the ID proteins, are lower expressed in 2i ESCs. The expression of lineage specifiers is however higher in serum ESCs and might be hold accountable for the fact that serum ESCs are more heterogeneous than 2i ESCs in terms of morphology[53,54]. Along with data on the epigenetic landscape of both ESC populations that indicate a higher number of bivalent domains which are known to concur with developmental genes suggesting that serum ESCs are primed for differentiation. Assessment of genome-wide DNA methylation, an epigenetic modification that correlates with differentiation, revealed that ground state 2i ESCs have low DNA methylation levels that resemble the pre-implantation ICM whereas naive serum ESCs have elevated levels of DNA methylation[55–57]. Together these results have led scientists to believe that naive serum and ground state 2i ESCs reflect pluripotent cells from pre- and post-implantation stages, respectively[54]. Another striking difference between serum and 2i ESCs is the expression of proteins of the cell cycle machinery[53]. In general, cyclin-dependent kinases are higher expressed in serum ESCs and the CDK-inhibitors P16, P19 and P21 are exclusively present in 2i ESCs. The studies described in this thesis will shed light on how these differences affect the ESC cell cycle and relate to developmental stages in vivo.

The Cell Cycle and Colorectal Cancer

Not only during development but also in cells of the fully developed body, a tightly regulated cell cycle is of utmost importance. In most tissues a population of adult stem cells exists that is needed for tissue homeostasis and replenishment of lost cells. Although these cells in general reside in a non-proliferative quiescent state they can be stimulated to produce cells that differentiate and repopulate the organ[24]. Dysregulated expression of genes involved in cell cycle progression often cause uncontrolled proliferation. Uncontrolled proliferation is one of the hallmarks of cancer and mutations that contribute to cancer development are often found in genes that regulate progression through G1-phase[58]. Although mutations

in these coding regions often result in mis-expression of the affected gene, Genome-Wide Association Studies (GWAS) have revealed that mutations linked to an increased risk of cancer are often located in genomic regions that are not transcribed. Possibly, these non-coding regions regulate the expression of genes located down- or up-stream. Although a great number of Single Nucleotide Polymorphisms (SNPs) in regulatory regions associated with cancer predisposition has been identified, it is often unknown how they affect gene expression. Presumably, SNPs in non-coding regions affect the binding of TFs thereby affecting the expression of nearby genes.

Outline of this thesis

Considering the role of the CDK/Cyclin pathway in G1-phase progression recently observed differences between serum and 2i ESCs imply that the cell cycle of the latter is different from that of the serum ESC-specific short G1-phase. In **chapter 1** we describe how the differential expression of cyclins and CDKs impact on the cell cycle of 2i ESCs. We will also elucidate on the signaling pathways that drive these differences.

Although DNA damage pathways are abrogated in mouse ESCs, some of the results we describe in chapter 2 imply that the role of P53 in 2i ESCs is different from that in serum ESCs. In **chapter 2** we set out to investigate the role of the P53 pathway in 2i ESCs.

In **chapter 3** we will briefly review how the cell cycle of mouse serum ESCs is different from 2i ESCs. We will consider how the pluripotency network contributes to the abbreviated G1-phase. In addition, we will discuss how members of the G1-phase control pathway directly interact with the core pluripotency network. Many SNPs have been identified within non-coding regions that increase the risk of an individual to develop Colorectal Cancer (CRC). In **chapter 4** we set out to investigate how these mutations affect Transcription Factor (TF) binding. To assess the effect of SNPs in CRC risk loci we performed a Proteome-Wide Analysis of SNPs (PWAS). Using this quantitative mass spectrometry approach on a library of 116 susceptibility loci we identified several candidate TFs that showed favored binding in either the WT locus or in the one harboring the SNP. In addition, we describe several follow up experiments that unravel a potential mechanism by which SNP increases the risk on CRC development.

References

1. Wu, H. & Zhang, Y. Reversing DNA methylation: Mechanisms, genomics, and biological functions. *Cell* **156,** 45–68 (2014).

2. Frum, T. & Ralston, A. Cell signaling and transcription factors regulating cell fate during formation of the mouse blastocyst. *Trends Genet.* **31,** 402–410 (2015).

3. Beddington, R. S. P. & Robertson, E. J. Axis development and early asymmetry in mammals. *Cell* **96,** 195–209 (1999).

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

5. Chambers, I. & Tomlinson, S. R. The transcriptional foundation of pluripotency. *Development* **136,** 2311–2322 (2009).

6.   Bartfeld, S. & Clevers, H. Stem cell-derived organoids and their application for medical research and patient treatment. *J. Mol. Med.* **95,** 729–738 (2017).

7.   Wu, J. *et al.* Stem cells and interspecies chimaeras. *Nature* **540,** 51–59 (2016).

8.   Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32,** 267–273 (2014).

9.   Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems_Sup. *Science* **339,** 819–23 (2013).

10.   Driehuis, E. & Clevers, H. CRISPR/Cas 9 genome editing and its applications in organoids. *Am. J. Physiol. - Gastrointest. Liver Physiol.* **312,** G257–G265 (2017).

11.   Kareta, M. S., Sage, J. & Wernig, M. Crosstalk between stem cell and cell cycle machineries. *Curr. Opin. Cell Biol.* **37,** 68–74 (2015).

12.   Evans, T., Rosenthal, E. T., Youngblom, J., Distel, D. & Hunt, T. Cyclin: A protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. *Cell* **33,** 389–396 (1983).

13.   Malumbres, M. & Barbacid, M. Mammalian cyclin-dependent kinases. *Trends Biochem. Sci.* **30,** 630–41 (2005).

14.   Malumbres, M. Cyclin-dependent kinases. *Genome Biol.* **15,** 1–10 (2014).

15.   Santamaría, D. *et al.* Cdk1 is sufficient to drive the mammalian cell cycle. *Nature* **448,** 811–815 (2007).

16.   Liu, L. *et al.* G1 cyclins link proliferation, pluripotency and differentiation of embryonic stem cells. *Nat. Cell Biol.* **19,** (2017).

17.   Cobrinik, D. Pocket proteins and cell cycle control. *Oncogene* **24,** 2796–2809 (2005).

18.   Chong, J.-L. *et al.* E2f1-3 switch from activators in progenitor cells to repressors in differentiating cells. *Nature* **462,** 930–4 (2009).

19.   Schwarz, C. *et al.* A Precise Cdk Activity Threshold Determines Passage through the Restriction Point. *Mol. Cell* **69,** 253–264.e5 (2018).

20.   Hsu, J. & Sage, J. Novel functions for the transcription factor E2F4 in development and disease. *Cell Cycle* **15,** 3183–3190 (2016).

21.   Spencer, S. L. *et al.* The proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit. *Cell* **155,** 369–383 (2013).

22.   Yao, G., Lee, T. J., Mori, S., Nevins, J. R. & You, L. A bistable Rb-E2F switch underlies the restriction point. *Nat. Cell Biol.* **10,** 476–482 (2008).

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

24.   Cheung, T. H. & Rando, T. a. Molecular regulation of stem cell quiescence. *Nat. Rev. Mol. Cell Biol.* **14,** 329–40 (2013).

25.   Nigro, J. M. *et al.* Mutations in the p53 gene occur in diverse human tumour types. *Nature* **342,** 705–708 (1989).

26.   Linzer, D. I. H. & Levine, A. J. Characterization of a 54K Dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell* **17,** 43–52 (1979).

27.   Lane, D. P. & Crawford, L. V. T antigen is bound to a host protein in SV40-transformed cells. *Nature* **278,** 261–263 (1979).

28.   Liu, Y. *et al.* p53 Regulates Hematopoietic Stem Cell Quiescence. *Cell Stem Cell* **4,** 37–48 (2009).

29.   Sancar, A., Lindsey-Boltz, L. A., Ünsal-Kaçmaz, K. & Linn, S. Molecular Mechanisms of Mammalian DNA Repair and the DNA Damage Checkpoints. *Annu. Rev. Biochem.* **73,** 39–85 (2004).

30.   Yang, H. W., Chung, M., Kudo, T. & Meyer, T. Competing memories of mitogen and p53 signalling control cell-cycle entry. *Nature* **549,** 404–408 (2017).

31.   Abbas, T. & Dutta, A. P21 in Cancer: Intricate Networks and Multiple Activities. *Nat. Rev. Cancer* **9,** 400–14 (2009).

32.   Rhind, N. & Gilbert, D. M. DNA replication timing. *Cold Spring Harb. Perspect. Biol.* **5,** (2013).

33.	Zeman, M. K. & Cimprich, K. A. Causes and consequences of replication stress. *Nat. Cell Biol.* **16,** 2–9 (2014).

34.	Técher, H., Koundrioukoff, S., Nicolas, A. & Debatisse, M. The impact of replication stress on replication dynamics and DNA damage in vertebrate cells. *Nat. Rev. Genet.* **18,** 535–550 (2017).

35.	Schmidt, M. *et al.* Regulation of G2/M transition by inhibition of WEE1 and PKMYT1 Kinases. *Molecules* **22,** 1–17 (2017).

36.	Matson, J. P. & Cook, J. G. Cell cycle proliferation decisions: the impact of single cell analyses. *FEBS J.* **284,** 362–375 (2017).

37.	Naetar, N. *et al.* PP2A-mediated regulation of ras signaling in G2 is essential for stable quiescence and normal G1 length. *Mol. Cell* **54,** 932–945 (2014).

38.	Savatier, P., Huang, S., Szekely, L., Wiman, K. G. & Samarut, J. Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. *Oncogene* **9,** 809–818 (1994).

39.	Stead, E. *et al.* Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. *Oncogene* **21,** 8320–33 (2002).

40.	Singh, A. *et al.* Cell-Cycle Control of Bivalent Epigenetic Domains Regulates the Exit from Pluripotency. *Stem Cell Reports* **5,** 1–14 (2015).

41.	Singh, A. *et al.* Cell-cycle control of developmentally regulated transcription factors accounts for heterogeneity in human pluripotent cells. *Stem cell reports* **1,** 532–44 (2013).

42.	Dalton, S. Linking the Cell Cycle to Cell Fate Decisions. *Trends Cell Biol.* **25,** 592–600 (2015).

43.	Li, V. C., Ballabeni, A. & Kirschner, M. W. Gap 1 phase length and mouse embryonic stem cell self-renewal. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 12550–5 (2012).

44.	Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. & Brivanlou, A. H. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat. Med.* **10,** 55–63 (2004).

45.	De Jaime-Soguero, A. *et al.* Wnt/Tcf1 pathway restricts embryonic stem cell cycle through activation of the Ink4/Arf locus. *PLoS Genet.* **13,** (2017).

46.	Gonzales, K. A. U. *et al.* Deterministic Restriction on Pluripotent State Dissolution by Cell-Cycle Pathways. *Cell* **162,** 564–579 (2015).

47.	Smith, A. G. & Hooper, M. L. Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Dev Biol* **121,** 1–9 (1987).

48.	Smith, A. G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336,** 688–690 (1988).

49.	Ying, Q.-L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115,** 281–92 (2003).

50.	Kunath, T. *et al.* 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 (2007).

51.	Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–23 (2008).

52.	Ying, Q.-L. & Smith, A. The Art of Capturing Pluripotency: Creating the Right Culture. *Stem Cell Reports* **8,** 1457–1464 (2017).

53.	Marks, H. *et al.* The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. *Cell* **149,** 590–604 (2012).

54.	Marks, H. & Stunnenberg, H. G. Transcription regulation and chromatin structure in the pluripotent ground state. *Biochim. Biophys. Acta* 1–9 (2013). doi:10.1016/j.bbagrm.2013.09.005

55.	Habibi, E. *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13,** 360–9 (2013).

56.	Seisenberger, S. *et al.* The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48,** 849–62 (2012).

57.   Leitch, H. G. *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20,** 311–316 (2013).

58.   Ho, A. & Dowdy, S. F. Regulation of G1 cell-cycle progression by oncogenes and tumor suppressor genes. *Curr. Opin. Genet. Dev.* **12,** 47–52 (2002).

0
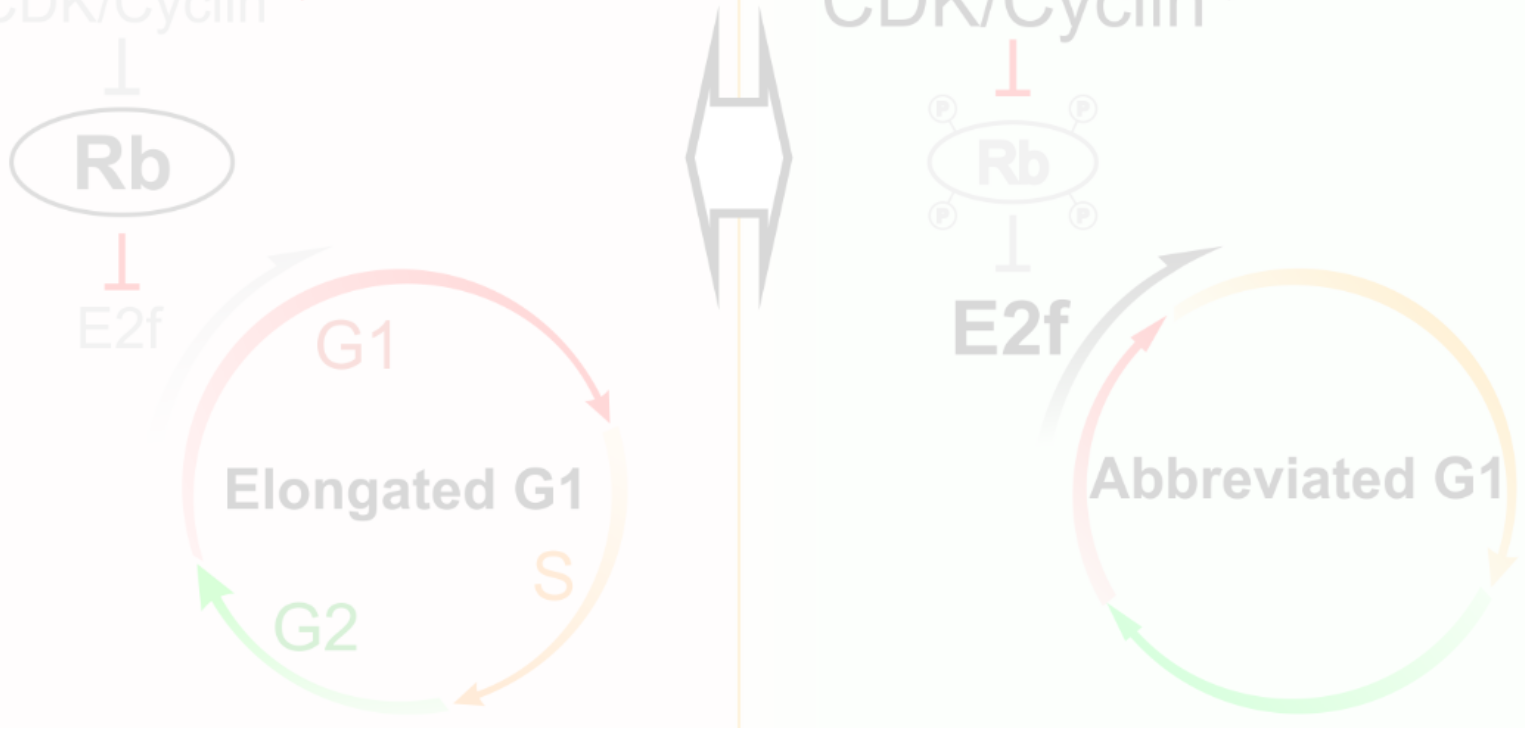
Chapter 1

Distinct cell cycle control in two different states of mouse pluripotency

Menno ter Huurne[1], James Chappell[2], Stephen Dalton[2], Hendrik G. Stunnenberg[1]

[1] Radboud University, Faculty of Science, Department of Molecular Biology, 6525GA Nijmegen, the Netherlands

[2] The University of Georgia, Paul D. Coverdell Center for Biomedical and Health Sciences, Department of Biochemistry and Molecular Biology, Athens, GA 30602, USA
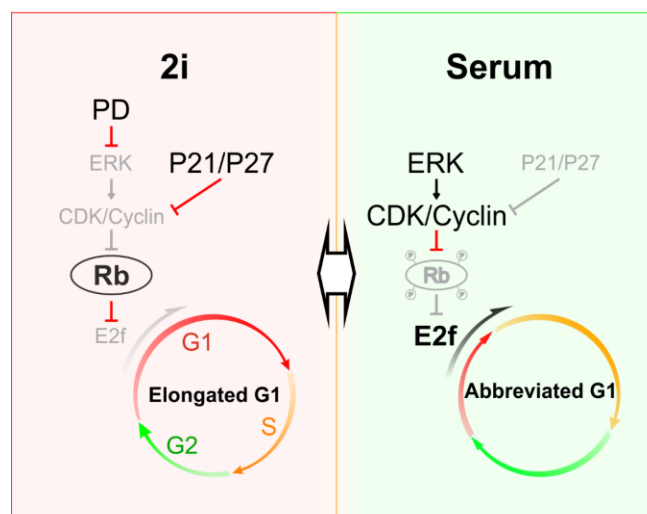
mouse embryonic stem cells

Summary

Mouse embryonic stem cells (ESCs) cultured in serum are characterized by hyper-phosphorylated RB protein, lack of G1 control and rapid progression through the cell cycle. Here, we show that ESCs grown in the presence of two small molecule inhibitors (2i ESCs) have a longer G1-phase with hypo-phosphorylated RB, implying that they have a functional G1 checkpoint. Deletion of RB, p107 and p130 in 2i ESCs results in a G1-phase similar to that of serum ESCs. Inhibition of the ERK signaling pathway in serum ESC results in the appearance of hypo-phosphorylated RB and reinstatement of a G1 checkpoint. In addition, induction of a dormant state by the inhibition of MYC, resembling diapause, requires the presence of the RB family proteins. Collectively, our data show that RB-dependent G1 restriction point signaling is active in mouse ESCs grown in 2i but abrogated in serum by ERK-dependent phosphorylation.

Main Text

One characteristic of ESCs is that they proliferate at an unusually rapid pace, with a doubling time of ~12-14hr. The highly proliferative nature of pluripotent ESCs is attributed to the lack of G1-phase control, a feature that has been intimately linked to pluripotency[1]. Constitutively active kinases and the absence of CDK-inhibitor proteins ensure rapid progression into S-phase, resulting in an extremely short G1-phase. This model is, however, based on studies performed in ESCs grown in serum (hereafter serum ESCs). Recent studies have indicated that ESCs cultured in defined medium in the presence of two small molecule inhibitors, PD0325901 and CHIR99021, (hereafter 2i ESCs) more closely reflect pluripotent ESCs from the inner cell mass (often referred to as ground state) whereas ESCs cultured in serum are more similar to pluripotent cells from post-implantation embryonic stages[2–4]. Our analysis of the cell cycle in 2i ESCs indicates that G1 control in ground state pluripotent ESCs is distinct from that in serum ESCs.

To assess the global effect of the 2i culture conditions on the mouse ESC cell cycle, we used BrdU incorporation and Propidium Iodide (PI) staining in combination with flow cytometry to determine the distribution of cells over the different phases of the cell cycle. In line with literature, the majority of serum ESCs reside in the S-phase whereas the rest are roughly equally distributed between G1 and G2 phase (Figure 1A). Conversely, the number of 2i ESCs in S-phase is reduced whereas the number of cells in the G1-phase is strongly increased which is largely effectuated already within 24 hours of adaptation from serum-to-2i or vice versa (Figure 1B and S1A). The adaptation of serum ESCs to 2i conditions does not affect the expression of key pluripotency factors[3], indicating that these changes in cell cycle are not the result of differentiation (Figure S1B). Similar results were obtained in multiple karyotyped WT ESC lines, both male and female with several different genetic backgrounds and in iPS cells[5] (Table S1, Figure S1C and Figure S1D).

Serum grown ESCs have been reported to contain Nanog-high and Nanog–low subpopulations with the former exhibiting a higher degree of pluripotency[6,7]. Both populations display a serum-type cell cycle distribution (Figure S1E), indicating that the elongated G1-phase is characteristic of ground state 2i ESCs. To gain further insight, we generated mESCs expressing the FUCCI reporters[8]. The FUCCI system enables FACS sorting of cells from different phases of the cell cycle and importantly determination of the length of the different phases, because the reporter activity increases proportionally to the time spent in a specific phase (Figure S1F). The analysis confirms that 2i ESCs have a markedly longer G1-phase and shorter S-/G2-phases, relative to serum ESCs (Figure 1C and S1G). As serum and 2i ESCs proliferate at roughly similar pace (Figure S1H)[9] these changes appear to have a compensatory effect on the length of the cell cycle.
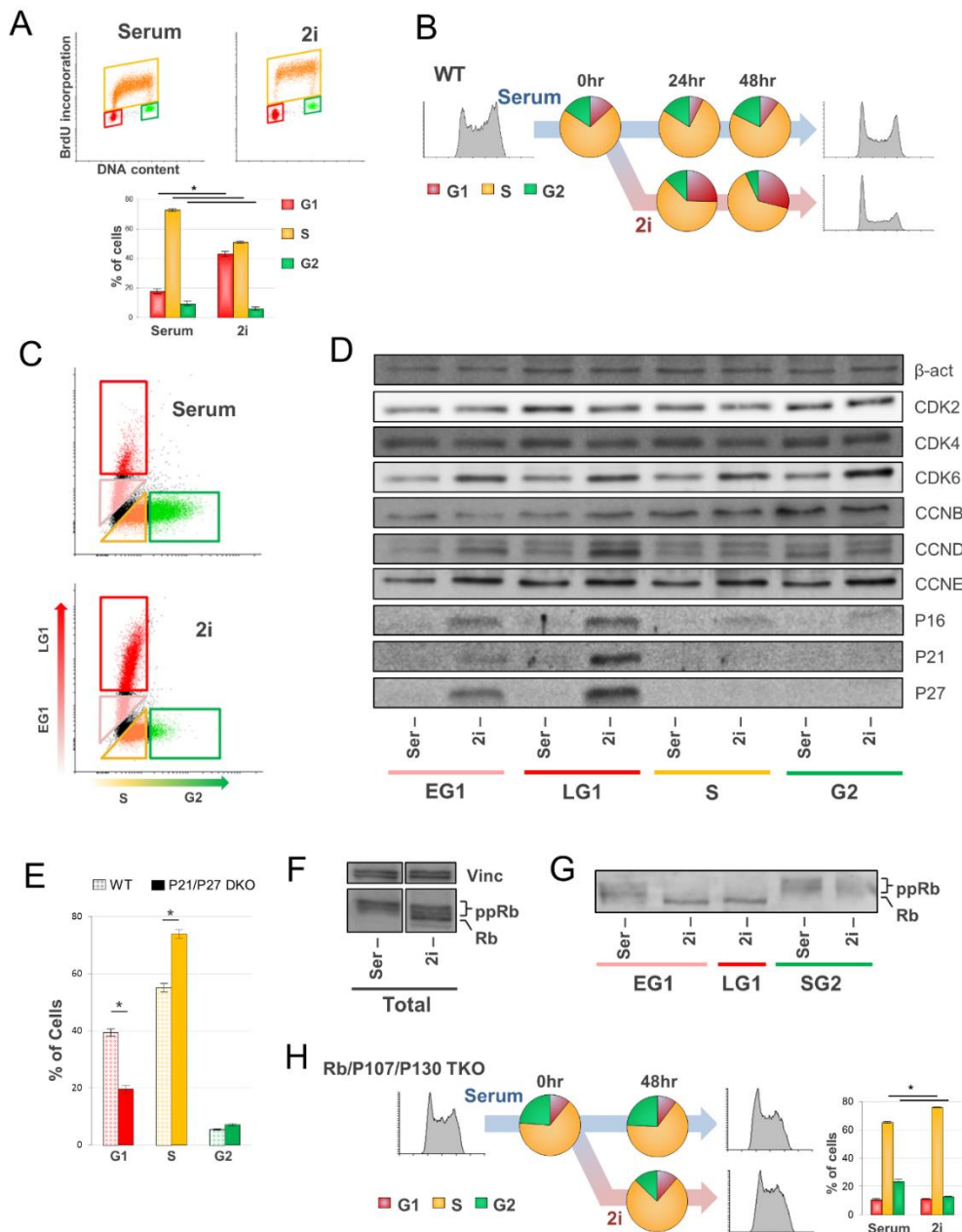
1

Figure 1. G1-phase is elongated upon adaptation of ESCs to 2i conditions. (A) DNA staining using PI in combination with BrdU incorporation shows a higher number of 2i R1 ESCs residing in G1 phase when compared to serum R1 ESCs. Significance was tested using the two tailed Student's T-test; *p-val<0.05. **(B)** Cell cycle analysis using BrdU/PI shows that the rapid increase in the number of cells residing in G1-phase upon transition from serum to 2i occurs within 48 hours. Similar results were observed in several different male and female ESC lines. **(C)** FUCCI reporter expression shows that 2i ESCs reside longer in G1 phase, whereas the combined S-/G2-phase is shortened in 2i ESCs compared to serum ESCs. Indicated are the gates as used for sorting. **(D)** Western Blot analysis of cell cycle proteins involved in progression into S-phase in serum and 2i ESCs showing specific upregulation of the CDK-inhibitors in 2i ESCs during both Early as well as Late G1-phase (EG1 and LG1, respectively). Three independent experiments showed similar results. **(E)** Distribution of cells over different phases of the cell cycle as determined by BrdU/PI staining performed in triplicate using three independent P21/P27 DKO clones. A two tailed Student's T-test was used to determine significance; *p-val<0.05. **(F)** Western blot analysis of total cell populations showing only phosphorylated RB in serum ESCs. Western blot is representative for three independent experiments. **(G)** As in C, using sorted Early and Late G1 as well as S-/G2-phase of ESCs, showing hypo-phosphorylated RB in G1-phase and phopho-RB in S-/G2 in 2i ESCs. At least two independent experiments showed similar results. **(H)** Cell cycle analysis on RB/P107/P130 TKO ESCs using BrdU in combination with PI showing unaltered number of cells in G1-phase upon adaptation from serum to 2i. A decrease in the number of cells in G2-phase was observed. Results are representative of at least two independent experiments (*p-val<0.05, two tailed Student's T-test). See also Table S1 and Figure S1 and S2. All values and error bars represent the mean ± SD.

The abbreviated G1-phase in serum ESCs is accompanied by the accelerated expression of histones during G1-/S-phase transition[10]. To assess whether serum and 2i ESCs differ in this respect, we performed RNA-seq on 2i and serum ESCs in G1-phase. GO-term analysis on differentially expressed genes (Figure S1I; Fold Change >2, p-val<0.05) revealed higher expression of genes associated with S-phase entry and chromatin assembly in serum ESCs (Figure S1J). Replication-dependent histone genes are more highly expressed in serum ESCs when compared to 2i ESCs both in late G1 as well as in early G1-phase (Figure S1K and not shown). The accelerated progression through G1-phase, upregulation of histone gene expression, enrichment of the E2F binding motif at highly expressed genes (Figure S1L) as well as increased expression of the E2F family members and known E2F target genes (Figure S1M) suggest elevated activity of the RB-E2F pathway in serum ESCs. E2F1 ChIP-seq revealed that ~5% of the binding sites displayed differential binding (2-fold change, p-val<0.05) between serum and 2i (Figure S1N). However, E2F1 occupancy did not correlate with differential RNA expression ($R2 = 0.08$). Thus differential E2F binding is very unlikely to drive differential gene expression in serum and 2i.

To assess whether cyclin/CDK activity upstream of the RB-E2F pathway explains the distinct cell cycle patterns, we determined the expression of key regulators of cell cycle progression through G1-phase. Quantitation of their protein levels throughout the cell cycle revealed that CDK2 and CDK4 are slightly lower, whereas CDK6, CCND and CCNE are slightly higher expressed in 2i ESCs (Figure 1D). The most notable difference, however, is the higher protein levels of the CDK-inhibitors: P16, P21 and P27 are nearly exclusively expressed only in 2i ESCs and primarily during the late G1-phase (Figure 1D). Importantly, differential expression of CDK-inhibitors in 2i and serum ESCs parallels their higher RNA levels in Inner Cell Mass as compared to the Epiblast[11] (Figure S2A).

Deletion of P21 or P27 alone did not significantly alter the cell cycle, however, the combined deletion of P21 and P27 resulted in a decrease in the number of cells in G1-phase as well as lowered expression of the FUCCI reporters. The deletion of P16, alone or in combination with P21, did not significantly affect the number of cells in G1-phase (data not shown). These results indicate that the CDK-inhibitor proteins have partially overlapping functions and that the length of G1-phase in 2i ESCs is controlled by the sum of P21/P27 inhibitory action (Figure 1E, S2B, S2C and S2D).

To assess the net effect of differential CDK-activity between serum and 2i ESCs, we determined the expression and phosphorylation status of their downstream targets, the RB family proteins RB, P107 and P130. In the hypo-phosphorylated form, these proteins bind to and inhibit the activity of DNA-bound E2F transcription factors, thereby slowing down cell cycle progression during G1-phase. Upon successive phosphorylation of RB, its interaction with E2F transcription factors is lost leading to the activation of genes involved in progression into S-phase[12,13]. Western blot analysis shows that hyper-phosphorylated forms of RB are detected in serum ESCs whereas RB appears hyper- as well as hypo-phosphorylated in 2i FUCCI ESCs, which was confirmed by phosphatase treatment (Figure 1F and S2E). Strikingly, in both early and late G1 phase of 2i FUCCI ESCs only hypo-phosporylated RB is detected (Figure 1G). These results were confirmed in G1-phase sorted Suv39h WT ESCs[14] and in E14 ESCs (Figure S2F and data not shown). Phosphorylated RB is, however, present in S-/G2-phase in both 2i and serum ESCs (Figure 1G and S2F). In P21/P27 DKO ESCs cultured in 2i, the hypo-phosphorylated form of RB is hardly detected. This indicates that the expression of CDK-inhibitors in 2i prevents phosphorylation of RB (Figure S2G). Besides hypo-phosphorylated RB, P107 protein levels are higher in 2i ESCs which could aid in reinstating the G1 checkpoint (Figure S2H). Together, these results suggest that E2F-activity is higher in serum and that E2F activity is inhibited by hypo-phosphorylated RB family proteins in 2i.

We used RB KO and RB/P107/P130 triple knock out (TKO) ESCs[15] to corroborate and extend our hypothesis that the elongated G1-phase in 2i ESCs is mediated by the RB family proteins. In TKO ESCs, the number of cells in G1-phase remains the same when shifting from serum to 2i conditions indicating that the RB family proteins are essential for elongation of G1-phase in 2i ESCs (Figure 1H). In 2i ESCs, the cell cycle distribution of only RB KO cells is changed, but to a lesser extent than in 2i TKO ESCs (Figure S2I). Note that the expression of key pluripotency genes and colony formation is not affected in TKO ESCs (Figure S2J and S2K). Taken together, the RB family proteins are involved in the control of the G1-phase in 2i ESCs.

To determine which signaling pathway regulates the phosphorylation status of RB, ESCs were cultured in medium + LIF supplemented either with PD0325901 ("PD"), CHIR99021 ("Chiron") or both ("2i"). When taken of 2i ESCs do proliferate for at least 1-2 weeks in Ndiff 227 media (Takara, formerly Ndiff N2B27 – StemCells) + LIF. FUCCI reporter assays and PI FACS analysis indicate that the addition of PD, which blocks the ERK-signaling pathway, drives the elongation of G1-phase whereas the GSK3-inhibitor Chiron has no effect (Figure 2A-C). Western Blot analysis detected hypo-phosphorylated RB only in ESCs cultured in the presence of PD (Figure 2D). Similarly, PD addition to serum ESCs gives rise to elevated, mostly

hypo-phosphorylated RB protein (Figure 2E and S2F), possibly because of lowered cyclin D levels (Figure S2L). These effects are accompanied by an increased number of cells in the G1-phase in wild type ESCs (Figure 2F and S2M), which is strongly reduced in TKO ESCs (Figure 2F). Chiron had again only a marginal effect on the length of G1 phase in serum ESCs as in serum free conditions (Figure S2M and Figure 2B). We conclude that pluripotent 2i ESCs do possess an active G1 restriction point that in serum ESCs is abrogated by ERK signaling and RB hyper-phosphorylation.

It has recently been shown that inhibition of C-MYC and N-MYC in 2i ESCs results in quiescence, mimicking in vivo diapause[16]. To test whether the RB family proteins are required for G1-arrest, we treated WT and RB/P107/P130 TKO ESCs with the MYC-inhibitor (MYCi). Upon MYC inhibition, WT 2i ESCs stall in the G1-phase in line with a recent report[16]. In RB/P107/P130 TKO ESCs, however, the number of cells in G1-phase did not significantly increase upon MYC inhibition (Figure 2G). We conclude that the RB/P107/P130-mediated restriction point is required to stall the cells in G1 in a dormancy-like state.
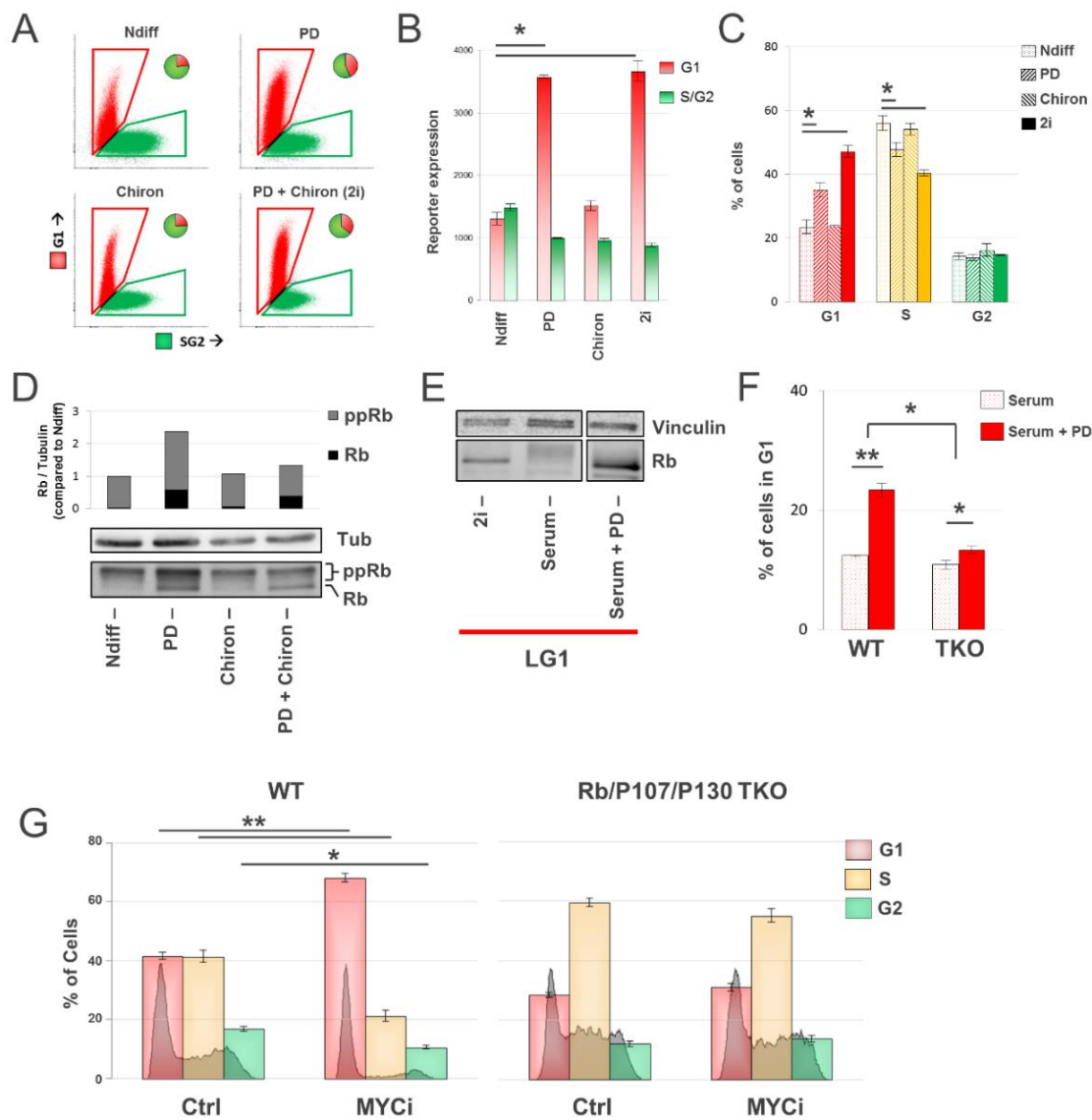
Figure 2. Role of ERK signaling and RB family proteins in cell cycle regulation in ground state ESCs. (A) Quantification of G1- and SG2-phases in FUCCI cells in Ndiff medium supplemented with LIF and either PD, Chiron or both inhibitors. **(B)** FUCCI reporter expression shows that inhibition of the ERK signaling pathway by PD results in elongation of G1-phase. Significance was assessed by a two-tailed Student's T test, *p-val<0.05. At least two independent experiments showed these results. **(C)** Cell cycle analysis using PI shows that the addition of PD results in an increase of cells in G1-phase, whereas Chiron had no significant effect (Student's T-test, *p-val<0.05). **(D)** ESCs cultured in serum free NDiff medium + LIF supplemented with either one or both inhibitors shows the presence of hypo-phosphorylated RB upon inhibition of the ERK signaling pathway by PD and not by Chiron. **(E)** Exposure of serum ESCs to PD results in hypo-phosphorylated and increased expression of RB. Similar results were observed in two independent experiments. **(F)** Quantitation of the number of cells in G1-phase of WT serum ESCs and RB/P107/P130 triple knock out serum ESCs incubated in the presence of PD showing that elongation of G1-phase required the RB family proteins. Bar graphs indicate the mean ± SD. Comparison was performed by two-tailed Student's T-test, *p-val < 0.01; **p-val < 0.001; n = 3. **(G)** Inhibition of MYC by a small molecule inhibitor, 10058-F4, results in a near complete block of WT ESC but not of RB/P107/P130 triple knock out ESCs cultured in 2i. Error bars are means ± SD from triplicates, representative of two independent experiments. Significance was assessed by two-tailed Student's T-test, *p-val <0.001, **p-val < 0.0001. See also Figure S2. The bar charts represent the means ± SD.

It has been postulated that an abbreviated G1-phase is characteristic of ESCs and ensures maintenance of pluripotency. Our study shows that elongation of the G1-phase in 2i ESCs is the most prominent difference in the cell cycle when compared to serum ESCs and yet 2i ESCs are highly pluripotent and perform even better in chimera formation then serum ESCs[17]. Together with the observations that 2i ESCs have a lower propensity to differentiate[3,18], our findings imply that a short G1-phase is not an intrinsic property of pluripotent ESCs. Our data corroborate and significantly extend previous observations in serum ESCs[19,20], by revealing the mechanism of the fundamentally different G1 control in 2i versus serum ESCs. The FGF/ERK signaling plays an important role during early embryonic development and loss of FGF4 results in impaired cell proliferation of ICM cells (reviewed in Lanner, F. & Rossant, J.[21] and Dorey, K. & Amaya, E.[22]). Our results indicate that signaling via the ERK pathway results in a decrease in hypo-phosphorylated RB, loss of the restriction point and an abbreviated G1-phase in serum ESCs. As the FGF/ERK pathway plays a crucial role in early fate-specification during the implantation stage[21] that coincides with bursts of cell proliferation[23], we lend further support to the model in which 2i ESCs resemble naive pluripotent (ICM) cells by showing that FGF/ERK signaling results in shortening of G1-phase in epiblast (serum) ESCs.

Besides the difference in G1-phase between serum and 2i ESCs, a shortening of S/G2-phase was observed as well. These observations are in concordance with marked changes in cell cycle during early embryonic development (reviewed by Duronio, R. J.[24] and Bouldin, C. M. & Kimelman, D.[25]). It is tempting to speculate that this is caused by observed discrepancies in the expression of cell cycle regulators, such as differences in Cyclin E expression during S-phase and the higher expression of all Cdc25 isoforms during G2-phase (unpublished data)[25–27]. Moreover, differences in epigenetic make up and chromatin landscape between serum and 2i ESCs[28,29] could possibly affect the length of these phases as well[30].

Collectively, our results provide a paradigm shift in how the cell cycle is regulated in pluripotent stem cells. Surprisingly, 2i ESCs do harbor an intact G1 control, which is lost upon adaptation to serum conditions due to increased ERK signaling and RB hyper-phosphorylation. Our data implies a revised conceptual framework for cell cycle regulation in pluripotent ESCs and during early embryonic development.

Author Contributions

Acknowledgements

References

1.   Singh, A. & Dalton, S. The Cell Cycle and Myc Intersect with Mechanisms that Regulate Pluripotency and Reprogramming. *Cell Stem Cell* **5,** 141–149 (2009).

2.   Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4,** 487–92 (2009).

3.   Marks, H. *et al.* The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. *Cell* **149,** 590–604 (2012).

4.   Weinberger, L., Ayyash, M., Novershtern, N. & Hanna, J. H. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat. Rev. Mol. Cell Biol.* 030676 (2016). doi:10.1101/030676

5.   Li, H. *et al.* The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* **460,** 1136–9 (2009).

6.   Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230–1234 (2007).

7.   Kalmar, T. *et al.* Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol.* **7,** 33–36 (2009).

8.   Sakaue-Sawano, A. *et al.* Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132,** 487–98 (2008).

9.   Tamm, C., Pijuan Galitó, S. & Annerén, C. A comparative study of protocols for mouse embryonic stem cell culturing. *PLoS One* **8,** e81156 (2013).

10.   Medina, R. *et al.* Epigenetic control of cell cycle-dependent histone gene expression is a principal component of the abbreviated pluripotent cell cycle. *Mol. Cell. Biol.* **32,** 3860–71 (2012).

11.   Boroviak, T. *et al.* Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. *Dev. Cell* **35,** 366–382 (2015).

12.   Hirschi, A. *et al.* An overlapping kinase and phosphatase docking site regulates activity of the retinoblastoma protein. *Nat. Struct. Mol. Biol.* **17,** 1051–7 (2010).

13.   Dick, F. A. & Rubin, S. M. Molecular mechanisms underlying RB protein function. *Nat. Rev. Mol. Cell Biol.* **14,** 297–306 (2013).

14.   Lehnertz, B. *et al.* Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.* **13,** 1192–1200 (2003).

15.   Dannenberg, J. *et al.* causing immortalization and increased cell turnover under Ablation of the Retinoblastoma gene family deregulates G 1 control causing immortalization and increased cell turnover under growth-restricting conditions. *Genes Dev.* 3051–3064 (2000). doi:10.1101/gad.847700

16.   Scognamiglio, R. *et al.* Myc Depletion Induces a Pluripotent Dormant State Mimicking Diapause. *Cell* 668–680 (2016). doi:10.1016/j.cell.2015.12.033

17.   Alexandrova, S. *et al.* Selection and dynamics of embryonic stem cell integration into early mouse embryos. *Development* dev.124602- (2015). doi:10.1242/dev.124602

18.   Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–23 (2008).

19.   Li, V. C., Ballabeni, A. & Kirschner, M. W. Gap 1 phase length and mouse embryonic stem cell self-renewal. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 12550–5 (2012).

20.   Gonzales, K. A. U. *et al.* Deterministic Restriction on Pluripotent State Dissolution by Cell-Cycle Pathways. *Cell* **162,** 564–579 (2015).

21.   Lanner, F. & Rossant, J. The role of FGF/Erk signaling in pluripotent cells. *Development* **137,** 3351–3360 (2010).

22.   Dorey, K. & Amaya, E. FGF signalling: diverse roles during early vertebrate embryogenesis. *Development* **137,** 3731–42 (2010).

23.   Snow, M. H. L. Gastrulation in the mouse: growth and regionalization of the epiblast. *J. Embryol. Exp. Morphol.* **Vol. 42,** 293–303 (1977).

24.   Duronio, R. J. Developing S-phase control. 746–750 (2012). doi:10.1101/gad.191171.112.The

25.   Bouldin, C. M. & Kimelman, D. Cdc25 and the importance of G2 control: Insights from developmental biology. *Cell Cycle* **13,** 2165–2171 (2014).

26.   Farrell, J. A., Shermoen, A. W., Yuan, K. & O'Farrell, P. H. Embryonic onset of late replication requires Cdc25 down-regulation. *Genes Dev.* **26,** 714–725 (2012).

27.   Sansam, C. G., Goins, D., Siefert, J. C., Clowdus, E. a & Sansam, C. L. Cyclin-dependent kinase regulates the length of S phase through TICRR / TRESLIN phosphorylation. *Genes Dev.* **29,** 555–566 (2015).

28.   Habibi, E. *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13,** 360–9 (2013).

29.   Marks, H. & Stunnenberg, H. G. Transcription regulation and chromatin structure in the pluripotent ground state. *Biochim. Biophys. Acta* 1–9 (2013). doi:10.1016/j.bbagrm.2013.09.005

30.   Shermoen, A. W., McCleland, M. L. & O'Farrell, P. H. Developmental control of late replication and s phase length. *Curr. Biol.* **20,** 2067–2077 (2010).

31.   Fedr, R. *et al.* Automatic cell cloning assay for determining the clonogenic capacity of cancer and cancer stem-like cells. *Cytom. Part A* **83 A,** 472–482 (2013).

32.   Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems_Sup. *Science* **339,** 819–23 (2013).

33.   Anders, S. & Huber, W. Differential expression of RNA-Seq data at the gene level–the DESeq package. *EMBL, Heidelberg, Ger.* (2012). at <http://gga01.med.wayne.edu/online_help/help_regionminer/DESeq.pdf>

34.   Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

1

Supplemental Figures

A WT
0hr 24hr 48hr
2i
Serum
G1 S G2

B Expression (relative to GAPDH)
Serum 2i
Oct-4 Nanog Rex1 Myc Prdm16

C iPS
% of cells
Serum 2i

D % of genomic sequence tags (compared to TNGA)
EB5 iPS E14
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

E
GFP →
DNA content →

	Nanog⁻	Nanog⁺
G1	21.13 ± 0.00	19.65 ± 0.66
S	65.25 ± 0.75	65.84 ± 0.46
G2	13.62 ± 0.74	14.50 ± 1.12

F DNA content →

G
Serum
2i
2268 ±265
6625 ±289
3034 ±414
1307 ±58
Kasubari-Orange (G1) →
Azami-Green (S/G2) →

H Growth Curve
No. of cells (in millions)
Serum – Td = 12.1hr
2i – Td = 12.7hr
Time (in hours)

I Serum vs 2i
Log2 Fold Change RPKM (Serum versus 2i)
Serum-Hi (N=547)
2i-Hi (N=905)
Histone Genes

J GO-term Analysis

	Term	PValue	Enrichment
Serum-High	Chromosome organization	6.26E-29	3.71
	Cell cycle	1.74E-28	3.21
	DNA replication	2.86E-13	5.22
	Nucleosome assembly	6.45E-07	5.44
	DNA methylation	2.28E-06	6.45

K RNA expression (Log2 RPKfil)
Serum 2i Serum 2i
Histones Random

L Motif Enrichment

	Rank	Motif	P-value	Best Match
Serum-High	1		1.00E-13	E2F2
	2		1.00E-12	Zbtb12
	3		1.00E-12	E2F3
	4		1.00E-12	NFY
	5		1.00E-12	NFkB

M

		RPKM	
		Serum	2i
Cell Cycle	Ccnd2	4.53	0.62
	Ccne1	87.73	31.34
	Ccne2	22.97	7.23
	Myc	8.34	0.12
	Cdc25a	40.65	19.71
	E2f2	4.73	1.40
	E2f3	8.02	2.35
	E2f7	4.23	1.40
	E2f8	7.97	1.03
	Jun	1.56	8.12
	Myb	0.85	0.27
DNA replication	Hist1h2ba	126.33	31.34
	Hist1h2ad	245.90	118.80
Checkpoint	Brca1	28.83	13.98
	Brca2	19.86	8.52
	Bub1	31.86	12.71
	Bub1B	54.57	16.42
	Cenpe	15.07	3.22
	Tfdp1	63.81	31.32
	Chek1	20.98	9.72

N E2f1 Occupancy
Serum
2i
Serum-Hi (N=121)
2i-Hi (N=85)
Common (N=4317)

Promoter-TSS TTS
Intronic 5'UTR
Intergenic 3'UTR
Exonic Non-coding

E2f1 Occupancy (Serum vs 2i)
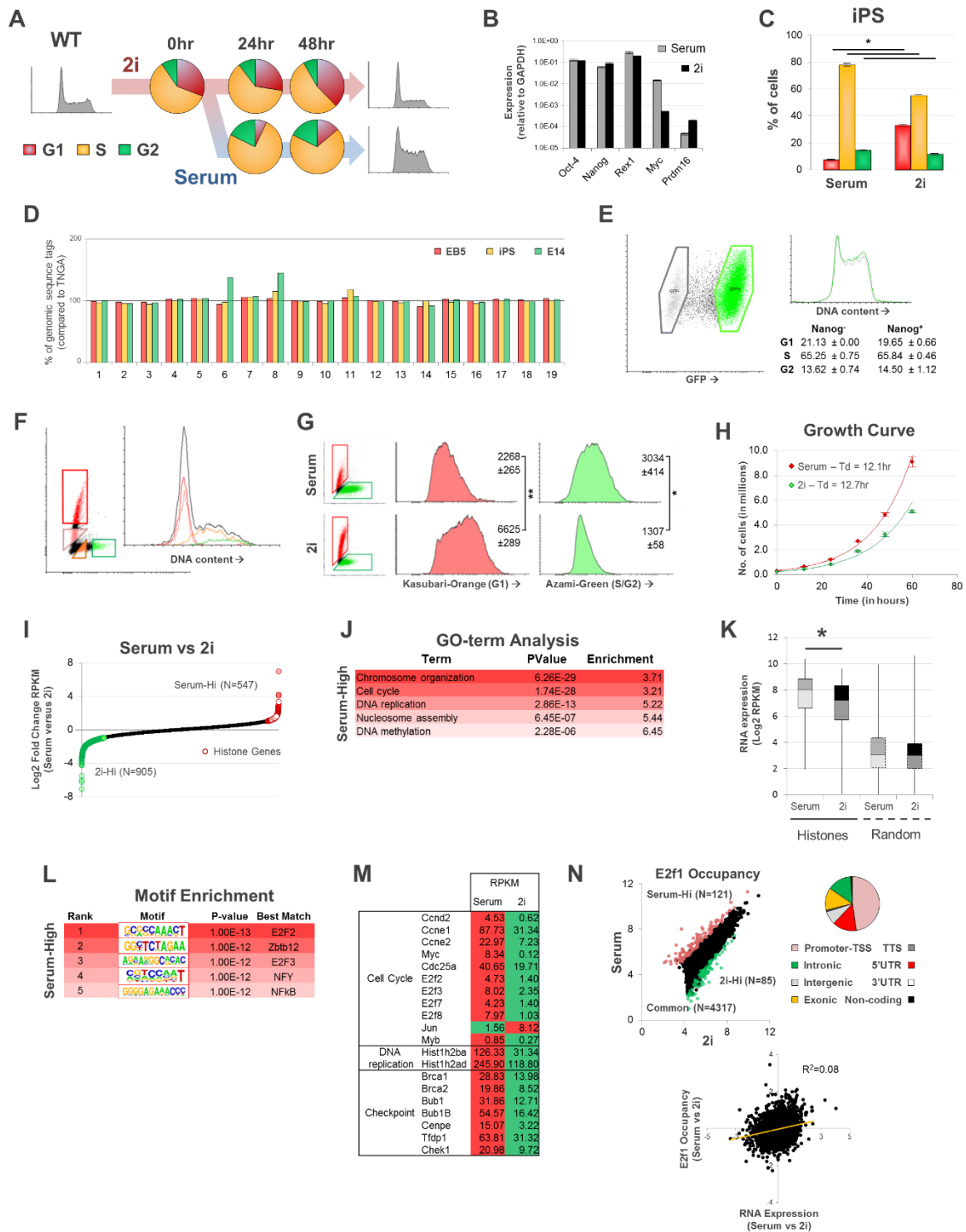RNA Expression (Serum vs 2i)
$R^2=0.08$

Figure S1 related to figure 1. G1-phase is elongated in 2i ESCs and is accompanied by lowered E2F activity and reduced expression of genes involved in S-phase entry. (A) Cell cycle analysis using BrdU/PI of WT 2i R1 ESCs upon adaptation to serum conditions. **(B)** Expression analysis using RT-qPCR on pluripotency, serum-specific and 2i-specific genes upon adaptation of serum ESCs to 2i for 48 hours. **(C)** Quantification of distribution of fully adapted serum and 2i iPS cells over the different phases of the cell cycle (Student's T-test, *p-val<0.05). **(D)** Karyotype analysis (sequencing of genomic DNA) of two cell lines (EB5 and E14) and one WT iPS line. EB5 and iPS line display a normal karyotype whereas E14 shows trisomy at chromosome 6 and 8. Shown is the distribution of all sequence tags from genomic DNA of indicated cell lines over chromosomes compared to the karyotypic-normal TNG-A ESC line[3]. **(E)** Cell cycle analysis on Nanog⁻ and Nanog⁺ serum ESCs using Hoechst. Overlay histogram showing DNA content from both populations as indicated by gates in the dot plot (grey line Nanog⁻ and green line Nanog⁺). Numbers represent means ± standard deviations from triplicates. The experiment has been performed twice showing similar results. **(F)** Determination of DNA content of 2i FUCCI ESCs using vibrant violet staining. Colored lines indicate DNA content of the cell populations sorted as indicated in dot plot. **(G)** Dot plot and histogram of FUCCI reporter expression in serum and 2i ESCs. Indicated are the mean fluorescence intensities and standard deviations of triplicates. At least two independent experiments were performed showing similar results. Student's T-test, **p-val<0.001, *p-val<0.05. **(H)** Growth curves and doubling times of fully adapted E14 serum and 2i ESCs. For both conditions $1.5 \cdot 10^5$ cells were seeded in triplicates and aliquots were counted at indicated times and presented as means ± SD. Two independent experiments showed similar results. **(I)** Ratio of gene transcription in serum ESCs versus 2i ESCs in late G1-phase using two biological replicates. **(J)** GO analysis on differentially expressed genes in serum ESCs versus 2i ESCs in late G1-phase. **(K)** Boxplots of RNA expression (log2 RPKM) of histone and random chosen genes in late G1-phase showing lower expression of histone genes in 2i ESCs compared to serum. Statistical analysis was performed using the paired two tailed Student's T-test, *p-val <0.001. **(L)** Motif enrichment analysis using Homer on promoter regions of genes higher expressed in serum ESCs. **(M)** Expression values (RPKM) of known E2F target genes clustered based on function. **(N)** ChIP-seq analysis of E2F1 occupancy is highly correlated between serum and 2i ($R^2 = 0.79$). The location of the E2F1 binding sites is presented as a PIE diagram: The majority of peaks is located in either the promoter-TSS or the 5'UTR and ~5% shows differential E2F1 occupancy between serum and 2i (Log2-fold change > 1, p-val <0.05). Lower panel: correlation analysis revealing absence of a correlation ($R^2= 0.08$) between differential E2F1 occupancy in ChIP-seq and RNA expression (RNA-seq) suggests that E2F binding does not drive the differential gene expression. All values represent the mean ± SD.
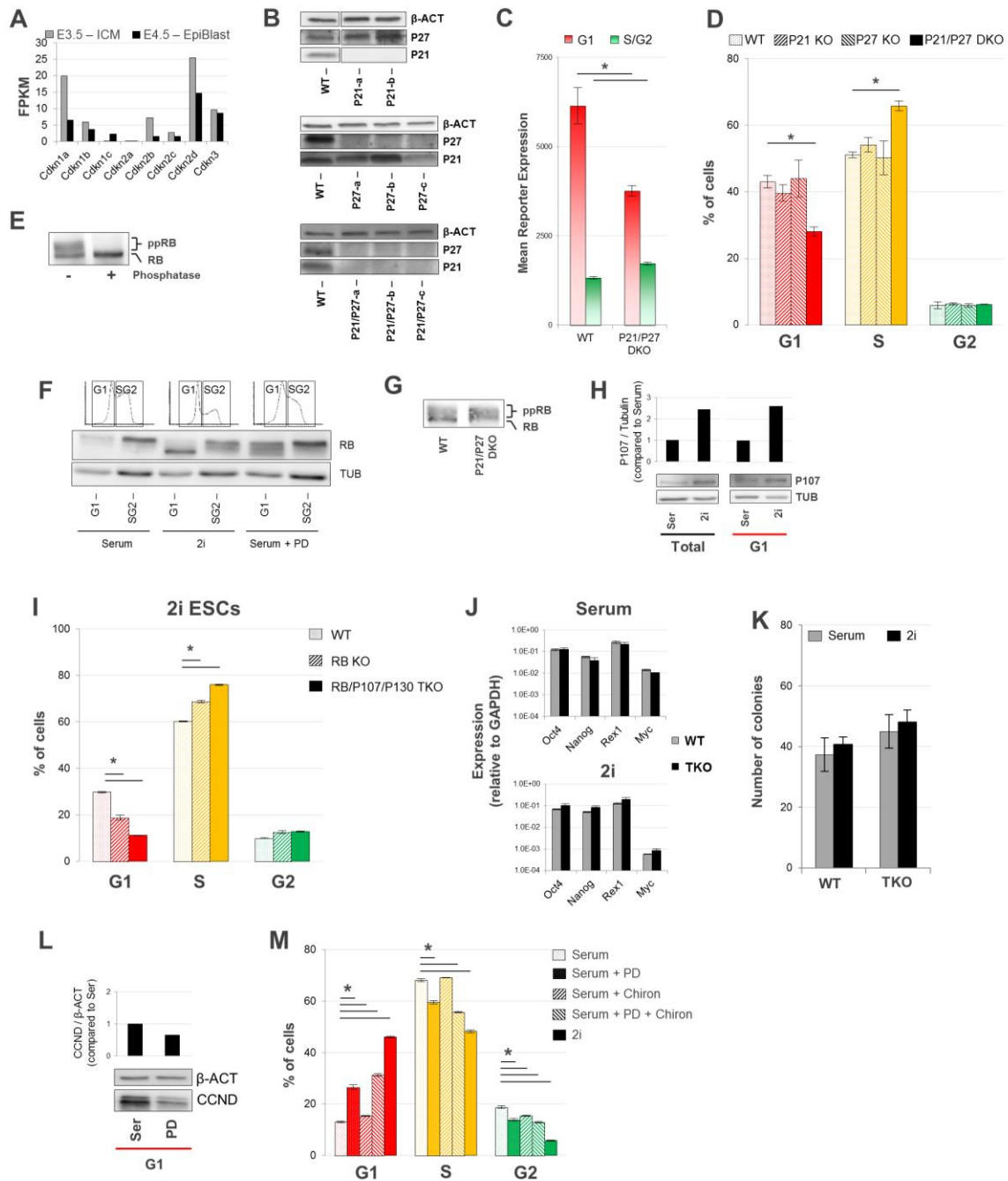
Figure S2 related to figure 1. The elevated expression of the CDK-inhibitors P21 and P27, hypo-phosphorylation of RB and increased number of cells in G1-phase in 2i as compared to serum ESCs. (A) RNA levels of CDK-inhibitors during early embryonic development in fragments per kilobase of exon per million fragments mapped (FPKM), adapted from [11] (https://doi.org/10.1016/j.devcel.2015.10.011). **(B)** Western blot analysis of P21, P27 in WT, single and combined KO ESCs fully adapted to 2i. **(C)** Expression of the FUCCI reporters in WT and P21/P27 DKO 2i ESCs indicates that combined knock out of P21 and P27 results in shortening of G1-phase (Student's T-test, *p-val<0.1·10⁻³). **(D)** Quantitation of Flow cytometry analysis using BrdU/PI showing a significant decrease in G1-phase (Student's T-test, *p-val<0.05) of P21/P27 DKO as compared to WT 2i ESCs. No significant changes were observed in P21 KO or P27 KO 2i ESCs compared to WT 2i ESCs. Experiment performed in triplicate, using two independent P21 KO clones, three independent P27 KO clones and three independent P21/P27 DKO clones. **(E)** Phosphatase treatment confirmed that the lower band present in 2i ESCs is hypo-phosphorylated RB. **(F)** The Suv39h WT ESC line [14] shows an elevated level of hypo-phosphorylated RB in G1-phase both in 2i as well as in serum plus PD0325901 (hereafter PD) conditions. **(G)** Western blot analysis phosphorylation of RB in P21/P27 DKO 2i ESCs compared to WT 2i ESCs. Two independent clones were assessed showing similar results. **(H)** Western blot analysis showing higher levels of P107 in 2i versus serum ESCs both in total cell lysates and lysates from G1-sorted cells. **(I)** Cell cycle analysis using BrdU/PI on RB KO and RB/P107/P130 TKO ESCs grown in 2i. A two-tailed Student's T-test was used to assess significance of the differences (*p-val<0.05). **(J)** Expression levels of pluripotency genes in RB/P107/P130 TKO ESCs cultured in 2i and serum conditions. **(K)** Colony formation assay as described by [31] performed in triplicate. **(L)** Western blot analysis and its quantitation of cyclin D protein in cell lysate from G1-phase sorted cells cultured in serum plus minus the PD inhibitor. **(M)** Quantitation of the distribution of serum cells upon adaptation to serum, serum plus PD, serum plus CHIR90221 (hereafter Chiron), serum plus PD and Chiron or 2i medium as assessed by BrdU/PI staining. Significance was tested using a two-tailed Student's T-test, *p-val<0.05. Bar charts and error bars represent the mean ± SD.

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Henk Stunnenberg (h.stunnenberg@ncmls.ru.nl).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines and culture conditions

All ESC lines, described including sex in Supplemental Table S1, were cultured in either serum medium, containing 15% fetal calf serum (Hyclone), penicillin/streptomycin, sodium pyruvate, 0.1mM 2-mercaptoethanol, and 1000 U/mL LIF or Ndiff 227 medium (Takara, formerly Ndiff B2N27 - StemCells), supplemented with CHIR99021 at 3 mM(Axon), PD0325901 at 1mM(Axon) and 1000 U/mL LIF (Millipore) ("2i") in a 37°C humidified incubator with 5% CO2. Prior to transition from serum medium to 2i medium or vice versa cells were washed twice with PBS.

METHOD DETAILS

Establishment of R1 FUCCI ESCs

mKO2-CDT1 (30-120) and AZ1-GEMININ (1-110) gene fragments were PCR amplified from 100 ng pcDNA3 backbone (gift from Miyawaki and colleagues [8]) using modified T7 and SP6 primers with EcoR1 restriction sites at their 5'-ends. PCR amplification was performed using Platinum Pfx DNA polymerase (Invitrogen) according to manufacturer protocol. After purification using the PCR purification Kit (Qiagen) these amplicons were digested with EcoR1 (NEB) according to manufacturer guidelines for 8 hours at 37°C. In addition, 5 µg expression vector containing the pCAG promoter and either the puromycin or the neomycin selection marker (a gift from Austin Smith and colleagues), was digested and treated with calf intestinal phosphatase (NEB). Next, the inserts and the expression plasmids were verified and purified by gel electrophoresis. The mKO2-CDT1 (30-110) fragment was ligated into pCAG-IRES-PURO and the AZ1-GEMININ (1-110) into pCAG-IRES-NEO expression plasmid overnight at 16°C using the DNA Ligation Mighty Mix Kit (Takara). The ligation products were transformed into DH5α E. Coli and plated onto agar plates containing carbenicillin (100 µg/mL). Next day colonies were picked and DNA was sent for sequencing. Clones of interest were amplified in LB containing carbenicillin (100 µg/mL) and DNA was isolated using the HP Endotoxin-free Maxi Prep kit (Sigma). For transfection, we used a two-step transfection selection strategy. First, pCAG-CDT1-PURO was transfected in mouse R1 ESCs and selected with puromycin (1 µg/ml) for approximately 2 weeks, and stable cell lines were isolated. Next, pCAG-GEMININ-NEO was transfected into the cell lines, and selected with G418 (200 µg/ml) for 2 weeks, and stable lines were isolated.

Western Blot

Cells were lysed in RIPA buffer with fresh cOmplete, EDTA-free protease inhibitor Cocktail (Roche) and PhosSTOP (Roche). Cell extracts were separated by SDS– PAGE and then transferred to nitrocellulose membranes in 20 mM Tris- HCl [pH 8.0], 150mM glycine, 20% (v/v)d methanol. Membranes were blocked with 5% (w/v) nonfat dry milk in Tris-buffered saline with Tween 20 (TBST; 20 mM Tris-HCl [pH 7.6], 0.1% Tween 20, 137 mM NaCl), incubated with primary antibodies, then secondary antibodies, and detected with ECL reagents (Amersham Biosciences).

Flow cytometry

The BD FACS Aria cell sorter was used to analyze and sort FUCCi ESCs. For cell cycle analysis cells were pulsed with 20 µM BrdU for 30 minutes, harvested by trypsinisation and fixed over night in 70% ethanol at 4°C. After denaturation of the DNA using 2N HCl + 0.5% Triton-X100 for 30 minutes at room temperature, neutralization with 0.1M Na2B4O7 (pH 8.5), samples were incubated with the anti-BrdU antibody over night at 4°C. Next day, the samples were stained with Propidium Iodide staining solution (10 ug/ml PI [Sigma, P4170] and 0.2mg/mL RNAse A in PBS) over night at 4°C. Samples of at least 10000 cells were acquired using a FACScalibur flow cytometre (Becton Dickinson). Subsequent analysis was done with Flowing

Software. For the cell cycle analysis of Nanog-GFP ESCs cells were incubated with Hoechst 33342 (Invitrogen) for one hour at 37°C and analyzed on the BD FACS Aria.

Colony formation assay

Serum grown ESCs were seeded at a dilution of one cell per well into 96-well plates using the BD FACS Aria, as described in [31], containing either serum or 2i medium. The cells were cultured for two weeks and the number of colonies was assessed in a blind manner.

ChIP-seq

Serum and 2i ESCs were fixed using 1% PFA for 10 min at room temperature. Subsequently, fixation was quenched by adding 1.25M glycine to a final concentration of 0.125M. After sonication the material of 4-5 million cells and 5 µL E2F1 antibody were used per ChIP. ChIP enrichment was assessed by qPCR and 2ng of DNA was used for library construction. Paired-end 43bp deep sequencing was performed using Illumina's NextSeq 500 sequencer.

Karyotype Analysis

For karyotype analysis chromatin of 0.5 million cells was sonicated and decrosslinked over night. Next genomic DNA was purified and used for library construction. Paired-end 43bp deep sequencing was performed using Illumina's NextSeq 500 sequencer. Reads were mapped using BWA and unique reads per chromosome were normalized by total reads.

RNA-seq

Cells were sorted and total RNA was extracted using Trizol (Life Technologies) according to manufacturer's instructions. After DNAse treatment, 5 µg of extracted RNA was depleted from ribosomal RNA using Ribo-Zero Gold Kit (Epicentre Madison, Winsconsin, USA). After fragmentation of the rRNA-depleted RNA, 500ng was reverse-transcribed using Super Script III Reverse Transcriptase (Invitrogen) and random primers (Invitrogen) following the manufacturer's instructions. Next, libraries were prepared using the TruSeq RNA Sample Prep Kit (Illumina) following the manufacturer's instructions. Libraries were indexed using NEXTflex adapters (Bioo- Scientific Corporation, Austin, TX, USA), and the quality was assessed by qPCR and Bioanalyzer (BioRad). Single-end 43bp deep sequencing was performed on Illumina instruments using TruSeq reagents (Illumina, San Diego, CA, USA), according to manufacturer's instructions.

Genome editing with CRISPR-Cas9

CRISPR-Cas9 gene editing was used to knock out *cdkn1a* (P21) and *cdkn1b* (P27). In brief, gRNAs were designed using the online tool (crispr.mit.edu) and cloned into the plasmid Cas9(BB)-2A-GFP (Addgene plasmid 48138) using the Bpi1 restriction sites as described previously[32]. FUCCI serum ESCs were transfected using lipofectamine-LTX (life technologies). After 48 h, GFP+ cells were sorted with a BD FACS Aria. Cells were split at clonal density and after approximately 7 days colonies were picked for expansion. Genomic DNA from individual clones was extracted using the Wizard Genomic DNA extraction kit. The

targeted region was PCR amplied and Sanger Sequenced. gRNA oligonucleotides were as follows: Cdkn1a-01_Fwd: CACCGTTGTCTCTTCGGTCCCG, Cdkn1a-01_Rev: AAACCGGGACCGAAGAGACAAC, Cdkn1a-02_Fwd: CACCGTCCGACCTGTTCCGCAC, Cdkn1a-02_Rev: AAACGTGCGGAACAGGTCGGAC Cdkn1b_Fwd: CACCGCGGATGGACGCCAGACAAG, Cdkn1b_Rev: AAACCTTGTCTGGCGTCCATCCGC.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bar charts represent the mean ± standard deviation of the mean (SD). When comparing two conditions, statistical differences were assessed in Microsoft Excel with a paired two-tailed Student's T test unless otherwise indicated in the legends. A p-value of < 0.05 was considered significant unless stated differently and the exact degree of significance as indicated by asterisks is stated in the legends. Pie charts display the means of an experiment performed in triplicate representative for at least two independent experiments. Quantification and statistics belonging to the pie charts are included in the figure or Supplemental Table S1.

DATA AND SOFTWARE AVAILABILITY

Software

BWA and bowtie were used for ChIP-seq and RNA-seq, respectively, to align sequencing reads to the mouse genome (mm9) using default parameters. For RNA-seq transcript quantification was performed using the MMSeq package and after setting a threshold of at least 50 reads over the gene body in either serum or 2i the DESeq2-package was used to call differentially expressed genes (log2-fold change >1 and a p-value < 0.05)[33]. Normalized read counts were subsequently used to calculate RPKM values. For ChIP-seq picard tools was used to remove duplicates (http:/broadinstitute.github.io/picard) and the Encode blacklist was used to filter out artifact regions[34]. Next, macs2 was used to call peaks in individual files and bedtools was used to intersect the peak-files of biological replicates. The peak-files of serum and 2i were merged using bedtools and reads over the genomic regions in resulting file were counted using bedtools multicov. The DESeq2-package was used to call differential peaks (log2-fold change >1 and a p-value < 0.05). GO-term analysis was performed with DAVID (http:// david.abcc.ncifcrf.gov/). Homer software was utilized to find de novo enriched motifs in the promoters of differentially expressed genes using default settings.

Data resources

The accession number for the RNA-seq data of serum and 2i ESCs in G1-phase as well as the E2F1 ChIP-seq data reported in this paper is Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) Database: GSE85690. The original unprocessed data is available through a Mendeley Database: http://dx.doi.org/10.17632/9hcdttwzyb.1.
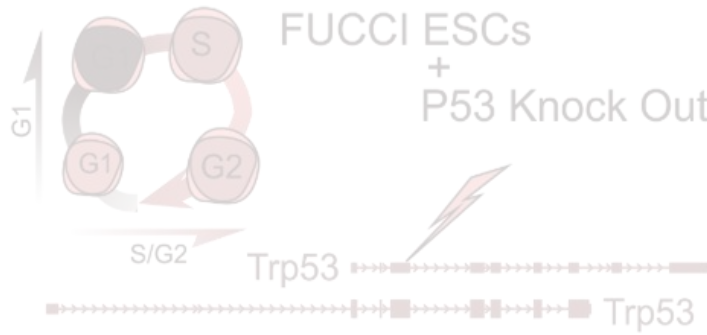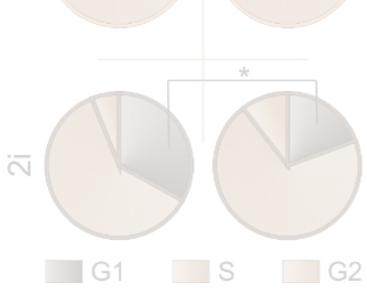
Chapter 2

The role of P53 in ground state embryonic stem cells

Menno ter Huurne[1], Hendrik G. Stunnenberg[1]

[1] Radboud University, Faculty of Science, Department of Molecular Biology, 6525GA Nijmegen, the Netherlands

Manuscript under preparation

Summary

In somatic cells P53 is involved in a plethora of regulatory pathways that act upon different types of stress and result in cell cycle arrest. On the contrary, in mouse Embryonic Stem Cells (ESCs) grown in serum-rich conditions the role of P53 in facilitating G1-phase arrest upon DNA damage has been reported to be compromised. Here we provide evidence that P53 modulates G1-phase progression upon transition of naive serum ESCs to their ground state. The expression of hypo-phosphorylated RB and elongation of G1-phase characteristic to ground state ESCs is dependent on P53. Moreover, in contrast to serum ESCs ground state ESCs accumulate in G1-phase upon induction of DNA damage suggesting that P53 elicits the DNA damage response pathway in ground state but not in serum ESCs. Furthermore, transcriptome analysis reveals that P53 activity induces the expression genes involved in early developmental processes specifically in ground state ESCs. Together, these findings suggest that P53 is active in ground state ESCs and that its activity is abolished in serum ESCs.

Introduction

Compared to somatic cells the cell cycle of mouse Embryonic Stem Cells (ESCs) cultured in the presence of serum and LIF is extremely short, mainly due to short Gap- (G-) phases. In particular the gap phase between mitosis and DNA replication, G1-phase, is very short in these mouse ESCs[1]. The continuous activity of Cyclin Dependent Kinases (CDKs), mainly CDK2, results in constitutively hyper-phosphorylated inactive RB and consequently in high E2F transcription factor activity and rapid progression into S-phase[2]. On the other hand, proteins that delay G1-progression, e.g. the CDK2-inhibitors P21 and P27, are not expressed in ESCs[2–5]. Because a key function of G1-phase is to ensure genome integrity prior to DNA replication these observations have raised questions about DNA damage response in ESCs. In particular, ESCs will give rise to the germline and mutations acquired in the early embryo will therefore be passed on to its progeny, stressing the importance of DNA damage control in the early embryo. The fact that spontaneous mutations occur at much lower frequency in ESCs when compared to somatic cells ($1/10^6$ versus $1/10^4$, respectively) suggests that ESCs have employed alternative ways to cope with DNA damage[6]. In somatic cells, DNA damage response pathways that are active during G1-phase induce either apoptosis or G1-arrest, thereby preventing progression into S-phase. Interestingly, ESCs do in contrast to somatic cells not go into G1 arrest upon DNA damaging treatments[7,8].

Bypass of the G1 checkpoint in ESCs has been attributed to high levels of the CDK2 activator CDC25A[9] and the lack of a P53-mediated DNA damage response[7,10]. P53 is the most frequently mutated gene in cancer[11] and plays a pivotal role in DNA damage induced stress responses. Activated P53 can induce either apoptosis, via canonical Bak / Bax, or induce a G1-arrest via its downstream effector P21[12]. The response of ESCs to treatment with anti-metabolites or ionizing radiation was similar to Mouse Embryonic

Fibroblasts (MEFs) that lacked P53, in that they both progressed into S-phase. Despite phosphorylation and activation of P53 upon irradiation, no increase in P21 protein levels was observed in ESCs[10]. Some studies have suggested that the lack of a functional P53/P21 DNA damage response is the result of cytoplasmic localization and inefficient translocation to the nucleus of P53 or the activity of micro RNAs[7,13]. More recent studies have however indicated that proteosomal degradation of P21 inhibits the accumulation of P21 protein upon DNA damage and thereby prevents induction of G1-arrest[8].

Although the P53-mediated DNA damage response pathways seems to be abolished in in ESCs in vivo data has indicated that P53 is active in the Inner Cell Mass (ICM) during early embryonic development[14]. These observations are in line with growing evidence that P53 plays in important role in embryonic development and differentiation. The exact role of P53 in embryonic development is however still unclear. Although P53 is highly expressed in the early embryo, P53 null mutant mice do not show developmental defects, strongly suggesting that P53 is not essential for embryonic development[15]. However, other studies have shown that loss of P53 can result in subtle developmental defects (reviewed by Shin, M. H., He, Y. & Huang, J.[16]) and that P53 is essential for normal embryonic development in *Xenopus laevis*[17]. These seemingly contradictory results could be due to the functional redundancy of P53 with its family members P63 and P73[18,19]. More recent studies in ESCs have shown that P53 plays an important role in the regulation of developmental genes during differentiation[20] and that P53 orchestrates mesoderm differentiation by modulating Wnt signaling[21]. Interestingly, a role of P53 in stem cell maintenance has also reported in the literature, either by regulating Wnt signaling[22] or by enhancing TGF-β signaling[23]. Together these reports indicate that the function of P53 depends on the cellular state and its environmental factors.

Although ESCs are classically cultured in the presence of serum, in the last decade new culture conditions have been developed[24,25]. Studies on the epigenome, transcriptome have revealed that these culture conditions give rise to different flavors of ESCs reflecting different developmental stages[5,26]. In this regard, we have previously shown that the short G1-phase is characteristic to naive serum ESCs and is the result of ERK signaling. The latter pathway is inhibited in ground state ESCs cultured in serum-free 2i conditions leading to an elongated G1-phase[3]. Because the shortened G1 is accompanied by abrogated DNA damage response pathways these observations imply reactivation of DNA damage response pathways in 2i ESCs. Therefore, we set out to decipher the role of P53 in ground state 2i conditions. To gain a better understanding of the role of P53 in 2i conditions we have deleted P53 in a mouse ESC cell line expressing the FUCCI reporters. These reporters allow the designation of cells throughout the different phases of the cell cycle and subsequent analysis of specific populations. The data presented here imply that the function of P53 is compromised during the abbreviated G1-phase in naive serum ESCs and that P53 is plays a role in G1-phase progression in ground state 2i ESCs. Moreover, data on genome-wide P53 binding and the transcriptome of P53$^{-/-}$ 2i ESCs suggests that P53 regulates the expression of developmental genes in ground state ESCs.

2

Results

P53 regulates G1-phase progression in 2i ESCs

Our previous studies revealed lengthening of G1-phase in ground state 2i ESCs when compared to naive serum ESCs. Moreover, an increase in the expression of several Cyclin-Dependent Kinase (CDK)-inhibitors (CDKi), among which the P53-target gene Cdkn1a (coding for P21), during G1-phase was observed in ground state 2i ESCs[3,5]. The deletion of P21 resulted however only a minor effect on the cell cycle of 2i ESCs (Figure 1A and ter Huurne, M.[3]). To assess the effect on gene expression we sorted WT and P21[-/-] 2i ESCs in G1-phase and performed whole genome RNA-seq. In line with the limited effect on the cell cycle, only a small set of genes (~0.2%) was either significantly up- or down-regulated (Log2-fold Change > 1 or < -1, Adjusted P-value <0.05). The deletion of P21 did however result in a moderate increase in the expression of genes involved in chromatin assembly, mainly histone genes, suggesting that P21 does prevent the induction of DNA replication to some extent (Figure 1B). Accordingly, the combined deletion of Cdkn1a and Cdkn1b, coding for P27, resulted in shortening of G1-phase[3], indicating that these proteins prevent S-phase entry in 2i ESCs. The elevated expression of P21 and elongated G1-phase led us to hypothesize that P53 activity is increased in 2i ESCs and contributes to cell cycle regulation of ground state ESCs. Although previous studies have not identified Trp53 as being differentially expressed between serum and 2i ESCs[3,5] P53 binding and activity is regulated by several phosphorylation and acetylation events[27,28]. Quantification of P53 protein levels in several fractions of the cell indicated that the level of chromatin-bound P53 is higher in ground state 2i ESCs when compared to naive serum ESCs (Figure 1C). To determine whether the expression of P21 is dependent on P53 we used the Crispr-Cas gene editing system to create three independent P53[-/-] clones in R1 ESCs that express the FUCCI reporter constructs (Figure 1D). Western blot analysis showed successful targeting of P53 as well as a decrease in P21 levels in 2i P53[-/-] ESCs (Figure 1E)[29].

To determine the effect of P53 deletion on the cell cycle we next measured the expression of the FUCCI reporters. As we reported previously the number of the WT 2i ESCs in late G1-phase was much higher when compared to WT serum ESCs[30]. Upon deletion of P53 there was however a dramatic decrease in the number of 2i ESCs in late G1-phase (Figure 1F and 1G). Serum ESCs enter S-phase prematurely and therefor lack cells in late G1-phase. Accordingly, virtually no effect on the cell cycle was observed in P53[-/-] serum ESCs. We next made use of an independent P53[-/-] cell line to assess the distribution of cells over the different phases of the cell cycle using BrdU / PI staining[20]. In 2i conditions the number of cells in G1-phase was significantly lower in P53[-/-] ESCs when compared to WT, confirming our previous observations. Again no effect of loss of P53 was observed in serum ESCs (Figure 1H). Because P21 is primarily expressed during G1-phase in 2i ESCs [3] the decreased expression in P53[-/-] ESCs could also be the result of the diminished number of cells in G1-phase. A western blot on G1-phase sorted cells shows however that the expression of P21 is lowered specifically in 2i G1-phase cells (Figure 1I). Together these results suggest that P53 regulates P21 expression and G1-phase progression in ground state 2i ESCs whereas it

has a minor role in serum ESCs. To test whether the P53/P21-mediated DNA damage checkpoint is activated in 2i cells we used doxorubicin to induce DNA damage in serum and 2i ESCs. As expected, we did not observe an increase in the number of serum cells in G1-phase upon doxorubicin treatment due to the lack of the G1-checkpoint[10]. In 2i cells however an increase in the number of cells in G1 was observed, suggesting that these cells have an activated DNA damage response pathway (Figure 1J).
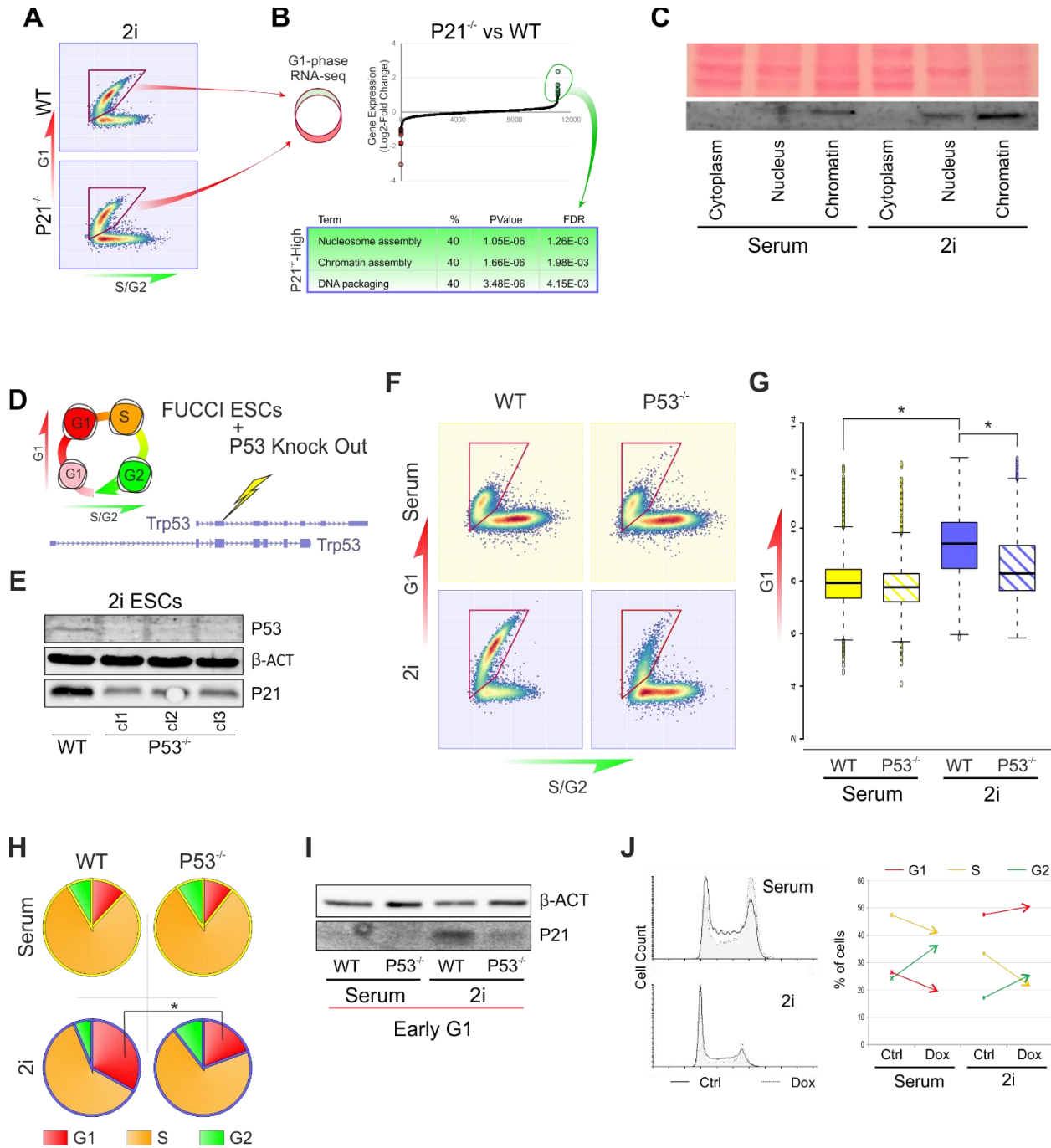
2

Figure 01: P53 is essential for the elongated G1-phase in 2i ESCs. (A) FUCCI reporter expression in WT and P21[-/-] 2i ESCs showing only minor differences in the distribution of cells in G1- and S-/G2-phase. **(B)** The transcriptome of P21[-/-] 2i ESCs displays increased expression of genes involved in nucleosome assembly and DNA packaging. **(C)** Higher P53 protein levels are observed in the nuclear fraction as well as on the chromatin in 2i ESCs when compared to serum ESCs. **(D + E)** Construction of P53[-/-] R1 FUCCI ESCs and a WB on P53 and P21 in three independent P53[-/-] clones in 2i. **(F)** When compared to WT the P53[-/-] ESCs have lowered expression of the G1-phase specific FUCCI reporter in 2i. In serum no differences between WT and P53[-/-] ESCs are observed. **(G)** Quantification of the G1-phase reporter expression using three independent P53[-/-] clones in serum and 2i. For WT two biological replicates were used. **(H)** Quantification of WT and P53[-/-] EB5 cells over the different phases of the cell cycle using Bromo-deoxy-Uridine (BrdU) incorporation combined with Propidium Iodide (PI) staining. **(I)** WB on P21 showing decreased expression of P21 during G1-phase in P53[-/-] ESCs. Significance was assessed using a two-tailed students T-test, * P-value < 0.05. **(J)** The doxycycline treatment of 2i ESCs results in an increase in the relative amount of cells in G1-phase whereas only an increase in number of G2-phase cells is observed in serum ESCs.

Genes involved in cell cycle arrest and apoptosis are downregulated upon deletion of P53 in ground state 2i and serum ESCs

To determine the impact of loss of P53 on the transcriptome of serum and 2i ESCs we performed RNA-seq on G1-phase sorted WT and P53[-/-] ESCs in serum and 2i. The PCA plot in figure 2A shows that the variation between WT and P53[-/-] is larger in 2i conditions when compared to serum conditions (Figure 2A). In addition, the percentage of differentially expressed genes between WT and P53[-/-] ESCs was higher in 2i when compared to serum (Figure 2B). Although the cell cycle of P53[-/-] serum ESCs was hardly affected, genes lower expressed in these cells when compared to WT cells were involved in apoptosis and cell cycle arrest both in serum as well as 2i (Figure 2C). Surprisingly, genes that showed lowered expression in 2i P53[-/-] ESCs when compared to 2i WT ESCs specifically were not only involved in DNA response but also in developmental processes (Figure 2D). Although previous reports have described a role of P53 upon differentiation, this is the first time that such an effect is observed in pluripotent ESCs, suggesting that serum conditions conceal this function of P53 in ESCs. Although P53 has been shown to be able to drive differentiation by inhibiting the expression of Nanog[31], we did not observe an increase in the expression of Nanog in 2i nor in serum P53[-/-] ESCs (Figure 2E). These findings are consistent with a previous report showing no effect of loss of P53 on Nanog expression[20].
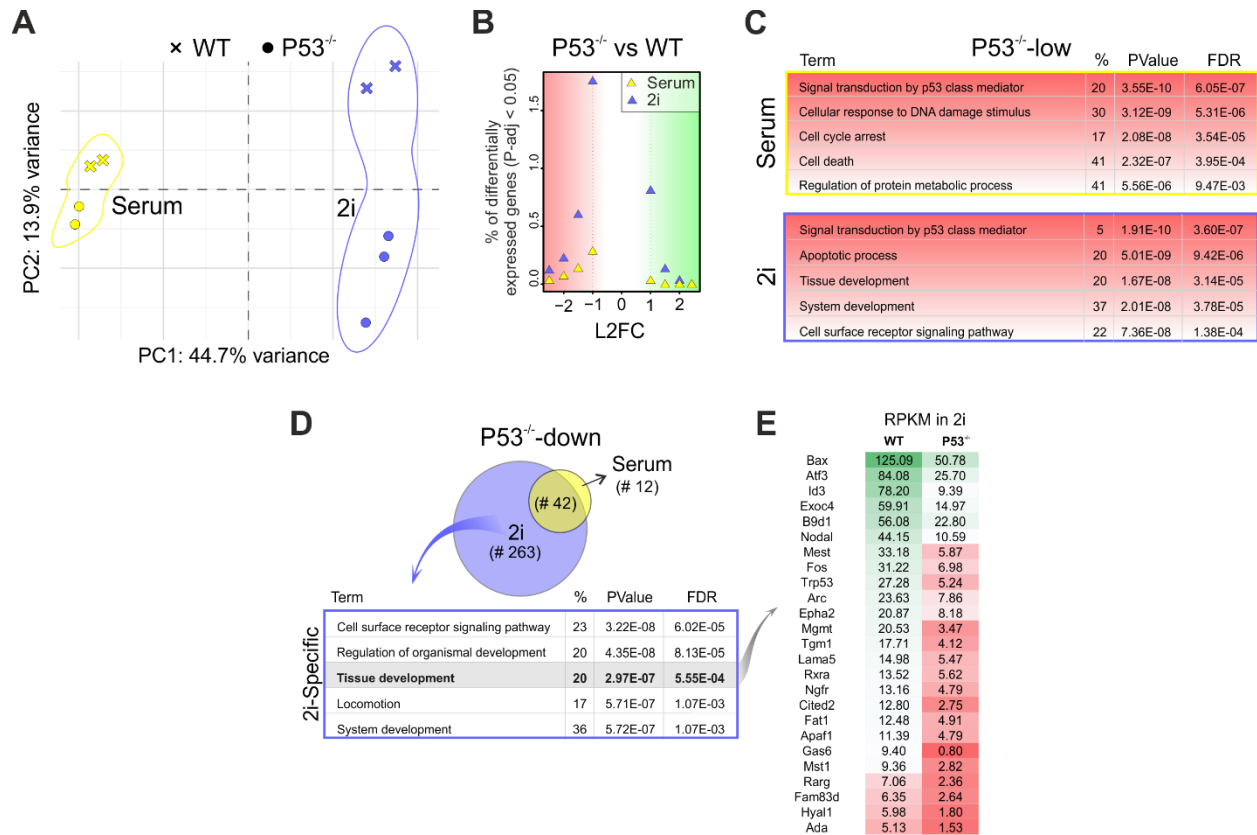
Figure 2: Deletion of P53 results in lowered expression of genes involved in apoptosis and organismal development in 2i ESCs. (A) Principal Component Analysis on gene expression in WT and P53[-/-] ESCs in serum and in 2i. **(B)** The number of genes differentially expressed between WT and P53[-/-] ESCs is higher in 2i than in serum conditions. **(C)** Gene Ontology analysis revealing lowered expression of genes involved in apoptosis and development upon deletion of P53 in 2i ESCs. **(D)** Genes involved in organismal and tissue development are specifically downregulated in P53[-/-] ESCs in 2i conditions. **(E)** Table containing the expression values of genes involved in tissue development that are down-regulated in P53[-/-] 2i ESCs.

The pocket proteins Rb and P130 are direct transcriptional targets of P53

The most prominent difference between WT and P53[-/-] 2i ESCs is the shortened G1-phase in P53[-/-] cells, resembling the G1-phase of WT serum ESCs. Although the expression of P21 is lowered in P53[-/-] 2i ESCs we did not observe lowered expression of the CDK-inhibitor Cdkn1b (P27) (Figure 3A). Since deletion of P21 alone has only a minor effect on the cell cycle (Figure 1A and[3]) these results suggest that the shortened G1-phase as the result of loss of P53 is mediated via other cell cycle regulators. We therefore examined the expression of other genes involved in G1-phase progression (Figure 3A). The downstream targets of the CDK/CYCLIN pathway are the pocket proteins RB, P107 and P130. The abbreviated G1-phase in serum ESCs is characterized by hyper-phosphorylated and therefor inactive RB. Inhibition of the ERK-signaling pathway however results in lowered CYCLIN-D levels, the appearance of hypo-phosphorylated

RB and reinstatement of the G1-checkpoint in 2i ESCs[3]. The pocket proteins, among others RB and P130, control G1-phase progression in 2i ESCs by inhibiting the E2F transcription factors[3].

In P53[-/-] ESCs, both Rb1 and Rbl2 (coding for RB and P130, respectively) RNA levels are decreased when compared to WT ESCs (L2FC -0.88 and -0.76, respectively) (Figure 3B). Since these proteins are active in 2i but not in serum ESCs this could explain why the effect of deletion of P53 has a much stronger effect in 2i ESCs when compared to serum ESCs. To assess how P53 regulates the expression of Rb1 we consulted ChIP-seq data on P53 in serum ESCs[32]. In serum and 2i ESCs P53 binds to the Transcriptional Start Site (TSS) and gene body of Rb1 and P130, suggesting a direct transcriptional regulation (Figure 3C). In line with literature only inactive hyper-phosphorylated RB was expressed in serum ESCs, whereas hyper- as well as active hypo-phosphorylated RB was expressed in 2i ESCs[3]. In both conditions the expression was drastically reduced in the P53[-/-] ESCs (Figure 3D). Since cell cycle distribution of asynchronously growing WT and P53[-/-] 2i ESCs is different, we next determined RB expression levels throughout the cell cycle in WT and P53[-/-] serum and 2i ESCs. Again, a clear reduction of RB expression in P53[-/-] ESCs was observed (Figure 3E). These results suggest that P53 regulates the transcription of the pocket proteins RB and P130 thereby elongating G1-phase in 2i ESCs. Due to increased ERK-/CDK/Cyclin-signaling RB is constitutively hyper-phosphorylated in serum ESCs and the deletion of P53 has hardly any effect in these cells (Figure 3F).
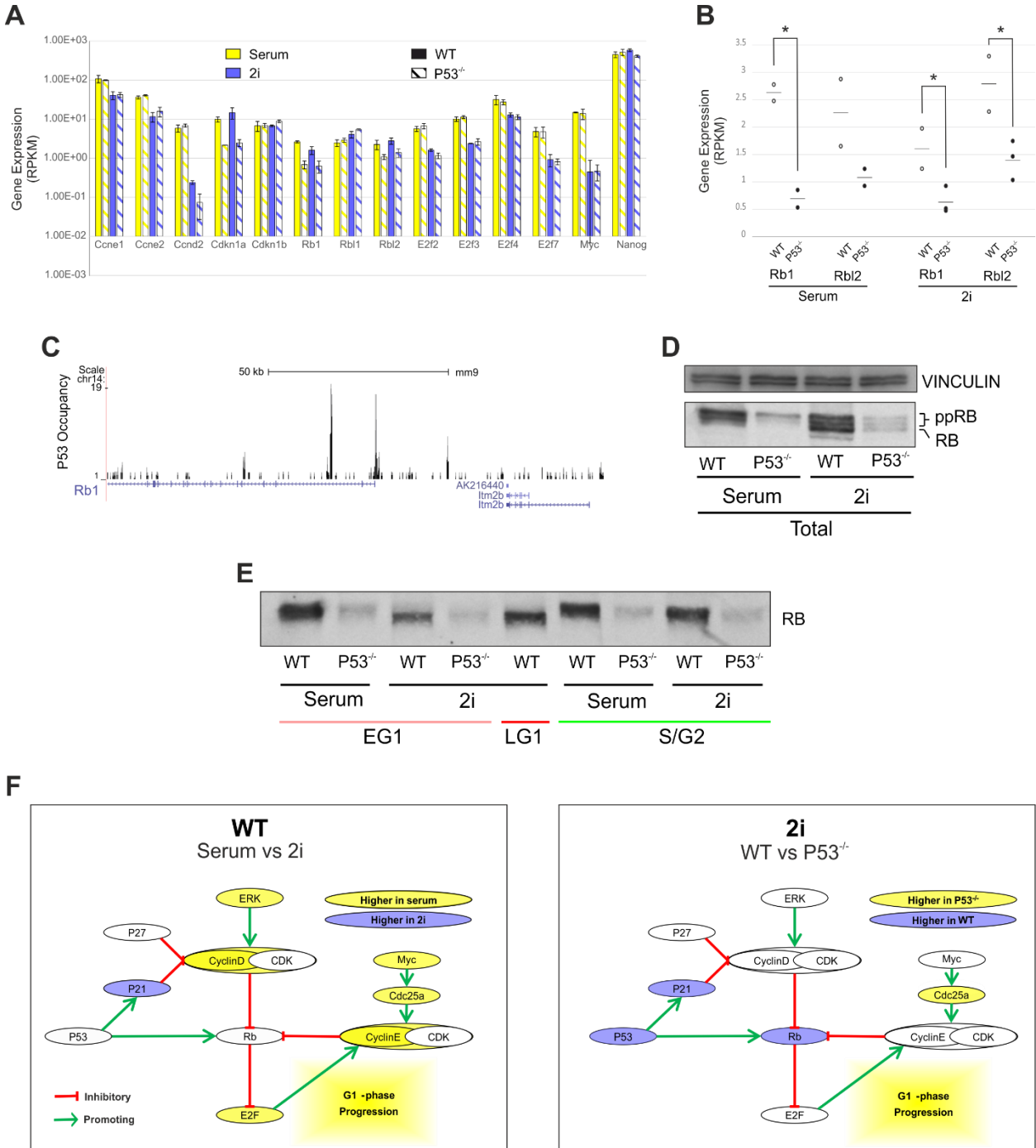
Figure 3: P53 regulates G1-phase progression via the expression of the pocket proteins Rb and P130. (A) Bar graph showing differential expression of cell cycle regulators in serum and 2i WT and P53[-/-] ESCs during G1-phase. **(B)** The expression of Rb1 and Rbl2 is down in P53[-/-] both in serum and 2i ESCs. *P-value<0.05. **(C)** ChIP-seq data shows binding of P53 at the Transcriptional Start Site (TSS) of Rb. **(D)** The expression of hyper-phosphorylated Rb and hypo-phosphorylated Rb is reduced in P53[-/-] ESCs in serum and 2i, respectively. **(E)** The expression of Rb is lowered throughout the cell cycle in P53[-/-] ESCs when compared to WT ESCs. **(F)** Proposed model that displays the differences in cell cycle control between serum and 2i WT ESCs (left panel) and the differences between WT and P53[-/-] ESCs in 2i (right panel).

Although our data indicate that loss P53 has a bigger impact on the cell cycle of 2i cells than on that of serum ESCs, it remains largely elusive what causes this difference. To our surprise, the expression of P53 was higher in serum ESCs (L2FC ≈ 1.23, Adjusted P-value < 0.05), suggesting that regulation takes place on protein level. Firstly, GSK3 promotes the degradation of P53[33] and the elevated activity of P53 in 2i might therefor be the result of the inhibition of this pathway by chiron. Secondly, the activity of P53 can be influenced by a range of posttranslational modifications, like phosphorylation, acetylation and sumo modifications[34]. The Aurora A kinase that can inhibit P53-activity[35] is significantly lower expressed in 2i when compared to serum ESCs (L2FC ≈ -1.10, Adjusted Pvalue < 0.05, data not shown), whereas its negative regulator is higher expressed in 2i (L2FC ≈ 0.7, data not shown). In addition, Sirt1, a deacetylase that has shown to block translocation of P53 into the nucleus in ESCs[36], is lower expressed in 2i ESCs (L2FC ≈ -1.29, Adjusted Pvalue < 0.05). Together these processes might contribute to the increased activity of P53 in 2i ESCs. To what extent these and other mechanisms impact on the role of P53 in 2i ESCs will be an interesting subject for future studies.

Discussion

Although P53 is highly expressed in ESCs and active in the early embryo its function is still subject of debate. Early studies in serum ESCs have suggested that it cannot act as a regulator of G1-phase progression due to functional uncoupling of the P53/P21 axe[8]. We have previously shown that the abbreviated G1-phase in serum ESCs is the result of ERK signaling and Rb hyper-phosphorylation[3]. The results presented here suggest that the abbreviated G1-phase in serum ESCs obscures an unexplored function of P53 in the cell cycle of ESCs. In ground state 2i ESCs, P53 regulates the expression of P21 and the pocket proteins Rb and P130. This function of P53 is required for the elongated G1-phase in 2i conditions. These results suggest that in contrast to serum ESCs DNA damage control pathways are functional in 2i ESCs [7]. On the other hand, the activation of the P53/P21 pathway is possibly the result of decreased DNA methylation and lowered genomic integrity in 2i ESCs[37]. Follow up studies using DNA damaging agents are however needed to test whether 2i ESCs are indeed less prone to DNA damage.
Our findings suggest that in the early embryo where ERK signaling is absent, P53 is plays a role during G1-phase and to restrict cell proliferation. Interestingly, the cellular senescent state that resembles diapause in vivo is dependent on the presence of the pocket proteins[3,38]. The extension of our observations is that

P53 plays a role during diapause. Possibly, P53 is highly expressed in early rodent embryos in order to induce diapause in response to stressful conditions.

Besides the differences in cell cycle control between WT and P53[-/-] 2i ESCs, the RNA-seq data uncovered a large number of genes affected by loss of P53 in 2i conditions (Figure 2D). As opposed to serum conditions, a significant proportion of the differentially expressed genes between WT and P53[-/-] ESCs in 2i is involved in systems and tissue development. These findings are in line with previous reports that suggest an important role of P53 in differentiation, although the underlying mechanisms might not be the same. How P53 regulates the expression developmental genes in 2i remains to be determined, possibly the differential regulation of pocket proteins plays a role, considering their role in development and differentiation[39,40].

Altogether we show that in ESCs the function of P53 depends on their cellular state. In ground state 2i ESCs P53 is involved in cell cycle control and regulates the expression of developmental genes.

Materials and Methods

Cell lines and culture conditions

All ESC lines were cultured in either serum medium, containing 15% fetal calf serum (Hyclone), penicillin/streptomycin, sodium pyruvate, 0.1mM 2-mercaptoethanol, and 1000 U/mL LIF or Ndiff 227 medium (Takara, formerly Ndiff B2N27 - StemCells), supplemented with CHIR99021 at 3 µM (Axon), PD0325901 at 1 µM (Axon) and 1000 U/mL LIF (Millipore) ("2i") in a 37°C humidified incubator with 5% CO_2. Prior to transition from serum medium to 2i medium or vice versa cells were washed twice with PBS.

Western Blot

Cells were lysed in RIPA buffer with fresh cOmplete, EDTA-free protease inhibitor Cocktail (Roche) and PhosSTOP (Roche). Cell extracts were separated by SDS– PAGE and then transferred to nitrocellulose membranes in 20 mM Tris- HCl [pH 8.0], 150mM glycine, 20% (v/v)d methanol. Membranes were blocked with 5% (w/v) nonfat dry milk in Tris-buffered saline with Tween 20 (TBST; 20 mM Tris-HCl [pH 7.6], 0.1% Tween 20, 137 mM NaCl), incubated with primary antibodies, then secondary antibodies, and detected with ECL reagents (Amersham Biosciences).

Flow cytometry

The BD FACS Aria cell sorter was used to analyze and sort FUCCi ESCs. For FUCCI reporter quantification CSV files were analyzed using the Flowcore Package. For cell cycle analysis cells were pulsed with 20 µM BrdU for 30 minutes, harvested by trypsinisation and fixed over night in 70% ethanol at 4°C. After denaturation of the DNA using 2N HCl + 0.5% Triton-X100 for 30 minutes at room temperature, neutralization with 0.1M $Na_2B_4O_7$ (pH 8.5), samples were incubated with the anti-BrdU antibody over night at 4°C. Next day, the samples were stained with Propidium Iodide staining solution (10 ug/ml PI [Sigma, P4170] and 0.2mg/mL RNAse A in PBS) over night at 4°C. Samples of at least 10000 cells were acquired

using a FACScalibur flow cytometre (Becton Dickinson). Subsequent analysis was done with Flowing Software.

ChIP-seq

Serum and 2i ESCs were fixed using 1% PFA for 10 min at room temperature. Subsequently, fixation was quenched by adding 1.25M glycine to a final concentration of 0.125M. After sonication the material of 4-5 million cells and 5 μL E2F1 antibody were used per ChIP. ChIP enrichment was assessed by qPCR and 2ng of DNA was used for library construction. Paired-end 43bp deep sequencing was performed using Illumina's NextSeq 500 sequencer.

RNA-seq

Cells were sorted and total RNA was extracted using Trizol (Life Technologies) according to manufacturer's instructions. After DNAse treatment, 5 μg of extracted RNA was depleted from ribosomal RNA using Ribo-Zero Gold Kit (Epicentre Madison, Winsconsin, USA). After fragmentation of the rRNA-depleted RNA, 500ng was reverse-transcribed using Super Script III Reverse Transcriptase (Invitrogen) and random primers (Invitrogen) following the manufacturer's instructions. Next, libraries were prepared using the TruSeq RNA Sample Prep Kit (Illumina) following the manufacturer's instructions. Libraries were indexed using NEXTflex adapters (Bioo- Scientific Corporation, Austin, TX, USA), and the quality was assessed by qPCR and Bioanalyzer (BioRad). Paired-end 36bp deep sequencing was performed on Illumina instruments using TruSeq reagents (Illumina, San Diego, CA, USA), according to manufacturer's instructions.

Genome editing using CRISPR-Cas9

CRISPR-Cas9 gene editing was used to knock out *Trp53* (P53). In brief, gRNAs were designed using the online tool (crispr.mit.edu) and cloned into the plasmid Cas9(BB)-2A-GFP (Addgene plasmid 48138) using the Bpi1 restriction sites as described previously[41]. FUCCI serum ESCs were transfected using lipofectamine-LTX (life technologies). After 48 h, GFP+ cells were sorted with a BD FACS Aria. Cells were split at clonal density and after approximately 7 days colonies were picked for expansion. Genomic DNA from individual clones was extracted using the Wizard Genomic DNA extraction kit. The targeted region was PCR amplified and Sanger Sequenced. gRNA oligonucleotides were as follows: P53_gRNA_Fwd: CACCGGAGCTCCTGACACTCGG, P53_gRNA_Rev: AAACCCGAGTGTCAGGAGCTCC.

Quantification And Statistical Analysis

For boxplots displaying reporter expression all events of the indicated number of biological replicates were combined. Subsequently the different conditions were compared and statistical differences was assessed in Microsoft Excel with a two-tailed Student's T test. Pie charts display the means of an experiment performed in triplicate.

Software

BWA and bowtie were used for ChIP-seq and RNA-seq, respectively, to align sequencing reads to the mouse genome (mm9) using default parameters. For RNA-seq transcript quantification was performed using the MMSeq package and after setting a threshold of an average of at least 5 reads over the gene body in all conditions the DESeq2-package was used to call differentially expressed genes (log2-fold change >1 and a adjusted p-value < 0.05)[42]. Normalized read counts were subsequently used to calculate RPKM values.

References

1.   Abe, T. *et al.* Visualization of cell cycle in mouse embryos with Fucci2 reporter directed by Rosa26 promoter. *Development* **140,** 237–46 (2013).

2.   Stead, E. *et al.* Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. *Oncogene* **21,** 8320–33 (2002).

3.   ter Huurne, M., Chappell, J., Dalton, S. & Stunnenberg, H. G. Distinct Cell-Cycle Control in Two Different States of Mouse Pluripotency. *Cell Stem Cell* **21,** 449–455.e4 (2017).

4.   Savatier, P., Huang, S., Szekely, L., Wiman, K. G. & Samarut, J. Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. *Oncogene* **9,** 809–818 (1994).

5.   Marks, H. *et al.* The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. *Cell* **149,** 590–604 (2012).

6.   Cervantes, R. B., Stringer, J. R., Shao, C., Tischfield, J. A. & Stambrook, P. J. Embryonic stem cells and somatic cells differ in mutation frequency and type. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 3586–90 (2002).

7.   Aladjem, M. I. *et al.* ES cells do not activate p53-dependent stress responses and undergo p53-independent apoptosis in response to DNA damage. *Curr. Biol.* **8,** 145–155 (1998).

8.   Suvorova, I. I., Grigorash, B. B., Chuykin, I. A., Pospelova, T. V & Pospelov, V. A. G1 checkpoint is compromised in mouse ESCs due to functional uncoupling of p53-p21Waf1 signaling. *Cell Cycle* **15,** 52–63 (2016).

9.   van der Laan, S., Tsanov, N., Crozet, C. & Maiorano, D. High Dub3 expression in mouse ESCs couples the G1/S checkpoint to pluripotency. *Mol. Cell* **52,** 366–79 (2013).

10.   Hong, Y. & Stambrook, P. J. Restoration of an absent G1 arrest and protection from apoptosis in embryonic stem cells after ionizing radiation. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 14443–8 (2004).

11.   Vähäkangas, K. Molecular epidemiology of human cancer risk. Gene-environment interactions and p53 mutation spectrum in human lung cancer. *Methods Mol. Med.* **74,** 43–59 (2003).

12.   Dulić, V. *et al.* p53-dependent inhibition of cyclin-dependent kinase activities in human fibroblasts during radiation-induced G1 arrest. *Cell* **76,** 1013–1023 (1994).

13.   Dolezalova, D. *et al.* MicroRNAs regulate p21(Waf1/Cip1) protein expression and the DNA damage response in human embryonic stem cells. *Stem Cells* **30,** 1362–72 (2012).

14.   Goh, a. M. *et al.* Using targeted transgenic reporter mice to study promoter-specific p53 transcriptional activity. *Proc. Natl. Acad. Sci.* **109,** 1685–1690 (2012).

15.   Donehower, L. A. *et al.* Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356,** 215–221 (1992).

16.   Shin, M. H., He, Y. & Huang, J. Embryonic stem cells shed new light on the developmental roles of p53. *Cell Biosci.* **3,** 42 (2013).

17.   Wallingford, J. B., Seufert, D. W., Virta, V. C. & Vize, P. D. p53 activity is essential for normal development in Xenopus. *Curr. Biol.* **7,** 747–757 (1997).

2

18.   Flores, E. R., Tsai, K. Y. & Crowley, D. p63 and p73 are required for p53- dependent apoptosis in response to DNA damage. *Nature* **416,** 560–565 (2002).

19.   Dötsch, V., Bernassola, F., Coutandin, D., Candi, E. & Melino, G. P63 and P73, the Ancestors of P53. *Cold Spring Harb. Perspect. Biol.* **2,** 1–15 (2010).

20.   Shigeta, M. *et al.* Maintenance of pluripotency in mouse ES cells without Trp53. *Sci. Rep.* **3,** 2944 (2013).

21.   Wang, Q. *et al.* The p53 Family Coordinates Wnt and Nodal Inputs in Mesendodermal Differentiation of Embryonic Stem Cells. *Cell Stem Cell* 1–17 (2016). doi:10.1016/j.stem.2016.10.002

22.   Lee, K.-H. *et al.* A genomewide study identifies the Wnt signaling pathway as a major target of p53 in murine embryonic stem cells. *Proc. Natl. Acad. Sci.* **107,** 69–74 (2010).

23.   Gonzales, K. A. U. *et al.* Deterministic Restriction on Pluripotent State Dissolution by Cell-Cycle Pathways. *Cell* **162,** 564–579 (2015).

24.   Kolodziejczyk, A. A. *et al.* Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell* **17,** 471–485 (2015).

25.   Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–23 (2008).

26.   Habibi, E. *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13,** 360–9 (2013).

27.   Sakaguchi, K. *et al.* DNA damage activates p53 through a phosphorylation – acetylation cascade. *Genes Dev.* **12,** 2831–2841 (1998).

28.   Appella, E. & Anderson, C. W. Post-transitional modifications and activation of p53 by gentoxic stress. *Eur. J. Biochem.* **2772,** 2764–2772 (2001).

29.   El-Deiry, W. S. *et al.* WAF1, a Potential Mediator of p53 Tumor Suppression. *Cell* **75,** 817–825 (1993).

30.   ter Huurne, M. *et al.* Distinct Cell-Cycle Control in Two Different States of Mouse Pluripotency Brief Report Distinct Cell-Cycle Control in Two Different States of Mouse Pluripotency. *Cell Stem Cell* **21,** 449–455.e4 (2017).

31.   Lin, T. *et al.* p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. *Nat Cell Biol* **7,** 165–171 (2005).

32.   Lee, K.-H. *et al.* A genomewide study identifies the Wnt signaling pathway as a major target of p53 in murine embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 69–74 (2010).

33.   Proctor, C. J. & Gray, D. A. GSK3 and p53 - Is there a link in Alzheimer's disease? *Mol. Neurodegener.* **5,** 1–15 (2010).

34.   Marouco, D., Garabadgiu, A. V, Melino, G. & Barlev, N. A. Lysine-specific modifications of p53: a matter of life and death? *Oncotarget* **4,** 1556–1571 (2013).

35.   Lee, D.-F. *et al.* Regulation of embryonic and induced pluripotency by aurora kinase-p53 signaling. *Cell Stem Cell* **11,** 179–94 (2012).

36.   Han, M. K. *et al.* SIRT1 Regulates Apoptosis and Nanog Expression in Mouse Embryonic Stem Cells by Controlling p53 Subcellular Localization. *Cell Stem Cell* **2,** 241–251 (2008).

37.   Choi, J. *et al.* Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature* **548,** 219–223 (2017).

38.   Scognamiglio, R. *et al.* Myc Depletion Induces a Pluripotent Dormant State Mimicking Diapause. *Cell* 668–680 (2016). doi:10.1016/j.cell.2015.12.033

39.   Julian, L. M. & Blais, A. Transcriptional control of stem cell fate by E2Fs and pocket proteins. *Front. Genet.* **6,** 1–15 (2015).

40.   Calo, E. *et al.* Rb regulates fate choice and lineage commitment in vivo. *Nature* **466,** 1110–1114 (2010).

41.   Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems_Sup. *Science* **339,** 819–23 (2013).

42.   Anders, S. & Huber, W. Differential expression of RNA-Seq data at the gene level–the DESeq package. *EMBL, Heidelberg, Ger.* (2012). at <http://gga01.med.wayne.edu/online_help/help_regionminer/DESeq.pdf>

Chapter 3
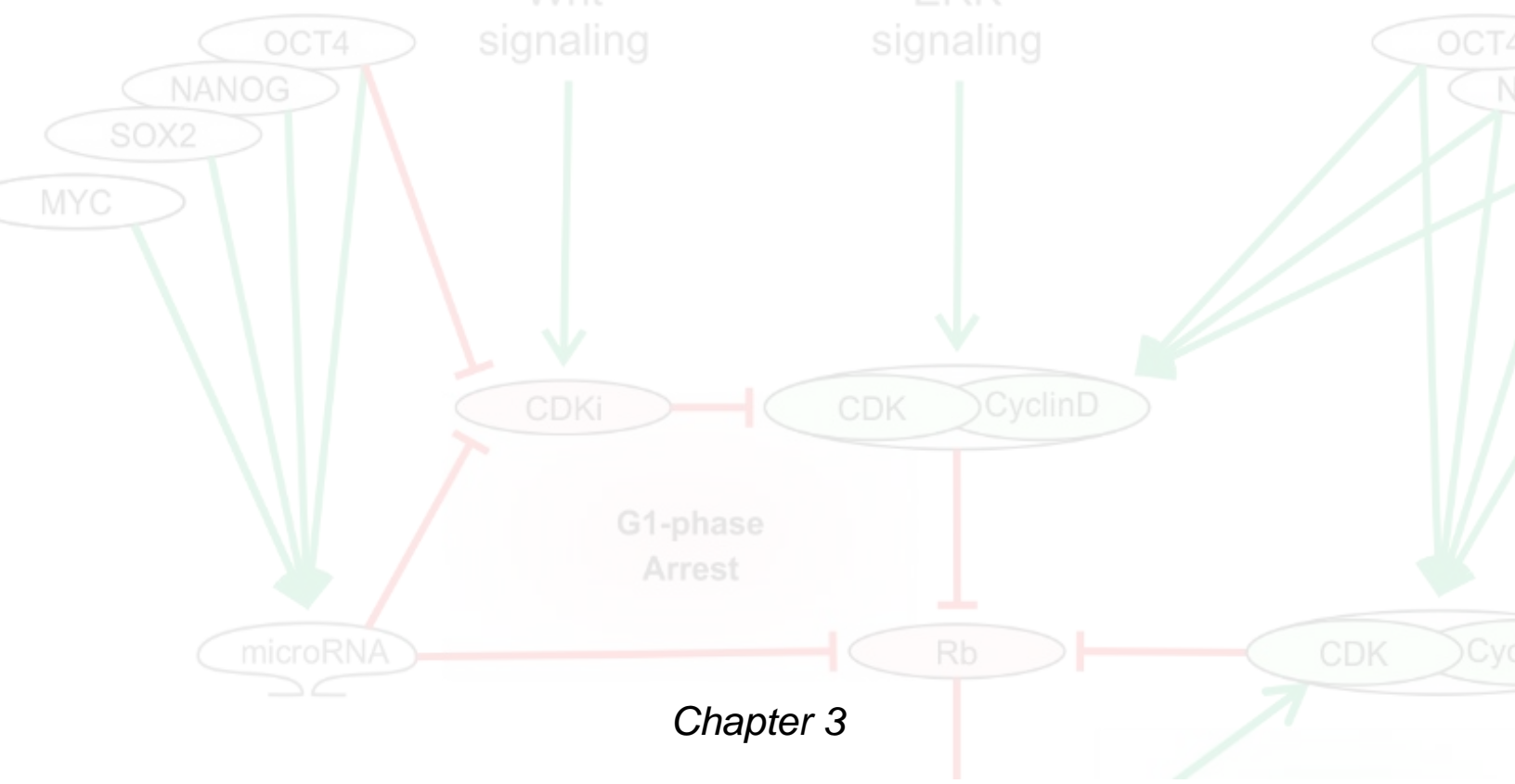
The pluripotency network and the cell cycle

Menno ter Huurne[1], Hendrik G. Stunnenberg[1]

[1] Radboud University, Faculty of Science, Department of Molecular Biology, 6525GA Nijmegen, the Netherlands

Manuscript under preparation

Introduction

Embryonic Stem Cells (ESCs) are subject of intense research. These cells derived from the Inner Cell Mass (ICM) of the early blastocyst can be propagated indefinitely in vitro while retaining the transcriptional and epigenetic properties of their in vivo counterparts. As a result these in vitro cultured cells maintain the ability differentiate into all cells of the fully developed body plan (termed "pluripotent"). Unlimited in vitro expansion allows scaling necessary for molecular analysis and has facilitated unraveling the molecular mechanisms that dictate maintenance of pluripotency and differentiation. The transcriptional network that ensures the pluripotent phenotype of Embryonic Stem Cells is at the core controlled by three Transcription Factors (TFs), OCT4, SOX2 and NANOG. These TFs have overlapping genomic binding sites and uphold an transcriptional network that favors the expression of pluripotency genes and inhibits lineage specifying genes[1]. Cell type specification on the other hand is driven by lineage specific signaling pathways and TFs that regulate epigenetic modifications that alter the cells' genomic structure and transcriptional program[2,3]. This knowledge has allowed scientists to recapitulate these processes in vitro and to develop a wide range of protocols that can be used to create specific cell types and even fully developed and functional organs in vitro[4]. These methods not only hold great promise to replace lost tissue but has also opened new exciting opportunities to study development and disease[5].

The identification of pathways that initiate differentiation has also resulted in distinct culture methods that allow self-renew and maintenance of pluripotency. Up to almost one decade ago mouse ESCs were cultured in conditions that included serum components or Bone Morphogenetic Proteins (BMPs) to prevent differentiation. The realization that the BMP acts through the inhibition of Fibroblast Growth Factor (FGF) / Extracellular signal-Regulated Kinase (ERK) signaling paved the way for the development of a serum-free condition that includes two small molecule inhibitors (2i) to support in vitro propagation of pluripotent mouse ESCs[6]. Transcriptional and epigenetic profiling implied that serum ESC populations are naive pluripotent cells (hereafter referred to as "naive ESCs") but rather heterogeneous in terms of morphology and the expression of core pluripotency genes. ESCs grown in 2i conditions (hereafter referred to as "ground state ESCs") on the other hand are more homogeneous and are considered to "occupy a ground state in which the pluripotency gene regulatory circuitry is maximally operative" (reviewed by Ying, Q.-L. & Smith, A.[7]).

During embryonic development lineage specification is paralleled by a steady increase in cell number. In between mitotic cell division proliferating cells pass consecutively through Gap1 (G1)-phase, Synthesis (S)-phase and G2-phase. Pluripotent cells from the ICM are characterized by extremely short Gap-phases that lengthen upon lineage specification during time of implantation (reviewed by Boward, B., Wu, T. & Dalton, S.[8]). Studies on the 3D chromatin conformation revealed that cell type-specific chromatin configurations are detected most prominently during G1-phase and that genomic interactions that drive gene expression, like enhancer-promoter interactions, are taking place during G1-phase[9]. Similarly, in ESCs the expression of lineage specifiers as well as the establishment of bivalent domains at the promoters of developmental genes occurs specifically during G1-phase and are accompanied by chromatin reorganization and enhancer-promoter interactions[10,11]. These findings have led scientists to believe that the G1-phase serves

as an window of opportunity for ESCs to initiate differentiation (reviewed by Dalton, S.[12]). In line with this hypothesis several studies have indicated that mouse and human ESCs are more sensitive to differentiation-inducing conditions in G1-phase when compared to either S- or G2-phase and that cells residing in G1-phase show reduced colony forming capacities[13,14].

The G1-phase of cells both from the ICM as well as from in vitro cultured naive mouse ESCs is extremely short. If differentiation associated with chromatin remodeling is indeed specifically initiated during G1-phase this characteristic short G1-phase of ESCs could potentially serve as a barrier to prevent differentiation. Although elongation of G1-phase correlates well with differentiation and delaying S-phase entry has been shown to induce differentiation in some cases[15,16] fast progression through the G1-phase is not a prerequisite for ESCs to maintain pluripotent. The most compelling evidence for this is in vivo diapause. During a period that in mice can take up to 20 days, transcriptional activity of the pluripotent cells in blastocyst is severely reduced and there is minimal cell division[17]. In vitro diapause can be recapitulated by inhibiting either MYC or the mTOR pathway[18,19]. Although the former stalls ground state ESCs in G1-phase it does not affect their pluripotent potential[18]. In line with these observations changes in the distribution of cells over the different phases of the cell cycle upon adaptation of ESCs to ground state conditions imply that the extremely short G1-phase is not an intrinsic property of ESCs but is at least in part the result of abundant external stimuli[20,21]. In spite of these differences in G1-phase, cells in both states of pluripotency proliferate at roughly the same pace implying that rapid proliferation is a characteristic feature of ESCs. What causes these cells to proliferate at such a rapid pace is still not entirely clear.

Accumulating evidence indicates that signaling pathways and transcriptions factors of the core pluripotency network impact on the cell cycle. In this communication we will review how the pluripotency network impacts on the cell cycle, mainly focusing on OCT4, SOX2 and NANOG. Most of the studies on cell cycle regulation in ESCs so far have been performed however using naive ESCs. Although the number of reports on cell cycle control in ground state ESCs is still limited they have revealed some striking differences regarding the cell cycle when compared to naive ESCs. Throughout the review we will also discuss these differences and elaborate on their implications. Finally, we will discuss how cell cycle regulators modulate pluripotency and differentiation.

3

Cell Cycle Control in Naive and Ground State ESCs

Characteristic of classically cultured naive ESCs is that they lack the control mechanisms that prevent S-phase entry. The canonical pathway that regulates G1-phase progression culminates in E2F activity that drives progression into S-phase. The E2F transcription factors are inhibited by the pocket protein family - consisting of RB, P107 and P130- that in turn are inhibited by CDK/cyclin-mediated phosphorylation. Two families of CDK-inhibitors exist, the CIP/KIP and the INK4/ARF that can inhibit CDK/cyclin complex formation. One of the first observations that explained the characteristic fast ESC cell cycle was the absence of hypo-phosphorylated RB in naive ESCs[22]. Subsequent studies revealed that the RB pathway

is compromised in these cells and not activated in growth-inhibitory conditions (reviewed by Burdon, T.[23]). Several mechanisms might act in concert to prevent the expression of hypo-phosphorylated RB.

One possible explanation is that in naive ESCs not only the relative amount of hypo-phosphorylated active RB but also the total RB protein level is low in naive ESCs compared to MEF[22]. Low RB levels imply that minimal CDK/cyclin activity suffices to induce S-phase entry and could explain rapid proliferation despite the low expression of CDKs in ESCs[24]. The expression of the pocket proteins is inhibited by the expression of an ESC-specific set of microRNAs[25]. Besides a low basal pocket protein expression level, the short G1-phase and the inability of ESCs to arrest in G1-phase upon stress has also been attributed to precocious CDK/Cyclin activity resulting in hyper-phosphorylation and inactivation of RB. In contrast to somatic cells ESCs display a non-cyclic expression of the cyclins and early studies have indicated that fast progression through G1-phase in naive ESCs is mediated by constitutive CDK/cyclin expression and the absence of CDK-inhibitors[26]. Naive ESCs in general do not express these molecules and in part seem insensitive to ectopic expression[24,27,28]. It is not entirely clear why CDKi's are not expressed in naive ESCs but the activity of microRNAs does contributes by repressing the expression of the CIP/KIP family of CDK-inhibitors[29]. Although this characteristic ESC-specific cell cycle is fundamentally different from the somatic cell cycle the underlying molecular mechanism remained elusive for more than three decades. Below we will discuss how certain signaling pathways and TFs of the core pluripotency network impact on the cell cycle of naive and ground state ESCs.

FGF/ERK and WNT signaling

The FGF/ERK-pathway is involved in primitive endoderm formation and germ line specification in the early embryo and its inhibition allows to maintain ESCs in ground state pluripotency[6,30]. Activation of the ERK pathway results in elevated activity of CDK/cyclin complexes and drives progression through G1-phase (reviewed by Meloche, S. & Pouysségur, J.[31]). We and others have recently shown that ERK-signaling plays a major role phosphorylation of RB and S-phase entry in naive mouse ESCs[32,33]. How ERK exactly controls phosphorylation is not entirely clear although it is likely that it involves the expression of CYCLIN D[32,34]. In ground state ESCs FGF/ERK signaling is inhibited leading to lowered expression of Cyclin D and more cells in G1-phase[21]. The fact that inhibition of ERK signaling in naive ESCs did not fully restore the G1-checkpoint as present in ground state ESCs suggests that other processes contribute to the shortening of G1-phase in naive ESCs[32].

Besides ERK signaling, the activity of cell cycle regulators is also affected by Wnt signaling which is a crucial mediator of pluripotency[35]. Positive regulation of both Cyclin D and Cyclin E is mediated via direct transcriptional control by β-catenin as well as through inhibition of GSK3, which targets both cyclins for degradation. Wnt signaling thereby is an important positive regulator of G1- to S-phase progression in somatic cells (reviewed by Niehrs, C. & Acebron, S. P.[36]). In contract to somatic cells, active Wnt signaling has recently been shown to reduce the speed with which ground ESCs progress through G1-phase[37]. Upon active Wnt signaling one of the two major downstream TFs, TCF1, was recruited to the INK4/Arf tumor

suppressor locus and induced the expression of the CDKi's P16 and P19. Corollary S-phase entry was delayed and proliferation of naive ESCs was slowed down without affecting their undifferentiated state. The differences in cell cycle between naive and ground state ESCs are therefor at least in part the result of lowered CYCLIN D and elevated CDKi expression mediated by FGF/ERK inhibition and stimulation of Wnt signaling, respectively. The addition of the 2i inhibitors to naive ESCs culturing in serum had however a less pronounced effect on the cell cycle when compared to adaptation to ground state serum-free conditions[21]. In ground state ESCs two members of the CIP/KIP family of CDKi's, P21 and P27, are higher expressed when compared to naive ESCs. What causes the higher expression of these proteins is not been determined yet. Several ESC-specific mechanisms have shown to contribute to the shortened G1-phase in naive ESCs. In the next section we will consider these mechanisms and elaborate on their influence on the cell cycle of naive and ground state ESCs.

OCT4/SOX2/NANOG

Several lines of evidence have indicated that members of the pluripotency network employ multiple mechanisms to modulate progression through the G1-phase. Firstly, both OCT4 and SOX2 directly control the expression of ESC-specific microRNAs that inhibit key regulators of G1-phase progression and thereby contribute to the abbreviated G1-phase[25,38]. These microRNAs repress the expression of the pocket proteins as well as CDKi's and desensitize ESCs to serum starvation[39]. Whether a similar mechanism is employed by OCT4 and SOX2 to repress the expression of the pocket proteins in ground state ESCs remain to be determined. However, a recent study has indicated that self-renewal of ground state ESCs has a higher dependence on microRNA than naive ESCs, which could be explained by the fact that ground state but not naive ESCs express both CDKi's as well hypo-phosphorylated RB[40].

Secondly, CDK/Cyclin-mediated phosphorylation of RB is promoted by both SOX2 and NANOG. SOX2 can stimulate cell proliferation either directly as a transcriptional activator of several cyclins[41–43] or as a repressor of P21[44,45]. NANOG contributes to accelerated S-phase entry through the transcriptional activation of CDK6 and CDC25A[46,47]. Similar to SOX2, NANOG is able to suppress the expression of CDKi's, NANOG binds upstream of the Cdkn1b gene and its expression correlates with repression of P27[48]. Interestingly, two independent OCT4-driven pathways mediate RB hyper-phosphorylation and G1-phase progression in naive ESCs[49,50]. Altogether, these results imply that the pluripotency network contributes to the phosphorylation of the pocket proteins and abrogation of the G1/S checkpoint. Despite the fact that OCT4, SOX2 and NANOG are expressed at similar levels in naive and ground state ESCs the levels of CDKi's were significantly increased and no hyper-phosphorylated RB was observed during the G1-phase of ground state ESCs, suggesting that the core pluripotency TFs do not facilitate G1-phase progression in ground state ESCs. Knock down studies on these TFs combined with a more comprehensive analysis of the expression and phosphorylation levels of the pocket proteins in ground state ESCs are needed to determine whether the core pluripotency TFs do contribute to G1-phase progression in ground state ESCs.

3

MYC

Although MYC is not considered to be part of the core pluripotency network, the MYC family of TFs has a well-established role in stem cell maintenance as well as cell cycle regulation (reviewed by Singh, A. & Dalton, S.[51]). MYC is a direct transcriptional regulator of both CYCLIN D and E and the MYC-mediated activation of CDK2/Cyclin E complexes has shown to overcome the restriction point in naive ESCs[52]. Although MYC is only lowly expressed in ground state ESCs when compared to naive ESCs no major differences in CYCLIN D and E expression were observed[6,21]. MYC however also antagonizes the expression of P21 and the decreased expression of MYC might therefor (next to inhibition of ERK and Wnt activation) contribute to the reinstatement of G1-phase control in ground state ESCs[21,53]. Interestingly, MYC is necessary for ground state ESCs to proliferate and its loss results in G1-phase arrest in ground state ESCs but not in naive ESCs ([18] and unpublished observations). Although MYC is well known for its role as a transcriptional activator of G1-cyclins the absence of hyper-phoshorylated pocket proteins in ground state ESCs during late G1 phase suggests that MYC-mediated S-phase entry is not CDK/cyclin-dependent. Previous studies have shown that MYC also mediates the expression of a microRNA cluster that targets the RB family protein P130 and thereby contributes positively to cell cycle progression[54]. It would therefor be interesting to assess the expression of hypo-phosphorylated pocket proteins in ground state ESCs upon MYC-deprivation. Such studies could provide valuable insights into the role of MYC in the ESC cell cycle. The fact that inhibition of MYC does not result in G1-arrest in naive ESCs confirms that not MYC but CDK/cyclin-driven activation of E2Fs initiates S-phase entry in naive ESCs[6,32].
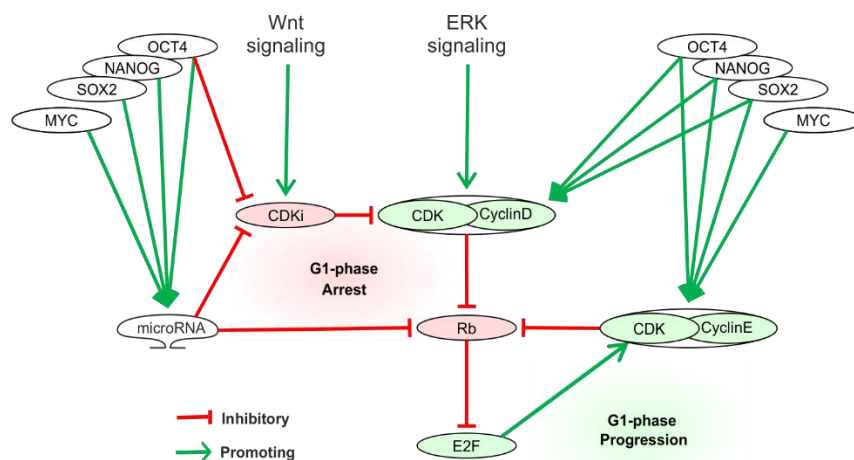

Figure 1. Schematic of the mechanistic link between the pluripotency and the cell cycle regulatory network. Members of the core pluripotency network and MYC promote G1-phase progression by the activation of CDK/Cyclin complexes. In addition, transcriptional activation of microRNAs results in silencing of CDK-inhibitors and members of the pocket proteins that delay S-phase entry.

Together, these results suggest that the abrogated G1-phase in naive ESCs is the result of high FGF/ERK activity and low Wnt signaling. Activation of Wnt signaling leads to transcription of the INK4/ARF locus and P16/P19-mediated elongation of G1-phase. Although inhibition of FGF/ERK induces the expression of P21 and P27 and reinstates the RB/E2F-mediated G1-checkpoint its exact molecular mechanism has not been deciphered yet. Possibly the lowered MYC expression in ground state compared to naive ESCs contributes to reinstatement of the G1-checkpoint.

Effect of Cell Cycle Regulators on Pluripotency

The discovery that somatic cells can be reprogrammed to induced pluripotent Stem (iPS) cells inspired many to combine this protocol with genome wide knock down or overexpression screens to identify genes that modulate the pluripotency network. Amongst others many cell cycle regulators were frequently identified to improve reprogramming[55,56]. Although reprogramming is a stochastic process that is highly susceptible to changes in cell proliferation rates[57], several studies have indicated that there are direct mechanistic link between certain cell cycle regulators and the pluripotency network (reviewed by White, J. & Dalton, S.[58]). In the following section we will review recent work that has uncovered a direct mechanistic links between the proteins that drive G1-phase progression and the pluripotency network.

CYCLINS

Although there is a general consensus that the main function of CYCLIN D and E is to drive progression through G1-phase, a recent study revealed that the cell cycle of ground state mouse ESCs does not depend on either one of them. Surprisingly, the combined depletion of Cyclin E1-2 and D1-3 does neither abrogate G1-phase progression nor successful S-phase entry in ground state ESCs[59]. Interestingly though, these cyclins elicit an important role in maintenance of pluripotency of ground state mouse ESCs through the stabilization of proteins of the core pluripotency network, NANOG, OCT4 and SOX2. ESCs that lack both Cyclin D and Cyclin E show increased expression of Cdx2 and Eomes, two transcription factors that confer trophectodermal differentiation[59]. The ability of G1-cyclins to prevent differentiation had been observed before, although employing a different mechanism. Cyclin D1-3 and associated CDKs contribute to maintenance of the undifferentiated state of ESCs by blocking the TGF-β signaling pathway that induces endoderm differentiation[60]. In addition, Cyclin D1 acts not only in partnership with CDKs but is also able to directly repress the expression of endoderm genes by binding and recruitment of transcriptional repressors to their promoters[61]. Together with the fact that endoderm differentiation correlates with loss of CYCLIN D1 expression these results suggest that CYCLIN D and E are essential for upholding the core pluripotency network in epiblast cells of the ICM whereas trophectoderm cells and cells destined to contribute to the primitive endoderm lose the expression of these cyclins. In line with these results, CYCLIN E not only regulates G1 progression but impairs the differentiation towards extra-embryonic lineages as well[62]. Similarly, in the early blastocyst CYCLIN E colocalizes with NANOG in the epiblast whereas it is downregulated in trophectoderm cells that proliferate at a much slower pace[8,63]. The fact that CYCLIN D and E knock out ESCs retain the ability to differentiate to either of the three major germ layers indicating that G1-cyclins are critical for maintenance of pluripotency but not required for germ layer specification[59].

CDKs

Not only the cyclins but also their catalytically active partners have been functionally linked to regulation of gene expression in embryonic stem cells and during lineage specification. Similar to CYCLIN D and E the CDKs that promote progression through the G1-phase, CDK2 and CDK4, are dispensable during embryonic

3

development until midgestation indicating that they are not required for the cell cycle of undifferentiated cells of the ICM. The loss of CDK1 on the other hand abrogates the first cell division upon conception and is considered to be the only CDK essential for undifferentiated cells to proliferate (reviewed in Malumbres, M. & Barbacid, M.[64]). The fact that CDK1 expression parallels the expression of pluripotency genes and that downregulation results in differentiation underpins its importance in pluripotency[65]. CDK1 is not only essential for proliferation of ICM cells but can also interact with OCT4 and thereby promote the repression of specifiers of the trophectoderm lineage[66,67]. Moreover, CDK1 is involved in a cell cycle-independent pathway that suppresses ERK signaling-mediated differentiation[65]. Together these findings suggests that CDK1 plays a crucial role in the early blastocyst that resembles the role of CYCLIN D and E. Besides CDK1 an important role in maintaining pluripotency has been attributed to CDK2. Although CDK2 is dispensable in ICM cells CDK2-mediated phosphorylation of SOX2 enhances reprogramming MEFs[68]. Moreover, the de-repression of CDK2 prevents differentiation upon LIF withdrawal[69]. A recent report identified not only SOX2 but also NANOG and OCT4 as targets for CDK2-mediated phosphorylation[59]. The CDK2-mediated phosphorylation of these core pluripotency factors inhibits ubiquitination and subsequent degradation and appears to be essential for maintenance of pluripotency. CDK-mediated phosphorylation of SOX2 also promotes the ability of a truncated form of SOX2 to negatively regulate neuroectodermal differentiation[70]. A recent study has uncovered how CDK4 and CDK6 act in concert with D cyclins to prevent TGF-β signaling-driven endoderm differentiation through phosphorylation of the SMAD proteins[60]. On the contrary, the CDK4-6/CYCLIN D activity promotes neuroectoderm differentiation. Besides the core pluripotency factors also the activity of MYC is influenced by CDK activity. CDK8 phosphorylates and protects MYC from degradation. By doing so CDK8 enhances MYC target gene expression and keeps ESCs in their pluripotent state[71]. CDK9, a less well characterized member of the CDK family binds KLF4 and contributes to Polymerase II release at the promoters of genes belonging to the pluripotency network[72]. Altogether these results indicate that the high activity of CDKs not only results in rapid cell cycle progression in ESCs, but also contributes to a transcriptional program that favors maintenance of pluripotency. However, the facts that only CDK1 is essential in the blastocyst, that most CDKs are expressed upon differentiation and that CDK-mediated modulation of the pluripotency network merely inhibits the differentiation into specific lineages suggests that CDKs are involved in lineage specification.

CDK-inhibitors

Considering the role of CDKs in ESCs it seems evident that their naturally occurring counterparts have the opposite function. This assumption is underscored by a large body of literature that indicates that CDKi's are upregulated upon differentiation (reviewed by Sage, J.[73]). Apart from blocking CDK/Cyclin interactions several CDKi's can both repress and activate gene expression by directly interacting with TFs (reviewed by Abbas, T. & Dutta, A.[74]). The convincing evidence that CDKi's can negative influence the pluripotency network comes from two studies on the role of the CIP/KIP proteins. P21 is able to repress Sox2 expression by directly binding to a Sox2 enhancer[75] and P27 in complex with amongst others P130 is able to

epigenetically silence the *Sox2* gene resulting in differentiation of ESCs[76–78]. Presumably, P21 and P27 bind the same repressive complex and prevent the expression of target genes during G1-phase.

Although these results imply that the expression of CDKi's correlates with differentiation, two recent reports show that the picture is more complex and context dependent. Active Wnt signaling via TCF1 controls the expression of the Cdkn1a tumor suppressor locus, that encodes the cell cycle inhibitors P15, P16 and P19 resulting in decreased cell proliferation while leaving the expression of pluripotency genes unchanged[37]. In a recent report we have shown that naive ground state ESCs also express active P21 and P27 resulting in an elongated G1 phase. Despite the higher levels of P16, P21 and P27 protein, no difference in the expression of Sox2 nor other core pluripotency genes was observed[21,79].

RB/E2F

The general notion that arises from studies on the role of RB family proteins during embryonic development and in ESCs is that these proteins mainly play a role in differentiation. Mouse embryos lacking either Rb or both P107 and P130 die during peri-implantation stages and display hyper-proliferation and deregulated differentiation. Conform these phenotypes, in vitro cultured ESCs RB$^{-/-}$ cells show diminished differentiation potential (reviewed by Julian, L. M. & Blais, A.[80]). These results might be due to the fact that RB is a transcriptional repressor of Sox2 and Oct4 in somatic cells [78]. In mouse ESCs the pocket proteins are not functional and, hence, the ablation has no effect on their proliferative capacities or potential to form alkaline[+] colonies[81]. On the other hand ground state ESCs do have a functional RB/E2F axis and the loss of RB, P107 and P130 in these cells therefore results in premature S-phase entry[21]. These results imply that the activity of pocket proteins might not only control G1-phase progression but negatively affect the pluripotency network in ground state ESCs as well. Despite the differential activity of these pocket proteins in naive and ground state ESCs these cells do not display major differences in expression of core pluripotency factors[6,79]. Furthermore, when comparing WT ESCs and ESCs lacking the pocket proteins no major differences in the expression of Oct4 were observed either[21]. It remains to be determined what causes these seemingly contradictory findings. Possibly one of the other proteins of the repressive complex is not available in ground state ESCs[76,78]. Whether the core pluripotency network of ground state ESCs is de-sensitized to Rb-mediated repression or that depletion of the pocket proteins could even amplify the pluripotency network remains elusive. Like the pocket proteins the transcription of Sox2 gene can also be regulated by E2F transcription factors. Two closely related E2F proteins, E2F 3A and 3B, have opposing effects on the expression of Sox2. Both transcription factors bind within close distance of the Sox2 Transcription Start Site (TSS) but E2F3B results in lowered deposition of H3K27me3 whereas replacement of E2F3B with E2F3A resulted in the recruitment of a repressor complex[82]. How they exactly contribute to the regulation of the pluripotency network is however unclear. The combined deletion of all three "activator" E2Fs, E2F1-3, did not abolish self-renewal of naive ESCs nor efficient teratoma-formation[83]. Only a few specific tissues showed impaired proliferation and increased apoptosis suggesting that they might contribute to

3

differentiation. This was confirmed by the finding that none of studied E2F family members is essential for cells of the ICM[84].

P53

If cells suffer from stressful conditions progression through the cell cycle can be temporarily halted during G1-phase. The pathways that are responsible for G1-arrest not only act in response to stress but have also been shown to mediate developmentally programmed cell senescence[85]. An important regulator of G1-phase progression is P53, an tumor suppressor and master regulator of stem cell quiescence[86]. Although highly expressed and active at early embryonic stages[87,88] P53[-/-] mice are viable[89] and a debate on the role of P53 in embryonic development is still ongoing[90]. A large number of studies have indicated that P53 in ESCs is transcriptionally inactive under normal circumstances but can mediate differentiation under stressful conditions by both inducing the expression of lineage specifiers and downregulation of the core pluripotency network (reviewed by Danilova, N.[91]). Upon DNA damage for example, P53 binds to the Nanog promoter and reduces the expression of NANOG[92]. Moreover, it can repress OCT4, NANOG and SOX2 by interfering with distal enhancer activity[93]. Moreover, microRNAs induced by P53 target OCT4 and SOX2[94,95] and facilitate extraembryonic endoderm specification[96]. Recently two studies have highlighted the role of P53 in mesendoderm differentiation. P53 inhibits a transcriptional network that promotes pluripotency through the expression of long non-coding RNAs resulting in differentiation into the mesoderm and endoderm[97]. These results were confirmed by the finding that the combined deletion of P53, P63 and P73 resulted in abrogated mesendodermal differentiation. The P53 family tweaks the collaborative WNT/TGF-β pathway by inducing the expression of several WNT ligands[98].

In contrast to the above mentioned literature, a few studies have shown that P53 can also inhibit lineage commitment and contribute to an undifferentiated phenotype. Stress-induced P53 activity was shown to mediate the transcription of Wnt ligands that contribute to an undifferentiated phenotype[99]. Moreover, P53 can elevate TGF-β signaling and concomitantly NANOG expression prevent exit of the pluripotent state by elevating TGF-beta signaling and accompanying NANOG expression[100].

Concluding remarks

Early studies on ESC derived from the inner cell mass revealed an unusual short cell cycle structure that was subject to dramatic changes during cell fate specification[23,101]. The cell cycle regulated differentiation and accompanied chromatin remodeling led scientist to believe that the ESC-specific cell cycle structure and differentiation are mechanistically linked. Whether the ESC-specific transcriptional program that coordinates pluripotency also facilitates their highly proliferative phenotype and characteristic short G1-phase has been subject of debate[58]. In this review we have summarized how G1-phase progression is affected by signaling pathways that dictate early embryonic development and the core pluripotency TF machinery plus MYC. In addition, we have tried to rationalize to what extent these biological processes

explain the cell cycle structure of ESCs in two distinct states of pluripotency[6,102]. Although ESCs in both states are highly proliferative, signaling pathways that ensure maintenance of pluripotency do not mediate uncontrolled G1-phase progression, indicating that the relative short G1-phase is not a feature of pluripotent ESCs[21,37]. In line with these observation, the presence of hypo- but not hyper-phosphorylated pocket proteins during G1-phase in ground state ESCs suggest that pluripotency factor-driven CDK/Cyclin activity is not characteristic of ground state ESCs and therefore not an intrinsic feature of pluripotent ESCs. Together these results imply that fast G1-phase progression and cell proliferation is not driven by pluripotency but is the default cell cycle, whereas cell cycle remodeling, like quiescence, is driven by either situations of stress or differentiation. Furthermore, we have discussed how members of the canonical G1-restriction point modulate pluripotency and differentiation. In general, pro-proliferative members, like cyclins and CDKs, have a positive effect on maintenance of pluripotency. Conversely, the activity of members that delay G1-phase progression correlates with differentiation[64]. Many members that promote proliferation and pluripotency are however dispensable in the early ICM and rather prevent differentiation than promote pluripotency[59,60]. The latter can be explained by the fact that the initial phase of tissue development not only requires proper cell differentiation but sufficient cell numbers as well.

References

1. Chambers, I. & Tomlinson, S. R. The transcriptional foundation of pluripotency. *Development* **136,** 2311–2322 (2009).

2. Gifford, C. A. *et al.* Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* **153,** 1149–1163 (2013).

3. Atlasi, Y. & Stunnenberg, H. G. The interplay of epigenetic marks and development. *Nat. Publ. Gr.* (2017). doi:10.1038/nrg.2017.57

4. Wu, J. *et al.* Stem cells and interspecies chimaeras. *Nature* **540,** 51–59 (2016).

5. Bartfeld, S. & Clevers, H. Stem cell-derived organoids and their application for medical research and patient treatment. *J. Mol. Med.* **95,** 729–738 (2017).

6. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–23 (2008).

7. Ying, Q.-L. & Smith, A. The Art of Capturing Pluripotency: Creating the Right Culture. *Stem Cell Reports* **8,** 1457–1464 (2017).

8. Boward, B., Wu, T. & Dalton, S. Control of cell fate through cell cycle and pluripotency networks. *Stem Cells* n/a-n/a (2016). doi:10.1002/stem.2345

9. Dekker, J. Two ways to fold the genome during the cell cycle: Insights obtained with chromosome conformation capture. *Epigenetics and Chromatin* **7,** 1–12 (2014).

10. Singh, A. *et al.* Cell-Cycle Control of Bivalent Epigenetic Domains Regulates the Exit from Pluripotency. *Stem Cell Reports* **5,** 323–336 (2015).

11. Singh, A. *et al.* Cell-cycle control of developmentally regulated transcription factors accounts for heterogeneity in human pluripotent cells. *Stem cell reports* **1,** 532–44 (2013).

12. Dalton, S. Linking the Cell Cycle to Cell Fate Decisions. *Trends Cell Biol.* **25,** 592–600 (2015).

13. Pauklin, S. & Vallier, L. The cell-cycle state of stem cells determines cell fate propensity. *Cell* **155,** 135–47 (2013).

3

14.   Calder, A. *et al.* Lengthened G1 phase indicates differentiation status in human embryonic stem cells. *Stem Cells Dev.* **22,** 279–95 (2013).

15.   Neganova, I., Zhang, X., Atkinson, S. & Lako, M. 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 (2009).

16.   Ruiz, S. *et al.* A High Proliferation Rate Is Required for Cell Reprogramming and Maintenance of Human Embryonic Stem Cell Identity. *Curr. Biol.* **21,** 45–52 (2011).

17.   Fenelon, J. C., Banerjee, A. & Murphy, B. D. Embryonic diapause: Development on hold. *Int. J. Dev. Biol.* **58,** 163–174 (2014).

18.   Scognamiglio, R. *et al.* Myc Depletion Induces a Pluripotent Dormant State Mimicking Diapause. *Cell* 668–680 (2016). doi:10.1016/j.cell.2015.12.033

19.   Bulut-Karslioglu, A. *et al.* Inhibition of mTOR induces a paused pluripotent state. *Nature* (2016). doi:10.1038/nature20578

20.   He, H. *et al.* p53 and p73 Regulate Apoptosis but Not Cell-Cycle Progression in Mouse Embryonic Stem Cells upon DNA Damage and Differentiation. *Stem Cell Reports* **7,** 1087–1098 (2016).

21.   ter Huurne, M., Chappell, J., Dalton, S. & Stunnenberg, H. G. Distinct Cell-Cycle Control in Two Different States of Mouse Pluripotency. *Cell Stem Cell* **21,** 449–455.e4 (2017).

22.   Savatier, P., Huang, S., Szekely, L., Wiman, K. G. & Samarut, J. Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. *Oncogene* **9,** 809–818 (1994).

23.   Burdon, T., Smith, A. & Savatier, P. Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol.* **12,** 432–8 (2002).

24.   Savatier, Pierre; van Grunsven, L; Rudkin, B; Samarut, J. Withdrawal of differentiation inhibitory activity/leukemia inhibitory factor up-regulates D-type cyclins and cyclin-dependent kinase inhibitors in mouse embryonic stem cells. (1996).

25.   Wang, Y. *et al.* Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat. Genet.* **40,** 1478–1483 (2008).

26.   Stead, E. *et al.* Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. *Oncogene* **21,** 8320–33 (2002).

27.   White, J. *et al.* Developmental Activation of the Rb – E2F Pathway and Establishment of Cell Cycle-regulated Cyclin-dependent Kinase Activity during Embryonic Stem Cell Differentiation. **16,** 2018–2027 (2005).

28.   Faast, R. *et al.* Cdk6-cyclin D3 activity in murine ES cells is resistant to inhibition by p16(INK4a). *Oncogene* **23,** 491–502 (2004).

29.   Li, J. *et al.* MicroRNA-221 is Required for Proliferation of Mouse Embryonic Stem Cells via P57 Targeting. *Stem Cell Rev. Reports* **11,** 39–49 (2015).

30.   Lanner, F. & Rossant, J. The role of FGF/Erk signaling in pluripotent cells. *Development* **137,** 3351–3360 (2010).

31.   Meloche, S. & Pouysségur, J. The ERK1/2 mitogen-activated protein kinase pathway as a master regulator of the G1- to S-phase transition. *Oncogene* **26,** 3227–3239 (2007).

32.   ter Huurne, M. *et al.* Distinct Cell-Cycle Control in Two Different States of Mouse Pluripotency Brief Report Distinct Cell-Cycle Control in Two Different States of Mouse Pluripotency. *Cell Stem Cell* **21,** 449–455.e4 (2017).

33.   Mayor-ruiz, C. *et al.* ERF deletion rescues RAS deficiency in mouse embryonic stem cells. 568–576 (2018). doi:10.1101/gad.310086.117.568

34.   Chambard, J., Lefloch, R., Pouysségur, J. & Lenormand, P. ERK implication in cell cycle regulation. **1773,** 1299–1310 (2007).

35.   ten Berge, D. *et al.* Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nat. Cell Biol.* **13,** 1070–5 (2011).

36.   Niehrs, C. & Acebron, S. P. Mitotic and mitogenic Wnt signalling. *EMBO J.* **31,** 2705–2713 (2012).

37.   De Jaime-Soguero, A. *et al.* Wnt/Tcf1 pathway restricts embryonic stem cell cycle through activation of the Ink4/Arf locus. *PLoS Genet.* **13,** (2017).

38.   Card, D. a G. *et al.* Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells. *Mol. Cell. Biol.* **28,** 6426–38 (2008).

39.   Wang, Y. *et al.* miR-294/miR-302 promotes proliferation, suppresses G1-S restriction point, and inhibits ESC differentiation through separable mechanisms. *Cell Rep.* **4,** 99–109 (2013).

40.   Yan, Y. *et al.* Significant differences of function and expression of microRNAs between ground state and serum-cultured pluripotent stem cells. *J. Genet. Genomics* **44,** 179–189 (2017).

41.   Lin, S. C. *et al.* Epigenetic switch between SOX2 and SOX9 regulates cancer cell plasticity. *Cancer Res.* **76,** 7036–7048 (2016).

42.   Li, C. *et al.* Sex-Determining Region Y-box 2 Promotes Growth of Lung Squamous Cell Carcinoma and Directly Targets Cyclin D1. *DNA Cell Biol.* **36,** 264–272 (2017).

43.   Lee, S. H. *et al.* SOX2 regulates self-renewal and tumorigenicity of stem-like cells of head and neck squamous cell carcinoma. *Br. J. Cancer* **111,** 2122–2130 (2014).

44.   Yamawaki, K. *et al.* Sox2-dependent inhibition of p21 is associated with poor prognosis of endometrial cancer. *Cancer Sci.* **108,** 632–640 (2017).

45.   Fukazawa, T. *et al.* SOX2 suppresses CDKN1A to sustain growth of lung squamous cell carcinoma. *Sci. Rep.* **6,** 1–12 (2016).

46.   Zhang, X. *et al.* A role for NANOG in G1 to S transition in human embryonic stem cells through direct binding of CDK6 and CDC25A. *J. Cell Biol.* **184,** 67–82 (2009).

47.   Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122,** 947–956 (2005).

48.   Münst, B. *et al.* Nanog induces suppression of senescence through downregulation of p27KIP1 expression. *J. Cell Sci.* **129,** 912–20 (2016).

49.   Schoeftner, S., Scarola, M., Comisso, E., Schneider, C. & Benetti, R. An Oct4-pRb axis, controlled by miR-335, integrates stem cell self-renewal and cell cycle control. *Stem Cells* **31,** 717–728 (2013).

50.   Comisso, E. *et al.* OCT4 controls mitotic stability and inactivates the RB tumor suppressor pathway to enhance ovarian cancer aggressiveness. *Oncogene* **36,** 4253–4266 (2017).

51.   Singh, A. & Dalton, S. The Cell Cycle and Myc Intersect with Mechanisms that Regulate Pluripotency and Reprogramming. *Cell Stem Cell* **5,** 141–149 (2009).

52.   Bartek, J. & Lukas, J. Mammalian G1- and S-phase checkpoints in response to DNA damage. *Curr. Opin. Cell Biol.* **13,** 738–747 (2001).

53.   Bretones, G., Delgado, M. D. & León, J. Myc and cell cycle control. *Biochim. Biophys. Acta* **1849,** 506–516 (2015).

54.   Smith, K. N., Singh, A. & Dalton, S. Myc represses primitive endoderm differentiation in pluripotent stem cells. *Cell Stem Cell* **7,** 343–54 (2010).

55.   Li, H. *et al.* The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* **460,** 1136–1139 (2009).

56.   Oh, H., Kim, J. & Kim, J. Critical roles of Cyclin D1 in mouse embryonic fibroblast cell reprogramming. *FEBS J.* **283,** 4549–4568 (2016).

57.   Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462,** 595–601 (2009).

58.   White, J. & Dalton, S. Cell cycle control of embryonic stem cells. *Stem Cell Rev.* **1,** 131–138 (2005).

59.   Liu, L. *et al.* G1 cyclins link proliferation, pluripotency and differentiation of embryonic stem cells. *Nat. Cell Biol.* **19,** (2017).

60.   Pauklin, S. & Vallier, L. The cell-cycle state of stem cells determines cell fate propensity. *Cell* **155,** 135–47 (2013).

61.   Pauklin, S., Madrigal, P., Bertero, A. & Vallier, L. Initiation of stem cell differentiation involves cell cycle-dependent regulation of developmental genes by Cyclin D. *Genes Dev.* **30,** 421–433 (2016).

62.   Coronado, D. A short G1 phase is an intrinsic determinant of naïve embryonic stem cell pluripotency. *Stem Cell Res.* **10,** 118–31 (2013).

3

63.   Krivega, M. V. *et al.* Cyclin E1 plays a key role in balancing between totipotency and differentiation in human embryonic cells. *Mol. Hum. Reprod.* **21,** 942–956 (2015).

64.   Malumbres, M. & Barbacid, M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat. Rev. Cancer* **9,** 153–166 (2009).

65.   Wang, X. Q. *et al.* CDK1-PDK1-PI3K/Akt signaling pathway regulates embryonic and induced pluripotency. *Cell Death Differ.* **24,** 38–48 (2017).

66.   Neganova, I. *et al.* CDK1 plays an important role in the maintenance of pluripotency and genomic stability in human pluripotent stem cells. *Cell Death Dis.* **5,** e1508 (2014).

67.   Li, L. *et al.* Cdk1 interplays with Oct4 to repress differentiation of embryonic stem cells into trophectoderm. *FEBS Lett.* **586,** 4100–4107 (2012).

68.   Ouyang, J. *et al.* Cyclin-dependent kinase-mediated Sox2 phosphorylation enhances the ability of Sox2 to establish the pluripotent state. *J. Biol. Chem.* **290,** 22782–22794 (2015).

69.   Kim, Y. *et al.* Cyclin-dependent kinase 2-associating protein 1 commits murine embryonic stem cell differentiation through retinoblastoma protein regulation. *J. Biol. Chem.* **284,** 23405–14 (2009).

70.   Lim, S. *et al.* Cyclin-Dependent Kinase-Dependent Phosphorylation of Sox2 at Serine 39 Regulates Neurogenesis Shuhui. 1–24 (2017).

71.   Adler, A. S. *et al.* CDK8 maintains tumor dedifferentiation and embryonic stem cell pluripotency. *Cancer Res.* **72,** 2129–2139 (2012).

72.   Liu, L. *et al.* Transcriptional pause release is a rate-limiting step for somatic cell reprogramming. *Cell Stem Cell* **15,** 574–588 (2014).

73.   Sage, J. The retinoblastoma tumor suppressor and stem cell biology. *Genes Dev.* 1409–1420 (2012). doi:10.1101/gad.193730.112.control

74.   Abbas, T. & Dutta, A. P21 in Cancer: Intricate Networks and Multiple Activities. *Nat. Rev. Cancer* **9,** 400–14 (2009).

75.   Marqués-Torrejón, M. Á. *et al.* Cyclin-dependent kinase inhibitor p21 controls adult neural stem cell expansion by regulating Sox2 gene expression. *Cell Stem Cell* **12,** 88–100 (2013).

76.   Li, H. *et al.* p27(Kip1) directly represses Sox2 during embryonic stem cell differentiation. *Cell Stem Cell* **11,** 845–52 (2012).

77.   Yamamizu, K., Schlessinger, D. & Ko, M. S. H. SOX9 accelerates ESC differentiation to three germ layer lineages by repressing SOX2 expression through P21 (WAF1/CIP1). *Development* **141,** 4254–4266 (2014).

78.   Kareta, M. S. *et al.* Inhibition of pluripotency networks by the Rb tumor suppressor restricts reprogramming and tumorigenesis. *Cell Stem Cell* **16,** 39–50 (2015).

79.   Marks, H. *et al.* The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. *Cell* **149,** 590–604 (2012).

80.   Julian, L. M. & Blais, A. Transcriptional control of stem cell fate by E2Fs and pocket proteins. *Front. Genet.* **6,** 1–15 (2015).

81.   Conklin, J. F., Baker, J. & Sage, J. The RB family is required for the self-renewal and survival of human embryonic stem cells. *Nat. Commun.* **3,** 1244 (2012).

82.   Julian, L. M. *et al.* Opposing Regulation of Sox2 by Cell-Cycle Effectors E2f3a and E2f3b in Neural Stem Cells. *Cell Stem Cell* **12,** 440–452 (2013).

83.   Chong, J.-L. *et al.* E2f1-3 switch from activators in progenitor cells to repressors in differentiating cells. *Nature* **462,** 930–4 (2009).

84.   Ouseph, M. M. *et al.* Atypical E2F Repressors and Activators Coordinate Placental Development. *Dev. Cell* **22,** 849–862 (2012).

85.   Muñoz-Espín, D. *et al.* Programmed cell senescence during mammalian embryonic development. *Cell* **155,** 1104–18 (2013).

86.   Liu, Y. *et al.* p53 Regulates Hematopoietic Stem Cell Quiescence. *Cell Stem Cell* **4,** 37–48 (2009).

87.   Goh, a. M. *et al.* Using targeted transgenic reporter mice to study promoter-specific p53 transcriptional activity. *Proc. Natl. Acad. Sci.* **109,** 1685–1690 (2012).

88.   Schmid, P., Lorenz, A., Hameister, H. & Montenarh, M. Expression of p53 during mouse embryogenesis. *Development* **113,** 857–65 (1991).

89.   Donehower, L. A. *et al.* Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356,** 215–221 (1992).

90.   Danilova, N., Sakamoto, K. M. & Lin, S. P53 Family in Development. *Mech. Dev.* **125,** 919–931 (2008).

91.   Levine, A. J. & Berger, S. L. The interplay between epigenetic changes and the p53 protein in stem cells. *Genes Dev.* **31,** 1195–1201 (2017).

92.   Lin, T. *et al.* p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. *Nat Cell Biol* **7,** 165–171 (2005).

93.   Li, M. *et al.* Distinct regulatory mechanisms and functions for p53-activated and p53-repressed DNA damage response genes in embryonic stem cells. *Mol. Cell* **46,** 30–42 (2012).

94.   Jain, A. K. *et al.* p53 regulates cell cycle and microRNAs to promote differentiation of human embryonic stem cells. *PLoS Biol.* **10,** e1001268 (2012).

95.   Singh, S. K. *et al.* Antithetical NFATc1-Sox2 and p53-miR200 signaling networks govern pancreatic cancer cell plasticity. *EMBO J.* **34,** 517–530 (2015).

96.   Li, C., Finkelstein, D. & Sherr, C. J. Arf tumor suppressor and miR-205 regulate cell adhesion and formation of extraembryonic endoderm from pluripotent stem cells. (2013). doi:10.1073/pnas.1302184110

97.   Jain, A. K. *et al.* LncPRESS1 Is a p53-Regulated LncRNA that Safeguards Pluripotency by Disrupting SIRT6-Mediated De-acetylation of Histone H3K56. *Mol. Cell* **64,** 967–981 (2016).

98.   Wang, Q. *et al.* The p53 Family Coordinates Wnt and Nodal Inputs in Mesendodermal Differentiation of Embryonic Stem Cells. *Cell Stem Cell* 1–17 (2016). doi:10.1016/j.stem.2016.10.002

99.   Lee, K.-H. *et al.* A genomewide study identifies the Wnt signaling pathway as a major target of p53 in murine embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 69–74 (2010).

100.   Gonzales, K. A. U. *et al.* Deterministic Restriction on Pluripotent State Dissolution by Cell-Cycle Pathways. *Cell* **162,** 564–579 (2015).

101.   Filipczyk, A. a, Laslett, A. L., Mummery, C. & Pera, M. F. Differentiation is coupled to changes in the cell cycle regulatory apparatus of human embryonic stem cells. *Stem Cell Res.* **1,** 45–60 (2007).

102.   Smith, A. G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336,** 688–690 (1988).

3
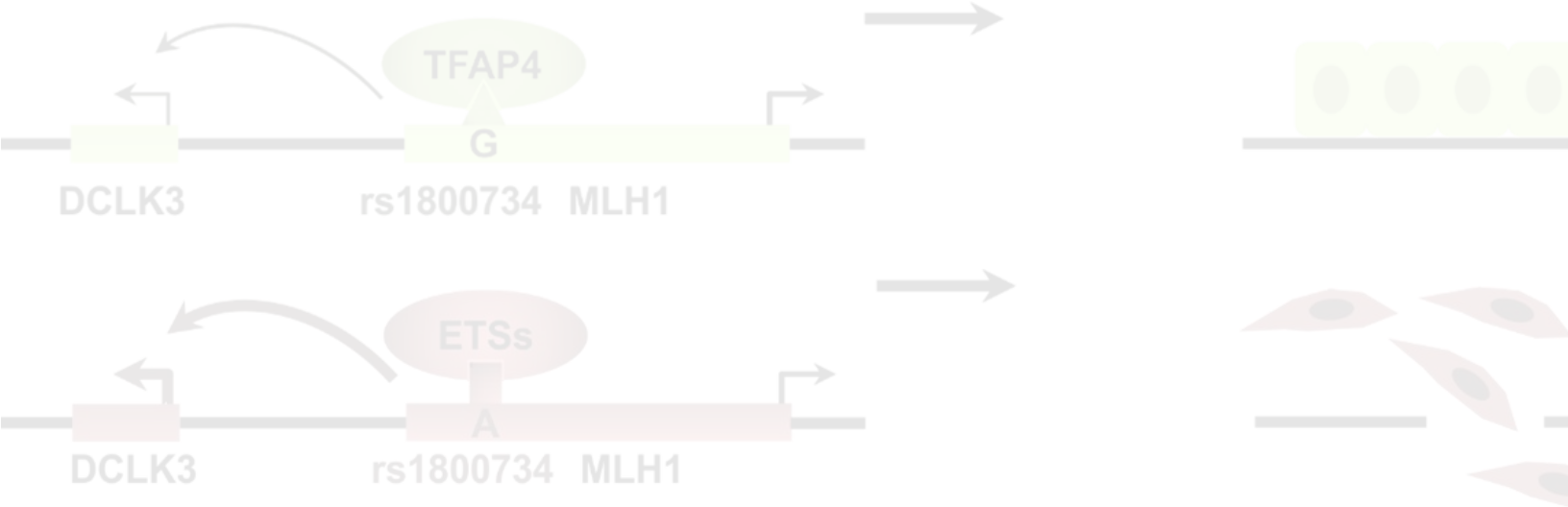
Chapter 4

The non-coding variant rs1800734 enhances DCLK3 expression through long-range interaction and promotes colorectal cancer progression

Ning Qing Liu[1,*], Menno ter Huurne[1], Luan N. Nguyen[1], Tianran Peng[1], Shuang-Yin Wang[1], James B Studd[2], Onkar Joshi[1], Halit Ongen[3], Jesper B. Bramsen[4], Jian Yan[5,6], Claus L. Andersen[4], Jussi Taipale[5], Emmanouil T. Dermitzakis[3], Richard S. Houlston[2], Nina C. Hubner[1], Hendrik G. Stunnenberg[1,*]

[1] Department of Molecular Biology, Faculty of Science, Radboud University, PO BOX 9101, 6500HB, Nijmegen, The Netherlands.

[2] Division of Genetics and Epidemiology, Institute of Cancer Research, 15 Cotswold Road, SM2 5NG, Sutton, Surrey, UK.

[3] Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva 1211, Switzerland.

[4] Department of Molecular Medicine, Aarhus University Hospital, Palle Juul-Jensens Boulevard 99, DK-8200, Aarhus, Denmark.

[5] Division of Functional Genomics and Systems Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE 141 83, Stockholm, Sweden.

[6] Ludwig Institute for Cancer Research, 9500 Gilman Dr, La Jolla, CA 92093, USA.

Summary

Genome-wide association studies have identified a great number of non-coding risk variants for colorectal cancer (CRC). To date, the majority of these variants have not been functionally studied. Identification of allele-specific transcription factor (TF) binding is of great importance to understand regulatory consequences of such variants. A recently developed proteome-wide analysis of disease-associated SNPs (PWAS) enables identification of TF-DNA interactions in an unbiased fashion. Here, we perform a large-scale PWAS study to comprehensively characterize TF binding landscape that is associated with CRC, which identifies 731 allele-specific TF binding at 116 CRC risk loci. This screen identifies the A-allele of rs1800734 within the promoter region of MLH1 as perturbing the binding of TFAP4 and consequently increasing DCLK3 expression through a long-range interaction, which promotes cancer malignancy through enhancing expression of the genes related to epithelial-to-mesenchymal transition.

Introduction

An individual's risk to develop colorectal cancer (CRC) is affected by a broad spectrum of genetic variants that abolish the functions and/or alter the expression of target genes. In CRC, two types of genetic variants have been extensively discussed to contribute to disease onset and progression: (1) protein-coding mutations; (2) non-coding variants in particular in DNA regulatory elements (DREs). To date, the majority of studies have focused on protein-coding mutations. Key coding mutations, such as APC/CTNNB1, KRAS/BRAF, PIK3CA, TP53, and SMAD4, have been intensively characterized[1]. However, even though a great number of non-coding risk variants for CRC have been identified in genome-wide association studies (GWAS)[2–9], their molecular functions have rarely been determined.

Functional genetic variants in distal DREs may alter transcription networks in several ways such as by affecting TF binding. SNPs within TF binding sites may affect the local chromatin accessibility[10–12] and/or alter the expression of gene targets through mediating different chromatin interactions[13–15]. Therefore, identification of variant-specific TF interactors is of great importance to understand regulatory consequences of the variants. However, DNA oriented methods, such as DNase I sequencing (DNase I-seq)[16], systematic evolution of ligands by exponential enrichment sequencing (SELEX-seq)[17] and chromatin immunoprecipitation with massively parallel DNA sequencing (ChIP-seq)[18], are biased by DNA motif knowledge or limited by the availability of antibodies, resulting in a biased identification of TF binding dynamics at disease associated loci. A recently developed proteome-wide analysis of disease-associated SNPs (PWAS) enables to identify DNA-TF interactions in an unbiased fashion[19]. A similar approach has been used to characterize a protein-DNA interaction map for ultra-conserved elements[20]. Therefore, we performed a large-scale PWAS studies to comprehensively understand TF binding landscape related to CRC.

As the outcome from our PWAS screen, we further investigated the functions of a SNP located in the promoter region of MLH1 gene (MLH1-93G>A or rs1800734), which has been associated with the risk of several cancer types including CRC[9,21,22], endometrial cancer[23], glioblastoma[24] and lung cancer[25]. Here, we identified a molecular function of this SNP in promoting cancer malignancy through a novel gene target named DCLK3.
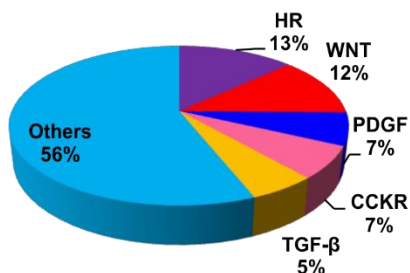
Results

PWAS identified TF occupancy switching at the 116 loci

We selected 116 SNPs associated with CRC risk[2–9] for PWAS analysis including (1) typed and imputed GWAS significant SNPs (for imputed SNPs, LD $r^2 \geq 0.2$) from 8q24.21 (MYC-335), 15q13.3 (GREM1) and 18q21.1 (SMAD7), (2) SNPs with functional evidence rs16969681/15q13.3[26], rs58920878/18q21.1[27], rs16888589/8q23.3[28] and rs4444235/14q22.2[29], (3) 3 SNPs from 3p22.2 (MLH1 region)[9], and (4) GWAS significant SNPs in DNase I-seq peaks in minimal one of 15 fetal large intestine tissues and 12 CRC cell lines (Supplementary Data 1). The PWAS analysis identified 731 TF binding alterations between reference (Ref) and alternate (Alt) alleles (P<0.01, A/B significance test) (Supplementary Data 2). Compared to proteome data, TF-DNA interactome data showed a clear enrichment for TFs (Supplementary Fig. 1a), and the altered binding events mediated by known TFs (Supplementary Data 3) showed stronger allele preference than other interactions (P= 4.4×10^{-6}, Mann-Whitney U test) (Supplementary Fig. 1b). Overlay of the pull-downs showed a consistent allele preference between 2 replicate experiments (Supplementary Fig. 1c). Many of the 731 TFs showed >8 fold affinity to one of the alleles at these loci (Fig. 1a), e.g. TFAP4 at rs1800734 and RUNX1/RUNX2/CBFB at rs1741640. As expected, top pathways associated with the 731 TFs included key CRC drivers such as WNT and TGF-β pathways (Fig. 1b).

It is well-known that DNase I hypersensitive sites coincide with regulatory elements and are the hotspots for TF binding[16]. To better predict in vivo TF binding, it is necessary to consider the hypersensitivity of tested regions. Therefore, we ranked all the SNP-TF interactions based on PWAS fold change (Ref/Alt allele) and DNase I hypersensitivity (DHS) of the SNP loci. A total of 27 significant allele-specific SNP-TF binding events were considered to be important (Fig. 1c and Supplementary Table 1). Many selected TF-SNP interactions (Fig. 1d and Supplementary Fig. 1d) were validated using ChIP-seq data (Supplementary Fig. 1e). Based on this selection, the top candidate SNP is rs1800734 (MLH1 -93 G>A) in the 5'-UTR region of MLH1 gene. Therefore, we decided to focus on this SNP.

4

a Allele specific TF binding landscape at the 116 loci
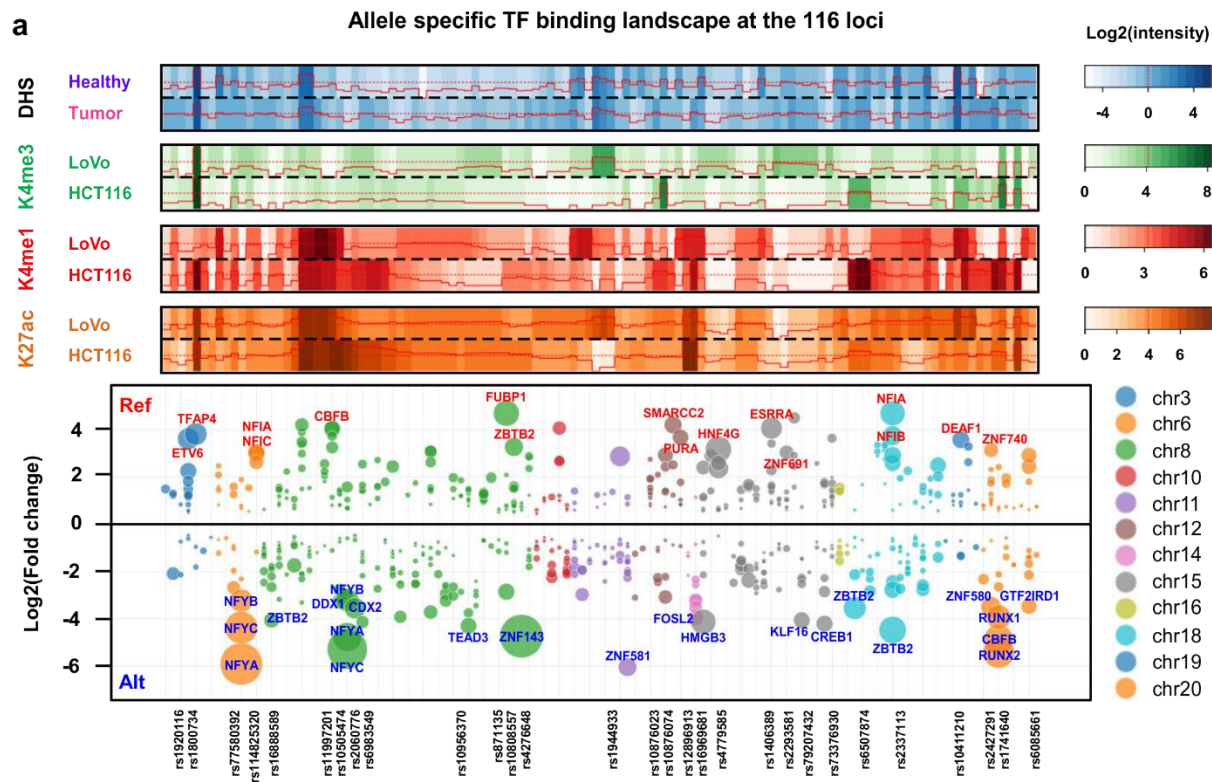
b TF interactor associated pathways

c

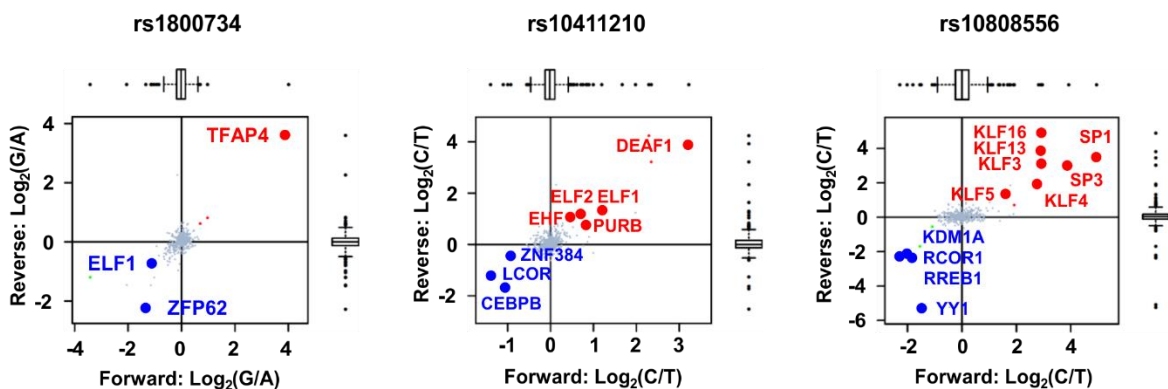**d** rs1800734   rs10411210   rs10808556
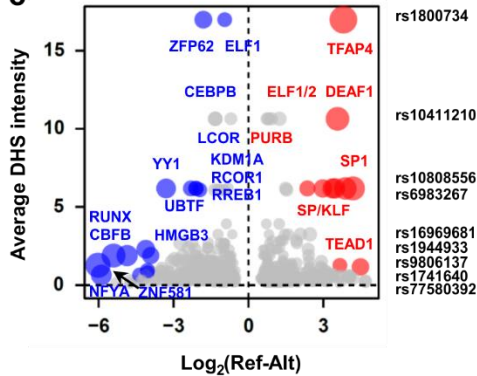
Fig. 1 PWAS screen systematically identified allele-specific TF binding at the selected CRC risk loci. (a) Allele-specific binding of the 731 candidate TFs at the 116 CRC risk loci. Chromatin environment of the 116 SNPs was described by DNase I hypersensitivity (DHS) of these loci in the 15 fetal large intestine tissues and 12 CRC cell lines, and histone modifications (H3K4me3, H3K4me1, and H3K27ac) within ±1 kb regions around these loci in the LoVo and HCT116 cell lines. The TFs with P-value$<10^{-30}$ and absolute $\log_2$(fold change)>3 was listed in the bubble plot. Bubble size represents -$\log_{10}$(P-values) of the interactors in the pull-down screen (n=2 pull-downs per SNP, P-values: A/B significance test). **(b)** Pathway annotation of the 731 TFs (HR: Gonadotropin releasing hormone receptor pathway, WNT/PDGF/CCKR/TGF-β: Wnt/PDGF/Cholecystokinin/TGF-beta signaling pathways). **(c)** TF-SNP interactions ranked by fold changes in the PWAS screen and DHS at the SNP loci. Bubble size indicates the -$\log_{10}$(P-values) of the TF-SNP interactions (n=2 pull-downs per SNP, red and blue bubbles: P-value<0.05, Z-test). **(d)** The top 3 candidate TF-SNP interactions (n=2 pull-downs per SNP, red and blue dots: P-values<0.01, A/B significance test).

PWAS identified specific interactors of rs1800734

Our PWAS screen identified TFAP4 as an allele-specific interactor with an almost 16-fold higher affinity for G-allele, while ELF1 showed higher affinity for the A-allele of rs1800734 in LoVo cells (Fig. 1d). The A-allele abolishes TFAP4 binding due to a point mutation at the last position of the E-box (Fig. 2a), which simultaneously creates an ETS family binding motif (Supplementary Fig. 2a). We corroborated and extended this finding by label-free based DNA pull-down with SNU175 and COLO320 extracts (Fig. 2b & supplementary Fig. 2b). Notably, other bHLH and ETS family members also displayed allele-specific binding, including MYC and ELF2 (Supplementary Fig. 2b/c), indicating that these TFs can compete with TFAP4 and ELF1 at this locus. Using ChIP-seq as an orthogonal technique, we validated TFAP4 and ELF1 binding at the rs1800734 in the SNU175 and COLO320, cell lines heterozygous for this locus. Consistent with PWAS, TFAP4 ChIP-seq showed a dominant preference for G-allele binding in the two cell lines (Fig. 2c). ELF1 did not show significant allele binding preference (Supplementary Fig. 2d), which may be due to the competitive binding interference by other ETS proteins.

DCLK3 is a novel target of rs1800734

Given the position of the SNP in the promoter of MLH1, we further investigated whether predisposition of rs1800734 in CRC is due to DNA methylation of MLH1 promoter as proposed[9,30]. We tested hypersensitivity and transcription of G- and A-allele in the two heterozygous cell lines, which showed the two alleles are equally accessible and transcribed (Fig. 2d). The neighbouring gene EPM2AIP1, also reported to be regulated by rs1800734[31], was similarly unaffected by rs1800734 (Fig. 2d). Hence we conclude that rs1800734 does not result in allele-specific epigenetic silencing of either MLH1 or EPM2AIP1 in these cell lines. We sought to confirm our observation in the Systems Biology of Colorectal Cancer (SYSCOL) cohort of paired healthy and tumor tissues (healthy: n=288, tumor: n =289). A strong eQTL between rs1800734 and MLH1 was observed in the healthy tissues (P=0.001, linear regression model) but the A-allele was associated with increased MLH1 expression. This eQTL was lost in the tumors and remained only a weak association in microsatellite stable (MSS) tumors (P=0.025, n=236, linear regression model) (Supplementary Fig. 3a). Intriguingly, we identified a correlation between rs1800734 and DCLK3

expression in the healthy (P=0.029, linear regression model) and tumor (P=0.031, linear regression model) samples, and this association was highly significant in the MSS patients (P=0.004, linear regression model) (Fig. 2e), indicating this locus may act as a distal enhancer and regulate DCLK3.
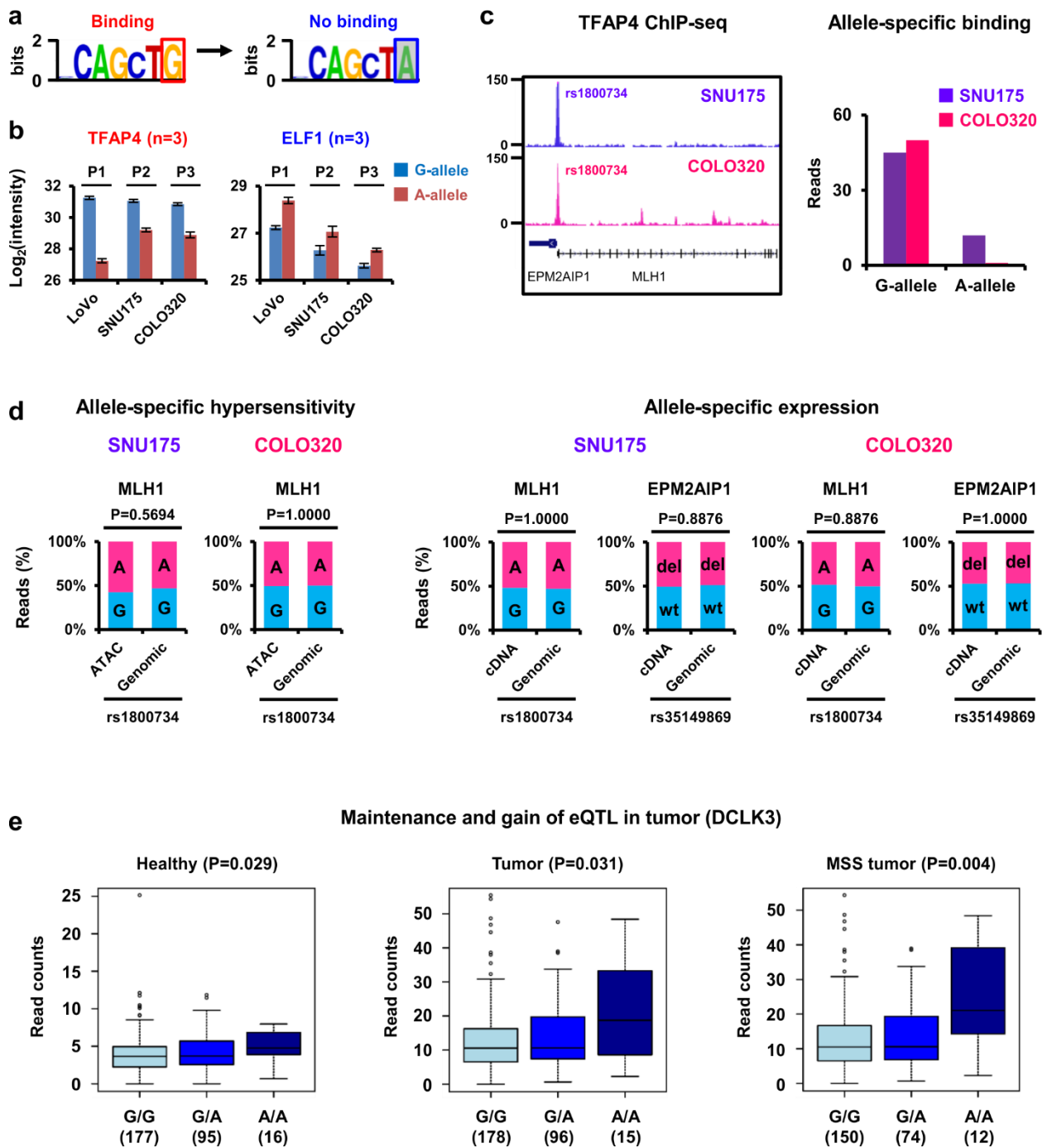
Fig. 2 Identification of allele-specific interactors for rs1800734 and its potential gene targets. (a) Motif analysis interpreted that A-allele of the loci perturbs E-box motif. **(b)** The specific binding of TFAP4 (P1=2.0×10^{-6}, P2=3.1×10^{-5}, P3=7.8×10^{-5}, Student's t-test) and ELF1 (P1=2.1×10^{-4}, P2=0.011, P3=6.9×10^{-4}, Student's t-test) on G- and A-allele was observed in the three different cell lines (Data are represented as mean and error bars indicate s.d., n=3 pull-downs per SNP). ZFP62 binding on A-allele, as shown in Fig. 1d, was not consistent in all the three cell lines, and was hence not considered as the general TF regulator at this locus. **(c)** This allele-specific binding was validated by ChIP-seq of TFAP4; This SNP did not change **(d)** local chromatin accessibility and the expression of two cis-regulated genes (MLH1 and EPM2AIP1) (G (G-allele) and A (A-allele) of rs1800734, wt (wild-type allele) and del (deletion allele) of rs35149869, P-values: Fisher's exact test). **(e)** eQTL analysis revealed a novel gene targets DCLK3 of this SNP (P-values were calculated in FastQTL software based on linear regression model).

The A-allele positively regulates transcription of DCLK3

To establish the functional relation between rs1800734 and DCLK3 expression, two isogenic cell lines homozygous for G- or A-allele were generated from COLO320 using CRISPR-CAS9 technique (Fig. 3a). Successful targeting of rs1800734 was confirmed by Sanger sequencing (Fig. 3b). No other mutation was observed in the surrounding region. ChIP-qPCR confirmed higher TFAP4 binding to the G-allele (Supplementary Fig. 3b), and the level of MLH1 and EPM2AIP1 transcription was identical in the three isogenic lines (Supplementary Fig. 3c). Importantly, the eQTL association between rs1800734 and DCLK3 expression was replicated, and a 5-fold difference in DCLK3 expression was observed between G- and A-homozygotes (Fig. 3c). By sequencing the inter-exonic RT-qPCR products of DCLK3, we confirmed the transcription of DCLK3 in these cell lines (Supplementary Fig. 3d).

rs1800734 regulates DCLK3 through long-range interactions

A capture HiC study has suggested that rs3806624 in the promoter of EOMES affects AZI2, a gene 640 kbp down-stream to this SNP, through a long-range chromatin interaction[14]. A similar constellation may apply to rs1800734. Therefore, 4C-seq was employed to search for long-range interactions between rs1800734 and other potential targets in the three isogenic cell lines. Using rs1800734 as a view point, a significantly enhanced interaction was observed in A-homozygote with DCLK3 region (Fig. 3d). Increased interactions were found in the promoter and 3'-UTR region of DCLK3 in two independent experiments (Fig. 3e), and additionally the 3'-UTR interaction appeared to increase chromatin accessibility (Fig. 3f). In conclusion, the A-allele of rs1800734 increases the DCLK3 transcription through increased chromatin interaction and enhanced chromatin accessibility.
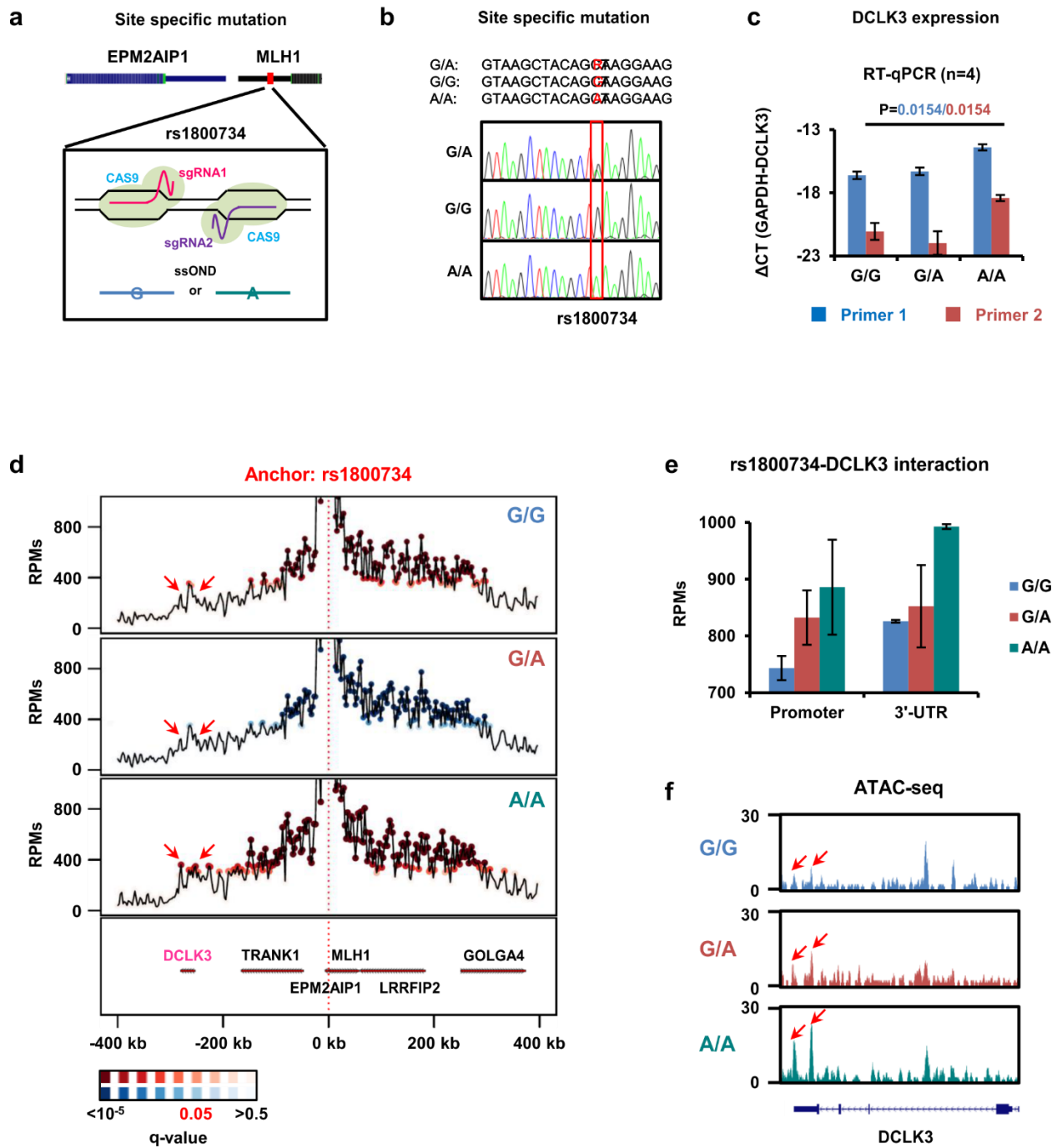
4

a Site specific mutation

b Site specific mutation

c DCLK3 expression

d Anchor: rs1800734

e rs1800734-DCLK3 interaction

f ATAC-seq

Fig. 3 Molecular regulatory mechanisms of rs1800734 on DCLK3. (a) Isogenic cell lines possessing homozygous mutations (G/G and A/A) at rs1800734 were generated and **(b)** confirmed, and (**c**) the expression of DCLK3 was elevated in the A/A cell line (Data are represented as mean and error bars indicate s.d., n=4 biological replicates per cell line, P-values: Student's t-test). **(d)** 4C-seq identified chromatin interactions between rs1800734 and the DCLK3 region. **(e)** The A/A cell line showed stronger chromatin interaction than other cell lines (Data are represented as mean and error bars indicate s.d., n=2 biological replicates per cell line). **(f)** The enhanced chromatin interaction results in increased chromatin accessibility at 3'-UTR region of DCLK3 in the A/A line.

DCLK3 is a potential oncogenic and tumor progressive factor

DCLK3 is one of three doublecortin-like kinases (DCLK1, DCLK2, and DCLK3). In this family, DCLK1 has been shown to be a cancer stem cell marker in intestinal tumors[32]. The molecular function of the DCLK3 has not been characterized in depth. We therefore performed gene set enrichment analysis (GSEA) to identify DCLK3 associated gene sets in the SYSCOL RNA-seq cohort. Interestingly, we found that epithelial-to-mesenchymal transition (EMT) related genes were highly correlated to the expression of DCLK3 in the healthy tissues (normalized enrichment score/NES=2.50), which was enhanced in the tumor samples (NES=3.10) (Fig. 4a and Supplementary Table 2). Using a cutoff of NES at 2.00, we also identified gene sets that are preferentially enriched in the healthy versus tumor tissues (Supplementary Table 2), e.g. "MYC targets" and "G2/M checkpoint" gene sets were enriched in the healthy tissue, whereas "Angiogenensis" and "TNFA signaling via NF-κB" gene sets were enriched in the tumor tissues (Fig. 4a). As expected, common EMT markers, such as CALD1, FN1, SNAI1, SNAI2, TWIST1, VIM, ZEB1 and ZEB2, were highly significantly co-expressed with DCLK3 in the tumors (Fig. 4a), indicating that DCLK3 is an EMT regulator. Furthermore, we observed elevated DCLK3 expression in the tumor compared to healthy tissues ($P=2.2\times10^{-16}$, Mann-Whitney U test) (Fig. 4b), but no difference between the MSI and MSS tumors (Fig. 4b). Interestingly, this elevation appeared to be correlated with the CRC progression: DCLK3 expression was at a comparable level in the healthy and adenoma tissues, and increased ~2-fold in the stage I tumors and remained at this level during tumor malignancy ($P=2.2\times10^{-16}$, Kruskal-Wallis test) (Fig. 4b). Therefore, DCLK3 may promote EMT events and consequently drive tumor malignancy.

4

a DCLK3 associated pathways (SYSCOL cohort)

b DCLK3 expression v.s. tumor onset/progression (SYSCOL cohort)

MSI: instable
MSS: stable
H: healthy
A: adenoma
I: stage 1
II: stage 2
III: stage 3
IV: stage 4

c

Protective allele:
TFAP4 binding
Decreased chromatin interaction
Decreased DCLK3 expression

Low-grade tumor:
Epithelial feature
Intact basal layer
Well-differentiated cells

Risk allele:
ETS family binding
Increased chromatin interaction
Increased DCLK3 expression

High-grade tumor:
Mesenchymal feature
Leaking basal layer
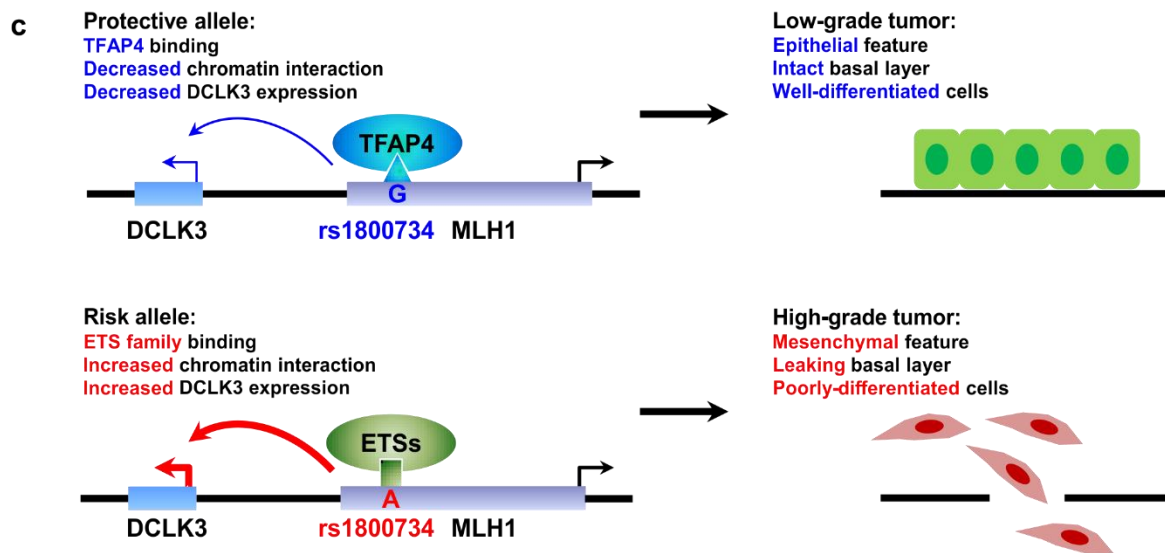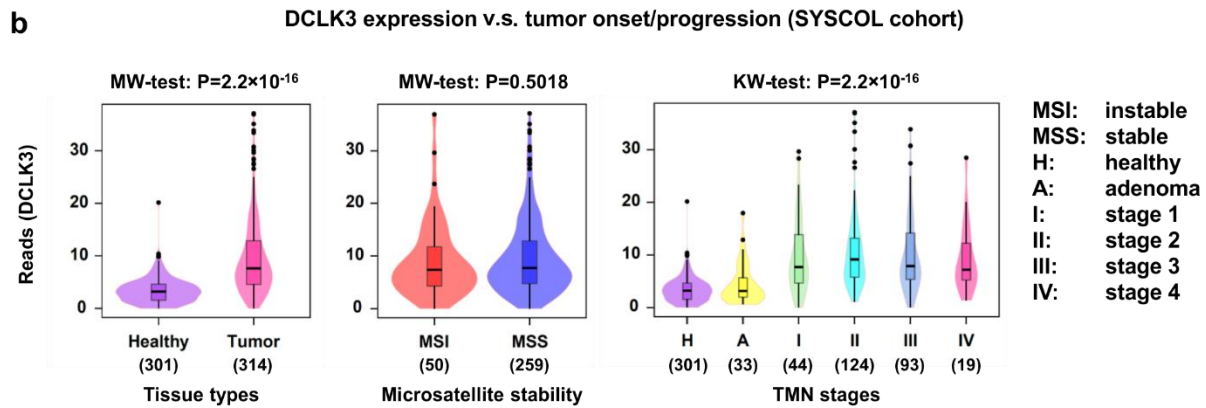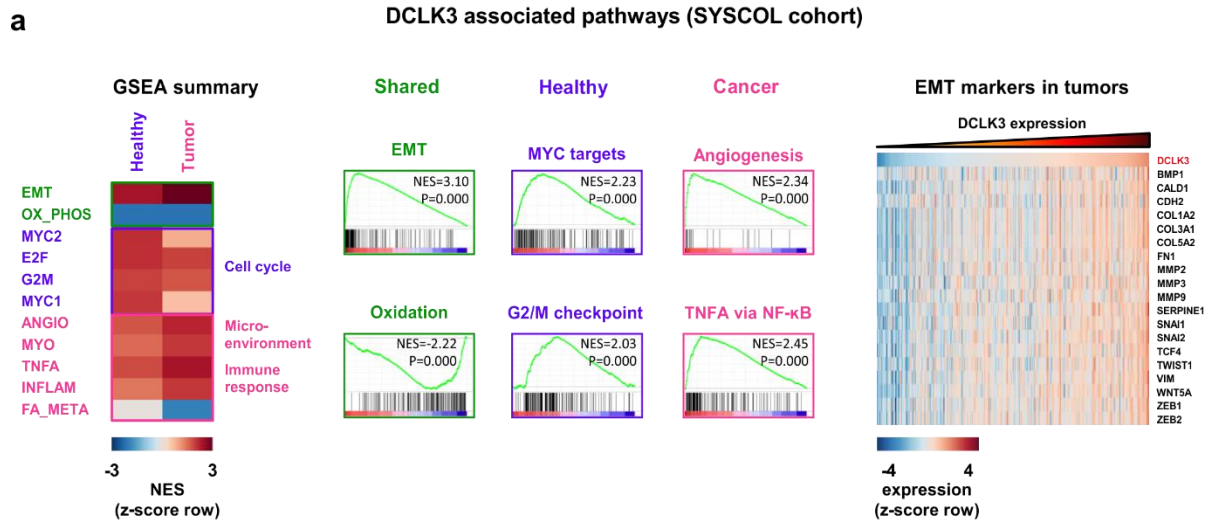**Poorly-differentiated** cells

Fig. 4 Identification of DCLK3 associated gene sets and clinical parameters in the SYSCOL RNA-seq cohort. (a) GSEA analysis revealed DCLK3 correlated cancer hallmark gene sets. EMT and oxidative phosphorylation associated gene sets were highly significantly associated with the DCLK3 expression in both healthy and tumor tissues. Also, some DCLK3 associated gene sets showed healthy or tumor tissue specificity (P-values were calculated in GSEA based on Pearson correlation). The expression of key EMT markers showed good correlation with the DCLK3 expression. **(b)** The DCLK3 expression was preferentially elevated in the tumors and especially the malignant tissues. However, we did not observe a clear difference between MSI and MSS tumors (P-values: Mann-whitney U test/MW-test and Kruskal-Wallis test/KW-test). **(c)** A proposed model describing how rs1800734 modifies the risk of CRC malignancy. TFAP4 and ETS family members specifically bind to the protective G- or risk A-allele, respectively. The rs1800734-ETS interaction increases the enhancer activity of the rs1800734 locus and enhances the expression of DCLK3 through an increased chromatin interaction. Cancer cells with the elevated DCLK3 expression undergo EMT and therefore metastasize to distal sites.

Discussion

Our study generated the first TF-SNP interaction map at presumed disease relevant loci of CRC and determined TF binding occupancy at the 116 upmost relevant CRC risk loci. Together with GWAS and epigenetic profiling data, our PWAS screens provide a comprehensive TF binding landscape of these loci and yielded candidate interactions for further functional investigations (Fig. 1a/c/d and Supplementary Fig. 1d/e). As an alternative tool to investigate allele-specific TF binding events, the PWAS approach shows several advantages over the DNA-centric methods. In contrast to ChIP-seq technique, PWAS based identification is a hypothesis-free approach which does not require any knowledge on the possible binding TFs at risk loci. Also, application of PWAS approach is not limited by the availability of high quality ChIP grade antibodies. Even though large ChIP-seq datasets have been generated[18,33], these data nevertheless cover only a small proportion of the entire repertoire of TFs. Other DNA-centric methods, such as DNase I-seq or ATAC-seq, are only capable of identifying DNA motifs present at risk loci, which makes it difficult to predict actual binding TF(s) from a family sharing similar motifs in a specific cell type. The outcome of PWAS is a reflection of TF abundance, TF ability to bind to the sequence (affinity) as well as synergistic and antagonist effects due to binding of other TFs to adjacent or overlapping sequences. Therefore, actual binding TFs in a given cell type can be predicted using PWAS method. In addition, PWAS approach also help to identify binding TFs at the loci only partially matching consensus motif sequences, and therefore cannot be predicted by motif prediction based methods.

Based on the PWAS screen, we investigated in great depths for a well-known CRC associated SNP: MLH1 -93G>A or rs1800734. It has been postulated that A-allele (risk allele) of the rs1800734 recruits repressive TFs which subsequently results in promoter methylation of MLH1 gene[9,30], supported by an association between the A-allele of the rs1800734 and promoter methylation[34] or decreased expression of MLH1[35,36]. However, Suter and colleagues showed contradictory results that A-allele of this SNP is associated with lower promoter methylation and higher transcription of MLH1[37]. Our results suggested that TFAP4 preferentially binds to T-allele of the rs1800734 but does not change promoter accessibility and transcription

of MLH1. Analysis of the SYSCOL cohort strengthened the findings in the cell lines (Supplementary Fig. 3a) and identified a new gene target DCLK3 in the MSS patients (Fig. 2e). Furthermore, CRISPR-CAS9 facilitated to generate fully comparable isogenic lines carrying G- or A-point mutation at rs1800734 locus. Therefore, mild changes in chromatin interaction, accessibility and consequently gene expression can be monitored by different genomic techniques. It has been shown that some functional SNPs in enhancer regions result in subtle changes in expression of their target genes, e.g. the G-allele of rs356168 increased SNCA expression by 1.06 times in neurons and 1.18 times in neuron precursors[38]. Hence, accurate genome editing is required to distinguish these subtle changes. Using this model system, an increased DCLK3 transcription was detected in the A/A homozygous line (Fig. 3c), which is due to increased chromatin interaction between the two locus and consequently elevated chromatin accessibility in the DCLK3 region (Fig. 3d/e/f). These data confirmed that rs1800734, even though locates in the promoter region of MLH1, serves as a distal enhancer for the DCLK3 gene. Notably, our findings are based on the genetic background of MSS tumors, which is likely responsible for the contradiction between our data and some of the literature. Moreover, DCLK3 has been shown to be associated with EMT process in this study (Fig. 4a). Although the full molecular mechanism of DCLK3 in regulating EMT has not been characterized, this protein has been shown to directly interact with CDK5[39], and the latter promotes breast cancer metastasis through regulating TGF-β1 induced EMT[40]. Alternatively, CDK5 also prevents phosphorylation and degradation of a EMT regulator CALD1[41] and hence promotes the EMT process[42]. Furthermore, DCLK3 may perform a similar function as another doublecortin-like kinase DCLK1, since it possesses the similar protein kinase domain and one of the two doublecortin domains[43] as DCLK1. In intestinal tumors, DCLK1 often co-expresses with LGR5 at crypt base, and DCLK1+LGR5+ stem cells are able to continuously produce tumor progeny under the APC+/- mice[32]. A further study showed that DCLK1+ cells are long-lived and quiescent population, that is only activated and display carcinogenesis properties upon oncogenic mutation and tissue injury[44].

In summary, we conclude that ETS family TFs preferentially bind to the A-allele of rs1800734 and increase chromatin interaction between the rs1800734 locus and the DCLK3 region. This enhanced chromatin interaction in turn increases the expression of DCLK3. Consequently, the risk of tumor metastasis is increased due to increased EMT feature of cancer cells (Fig. 4c). In addition, our study systematically identified changes in TF binding at regulatory CRC risk loci, which provide candidates for functional follow-up.

Methods

Cell culture and extraction of nuclear soluble fraction

Human colorectal cancer cell lines were cultured in DMEM or RPMI medium supplemented with 10% fetal bovine, 100 U/ml penicillin and 100 μg/ml streptomycin. LoVo and SNU175 cell lines were purchased from American Type Culture Collection (ATCC) and Korean Cell Line Bank (KCLB), respectively. COLO320 cell line was a generous gift of Dr. Riccardo Fodde (Erasmus Medical Center Rotterdam, The Netherlands),

which is originally from ATCC. The authenticity of the cell lines were confirmed using microsatellite STR assay by the suppliers of the cell lines. Mycoplasma infection was routinely tested in house to ensure that all the cell lines used for this study were free of mycoplasma contamination. Nuclear soluble fraction of LoVo, SNU175 and COLO320 was performed using a published protocol[45]. Protein concentration of the obtained nuclear extract was quantified using Branford assay. Each of the 3 mg extract (12 individual DNA pull-downs) was aliquoted, snap-frozen in liquid nitrogen and stored at -80°C.

High throughput DNA pull-down

High throughput DNA pull-down experiments was performed on 96-well filter plate format using our published method[46] with minor modifications. To synthesize biotinylated dsDNA oligo, we attached a non-genomic 15 bp sequence at the 3' end of the anti-sense strands. Subsequently, a reverse complement biotinylated primer was used to extend single-stranded DNA (ssDNA) templates into dsDNA oligos. For each synthesis, 150 pmol of the biotinylated primer and 200 pmol template were subjected to a PCR reaction using Herculase II Fusion Enzyme kit under the following conditions: 95°C for 3 min; thermocycling (n=20) at 95°C for 1 min, 45°C for 1 min and 72°C for 1 min; 72°C for 3 min; infinite hold at 12°C. Briefly, high throughput DNA pull-down was performed using a Multiscreen filter plate with 1.2 μm pores (Millipore, MSBVN1210). The biotinylated dsDNA oligos were immobilized on 20 μl of high-performance streptavidin sepharose (GE Healthcare, 17511301). Two-hundred fifty μg nuclear extracts and 15 μg of competitors (5 μg of poly(deoxyinosinic-deoxycytidylic) acid sodium salt (Sigma-Aldrich, P4929), poly(deoxyadenylic-thymidylic) acid sodium salt (Sigma-Aldrich, P0883) and Bakers yeast RNA (Sigma-Aldrich, R6750)) were added and incubated with immobilized oligos for 1.5 h at 4°C on a plate shaker for each of the pull-down experiments. The components of the competitors were sonicated into ~300 bp fragments prior to use. The proteins unbound to DNA oligos were washed off using different washing buffers and the bound TFs were on-bead digested overnight using trypsin/lysC. The pull-down duplicates underwent dimethyl label swapping and measured by nLC-MS/MS in a 2h gradient.

Nuclear extract proteome

Deep proteome profile was generated from the nuclear extracts used for the pull-down experiments. An absolute quantification strategy was taken following a published method[46]. In brief, 3.3 μg of universal protein standard 2 (UPS2, Sigma-Aldrich) and 10 μg of the nuclear extracts were mixed and subjected to a filter-aided sample preparation (FASP) digestion. In parallel, 30 μg of the nuclear extracts were also digested using the FASP protocol and subsequently separated into 6 fractions using strong anion exchange. The UPS2 sample and the fractions was purified on C18 stage-tips and profiled by nLC-MS/MS in a 4h gradient.

DNase I-seq

DNase I library of the LoVo cell line was constructed following a reported protocol with some minor modifications[47]. In short, 5×10^6 nuclei were isolated using a buffer (15 mM Tris-HCl (pH=8.0), 15 mM NaCl, 60 mM KCl, 1 mM EDTA (pH=8.0), 0.5 mM EGTA (pH=8.0) and 0.5 mM Spermidine) supplemented with 0.05% IGEPAL CA-630 detergent. Subsequently, the isolated nuclei were digested with 80 U DNase I (Sigma-Aldrich, D4527) for 3 min and the digestion was quenched by a stop buffer (50 mM Tris-HCl (pH=8.0), 100 mM NaCl, 0.1% SDS, 100 mM EDTA (pH=8.0), 1 mM Spermidine and 0.3 mM Spermine). A 9% Sucrose gradient was applied to fractionate the samples for 24 h at 25,000 rpm at 16°C, and the fractions with <1 kb fragments were further purified and prepared according to the Illumina library preparation protocol.

DNase I library of other cell lines were prepared using a published protocol[18] as described below. The cell lines were harvested under the confluence of 60% and washed with PBS. Nuclei were isolated with RSB lysis buffer (10 mM Tris-Cl, pH=7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL CA-630) at 4°C for 10 min. Then the nuclei were treated with 0.12 unit of DNase I (Roche) in the provided buffer at 37 °C for 15 min before being quenched by 50 mM of EDTA. Following RNase A (Sigma) treatment at 37 °C for 15 min, proteinase K (NEB) was added for an additional hour at 56°C. DNA was extracted using Phenol:Chloroform:Isoamyl-Alcohol. Agarose Gel (2%) electrophoresis was applied to separate the released fragments (~100 bp) that were purified (Qiagen, MinElute Gel Extraction Kit), followed by Illumina TruSeq library preparation and Sequencing (HiSeq2000).

ChIP-seq and ChIP-qPCR

ChIP assays were performed following a standard protocol. Cell lines were cross-linked by a final concentration of 1% paraformaldehyde for 10 min and subsequently cross-linking reaction was quenched using 1.5 M glycine. The harvested cell lines were then lysed and sonicated to obtain ~300 bp chromatin using Bioruptor Plus sonication device (Diagenode). The sonicated chromatin was pre-cleared by Protein A/G magetic beads (ThermoFisher Scientific, 88802) and then incubated together with antibody conjuncated beads overnight at 4°C. Antibodies against H3K27ac (Diagenode, C15410196, 1 µg per ChIP assay), TFAP4 (Santa Cruz Biotechnology, sc-18593X, 6 µg per ChIP assay) and ELF1 (Santa Cruz Biotechnology, sc-631X, 4 µg per ChIP assay) were used in our ChIP experiments. Posterior to the incubation, captured chromatin was washed, eluted and de-crosslinked. Resulted DNA fragments were purified and prepared according to Illumina library preparation (H3K27ac ChIP) or KAPA Hyper Prep (TFAP4 and ELF1 ChIP) protocols prior to sequecning, or directly quantified using SYBR Green based qPCR assays (Supplementary Table 3).

ATAC-seq

ATAC libraries of the SNU175 and COLO320 cell lines were prepared by a documented protocol[48] with some modifications. In brief, nuclei were isolated using a lysis buffer consisting of 10 mM Tris-HCl (pH=7.5),

10 mM NaCl, 3 mM MgCl$_2$ and 0.1% IGEPAL CA-630 detergent and then tagmentated[49] using 2 µl of Tn5 transposase and 12.5 ul 2X TD buffer (Illumina, Nextera DNA Library Preparation Kit). The resulted DNA fragments underwent 2 sequential 9-cycle PCR amplification, and in between two PCR reactions the libraries were selected for <600 bp fragments using SPRI beads. The final PCR products were purified and quantified by KAPA Library Quantification Kits prior to sequencing.

Targeted RNA-seq and RT-qPCR

Total RNA was isolated from the SNU175 and COLO320 cell lines using a TRIzol reagent (ThermoFisher Scientific, 15596018) based method. The yielded RNA was treated using DNase and then reversely transcribed into cDNA using random hexamers (ThermoFisher Scientific, SO142). The cDNA was amplified using targeted primers and followed by standard KAPA Hyper Prep protocols, or directly quantified using SYBR Green based qPCR assays (Supplementary Table 3).

CRISPR-CAS9 based SNP editing

CRISPR-CAS9 based SNP editing were performed according to a previously reported method[50]. Guide RNAs (sgRNAs) were designed using an online tool (http://crispr.mit.edu/) and double-nicking strategy was taken to reduce undesirable off-target mutagenesis. The sgRNAs were then cloned into a U6-driven plasmid containing a GFP marker and the D10A mutant Cas9 nickase (Addgene, pSpCas9n(BB)-2A-GFP, Plasmid#48140). The plasmid was then transformed into the DH5α competent E.coli strain and the products were purified for transfection. Subsequently, 2 sgRNAs (400 ng per each) and an 199 bp single-stranded oligonucleotides (10 pmol ssODNs, possessing G or A point mutations at the rs1800734 locus) were co-transfected into COLO320 cell lines following a standard lipofectamine LTX Reagent protocol (ThermoFisher Scientific, 15338100). Fluorescence-activated cell sorting (FACS) was used to sort 192 GFP-positive cells per cell line (G or A point mutations) into 96-well plates in 36 h after the transfection. In parallel, a wild-type cell line was treated in the same manner by without using sgRNAs, and this mocked cell line was then sorted by FACS and served as the control. The desirable genotype at the rs1800734 was confirmed by Sanger sequencing. The oligos (sgRNAs and ssODNs) used in these experiments were listed in Supplementary Table 3.

SNP genotyping

SNP genotyping was performed by a standard Sanger sequencing based method. The regions containing mutations were PCR amplified into ~500 bp fragments using specific primers (Supplementary Table 3) and the PCR products were purified using 1.5% agarose gel. A mixture of 10 ng purified PCR products and 6 pmol primers was used for Sanger sequencing and the data were analyzed using CodonCode Aligner (V.5.0.2).

4

4C-seq

4C experiments were carried out using a published protocol[51] with some modifications. For each assay, 1×10^7 cells were cross-linked and quenched as in ChIP assays. Nuclei were isolated in a 50 ml of lysis buffer (10 mM Tris-HCl (pH=7.5), 10 mM NaCl, 0.2% IGEPAL CA-630 detergent and 1X protease inhibitors). Subsequently, the nuclei were digested with 240 U NlaIII enzyme (NEW ENGLAND BioLabs Inc., R0215L) followed by an overnight in-nuclei ligation with 4,000 U T4 ligase (NEW ENGLAND BioLabs Inc., M0202M) at 16°C. The ligated DNA was de-crosslinked, purified, digested with 90 U CviQI enzyme (NEW ENGLAND BioLabs Inc., R0639S) and circularized by 5,000 U T4 ligase. The circularized DNA (16 × 300 ng) was amplified with bait-specific inverse primers (Supplementary Table 3), pooled and purified, followed by KAPA Hyper Prep protocols.

Proteomics data processing

Recorded MS files were analyzed by MaxQuant software (version 1.3.5.7)[52] using standard settings for dimethyl or label-free quantification analysis. All the files were searched against UniProtKB/Swiss-Prot human database (generated from version 06-2012). Batch effects of the pull-down data were removed using "Combat" algorithm[53]. In dimethyl analysis, some proteins were consistently quantified in the pull-downs of only one-allele (Ref- or Alt-allele) in two replicate experiments, and the protein ratios can therefore not be obtained from MaxQuant output. Therefore, we imputed the missing value in these pull-down using "Replace missing values from normal distribution" option in Perseus software (version 1.3.9.18)[54], which allows to further calculate protein ratios and perform down-stream analysis. TF interactors for each of the pull-down were identified using 'Significance B' function in Perseus software.

DNase I-seq ChIP-seq ATAC-seq data processing

Read mapping was performed using BWA-ALN (DNase I-seq and ChIP-seq) and BWA-MEM (ATAC-seq)[55] against the hg19 reference human genome. PCR-duplicates were removed for further data analysis. Peak calling was carried out by MASC2[56] with default settings, except H3K27ac peaks were called using "broad" option. Peaks were called at a q-value cutoff of 0.01. Overlapping peaks were merged for each of the different experiments before further analysis. Integrative Genomics Viewer (IGV)[57] was used to detect bi-allelic differential binding and hypersensitivity.

Targeted RNA-seq and RNA-seq data processing

Targeted RNA-seq data were mapped to the hg19 reference human genome using BWA-MEM[55]. IGV[57] was used to visualize the targeted RNA-seq data and detect bi-allelic differential expression. The SYSCOL RNA-seq cohort was mapped to the human reference genome sequence (GRCh37 autosomes + X + Y + M) using GEM mapper[58]. The reads with mapping quality <150 were excluded for further analysis. The genes were annotated using Ensemble 75. The reads of the genes were counted by "HTSeq" framework[59] and normalized using "DESeq" package[60].

Calculate significance of the TF-SNP interactions

Importance of the TF-SNP interaction in our screen study was considered by combining DNase I hypersensitivity of the SNPs (read counts at the SNP position in 15 fetal large intestine tissues and 12 CRC cell lines) and the interaction strength (Log$_2$(fold change) of a TF between Ref and Alt allele pull-down). The detailed P-value calculation was performed using z-test based on following steps: (1) Calculate average DNase I hypersensitivity (average DNase I-seq reads) of all the 116 loci in the DNase I-seq data from 15 fetal large intestine tissues and 12 CRC cell lines, and then z-score transform the hypersensitivity data of these loci; (2) Calculate absolute Log$_2$(fold change) of every TF-SNP interaction between Ref and Alt alleles, and then z-score transform the absolute Log$_2$(fold change) of all the interactions; (3) Average the z-scores of DNase I hypersensitivity and absolute Log$_2$(fold change), and convert the average z-scores into P-values based on normal statistical distribution.

eQTL analysis

The eQTL analysis was performed as described in a previous publication[61]. Germline genotypes of these patients were genotyped on the Illumina 2.5M Exome v1.0 and imputed to 1000 genomes phase 3 release using IMPUTE2. For cis-eQTL analysis, we normalized gene quantification separately for healthy and tumour samples. Technical covariates were discovered using the PEER program[62] and 20 PEER factors were used in normalization. The cis region was defined as ± 1 Mb from the transcription start site for each gene. The associations between genotypes and gene quantification were obtained using the FastQTL software[63].

4C-seq data processing

A reduced genome was generated by extracting the sequences flanking the NlaIII restriction sites (30 bp on each strand from the NlaIII restriction sites to downstream) using the hg19 reference human genome in order to improve the mappability of our 4C-seq data. Subsequently, the mappability of the reduced sequences from each strand was evaluated and the uniquely mappable NlaIII restriction sites were kept for downstream analysis.

The reads from each library were parsed based on the bait-specific primer sequence and mapped to the reduced hg19 genome using BWA-ALN with the default parameters. A Bioconductor package "r3Cseq"[64] with 2.5 kb sliding window was used to determine significant interactions and calculate interaction difference.

Pathway annotation and Gene set enrichment analysis

TFs binding differentially to the 116 loci were annotated to PANTHER pathways using PANTHER Classification System[65].

GSEA analysis[66] was applied to identify the gene sets correlated to DCLK3 expression. The search was performed against hallmark gene sets in Molecular Signatures Database (MSigDB, v5.1)[67].

Data availability

Raw and processed LC-MS/MS data and sequencing data are available at ProteomeXchange (http://www.ebi.ac.uk/pride/) and Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) under the accession numbers PXD004435 and GSE83968, respectively. The publically available data used in this study were downloaded from Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) using the following accession numbers: (1) H3K4me1 and H3K4me3 ChIP-seq: GSM1240111, GSM945304, GSM1208810, GSM1208811; (2) DNase I-seq: GSM736493, GSM736600, GSM736500, GSM736587, GSM665815, GSM665818, GSM665826, GSM701490, GSM701495, GSM701514, GSM701531, GSM774213, GSM774214, GSM774217, GSM774220, GSM774228, GSM774233, GSM817162, GSM817188; (3) TF ChIP-seq: GSM1010902, GSM1208683, GSM1208642, GSM1240820, GSM803354, GSM1010847, GSM1208763, GSM1010765, GSM1010790, GSM1010852, GSM1208598, GSM791411, GSM791412, GSM782123, GSM1122306, GSM722708, GSM1122302, GSM1122303. The detailed information of in-house generated and public datasets used in this study are listed in Supplementary Table 4. All other data used in the article and its supplementary files are available from the authors on request.

References

1.   TCGA. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487,** 330–337 (2012).

2.   Broderick, P. *et al.* A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat. Genet.* **39,** 1315–1317 (2007).

3.   Houlston, R. S. *et al.* Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat. Genet.* **40,** 1426–1435 (2008).

4.   Tomlinson, I. P. M. *et al.* A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.* **40,** 623–630 (2008).

5.   Tenesa, A. *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat. Genet.* **40,** 631–637 (2008).

6.   Jaeger, E. *et al.* Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat. Genet.* **40,** 26–28 (2008).

7.   Houlston, R. S. *et al.* Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat. Genet.* **42,** 973–977 (2010).

8.   Tomlinson, I. P. M. *et al.* Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet.* **7,** 2–12 (2011).

9.   Whiffin, N. *et al.* MLH1-93G > a is a risk factor for MSI colorectal cancer. *Carcinogenesis* **32,** 1157–1161 (2011).

10.   Degner, J. F. *et al.* DNase-I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482,** 390–394 (2012).

11.   del Rosario, R. C.-H. *et al.* Sensitive detection of chromatin-altering polymorphisms reveals autoimmune disease mechanisms. *Nat. Methods* **12,** 458–464 (2015).

12.   Maurano, M. T. *et al.* Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat. Genet.* **47,** 1393–1401 (2015).

13.   Duggal, G., Wang, H. & Kingsford, C. Higher-order chromatin domains link eQTLs with the expression of far-away genes. *Nucleic Acids Res.* **42,** 87–96 (2014).

14.   Martin, P. *et al.* Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nat. Commun.* **6,** 1–7 (2015).

15.   Jäger, R. *et al.* Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat. Commun.* **6,** 6178 (2015).

16.   Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489,** 83–90 (2012).

17.   Jolma, A. *et al.* DNA-binding specificities of human transcription factors. *Cell* **152,** 327–339 (2013).

18.   Yan, J. *et al.* Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. *Cell* **154,** 801–813 (2013).

19.   Butter, F. *et al.* Proteome-Wide Analysis of Disease-Associated SNPs That Show Allele-Specific Transcription Factor Binding. *PLoS Genet.* **8,** (2012).

20.   Viturawong, T., Meissner, F., Butter, F. & Mann, M. A DNA-Centric Protein Interaction Map of Ultraconserved Elements Reveals Contribution of Transcription Factor Binding Hubs to Conservation. *Cell Rep.* **5,** 531–545 (2013).

21.   Raptis, S. *et al.* MLH1 -93G>A promoter polymorphism and the risk of microsatellite-unstable colorectal cancer. *J. Natl. Cancer Inst.* **99,** 463–474 (2007).

22.   Allan, J. M. *et al.* MLH1 -93G>A promoter polymorphism and risk of mismatch repair deficient colorectal cancer. *Int. J. Cancer* **123,** 2456–2459 (2008).

23.   Poplawski, T., Sobczuk, A., Sarnik, J., Pawlowska, E. & Blasiak, J. Polymorphism of DNA mismatch repair genes in endometrial cancer. *Exp. Oncol.* **37,** 44–47 (2015).

4

24.   Rodriguez-Hernandez, I. *et al.* Analysis of DNA repair gene polymorphisms in glioblastoma. *Gene* **536,** 79–83 (2014).

25.   Lo, Y.-L. *et al.* Polymorphisms of MLH1 and MSH2 genes and the risk of lung cancer among never smokers. *Lung Cancer* **72,** 280–286 (2011).

26.   Lewis, A. *et al.* A polymorphic enhancer near GREM1 influences bowel cancer risk through differential CDX2 and TCF7L2 binding. *Cell Rep.* **8,** 983–990 (2014).

27.   Fortini, B. K. *et al.* Multiple functional risk variants in a SMAD7 enhancer implicate a colorectal cancer risk haplotype. *PLoS One* **9,** (2014).

28.   Pittman, A. M. *et al.* Allelic variation at the 8q23.3 colorectal cancer risk locus functions as a Cis-acting regulator of EIF3H. *PLoS Genet.* **6,** (2010).

29.   Lubbe, S. J. *et al.* The 14q22.2 colorectal cancer variant rs4444235 shows cis-acting regulation of BMP4. *Oncogene* **31,** 3777–3784 (2012).

30.   Chen, H. *et al.* Evidence for heritable predisposition to epigenetic silencing of MLH1. *Int. J. Cancer* **120,** 1684–1688 (2007).

31.   Perera, S., Mrkonjic, M., Rawson, J. B. & Bapat, B. Functional effects of the MLH1-93G>A polymorphism on MLH1/EPM2AIP1 promoter activity. *Oncol. Rep.* **25,** 809–815 (2011).

32.   Nakanishi, Y. *et al.* Dclk1 distinguishes between tumor and normal stem cells in the intestine. *Nat. Genet.* **45,** 98–103 (2013).

33.   Bernstein, B. E. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

34.   Miyakura, Y., Tahara, M., Lefor, A. T., Yasuda, Y. & Sugano, K. Haplotype defined by the MLH1-93G/A polymorphism is associated with MLH1 promoter hypermethylation in sporadic colorectal cancers. *BMC Res. Notes* **7,** 1–11 (2014).

35.   Funck, A. *et al.* Effect of MLH1 -93G>A on gene expression in patients with colorectal cancer. *Med. Oncol.* **31,** 160 (2014).

36.   Ma, G. *et al.* Functional annotation of colorectal cancer susceptibility loci identifies MLH1 rs1800734 associated with MSI patients. *Gut* **65,** 1227–1228 (2016).

37.   Suter, C. M., Martin, D. I. K. & Ward, R. L. Germline epimutation of MLH1 in individuals with multiple cancers. *Nat. Genet.* **36,** 497–501 (2004).

38.   Soldner, F. *et al.* Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression. *Nature* **533,** 95–99 (2016).

39.   Varjosalo, M. *et al.* The Protein Interaction Landscape of the Human CMGC Kinase Group. *Cell Rep.* **3,** 1306–1320 (2013).

40.   Liang, Q. *et al.* CDK5 is essential for TGF-β1-induced epithelial-mesenchymal transition and breast cancer progression. *Sci. Rep.* **3,** 2932 (2013).

41.   Quintavalle, M., Elia, L., Price, J. H., Heynen-Genel, S. & Courtneidge, S. A. A cell-based high-content screening assay reveals activators and inhibitors of cancer cell invasion. *Sci. Signal.* **4,** ra49 (2011).

42.   Pitts, T. M. *et al.* Association of the epithelial-to-mesenchymal transition phenotype with responsiveness to the p21-activated kinase inhibitor, PF-3758309, in colon cancer models. *Front. Pharmacol.* **4,** 35 (2013).

43.   Reiner, O. *et al.* The evolving doublecortin (DCX) superfamily. *BMC Genomics* **7,** 188 (2006).

44.   Westphalen, C. B. *et al.* Long-lived intestinal tuft cells serve as colon cancer-initiating cells. *J. Clin. Invest.* **124,** 1283–1295 (2014).

45.   Vermeulen, M. Identifying chromatin readers using a SILAC-based histone peptide pull-down approach. *Methods Enzym.* **512,** 137–160 (2012).

46.   Hubner, N. C., Nguyen, L. N., Hornig, N. C. & Stunnenberg, H. G. A quantitative proteomics tool to identify DNA-protein interactions in primary cells or blood. *J. Proteome Res.* **14,** 1315–1329 (2015).

47.   Saeed, S. *et al.* Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science (80-.).* **345,** 1251086 (2014).

48.   Lara-Astiaso, D. *et al.* Chromatin state dynamics during blood formation. *Science (80-.).* **345,** 943–949 (2014).

49.   Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* **11,** R119 (2010).

50.   Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8,** 2281–2308 (2013).

51.   Joshi, O. K. *et al.* Dynamic reorganization of Extremely Long-Range promoter-promoter Interactions between two states of pluripotency. *Cell Stem Cell* **17,** 748–757 (2015).

52.   Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–1372 (2008).

53.   Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8,** 118–127 (2007).

54.   Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13,** 731–740 (2016).

55.   Li, H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30,** 2843–2851 (2014).

56.   Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).

57.   Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29,** 24–26 (2011).

58.   Marco-Sola, S., Sammeth, M., Guigó, R. & Ribeca, P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat. Methods* **9,** 1185–1188 (2012).

59.   Anders, S., Pyl, P. T. & Huber, W. HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2014).

60.   Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).

61.   Ongen, H. *et al.* Putative cis-regulatory drivers in colorectal cancer. *Nature* **512,** 87–90 (2014).

62.   Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7,** 500–507 (2012).

63.   Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32,** 1479–1485 (2016).

64.   Thongjuea, S., Stadhouders, R., Grosveld, F. G., Soler, E. & Lenhard, B. r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic Acids Res.* **41,** e132 (2013).

65.   Mi, H., Guo, N., Kejariwal, A. & Thomas, P. D. PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. *Nucleic Acids Res.* **35,** D247-252 (2007).

66.   Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102,** 15545–15550 (2005).

67.   Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1,** 417–425 (2015).

4

Acknowledgements

Author contributions

N.Q.L., N.C.H. and H.G.S. conceived and designed this study. N.Q.L., M.t.H., L.N.N., T.P., and O.J. performed the proteomics, genomics and cell biology experiments. N.Q.L., S.Y.W., J.B.S., H.O., E.T.D., R.S.H., and N.C.H. carried out statistical and bioinformatics analyses. J.B.B. and C.L.A. provided the SYSCOL RNA-seq data, and J.Y. and J.T. provided some sequencing data of the LoVo cell line for this study. H.G.S. supervised the whole study. N.Q.L. and H.G.S. drafted the manuscript, which was reviewed by all the authors.

Competing financial interests

The authors declare no competing financial interests.

Supplementary data
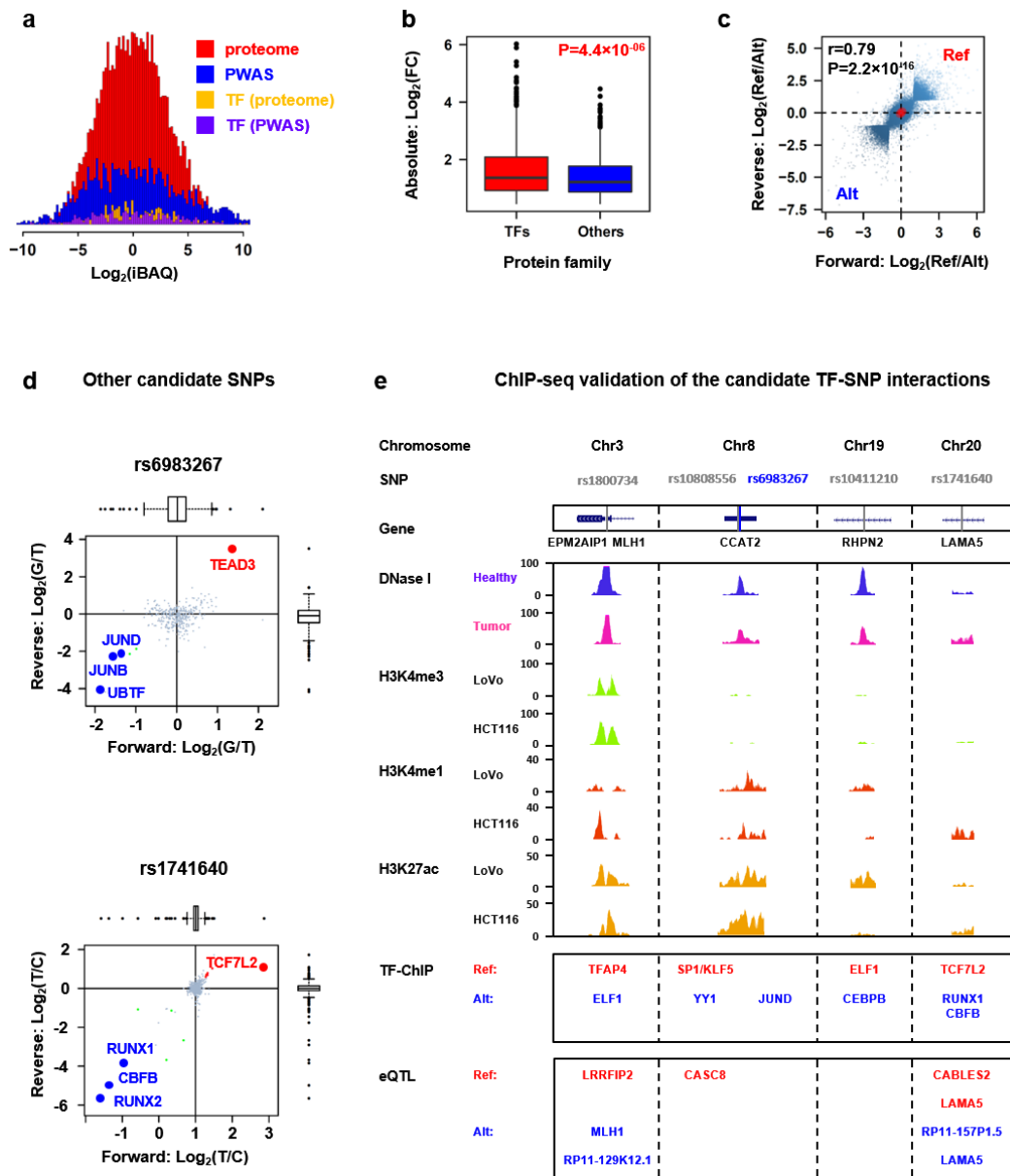
Supplementary Fig. 1 Summary of PWAS screen and ChIP-seq validation. (**a**) A histogram shows number and abundance distribution of identified proteins and TFs in our screen. We could clea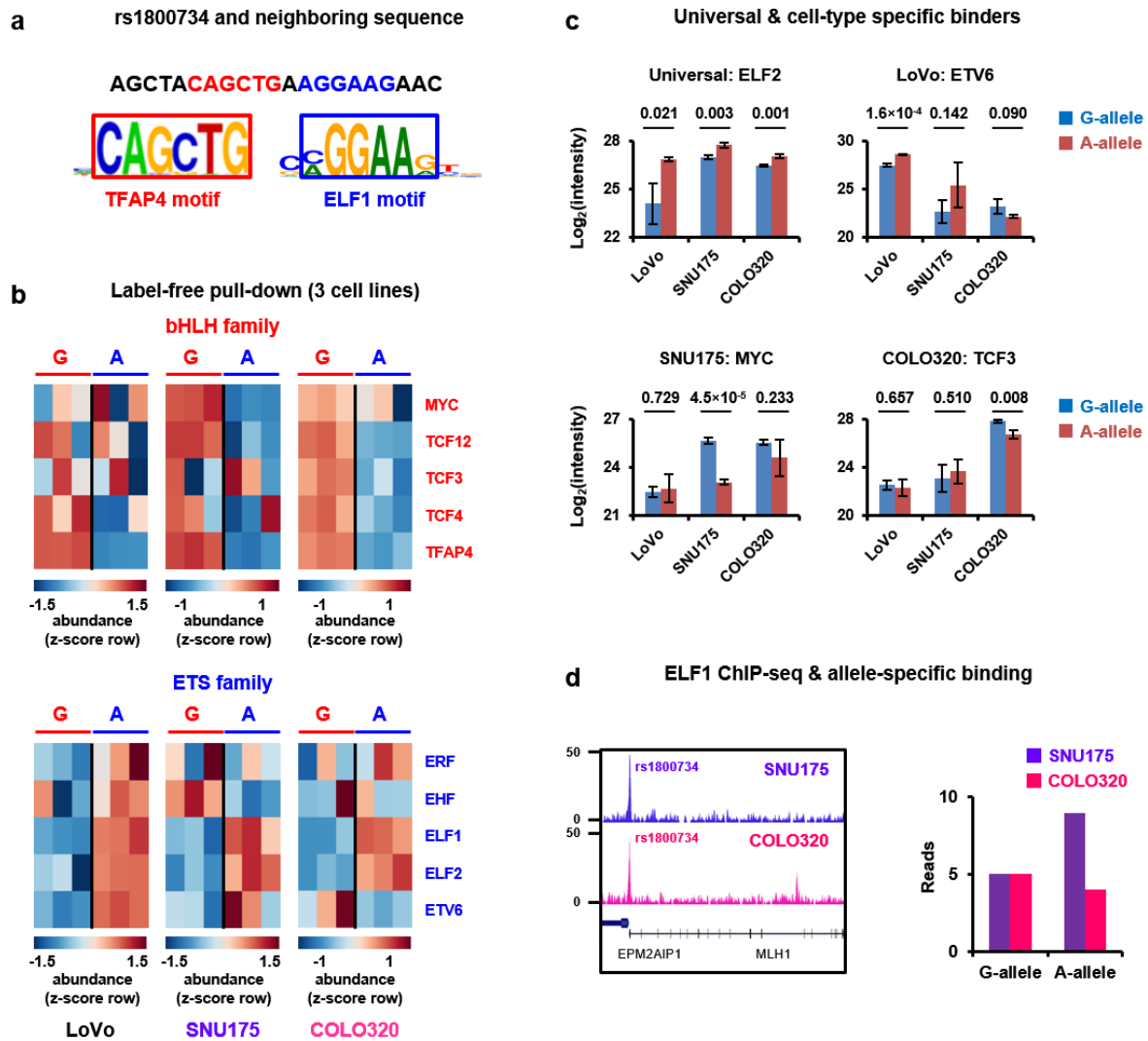rly observe a TF enrichment in our PWAS data compared to nuclear extract proteome from the LoVo cell line; (**b**) Affinity changes in respect to the TFs and other proteins. In general, the TF interactors showed higher allele preference than other proteins (P-values: Mann-whitney U test); (**c**) Overlay of the 116 PWAS experiments. The red density dot indicates the central cloud of the PWAS screen (P-values: Spearman correlation test); (**d**) Other candidate TF-SNP interactions (n=2 pull-downs per SNP, red and blue dots: P-values<0.01, A/B significance test); (**e**) ChIP-seq validated TF-SNP interactions, and the chromatin environment (DNase hypersensitivity and histone markers) and eQTL information of these SNPs. The eQTL information was obtained from GTEx Portal website (http://www.gtexportal.org/home/).

a rs1800734 and neighboring sequence

AGCTACAGCTGAAGGAAGAAC

TFAP4 motif   ELF1 motif

b Label-free pull-down (3 cell lines)

bHLH family

G A   G A   G A

MYC
TCF12
TCF3
TCF4
TFAP4

-1.5 1.5   -1 1   -1 1
abundance   abundance   abundance
(z-score row) (z-score row) (z-score row)

ETS family

G A   G A   G A

ERF
EHF
ELF1
ELF2
ETV6

-1.5 1.5   -1.5 1.5   -1 1
abundance   abundance   abundance
(z-score row) (z-score row) (z-score row)

LoVo   SNU175   COLO320

c Universal & cell-type specific binders

Universal: ELF2
0.021 0.003 0.001

LoVo: ETV6
1.6×10⁻⁴ 0.142 0.090

SNU175: MYC
0.729 4.5×10⁻⁵ 0.233

COLO320: TCF3
0.657 0.510 0.008

G-allele
A-allele

d ELF1 ChIP-seq & allele-specific binding

rs1800734   SNU175
rs1800734   COLO320
EPM2AIP1   MLH1

SNU175
COLO320

Reads
G-allele A-allele

Supplementary Fig. 2 Identification of allele-specific TF binding at the rs1800734 loci. (**a**) Motif searching revealed bHLH and ETS motifs around the rs1800734 locus; (**b**) Significant allele-specific bHLH and ETS interactors of rs1800734 identified from the LoVo, SNU175, and COLO320 cell lines. Some of these interactors displayed cell-type specificity; (**c**) Examples of universal and cell-type specific interactors (Data are represented as mean and error bars indicate s.d., n=3 pull-downs per SNP, P-values: Student's t-test); (**d**) ELF1 ChIP-seq showed a stronger binding site at rs1800734. However, no clear allele-specific binding was observed, which might be caused by ETS family binding competition at rs1800734.
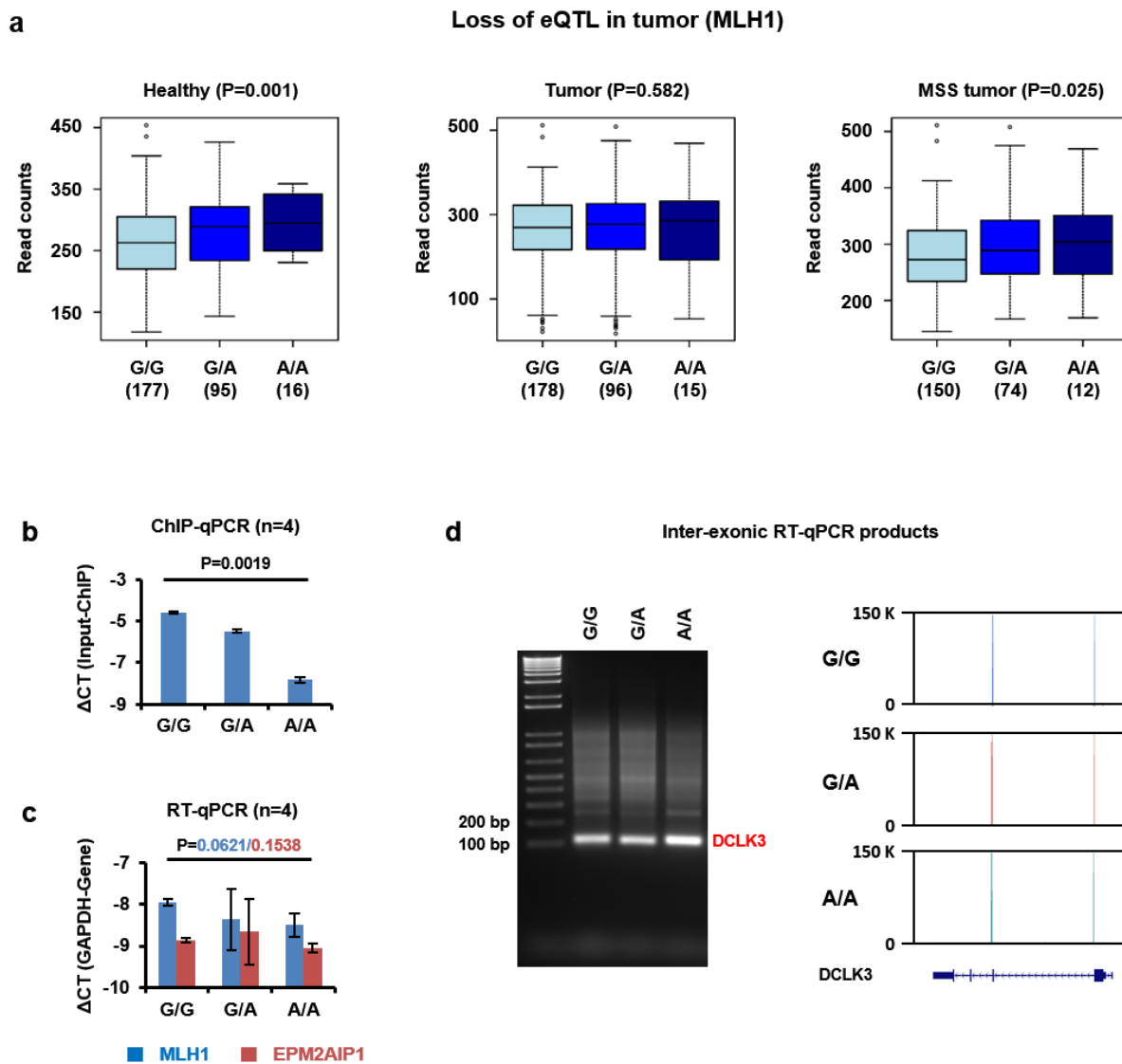
Supplementary Fig. 3 Potential gene targets of the rs1800734. (**a**) eQTL analysis showed a significant association between rs1800734 and the MLH1 expression in the healthy samples, which was lost in the tumor samples (P-values were calculated in FastQTL software based on Pearson correlation); (**b**) ChIP-qPCR analysis confirmed TFAP4 as a specific binder for the G-allele of rs1800734 (Data are represented as mean and error bars indicate s.d., n=4 ChIP assays per cell line, P-values: Student's t-test); (**c**) RT-qPCR analysis confirmed no expression change of MLH1 and EPM2AIP1 between the three isogenic cell lines (Data are represented as mean and error bars indicate s.d., n=4 biological replicates per cell line, P-values: Student's t-test); (**d**) RNA-seq of DCLK3 qPCR products confirmed the expression of this gene in these isogenic cell lines.

Chapter 5

General discussion

Over the past decades ESCs have been subject of intense investigation. ESCs are not only one of the favored systems to model mammalian embryology but they also share some striking characteristics with tumor cells. The results presented in this thesis will therefore not only fundamentally change our view on G1-phase progression in the early embryo but will also provide new perspectives on how changes in signaling networks and TF binding contribute to cell proliferation, perturbation of differentiation and cancer development.

The Cell Cycle in Two Distinct States of Pluripotency

During close to four decades of research on in vitro-grown ESCs scientist have gained a vast amount of knowledge on the molecular pathways that ensure maintenance of pluripotency. Almost one decade ago this led to the development of a protocol that allowed scientists to propagate ESCs by solely shielding them from differentiation cues[1]. One of the impeded pathways, the FGF/ERK signaling pathway, plays a key role in epiblast and primitive endoderm specification in the ICM[2]. In ground state conditions the ERK pathway is abrogated and, combined with GSK3 inhibition, results in the propagation of "ground state ESCs" that are devoid of DNA methylation similar to epiblast cells in the early blastocyst[3,4]. Because of amongst others the higher expression of lineage specifiers it has been postulated that classically cultured serum ESCs (so called "naive ESCs") reflect later stage epiblast cells[5–7]. Another striking difference was the elevated expression of CDKi's in ground state ESCs[5]. These observations seemed to contradict the prevailing model in which the absence of CDKi's ensured rapid progression through the G1-phase.

In **chapter 1** we set out to determine what causes these differences in expression of cell cycle regulators and what the consequences are. To obtain a global view on how the cell cycle differed between naive and ground state ESCs we first determined the distribution of ESCs throughout the cell cycle. A clear increase in the number of cells in G1-phase was observed upon adaptations to ground state conditions already within 24 hours. We also observed elevated CDKi expression levels. Together these two observations lead us to believe that the increased expression of CDKi's slows down G1-phase progression. Notably, these changes in cell cycle are fully reversible as previously reported for gene expression and DNA methylation. The construction of an ESC line expressing cell cycle phase-specific reporters (termed "FUCCI ESCs") and allowed us to sort and investigate cell cycle phase-specific populations. Cell sorting, RNA and protein analysis of G1-phase cells confirmed that genes involved in S-phase entry are higher expressed in naive ESCs whereas several CDKi's are upregulated in ground state ESCs. To investigate the role of the CDKi's in driving the observed differences in cell cycle, we made use of genetic engineering to delete a number of CDKi's, P16, P21 and P27. The combined deletion of the CDKi's P21 and P27 was required to partially revert the changes in cell cycle that occurs upon adaptation to ground state conditions indicating that these proteins prevent CDK/Cyclin-mediated phosphorylation of RB. The fact that deletion of P21 and P27, both well-known inhibitors of CYCLIN E (reviewed by Foster, D. A. et al.[8]), enhanced G1-progression suggests that the elongated G1-phase in ground state ESCs is not the result of lack of CDK/cyclin activity but the

result of active suppression of these complexes by CDKi's. Because CDKi's are higher expressed in the ICM when compared to the epiblast (in contrast to CDKs) these findings suggest that the progression through G1-phase is also slower in the ICM. Because the downstream RB family proteins were supposedly inactive in serum ESCs we next assessed the expression of RB throughout the cell cycle[9]. The use of the FUCCI ESCs clearly showed that hypo-phosphorylated and thus active RB is present during G1-phase in ground state ESCs implying that G1-phase progression is actively controlled by the CDKi/CDK/Cyclin and RB/E2F pathways. These findings were confirmed using RB/P107/P130 Triple Knock Out (TKO) ESCs. To determine what abrogates the RB/E2F pathway in naive ESCs we tested both inhibitors individually and observed that blocking the ERK pathway reinstates the control of G1-phase progression. FGF signaling in the ICM is mediated via two receptors, FGFR1 and FGFR2, and dictates either epiblast or primitive endoderm specification[2]. It would be interesting to investigate which receptor is responsible for downstream RB phosphorylation and abrogation of G1-phase and whether this is essential for lineage specification. The balance between FGFR1 and FGFR2 expression might also explain the observation that the degree to which the number of cells in G1-phase increased upon adaptation to ground state conditions varied between different ESC lines. The finding that ERK-signaling drives G1-phase progression in naive serum but not in ground state 2i ESCs taken together with previous observations that FGF-signaling and cell proliferation are upregulated during later stages of embryonic development supports the model in which ground state 2i ESCs reflect ICM cells and naive ESCs are more similar to epiblast cells from a later embryonic stage[6]. Surprisingly, the length of the cell cycle in naive serum and ground state 2i ESCs was similar (12-13hrs), which could be due to the elongation of S-/G2-phase upon adaptation of ground state ESCs to naive serum conditions. We have also observed shortening of G1-phase upon FGF4/Activin A-driven differentiation of ground state ESCs into EPIblast Stem Cells (EpiSC). Although doubling times observed in vivo (~4.4 hr) were significantly shorter than those in vitro of either ground state or naive ESCs we did not assess the doubling time of EpiSCs. Possibly the differentiation of ground state ESCs to EpiSCs is not accompanied by elongation of S-/G2-phase and results in doubling times close to 4.4 hours. It would be interesting to do these measurements in order to draw a comprehensive model of G1-phase progression and cell proliferation of ESCs in the ICM and during epiblast formation.

The discovery of an active RB/E2F pathway in ground state ESCs made us curious whether this pathway could be of importance during early embryonic development. Although cells in the early embryo proliferate continuously under normal circumstances, stress to the mother results in diapause, a biological response that delays implantation of the embryo and is characterized by a severe reduction in proliferation of the ICM cells. Although it has to our knowledge not been determined in which phase of the cell cycle the ICM cells in diapaused embryos reside, in at least one out of two in vitro systems that mimic diapause ESCs reside in G1-phase[10]. Using this system, we have shown that the RB/E2F pathway regulates in vitro diapause.

Altogether these findings provide evidence that shortened G1-phase is not an intrinsic property of pluripotent ESCs. Moreover, they indicate that pluripotency can co-occur with G1-phase stalling nor is compromised by elongation of G1-phase. These findings are in line with a previous study that shows that

5

artificial elongation of G1-phase of naive ESCs does not result in differentiation[11]. From these experiments it can however not be excluded that the absence of differentiation signals is a prerequisite for maintenance of pluripotency while elongating G1-phase. Whether an extremely short G1-phase can prevent differentiation is not entirely clear. In order to proof that a short G1-phase can prevent differentiation and ensures maintenance of pluripotency one would have to promote G1-phase progression while inducing differentiation. Such experiments have been performed[12], however, considering the fact that proteins that positively promote cell cycle progression have a direct effect on the pluripotency network the interpretation of these experiments remains ambiguous[13].

The discovery of an RB-mediated active G1-control in ground state ESCs raises several important questions. RB was initially discovered as tumor suppressor gene and subsequent research has attributed a central role to the pocket proteins in preventing unrestricted G1-phase progression to protect genomic integrity[14,15]. Conform this model, loss of RB in ESCs results in genomic integrity[16]. The difference between naive and ground state ESCs therefore implies that, in contrast to ground state ESCs, naive ESCs are unable to elongate G1-phase in order to repair damaged DNA during G1-phase. Indeed, previous studies have shown that the DNA damage response pathway is compromised in naive ESCs and that they do not enter G1-arrest[17]. Whether ground state ESCs cope differently with DNA damage during G1-phase remains however to be determined. The outcome of such studies could have important implications for the clinical application of ESCs in organoid formation and tissue replacement. Besides their role in constricting G1-phase progression, the pocket proteins have been shown to regulate differentiation and the expression of these proteins in highly pluripotent ESCs is therefore counter intuitive[18]. Although colony formation was not affected in RB/P107/P130 Triple Knock Out (TKO) ESCs and we did not find differences in the expression of genes the core pluripotency network, it would be interesting to study the transcriptional program of ground state TKO ESCs in more detail.

Besides the unexpected presence of a G1-checkpoint in ESCs, the notion that ground state 2i ESCs in late G1-phase express hypo-phosphorylated RB is also interesting from a fundamental biological point of view. In the current widely accepted model passage through the R point is initiated by mitogenic stimuli that induce CYCLIN D-mediated mono-phosphorylation of RB and subsequent hyper-phosphorylation by CDK2/CYCLIN E[19]. Strikingly, our results suggest that hyper-phosphorylation of RB is not needed in ESCs for them to passage the R point and is not even a prerequisite for 2i ESCs to enter S-phase. The conventional western blot technique employed in our study does however not allow to discriminate between un-phosphorylated and hypo-phosphorylated RB and it would be interesting to determine to what extent RB is phosphorylated in late G1-phase ground state 2i ESCs using phospho-specific antibodies or more sensitive technique like 2-Dimensional IsoElectric Focussing (2D-IEF)[19]. In addition, the use of genetically engineered constitutively active RB could be used to assess whether phosphorylation is a required for 2i ESCs to enter S-phase. In line with previous studies our results show that inhibition of MYC results in

reversible G1-arrest in serum-free conditions[10]. Although MYC acts as a transcriptional activator of G1 cyclins the absence of hyper-phosphorylated RB in ground state conditions suggests that MYC-mediated S-phase entry is independent of CDK/cyclin activity. These findings are in line with a previous report that indicated that MYC controls cell cycle commitment whereas the CDK/cyclin complexes tune the length of G1-phase[20]. In naive serum conditions, depletion of MYC does not result in G1-arrest of ESCs (personal observations and Hölzel, M. *et al.*[21]) indicating that other stimuli, presumably ERK-mediated cell cycle control, overcome the absence of MYC. Together our results suggest a model in which the R points is controlled by the MYC and the RB/E2F pathway. In addition, serum components overcome this checkpoint due to ERK-mediated activation of CDK/cyclin complexes. Because of its serum-free nature and the ability to reversibly arrest immortal karyotypically normal cells in G1-phase we believe that the 2i system could serve as an attractive system to study the fundamentals of G1-phase progression and the R point.

The Role of P53 in Ground State ESCs

Naive serum ESCs progress rapidly through the G1-phase and have lost cell cycle checkpoints that allow them to stall in G1-phase upon stress. P53 plays a key role in inducing G1-arrest and it had been reported that the P53/P21 pathway was inactive in naive ESCs. The observed increase in P21 expression in 2i ESCs and our observations described in chapter 1 made us hypothesize that the P53/P21 pathway is activated in ESCs upon adaptation to ground state 2i conditions.

In **chapter 2** we describe the experiments we did to test this hypothesis. We reasoned that P53 was a driving factor behind P21/P27-mediated elongation of G1-phase and therefore deleted P53 in the FUCCI ESCs. We indeed observed lowered expression of P21 and a shortening of G1-phase in ground state P53[-/-] ESCs when compared to WT. Surprisingly, whole genome RNA-seq revealed that the expression of P21 but not P27 is affected upon deletion of P53. However, the expression of both Rb1 and Rbl2 (coding for RB and P130, respectively) is significantly lower in P53KO ESCs. ChIP-seq data on P53 revealed that P53 binds the Rb1 promoter and is therefore likely to directly control Rb1 expression. Despite the difference in the impact of P53 loss on naive serum and ground state 2i ESCs, the RNA-seq data clearly shows that the expression of target genes involved in G1-arrest is affected in both conditions. Considering the fact that active ERK signaling renders the pocket proteins inactive in naive serum ESCs but not in ground state 2i ESCs the lowered expression of Rb1 and Rbl2 could explain the discrepancy in the effect of loss of P53 between naive and ground state ESCs.

Whereas pathways involved in apoptosis and G1-arrest were affected in both naive and ground state ESCs upon loss of P53, genes involved in developmental processes were downregulated solely in ground state conditions. Although P53 is highly expressed in the early mouse epiblast and loss of P53 results in severe developmental defects the molecular mechanism behind its function is not entirely clear (reviewed by Jain, A. K. & Barton, M. C.[22]). A possible explanation for the differences between naive and ground state conditions are the distinct signaling pathways that are activated in both conditions. In ground state 2i ESCs,

5

enhanced Wnt signaling is critical and β-catenin relieves the pluripotency network from a repressive complex[23]. Interestingly, P53 has shown to govern a Wnt/β-catenin-mediated transcriptional program that specifies the mesoderm lineage[24]. These findings imply that the elevated Wnt signaling in ground state ESCs results in P53-mediated lineage priming.

Two recent studies have shown that the loss of DNA methylation in ground state 2i conditions also extends to Imprinting Control Regions (ICRs) and results in genomic instability[25,26]. It is tempting to speculate that global DNA demethylation results in genomic instability and leads to P53 activation in ground state 2i ESCs. Another interesting possibility is that P53 regulates the expression of developmental genes through DNA methylation. A very recent study has provided evidence that P53 is crucial for DNA methylation homeostasis in ESCs[27]. Interestingly, our RNA-seq analysis revealed that the expression of DNMT1 was severely reduced in P53$^{-/-}$ ESCs in ground state conditions. Since DNMT1, but not DNMT3A and DNMT3B, is expressed in the pre-implantation embryo and is required for the maintenance of DNA methylation at imprinted loci[28] these results indicate that loss of P53 in ground state ESCs might affect DNA methylation at those regions. Interestingly like loss of P53, reduced DNA methylation at imprinted region results in developmental defects[29]. Naive ESCs display higher DNA methylation levels due to higher expression of DNMT3A and DNMT3B and might therefore remain unaffected by the loss of DNMT1 [4]. In the early blastocyst the DNA methylation is actively erased and high P53 expression might therefore be crucial for preservation of DNA methylation at ICRs[30,31]. The elucidation of the role of P53 in preserving DNA methylation at ICRs in the early blastocyst may be an attractive line of investigation. Together, these data indicate that P53 is essential in early embryonic development and suggest that ground state ESCs are better suited to study the role of P53 during development when compared to naive ESCs.

Pluripotency and the Cell Cycle

Focusing on in vitro serum-grown ESCs, several studies revealed that the ESC cell cycle structure was characterized by a short G1-phase. Together with studies that indicated that ESCs are more sensitive to fate specification during G1-phase, these findings led scientists to hypothesize that the short G1-phase is an intrinsic property of ESCs and is ensured by the ESC-specific transcriptional program to prevent differentiation (reviewed in Dalton, S.[32]). In line with this model several studies on naive serum ESCs have indicated that OCT4, SOX2 and NANOG can positively affect the CDK/cyclin activity or suppress the RB/E2F pathway and thereby contribute to the short G1-phase. The data presented in the first two chapters have however fundamentally changed our perspective on the ESC-specific cell cycle. In **chapter 3** we have reviewed whether and how members of the core pluripotency network modulate the unusual cell cycle structure of ESCs. When comparing naive serum to ground state 2i ESCs it is clear that the combined activity of ERK signaling and the lack of TCF1-driven Wnt signaling largely drives the short G1-phase of naive serum ESCs[33,34]. Since ERK signaling drives lineage specification and Wnt signaling through TCF1 had no effect on pluripotency these findings indicate that ESCs completely shielded from differentiation

cues do have an intact G1-phase control pathway. Despite the elevated levels of CDKi's and the presence of hypo-phosphorylated RB, ground state ESCs proliferate fast. What exactly causes rapid proliferation needs to be determined but the high dependence on the microRNA-processing machinery suggests that microRNA-mediated downregulation of the pocket proteins plays a pivotal role[35]. MicroRNAs play a crucial role during development and mature oocytes that are depleted from microRNAs fail to progress through the initial cell divisions[36,37]. Possibly, G1-phase progression in the early embryo is driven by OCT4/SOX2-induced microRNAs expression whereas CDK/cyclin-mediated phosphorylation of RB takes over during later stages[38,39].

The pluripotency network does not only affect the cell cycle, but recent research has highlighted a role of cell cycle regulators in pluripotency and differentiation. Evidence that TFs from the pluripotency network are stabilized by cell cycle regulators and that upon loss of the latter ICM cells differentiate indicates that cell cycle regulators are essential for maintenance of pluripotency. Differentiation on the other hand correlates with the activity of cell cycle inhibitors and several lines of evidence indicate that the latter can inhibit the pluripotency network.

ESCs and Cancer

Both naive ESCs and tumor cells are immortal and proliferate uncontrollably. The latter have often acquired mutations that either result in the enhanced activity of proteins that drive G1-phase progression or prevents the tumor cell from entering G1 arrest[40]. Similarly, proteins that mediate G1-phase progression are constitutively activated in ESCs whereas pathways that induce G1-arrest are abolished[41]. In line with these data, the results presented in this thesis suggest that the G1-phase of tumor cells that proliferate uncontrollably due to constitutively activated ERK-signaling resembles that of naive ESCs (**chapter 1** and Malumbres, M. & Barbacid, M.[42]). The use of the two small molecule inhibitors (2i) might therefore restore the RB/E2F pathway in tumor cells and restrict proliferation. Moreover, our findings suggest that P53 is not able to elicit its function in naive ESCs where CDK/Cyclin activity is high. Because P53 is the most frequently mutated gene in cancer these findings might have important implications in cancer. Our findings described in **chapter 2** indicate that constitutive ERK signaling impacts on downstream targets of P53 leading to the inability of ESCs to engage a proper P53-mediated G1 checkpoint. It would therefore be interesting to study regulators of the ERK-signaling pathway in tumor cells that do not undergo P53-mediated cell cycle arrest. Furthermore, we have identified RB as a direct transcriptional target for P53. Although P53 is in general thought to prevent genomic instability via P21-mediated cell cycle arrest, our results suggest that P53 may prevent tumorigenesis by directly upregulating the expression RB. It would be worthwhile to assess whether the induced expression of RB can lower the tumorigenicity of P53$^{-/-}$ tumors. Recently developed CRISPR-based methods to induce expression of RB could possibly be employed to treat loss of P53-driven tumors.

5

Although these findings assume that their unrestricted proliferative nature underlies the tumorigenic phenotype of cancer cells, accumulating evidence over the past decades has indicated that tumorigenicity is often the result of flawed differentiation and that the pluripotent ESC-like transcriptional profile is often found in poorly differentiated tumors[43]. Regarding the fact that the cell cycle machinery is able to directly impact on the pluripotent phenotype of cells (**chapter 3**) it might be feasible to target both these features of cancer cells. Blocking certain CDK/cyclin complexes could for instance result in simultaneous lowering of proliferation and degradation of pluripotency factors. The elucidation of pathway that link the cell cycle to pluripotency might therefore identify new therapeutic targets in cancer treatment.

Cancer is often driven by mutations that lead to aberrant signaling and upregulated transcription of genes that promote cell cycle progression. These mutations do not only occur in coding regions but are often found in regulatory regions like enhancers. Similarly, Single Nucleotide Polymorphisms (SNPs) at enhancer sites have shown to affect TF binding and gene expression[44]. Genome-Wide Association Studies (GWAS) have identified SNPs that are associated with increased susceptibility of cancer within enhancers regions[45,46]. It is expected that these SNPs modulate TF binding and gene expression and thereby increase the risk for cancer. In **chapter 4** we set out to study how SNPs within TF binding sites could contribute to ColoRectal Cancer (CRC). We have performed an unbiased Proteome-Wide Analysis of disease-associated SNPs (PWAS) on 116 SNPs associated with CRC from publicly available GWAS databases. Our approach identified several TFs that showed preferential binding to the alternative allele over the reference allele. The replacement of the Adenine with a Guanine at rs1800734 resulted in an increased binding affinity of TFAP4 whereas ETS-like factors showed preferred binding to the Adenine-(A-)Allele. Subsequent experiments revealed that the A-allele of rs1800734 correlated with a higher expression of DCLK3 and that this was mediated by a long-range interaction between rs1800734 and the promoter of DCLK3. Interestingly, RNA-seq data from a cohort study revealed that the expression of DCLK3 correlated with the expression of genes involved in Epithelial-to-Mesenchymal Transition (EMT), suggesting that DCLK3 is an important regulator of EMT. Together these finding led us to propose a model in which the G-allele of rs1800734 results in binding of ETS-factors that drive the expression DCLK3 through a long range genomic interaction and that the elevated level of DCLK3 promotes EMT and tumor initiation.

References

1.   Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–23 (2008).

2.   Kang, M., Garg, V. & Hadjantonakis, A. K. Lineage Establishment and Progression within the Inner Cell Mass of the Mouse Blastocyst Requires FGFR1 and FGFR2. *Dev. Cell* **41,** 496–510.e5 (2017).

3.   Leitch, H. G. *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20,** 311–316 (2013).

4.   Habibi, E. *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13,** 360–9 (2013).

5.   Marks, H. *et al.* The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. *Cell* **149,** 590–604 (2012).

6.   Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4,** 487–92 (2009).

7.   Joshi, O. K. *et al.* Dynamic reorganization of Extremely Long-Range promoter-promoter Interactions between two states of pluripotency. *Cell Stem Cell* **17,** 748–757 (2015).

8.   Foster, D. A., Yellen, P., Xu, L. & Saqcena, M. Regulation of G1 cell cycle progression: Distinguishing the restriction point from a nutrient-sensing cell growth checkpoint(s). *Genes and Cancer* **1,** 1124–1131 (2011).

9.   Savatier, P., Huang, S., Szekely, L., Wiman, K. G. & Samarut, J. Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. *Oncogene* **9,** 809–818 (1994).

10.   Scognamiglio, R. *et al.* Myc Depletion Induces a Pluripotent Dormant State Mimicking Diapause. *Cell* 668–680 (2016). doi:10.1016/j.cell.2015.12.033

11.   Li, V. C., Ballabeni, A. & Kirschner, M. W. Gap 1 phase length and mouse embryonic stem cell self-renewal. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 12550–5 (2012).

12.   Coronado, D. A short G1 phase is an intrinsic determinant of naïve embryonic stem cell pluripotency. *Stem Cell Res.* **10,** 118–31 (2013).

13.   Liu, L. *et al.* G1 cyclins link proliferation, pluripotency and differentiation of embryonic stem cells. *Nat. Cell Biol.* **19,** (2017).

14.   Nevins, J. R. The Rb/E2F pathway and cancer. *Hum. Mol. Genet.* **10,** 699–703 (2001).

15.   Dick, F. A. & Rubin, S. M. Molecular mechanisms underlying RB protein function. *Nat. Rev. Mol. Cell Biol.* **14,** 297–306 (2013).

16.   Zheng, L., Flesken-Nikitin, A., Chen, P. L. & Lee, W. H. Deficiency of Retinoblastoma gene in mouse embryonic stem cells leads to genetic instability. *Cancer Res.* **62,** 2498–2502 (2002).

17.   Suvorova, I. I., Grigorash, B. B., Chuykin, I. A., Pospelova, T. V & Pospelov, V. A. G1 checkpoint is compromised in mouse ESCs due to functional uncoupling of p53-p21Waf1 signaling. *Cell Cycle* **15,** 52–63 (2016).

18.   Sage, J. The retinoblastoma tumor suppressor and stem cell biology. *Genes Dev.* 1409–1420 (2012). doi:10.1101/gad.193730.112.control

19.   Narasimha, A. M. *et al.* Cyclin D activates the Rb tumor suppressor by mono-phosphorylation. 1–21 (2014). doi:10.7554/eLife.02872

20.   Dong, P. *et al.* Division of labour between Myc and G1 cyclins in cell cycle commitment and pace control. *Nat. Commun.* **5,** 4750 (2014).

21.   Hölzel, M. *et al.* Myc/Max/Mad regulate the frequency but not the duration of productive cell cycles. *EMBO Rep.* **2,** 1125–1132 (2001).

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

23.   Marks, H. & Stunnenberg, H. G. Transcription regulation and chromatin structure in the pluripotent ground state. *Biochim. Biophys. Acta* 1–9 (2013). doi:10.1016/j.bbagrm.2013.09.005

24.   Wang, Q. *et al.* The p53 Family Coordinates Wnt and Nodal Inputs in Mesendodermal Differentiation of Embryonic Stem Cells. *Cell Stem Cell* 1–17 (2016). doi:10.1016/j.stem.2016.10.002

25.   Choi, J. *et al.* Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature* **548,** 219–223 (2017).

26.   Yagi, M. *et al.* Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. *Nature* **548,** 224–227 (2017).

27.   Tovy, A. *et al.* p53 is essential for DNA methylation homeostasis in naïve embryonic stem cells , and its loss promotes clonal heterogeneity. *Genes Dev.* 1–14 (2017). doi:10.1101/gad.299198.117.5

28.   Hirasawa, R. *et al.* Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev.* **22,** 1607–1616 (2008).

5

29.   Plasschaert, R. N. & Bartolomei, M. S. Genomic imprinting in development , growth , behavior and stem cells. 1805–1813 (2014). doi:10.1242/dev.101428

30.   Wang, L. *et al.* Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157,** 979–91 (2014).

31.   Wu, H. & Zhang, Y. Reversing DNA methylation: Mechanisms, genomics, and biological functions. *Cell* **156,** 45–68 (2014).

32.   Dalton, S. Linking the Cell Cycle to Cell Fate Decisions. *Trends Cell Biol.* **25,** 592–600 (2015).

33.   De Jaime-Soguero, A. *et al.* Wnt/Tcf1 pathway restricts embryonic stem cell cycle through activation of the Ink4/Arf locus. *PLoS Genet.* **13,** (2017).

34.   ter Huurne, M., Chappell, J., Dalton, S. & Stunnenberg, H. G. Distinct Cell-Cycle Control in Two Different States of Mouse Pluripotency. *Cell Stem Cell* **21,** 449–455.e4 (2017).

35.   Yan, Y. *et al.* Significant differences of function and expression of microRNAs between ground state and serum-cultured pluripotent stem cells. *J. Genet. Genomics* **44,** 179–189 (2017).

36.   Svoboda, P. & Flemr, M. The role of miRNAs and endogenous siRNAs in maternal-to-zygotic reprogramming and the establishment of pluripotency. **11,** (2010).

37.   Tang, F. *et al.* Maternal microRNAs are essential for mouse zygotic development. 644–648 (2007). doi:10.1101/gad.418707.644

38.   Card, D. a G. *et al.* Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells. *Mol. Cell. Biol.* **28,** 6426–38 (2008).

39.   Wang, Y. *et al.* Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat. Genet.* **40,** 1478–1483 (2008).

40.   Ho, A. & Dowdy, S. F. Regulation of G1 cell-cycle progression by oncogenes and tumor suppressor genes. *Curr. Opin. Genet. Dev.* **12,** 47–52 (2002).

41.   Kareta, M. S., Sage, J. & Wernig, M. Crosstalk between stem cell and cell cycle machineries. *Curr. Opin. Cell Biol.* **37,** 68–74 (2015).

42.   Malumbres, M. & Barbacid, M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat. Rev. Cancer* **9,** 153–166 (2009).

43.   Ben-Porath, I. *et al.* An embryonic stem cell–like gene expression signature in poorly differentiated aggressive human tumors. *Nat. Genet.* **40,** 499–507 (2008).

44.   Maurano, M. T. *et al.* Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat. Genet.* **47,** 1393–1401 (2015).

45.   Houlston, R. S. *et al.* Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat. Genet.* **42,** 973–977 (2010).

46.   Tomlinson, I. P. M. *et al.* A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.* **40,** 623–630 (2008).

Summary

At an early stage during organismal development Embryonic Stem Cells (ESCs) of the Inner Cell Mass (ICM) transiently reside in a pluripotent state, which means that they do not show any sign of cell fate specification and are able to differentiate in any desired cell type. In this thesis we have studied the cell cycle of pluripotent mouse ESCs and determined how extra-cellular signaling can affect G1-phase control. Furthermore, we have investigated how Single Nucleotide Polymorphisms (SNPs) in non-coding regions of the genome can affect gene expression and pre-disposition to cancer.

Cells of the ICM proliferate fast and classically cultured serum ESCs lack G1-phase control. In chapter 1 we have shown that ground state pluripotent ESCs deprived of ERK signaling express hypo-phosphorylated RB and have an elongated G1-phase when compared to pluripotent ESCs cultured in serum-rich conditions. The expression of the CDK-inhibitors P21 and P27 prevents phosphorylation of RB and mediates G1-phase elongation in ground state ESCs. On the other hand, we have shown that ERK signaling in naive serum ESCs results in hyper-phosphorylation of RB and fast progression through G1-phase. Although cells of the early developing embryo proliferate fast, proliferation is halted upon stressful conditions during a process termed "diapause". The G1-checkpoint observed in 2i ESCs is actually required for an in vitro condition that mimics diapause.

The reinstatement of the G1-checkpoint in ground state mouse ESCs implies that molecular pathways that control this checkpoint are active as well. An important cell cycle regulator that impacts on the CDK/Cyclin and RB/E2F pathway is P53. P53 is a well-studied tumor suppressor gene and plays a pivotal role in the DNA damage response pathway. P21 is an important target gene of P53 and one of the CDK-inhibitors up-regulated in ground state ESCs when compared to naive serum ESCs. In chapter 2 we have therefor set out to study the role of P53 in ground state ESCs. Deletion of P53 resulted in a decreased expression of P21 and shortening of G1-phase in ground state ESCs. In contrast, in naive serum ESCs hardly any effect of the loss of P53 was detected. Together these results imply that the expression of P53 in ground state ESCs is necessary for the G1-checkpoint whereas the G1-checkpoint is absent in naive serum ESCs. Indeed, the treatment with DNA damaging agents resulted in an increase in the number of cells in G1-phase in ground state 2i ESCs whereas a decrease in G1-phase cells was observed in naive ESCs. Surprisingly, transcriptome analysis revealed that the loss of P53 in both serum as well as 2i ESCs results in lowered expression of genes involved in cell cycle arrest and DNA damage response. It came to our notion that the expression of the pocket proteins RB and P130, which are inactive in serum but active in 2i ESCs, was severely reduced upon deletion of P53. These results were confirmed on protein level in cell cycle phase-sorted serum and 2i ESCs and explain the phenotype observed in 2i but not serum ESCs upon deletion of P53. Interestingly, in ground state ESCs the deletion of P53 was accompanied by a decreased expression of genes involved in developmental processes. These findings suggest a role for P53 in early

embryonic development and may explain the high expression of P53 in the developing embryo. Because such an effect was not observed in serum ESCs these results might encourage scientists to use ground state 2i ESCs to study the function of P53 in ESCs.

Collectively, the results described in chapter one and two fundamentally changed the view on cell cycle progression and the G1-checkpoint in pluripotent ESCs. In chapter 3 of this thesis we have summarized existing literature on the cell cycle of pluripotent ESCs and we have discussed how our newly acquired insights relate to these literature. In line with the highly proliferative nature of ESCs, Transcription Factors (TFs) of the pluripotency network favor the expression of CDK/Cyclin complexes and microRNAs. These promote cell cycle progression and prevent the expression of cell cycle inhibitors, respectively. The observed G1-checkpoint underscores however the fact that fast cell cycle progression is not a pre-requisite for maintenance of pluripotency. The pluripotency network stimulates cell cycle progression, but cyclins and CDKs are also able to prevent degradation of SOX2, OCT4 and NANOG as well. By doing so these complexes prevent differentiation and contribute to the undifferentiated phenotype of ESCs.

Cell cycle progression and differentiation are tightly linked and when either one of these biological processes go awry this can result in developmental defects and disease. In particular, mis-regulation of genes involved in cell cycle progression often result in unrestricted proliferation, a hallmark of cancer. Although mutations in coding regions of genes affect protein expression and often underlie the development of cancer, the aberrant regulation of gene expression leading to cancer development is often caused by mutations that do not affect coding regions but affect genomic regions that regulate genes involved in cancer development. In chapter 4 we have used a novel proteome-based method to assess how Single Nucleotide Polymorphisms within enhancer elements affect gene expression and can lead to increased susceptibility to cancer. We found that a base-pair mutation within the promoter region of MLH1, resulting in an Adenine instead of a Guanine, perturbs the binding of TFAP4 resulting in an increased expression of DCLK3 through a long range interaction. The analysis of RNA-sequencing data of healthy and tumor tissue revealed that the expression of DCLK3 correlated with Epithelial-to-Mesenchymal Transition (EMT)-related genes. Together, these findings suggest that the SNP affects TF binding, regulates EMT and thereby promotes cancer development.

Samenvatting

Vroeg tijdens de embryonale ontwikkeling van een organisme bevinden de Embryonale Stam Cellen (ESCs) van de zogeheten Inner Cell Mass (ICM) zich tijdelijk in een pluripotente staat, wat betekent dat zij geen kenmerken van cel specificatie vertonen en dat zij nog in staat zijn in elk willekeurig celtype te differentiëren. In dit manuscript hebben wij de cel cyclus van pluripotente muis ESCs bestudeerd en bepaald hoe extracellulaire signalering routes de lengte van de G1-fase controleren. Bovendien hebben we onderzocht hoe mutaties in niet-coderende delen van het genoom genexpressie beïnvloeden en hoe dit predispositie tot kanker tot gevolg kan hebben.

Cellen van de ICM prolifereren snel en ESCs gecultiveerd in serum-rijk medium (serum ESCs) missen G1-fase controle mechanismen. In hoofdstuk 1 hebben we laten zien dat zogenaamd "ground state" pluripotente ESCs waarbij ERK-signalering is platgelegd (2i ESCs) een verlengde G1-fase hebben vergeleken met serum ESCs. De expressie van de CDK-remmers P21 en P27 verhinderd fosforylering van RB en medieert verlenging van G1-fase in 2i ESCs. Bovendien hebben we laten zien dat ERK signalering in serum ESCs resulteert in hyper-fosforylering van RB en snelle progressie door de G1-fase. Hoewel cellen in het vroege embryo snel delen wordt proliferatie tot een halt geroepen tijdens stressvolle omstandigheden, een proces dat "diapauze" wordt genoemd. Het G1-checkpoint dat we observeerden in 2i ESCs is noodzakelijk voor een in vitro conditie welke diapauze nabootst.

De activatie van het G1-checkpoint in ground state ESCs impliceert dat de moleculaire signalerings mechanismen welke hierbij betrokken zijn geactiveerd worden. Een belangrijke cel cyclus regulator welke CDK/Cycline en RB/E2F signalering beïnvloedt is P53. P53 is uitgebreid bestudeerd en speelt een belangrijke rol tijdens herstel van DNA schade. P21 is een belangrijk eiwit dat wordt geactiveerd door P53 en één van de CDK-remmers opgereguleerd in 2i ESCs ten opzichte van serum ESCs. In hoofdstuk 2 hebben we de rol van P53 in 2i ESCs bestudeerd. De deletie van P53 resulteerde in verminderde expressie van P21 en een verkorting van de G1-fase in 2i ESCs. In serum ESCs had de deletie van P53 echter nauwelijks een effect. Deze resultaten impliceren dat de expressie van P53 in 2i ESCs is noodzakelijk ten behoeve van het G1-checkpoint, welke afwezig is in serum ESCs. De behandeling van 2i ESCs met chemicaliën welke DNA schade tot gevolg hebben leidde tot een toename van het aantal cellen in G1-fase. De behandeling van serum ESCs met dezelfde chemicaliën had echter geen toename van het aantal cellen in G1-fase tot gevolg. Transcriptoom analyse onthulde echter dat het verlies van P53 in zowel serum als 2i ESCs resulteert in de verminderde expressie van genen betrokken bij DNA schade herstel en cel cyclus arrest. Twee genen waarvan de expressie lager was in P53KO ESCs ten opzichte van WT ESCs waren Rb1 en Rbl1, coderend voor RB en P130. Omdat deze genen inactief zijn in serum ESCs als gevolg van fosforylatie (hoofdstuk 1) zal de verminderde expressie in P53KO serum ESCs geen effect hebben. In 2i ESCs daarentegen zijn deze eiwitten niet gefosforyleerd en dus actief, wat verklaart waarom de deletie van

P53 in 2i ESCs een drastische effect heeft op de lengte van de G1 fase. De verlaagde expressie van RB werd bevestigd door verlaagde eiwit levels in zowel G1-, S- en G2-phase P53 KO cellen ten opzichte van WT cellen. Een andere interessante observatie was de verlaagde expressie van ontwikkelings-gerelateerde genen in P53 KO ESCs in 2i condities. Dit suggereerd dat P53 een rol speelt tijdens de vroege ontwikkeling, wat een verklaring zou zijn voor de hoge expressie van P53 in het vroege embryo. Omdat dit effect niet waargenomen wordt in P53 KO ESCs in serum condities zullen deze resultaten mogelijk andere wetenschappers ertoe aansporen om 2i ESCs te gebruiken om de rol van P53 tijdens de vroege ontwikkeling te bestuderen.

De regulatie van de cell cyclus en van differentiatie zijn verbonden en wanneer er iets fout gaat in een van deze twee processen dank an dit leiden tot ontwikkelingsstoornissen dan wel ziektes. De mis-regulatie van genen betrokken bij de cel cyclus leidt vaak tot ongecontroleerde proliferatie, een kenmerk van kanker. Mutaties welke leiden tot verstoorde genexpressie vaak ten grondslag liggen aan het ontstaan van kanker, zijn deze mutaties niet altijd gelegen in genen zelf. In hoofdstuk 4 hebben we een nieuwe methode gebruikt om te bestuderen hoe SNPs in enhancer elementen genexpressie beinvloeden en hoe dit tot een verhoogde kans op kanker kan leiden. We vonden dat een base paar mutatie in de enhancer van MLH1, resulterend in een Adenine in plaats van een Guanine, leidt tot verlaagde binding van de transcriptie factor TFAP4. Bovendien was de expressive van DCLK3 verhoogd in cellijnen met de desbetreffende SNP. De analyse van RNA-seq data van gezond en tumor weefsel onthulde dat de expressie van DCLK4 correleert met die van Epithelial-to-Mesenchymal Transition (EMT)-gerelateerde genen. Tesamen suggereren deze bevindingen dat de SNP de binding van TFAP4 verhindert, EMT bevordert en zodoende de kans op het ontstaan van kanker vergroot.

Dankwoord

Het einde van mijn promotie-traject is eindelijk in zicht. Via deze wil ik graag iedereen bedanken die (in)direct bijgedragen heeft aan de totstandkoming van dit boekje, in het bijzonder:

Allereerst uiteraard Henk. Na gestopt te zijn bij reumatologie zocht ik een fundamenteel-biologisch project met de vrijheid om mijn eigen weg te vinden, en jij bood me die mogelijkheid. Jij gaf mij de voorzieningen en vrijheid om mijn project te leiden, gepaard met zo nu en dan een (goede) discussie om te voorkomen dat ik te ver afdwaalde. Toen ik eenmaal mijn draai had gevonden had en de eerste resultaten behaalde stond jij altijd klaar voor advies en een kritische blik tov mijn werk. Uiteindelijk heeft dit alles ervoor gezorgd dat mijn promotie een erg leerzame tijd was met dit proefschrift als mooie resultaat. Dank hiervoor! Veel succes en geluk in de toekomst!

The FUCCI cells have played an important role in my major paper and throughout this thesis. For the construction of these cells and help with sorting I'd like to thank our collaborators James Chappell and Stephen Dalton.

Besides the work on ESCs an important part of the thesis is focused on the role of SNPs during colorectal cancer development, for which I would like to thank you, dr. Robert. I really admire your enthousiasm and it was always a pleasure collaborating with you. Hopefully we'll find new opportunities to collaborate (or just drink a beer) in the future.

Ms. Peng, it was also nice to continue working with you on the P53 story during the last couple of months, hopefully we'll be able to get it published soon! Good luck finishing your thesis!

Graag wil ik ook prof. Ad Geurts van Kessel, prof. Gert Flik en prof. Hein te Riele bedanken voor het beoordelen van mijn manuscript en het oppositie voeren tijdens de promotiegelegenheid.

Besides those that were directly involved in the work described in this thesis I would like to thank all members of the department that helped me along the way via useful discussions, technical assistance or administrative stuff.

First of all would like to thank all members of the ESC-group. Hendrik, as founding father of the ESC-group you've played an important role and I've always appreciated your advice / critical comments on my project (besides spontaneous life lessons..). For fruitful discussions during the ESC meetings I'd also like to thank Ehsan, Guido, Jani, Onkar, René, Shuang Yin, Tianran, Yaser & -ouwe trekpop- Wout. As members of the ESC-group we shared, besides scientific interests, also the lab at the second floor. I've always really enjoyed the atmosphere at our floor, everyone was always willing to give advice, help me out on practical stuff, or available for just a quick chat. I wish you all the best for your (scientific) careers and the future in general!

Besides the ESC meetings the general lab meetings formed always good opportunities to evaluate the project and to receive critical comments and usefull suggestions. I'd like to thank all of those that actively contributed to these lab meetings.

I would also like to thank those that provided me with technical support (for sequencing Eva, Nilofar, Bowon an Martijn) or bioinformatics assistance / computer-related stuff (mn Kees-Jan en Rita).

Voor (veelal administratief) advies / vragen / werk kon ik altijd terecht bij de dames van het secretariaat. Maria, Josefien, Anita en Lidwien, dank voor jullie hulp gedurende mijn promotie. En voor general PhD student-support wil ik ook graag Marion bedanken.

Although the science itself is interesting & fun, the thing that I appreciated most from doing my PhD is meeting a lot of nice and interesting people. During the last ~6 years I've gotten to know many cool people from all over the world and I have made friends with whom I hold dear memories.

From the start of the project I was condemned to work with you, Ehsan. For some reason however, we were a perfect match! I really enjoyed working with you and I had a great time sharing a unit with you (incl. endless "discussions" on science-(un)related stuff, classes on Farsi and so on.. :)). Thanks Habibi! Hope to see you and Roya again some day.

Next, I'd like to thank my Indian friends and SY. Jani, Onkar and Shuang Yin, starting our PhDs around the same time we have shared a lot of good times together. I've always enjoyed hanging out with you guys (ranging from having lunch to memorable nights out in het Doornroosje) and I'm sure we'll all find our way in the future. Onkar, thank you for being one of my paranimphs! Amit, besides the countless times you helped me out with advice on practical science-related stuff I best remember the splendid dinners at Muzenplats (later on joined by Raghu and whiskey :)).

I am really thankful for getting to know all of you, hopefully we'll catch up soon and keep on seeing eachother in the future.

I was working at the department for half a year when the Friday afternoon-borrel was introduced (courtesy of Nina). These borrels were always a good opportunity to relax, drink a beer and complain about failed experiments, i.e. a great way to start the weekend. I'd like to thank fellow borrel mates Bilge, Boris & Bowon (also many thanks for your help on me getting my new position in Melbourne, see you soon!), Evelyn, George, Giovanni & Nuria, Guido, Ila, Jacob, Jessie, Kim, Koen & Tanya, Luan, Matteo, Naomi, NQ, Roderick, René, Sabine T, and later on people from the Vermeulen lab for, among others things, spending many Friday afternoons at the coffee table. I've also always enjoyed the company of other great colleagues in the lab, I'd like to thank Cheng, Christa, Eduardo, Farid (Farsi-teacher V2.0), Georgina & Ann, Laura, in memory Pankaj, Sabine F and many others. You all made the lab a great place to be! Good luck to those that still have to finish their PhD!

Ondanks dat de overstap van reumatologie naar moleculaire biologie ogenschijnlijk een grote is, viel dit in mijn geval behoorlijk mee. Voor een antilichaam of voor een biertje in de Aesculaaf kon ik dan ook altijd nog bij jullie, Arjan & andere oud-collega's van Reumatologie, terecht, dank daarvoor.

Behalve de mensen van het lab hebben ook vrienden en familie een belangrijke rol gespeeld gedurende de afgelopen 6 jaar, hun steun was dan ook onmisbaar tijdens mijn promotie-traject.

Mijn cariere als bioloog begon ongeveer 15 jaar geleden, en als mede-biologen hebben jullie een belangrijke invloed gehad op mij sindsdien. Niet alleen tijdens de studie maar ook gedurende mijn promotie heeft onze vriendschap er bovendien voor gezorgd dat ik me altijd thuis gevoeld heb in Nijmegen. Bedankt Bas (uiteraard ook voor (wetenschappelijke) input tijdens koffie op de 1^e), Bertolt, Job, Roel, Sid en Sjoerd , (+♀)!

Tijdens m'n promotie heb ik altijd met veel plezier gewoond in Mater Dei. Catalaanse kip op dinsdag (Johan) gepaard met goede "discussies" over wetenschap en politiek (Diederik, Noud). Maar daarnaast ook veel gezelligheid en zo af en toe een feestje. Voor dit wil ik deze en andere (oud)huisgenoten heel erg bedanken (mn Bastiaan, Bente, Dominic, Els, Felicia, Ineke & Adnan, Kyara & Maureen, Lisa, Patricia, Rob en Willem).

Ana, in many ways you've also contributed to this thesis, thanks for that! Hay que vivir bonito!

Erik, meer dan 15 jaar na ons eerste wetenschappelijke succes ("de Magnetron", Vrij en ter Huurne, 2002) is het nog steeds met geen ander zo leuk om over wetenschap te praten, maar vooral ook over minder wetenschappelijke zaken.. :). Thanks! Fijn dat jij mijn paranimpf wilde zijn!

Ondanks dat ik nu al vrijwel net zo lang in Nijmegen woon als in Buurse gewoond heb ben ik ook gedurende mijn promotie nog vaak naar het oosten des lands afgereisd voor vrienden en familie.

Bart-Jan, Clemens, Huub, Laurens, Marc, Remco, Ronnie, Ruben, Ruurd, (+ aanhang uiteraard), het feit dat wij meer dan 25 jaar bevriend zijn zegt eigenlijk alles. Het was altijd fijn om even ver weg te zijn van het lab en het in goed gezelschap niet over stamcellen te hebben, dank daarvoor!

Ook ging een bezoek aan Twente vaak (en al jarenlang) gepaard met een bezoek aan de buren, Robbert en de rest van fam. Lieven bedankt!

Elke en Inge, jullie zijn altijd betrokken geweest bij wat ik deed en ik kon ook altijd bij jullie terecht voor advies. En Thomas, ook jouw nuchtere kijk op dingen heb ik altijd gewaardeerd! (al kwam ik vaak vooral

ook voor afleiding door Kiki, Yara en Lune.. :)). Inge, de VS waren een erg welkome break tijdens mijn promotie, goed dat jij doorzette (en dank voor het geduld :))!

Pa en Ma, dit boekje en mijn promotie zijn natuurlijk vooral ook de verdienste van jullie. Onvoorwaardelijke steun en advies hebben mij gebracht waar ik nu ben, en daarvoor ben ik jullie heel erg dankbaar!

Menno

Curriculum Vitae

Menno Christiaan ter Huurne was born in Enschede, the Netherlands, on October the 12th, 1983. After attending primary school "De Noor" in Buurse he went to "Het Assink" in Haaksbergen. During pre-university education Menno chose the direction Nature and Health.

After finishing high school he studied Biology at the Radboud University Nijmegen with a focus on Immunology and Stem Cell Biology. During the first year of his master he studied the intracellular pathway of a dendritic cell-specific receptor at the department of Tumor Immunology (prof. Carl Figdor) at the Radboud Institute of Molecular Life Sciences. Next, he moved to Amsterdam for one year to perform an internship focusing on non-coding RNA-mediated recruitment of Polycomb proteins at the department Molecular Genetics (prof. Maarten van Lohuizen) at the Netherlands Cancer Institute.

After finishing his studies, he started as a researcher at the department of Experimental Rheumatology (prof. Wim van den Berg) at the Radboud University Medical Centre where he assessed the use of adipose tissue-derived mesenchymal stem cells as a therapeutic tool in different mouse models of osteoarthritis.

In 2012 he started his PhD at the department of Molecular Biology headed by prof. Henk Stunnenberg studying the cell cycle of mouse embryonic stem cells. The results obtained during his PhD are described in this thesis.

Menno will continue his career at the department of Heart Regeneration (headed by dr. Enzo Porrello) at the Murdoch Childrens Research Institute in Melbourne, Australia, studying directed (de-)differentiation of cardiomyocytes.

Publications

Part of this thesis:

- **ter Huurne, M.**, Chappell, J., Dalton, S. and Stunnenberg, H. G. (2017) 'Distinct Cell-Cycle Control in Two Different States of Mouse Pluripotency', *Cell Stem Cell*, 21(4), p. 449–455.e4. doi: 10.1016/j.stem.2017.09.004.
- Liu, N. Q., **ter Huurne, M.**, Nguyen, L. N., Peng, T., Wang, S.-Y., Studd, J. B., Joshi, O., Ongen, H., Bramsen, J. B., Yan, J., Andersen, C. L., Taipale, J., Dermitzakis, E. T., Houlston, R. S., Hubner, N. C. and Stunnenberg, H. G. (2017) 'The non-coding variant rs1800734 enhances DCLK3 expression through long-range interaction and promotes colorectal cancer progression', *Nature Communications*, 8, p. 14418. doi: 10.1038/ncomms14418.

Other:

- Mandoli, A.*, Singh, A. A.*, Prange, K. H. M., Tijchon, E., Oerlemans, M., Dirks, R., **Ter Huurne, M.**, Wierenga, A. T. J., Janssen-Megens, E. M., Berentsen, K., Sharifi, N., Kim, B., Matarese, F., Nguyen, L. N., Hubner, N. C., Rao, N. A., van den Akker, E., Altucci, L., Vellenga, E., Stunnenberg, H. G. and Martens, J. H. A. (2016) 'The Hematopoietic Transcription Factors RUNX1 and ERG Prevent AML1-ETO Oncogene Overexpression and Onset of the Apoptosis Program in t(8;21) AMLs', *Cell Reports*, 17(8), pp. 2087–2100. doi: 10.1016/j.celrep.2016.08.082.
- Saeed, S.*, Quintin, J.*, Kerstens, H. H. D.*, Rao, N. A.*, Aghajanirefah, A., Matarese, F., Cheng, S.-C., Ratter, J., Berentsen, K., van der Ent, M. a., Sharifi, N., Janssen-Megens, E. M., **ter Huurne, M.**, Mandoli, A., van Schaik, T., Ng, A., Burden, F., Downes, K., Frontini, M., Kumar, V., Giamarellos-Bourboulis, E. J., Ouwehand, W. H., van der Meer, J. W. M., Joosten, L. a. B., Wijmenga, C., Martens, J. H. a., Xavier, R. J., Logie, C., Netea, M. G. and Stunnenberg, H. G. (2014) 'Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity', *Science*, 345(6204), pp. 1251086–1251086. doi: 10.1126/science.1251086.

- Schelbergen, R. F., van Dalen, S., **ter Huurne, M.**, Roth, J., Vogl, T., Noël, D., Jorgensen, C., van den Berg, W. B., van de Loo, F. A., Blom, A. B. and van Lent, P. L. E. M. (2014) 'Treatment efficacy of adipose-derived stem cells in experimental osteoarthritis is driven by high synovial activation and reflected by S100A8/A9 serum levels', *Osteoarthritis and Cartilage*, 22(8), pp. 1158–1166. doi: 10.1016/j.joca.2014.05.022.
- **ter Huurne, M.**, Schelbergen, R., Blattes, R., Blom, A., De Munter, W., Grevers, L. C., Jeanson, J., Noël, D., Casteilla, L., Jorgensen, C., Van Den Berg, W. and Van Lent, P. L. E. M. (2012) 'Antiinflammatory and chondroprotective effects of intraarticular injection of adipose-derived stem cells in experimental osteoarthritis.', *Arthritis and rheumatism*, 64(11), pp. 3604–13. doi: 10.1002/art.34626.

- Tacken, P. J.*, **ter Huurne, M.***, Torensma, R. and Figdor, C. G. (2012) 'Antibodies and carbohydrate ligands binding to C-type lectin DC-SIGN differentially modulate receptor trafficking.', *European Journal of Immunology*, 42(8), pp. 1989–98. doi: 10.1002/eji.201142258.
- **ter Huurne, M.**, Figdor, C. G. and Torensma, R. (2010) 'Hematopoietic stem cells are coordinated by the molecular cues of the endosteal niche.', *Stem Cells and Development*, 19(8), pp. 1131–1141. doi: 10.1089/scd.2010.0038.

* Authors contributed equally