UNIVERSITY OF CALIFORNIA

Los Angeles

# Proteomic and transcriptional analysis of components of the DNA replication and repair machinery in mouse embryonic stem cells

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy in Biological Chemistry

by Nadia Sellami

2013

ABSTRACT OF THE DISSERTATION

**Proteomic and transcriptional analysis of components of the DNA replication and repair machinery in mouse embryonic stem cells**

by

Nadia Sellami

Doctor of Philosophy in Biological Chemistry

University of California, Los Angeles, 2013

Professor Kathrin Plath, Chair

DNA replication timing, chromatin state and transcriptional activity of a genomic locus are tightly correlated and undergo vast changes during cell fate changes underlying differentiation and reprogramming processes. It has been proposed that the correlation of replication timing and transcriptional activity is mediated by the interaction of pre-replication complex (preRC) components with the transcription machinery. We addressed this question by investigating the influence of preRC components on transcriptional activity in embryonic stem cells (ESCs) by depleting components of the preRC. Interestingly, we did not find a global change of transcriptional activity. In addition, we tested the interaction of the preRC with the COMPASS chromatin modification complex. Contrary to results in the yeast system, we found no interaction

between preRC components and the COMPASS complex in mouse ESCs do, and thus the recruitment of the preRC to origins of replication via this interacton is unlikely.

As the replication fork clamp proliferacting cell nuclear antigen (PCNA) is conveniently located at the replication fork and has a plethora of interaction partners, including a number of chromatin-modifying enzymes, PCNA might mediate the replication of chromatin states by recruiting chromatin-modifying enzymes to the replication fork at appropriate times during S-phase. We addressed this hypothesis by determining the PCNA interactome in unsynchronized ESCs as well as ESCs specifically in early or late S-phase, respectively, and identified a number of chromatin-modifying complexes as PCNA interactors some of which interacted specifically in early or late S-phase, supporting our model.

Interestingly, among the PCNA interaction partners identified, we found the annealing helicase Zinc finger Ran domain containing protein 3 (Zranb3), which we found to be particularly highly expressed in ESC. We further investigated the function of Zranb3 and identified the MutS$\alpha$ complex of the DNA mismatch repair machinery (MMR), a process especially upregulated in ESC, as interactor of Zranb3. Given the interaction of Zranb3 with the MutS$\alpha$ complex as well as PCNA and its high levels in ESCs, we suggest a model in which MutS$\alpha$ recruits PCNA to sites of MMR and both are then bound by Zranb3 to repair DNA lesions

The dissertation of Nadia Sellami is approved.

James A. Wohlschlegel

Michael F. Carey

Stephen T. Smale

Steven E. Jacobson

Kathrin Plath, Committee Chair

University of California, Los Angeles

2013

**Table of Contents**

## Figures and Tables

### Figures

**Supplementary Figures**

**Supplementary Tables**

**Acknowledgement**

This work was supported by the generous funding of the

Boehringer Ingelheim Fonds

and the

UCLA Dissertation Year Fellowship.

The Appendix is a reprint of a manuscript as published in Cell Reports:

Alissa Minkovsky, Tahsin Stefan Barakat, Nadia Sellami, Mark Henry Chin, Nilhan Gunhanlar, Joost Gribnau,and Kathrin Plath: The Pluripotency Factor-Bound Intron 1 of Xist Is Dispensable for X Chromosome Inactivation and Reactivation In Vitro and In Vivo.

Cell Rep. 2013 Mar 28; 3(3) :905-18.

It is reprinted with permission from Elsevier, license number: 3214280913424.

## VITA

EDUCATION

2003-2008 Diplom Biochemikerin

Christian-Albrechts-Universitaet zu Kiel, Germany

Major: Biochemistry and Molecular Biology

PUBLICATIONS

Minkovsky A, Barakat TS, Sellami N, Chin MH, Gunhanlar N, Gribnau J, Plath K. (2013) The pluripotency factor-bound intron 1 of Xist is dispensable for X chromosome inactivation and reactivation in vitro and in vivo. Cell Rep. 3(3): 905-18.

Koop A, Sellami N, Adam-Klages S, Lettau M, Kabelitz D, Janssen O, Heidebrecht HJ (2013) Down-regulation of the cancer/testis antigen 45 (CT45) is associated with altered tumor cell morphology, adhesion and migration. Cell Commun Signal. 11(1): 41.

AWARDS AND HONORS

UCLA dissertation year fellowship

Boehringer Ingelheim PhD fellowship

Studienstiftung des Deutschen Volkes undergraduate fellowship

# 1 Introduction

## Proteomic and transcriptional analysis of components of the DNA replication and repair machinery in mouse embryonic stem cells

The timing of DNA replication of a particular locus is tightly correlated with its chromatin state and transcriptional activity and together they represent distinguishing features of cell types. All three features, DNA replication timing, chromatin state and transcriptional activity, undergo vast changes during differentiation from embryonic stem cells (ESCs) to specialized cell types (Hiratani et al., 2008) (Jørgensen et al., 2007) (Ryba et al., 2011). ESCs can be differentiated into any cell type and differentiated cells can be reprogrammed to an ESC-like state coined induced pluripotent stem cells (iPSC) by overexpression of the four transcription factors Oct4, cMyc, Klf4 and Sox2 (Takahashi and Yamanaka, 2006). As during differentiation, in the reverse process of reprogramming of differentiated cells to iPSCs, similarly dramatic changes to chromatin states, replication timing and transcription occur (Shufaro et al., 2010) (Hiratani et al., 2010) (Gondor and Ohlsson, 2009). iPSCs represent a promising cell type for medical applications but are not yet widely used since reprogramming is a very inefficient and slow process. If we understand which factors contribute to the inheritance and changes of chromatin modifications and transcriptional activity, we could try to use this knowledge and manipulate to influence both differentiation and reprogramming to be more efficient and faster. This work aims to investigate components of the DNA replication machinery in mouse ESCs by using transcriptional and proteomic analysis tools. Furthermore, since stem cells give rise to all cells of the organism, they need to replicate their genome more faithfully than differentiated cells in order to prevent detrimental mutations. Hence, DNA repair pathways play an important role in the maintenance of pluripotent stem cells and components of the DNA repair machinery will be also included in our study.

DNA replication is one part of the cell division cycle. The cell cycle is the term used to describe the successive events that take place in order for cells to duplicate and divide into two daughter cells. It is divided into the four main parts: a typically long gap phase (G1-phase), a synthesis phase during which the genome is duplicated (S-phase), a second gap phase (G2-phase) and mitosis (M-phase), during which the chromosomes and cellular materials are divided to the daughter cells. The progression through these different cell cycle phases is tightly controlled by cyclins and cyclin-dependent kinases (Graña and Reddy, 1995). ESCs are a fast-dividing cell type. To this end, they show a shorter total progression time through the cell cycle as a result of a short G1-phase such that a high percentage of cells in an unsynchronized population is in S-phase. Stem cells regulate this fast progression through the cell cycle in part by the constitutive activity of cyclins and cyclin-dependent kinases (White and Dalton, 2005). The importance of those cell cycle differences is emphasized by observations that DNA synthesis is required for the reprogramming to the pluripotent state, since ESCs in S/G2-phase have a higher propensity of reprogramming differentiated cells to the pluripotent state through cell fusion (Tsubouchi et al., 2013). Furthermore, the capacity of stem cells to differentiate into different lineages varies during their progression through the cell cycle and requires DNA replication (Pauklin and Vallier, 2013). The importance of S-phase and DNA replication for differentiation and reprogramming shows the need to understand the components involved in the replication of the genome in ESCs in particular.

Faithful DNA replication is required to produce two identical daughter cells both carrying the complete genetic information of the original cell. DNA replication is a semi-conservative process in which the DNA double strand is unwound and identical copies of both strands generated to be passed on to the two daughter cells (Davey and O'Donnell, 2000).

Replication of the genome starts at origins of replication (ORI), which are well defined by their DNA sequence in yeast but not in mammalian cells (Cayrou et al., 2012). The origin recognition

complex (ORC, consisting of the subunits Orc1-6) binds constitutively to the replication origins and upon entry into G1-phase, the replicative helicase minichromosome maintenance complex (MCM, consisting of the subunits Mcm2-7) is loaded to the ORC to form the pre-replicative complex (preRC). As cells enter S-phase of the cell cycle, the preRC is activated and the primase DNA polymerase alpha loaded to the origin, after which the loading clamp proliferating cell nuclear antigen (PCNA) is placed around each of the two unwound DNA strands by the replication factor C (RFC) complex. After this is completed, the replicative DNA polymerases delta and epsilon are loaded to this complex, upon which duplication of the genome can begin. PCNA has a plethora of interaction partners and not only serves to make the DNA polymerases more processive but also serves as a 'landing pad', guiding its interactors to the replication fork (Moldovan et al., 2007).

All cells of an organism have the same genetic information but not all of the information is used in every cell type. Which parts of the genome are being actively transcribed is not only regulated by the concerted action of transcription factors, but also by chromatin modifications such as DNA methylation and covalent histone modifications (Holliday, 2006). Certain marks such as DNA methylation and trimethylated lysine 9 histone 3 define silent regions of the chromosome, termed heterochromatin, and other marks such as trimethylated lysine 4 histone 3 or histone acetylation define the actively transcribed, euchromatic, region of the genome. The transcriptional and chromation profile of a cell defines cellular identity, and these features need to be copied and passed along to daughter cell generations equally faithfully as the DNA molecule itself. However, since a semi-conservative mechanism to replicate histone marks is not likely, it has remained an important question as to how this could be accomplished (Alabert and Groth, 2012).

Interestingly, the time at which a locus is replicated during S-phase correlates with global chromatin states as euchromatic, actively transcribed regions are generally replicated in early S-

phase and heterochromatic, silent, regions late in S-phase (Hiratani et al., 2009). Since both transcriptional profile and chromatin state represent measures of cellular identity and are linked with the replication timing of the respective loci, it does not surprise that the replication timing profile is also tightly correlated with cell type and has even been suggested to be considered as another epigenetic mark (Hiratani and Gilbert, 2009).

It has been proposed that the correlation of replication timing and transcriptional activity is mediated by the interaction of preRC components with the transcription machinery. Chapter 2 addresses this question by investigating the direct influence of preRC components on transcriptional activity and furthermore testing the interaction of the preRC with a chromatin-modifying complex responsible for the establishment of the euchromatic mark of trimethylated lysine 4 at histone 3 (H3K4me3).

As PCNA is conveniently located at the replication fork and has a plethora of interaction partners, amongst which are a number of chromatin-modifying enzymes, PCNA might mediate the replication of chromatin states by recruiting chromatin-modifying enzymes to the replication fork, which may change throughout the progression of S-phase to allow the replication of eu- and heterochromatin. Chapter 3 describes the results of a study addressing this hypothesis by determining the PCNA interactome in ESC in unsynchronized cells as well as specifically in early or late S-phase, respectively.

To faithfully maintain the integrity of the genome through cell divisions, cells have a number of DNA repair mechanisms in place. There are several DNA repair mechanisms such as base excision repair, nucleotide excision repair and mismatch repair which respond to defects on individual strands of the DNA molecule, as well as homologous recombination and non-homologous end-joining, which respond to double strand breaks. Since pluripotent cells that are present in early embryonic development give rise to every cell in the organism, it is perceivable

that the integrity of their genome must be controlled more tightly than that of differentiated cells (Stambrook and Tichy, 2010). Accordingly, it has been shown that components of the nucleotide excision repair pathway are crucial for reprogramming to iPSCs (Fong et al., 2011) and that DNA mismatch repair is 20 times more active in ESCs than differentiated cells (Tichy et al., 2011).

Interestingly, among the PCNA interaction partners identified in chapter 3, we found the annealing helicase Zinc finger Ran domain containing protein 3 (Zranb3), which is particularly highly expressed in ESCs. DNA helicases are mostly known for their DNA-unwinding activity. There are only two known mammalian proteins, HARP (HepA related protein) and Zranb3, with the inverse ability to anneal single DNA strands to a double strand (Yusufzai and Kadonaga, 2011). Both HARP and Zranb3 have a high affinity for DNA fork structures as generated at DNA replication forks or DNA repair intermediates (Yusufzai and Kadonaga, 2008) (Yusufzai and Kadonaga, 2010). HARP is recruited to DNA repair sites through interaction with the single strand binding protein RPA (Ciccia et al., 2009). This suggested that Zranb3 might potentially also play a role in DNA repair and that it might be recruited to DNA repair sites by its direct interaction with PCNA. Chapter 4 addresses this question and furthermore identifies the MutS$\alpha$ complex of the DNA mismatch repair machinery (MMR) as interactor of Zranb3.

MMR is a highly conserved repair mechanism that repairs single base mismatches, insertions or deletions during DNA replication (Kolodner and Marsischky, 1999). In eukaryotes, the MutS$\alpha$ complex, consisting of the proteins Msh2 and Msh6 (MutS protein homolog 2 and 6, respectively), recognizes and binds to the damaged nucleotide. Upon binding, it recruits the MutL$\alpha$ complex (consisting of Mlh1 and Pms2) to the lesion, upon which a nick is introduced into the damaged strand (exact mechanism in eukaryotes not known). In the next step, the exonuclease Exo1 cleaves the nicked strand. The generated single stranded DNA is bound by single strand binding proteins such as RPA (replication protein A) and PCNA is recruited through

direct interaction with Msh6. Upon successful loading of these factors, the single strand will be resynthesized by DNA polymerase delta and re-ligated by DNA ligase 1 (Li, 2008). Proteins involved in MMR are highly expressed in ESCs (Tichy et al., 2011). Given the interaction of Zranb3 with the MutS$\alpha$ complex as well as PCNA and its high levels in ESCs, we suggest that Zranb3 is required for MMR (further addressed in chapter 4).

Cellular division does not only require faithful copying of the genetic material, but also the heritable maintenance of gene dosage compensation of chromosomal regions that are differentially expressed between male and female cells. Mammals regulate the gene dosage difference imposed by different numbers of sex chromosomes (XX in females, XY in males) by silencing one of the X chromosomes transcriptionally in female cells (Minkovsky et al., 2012). ESCs have two active X chromosomes and silence one of them during differentiation. This silencing is mediated by the long non-coding RNA XIST and followed by dramatic changes in the chromatin state of the chromosome. Silencing of the X-chromosome in female cells causes the most dramatic changes in replication timing in the genome, switching from replication in early in S-phase in ESC to late in S-phase in differentiated cells (Hiratani et al., 2010). The Appendix represents a reprint of a published study performed on the influence of intron 1 of XIST on the silencing of the X-chromosome during differentiation (Minkovsky et al., 2013).

Taken together, given the importance of DNA replication and repair for the pluripotent state and differentiation, this study aims to understand components of these pathways in mouse ESCs. The ease of genetic manipulation of ESCs and their scalability makes them an ideal system for proteomic and transcriptional approaches. To this end, we investigated the influence of components of the preRC on transcriptional activity, determined the PCNA interactome at different time points in S-phase and investigated the role of Zranb3 in ESCs.

## 1.1 References


Alabert, C., and Groth, A. (2012). Chromatin replication and epigenome maintenance. Nat. Rev. Mol. Cell Biol. *13*, 153–167.

Cayrou, C., Coulombe, P., Puy, A., Rialle, S., Kaplan, N., Segal, E., and Méchali, M. (2012). New insights into replication origin characteristics in metazoans. Cell Cycle Georget. Tex *11*, 658–667.

Ciccia, A., Bredemeyer, A.L., Sowa, M.E., Terret, M.-E., Jallepalli, P.V., Harper, J.W., and Elledge, S.J. (2009). The SIOD disorder protein SMARCAL1 is an RPA-interacting protein involved in replication fork restart. Genes Dev. *23*, 2415–2425.

Davey, M.J., and O'Donnell, M. (2000). Mechanisms of DNA replication. Curr. Opin. Chem. Biol. *4*, 581–586.

Fong, Y.W., Inouye, C., Yamaguchi, T., Cattoglio, C., Grubisic, I., and Tjian, R. (2011). A DNA Repair Complex Functions as an Oct4/Sox2 Coactivator in Embryonic Stem Cells. Cell *147*, 120–131.

Gondor, A., and Ohlsson, R. (2009). Replication timing and epigenetic reprogramming of gene expression: a two-way relationship? Nat Rev Genet *10*, 269–276.

Graña, X., and Reddy, E.P. (1995). Cell cycle control in mammalian cells: role of cyclins, cyclin dependent kinases (CDKs), growth suppressor genes and cyclin-dependent kinase inhibitors (CKIs). Oncogene *11*, 211–219.

Hiratani, I., and Gilbert, D.M. (2009). Replication timing as an epigenetic mark. Epigenetics Off. J. DNA Methylation Soc. *4*, 93–97.

Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.-W., Lyou, Y., Townes, T.M., Schübeler, D., and Gilbert, D.M. (2008). Global Reorganization of Replication Domains During Embryonic Stem Cell Differentiation. PLoS Biol *6*, e245.

Hiratani, I., Takebayashi, S., Lu, J., and Gilbert, D.M. (2009). Replication timing and transcriptional control: beyond cause and effect—part II. Curr. Opin. Genet. Dev. *19*, 142–149.

Hiratani, I., Ryba, T., Itoh, M., Rathjen, J., Kulik, M., Papp, B., Fussner, E., Bazett-Jones, D.P., Plath, K., Dalton, S., et al. (2010). Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. Genome Res. *20*, 155 –169.

Holliday, R. (2006). Epigenetics: a historical overview. Epigenetics Off. J. DNA Methylation Soc. *1*, 76–80.

Jørgensen, H.F., Azuara, V., Amoils, S., Spivakov, M., Terry, A., Nesterova, T., Cobb, B.S., Ramsahoye, B., Merkenschlager, M., and Fisher, A.G. (2007). The impact of chromatin modifiers on the timing of locus replication in mouse embryonic stem cells. Genome Biol. *8*, R169–R169.

Kolodner, R.D., and Marsischky, G.T. (1999). Eukaryotic DNA mismatch repair. Curr. Opin. Genet. Dev. *9*, 89–96.

Li, G.-M. (2008). Mechanisms and functions of DNA mismatch repair. Cell Res. *18*, 85–98.

Minkovsky, A., Patel, S., and Plath, K. (2012). Concise review: Pluripotency and the transcriptional inactivation of the female Mammalian X chromosome. Stem Cells Dayt. Ohio *30*, 48–54.

Minkovsky, A., Barakat, T.S., Sellami, N., Chin, M.H., Gunhanlar, N., Gribnau, J., and Plath, K. (2013). The pluripotency factor-bound intron 1 of Xist is dispensable for X chromosome inactivation and reactivation in vitro and in vivo. Cell Reports *3*, 905–918.

Moldovan, G.-L., Pfander, B., and Jentsch, S. (2007). PCNA, the Maestro of the Replication Fork. Cell *129*, 665–679.

Pauklin, S., and Vallier, L. (2013). The Cell-Cycle State of Stem Cells Determines Cell Fate Propensity. Cell *155*, 135–147.

Ryba, T., Hiratani, I., Sasaki, T., Battaglia, D., Kulik, M., Zhang, J., Dalton, S., and Gilbert, D.M. (2011). Replication timing: a fingerprint for cell identity and pluripotency. PLoS Comput. Biol. *7*, e1002225.

Shufaro, Y., Lacham-Kaplan, O., Tzuberi, B.-Z., McLaughlin, J., Trounson, A., Cedar, H., and Reubinoff, B.E. (2010). Reprogramming of DNA replication timing. Stem Cells Dayt. Ohio *28*, 443–449.

Stambrook, P.J., and Tichy, E.D. (2010). Preservation of genomic integrity in mouse embryonic stem cells. Adv. Exp. Med. Biol. *695*, 59–75.

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

Tichy, E.D., Liang, L., Deng, L., Tischfield, J., Schwemberger, S., Babcock, G., and Stambrook, P.J. (2011). Mismatch and base excision repair proficiency in murine embryonic stem cells. DNA Repair *10*, 445–451.

Tsubouchi, T., Soza-Ried, J., Brown, K., Piccolo, F.M., Cantone, I., Landeira, D., Bagci, H., Hochegger, H., Merkenschlager, M., and Fisher, A.G. (2013). DNA Synthesis Is Required for Reprogramming Mediated by Stem Cell Fusion. Cell *152*, 873–883.

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

Yusufzai, T., and Kadonaga, J.T. (2008). HARP Is an ATP-Driven Annealing Helicase. Science *322*, 748–750.

Yusufzai, T., and Kadonaga, J.T. (2010). Annealing helicase 2 (AH2), a DNA-rewinding motor with an HNH motif. Proc. Natl. Acad. Sci. U. S. A. *107*, 20970–20973.

Yusufzai, T., and Kadonaga, J.T. (2011). Branching out with DNA helicases. Curr. Opin. Genet. Dev. *21*, 214–218.

# 2 Involvement of the PreRC complex in the regulation of gene expression and the histone 3 lysine 4 methylatransferase complex COMPASS

## 2.1 Introduction

Every dividing eukaryotic cell has to copy its genetic material during S-phase before it can give rise to two new daughter cells. Proper copying of the genetic material includes not only the DNA itself, but also the epigenetic information, which is encoded in chromatin, which is in turn tightly linked to the transcriptional activity of a given locus (Farkash-Amar et al., 2008).

As the cell has to make sure that the whole genome gets faithfully copied once and only once per cell cycle, the initiation of replication is a highly regulated process. Prior to initiation, the origin recognition complex (ORC) will bind to select ORIs during G1-phase of the cell cycle. It is thought that the timing and frequency of recruitment determine the efficiency and timing of replication. After activation of ORC, the MCM-complex (minichromosome maintenance complex) is recruited, which is believed to be the replicative helicase, forming the pre-replicative complex (preRC). Upon entry of the cell into S-phase, DNA polymerases and other replication factors such as the proliferating cell nuclear antigen (PCNA) are loaded onto the ORI to start replication (Kneissl et al., 2003).

The replication fork will proceed in both directions and replicate the DNA and prior chromatin state. The latter is facilitated by the chromatin assembly factor CAF-1, which deposits new histone molecules at the newly replicated DNA. Newly incorporated histones and DNA then acquire the chromatin structure including all its modifications through the function of chromatin modifiers and chromatin remodeling enzymes. After successful replication, chromosomes will condense and be divided onto the two daughter cells and a new cell cycle can begin.

Replication does not occur in a synchronous manner throughout the genome, but large domains are replicated at rather discrete points of time in S-Phase. The time at which a specific region

gets replicated is inherited through cell division and the genome-wide replication timing profile is cell type-specific (Hiratani et al., 2008). The replication timing profile of a given locus can change during development (for example as genes are shut down upon differentiation they shift from early replication in S phase to late) (Hiratani et al., 2010), but the underlying mechanism for this shift remains unclear.

It is established that ORC proteins are involved in the formation of heterochromatic, transcriptionally silent domains in yeast (Kelly et al., 1994) and might be involved in transcriptional repression by interaction with the heterochromatin protein HP1 in mouse (Auth et al., 2006). In agreement with the notion of a link between ORC and transcriptional regulation, a recent study defined ORIs in mouse cells and demonstrated that they are often found in promoter regions and that the transcription initiation activity determines the frequency of origin activation (Sequeira-Mendes et al., 2009). Furthermore, transcription factor binding sites are enriched in candidate ORIs (Cadoret et al., 2008), and ORC has been shown to interact with transcription factors such as the retinoblastoma protein Rbf1 in *drosophila* (Ahlander et al., 2008), AIF-C in rat (Saitoh et al., 2002), c-Myc in humans (Takayama et al., 2000). Additionally, roles in transcription have been demonstrated for proteins of the MCM complex, which interacts with the transcription factor Stat1 (Snyder et al., 2005) and importantly also with RNA polymerase II (Snyder et al., 2009). An important question is whether ORC/MCM proteins actively regulate transcription, independently of their role in DNA replication. For instance, they could be recruited directly by the transcription machinery.

Hiratani et al furthermore demonstrated that early replicating regions are enriched for active chromatin marks including histone H3 lysine 4 trimethylation (H3K4me3), and late replicating regions for repressive chromatin marks (Hiratani et al., 2008). This raises the question if there might be a functional link between chromatin state and replication timing, which in turn is determined by the activation of the preRC.

Interestingly, the ORC subunit Orc2 binds to histone H3 trimethylated at Lysine 4 (H3K4me3) and might mediate histone 3 lysine 4 methylation by interaction with a Spp1, a member of an H3K4 methyl transferase (HMTase) complex in yeast (Kan et al., 2008). The mammalian homolog of Spp1 is CGBP (Voo et al., 2000), which is crucial in early embryonic development (Carlone and Skalnik, 2001) and also part of a H3K4-HMTase complex (Lee and Skalnik, 2005). As H3K4me3 is found at the promoters of all active and bivalent genes, these findings suggest that this particular chromatin modification (and maybe others as well) might affect how ORC gets targeted or activated. A possible explanation could be that, just as seen in yeast, ORC might directly interact with CGBP in mammalian cells and thus be recruited to preferentially to euchromatic regions and therefore, preference would be given to activation of ORIs containing high levels of H3K4me3.

Here, we wanted to further explore the functional link between the preRC complex and transcriptional regulation. Furthermore, as H3K4me3 is correlated to active transcription and replication timing and since there may be direct interactions between the ORC and H3K4HMTase complexes, we also wanted to investigate the link between the H3K4 HMT complexes and the ORC in more detail.

For these studies, I used ESCs because of the ease of manipulation and their rapid growth and indefinite life span, making them an ideal model system for transcriptional and biochemical approaches.

## 2.2 Results

The influence of the preRC on transcription

Given the tight correlation of the time at which a locus is replicated and its transcriptional status, we wanted to explore this relationship further. Since the ORC and MCM complexes are the essential components of the preRC and interact with transcription factors, we wanted to explore if the preRC might directly recruit the transcription machinery and thus if a depletion of its components might lead to effects on gene expression levels. Therefore, we chose components of the ORC and MCM that had previously been shown to be interact with either transcription factors or RNA polymerases to be depleted by siRNA-mediated knockdown in mouse ESCs and to determine the effect of this perturbation on transcription by expression microarrays. In particular, Orc1 had been demonstrated to interact with the transcription factor cMyc in human cells and Mcm2 and Mcm5 with Stat1 and RNA Polymerase II (Takayama et al., 2000) (Snyder et al., 2009).

Expression levels of the preRC components Orc1, Mcm2 and Mcm5 were reduced by at least 80 % as measured on the protein as well the RNA level (Figure 2.1).

The effect of the knockdown on transcription was then measured by RNA microarrays, first from RNA preparations corresponding to the samples shown in Figure 2.1 A (Figure 2.2). Interestingly, the results revealed that the transcriptional changes were only minor upon knockdown of either Mcm5 or Orc1. We observed no upregulation greater than two fold of any of the 16331 genes tested in either sample when compared to the controls, and the only genes downregulated more than twofold were Mcm5 and Orc1 themselves (to 13 % and 32 % of control transfected cells, respectively).

To determine if the small transcriptional changes observed upon Mcm5 or Orc1 knockdown might be still functionally relevant, we performed unsupervised hierarchical clustering analysis (Figure 2.2 A). In the case that the same genes are consistently up or down-regulated in a

particular condition, this might argue that despite small, those changes might still be relevant. Therefore, all samples with the same condition (i.e. the same RNA being knocked down) should cluster together. Interestingly, this is not the case as one of the samples depleted for Orc1 (highlighted in green, Figure 2.2 A) clusters closer to the control samples than the other samples treated with siOrc1. This indicates that the differences between the samples really small and that the Mcm5 or Orc1 depleted samples are not significantly different from the control samples. Indeed, a pearson correlation analysis (Figure 2.2 B) reveals that all samples correlate very well to the controls (pearson coefficient 0.998 or higher) further supporting this argument.

To increase the sample number and confirm these initial findings, a second set of RNAs corresponding to the samples in Figure 2.1 B was analyzed by microarrays and compared to the first set of samples (Figure 2.3). Transcript levels were globally normalized for all samples and compared to each other. As in the last set of samples, there were no changes of transcript levels higher than two fold observed in this experiment, other than the gene that was knocked down. To highlight very small changes in transcription and determine if the respective minimally changing transcripts follow the same trends in both sample sets, we are showing the left panel of Figure 2.3 A in a color scheme representing a log2 scale from -0.2 to +0.2 (corresponding to a +/- 1.15 fold change). The first sample set is highlighted in green, the second in purple text. Surprisingly, the small changes in each set do not follow the same trend. When comparing the samples in a color scheme representing a log2 scale from -1 to +1 (corresponding to +/- 2 fold change), it is apparent that any changes in expression observed are miniscule (Figure 2.3 A, right). Those genes experiencing minimal changes upon Mcm5 or Orc1 knockdown tend to be the ones with medium gene expression levels but do not appear to be particularly correlated with replication timing or RNA polymerase II binding. Interestingly, they seem to be slightly correlating with both trimethylation at lysine 4 or lysine 27 of histone H3. The combination of these chromatin marks signifies bivalent genes, which experience the most drastic changes during

differentiation from ESC to specialized cell types (Bernstein et al., 2006). However, the lack of distinctive effects of the Mcm5 or Orc1 knockdown is further demonstrated by a clustering of all samples analyzed (Figure 2.3 B) in which samples cluster by the date they were processed (indicated by coloring in either purple or green) rather than by the gene that was depleted.

To validate the, albeit small, effect of the knockdown of either Mcm5 or Orc1 on transcript levels, we confirmed the expression changes of a number of the genes that were amongst the most affected (top ten) upon knockdown by qRT-PCR (Figure 2.4). We chose Cdo1 and Atf4, which slightly changed expression upon Mcm5 knockdown, Pded6, Mthfd1l and Ctdsp1 which showed small changes upon Orc1 knockdown and Wfdc2 and Icam1, for which expression was slightly altered in both experimental conditions. Since Orc1 had been shown to interact with the transcription factor c-Myc, we were particularly interested in seeing if Orc1 (or preRC) depletion would have a differential (if small) effect on genes regulated by c-Myc or not. The qRT-PCR confirmed the results of the microarray seeing that the results were reproducible and that the effect was minor and less than a twofold change and both c-Myc regulated and independent genes responded similarly to preRC depletion.

The possibility remained that the effect on transcription would only be observable if the cells had to undergo a change in gene expression patterns, which might be only then be able to be perturbed through depletion of the MCM complex. This hypothesis is in accordance with the finding that those genes experiencing some level of expression changes tend to correlate with bivalent chromatin modifications, which undergo the most drastic changes in gene expression during differentiation.

To test this hypothesis, we created a condition in which ESCs are induced to rapidly differentiate by removal of the stem cell transcription factor Oct-4. To achieve this, we used a cell line in which both endogenous alleles of Oct-4 were inactivated and replaced by an Oct-4 transgene under control of the tet operon, which is activated in the absence of tetracycline (or doxycycline).

Upon addition of doxycycline (DOX), the cells will lose Oct-4 expression within 24 hours and differentiate (Niwa et al., 2000). To create an even more rapid expression change by a more thorough differentiation, we additionally removed the factor LIF (leukemia inhibitory factor) from the medium in the differentiating condition, which is crucial for the maintenance of pluripotency of mouse ESC in culture. Differentiating and pluripotent cells were then depleted for either Mcm2 or Mcm5 and the expression changes analyzed by microarrays (Figure 2.5). The successful differentiation was confirmed by morphology (Figure 2.5 A) and did not show an obvious difference between Mcm2 knockdown, Mcm5 knockdown and a control.

We normalized the samples to the respective scramble-siRNA transfected control in order to be able to visualize changes that depend on the specific knockdown of either Mcm2 or Mcm5 under each condition (differentiated or pluripotent).

As in the previous analysis, we only observed minor changes in expression levels. In the Mcm2 knockdown (KD) samples, we only observe 11 genes in the differentiated condition and 16 genes in the pluripotent condition change twofold or more. Similarly, for the samples depleted for Mcm5, a twofold or greater change is seen for 26 or 28 genes in the differentiated or pluripotent condition, respectively.

Since we did not observe any changes to cell proliferation and the MCM (as well as the ORC complex) is crucial for cell division, we believe that the rather small changes we observed might be due to a not sufficient depletion of the preRC by siRNA. However, a complete knockout is technically very challenging since all components of the preRC are essential.

In order to understand the albeit small changes we observe better, we compared our data to published transcription factor binding data, especially since Mcm5 had been shown to interact with the transcription factor Stat1 and might interact with others. Ng *et al* had classified major ESC transcription factors into different groups according to their co-occupancy of promoters in ESC (Chen et al., 2008). When grouping our data according the groups classified by Ng

(Figure 2.5 B), we observe opposing trends between differentiating and pluripotent cells that were depleted for MCM components (highlighted with a red box) in Ng group II, which most prominent feature is a high level of binding of c-Myc and n-Myc. This may suggest a differential role of c-Myc and n-Myc in pluripotent cells and differentiated cells and its interaction with the MCM complex should be further investigated in the future.

To additionally classify the small changes observed upon knockdown of either Mcm2 or Mcm5 in differentiating cells, we also performed an unsupervised K-means clustering analysis and searched for gene ontology (GO) terms that were enriched in each of the clusters (Figure 2.5 C). Interestingly, in the cluster that is specifically upregulated upon MCM knockdown in differentiation, the most enriched GO categories are 'transcription, regulation of transcription and RNA processing', in accordance with our initial hypothesis. Since all changes we observed were relatively small, it is not possible to determine the significance of this enrichment. We suggest that the MCM complex may indeed have a role in the regulation of transcription but that future work with a different approach to this question will be required to conclusively answer this question.

<u>The relationship of the histone methyltransferase complex COMPASS and ORC</u>

Since Orc1 had been shown to interact with the yeast protein Spp1, which is a compontent of the COMPASS complex, there remained another possibility for the link between transcription and replication timing. The COMPASS complex is the histone methyl transferase (HMT) to create trimethylate lysine 4 histone H3 (H3K4me3), which is a chromatin mark associated with active transcription. Early replicating regions are enriched for this mark, so we hypothesized that the mammalian Spp1 homolog CGBP might also interact with Orc1 and Orc2 and thus recruit the ORC to sites of active transcription.

To test this, we created an ESC line that can be induced to express flag-tagged CGBP to conduct flag-immunoprecipitation and analyze the mammalian interactome of this COMPASS subunit by mass spectrometry (Wolters et al., 2001). A C-terminally Flag-tagged CGBP transgene was inserted into the collagen locus under control of the tet operon and is thus inducible by doxycycline (Figure 2.6 A). We generated nuclear extracts from the Flag-CGBP expressing cell line and the original, untargeted, cell line. Flag-immunoprecipitation from both extracts was then performed and the successful immunoprecipitation confirmed by western blot analysis (Figure 2.6 B) and silver stain (Figure 2.6 C). We confirmed the co-immunoprecipitation of a known interactor (Wdr5) as a positive control. Analysis of the complete interactome of CGBP in mouse ESC by MudPIT revealed the enrichment of all components of the COMPASS complex but no subunit of the ORC or MCM complexes in the immunoprecipitates from the Flag-CGBP line compared to those from the control line (BG=background) (Figures 2.6 D and E). In addition to the COMPASS subunits Ash2L, Wdr5, Setd1a, Rbbp5, Wdr82 and Hcf-1 we also identify Ogt (O-linked N-acetylglucosamine (GlcNAc) transferase) in the immunoprecipitates from Flag-CGBP expressing cells. Hcf-1 has been shown to be a target of Ogt and thus might be only indirectly associated with CGBP. Additionally identified proteins have not yet been associated with either the COMPASS complex or the preRC. Interestingly, we find Bod1l (biorientation of chromosomes in cell division protein 1 like) amongst the putative CGBP interaction partners. This is noteworthy since Bod1l is homologous to Bod1, which is required for cell division as part of the kinetochore (Porter et al., 2007).

We conclude that, under our experimental conditions, it is therefore unlikely, that the same mechanism that has been demonstrated in yeast also applies to the mammalian system. Taken that origins of replication are well defined by sequence in yeast but not in mammals, replication timing and preRC recruitment must be regulated differently in the two systems. The results shown here are further evidence for this notion.

**Figure 2.1: Establishment of ORC and MCM knockdown.**

ESC were transfected with siRNA at a final concentration as stated. A second transfection was performed 24 h later and cells harvested at 48 h after the first transfection. The effectiveness of the knockdown was assessed by western blot (A and B) or qRT-PCR (C and D).

**Figure 2.1: Establishment of ORC and MCM knockdown**

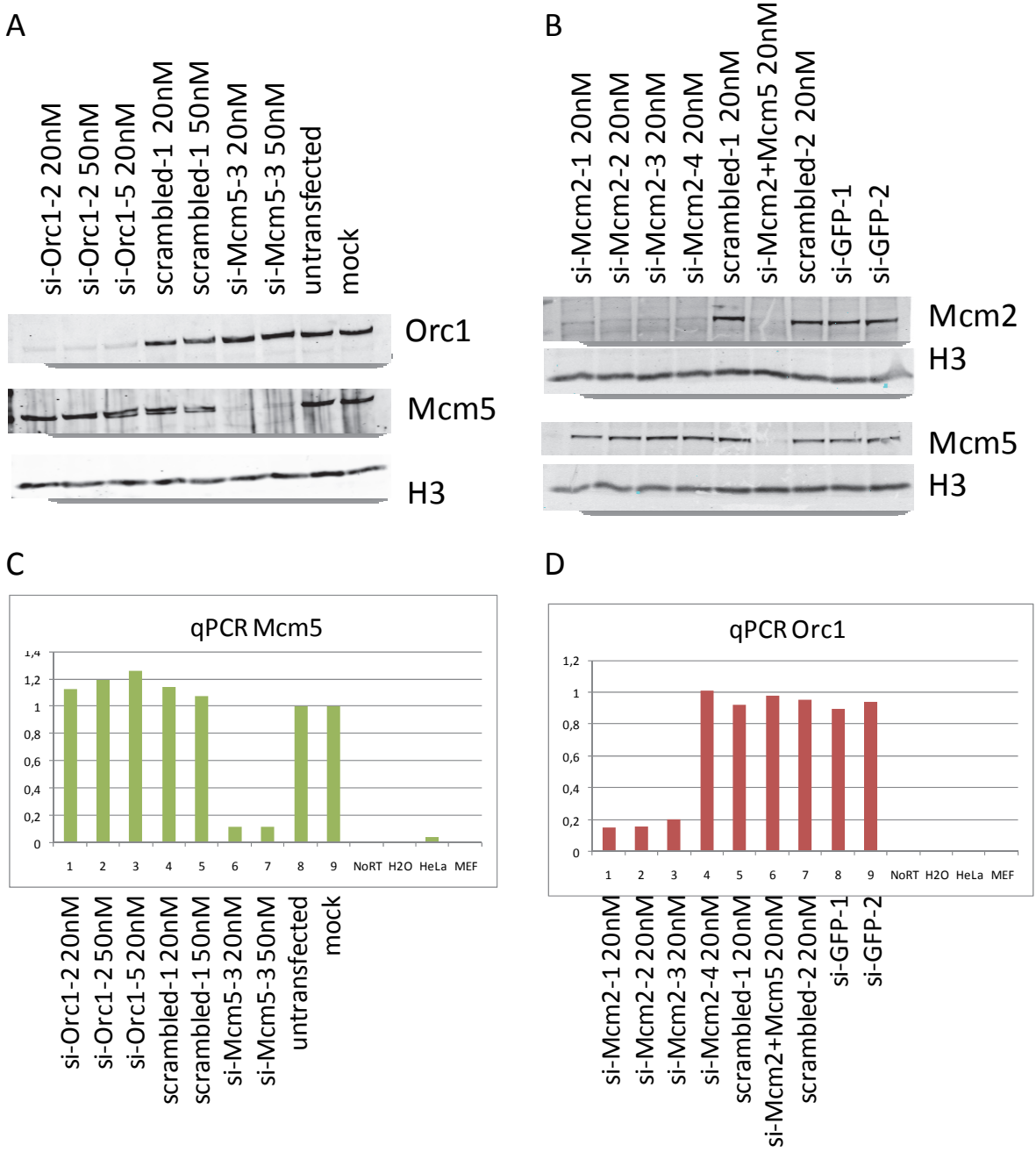
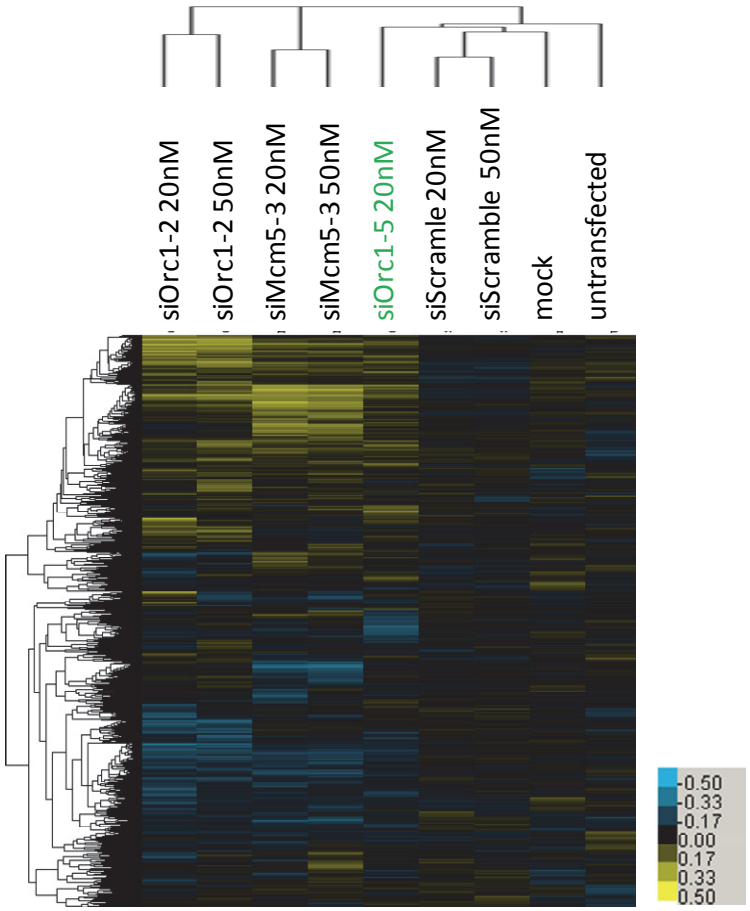


**Figure 2.2: Knockdown of ORC or MCM does not lead to transcriptional changes compared to controls.**

A) ESC were transfected siRNA at a final concentration as stated. A second transfection was performed 24 h later and cells harvested at 48 h after the first transfection. Affymetrix Mouse 430 plus 2.0 expression microarrays were performed to obtain transcript levels. Transcripts were averaged by gene and unsupervised clustering performed. Color schemes are on a log2 scale relative to expression changes. B) Pearson correlation coefficients of the data from the samples shown in A).

**Figure 2.2: Knockdown of ORC or MCM does not lead to transcriptional changes compared to controls.**

A




B

| | siMcm5-3 20nM | siMcm5-3 50nM | siOrc1-5 20nM | siOrc1-2 20nM | siOrc1-2 50nM | siScramle 20nM | siScramble 50nM | mock | untransfected |
|---|---|---|---|---|---|---|---|---|---|
| siMcm5-3 20nM | 1 | 0.99964 | 0.99894 | 0.99822 | 0.99856 | 0.99831 | 0.99838 | 0.99796 | 0.99868 |
| siMcm5-3 50nM | | 1 | 0.99881 | 0.99832 | 0.99849 | 0.99837 | 0.99842 | 0.99804 | 0.99851 |
| siOrc1-5 20nM | | | 1 | 0.9988 | 0.99904 | 0.99904 | 0.99907 | 0.99881 | 0.99911 |
| siOrc1-2 20nM | | | | 1 | 0.9988 | 0.99904 | 0.99904 | 0.99907 | 0.99881 |
| siOrc1-2 50nM | | | | | 1 | 0.99918 | 0.99827 | 0.99812 | 0.99831 |
| siScramle 20nM | | | | | | 1 | 0.99805 | 0.99816 | 0.99807 |
| siScramble 50nM | | | | | | | 1 | 0.99979 | 0.99905 |
| mock | | | | | | | | 1 | 0.99912 |
| untransfected | | | | | | | | | 1 |

Figure 2.3**: Minor changes in gene expression of ORC or MCM correlate slightly with medium expression in ESC and bivalent chromatin modification.**

A) ESC were transfected siRNA at a final concentration as stated. A second transfection was performed 24 h later and cells harvested at 48 h after the first transfection. Affymetrix Mouse 430 plus 2.0 expression microarrays were performed to obtain transcript levels. Transcripts were averaged by gene and average gene expression levels in ESC, DNA replication timing data (early replicating in S-phase versus late in S-phase), chromatin Immunoprecipitation data for RNA Polymerase II, H3K4me3 and H3K27me3 was added for each gene as comparison. The data were then sorted according to average gene expression level in mESC. All data are depicted on a log2 scale. Samples are colored according to the group (experiment) they are a part of (purple and green). The left panel is a depiction of the same data as right in higher contrast as indicated by the legend. Color schemes are on a log2 scale relative to expression changes.

B) Unsupervised clustering of the Affymetrix transcription data

**Figure 2.3: Minor changes in gene expression of ORC or MCM correlate slightly with medium expression in ESC and bivalent chromatin modification.**

A

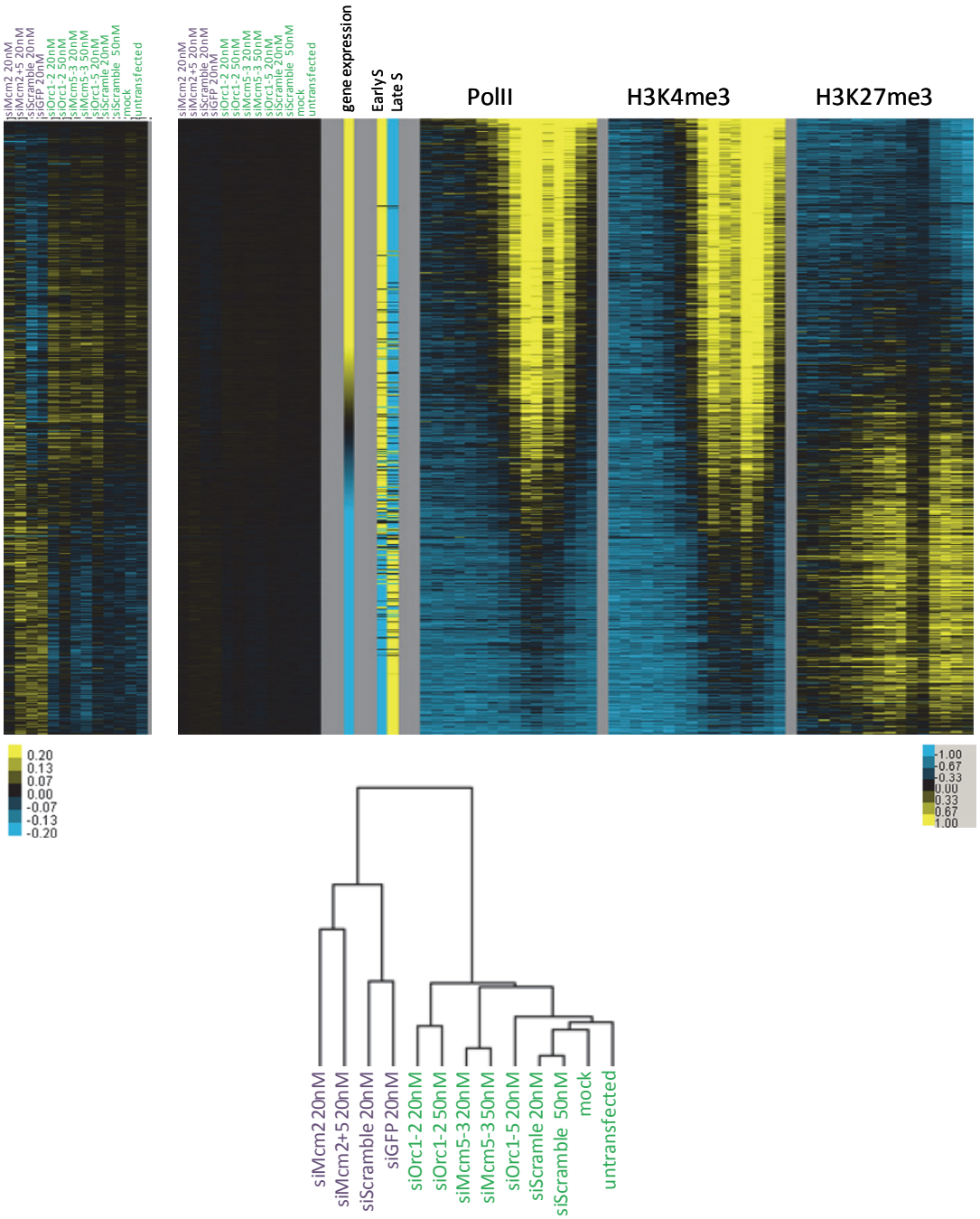
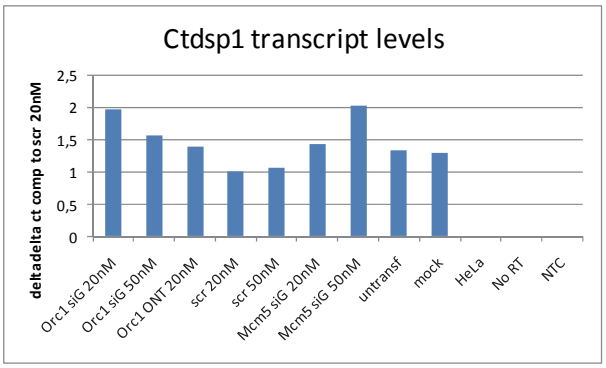


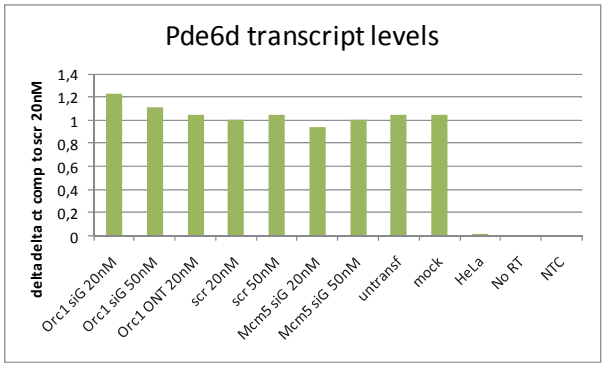
**Figure 2.4: Validation of microarray results by qPCR.**

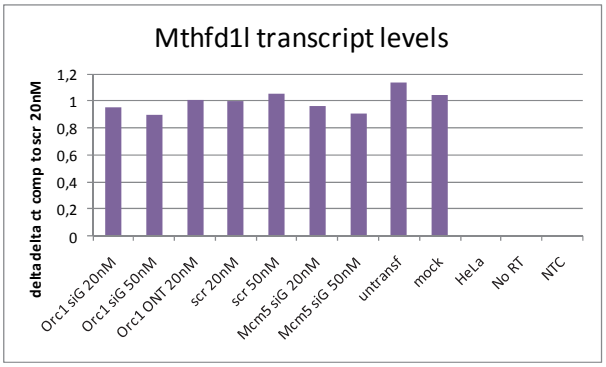
Genes among the top ten genes that change transcription upon knockdown as indicated in the table were selected to perform qRT-PCR analyses. RNA samples were taken from the samples used in the microarray experiment.

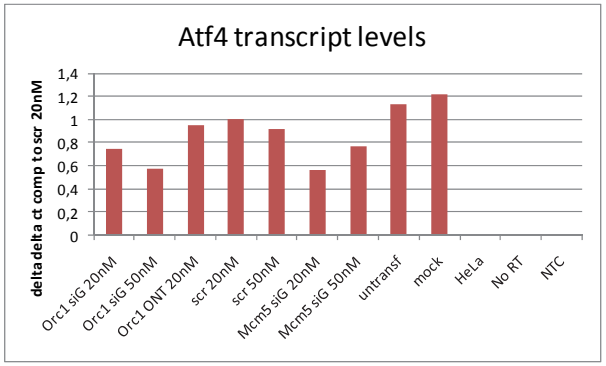
**Figure 2.4: Validation of microarray results by qPCR.**

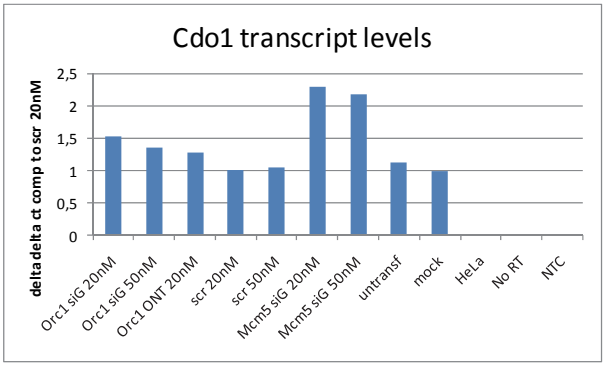
| Gene | Differentially regulated in knockdown | Myc-target? |
|---|---|---|
| Cdo1 | Mcm5 | No |
| Atf4 | Mcm5 | Yes |
| Pde6d | Orc1 | ND |
| Mthfd1l | Orc1 | Yes |
| Ctdsp1 | Orc1 | No |
| Wfdc2 | Orc1+Mcm5 | No |
| Icam1 | Orc1+Mcm5 | Yes |

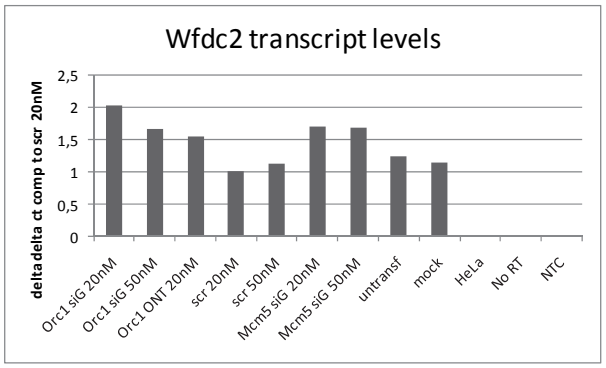


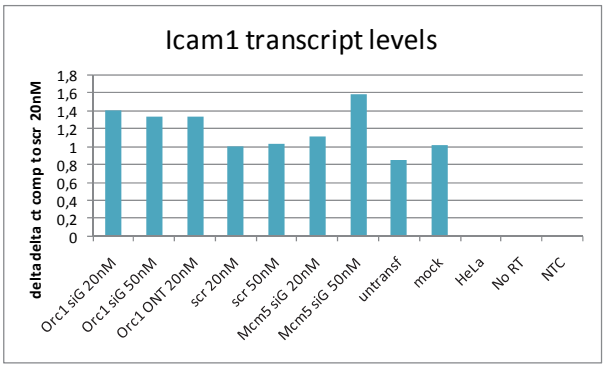
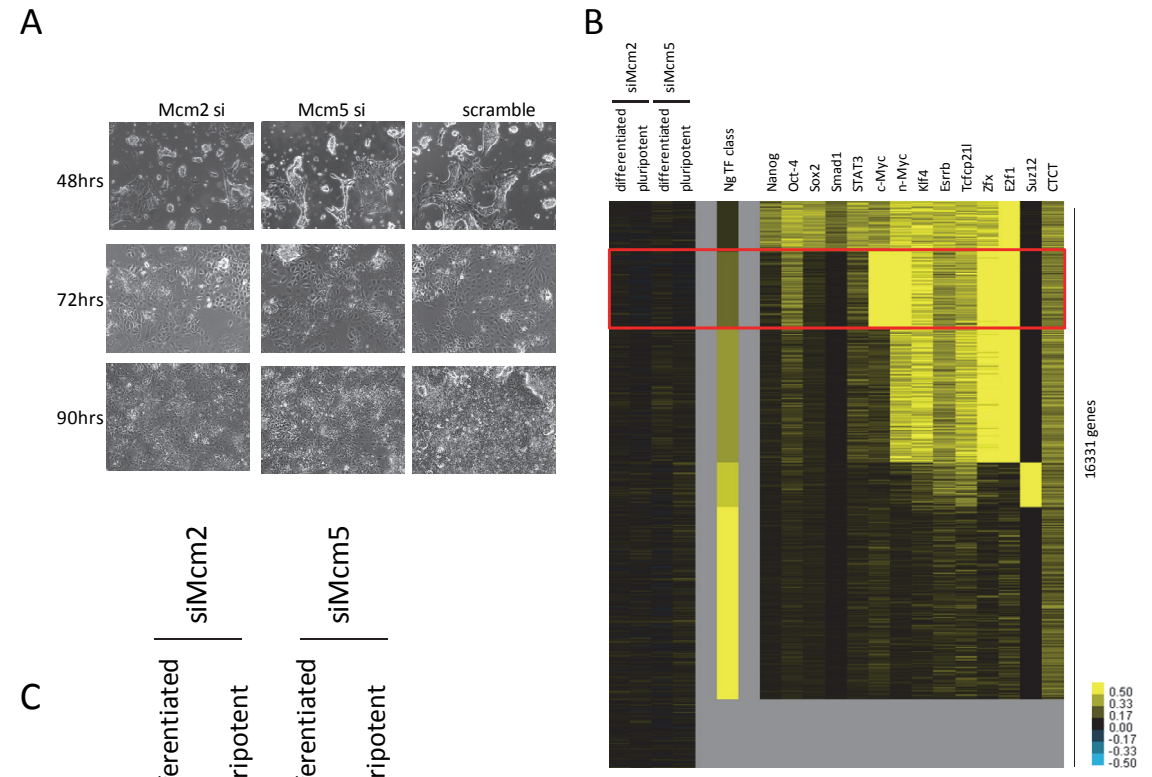



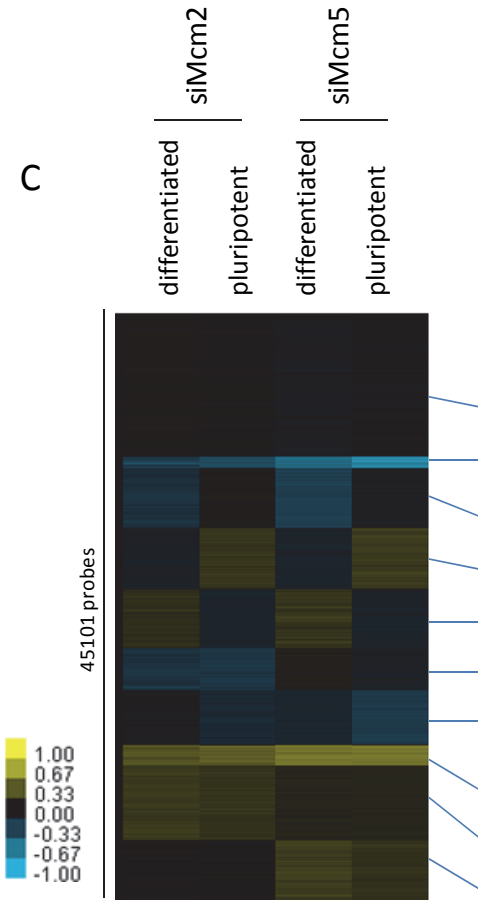
**Figure 2.5: Analysis of minor changes in transcript levels upon MCM knockdown in differentiating cells.**

ZHBTc4 ESC were transfected siRNA at a final concentration of 20 nM. A second transfection was performed 24 h later and cells harvested at 90 h. Affymetrix Mouse 430 plus 2.0 microarrays were performed to obtain transcript levels. During the course of the experiments the cells were either kept under conditions required fror the maintenance of pluripotency, or induced to differentiate by withdrawal of Oct-4 and LIF. A) During the course of the experiment, pictures of the cultures were taken at 10x magnification at the indicated time points to observe morphological changes. B) The data was averaged by gene and additionally, transcription factor binding data was obtained from (Chen et al., 2008) and the dataset was sorted according to Ng transcription factor binding class (Ng class). Ng transcription factor class II is highlighted with a red box. All data is depicted on a log2 scale..

C) K-means clustering was performed at the probe level (data not averaged by gene) and gene ontology analysis (GO) for biological function on each cluster performed using the online tool DAVID. The most enriched terms are shown in the table on the right.

**Figure 2.5: Analysis of minor changes in transcript levels upon MCM knockdown in differentiating cells**.



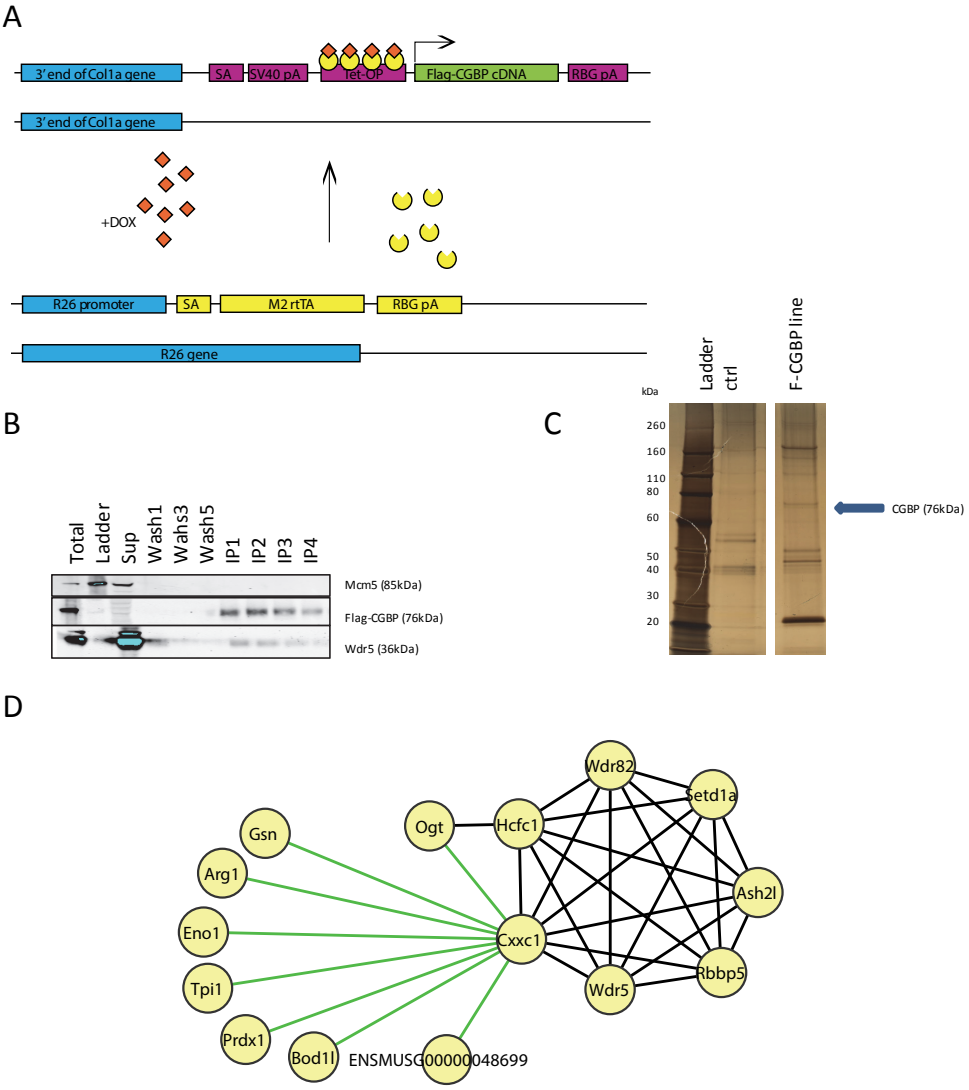




| clusters | enriched GO terms (biological function) |
|---|---|
| Unchanged | translation, protein localization, RNA processing |
| MCM specific DOWN | chromosome organization, chromatin organization, chromatin modification, regulation of transcription |
| DOWN in KD upon differentiation | cellular component morphogenesis, cell morphogenesis, behavior |
| UP in KD without differentiation | cell adhesion, biological adhesion, neuron development |
| UP in KD upon differentiation | transcription, regulation of transcription, RNA processing |
| DOWN specific to Mcm2 KD | transcription, mRNA metabolic process, mRNA processing |
| DOWN in KD without differentiation | transcription, regulation of transcription, regulation of transcription from RNA polymerase II promoter |
| MCM specific UP | cell morphogenesis, cellular component morphogenesis, cell morphogenesis involved in differntiation |
| UP specific to Mcm2 KD | protein localization, establishment of protein localization, protein transport |
| UP specific to Mcm5 KD | phosphate metabolic process, phosphorous metabolic process, phosphorylation |

**Figure 2.6: CGBP interacts with the COMPASS histone methyltransferase complex but not the preRC.**

A) Schematic of the generation of a doxycycline-inducible Flag-CGBP cell line. Flag-tagged CGBP was integrated under the control of the Tet-operon (Tet-OP) at the 3' end of the Collagen1a (Col1a) locus. Upon addition of doxycycline (DOX), rtTA that is expressed under control of the ROSA26 (R26) promoter binds to the Tet-operon and induces expression of the integrated construct. SA = splice acceptor, SV40 pA = SV40 polyadenylation signal, RBG pA = RBG polyadenylation signal, M2 rtTA = M2 tetracycline reverse transcriptional activator. B) Western blot of a CGBP immunoprecipitation. Nuclear extracts (total) were immunoprecipitated with Flag-beads. After 3 hours of rotation at 4C, the supernatant (Sup) was removed and the beads washed 5 times (Wash1 – 5) before sequential elutions with flag peptide (IP1 – 4). C) Silver stain of pooled immunoprecipitates of control and Flag-CGBP-expressing cells. D) Flag-CGBP immunoprecipitates were analyzed by MudPIT and the identified interactors are shown in a network view. Black edges represent known interactions, green edges interactions identified in this study. E) Summary of normalized spectral absorbance factors (NSAF) of the identified CGBP interaction partners. BG = control cells that do not express a flag-tagged protein, which were processed simultaneously.

**Figure 2.6: CGBP interacts with the COMPASS histone methyltransferase complex but not the preRC.**




E

| RefSeq ID | name | CGBP NSAFe5 | BG NSAFe5 | CGBP-BG | description |
|---|---|---|---|---|---|
| NP_083144.1 | Cxxc1 (CGBP) | 962.44 | 0.00 | 962.44 | CXXC finger 1 (PHD domain) |
| NP_543124.1 | Wdr5 | 887.52 | 70.15 | 817.36 | WD repeat domain 5 |
| NP_808385.2 | | 1062.62 | 479.09 | 583.53 | hypothetical protein 4732456N10 |
| NP_001074262.1 | Ash2l | 515.46 | 0.00 | 515.46 | ash2-like isoform b |
| NP_766105.2 | Rbbp5 | 472.28 | 0.00 | 472.28 | retinoblastoma binding protein 5 |
| NP_035164.1 | Hcfc1 | 532.00 | 141.30 | 390.71 | peroxiredoxin 1 |
| NP_032250.2 | Prdx1 | 331.32 | 6.87 | 324.45 | host cell factor C1 |
| NP_821172.2 | Setd1a | 320.81 | 0.00 | 320.81 | SET domain containing 1A |
| NP_084172.1 | Wdr82 | 270.59 | 0.00 | 270.59 | WD repeat domain containing |
| NP_031508.1 | Arg1 | 262.21 | 43.53 | 218.69 | arginase 1, liver |
| NP_033441.2 | Tpi1 | 170.07 | 0.00 | 170.07 | triosephosphate isomerase 1 |
| NP_001074891.1 | Bod1l | 167.60 | 0.00 | 167.60 | biorientation of chromosomes |
| NP_035229.2 | Pkm2 | 159.50 | 0.00 | 159.50 | pyruvate kinase, muscle |

## 2.3 Discussion

It has become evident that transcription, chromatin state and replication timing are tightly correlated. There have been several anecdotal reports linking components of the preRC to the transcription machinery. The ORC has been shown to interact with transcription factors such as c-Myc in mouse and the MCM complex with factors such as Stat1 and importantly RNA Polymerase II. This suggested that the components of the preRC might have a role outside of DNA replication initiation to also control transcription, which might explain the tight correlation between transcription and replication timing.

To test this we knocked down several components of the two complexes and assessed transcriptional changes by microarray experiments. Interestingly, there were no major differences beyond a two-fold change observed, even when expression changes were forced by differentiation. Since we did not observe an effect of the knockdown of these essential cellular components on the ability of cells to proliferate, we suggest that an siRNA mediated knockdown might not deplete enough protein to observe the full effect of the preRC on transcription. A complete knockout is not possible since the preRC components are required for cell division. Hence, a more sophisticated approach will be required to answer the question of an involvement of the preRC in transcriptional regulation in the future..

A second explanation could be that there is no causal relationship between the two processes but that the correlation is secondary to another process that affects both transcription and DNA replication. One such process could be the regulation of chromatin marks, which are a key component in gene regulation.

There had been reports that in ESC, in early replicating origins of replication there is an enrichment of H3K4me3, which is a mark of active transcription.

Interestingly, ORC interacts with Spp1 in yeast, which is a component of the COMPASS histone methyltransferase complex that deposits this modification. To test if this interaction is also occurring in mouse ESC, we co-immunoprecipitated CGBP (the mammalian Spp1 homolog) complexes and assessed the interactome by MudPIT mass spectrometry and western blot. In contrast to its yeast homolog, CGBP did not immunoprecipitate the ORC complex. Apart from the possibility that technical reasons prevented the co-immunoprecipitation in our system, one explanation for this observation might also be that the initiation of replication is regulated very differently in yeast than in mammals since in the single cell organism, origins are defined by sequence, which is not the case in mammalian cells. How the origin is targeted to specific sites in the genome thus remains unclear. There have been several large-scale studies (Hiratani et al., 2010) (Ryba et al., 2011) to define origins of replication in human and mouse cells which have demonstrated that there is not a simple rule by which those genetic features can be defined. It remains a major unanswered question in biology to understand how DNA replication is initiated and timed in mammals.

## 2.4 Material and Methods

Cell culture, cell extraction and western blot

V6.5 mouse ESC were maintained on a feeder cell layer in ESC medium containing leukemia inhibitory factor (LIF), 15% fetal bovine serum, non-essential amino acids, glutamate and penicillin/streptomycin in knockout DMEM medium on gelatinized cell culture dishes. For differentiation, Zhbtc4 cells (Niwa et al., 2000) were induced with 2ug/ml doxycycline in LIF-free ESC medium for 96 hours on gelatinized cell culture dishes without feeders.

To generate whole cell extracts, cells were trypsinized and resuspended in lysis buffer (300mM Tris-HCl pH 6.8, 10 % SDS, 15 % glycerol, 1 mM DTT). The proteins were separated by gel electrophoresis on 4-12 % Acrylamide gels at 140 V at room temperature for 60 mins and transferred to nitrocellulose membranes. The membranes were blocked with Odyssey blocking buffer (LiCOR) for one hour at room temperature and incubated with primary antibodies as indicated over night at 4 C. After washing with PBS 0.1% Tween three times, membranes were incubated with secondary antibodies (1:10000, LiCOR) for one hour at room temperature. Protein bands were visualized using the LiCOR infrared scanner and analyzed with the corresponding software.

siRNA transfection, RNA isolation and qRT-PCR

V6.5 or ZHBTc4 ESC (Niwa et al., 2000)were plated at 10^5 cells per well in 6-well plates on gelatine without feeders and reverse transfected with 20 or 50 nm final siRNA (Dharmacon) concentration using invitrogen lipofectamine RNAimax according to manufacturer's instructions. After 24 hours, the medium was exchanged and forward transfected. The cells were trypsinized 48 hours after plating (unless stated otherwise) and resuspended in 500 ul Trizol (invitrogen). RNA was isolated with Qiagen RNAeasy kits according to manufacturer's instructions.

Superscript III Supermix (invitrogen) was used to generate random hexamer primed cDNA. The cDNA was diluted 1:5 with water and used in qRT-PCR with Roche Lightcycler reaction mix.

Microarray analysis

RNA was harvested as described and 2 ug RNA were submitted to a core facility for use on Affymetrix Mouse 430 plus 2.0 transcription microarrays.

Expression data was RMA normalized and Affy IDs were matched to corresponding RefSeq sequences. IDs belonging to the same RefSeq sequence were averaged to obtain expression levels by gene.

Clustering was performed using the cluster 3.0 software by uncentered hierarchical clustering and visualized using the javatree software package. Gene ontology analysis was performed using the online tool DAVID (Huang et al., 2008).

Imunoprecipitation and mass spectrometry

3xFlag-6xHis-tagged CGBP was targeted to the Collagen locus under control of the Tet operon of mouse ESC by FRT/FLPase mediated recombination as previously described (Beard et al., 2006). Cells were treated with 2 ug/ml doxycycline in ESC medium for 48 hours on gelatinized culture plates without feeders before harvesting. Nuclear extracts were generated by dignam extraction with 420mM salt as previously described (Carey et al., 2009). The extracts were dialized to 100 mM salt concentration and immunoprecipitation performed with anti-Flag M2 agarose beads (invitrogen) for 3 hours at 4 C. Complexes were eluted with 25 ug/ml flag peptide in 4 sequential elutions for 15 mins at room temperature. Each eluate was verified by western blot and silver stain before they were pooled. Pooled eluates were subjected to MudPIT mass spectrometry as previously described (Wolters et al., 2001).

## 2.5 References


Ahlander, J., Chen, X.-B., and Bosco, G. (2008). The N-terminal domain of the Drosophila retinoblastoma protein Rbf1 interacts with ORC and associates with chromatin in an E2F independent manner. PloS One *3*, e2831.

Auth, T., Kunkel, E., and Grummt, F. (2006). Interaction between HP1alpha and replication proteins in mammalian cells. Exp. Cell Res. *312*, 3349–3359.

Beard, C., Hochedlinger, K., Plath, K., Wutz, A., and Jaenisch, R. (2006). Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. Genes. New York N 2000 *44*, 23–28.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Cadoret, J.-C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L., Quesneville, H., and Prioleau, M.-N. (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc. Natl. Acad. Sci. U. S. A. *105*, 15837–15842.

Carey, M.F., Peterson, C.L., and Smale, S.T. (2009). Dignam and Roeder Nuclear Extract Preparation. Cold Spring Harb Protoc *2009*, pdb.prot5330.

Carlone, D.L., and Skalnik, D.G. (2001). CpG binding protein is crucial for early embryonic development. Mol. Cell. Biol. *21*, 7601–7606.

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

Farkash-Amar, S., Lipson, D., Polten, A., Goren, A., Helmstetter, C., Yakhini, Z., and Simon, I. (2008). Global organization of replication time zones of the mouse genome. Genome Res. *18*, 1562–1570.

Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.-W., Lyou, Y., Townes, T.M., Schübeler, D., and Gilbert, D.M. (2008). Global Reorganization of Replication Domains During Embryonic Stem Cell Differentiation. PLoS Biol *6*, e245.

Hiratani, I., Ryba, T., Itoh, M., Rathjen, J., Kulik, M., Papp, B., Fussner, E., Bazett-Jones, D.P., Plath, K., Dalton, S., et al. (2010). Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. Genome Res. *20*, 155 –169.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Kan, J., Zou, L., Zhang, J., Wu, R., Wang, Z., and Liang, C. (2008). Origin recognition complex (ORC) mediates histone 3 lysine 4 methylation through cooperation with Spp1 in Saccharomyces cerevisiae. J. Biol. Chem. *283*, 33803–33807.

Kelly, T.J., Jallepalli, P.V., and Clyne, R.K. (1994). Replication and transcription. Silence of the ORCs. Curr. Biol. CB *4*, 238–241.

Kneissl, M., Pütter, V., Szalay, A.A., and Grummt, F. (2003). Interaction and assembly of murine pre-replicative complex proteins in yeast and mouse cells. J. Mol. Biol. *327*, 111–128.

Lee, J.-H., and Skalnik, D.G. (2005). CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. J. Biol. Chem. *280*, 41725–41731.

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

Porter, I.M., McClelland, S.E., Khoudoli, G.A., Hunter, C.J., Andersen, J.S., McAinsh, A.D., Blow, J.J., and Swedlow, J.R. (2007). Bod1, a novel kinetochore protein required for chromosome biorientation. J. Cell Biol. *179*, 187–197.

Ryba, T., Hiratani, I., Sasaki, T., Battaglia, D., Kulik, M., Zhang, J., Dalton, S., and Gilbert, D.M. (2011). Replication timing: a fingerprint for cell identity and pluripotency. PLoS Comput. Biol. *7*, e1002225.

Saitoh, Y., Miyagi, S., Ariga, H., and Tsutsumi, K. (2002). Functional domains involved in the interaction between Orc1 and transcriptional repressor AIF-C that bind to an origin/promoter of the rat aldolase B gene. Nucleic Acids Res. *30*, 5205–5212.

Sequeira-Mendes, J., Díaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N., and Gómez, M. (2009). Transcription Initiation Activity Sets Replication Origin Efficiency in Mammalian Cells. PLoS Genet. *5*.

Snyder, M., He, W., and Zhang, J.J. (2005). The DNA replication factor MCM5 is essential for Stat1-mediated transcriptional activation. Proc. Natl. Acad. Sci. U. S. A. *102*, 14539–14544.

Snyder, M., Huang, X.-Y., and Zhang, J.J. (2009). The minichromosome maintenance proteins 2-7 (MCM2-7) are necessary for RNA polymerase II (Pol II)-mediated transcription. J. Biol. Chem. *284*, 13466–13472.

Takayama, M.A., Taira, T., Tamai, K., Iguchi-Ariga, S.M., and Ariga, H. (2000). ORC1 interacts with c-Myc to inhibit E-box-dependent transcription by abrogating c-Myc-SNF5/INI1 interaction. Genes Cells Devoted Mol. Cell. Mech. *5*, 481–490.

Voo, K.S., Carlone, D.L., Jacobsen, B.M., Flodin, A., and Skalnik, D.G. (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol. Cell. Biol. *20*, 2108–2121.

Wolters, D.A., Washburn, M.P., and Yates, J.R. (2001). An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem. *73*, 5683–5690.

# 3 Proteomic analysis of the time-specific interactome of proliferating cell nuclear antigen in embryonic stem cells

## 3.1 Abstract

In a comprehensive approach to identify PCNA-interacting proteins by multi-dimensional protein identification technology (MudPIT), we identified novel PCNA interaction motif (PIP box)-containing interaction partners in mouse embryonic stem cells (ESCs) amongst which a number are chromatin-modifying enzymes. The PCNA interactome from synchronized ESCs was determined in early/mid and late S-phase cells and shows significant differences. Wdhd1 is specifically associated with PCNA in late S-phase. Interestingly, the interaction of Wdhd1 with PCNA may be required for its localization. The interactome identified for Wdhd1 confirms its interaction with components of the DNA replication machinery and further suggests that it may also play a role in other molecular processes, such as chromatin state regulation, through interaction with components of the chromatin remodeling complex NURD. This study thus provides evidence of the coupling of the DNA replication machinery with chromatin modifying complexes.

## 3.2 Introduction

Faithful replication of the genome and transmission of chromatin states to daughter generations of cells is crucial for development and the prevention of aberrations (Alabert and Groth, 2012).

The replication of the genome occurs during S-phase of the cell cycle. In a first step, origins of replication are bound by the prereplicative complex (pre-RC). This complex contains the origin recognition complex (ORC) and the replicative helicase minichromosome maintenance complex (MCM). Upon activation of origin firing, DNA polymerases alpha, delta and epsilon and other factors are recruited to the origin to start DNA synthesis. A crucial factor for the processivity of

the DNA polymerase is proliferating cell nuclear antigen (PCNA), which is loaded on to the DNA strand by the replication factor C (RFC) complex. PCNA is a conserved member of the sliding clamp protein family and forms a homotrimeric ring encircling DNA (Moldovan et al., 2007). PCNA has a plethora of interaction partners, many of which interact through a conserved PCNA interacting protein motif (PIP box) with the consensus amino acid motif sequence Q-x-[x]-I/L/V-x-[x]-F/Y/W/H-F/Y/W/H where x stands for any amino acid, a dash signifies alternative amino acids and brackets stands for the optional insertion of an amino acid (Warbrick, 1998). Among the known interaction partners of PCNA are proteins involved in DNA replication and repair, cell cycle control and survival, transcription and chromatin modification (De Biasio and Blanco, 2013). This suggest that PCNA coordinates different functions at the replication fork by temporally controlled interactions with various complex, prompting the term ‘maestro of the replication fork’ (Moldovan et al., 2007).

In this work, we demonstrate that the DNA replication factor WD repeat and HMG-box DNA binding protein 1 Wdhd1 (also known as Ctf4 or And-1 for acidic nucleoplasmic DNA-binding protein) contains a PIP box and interacts with PCNA.

Wdhd1 had originally been identified in a screen for yeast mutants involved in chromosome transmission fidelity (Ctf4) (Kouprina et al., 1992). It is required for the recruitment of DNA polymerase alpha to chromatin (Zhu et al., 2007) and stimulates its activity (Bermudez et al., 2010). It is furthermore involved in pre-RC assembly (Li et al., 2012a) and accordingly knockdown of Wdhd1 leads to delays in the progression through S-phase (Yoshizawa-Sugata and Masai, 2009). In addition to these functions, Wdhd1 is also required for the replication of centromeres (Jaramillo-Lambert et al., 2013)(Hsieh et al., 2011). Besides its role as a DNA replication factor, it also regulates the stability of the histone acetyl transferase Gcn5 (Li et al., 2012b). We identify further chromatin-modifying enzymes as interaction partners of Wdhd1,

which suggests that it could be an additional link in the coupling of DNA replication with the replication of chromatin states through the cell cycle.

## 3.3 Results

Creation of an embryonic stem cell line inducibly overexpressing Flag-PCNA

To be able to better study the interactome of PCNA, we created an ESC line by stably integrating a C-terminally 3xFlag-tagged PCNA construct into the 3' end of the Collagen locus under the control of the Tet-inducible promoter, as was previously described (Beard et al., 2006) (Figure 3.1 A). This creates a cell line that can be induced to express Flag-PCNA by addition of doxycycline (DOX), which we termed 'PC'. The successful induction of Flag-PCNA expression was tested by western blotting (Figure 3.1 B) and total PCNA levels in the induced ESC were quantified to be about 1.5 fold higher than the endogenous levels in un-induced cells. The cellular level of PCNA is regulated dependent on the cell cycle (Kurki et al., 1986). This makes it a critical requirement that exogenous PCNA levels remain similar to the endogenous levels to not perturb the function of DNA replication and cell cycle progression. To determine if the localization pattern of PCNA is affected by either the overexpression or the addition of the Flag-tag, we performed immunofluorescence staining on PC ESCs that were induced with DOX (Figure 3.1 C), and found the typical PCNA staining pattern with the staining patterns of anti-Flag and anti-PCNA antibodies overlapping. This together with the relatively minor PCNA overexpression demonstrates that Flag-PCNA follows the behavior of endogenous PCNA as measured by protein levels and localization, making this system a good tool to study the protein interaction network of the replication fork clamp in ESCs.

PCNA interactome in unsynchronized cells

To define the PCNA interactome in our ESC line, we performed Flag-immunoprecipitation on nuclear extracts. The successful purification of PCNA was confirmed by silver staining of

immunoprecipitates, which showed bands corresponding to the sizes of endogenous and Flag-PCNA (marked by stars, Figure 3.2 A), which do not appear in a control Immunoprecipitation from a ESC line not expressing any flag-tagged protein. Furthermore, this result is also confirmed by western blotting with a PCNA antibody (Figure 3.2 B). As both the silver stain and western blot showed both bands for endogenous and Flag-tagged PCNA, it is confirmed that Flag-PCNA interacts with endogenous PCNA, which is known to form a homotrimer. This demonstrates further, that C-terminally tagging PCNA is incorporated into the replication clamp and likely does not interfere with its function.

To identify interaction partners, the flag-immunoprecipitates were subjected to multidimensional protein identification technique (MudPIT) by mass spectrometry (Wolters et al., 2001). We identified 93 proteins that are more abundant than the average protein identified in the mass spectrometric run in the Flag-PCNA immunoprecipitates than in the control (Supplementary Table 3.1). Since many proteins interact with PCNA through the conserved PIP-box motif Q-x-[x]-I/L/V-x-[x]-F/Y/W/H-F/Y/W/H (Moldovan et al., 2007), we performed a motif search in the reference sequence mouse proteome and identified 5279 proteins that contain this motif (Supplementary table 1.2, provided electronically). 23 proteins (about 25%) that were found in our Flag-PCNA co-immunoprecipitation contain such a sequence motif, making them interesting candidates as direct PCNA interactors.

To better understand which of the co-immunoprecipitating proteins identified in our work were known interactors of PCNA, we compared our results to known PCNA interacting protein and protein-protein interactions that are deposited in the STRING database (Supplementary Table 3.2) (Franceschini et al., 2013). The result of this comparison is shown in Figure 3.2 C. Known interaction partners are depicted as diamond-shaped nodes. Protein-protein interactions are presented as black edges if they are present in the STRING database and as green edges if they were identified in this experiment. Proteins of which we identified a PIP motif in their amino

acid sequence are shown with a red edge around the node. If a PIP box has been described in the literature, the protein's node shown with a wide boarder (Moldovan et al., 2007). Furthermore, proteins that are annotated in the gene ontology term 'covalent chromatin modification' are highlighted with the blue node fill color.

Amongst the 93 identified PCNA interactors, we find 12 proteins associated with the GO term covalent chromatin modification and 23 known PCNA interaction partners. Of the 13 known PCNA interactors that contain a PIP box described in the literature (any organism was included since most studies were done in human or yeast cells), our script assigned 7 with a PIP box motif in the mouse amino acid sequence.

This analysis demonstrated that several well-described PCNA interactors are among our FLAG-PCNA co-immunoprecipitating proteins, which validates our comprehensive approach. Examples are the replication factor C (RFC) complex that loads PCNA to DNA and which consists of the subunits RFC1-5 (Majka and Burgers, 2004), the leading and lagging strand DNA polymerases delta and epsilon which require PCNA for increased processivity (Chilkova et al., 2007), the cell cycle regulators CDK2 and CDK4 (He et al., 2013) (Koundrioukoff et al., 2000), the mismatch repair complex MutS$\alpha$ consisting of MSH2 and MSH6 (Zlatanou et al., 2011). Consistent with a role of PCNA as a mediator for the re-establishment of chromatin modifications through its various interaction partners, we also identified a number of co-immunoprecipitating proteins that are annotated in the gene ontology class 'covalent chromatin modification'(Gondor and Ohlsson, 2009)(Probst et al., 2009). Amongst those is a very well researched PCNA interaction partner, DNA methyltransferase 1, which maintains DNA methylation patterns after replication (Schermelleh et al., 2007)(Schneider et al., 2013). Interestingly, we also identify O-GlcNAc transferase (Ogt) as a chromatin modifying enzyme with a previously undescribed PIP box motif. Ogt has been shown to target PCNA for O-GlcNAcylation (Dehennaut et al., 2008). Therefore, our results indicate that this enzymatic activity may be targeted through the PCNA interaction

motif. Another interesting chromatin remodeling enzyme with a previously unidentified PIP box is the histone methyl transferase Ehmt2 (also known as G9a). It has been shown to co-localize with PCNA at replication foci and to directly interact with DNMT1 (Estève et al., 2006). An interaction of Ehmt2 with PCNA has recently benn demonstrated, however it remained an open question through which domain this interaction is being facilitated (Yu et al., 2012).Our results further suggest that Ehmt2 may be directly targeted to replication foci through the PIP box motif. Another interesting protein that has been thought to indirectly interact with PCNA through DNMT1 is the histone demethylase Kdm1a (LSD1), for which we also identified a putative PIP box motif that may facilitate a direct interaction. We also found a number of interaction partners that have known interactions with each other in our set of PCNA interaction partners, indicating that they may interact with the replication clamp as complexes. One of these examples is a network of proteins containing Wdr5, Ruvbl1, Ruvbl2, Uchl5, Wdr1 and Smarca5 (also known as ISWI or SNF2H). Smarca5 is a chromatin remodeling factor involved in the replication of heterochromatin and had previously been identified to interact with PCNA, but a PIP box had not been described (Poot et al., 2004). A second complex we identify is composed of Clathrin heavy chain (Cltc) and the AP2 complex (Ap2a1). Clathrin and the AP2 complex are involved in endocytosis and apoptosis and have not yet been shown to interact with PCNA (Chen et al., 2013). The observation of a PIP box motif in the amino acid sequence of clathrin heavy chain indicates that this interaction might be direct, posing the question if PCNA and the replication fork are also linked to apoptotic and endocytotic processes. We furthermore identified a network of proteins consisting of Pbrm1, Smarce1, Arid1a and Smarcb1, corresponding to components of the SWI/SNF-B (PBAF) chromatin-remodeling complex (Xue et al., 2000), which has also not yet been directly linked to DNA replication.

In summary, our approach confirmed known PCNA interactors, but more importantly may have identified interesting new interaction partners for future research of how DNA replication might be integrated with other cellular functions.

Identification of time-specific PCNA interaction partners

PCNA has a plethora of interaction partners and for some of them, a time-specific association has been demonstrated (Gondor and Ohlsson, 2009). It has been proposed that this time-dependent interaction may contribute to the replication of chromatin states with euchromatic regions being replicated in early S-phase whereas heterochromatic regions are replicated in late S-phase.

To gain more insight in the time-dependent PCNA interactome, we determined PCNA interaction partners in early/mid S-phase and in late S-phase. To this end, we synchronized ESCs by treatment with the microtubule inhibitor Nocodazole, which arrests cells reversibly in G2/M phase (Baumgartner et al., 1999). Cells were harvested at 5 hours and 9 hours after release from the cell cycle block, corresponding to early/mid and late S-phase. The successful synchronization was confirmed by flow cytometric analysis (Figure 3.3 A). Nuclear extracts and Flag-immunoprecipitation were performed for Flag-PCNA expressing cells in both stages (Figure 3.3 B) and the immunoprecipitates were then subjected to MudPIT analysis as described above. We identified 104 co-purifying proteins in the sample collected at 5 hours post-release and 107 at 9 hours post-release (Supplementary Tables 3.3 and 3.4 and Supplementary Figures 3.1 and 3.2). Amongst those, 76 were found in both samples (Figure 3.3 C) and represent the proteins that associate with PCNA throughout most of S-phase. Network analysis of these confirms the findings from the analysis of the interactome from unsynchronized ESCs (Figure 3.4). Interestingly, only 28 proteins co-purified with PCNA specifically in early/mid S-phase (5 hours) and 31 proteins were specific to late S-phase (9 hours) (Figure 3.3C, Figure 3.4). In accordance

with previous reports, the chromatin modifying enzyme DNA methyltransferase 1 (DNMT1) is permanently associated with PCNA as it was found for both time points (Rountree et al., 2000) (Estève et al., 2006). Consistent with the idea that PCNA associates primarily with chromatin-modifying enzymes required for the establishment of euchromatin in early S-phase, we identify Wdr5 and Tet1 specifically in the early/mid S-phase sample. Wdr5 is part of the Set1/COMPASS-histone methyltransferase complex which establishes the euchromatic mark of trimethylated histone 3 lysine 4 (Takahashi et al., 2011). Tet1 is an enzyme involved in the hydroxylation of methylated DNA, which has been linked to DNA demethylation processes, and has recently been described to be regulated by Ogt, which we identified as novel PCNA interactor in unsynchronized ESCs (Ito et al., 2010)(Shi et al., 2013).

Interestingly we also identified a number of chromatin-modifying enzymes required for the establishment of heterochromatin in the immunoprecipitates that are specifically co-purifying with PCNA in late S-phase (9 hours post-release). Amongst those were Cbx3 and Smarcad1. Cbx3 is heterochromatin protein 1 gamma (HP1), which binds to the heterochromatic histone mark of trimethylated histone 3 lysine 9 and plays a critical role for reprogramming of differentiated cells to induced pluripotent stem cells (Sridharan et al., 2013). Smarcad1 has been shown to be required for the replication of heterochromatin and to directly interact with PCNA (Rowbotham et al., 2011).

In addition to these chromatin-modifying enzymes, we also identified a number of novel interaction partners with a time-specific association with PCNA. To start validating some of those, we decided to focus on those that contained a putative PIP box motif. With these follow up studies, we and others (data to be published elsewhere, (Yuan et al.)) were able to confirm the PIP box dependent interaction of Zranb3 and Wdhd1 with PCNA (next chapter and see below). Both proteins were isolated along with PCNA at the 9hrs release sample.

Validation of the interaction of PCNA with Wdhd1

Wdhd1 (also known as Ctf4 or And-1) is a 126 kDa protein that is critical for DNA replication (Li et al., 2012a)(Zhu et al., 2007) and has not yet been shown to directly interact with the replication factor PCNA.

Wdhd1 interacts with a number of DNA replication factors and knockdown of Wdhd1 leads to delays in S-phase (Yoshizawa-Sugata and Masai, 2009), highlighting its importance for DNA replication. Interestingly, it was observed that Wdhd1 shows a fine granular pattern in early S-phase and a coarse clumpy pattern in late S-phase (Zhu et al., 2007). Wdhd1 is also thought to be involved in regulating histone acetylation levels by stabilizing the histone acetyl transferase Gcn5, which demonstrates that Wdhd1 may have additional roles outside of DNA replication (Li et al., 2012b).

To validate the interaction with PCNA and to gain additional insight into the molecular function of Wdhd1, we created a ESC line that expressed C-terminally 3xFlag-tagged Wdhd1 upon addition of doxycycline, similar to the ESC line described above for PCNA expression, and performed Flag-immunoprecipitation from nuclear extracts of this cell line. Silver staining of the immunoprecipitates confirms the successful purification of Wdhd1 as a band corresponding to the size of Wdhd1 appears in Flag-Wdhd1 but not the control sample (the untargeted cell line) (Figure 3.5 A). Notably, the interaction with PCNA was confirmed by western blotting (Figure 3.5 B). Additionally, we also identified the DNA mismatch repair proteins Msh2 and Msh6 as novel Wdhd1 interaction partners, implying a novel role of Wdhd1 not only in DNA replication and chromatin modification but also in DNA repair. To gain insight into additional functions and further validate the PCNA interaction, we also subjected the Flag-Wdhd1 immunoprecipitates to MudPIT analysis (Figure 3.5 C and Supplementary Table 3.5). We identified 161 proteins that co-immunoprecipitated with Wdhd1. Amongst them are known interaction partners such as DNA

polymerase alpha, delta and epsilon, the RFC complex and the MCM complex, and PCNA, further confirming the interaction of PCNA with Wdhd1. In accordance with the result of the western blotting, Msh2 and Msh6 are also identified by MudPIT.

Gene ontology (GO) analysis of the identified Wdhd1 interactome shows a number of enriched terms. Among the most significantly enriched GO term were cell cycle and cell division, which matches Wdhd1's described role as crucial component for DNA replication. It is interesting to note that we observed components of the chromatin remodeling NURD complex such as Chd4, Mta2, Mta3, Sin3a and Rbbp4 as well as subunits of other chromatin-modifying enzymes such as Suz12 and Wdr5 co-purifying with Wdhd1. This may suggest that Wdhd1 also plays a role in the maintenance of chromatin states in ESCs, as seen by its interaction with the histone acetyl transferase Gcn5.

To evaluate the requirement of the PIP box motif in Wdhd1 for its localization, we generated a plasmid carrying Flag-Wdhd1 with PIP-mutation that had previously been shown to completely abrogate PCNA binding (Q→A, as in (Chuang et al., 1997)).

It is surprising that immunofluorescence staining of transiently transfected Flag-tagged wildtype or PIP-mutant Wdhd1 in either ESCs or mouse embryonic fibroblasts (MEFs) does not show co-localization with PCNA in the majority of the cells (Figure 3.6 A). However, it had been speculated that Wdhd1 binding to DNA templates may compete with PCNA and RFC binding (Bermudez et al., 2010). This indicates that Wdhd1 binding to DNA is not dependent on its interaction with PCNA and that this interaction could serve another molecular function such as the regulation of protein stability or covalent modification. Wdhd1 shows staining patterns from a fine granular staining to larger foci as was observed previously (Zhu et al., 2007). The foci shown in Figure 3.6 A co-localize with DAPI-dense regions that correspond to heterochromatic regions. It should be noted that since PCNA typically also colocalizes with DAPI-dense regions

especially late in S-phase when heterochromatin is replicated, a possible explanation for the lack of colocalization in our experiments could also be that the PCNA staining protocol was not ideal and thus, for technical reasons, we did not observe a colocalization of these two proteins.

The percentage of cells that show the punctuate Wdhd1 staining pattern is independent of the presence of the PIP box motif as there is not a drastic difference between cells expressing wildtype or PIP mutant Wdhd1 (Figure 3.6 B). However, the percentage of cells showing this staining pattern increases from early to late S-phase (Figure 3.6 C) which suggests that these foci correspond to late-replicating heterochromatic regions such as pericentromeric heterochromatin, in accordance to Wdhd1's function in CENP-A recruitment (Jaramillo-Lambert et al., 2013).

**Figure 3.1: Creation of the ESC line 'PC' that expresses 3xFlag-PCNA upon induction with doxycycline.**

A) Schematic of the targeted locus. C-terminally Flag-tagged PCNA was integrated under the control of the Tet-inducible promoter at the 3' end of the Collagen1a (Col1a) locus using site specific Flp-mediated recombination. Upon addition of doxycycline (DOX), rtTA expressed under control of the constitutive ROSA26 (R26) promoter, binds to the tet-inducible promoter and induces expression of the integrated construct. SA = splice acceptor, SV40 pA = SV40 polyadenylation signal, RBG pA = rabbit beta globin polyadenylation signal, M2rtTA = M2 tetracycline reverse transcriptional activator. B) Confirmation of Flag-PCNA induction by western blot. PC ESCs were induced with either 2 ug/ml doxycycline or no drug for 24 hours and whole cell extracts were analyzed by western blot with a Flag or PCNA specific antibody. Left: western blot. Right: quantification of the western blot relative to PCNA levels in un-induced cells. C) Top: Immunofluorescence image of an undifferentiated ESC colony that was induced with 2 ug/ml doxycycline for 48 hours, stained with antibodies targeting Flag (green) and PCNA (red), respectively. DNA was visualized using DAPI. Bottom: zoom in to a cell as shown in the box in picture of colony.

**Figure 3.1: Creation of the ESC line 'PC' that expresses 3xFlag-PCNA upon induction with doxycycline.**

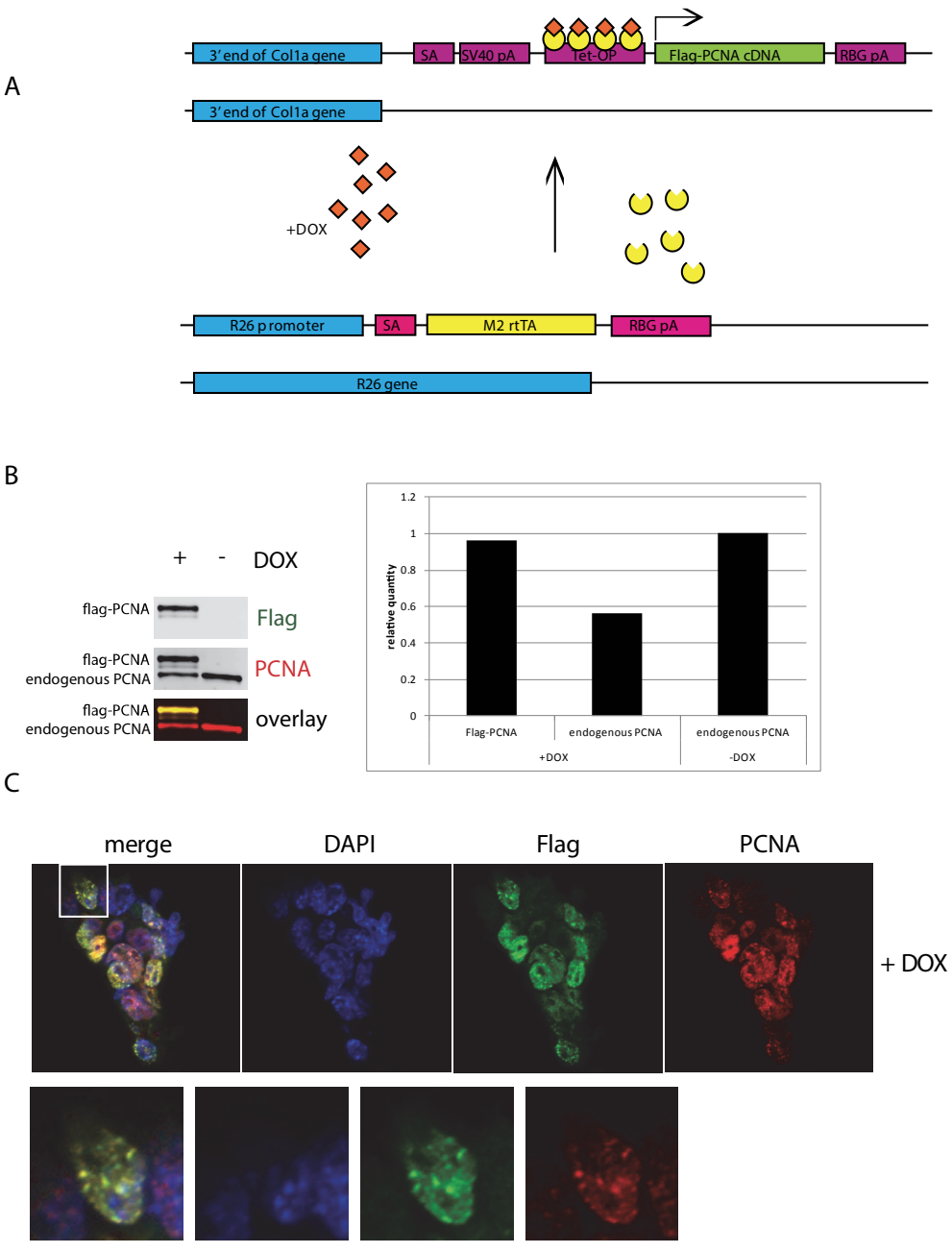


**Figure 3.2: Immunoprecipitation and MudPIT analysis from unsynchronized PC ESC.**

A) Silver stain of fractions of the Flag-immunoprecipitation form either PC (Flag-PCNA expressing) or control cells (un-targeted control ESC line) that were induced with 2 ug/ml doxycycline for 48 hours. L: ladder, INP: input, SUP: supernatant, WASH3/WASH4: flow-through wash fractions, IP1-IP4: Immunoprecipitation fractions obtained by sequential elution with flag-peptide. Asterix's mark bands with the expected sizes of Flag-PCNA and endogenous PCNA. B) Western blot of the same fractions as in A) probed with an antibody targeting PCNA. C) Network graphic of all specific PCNA-interacting proteins identified by MudPIT. Known PCNA interaction partners as found in the STRING database are depicted as diamond-shaped nodes, novel interactors identified in this study as circles. Known interactions as found in the STRING database are represented as black edges, novel interactions as green edges. If an interaction partner has a PIP box motif in its amino acid sequence, it will be shown as red node boarder and otherwise black. Proteins that contain validated PIP boxes described in the literature will have a bold node boarder. Interaction partners that fall under the gene ontology term 'covalent chromatin modification' are filled with blue node color, all others yellow.

**Figure 3.2: Immunoprecipitation and MudPIT analysis from unsynchronized PC ESC.**

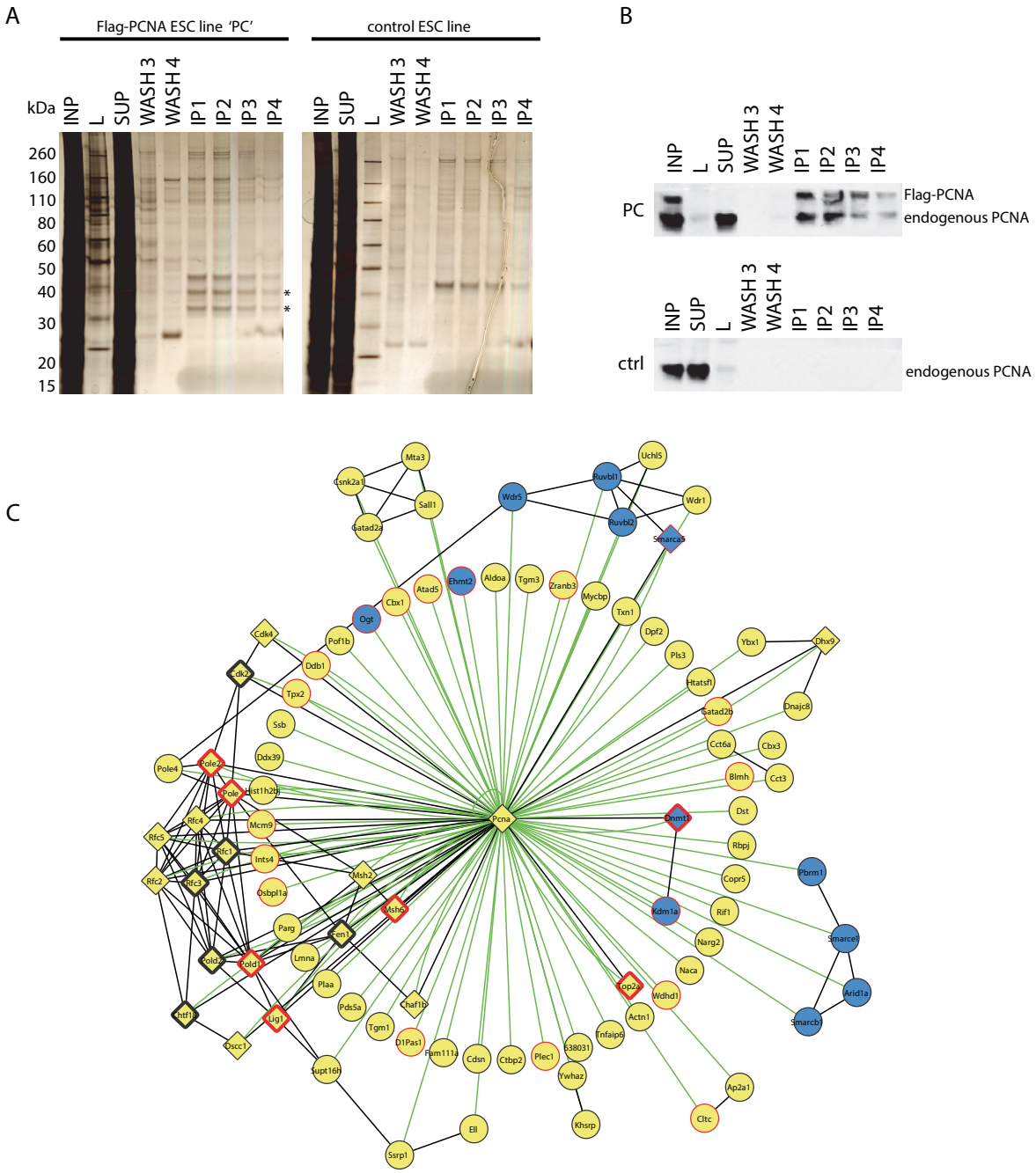
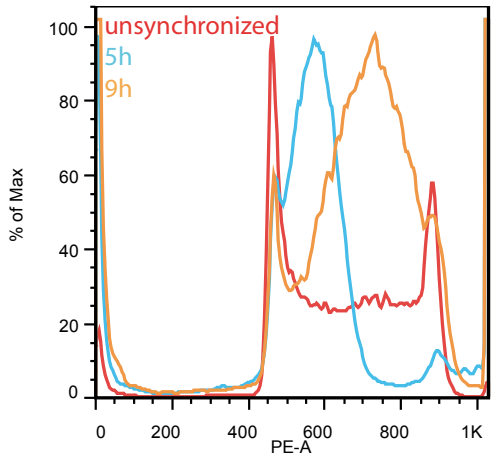


**Figure 3.3: Synchronization and Immunoprecipitation from PC ESC cells.**

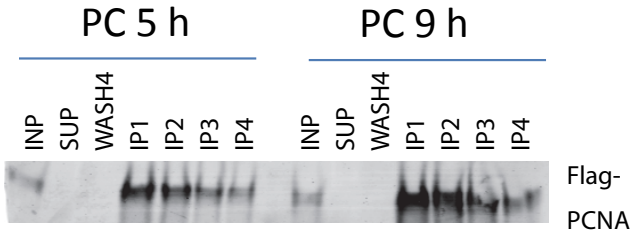
A) Flow cytometric analysis of DNA content as measured by incorporation of propidium iodide (PE) of unsynchronized (red) PC ESCs, and cells that were released from nocodazole arrest 5 hours (blue) or 9 hours (orange), respectively. All cells had been induced to overexpress Flag-PCNA by incubation with 2 ug/ml doxycycline for 48 hours. B) Western blot analysis of immunoprecipitation fractions from synchronized cells (5 and 9 hours) with a flag-specific antibody. INP: input, SUP: supernatant, WASH3/WASH4: flow-through wash fractions, IP1-IP4: Immunoprecipitation fractions obtained by sequential elution with flag-peptide. C) Venn diagram depicts the the overlap proteins co-immunoprecipitating with Flag-PCNA in the 5 hours and 9 hours post nocodazole-release samples.

**Figure 3.3: Synchronization and Immunoprecipitation from PC ESC cells.**

A




B




C

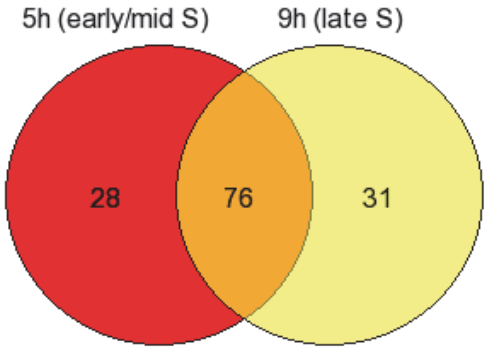


**Figure 3.4: Network analysis of Flag-PCNA co-immunoprecipitates from synchronized cells in early/mid or late S-phase.**

Top: Representation of PCNA interaction partners found to be associated with Flag-PCNA in both the early/mid (5 hours) and late (9 hours) S-phase samples. Bottom: Interaction networks of proteins identified to be specifically associated with PCNA in only one of the samples at 5h or 9h post nocodazole release. The same legend as in Figure 3.2 applies.

**Figure 3.4: Network analysis of Flag-PCNA co-immunoprecipitates from synchronized cells in early/mid or late S-phase.**

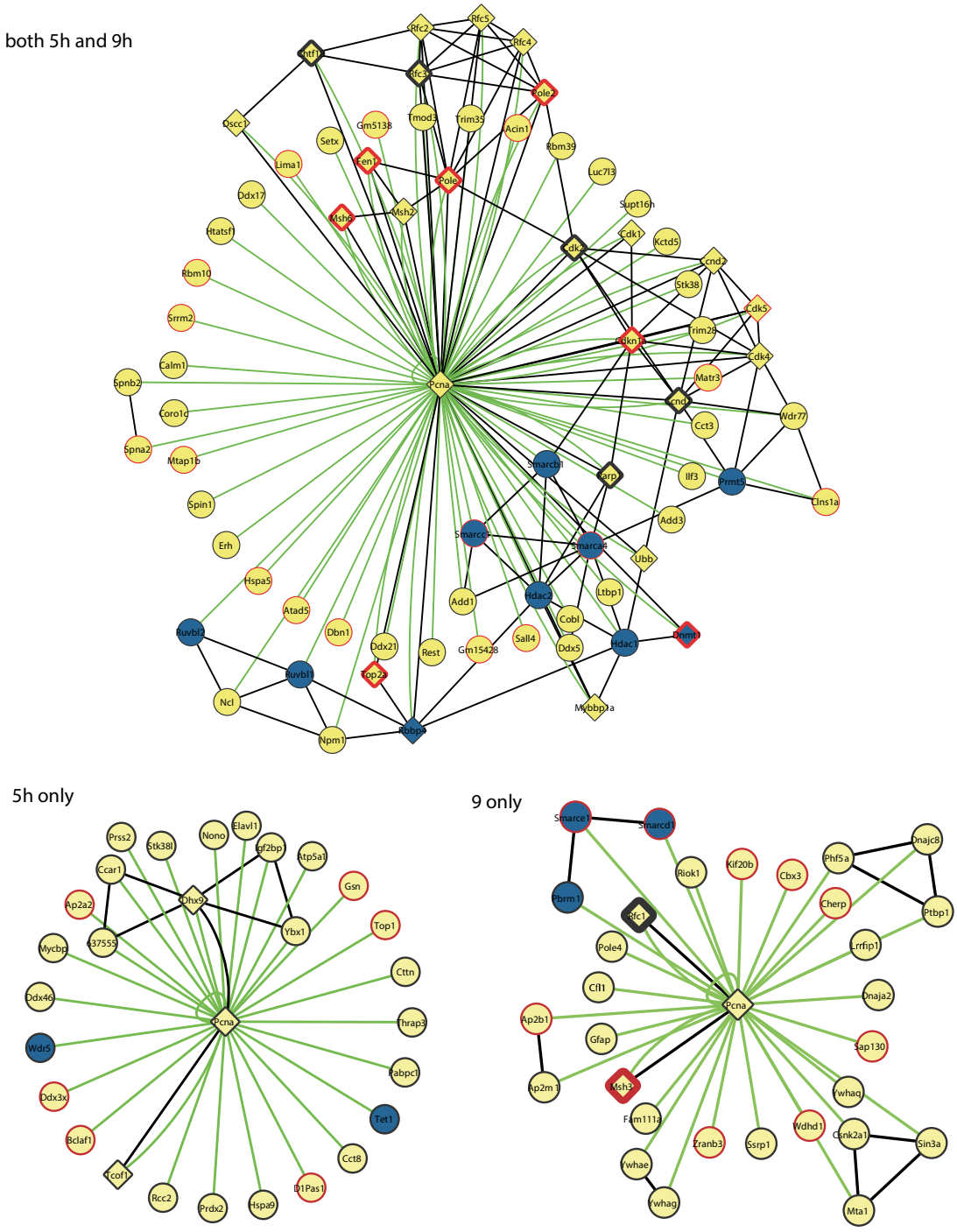


**Figure 3.5: Immunoprecipitation and network analysis of interaction partners identified for Flag-Wdhd1.**

A) Silver stain of fractions of the Flag-immunoprecipitation form either Flag-Wdhd1 expressing ESCs or control ESCs (un-targeted control cell line) that were induced with 2 ug/ml doxycycline for 48 hours. L: ladder, INP: input, SUP: supernatant, IP: pooled immunoprecipitation fractions obtained by sequential elution with Flag-peptide. The arrow marks the band with the expected sizes of Flag-Wdhd1 and endogenous Wdhd1 (126 kDa). B) Western blot of the same fractions as in A) probed with an antibodies targeting Flag, PCNA, Msh2 and Msh6. C) Network graphic of all Flag-Wdhd1-interacting proteins identified by MudPIT. The same legend as in Figure 3.2 applies. D) Analysis of condensed GO term (GOslim term) enrichment in the Wdhd1-interacting protein set, ordered by significance.

**Figure 3.5: Immunoprecipitation and network analysis of interaction partners identified for Flag-Wdhd1.**

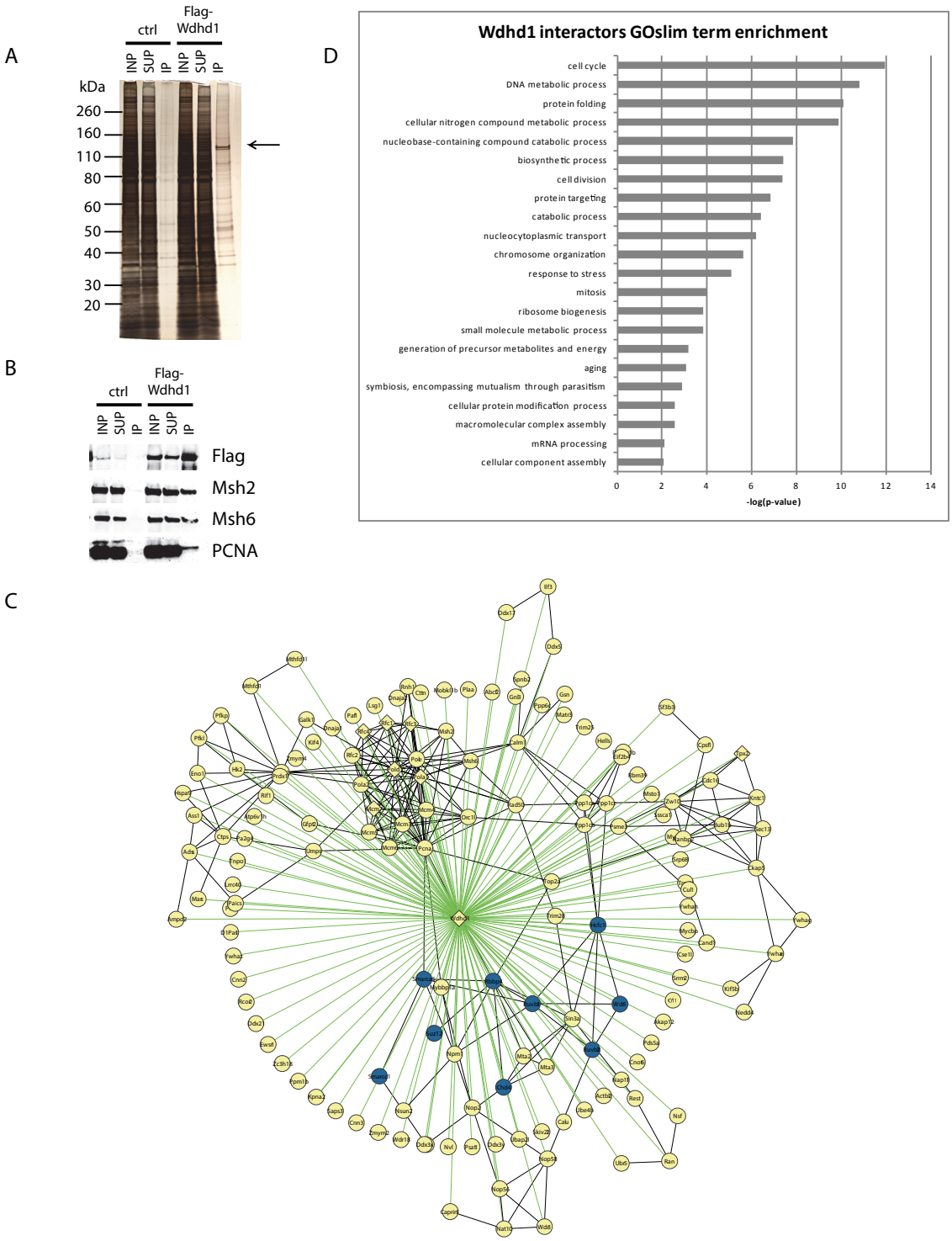
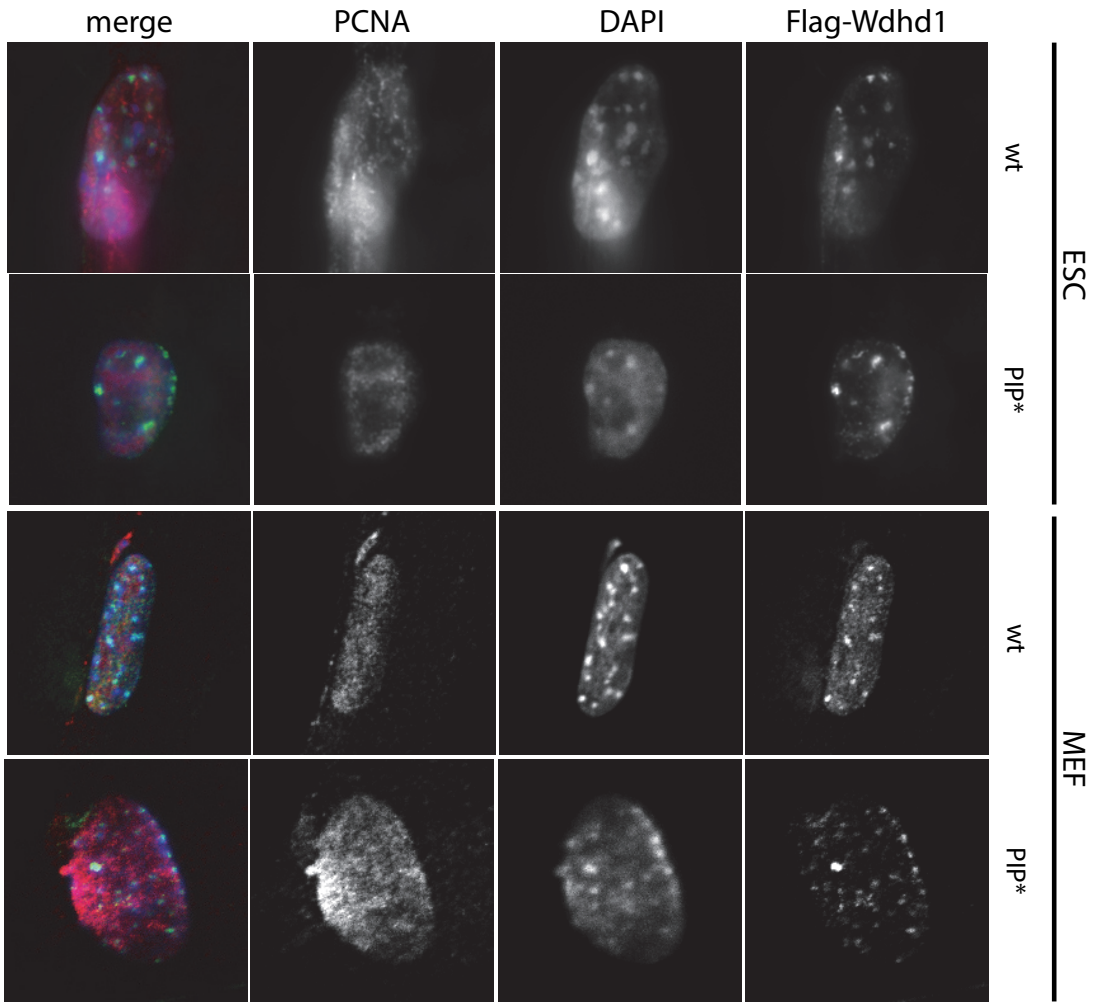


**Figure 3.6:-Wdhd1 enriches in DAPI-intense regions of the ESC and MEF nucleus.**

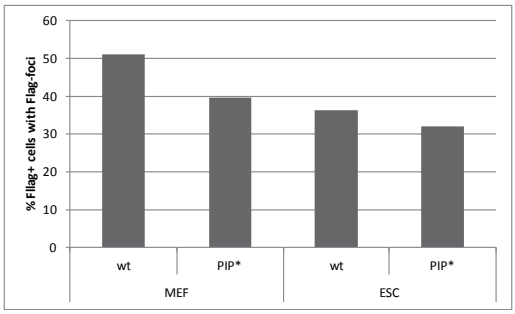
A) Immunofluorescence images of ESCs or MEFs that transiently express either wildtype (wt) or PIP box mutant (PIP*) Flag-Wdhd1, were stained with antibodies against PCNA or Flag. DNA was visualized using DAPI. In the merged image, PCNA is represented in red, Flag in green and DAPI in blue. B) Quantification of the percentage of cells exhibiting the focal nuclear staining pattern as shown in A). All cells staining positive for Flag-Wdhd1 after extraction of soluble proteins were scored and the percentage of cells showing the focal staining pattern is depicted as a percentage of all cells that stain positive for Flag-Wdhd1, for MEFs and ESCs. C) Quantification of the percentage of cells showing the focal staining pattern (scored as in B) in unsynchronized ESC that express Flag-Wdhd1 (upon addition of doxycycline) and in ESCs that were released from nocodazole arrest block at either 5 hours (corresponding to early/mid S-phase) or 9 hours (corresponding to late S-phase).

**Figure 3.6: Wdhd1 enriches in DAPI-intense regions of the ESC and MEF nucleus.**

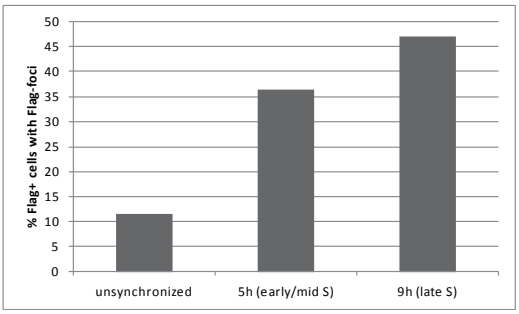
A




B



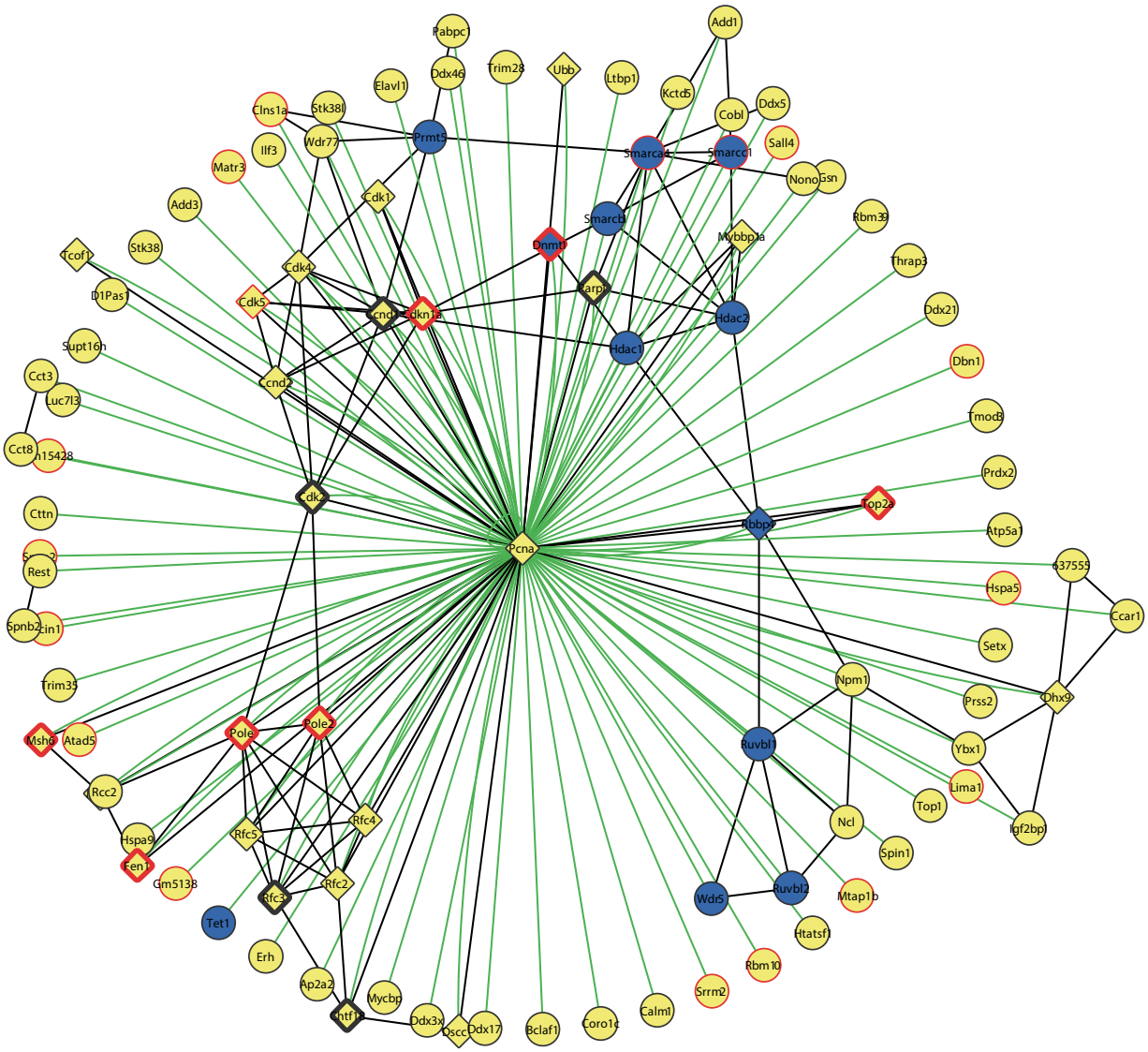

C



**Supplementary Figure 3.1: Network analysis of Flag-PCNA protein network from synchronized cells in early/mid S-phase (5 hours).**

Representation of all PCNA interaction partners found to be associated with Flag-PCNA in the early/mid S-phase samples. The same legend as in Figure 3.2 applies.

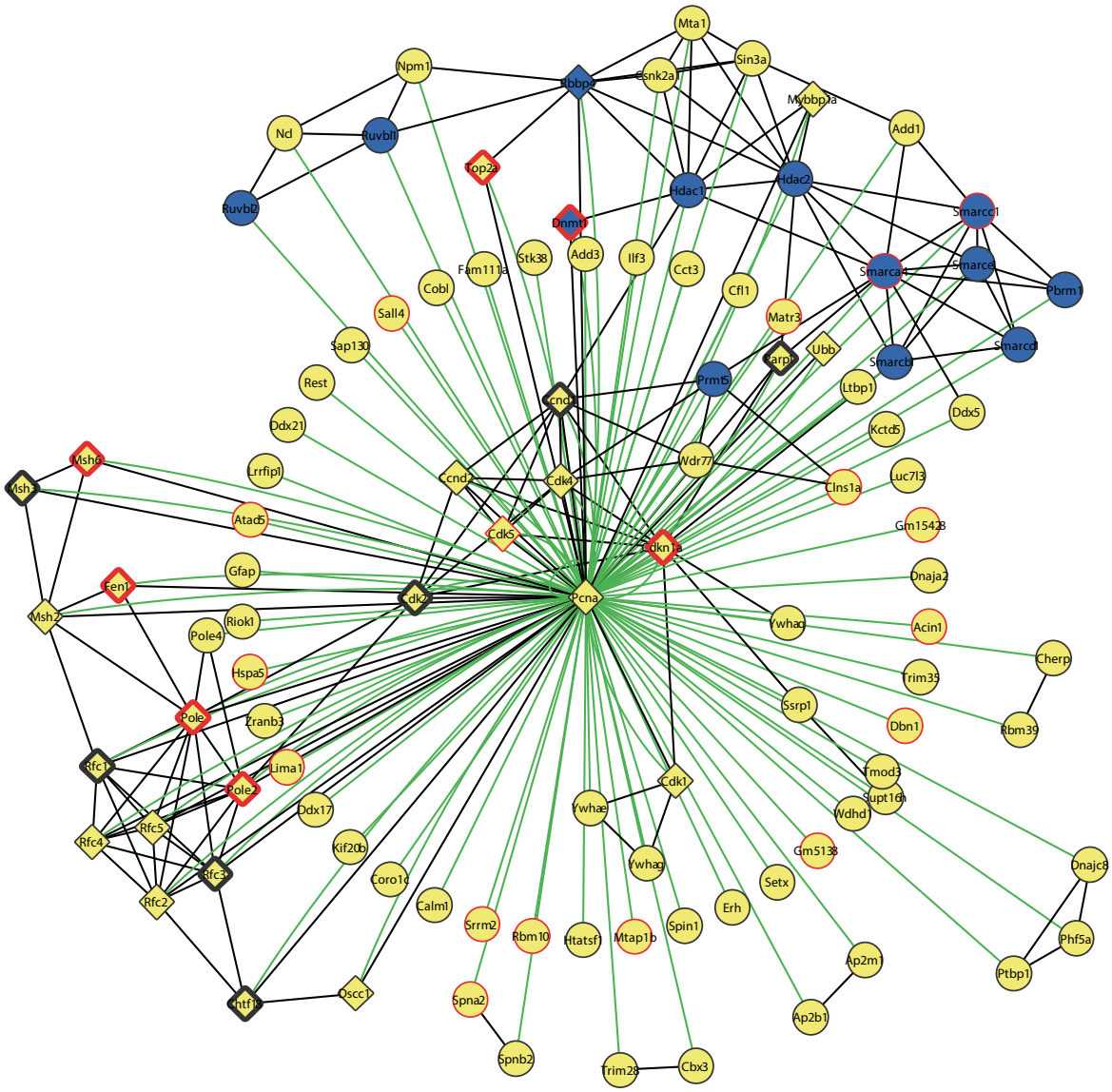
**Supplementary Figure 3.1: Network analysis of Flag-PCNA protein network from synchronized cells in early/mid S-phase (5 hours).**



**Supplementary Figure 3.2: Network analysis of Flag-PCNA protein network from synchronized cells in late S-phase (9 hours).**

Representation of all PCNA interaction partners found to be associated with Flag-PCNA in the late S-phase samples. The same legend as in Figure 3.2 applies.

**Supplementary Figure 3.2: Network analysis of Flag-PCNA protein network from synchronized cells in late S-phase (9 hours).**



## Supplementary Table 3.1: Unsynchronized PCNA MudPIT results.

| name | description | Uniprot ID | PCNA unsynch NSAFe5 | BG NSAFe5 | PCNA-BG |
|---|---|---|---|---|---|
| **Pcna** | Proliferating cell nuclear antigen | P17918 | 18891.73 | 310.11 | 18581.62 |
| **Rfc4** | Replication factor C subunit 4 | Q99J62 | 925.50 | 0.00 | 925.50 |
| **Rfc3** | Replication factor C subunit 3 | Q8R323 | 924.79 | 0.00 | 924.79 |
| **Rfc2** | Replication factor C subunit 2 | Q9WUK4 | 921.40 | 0.00 | 921.40 |
| **Rfc5** | Replication factor C subunit 5 | Q9D0F6 | 654.97 | 0.00 | 654.97 |
| **Rfc1** | Replication factor C subunit 1 | P35601 | 643.11 | 0.00 | 643.11 |
| **Msh2** | DNA mismatch repair protein Msh2 | P43247 | 548.64 | 0.00 | 548.64 |
| **Msh6** | DNA mismatch repair protein Msh6 | P54276 | 462.32 | 0.00 | 462.32 |
| **Tnfaip6** | Tumor necrosis factor-inducible gene 6 protein | O08859 | 361.94 | 0.00 | 361.94 |
| **Mycbp** | C-Myc-binding protein | Q9EQS3 | 297.34 | 0.00 | 297.34 |
| **Cbx3** | Chromobox protein homolog 3 | P23198 | 251.03 | 0.00 | 251.03 |
| **Fen1** | Flap endonuclease 1 | P39749 | 202.55 | 0.00 | 202.55 |
| **Pole4** | DNA polymerase epsilon subunit 4 | Q9CQ36 | 194.66 | 0.00 | 194.66 |
| **Top2a** | DNA topoisomerase 2-alpha | Q01320 | 190.41 | 0.00 | 190.41 |
| **Chtf18** | Chromosome transmission fidelity protein 18 homolog | Q8BIW9 | 189.63 | 0.00 | 189.63 |
| **Ppia** | Peptidyl-prolyl cis-trans isomerase A | P17742 | 186.74 | 0.00 | 186.74 |
| **Pole** | DNA polymerase epsilon catalytic subunit A | Q9WVF7 | 167.68 | 0.00 | 167.68 |
| **Csnk2a1** | Casein kinase II subunit alpha | Q60737 | 156.65 | 0.00 | 156.65 |
| **DSCC1** | Sister chromatid cohesion protein DCC1 | Q14AI0 | 153.51 | 0.00 | 153.51 |
| **Txn** | Thioredoxin | P10639 | 145.84 | 0.00 | 145.84 |
| **Pole2** | DNA polymerase epsilon subunit 2 | O54956 | 145.28 | 0.00 | 145.28 |
| **Lig1** | DNA ligase 1 | P37913 | 125.38 | 0.00 | 125.38 |
| **Hist1h2bf** | Histone H2B type 1-F/J/L | P10853 | 546.89 | 428.25 | 118.64 |
| **Zranb3** | Zinc finger Ran-binding domain-containing protein 3 | Q6NZP1 | 114.60 | 0.00 | 114.60 |
| **Dnmt1** | DNA (cytosine-5)-methyltransferase 1 | P13864 | 103.98 | 0.00 | 103.98 |
| **Ywhaz** | 14-3-3 protein zeta/delta | P63101 | 93.75 | 0.00 | 93.75 |

| name | description | Uniprot ID | PCNA unsynch NSAFe5 | BG NSAFe5 | PCNA-BG |
|---|---|---|---|---|---|
| **Ruvbl2** | RuvB-like 2 | Q9WTM5 | 148.83 | 58.27 | 90.56 |
| **Copr5** | Cooperator of PRMT5 | Q9CQ13 | 88.51 | 0.00 | 88.51 |
| **Fam111a** | Protein FAM111A | Q9D2L9 | 87.43 | 0.00 | 87.43 |
| **Cbx1** | Chromobox protein homolog 1 | P83917 | 82.77 | 0.00 | 82.77 |
| **Pold2** | DNA polymerase delta subunit 2 | O35654 | 81.63 | 0.00 | 81.63 |
| **Ybx1** | Nuclease-sensitive element-binding protein 1 | P62960 | 71.33 | 0.00 | 71.33 |
| **Uchl5** | Ubiquitin carboxyl-terminal hydrolase isozyme L5 | Q9WUP7 | 69.82 | 0.00 | 69.82 |
| **Cdsn** | Corneodesmosin | Q7TPC1 | 68.24 | 0.00 | 68.24 |
| **Cdk2** | Cell division protein kinase 2 | P97377 | 66.39 | 0.00 | 66.39 |
| **Mta3** | Metastasis-associated protein MTA3 | Q924K8 | 64.78 | 0.00 | 64.78 |
| **Dnajc8** | DnaJ homolog subfamily C member 8 | Q6NZB0 | 60.53 | 0.00 | 60.53 |
| **Smarcb1** | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1 | Q9Z0H3 | 59.66 | 0.00 | 59.66 |
| **Dpf2** | Zinc finger protein ubi-d4 | Q61103 | 58.74 | 0.00 | 58.74 |
| **Rbpj** | Recombining binding protein suppressor of hairless | P31266 | 58.22 | 0.00 | 58.22 |
| **Atad5** | ATPase family AAA domain-containing protein 5 | Q4QY64 | 54.51 | 0.00 | 54.51 |
| **Htatsf1** | HIV Tat-specific factor 1 homolog | Q8BGC0 | 50.57 | 0.00 | 50.57 |
| **Cdk4** | Cell division protein kinase 4 | P30285 | 50.54 | 0.00 | 50.54 |
| **Blmh** | Bleomycin hydrolase | Q8R016 | 50.48 | 0.00 | 50.48 |
| **Pold1** | DNA polymerase delta catalytic subunit | P52431 | 48.50 | 0.00 | 48.50 |
| **Parg** | Poly(ADP-ribose) glycohydrolase | O88622 | 47.41 | 0.00 | 47.41 |
| **Ddb1** | DNA damage-binding protein 1 | Q3U1J4 | 47.01 | 0.00 | 47.01 |
| **Lmna** | Lamin-A/C | P48678 | 46.05 | 0.00 | 46.05 |
| **Wdr5** | WD repeat-containing protein 5 | P61965 | 45.85 | 0.00 | 45.85 |
| **Ssrp1** | FACT complex subunit SSRP1 | Q08943 | 43.26 | 0.00 | 43.26 |
| **Aldoa** | Fructose-bisphosphate aldolase A | P05064 | 42.07 | 0.00 | 42.07 |
| **Chaf1b** | Chromatin assembly factor 1 subunit B | Q9D0N7 | 40.16 | 0.00 | 40.16 |
| **Ell** | RNA polymerase II elongation factor ELL | O08856 | 38.15 | 0.00 | 38.15 |

| name | description | Uniprot ID | PCNA unsynch NSAFe5 | BG NSAFe5 | PCNA-BG |
|---|---|---|---|---|---|
| **Wdr1** | WD repeat-containing protein 1 | O88342 | 37.90 | 0.00 | 37.90 |
| **Smarce1** | SWI/SNF-related matrix-associated actin-dependent regulator chromatin subfamily E member 1 | O54941 | 37.26 | 0.00 | 37.26 |
| **Ssb** | Lupus La protein homolog | P32067 | 36.90 | 0.00 | 36.90 |
| **Gatad2a** | Transcriptional repressor p66 alpha | Q8CHY6 | 36.52 | 0.00 | 36.52 |
| **Ddx39** | ATP-dependent RNA helicase DDX39 | Q8VDW0 | 35.86 | 0.00 | 35.86 |
| **Ctbp2** | C-terminal-binding protein 2 | P56546 | 34.41 | 0.00 | 34.41 |
| **Osbpl1a** | Oxysterol-binding protein-related protein 1 | Q91XL9 | 32.24 | 0.00 | 32.24 |
| **Cct6a** | T-complex protein 1 subunit zeta | P80317 | 28.84 | 0.00 | 28.84 |
| **Tgm1** | Protein-glutamine gamma-glutamyltransferase K | Q9JLF6 | 28.18 | 0.00 | 28.18 |
| **Cct3** | T-complex protein 1 subunit gamma | P80318 | 28.10 | 0.00 | 28.10 |
| **Kdm1a** | Lysine-specific histone demethylase 1A | Q6ZQ88 | 26.93 | 0.00 | 26.93 |
| **Pof1b** | Protein POF1B | Q8K4L4 | 25.82 | 0.00 | 25.82 |
| **Gatad2b** | Transcriptional repressor p66-beta | Q8VHR5 | 25.78 | 0.00 | 25.78 |
| **Actn1** | Alpha-actinin-1 | Q7TPR4 | 25.75 | 0.00 | 25.75 |
| **Pls3** | Plastin-3 | Q99K51 | 24.31 | 0.00 | 24.31 |
| **Ruvbl1** | RuvB-like 1 | P60122 | 201.49 | 177.50 | 23.99 |
| **D1Pas1** | Putative ATP-dependent RNA helicase Pl10 | P16381 | 23.20 | 0.00 | 23.20 |
| **Cltc** | Clathrin heavy chain 1 | Q68FD5 | 22.86 | 0.00 | 22.86 |
| **Tgm3** | Protein-glutamine gamma-glutamyltransferase E | Q08189 | 22.10 | 0.00 | 22.10 |
| **Supt16h** | FACT complex subunit SPT16 | Q920B9 | 21.94 | 0.00 | 21.94 |
| **Tpx2** | Targeting protein for Xklp2 | A2APB8 | 20.55 | 0.00 | 20.55 |
| **Khsrp** | Far upstream element-binding protein 2 | Q3U0V1 | 20.47 | 0.00 | 20.47 |
| **Plaa** | Phospholipase A-2-activating protein | P27612 | 19.29 | 0.00 | 19.29 |
| **Sall1** | Sal-like protein 1 | Q9ER74 | 17.37 | 0.00 | 17.37 |
| **Dhx9** | ATP-dependent RNA helicase A | O70133 | 16.64 | 0.00 | 16.64 |
| **Ints4** | Integrator complex subunit 4 | Q8CIM8 | 15.88 | 0.00 | 15.88 |
| **Rif1** | Telomere-associated protein RIF1 | Q6PR54 | 15.83 | 0.00 | 15.83 |

| name | description | Uniprot ID | PCNA unsynch NSAFe5 | BG NSAFe5 | PCNA-BG |
|---|---|---|---|---|---|
| **Ap2a1** | AP-2 complex subunit alpha-1 | P17426 | 15.67 | 0.00 | 15.67 |
| **Narg2** | NMDA receptor-regulated protein 2 | Q3UZ18 | 15.50 | 0.00 | 15.50 |
| **Ogt** | UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit | Q8CGY8 | 14.64 | 0.00 | 14.64 |
| **Smarca5** | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5 | Q91ZW3 | 14.57 | 0.00 | 14.57 |
| **Wdhd1** | WD repeat and HMG-box DNA-binding protein 1 | P59328 | 13.71 | 0.00 | 13.71 |
| **Mcm9** | DNA replication licensing factor MCM9 | Q2KHI9 | 13.50 | 0.00 | 13.50 |
| **Ehmt2** | Histone-lysine N-methyltransferase, H3 lysine-9 specific 3 | Q9Z148 | 12.12 | 0.00 | 12.12 |
| **Pds5a** | Sister chromatid cohesion protein PDS5 homolog A | Q6A026 | 11.50 | 0.00 | 11.50 |
| **Arid1a** | AT-rich interactive domain-containing protein 1A | A2BH40 | 10.06 | 0.00 | 10.06 |
| **Pbrm1** | Protein polybromo-1 | Q8BSQ9 | 9.37 | 0.00 | 9.37 |
| **Naca** | Nascent polypeptide-associated complex subunit alpha, muscle-specific form | P70670 | 7.00 | 0.00 | 7.00 |
| **Plec1** | Plectin-1 | Q9QXS1 | 4.90 | 0.00 | 4.90 |
| **Dst** | Bullous pemphigoid antigen 1 | Q91ZU6 | 2.07 | 0.00 | 2.07 |

## Supplementary Table 3.2: Known PCNA interaction partners extracted from STRING.

| Name | ENSEMBLE ID | description |
|---|---|---|
| **Rbl1** | ENSMUSP00000029170 | retinoblastoma-like 1 (p107) Gene; Key regulator of entry into cell division. Directly involved in heterochromatin formation by maintaining overall chromatin structure and, in particular, that of constitutive heterochromatin by stabilizing histone methylation. Recruits and targets histone methyltransferases SUV420H1 and SUV420H2, leading to epigenetic transcriptional repression. Controls histone H4 'Lys-20' trimethylation. Probably acts as a transcription repressor by recruiting chromatin-modifying enzymes to promoters. Potent inhibitor of E2F-mediated trans-activation. Forms a complex [...] |
| **Apex2** | ENSMUSP00000108341 | apurinic/apyrimidinic endonuclease 2 Gene; May participate in both nuclear and mitochondrial post- replicative base excision repair (BER). In the nucleus functions in the PCNA-dependent BER pathway |
| **Atm** | ENSMUSP00000113388 | ataxia telangiectasia mutated homolog (human) Gene; Serine/threonine protein kinase which activates checkpoint signaling upon double strand breaks (DSBs), apoptosis and genotoxic stresses such as ionizing ultraviolet A light (UVA), thereby acting as a DNA damage sensor. Recognizes the substrate consensus sequence [ST]-Q. Phosphorylates 'Ser-139' of histone variant H2AX/H2AFX at double strand breaks (DSBs), thereby regulating DNA damage response mechanism. Also involved in signal transduction and cell cycle control. May function as a tumor suppressor. Necessary for activation of ABL1 an [...] |
| **Ccna1** | ENSMUSP00000029368 | cyclin A1 Gene; May be involved in the control of the cell cycle at the G1/S (start) and G2/M (mitosis) transitions. May primarily function in the control of the germline meiotic cell cycle and additionally in the control of mitotic cell cycle in some somatic cells |
| **Ccnd1** | ENSMUSP00000091495 | cyclin D1 Gene; Essential for the control of the cell cycle at the G1/S (start) transition |
| **Ccnd2** | ENSMUSP00000000188 | cyclin D2 Gene; Essential for the control of the cell cycle at the G1/S (start) transition |
| **Ccne1** | ENSMUSP00000054023 | cyclin E1 Gene; Essential for the control of the cell cycle at the G1/S (start) transition |
| **Cdc25c** | ENSMUSP00000055427 | cell division cycle 25 homolog C (S. pombe) Gene; Functions as a dosage-dependent inducer in mitotic control. It is a tyrosine protein phosphatase required for progression of the cell cycle. It directly dephosphorylates CDC2 and activate its kinase activity. May be involved in regulating the proliferation of T-lymphocytes following cytokine stimulation |
| **Cdc45l** | ENSMUSP00000111249 | cell division cycle 45 homolog (S. cerevisiae)-like Gene; Required for initiation of chromosomal DNA replication |
| **Cdc6** | ENSMUSP00000091469 | cell division cycle 6 homolog (S. cerevisiae) Gene; Involved in the initiation of DNA replication. Also participates in checkpoint controls that ensure DNA replication is completed before mitosis is initiated |
| **Cdk1** | ENSMUSP00000020099 | cyclin-dependent kinase 1 Gene; Plays a key role in the control of the eukaryotic cell cycle. It is required in higher cells for entry into S-phase and mitosis. p34 is a component of the kinase complex that phosphorylates the repetitive C-terminus of RNA polymerase II |
| **Cdk2** | ENSMUSP00000026416 | cyclin-dependent kinase 2 Gene; Involved in the control of the cell cycle. Interacts with cyclins A, B1, B3, D, or E. Activity of CDK2 is maximal during S phase and G2 (By similarity) |
| **Cdk4** | ENSMUSP00000006911 | cyclin-dependent kinase 4 Gene; Probably involved in the control of the cell cycle |
| **Cdk6** | ENSMUSP00000037925 | cyclin-dependent kinase 6 Gene; Probably involved in the control of the cell cycle. Interacts with D-type G1 cyclins (By similarity) |
| **Cdkn1a** | ENSMUSP00000023829 | cyclin-dependent kinase inhibitor 1A (P21) Gene; May be the important intermediate by which p53 mediates its role as an inhibitor of cellular proliferation in response to DNA damage. Binds to and inhibits cyclin-dependent kinase activity, preventing phosphorylation of critical cyclin-dependent kinase substrates and blocking cell cycle progression |
| **Cdkn1c** | ENSMUSP00000037302 | cyclin-dependent kinase inhibitor 1C (P57) Gene; Potent tight-binding inhibitor of several G1 cyclin/CDK complexes (cyclin E-CDK2, cyclin D2-CDK4, and cyclin A-CDK2) and, to lesser extent, of the mitotic cyclin B-CDC2. Negative regulator of cell proliferation. May play a role in maintenance of the non- proliferative state throughout life |

| Name | ENSEMBLE ID | description |
|---|---|---|
| **Cdt1** | ENSMUSP00000006760 | chromatin licensing and DNA replication factor 1 Gene; Cooperates with CDC6 to promote the loading of the mini- chromosome maintenance complex onto chromatin to form the pre- replication complex necessary to initiate DNA replication. Binds DNA in a sequence-, strand-, and conformation-independent manner. Potential oncogene |
| **Dhfr** | ENSMUSP00000022218 | dihydrofolate reductase Gene |
| **Dna2** | ENSMUSP00000115750 | DNA replication helicase 2 homolog (yeast) Gene; May function in chromosomal DNA replication (By similarity) |
| **Dnmt1** | ENSMUSP00000004202 | DNA methyltransferase (cytosine-5) 1 Gene; Methylates CpG residues. Preferentially methylates hemimethylated DNA. Associates with DNA replication sites in S phase maintaining the methylation pattern in the newly synthesized strand, that is essential for epigenetic inheritance. Associates with chromatin during G2 and M phases to maintain DNA methylation independently of replication. It is responsible for maintaining methylation patterns established in development. DNA methylation is coordinated with methylation of histones. Mediates transcriptional repression by direct binding to HDAC2. [...] |
| **Dntt** | ENSMUSP00000062078 | deoxynucleotidyltransferase, terminal Gene; Template-independent DNA polymerase which catalyzes the random addition of deoxynucleoside 5'-triphosphate to the 3'-end of a DNA initiator. One of the in vivo functions of this enzyme is the addition of nucleotides at the junction (N region) of rearranged Ig heavy chain and T-cell receptor gene segments during the maturation of B- and T-cells |
| **E2f1** | ENSMUSP00000099434 | E2F transcription factor 1 Gene; Transcription activator that binds DNA cooperatively with dp proteins through the E2 recognition site, 5'-TTTC[CG]CGC- 3' found in the promoter region of a number of genes whose products are involved in cell cycle regulation or in DNA replication. The DRTF1/E2F complex functions in the control of cell-cycle progression from G1 to S phase. E2F-1 binds preferentially RB1 protein, in a cell-cycle dependent manner. It can mediate both cell proliferation and p53-dependent apoptosis |
| **E2f4** | ENSMUSP00000015003 | E2F transcription factor 4 Gene; Transcription activator that binds DNA cooperatively with DP proteins through the E2 recognition site, 5'-TTTC[CG]CGC- 3' found in the promoter region of a number of genes whose products are involved in cell cycle regulation or in DNA replication. The DRTF1/E2F complex functions in the control of cell-cycle progression from G1 to S phase. E2F-4 binds with high affinity to RBL1 and RBL2. In some instances, can also bind RB protein (By similarity) |
| **E2f5** | ENSMUSP00000029069 | E2F transcription factor 5 Gene; Transcriptional activator that binds to E2F sites, these sites are present in the promoter of many genes whose products are involved in cell proliferation. May mediate growth factor- initiated signal transduction |
| **Egf** | ENSMUSP00000029653 | epidermal growth factor Gene; EGF stimulates the growth of various epidermal and epithelial tissues in vivo and in vitro and of some fibroblasts in cell culture. Magnesiotropic hormone that stimulates magnesium reabsorption in the renal distal convoluted tubule via engagement of EGFR and activation of the magnesium channel TRPM6 (By similarity) |
| **Ercc5** | ENSMUSP00000027214 | excision repair cross-complementing rodent repair deficiency, complementation group 5 Gene; Single-stranded structure-specific DNA endonuclease involved in DNA excision repair. Makes the 3'incision in DNA nucleotide excision repair (NER). Acts as a cofactor for a DNA glycosylase that removes oxidized pyrimidines from DNA. May also be involved in transcription-coupled repair of this kind of damage, in transcription by RNA polymerase II, and perhaps in other processes too (By similarity) |
| **Fancd2** | ENSMUSP00000045667 | Fanconi anemia, complementation group D2 Gene; Required for maintenance of chromosomal stability. Promotes accurate and efficient pairing of homologs during meiosis. Involved in the repair of DNA double-strand breaks, both by homologous recombination and single-strand annealing. May participate in S phase and G2 phase checkpoint activation upon DNA damage. Promotes BRCA2/FANCD1 loading onto damaged chromatin. May also be involved in B-cell immunoglobulin isotype switching |
| **Fbxo5** | ENSMUSP00000019907 | F-box protein 5 Gene; Regulates progression through early mitosis by inhibiting the anaphase promoting complex/cyclosome (APC). Binds to the APC activators CDC20 and |

| Name | ENSEMBLE ID | description |
|---|---|---|
| | | FZR1/CDH1 to prevent APC activation. Can also bind directly to the APC to inhibit substrate-binding |
| **Fen1** | ENSMUSP00000025651 | flap structure specific endonuclease 1 Gene; Endonuclease that cleaves the 5'-overhanging flap structure that is generated by displacement synthesis when DNA polymerase encounters the 5'-end of a downstream Okazaki fragment. It fails to cleave other DNA structures, including 3'-flaps and single stranded DNA |
| **Gadd45a** | ENSMUSP00000044034 | growth arrest and DNA-damage-inducible 45 alpha Gene; Binds to proliferating cell nuclear antigen. Might affect PCNA interaction with some CDK (cell division protein kinase) complexes; stimulates DNA excision repair in vitro and inhibits entry of cells into S phase |
| **Gadd45b** | ENSMUSP00000015456 | growth arrest and DNA-damage-inducible 45 beta Gene; Involved in the regulation of growth and apoptosis. Mediates activation of stress-responsive MTK1/MEKK4 MAPKKK (By similarity) |
| **Gadd45g** | ENSMUSP00000021903 | growth arrest and DNA-damage-inducible 45 gamma Gene; Involved in the regulation of growth and apoptosis. Mediates activation of stress-responsive MTK1/MEKK4 MAPKKK |
| **Hras1** | ENSMUSP00000091515 | Harvey rat sarcoma virus oncogene 1 Gene; Ras proteins bind GDP/GTP and possess intrinsic GTPase activity |
| **Igf1** | ENSMUSP00000056668 | insulin-like growth factor 1 Gene; The insulin-like growth factors, isolated from plasma, are structurally and functionally related to insulin but have a much higher growth-promoting activity |
| **Kras** | ENSMUSP00000032397 | v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog Gene; Ras proteins bind GDP/GTP and possess intrinsic GTPase activity |
| **Lig1** | ENSMUSP00000096411 | ligase I, DNA, ATP-dependent Gene; DNA ligase that seals nicks in double-stranded DNA during DNA replication, DNA recombination and DNA repair |
| **Lin9** | ENSMUSP00000082959 | lin-9 homolog (C. elegans) Gene; Acts as a tumor suppressor. Inhibits DNA synthesis. Its ability to inhibit oncogenic transformation is mediated through its association with RB1. Plays a role in the expression of genes required for the G1/S transition (By similarity) |
| **Msh3** | ENSMUSP00000022220 | mutS homolog 3 (E. coli) Gene; Component of the post-replicative DNA mismatch repair system (MMR). Heterodimerizes with MSH2 to form MutS beta which binds to DNA mismatches thereby initiating DNA repair. When bound, the MutS beta heterodimer bends the DNA helix and shields approximately 20 base pairs. MutS beta recognizes large insertion- deletion loops (IDL) up to 13 nucleotides long. After mismatch binding, forms a ternary complex with the MutL alpha heterodimer, which is thought to be responsible for directing the downstream MMR events, including strand discrimination, excision, and [...] |
| **Msh6** | ENSMUSP00000005503 | mutS homolog 6 (E. coli) Gene; Component of the post-replicative DNA mismatch repair system (MMR). Heterodimerizes with MSH2 to form MutS alpha, which binds to DNA mismatches thereby initiating DNA repair. When bound, MutS alpha bends the DNA helix and shields approximately 20 base pairs, and recognizes single base mismatches and dinucleotide insertion-deletion loops (IDL) in the DNA. After mismatch binding, forms a ternary complex with the MutL alpha heterodimer, which is thought to be responsible for directing the downstream MMR events, including strand discrimination, excision, and [...] |
| **Orc1l** | ENSMUSP00000099805 | origin recognition complex, subunit 1-like (S.cereviaiae) Gene; Component of the origin recognition complex (ORC) that binds origins of replication. Binds to the ARS consensus sequence (ACS) of origins of replication in an ATP-dependent manner (By similarity) |
| **Pcna** | ENSMUSP00000028817 | proliferating cell nuclear antigen Gene; This protein is an auxiliary protein of DNA polymerase delta and is involved in the control of eukaryotic DNA replication by increasing the polymerase's processibility during elongation of the leading strand (By similarity) |
| **Pola1** | ENSMUSP00000006856 | polymerase (DNA directed), alpha 1 Gene; Plays an essential role in the initiation of DNA replication. During the S phase of the cell cycle, the DNA polymerase alpha complex (composed of a catalytic subunit POLA1/p180, a regulatory subunit POLA2/p70 and two primase subunits PRIM1/p49 and PRIM2/p58) is recruited to DNA at the replicative forks via direct interactions with MCM10 and WDHD1. The primase subunit of the polymerase alpha complex initiates DNA synthesis by oligomerising short RNA primers on both |

| Name | ENSEMBLE ID | description |
|---|---|---|
| | | leading and lagging strands. These primers are initially extended by the polymeras [...] |
| **Pola2** | ENSMUSP00000025752 | polymerase (DNA directed), alpha 2 Gene; May play an essential role at the early stage of chromosomal DNA replication by coupling the polymerase alpha/primase complex to the cellular replication machinery (By similarity) |
| **Polb** | ENSMUSP00000033938 | polymerase (DNA directed), beta Gene; Repair polymerase. Conducts "gap-filling" DNA synthesis in a stepwise distributive fashion rather than in a processive fashion as for other DNA polymerases. Has a 5'-deoxyribose-5- phosphate lyase (dRP lyase) activity (By similarity) |
| **Pold1** | ENSMUSP00000039776 | polymerase (DNA directed), delta 1, catalytic subunit Gene; Possesses two enzymatic activities: DNA synthesis (polymerase) and an exonucleolytic activity that degrades single stranded DNA in the 3'- to 5'-direction. Required with its accessory proteins (proliferating cell nuclear antigen (PCNA) and replication factor C (RFC) or activator 1) for leading strand synthesis. Also involved in completing Okazaki fragments initiated by the DNA polymerase alpha/primase complex |
| **Pold2** | ENSMUSP00000099986 | polymerase (DNA directed), delta 2, regulatory subunit Gene; The function of the small subunit is not yet clear |
| **Pold3** | ENSMUSP00000102694 | polymerase (DNA-directed), delta 3, accessory subunit Gene; Required for optimal DNA polymerase delta activity (By similarity) |
| **Pold4** | ENSMUSP00000025773 | polymerase (DNA-directed), delta 4 Gene; Required for optimal DNA polymerase delta activity. May contribute to PCNA-dependent activity of DNA polymerase delta (By similarity) |
| **Pole** | ENSMUSP00000007296 | polymerase (DNA directed), epsilon Gene; Participates in DNA repair and in chromosomal DNA replication |
| **Pole2** | ENSMUSP00000021359 | polymerase (DNA directed), epsilon 2 (p59 subunit) Gene; Participates in DNA repair and in chromosomal DNA replication |
| **Polh** | ENSMUSP00000024749 | polymerase (DNA directed), eta (RAD 30 related) Gene; DNA polymerase specifically involved in DNA repair. Plays an important role in translesion synthesis, where the normal high fidelity DNA polymerases cannot proceed and DNA synthesis stalls. Plays an important role in the repair of UV-induced pyrimidine dimers. Depending on the context, it inserts the correct base, but causes frequent base transitions and transversions. May play a role in hypermutation at immunoglobulin genes. Forms a Schiff base with 5'-deoxyribose phosphate at abasic sites, but does not have lyase activity. Targets [...] |
| **Poll** | ENSMUSP00000026239 | polymerase (DNA directed), lambda Gene; Repair polymerase. Involved in base excision repair (BER) responsible for repair of lesions that give rise to abasic (AP) sites in DNA. Has both DNA polymerase and terminal transferase activities. Has a 5'-deoxyribose-5-phosphate lyase (dRP lyase) activity (By similarity) |
| **Prim1** | ENSMUSP00000026461 | DNA primase, p49 subunit Gene; DNA primase is the polymerase that synthesizes small RNA primers for the Okazaki fragments made during discontinuous DNA replication |
| **Rbbp4** | ENSMUSP00000099658 | retinoblastoma binding protein 4 Gene; Core histone-binding subunit that may target chromatin assembly factors, chromatin remodeling factors and histone deacetylases to their histone substrates in a manner that is regulated by nucleosomal DNA. Component of several complexes which regulate chromatin metabolism. These include the chromatin assembly factor 1 (CAF-1) complex, which is required for chromatin assembly following DNA replication and DNA repair; the core histone deacetylase (HDAC) complex, which promotes histone deacetylation and consequent transcriptional repression; the nucle [...] |
| **Rbl2** | ENSMUSP00000034091 | retinoblastoma-like 2 Gene; Key regulator of entry into cell division. Directly involved in heterochromatin formation by maintaining overall chromatin structure and, in particular, that of constitutive heterochromatin by stabilizing histone methylation. Recruits and targets histone methyltransferases SUV420H1 and SUV420H2, leading to epigenetic transcriptional repression. Controls histone H4 'Lys-20' trimethylation. Probably acts as a transcription repressor by recruiting chromatin-modifying enzymes to promoters. Potent inhibitor of E2F-mediated trans-activation, associates preferentia [...] |
| **Rev1** | ENSMUSP00000027251 | REV1 homolog (S. cerevisiae) Gene; Deoxycytidyl transferase involved in DNA repair. Transfers a dCMP residue from dCTP to the 3'-end of a DNA primer in a template-dependent reaction. May assist in the first step in the bypass of abasic lesions by the |

| Name | ENSEMBLE ID | description |
|---|---|---|
| | | insertion of a nucleotide opposite the lesion. Required for normal induction of mutations by physical and chemical agents |
| **Rev3l** | ENSMUSP00000019986 | REV3-like, catalytic subunit of DNA polymerase zeta RAD54 like (S. cerevisiae) Gene |
| **Rfc1** | ENSMUSP00000031092 | replication factor C (activator 1) 1 Gene; The elongation of primed DNA templates by DNA polymerase delta and epsilon requires the action of the accessory proteins PCNA and activator 1. This subunit binds to the primer-template junction |
| **Rfc2** | ENSMUSP00000023867 | replication factor C (activator 1) 2 Gene; The elongation of primed DNA templates by DNA polymerase delta and epsilon requires the action of the accessory proteins proliferating cell nuclear antigen (PCNA) and activator 1. This subunit binds ATP (By similarity) |
| **Rfc3** | ENSMUSP00000039621 | replication factor C (activator 1) 3 Gene; The elongation of primed DNA templates by DNA polymerase delta and epsilon requires the action of the accessory proteins proliferating cell nuclear antigen (PCNA) and activator 1 |
| **Rfc4** | ENSMUSP00000110995 | replication factor C (activator 1) 4 Gene; The elongation of primed DNA templates by DNA polymerase delta and epsilon requires the action of the accessory proteins proliferating cell nuclear antigen (PCNA) and activator 1. This subunit may be involved in the elongation of the multiprimed DNA template (By similarity) |
| **Rfc5** | ENSMUSP00000083652 | replication factor C (activator 1) 5 Gene; The elongation of primed DNA templates by DNA polymerase delta and epsilon requires the action of the accessory proteins proliferating cell nuclear antigen (PCNA) and activator 1 |
| **Rpa1** | ENSMUSP00000000767 | replication protein A1 Gene; Plays an essential role in several cellular processes in DNA metabolism including replication, recombination and DNA repair. Binds and subsequently stabilizes single-stranded DNA intermediates and thus prevents complementary DNA from reannealing (By similarity) |
| **Rpa2** | ENSMUSP00000099621 | replication protein A2 Gene; Required for DNA recombination, repair and replication. The activity of RP-A is mediated by single-stranded DNA binding and protein interactions |
| **Rpa3** | ENSMUSP00000012627 | replication protein A3 Gene; Required for DNA recombination, repair and replication. The activity of RP-A is mediated by single-stranded DNA binding and protein interactions (By similarity) |
| **Rrm2** | ENSMUSP00000020980 | ribonucleotide reductase M2 Gene; Provides the precursors necessary for DNA synthesis. Catalyzes the biosynthesis of deoxyribonucleotides from the corresponding ribonucleotides. Inhibits Wnt signaling (By similarity) |
| **Tfdp1** | ENSMUSP00000078078 | transcription factor Dp 1 Gene; Can stimulate E2F-dependent transcription. Binds DNA cooperatively with E2F family members through the E2 recognition site, 5'-TTTC[CG]CGC-3', found in the promoter region of a number of genes whose products are involved in cell cycle regulation or in DNA replication. The DP2/E2F complex functions in the control of cell-cycle progression from G1 to S phase. The E2F-1/DP complex appears to mediate both cell proliferation and apoptosis |
| **Tfdp2** | ENSMUSP00000034982 | transcription factor Dp 2 Gene; Can stimulate E2F-dependent transcription. Binds DNA cooperatively with E2F family members through the E2 recognition site, 5'-TTTC[CG]CGC-3', found in the promoter region of a number of genes whose products are involved in cell cycle regulation or in DNA replication. The DP2/E2F complex functions in the control of cell-cycle progression from G1 to S phase. The E2F-1/DP complex appears to mediate both cell proliferation and apoptosis |
| **Top2a** | ENSMUSP00000068896 | topoisomerase (DNA) II alpha Gene; Control of topological states of DNA by transient breakage and subsequent rejoining of DNA strands. Topoisomerase II makes double-strand breaks (By similarity) |
| **Trp53** | ENSMUSP00000104298 | transformation related protein 53 Gene; Acts as a tumor suppressor in many tumor types; induces growth arrest or apoptosis depending on the physiological circumstances and cell type. Involved in cell cycle regulation as a trans-activator that acts to negatively regulate cell division by controlling a set of genes required for this process. One of the activated genes is an inhibitor of cyclin-dependent kinases. Apoptosis induction seems to be mediated either by stimulation of BAX and FAS antigen expression, or by repression of Bcl-2 expression (By similarity) |
| **Tyms** | ENSMUSP00000026846 | thymidylate synthase Gene |

| Name | ENSEMBLE ID | description |
|---|---|---|
| **Ung** | ENSMUSP00000031587 | uracil DNA glycosylase Gene; Excises uracil residues from the DNA which can arise as a result of misincorporation of dUMP residues by DNA polymerase or due to deamination of cytosine (By similarity) |
| **Wrn** | ENSMUSP00000033990 | Werner syndrome homolog (human) Gene; Essential for the formation of DNA replication focal centers; stably associates with foci elements generating binding sites for RP-A. Exhibits a magnesium-dependent ATP-dependent DNA- helicase activity. May be involved in the control of genomic stability |

**Supplementary Table 3.3: PCNA interactors at 5 hours post release (early S-phase).**

| name | description | Uniprot ID | PCNA 5h NSAFe5 | BG NSAFe5 | 5h PCNA-BG |
|---|---|---|---|---|---|
| **Pcna** | Proliferating cell nuclear antigen | P17918 | 8981.96 | 332.82 | 8649.13 |
| **Paf** | PCNA-associated factor | Q9CQX4 | 2228.68 | 0.00 | 2228.68 |
| **Stk38** | Serine/threonine-protein kinase 38 | Q91VJ4 | 1252.13 | 0.00 | 1252.13 |
| **Ddx5** | Probable ATP-dependent RNA helicase DDX5 | Q61656 | 923.32 | 0.00 | 923.32 |
| **Tmod3** | Tropomodulin-3 | Q9JHJ0 | 914.11 | 0.00 | 914.11 |
| **Prmt5** | Protein arginine N-methyltransferase 5 | Q8CIG8 | 649.45 | 0.00 | 649.45 |
| **Wdr77** | Methylosome protein 50 | Q99J09 | 627.22 | 0.00 | 627.22 |
| **Ncl** | Nucleolin | P09405 | 736.85 | 122.87 | 613.98 |
| **Rfc3** | Replication factor C subunit 3 | Q8R323 | 602.56 | 0.00 | 602.56 |
| **Cdkn1a** | Cyclin-dependent kinase inhibitor 1 | P39689 | 578.19 | 0.00 | 578.19 |
| **Rfc2** | Replication factor C subunit 2 | Q9WUK4 | 570.74 | 0.00 | 570.74 |
| **Ddx17** | Probable ATP-dependent RNA helicase DDX17 | Q501J6 | 565.74 | 0.00 | 565.74 |
| **Rfc4** | Replication factor C subunit 4 | Q99J62 | 463.03 | 0.00 | 463.03 |
| **Clns1a** | Methylosome subunit pICln | Q61189 | 454.47 | 0.00 | 454.47 |
| **Rfc5** | Replication factor C subunit 5 | Q9D0F6 | 451.98 | 0.00 | 451.98 |
| **Erh** | Enhancer of rudimentary homolog | P84089 | 441.99 | 0.00 | 441.99 |
| **Stk38l** | Serine/threonine-protein kinase 38-like | Q7TSE6 | 429.28 | 0.00 | 429.28 |
| **Ccnd1** | G1/S-specific cyclin-D1 | P25322 | 415.52 | 0.00 | 415.52 |
| **Sptan1** | Spectrin alpha chain, brain | P16546 | 415.29 | 0.00 | 415.29 |
| **Lima1** | LIM domain and actin-binding protein 1 | Q9ERG0 | 406.96 | 0.00 | 406.96 |
| **Rps27a** | Ubiquitin | P62991 | 403.21 | 0.00 | 403.21 |
| **Sptbn1** | Spectrin beta chain, brain 1 | Q62261 | 389.05 | 0.00 | 389.05 |
| **Kctd5** | BTB/POZ domain-containing protein KCTD5 | Q8VC57 | 327.40 | 0.00 | 327.40 |
| **Matr3** | Matrin-3 | Q8K310 | 326.00 | 0.00 | 326.00 |

| name | description | Uniprot ID | PCNA 5h NSAFe5 | BG NSAFe5 | 5h PCNA-BG |
|---|---|---|---|---|---|
| **Npm1** | Nucleophosmin | Q61937 | 314.84 | 0.00 | 314.84 |
| **Calm1** | Calmodulin | P62204 | 308.50 | 0.00 | 308.50 |
| **Rbbp4** | Histone-binding protein RBBP4 | Q60972 | 298.26 | 0.00 | 298.26 |
| **Mycbp** | C-Myc-binding protein | Q9EQS3 | 297.52 | 0.00 | 297.52 |
| **Hspa5** | 78 kDa glucose-regulated protein | P20029 | 280.71 | 0.00 | 280.71 |
| **Trim28** | Transcription intermediary factor 1-beta | Q62318 | 275.58 | 0.00 | 275.58 |
| **Prss2** | Anionic trypsin-2 | P07146 | 249.14 | 0.00 | 249.14 |
| **Nhp2l1** | NHP2-like protein 1 | Q9D0T1 | 239.41 | 0.00 | 239.41 |
| **Prdx2** | Peroxiredoxin-2 | Q61171 | 232.15 | 0.00 | 232.15 |
| **Ruvbl2** | RuvB-like 2 | Q9WTM5 | 231.65 | 0.00 | 231.65 |
| **Ddx21** | Nucleolar RNA helicase 2 | Q9JIK5 | 216.06 | 0.00 | 216.06 |
| **Trim35** | Tripartite motif-containing protein 35 | Q8C006 | 214.08 | 0.00 | 214.08 |
| **Cdc2** | Cell division control protein 2 homolog | P11440 | 206.36 | 0.00 | 206.36 |
| **Add3** | Gamma-adducin | Q9QYB5 | 195.32 | 0.00 | 195.32 |
| **DSCC1** | Sister chromatid cohesion protein DCC1 | Q14AI0 | 192.01 | 0.00 | 192.01 |
| **Top2a** | DNA topoisomerase 2-alpha | Q01320 | 190.52 | 0.00 | 190.52 |
| **Ybx1** | Nuclease-sensitive element-binding protein 1 | P62960 | 190.34 | 0.00 | 190.34 |
| **Wdr5** | WD repeat-containing protein 5 | P61965 | 183.50 | 0.00 | 183.50 |
| **Msh2** | DNA mismatch repair protein Msh2 | P43247 | 180.26 | 0.00 | 180.26 |
| **Cdk2** | Cell division protein kinase 2 | P97377 | 177.13 | 0.00 | 177.13 |
| **Fen1** | Flap endonuclease 1 | P39749 | 162.14 | 0.00 | 162.14 |
| **Ccnd2** | G1/S-specific cyclin-D2 | P30280 | 159.05 | 0.00 | 159.05 |
| **Hdac1** | Histone deacetylase 1 | O09106 | 158.19 | 0.00 | 158.19 |
| **Hdac2** | Histone deacetylase 2 | P70288 | 158.19 | 0.00 | 158.19 |
| **Cdk5** | Cell division protein kinase 5 | P49615 | 157.42 | 0.00 | 157.42 |

| name | description | Uniprot ID | PCNA 5h NSAFe5 | BG NSAFe5 | 5h PCNA-BG |
|---|---|---|---|---|---|
| **Cdk4** | Cell division protein kinase 4 | P30285 | 151.70 | 0.00 | 151.70 |
| **Pabpc1** | Polyadenylate-binding protein 1 | P29341 | 144.55 | 0.00 | 144.55 |
| **Rest** | RE1-silencing transcription factor | Q8VIG1 | 141.61 | 0.00 | 141.61 |
| **Cttn** | Src substrate cortactin | Q60598 | 140.31 | 0.00 | 140.31 |
| **Gapdh** | Glyceraldehyde-3-phosphate dehydrogenase | P16858 | 138.04 | 0.00 | 138.04 |
| **Cobl** | Protein cordon-bleu | Q5NBX1 | 137.52 | 0.00 | 137.52 |
| **Ruvbl1** | RuvB-like 1 | P60122 | 134.40 | 0.00 | 134.40 |
| **Ccar1** | Cell division cycle and apoptosis regulator protein 1 | Q8CH18 | 133.70 | 0.00 | 133.70 |
| **Srrm2** | Serine/arginine repetitive matrix protein 2 | Q8BTI8 | 130.38 | 0.00 | 130.38 |
| **Coro1c** | Coronin-1C | Q9WUM4 | 129.30 | 0.00 | 129.30 |
| **Sall4** | Sal-like protein 4 | Q8BX22 | 129.24 | 0.00 | 129.24 |
| **Add1** | Alpha-adducin | Q9QYC0 | 125.08 | 0.00 | 125.08 |
| **Htatsf1** | HIV Tat-specific factor 1 homolog | Q8BGC0 | 121.44 | 0.00 | 121.44 |
| **Gsn** | Gelsolin | P13020 | 117.86 | 0.00 | 117.86 |
| **L1td1** | LINE-1 type transposase domain-containing protein 1 | Q587J6 | 117.56 | 0.00 | 117.56 |
| **Spin1** | Spindlin-1 | Q61142 | 116.96 | 0.00 | 116.96 |
| **Setx** | Probable helicase senataxin | A2AKX3 | 110.02 | 0.00 | 110.02 |
| **Dbn1** | Drebrin | Q9QXS6 | 108.51 | 0.00 | 108.51 |
| **Igf2bp1** | Insulin-like growth factor 2 mRNA-binding protein 1 | O88477 | 106.22 | 0.00 | 106.22 |
| **Ilf3** | Interleukin enhancer-binding factor 3 | Q9Z1X4 | 102.38 | 0.00 | 102.38 |
| **Elavl1** | ELAV-like protein 1 | P70372 | 94.00 | 0.00 | 94.00 |
| **Ddx3x** | ATP-dependent RNA helicase DDX3X | Q62167 | 92.72 | 0.00 | 92.72 |
| **D1Pas1** | Putative ATP-dependent RNA helicase Pl10 | P16381 | 92.72 | 0.00 | 92.72 |
| **Parp1** | Poly [ADP-ribose] polymerase 1 | P11103 | 90.75 | 0.00 | 90.75 |
| **Rcc2** | Protein RCC2 | Q8BK67 | 88.40 | 0.00 | 88.40 |

| name | description | Uniprot ID | PCNA 5h NSAFe5 | BG NSAFe5 | 5h PCNA-BG |
|---|---|---|---|---|---|
| **Atad5** | ATPase family AAA domain-containing protein 5 | Q4QY64 | 83.91 | 0.00 | 83.91 |
| **Top1** | DNA topoisomerase 1 | Q04750 | 79.91 | 0.00 | 79.91 |
| **Mybbp1a** | Myb-binding protein 1A | Q7TPV4 | 79.80 | 0.00 | 79.80 |
| **Smarcb1** | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1 | Q9Z0H3 | 79.60 | 0.00 | 79.60 |
| **Chtf18** | Chromosome transmission fidelity protein 18 homolog | Q8BIW9 | 79.06 | 0.00 | 79.06 |
| **Map1b** | Microtubule-associated protein 1B | P14873 | 74.62 | 0.00 | 74.62 |
| **Luc7l3** | Luc7-like protein 3 | Q5SUF2 | 70.94 | 0.00 | 70.94 |
| **Bclaf1** | Bcl-2-associated transcription factor 1 | Q8K019 | 66.69 | 0.00 | 66.69 |
| **Nono** | Non-POU domain-containing octamer-binding protein | Q99K48 | 64.79 | 0.00 | 64.79 |
| **Pole2** | DNA polymerase epsilon subunit 2 | O54956 | 58.15 | 0.00 | 58.15 |
| **Rbm39** | RNA-binding protein 39 | Q8VH51 | 57.82 | 0.00 | 57.82 |
| **Cct3** | T-complex protein 1 subunit gamma | P80318 | 56.23 | 0.00 | 56.23 |
| **Cct8** | T-complex protein 1 subunit theta | P42932 | 55.92 | 0.00 | 55.92 |
| **Smarcc1** | SWI/SNF complex subunit SMARCC1 | P97496 | 55.52 | 0.00 | 55.52 |
| **Atp5a1** | ATP synthase subunit alpha, mitochondrial | Q03265 | 55.41 | 0.00 | 55.41 |
| **Ltbp1** | Latent-transforming growth factor beta-binding protein 1 | Q8CG19 | 53.70 | 0.00 | 53.70 |
| **Rbm10** | RNA-binding protein 10 | Q99KG3 | 49.43 | 0.00 | 49.43 |
| **Ap2a2** | AP-2 complex subunit alpha-2 | P17427 | 49.00 | 0.00 | 49.00 |
| **Dnmt1** | DNA (cytosine-5)-methyltransferase 1 | P13864 | 47.29 | 0.00 | 47.29 |
| **Hspa9** | Stress-70 protein, mitochondrial | P38647 | 45.13 | 0.00 | 45.13 |
| **Ddx46** | Probable ATP-dependent RNA helicase DDX46 | Q569Z5 | 44.54 | 0.00 | 44.54 |
| **Supt16h** | FACT complex subunit SPT16 | Q920B9 | 43.90 | 0.00 | 43.90 |
| **Tcof1** | Treacle protein | O08784 | 34.82 | 0.00 | 34.82 |

| name | description | Uniprot ID | PCNA 5h NSAFe5 | BG NSAFe5 | 5h PCNA-BG |
|---|---|---|---|---|---|
| **Acin1** | Apoptotic chromatin condensation inducer in the nucleus | Q9JIX8 | 34.35 | 0.00 | 34.35 |
| **Pole** | DNA polymerase epsilon catalytic subunit A | Q9WVF7 | 33.56 | 0.00 | 33.56 |
| **Dhx9** | ATP-dependent RNA helicase A | O70133 | 33.31 | 0.00 | 33.31 |
| **Thrap3** | Thyroid hormone receptor-associated protein 3 | Q569Z6 | 32.22 | 0.00 | 32.22 |
| **Tet1** | Methylcytosine dioxygenase TET1 | Q3URK3 | 22.90 | 0.00 | 22.90 |
| **Msh6** | DNA mismatch repair protein Msh6 | P54276 | 22.57 | 0.00 | 22.57 |
| **Smarca4** | Transcription activator BRG1 | Q3TKT4 | 19.00 | 0.00 | 19.00 |

**Supplementary Table 3.4: PCNA interactors at 9 hours post release (late S-phase).**

| name | description | Uniprot ID | PCNA-9h NSAFe5 | BG NSAFe5 | 9h PCNA-BG |
|---|---|---|---|---|---|
| **Pcna** | Proliferating cell nuclear antigen | P17918 | 13187.06 | 332.82 | 12854.24 |
| **Rfc4** | Replication factor C subunit 4 | Q99J62 | 1296.09 | 0.00 | 1296.09 |
| **Rfc5** | Replication factor C subunit 5 | Q9D0F6 | 1138.64 | 0.00 | 1138.64 |
| **Rfc3** | Replication factor C subunit 3 | Q8R323 | 1054.15 | 0.00 | 1054.15 |
| **Rfc2** | Replication factor C subunit 2 | Q9WUK4 | 1044.57 | 0.00 | 1044.57 |
| **Tmod3** | Tropomodulin-3 | Q9JHJ0 | 913.82 | 0.00 | 913.82 |
| **Stk38** | Serine/threonine-protein kinase 38 | Q91VJ4 | 853.16 | 0.00 | 853.16 |
| **Cdkn1a** | Cyclin-dependent kinase inhibitor 1 | P39689 | 809.22 | 0.00 | 809.22 |
| **Prmt5** | Protein arginine N-methyltransferase 5 | Q8CIG8 | 690.13 | 0.00 | 690.13 |
| **Cdk4** | Cell division protein kinase 4 | P30285 | 636.96 | 0.00 | 636.96 |
| **Paf** | PCNA-associated factor | Q9CQX4 | 584.85 | 0.00 | 584.85 |
| **Ddx5** | Probable ATP-dependent RNA helicase DDX5 | Q61656 | 558.81 | 0.00 | 558.81 |
| **Wdr77** | Methylosome protein 50 | Q99J09 | 532.97 | 0.00 | 532.97 |
| **Rfc1** | Replication factor C subunit 1 | P35601 | 492.97 | 0.00 | 492.97 |
| **Pole4** | DNA polymerase epsilon subunit 4 | Q9CQ36 | 454.33 | 0.00 | 454.33 |
| **Rps27a** | Ubiquitin | P62991 | 423.24 | 0.00 | 423.24 |
| **Cdk5** | Cell division protein kinase 5 | P49615 | 367.20 | 0.00 | 367.20 |
| **Msh2** | DNA mismatch repair protein Msh2 | P43247 | 366.96 | 0.00 | 366.96 |
| **Top2a** | DNA topoisomerase 2-alpha | Q01320 | 350.86 | 0.00 | 350.86 |
| **Fen1** | Flap endonuclease 1 | P39749 | 340.39 | 0.00 | 340.39 |
| **Trim28** | Transcription intermediary factor 1-beta | Q62318 | 334.26 | 0.00 | 334.26 |
| **Kctd5** | BTB/POZ domain-containing protein KCTD5 | Q8VC57 | 320.75 | 0.00 | 320.75 |
| **Atad5** | ATPase family AAA domain-containing protein 5 | Q4QY64 | 299.47 | 0.00 | 299.47 |
| **Ddx17** | Probable ATP-dependent RNA helicase DDX17 | Q501J6 | 296.92 | 0.00 | 296.92 |

| name | description | Uniprot ID | PCNA-9h NSAFe5 | BG NSAFe5 | 9h PCNA-BG |
|---|---|---|---|---|---|
| **Npm1** | Nucleophosmin | Q61937 | 293.76 | 0.00 | 293.76 |
| **Phf5a** | PHD finger-like domain-containing protein 5A | P83870 | 292.42 | 0.00 | 292.42 |
| **Rest** | RE1-silencing transcription factor | Q8VIG1 | 277.47 | 0.00 | 277.47 |
| **DSCC1** | Sister chromatid cohesion protein DCC1 | Q14AI0 | 268.73 | 0.00 | 268.73 |
| **Rbbp4** | Histone-binding protein RBBP4 | Q60972 | 268.35 | 0.00 | 268.35 |
| **Ccnd2** | G1/S-specific cyclin-D2 | P30280 | 259.71 | 0.00 | 259.71 |
| **Trim35** | Tripartite motif-containing protein 35 | Q8C006 | 256.82 | 0.00 | 256.82 |
| **Ccnd1** | G1/S-specific cyclin-D1 | P25322 | 254.42 | 0.00 | 254.42 |
| **Matr3** | Matrin-3 | Q8K310 | 253.48 | 0.00 | 253.48 |
| **Cdk2** | Cell division protein kinase 2 | P97377 | 247.91 | 0.00 | 247.91 |
| **Clns1a** | Methylosome subunit pICln | Q61189 | 227.16 | 0.00 | 227.16 |
| **Sptan1** | Spectrin alpha chain, brain | P16546 | 225.55 | 0.00 | 225.55 |
| **Msh6** | DNA mismatch repair protein Msh6 | P54276 | 221.08 | 0.00 | 221.08 |
| **Calm1** | Calmodulin | P62204 | 215.88 | 0.00 | 215.88 |
| **Erh** | Enhancer of rudimentary homolog | P84089 | 206.20 | 0.00 | 206.20 |
| **Sptbn1** | Spectrin beta chain, brain 1 | Q62261 | 199.65 | 0.00 | 199.65 |
| **Htatsf1** | HIV Tat-specific factor 1 homolog | Q8BGC0 | 198.30 | 0.00 | 198.30 |
| **Pole2** | DNA polymerase epsilon subunit 2 | O54956 | 183.11 | 0.00 | 183.11 |
| **Ncl** | Nucleolin | P09405 | 303.32 | 122.87 | 180.45 |
| **Pole** | DNA polymerase epsilon catalytic subunit A | Q9WVF7 | 173.77 | 0.00 | 173.77 |
| **Chtf18** | Chromosome transmission fidelity protein 18 homolog | Q8BIW9 | 165.98 | 0.00 | 165.98 |
| **Hspa5** | 78 kDa glucose-regulated protein | P20029 | 163.70 | 0.00 | 163.70 |
| **Fam111a** | Protein FAM111A | Q9D2L9 | 157.42 | 0.00 | 157.42 |
| **Hdac1** | Histone deacetylase 1 | O09106 | 154.98 | 0.00 | 154.98 |
| **Hdac2** | Histone deacetylase 2 | P70288 | 154.98 | 0.00 | 154.98 |

| name | description | Uniprot ID | PCNA-9h NSAFe5 | BG NSAFe5 | 9h PCNA-BG |
|---|---|---|---|---|---|
| **Cdc2** | Cell division control protein 2 homolog | P11440 | 144.41 | 0.00 | 144.41 |
| **Smarcb1** | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1 | Q9Z0H3 | 139.25 | 0.00 | 139.25 |
| **Coro1c** | Coronin-1C | Q9WUM4 | 135.72 | 0.00 | 135.72 |
| **Cfl1** | Cofilin-1 | P18760 | 129.18 | 0.00 | 129.18 |
| **Kctd2** | BTB/POZ domain-containing protein KCTD2 | Q8CEZ0 | 120.93 | 0.00 | 120.93 |
| **Cbx3** | Chromobox protein homolog 3 | P23198 | 117.18 | 0.00 | 117.18 |
| **Sall4** | Sal-like protein 4 | Q8BX22 | 110.54 | 0.00 | 110.54 |
| **Gfap** | Glial fibrillary acidic protein | P03995 | 109.90 | 0.00 | 109.90 |
| **Dbn1** | Drebrin | Q9QXS6 | 106.31 | 0.00 | 106.31 |
| **Add3** | Gamma-adducin | Q9QYB5 | 106.31 | 0.00 | 106.31 |
| **Mta1** | Metastasis-associated protein MTA1 | Q8K4B0 | 104.97 | 0.00 | 104.97 |
| **Add1** | Alpha-adducin | Q9QYC0 | 102.12 | 0.00 | 102.12 |
| **Gapdh** | Glyceraldehyde-3-phosphate dehydrogenase | P16858 | 96.60 | 0.00 | 96.60 |
| **Ruvbl1** | RuvB-like 1 | P60122 | 94.05 | 0.00 | 94.05 |
| **Srrm2** | Serine/arginine repetitive matrix protein 2 | Q8BTI8 | 91.24 | 0.00 | 91.24 |
| **Ywhaq** | 14-3-3 protein theta | P68254 | 87.53 | 0.00 | 87.53 |
| **Ywhag** | 14-3-3 protein gamma | P61982 | 86.82 | 0.00 | 86.82 |
| **Dnajc8** | DnaJ homolog subfamily C member 8 | Q6NZB0 | 84.76 | 0.00 | 84.76 |
| **Parp1** | Poly [ADP-ribose] polymerase 1 | P11103 | 84.68 | 0.00 | 84.68 |
| **Ywhae** | 14-3-3 protein epsilon | P62259 | 84.10 | 0.00 | 84.10 |
| **Spin1** | Spindlin-1 | Q61142 | 81.85 | 0.00 | 81.85 |
| **Rbm39** | RNA-binding protein 39 | Q8VH51 | 80.92 | 0.00 | 80.92 |
| **Smarcc1** | SWI/SNF complex subunit SMARCC1 | P97496 | 77.70 | 0.00 | 77.70 |
| **Ap2m1** | AP-2 complex subunit mu | P84091 | 73.95 | 0.00 | 73.95 |
| **Smarcd1** | SWI/SNF-related matrix-associated actin-dependent | Q61466 | 62.46 | 0.00 | 62.46 |

| name | description | Uniprot ID | PCNA-9h NSAFe5 | BG NSAFe5 | 9h PCNA-BG |
|---|---|---|---|---|---|
| | regulator of chromatin subfamily D member 1 | | | | |
| **Zranb3** | Zinc finger Ran-binding domain-containing protein 3 | Q6NZP1 | 60.18 | 0.00 | 60.18 |
| **Cct3** | T-complex protein 1 subunit gamma | P80318 | 59.02 | 0.00 | 59.02 |
| **Lima1** | LIM domain and actin-binding protein 1 | Q9ERG0 | 56.96 | 0.00 | 56.96 |
| **Csnk2a1** | Casein kinase II subunit alpha | Q60737 | 54.84 | 0.00 | 54.84 |
| **Dnmt1** | DNA (cytosine-5)-methyltransferase 1 | P13864 | 52.95 | 0.00 | 52.95 |
| **Setx** | Probable helicase senataxin | A2AKX3 | 52.68 | 0.00 | 52.68 |
| **Smarce1** | SWI/SNF-related matrix-associated actin-dependent regulator chromatin subfamily E member 1 | O54941 | 52.18 | 0.00 | 52.18 |
| **Dnaja2** | DnaJ homolog subfamily A member 2 | Q9QYJ0 | 52.05 | 0.00 | 52.05 |
| **Luc7l3** | Luc7-like protein 3 | Q5SUF2 | 49.64 | 0.00 | 49.64 |
| **Ruvbl2** | RuvB-like 2 | Q9WTM5 | 46.32 | 0.00 | 46.32 |
| **Ssrp1** | FACT complex subunit SSRP1 | Q08943 | 45.43 | 0.00 | 45.43 |
| **L1td1** | LINE-1 type transposase domain-containing protein 1 | Q587J6 | 41.13 | 0.00 | 41.13 |
| **Ptbp1** | Polypyrimidine tract-binding protein 1 | P17225 | 40.69 | 0.00 | 40.69 |
| **Riok1** | Serine/threonine-protein kinase RIO1 | Q922Q2 | 37.82 | 0.00 | 37.82 |
| **Ddx21** | Nucleolar RNA helicase 2 | Q9JIK5 | 37.80 | 0.00 | 37.80 |
| **Ilf3** | Interleukin enhancer-binding factor 3 | Q9Z1X4 | 35.82 | 0.00 | 35.82 |
| **Cherp** | Calcium homeostasis endoplasmic reticulum protein | Q8CGZ0 | 34.37 | 0.00 | 34.37 |
| **Sin3a** | Paired amphipathic helix protein Sin3a | Q60520 | 33.45 | 0.00 | 33.45 |
| **Smarca4** | Transcription activator BRG1 | Q3TKT4 | 33.24 | 0.00 | 33.24 |
| **Pbrm1** | Protein polybromo-1 | Q8BSQ9 | 32.81 | 0.00 | 32.81 |
| **Acin1** | Apoptotic chromatin condensation inducer in the nucleus | Q9JIX8 | 32.05 | 0.00 | 32.05 |
| **Supt16h** | FACT complex subunit SPT16 | Q920B9 | 30.72 | 0.00 | 30.72 |
| **Map1b** | Microtubule-associated protein 1B | P14873 | 30.46 | 0.00 | 30.46 |

| name | description | Uniprot ID | PCNA-9h NSAFe5 | BG NSAFe5 | 9h PCNA-BG |
|---|---|---|---|---|---|
| **Msh3** | DNA mismatch repair protein Msh3 | P13705 | 29.48 | 0.00 | 29.48 |
| **Lrrfip1** | Leucine-rich repeat flightless-interacting protein 1 | Q3UZ39 | 29.42 | 0.00 | 29.42 |
| **Kif20b** | Kinesin-like protein KIF20B | Q80WE4 | 24.18 | 0.00 | 24.18 |
| **Rbm10** | RNA-binding protein 10 | Q99KG3 | 23.06 | 0.00 | 23.06 |
| **Ap2b1** | AP-2 complex subunit beta | Q9DBG3 | 22.89 | 0.00 | 22.89 |
| **Sap130** | Histone deacetylase complex subunit SAP130 | Q8BIH0 | 20.29 | 0.00 | 20.29 |
| **Wdhd1** | WD repeat and HMG-box DNA-binding protein 1 | P59328 | 19.20 | 0.00 | 19.20 |
| **Cobl** | Protein cordon-bleu | Q5NBX1 | 16.04 | 0.00 | 16.04 |
| **Mybbp1a** | Myb-binding protein 1A | Q7TPV4 | 15.96 | 0.00 | 15.96 |
| **Ltbp1** | Latent-transforming growth factor beta-binding protein 1 | Q8CG19 | 12.53 | 0.00 | 12.53 |

**Supplementary Table 3.5: Wdhd1 interaction partners identified by MudPIT.**

| name | description | Uniprot ID | Wdhd1 NSAFe5 | BG NSAFe5 | Wdhd1 - BG |
|---|---|---|---|---|---|
| **Wdhd1** | WD repeat and HMG-box DNA-binding protein 1 | P59328 | 2754.81 | 0.00 | 2754.81 |
| **Msh2** | DNA mismatch repair protein Msh2 | P43247 | 1694.54 | 77.64 | 1616.89 |
| **Pkm2** | Pyruvate kinase isozymes M1/M2 | P52480 | 1356.27 | 341.79 | 1014.48 |
| **Pcna** | Proliferating cell nuclear antigen | P17918 | 551.86 | 0.00 | 551.86 |
| **Ruvbl2** | RuvB-like 2 | Q9WTM5 | 395.94 | 0.00 | 395.94 |
| **Cfl1** | Cofilin-1 | P18760 | 394.40 | 0.00 | 394.40 |
| **Ddx5** | Probable ATP-dependent RNA helicase DDX5 | Q61656 | 1087.63 | 709.41 | 378.21 |
| **Mcm3** | DNA replication licensing factor MCM3 | P25206 | 354.77 | 0.00 | 354.77 |
| **Actbl2** | Beta-actin-like protein 2 | Q8BFZ3 | 348.25 | 0.00 | 348.25 |
| **Rfc4** | Replication factor C subunit 4 | Q99J62 | 323.76 | 0.00 | 323.76 |
| **Cct8** | T-complex protein 1 subunit theta | P42932 | 310.63 | 0.00 | 310.63 |
| **Msh6** | DNA mismatch repair protein Msh6 | P54276 | 308.55 | 0.00 | 308.55 |
| **Mobkl1b** | Mps one binder kinase activator-like 1B | Q921Y0 | 303.11 | 0.00 | 303.11 |
| **Ass1** | Argininosuccinate synthase | P16460 | 286.04 | 0.00 | 286.04 |
| **Nvl** | Nuclear valosin-containing protein-like | Q9DBY8 | 275.67 | 0.00 | 275.67 |
| **Calm1** | Calmodulin | P62204 | 263.64 | 0.00 | 263.64 |
| **Prdx1** | Peroxiredoxin-1 | P35700 | 263.20 | 0.00 | 263.20 |
| **Cct6a** | T-complex protein 1 subunit zeta | P80317 | 246.59 | 0.00 | 246.59 |
| **Abcf2** | ATP-binding cassette sub-family F member 2 | Q99LE6 | 229.36 | 0.00 | 229.36 |
| **Nop56** | Nucleolar protein 56 | Q9D6Z1 | 225.76 | 0.00 | 225.76 |
| **Rfc2** | Replication factor C subunit 2 | Q9WUK4 | 225.11 | 0.00 | 225.11 |
| **Dnaja2** | DnaJ homolog subfamily A member 2 | Q9QYJ0 | 222.47 | 0.00 | 222.47 |
| **Cct3** | T-complex protein 1 subunit gamma | P80318 | 216.23 | 0.00 | 216.23 |
| **Rbbp4** | Histone-binding protein RBBP4 | Q60972 | 215.67 | 0.00 | 215.67 |
| **Ddx21** | Nucleolar RNA helicase 2 | Q9JIK5 | 215.42 | 0.00 | 215.42 |
| **Ddx17** | Probable ATP-dependent RNA helicase DDX17 | Q501J6 | 604.35 | 390.90 | 213.44 |
| **Ywhae** | 14-3-3 protein epsilon | P62259 | 205.40 | 0.00 | 205.40 |
| **Sssca1** | Sjoegren syndrome/scleroderma autoantigen 1 homolog | P56873 | 197.40 | 0.00 | 197.40 |
| **Nop58** | Nucleolar protein 58 | Q6DFW4 | 195.44 | 0.00 | 195.44 |
| **Mycbp** | C-Myc-binding protein | Q9EQS3 | 889.89 | 704.82 | 185.07 |
| **Rfc3** | Replication factor C subunit 3 | Q8R323 | 183.91 | 0.00 | 183.91 |
| **Ran** | GTP-binding nuclear protein Ran | P62827 | 181.86 | 0.00 | 181.86 |
| **Trim28** | Transcription intermediary factor 1-beta | Q62318 | 439.61 | 261.14 | 178.47 |
| **Myl6** | Myosin light polypeptide 6 | Q60605 | 173.43 | 0.00 | 173.43 |
| **Nsun2** | tRNA (cytosine-5-)-methyltransferase NSUN2 | Q1HFZ0 | 172.97 | 0.00 | 172.97 |
| **Ruvbl1** | RuvB-like 1 | P60122 | 172.29 | 0.00 | 172.29 |
| **Caprin1** | Caprin-1 | Q60865 | 166.69 | 0.00 | 166.69 |

| name | description | Uniprot ID | Wdhd1 NSAFe5 | BG NSAFe5 | Wdhd1 - BG |
|---|---|---|---|---|---|
| **Ywhaz** | 14-3-3 protein zeta/delta | P63101 | 160.34 | 0.00 | 160.34 |
| **Ywhah** | 14-3-3 protein eta | P68510 | 159.68 | 0.00 | 159.68 |
| **Ywhag** | 14-3-3 protein gamma | P61982 | 159.04 | 0.00 | 159.04 |
| **Matr3** | Matrin-3 | Q8K310 | 154.78 | 0.00 | 154.78 |
| **Kpna2** | Importin subunit alpha-2 | P52293 | 148.52 | 0.00 | 148.52 |
| **Mcm5** | DNA replication licensing factor MCM5 | P49718 | 142.91 | 0.00 | 142.91 |
| **Ddx3x** | ATP-dependent RNA helicase DDX3X | Q62167 | 138.75 | 0.00 | 138.75 |
| **Ddx3y** | ATP-dependent RNA helicase DDX3Y | Q62095 | 138.75 | 0.00 | 138.75 |
| **D1Pas1** | Putative ATP-dependent RNA helicase Pl10 | P16381 | 138.75 | 0.00 | 138.75 |
| **Wdr5** | WD repeat-containing protein 5 | P61965 | 352.84 | 217.35 | 135.48 |
| **Npm1** | Nucleophosmin | Q61937 | 134.53 | 0.00 | 134.53 |
| **Gfpt2** | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] 2 | Q9Z2Z9 | 134.40 | 0.00 | 134.40 |
| **Pola1** | DNA polymerase alpha catalytic subunit | P33609 | 134.07 | 0.00 | 134.07 |
| **Dnaja1** | DnaJ homolog subfamily A member 1 | P63037 | 131.93 | 0.00 | 131.93 |
| **Pole** | DNA polymerase epsilon catalytic subunit A | Q9WVF7 | 126.18 | 0.00 | 126.18 |
| **Sec13** | Protein SEC13 homolog | Q9D1M0 | 122.00 | 0.00 | 122.00 |
| **Ppp1ca** | Serine/threonine-protein phosphatase PP1-alpha catalytic subunit | P62137 | 120.72 | 0.00 | 120.72 |
| **Ppp1cb** | Serine/threonine-protein phosphatase PP1-beta catalytic subunit | P62141 | 120.72 | 0.00 | 120.72 |
| **Ppp1cc** | Serine/threonine-protein phosphatase PP1-gamma catalytic subunit | P63087 | 120.72 | 0.00 | 120.72 |
| **Cnn3** | Calponin-3 | Q9DAW9 | 119.04 | 0.00 | 119.04 |
| **Mars** | Methionyl-tRNA synthetase, cytoplasmic | Q68FL6 | 116.13 | 0.00 | 116.13 |
| **Psat1** | Phosphoserine aminotransferase | Q99K85 | 106.17 | 0.00 | 106.17 |
| **Psme3** | Proteasome activator complex subunit 3 | P61290 | 103.10 | 0.00 | 103.10 |
| **Ilf3** | Interleukin enhancer-binding factor 3 | Q9Z1X4 | 102.07 | 0.00 | 102.07 |
| **Nap1l1** | Nucleosome assembly protein 1-like 1 | P28656 | 100.47 | 0.00 | 100.47 |
| **Rcor2** | REST corepressor 2 | Q8C796 | 100.15 | 0.00 | 100.15 |
| **Eif2b4** | Translation initiation factor eIF-2B subunit delta | Q61749 | 99.96 | 0.00 | 99.96 |
| **Ampd2** | AMP deaminase 2 | Q9DBT5 | 98.45 | 0.00 | 98.45 |
| **Cct2** | T-complex protein 1 subunit beta | P80314 | 97.90 | 0.00 | 97.90 |
| **Pola2** | DNA polymerase alpha subunit B | P33611 | 87.29 | 0.00 | 87.29 |
| **Rnh1** | Ribonuclease inhibitor | Q91VI7 | 86.15 | 0.00 | 86.15 |
| **Cnn2** | Calponin-2 | Q08093 | 85.86 | 0.00 | 85.86 |
| **Ppp6c** | Serine/threonine-protein phosphatase 6 catalytic subunit | Q9CQR6 | 85.86 | 0.00 | 85.86 |
| **Mthfd1** | C-1-tetrahydrofolate synthase, cytoplasmic | Q922D8 | 84.03 | 0.00 | 84.03 |
| **Pfkl** | 6-phosphofructokinase, liver type | P12382 | 83.94 | 0.00 | 83.94 |
| **Calu** | Calumenin | O35887 | 83.14 | 0.00 | 83.14 |
| **Trim25** | E3 ubiquitin/ISG15 ligase TRIM25 | Q61510 | 82.61 | 0.00 | 82.61 |
| **Rfc1** | Replication factor C subunit 1 | P35601 | 81.04 | 0.00 | 81.04 |

| name | description | Uniprot ID | Wdhd1 NSAFe5 | BG NSAFe5 | Wdhd1 - BG |
|---|---|---|---|---|---|
| **Orc1l** | Origin recognition complex subunit 1 | Q9Z1N2 | 77.94 | 0.00 | 77.94 |
| **Saps3** | Serine/threonine-protein phosphatase 6 regulatory subunit 3 | Q922D4 | 77.57 | 0.00 | 77.57 |
| **Rbm39** | RNA-binding protein 39 | Q8VH51 | 74.12 | 0.00 | 74.12 |
| **Gnl3** | Guanine nucleotide-binding protein-like 3 | Q8CI11 | 73.02 | 0.00 | 73.02 |
| **Cct5** | T-complex protein 1 subunit epsilon | P80316 | 72.61 | 0.00 | 72.61 |
| **Cct7** | T-complex protein 1 subunit eta | P80313 | 72.21 | 0.00 | 72.21 |
| **Hk2** | Hexokinase-2 | O08528 | 71.40 | 0.00 | 71.40 |
| **Tcp1** | T-complex protein 1 subunit alpha | P11983 | 70.65 | 0.00 | 70.65 |
| **Tpx2** | Targeting protein for Xklp2 | A2APB8 | 70.30 | 0.00 | 70.30 |
| **Rad50** | DNA repair protein RAD50 | P70388 | 69.86 | 0.00 | 69.86 |
| **Ppm1b** | Protein phosphatase 1B | P36993 | 67.15 | 0.00 | 67.15 |
| **Galk1** | Galactokinase | Q9R0N0 | 66.98 | 0.00 | 66.98 |
| **Ctps** | CTP synthase 1 | P70698 | 66.47 | 0.00 | 66.47 |
| **Mta3** | Metastasis-associated protein MTA3 | Q924K8 | 66.47 | 0.00 | 66.47 |
| **Pa2g4** | Proliferation-associated protein 2G4 | P50580 | 66.47 | 0.00 | 66.47 |
| **Lrrc40** | Leucine-rich repeat-containing protein 40 | Q9CRC8 | 65.25 | 0.00 | 65.25 |
| **Srp68** | Signal recognition particle 68 kDa protein | Q8BMA6 | 62.85 | 0.00 | 62.85 |
| **Paics** | Multifunctional protein ADE2 | Q9DCL9 | 61.62 | 0.00 | 61.62 |
| **Sall4** | Sal-like protein 4 | Q8BX22 | 61.36 | 0.00 | 61.36 |
| **Lsg1** | Large subunit GTPase 1 homolog | Q3UM18 | 61.00 | 0.00 | 61.00 |
| **Wdr18** | WD repeat-containing protein 18 | Q4VBE8 | 60.76 | 0.00 | 60.76 |
| **Eno1** | Alpha-enolase | P17182 | 60.34 | 0.00 | 60.34 |
| **Top2a** | DNA topoisomerase 2-alpha | Q01320 | 59.99 | 0.00 | 59.99 |
| **Smarcc1** | SWI/SNF complex subunit SMARCC1 | P97496 | 59.30 | 0.00 | 59.30 |
| **Mta2** | Metastasis-associated protein MTA2 | Q9R190 | 58.81 | 0.00 | 58.81 |
| **Hspa9** | Stress-70 protein, mitochondrial | P38647 | 57.85 | 0.00 | 57.85 |
| **Adss** | Adenylosuccinate synthetase isozyme 2 | P46664 | 57.43 | 0.00 | 57.43 |
| **Zmym2** | Zinc finger MYM-type protein 2 | Q9CU65 | 57.10 | 0.00 | 57.10 |
| **Umps** | Uridine 5'-monophosphate synthase | P13439 | 54.45 | 0.00 | 54.45 |
| **Atp6v1h** | V-type proton ATPase subunit H | Q8BVE3 | 54.22 | 0.00 | 54.22 |
| **Rif1** | Telomere-associated protein RIF1 | Q6PR54 | 54.13 | 0.00 | 54.13 |
| **Trm1l** | TRM1-like protein | A2RSY6 | 53.96 | 0.00 | 53.96 |
| **Sf3b3** | Splicing factor 3B subunit 3 | Q921M3 | 53.80 | 0.00 | 53.80 |
| **Nat10** | N-acetyltransferase 10 | Q8K224 | 51.15 | 0.00 | 51.15 |
| **L1td1** | LINE-1 type transposase domain-containing protein 1 | Q587J6 | 50.23 | 0.00 | 50.23 |
| **Pfkp** | 6-phosphofructokinase type C | Q9WUA3 | 50.11 | 0.00 | 50.11 |
| **Paf1** | RNA polymerase II-associated factor 1 homolog | Q8K2T8 | 48.95 | 0.00 | 48.95 |
| **Cct4** | T-complex protein 1 subunit delta | P80315 | 48.59 | 0.00 | 48.59 |
| **Rest** | RE1-silencing transcription factor | Q8VIG1 | 48.41 | 0.00 | 48.41 |

| name | description | Uniprot ID | Wdhd1 NSAFe5 | BG NSAFe5 | Wdhd1 - BG |
|---|---|---|---|---|---|
| **Cttn** | Src substrate cortactin | Q60598 | 47.96 | 0.00 | 47.96 |
| **Hells** | Lymphocyte-specific helicase | Q60848 | 47.85 | 0.00 | 47.85 |
| **Pold1** | DNA polymerase delta catalytic subunit | P52431 | 47.40 | 0.00 | 47.40 |
| **Msto1** | Protein misato homolog 1 | Q2YDW2 | 47.10 | 0.00 | 47.10 |
| **Cnot6** | CCR4-NOT transcription complex subunit 6 | Q8K3P5 | 47.02 | 0.00 | 47.02 |
| **Nedd4** | E3 ubiquitin-protein ligase NEDD4 | P46935 | 44.29 | 0.00 | 44.29 |
| **Tnpo1** | Transportin-1 | Q8BFY9 | 43.74 | 0.00 | 43.74 |
| **Ppp2r1b** | Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A beta isoform | Q7TNP2 | 43.57 | 0.00 | 43.57 |
| **Cand1** | Cullin-associated NEDD8-dissociated protein 1 | Q6ZQ38 | 42.58 | 0.00 | 42.58 |
| **Cdc16** | Cell division cycle protein 16 homolog | Q8R349 | 42.24 | 0.00 | 42.24 |
| **Wdr3** | WD repeat-containing protein 3 | Q8BHB4 | 41.70 | 0.00 | 41.70 |
| **Ewsr1** | RNA-binding protein EWS | Q61545 | 39.98 | 0.00 | 39.98 |
| **Mybbp1a** | Myb-binding protein 1A | Q7TPV4 | 38.97 | 0.00 | 38.97 |
| **Bub1b** | Mitotic checkpoint serine/threonine-protein kinase BUB1 beta | Q9Z1S0 | 37.34 | 0.00 | 37.34 |
| **Mcm7** | DNA replication licensing factor MCM7 | Q61881 | 36.42 | 0.00 | 36.42 |
| **Ubap2l** | Ubiquitin-associated protein 2-like | Q80X50 | 35.49 | 0.00 | 35.49 |
| **Suz12** | Polycomb protein Suz12 | Q80U70 | 35.34 | 0.00 | 35.34 |
| **Nsf** | Vesicle-fusing ATPase | P46460 | 35.20 | 0.00 | 35.20 |
| **Cul1** | Cullin-1 | Q9WTX6 | 33.75 | 0.00 | 33.75 |
| **Zw10** | Centromere/kinetochore protein zw10 homolog | O54692 | 33.62 | 0.00 | 33.62 |
| **Gsn** | Gelsolin | P13020 | 33.57 | 0.00 | 33.57 |
| **Nop2** | Putative ribosomal RNA methyltransferase NOP2 | Q922K7 | 33.02 | 0.00 | 33.02 |
| **Plaa** | Phospholipase A-2-activating protein | P27612 | 32.98 | 0.00 | 32.98 |
| **Mcm6** | DNA replication licensing factor MCM6 | P97311 | 31.90 | 0.00 | 31.90 |
| **Sin3a** | Paired amphipathic helix protein Sin3a | Q60520 | 30.64 | 0.00 | 30.64 |
| **Mcm4** | DNA replication licensing factor MCM4 | P49717 | 30.38 | 0.00 | 30.38 |
| **Pml** | Probable transcription factor PML | Q60953 | 29.59 | 0.00 | 29.59 |
| **Pds5a** | Sister chromatid cohesion protein PDS5 homolog A | Q6A026 | 29.49 | 0.00 | 29.49 |
| **Zc3h18** | Zinc finger CCCH domain-containing protein 18 | Q0P678 | 27.62 | 0.00 | 27.62 |
| **Kif5b** | Kinesin-1 heavy chain | Q61768 | 27.19 | 0.00 | 27.19 |
| **Cse1l** | Exportin-2 | Q9ERK4 | 26.97 | 0.00 | 26.97 |
| **Mthfd1l** | Monofunctional C1-tetrahydrofolate synthase, mitochondrial | Q3V3R1 | 26.80 | 0.00 | 26.80 |
| **Skiv2l2** | Superkiller viralicidic activity 2-like 2 | Q9CZU3 | 25.18 | 0.00 | 25.18 |
| **Smarca5** | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5 | Q91ZW3 | 24.92 | 0.00 | 24.92 |
| **Akap12** | A-kinase anchor protein 12 | Q9WTQ5 | 23.33 | 0.00 | 23.33 |
| **Ube4b** | Ubiquitin conjugation factor E4 B | Q9ES00 | 22.33 | 0.00 | 22.33 |
| **Kif4** | Chromosome-associated kinesin KIF4 | P33174 | 21.27 | 0.00 | 21.27 |

| name | description | Uniprot ID | Wdhd1 NSAFe5 | BG NSAFe5 | Wdhd1 - BG |
|---|---|---|---|---|---|
| **Chd4** | Chromodomain-helicase-DNA-binding protein 4 | Q6PDQ2 | 20.51 | 0.00 | 20.51 |
| **Cpsf1** | Cleavage and polyadenylation specificity factor subunit 1 | Q9EPU4 | 18.17 | 0.00 | 18.17 |
| **Kntc1** | Kinetochore-associated protein 1 | Q8C3Y4 | 17.80 | 0.00 | 17.80 |
| **Zmym4** | Zinc finger MYM-type protein 4 | A2A791 | 16.91 | 0.00 | 16.91 |
| **Ubr5** | E3 ubiquitin-protein ligase UBR5 | Q80TP3 | 14.07 | 0.00 | 14.07 |
| **Ckap5** | Cytoskeleton-associated protein 5 | A2AGT5 | 12.89 | 0.00 | 12.89 |
| **Ranbp2** | E3 SUMO-protein ligase RanBP2 | Q9ERU9 | 12.87 | 0.00 | 12.87 |
| **Hcfc1** | Host cell factor 1 | Q61191 | 12.81 | 0.00 | 12.81 |
| **Sptbn1** | Spectrin beta chain, brain 1 | Q62261 | 11.08 | 0.00 | 11.08 |
| **Srrm2** | Serine/arginine repetitive matrix protein 2 | Q8BTI8 | 9.69 | 0.00 | 9.69 |

## 3.4 Discussion

In summary, this work demonstrated a comprehensive approach to identify PCNA-interacting proteins and allowed the identification of PIP box-containing novel interaction partners, of which we were able to validate the DNA replication factor Wdhd1 as an example. Furthermore, we were able to demonstrate the S-phase-specific interactions of PCNA with various proteins, amongst which are notably a number of chromatin-modifying enzymes as suggested by earlier studies (Gondor and Ohlsson, 2009; Moldovan et al., 2007; Ohta et al., 2002). Interestingly, under our experimental conditions, the interaction of Wdhd1 with PCNA does not appear to be required for its localization. The interactome identified for Wdhd1 confirms its interaction with components of the DNA replication machinery and further suggests that it may also play a role in other molecular processes such as chromatin state regulation.

Previously, there have been attempts to understand the PCNA interactome by proteomics approaches such as by Ohta and colleagues (Ohta et al., 2002) and reviewed by S. N. Naryzhny (Naryzhny, 2008), G-L Moldovan (Moldovan et al., 2007) and A. De Biasio (De Biasio and Blanco, 2013), but none of them addressed the temporal differences of the various interactions at different stages of reprogramming. In fact, De Biasio and Blanco even suggested that more attention should be paid to transient PCNA interactions, which is exactly what this work attempted to do. Especially the alignment of the identified interaction partners with a systematic search of PIP box motifs in their amino acid should highlight potential direct interaction partners and such this work will help guide future research and can serve as a resource for other investigators. Alternative ways to directly interact with PCNA through for example through an AlkB homologue 2 PCNA-interacting motif (APIM) have been described in the literature (Gilljam et al., 2009), but due to the relatively limited knowledge on those, was omitted in this work. It will be interesting to understand more about alternative ways in which PCNA can bind other proteins and integrate this knowledge with our data collection on the time-specific PCNA-interactome.

It is still unclear how chromatin marks are inherited after the passing of the replication fork has displaced nucleosomes from the DNA template. It is likely that multiple independent mechanisms are in place for the various chromatin marks, one of which might be the PCNA-mediated recruitment of chromatin-modifying enzymes to the replication fork, possibly in a temporally regulated fashion (Probst et al., 2009). The identification of several chromatin-modifying enzymes as PCNA interaction partners described in this work, amongst which we find euchromatin-establishing enzymes associated specifically early in S-phase and enzymes required for the establishment of heterochromatin late in S-phase, supports this model.

We validated the interaction with the DNA replication factor Wdhd1, which has been shown to also have a role in the regulation of the stability of the chromatin-modifying enzyme Gcn5 (Li et al., 2012b). When analyzing Wdhd1-associated proteins, we not only observed the expected DNA replication factors but additionally also chromatin modifiers such as components of the NURD chromosome remodeling complex. This further establishes a close connection between the DNA replication machinery and the regulation of chromatin state. Mass spectrometric approaches to identify Wdhd1 interaction partners have been performed before but were less comprehensive and unbiased than our approach and thus did not reach the same depth in interaction protein discovery that our study demonstrates (Li et al., 2012a).

It was interesting to observe that the PCNA interaction was not required for localization of Wdhd1 under our experimental conditions and implies that this interaction might serve a different purpose than recruitment to replication forks. Alternative molecular functions of this interaction might be the regulation of protein stability. PCNA is known to act as a targeting factor for the ubiquitin-mediated degradation of interaction partners, a hypothesis that will be interesting to test in future studies.

## 3.5 Material and Methods

Cell culture, generation of cell lines, cell extraction and western blot

All mouse ESC lines were maintained on a feeder cell layer in ESC medium containing leukemia inhibitory factor (LIF), 15% fetal bovine serum, nonessential amino acids, glutamate and penicillin/streptomycin in knockout DMEM medium on gelatinized cell culture dishes.

3xFlag-6xHis-tagged proteins were targeted to the Collagen locus under control of the Tet operon of mouse ESC by FRT/FLPase mediated recombination as previously described (Beard et al., 2006).

To generate whole cell extracts, cells were trypsinized and resuspended in lysis buffer (300mM Tris-HCl pH 6.8, 10 % SDS, 15 % glycerol, 1 mM DTT). The proteins were separated by gel electrophoresis on 4-12 % Acrylamide gels at 140 V at room temperature for 60 mins. For Western blot, the proteins were transferred to nitrocellulose membranes. The membranes were blocked with Odyssey blocking buffer (LiCOR) for one hour at room temperature and incubated with primary antibodies as indicated over night at 4 C. After washing with PBS 0.1% Tween three times, membranes were incubated with secondary antibodies (1:10000, LiCOR) for one hour at room temperature. Protein bands were visualized using the LiCOR infrared scanner and quantified with the corresponding software.

For silver stains, the membranes were fixed and stained using the SilverQuest Staining Kit (Invitrogen) according to the manufacturer's instructions.

Cell synchronization and flow cytometric analysis

Cells were synchronized in G2/M-phase by incubation with the microtubule inhibitor nocodazole at 45 ng/mL of culture medium over night. Following the incubation, cells were gently washed

with phosphate buffered saline (PBS) and released into nocodazole-free culture medium for either 5 or 9 hours and flow cytometric analysis of their DNA content performed.

To this end, cells were trypsinized, washed with PBS and fixed in 50 % ice-cold ethanol for one hour at -20 C. The fixed cells were then spun at 13000 rpm at room temperature in a table top centrifuge and incubated for 30 mins at room temperature in the dark in propidium-iodide buffer (3mM EDTA pH8.0, 0.05% NP40, 50ug/ml propidium iodide, 1mg/ml RNAseA in PBS).

Flow cytometric analysis was then performed using a BD FACSDiva flow cytometer and absorbance in the PE channel measuered. Data analysis was performed with the corresponding FlowJo software.

Immunoprecipitation and mass spectrometry

Cells were treated with 2 ug/ml doxicycline in ESC medium for 48 hours on gelatinized culture plates without feeders before harvesting. Nuclear extracts were generated by dignam extraction with 420mM salt as previously described (Carey et al., 2009). The extracts were dialized to 100 mM salt concentration and immunoprecipitation performed with antiFlag M2 agarose beads (invitrogen) for 3 hours at 4 C. Complexes were eluted with 25 ug/ml flag peptide in 4 sequential elutions for 15 mins at room temperature. Each eluate was verified by western blot and silver stain before they were pooled. Pooled eluates were subjected to MudPIT mass spectrometry as previously described (Wolters et al., 2001).

MudPIT data analysis

Lists of proteins identified by MudPIT were manually cleared of common contaminants such as ribosomal proteins, heat shock proteins, keratin, proteins involved in mRNA processing, proteasome as well as actin, myosin and immunoglobulin. Natural spectral absorbance factors (NSAF) which represent the abundance of the identified protein in the sample were then

adjusted by substracting the value obtained from a control Immunoprecipitation from that obtained from Immunoprecipitation from Flag-tagged protein expressing cells to allow for identification of differential enrichment of proteins between experimental condition and control. All proteins that scored a higher differential NSAF than the average NSAF of the proteins not identified in the control were included for further analysis.

The list of proteins obtained in this way was then subjected to network identification using the STRING database (Franceschini et al., 2013) with the criteria of only allowing protein-protein interactions that were identified either by experiments or databases at at least a medium confidence score.

The such obtained network file was then adjusted to contain the additionally identified interactions from this study and visualized using the cytoscape analysis software (Shannon et al., 2003). Edges were annotated corresponding to identification of the interaction either through STRING ('known') or this study ('new') and nodes annotated for known PCNA interaction partners also obtained from the STRING database under the same criteria, proteins with validated PIP box motifs as described in Moldovan *et al.* (Moldovan et al., 2007) ('known' or 'not described'), proteins in which we observed a PIP box motif in their amino acid sequence (PIP motif match 'yes' or 'no') as well as for proteins that fall under the gene ontology term 'covalent chromatin modification'.

Immunofluorescence

Glass coverslips were washed with 70% Ethanol and coated with porcupine gelatine. Cells were plated directly on the coverslips and induced with 2 uM doxicycline for 24 hours. Cells were then extracted on ice with CSK buffer (100 mM PIPES pH6.8, 300 mM sucrose, 3 mM $MgCl_2$, 100 mM NaCl), CSK-0.5% Triton and CSK for 5 minutes each before they were fixed with 4% paraformaldehyde. The samples were incubated in blocking buffer (5% normal goat serum, 0.2

% Tween20, 2 % fish skin gelatine in PBS) for 30 minutes at room temperature, in primary antibody dilution (mouse a Flag, rab a PCNA, 1:200 each) at 4 C over night, washed 3 x 5 minutes in PBS 0.2% tween20, incubated in secondary antibody (alexafluor 488 anti mus, alexafluor 546 anti rabbit, 1:2000 each) for 1 hour at room temperature, washed 3x 5 mins in PBS 0.2% tween20 in which the middle step was supplemented with DAPI for visualization of DNA. Pictures were obtained using either a Zeiss Epifluorescence or Confocal microscope and adjusted using the corresponding software.

## 3.6 References


Alabert, C., and Groth, A. (2012). Chromatin replication and epigenome maintenance. Nat. Rev. Mol. Cell Biol. *13*, 153–167.

Baumgartner, M., Tardieux, I., Ohayon, H., Gounon, P., and Langsley, G. (1999). The use of nocodazole in cell cycle analysis and parasite purification from Theileria parva-infected B cells. Microbes Infect. Inst. Pasteur *1*, 1181–1188.

Beard, C., Hochedlinger, K., Plath, K., Wutz, A., and Jaenisch, R. (2006). Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. Genes. New York N 2000 *44*, 23–28.

Bermudez, V.P., Farina, A., Tappin, I., and Hurwitz, J. (2010). Influence of the Human Cohesion Establishment Factor Ctf4/AND-1 on DNA Replication. J. Biol. Chem. *285*, 9493 –9505.

De Biasio, A., and Blanco, F.J. (2013). Chapter One - Proliferating Cell Nuclear Antigen Structure and Interactions: Too Many Partners for One Dancer? In Advances in Protein Chemistry and Structural Biology, Rossen Donev, ed. (Academic Press), pp. 1–36.

Carey, M.F., Peterson, C.L., and Smale, S.T. (2009). Dignam and Roeder Nuclear Extract Preparation. Cold Spring Harb Protoc *2009*, pdb.prot5330.

Chen, D., Jian, Y., Liu, X., Zhang, Y., Liang, J., Qi, X., Du, H., Zou, W., Chen, L., Chai, Y., et al. (2013). Clathrin and AP2 Are Required for Phagocytic Receptor-Mediated Apoptotic Cell Clearance in Caenorhabditis elegans. PLoS Genet *9*, e1003517.

Chilkova, O., Stenlund, P., Isoz, I., Stith, C.M., Grabowski, P., Lundström, E.-B., Burgers, P.M., and Johansson, E. (2007). The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. Nucleic Acids Res. *35*, 6588–6597.

Chuang, L.S.-H., Ian, H.-I., Koh, T.-W., Ng, H.-H., Xu, G., and Li, B.F.L. (1997). Human DNA-(Cytosine-5) Methyltransferase-PCNA Complex as a Target for p21WAF1. Science *277*, 1996 –2000.

Dehennaut, V., Slomianny, M.-C., Page, A., Vercoutter-Edouart, A.-S., Jessus, C., Michalski, J.-C., Vilain, J.-P., Bodart, J.-F., and Lefebvre, T. (2008). Identification of Structural and Functional O-Linked N-Acetylglucosamine-bearing Proteins in Xenopus laevis Oocyte. Mol. Cell. Proteomics *7*, 2229–2245.

Estève, P.-O., Chin, H.G., Smallwood, A., Feehery, G.R., Gangisetty, O., Karpf, A.R., Carey, M.F., and Pradhan, S. (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. Genes Dev. *20*, 3089–3103.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. *41*, D808–815.

Gilljam, K.M., Feyzi, E., Aas, P.A., Sousa, M.M.L., Müller, R., Vågbø, C.B., Catterall, T.C., Liabakk, N.B., Slupphaug, G., Drabløs, F., et al. (2009). Identification of a novel, widespread, and functionally important PCNA-binding motif. J. Cell Biol. *186*, 645 –654.

Gondor, A., and Ohlsson, R. (2009). Replication timing and epigenetic reprogramming of gene expression: a two-way relationship? Nat Rev Genet *10*, 269–276.

He, G., Kuang, J., Koomen, J., Kobayashi, R., Khokhar, A.R., and Siddik, Z.H. (2013). Recruitment of trimeric proliferating cell nuclear antigen by G1-phase cyclin-dependent kinases following DNA damage with platinum-based antitumour agents. Br. J. Cancer.

Hsieh, C.-L., Lin, C.-L., Liu, H., Chang, Y.-J., Shih, C.-J., Zhong, C.Z., Lee, S.-C., and Tan, B.C.-M. (2011). WDHD1 modulates the post-transcriptional step of the centromeric silencing pathway. Nucleic Acids Res. *39*, 4048–4062.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129–1133.

Jaramillo-Lambert, A., Hao, J., Xiao, H., Li, Y., Han, Z., and Zhu, W. (2013). Acidic Nucleoplasmic DNA-binding Protein (And-1) Controls Chromosome Congression by Regulating the Assembly of Centromere Protein A (CENP-A) at Centromeres. J. Biol. Chem. *288*, 1480–1488.

Koundrioukoff, S., Jónsson, Z.O., Hasan, S., de Jong, R.N., van der Vliet, P.C., Hottiger, M.O., and Hübscher, U. (2000). A Direct Interaction between Proliferating Cell Nuclear Antigen (PCNA) and Cdk2 Targets PCNA-interacting Proteins for Phosphorylation. J. Biol. Chem. *275*, 22882 –22887.

Kouprina, N., Kroll, E., Bannikov, V., Bliskovsky, V., Gizatullin, R., Kirillov, A., Shestopalov, B., Zakharyev, V., Hieter, P., and Spencer, F. (1992). CTF4 (CHL15) mutants exhibit defective DNA metabolism in the yeast Saccharomyces cerevisiae. Mol. Cell. Biol. *12*, 5736–5747.

Kurki, P., Vanderlaan, M., Dolbeare, F., Gray, J., and Tan, E.M. (1986). Expression of proliferating cell nuclear antigen (PCNA)/cyclin during the cell cycle. Exp. Cell Res. *166*, 209–219.

Li, Y., Xiao, H., Renty, C. de, Jaramillo-Lambert, A., Han, Z., DePamphilis, M.L., Brown, K.J., and Zhu, W. (2012a). The Involvement of Acidic Nucleoplasmic DNA-binding Protein (And-1) in the Regulation of Prereplicative Complex (pre-RC) Assembly in Human Cells. J. Biol. Chem. *287*, 42469–42479.

Li, Y., Jaramillo-Lambert, A.N., Yang, Y., Williams, R., Lee, N.H., and Zhu, W. (2012b). And-1 is required for the stability of histone acetyltransferase Gcn5. Oncogene *31*, 643–652.

Majka, J., and Burgers, P.M.J. (2004). The PCNA-RFC families of DNA clamps and clamp loaders. Prog. Nucleic Acid Res. Mol. Biol. *78*, 227–260.

Moldovan, G.-L., Pfander, B., and Jentsch, S. (2007). PCNA, the Maestro of the Replication Fork. Cell *129*, 665–679.

Naryzhny, S.N. (2008). Proliferating cell nuclear antigen: a proteomics view. Cell. Mol. Life Sci. *65*, 3789–3808.

Ohta, S., Shiomi, Y., Sugimoto, K., Obuse, C., and Tsurimoto, T. (2002). A Proteomics Approach to Identify Proliferating Cell Nuclear Antigen (PCNA)-binding Proteins in Human Cell Lysates. J. Biol. Chem. *277*, 40362 –40367.

Poot, R.A., Bozhenok, L., van den Berg, D.L.C., Steffensen, S., Ferreira, F., Grimaldi, M., Gilbert, N., Ferreira, J., and Varga-Weisz, P.D. (2004). The Williams syndrome transcription factor interacts with PCNA to target chromatin remodelling by ISWI to replication foci. Nat. Cell Biol. *6*, 1236–1244.

Probst, A.V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. Nat Rev Mol Cell Biol *10*, 192–206.

Rountree, M.R., Bachman, K.E., and Baylin, S.B. (2000). DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nat. Genet. *25*, 269–277.

Rowbotham, S.P., Barki, L., Neves-Costa, A., Santos, F., Dean, W., Hawkes, N., Choudhary, P., Will, W.R., Webster, J., Oxley, D., et al. (2011). Maintenance of Silent Chromatin through Replication Requires SWI/SNF-like Chromatin Remodeler SMARCAD1. Mol. Cell *42*, 285–296.

Schermelleh, L., Haemmer, A., Spada, F., Rösing, N., Meilinger, D., Rothbauer, U., Cardoso, M.C., and Leonhardt, H. (2007). Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. Nucleic Acids Res. *35*, 4301 – 4312.

Schneider, K., Fuchs, C., Dobay, A., Rottach, A., Qin, W., Wolf, P., Álvarez-Castro, J.M., Nalaskowski, M.M., Kremmer, E., Schmid, V., et al. (2013). Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling. Nucleic Acids Res. *41*, 4860–4876.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Shi, F.-T., Kim, H., Lu, W., He, Q., Liu, D., Goodell, M.A., Wan, M., and Songyang, Z. (2013). Ten-Eleven Translocation 1 (Tet1) Is Regulated by O-Linked N-Acetylglucosamine Transferase (Ogt) for Target Gene Repression in Mouse Embryonic Stem Cells. J. Biol. Chem. *288*, 20776–20784.

Sridharan, R., Gonzales-Cope, M., Chronis, C., Bonora, G., McKee, R., Huang, C., Patel, S., Lopez, D., Mishra, N., Pellegrini, M., et al. (2013). Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. Nat. Cell Biol. *15*, 872–882.

Takahashi, Y., Westfield, G.H., Oleskie, A.N., Trievel, R.C., Shilatifard, A., and Skiniotis, G. (2011). Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human. Proc. Natl. Acad. Sci. U. S. A. *108*, 20526–20531.

Warbrick, E. (1998). PCNA binding through a conserved motif. BioEssays *20*, 195–199.

Wolters, D.A., Washburn, M.P., and Yates, J.R. (2001). An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem. *73*, 5683–5690.

Xue, Y., Canman, J.C., Lee, C.S., Nie, Z., Yang, D., Moreno, G.T., Young, M.K., Salmon, E.D., and Wang, W. (2000). The human SWI/SNF-B chromatin-remodeling complex is related to yeast rsc and localizes at kinetochores of mitotic chromosomes. Proc. Natl. Acad. Sci. U. S. A. *97*, 13015–13020.

Yoshizawa-Sugata, N., and Masai, H. (2009). Roles of human AND-1 in chromosome transactions in S phase. J. Biol. Chem. *284*, 20718–20728.

Yu, Y., Song, C., Zhang, Q., DiMaggio, P.A., Garcia, B.A., York, A., Carey, M.F., and Grunstein, M. (2012). Histone H3 lysine 56 methylation regulates DNA replication through its interaction with PCNA. Mol. Cell *46*, 7–17.

Yuan, J., Ghosal, G., and Chen, J. The HARP-like Domain-Containing Protein AH2/ZRANB3 Binds to PCNA and Participates in Cellular Response to Replication Stress. Mol. Cell.

Zhu, W., Ukomadu, C., Jha, S., Senga, T., Dhar, S.K., Wohlschlegel, J.A., Nutt, L.K., Kornbluth, S., and Dutta, A. (2007). Mcm10 and And-1/CTF4 recruit DNA polymerase α to chromatin for initiation of DNA replication. Genes Dev. *21*, 2288–2299.

Zlatanou, A., Despras, E., Braz-Petta, T., Boubakour-Azzouz, I., Pouvelle, C., Stewart, G.S., Nakajima, S., Yasui, A., Ishchenko, A.A., and Kannouche, P.L. (2011). The hMsh2-hMsh6 Complex Acts in Concert with Monoubiquitinated PCNA and Pol η in Response to Oxidative DNA Damage in Human Cells. Mol. Cell *43*, 649–662.

# 4 Zranb3 is a novel PCNA interaction partner and involved in DNA mismatch repair in embryonic stem cslls

## 4.1 Abstract

Zranb3 is a novel player in DNA repair that was first described in human cancer cell lines. It is one of the few proteins with annealing helicase function and interacts with ubiquitinylated PCNA. We demonstrate that *Zranb3* is highly expressed in mouse embryonic stem cells (mESCs) and describe the interaction of Zranb3 with PCNA in ESCs by mass spectrometry (MudPIT) and co-immunoprecipitation (Co-IP), which depends on the PIP_box, a conserved PCNA-interaction motif. We report that Zranb3, in contrast to other DNA repair proteins, does not appear to play a role in the regulation of pluripotency and self-renewal of ESCs. However, the mass spectrometry results reveal novel interaction partners of Zranb3 such as the MutS$\alpha$ complex, suggesting a link of Zranb3 to the MMR pathway. We demonstrate that Zranb3 shows a punctate nuclear staining pattern by immunofluorescence in ESCs, reminiscent of other repair proteins that depend on the presence of MutS$\alpha$. is the involvement of Zranb3 in MMR is supported by cytotoxicity assays in which the depletion of Zranb3 confers higher resistance to Temezolomide and sensitivity to Camptothecin, which is typical for proteins involved in MMR (Liu et al., 1996). Our data suggest a model in which the activity of the MutS$\alpha$ complex leads to ubiquitinylation and recruitment of PCNA to DNA repair sites, which is required for Zranb3 localization at these sites.

## 4.2 Introduction

Zranb3 (Zinc finger Ran binding 3, also referred to as annealing helicase AH2) is a recently identified protein that has annealing helicase function (Yusufzai and Kadonaga, 2010). Zranb3 and the protein HARP (also known as SmarcaL1) share high sequence similarity in their

helicase domain and are the only two known mammalian proteins known to solely possess annealing helicase activity but not unwinding activity for double-stranded DNA like other helicase enzymes (Yusufzai and Kadonaga, 2008). Annealing helicases convert single stranded DNA into double strands in *in vitro* assays and might exert this activity at any site where DNA-fork structures occur *in vivo,* such as DNA replication forks and repair sites (Wu, 2012)(Yusufzai and Kadonaga, 2011). HARP is recruited to stalled replication forks by the single strand binding protein RPA through a specific interaction domain (Driscoll and Cimprich, 2009)(Ciccia et al., 2009). Zranb3, however, lacks a RPA-interaction domain and, instead, interacts with ubiquitinated proliferating cell nuclear antigen (PCNA) through both a conserved PIP box (PCNA interacting protein) domain and a ubiquitin binding zinc finger (Ciccia et al., 2012) (see our findings described below). The biological significance of the activity of annealing helicases is unclear but the current view is that they act in concert with repair pathways to protect stalled replication forks and help them restart after stalling to protect the integrity of the genome (Yuan et al., 2012). Biochemical evidence confirms this activity for Zranb3 but suggests that HARP and Zranb3 may act in different pathways, and that, like HARP activity is guided by its interaction with RPA, Zranb3's activity and recruitment might be regulated by the interaction with PCNA (Bétous et al., 2013).

Zranb3 shows a punctuate nuclear pattern by immunofluorescence staining pattern that is dependent on its interaction with PCNA, a staining pattern that is characteristic of proteins involved in DNA repair (Buonomo et al., 2009). Knockdown of Zranb3 in human cancer cell lines leads to higher sensitivity to DNA damage inducing agents such as camptothecin, demonstrating its putative involvement in the DNA damage repair pathway. Upon UV laser-induced DNA damage, Zranb3 localizes to the sites of damage, further indicating a role in the response to such treatments (Weston et al., 2012). However, the exact role of Zranb3 in this process remains unclear. Therefore, it would be informative to not only understand which repair

processes Zranb3 acts in, but also how this is affected by its interaction with PCNA and whether other mechanisms exist for Zranb3 targeting.

PCNA is a homotrimeric protein that acts as a loading clamp for DNA polymerase during DNA replication and is also involved in several DNA repair pathways (Moldovan et al., 2007). It has a plethora of interaction partners, most of which interact with it through a conserved PCNA-interacting protein motif (PIP box) (Warbrick, 1998). Posttranslational modifications of PCNA are believed to regulate which exact set of interactors can bind. The process PCNA is acting in is often defined by its protein interactors and therefore regulated by its posttranslational modifications (Mailand et al., 2013).

Interestingly, DNA damage leads to the ubiquitinylation of PCNA. Ubiquitinylated PCNA has not only been demonstrated to recruit other factors such as the repair DNA polymerase $\eta$, but also is specifically recognized and bound by the PIP box and Zinc finger (Znf) of Zranb3 (van der Kemp et al., 2009).

The ubiquitinylation of PCNA is dependent on the activity of the DNA mismatch repair (MMR) complex MutS$\alpha$ that is comprised of the two components Msh2 and Msh6. The depletion of Msh2 has been shown to dramatically reduce the levels of ubiquitinylated PCNA upon induction of DNA damage (Zlatanou et al., 2011). Importantly, MutS$\alpha$ is also particularly abundant in ESCs, suggesting that ESCs support high levels of MMR (Tichy et al., 2011). As MutS$\alpha$ is a key component of the MMR pathway and is required for ubiquitinylation of PCNA, we hypothesized that Zranb3 might also play a role in the MMR process. Furthermore, since the MMR pathway is especially active in ESCs, we determined whether Zranb3 is required for the maintencance of pluripotency and its establishment by transcription factor-induced reprogramming to the pluripotent state (Fong et al., 2011).

## 4.3 Results

Zranb3 is highly expressed in embryonic stem cells (ESCs)

While studying the interactome of PCNA in mouse ESCs, we identified Zranb3 as a novel interaction partner of Flag-PCNA by MudPIT analysis (Figure 4.1 A, see (chapter 3). At this time, the only publications on Zranb3 described Zranb3 as a novel protein that can function as annealing helicase and therefore has a putative role in DNA repair and replication mechanisms. DNA repair pathways are particularly important in ESCs and many components of those pathways are highly expressed in this cell type, stressing their importance for the faithful propagation of the genomic information from pluripotent cells to more differentiated cell types. Therefore, we sought to determine the levels of Zranb3 protein and RNA in ESCs relative to more differentiated mouse embryonic fibroblasts (MEFs) and to explore its biological function specifically in ESCs. We tested Zranb3 expression levels by western blotting and qPCR (Figure 4.1 B and C). There is a markedly higher level (about 5 fold) of both the protein as well as the RNA in ESCs, confirming that it might be a general phenomenon that proteins that are associated with DNA repair and likely also DNA replication, are more highly expressed in ESC. These findings were further validated by expression data from RNA microarray and RNA sequencing data (Supplementary Figure 4.1 A). We conclude that Zranb3 is a highly expressed factor in mouse ESCs and that its levels decrease with differentiation, which made ESCs a good system to further study the biological function of Zranb3.

PCNA is required for recruitment of Zranb3 to repair sites

By screening the amino acid sequence of Zranb3, we found that Zranb3 contains a conserved PCNA (proliferating cell nuclear antigen) interaction motif, the PIP box (Figure 4.2 A). To be able to study the role of Zranb3 and is interaction with the DNA replication and repair factor PCNA, we generated mouse ESC lines that express either c-terminally 3xFlag-tagged wt or PIP-box

mutant Zranb3. The mutation that was introduced is indicated in Figure 4.2 A and has been shown to completely abrogate PCNA binding in other proteins (Chuang et al., 1997). To generate these lines, we used site-directed integration into a FRT recombination site downstream of the Collagen1a locus, to integrate a single copy of the Zranb3/Flag cDNA under the control of a tet-inducible promoter, allowing us to express the transgene by addition of DOX (Beard et al., 2006) (Supplementary Figure 4.2 A). Successful expression upon DOX induction was verified by western blotting (Supplementary Figure 4.2 B and data not shown). To obtain the best enrichment in our flag-purifications, we chose to use the highest concentration of doxycycline (2 ug/ml) as this produced the most robust overexpression of flag-Zranb3.

To confirm the interaction between Zranb3 and PCNA, Flag-Zranb3 immunoprecipitation (IP) with Flag-sepharose beads was carried out upon generation of nuclear extracts. Successful protein extraction and immunoprecipitation of Flag-Zranb3 was verified by western blotting and silver staining (Supplementary Figure 4.2 C and D). Western blot analysis of the immunoprecipitates confirmed the interaction of Zranb3 with PCNA in mouse ESC and it's dependence on the PIP box motif (Figure 4.2 C), validating our mass spectrometry findings and the presence of a functional PIP box.

We also found that wild-type Flag-Zranb3 shows a punctate pattern in ESC nuclei in immunofluorescence staining assays and co-localized with PCNA. Notably, quantification of the percentage of ESCs containing foci demonstrated that foci number is reduced when the PIP mutant Zranb3 is expressed, suggesting that this localization pattern is dependent on the PCNA-interaction (Figures 4.2 C and D). Specifically, when co-staining ESCs for Flag (detecting the overexpressed Flag-Zranb3) and PCNA, we observed about 45% of ESCs expressing wild-type ESCs with a punctate staining, which is reduced to 5 % for the PIP mutant (Figure 4.2 C and D). We therefore concluded that Zranb3 interacts with PCNA and that PCNA is required for Zranb3-foci formation in ESCs.

While our work was in progress, several manuscripts where published that described the interaction between Zranb3 and PCNA in human cancer cell lines, and its dependence on the PIP box (PCNA interacting protein) domain as well as a ubiquitin binding zinc finger of Zranb3(Ciccia et al., 2012; Weston et al., 2012; Yuan et al., 2012). Notably, these papers demonstrated that the percentage of human cancer cells containing PCNA-overlapping foci of Zranb3 is increased upon induction of DNA damage, an observation typical for proteins involved in DNA repair pathways (Weston et al., 2012). Our findings therefore extend these findings to ESCs.

Zranb3 interacts with the MutSα complex and requires MutSα for localization

Despite the evidence that Zranb3 is involved in DNA repair from the human cancer cell studies (Ciccia et al., 2012; Weston et al., 2012; Yuan et al., 2012), it has remained unclear in which of the multiple repair pathways it may be involved. To better understand the exact role of Zranb3 and determine which repair complexes it might interact with, we subjected nuclear extracts from ESCs expressing Flag-wt and Flag-PIP-mutant Zranb3 to immunoprecipitations with anti-FLAG antibodies and performed multi-dimensional protein identification by mass spectrometry (MudPIT, (Wolters et al., 2001)). The identified interaction partners from three independent mass spec experiments are shown in Figure 4.5 and Supplementary Table 4.1.

As expected, we found PCNA in the interaction list of wild-type Zranb3, which was absent or reduced in the PIP-mutant immunoprecipitates, validating the PCNA-Zranb3 interaction in ESCs (Figure 4.5 A and Supplementary table 4.1). We furthermore identified a number of novel interaction partners that interact with Zranb3 in at least 2 out of 3 experiments (Figure 4.5 A), including the RNA binding protein Pcbp1, the elongation factor Eef1a1, the matrix-associated RNA binding protein Matrin 3, the RNA helicases D1Pas1 and Ddx3x, the pluripotency factor Sall4, the repair factor Rif1 that has also been shown to play a role in the regulation of

pluripotency (Virgilio et al., 2013) (Loh et al., 2006), and Msh6 and Msh2, which form the DNA mismatch repair complex MutSα.

As Zranb3 had been proposed to be involved in DNA repair processes, we were particularly interested in the interaction with the MutSα complex that is a critical component of the DNA mismatch repair pathway. Incidentally, this pathway is especially upregulated in ESC (Tichy et al., 2011). We found that the interaction between Zranb3 and MutSα is PIP box-independent since both Msh6 and Msh2 were found in immunoprecipitations of PIP-mutant Zranb3, with similar abundance to those of wild-type Zranb3. This finding was confirmed by western blot analysis of co-immunoprecipitates (Figure 4.3 A). Since the MutSα complex is involved in DNA repair and required for the recruitment of ubiquitinylated PCNA to repair sites (Zlatanou et al., 2011), with which Zranb3 specifically interacts (Ciccia et al., 2012), we hypothesized that a reduction of MutSα would also lead to less nuclear foci of Zranb3 in immunofluorescence staining. The quantification of the punctate immunofluorescence staining pattern for wild-type Flag-Zranb3 indeed revealed a, on average, 40 % reduction in the percentage of nuclear Zranb3 foci containing cells upon siRNA-mediated knockdown of the MutSα complex (Figure 4.3 B). We therefore conclude that that MutSα is required for Zranb3 localization.

Zranb3 is involved in mismatch repair

The knockdown of proteins involved in the DNA mismatch repair pathway results in a higher sensitivity of cells to the damage-inducing agents camptothecin and cisplatin, and resistance to temezolomide (Liu et al., 1996). It had previously been demonstrated that the knockdown of Zranb3 in human cancer cells confers a higher sensitivity to camptothecin and hydroxyurea (Ciccia et al., 2012). Those compounds are known to induce DNA damage by stalling replication forks but are not specific to mismatch repair. As the resistance to temezolomide, which is an alkylating agent, is more specific to MMR, we tested the sensitivity of ESCs to the previously

described compounds as well as temezolomide, upon treatment with siRNA against Zranb3 and control siRNA. We found that ESCs treated with siZranb3 have a higher sensitivity to camptothecin (Figure 4.4 A) and hydroxyurea (Supplementary Figure 4.3 A), and to a lesser extent also cisplatin that also stalls replication forks (Supplementary Figure 4.3 B), in accordance with the finding in human cancer cells (Ciccia et al., 2012; Weston et al., 2012; Yuan et al., 2012). However, in addition, we uncovered that siZranb3-treated ESCs have a higher resistance to temezolomide than control cells (Figure 4.4 B).

Together, these findings indicate an involvement of Zranb3 in the DNA mismatch repair process and a functional role of the interaction with MutS$\alpha$ and Zranb3.

**Figure 4.1: Zranb3 is highly expressed mouse embryonic stem cells.**

A) Result of the Flag-PCNA MudPIT experiment. Values are NSAF values that were normalized to the bait (PCNA). B) Zranb3 expression levels were measured on the protein level by western blot with antibody specific to Zranb3 in embryonic stem cells (ESC) or differentiatied cells (mouse embryonic fibroblasts, MEF). C) qRT-PCR comparison of expression levels of Zranb3 transcript levels in ESC and MEF.

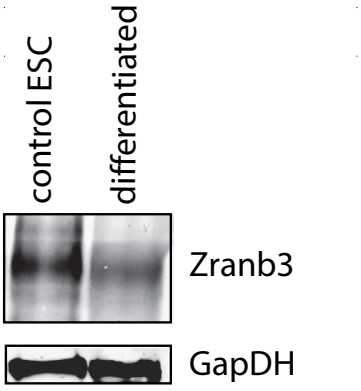
**Figure 4.1: Zranb3 is highly expressed mouse embryonic stem cells.**

A

Flag-PCNA mass spectromety

| interactor | normal. NSAF |
|---|---|
| PCNA | 100 |
| Zranb3 | 0.456 |

B




C

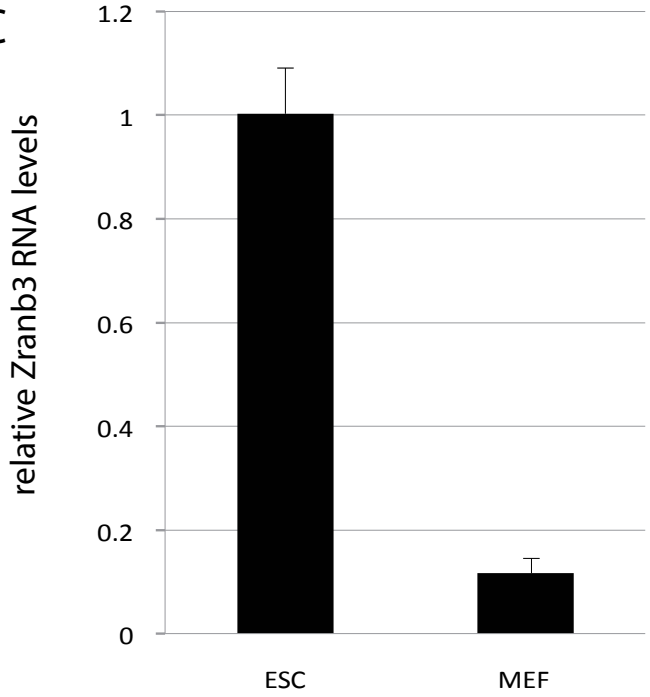


**Figure 4.2: Zranb3 directly interacts with PCNA in ESC.**

A) Schematic of the domain structure of Zranb3 and the conserved PIP box. The mutation that was introduced for the PIP box mutant is highlighted in red. B) Westernblot of Flag-Zranb3 (wt or PIP mutant) Immunoprecipitations from nuclear extracts. Sup = supernatant, IP = immunoprecipitates. C) Immunofluorescence pictures of Flag-Zranb3 (wt or PIP mutant) (green) and endogenous PCNA (red). Nuclei are stained with DAPI (blue). The punctuate pattern of Flag-wt Zranb3 is shown in multiple examples with different punctuation pattern. Flag-PIP mutant Zranb3 does rarely show the punctuate pattern. D) Quantification of cells with punctuate staining patterns shown as a percentage of about 200 nuclei that were counted for each sample.

**Figure 4.2: Zranb3 directly interacts with PCNA in ESC.**

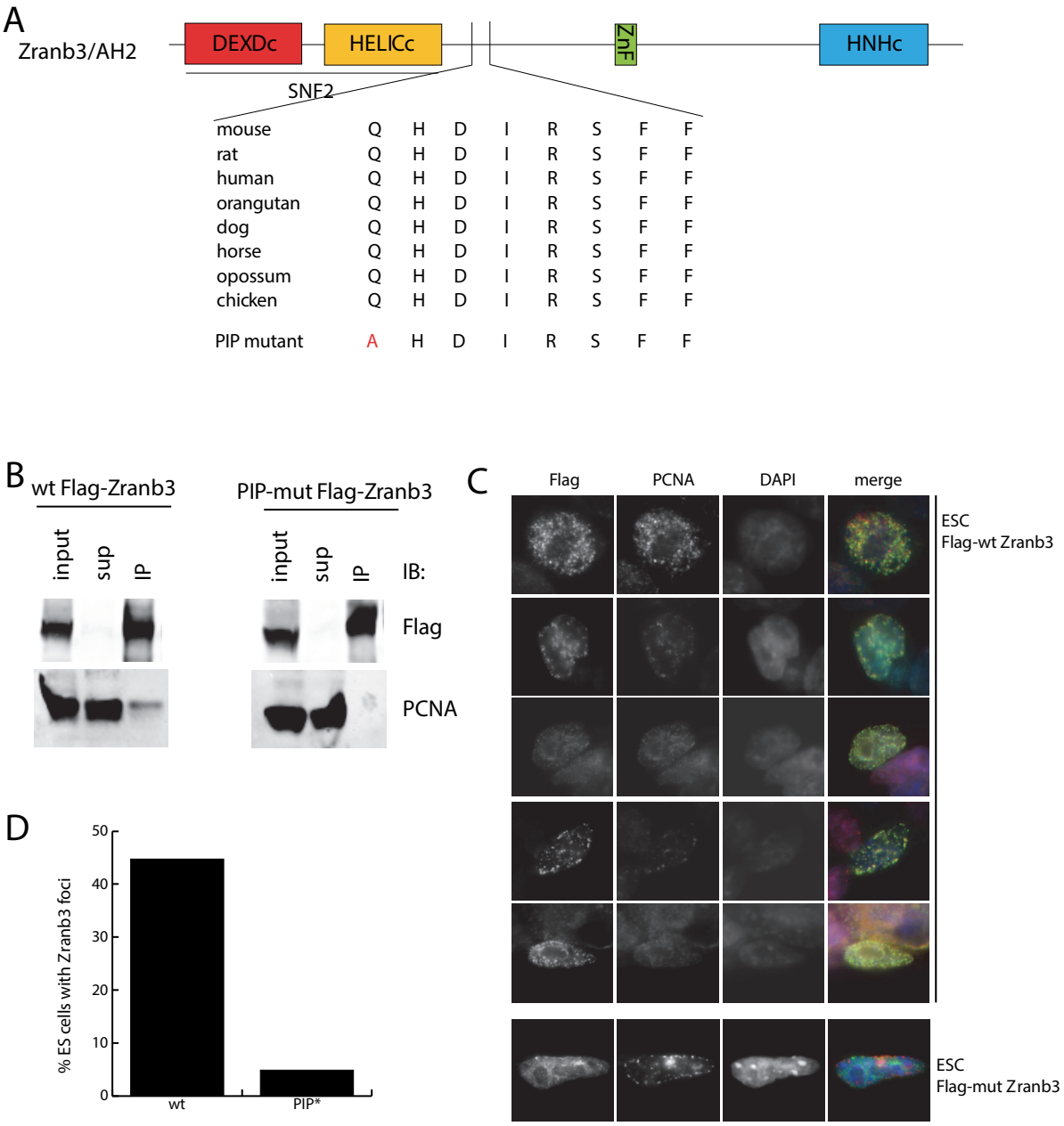


**Figure 4.3: wt and PIP mutant Zranb3 interact with the MutSα complex which is required for the punctuate staining pattern.**

A) Westernblot of immunoprecipitates from cells expressing Flag-Zranb3 (wt or PIP mutant Zranb3 and control cells. INP = input, SUP = supernatant, IP = Immunoprecipitation. B) Quantification of the Flag-wt Zranb3 punctuate staining pattern after knockdown of both MutSα subunits Msh2 and Msh6 (siMutS). Results of two independent experiments are shown a percentage of about 200 nuclei that were counted for each sample.

**Figure 4.3: wt and PIP mutant Zranb3 interact with the MutSα complex which is required for the punctuate staining pattern.**

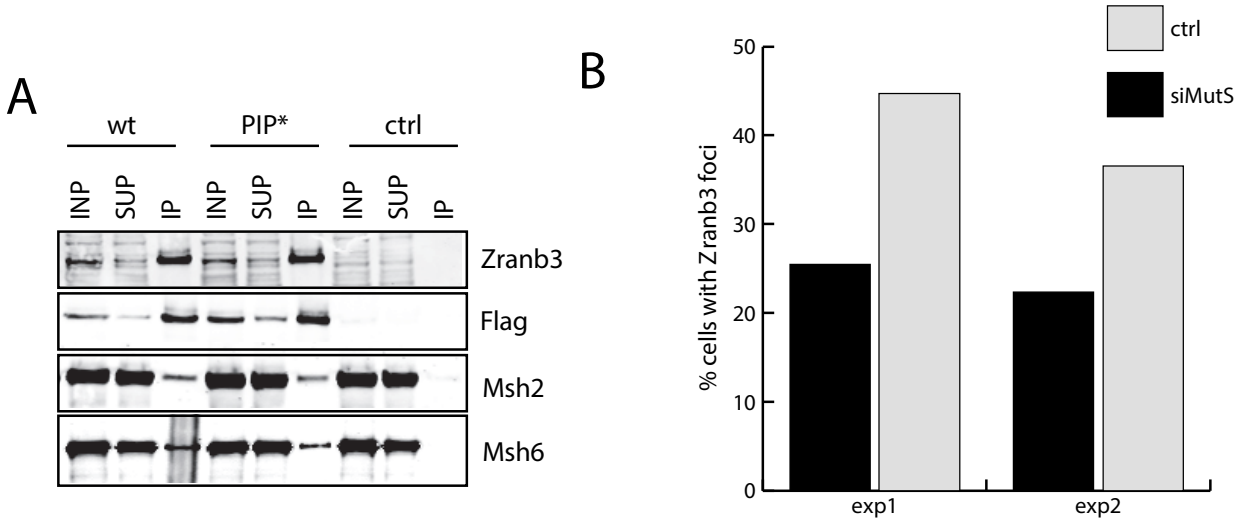


**Figure 4.4: Zranb3 plays a role in the DNA mismatch repair pathway.**

A) Relative cell density was measured in cytotoxicity assays with a range of concentrations of Camptothecin and B) Temezolomide upon treatment of ESC with siRNA targeted to Zranb3 or a non-targeting control siRNA (ctrl). C) Model of how the DNA mismatch repair pathway is integrated with Zranb3 and ubiquitinylated PCNA. Upon DNA damage (star), the MutSα complex comprised of Msh2 and Msh6 binds the lesion. This recruits ubiquitinylated PCNA (ub-PCNA) to the site of damage. MutSα and ub-PCNA are then bound by Zranb3.

**Figure 4.4: Zranb3 plays a role in the DNA mismatch repair pathway.**

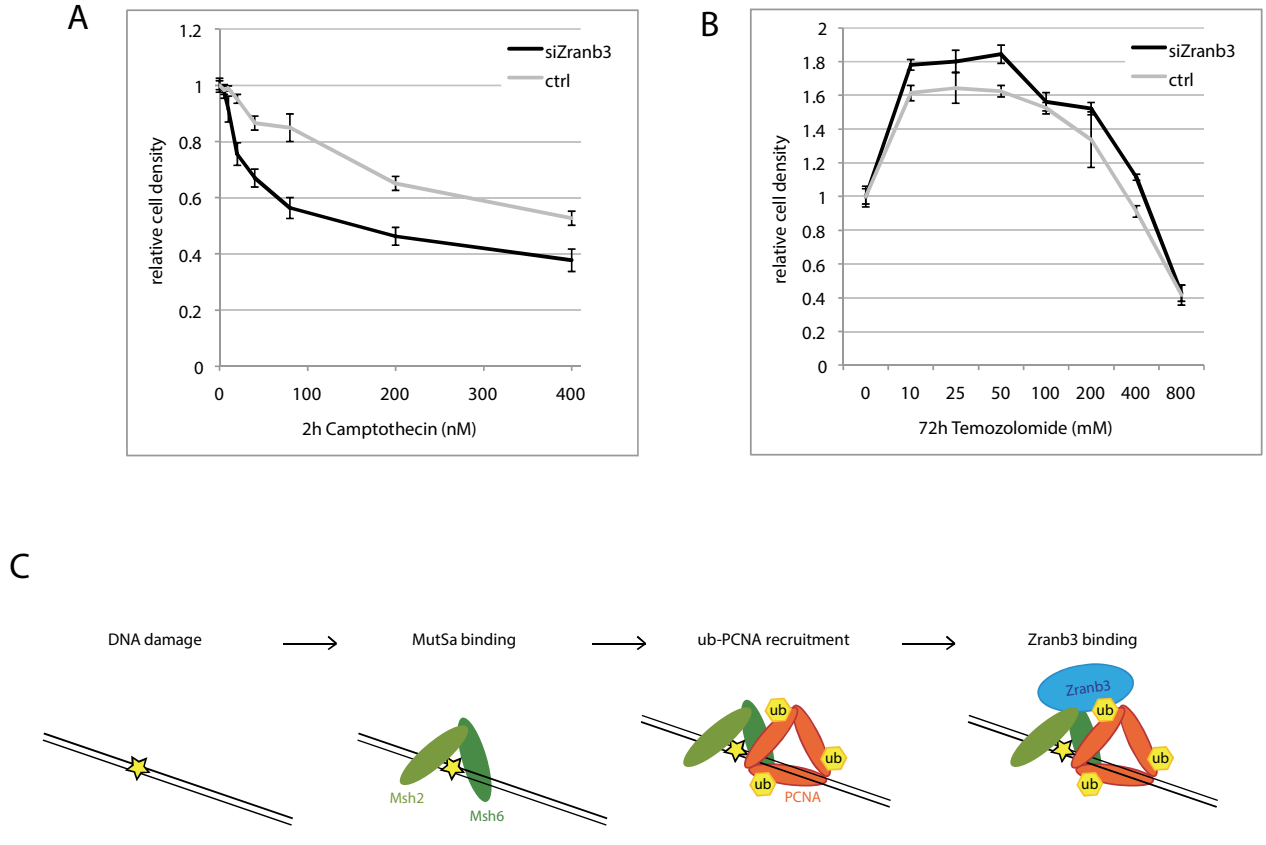
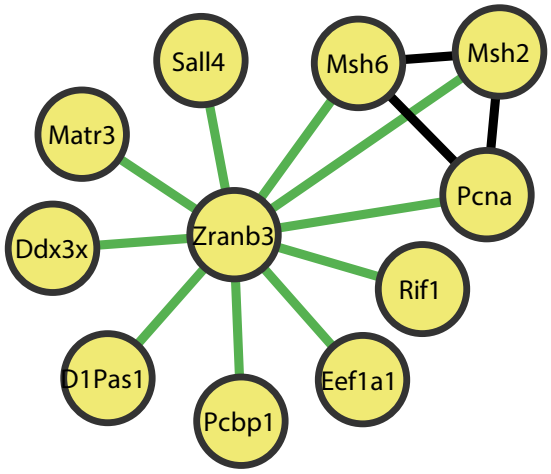


**Figure 4.5: Zranb3 interacts with PCNA and the MutSα complex.**

Top: Excerpt of the MudPIT results of Flag-Zranb3 immunoprecipitates (wt or PIP mutant – PIP*) were normalized for the bait (Zranb3) of 3 independent experiments. All interaction partners that were not found in at least two out of three experiments are excluded from this list. For a full list of MudPIT results see Supplementary Table. Bottom: visualization of known and newly identified interactions as extracted from STRING. Known interactions: black edge. New interactions: black edge.

**Figure 4.5: Zranb3 interacts with PCNA and the MutSα complex.**

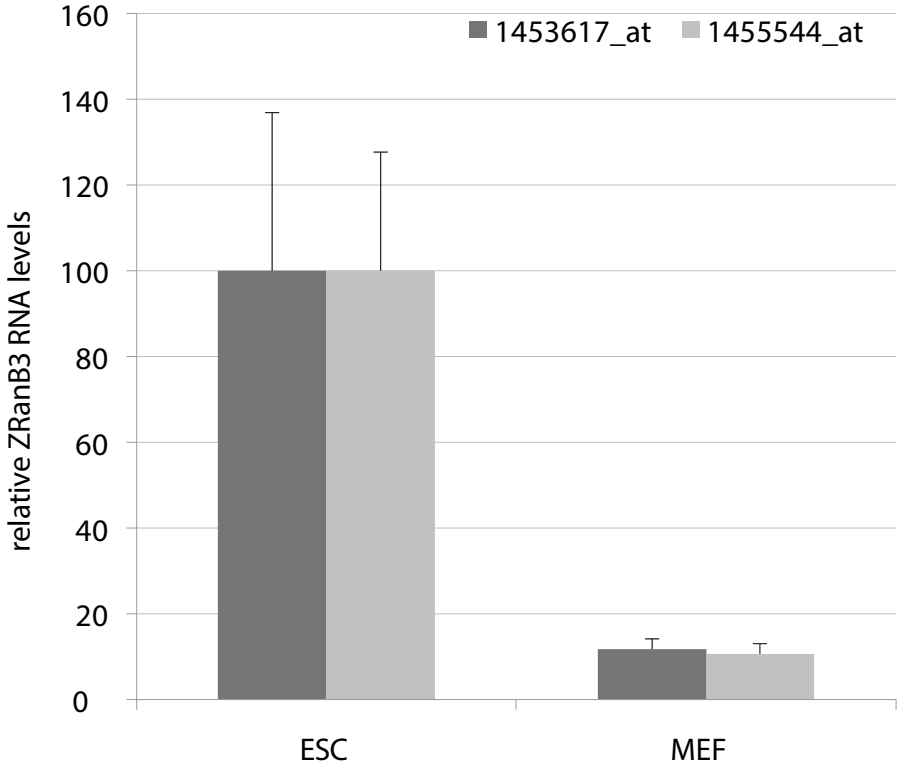
| LocusID | Name | selected GO categories | WT exp1 | WT exp2 | PIP* exp2 | WT exp3 | PIP* exp3 |
|---|---|---|---|---|---|---|---|
| Q6NZP1 | **Zranb3** | DNA repair, replication fork protection | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| P17918 | **PCNA** | DNA replication, DNA repair, cell proliferation | 42.16 | 89.68 | 0.00 | 32.51 | 19.69 |
| P60335 | **Pcbp1** | RNA splicing | 0.00 | 2.18 | 2.48 | 28.60 | 21.65 |
| P54276 | **Msh6** | DNA repair, mismatch repair | 0.00 | 1.20 | 0.47 | 19.99 | 12.87 |
| P43247 | **Msh2** | DNA repair, mismatch repair | 13.45 | 2.24 | 0.51 | 16.33 | 14.29 |
| Q8BX22 | **Sall4** | stem cell maintenance | 0.00 | 6.55 | 3.13 | 7.95 | 5.78 |
| P10126 | **Eef1a1** | translational elongation | 187.15 | 7.56 | 8.26 | 7.35 | 3.34 |
| Q8K310 | **Matr3** | nuclear matrix, RNA binding | 0.00 | 0.83 | 1.50 | 6.02 | 11.54 |
| P16381 | **D1Pas1** | cell differentiation | 0.00 | 0.00 | 1.68 | 5.14 | 10.89 |
| Q62167 | **Ddx3x** | intrinsic apoptotic signaling pathway, RNA secondary structure unwinding | 0.00 | 0.00 | 1.68 | 5.14 | 10.89 |
| Q6PR54 | **Rif1** | response to DNA damage stimulus, cell cycle, stem cell maintenance | 3.90 | 0.82 | 0.20 | 4.91 | 3.82 |



**Supplementary Figure 4.1: Zranb3 is developmentally regulated.**

A) Zranb3 transcript levels in ESC and MEF as measured in by RNA sequencing experiments. B) qRT-PCR of the pluripotency markers Nanog and Oct4 during the differentiation of Zhbtc4 cells in the experiment shown in Figure 1 C.

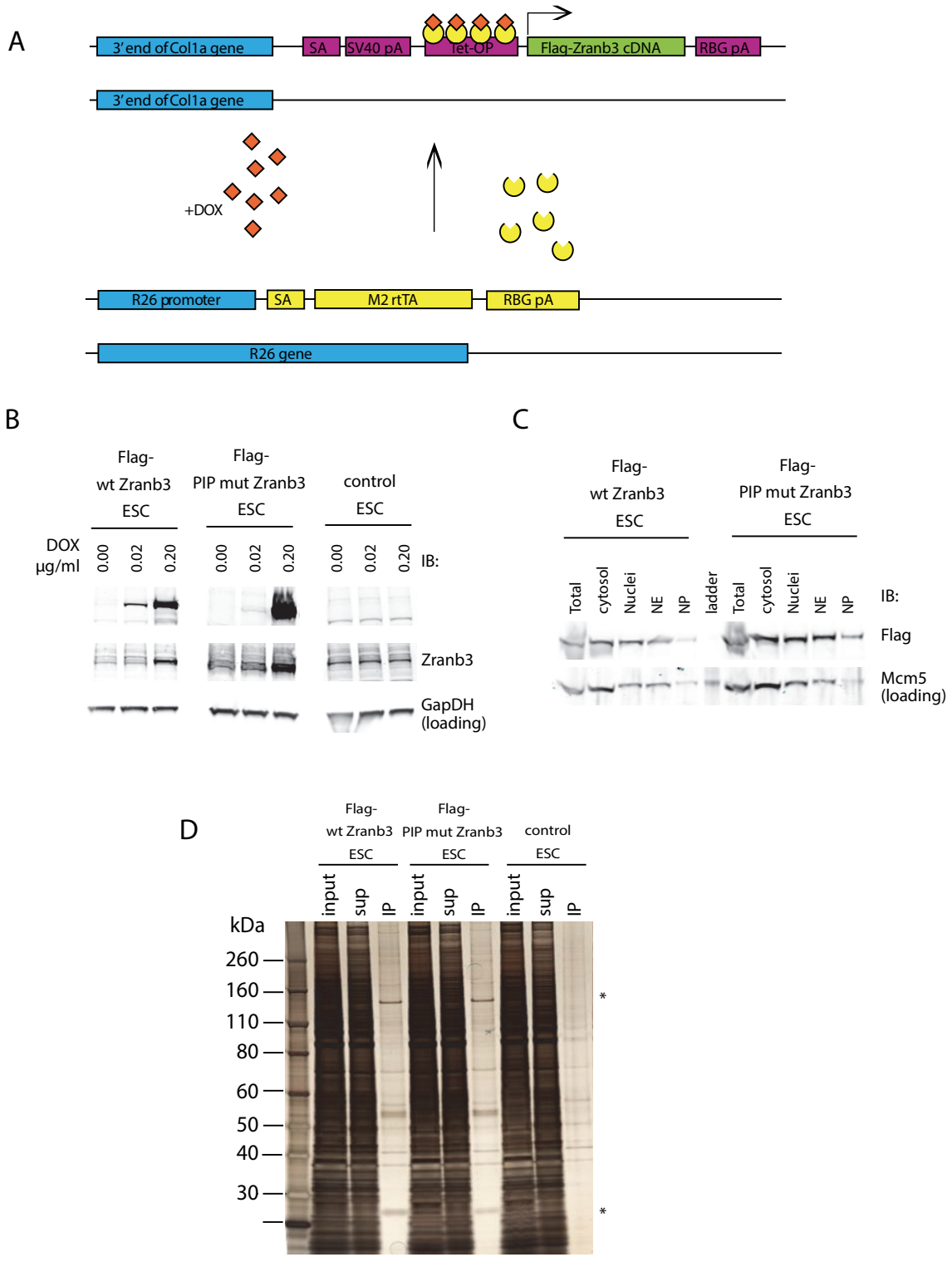
**Supplementary Figure 4.1: Zranb3 is developmentally regulated.**



**Supplementary Figure 4.2: Generation of cell lines that inducibly express Flag-tagged constructs, nuclear extraction and PCNA Immunoprecipitation.**

A) Schematic of the targeted locus. Flag-tagged constructs (such as Flag-Zranb3, shown here) are integrated under the control of the Tet-operon (Tet-OP) at the 3' end of the Collagen1a (Col1a) locus. Upon addition of doxicycline (DOX), rtTA that is expressed under control of the ROSA26 (R26) promoter binds to the Tet-operon and induces expression of the integrated construct. SA = splice acceptor, SV40 pA = SV40 polyadenylation signal, RBG pA = RBG polyadenylation signal, M2 rtTA = M2 tetracycline reverse transcriptional activator. B) Induction of Flag-Zranb3 by various doxicycline (DOX) concentrations as shown in western blots from whole cell extracts. IB = immunoblot. C) Western blot confirmation of successful extraction of Flag- wt or PIP-mutant Zranb3 into the nuclear extract fraction. D) Validation of successful Flag-Zranb3 IP. NE of Flag-Zranb3 (wt or PIP mutant) expressing cells and control cells (ctrl) were immunoprecipitated for Flag and successful Immunoprecipitation validated by silver stain. Bands with the size of Zranb3 and PCNA are highlighted with an asterix. sup = supernatant, IP = immunoprecipitates.

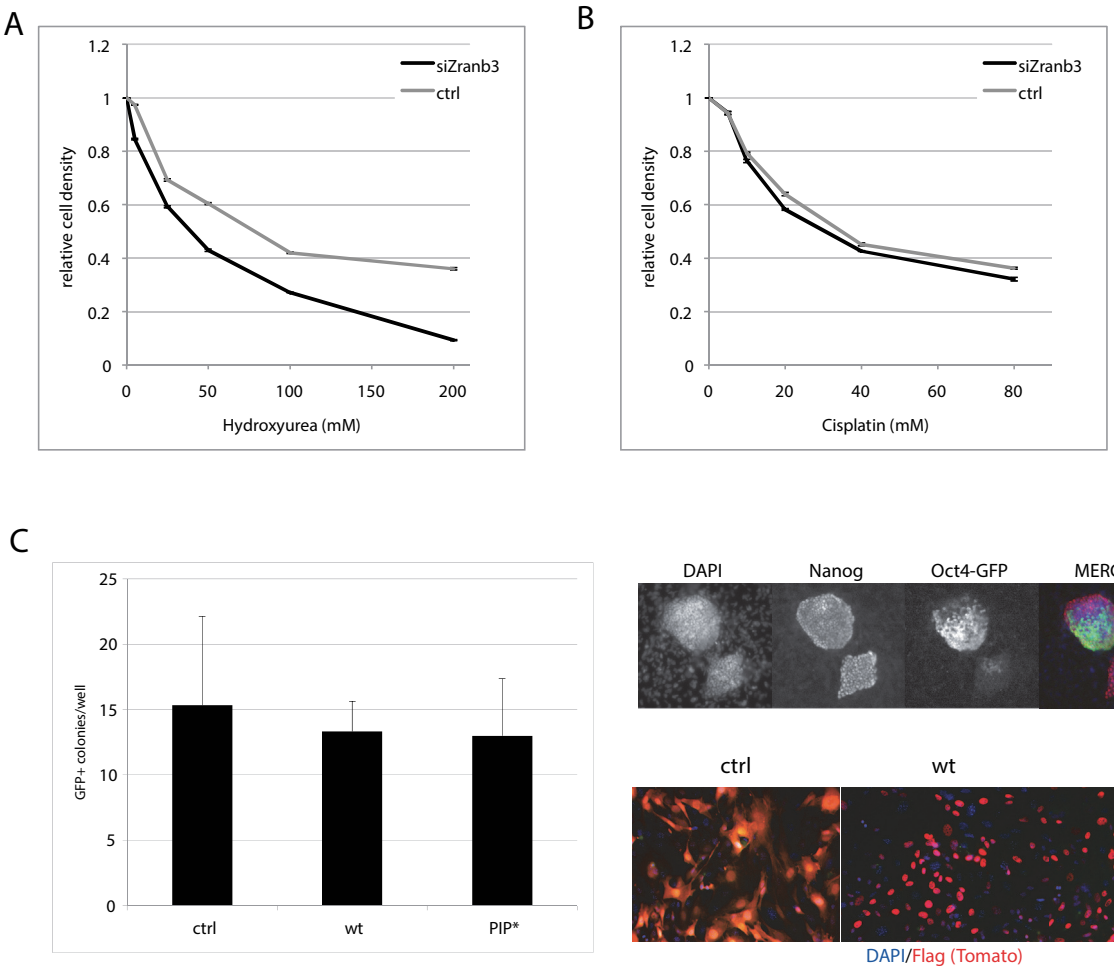
**Supplementary Figure 4.2: Generation of cell lines that inducibly express Flag-tagged constructs, nuclear extraction and PCNA Immunoprecipitation.**



**Supplementary Figure 4.3: Zranb3 confers sensitivity to Hydroxyurea but not to Cisplatin and does not increase nuclear reprogramming efficiency and is not required for the maintenance of pluripotency.**

A) Relative cell density was measured in cytotoxicity assays with a range of concentrations of Hydroxyurea and B) Cisplatin upon treatment of ESC with siRNA targeted to Zranb3 or a non-targeting control siRNA (ctrl). C) Nuclear reprogramming experiment with MEF cells containing a GFP reporter under the endogenous Oct4 promoter which are inducibly expressing the pluripotency factors Oct4, Sox2, Nanog and Klf4 (Stadtfeld et al., 2010) that were infected with virus expressing a tomato control construct (ctrl), wildtype (wt) or PIP-mutant (PIP*) Flag-Zranb3. Left: The efficiency of reprogramming of reprogramming was quantified by counting the number of GFP positive induced pluripotent cell (iPSC) colonies obtained in each condition. Right top: an example of a colony of iPSCs resulting from the reprogramming that was stained for Nanog (red), Oct4 (green) and nuclei (DAPI, blue) by immunofluorescence (IF). Right bottom: successful transfection was confirmed by Flag-IF (red, also shows tomato) on day five of reprogramming. D) qRT-PCR result of several pluripotency and germline markers upon knockdown of Zranb3 by shRNA (shZranb3, top panel) or siRNA (siZranb3, bottom panel) for 48 hours in ESC compared to a knockdown with a non-targeting control shRNA or siRNA (ctrl). E) Alkaline phosphatase staining of cells that were treated with siRNA targeted to Zranb3 or a non-targeting control siRNA (ctrl) and the corresponding validation of knockdown by qRT-PCR.

**Supplementary Figure 4.3: Zranb3 confers sensitivity to Hydroxyurea but not to Cisplatin and does not increase nuclear reprogramming efficiency and is not required for the maintenance of pluripotency.**



**Supplementary Table 4.1: Complete MudPIT result of Flag-Zranb3 (wt or PIP mutant) immunoprecipitations.**

MudPIT results of Flag-Zranb3 immunoprecipitates (wt or PIP mutant – PIP*) were normalized for the bait (Zranb3) of 3 independent experiments.

| | | NSAF scores | | | | | normalized scores (to bait) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LocusID | name | WT exp3 | PIP* exp3 | WT exp2 | PIP* exp2 | WT exp1 | WT exp3 | PIP* exp3 | WT exp2 | PIP* exp2 | WT exp1 |
| **Q6NZP1** | Zranb3 | 1973.65 | 2000.62 | 6523.94 | 5610.31 | 745.14 | 100 | 100 | 100 | 100 | 100 |
| **P17918** | Pcna | 641.56 | 393.95 | 5850.61 | 0 | 314.17 | 32.51 | 19.69 | 89.68 | 0 | 42.16 |
| **P60335** | Pcbp1 | 564.43 | 433.23 | 142.27 | 139.27 | 0 | 28.6 | 21.65 | 2.18 | 2.48 | 0 |
| **P54276** | Msh6 | 394.57 | 257.43 | 78.32 | 26.29 | 0 | 19.99 | 12.87 | 1.2 | 0.47 | 0 |
| **P43247** | Msh2 | 322.36 | 285.92 | 146.25 | 28.64 | 100.23 | 16.33 | 14.29 | 2.24 | 0.51 | 13.45 |
| **Q8BX22** | Sall4 | 156.93 | 115.64 | 427.2 | 175.65 | 0 | 7.95 | 5.78 | 6.55 | 3.13 | 0 |
| **P10126** | Eef1a1 | 144.98 | 66.77 | 493.32 | 463.62 | 1394.53 | 7.35 | 3.34 | 7.56 | 8.26 | 187.15 |
| **Q8K310** | Matr3 | 118.76 | 230.92 | 53.88 | 84.39 | 0 | 6.02 | 11.54 | 0.83 | 1.5 | 0 |
| **P16381** | D1Pas1 | 101.39 | 217.91 | 0 | 94.51 | 0 | 5.14 | 10.89 | 0 | 1.68 | 0 |
| **Q62167** | Ddx3x | 101.39 | 217.91 | 0 | 94.51 | 0 | 5.14 | 10.89 | 0 | 1.68 | 0 |
| **Q6PR54** | Rif1 | 96.91 | 76.51 | 53.39 | 11.07 | 29.06 | 4.91 | 3.82 | 0.82 | 0.2 | 3.9 |
| **Q7TSF1** | Dsg1b | 94.92 | 38.86 | 21.53 | 0 | 0 | 4.81 | 1.94 | 0.33 | 0 | 0 |
| **Q61495** | Dsg1a | 94.92 | 38.86 | 21.53 | 0 | 0 | 4.81 | 1.94 | 0.33 | 0 | 0 |
| **Q99PV0** | Prpf8 | 43.03 | 35.23 | 9.76 | 0 | 0 | 2.18 | 1.76 | 0.15 | 0 | 0 |
| **P62806** | Hist1h4a | 0 | 199.65 | 811.34 | 0 | 0 | 0 | 9.98 | 12.44 | 0 | 0 |
| **Q9D0M5** | Dynll2 | 0 | 346.58 | 597.52 | 902.5 | 0 | 0 | 17.32 | 9.16 | 16.09 | 0 |
| **Q60972** | Rbbp4 | 0 | 266.12 | 341.4 | 353.88 | 340.62 | 0 | 13.3 | 5.23 | 6.31 | 45.71 |
| **O09106** | Hdac1 | 0 | 85.33 | 235.4 | 165.92 | 193.58 | 0 | 4.27 | 3.61 | 2.96 | 25.98 |
| **Q60973** | Rbbp7 | 0 | 217.74 | 232.38 | 230.99 | 0 | 0 | 10.88 | 3.56 | 4.12 | 0 |
| **P97496** | Smarcc1 | 0 | 37.25 | 213.32 | 177.85 | 106.11 | 0 | 1.86 | 3.27 | 3.17 | 14.24 |
| **P59708** | Sf3b14 | 0 | 164.51 | 182.33 | 0 | 0 | 0 | 8.22 | 2.79 | 0 | 0 |
| **Q9Z2N8** | Actl6a | 0 | 119.84 | 159.38 | 166.43 | 245.75 | 0 | 5.99 | 2.44 | 2.97 | 32.98 |
| **Q921M3** | Sf3b3 | 0 | 59.14 | 124.85 | 132 | 0 | 0 | 2.96 | 1.91 | 2.35 | 0 |
| **Q6ZQ88** | Kdm1a | 0 | 36.16 | 106.88 | 52.31 | 0 | 0 | 1.81 | 1.64 | 0.93 | 0 |
| **Q8BTI8** | Srrm2 | 0 | 11.41 | 84.32 | 39.62 | 0 | 0 | 0.57 | 1.29 | 0.71 | 0 |
| **Q61990** | Pcbp2 | 0 | 426.05 | 83.95 | 0 | 0 | 0 | 21.3 | 1.29 | 0 | 0 |
| **Q9R190** | Mta2 | 0 | 92.35 | 79.61 | 53.44 | 0 | 0 | 4.62 | 1.22 | 0.95 | 0 |
| **Q6PDQ2** | Chd4 | 0 | 37.58 | 75.38 | 27.96 | 0 | 0 | 1.88 | 1.16 | 0.5 | 0 |
| **Q921F2** | Tardbp | 0 | 124.18 | 73.4 | 43.11 | 0 | 0 | 6.21 | 1.13 | 0.77 | 0 |
| **Q8VH51** | Rbm39 | 0 | 155.2 | 71.67 | 117.87 | 0 | 0 | 7.76 | 1.1 | 2.1 | 0 |
| **P13020** | Gsn | 0 | 65.91 | 68.18 | 0 | 0 | 0 | 3.29 | 1.05 | 0 | 0 |

| LocusID | name | NSAF scores | | | | | normalized scores (to bait) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | WT exp3 | PIP* exp3 | WT exp2 | PIP* exp2 | WT exp1 | WT exp3 | PIP* exp3 | WT exp2 | PIP* exp2 | WT exp1 |
| **Q9Z1X4** | Ilf3 | 0 | 114.5 | 67.68 | 29.82 | 0 | 0 | 5.72 | 1.04 | 0.53 | 0 |
| **Q8CCF0** | Prpf31 | 0 | 41.21 | 60.9 | 0 | 0 | 0 | 2.06 | 0.93 | 0 | 0 |
| **Q9CXY6** | Ilf2 | 0 | 79.09 | 58.44 | 0 | 0 | 0 | 3.95 | 0.9 | 0 | 0 |
| **Q3TKT4** | Smarca4 | 0 | 19.12 | 56.52 | 71.93 | 29.05 | 0 | 0.96 | 0.87 | 1.28 | 3.9 |
| **O08810** | Eftud2 | 0 | 42.36 | 46.94 | 45.96 | 0 | 0 | 2.12 | 0.72 | 0.82 | 0 |
| **Q99NB9** | Sf3b1 | 0 | 23.65 | 46.61 | 47.91 | 0 | 0 | 1.18 | 0.71 | 0.85 | 0 |
| **Q9ER74** | Sall1 | 0 | 23.33 | 45.97 | 0 | 0 | 0 | 1.17 | 0.7 | 0 | 0 |
| **Q99K85** | Psat1 | 0 | 222.31 | 41.07 | 0 | 0 | 0 | 11.11 | 0.63 | 0 | 0 |
| **Q99KG3** | Rbm10 | 0 | 33.17 | 40.84 | 0 | 0 | 0 | 1.66 | 0.63 | 0 | 0 |
| **Q3U1J4** | Ddb1 | 0 | 18.04 | 39.98 | 0 | 0 | 0 | 0.9 | 0.61 | 0 | 0 |
| **Q924K8** | Mta3 | 0 | 104.39 | 39.47 | 26.49 | 34.78 | 0 | 5.22 | 0.6 | 0.47 | 4.67 |
| **Q8K4B0** | Mta1 | 0 | 57.52 | 39.47 | 26.49 | 34.78 | 0 | 2.88 | 0.6 | 0.47 | 4.67 |
| **Q80VC9** | Kiaa1543 | 0 | 16.42 | 36.41 | 57.03 | 0 | 0 | 0.82 | 0.56 | 1.02 | 0 |
| **A2AGT5** | Ckap5 | 0 | 10.12 | 18.69 | 0 | 0 | 0 | 0.51 | 0.29 | 0 | 0 |
| **Q6NV83** | Sr140 | 0 | 19.98 | 14.77 | 0 | 0 | 0 | 1 | 0.23 | 0 | 0 |
| **P70288** | Hdac2 | | | 235.4 | 165.92 | 193.58 | 0 | 0 | 3.61 | 2.96 | 25.98 |

## 4.4 Discussion

Zranb3 is an annealing helicase (Yusufzai and Kadonaga, 2010) with helicase properties similar to HARP (Bétous et al., 2013), that has recently been implied to function in DNA repair pathways in human cancer cells cells (Ciccia et al., 2012; Weston et al., 2012). However, the repair pathway that Zranb3 acts in had not been defined. Here, we demonstrate that Zranb3 is highly expressed in mouse ESCs and that it's expression is developmentally regulated. We extend previous findings from human cells that Zranb3 interacts directly with PCNA to mouse ESCs, and provide evidence that Zranb3 is involved in the DNA mismatch repair pathway, which leads us to propose a model of how the MutSα complex recruits Zranb3 to sites of DNA damage.

The high expression of Zranb3 in ESCs (Figure 4.1) could imply a particular importance of the protein in stem cells. Interestingly, a number of repair proteins have been shown to be required for the establishment and maintenance of pluripotency (Fong et al., 2011) (Stambrook and Tichy, 2010), and, importantly, the mismatch repair pathway is highly active in mouse ESC (Tichy et al., 2011). In fact, the MutSα complex is expressed about tenfold higher in ESCs than MEFs, which is about the same ratio that we observe for Zranb3. Interestingly, despite the high expression level of Zranb3 in ESCs and the indication that repair processes might be required for pluripotency (Fong et al., 2011), we did not find a role for Zranb3 in the maintenance or establishment of pluripotency as the depletion of Zranb3 did neither alter the ability of ESC to maintain the pluripotent state nor the the efficacy of transcription factor induced reprogramming differentiated cells to induced pluripotent stem cells (Supplementary Figures 4.3 C and D). We can not exclude that the lack of an effect on maintenance or establishment of pluripotency is due to the incomplete knockdown of Zranb3 by siRNA. A knockout model would be required to further test Zranb3's role in pluripotency. Aside from such technical considerations, our results could imply that there are either compensatory mechanisms that mask the importance of this

factor in the establishment and maintenance of pluripotency or that not all repair pathways have the same importance for these processes.

We observed a punctuate nuclear staining pattern of Zranb3 that overlaps with PCNA in mouse ESCs, and showed that this pattern is dependent on the PIP box and thus dependent on the interaction of Zranb3 with PCNA (Figure 4.2). Immunoprecipitation experiments both with Flag-PCNA and Flag-Zranb3 confirmed the interaction in ESCs.

The MudPIT result of Flag-Zranb3 interaction partners revealed the MutSα complex as a novel interaction partner and thus provides evidence for the specific pathway that Zranb3 might be involved in. This demonstrates the usefulness of comprehensive approaches for protein characterizations that should be more broadly applied to other studies. Since this interaction had not been observed previously (Yuan et al., 2012) this might mean that the more comprehensive approach is more useful in identifying novel interaction partners or that this interaction is in fact cell type -pecific. Given the fact that MutSα is also more highly expressed in ESC than in differentiated cells, this seems likely but will need to be tested in future experiments.

We confirmed previous findings that Zranb3 confers resistance to camptothecin (Figure 4.4 A) (Yuan et al., 2012), demonstrating its requirement for the repair of damage incurred during DNA replication. An involvement of Zranb3 in the mismatch repair pathway is likely as knockdown of Zranb3 lead to lower sensitivity to the chemical Temezolomide (Figure 4.4 B) since cells deficient in mismatch repair tend to have a higher resistance to this drug (Liu et al., 1996). Furthermore, when the interaction with MutSα is interrupted by siRNA mediated knockdown (Figure 4.3), Zranb3 loses it's typical punctuate staining pattern.This punctuate staining pattern is reminiscent for other proteins associated with DNA repair, such as phoshprylated histone 2 A (γH2AX). γH2AX is widely used as a marker for cells undergoing DNA repair in differentiated cells. In such assays, the percentage of cells showing a punctuate γH2AX staining pattern is

usually quantified as a measure of DNA repair activity (Kuo and Yang, 2008). ESC in contrast to more differentiated cells tend to have a much higher percentage staining positive for this marker even without the induction of DNA damage, which is making a quantification of DNA repair activity in this cell type very technically challenging.but is also confirming that genome surveillance and repair processes are much more active in stem cells (Turinetto et al.). The high percentage of ESCs showing the punctuate Zranb3 staining pattern without further stimulation of repair is thus in line with previous observations for the DNA damage marker γH2AX in murine embryonic stem cells and provides further evidence that Zranb3 is likely a part of the cellular DNA repair machinery.

Since MutSα is required for recruitment of ubiquitinylated PCNA to sites of DNA mismatch repair (Zlatanou et al., 2011) and Zranb3 binds to ub-PCNA through both it's PIP box and it's zinc finger domain (our study and Ciccia et al.), we propose a mechanism (Figure 4.4 C) in which Msh2 and Msh6 bind to DNA lesions which leads to the recruitment of ubiquitinylated PCNA to the site of damage. Zranb3 is then recruited through its interaction with both proteins to exert its function in the repair process.

## 4.5 Material and Methods

Cell culture, cell extraction and western blot

V6.5 mouse ESC were maintained on a feeder cell layer in ESC medium containing leukemia inhibitory factor (LIF), 15% fetal bovine serum, nonessential amino acids, glutamate and penicillin/streptomycin in knockout DMEM medium on gelatinized cell culture dishes. For differentiation, Zhbtc4 cells (Niwa et al., 2000) were induced with 2ug/ml doxicycline in LIF-free ESC medium for up to 8 days on gelatinized cell culture dishes without feeders.

To generate whole cell extracts, cells were trypsinized and resuspended in lysis buffer (300mM Tris-HCl pH 6.8, 10 % SDS, 15 % glycerol, 1 mM DTT). The proteins were separated by gel electrophoresis on 4-12 % Acrylamide gels at 140 V at room temperature for 60 mins. For Western blot, the proteins were transferred to nitrocellulose membranes. The membranes were blocked with Odyssey blocking buffer (LiCOR) for one hour at room temperature and incubated with primary antibodies as indicated over night at 4 C. After washing with PBS 0.1% Tween three times, membranes were incubated with secondary antibodies (1:10000, LiCOR) for one hour at room temperature. Protein bands were visualized using the LiCOR infrared scanner and analyzed with the corresponding software.

For silver stains, the membranes were fixed and stained using the SilverQuest Staining Kit (invitrogen) according to the manufacturer's instructions.

siRNA transfection, RNA isolation and qRT-PCR

V6.5 cells were plated at 10^5 cells per well in 6-well plates on gelatine without feeders and reverse transfected with 20 or 50 nm final siRNA (Dharmacon) concentration using invitrogen lipofectamine RNAimax according to manufacturer's instructions. After 24 hours, the medium was exchanged and forward transfected. The cells were trypsinized 48 hours after plating

(unless stated otherwise) and resuspended in 500 ul Trizol (invitrogen). RNA was isolated with Qiagen RNAeasy kits according to manufacturer's instructions. Superscript III Supermix (invitrogen) was used to generate random hexamer primed cDNA. The cDNA was diluted 1:5 with water and used in qRT-PCR with Roche Lightcycler reaction mix.

Imunoprecipitation and mass spectrometry

3xFlag-6xHis-tagged proteins were targeted to the Collagen locus under control of the Tet operon of mouse ESC by FRT/FLPase mediated recombination as previously described (Beard et al., 2006). Cells were treated with 2 ug/ml doxicycline in ESC medium for 48 hours on gelatinized culture plates without feeders before harvesting. Nuclear extracts were generated by dignam extraction with 420mM salt as previously described (Carey et al., 2009). The extracts were dialized to 100 mM salt concentration and immunoprecipitation performed with antiFlag M2 agarose beads (invitrogen) for 3 hours at 4 C. Complexes were eluted with 25 ug/ml flag peptide in 4 sequential elutions for 15 mins at room temperature. Each eluate was verified by western blot and silver stain before they were pooled. Pooled eluates were subjected to MudPIT mass spectrometry as previously described (Wolters et al., 2001). Hits that were also identified in a control immunoprecipitation were excluded from further analysis.

Immunofluorescence

Glass coverslips were washed with 70% Ethanol and coated with porcupine gelatine. Cells were plated directly on the coverslips and induced with 2 uM doxicycline for 24 to 72 hours. Cells were then extracted on ice with CSK buffer (100 mM PIPES pH6.8, 300 mM sucrose, 3 mM $MgCl_2$, 100 mM NaCl), CSK-0.5% Triton and CSK for 5 minutes each before they were fixed with 4% paraformaldehyde. The samples were incubated in blocking buffer (5% normal goat serum, 0.2 % Tween20, 2 % fish skin gelatine in PBS) for 30 minutes at room temperature, in primary antibody dilution (mouse a Flag, rab a PCNA, 1:200 each) at 4C over night, washed 3 x 5

minutes in PBS 0.2% tween20, incubated in secondary antibody (alexafluor 488 anti mus, alexafluor 546 anti rabbit, 1:2000 each) for 1 hour at room temperature, washed 3x 5 mins in PBS 0.2% tween20 in which the middle step was supplemented with DAPI for visualization of DNA.

Reprogramming to the pluripotent state

Mouse embryonic fibroblasts (MEF) that can be induced to express Oct4, Sox2, Nanog and Klf4 upon doxicycline induction (Stadtfeld et al., 2010) plated on cell culture plates MEF medium containing 10 % fetal bovine serum, nonessential amino acids, glutamate and penicillin/streptomycin in knockout DMEM medium. One day after plating, cells were incubated over night with a 500 ul supernatant from cells expressing lentivirus to overexpress Flag-tagged wt or PIP mutant Zranb3 or a control construct (tomato), 1 ul polybreme reagent and 500 ul ESC medium containing leukemia inhibitory factor (LIF), 15% fetal bovine serum, nonessential amino acids, glutamate and penicillin/streptomycin in knockout DMEM medium. After incubation, the infection mix was removed and replaced with ESC medium containing 2 ug/ml doxicycline. On day three, cells were split at a 1:5 ratio onto wells with or without gelatinized glass coverslips. At day five, the medium was changed to KSR medium (containing leukemia inhibitory factor (LIF), 15% knockout serum, nonessential amino acids, glutamate and penicillin/streptomycin in knockout DMEM medium) containing 2 ug/ml doxicycline. On day twelve, cells were fixed and immunofluorescence or alkaline phosphatase staining performed and colonies counted.

Cytotoxicity assays

Cytotoxicity assays were performed in collaboration with Timur Yusufzai (Harvard University) on V6.5 ESCs that were either depleted for Zranb3 with an siRNA or that were transfected with a non-targeting construct as described in (Yuan et al., 2012).

## 4.6 References


Beard, C., Hochedlinger, K., Plath, K., Wutz, A., and Jaenisch, R. (2006). Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. Genesis *44*, 23–28.

Bétous, R., Couch, F.B., Mason, A.C., Eichman, B.F., Manosas, M., and Cortez, D. (2013). Substrate-Selective Repair and Restart of Replication Forks by DNA Translocases. Cell Reports *3*, 1958–1969.

Buonomo, S.B.C., Wu, Y., Ferguson, D., and de Lange, T. (2009). Mammalian Rif1 contributes to replication stress survival and homology-directed repair. J. Cell Biol. *187*, 385–398.

Carey, M.F., Peterson, C.L., and Smale, S.T. (2009). Dignam and Roeder Nuclear Extract Preparation. Cold Spring Harb Protoc *2009*, pdb.prot5330.

Chuang, L.S.-H., Ian, H.-I., Koh, T.-W., Ng, H.-H., Xu, G., and Li, B.F.L. (1997). Human DNA-(Cytosine-5) Methyltransferase-PCNA Complex as a Target for p21WAF1. Science *277*, 1996 – 2000.

Ciccia, A., Bredemeyer, A.L., Sowa, M.E., Terret, M.-E., Jallepalli, P.V., Harper, J.W., and Elledge, S.J. (2009). The SIOD disorder protein SMARCAL1 is an RPA-interacting protein involved in replication fork restart. Genes Dev *23*, 2415–2425.

Ciccia, A., Nimonkar, A.V., Hu, Y., Hajdu, I., Achar, Y.J., Izhar, L., Petit, S.A., Adamson, B., Yoon, J.C., Kowalczykowski, S.C., et al. (2012). Polyubiquitinated PCNA recruits the ZRANB3 translocase to maintain genomic integrity after replication stress. Mol. Cell *47*, 396–409.

Ciccia, A., Nimonkar, A.V., Hu, Y., Hajdu, I., Achar, Y.J., Izhar, L., Petit, S.A., Adamson, B., Yoon, J.C., Kowalczykowski, S.C., et al. Polyubiquitinated PCNA Recruits the ZRANB3 Translocase to Maintain Genomic Integrity after Replication Stress. Molecular Cell.

Driscoll, R., and Cimprich, K.A. (2009). HARPing on about the DNA damage response during replication. Genes Dev *23*, 2359–2365.

Fong, Y.W., Inouye, C., Yamaguchi, T., Cattoglio, C., Grubisic, I., and Tjian, R. (2011). A DNA Repair Complex Functions as an Oct4/Sox2 Coactivator in Embryonic Stem Cells. Cell *147*, 120–131.

Van der Kemp, P.A., de Padula, M., Burguiere-Slezak, G., Ulrich, H.D., and Boiteux, S. (2009). PCNA monoubiquitylation and DNA polymerase η ubiquitin-binding domain are required to prevent 8-oxoguanine-induced mutagenesis in Saccharomyces cerevisiae. Nucleic Acids Res *37*, 2549–2559.

Kuo, L.J., and Yang, L.-X. (2008). Gamma-H2AX - a novel biomarker for DNA double-strand breaks. In Vivo *22*, 305–309.

Liu, L., Markowitz, S., and Gerson, S.L. (1996). Mismatch Repair Mutations Override Alkyltransferase in Conferring Resistance to Temozolomide but not to 1,3-Bis(2-chloroethyl)nitrosourea. Cancer Res *56*, 5375–5379.

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

Mailand, N., Gibbs-Seymour, I., and Bekker-Jensen, S. (2013). Regulation of PCNA–protein interactions for genome stability. Nat Rev Mol Cell Biol *14*, 269–282.

Moldovan, G.-L., Pfander, B., and Jentsch, S. (2007). PCNA, the Maestro of the Replication Fork. Cell *129*, 665–679.

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

Stadtfeld, M., Maherali, N., Borkent, M., and Hochedlinger, K. (2010). A reprogrammable mouse strain from gene-targeted embryonic stem cells. Nat. Methods *7*, 53–55.

Stambrook, P.J., and Tichy, E.D. (2010). Preservation of genomic integrity in mouse embryonic stem cells. Adv. Exp. Med. Biol. *695*, 59–75.

Tichy, E.D., Liang, L., Deng, L., Tischfield, J., Schwemberger, S., Babcock, G., and Stambrook, P.J. (2011). Mismatch and base excision repair proficiency in murine embryonic stem cells. DNA Repair *10*, 445–451.

Turinetto, V., Orlando, L., Sanchez-Ripoll, Y., Kumpfmueller, B., Storm, M.P., Porcedda, P., Minieri, V., Saviozzi, S., Accomasso, L., Rocchietti, E.C., et al. High Basal γH2AX Levels Sustain Self-Renewal of Mouse Embryonic and Induced Pluripotent Stem Cells. STEM CELLS.

Virgilio, M.D., Callen, E., Yamane, A., Zhang, W., Jankovic, M., Gitlin, A.D., Feldhahn, N., Resch, W., Oliveira, T.Y., Chait, B.T., et al. (2013). Rif1 Prevents Resection of DNA Breaks and Promotes Immunoglobulin Class Switching. Science *339*, 711–715.

Warbrick, E. (1998). PCNA binding through a conserved motif. BioEssays *20*, 195–199.

Weston, R., Peeters, H., and Ahel, D. (2012). ZRANB3 is a structure-specific ATP-dependent endonuclease involved in replication stress response. Genes Dev. *26*, 1558–1572.

Wolters, D.A., Washburn, M.P., and Yates, J.R. (2001). An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem *73*, 5683–5690.

Wu, Y. (2012). Unwinding and Rewinding: Double Faces of Helicase? Journal of Nucleic Acids *2012*, 1–14.

Yuan, J., Ghosal, G., and Chen, J. (2012). The HARP-like domain-containing protein AH2/ZRANB3 binds to PCNA and participates in cellular response to replication stress. Mol. Cell *47*, 410–421.

Yusufzai, T., and Kadonaga, J.T. (2008). HARP Is an ATP-Driven Annealing Helicase. Science *322*, 748–750.

Yusufzai, T., and Kadonaga, J.T. (2010). Annealing helicase 2 (AH2), a DNA-rewinding motor with an HNH motif. Proc Natl Acad Sci U S A *107*, 20970–20973.

Yusufzai, T., and Kadonaga, J.T. (2011). Branching out with DNA helicases. Current Opinion in Genetics & Development *21*, 214–218.

Zlatanou, A., Despras, E., Braz-Petta, T., Boubakour-Azzouz, I., Pouvelle, C., Stewart, G.S., Nakajima, S., Yasui, A., Ishchenko, A.A., and Kannouche, P.L. (2011). The hMsh2-hMsh6 Complex Acts in Concert with Monoubiquitinated PCNA and Pol η in Response to Oxidative DNA Damage in Human Cells. Molecular Cell *43*, 649–662.

# 5 Summary and Discussion

DNA replication timing, chromatin state and transcriptional activity are tightly correlated and change during differentiation or reprogramming to the pluripotent state. Furthermore, DNA repair is a crucial mechanism required to faithfully maintain the integrity of the genome and has also been implicated to play a role in the establishment and maintenance of pluripotency. This study therefore set out to understand components of the DNA replication and repair pathways in mouse ESCs. The ease of genetic manipulation of ESCs and their scalability makes them an ideal system for proteomic and transcriptional approaches.

As components of the preRC, namely the ORC and MCM complex, have been shown to interact with a number of transcription factors and might contribute to the transcriptional activity of a locus, we knocked down several components of these complexes in ESCs and analyzed the effect of this manipulation by expression microarrays. We were surprised to find that the knockdown had no strong effect on transcription, even under differentiating conditions. This finding therefore fails to explain why transcriptional activity and replication timing are so tightly correlated and future work will be required to elucidate the mechanism behind this correlation. A possible explanation could be that transcription and replication timing are both regulated by the chromatin state and 3D genome organization of the particular locus, which is supported by more recent studies (Chakraborty et al., 2011) (Moindrot et al., 2012) (MacAlpine and Almouzni, 2013).

Since not only the DNA molecule but chromatin marks need to be replicated during DNA replication, we further investigated whether PCNA could be a contributing factor in this process by linking chromatin-modifying enzymes to the replication fork in a S-phase specific manner, thus allowing euchromatin to be replicated in early S-phase and heterochromatin in late S-phase. In order to do so, we determined the PCNA interactome both in early/mid and late S-

phase and identified a number of interaction partners that support this hypothesis. Furthermore, we were able to detect novel PCNA interaction partners, of which we were particularly interested in the DNA replication factor Wdhd1, that had not been previously directly linked to PCNA. Interestingly, similar to PCNA, Wdhd1 also interacts with a number of proteins involved in the covalent modification of chromatin, further supporting the hypothesis that DNA and chromatin replication machineries are tightly linked to each other. It will be interesting to determine in future studies which effect the interruptions of such interactions may have on the inheritance of chromatin.

As one additional novel interaction partner of PCNA, we also identified the annealing helicase Zranb3. In accordance with studies that were published while our investigation was in progress, we confirmed the interaction of Zranb3 with PCNA through the conserved PIP box motif and demonstrated that Zranb3 is involved in DNA repair. Interestingly, Zranb3 is ten times more highly expressed in ESCs than differentiated cells, which suggests that it may have a role in the establishment or maintenance of pluripotency. Surprisingly, we did not find evidence for this hypothesis, but through identification of the MutS$\alpha$ complex amongst Zranb3's interaction partners by mass spectrometry, we were able to demonstrate that Zranb3 is involved in mismatch repair, a process that is critical in stem cells. This finding was then confirmed using cytotoxicity assays and led us to propose a mechanism in which MutS$\alpha$ recruits both PCNA and Zranb3 to sites of MMR in order to repair lesions in the DNA molecule. An important avenue of research will be to determine how this and other DNA repair pathways are involved in the establishment and maintenance of pluripotency, as suggested by other studies (Fong et al., 2011).

Taken together, we have gained insight on crucial components of the DNA replication and repair system in ESC through the use of proteomic and transcriptional analysis, combined with

molecular biology approaches. This work demonstrates the intricate interplay between these pathways and should encourage future researchers to also utilize such powerful approaches as MudPIT in order to understand the functional relevance of identified interactions.

## 5.1 References

Chakraborty, A., Shen, Z., and Prasanth, S.G. (2011). “ORCanization” on heterochromatin: Linking DNA replication initiation to chromatin organization. Epigenetics *6*, 665–670.

Fong, Y.W., Inouye, C., Yamaguchi, T., Cattoglio, C., Grubisic, I., and Tjian, R. (2011). A DNA Repair Complex Functions as an Oct4/Sox2 Coactivator in Embryonic Stem Cells. Cell *147*, 120–131.

MacAlpine, D.M., and Almouzni, G. (2013). Chromatin and DNA Replication. Cold Spring Harb. Perspect. Biol. a010207.

Moindrot, B., Audit, B., Klous, P., Baker, A., Thermes, C., de Laat, W., Bouvet, P., Mongelard, F., and Arneodo, A. (2012). 3D chromatin conformation correlates with replication timing and is conserved in resting cells. Nucleic Acids Res. *40*, 9470–9481.

# 6 Appendix

The Appendix is a reprint of a manuscript as published in Cell Reports:

Alissa Minkovsky, Tahsin Stefan Barakat, Nadia Sellami, Mark Henry Chin, Nilhan Gunhanlar, Joost Gribnau,and Kathrin Plath: The Pluripotency Factor-Bound Intron 1 of Xist Is Dispensable for X Chromosome Inactivation and Reactivation In Vitro and In Vivo.

Cell Rep. 2013 Mar 28; 3(3) :905-18.

# The Pluripotency Factor-Bound Intron 1 of *Xist* Is Dispensable for X Chromosome Inactivation and Reactivation In Vitro and In Vivo

Alissa Minkovsky,[1,2] Tahsin Stefan Barakat,[3] Nadia Sellami,[1,2] Mark Henry Chin,[1,2] Nilhan Gunhanlar,[3] Joost Gribnau,[3] and Kathrin Plath[1,2,*]

[1]Department of Biological Chemistry, David Geffen School of Medicine, Jonsson Comprehensive Cancer Center, Molecular Biology Institute
[2]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research
University of California, Los Angeles, Los Angeles, CA 90095, USA
[3]Department of Reproduction and Development, Erasmus MC, University Medical Center, 3015 Rotterdam, The Netherlands
*Correspondence: kplath@mednet.ucla.edu
http://dx.doi.org/10.1016/j.celrep.2013.02.018

## SUMMARY

X chromosome inactivation (XCI) is a dynamically regulated developmental process with inactivation and reactivation accompanying the loss and gain of pluripotency, respectively. A functional relationship between pluripotency and lack of XCI has been suggested, whereby pluripotency transcription factors repress the master regulator of XCI, the noncoding transcript *Xist*, by binding to its first intron (intron 1). To test this model, we have generated intron 1 mutant embryonic stem cells (ESCs) and two independent mouse models. We found that *Xist*'s repression in ESCs, its transcriptional upregulation upon differentiation, and its silencing upon reprogramming to pluripotency are not dependent on intron 1. Although we observed subtle effects of intron 1 deletion on the randomness of XCI and in the absence of the antisense transcript *Tsix* in differentiating ESCs, these have little relevance in vivo because mutant mice do not deviate from Mendelian ratios of allele transmission. Altogether, our findings demonstrate that intron 1 is dispensable for the developmental dynamics of *Xist* expression.

## INTRODUCTION

To balance the expression of X-linked genes between males and females, female mammals silence one of the two X chromosomes in a developmentally regulated process called X chromosome inactivation (XCI). XCI occurs in two waves in the course of mouse embryogenesis. The earliest form of XCI begins at the two- to four-cell stage in preimplantation embryo and is imprinted, selectively occurring on the paternally inherited X chromosome (Xp) (Huynh and Lee, 2003; Kalantry et al., 2009; Namekawa et al., 2010; Patrat et al., 2009). At the preimplantation blastocyst stage, imprinted XCI is retained in the trophectoderm and primitive endoderm lineages but reversed in arising pluripotent epiblast cells yielding a state with two active X chromosomes (XaXa) (Mak et al., 2004; Okamoto et al., 2004; Silva et al., 2009; Williams et al., 2011). Upon implantation, these epiblast cells establish a random form of XCI that stochastically initiates on the maternal or paternal X chromosome and is retained through the lifetime of mitotic divisions (Kay et al., 1993; Rastan and Robertson, 1985). Similarly, mouse embryonic stem cells (ESCs), which are derived from epiblast cells of the preimplantation blastocyst, undergo random XCI when induced to differentiate ex vivo. The only exception to somatic maintenance of random XCI is inactive X (Xi) reactivation in the germline, which is assumed to be essential for female fertility and occurs in primordial germ cells as they traverse the hindgut to seed the genital ridges (Chuva de Sousa Lopes et al., 2008; de Napoles et al., 2007; Sugimoto and Abe, 2007). Xi reactivation is also a feature of experimentally induced acquisition of pluripotency via transcription factor-mediated reprogramming to induced pluripotent stem cells (iPSCs), fusion of somatic cells with ESCs, or somatic cell nuclear transfer (Eggan et al., 2000; Maherali et al., 2007; Tada et al., 2001).

The cycles of XCI and Xi reactivation are associated with changes in *Xist* RNA coating, where cells with a Xi display coating by the noncoding *Xist* RNA on the Xi chromosome, and those with two active X chromosomes lack *Xist* RNA expression (Brockdorff et al., 1991; Brown et al., 1991). *Xist*'s function has been most studied in the random form of XCI in the mouse system, where it is shown to be the critical trigger of XCI. The upregulation of *Xist* RNA and coating of the X at the onset of random XCI immediately lead to transcriptional silencing of X-linked genes and result in the exclusion of RNA polymerase II and the recruitment of repressive chromatin-modifying protein complexes such as the Polycomb complex PRC2, which establishes an accumulation of H3K27me3 (Chaumeil et al., 2006; Chow and Heard, 2009; Plath et al., 2003; Silva et al., 2003). A stereotypic order of changes in chromatin structure culminates in heritable silencing of either the maternally or paternally transmitted X chromosome in each cell of the female adult mammal. *Xist* is essential for XCI to occur in *cis* because its deletion leads to silencing of the other X chromosome carrying an intact *Xist* allele, regardless of parent of origin (Marahrens et al., 1997; Penny et al., 1996). Moreover, the importance of *Xist* regulation for the developmental and sex-specific context of XCI is

demonstrated by its sufficiency: overexpression of a X-linked *Xist* cDNA transgene in male mouse ESCs (XY:tetOP-*Xist*) initiates XCI and cell death due to silencing of the single X chromosome (Wutz and Jaenisch, 2000).

*Xist* is transcribed from a larger locus on the X chromosome that has been defined as the minimal critical region for XCI and besides housing *Xist*, contains other protein-coding and non-coding activators and repressors of *Xist*, some of which act in *cis* and others in *trans* (Rastan and Robertson, 1985; reviewed in Minkovsky et al., 2012). The best-characterized repressor of *Xist* is its antisense transcript, *Tsix*, which is highly transcribed in epiblast cells of the preimplantation blastocyst and in undifferentiated mouse ESCs/iPSCs, where *Xist* is repressed (Lee et al., 1999; Sado et al., 2001; Maherali et al., 2007). Deletion of *Tsix* leads to only slight *Xist* upregulation without causing precocious XCI or *Xist* RNA coating in self-renewing, undifferentiated ESCs. However, upon differentiation, XCI is skewed to the *Tsix*-deleted X in female cells heterozygous for the mutant *Tsix* allele (Lee et al., 1999; Lee, 2000; Luikenhuis et al., 2001; Sado et al., 2001). The effect of *Tsix* deletion on *Xist* indicates that it participates in parallel pathways with other regulators of *Xist* repression or activation.

Interestingly, the pluripotency factors Oct4, Sox2, and Nanog have been implicated in the control of *Xist* expression in pluripotent cells. Navarro and colleagues found that in mouse ESCs, Oct4, Sox2, and Nanog bind the first intron of the *Xist* gene (intron 1) (Navarro et al., 2008), a finding that has been recapitulated in many genomic data sets and extends to additional pluripotency regulators such as Tcf3 and Prdm14, and early developmental regulators such as Cdx2 (Figure S1A; Loh et al., 2006; Marson et al., 2008; Ma et al., 2011; Erwin et al., 2012). Such genomic regions of extensive pluripotency transcription factor co-occupancy in the ESC genome occur more commonly than would be expected by chance (Chen et al., 2008). It is thought that these cobound genomic regions represent functionally important sites and often represent enhancer elements (Chen et al., 2008). Further support for a gene regulatory role of intron 1 is that, in ESCs, the intron 1 region has a propensity to be in the three-dimensional proximity to the promoter of *Xist* and adopts a DNase hypersensitive state (Tsai et al., 2008). Additionally, pluripotency factors appear directly linked to *Xist* regulation. Upon *Nanog* deletion or inducible repression of *Oct4*, *Xist* is upregulated, and binding of the pluripotency factors to intron 1 is lost (Navarro et al., 2008). In male ESCs, which normally do not upregulate *Xist*, experimentally forced *Oct4* repression can even induce *Xist* RNA coating in up to 10% of the cells (Navarro et al., 2008). Another study could not replicate *Xist* RNA coating upon *Oct4* knockdown in male ESCs but observed biallelic XCI in differentiating female ESCs upon *Oct4* depletion (Donohoe et al., 2009). A role for Nanog in *Xist* suppression is also supported by its expression pattern with regard to domains of Xi reactivation in the preimplantation blastocyst, where the restriction of Nanog expression demarcates the fraction of cells undergoing reactivation of the imprinted Xi (Silva et al., 2009). Furthermore, preimplantation embryos lacking Nanog are unable to specify epiblast cells and to lose *Xist* RNA, whereas forced expression of Nanog induces a more rapid loss of *Xist* RNA coating in developing preimplantation embryos (Silva et al., 2009; Williams et al., 2011).

Together, these findings led to the model that pluripotency factor binding to intron 1 is critical for repression of *Xist* in undifferentiated XaXa ESCs. However, in the experiments leading to this conclusion, cell identity and, therefore, likely the expression of many genes were modulated by experimental changes in pluripotency factor expression, which could confound the interpretation that Oct4, Nanog, and other pluripotency factors act directly on intron 1 of *Xist* to regulate XCI. It has also been suggested that the pluripotency transcription factors control the levels of positive and negative regulators of *Xist* because they are binding to *Tsix* and the *trans*-acting activator of XCI, *Rnf12* (Donohoe et al., 2009; Gontan et al., 2012; Navarro et al., 2010, 2011). Accordingly, an experiment directly addressing the functional importance of binding to intron 1 showed only subtle dysregulation of XCI: in female ESCs carrying a heterozygous deletion of intron 1 of *Xist*, XCI remained suppressed in the undifferentiated state. However, upon differentiation, *Xist* appeared more highly expressed from the chromosome carrying the mutation, supporting a role for intron 1 in suppressing *Xist* during differentiation (Barakat et al., 2011). Furthermore, deletion of intron 1 in the context of a transgene carrying the extended *Xist* locus moderately increased expression of *Xist* in undifferentiated ESCs, which was amplified by simultaneous deletion of the antisense transcript *Tsix* (Nesterova et al., 2011). Notably, these results were very variable between clones potentially reflecting the effect of transgene copy number and variations (Nesterova et al., 2011). Binding to *Xist* intron 1 has also been proposed to govern the switch from imprinted to random XCI in preimplantation development (Erwin et al., 2012). In vitro, gel shift assays suggest that the binding events between *Xist*'s intron 1 and the pluripotency regulator Oct4 and the trophectoderm regulator Cdx2 are direct but mutually exclusive (Erwin et al., 2012). Collectively, these findings motivated us to examine the role of *Xist* intron 1 further to test the model wherein pluripotency factor binding silences *Xist* to prevent XCI in pluripotent cells and to determine the role of the intronic region in X chromosome reactivation events, both in vivo and in vitro.

## RESULTS

### Generation of Conditional *Xist* Intron 1 ESC Lines

To further define the role of *Xist* intron 1, we used gene targeting to generate a conditional allele in male and female mouse ESCs. We tested the requirement of intron 1 in both sexes because male ESCs are able to undergo XCI upon forced expression of *Xist*, providing a sensitive background for monitoring *Xist* regulation independently of other X chromosomes present in a cell (Wutz and Jaenisch, 2000). By contrast, heterozygous female ESCs permit investigation of kinetics of XCI upon induction of differentiation and insight into potential effects on skewing of XCI between the targeted and wild-type (WT) chromosome.

To delineate the region of intron 1 involved in *Xist* repression, we inspected where pluripotency transcription factors bind within the intron 1 region as detected by published chromatin immunoprecipitation sequencing (ChIP-seq) data sets (Marson et al., 2008). We also determined the localization of pluripotency

factor DNA binding motifs and considered sequence conservation across mammals (Figure S1). We found that co-occupancy of pluripotency factors occurs in a 600 bp region within the full 2.8 kb sequence of intron 1. Most of the intron 1 sequence is not conserved in placental mammals; however, two highly conserved composite Oct4-Sox2 DNA binding motifs, which are found to stabilize a ternary Oct4-Sox2-DNA complex in the expression of many ESC-specific genes, underlie the ChIP-seq binding peaks of Oct4 and Sox2 (Figure S1; Reményi et al., 2003; Marson et al., 2008; Mason et al., 2010; UCSC genome browser phastCons conserved-elements track, http://genome.ucsc.edu). On the basis of these data, we decided to delete 800 bp of intron 1 and, subsequently, refer to this mutation as "intron 1" (Minkovsky/Plath allele; Figure S1).

We flanked the 800 bp intron 1 region with loxP sites, simultaneously inserting a hygromycin-resistance cassette (yielding a targeted allele with 3loxP sites), and, subsequently, generated experimental (1lox) and control (2lox) alleles by transient expression of Cre recombinase in hemizygously targeted male and heterozygous female ESCs (Figures 1 and S2). To be able to monitor the effects of the deletion of intron 1 on *Xist* in *cis* in female cells, we employed genetically polymorphic F1 2-1 female ESCs (129/Cas) carrying a MS2 RNA tag in exon 7 of *Xist* on the 129 allele (Jonkers et al., 2008). Southern blotting and PCR analysis confirmed that intron 1 was targeted in *cis* to the MS2 RNA tag in female ESCs (Figure S2). Male- and female-targeted ESC lines showed normal chromosome complement upon karyotyping (Figure S2; data not shown).

To confirm that deletion of 800 nucleotides from intron 1 sufficiently removes pluripotency factor binding, we performed ChIP against Oct4 and Sox2 coupled to quantitative PCR for the targeted region of intron 1, neighboring intronic regions, the *Xist* promoter, and previously validated control regions (Navarro et al., 2008). Importantly, we did not observe an increase in Oct4 or Sox2 binding in these regions upon deletion of intron 1 (Figures 1G–1I). Thus, compensatory binding at cryptic binding sites upon intron 1 deletion appears unlikely.

### Ectopic *Xist* RNA Coating Is Not Observed in Intron 1-Deleted Undifferentiated and Differentiating Male and Female ESCs

To understand the role of intron 1 in the regulation of XCI, we first performed fluorescence in situ hybridization (FISH) to analyze the expression and localization of *Xist* and *Tsix* RNA at the single-cell level using strand-specific RNA probes. Undifferentiated male and female ESC lines displayed no significant *Xist* RNA cloud or pinpoint signal in the presence or absence of intron 1 (Figures 2A and 2B). The absence of *Xist* RNA coating in the undifferentiated ESC state was confirmed by the lack of a Xi-like enrichment of H3K27me3 in Nanog+ cells, which is known to occur on the Xi when *Xist* RNA coats (Plath et al., 2003; Silva et al., 2003) (Figures S3A and S3B). In agreement with this finding, the signal for *Tsix* was present in the majority of cells in each case and indistinguishable among all tested genotypes (Figure 2A).

Upon induction of differentiation by embryoid body (EB) formation, the lack of intron 1 did not induce *Xist* RNA in male ESCs to a level detectable by FISH (data not shown) and yielded no Xi-like enrichment of H3K27me3 (Figures S3C and S3D), indicating that intron 1 is not an essential regulator of *Xist* suppression in differentiating male ESCs when all other regulators of XCI are intact. Heterozygous 1lox/WT female ESCs formed *Xist* RNA clouds and H3K27me3 Xi foci at comparable rates to 2lox/WT control ESCs (Figures 2C, 2D, S3C, and S3D). *Xist* RNA levels were also similar between undifferentiated and differentiating male and female ESCs, with or without intron 1, in RT-PCR experiments (Figure 2E). Proper differentiation was confirmed by decrease in *Nanog* transcript levels (Figure 2F). Furthermore, the use of *Xist* intron 1-spanning PCR primer pairs ruled out dramatic secondary effects of intron 1 deletion on *Xist* splicing (data not shown).

Next, we assessed whether XCI is skewed upon intron 1 deletion in differentiating female ESCs. The polymorphic 129/cas F1 2-1 female ESC line is known to have a baseline skewing of XCI toward the 129 allele such that approximately 70% of the cells will silence the 129 allele, due to strain-specific haplotypes (Cattanach and Isaacson, 1967). Due to the integration of the MS2 RNA tag on the intron-targeted 129 X chromosome, combined RNA-FISH for MS2 and *Xist* sequences can distinguish between *Xist* being expressed from the targeted chromosome (positive for both *Xist* and MS2 signals) and the untargeted X (only marked by the *Xist* probe) (Figure 2C; Jonkers et al., 2008). We found that, at the single-cell level, female 1lox intron/WT ESCs consistently had ~15% more cells expressing the MS2-tagged *Xist* than their 2lox/WT counterparts, in three of four ex vivo differentiation methods (Figures 2G and S4). This mild skewing effect in differentiating female ESCs is consistent with published results (Barakat et al., 2011).

### Genetic Interaction of *Xist* Intron 1 with *Tsix*

Next, we investigated the possibility that the intron 1-dependent skewing of XCI in differentiating female ESCs represents a mild effect on the intron 1-deleted X chromosome at the transition to the differentiated state. We reasoned that such an effect may be more strongly revealed in the absence of other regulators of *Xist* and sought to assay such an effect on a "sensitized" background for *Xist* transcription. *Tsix* represents the prime candidate for a redundant *Xist* repressor that could compensate to repress *Xist* in the absence of intron 1. One study supports the view that a functional role for the intron can be uncovered in the absence of *Tsix* because male ESCs with randomly integrated genomic *Xist* transgenes lacking intron 1 and a functional *Tsix* allele dysregulated the expression of the transgenic *Xist* (Nesterova et al., 2011). We therefore performed the aforementioned analyses in male ESCs lacking intron 1 in the endogenous *Xist* allele on the background of a previously characterized *Tsix* loss-of-function mutation at the endogenous locus (Figure 3; Lee et al., 1999; Luikenhuis et al., 2001; Sado et al., 2001). We targeted the disruption of *Tsix* to both 2lox and 1lox intron male ESCs using a construct that inserts a splice acceptor-IRES*βGeo* cassette in exon 2 of *Tsix* resulting in an early transcriptional stop (Figure 3A) (Sado et al., 2002). Correct targeting and loss of the *Tsix* transcript were confirmed by Southern blot (Figure 3B) and absence of FISH signal for *Tsix* (Figure 3C).

As expected, in the presence of intron 1 (2lox intron 1), *Tsix* deletion in male ESCs induced a mild transcriptional

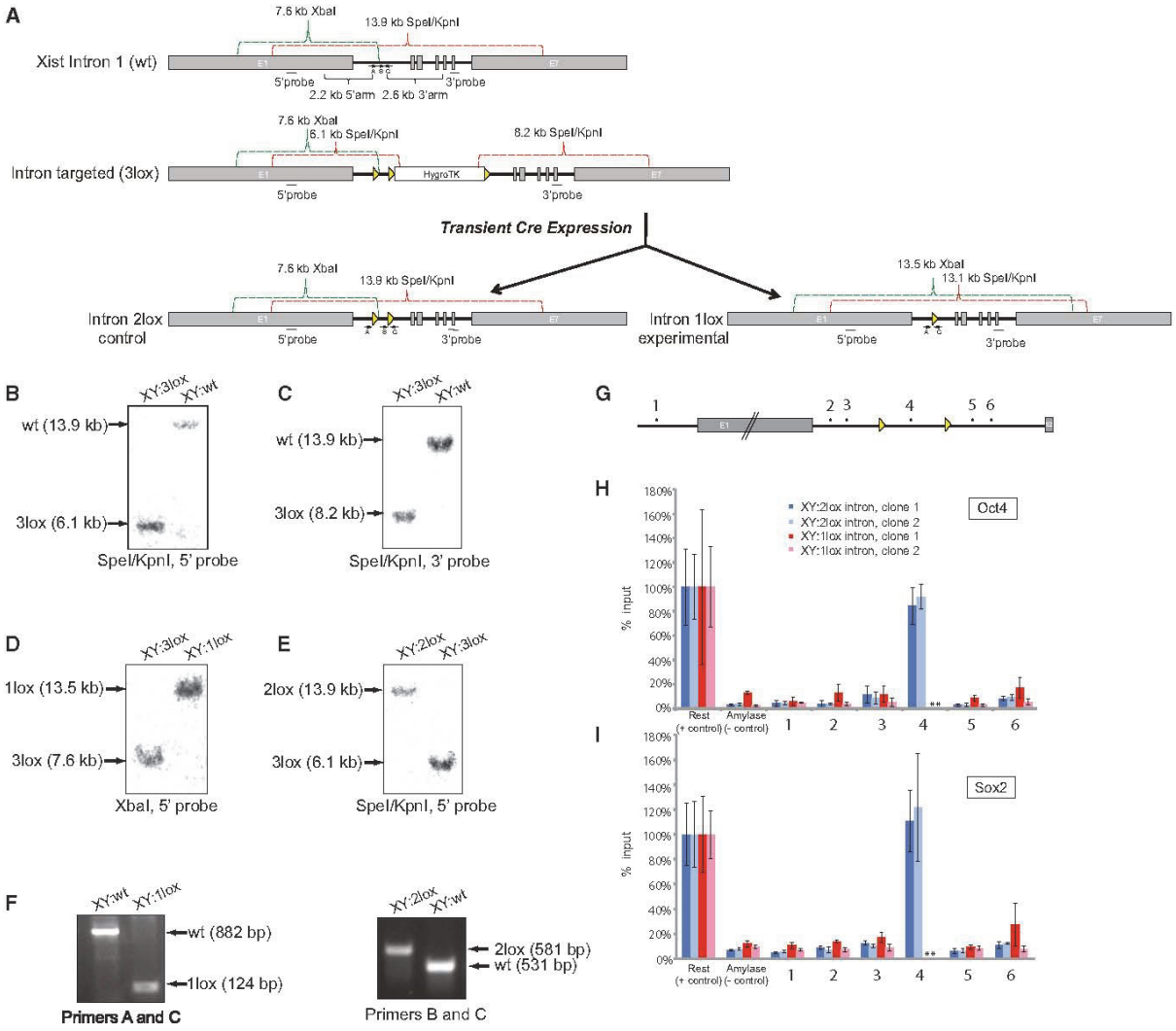



**Figure 1. Generation of Male mouse ESCs Carrying a Conditional *Xist* Intron 1 Allele**

(A) Gene targeting and Southern blotting strategy schematic for male ESCs. Transient expression of Cre recombinase in properly targeted 3lox clones yielded both 2lox (control) and 1lox (experimental) ESC lines.

(B–E) Representative images of correctly targeted clones from Southern blot analysis.

(F) PCR genotyping with primers A and C shows the presence of the 1lox allele, and genotyping with primers B and C shows that of the 2lox allele.

(G) Location of ChIP-qPCR primer sets within the Xist locus used in the subsequent figures. Primer pair 1 is located within the Xist promoter.

(H and I) ChIP-qPCR analysis of Oct4 (H) and Sox2 (I) binding to regions indicated in (G) and a known positive and negative control (within Rest and Amylase, respectively) for Oct4 and Sox2 binding (van den Berg et al., 2008) in 2lox and 1lox intron 1 male ESCs (two clones each). Values represent the amount of DNA precipitated after normalization to input chromatin and are given relative to binding within the positive control region. Error bars represent SD from triplicate qPCR measurements. Asterisk (*) indicates high Ct values for the input samples in the genetically deleted regions, probably arising from support feeder cells.

See also Figures S1 and S2.

upregulation of *Xist* RNA compared to XY:2lox/*Tsix*-WT ESCs in RT-PCR experiments, reaching a level found in female ESCs (Figure 3D). Upon differentiation, XY:2lox intron/*Tsix*-Stop ESCs further upregulated *Xist* transcript levels ~5-fold (Figure 3D). However, this induction was rarely correlated with a *Xist* RNA cloud signal detectable by RNA FISH or a Xi-like H3K27me3 accumulation (Figures 3E–3H) before and after induction of differentiation, in agreement with previous reports by Luikenhuis et al. (2001) and Sado et al. (2002). Combined deletion of intron 1 and *Tsix* did not alter the *Xist* status in undifferentiated ESCs but upon induction of differentiation, resulted in a *Xist* RNA cloud-like signal in FISH experiments in 3%–6% of

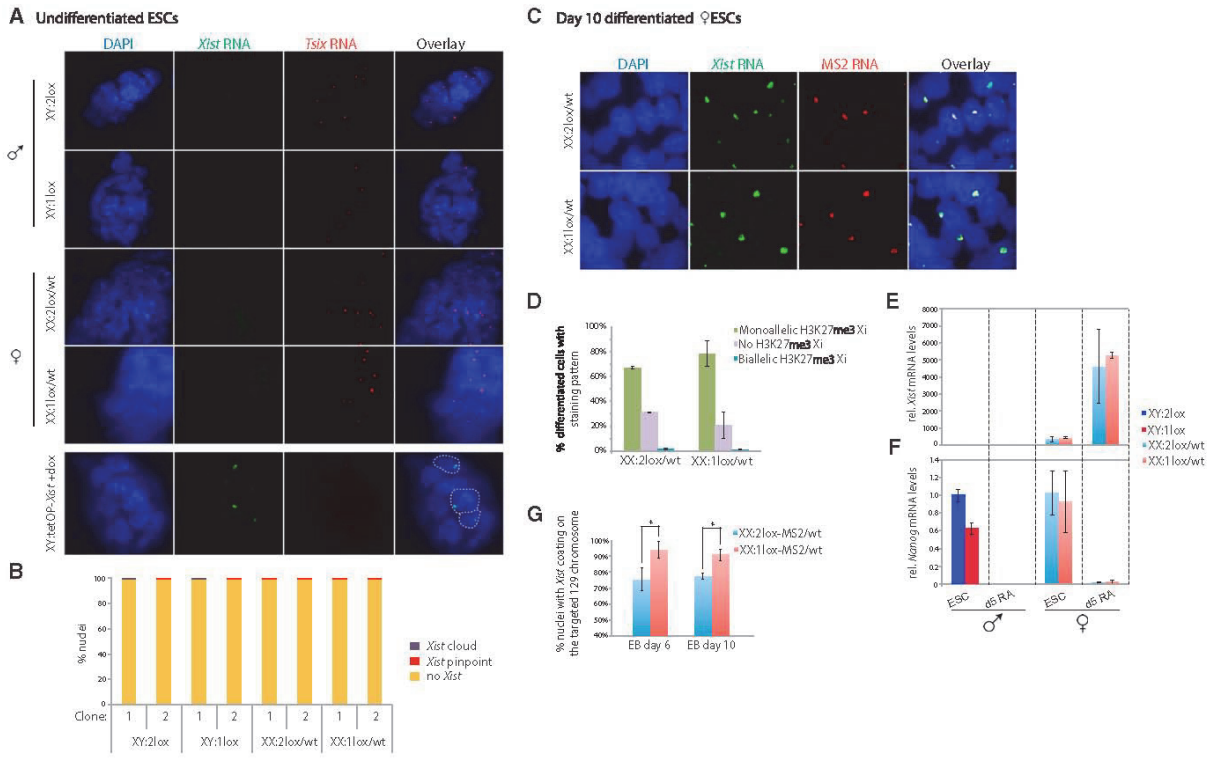



**Figure 2. Analysis of *Xist* Expression in Undifferentiated and Differentiating Female and Male ESC Lines in the Presence and Absence of Intron 1**

(A) Strand-specific FISH for *Xist* RNA (green) and *Tsix* RNA (red) in undifferentiated male and female ESCs of the indicated genotypes, using RNA probes. DAPI staining (blue) indicates nuclei. Representative images are shown. A male ESC line carrying a doxycycline (dox)-inducible *Xist* allele in the endogenous locus was used as positive control for the *Xist*-staining pattern, at 24 hr of dox addition.

(B) Graph summarizes the proportion of DAPI-stained nuclei with indicated patterns of *Xist* RNA based on an experiment as described in (A). Pairs of independent ESC clones of the given genotype were stained and counted. In each case, 500 nuclei were assessed.

(C) FISH with DNA probes targeting *Xist* RNA (green) and the *MS2* tag (red), respectively, in female ESCs of indicated genotypes at day 10 of EB differentiation.

(D) Graph summarizing the proportion of Nanog-negative cells in day 10 EB-differentiated female ESCs with no, one, or two H3K27me3 Xi-like accumulations. Notably, the number of cells within each H3K27me3 pattern is not statistically different (by Student's t test) between 2lox/wt and 1lox/wt ESC lines. Values are means of counts of independent clones as shown in Figure S4D; in each case, at least 500 nuclei were assessed.

(E) RT-PCR for *Xist* RNA levels normalized to *Gapdh* expression from one representative clone of indicated ESC genotypes in the undifferentiated state (ESC) and at day 5 of RA differentiation (d5 RA). Error bars indicate SD from triplicate RT-PCR measurements in one experiment.

(F) As in (E), except that *Nanog* transcript levels were analyzed.

(G) Quantification of allele-specific *Xist* RNA cloud patterns from the experiment shown in (C) at days 6 and 10 of EB differentiation, given as mean of values from counts of two independent ESC clones of the indicated genotype. *Xist* expression from the 129 chromosome (targeted chromosome) is detected by both the *Xist* and MS2 probes, whereas *Xist* expression from the CAST chromosome is only detected by the *Xist* probe. The graph depicts the percentage of cells where the intron 1-targeted 129 chromosome is coated by *Xist* RNA, as identified by colocalization of the *Xist* and MS2 signals. *$p < 0.05$ by Student's t test with 500 *Xist* clouds analyzed for each sample.

See also Figures S3 and S4.

cells compared to 0.2%–0.8% in differentiating XY:2lox intron/*Tsix*-Stop cells (Figures 3E and 3G). We did not, however, see any significant intron 1-dependent effect on *Xist* RNA levels by RT-PCR comparing XY:2lox intron/*Tsix*-Stop and XY:1lox intron/*Tsix*-Stop cells (Figure 3D) or an increase in the number of H3K27me3 Xi-like accumulations (Figures 3F and 3H). Thus, even though *Xist* RNA was induced in a slightly larger proportion of differentiating cells in the absence of both *Tsix* and intron 1 than in the absence of either *Tsix* or intron 1, this upregulation does not appear to be sufficient to mediate H3K27me3 enrichment on the targeted X chromosome, suggesting that the RNA does not efficiently coat the chromosome in these cells or that the recruitment of Polycomb proteins is affected. We conclude that these experiments reveal a subtle role of intron 1 in the control in *Xist* expression, which may be related to the weak skewing phenotype of XCI described above for differentiating intron 1 mutant heterozygous female ESCs (Figures 2 and S4).

In a second assay, we tested the consequence of intron 1 deletion upon modulation of global *Oct4* transcript levels. We first confirmed the previously reported relationship between

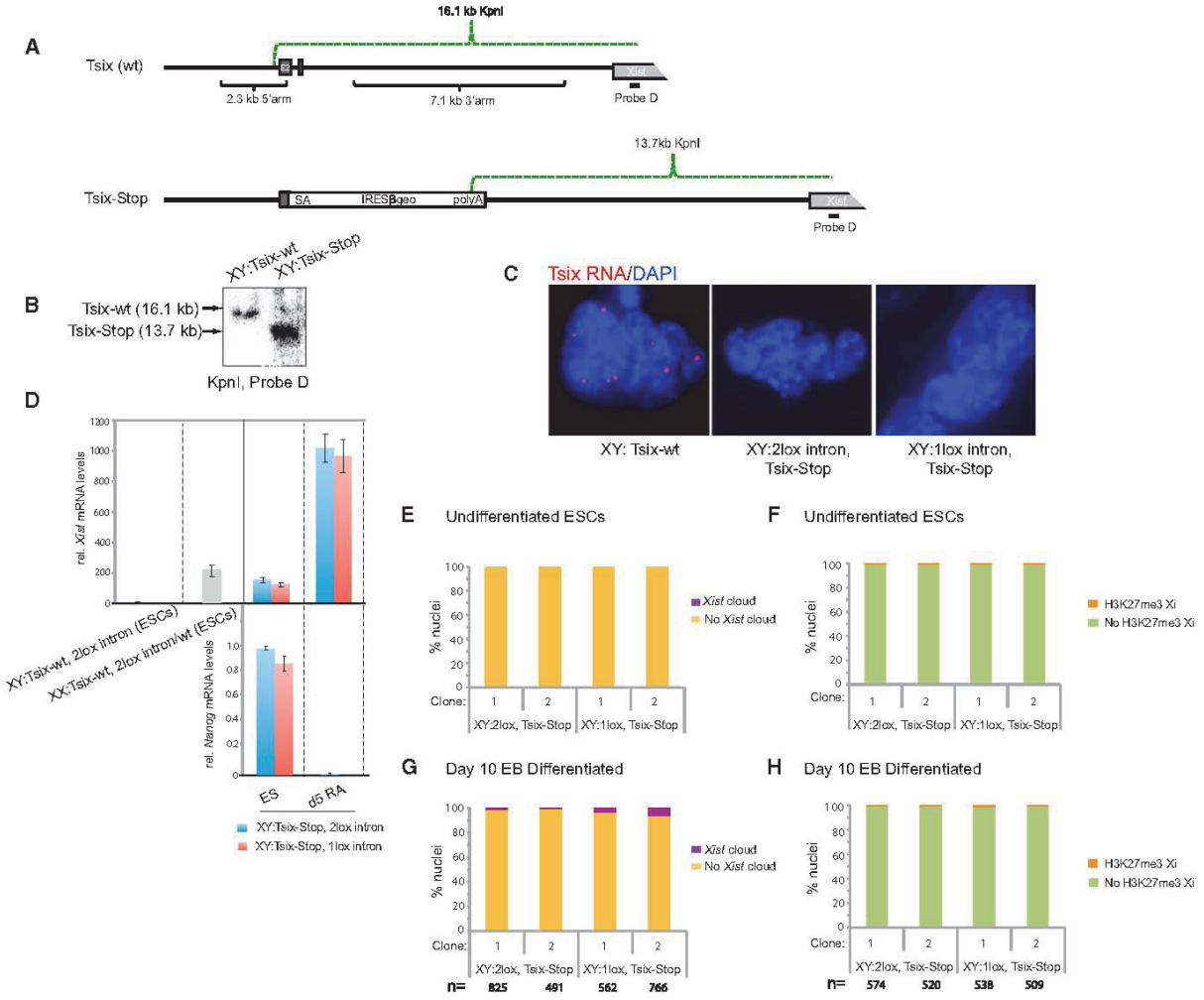



**Figure 3. *Xist* RNA Pattern in Male ESCs Lacking *Tsix* and Intron 1**

(A) Gene targeting and Southern blotting strategy schematic for the generation of the *Tsix*-Stop allele in male ESCs according to Sado et al. (2001) using the pAA2Δ1.7-targeting vector.

(B) Male 2lox and 1lox intron 1 ESC clones were targeted with the *Tsix*-Stop allele. A correctly targeted 2lox intron 1/*Tsix*-Stop male ESC clone is shown in this Southern blot analysis.

(C) Strand-specific FISH for *Tsix* RNA (red) in undifferentiated male ESCs of the indicated genotypes, using an RNA probe, indicates the absence of the *Tsix* FISH signal in *Tsix*-Stop-targeted clones.

(D) Graph summarizing the transcript levels for *Nanog* and *Xist* normalized to *Gapdh* transcript levels as determined by RT-PCR from a representative clone of each genotype in the undifferentiated state (ES) and at day 5 of RA differentiation. Control *Xist* RNA levels from WT undifferentiated male and female ESCs are shown on the left. Error bars indicate SD from triplicate RT-PCR measurements in one experiment.

(E) Graph summarizing the percentage of undifferentiated ESCs of the given genotype with and without a *Xist* RNA cloud-like pattern. Two independent male ESC clones for each genotype were analyzed by *Xist* RNA FISH with a RNA probe, and 500 nuclei were assessed.

(F) As in (E), except that the percentage of undifferentiated ESCs with and without a H3K27me3 Xi-like accumulation is given.

(G) *Xist* RNA cloud quantification as in (E), except that *Nanog*-negative cells were quantified upon day 10 of EB differentiation.

(H) As in (F) for H3K27me3 patterns at day 10 of EB differentiation in indicated ESC lines.

See also Figure S5.

the decrease of *Oct4* levels and *Xist* RNA induction (Navarro et al., 2008; Donohoe et al., 2009). Specifically, upon *Oct4* depletion in the male ZHBTc4 ESC line, in which *Oct4* expression can be silenced acutely by the addition of doxycycline (Niwa et al., 2000), *Xist* RNA levels were induced almost 100-fold 96 hr post-induction of *Oct4* repression (Figure S5A), and *Xist* RNA could be detected by FISH in a small number of cells (Figures S5B and S5C). Notably, we observed that *Oct4* transcript levels drop

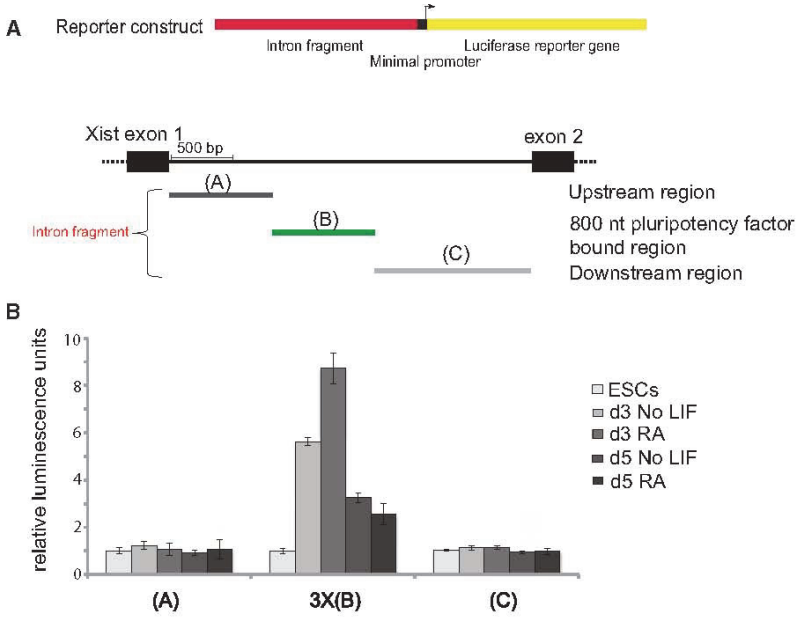



**Figure 4. Enhancer Assay of *Xist* Intron 1**

(A) Schematic representation of genomic fragments of the entire genomic intron 1 region cloned upstream of a luciferase reporter gene driven by a minimal promoter. The intronic region was broken up in three parts, with (B) representing the part bound by pluripotency factors and flanked by loxP sites as described in Figure 1 (Minkovsky/Plath allele), and (A) and (C) representing regions not bound by the pluripotency factors in ChIP-seq experiments (Figure S1). Note that (B) was concatenated 3× in the reporter construct.

(B) Three independent stable cell lines were generated by electroporation of male ESCs with the three constructs described in (A). Cells carrying the reporter constructs were selected with hygromycin, and an equal number of cells was plated and maintained in the undifferentiated state or differentiated for days 3 and 5 by LIF withdrawal and/or RA addition as indicated. After treatment, one-tenth of the cells in the well was analyzed by luciferase assay. For each reporter construct, values represent mean luminescence units normalized to values from the respective cell line in the undifferentiated state (n = 3, ± 1 SD). See also Figure S6.

with faster kinetics than *Xist* RNA levels increase, suggesting that the effect of *Oct4* on intron 1 is indirect and may require efficient differentiation, which occurs at 96 hr post-*Oct4* repression, as indicated by the loss of the pluripotency factor Nanog (Figure S5D). In agreement with this conclusion, siRNA-mediated knockdown of *Oct4* in ESCs did not increase *Xist* RNA levels more than 2-fold after 72 hr, confirming a previous report by Donohoe et al. (2009) (Figure S5E). Furthermore, the absence of intron 1 did not significantly alter *Xist* RNA levels in female ESCs or in male ESCs lacking *Tsix* in *Oct4* knockdown conditions (Figure S5E). These data indicate that the slight increase in *Xist* levels immediately upon *Oct4* depletion is independent of intron 1.

## Intron 1 Acts as an Enhancer in a Reporter Assay in Differentiating ESCs

The model of pluripotency factor binding to intron 1 to repress *Xist* motivated us to directly assess whether intron 1 behaved as a silencer in ESCs in a reporter assay. We transfected constructs with intron 1 or control sequences upstream of a minimal promoter driving luciferase and did not see intron 1-dependent decreases in reporter activity (data not shown). The small effect of intron 1 deletion on *Xist* RNA levels detected in differentiating ESCs in the absence of *Tsix* motivated us to revisit these experiments and instead investigate whether *Xist* intron 1 represents a developmentally regulated enhancer that becomes active upon induction of differentiation. We therefore tested transactivation activity of intron 1 in undifferentiated and differentiating ESCs using stably integrated luciferase reporter constructs (Figure 4). Male ESCs were electroporated with hygromycin resistance-bearing constructs containing either the part of intron 1 that we deleted in our experimental cell lines or two control sequences representing the upstream and downstream flanking regions of the intron 1 region (Figure 4A). The experimental intron 1 region ((B) in Figure 4B) was cloned in triple copy to amplify any putative enhancer activity of this region. Pooled clones were subjected to monolayer differentiation by LIF withdrawal with and without retinoic acid (RA) treatment. Only cells bearing the intron 1 construct covering the pluripotency factor binding site showed a robust increase in luciferase activity upon differentiation (Figure 4B). In agreement with the notion that intron 1 does not act as an active enhancer in undifferentiated ESCs, we did not find a histone acetylation mark characteristic of active enhancers, namely H3K27ac, examining our own and published ChIP-seq data sets from ESCs, despite binding of intron 1 by a battery of pluripotency factors and p300 in undifferentiated ESCs (mouse ENCODE; Creyghton et al., 2010; data not shown).

We also considered recently published spatial organization data that demonstrated that the *Xist* gene lies in a topologically associating domain (TAD) with genes encoding the noncoding RNAs *Ftx* and *Jpx/Enox* and the protein-coding genes *Rnf12/Rlim*, *Zcchc13*, and *Slc16a2* (Nora et al., 2012). It has been proposed that promoters and enhancers predominantly interact (loop) within TADs (Dixon et al., 2012; Nora et al., 2012). Notably, significant intra-TAD contacts originating from within intron 1 of *Xist*, indicative of putative enhancer/promoter looping, were only found in differentiated and not in undifferentiated ESCs (Nora et al., 2012) (Figure S6A), consistent with our finding of reporter activity upon differentiation. However, similar to our result that *Xist* levels in female and male ESCs did not significantly change in the absence of intron 1, we also did not see intron 1-dependent transcriptional differences in the three genes that come in contact with intron 1 within the *Xist*-containing TAD, before and during differentiation (Figure S6B). Thus, even though intron 1 is pluripotency factor bound in ESCs, it may only gain significant enhancer activity upon differentiation

though still not to an extent where deletion affects transcription of *Xist* or of neighboring protein-coding genes.

Together, these ex vivo studies in undifferentiated and differentiating male and female ESCs point to a minor role for intron 1 in the regulation of *Xist* expression, uncovered only when another *Xist* repressor is deleted, and some aspect of X chromosome choice (potentially also through slight modulation of *Xist* RNA levels). These data do not support intron 1 as a main aspect of the mechanism of transcriptional repression of *Xist*, at least in this tissue culture model.

### Mice Are Normal in the Absence of Intron 1

Next, we assayed the significance of intron 1 in vivo. Our male ESCs deleted for intron 1 (1lox) were injected into C57BL/6 blastocysts. Chimeras were obtained at high efficiency and bred with C57BL/6 females to obtain germline transmission of the mutant allele. Subsequently, the 1lox intron 1 allele showed normal propagation through the maternal or paternal germline, and mice completely lacking intron 1 (crossing 1lox/1lox females with 1lox males) could be efficiently bred without any female-specific defect (Figure 5A). Because X chromosome reactivation occurs in the female germline and is likely essential for female fertility, we assessed litter size of the F2 generation of female homozygous knockout mice, and we found their litter sizes unaffected (data not shown), indicating that intron 1 is not essential in mice.

To strengthen these observations of normal transmission of the intron 1 mutation and rule out that genetic background obscured a potential intron 1 phenotype in vivo, we generated a second mouse model carrying an independent intron 1 mutation. We generated mice using previously published 129/CAST F1 female ESCs in which a larger (1.815 kb) region harboring most of intron 1 was deleted on the 129 X chromosome (Barakat/Gribnau allele) (Figure S1A; Barakat et al., 2011). Using these ESCs, we previously observed a slight upregulation of *Xist* RNA levels on the deleted chromosome upon induction of differentiation (Barakat et al., 2011), in agreement with results obtained using the Minkovsky/Plath allele, indicating skewing of X inactivation toward the intron 1-deleted chromosome. Importantly, this second mouse model also displayed normal Mendelian transmission of the intron 1 lox allele (Figure 5B).

To assay whether random XCI has occurred in female mice carrying a Xp lacking intron 1 and a maternally inherited WT X chromosome, and whether the lack of the intron leads to any skewing of XCI in vivo, we analyzed the allele-specific expression of *Xist* and two X-linked genes, *Mecp2* and *G6pdx*, in polymorphic heterozygous females ($1lox^{C57BL/6}/WT^{CAST/Ei}$) and a WT control ($WT^{C57BL/6}/WT^{CAST/Ei}$). In these mice, the C57BL6 X chromosome was transmitted from the father and the CAST/Ei WT X from the mother. Allele-specific expression analysis was performed using semiquantitative RT-PCR on RNA isolated from various tissues (Figures 5C–5E). In these experiments, we used the Barakat/Gribnau mouse model described in Figures 5B and S1A. Normally, the paternal X chromosome initially undergoes imprinted XCI, which is reversed in the epiblast cells of the preimplantation blastocyst to allow subsequent random XCI. The intron 1 region has been implicated to be important for Xi reactivation in the ICM, and thus, if the absence of intron 1 prevents reactivation of imprinted XCI, we may observe nonrandom XCI in the adult mouse (Navarro et al., 2008).

However, we did not find differences in allele-specific expression pattern in the presence and absence of intron 1 in heterozygous female mice (Figures 5C–5E). As expected, the C57BL/6 Xist allele is more often expressed than the CAST/Ei X, consistent with a modifier effect, likely resulting in more cells with an inactivated C57BL/6 X (Cattanach and Isaacson, 1967). Because of the stochastic and clonal nature of XCI patterns in the adult mouse, variations in skewing toward *Xist* RNA from the C57BL/6 allele ranged from 50% to 90% (Figures 5C and 5E). Notably, we did not see a preference of *Xist* upregulation on the intron 1-deleted X chromosome in tissues of the adult mouse in vivo, albeit we observed slightly skewed *Xist* RNA levels in heterozygous-differentiating female ESCs carrying the same mutant intron 1 allele in vitro (Barakat et al., 2011). In agreement with this notion, the X-linked genes *Mecp2* and *G6pdx*, both subject to silencing on the Xi, showed reciprocal and intron 1-independent levels of expression from the C57BL/6 chromosome compared to *Xist*, as would be expected from the fact that the *Xist*-expressing chromosome is more likely to be silent (Figures 5D and 5E). These data suggest that the paternal transmission of the intron 1 mutation does not interfere with reactivation of imprinted XCI and subsequent random XCI. A reverse cross in which the maternal allele lacked intron 1 also resulted in random XCI (data not shown). In summary, the intron 1 genomic region is dispensable in the mouse and does not critically control *Xist* expression and skewing of XCI in vivo.

### Intron 1 Is Not Required for Loss of *Xist* RNA upon Reprogramming to iPSCs

Although there was no dramatic effect on XCI state in vivo, we sought to understand the requirement for intron 1 in *Xist* silencing associated with reprogramming to iPSCs. We have shown previously that female iPSCs derived from mouse embryonic fibroblasts (MEFs) carry two active X chromosomes, where *Xist* is efficiently repressed and *Tsix* upregulated, as seen in mouse ESCs (Maherali et al., 2007). Another study suggested that Xi reactivation occurs late in reprogramming at around the time pluripotency genes become expressed, again suggesting that pluripotency transcription factors could contribute to Xi reactivation and the silencing of *Xist*, potentially via binding to intron 1 (Stadtfeld et al., 2008). To test the role of intron 1 in the *Xist*-silencing process during reprogramming, we bred male mice carrying the 2lox intron 1 allele (obtained upon blastocyst injection of our male 2lox ESCs described in Figure 1, Minkovsky/Plath allele) with female mice heterozygous for a *Xist* knockout allele (Marahrens et al., 1997), yielding female XX:2lox intron/*ΔXist* MEFs. Due to the presence of the *Xist* knockout allele, the X chromosome bearing the conditional intron 1 allele is exclusively inactivated in vivo by normal developmental mechanisms (Marahrens et al., 1998). MEFs isolated from E14.5 embryos had uniform *Xist* coating (Figure 6C) and were transduced with retroviruses encoding the reprogramming factors Oct4, Sox2, and Klf4, and subsequently infected with adenovirus encoding Cre recombinase at day 4 of reprogramming to efficiently delete the intron 1 region or with titer-matched empty adenovirus in control samples (Figure 6A). This

A

Minkovsky/Plath Allele

| Paternal Genotype | Maternal Genotype | Average Litter Size | Offspring ♂ | | Offspring ♀ | | p-Value |
|---|---|---|---|---|---|---|---|
| XY:+ | XX:1lox/+ | 7.2 | XY:+ | XY:1lox | XX:+/+ | XX:1lox/+ | >0.25[a] |
| | | | 10 | 10 | 6 | 10 | |
| XY:1lox | XX:1lox/+ | 6.7 | XY:+ | XY:1lox | XX:1lox/+ | XX:1lox/1lox | >0.25[a] |
| | | | 10 | 12 | 14 | 11 | |
| XY:1lox | XX:1lox/1lox | 8.0 | XY:1lox | | XX:1lox/1lox | | >0.25[b] |
| | | | 23 | | 17 | | |

B

Barakat/Gribnau Allele

| Paternal Genotype | Maternal Genotype | Average Litter Size | Offspring ♂ | | Offspring ♀ | | p-Value |
|---|---|---|---|---|---|---|---|
| XY:+ | XX:1lox/+ | 6.7 | XY:+ | XY:1lox | XX:1lox/1lox | XX:1lox/+ | >0.10[a] |
| | | | 2 | 6 | 3 | 9 | |
| XY:1lox | XX:+/+ | 6.5 | XY:+ | | XX:1lox/+ | | >0.25[b] |
| | | | 11 | | 15 | | |
| XY:1lox | XX:1lox/+ | 6.1 | XY:+ | XY:1lox | XX:1lox/+ | XX:1lox/1lox | >0.25[a] |
| | | | 13 | 15 | 8 | 13 | |
| XY:1lox | XX:1lox/1lox | 5.3 | XY:1lox | | XX:1lox/1lox | | >0.25[b] |
| | | | 32 | | 37 | | |

[a-b] Chi-square test for expected 1:1:1:1 (a) or 1:1 (b) ratio

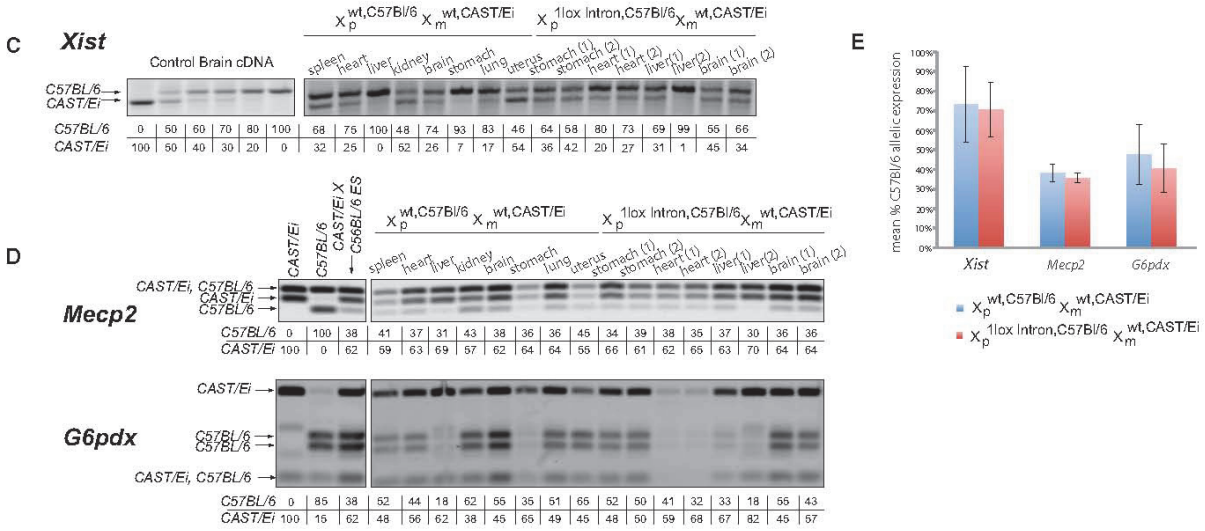



**Figure 5. Transmission of the Intron 1 Mutation In Vivo**

(A) Table summarizing the number and genotypes of offspring from indicated mouse crosses using the intron 1 allele generated in the Plath lab (see Figure 1, Minkovsky/Plath allele).

(B) As in (A), except that mice carrying a second, independent intron 1 deletion generated by the Gribnau lab were crossed (see Figure S1 for comparison of alleles; Barakat et al., 2011).

(C) Allele-specific RT-PCR analysis of *Xist* RNA detecting a length polymorphism that distinguishes *Xist* RNA originating from the C57BL/6 and CAST X chromosome in organs of one female WT mouse and two littermate heterozygous 1lox/WT mice obtained by crossing a C57BL/6 male (with and without the intron 1 1lox allele) with a WT CAST/Ei female. Panel includes controls on the left mixing pure C57BL/6 and CAST/Ei brain cDNA template in given ratios. Numbers below the tissue samples represent the relative band intensity for the C57BL/6 and CAST/Ei *Xist* allele determined by comparison with the control samples.

(D) Examination of tissues as in (C) for allelic expression of X-linked genes *MeCP2* (top) and *G6pdX* (bottom) by RFLP RT-PCR. Panel includes controls (left) from pure C57BL/6 or CAST/Ei mice as well as RNA isolated from a polymorphic C57BL/6 and CAST/Ei ESC line.

(E) Graph averaging the allele-specific expression data in (C) and (D) across all tissue and mice per genotype ± 1 SD.

experimental setup allowed us to test the role of intron 1 in reprogramming efficiency for the same infected fibroblast population. Genotyping confirmed that Ad-Cre addition resulted in efficient deletion of the intron 1 region (Figure 6B). To test whether intron 1 deletion affects the efficiency of reprogramming, we determined the number of Nanog-expressing colonies at day 13 after reprogramming factor introduction because Nanog expression has been shown to mark faithfully reprogrammed cells in retroviral reprogramming experiments (Maherali et al., 2007). We found a comparable number of Nanog+ colonies in the presence and absence of intron 1 (Figure 6D). Normally, at this point of reprogramming, *Xist* RNA coating is just lost in Nanog+ cells (K.P. and J. Tchieu, unpublished data). In agreement with this notion, an examination of all Nanog+ cells for

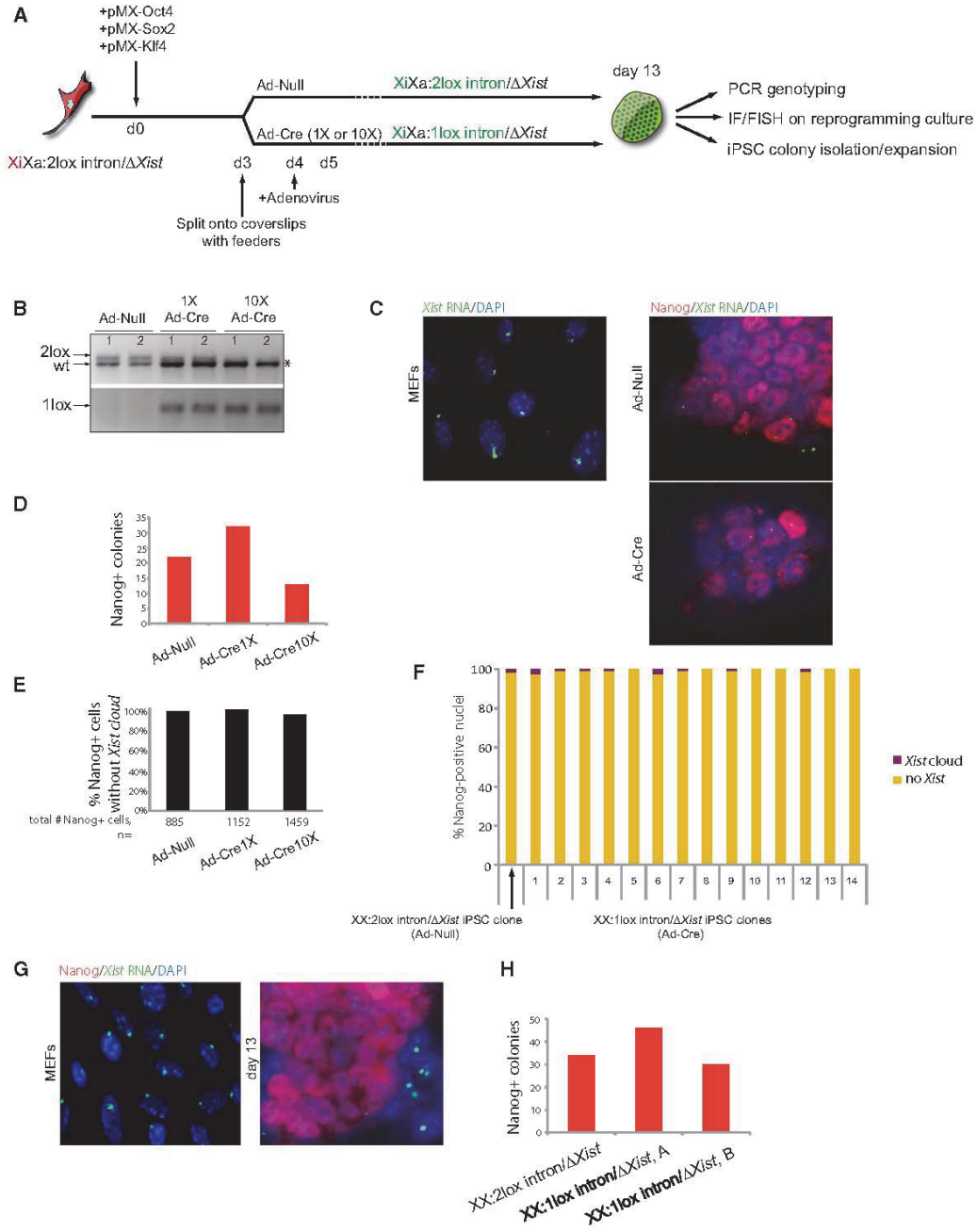



**Figure 6. The Absence of Intron 1 on the Xi Does Not Interfere with Loss of *Xist* RNA Coating upon Reprogramming of MEFs to iPSCs**

(A) Schematic representation of the reprogramming experiment with female MEFs bearing the conditional intron 1 allele on the Xi and a *Xist* knockout allele on the Xa. Reprogramming was induced by infection with pMX retroviruses encoding the reprogramming factors, and the reprogramming culture was split at day 3 postinfection. Deletion of the conditional intron 1 allele was induced by delivery of 1× or 10× adenoviral particles carrying Cre recombinase, performed at day 4. Control 1× Ad Null treatment was done in parallel. At day 13 of reprogramming, efficient deletion of intron 1 was assessed by genotyping, reprogramming efficiency was determined by Nanog+ colony count, and loss of *Xist* RNA coating in Nanog+ cells was examined by IF/FISH. In addition, individual colonies were picked, expanded, and analyzed further.

(B) PCR genotyping for the presence of the 2lox and 1lox intron 1 alleles in reprogramming cultures at day 13, using primers pairs A and C (top panel) or B and C (bottom panel) (as in Figure 1F), indicates efficient deletion of intron 1 upon Ad-Cre treatment. "1" and "2" represent independent reprogramming samples. The asterisk marks the WT allele, which is attributed to the presence of feeder cells in reprogramming cultures.

the presence or absence of a *Xist* RNA cloud demonstrated that nearly all Nanog+ cells carrying the 2lox intron 1 allele (Ad Null-reprogramming cultures) lack a *Xist* RNA cloud at day 13 of reprogramming (Figures 6C and 6E). Importantly, even in the absence of intron 1 (Ad-Cre samples), Nanog+ cells displayed loss of the *Xist* RNA cloud (Figures 6C and 6E) and of the Xi-like H3K27me3 focus (data not shown). Furthermore, from the Ad-Cre-treated reprogramming cultures, 14 iPSC clones were isolated and clonally propagated and all confirmed to have lost both intron 1 and the *Xist* RNA cloud, demonstrating the efficient deletion of the intronic sequence early in reprogramming (Figure 6F). To ensure that the ability of an intron 1-deleted Xi chromosome to downregulate *Xist* was not due to intron 1-dependent events occurring within the first 4 days of reprogramming, i.e., prior to Cre-mediated deletion, we also reprogrammed MEFs carrying a germline-transmitted 1lox intron allele. These XX:1lox intron/*ΔXist* MEFs displayed normal *Xist* RNA coating before reprogramming (detectable in 95% of the cells) and lost *Xist* RNA in Nanog+ colonies (Figure 6G). When comparing to XX:2lox intron/*ΔXist* MEFs, MEFs lacking intron 1 form Nanog+ colonies with similar efficiencies (Figure 6H). Together, these studies rule out that *Xist* intron 1 is necessary for the downregulation of *Xist* in reprogramming to pluripotency.

## DISCUSSION

In summary, our data argue that *Xist* intron 1 does not represent an essential tether-coupling repression of both *Xist* and XCI to the pluripotent state. ESCs lacking intron 1 do not dysregulate *Xist* expression in the undifferentiated state nor upon in vitro differentiation, reprogramming to the iPSC state leads to *Xist* repression on a Xi lacking intron 1, and mice lacking intron 1 do not display any of the gross reproductive abnormalities that would be expected if XCI was perturbed.

The deletion of intron 1 represents a clean experimental system to probe the functional role of a genomic element that displays very strong pluripotency transcription factor binding, unhampered by the secondary effects on initiation of XCI associated with global modulation of protein factors implicated in the maintenance of the pluripotent state. Although correlative binding studies were supported in part by *Xist* dysregulation in ESC lines with inducible deletions of the pluripotency factors Nanog and Oct4, our study cautions against extrapolating these findings to the behavior of WT ESCs and mice. In the case of the ZHBTc4 Oct4-repressable cell line, a compromised pluripotency network may result in Rnf12 upregulation followed by downregulation of the pluripotency factor Rex1, sufficient to trigger XCI in male cells independent of intron 1 (Barakat et al., 2011; Gontan et al., 2012). We also noted that ZHBTc4 ESCs lack pinpoint *Tsix* signal and draw a corollary between their *Xist* upregulation and our male ESCs deleted for *Tsix* (that, when differentiating, have a significantly greater number of *Xist* clouds upon deletion of intron 1).

In light of the two mild phenotypes (skewing effect of deleting intron 1 in female ESCs heterozygous for the allele and the slight increase in *Xist* clouds in *Tsix* and intron 1-deleted differentiating male ESCs), we hypothesize that intron 1 loss leads to mild destabilization of *Xist* transcriptional repression at the transition to the differentiated state, in the narrow development window of XCI initiation. Unable to capture a transcriptional difference in *Xist* levels at the onset of in vitro differentiation, we believe that more sensitive methods of transcript quantitation or investigation of chromatin state may address this hypothesis.

We noted a discrepancy between the ex vivo XCI-skewing phenotype and the normally occurring in vivo XCI choice in the absence of the intron. The lack of an intron 1 deletion effect in adult mice and in ESC differentiation induced by bFGF/Activin (Figure S5), which is sensitive to clonogenic skewing of XCI because of serial passage and outgrowth of few cells (unlike monolayer differentiation; Chenoweth and Tesar, 2010), suggests that *Xist* regulation is more robust in vivo than in vitro in the absence of the intron 1. For instance, slightly different *cis*-acting elements could be used in vivo and in vitro for regulating *Xist* expression. Thus, the cell culture-observed favoring of the intron-deleted *Xist* could not be organismally relevant, or the stochastic developmental nature of XCI could overshadow the effect.

It seems that the regulation of *Xist*, at the helm of a chromosome-wide program of gene expression, is genetically ensured by a complex multifactor mechanism. The dispensability of intron 1 for repression of *Xist* may be mouse specific because mice appear to be unique in the functionality of *Tsix* and also in the sufficiency of *Xist* activators such as Rnf12 to elicit *Xist* upregulation: addition of one copy of Rnf12 is sufficient to drive *Xist* expression in undifferentiated female ESCs (Jonkers et al., 2009). Other eutherians such as bovines and humans, with truncated and likely nonfunctional *TSIX*, may rely more on intron 1-dependent mechanisms for *Xist* repression (Chureau et al., 2002). Therefore, evolution of the overlapping *Tsix* gene and

(C) FISH of starting MEFs before introduction of pMX retrovirus displaying *Xist* RNA coating (left) and immunostaining/FISH images (right) of representative Nanog+ colonies in reprogramming cultures treated with Ad Null and Ad-Cre, respectively, at day 13 of reprogramming, showing Nanog expression (red), FISH for *Xist* RNA using a DNA probe (green), and DAPI (blue). Note that Nanog+ cells at this stage of reprogramming display a biallelic pinpoint signal when using the double-stranded DNA probe, which can be attributed to *Tsix* expression.

(D) Graph summarizing reprogramming efficiency by counting Nanog+ colonies at day 13.

(E) Graph showing the percentage of Nanog+ cells without a *Xist* RNA cloud at day 13 of reprogramming. All Nanog+ cells (number is given) on the reprogramming culture coverslip were counted and analyzed for the *Xist* signal.

(F) The graph summarizes the percentage of Nanog+ nuclei with and without *Xist* RNA clouds in individually expanded iPSC clones from Ad Null or Ad-Cre reprogramming cultures (200 nuclei counted for each iPSC line). Genotyping of all iPSC clones confirmed that intron 1 was deleted in all iPSCs expanded from Ad-Cre reprogramming cultures.

(G) MEFs were obtained from XX:1lox intron/*ΔXist* embryos and reprogrammed with Oct4, Sox2, and Klf4. Immunostaining/FISH images show the presence of normal *Xist* RNA coating in the starting MEFs and the absence of *Xist* RNA coating in resulting Nanog+ colonies at day 13 of reprogramming.

(H) Graph showing counts of Nanog+ colonies at day 13 of reprogramming for two different XX:1lox intron/*ΔXist* MEF preparations (A and B, from different matings) and one XX:2lox intron/*ΔXist* line that was reprogrammed in parallel.

the network of XCI activators in mice may have become the dominant mechanism in *Xist* repression.

## EXPERIMENTAL PROCEDURES

### Generation of Mutant mESCs and Mice

*Xist* intron 1 transgenic mice analyzed in Figures 5B–5E were generated from polymorphic $X^{Xist:2loxneo\ intron\ (129/Sv)}$ $X^{Xist:\ WT\ (CAST/Ei)}$ ESC line 29, in which a 1.8 kb region of *Xist* intron 1 was replaced by a floxed neomycin cassette (Barakat et al., 2011). Germline transmission was verified by genotyping for the presence of the neomycin cassette integrated in the intron 1 region of *Xist*, and XX:2lox-neo/WT females were bred to males expressing pCAGGS-Cre, to loop out the selection cassette. Loopout of the selection cassette was verified by PCR on genomic tail tip-derived DNA. All other intron 1-mutant ESC lines and mice carrying the Plath/Minkovsky allele were derived from a targeting construct generated by cloning the respective genomic fragments representing the 5′ and 3′ homology regions into the pCRII plasmid vector upon PCR amplification (see Table S1 for list of primers used). The 800 bp of intron sequence with a 5′ loxP site was ligated between a 2.2 kb 5′ homology arm and 3′ 2.6 kb homology arm by AgeI/NotI subcloning. A positive-negative CMV-HygroTK selection cassette flanked by loxP sites was inserted into the unique NotI site. A diphtheria toxin gene (PGK-DTA) was inserted into a unique backbone EcoRI site. A total of 40 μg of plasmid was linearized by MluI digestion and electroporated into male ESCs (V6.5 line; F1 between C57BL/6 and 129SV/Jae) and into female F1 2-1 ESCs carrying the MS2 tag in the final large exon of *Xist* (F1 between C57BL/6 and CAST/Ei) cells cocultured with drug-resistant DR4 MEFs (Jonkers et al., 2008; Tucker et al., 1997). Hygromycin selection (140 μg/ml) was started 1 day after, and clones were screened by SpeI/KpnI digest and both 5′ and 3′ external probes. BmtI digest and 3′ external probe were used to assess allelism of targeting in F1 2-1 clones. Targeting efficiency was 30% in V6.5 and 1% in F1 2-1 cells. Two independent V6.5 and one F1 2-1 clones were expanded, electroporated with pPAC-Cre plasmid, and selected with G418 (300 μg/ml) for 8 days. Southern blot screening was performed with a 5′ probe and XbaI digest for 1lox and SpeI/KpnI for 2lox clones. All subsequent intron 1 genotyping was performed by PCR. For intron 1/*Tsix*-Stop double-transgenic ESC clones, XY:2lox and XY:1lox V6.5 clones were targeted with pAA2Δ1.7 and screened by Southern blot as previously described by Sado et al. (2001). XY:1lox and XY:2lox V6.5 ESCs were microinjected into C57BL/6 blastocysts to produce chimeric mice following standard procedures. High-agouti coat color-contributing chimeras were bred with C57BL/6 females for germline transmission. All animal experiments were in accordance with the legislation of the Erasmus MC Animal Experimental Commission and the UCLA Animal Research Committee.

### Cell Culture, Differentiation, and Reprogramming Methods

ESCs were grown on irradiated DR4 MEFs in standard media (DMEM supplemented with 15% FBS, nonessential amino acids, L-glutamine, penicillin-streptomycin, β-mercaptoethanol, and 1,000 U/ml LIF). Prior to induction of RA differentiation, cells were feeder depleted for 45 min on gelatinized plates and plated at a density of $5.0 \times 10^4$ cells/6-well in MEF media (same as ESC media except 10% FBS and excluding LIF). One day later, MEF medium was supplemented with 1 μM all *trans* RA (Sigma-Aldrich) or with DMSO only (LIF withdrawal) and refreshed every 2 days. For EB differentiation, ESCs were preplated on gelatin overnight to feeder deplete, briefly trypsinized, and put in MEF media for suspension culture on bacterial culture plates for 4 days, then plated on gelatinized coverslips for another 2 or 6 days. For FGF/Activin differentiation, ESCs were feeder depleted and $2.0 \times 10^4$ cells plated on six wells pretreated with fibronectin in DMEMF12/B-27/N-2 (Invitrogen) supplemented with FGF-2 (R&D Systems; 40 ng/ml) and Activin A (PeproTech; 20 ng/ml). Medium was changed daily, and colonies were manually passaged onto fibronectin several times then at passage 4 returned to feeder cells. ZHBTc4 ESCs were induced to differentiate with 1 μg/ml doxycycline (resulting in acute repression of Oct4) in standard ESC media (Niwa et al., 2000). For reprogramming, primary MEFs were derived at E14.5, and three-factor retroviral reprogramming was performed following previously published methods by Maherali et al. (2007).

### ChIP

ChIP was performed according to previously published methods by Maherali et al. (2007). In summary, formaldehyde-crosslinked chromatin fragments were generated by sonication, and 150 μg of material was precleared with Protein A Sepharose beads. IP was performed overnight with 5 μg antibodies targeting Oct4 (R&D Systems; AF1759) or Sox2 (R&D Systems; AF2018), or with normal goat IgG (Santa Cruz Biotechnology; sc-2028) and subsequent incubation with protein A Sepharose beads for 3 hr. Beads were washed and eluted in TE/0.67% SDS. Both IP and input samples were reverse cross-linked overnight at 65°C and treated with RNase A and Proteinase K before DNA phenol-chloroform purification. The proportion of input material immunoprecipitated was calculated using standard curves constructed from input serial dilutions and comparing fractional measurements in IP and input relative to a known region positive for Oct4 and Sox2 binding (Rest; van den Berg et al., 2008). ChIP with goat IgG antibody did not find any enrichment (data not shown).

### Immunofluorescence and FISH Analysis

Cells were plated on glass coverslips (and in the case of EB differentiation, permeabilized with 5 min washes of ice-cold CSK buffer, followed by CSK buffer with 0.5% Triton X-100, and another wash in CSK buffer, washed once with PBS, and fixed for 10 min in 4% paraformaldehyde) (Plath et al., 2003). Immunostaining with antibodies against Nanog (BD Pharmingen; 560259) and H3K27me3 (Active Motif 39155) and combined immunostaining/FISH with double-strand *Xist* DNA probe labeled with FITC were performed as previously reported and mounted with Prolong Gold reagent with DAPI (Tchieu et al., 2010). *Xist* and *Tsix* strand-specific RNA probes were made by in vitro transcription of T3-ligated PCR products of cDNA templates using Riboprobe system T3 (Promega) with Cy3-CTP (VWR) or FITC-UTP (PerklinElmer) (Maherali et al., 2007).

### qRT-PCR Analysis and Allele-Specific qRT-PCR

Cells were harvested from a 6-well format in TRIzol (Invitrogen), and RNA purification was performed with the RNeasy kit (QIAGEN) according to manufacturer's instructions with on-column DNase treatment (QIAGEN). cDNA was prepared using SuperScript III (Invitrogen) with random hexamers, and qRT-PCR was performed using a Stratagene Mx3000 thermocycler with primers listed in Table S1. Results were normalized to Gapdh by the ΔCt method. To assess XCI skewing in adult mice, parts of organs were collected, snap frozen, and triturated using micropestles in 1 ml of TRIzol reagent. After an additional centrifugation to clear debris, 700 μl was added to 300 μl fresh TRIzol, and RNA was purified following manufacturer's instructions. RNA was reverse transcribed with SuperScript II (Invitrogen) using random hexamers. Allele-specific *Xist* expression was analyzed by RT-PCR amplifying a length polymorphism using primers *Xist* LP 1445 and *Xist* LP 1446. To determine allele-specific X-linked gene expression of *Mecp2* and *G6pdx* primers, MeCP2-*Dde*I-F and R and G6PD-*Scr*FI-F and R were used to amplify respective restriction fragment-length polymorphisms (RFLPs). PCR products were gel purified and digested with the indicated restriction enzymes and analyzed on a 2% agarose gel stained with ethidium bromide. Allele-specific expression was determined by measuring relative band intensities using a Typhoon image scanner and ImageQuant software.

### Luciferase Enhancer Assay

XY:2lox ESCs were transfected by electroporation with 40 μg of one of three BamHI-linearized pgl4.27-cloned constructs carrying different intron fragments (Promega; Table S1) and transferred to hygromycin selection (140 μg/ml) 1 day later. After serial passaging and outgrowth of stable transfectants, $1.0 \times 10^5$ or $2.0 \times 10^4$ ESCs were seeded for differentiation with and without (no LIF) RA for 3 and 5 days and harvested along with $2.0 \times 10^5$ ESCs and measured for luciferase activity with the luciferase assay system (Promega).

## SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.02.018.

## ACKNOWLEDGMENTS

K.P. is supported by the NIH (DP2OD001686 and P01 GM099134), CIRM (RN1-00564 and RB3-05080), and the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA. A.M. is supported by National Research Service Award AG039179. K.P. and A.M. are supported by funds from the Iris Cantor-UCLA Women's Health Center Executive Advisory Board. J.G. was supported by NWO VICI and ERC starting grants. We thank Ying Wang of the UCLA Transgenic Core Facility for blastocyst injections; Takashi Sado for his pAA2Δ2.17 Tsix-targeting construct and helpful input; and Elyse Rankin-Gee, Dana Case, Matthew Denholtz, Rebecca Rojansky, Konstantinos Chronis, Ritchie Ho, Bernadett Papp, and Sanjeet Patel for advice and experimental support.

Received: November 30, 2012
Revised: February 12, 2013
Accepted: February 14, 2013
Published: March 21, 2013

## REFERENCES

Barakat, T.S., Gunhanlar, N., Pardo, C.G., Achame, E.M., Ghazvini, M., Boers, R., Kenter, A., Rentmeester, E., Grootegoed, J.A., and Gribnau, J. (2011). RNF12 activates Xist and is essential for X chromosome inactivation. PLoS Genet. *7*, e1002001.

Brockdorff, N., Ashworth, A., Kay, G.F., Cooper, P., Smith, S., McCabe, V.M., Norris, D.P., Penny, G.D., Patel, D., and Rastan, S. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature *351*, 329–331.

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38–44.

Cattanach, B.M., and Isaacson, J.H. (1967). Controlling elements in the mouse X chromosome. Genetics *57*, 331–346.

Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes Dev. *20*, 2223–2237.

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

Chenoweth, J.G., and Tesar, P.J. (2010). Isolation and maintenance of mouse epiblast stem cells. Methods Mol. Biol. *636*, 25–44.

Chow, J., and Heard, E. (2009). X inactivation and the complexities of silencing a sex chromosome. Curr. Opin. Cell Biol. *21*, 359–366.

Chureau, C., Prissette, M., Bourdet, A., Barbe, V., Cattolico, L., Jones, L., Eggen, A., Avner, P., and Duret, L. (2002). Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. Genome Res. *12*, 894–908.

Chuva de Sousa Lopes, S.M., Hayashi, K., Shovlin, T.C., Mifsud, W., Surani, M.A., and McLaren, A. (2008). X chromosome activity in mouse XX primordial germ cells. PLoS Genet. *4*, e30.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.

de Napoles, M., Nesterova, T., and Brockdorff, N. (2007). Early loss of Xist RNA expression and inactive X chromosome associated chromatin modification in developing primordial germ cells. PLoS One *2*, e860.

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

Donohoe, M.E., Silva, S.S., Pinter, S.F., Xu, N., and Lee, J.T. (2009). The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. Nature *460*, 128–132.

Eggan, K., Akutsu, H., Hochedlinger, K., Rideout, W., 3rd, Yanagimachi, R., and Jaenisch, R. (2000). X-Chromosome inactivation in cloned mouse embryos. Science *290*, 1578–1581.

Erwin, J.A., del Rosario, B., Payer, B., and Lee, J.T. (2012). An ex vivo model for imprinting: mutually exclusive binding of Cdx2 and Oct4 as a switch for imprinted and random X-inactivation. Genetics *192*, 857–868.

Gontan, C., Achame, E.M., Demmers, J., Barakat, T.S., Rentmeester, E., van IJcken, W., Grootegoed, J.A., and Gribnau, J. (2012). RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. Nature *485*, 386–390.

Huynh, K.D., and Lee, J.T. (2003). Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos. Nature *426*, 857–862.

Jonkers, I., Monkhorst, K., Rentmeester, E., Grootegoed, J.A., Grosveld, F., and Gribnau, J. (2008). Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. Mol. Cell. Biol. *28*, 5583–5594.

Jonkers, I., Barakat, T.S., Achame, E.M., Monkhorst, K., Kenter, A., Rentmeester, E., Grosveld, F., Grootegoed, J.A., and Gribnau, J. (2009). RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. Cell *139*, 999–1011.

Kalantry, S., Purushothaman, S., Bowen, R.B., Starmer, J., and Magnuson, T. (2009). Evidence of Xist RNA-independent initiation of mouse imprinted X-chromosome inactivation. Nature *460*, 647–651.

Kay, G.F., Penny, G.D., Patel, D., Ashworth, A., Brockdorff, N., and Rastan, S. (1993). Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. Cell *72*, 171–182.

Lee, J.T. (2000). Disruption of imprinted X inactivation by parent-of-origin effects at Tsix. 103, *17*–27.

Lee, J.T., Davidow, L.S., and Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. Nat. Genet. *21*, 400–404.

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

Luikenhuis, S., Wutz, A., and Jaenisch, R. (2001). Antisense transcription through the Xist locus mediates Tsix function in embryonic stem cells. Mol. Cell. Biol. *21*, 8512–8520.

Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A., and Wysocka, J. (2011). Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. Nat. Struct. Mol. Biol. *18*, 120–127.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell *1*, 55–70.

Mak, W., Nesterova, T.B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A.P., and Brockdorff, N. (2004). Reactivation of the paternal X chromosome in early mouse embryos. Science *303*, 666–669.

Marahrens, Y., Panning, B., Dausman, J., Strauss, W., and Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. *11*, 156–166.

Marahrens, Y., Loring, J., and Jaenisch, R. (1998). Role of the Xist gene in X chromosome choosing. Cell *92*, 657–664.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al.

(2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521–533.

Mason, M.J., Plath, K., and Zhou, Q. (2010). Identification of context-dependent motifs by contrasting ChIP binding data. Bioinformatics *26*, 2826–2832.

Minkovsky, A., Patel, S., and Plath, K. (2012). Concise review: pluripotency and the transcriptional inactivation of the female Mammalian X chromosome. Stem Cells *30*, 48–54.

Namekawa, S.H., Payer, B., Huynh, K.D., Jaenisch, R., and Lee, J.T. (2010). Two-step imprinted X inactivation: repeat versus genic silencing in the mouse. Mol. Cell. Biol. *30*, 3187–3205.

Navarro, P., Chambers, I., Karwacki-Neisius, V., Chureau, C., Morey, C., Rougeulle, C., and Avner, P. (2008). Molecular coupling of Xist regulation and pluripotency. Science *321*, 1693–1695.

Navarro, P., Oldfield, A., Legoupi, J., Festuccia, N., Dubois, A., Attia, M., Schoorlemmer, J., Rougeulle, C., Chambers, I., and Avner, P. (2010). Molecular coupling of Tsix regulation and pluripotency. Nature *468*, 457–460.

Navarro, P., Moffat, M., Mullin, N.P., and Chambers, I. (2011). The X-inactivation trans-activator Rnf12 is negatively regulated by pluripotency factors in embryonic stem cells. Hum. Genet. *130*, 255–264.

Nesterova, T.B., Senner, C.E., Schneider, J., Alcayna-Stevens, T., Tattermusch, A., Hemberger, M., and Brockdorff, N. (2011). Pluripotency factor binding and Tsix expression act synergistically to repress Xist in undifferentiated embryonic stem cells. Epigenetics Chromatin *4*, 17.

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

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

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. Science *303*, 644–649.

Patrat, C., Okamoto, I., Diabangouaya, P., Vialon, V., Le Baccon, P., Chow, J., and Heard, E. (2009). Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice. Proc. Natl. Acad. Sci. USA *106*, 5198–5203.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. Nature *379*, 131–137.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131–135.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877–879.

Rastan, S., and Robertson, E.J. (1985). X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. J. Embryol. Exp. Morphol. *90*, 379–388.

Reményi, A., Lins, K., Nissen, L.J., Reinbold, R., Schöler, H.R., and Wilmanns, M. (2003). Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. Genes Dev. *17*, 2048–2059.

Sado, T., Wang, Z., Sasaki, H., and Li, E. (2001). Regulation of imprinted X-chromosome inactivation in mice by Tsix. Development *128*, 1275–1286.

Sado, T., Li, E., and Sasaki, H. (2002). Effect of TSIX disruption on XIST expression in male ES cells. Cytogenet. Genome Res. *99*, 115–118.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H.F.M., Jenuwein, T., Otte, A.P., and Brockdorff, N. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev. Cell *4*, 481–495.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722–737.

Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell Stem Cell *2*, 230–240.

Sugimoto, M., and Abe, K. (2007). X chromosome reactivation initiates in nascent primordial germ cells in mice. PLoS Genet. *3*, e116.

Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., and Tada, T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Curr. Biol. *11*, 1553–1558.

Tchieu, J., Kuoy, E., Chin, M.H., Trinh, H., Patterson, M., Sherman, S.P., Aimiuwu, O., Lindgren, A., Hakimian, S., Zack, J.A., et al. (2010). Female human iPSCs retain an inactive X chromosome. *7*, 329–342.

Tsai, C.L., Rowntree, R.K., Cohen, D.E., and Lee, J.T. (2008). Higher order chromatin structure at the X-inactivation center via looping DNA. Dev. Biol. *319*, 416–425.

Tucker, K.L., Wang, Y., Dausman, J., and Jaenisch, R. (1997). A transgenic mouse strain expressing four drug-selectable marker genes. Nucleic Acids Res. *25*, 3745–3746.

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

Williams, L.H., Kalantry, S., Starmer, J., and Magnuson, T. (2011). Transcription precedes loss of Xist coating and depletion of H3K27me3 during X-chromosome reprogramming in the mouse inner cell mass. Development *138*, 2049–2057.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol. Cell *5*, 695–705.