UNIVERSIDAD AUTÓNOMA DE MADRID

DEPARTMENT OF MOLECULAR BIOLOGY
FACULTY OF SCIENCES

REGULATORY MECHANISMS OF MAMMALIAN REPLICATION ORIGINS

Karolina Jodkowska

Madrid, 2017

DEPARTMENT OF MOLECULAR BIOLOGY

FACULTY OF SCIENCES

UNIVERSIDAD AUTÓNOMA DE MADRID

REGULATORY MECHANISMS OF MAMMALIAN REPLICATION ORIGINS

Karolina Jodkowska

Graduate in Biology

Thesis Director: Dr Juan Méndez Zunzunegui

Centro Nacional de Investigaciones Oncológicas

Madrid, 2017

SPANISH NATIONAL CANCER RESEARCH CENTRE

Dr Juan Méndez Zunzunegui, Head of the DNA Replication Group at the Molecular Oncology Programme, Spanish National Cancer Research Centre (CNIO),

CERTIFIES:

That Ms. **Karolina Jodkowska**, Graduate in Biology by the University of Warsaw (Poland), has completed her Doctoral Thesis entitled “Regulatory mechanisms of mammalian replication origins” under my supervision and meets the necessary requirements to obtain the PhD degree in Molecular Biosciences. To this purpose, she will defend her doctoral thesis at the Universidad Autónoma de Madrid. I hereby authorize its defense in front of the appropriate Thesis evaluation panel.

I issue this certificate in Madrid on October 14th 2017

Juan Méndez Zunzunegui
Thesis Director

This Thesis was supported by a "La Caixa"/CNIO Fellowship from the International PhD Programme and research grants from the Spanish Ministerio de Economía y Competitividad (BFU2013-49153-P and BFU2016-80402-R).

Acknowledgements

In the first place, I would like to thank Juan Méndez for giving me the opportunity to carry out my PhD project in his laboratory. I deeply appreciate your guidance and commitment. I am very grateful that your doors were always open to discuss scientific questions and doubts. Thank you for all your help with scientific writing, public presentations and teaching me self-discipline in laboratory work. Thank you for your encouragement and support during the development of my thesis.

I want to express my sincere gratitude to our collaborators without whom this work would have been impossible to carry out. I am very thankful to the Functional Organization of the Mammalian Genome Group (CBMSO), especially to María Gómez for the scientific discussions as well as for invaluable help with SNS-seq assay. I am very grateful to Ricardo Almeida for his support with SNS-seq experiments and to José Miguel Fernandez for his help with bioinformatic analysis.

I am deeply grateful to Vera Pancaldi for introducing me to the world of 3D genomics and for adding a completely new angle to our project. I greatly appreciate your dedication as well as all your brilliant ideas and help that I received from you while writing the thesis.

I am very thankful to Osvaldo Graña and Miriam Rubio for all the bioinformatic analysis. Your patience and effort were extremely important for the development of this work. Also I appreciate a lot the input received from David Pisano and Fátima Al-Shahrour as Heads of the CNIO Bioinformatics Unit. I would also like to acknowledge Enrique Carrillo and Luis Quintales, who contributed to the computational analysis.

My gratitude to the members of my Thesis Committee: to Ana Losada, Francisco Antequera and David Santamaría for your advice, ideas and expertise, which contributed to the development of this work.

I would like to thank all of the current and former members of the DNA Replication Group for making my PhD years memorable. I am especially grateful to Sara Rodríguez-Acebes for her constant scientific advice and for all the time that she dedicated to me. Thanks a lot Marcos for your support with DNA fibers and Dani for your help with CDC6/CDC7 project. And of course thanks to both of you for being my neighbours in the “pasillo”. I am also grateful to Sergio for being an exceptional “jefe

de estudiantes". Silvana, Silvia, Sabela, Elena, Patricia, I really appreciate having had the opportunity of spending my time with you in the lab.

I am also thankful to the Chromosome Dynamics Group, especially to Ana Losada for her constructive criticism, her input during the lab meetings and her willingness to help. I would like to thank Ana Cuadrado for her enthusiasm and great ideas and Alek for his support with bioinformatic analysis. Miri, Miguel, Magali, Carmen, Sam, Marta and Dani, I'm very lucky I could share the lab with you!

I would like to acknowledge the invaluable help of CNIO Core Units: to Lola, Ultan, Miguel Ángel from Flow Cytometry; to Manu and Diego from Confocal Microscopy; to Javier and Fernando from Proteomics; to Sagrario Ortega and Carmen from Transgenic Mice. I deeply appreciate your professional assistance.

Last but not least, I am truly grateful to the friends I met during my time in Madrid. Johanne, thank you for your scientific and personal support, your kindness and honesty. Of course I will remember your delicious cakes!

Koszonom Magdi! Thank you for being there whenever I've needed. I will never forget our surfing trips and I hope we will keep on chasing the sun together!

Kokosy kochane! Vero, Terka y Aneta, dziękuję Wam, że zawsze mogę się do Was zwrócić, zarówno w dobrych jak i trudnych chwilach. Dziękuję za wszystkie podróże, które odbyłyśmy razem i za te które odbędziemy. Vero, Ty wiesz dobrze, że bez Ciebie byłoby mi bardzo ciężko przebrnąć przez ten etap. Ďakujem!

Kuba, dziękuję za zaprojektowanie okładki!

Na koniec chciałabym podziękować całej mojej rodzinie i przyjaciołom z Polski, za ciągłe wsparcie, pomimo odległości jaka nas dzieliła. W szczególności jestem wdzięczna mojej Mamie i mojemu Tacie za to, że zawsze są kiedy ich potrzebuję i za nieskończoną cierpliwość. Dziękuję Kubie i Ali, na których zawsze mogę liczyć i którzy mnie motywowali w trudnych chwilach. Dziękuję!

Moim Rodzicom

Abstract/Resumen

Abstract

In mammalian cells, DNA replication starts from thousands of replication origins whose activation is tightly regulated in the cell division cycle. In this study, we have aimed at better understanding the regulation of origin selection and activation in mouse embryonic stem cells and human cancer cells.

In the first part of this dissertation, we set out to investigate the dynamics of origin activation in response to stress conditions that slow down or stall replication forks. This situation promotes the activation of otherwise 'dormant' origins, which provide a backup mechanism to complete replication and prevent genomic instability. Using the SNS-Seq technique based on deep-sequencing of short nascent DNA strands, we have mapped the genomic positions of origins in control growth conditions and two experimental settings that trigger the activation of extra origins: (a) exposure to DNA polymerase inhibitor aphidicolin; (b) overexpression of CDC6, a limiting factor for origin licensing and activation in primary murine cells. Using SNS-Seq data we also determined the efficiency of activation of each individual origin in the cell population. Constitutive origins, in contrast to stress-responsive ones, show a strong preference towards open, transcriptionally active chromatin and display higher efficiency. Our results strongly suggest that the main response to stress is mediated by modulating the activity of pre-existing origins rather than the activation of new ones. We have also carried out an unprecedented integration of linear origin maps into 3D chromatin networks that reveals how origins tend to group together in clusters that likely correspond to DNA replication factories. Origin connections are found within the same topologically associated domain (TAD), but also between origins located in different TADs. We report for the first time that the connectivity of an origin is directly proportional to its efficiency of activation.

In the second part, we aimed at dissecting a novel mechanism that regulates CDC6 protein stability. Downregulation of CDC7 kinase, known to activate the MCM helicase at replication origins, caused a drop in cellular CDC6 levels. CDC6 was phosphorylated by CDC7 *in vitro* and became destabilized *in vivo* when all possible CDC7-dependent phosphorylation sites were mutated. We report a previously unknown role of CDC7 in the regulation of CDC6 stability that is mediated by a combination of direct and indirect mechanisms.

Resumen

En células de mamífero, la replicación del DNA comienza en miles de puntos conocidos como orígenes de replicación, cuya actividad está regulada en el ciclo celular. En este trabajo, nuestro objetivo ha sido entender la regulación de la selección y activación de orígenes en células embrionarias pluripotentes de ratón.

En la primera parte de la Tesis, hemos investigado la dinámica de activación de orígenes en respuesta a situaciones que ralentizan o detienen las horquillas de replicación. En esta situación se activan orígenes "silentes" para completar la replicación y evitar la inestabilidad genómica. Mediante la técnica de SNS-Seq, basada en la secuenciación masiva de cadenas de DNA naciente, hemos determinado las posiciones genómicas de los orígenes en células control y en células sometidas a tratamientos que causan la activación de orígenes silentes: (a) afidicolina, un inhibidor de DNA polimerasas; (b) sobre-expresión de CDC6, un factor limitante para la activación de orígenes en células primarias. Además, hemos calculado la eficiencia de activación de cada origen dentro de la población celular. Los orígenes constitutivos correlacionan preferentemente con zonas de cromatina abierta, transcripcionalmente activa, y tienen mayor eficiencia. Nuestros resultados sugieren que la principal respuesta a estrés consiste en aumentar la eficiencia de activación de orígenes pre-existentes, más que en la activación de orígenes nuevos. También hemos integrado los mapas de orígenes en redes tridimensionales de cromatina, encontrando que los orígenes tienen tendencia a agruparse en 3D, formando "factorías de replicación". Si bien muchas de las conexiones entre orígenes se localizan dentro de una misma zona TAD ("dominio de asociación topológica") en la cromatina, también existen conexiones entre orígenes intra-TAD. En este trabajo describimos por primera vez que la conectividad de un origen con otros orígenes es directamente proporcional a su eficiencia de activación.

En la segunda parte, hemos estudiado un nuevo mecanismo de control sobre la estabilidad de la proteína CDC6. La pérdida de expresión o actividad de la quinasa CDC7 causa una disminución en los niveles celulares de CDC6. Hemos demostrado que CDC6 es sustrato de CDC7 *in vitro*, y que la mutación de todos los posibles sitios de fosforilación por CDC7 resulta en una proteína CDC6 con vida media más corta *in vivo*. Nuestros datos indican la existencia de una función hasta ahora desconocida de CDC7 en la regulación de CDC6, que está mediada por una combinación de efectos directos e indirectos.

Table of contents

Table of contents

Abbreviations

Abbreviations

Aph	Aphidicolin
APC/C	Anaphase promoting complex-cyclosome
BrdU	5-bromo-2'-deoxyuridine
Cdc7i	PHA-767491, an inhibitor of Cdc7 kinase
CDK	Cyclin-dependent protein kinase
CFS	Common fragile site
CGI	CpG island
ChAs	Chromatin assortativity
ChIA-PET	Chromatin interaction analysis with paired-end tagging
ChIP-seq	Chromatin immunoprecipitation and high-throughput sequencing
CHX	Cycloheximide
CKI	CDK inhibitors
CldU	5-Chloro-2'-deoxyuridine
CMG	Cdc45-MCM-GINS
DDK	DBF4-dependent CDC7 kinase
Dox	Doxycycline
DDR	DNA damage response
DSB	Double strand breaks
ESCs	Embryonic stem cells
FACS	Fluorescence activated cell sorting
FBS	Foetal bovine serum
FR	Fork rate
G4	G-quadruplex
Hi-C	High-throughput variation of 3C
LADs	Lamin B1-associated domains
MCM	Mini-chromosome maintenance complex
MEFs	Mouse embryonic fibroblasts
mESCs	Mouse embryonic stem cells
O	Other ends (in "P-O" network)
ORC	Origin Recognition Complex
ORIs	Origins of replication initiation
OriAs	Origin assortativity
OriEfAs	Assortativity of origin efficiency
P	promoter ends (in "P-P" and "P-O" networks)

PcG	Polycomb Group Proteins
PCHiC	Promoter-Capture HiC
PI	Propidium iodide
PIP	PCNA-interacting protein
RNAPII	RNA Polymerase II
Pre-RC	Pre-Replication Complex
RD	Replication domains
RS	Replication stress
RT	Room temperature
SD	Standard deviation
siRNA	Short interfering RNA
SNS	Short nascent strands
SNS-seq	SNS assay followed by high-throughput sequencing
TADs	Topologically associated domains
TSS	Transcription start site
TTS	Transcription termination site
RT-qPCR	Quantitative reverse transcription PCR
WT	Wild type
λ-exo	λ-exonuclease
3C	Chromatin conformation capture

Note about nomenclature. The standard HUGO nomenclature for human, mouse and yeast genes and proteins has been applied depending on the context.

Introduction

Introduction

Initiation of eukaryotic DNA replication: regulatory principles

In order to transmit genetic information from one cell generation to the next, cells prepare two identical copies of their genetic material before each cell division. The process of duplication of a genome is called **DNA replication**, and needs to be integrated with nucleosome assembly and restoration of chromatin structure (reviewed by Alabert and Groth, 2012). Because of the large size of eukaryotic genomes, which are organized in long linear chromosomes, DNA replication starts at thousands of sites called "**replication origins**" that are progressively activated throughout the S phase of the cell cycle. Accurate regulation of replication origins is fundamental for genomic stability. Indeed, under- or over-replication of chromosomal DNA may have deleterious consequences for the cells and lead to several human diseases, including cancer (reviewed by Boyer et al., 2016).

Initiation of DNA replication

Replication origins are recognized by **initiator proteins**, whose main role is to recruit the DNA helicase as a first step to assemble the '**replisome**' machinery (reviewed by Deegan and Diffley, 2016; Gilbert, 2004). The eukaryotic replicative helicase is a ring-shaped hexameric complex formed by six mini-chromosome maintenance proteins (MCM2-7), and is commonly referred to as **MCM**. The role of the MCM helicase is to unwind the double helix in order to provide the DNA primase and polymerases with single-stranded DNA (ssDNA) template.

Initiation of DNA replication is separated into two steps, which are coordinated with cell cycle progression. The first step is called '**origin licensing**' and consists in the loading of MCM helicases in inactive form. It takes place at late mitosis and during G1. The second step involves helicase activation and replisome assembly and it occurs in S phase (Fig. 1A). The temporal separation of helicase loading and activation ensures that the same origin cannot be activated twice within the same cell cycle (reviewed by Deegan & Diffley 2016; Bleichert et al. 2017; Masai et al. 2010).

Origin licensing

Origins are marked by the **Origin Recognition Complex** (ORC), an AAA+ ATPase formed by six subunits (ORC1-6). Origin licensing starts when ORC recruits another AAA+ ATPase called **CDC6** (Cell Division Cycle 6). Subsequently, ORC and CDC6 cooperate with **CDT1** (CDC10-Dependent Transcript 1) protein to facilitate the

deposition of two MCM complexes onto dsDNA as a head-to-head double hexamer (Evrin et al., 2009; Remus et al., 2009). Besides the ATPase activities of ORC and CDC6, ATP binding and hydrolysis by MCM itself seems to play a crucial role in the loading reaction (Deegan and Diffley, 2016). The assembly of ORC, CDC6, CDT1 and MCM is commonly known as **pre-replication complex** (pre-RC) formation. Pre-RCs are the signature of licensed origins (Fig. 1A). At least in yeast, Cdc6 and Cdt1 are released after the loading reaction, whereas ORC and MCM remain stably bound to the DNA (reviewed by Bleichert et al., 2017; Deegan and Diffley, 2016).

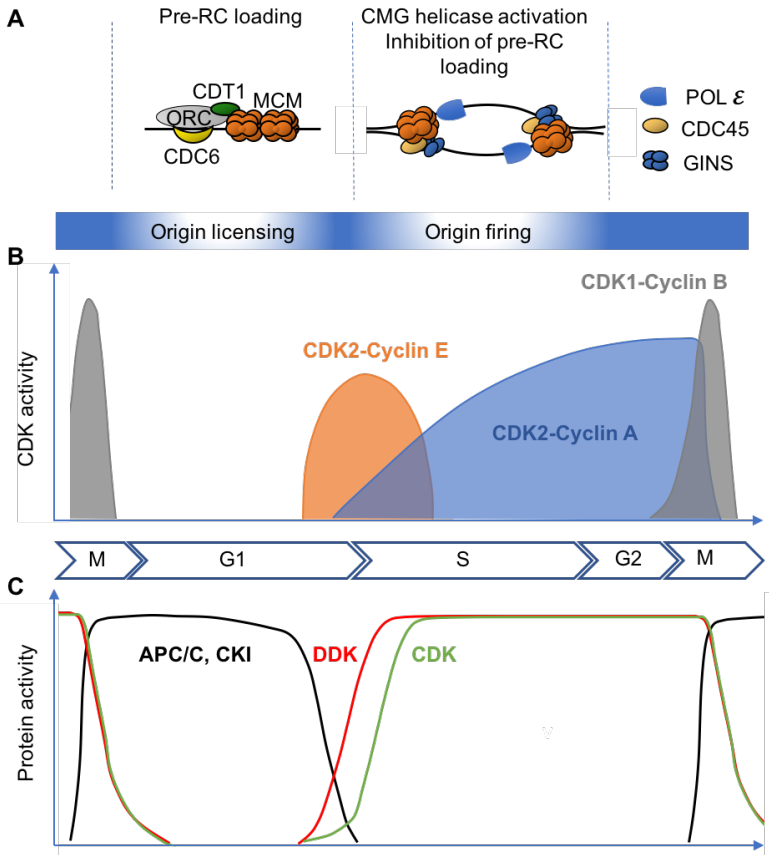


Figure 1. Cell cycle regulation of DNA replication. (A) Schematic of the main proteins that assemble at replication origins during origin licensing and firing. **(B)** Changes in the activity of CDK2–Cyclin E, CDK2-Cyclin A and CDK1-Cyclin B during cell cycle progression. **(C)** Changes in the activity of anaphase promoting complex-cyclosome (APC/C), CDK inhibitors (CKI) as CDK and DDK activity across the cell cycle. Figure inspired from Depamphilis et al. (2012). See text for details.

Helicase activation

Concomitant to origin activation, the MCM double hexamer is remodelled into two separate helicases that move away bidirectionally from the origin, translocating on the leading strand of DNA (Deegan and Diffley, 2016; Tanaka and Araki, 2013). The active form of each replicative helicase is composed of a MCM hexamer and two

accessory proteins called **CDC45** and **GINS** (a tetramer consisting of SLD5, PSF1, and PSF3 subunits). The resultant super-complex is called **CMG** (CDC45-MCM-GINS; Costa et al., 2011; Costantino et al., 2014; Ilves et al., 2010; Moyer et al., 2006).

In budding yeast, helicase activation requires the action of **DDK** (Dbf4-dependent Cdc7 kinase) and **CDK** (cyclin-dependent protein kinase). DDK phosphorylates MCM to facilitate the association of Sld3/7 and Cdc45 proteins to the pre-RC (Heller et al., 2011; Sheu and Stillman, 2006; Tanaka et al., 2011). In turn, CDK phosphorylates Sld2 and Sld3, facilitating their association to yet another factor called Dpb11 (Tanaka et al., 2007; Zegerman and Diffley, 2007). These phosphorylation events promote the formation of a complex composed of Sld2, Dbp11, GINS and DNA polymerase ε (Pol ε), (Heller et al., 2011; Tanaka and Araki, 2013) and the eventual assembly of two CMG complexes around ssDNA. An additional factor called Mcm10 is necessary for CMG activation but not for its assembly (Deursen et al., 2012; Yeeles et al., 2015). Recently, DDK, CDK, Sld3/7, Sld2, Cdc45, Dpb11, GINS, DNA Pol ε and Mcm10 were shown to be the minimal set of factors essential and sufficient for helicase activation *in vitro* (Yeeles et al., 2015). In metazoans, regulation of CMG assembly and activation operates under similar principles, and functional homologs of Dbp11, Sld2 and Sld3 have been identified (TOPBP1, RECQL4 and Treslin, respectively). Still, many details of the process may be slightly different from those described in unicellular organisms (Tanaka and Araki, 2013).

The replisome machinery that performs DNA replication includes three DNA polymerases. Polymerase α-primase is responsible for the generation of short RNA primers and their initial extension with dNTPs. Subsequently, DNA Polymerase ε (POL ε) carries out continuous DNA synthesis in the leading strand whereas DNA Polymerase δ (POL δ) synthetizes DNA at the lagging strand in a discontinuous manner (reviewed by Kunkel and Burgers, 2008).

Cell cycle regulation of DNA replication

The temporal separation between origin licensing and activation prevents origin re-initiation after the G1/S transition (Fig. 1A). This is achieved through the combined action of cell cycle regulators such as cyclin-dependent kinases (CDKs) and ubiquitin ligases (Fig 1B,C, reviewed by Depamphilis et al. 2012).

In mammalian cells, M-phase CDK (CDK1-Cyclin B) activates APC/C^{CDC20} ubiquitin

ligase, which promotes destruction of Cyclins A and B. When the levels of mitotic cyclins are reduced, APC/C^{CDC20} is inhibited and the alternative ubiquitin ligase APC/C^{CDH1} becomes active. APC/C^{CDH1} stabilizes CDK inhibitors (CKI) p21 and p27 (Manchado et al., 2010), effectively maintaining a temporal window of low CDK levels during G1-phase and creating a permissive state for origin licensing. After G1, the accumulation of Cyclin E activates CDK2 and inactivates CDH1, preventing its binding to APC/C. The activity of APC/C is also counteracted in S phase by its inhibitor EMI1 and CDH1 degradation (Manchado et al., 2010). This allows for the accumulation of S-phase CDK (CDK2-CyclinA). The increase in CDK levels at the G1/S transition, together with activation of DDK, promotes origin firing. At the same time, high CDK activity inhibits further origin licensing throughout S, G2 and early M phases (Depamphilis et al., 2012; Manchado et al., 2010).

Cell cycle regulation of DDK

DDK, an evolutionary conserved S/T kinase formed by cell division cycle 7 (Cdc7) protein and its activating subunit Dbf4, was originally identified in budding yeast (Masai and Arai, 2002). DDK activity fluctuates in the cell cycle in response to changes in the abundance of Dbf4, paralleling CDK regulation by oscillating cyclins. Actually, fluctuations of DDK activity in the cell cycle resemble the changes observed for CDKs (Fig. 1 B and 1C). In yeast, APC/C-dependent degradation of Dbf4 in M/G1 results in DDK inactivation. At the G1/S transition, the levels of Dbf4 start to increase leading to restoration of Cdc7 activity, which is maintained during S and G2 (Fig. 1C; Ferreira et al., 2000; Oshiro et al., 1999; Weinreich and Stillman, 1999). The human homolog of Dbf4, also known as ASK (activator of S-phase kinase), is a target of APC/C^{CDH1} as well (Yamada et al, 2013) and undergoes similar changes in the cell cycle (Kumagai et al., 1999). Mammalian DBF4 is transcriptionally regulated by E2F factors (Yamada et al., 2002). Both DDK and CDKs act together to restrict origin firing and DNA synthesis to the S phase.

Genome duplication: only once per cell cycle

In metazoa, at least three different systems ensure that licensing occurs once and only once per cell cycle: (1) CDT1 inhibition by Geminin, (2) destruction of CDT1 by $CLR4^{CDT2}$ and (3) CDK-dependent inhibition of several pre-RC components. These mechanisms provide multiple layers of protection against uncontrolled origin licensing (reviewed by Arias and Walter, 2007; Diffley, 2011; Siddiqui et al., 2013).

Geminin is a small coiled-coil protein that binds and inhibits CDT1 by preventing its interaction with MCM (Lee et al., 2004; Wohlschlegel et al., 2000). Both APC^{CDC20} and APC^{CDH1} mediate Geminin destruction in mitosis (Clijsters et al., 2013; McGarry and Kirschner, 1998), allowing CDT1 to perform its licensing function. Geminin re-accumulates throughout S and G2, sequestering CDT1 to impede uncontrolled pre-RC formation. Reducing the expression of Geminin is sufficient to induce re-replication in different cell lines (Melixetian et al., 2004; Quinn et al., 2001; Zhu et al., 2004). Of note, cancer cell lines seem to be especially vulnerable to geminin ablation (Zhu and DePamphilis, 2009).

Degradation of CDT1 by $CLR4^{CDT2}$ E3 ubiquitin ligase occurs only in S-phase as it depends on the binding of CDT1 to the PCNA (Arias and Walter, 2007). CDT1 binds PCNA via its PCNA-interacting protein (PIP) motif localized at its N-terminus. CLR4 is recruited to PCNA-CDT1 complex through CDT2 (Arias and Walter, 2006; Nishitani et al., 2006). Alterations in this pathway cause re-replication in metazoans (Arias and Walter, 2005; Jin et al., 2006; Lovejoy et al., 2006). In human cells, CDT1 is subject to yet another regulatory mechanism. Phosphorylation of Thr29 by CDK2-CycA is recognized by SKP2, a substrate receptor for SCF^{SKP2} E3 ubiquitin ligase that mediates CDT1 destruction in S and G2 phase (Li et al., 2003; Sugimoto et al., 2004). However, abrogation of this pathway does not cause significant re-replication (Nishitani et al., 2006; Takeda et al., 2005), suggesting that the mechanisms discussed before can regulate CDT1 even in the absence of SCF^{SKP2} activity.

CDK–dependent inhibition of MCM loading is the only system that prevents inappropriate pre-RC assembly in budding yeast. In contrast, this pathway seems to be less important in metazoa (Arias and Walter, 2007; Siddiqui et al., 2013) but it is critical at G2-M phase, when CDK1 inhibition can promote origin relicensing even in the presence of Geminin (Ballabeni et al., 2004; Coverley et al., 1998; Sugimoto et al., 2004). The specific targets of CDK1-mediated inhibition of licensing in G2-M remain largely unknown.

ORC might also be regulated by CDKs in mammalian cells. CDK1 inhibits ORC1 loading on chromatin in mitosis (Li et al., 2004). Moreover, ORC1 chromatin levels fluctuate in the cell cycle, being higher in G1 and lower in S and G2 (Méndez et al., 2002; Tatsumi et al., 2003). This regulation could depend on phosphorylation by CDK2-CycA (Méndez et al., 2002). Moreover, phosphorylation of ORC2 by S-CDK

was shown to promote disassociation of ORC from chromatin and replication origins (Lee et al., 2012).

Complex regulation of CDC6 throughout the cell cycle

CDC6 is one of the crucial targets for inhibiting origin relicensing in yeast. Intriguingly, mammalian CDC6 is subjected to multiple regulatory mechanisms, some of them with seemingly opposing effects. CDC6 levels and subcellular localization are regulated mostly by CDK and APC/C^{CDH1} activity. CDC6 contains three CDK target sites (S54, S74, S106) in its N-terminus, as well as a cyclin-binding motif and D- and KEN-boxes required for APC/C^{CDH1} dependent protein degradation (Borlado and Méndez, 2008). CDC6 is degraded at the M-G1 transition by APC/C^{CDH1} mediated proteolysis and re-accumulates in S, G2 and M phases (Méndez and Stillman, 2000; Petersen et al., 2000). $CLR4^{CDT2}$-dependent destruction of CDC6 at the G1-S transition has also been suggested (Clijsters and Wolthuis, 2014). More recently, CDC6 was shown to be targeted for degradation by $SCF^{CyclinF}$ in G2 and M phase and this regulation appears to prevent re-replication in the absence of Geminin (Walter et al., 2016).

Despite the described proteolytic mechanisms, a significant amount of CDC6 protein is bound to chromatin throughout the cell cycle (Fujita et al, 1999; Coverley et al, 2000; Méndez and Stillman, 2000; Alexandrow and Hamlin, 2004; Oehlmann et al, 2004). It has also been proposed that soluble CDC6 is exported from the nucleus to the cytoplasm after the G1-S transition in a CDK2-CycA dependent manner (Jiang et al., 1999; Petersen et al., 1999; Saha et al., 1998). The data on CDC6 translocation were initially obtained with exogenously expressed protein, although one report notes that it may also apply to endogenous CDC6 (Paolinelli et al., 2009). Considering the amount of CDC6 detected on chromatin, the functional significance of CDC6 subcellular localization remains unclear.

In quiescent cells, CDC6 is degraded by APC/C^{CDH1} (Mailand and Diffley, 2005; Petersen et al., 2000). Upon cell cycle re-entry, CDC6 is protected from APC^{CDH1} mediated proteolysis by phosphorylation by CDK2-CycE, allowing origin licensing (Mainland and Diffley, 2005). Interestingly, activation of p53 upon DNA damage destabilizes CDC6 due to the loss of its protective phosphorylation by CDK2-CycE at S54 (Duursma and Agami, 2005).

Overexpression of CDC6 enhances MCM loading on chromatin and induces higher activity of origins, both *in vitro and in vivo* (Búa et al., 2015; Muñoz et al., 2017). However, deregulation of CDC6 (either by affecting its stability, cellular localization or expression levels) is not sufficient to induce re-replication in mammalian cells (Muñoz et al., 2017; Paolinelli et al., 2009; Petersen et al., 1999, 2000). In line with this, recent reports from our laboratory show that individual overexpression of CDC6 or CDT1 in adult mice have very limited effects (Búa et al, 2015; Muñoz et al, 2017). Only when both proteins are overexpressed simultaneously, DNA re-replication causes lethal dysplasia of the intestinal epithelium and other tissues (Muñoz et al., 2017). This study underscores the importance of having multiple safeguard mechanisms against origin re-firing.

In summary, origin licensing starts at M-G1, when CDT1 is released from Geminin-mediated inhibition and CDC6 is not yet degraded by APC/C. It is somehow paradoxical that the levels of CDC6 are lower in early G1, when pre-RC assembly is permitted, than in the remaining phases of the cell cycle when licensing is inhibited. This regulation suggests that CDC6 may play additional roles. In this regard, depletion of CDC6 during S-phase in HeLa cells affects origin activation and leads to abnormal spindle formation and mitotic defects (Lau et al., 2006). Another report proposed that CDC6 prevents premature mitosis through mechanisms involving CHK1, the checkpoint kinase 1 (Clay-Farrace et al., 2003). Moreover, CDC6 centrosomal localization has been described, suggesting a possible role in mitotic chromosome segregation (Kalfalah et al., 2015; Kim et al., 2015).

Replication origins

According to the **replicon** model proposed over half a century ago, an origin of replication is a specific genetic element (initially called "**replicator**") that is recognized by a diffusible element called "**initiator**" to promote initiation of DNA duplication from this particular site (Jacob et al., 1963; reviewed by Gilbert, 2004). In contrast to bacteria and budding yeast, the metazoan initiator protein ORC does not bind to a specific consensus sequence, hindering our understanding of the positioning and regulation of replication origins. Hence, the nature of mammalian replication origins remains partially elusive, but extensive research has determined that they are specified by a combination of genetic and epigenetic features that shape up chromatin environment and nuclear organization (reviewed by Aladjem and Redon, 2016; Antequera, 2004; Fragkos et al., 2015; Méchali, 2010).

Flexibility in origin choice

In mammalian cells, approximately 50,000 origins are activated in every cell cycle (Cayrou et al., 2011; Huberman and Riggs, 1966). Of note, these activated origins represent only a subset of the many potential origins that are licensed during G1 phase (Gilbert, 2007; Taylor, 1977). It is estimated that cells use 10-20% of the potential origins and the rest is passively replicated by ongoing forks (McIntosh and Blow, 2012; Méchali, 2010). For this reason, DNA replication can occur even after a significant downregulation of MCM proteins in mammalian cells (Ge et al, 2007; Ibarra et al, 2008). An excess of potential origins provides cells with flexibility when it comes to origin choice. Theoretically, replication origins can be grouped into three categories depending on how often (i.e. how efficiently) they are activated within the population (Fig. 2; Fragkos et al. 2015; Aladjem & Redon 2016). In this view, **constitutive** origins are the most efficient, becoming activated in each cell in the population regardless of developmental stage or environmental conditions. **Flexible** origins would be activated in a subset of cells in the population and likely represent the most numerous group (Fragkos et al., 2015). The firing efficiency of each flexible origin is indicated by the percentage of cells in the population that activate it (Fig. 2). The last category consists of **dormant** origins (McIntosh and Blow, 2012), which become activated only when neighbouring forks are stalled (Fig. 2 and 4). It is postulated that the excess of licensed origins gives the cells replicative flexibility in response to changes associated to differentiation, or in response to replication stress (Alvarez et al., 2015; Méchali, 2010). Two major questions remaining in the field are: (1) whether each one of the origin types (constitute, flexible and dormant) have specific characteristics; (2) how a subset of the licensed origins is selected for activation.

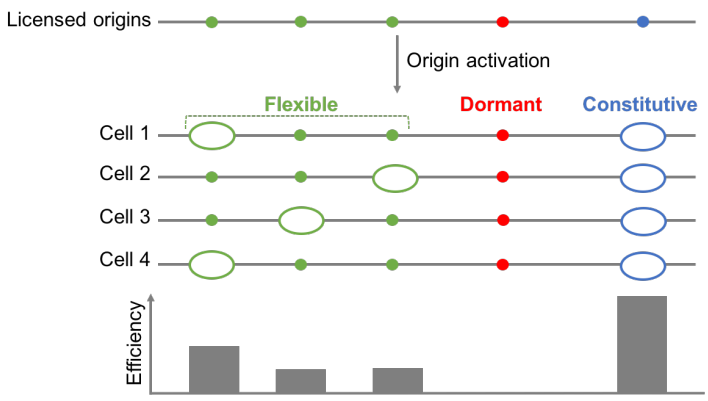


Figure 2. Different types of replication origins according to their usage. See text for details.

Organization of replication origins in the nucleus

The length of DNA replicated by forks generated from a single active origin is defined as a **replicon**. It has been proposed that each replicon contains several flexible origins, only one of which becomes activated while the others are inhibited (Cayrou et al., 2011). In mammalian cells, the size of replicons ranges between 50-150 kb (Cayrou et al., 2011; Jackson and Pombo, 1998). Groups of consecutive replicons can form a **cluster** encompassing 0.5-1 Mb of DNA (Berezney et al., 2000). Origins belonging to the same cluster tend to activate simultaneously in S phase (Jackson and Pombo, 1998; McIntosh and Blow, 2012). Replicons within clusters may form loops that are anchored to a **nuclear matrix** structure, and origins localized at the base of the loops are activated preferentially (Fig. 3 and Courbet et al. 2008). While the selection of an origin within a replicon is believed to be stochastic, activation of the clusters is tightly controlled to follow a spatial and temporal order of genome duplication (Blow et al., 2011).

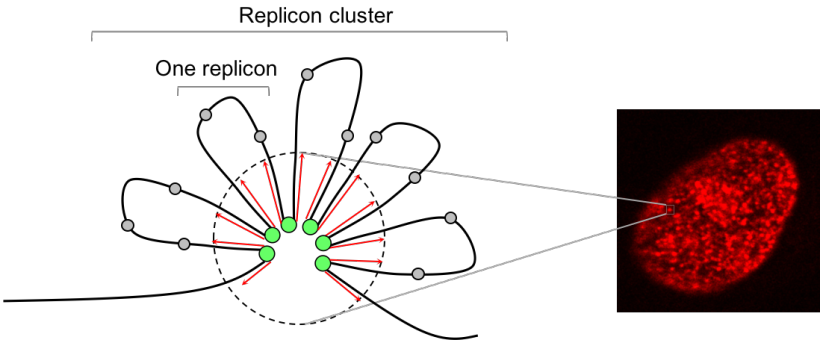


Figure 3. Organization of replication origins into a replication factory. Left, schematic of chromatin folding within a factory, where each loop corresponds to one replicon. Active origins located at the base of the loops are depicted in green. Dormant origins are depicted in grey. Red arrows indicate the directions of fork movement. Right, foci of BrdU incorporation in a HeLa cell nucleus. Each focus could correspond to a factory.

Replication foci

Replicon clusters (or groups of clusters) likely correspond to **replication foci**, which are discrete sites in the nucleus where DNA synthesis takes place (Fig. 3 and Berezney et al. 2000). Replication foci can be visualized microscopically after labelling cells with nucleotide analogues such as BrdU or by immunofluorescence detection of replisome components such as PCNA. Replication foci were proposed to be stable units of chromosome structure with constant dimensions in the cell cycle remain stable throughout many cell generations (Berezney et al., 2000; Jackson and Pombo, 1998).

The pattern of active replication foci changes during S phase (Berezney et al., 2000).

Early replicating sites are distributed throughout the nucleus as numerous, relatively separated, small foci. In mid-S, foci become bigger and localize towards the periphery of nucleoli and nuclear envelope. In late S, fewer, larger foci are formed that probably correspond to constitutive heterochromatin regions (Chagin et al., 2010). A domino-like model has been proposed to explain the progression of replication throughout S phase (Sporbert et al., 2002). In this model, ongoing replication in a particular cluster triggers subsequent initiation of DNA synthesis in neighbouring clusters. Replication would therefore spread from early replicating, easy–accessible euchromatin regions, through facultative heterochromatin to finally reach highly-condensed, constitutive heterochromatin regions in the late S phase (Chagin et al., 2010; Rivera-Mulia and Gilbert, 2016a).

Replication factories

Replication clusters/foci are also referred to in the literature as **replication factories,** a concept that integrates the organization of replicons with the DNA replication protein machinery. By analogy to the concept of transcriptional factories, it was initially suggested that replication factories are static centres, in which DNA synthesis would occur simultaneously at different replicons by template translocation through replication machineries 'fixed' to a nuclear skeleton (Cook, 1999; Hozák et al., 1993; Kitamura et al., 2006). However, recent super-resolution microscopy data support a model in which replication factories consist of a group of separated smaller foci localized in close proximity, each corresponding to DNA synthesis carried out at an individual replicon (Chagin et al., 2016). Spatial proximity of smaller foci would be dictated by the three-dimensional organization of chromatin structure (Chagin et al., 2016; Löb et al., 2016).

In any case, organization of origins into "replication factories" could facilitate local concentration of limiting replication factors and activator kinases at the sites of DNA synthesis. Interestingly, cohesin was proposed to play a role in the spatial organization of replication factories by mediating the formation and stabilization of chromatin loops (Guillou et al., 2010). Downregulation of cohesin did not affect the total number of replication factories but limited the efficiency of origin firing within each one. This effect is likely due to an increase in the size of chromatin loops, leading to a decrease in the number of origins localized at the bases of the loops, where their chance of activation is higher (Guillou et al., 2010).

Despite these antecedents, the connections between replicons, chromatin loops, foci

and replication factories still remain poorly understood. Integration of super-resolution microscopic observations with linear, high-throughput genomic data and the recently available high-resolution, three-dimensional (3D) chromatin interaction maps could provide a better understanding of their structure and function.

Activation of back-up origins upon replication stress

Replication stress (RS) is defined as the slow down or stalling of replication forks in the face of DNA lesions, hard-to-replicate DNA structures, collisions with the transcriptional machinery, and other circumstances. In many of these situations, the functional uncoupling of the helicase and DNA polymerase leads to the formation of stretches of ssDNA that are subsequently coated by RPA. This is sensed by the ATR checkpoint kinase, which is recruited to stalled forks and phosphorylates CHK1 that in turn regulates origin initiation both globally and locally (reviewed by Mcintosh et al., 2012; Muñoz and Méndez, 2017).

Fork stalling poses a threat for genomic stability as it may lead to under-replication of fragments of the genome. If a single fork stalls, the unreplicated DNA fragment will be eventually replicated by a second fork moving in the opposite direction (Fig. 4A). If, however, two converging forks stop, replication cannot in principle be completed. Work from our group and others revealed that in the latter situation, genome duplication relies on the presence of an excess of licensed origins referred to as “dormant” or “back-up” origins (Ge et al., 2007; Ibarra et al., 2008; Woodward et al., 2006). MCM complexes are loaded onto chromatin in excess relative to the number of origins that are activated. Strikingly, a strong downregulation of MCM proteins (up to 95%) does not have major consequences for cellular replication and proliferation in normal conditions (Ibarra et al., 2008). However, in situations of RS caused by DNA polymerase inhibitors or dNTP depletion, cells with limited licensing accumulate DNA damage and display reduced viability coupled to checkpoint activation (Ge and Blow, 2010; Ge et al., 2007; Ibarra et al., 2008). It was proposed that upon RS, ‘silent’ origins located in the proximity of stalled forks are activated as a back-up mechanism to replicate the stretch of DNA between stalled forks (Fig. 4B and Ge et al. 2007; Ibarra et al. 2008). There is evidence that these newly activated origins could belong either to the ‘dormant’ or ‘flexible’ categories (Fig. 5). On one hand, the activation of novel dormant origins in response to a partial depletion of the dNTP pool was shown by DNA combing assays at the GNAI3 locus of Chinese hamster cells (Anglana et al., 2003). On the other hand, an increased usage of flexible origins without activation of dormant initiation sites has been reported in response to DNA

replication slowdown caused by MUS81 deficiency (Fu et al., 2015). The activation of extra origins may require the action of CHK1 kinase. Interestingly, basal levels of CHK1 activity may negatively regulate the activation of potential origins within a single replicon in the absence of RS, because CHK1 downregulation in human cells leads to increased firing of neighbouring origins (Maya-Mendoza et al., 2007).

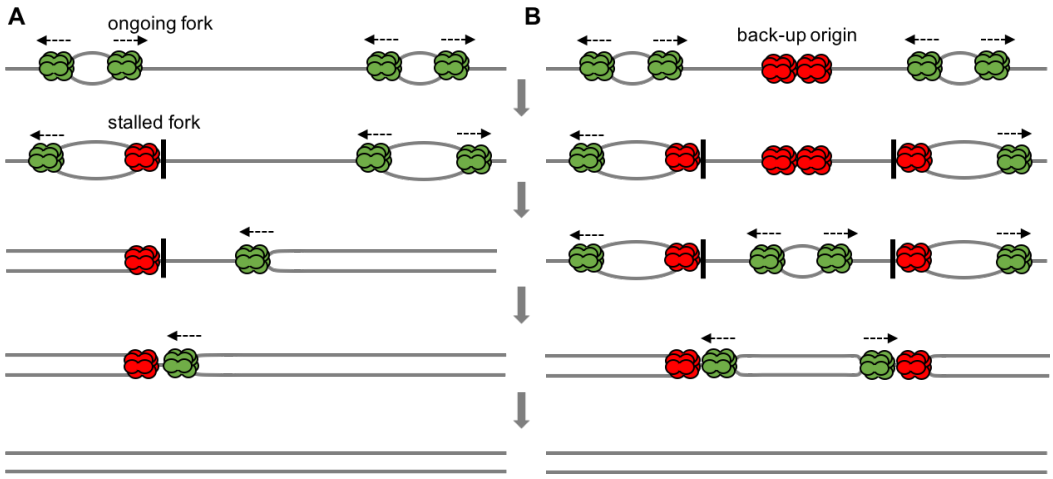


Figure 4. **Mechanisms to rescue stalled forks: (A)** a single stalled fork can be recovered from an adjacent fork; **(B)** two converging forks that stall simultaneously can only be rescued by the activation of a dormant origin located between them. See text for details.

In parallel, CHK1 activation globally inhibits the replication of clusters of late-replicating origins (Ge and Blow, 2010; Ge et al., 2007), through the inhibition of CDK and DDK (Fragkos et al. 2015). Interestingly, ATR-dependent inhibition of origin firing protects cells from deleterious DNA breakage caused by RPA exhaustion (Toledo et al., 2014). If the levels of RS are sufficiently high, the cell cycle is temporarily arrested due to the inhibition of CDK1, which blocks the entry into mitosis.

The importance of having a full complement of potential origins (constitutive and flexible/dormant) has been demonstrated *in vivo* in several mouse models. For instance, mice hypomorphic for the MCM2 or MCM4 subunits of the MCM licensing factor display higher levels of genomic instability and are prone to cancer (Pruitt et al., 2007; Shima et al., 2007). Mice with reduced levels of MCM3 suffer from lethal anaemia caused by defects in the differentiation of hematopoietic progenitors (Alvarez et al., 2015). Hence, proper use of potential origins is of the highest importance for the different cell types in an organism. The regulation of origin activity in response to RS and other contexts such as differentiation is still under investigation.

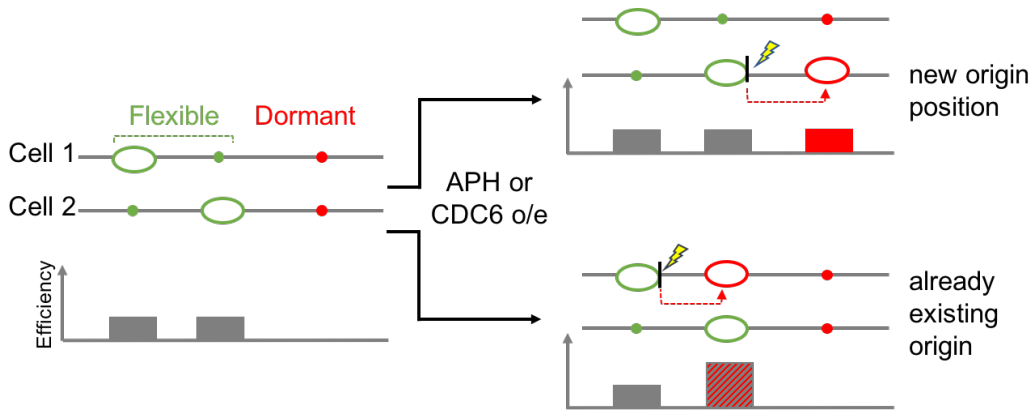


Figure 5. Flexibility in origin choice allows two models for additional origin activation: either an origin in a new genomic position (top right) or an increased frequency of activation of an existing origin in the population (bottom right). See text for details.

Replication timing

In eukaryotes, certain factors needed for initiation and progression of replication may be limiting and prevent the activation of all origins at the same time. As seen by the different patterns of replication foci in the nucleus, origins fire at different times throughout S phase. Still, a **replication timing program** governs the temporal and spatial organization of origin firing. This program is regulated by different determinants such as three-dimensional (3D) chromosomal organization, transcriptional programs and gradients of factors limiting for initiation (reviewed by Rhind & Gilbert 2013; Rivera-Mulia & Gilbert 2016a,b).

Three-dimensional genome architecture

Chromatin is hierarchically organized in the nucleus, starting with the formation of **clutches** and **fibers** by dynamic nucleosome-nucleosome interactions (reviewed by Bian and Belmont, 2012; Ricci et al., 2015). The next level of chromatin organization consists of formation of **loops**, which can bring together genomic elements separated by long distances in the linear genome, such as promoters and enhancers (Rao et al., 2014). The recent discovery of **topologically associated domains** (TADs) revealed yet another level of chromatin organization (Dixon et al., 2012). TADs are stable sub-megabase size domains that encompass strong chromatin interactions within them, but not with other regions. TADs often correlate with specific epigenomic signatures. TADs are then folded into higher-level **compartments** of multi-megabase scale, which are of two types, A and B, associated to active/open and inactive/closed chromatin, respectively. At the highest level of genome organization, different compartments of the same chromosome group together forming a **chromosome territory**. Inter-chromosomal interactions are rarely observed within the nucleus. In mammalian cells, 3D chromatin organization is

established and maintained by architectural proteins such as **mediator**, **CTCF** and **cohesin** (reviewed by Bonev and Cavalli, 2016; Phillips-Cremins et al., 2013; Pombo and Dillon, 2015). The nuclear lamina is also linked to chromatin organization since it anchors chromatin to the nuclear periphery (Shimi et al., 2010)

Recent advances in chromatin conformation capture (3C) techniques, including Hi-C (a high-throughput variation of 3C), have provided novel insights into 3D chromatin architecture, allowing for better understanding of its functional role in processes occurring in the nucleus (reviewed in Bonev & Cavalli 2016). Briefly, in 3C methods, chromatin is crosslinked, fragmented with restriction enzymes or sonication, and this is followed by ligation of adjacent DNA ends. The possible interaction between two fragments is detected by PCR or sequencing methods. Importantly, modifications of 3C assays enable for genome-wide detection of chromatin interaction networks mediated by specific proteins (e.g. cohesin or Polycomb) or occurring at specific genomic elements such as promoters (Bian and Belmont, 2012; Bonev and Cavalli, 2016; Schoenfelder et al., 2015a, 2015b).

Chromatin organization and timing of DNA replication

In eukaryotes, genome replication follows a defined order that seems to be evolutionary conserved between closely related species (Ryba et al., 2010). In mammalian cells, genome-wide replication timing profiles reveal the existence of 400-800 kb replication timing domains, frequently abbreviated **replication domains** (RD), which replicate simultaneously (Hiratani et al., 2008; Ryba et al., 2010). They are interspersed by **timing transition regions**, where the change in replication timing occurs. The approximately 5000 RD likely correspond to microscopically observed foci, as early RDs localize in the nuclear interior and late RDs in the nuclear envelope and nucleolar periphery (Fig. 6 and Rivera-Mulia and Gilbert, 2016a).

Replication timing correlates with chromatin accessibility, certain DNA sequence features, and transcriptional activity. Early replication occurs preferentially in euchromatin regions and is associated with high gene density, high GC content and active transcription. Late-replicating regions present an opposite pattern. They tend to localize in heterochromatin and correlate with gene-poor regions, low GC content and low transcriptional activity (Hiratani et al., 2008; Rhind and Gilbert, 2013; Ryba et al., 2010).

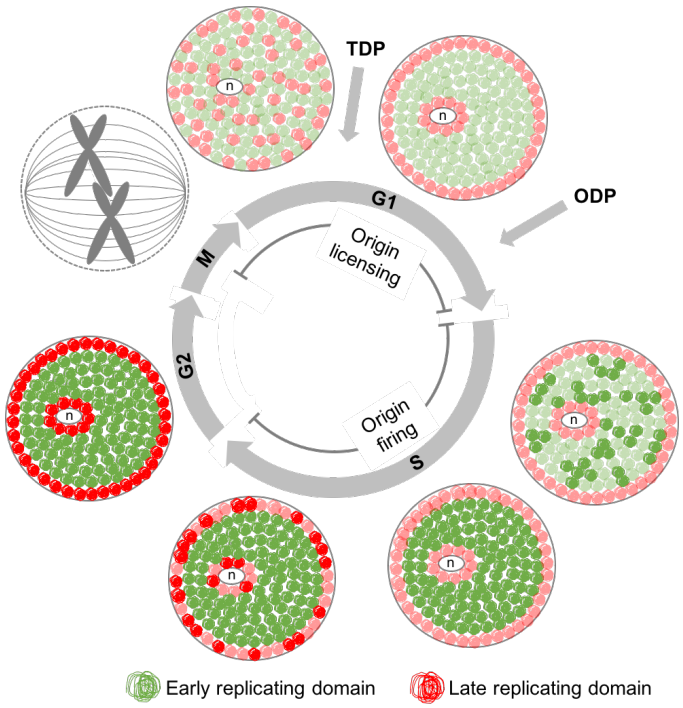


Figure 6. Model for the regulation of DNA replication in the nucleus. TDP, timing decision point, ODP, origin decision point; n-nucleolus. Figure inspired from Rivera-Mulia and Gilbert (2016a). See text for details.

Several observations have revealed a striking concordance between 3D chromatin architecture and the replication timing program. First, the strong correlation between early and late RDs with A- and B- Hi-C compartments, respectively (Ryba et al., 2010). In addition, a high degree of overlap has been observed between late RD and **Lamin B1-associated domains** (LADs), which are chromosome regions associated with nuclear periphery (Guelen et al., 2008). In a recent study, RDs were found to actually correspond to TADs (Pope et al., 2014) and it was therefore proposed that replication timing is controlled at the level of TADs (Dileep et al., 2015a; Rivera-Mulia and Gilbert, 2016b).

Replication timing is established early in G1, at the so-called **timing decision point** (TDP; Dimitrova & Gilbert 1998). At this moment, chromatin reorganizes spatially into RD/TADs allowing for segregation of early and late replication timing domains into different nuclear compartments (Dileep et al., 2015b). The selection of origins for activation takes place later in S phase, at the so-called **origin decision point** (ODP; Fig. 6).

Timing of replication is regulated during development. Changes in the temporal order of replication generally correlate with changes in transcriptional activity (Hiratani et al., 2008; Ryba et al., 2010). Importantly, developmentally regulated RDs remain

intact and undergo changes as whole units. RDs that maintain their temporal order of activation upon differentiation are called constitutive RDs (Rhind and Gilbert, 2013; Rivera-Mulia and Gilbert, 2016a). Intriguingly, constitutive and developmentally regulated domains present different chromatin properties (Dileep et al., 2015b).

Replication timing may be regulated by **competition for limiting factors** that would initially be directed to easily accessible, early-replicating origins. As S phase progresses, fewer origins compete for the pool of limiting factors, and their probability of firing increases in late S phase. This mechanism would ensure that late-replicating heterochromatin regions duplicate timely and efficiently (Rhind and Gilbert, 2013). In yeast, limiting proteins include Cdc45, Sld2, Sld3, Sld7, Dbp11 and Cdc7-Dbf4, and their overexpression increases the probability of late origins activation (Mantiero et al., 2011; Tanaka et al., 2011). In mammalian cells, it has been proposed that Cdc45 is the main limiting factor (Wong et al., 2011).

Determinants of replication origin choice

Although replication is tightly controlled temporally and spatially at the level of RD/TADs, the selection of replication origins within those domains appears to be flexible and stochastic. Decades of work have been devoted to investigate the genome-wide distribution of origins in different eukaryotic organisms, to understand what determines their localization and efficiency (reviewed in Fragkos et al. 2015; Prioleau & Macalpine 2016; Aladjem & Redon 2016).

Methods to map replication origins genome-wide

A straightforward origin mapping method consists in chromatin immunoprecipitation of pre-RC components, followed by hybridization to DNA microarrays (**ChIP-chip**) or high-throughput sequencing of the immunoprecipitated DNA (**ChIP-seq**). These techniques have been successful applied in yeast (Hayashi et al., 2007; Wyrick et al., 2001) and *Drosophila melanogaster* (MacAlpine et al., 2010; Powell et al., 2015). However, they have proven to be more challenging in mammalian genomes. Although ChIP-seq datasets have been recently published with ORC1 and ORC2 proteins (Dellino et al., 2013; Miotto et al., 2016), the complexity of mammalian genomes makes it difficult to convincingly map origin 'peaks' over background signal (reviewed by Gilbert, 2010).

Other methods aim at localizing the start sites of bidirectional DNA replication (Prioleau & MacAlpine 2016). Isolation and sequencing of **short nascent strands**

(SNS-Seq) synthetized on the leading strand at origins has been widely used to map initiation sites (see Fig. 7 in Materials and Methods, Cayrou et al., 2012a). RNA-primed SNS are isolated by sucrose gradient centrifugation and digested with lambda exonuclease (λ-exo) to eliminate fragments of broken genomic DNA, which are not protected by the presence of RNA primers at their 5' end. SNS-seq provides good resolution and can also be used to estimate origin efficiency (Besnard et al., 2012; Cayrou et al., 2011; Comoglio et al., 2015; Picard et al., 2014). Nevertheless, this method has been criticized because the λ-exonuclease step might enrich in GC-rich sequences and G-quadruplex (G4)-containing DNA (Foulk et al., 2015). A variation of this method in which SNS are first labelled with BrdU and then enriched by immunoprecipitation has also been applied (Mukhopadhyay et al., 2014).

DNA Sequencing of DNA-replication bubbles (**bubble-seq**) is an alternative method, in which replication bubbles are first 'trapped' in a 2D agarose gel and then sequenced (Mesner et al., 2013). Bubble-seq is very sensitive but it has limited resolution and is technically challenging (reviewed by Prioleau and MacAlpine, 2016; Urban et al., 2015).

Another recently developed approach for origin mapping consists in the isolation and sequencing of Okazaki fragments (**Oka-seq**), as initially described in yeast (Smith and Whitehouse, 2012). Oka-seq detects initiation sites by identifying the transition zones between leading (continuous) and lagging (discontinuous) strand synthesis. In yeast, this method has allowed for global analysis of initiation, origin efficiency, fork progression and termination events (McGuffee et al., 2013). In human cells, the resolution of Oka-seq allowed for detection of broad initiation zones rather than specific sites of origin firing (Petryk et al., 2016).

Attempts at genome-wide origin mapping using the techniques described have revealed the presence of tens of thousands of origin positions in a metazoan genome (Besnard et al., 2012; Cadoret et al., 2008; Cayrou et al., 2015; Comoglio et al., 2015; Martin et al., 2011; Mesner et al., 2013; Petryk et al., 2016; Picard et al., 2014; Sequeira-Mendes et al., 2009). Remarkably, the overlap between the different sets of results is limited, even when the same cell lines are used (Hyrien, 2015). Several factors could contribute to these discrepancies. First, the initiator ORC complex does not recognize a consensus sequence in higher eukaryotes (Miotto et al., 2016; Vashee et al., 2003). Second, origins are licensed in excess and each cell in the population may activate a different subset. Besides, at the time of origin licensing,

MCM complexes can slide along the DNA and change their localization from the ORC-binding site (Gros et al., 2015; Powell et al., 2015). Despite these difficulties, some features of mammalian replication origins and the factors that influence their probability of activation are progressively being discovered.

Genomic features influencing origin choice

S. cerevisiae origins are autonomously replicating sequences (ARS) that support DNA replication when placed into plasmids (reviewed by Gilbert, 2004). ARS contain a short consensus sequence that is specifically recognized by ORC (Eaton et al., 2011; Wyrick et al., 2001). However, not all the existing ACS are used as origins, indicating that sequence is not sufficient to determine replication start sites (Nieduszynski et al., 2006). Notably, recent reconstitution of DNA replication *in vitro* using chromatinized templates revealed that origin selection is highly dependent on chromatin (Devbhandari et al., 2017; Kurat et al., 2017).

Origin consensus sequences have not been identified in any other eukaryotic organism. In *S. pombe,* origins are determined by A/T-rich islands recognized by ORC (Hayashi et al., 2007; Heichinger et al., 2006; Segurado et al., 2003). Metazoan replication origins tend to localize in GC-rich regions and are enriched at CpG islands, methylation free regions frequently associated with promoters (Antequera, 2004; Cadoret et al., 2008; Cayrou et al., 2011; Delgado et al., 1998). The most efficient origins are located in regions with high GC content (Besnard et al., 2012; Sequeira-Mendes and Gómez, 2012).

A correlation of origins with G-rich repeated elements (OGRE) that have a potential to form G-quadruplexes (G4) has also been reported (Besnard et al., 2012; Cayrou et al., 2012b, 2015; Comoglio et al., 2015). The presence of G4 may have a role in origin positioning and efficiency (Valton et al., 2014). At least *in vitro,* ORC binds preferentially to ssDNA that forms G4s (Hoshina et al., 2013). However, potential G4-forming sequences are in large excess relative to the number of origins (Huppert and Balasubramanian, 2005), indicating that this element is not sufficient to define replication start sites.

Transcription, epigenetic features, and chromatin environment

The correlations between origins and genomic features such as DNAseI hypersensitive sites (Comoglio et al., 2015; Mesner et al., 2013) and nucleosome occupancy (Cayrou et al, 2015; Lombraña et al, 2013) suggest a direct link between replication and chromatin accessibility. Indeed, mammalian origins are preferentially

located at permissive chromatin environments (Fragkos et al., 2015; Sequeira-Mendes and Gómez, 2012). Early RDs show higher density and efficiency of origins than late RDs (Besnard et al., 2012; Cadoret et al., 2008; Miotto et al., 2016). The general view that emerges is that origins preferentially locate at TSS, promoters and gene-rich regions, indicating a positive correlation with active transcription. Of note, origins located around promoters and CpG islands are more efficient (Cadoret et al., 2008; Cayrou et al., 2010; Dellino et al., 2013; MacAlpine et al., 2010; Martin et al., 2011; Sequeira-Mendes et al., 2009).

While correlations are not absolute, several epigenetic marks have been associated to replication start sites (Aladjem and Redon, 2016; Prioleau and MacAlpine, 2016). Early origins tend to colocalize with euchromatin histone modifications, e.g. H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K18ac and H3K36me3. In contrast, late origins associate with heterochromatin marks, e.g. H3K9me, H3K9me3, H3K27me3 and hypoacetylation of H3 and H4 (Aladjem and Redon, 2016). H3K79me2-containing DNA has also been suggested to correlate with origin activity (Fu et al., 2013)

Factors involved in nucleosome modification and remodelling also influence origin usage (reviewed by Aladjem and Redon, 2016; Prioleau and MacAlpine, 2016). H4 specific acetylase HBO1 (Iizuka & Stillman, 1999) is recruited to origins through its interaction with ORC1 and CDT1 and is important for MCM loading (Miotto and Struhl, 2010). PR-Set7-mediated methylation of H4K20 also promotes pre-RC formation (Tardat et al., 2010). Interestingly, the bromo-adjacent homology (BAH) domain of ORC1 recognizes specifically H4K20me2, and its mutation leads to lower ORC occupancy at selected origins (Kuo et al., 2012).

Recent genome-wide origin studies in human and mouse cells have revealed a correlation of origins with marks associated to Polycomb (PcG) repressive complexes (PRC) 1 and 2. Replication start sites associate with PcG proteins (mESCs) and H3K27me3 (mESC and human K562 cells), an epigenetic mark deposited by PRC2 (Cayrou et al., 2015; Picard et al., 2014). In ESCs, origins localize preferentially at sites bound by both repressive H3K27me3 and activating H3K4me3 marks, a feature of bivalent domains that silence developmentally regulated genes in ESCs but keep them poised for transcriptional activation (Bernstein et al., 2006; Cayrou et al., 2015). Interestingly, PRC1 organizes a 3D network of promoter-promoter contacts in the nucleus, further emphasizing the

strong connection between origin localization, chromatin architecture and transcription (Schoenfelder et al., 2015a).

In this dissertation, we have aimed at better understanding the regulation of origin selection and activation in mammalian cells. In the first part we investigated the response of replication origins to stress and their spatial organization in the nucleus. In the second part we studied a novel mechanism that regulates CDC6 protein stability during the cell cycle.

Objectives

Objectives

1. Generate genome-wide maps of replication origins in mESCs growing in normal conditions or under situations of stress.
2. Define subsets of constitutive and stress-responsive origins, in order to determine their genomic and epigenomic characteristics, as well as their activation efficiency.
3. Integrate linear origin maps into three-dimensional chromatin structure.
4. Investigate the role of CDC7 kinase in the regulation of CDC6 licensing factor.

Objetivos

1. Generar mapas a nivel genómico de orígenes de replicación en mESCs en condiciones normales de crecimiento o en situaciones de estrés.
2. Definir subconjuntos de orígenes constitutivos y orígenes de respuesta a estrés, para determinar sus características genómicas y epigenómicas, así como su eficiencia de activación.
3. Integrar los mapas lineales de orígenes en la estructura tridimensional de la cromatina.
4. Investigar la función de la quinasa CDC7 en la regulación de la proteína replicativa CDC6

Materials and Methods

Materials and Methods

Cell lines and culture conditions

The TetO-CDC6 mouse model has been recently described (Muñoz et al., 2017). In this model, HA-tagged CDC6 is expressed upon induction with doxycycline. TetO-CDC6- mouse embryonic stem cells (TetO-CDC6-ESCs) were grown on 0.1% gelatin-coated plates in Dulbecco's Modified Eagle's Medium (DMEM) with Ultraglutamine 1 and 4.5 g/L glucose (Lonza) supplemented with 15% FBS (Sigma), 50 U/mL Penicillin - 50 mg/mL Streptomycin (Invitrogen), Minimum Essential Medium Non-Essential Aminoacids (MEM NEA; Invitrogen), 100 µM 2-Mercaptoethanol (Invitrogen) and 103 U/mL ESGRO mLIF Medium Supplement (Millipore). TetO-CDC6-mouse embryonic fibroblasts (TetO-CDC6-MEFs), U2OS and HeLa cells were grown in complete DMEM (Lonza) supplemented with 10% FBS (Sigma-Aldrich) and 10% penicillin/streptomycin (Invitrogen). All cells were incubated at 37°C in humid environment containing 5% CO_2.

To induce CDC6 overexpression, 1 µg/ml doxycycline (dox, Sigma) was added to the medium for 27h (MEFs) or 30h (ESCs). To induce mild replication stress cells were treated with 0.5 µM aphidicolin (Sigma-Aldrich) for 2.5h.

To synchronize cells in metaphase 250 ng/ml nocodazole (Sigma-Aldrich) was added to the medium for 18h. Cells were released from the block by washing twice with pre-warmed PBS and once in pre-warmed medium and then incubated in fresh medium for the indicated times.

The following drug treatments were used: 25 µg/ml cycloheximide (CHX, Sigma), 10-25 µM MG-132 (Sigma-Aldrich), nocodazole (Sigma-Aldrich), 25µg/ml roscovitine (Sigma-Aldrich), 20 µM PHA-767491 (Cdc7i, Sigma-Aldrich). We observed that PHA-767491 activity varied depending on the batch. Incubation times were adjusted empirically and therefore may vary among experiments, as reflected in the figure legends.

siRNA and plasmid transfections

DNA sequences encoding for WT, phospho-dead (S8A) and phospho-mimic (S8D) *CDC6* were ordered as GeneArt Strings DNA Fragments or GeneArt Gene Synthesis plasmids (Thermo-Fisher Scientific) and subsequently cloned into Gateway expression vectors introducing an N-terminal V5 tag (Invitrogen). Transient siRNA and plasmid transfection were performed using Lipotransfectin (Niborlab) and Lipofectamine 2000 (Invitrogen), respectively, according to manufacturer's instructions. Transfections with siRNA against *CDC7* (siCdc7) were performed with

100nM siRNA for 48h. siCdc7 oligonucleotides were synthesized to the following *CDC7* coding region: 5'-AAGCTCAGCAGGAAAGGTGTT-3' (GE Healthcare), as previously reported (Cdc7-A; Montagnoli et al., 2004).

Flow cytometry analysis

In order to analyse DNA synthesis, cells were pulse-labelled with 20μM BrdU for 30 min. Cells were trypsinized, washed in PBS and fixed with cold (-20°C) 70% ethanol for at least 24h. DNA denaturation and cell permabilization was achieved by incubation with 2N HCl for 20 min at RT. Cells were then washed twice with PBS, treated with blocking solution (1% bovine serum albumin in PBS, 0.05% Tween20) for 15 min at RT and incubated with FITC-conjugated anti-BrdU antibody (BD Biosciences Pharmigen) for 1h at 37°C. To monitor DNA content, cells were stained overnight with 50μg/ml propidium iodide (PI; Sigma-Aldrich) in the presence of RNaseA (10μg/ml, Qiagen). Flow cytometry was conducted in a FACS Canto II cytometer (BD) and data was analysed using FlowJo V 9.4 or V.10.1 (Three Star).
For cell sorting into different phases of cell cycle, U2OS cells were stained ($1x10^6$ cells/ml) with 2.5 μg/ml Hoechst (Invitrogen) for 30 min at 37°C in culture medium. Cells were washed in PBS, and resuspended in PBS + 0.3% FBS + 3mM EDTA and sorted by DNA content using a BD Influx sorter (BD, San Jose, CA) in the Flow Cytometry Unit at CNIO.

Single molecule analysis of DNA replication in stretched DNA fibers

Exponentially growing cells were pulse-labelled for 20 min with 50 μM CldU, washed 3 times with PBS and pulse-labelled with 250 μM IdU for another 20 min. Cells were trypsinized, washed and resuspended in ice-cold PBS ($0.5x10^6$ cell/ml). 2μl of cell suspension was placed on microscope slides and lysed with 0.2M Tris pH 7.4, 50μM EDTA, 0.5% SDS (10 μl) for 6min at RT. Slides were tilted 15 degrees to spread DNA fibers. Next, slides were air-dried, fixed in cold (-20°C) methanol:acetic acid (3:1) for 2 min and stored at 4°C overnight. DNA was denaturated in 2.5N HCl (30min/RT) and washed 3x in PBS. Blocking solution (1% bovine serum albumin in PBS, 0.1% Triton X-100) was added for 1h at RT. Slides were incubated with CldU, IdU and ssDNA primary antibodies (Table 2) diluted in blocking solution (1h, RT), washed in PBS and incubated with secondary antibodies for 30 min. Slides were mounted with Prolong (Invitrogen). Images were acquired with DM6000 B Leica microscope with an HCX PL APO 40x, 0.75 NA objective. ssDNA staining was performed to monitor DNA integrity. For fork rate analysis, the length of second-labelled (IdU, green) track was measured using ImageJ software. The conversion

factor 1µm= 2.59kb (Jackson and Pombo, 1998) was applied. At least 300 tracks were measured per condition. To assess origin firing activity, first-label origins (corresponding to CldU tracks flaked on both sides by IdU tracks, green-red-green structures) were counted and quantified as a percentage of all CldU (red) tracks.. At least 500 structures in total were counted per condition. Three biological replicates of each DNA fibre experiment were performed.

Short Nascent Strand assay

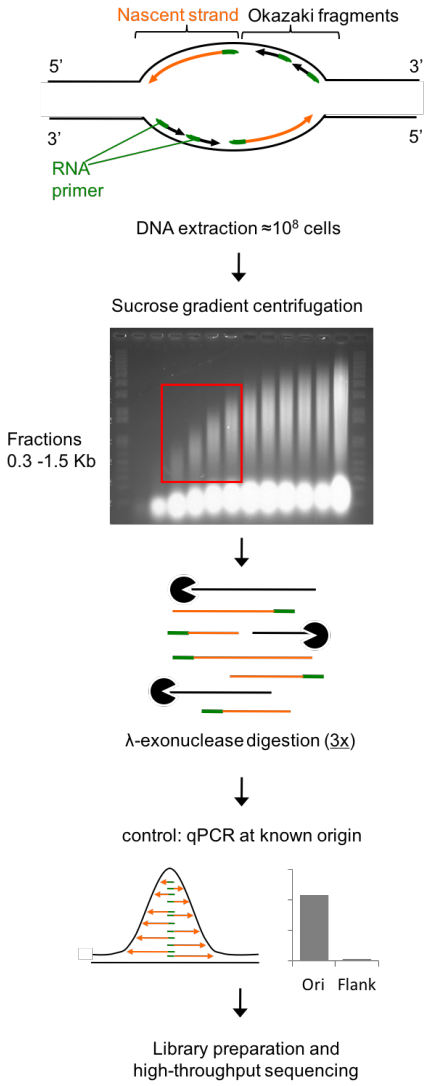


Figure 7. SNS assay workflow.

Short Nascent Strand assay was performed in collaboration with Dr. María Gómez's laboratory (CBMSO, Madrid), according to a previously described protocol (Sequeira-Mendes et al., 2009). This assay aims at the isolation of short nascent strands (SNS), which are leading strand replication intermediates synthetized at replication bubbles (Fig. 7).

Extraction of genomic DNA

Exponentially growing cells ($1X10^8$) were washed twice in PBS and lysed for 15 min in Lysis Buffer (50 mM Tris pH 8.0, 10 mM NaCl., 10mM EDTA pH 8.0, 0.5% SDS). 100 µg/ml Proteinase K (Roche) was added and samples were incubated overnight at 37°C. The DNA lysate was gently mixed with one volume of Tris-saturated Phenol (Sigma-Aldrich) and centrifuged (3500 rpm, 10 min). The aqueous phase was mixed gently with one volume of Phenol:chlorophorm:isoamyl alcohol (25:24:1) and centrifuged (3500 rpm, 10 min). DNA was precipitated with two volumes of cold (-20°C) 100% ethanol (Panreac AppliChem), washed with 70% ethanol, air dried and resuspended in TE buffer (10 mM Tris pH 8.0, 1mM EDTA) supplemented with 0.1 U/µl RNAseOUT (Invitrogen). Samples were stored at 4°C.

Purification of SNS

Seven-step discontinuous sucrose gradients (5-20% sucrose, 2,5% steps made in 10 mM Tris pH 8.0, 1 mM EDTA and 100 mM NaCl) were prepared in centrifuge tubes (Beckman Coulter 331374). Total genomic DNA was heat-denaturated (100°C / 10 min), cooled on ice and loaded on the top of gradient. Size-fractionation was performed by centrifugation at 78000 rcf in an Ultracentifuge (Beckman Coulter Optima L-100 XP), with a SW-40Ti rotor (Beckman Coulter) for 20 min at 20°C, as previously described (Gómez and Antequera, 2008). 12-13 1ml fractions were recovered and DNA from each of them was precipitated with ethanol. Analysis of DNA size in each fraction was performed by 1% agarose gel electrophoresis in alkaline conditions. Fractions 4-5, corresponding to the size of ≈ 300-1500 nt, were chosen for further analysis. Fractions containing shorter and larger DNA fragments were not used to avoid the selection of Okazaki fragments and large fragments of broken genomic DNA, respectively. Examples of DNA fractionation are depicted in Fig. 8A. For each experimental replicate, DNA samples were pooled from two separate gradients.

Samples were treated with 100 U of T4 Polynucleotide Kinase (PNK, Thermo Fisher) in the presence of 1mM dATP (Roche) and 40 U of RNAseOUT (Thermo Scientific) for 30 min at 37°C. PNK phosphorylates 5'-hydroxyl ends rendering DNA molecules that are preferentially digested by λ-exonuclease. PNK reaction was stopped with of Proteinase K (6.25 µg), 0.125% sarkosyl and 2.5 µM EDTA (30 min at 37°C). DNA was extracted, precipitated and resuspended in H_2O. To eliminate any broken genomic DNA from the samples, digestion with λ-exonuclease was carried out. This enzyme degrades DNA fragments from their 5'-end, and RNA-primed short nascent strands are protected by their 5'-RNA primers. PNK-treated samples were heat-denaturated (95°C, 5 min), cooled on ice, and incubated with 150 U of λ-exonuclease (Thermo Scientific) in λ-exonuclease digestion buffer (Thermo Scientific) in the presence of 40 U of RNAseOUT (37°C, over night). Reactions were heat-inactivated (75°C, 10 min). DNA was extracted, precipitated and resuspended in H_2O. To increase the purity of SNS in the fractions, PNK treatment and λ-exonuclease digestion were performed three times. Upon digestion, DNA was precipitated with two volumes of 100% ethanol (-20°C) in the presence of 0.3 M sodium acetate pH 5.2 and 20 µg/ml glycogen (Roche) for 30 min at -20°C. Samples were centrifuged (20000 rcf, 30 min, 4°C) and pellets were washed with 70% ethanol. After air-drying, DNA pellets were resuspended in TE buffer. Samples were subsequently processed for high-throughput sequencing.

Each λ-exonuclease digestion step was controlled by adding 50 ng of linearized pFRT-myc plasmid to 5% of digestion reaction and incubated in the same conditions. pFRT-myc (a kind gift from S. Gerbi, Brown University, Providence, USA) contains two G-quadruplex (G4) forming sequences, which are known to be digested less efficiently by λ-exonuclease (Foulk et al., 2015). Also, a reaction without enzyme was performed. These control reactions were loaded on 1% agarose gel to confirm full digestion of the plasmid (Fig. 8B).

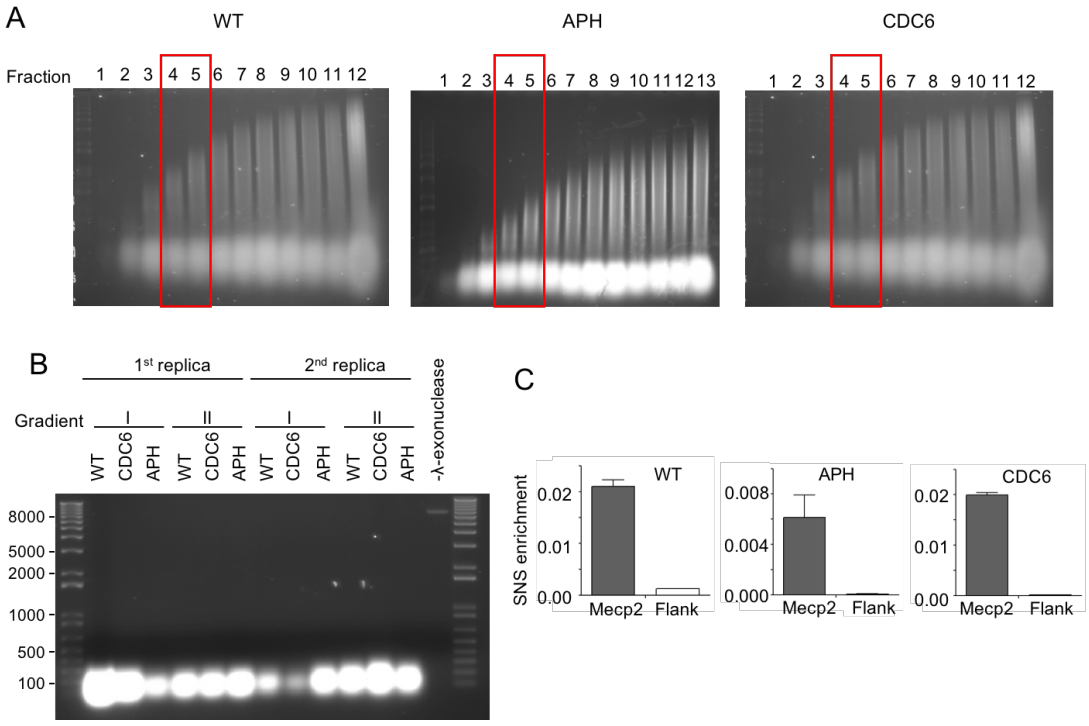


Figure 8. (A) Representative images of alkaline agarose gel electrophoresis following DNA fractionation by sucrose gradient centrifugation. Fractions 4 and 5 (marked in red) were used for further analysis. **(B)** Control reaction for λ-exonuclease digestion. 5% of the pool of fractions 4 and 5 was mixed with pFRT-Myc plasmid and digested with λ-exo. Two separate gradients (I and II) for each replica are shown. Note the presence of a band (≈7200 bp) corresponding to undigested linearized pFRT-myc vector only in the - λ-exo control. **(C)** Control qPCR reaction for one of the replicates of WT, APH and CDC6 samples at Mecp2 origin and its flanking region. Absolute SNS enrichment is depicted. Mean and SD is represented; n=2.

Before library preparation, a fraction of the samples was tested by quantitative real-time PCR (qPCR) to control for SNS enrichment at the known Mecp2 origin (Fig. 8C). Primer sequences for Mecp2 origin and Mecp2 flanking region are indicated in Table 1. qPCR reactions were performed in duplicates using ABI Prism 7900HT Detection System (Applied Biosystems) and HotStarTaq DNA polymerase (Qiagen) according to manufacturer's instructions. Data was analysed in Applied Biosystems Software

SDS v2.4. For absolute quantifications samples of known quantity were serially diluted and amplified to generate a standard curve.

Genomic DNA extraction for input sample

Exponentially growing mESCs (≈ $40x10^6$) were washed twice in PBS and collected by scraping and centrifugation (3000 rpm, 7 min, 4 °C). Cells were lysed in lysis buffer (1% SDS, 10mM EDTA, 50 mM Tris-HCl ph 8.1) at a $2x10^7$ cell/ml density. Sonication was performed in a Bioruptor device (Diagenode) for 25 min (30s sonication/30s pause mode). DNA was extracted with phenol/chloroform, ethanol-precipitated, air-dried and resuspended in 0.5x TE buffer. The size of DNA in the samples was monitored by 1% agarose gel electrophoresis.

Library preparation and high-throughput sequencing

RNA primers were first removed by treatment with RNAse A/T1 Mix (Roche) for 60 min at 37 °C. 100 µg/ml Proteinase K was added (30 min, 37 °C) and DNA was extracted and precipitated. ssDNA was converted to dsDNA by random priming using 50 pM random hexamer primer phosphate (Roche) for 5 min at 95°C and cooling gradually to 4°C, as previously described (Cadoret et al., 2008). Primer extension was performed by subsequent incubation (1 h, 37°C) with 10 mM dNTPs (Roche), 5 U of Klenow Fragment (3'→5' exo-; New Englands Biolabs) in 1x NEB buffer 2 (New Englands Biolabs). Reactions were inactivated by incubation at 75°C for 10 min. Ligation of adjacent DNA fragments was performed by incubation with 80 U of TaqDNA ligase (New England Biolabs) in 1x TaqDNA ligase buffer (30 min, 50°C), followed by enzyme inactivation (10 min, 75°C), DNA extraction with phenol/chloroform and ethanol precipitation. Samples were resuspended in TE buffer.

DNA libraries were prepared in Fundación Parque Científico de Madrid (FPCM) using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (New England Biolabs) following manufacturer's instructions and purified with Agencourt AMPure XP beads (Beckman Coulter). Each library was sequenced using a read length of single end 75 bp (120-140 $x10^6$ reads per sample) in NextSeq500 System (Illumina) at the FPCM.

SNS-seq data analysis

Note: Bioinformatic and computational analysis of SNS-seq data and its integration with genome-wide chromatin data was performed in collaboration with Dr. Vera Pancaldi (Structural Biology and Biocomputing group, CNIO, Madrid; Computational Biology Life Sciences Group, BSC, Barcelona), Osvaldo Graña and Miriam Rubio

(Bioinformatics Unit, CNIO, Madrid) with additional assistance from José Miguel Fernández (Genome Dynamics and Function Group, CBMSO, Madrid).

The quality of sequencing reads was analyzed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Short sequences, adaptor sequences and read duplicates were removed. Reads were then analyzed with the RUbioSeq pipeline v3.8 (Rubio-Camarillo et al., 2017), using BWA v0.7.10 (Li and Durbin, 2009), SAMtools v0.1.19 (Li et al., 2009), Picard tools v1.107 (http://broadinstitute.github.io/picard/) and MACS v2.0.10 (Feng et al., 2012), and the mouse reference genome GRCm38/mm10. When MACS algorithm was used, the peak calling was performed versus input. The algorithm described in Picard et al. (2014) was used to provide a second way to identify ORIs. Genome segmentation required by this algorithm was based on replication timing data from mESCs, which accurately matches the read coverage differences between segments. Common peaks were obtained using BedTools v2.23.0 (Quinlan, 2014) with parameters: -f 0.1 -r -wa –u. In stringent criteria, a peak was considered to be an origin when it was detected by both peak-calling algorithms (MACS/Picard) in both replicates. In relaxed criteria, a peak was considered to be an origin when it was detected by both peak-calling algorithms in at least one replicate. For common peaks between MACS and Picard the genomic coordinates defined by Picard were used. Read distribution around peak centres was generated using seqMINER v1.3.3e. For the comparative analysis with epigenomic features and chromatin states, genomic coordinates of origins were converted from mm10 to mm9 genome assembly with LiftOver (https://genome.ucsc.edu/util.html).

Genomic feature analysis

For analysis of genomic distribution, origins were intersected with the genomic features depicted in Table 3, using the indicated overlapping parameters. Randomized origin datasets were generated to facilitate interpretation of the data. This was done by reshuffling origin positions 1000 times along the genome, with preservation of the number and sizes of origin peaks for each chromosome. Origin density was calculated in 500 kb sliding windows; the size of the slide was 1 kb. This analysis was performed in mm10 mouse genome assembly in collaboration with Dr. María Gómez's Laboratory (CBMSO, Madrid).

Origin efficiency analysis

Read density divided by peak length was used as an approximation for origin efficiency. To calculate read density, each read within a peak was multiplied by the

number of nucleotides that it covers. Mean origin efficiency was calculated for each origin group or subgroup. In stringent datasets, efficiency of each origin was calculated as an average of efficiency between two replicates. In relaxed datasets their average efficiency was used when a peak was detected in both replicates, and this single efficiency value was used when the peak was detected only for one replicate. Comparison of efficiencies between different origin groups was performed using Wilcoxon signed-rank test.

Epigenomic feature analysis

We analysed the overlap of origin groups with a set of ≈ 70 epigenomic features previously compiled from >130 ChIP-seq experiments performed in ESCs (Juan et al., 2016). ChIP-seq datasets were 'discretized' in 200 bp windows: the presence of a given epigenetic mark within a 200bp window was scored as 1, and its absence as 0. The overlap between origin fragments and the genomic windows was calculated using findOverlaps in the genomicRanges R package. A score for each epigenomic feature was calculated as the number of 200bp windows with value=1 within each origin fragment, divided by the total number of windows within the origin. This produced a value between 0 and 1 for each epigenomic mark in each origin fragment. The overlap of origins with epigenetic features was visualized using heatmaps with rows and columns ordered by unsupervised hierarchical clustering, using the cluster package and gplots in R. The enrichment of any particular feature in origins was calculated as the ratio between its average value in origin fragments and its average genomic value. To further assess the significance of these overlaps origin positions were randomly shuffled along each chromosome 100 times (using Bedtools) and the average overlap in randomized datasets was also calculated. We observed that randomized datasets are mostly indistinguishable from the genome average. Therefore, enrichment is presented as fold-change between the origin overlap and the genome average.

Chromatin state analysis

Using the same approach, we calculated enrichment of origins at 20 chromatin states determined previously (Juan et al., 2016). In brief, these states were defined using a Hidden Markov Model applied to a collection of datasets including major histone modifications (ChromHMM, Ernst and Kellis, 2010), see Juan et al., 2016 for details. For simplification purposes, states corresponding to similar chromatin functions were merged according to the definitions indicated in Table 4.

Integration with chromatin interaction maps

Linear origin maps were integrated into the following 3D chromatin interaction maps for mESCs: Promoter-Capture HiC (PCHiC; Schoenfelder et al., 2015), SMC1 ChIA-PET (Dowen et al., 2014), RNA Polymerase II (RNAPII) ChIA-PET (Zhang et al., 2013) and another variation of promoter Capture Hi-C called Hi-Cap (Sahlén et al., 2015). The PCHiC, Hi-Cap and ChIA-PET networks were processed as described in Pancaldi et al. (2016). The PCHiC contact map was processed using the CHiCAGO pipeline, which identifies significant 3D contacts starting from the raw Capture HiC data. The other networks were generated starting from the contacts provided in the original references.

Origin positions were mapped to the chromatin fragments of 3D maps using findOverlaps (genomicRanges package). Origin efficiency of the chromatin fragment was calculated as the mean of the efficiencies of origins that overlap with this fragment.

Network analysis

All network analyses including correlations between degree, efficiency and replication timing (RT) were performed using the igraph package in R and standard R functions. TAD definitions were taken from Dixon et al. (2012). Networks were visualised using Cytoscape v3.4.0.

Replication timing data

Replication timing data was downloaded from Hiratani et al. (2010). Here RT is given as the ratio of probes of two colour tiling microarrays representing either early or late RT for 3 mESC cell lines, which were combined in our analysis. The data was processed in two separate ways: 1) the median value for probes overlapping each origin was taken and individual origins were ranked based on RT in the range 0 to 1 (from early to late); 2) the probes within each chromatin fragment in the 3D chromatin network were combined to give an RT for the entire chromatin fragments.

Assortativity analysis

Origin Assortativity (OriAs) and Assortativity of Origin Efficiency (OriEfAs) were calculated using the previously described measure of Chromatin Assortativity (Pancaldi et al., 2016).

Briefly, assortativity in a chromatin contact network is defined as the Pearson correlation coefficient of the presence of an origin (OriAs) or of the value of origin

efficiency (OriEfAs) across all pairs of nodes that are connected with each other. This value was calculated with the assortativity function in the igraph package for R.

OriAs is thus a value between -1 and 1, with negative values meaning that origins are preferentially in contact with non-origin fragments and positive values meaning that origins are preferentially in contact with other origins. To estimate the significance of OriAs values we performed 100 randomization of origin assignations on the network fragments, preserving the total number of origins (OriAs) or the mean origin efficiency (OriEfAs). As ChAs depends on how prevalent a particular feature is in the network fragments, it is important to analyse it in relation to the feature abundance. For example, if a particular mark is found in the majority of the fragments in the network, its localization in specific areas of the network cannot be observed and the value of ChAs will be low. On the contrary, when a certain feature is detected only in a small subset of fragments, but they are found to interact preferentially with each other, the ChAs measure is high.

OriAs and OriEfAs have very similar behaviour, as OriAs is just a discretization of efficiency values, where any efficiency >0 is rounded to 1 and fragments with no origins are assigned efficiency 0. Thus, instead of representing the number of fragments containing origins as in the OriAs plot, the OriEfAs plots show on the x-axis the mean value of their efficiency in the network fragments.

RNA expression analysis

RNA was isolated from cells using Trizol (Invitrogen) according to the manufacturer's instructions. To remove potential contamination with genomic DNA, DNAse I (Roche) digestion was performed (37°C, 20 min). Reverse transcription (RT) reaction to generate cDNA was carried out using Maxima First Strand Kit for RT-qPCR (ThermoFisher). qPCR reaction was performed in triplicates using SYBR Green PCR Master Mix (Thermo Fisher), using ABI Prism 7900HT Detection System (Applied Biosystems). Quantifications were normalized to endogenous GAPDH, using ΔΔCt method. Primer sequences are listed in Table 1.

Whole cell extract preparation

Cells were trypsinised, collected by centrifugation (1200 rpm, 5 min), washed in PBS and resuspended in Laemmli Sample Buffer (50 mM Tris-HCl pH 6.8, 10% glycerol, 3% SDS, 0.006% w/v bromophenol blue and 5% 2-mercaptoetanol at 1000 cells/µl.

Solutions were sonicated for 30 seconds at 15%-20% amplitude (Branson Digital Sonifier).

Biochemical fractionation

Biochemical fractionation assays were performed according to a previously described protocol (Méndez and Stillman, 2000). Briefly, cells were resuspended ($1x10^7$ cells/ml) in buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM MgCl2, 0.34 M sucrose, 10% glycerol, 1 mM DTT, 1mM NaVO4, 0.5mM NaF, 5mM β-glycerophosphate, 0.1mM PMSF, protease inhibitor cocktail-Roche). TritonX-100 was added at the concentration of 0.05 % and cells were incubated for 5 min on ice. Low-speed centrifugation (4min, 3600 rpm, 4°C) was used to separate the cytosolic fraction (supernatant) from the nuclei (pellet). Cytosolic fraction was centrifuged at high speed (14000rpm, 15min) to eliminate debris. Pellets were washed in buffer A and lysed for 30 min in buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT, protease inhibitors as described above). Centrifugation at 4000rpm for 5 min was performed to separate soluble nuclear proteins from chromatin-bound proteins. Chromatin fraction was resuspended in Laemmli Sample Buffer and sonicated. The distribution of proteins between equivalent amounts of the soluble and chromatin-enriched fractions was analysed by immunoblot.

Immunoblots

Standard protocols were used for SDS-polyacrylamide gel electrophoresis and immunoblotting (SDS-PAGE; Harlow and Lane, 1999). Primary antibodies used in this study are listed in Table 2. Horseradish peroxidase (HRP)-conjugated secondary antibodies (GE Healthcare) and ECL developing reagent (Amersham Biosciencies) were used.

Protein purification

pGEXCdc7/Dbf4$^{199-333}$ expression plasmid (Rainey et al., 2013) was kindly provided by Dr. C. Santocanale (National University of Irleand, Galway). pGST-TEV-hCdc6 expression plasmid to purify GST-CDC6 was previously described (Méndez and Stillman, 2000). Both pGEXCdc7/Dbf4$^{199-333}$ and GST-CDC6 proteins were expressed in Rosetta (DE3) *E. coli* strain and purified using gluthathione-sepharose beads according to standard protocols. GST tag was removed from CDC7/DBF4$^{199-333}$ using PreScission protease. CDK2-CycA was purified as described (Méndez et al., 2002).

In vitro kinase assay

The kinase reaction was performed in a total volume of 20 µl in the presence of 1µCi of [γ-32P]ATP in kinase buffer (50M HEPES pH 7.9, 10 mM $MgCl_2$, 2 mM DTT, 10 µM ATP) for 30 min at 30 °C, as described (Rainey et al., 2013). 1.5 µg GST-CDC6 was used as substrate and 0.5 g CDC7/DBF4$^{199-333}$ or 1.5 µg CDK2-CycA as kinases in separate reactions. To remove unbound [γ-32P]ATP, samples were purified through Ilustra MicroSpin G-25 Columns (GE Healthcare). The reaction was analysed by SDS-PAGE. After Coomassie staining and gel drying, signal was detected by exposing the gel to autoradiographic film.

Cold *in vitro* kinase assays were performed to obtain samples for mass spectrometry (MS) analysis, at the Proteomics Unit (CNIO). Samples (20 µL) were diluted with 180 µL of freshly prepared 8M urea in 100 mM Tris/HCl pH 8.0 containing 15 mM TCEP and 45 mM chloroacetamide, and incubated 1h at 25 ºC in the dark. Protein samples were desalted following the FASP protocol (Wiśniewski et al., 2009) and digested with 200 ng of trypsin for 5h at 37 ºC. Peptides were desalted using home-made Empore® C18 tips. All samples were analysed using a QqTOF Impact (Bruker Daltonics) coupled online to a NanoLC Ultra 1D+ system (Eksigent), equipped with a CaptiveSpray nanoelectrospray ion source supplemented with a CaptiveSpray NanoBooster operated at 0.2 bar/minute with isopropanol as dopant. Data was analysed and quantified using the Maxquant software v1.5.3.30.

Statistical analysis

Statistical analyses were performed using Prism v4.0 (GraphPad Software) or Microsoft Excel v15.38. For comparison of two data groups, two-tailed paired Student's t-test was used. In the analysis of fork rate in stretched DNA fibers, data distribution is normally not Gaussian and statistical differences were assessed with nonparametric Mann-Whitney rank sum test.

Table 1. Primers for qPCR.

Gene/region	primer name	primer sequence 5'- 3'
Mecp2 origin	Mecp2 43 Fwd	CTACCCGCCCCCCAGCAAG
	Mecp2 44 Rv	GTGAGTGGGACCGCCAAGG
Mecp2 flank	Mecp2 1 Fwd	GCATCCAATGCTCTTTGTGC
	Mecp2 14 Rv	GTCTCTTGTTGAGCATTTGT
GAPDH	*GAPDH*-Fwd	TGCACCACCAACTGCTTAGC
	GAPDH-Rv	GAGGGGCCATCCACAGTCTTC
CDC6	Cdc6-Fwd	TCTGATTCCCAAGAGGGTTG
	Cdc6-Rv	TTCTGCTGAAGAGGGAAGGA
MCM3	Mcm3-Fwd	AGCTTCTGCGGTATGTGCTT
	Mcm3-Rv	CCTGTTTCCTGGTCTGTGGT
ORC1	Orc1-Fwd	GCTCCTCAGATCCGTAGTCG
	Orc1-Rv	CGACAGGGAAGAGACTCAGG
p21	p21-Fwd	TTCATGCCAGCTACTTCCTC
	p21-Rv	GAATTTCATAACCGCCTGTG

Table 2. Antibodies used in this study.

Antibody	Use	Supplier	Ref/Catalogue #	Species/ Dilution
Aurora B	WB	Abcam	ab2254	Rb/1:1000
CDC45	WB	Méndez Lab	Aparicio et al. (2009)	Rb/1:400
CDC6	WB	Millipore	05-550	Ms/1:1000
CDC6 pS54	WB	Santa Cruz	sc-12920	Goat/1:400
CDC7	WB	Novus Biologicals	SPM171	Ms/1:1000
CDC20	WB	Proteintech	10252-1-AP	Rb/1:500
FZR1/CDH1	WB	Abcam	DH01 (DCS-266)	Ms/1:500
CDK9	WB	Cell signalling	C12F7	Rb/1:1000
CDT1	WB	Millipore	07-1383	Rb/1:1000
Cyclin A	WB	Santa Cruz	sc-751	Rb/1:1000
Cyclin B1	WB	Santa Cruz	sc-752	Rb/1:1000
Cyclin B1	WB	Santa Cruz	sc-245	Ms/1:1000

Cyclin E	WB	Santa Cruz	sc-247	Ms/1:1000
GAPDH	WB	Antibodies Unit,CNIO	220/EG3	Ms/1:2000
Geminin	WB	Santa Cruz	sc-13015	Rb/1:1000
H3	WB	Abcam	ab1791	Rb/1:50000
MCM2	WB	Méndez Lab	Ekholm-Reed et al. (2004)	Rb/1:2000
MCM2 pS40	WB	Abcam	ab133243	Rb/1:1000
MCM2 pS53	WB	Abcam	ab109133	Rb/1:1000
MCM3	WB	Méndez Lab	Ekholm-Reed et al. (2004)	Rb/1:5000
MCM4	WB	Méndez Lab	Ekholm-Reed et al. (2004)	Rb/1:4000
MCM6	WB	Méndez Lab	Ekholm-Reed et al. (2004)	Rb/1:4000
MEK2	WB	BD	610235	Ms/1:2000
ORC1	WB	Méndez Lab	769AP Mendez et al. (2002)	Rb/1:200
ORC 2	WB	Méndez Lab	205 S Mendez et al. (2002)	Rb/1:2000
p53 pS15	WB	Cell signalling	9284	Rb/1:1000
p21	WB	Santa Cruz	sc-397	Goat/1:1000
PSF1	WB	Méndez Lab	192B Aparicio et al. (2009)	Ms/1:500
PSF2	WB	Méndez Lab	AB34 Aparicio et al. (2009)	Rb/1:500
RNAPII pS2	WB	Abcam	ab5095	Rb/1:1000
γH2AX	WB	Millipore	05-636	Ms/1:500
BrdU-FITC	FACS	BD	556028	Ms/1:200
BrdU (CldU)	IF	Abcam	ab6326	Rat/1:100
BrdU (IdU)	IF	BD	347580	Ms/1:100

Rb-rabbit, Ms-mouse

Table 3. Genomic features analysed in this study.

Feature	Average size (bp)	% overlap	Source/definition
CGI	655	10	UCSC browser, (Gardiner-Garden and Frommer, 1987)
promoters	2000	10	RefSeq UCSC track, (Pruitt et al., 2004, 2014); +1.5 kb and -0.5 kb from TSS
exons	176.4	10	RefSeq UCSC track, (Pruitt et al., 2004, 2014)
introns	4619.8	10	RefSeq UCSC track, (Pruitt et al., 2004, 2014)
TTS	2000	10	RefSeq UCSC track, (Pruitt et al., 2004, 2014); +/- 1kb from TTS
Intergenic regions	92823.2	10	RefSeq UCSC track (Pruitt et al., 2004, 2014)
G4	34	1bp	Home-made script; G4 were defined as GGGN[1,7]GGGN[1,7]GGGN[1,7]GGG, intersecting matches were merged
Early timing regions	1.49 $x10^6$	100	Hiratani et al. (2010)
Late timing regions	2.21 $x10^6$	100	Hiratani et al. (2010)

Table 4. Definitions of chromatin states used in this study.

State	Definition from Juan et al. (2016)
Enhancers	1+2+3+11+12+13+14
Active promoters	15+16+17
Poised promoters	18
Transcriptional elongation	4+5
Insulators	20
PcG Repressed	19
Heterochromatin	6+7+8+10
Low signal	9

Results

Results

Chapter 1: Response of replication origins to stress and their spatial organization in the nucleus

Changes in origin usage in response to fork stalling and CDC6 overexpression

In this chapter, we set out to determine whether stress-responsive origins may differ from regular origins on the basis of their genomic position, correlation with epigenomic marks or their three-dimensional (3D) organization within chromosome structures. Mouse embryonic stem cells (ESCs) were chosen for these experiments for two reasons: their highly proliferative status, which allows the efficient isolation of short nascent strands, and the extensive information available about their genomic and epigenomic features.

Two approaches were used to 'activate' extra origins: (1) treatment with aphidicolin, a DNA polymerase inhibitor that triggers compensatory origin activation in response to replication fork slowdown; (2) overexpression of CDC6, recently found to be a limiting factor for origin licensing and activation in primary murine cells (Muñoz et al, 2017; Fig. 9A).

In the first approach, ESCs were treated with 0.5 μM aphidicolin for 2.5 h. At this concentration no major changes in the distribution of cells within the different cell cycle phases were observed (Fig. 9B, C). While the percentage of cells in S phase was not affected, aphidicolin caused a decrease in the intensity of BrdU incorporation, a partial accumulation of cells in early S and a modest activation of the DNA damage response, as assessed by p53-P and γH2AX protein levels (Fig. 9D). The dynamics of replication were assessed using stretched DNA fibers in cells labelled consecutively with CldU and IdU (Materials and Methods). As expected, aphidicolin decreased the rate of fork progression (Fig. 9E, F) and this was accompanied by an increase in the frequency of origin firing (Fig. 9G). We conclude that even mild levels of RS are sufficient to activate 'dormant' origins.

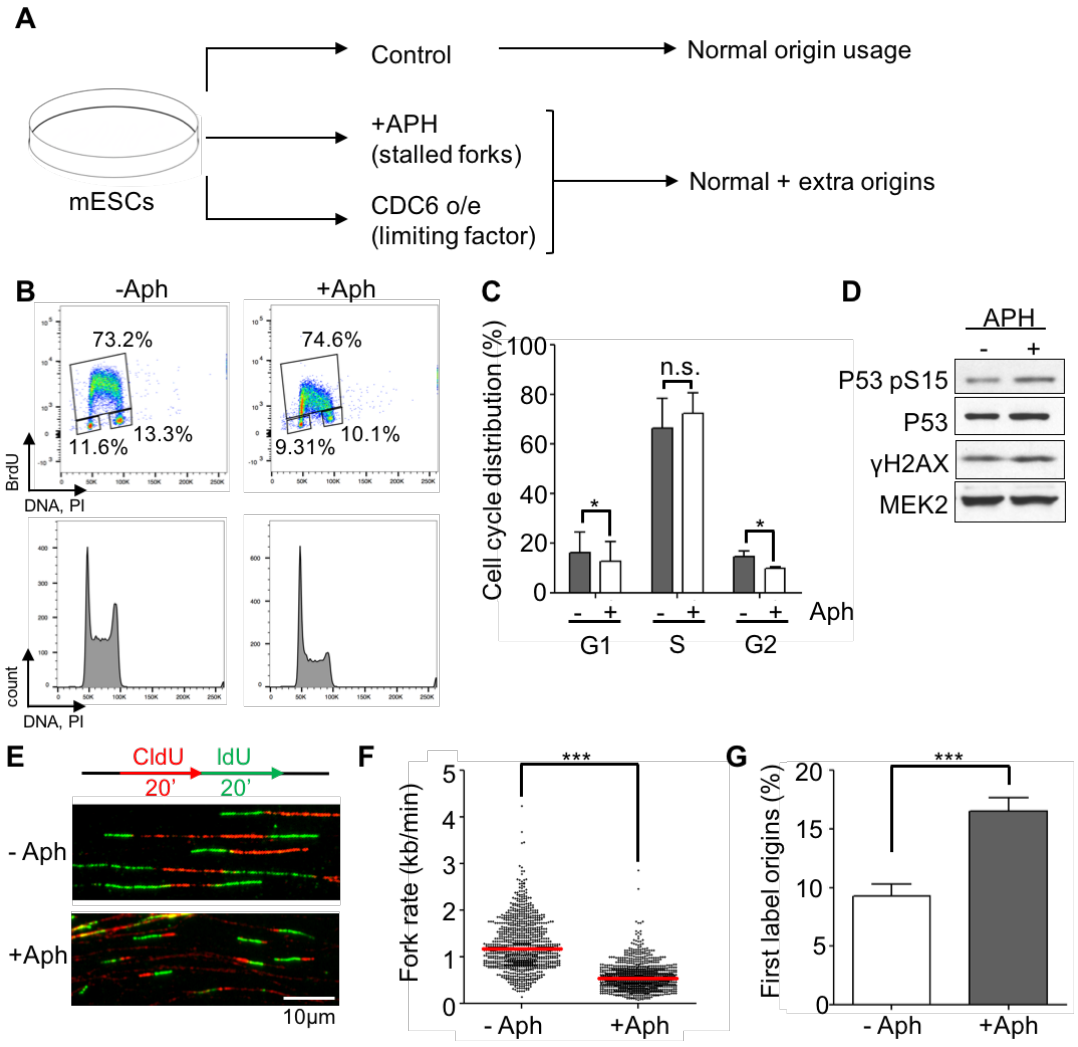


Figure 9. **Aphidicolin triggers the activation of extra origins in mESCs. (A)** Schematic of the experimental conditions used to compare origin usage in normal and stress conditions. **(B)** Top, flow cytometry plots showing BrdU incorporation profiles in control and Aph-treated mESCs. DNA content is assessed with propidium iodide (PI). Gates indicate the fractions of cells in G1, S or G2/M. The percentages of the total cell population in each gate are indicated. Bottom, DNA content profiles. **(C)** Histogram with the quantification of cell cycle distribution. Mean value and standard deviation, n=3 assays, *p < 0.1; n.s. = not significant in Student's t test. **(D)** Immunoblot detection of the indicated proteins in mESC whole cell extracts (WCE). MEK2 is shown as loading control. **(E)** DNA fibre analysis of ESCs cells. Cell labelling scheme and representative pictures of DNA fibers are shown. **(F)** Fork rate quantification; ***p<0.001 in Mann-Whitney test. **(G)** Origin activity quantification. First label origins (green-red-green structures) were counted as a percentage of all red-labelled fibres; ***p<0.001 in Student's t test.

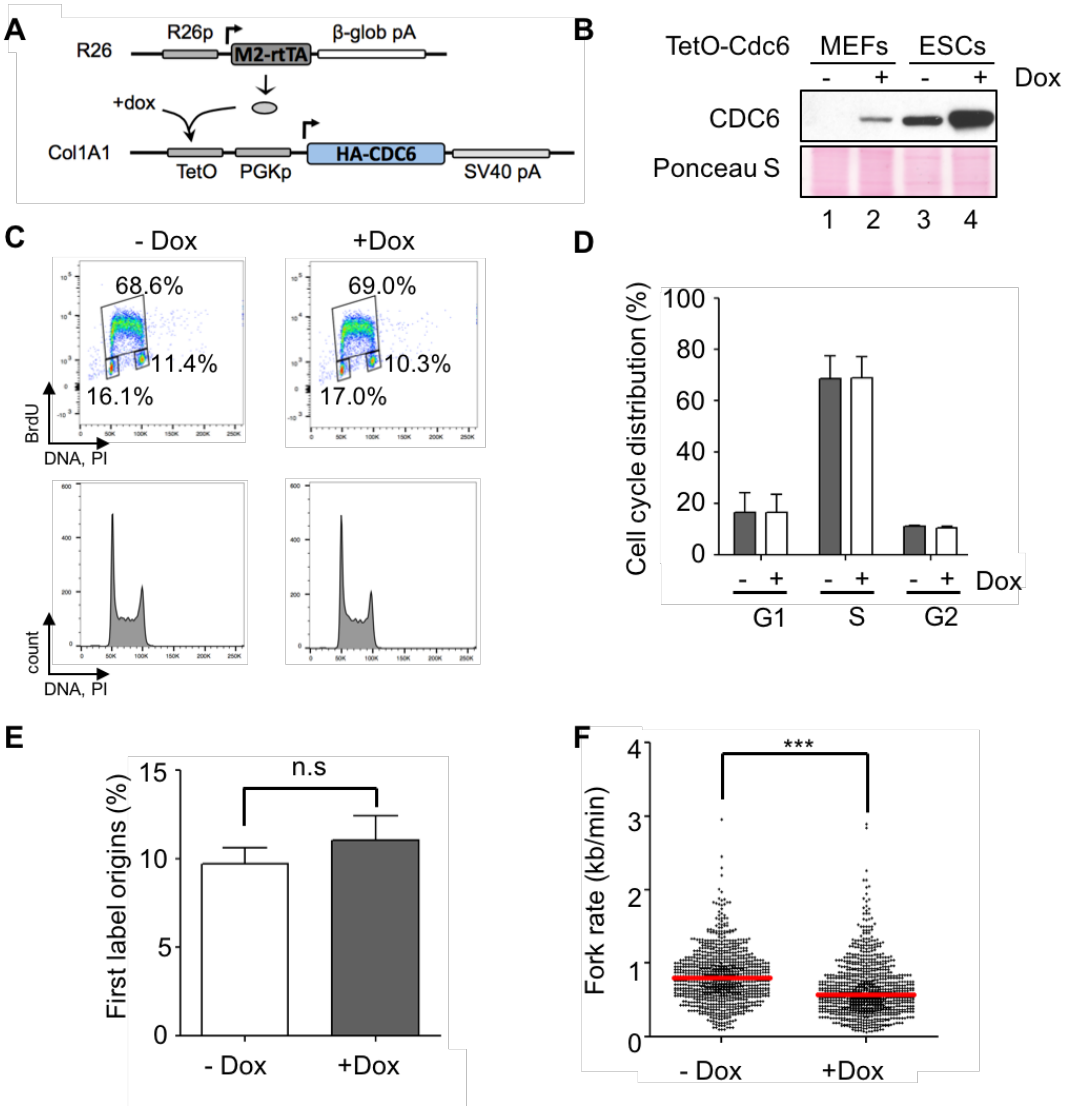


Figure 10. **CDC6 overexpression affects DNA replication in mESCs. (A)** Schematic of TetO-CDC6 cells genetic alterations. Dox addition triggers overexpression of CDC6. **(B)** Immunoblot of CDC6 protein in TetO-CDC6 MEFs and ESCs in the absence or presence of Dox. Ponceau S staining of the nitrocellulose membrane is shown as loading control. **(C)** Flow cytometry plots showing BrdU incorporation profiles of TetO-CDC6 ESCs. Bottom, cell cycle profiles determined by DNA content. **(D)** Quantification of the cell cycle distribution of the cell populations in (C). **(E)** Quantification of origin activity by DNA fibre analyses in the cells shown in (C); n.s., not significant in Student's t test. **(F)** Quantification of fork rate in the same experiment as in (E); ***p<0.001 in Mann-Whitney test.

The second approach relies on CDC6 overexpression and is based on the fact that in mouse embryonic fibroblasts (MEFs) this manipulation was sufficient to enhance origin licensing and firing, without triggering a DNA damage response (Muñoz et al., 2017). We took advantage of the available TetO-CDC6 ESC line in which HA-tagged CDC6 can be conditionally expressed with doxycycline (dox; Fig. 10A). CDC6 was efficiently overexpressed in ESCs upon 30h of induction with dox (Fig. 10B, compare lanes 3 and 4). Interestingly, the endogenous levels of CDC6 were much higher in ESCs than in MEFs (Fig. 10B, compare lanes 1 and 3). In ESCs, CDC6

overexpression did not alter the cell cycle profiles (Fig. 10C, D). DNA fiber analysis revealed a tendency towards higher origin firing following CDC6 overexpression in three separate replicates, even if the difference was less prominent than in MEFs and did not reach statistical significance (p-value 0.2; Fig. 10E; Muñoz et al., 2017). However, CDC6 overexpression led to a decrease in fork rate, which is consistent with increased origin firing (Fig. 10F). The quantitative differences between MEFs and ESCs in their response to CDC6 overexpression are likely due to the difference in endogenous CDC6 levels (Fig. 10B).

Generation of genome-wide origin maps by SNS-Seq

SNS-seq consists in the biochemical isolation of short nascent strands (SNS), followed by next-generation DNA sequencing. SNS are the first molecules synthetized in the leading strand at replication origins, and can be separated from total DNA by sucrose gradient fractionation. SNS in the 300-1500 nt size range are selected, while shorter fragments are excluded to avoid the interference of Okazaki fragments. SNS isolation can also be contaminated by small fragments of broken genomic DNA. To avoid sequencing these fragments, SNS preparations are subjected to extensive digestion with 5'-3' λ-exonuclease, which does not act on SNS because of the presence of 5' RNA primers. The efficiency of the digestion step is crucial and it was carefully controlled in our experiments using an exogenous plasmid containing two G4 motifs which are partially resistant to λ-exonuclease (Foulk et al., 2015, Materials and Methods). Before sequencing, RNA primers are removed and the remaining SNS are converted to dsDNA using random priming.

SNS were isolated from three experimental conditions: (1) control ESCs (hereafter referred to as "WT"), (2) aphidicolin-treated ESCs (hereafter "APH") and (3) CDC6-overexpressing ESCs (hereafter "CDC6"). For each condition, two independent replicates were prepared and subjected to high-throughput DNA sequencing. Genomic DNA from asynchronous ESCs was isolated, fragmented by sonication to the same size range as SNS preparations, and sequenced. This "input" sample provided a background noise line for subsequent peak calling analysis, which is performed by specific algorithms (see below).

Good quality sequencing tracks with well-defined read density peaks were obtained in five out of six samples (1xWT, 2xAPH and 2xCDC6). A third replicate of WT SNS was obtained, but the overall quality of the sequencing data was still low. Therefore we decided to incorporate one set of WT mouse ESCs SNS-seq data obtained

independently in the laboratory of Dr. M. Gómez (CBMSO, Madrid), a collaborator in this project. Despite the slight differences that could exist between both ESC lines, read density tracks looked remarkably similar (Fig. 11).

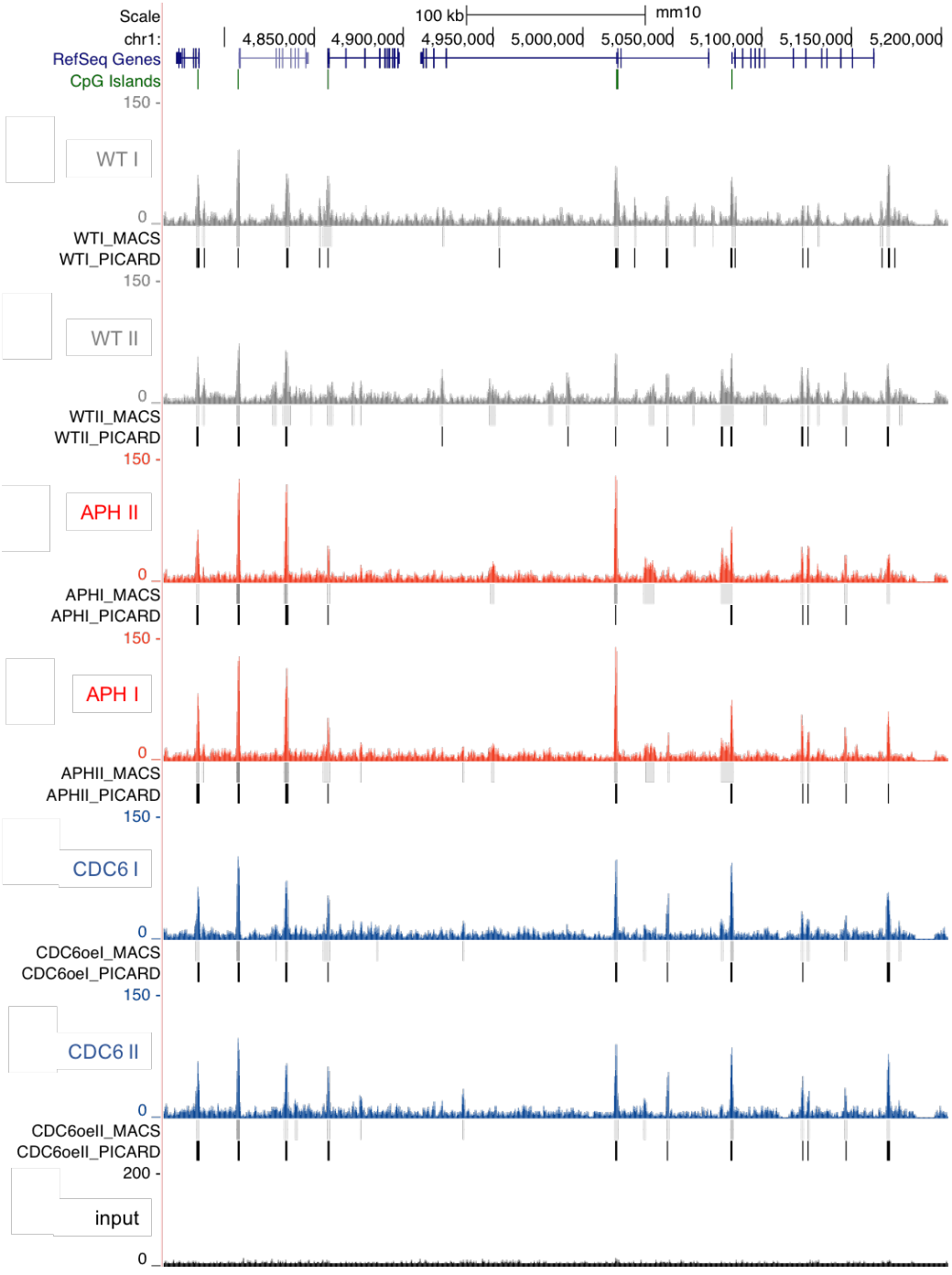


Figure 11. **Examples of origin distribution in normal and stress conditions**. Origin datasets defined in our study were uploaded to the UCSC genome browser. Images show read density tracks and peaks for the indicated datasets in a representative region of chromosome 1. Two biological replicates are shown per condition. Dashes below the tracks indicate the positions of peaks called by MACS (grey) and Picard (black) algorithms. The upper part of the graph indicates the positions of genes and CGI in this region.

SNS-seq sequencing reads were aligned to the reference mouse genome, duplicates were removed, and peaks were identified using two separate algorithms (Table 5). MACS is a widely used peak-finding algorithm developed for ChIP-seq that has also been applied for the analysis of SNS-seq data (Comoglio et al, 2015). In addition, we used a second algorithm optimized for SNS-seq analysis by the group of M.N. Prioleau (Institut Jacques-Monod, Paris), to which we refer as "Picard", after the first author in the study (Picard et al., 2014). The perceived advantage of the Picard algorithm is that the threshold of peak detection is adjusted to any variations in the local read coverage. A visual comparison in the genome browser of read density tracks revealed that both algorithms adequately detected regions of SNS enrichment (Fig. 11). Tens of thousands of putative origins were found in each condition, a number that is within the expected range (Table 5). In general, the Picard algorithm was more stringent than MACS (with the exception of WT-I replicate, in which the number of peaks were similar). In most cases, over 90% of the peaks detected by Picard were included within the peaks determined by MACS (Table 5).

Sample	No of aligned reads without duplicates	No of called peaks		No of common peaks Picard vs MACS	% of common peaks
		MACS	Picard		
WT I	57 789 135	94 758	94 924	65 623	69.1%
WT II	64 910 836	81 119	44 907	41 974	93.5%
APH I	33 994 540	74 644	49 097	42 536	86.6%
APH II	42 562 298	71 515	41 412	38 200	92.2%
CDC6 I	77 850 848	72 418	42 831	39 883	93.1%
CDC6 II	60 639 981	82 108	62 973	57 020	90.5%

Table 5. Summary of aligned reads, called peaks, and peaks overlap between MACS and Picard in separate SNS replicates. Replicate WT I was obtained in the laboratory of Dr. M. Gómez (CBMSO, Madrid).

Regardless of the peak algorithm used, a high overlap (≈80%) was obtained between the two experimental replicates in either APH and CDC6 conditions (Fig. 12A). Peak overlap was ≈50% between WT replicates, probably due to the fact that one of the two samples (WTI) displayed greater read coverage and a better signal-to-noise ratio resulting in a larger number of peaks (Fig. 12A).

As an alternative test of reproducibility, we compared the patterns of read distribution around peak centres in both replicates (Fig. 12B). Heatmaps consist of two panels, each centered on the peaks from replicate I. The panels on the left show read distribution from replicate I around the peak centres and the panel on the right represents read distribution from replicate II around the same peak centres (from

replicate I). The patterns of read distribution of both replicates were very similar in all experimental conditions, supporting the high degree of reproducibility of the SNS-seq approach (Fig. 12B).

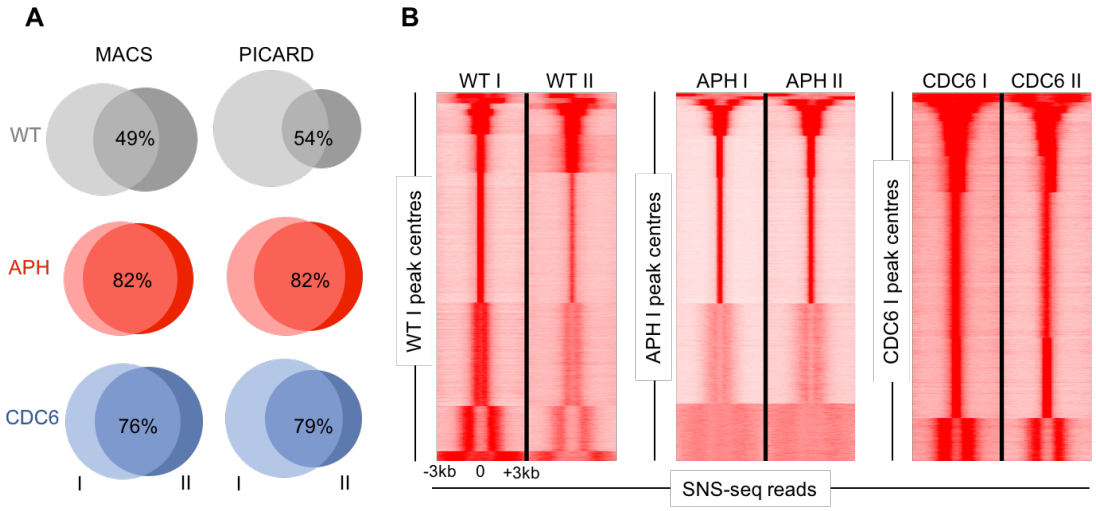


Figure 12. **Reproducibility of SNS-seq experiments**. **(A)** Venn diagrams showing the overlap of origins identified in replica I and II in each condition, either using MACS (left) or Picard (right) as the peak-calling algorithm. **(B)** Heatmaps show the distribution of SNS-seq reads for the indicated experimental replicates, around origin peak centres. See text for a detailed explanation.

Definition of WT, APH and CDC6 origin datasets

Next, we defined WT, APH and CDC6 origin datasets for further analyses. In our first approach, a peak was only considered a replication origin when it was "called" by both algorithms (MACS/Picard) and detected in both replicates. With these stringent criteria, 20178, 31653 and 31384 origins were defined in WT, APH and CDC6 conditions, respectively (Table 6, stringent). These values are below the estimated number of total origins in mammalian cells (≈50000; Méchali, 2010), but this can be expected given the criteria applied. The number of origins in the APH and CDC6 conditions was strikingly higher than in WT cells, suggesting the activation of a very large number (>10,000) of dormant origins.

Dataset	Number of origins	
	Stringent	Relaxed
WT	20178	88011
APH	31653	49922
CDC6	31384	65947
CONST	11998	29240
APH-R	17279	13843
CDC6-R	17655	27519
CDC6+APH-R	8242	6443

Table 6. Number of peaks defined for different groups of origins by stringent and relaxed criteria.

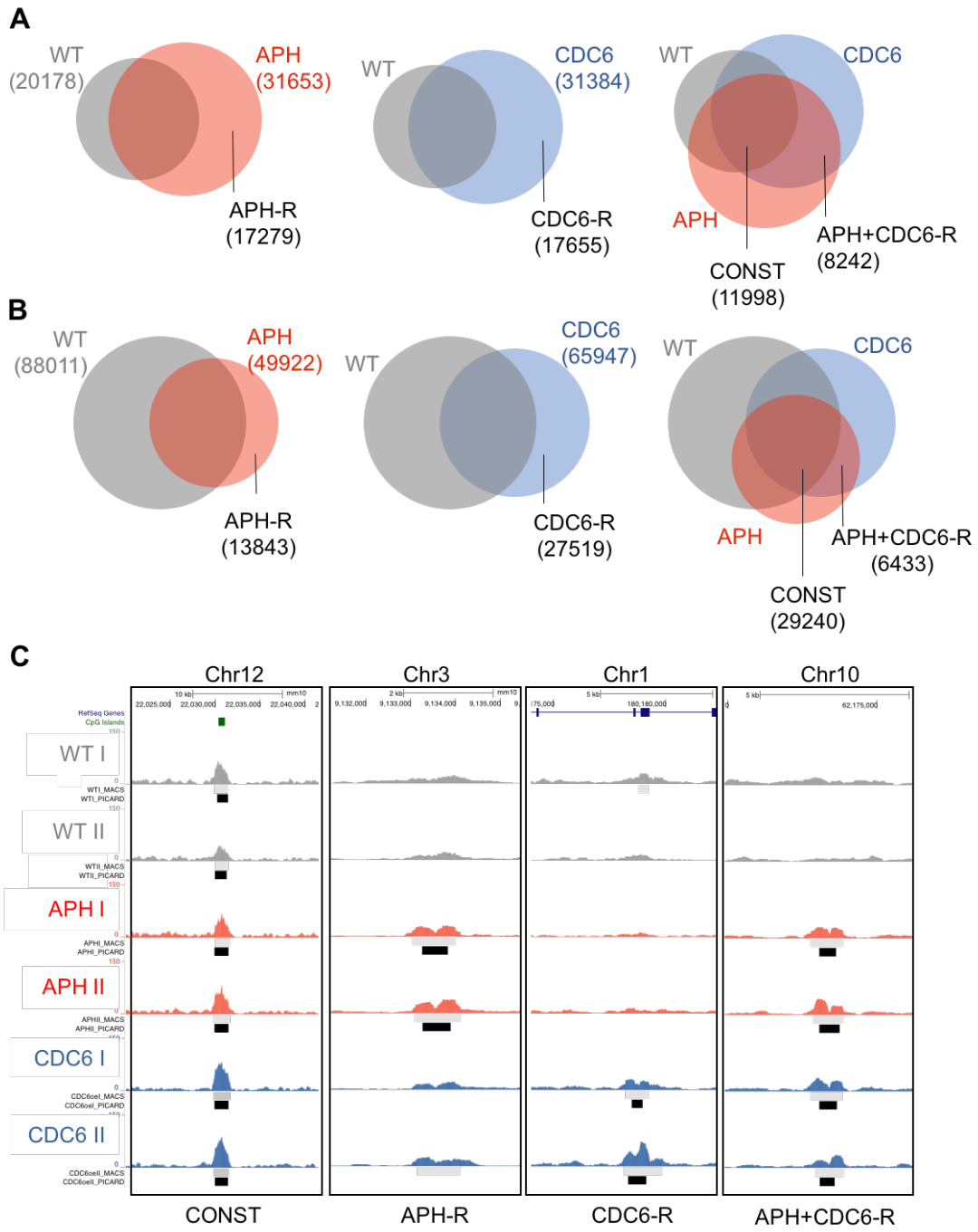


Figure 13. Definition of origin subgroups. Venn diagrams represent intersections of the indicated groups of origins defined by stringent **(A)** or relaxed **(B)** criteria. Total numbers of origins in each group or subgroup are indicated in parenthesis. **(C)** Examples of constitutive and stress-responsive origins. UCSC browser images from regions containing examples of indicated groups of origins. Read density tracks and peaks are represented. Grey bars, peaks called by MACS. Black bars, peaks called by Picard. CONST, constitutive origins; APH-R, aphidicolin-responsive origins; CDC6-R, CDC6 overexpression-responsive origins; APH+CDC6-R, common origins between APH-R and CDC6-R subgroups.

Actually, visual inspection of these origin datasets in the genome browser did not immediately support the notion that >10,000 dormant origins were activated *de novo*. In many cases, small peaks could be detected that were not called by both

algorithms in both replicates. For this reason, we defined a second group of datasets using less restrictive criteria: a peak was considered an origin when it was called by both algorithms in at least one replicate. Using this approach, "relaxed" sets of origins were defined obtaining 88011, 49922 and 65947 origins for WT, APH and CDC6 respectively (Table 6). In this case, the WT condition displayed a larger number of origins than the others, because a large number of peaks were called by both algorithms in one of the two WT replicates.

Constitutive vs responsive origins

In order to identify origins that are activated in response to APH or CDC6, we analysed the intersections between the WT, APH and CDC6 origin datasets. The subset of origins identified in all experimental conditions was termed "**constitutive**" (CONST). The subsets of **aphidicolin-responsive** (APH-R) and **CDC6-responsive** (CDC6-R) origins are those identified only upon aphidicolin treatment and CDC6-overexpression, respectively. Another subset was defined as **APH+CDC6-responsive** (APH+CDC6-R), including those origins that were activated upon both stimuli. Intersections were calculated with the stringent and relaxed origin datasets (Fig. 13 A, B; Table 6). Genome browser examples of replication origins belonging to the different categories are shown in Fig. 13C.

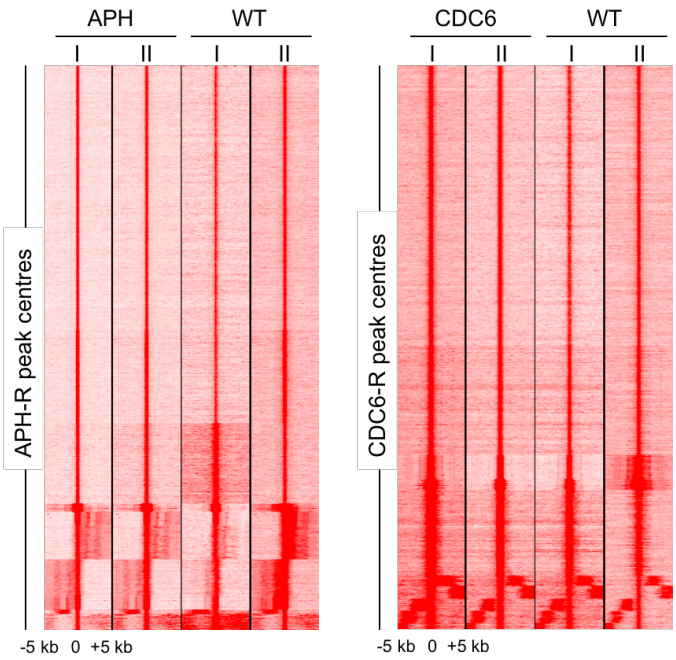


Figure 14. Heatmaps showing the distribution of SNS-seq reads for the indicated experimental replicates, around APH-R (left) or CDC6-R (right) peak centers.

As indicated before, visual inspection of the genomic browser suggests than many APH-R and CDC6-R origins are located at positions where certain activity in WT cells

is also detected (for examples see Fig. 13C). This impression was confirmed by aligning the reads of WT samples around peak centres in the APH-R and CDC6-R conditions (Fig. 14). This result immediately suggests that changes in origin efficiency are very common in the response to stress (see below).

Defining origin efficiency based on read density

The efficiency of a given origin is proportional to the percentage of cells in the population that activate it. The quantification of origin efficiency is useful because as shown above, origins already used in WT cells may display an increased frequency of activation upon stimuli. This effect may occur in parallel to the activation of new origins.

The efficiency of individual origins was determined as read density divided by peak length, as described in earlier studies (Besnard et al., 2012; Comoglio et al., 2015; Picard et al., 2014). Efficiencies were calculated using normalized read alignment files, e.g. read numbers were balanced so that the files corresponding to different replicates contained a similar number of reads. This step minimizes the risk that differences in the total number of reads are erroneously translated into differences in origin efficiency.

The possible changes in origin efficiency upon aphidicolin treatment and CDC6 overexpression were monitored. Interestingly, the average origin efficiency in the APH and CDC6 conditions was higher than in WT cells, in both stringent and relaxed datasets (Fig. 15A). This result indicates that aphidicolin and CDC6 overexpression further enhanced the use of origins that are already active in unperturbed conditions. The overall efficiency of responsive origins (APH-R and CDC6-R) was similar to the WT origins (compare APH-R and CDC6-R in Fig. 15C with WT in Fig. 15A), suggesting that the higher efficiency detected in the APH and CDC6 datasets is mainly mediated by origins that increase their frequency of firing. To further confirm this result, we calculated the average efficiency of the constitutive origins in the WT, APH, CDC6 origin groups (Fig. 15E). Indeed, both CDC6 overexpression and APH triggered an increase in the efficiency of the same origins used in WT cells.

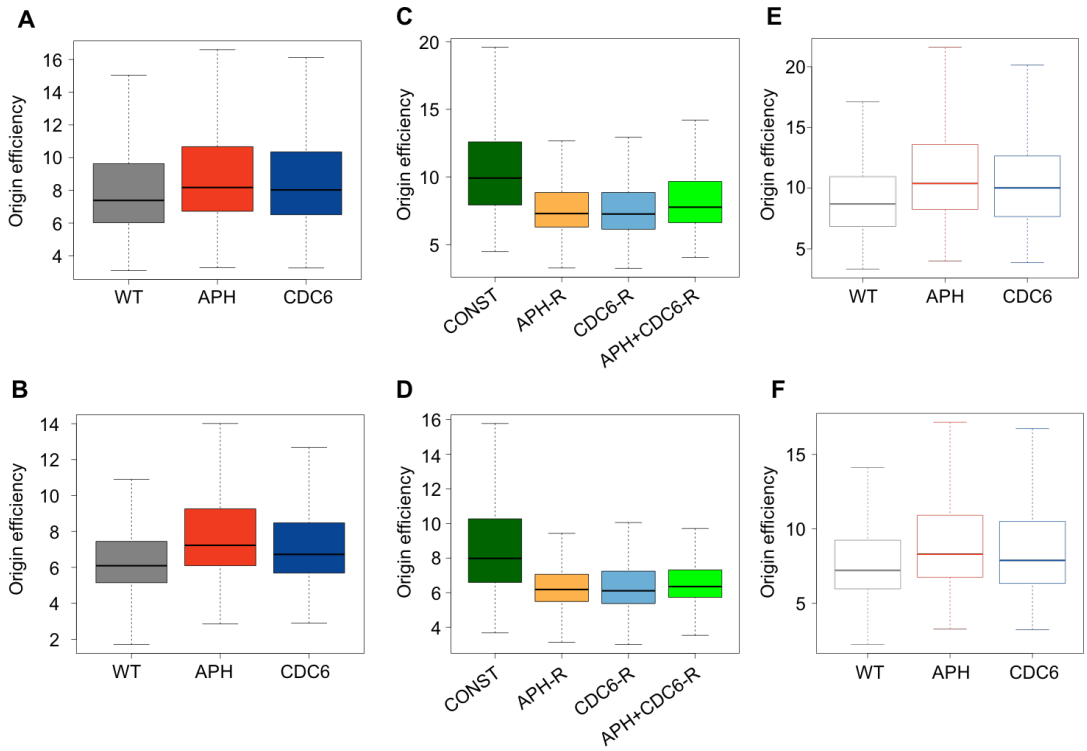


Figure 15. Relative efficiencies of different origin groups. Efficiency values were calculated as indicated in the main text for origin groups defined using stringent (A, C, E) and relaxed (B, D, F) criteria. **(A, B)** Efficiencies of main origin datasets. **(C, D)** Efficiencies of constitutive and stress-responsive origins. **(E, F)** Efficiencies of the origins common to WT, APH and CDC6 datasets in each group. Box plots indicate the distribution of efficiency (the box includes the two central quartiles and the whiskers the two extreme quartiles). Median value is indicated by a horizontal bar. For all comparisons, $p<10^{-16}$ in Wilcoxon signed-rank test.

From this analysis we conclude that aphidicolin and CDC6 overexpression enhanced the efficiency of flexible origins, i.e. the 'activation' of extra origins from already existing origin positions. Examples of increased origin efficiency in APH and CDC6 groups are shown in Fig. 16.

The efficiency of constitutive origins, calculated as the mean efficiency of common origins in the WT, APH and CDC6 conditions (Fig. 15C), was significantly higher than those of the WT, APH and CDC6 individual datasets (Fig. 15A, C), indicating that this group contains the strongest origins. By comparison, APH-R and CDC6-R origins displayed lower efficiency, indicating that they are only activated in a fraction of the population (Fig. 15A, C). All results regarding efficiency were reproduced using relaxed datasets (Fig. 15 B, D, F).

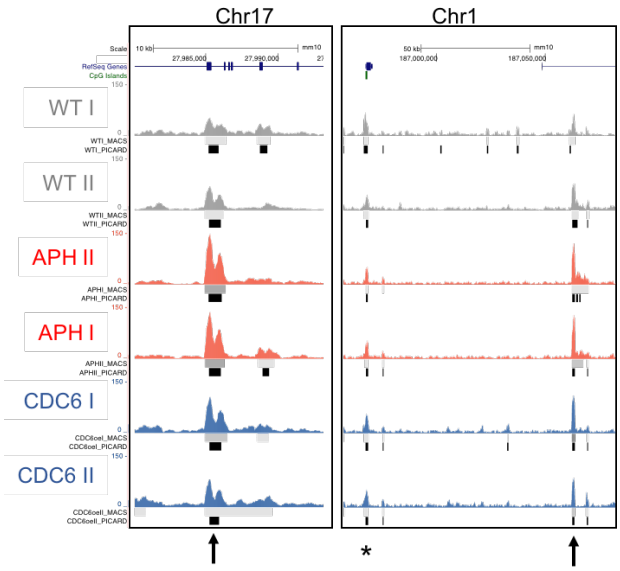


Figure 16. Examples of increased origin efficiency in response to stress. UCSC browser images from two genomic regions containing origins whose efficiency increased upon aphidicolin or CDC6 overexpression (indicated by arrows). The asterisk marks an origin that does not change its efficiency upon stimuli.

Characterization of origin datasets

Correlation with genomic features and replication timing

We next examined the correlation of WT, APH and CDC6 'stringent' origin datasets with specific genomic features. As an important control, 100-1000 datasets of randomized origins were generated in each case and intersected with the same genomic features. The percentages of overlap of randomized origins with the indicated features are depicted in all plots to facilitate the interpretation of the results.

In line with previous origin-mapping studies, origins in all datasets ('stringent' criteria) were found preferentially in promoters and exons. A strong association with CpG islands (CGI) was detected (Fig 17A). The overlap with introns was slightly higher than expected by random and no preference towards transcription termination sites (TTS) was found. Origins were slightly depleted from intergenic regions. The overlap with genomic features was quite comparable in WT, APH and CDC6 origin datasets. Analyses of the "relaxed" origin datasets produced almost identical results (Fig. 17B).

When we examined these correlations in the constitutive and responsive origin subsets, significant differences were found: constitutive origins showed much stronger associations to CGI, promoter and exon regions than responsive ones (Fig. 18A, B). Actually, APH and CDC6 responsive origins did not show a clear enrichment for any specific genomic features, compared to the randomized values. Of note, a modest but detectable enrichment was still found at CGI and promoters. Constitutive

origins showed a lower preference towards intergenic regions in comparison to responsive ones. Again, these effects were observed in the origin datasets defined with both stringent and relaxed criteria. A significant overlap of origins with potential G4 structures was detected in all origin groups analysed (Fig. 19).

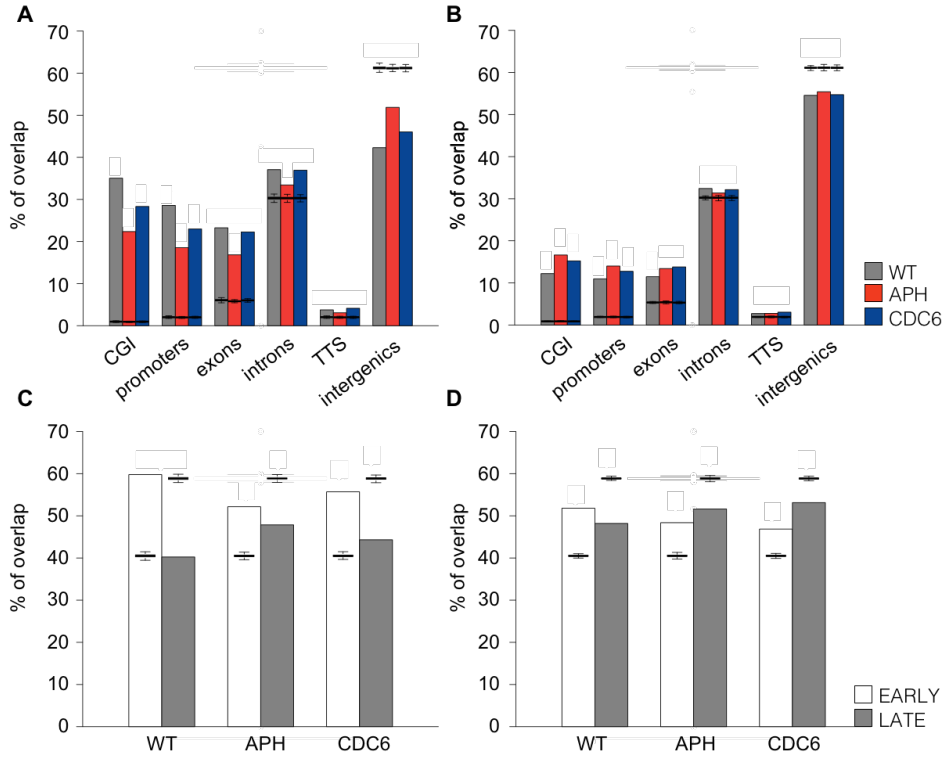


Figure 17. Overlap of origins with genomic features and replication timing regions. (A, B) Histograms show the percentage of overlap of stringent **(A)** or relaxed **(B)** origin datasets with the indicated genomic features. **(C, D)** Histograms show the percentage of overlap of the same origin groups with early and late replication timing regions for stringent and relaxed datasets, respectively. Throughout the figure, horizontal black lines indicate the percentage of overlap of randomized origin groups.

A replication timing (RT) map available for the genome of mouse ESCs (Hiratani et al., 2010) was used for correlative studies with the new origin datasets. WT, APH and CDC6 origins were distributed between early and late RT regions displaying a bias towards early RT domains (Fig. 17C, D). This effect was more pronounced in constitutive origins, which showed a stronger correlation with early RT than responsive ones (Fig 18C, D).

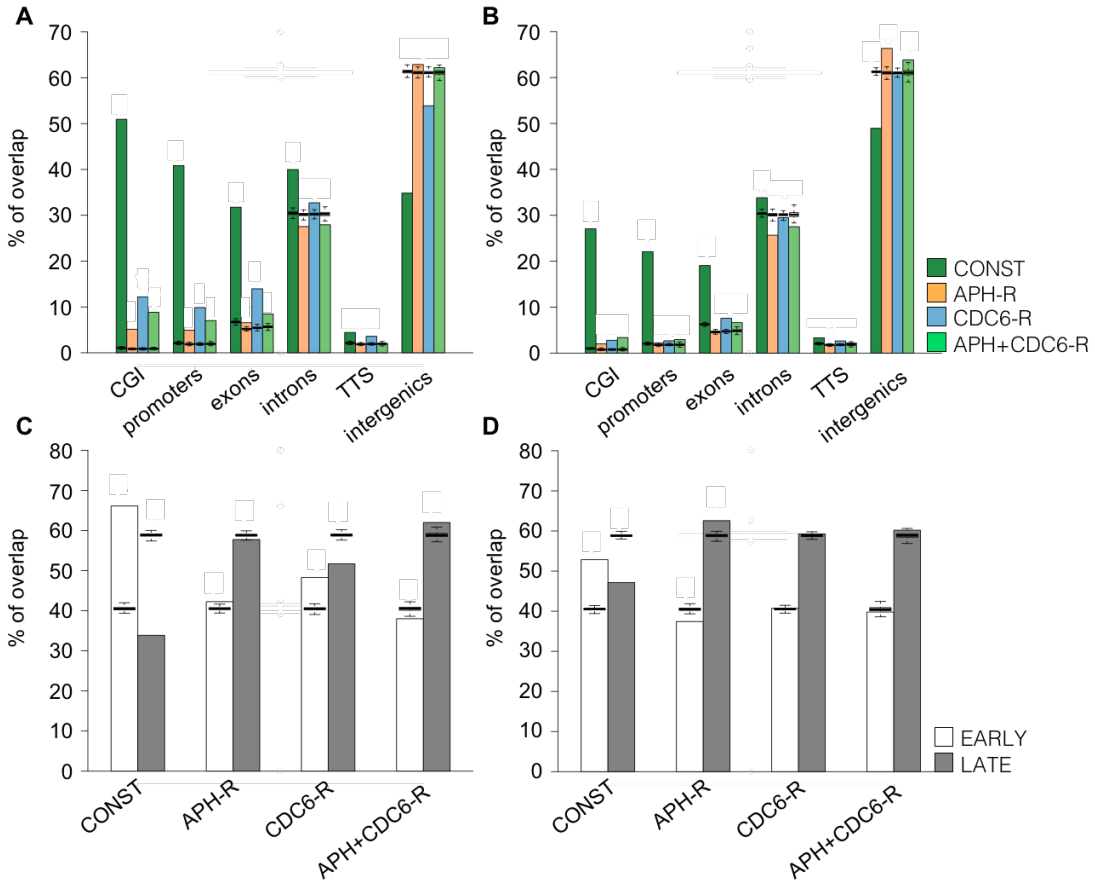


Figure 18. Overlap of constitutive and stress responsive origins with genomic features and replication timing regions. (A, B) Histograms show the percentage of overlap of stringent **(A)** or relaxed **(B)** origin subgroups with the indicated genomic features. **(C, D)** Histograms show the percentage of overlap of the same origin groups with early and late replication timing regions for stringent and relaxed datasets, respectively. Throughout the figure, horizontal black lines indicate the percentage of overlap of randomized origin groups.

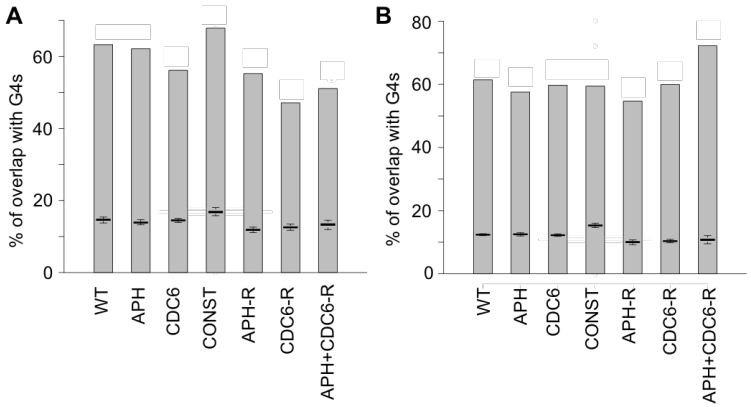


Figure 19. Overlap of origins with G4 structures. Histograms show the percentage of overlap of the stringent **(A)** and relaxed **(B)** origin datasets with potential G4-quadruplex-forming sequences.

In order to visualize the distribution of origins within RT domains, origin density was represented across the genome using 500 kb sliding windows. In the main origin groups (WT, APH and CDC6), a higher density of origins was found in early RT

regions, in line with existing studies (Besnard et al., 2012; Cayrou et al., 2011). Visualization of the data in a genome browser confirmed that constitutive origins tend to localize in early RT zones, whereas responsive ones lay predominantly in late RT domains (Fig. 20, compare CONST and APH+CDC6-R tracks with RT track).

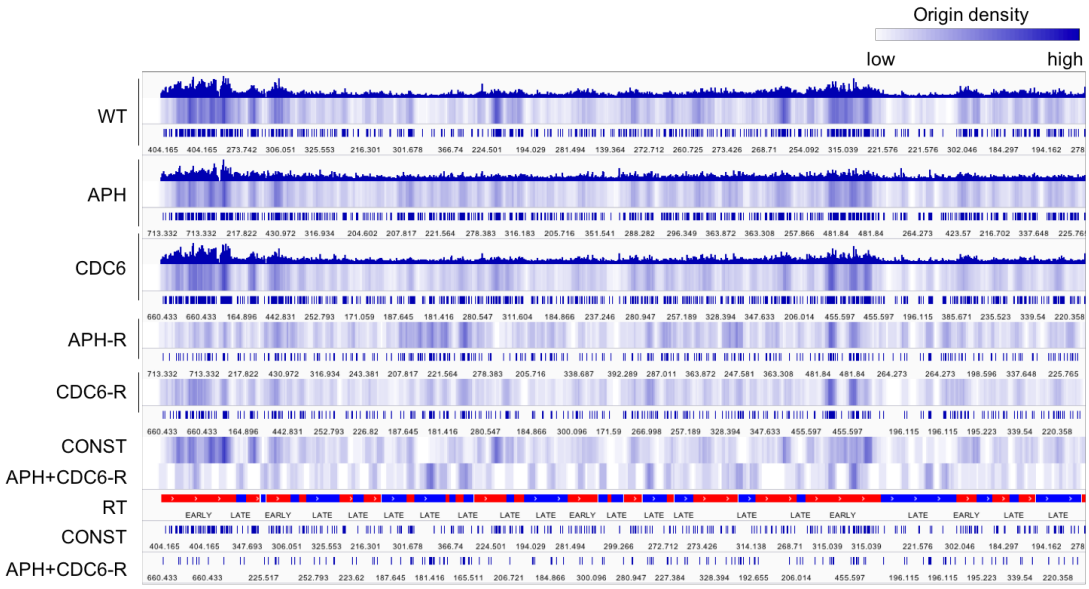


Figure 20. Stress-responsive origins display higher density in late replicating regions then constitutive ones. Image from IGV genome browser showing read density tracks, origin density and peaks (stringent criteria) for the indicated datasets in chromosome 19. Origin density correlates with colour intensity (e.g. darker blue indicates higher density). Density tracks for CONST and APH+CDC6-R origins are represented next to a replication timing (RT) track to facilitate the visualization of constitutive and stress-responsive origins in relation to RT regions.

With these analyses we conclude that constitutive origins can be distinguished from responsive origins by their different efficiency and distinct preference towards a set of genomic features and replication timing regions.

Correlation with epigenetic marks and chromatin states

In order to get novel insights into the characteristics and common features of replication start sites, the origin datasets were intersected with a comprehensive collection of epigenetic data corresponding to >70 chromatin marks recently compiled from >130 ChIP-seq experiments performed in ESCs (Juan et al., 2016). Methodologically, ChIP-seq datasets were 'discretized' in 200 bp windows: the presence of a given epigenetic mark within a window that overlapped with an origin was scored as 1, and its absence as 0. Then, the presence of each epigenomic mark at each origin fragment was quantified (a number between 0 and 1) and averaged

across the dataset. The overlap of origins with epigenetic features was visualized using heatmaps generated by unsupervised hierarchical clustering. The three main datasets (WT, APH, CDC6) showed similar patterns of overlap with a fraction of the ChIP-seq features (Fig. 21). The same pattern was stronger in the subset of constitutive origins. In contrast, the association of origins with these specific epigenomic marks was modest in CDC6-R and barely detected in the APH-R dataset. To establish the significance of these results, similar analyses were performed using randomized origin datasets. The pattern of enrichment was completely lost upon randomization (Fig. 21; Rand WT, Rand APH and Rand CDC6). As it occurred with the genomic features, similar results were obtained when origins where defined with relaxed criteria (Fig. 22). The different extent of overlap with epigenetic features suggests that constitutive and responsive origins have distinct preference towards different functional activities of chromatin.

Next, we calculated the enrichment of epigenomic marks at origins. The enrichment of any particular feature was calculated as the ratio between its average value at origins and its average genomic value. As expected, the enrichment of features at randomized origin datasets was very similar to their average genomic value (Fig. 23). Epigenomic marks were organized into different groups based on their function. Features within the groups were sorted according to their enrichment, ranking from highest to lowest. First, we focused our analyses on the correlations from the three main datasets (WT, APH and CDC6), concentrating on the ≈30 most enriched features (Fig. 24).

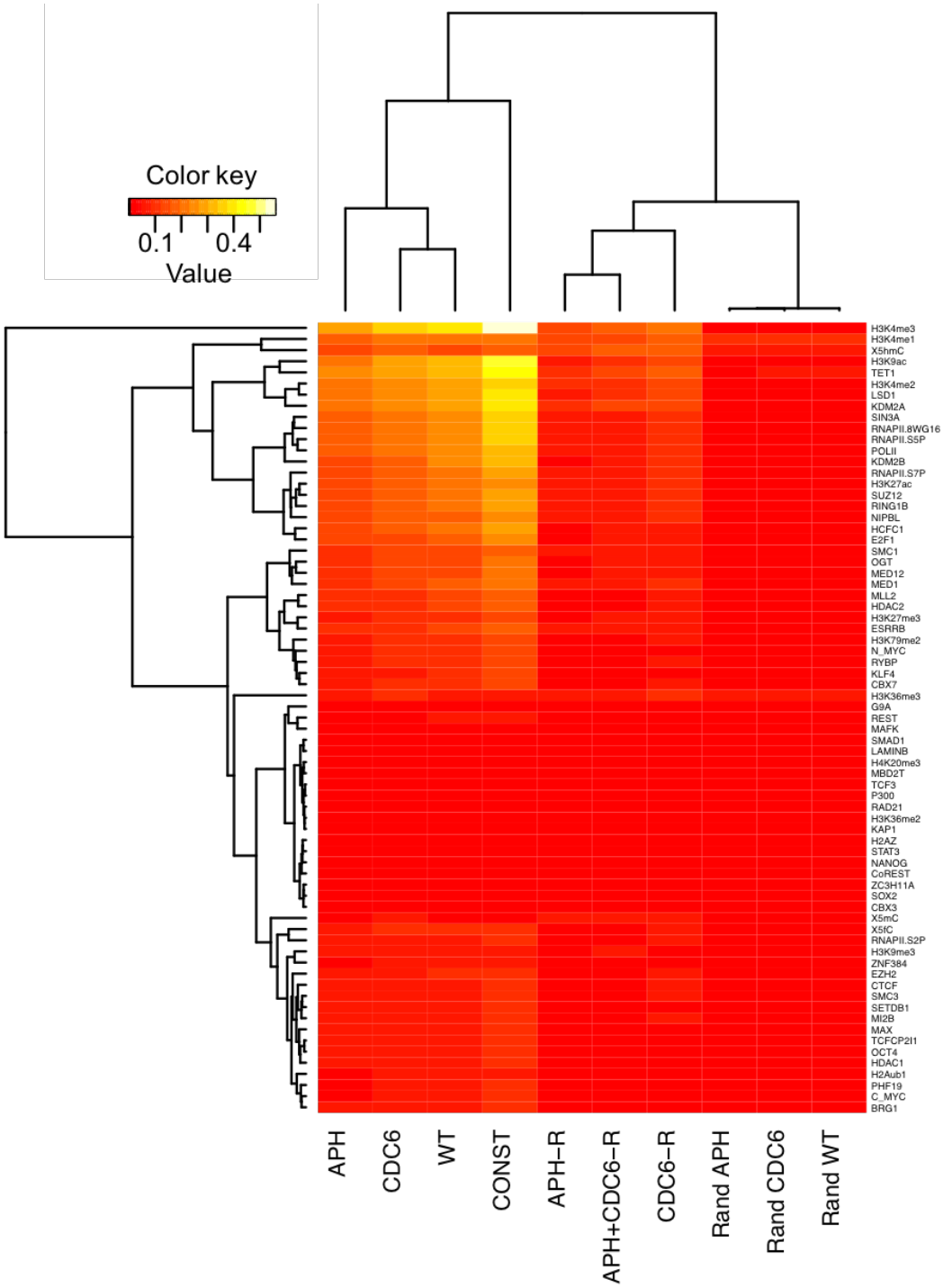


Figure 21. Overlap of origins with epigenomic marks. Heatmap representing the overlap of the indicated origin groups defined by **stringent** criteria with ChIP-seq epigenomic marks from Juan et al. (2016). Heatmap was generated by unsupervised hierarchical clustering. As a control, an overlap of randomized (Rand) origin groups is also represented.

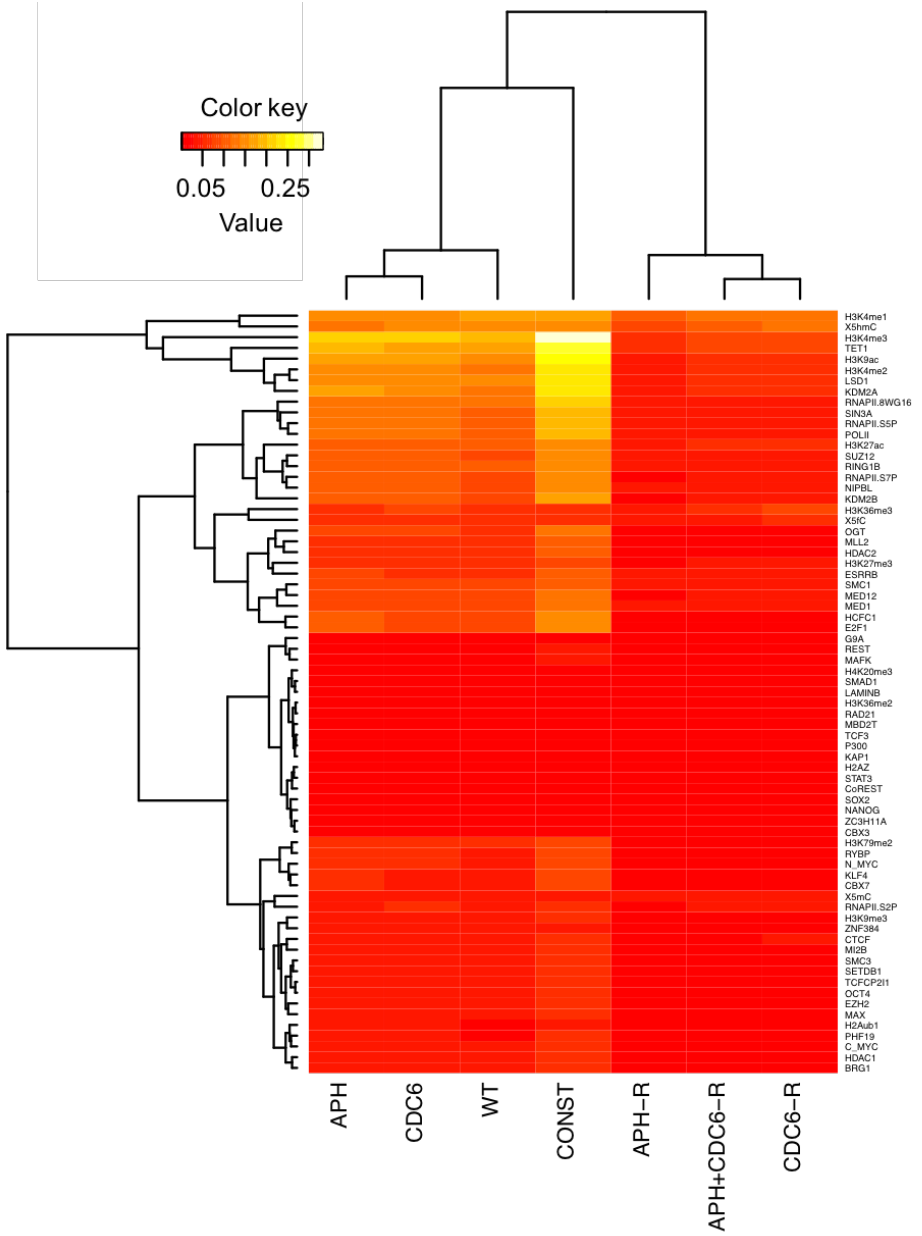


Figure 22. **Overlap of origins with epigenomic marks.** Heatmap representing the overlap of the indicated origin groups defined by **relaxed** criteria with ChIP-seq epigenomic marks from Juan et al. (2016). Heatmap was generated by unsupervised hierarchical clustering.

The overlap between replication origins and transcriptionally active chromatin has been described in several studies (Cayrou et al., 2011, 2015; Martin et al., 2011; Picard et al., 2014; Sequeira-Mendes et al., 2009). Indeed, our analyses indicate that origins from the main datasets (WT, APH, CDC6) correlated with open and transcriptionally active chromatin marks such as H3K9ac and H3K4me3. The latter is a well-established mark of promoters in ESCs (Mikkelsen et al., 2007; Xiao et al., 2012; Zhou et al., 2011). We also found strong correlations with RNA polymerase II (RNAPII-8WG16) as well as its variants RNAPII-S5 and RNAPII-S7, which are enriched at promoters and TSS.

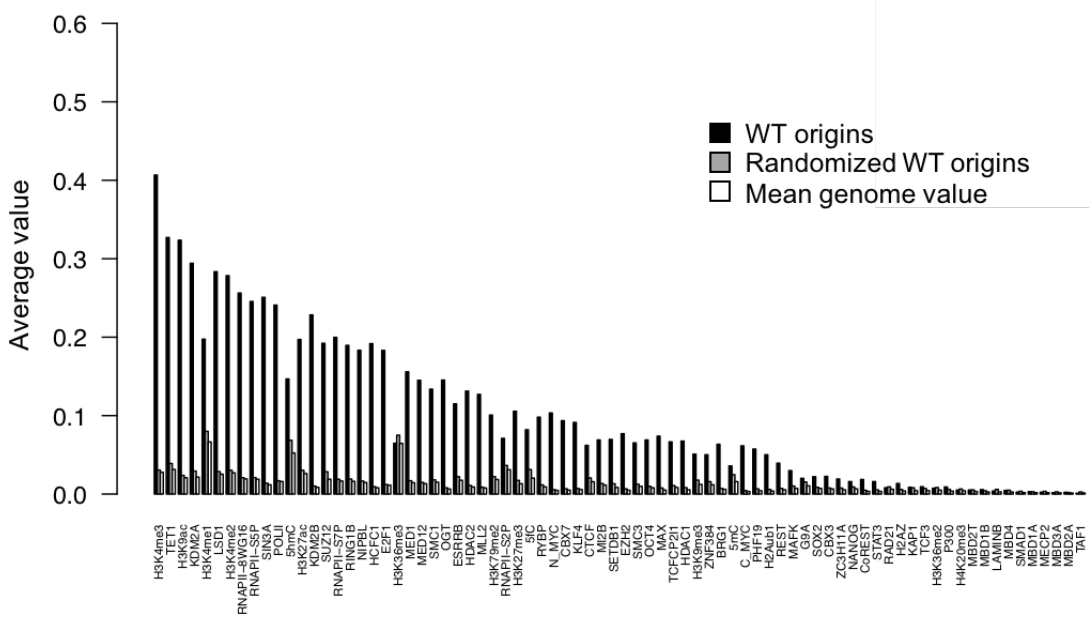


Figure 23. The presence of epigenomic features at randomized origins is similar to their average genome values. Histogram representing average values of epigenomic features at WT origins (black), randomized origins (grey) and the mean genome value of the features (white). The enrichment of any particular feature was calculated as the ratio between its average value at origins and its average genomic value.

Marks with the highest enrichment at origins include histone modifying enzymes known for their association with CpG islands (CGI) such as MLL (H3K4 methyltransferase) and KDM2A and KDM2B (H3K36 demethylases; Dimitrova et al., 2015; Zhang et al., 2015). Several proteins of the Polycomb group were also strongly associated with origins, such as members of Polycomb Repressive Complex 1 (PRC1): CBX7, RING1, RYBP. Monoubuquitination of H2AK119 (H2Aub1), a mark deposited by RING1 was also found. From the Polycomb Repressive Complex 2 (PRC2) complex, we detected enrichment of EZH2 and PHF19 (Sauvageau and Sauvageau, 2010).

Other features highly enriched at origins include TET1, a dioxygenase with the capacity of converting 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), as well as its interactors: OGT, SIN3A and HCFC1. TET-associated complexes are proposed to modulate chromatin conformation and gene expression by different mechanisms (Balasubramani and Rao, 2013).

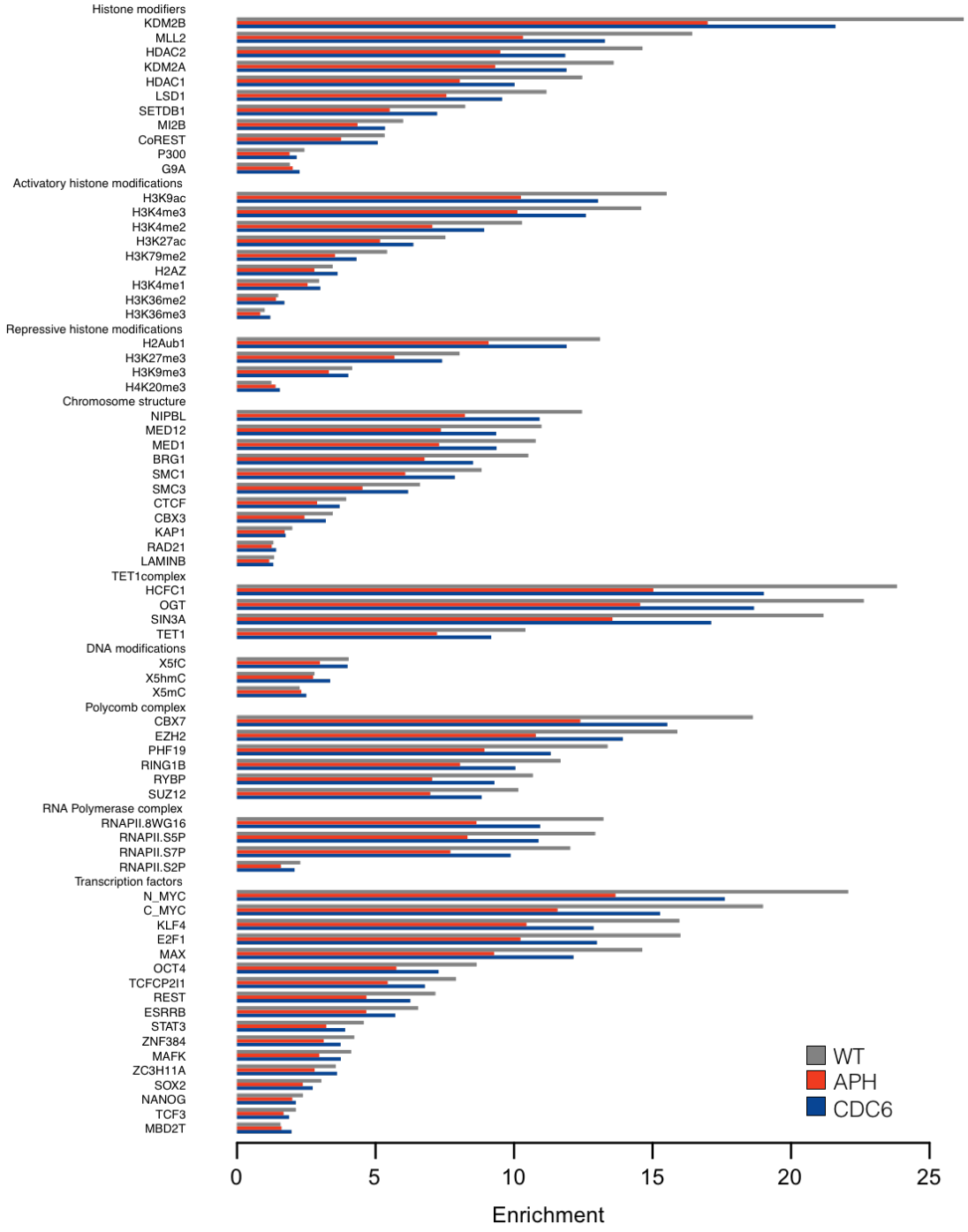


Figure 24. Enrichment of epigenomic marks at WT, APH and CDC6 origin datasets defined with stringent criteria. Epigenomic features (Juan et al., 2016) were organized into different groups based on their function. Features within the groups were sorted according to their enrichment, ranking from highest to lowest.

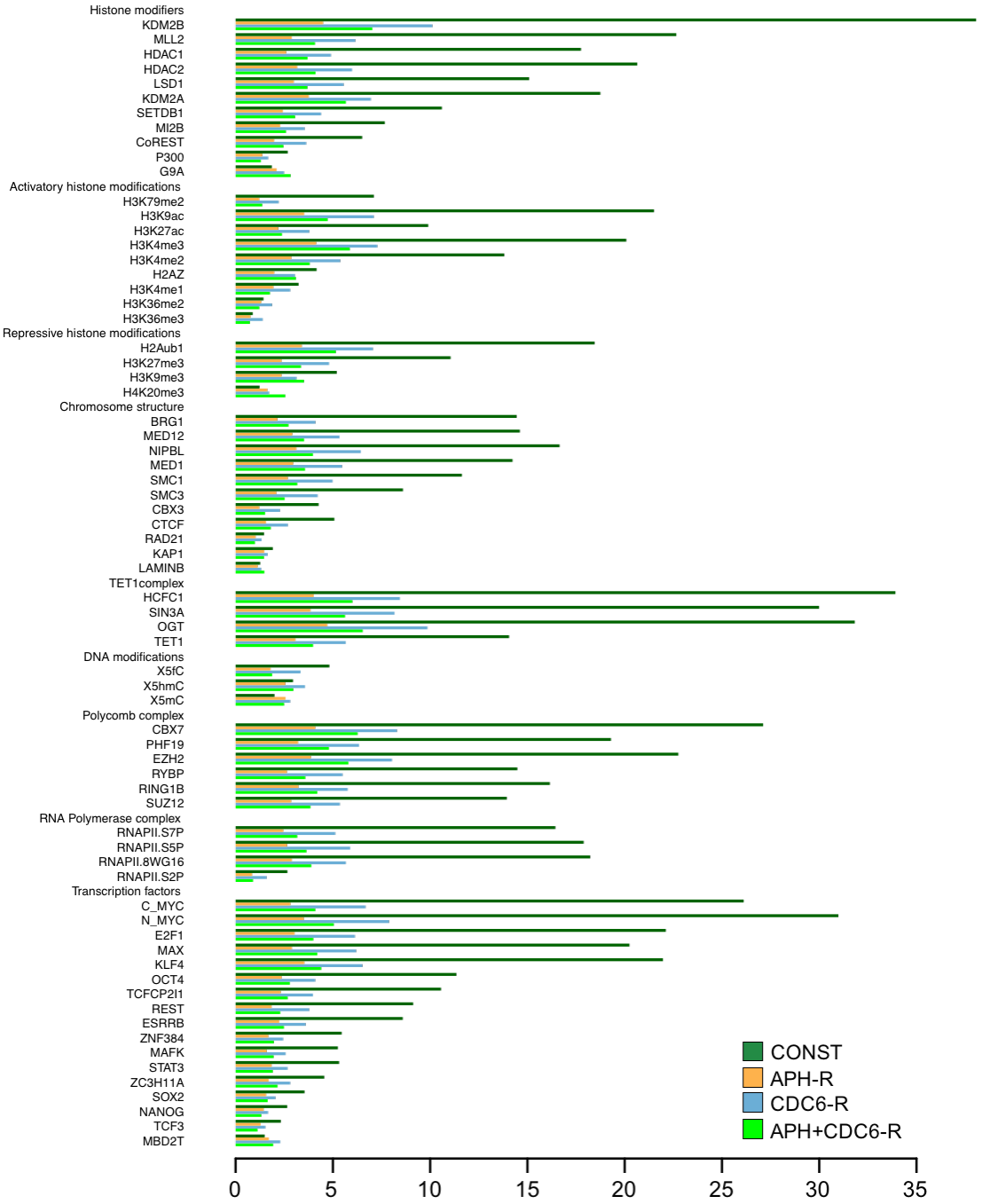


Figure 25. **Enrichment of epigenomic marks at constitutive and stress-responsive origins defined with stringent criteria.** Epigenomic features (Juan et al., 2016) were organized into different groups based on their function. Features within the groups were sorted according to their enrichment, ranking from highest to lowest.

We also found histone deacetylases HDAC1 and HDAC2, the members of the multiprotein complexes CoREST, NuRD and SIN3, which are responsible for transcriptional repression and chromatin remodelling. LSD1 H3K4/K9 demethylase, another member of CoREST and NuRD complexes (Whyte et al., 2012; Yang and Seto, 2008), was also enriched at replication start sites.

Interestingly, within the features highly associated with origins we detected two subunits of the mediator complex (MED1 and MED12) as well as NIPBL, a cohesin-loading factor. This is likely related to the fact that cohesin, mediator and NIPBL colocalize at enhancers and promoters of actively transcribed genes in ESCs (Kagey et al., 2010). Furthermore, transcription factors KLF4, E2F1, N-MYC, C-MYC and MAX colocalized with origins, further linking the choice of replication start sites to the elements that regulate transcription.

Next, we focused our analysis on the differences between constitutive and responsive origins. In Fig. 25, features were ordered by their ratio of enrichment in constitutive versus responsive origins, ranking from highest to lowest. In general, responsive origins displayed much lower enrichment in any ChIP-seq mark than constitutive ones. Therefore, the majority of the features enriched in constitutive vs responsive origins were the same as those features displaying enrichment in WT, APH and CDC6 datasets (Fig. 24 and 25). Among other features preferentially enriched at constitutive origins we detected BRG1 (SWI/SNF chromatin remodelling complex (Wilson and Roberts, 2011), H3K79me2 (a mark associated to transcriptional elongation), OCT4 (a transcription factor and pluripotency marker) and TCFCP2l1 (a transcription factor).

We wondered whether specific groups of origins might be associated with particular functional chromatin states, taking advantage of their recent definition in ESCs (Juan et al., 2016). Chromatin states can be defined as combinations of chromatin marks that can be linked to distinct biological function (Ernst and Kellis, 2010; Filion et al., 2010). In our case, we used a 20 chromatin state model generated using Hidden Markov Model applied to a set of core epigenomic features and genomic annotations. Origins in the main datasets (WT, APH and CDC6; stringent definition) were in general enriched at active and poised promoters, as well as enhancers (Fig. 26A). They showed moderate enrichment in heterochromatin and Polycomb-repressed regions and very limited association with transcriptional elongation and with insulators. In comparison to WT, the APH and CDC6 datasets displayed slightly lower association with active and poised promoters and transcriptional elongation. On the contrary, APH and CDC6 origins were slightly more enriched in heterochromatin (Fig. 26A).

Constitutive, but not responsive origins were strongly enriched with active and poised promoters and to a lesser extent with Polycomb-repressed chromatin (Fig. 26B). In contrast, responsive origins displayed a stronger overlap with heterochromatin.

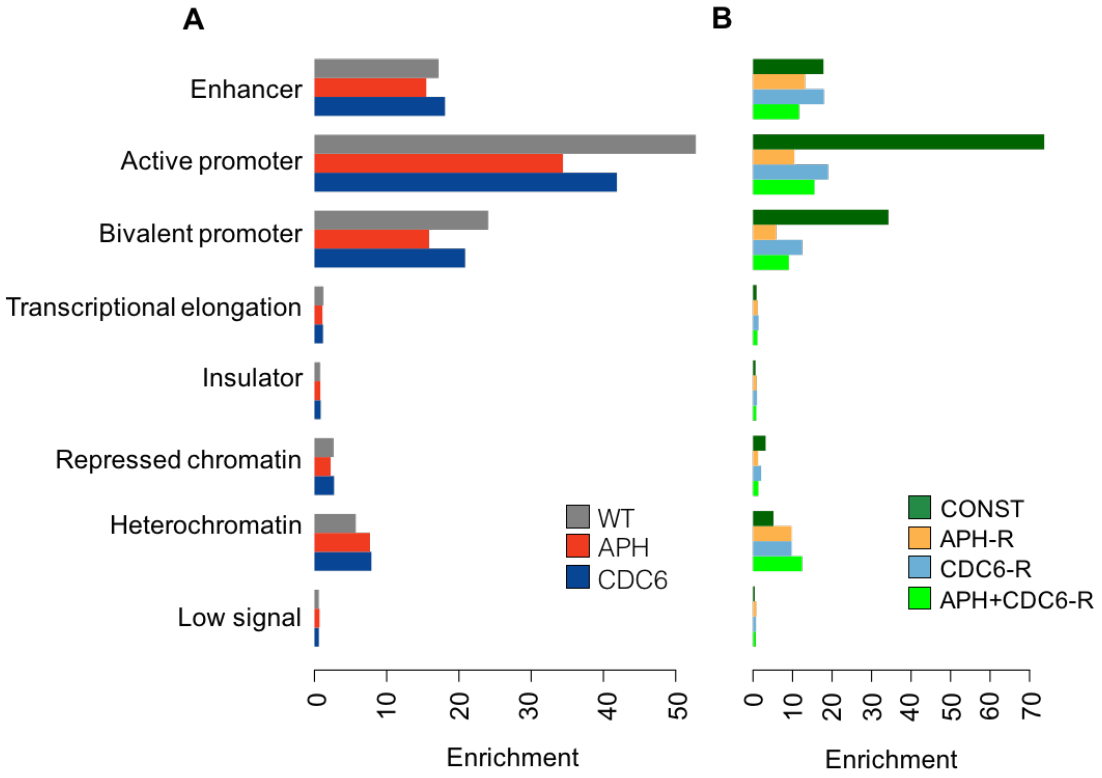


Figure 26. **Enrichment of different chromatin states at origins.** Chromatin states definitions were reported previously (Juan et al., 2016). Origins defined by **stringent** criteria were used for analysis**. (A)** Enrichment of chromatin states in WT, APH and CDC6 origin datasets. **(B)** Enrichment of chromatin states in constitutive and stress-responsive origin groups.

Integration of linear origin maps with 3D chromatin structure

Mapping origins to the promoter-centered chromatin interaction network

To provide a 3D context for the replication start sites identified in this study, origin datasets were integrated with published chromatin contact maps for mESCs. Given the significant correlation of origins with promoters and transcription-associated features, we focused our analysis on a promoter-centered chromatin contact map generated using Promoter-Capture HiC (PCHiC). Capture HiC is an adaptation of Hi-C that includes a step to enrich for interactions involving specific parts of the genome. In the case of PCHiC, these sequences correspond to 22,225 annotated gene promoters (Schoenfelder et al., 2015b), producing a network that includes gene promoters and the genomic regions that interact three dimensionally with them.

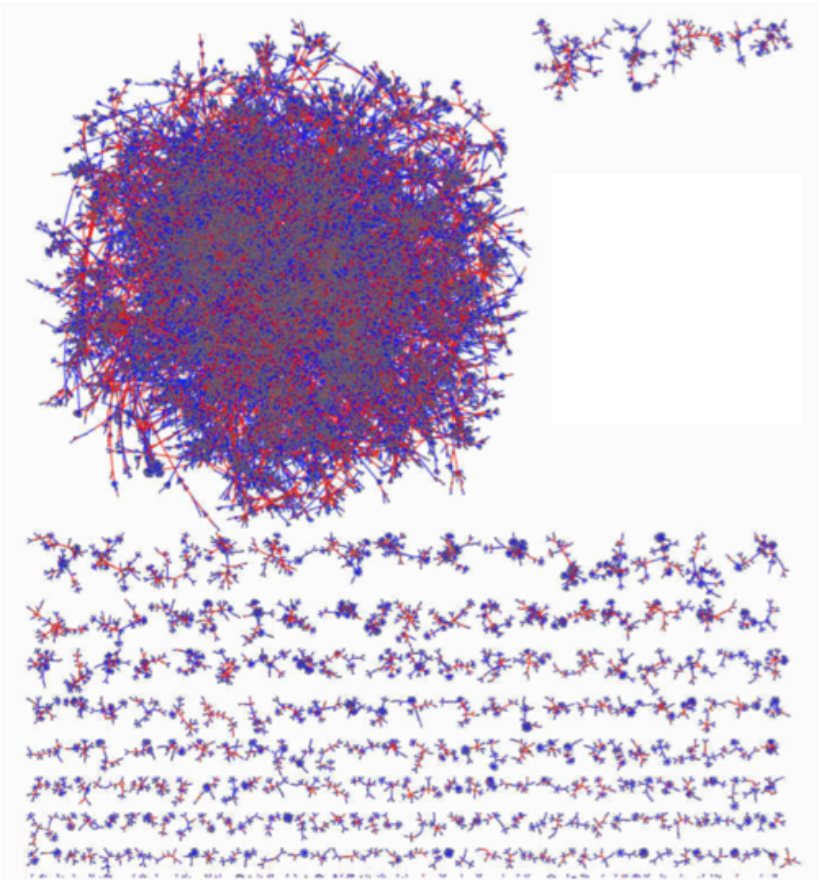


Figure 27. Promoter Capture HiC (PCHiC) chromatin interaction network. Figure taken from Pancaldi et al. (2016). The PCHiC dataset was published by Schoenfelder et al. (2015). Promoter (P) nodes and P-P contacts are depicted in red. Other end (O) nodes and P-O edges are depicted in blue. One large connected component of the network and many small, disconnected components are observed.

Network analysis is a very useful tool to study 3D chromatin interactions (Sandhu et al., 2012). The PCHiC chromatin interactions were processed with CHiCAGO, a method to extract significant long-range contacts from capture HiC datasets (Cairns et al., 2016) and the resulting interactions were represented as a network in which **nodes** are chromatin fragments and **edges** represent a significant 3D interaction between them (Fig. 27, 28; Pancaldi et al., 2016). The **degree** of each node is the number of connections it has with other nodes. PCHiC enriches for long-range interactions (on average 250 kb). Two types of chromatin fragments (nodes) are included: promoters (P) and other-ends (O), which are non-promoter genomic elements (Fig. 28A). Therefore contacts can be defined between two promoters (P-P) and between promoters and other ends (P-O; Fig. 28A). The network can actually be split into two subnetworks, one including only promoter-promoter contacts (P-P, Fig. 28B), the other containing only promoter-other ends contacts (P-O, Fig. 28C). The global PCHiC network consists of 55845 nodes with a median size of 3953 kb. It contains 72231 connections (edges), from which 22767 mediate promoter-promoter

contacts (P-P edges) and 49464 connect promoters with other elements (P-O edges).

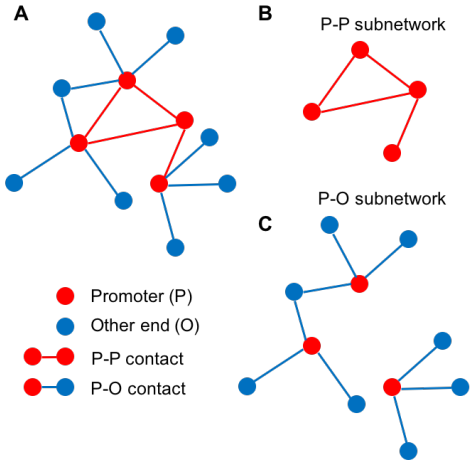


Figure 28. Schematic representing a PCHiC network **(A)** that can be de-composed into P-P subnetwork **(B)** and P-O subnetwork **(C)**. Nodes (chromatin fragments) are represented as circles and edges (contacts between two nodes) as lines.

Origin positions were mapped to the PCHiC network, to identify fragments containing one or more origins. In the stringent dataset, nearly 50% of WT origins and approximately 40% of APH and CDC6 origins localized to fragments contained within this network. The majority (60%) of constitutive origins were positioned within the nodes, as expected given their increased preference for promoters, CGI, and promoter-associated epigenomic marks. In contrast, only approximately 30% of CDC6-R and 22% of APH-R and APH+CDC6-R were located within this network. In all groups, the number of origins was comparable between the P-P and P-O subnetworks (Table 7). Taking into account that the number of P-O contacts in the network more than doubles the number of P-P interactions, origins show preference towards the P-P subnetwork.

	WT	APH	CDC6	CONST	APH-R	CDC6-R	APH+ CDC6-R
Total origin No	20164	31627	31360	11989	17264	17264	8234
PCHiC	9790 (49%)	11693 (37%)	13598 (43%)	7188 (60%)	3867 (22%)	5564 (32%)	1840 (22%)
PCHiC in P-P	6444 (32%)	6907 (22%)	8301 (26%)	5215 (43%)	1489 (9%)	2534 (14%)	841 (10%)
PCHiC in P-O	6906 (34%)	7924 (25%)	9341 (30%)	5254 (44%)	2288 (13%)	3472 (20%)	1146 (14%)

Table 7. Overlap of origins (stringent dataset) with PCHiC network and its P-P and P-O subnetworks. In brackets the percentage of overlap is annotated. The total number of origins for each origin group is indicated. These numbers differ slightly from the values found in Table 6 because of the conversion of the data from the mm10 to the mm9 mouse genome assembly.

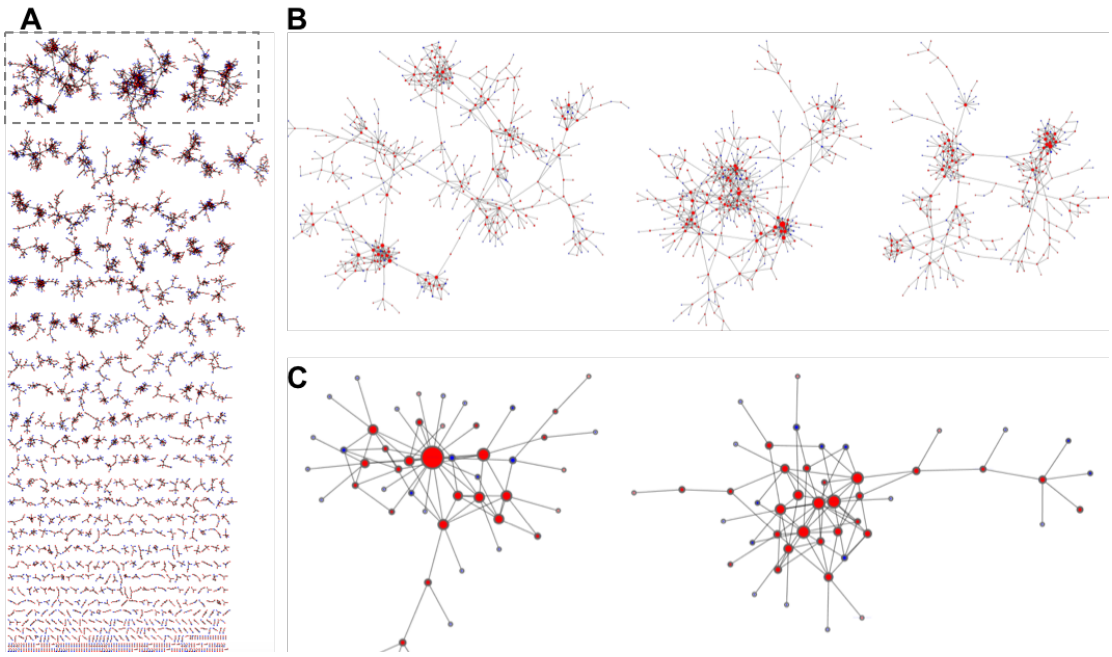


Figure 29. Origins tend to form clusters in 3D. (A) PCHiC origin subnetwork. **(B)** Magnification of a fragment of the origin subnetwork marked in (A) by dashed line. **(C)** Examples of two hubs from the PCHiC origin subnetwork. Red - promoter nodes, blue - other end nodes. The size of the node is proportional to its degree.

Next, we focused on the PCHiC origin subnetwork, i.e. the network where all nodes contain one or more origins (Fig. 29A). We noticed that this network is highly clustered, forming hubs (highly connected nodes) of chromatin fragments containing origins (Fig. 29A, B, C). We hypothesized that these clusters could correspond to replication factories, in which origins separated in the linear genome are brought together in 3D, to become activated simultaneously during S phase.

Nodes containing origins have a higher degree than average nodes

To establish whether there are any particular characteristics of chromatin fragments containing origins, we first calculated their degree (i.e. number of connections) distribution. Remarkably, nodes containing origins displayed higher connectivity (mean degree 4.6 for WT and 4.4 for APH and CDC6 origins) relative to the mean degree of 2.6 for the entire PCHiC network (Fig. 30).

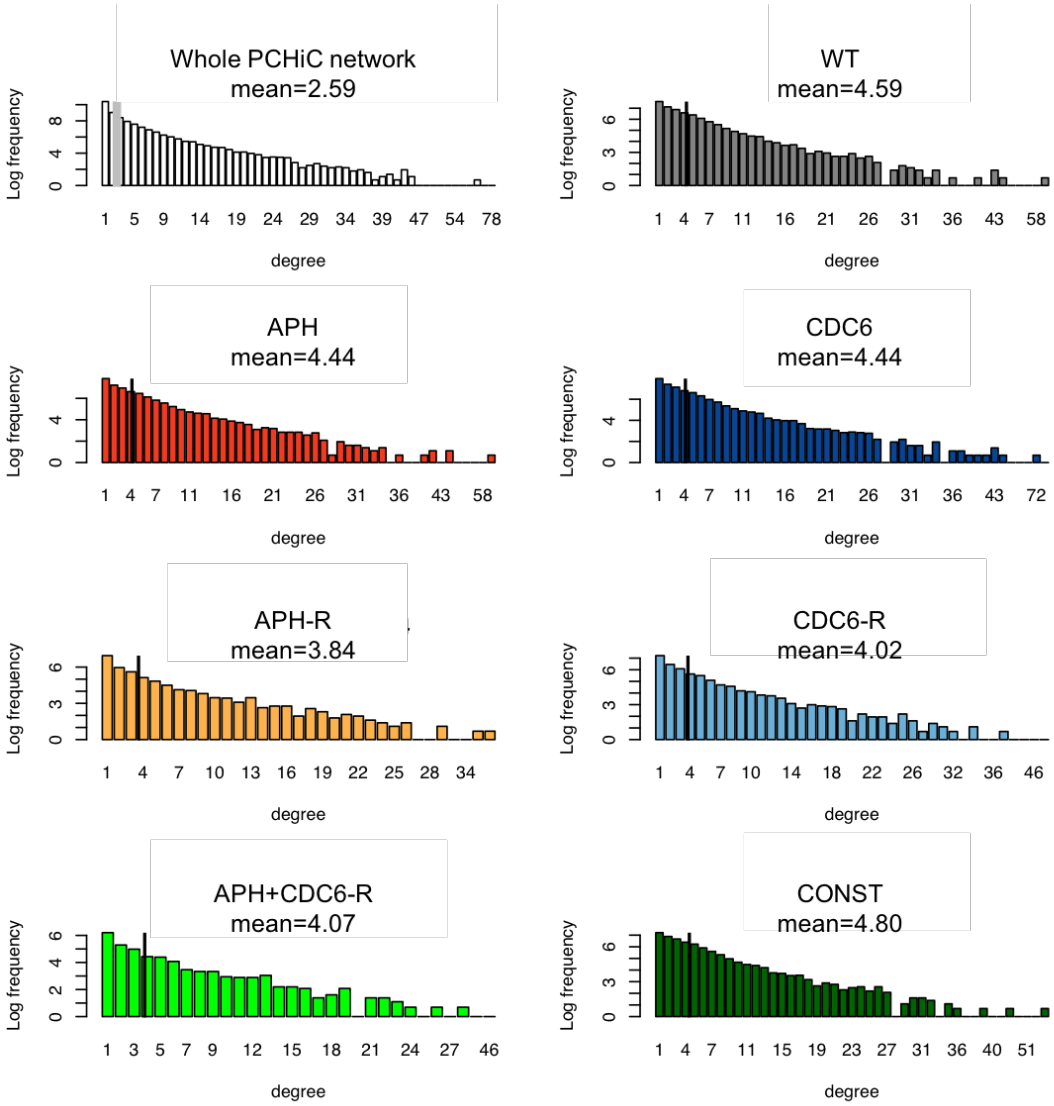


Figure 30. Chromatin fragments containing origins show higher connectivity. Distribution of degrees of the nodes containing origins in the PCHiC network. The mean degree value of chromatin fragments containing the indicated type of origins is depicted. The distribution and mean degree of the whole PCHiC network is also shown (top left).

In addition, those chromatin fragments overlapping with constitutive origins presented the highest mean degree (4.8), whereas fragments containing responsive origins showed lower connectivity (mean degree in the range 3.8 - 4; Fig. 30).

Origin connectivity correlates with efficiency and replication timing

Constitutive origins are also the most efficient (Fig. 15), suggesting a direct correlation between origin efficiency and connectivity. To test this hypothesis, the distribution of origin efficiencies was plotted in relation to the degree of fragments containing origins. In all datasets analysed, a significant positive correlation between efficiency and node degree was observed (Fig. 31). The same results were obtained with the relaxed origin datasets (not shown).

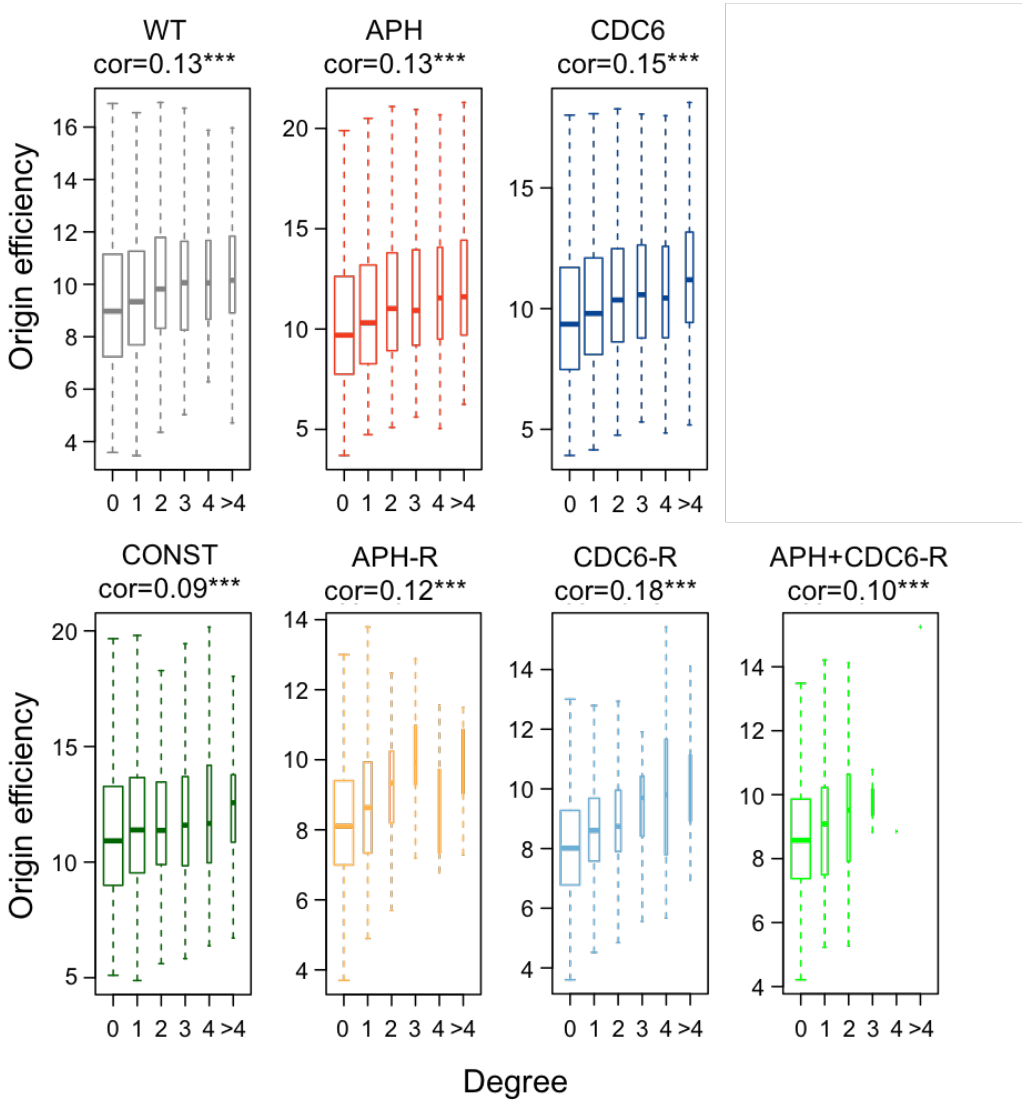


Figure 31. More connected origins are more efficient. Distribution of origin efficiency in relation to the degree of PCHiC fragments containing origins belonging to the indicated groups. Origins defined by **stringent** criteria were analysed. The width of the bars is proportional to the number of elements in the bin. ***$p<10^{-16}$ in Pearson correlation coefficient. cor, correlation.

Next we established the possible correlation with replication timing (RT) of the network fragments using the RT data available (Hiratani et al., 2010). When the RT of chromatin fragments in the network was represented relative to their degree of connectivity, a significant correlation was obtained: in all origin groups analysed, those more connected displayed earlier RT (Fig. 32).

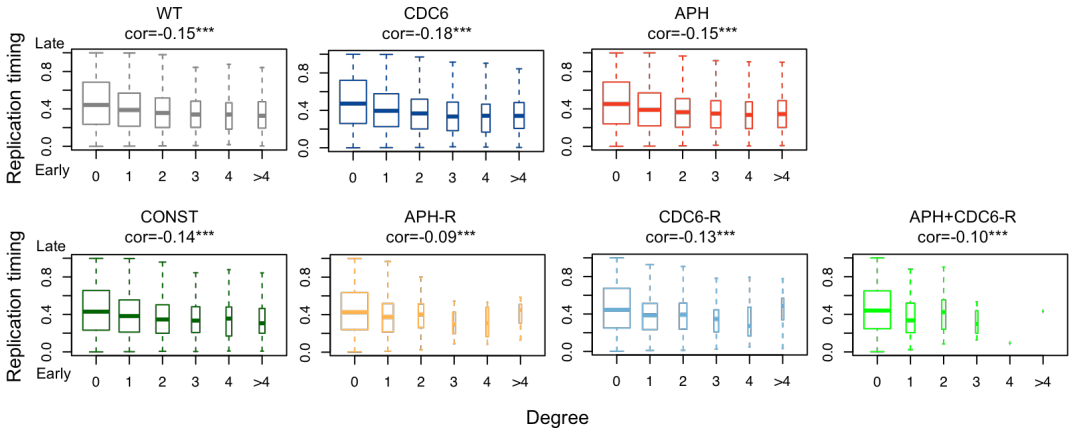


Figure 32. More connected origins replicate earlier. Distribution of replication timing of origins in relation to the degree of PCHiC fragments containing origins belonging to the indicated groups. Origins defined by **stringent** criteria were analysed. For replication timing, early=0 and late =1. The width of the bars is proportional to the number of elements in the bin. ***$p<10^{-16}$ in Pearson correlation coefficient.

A significant correlation was also observed between RT and origin efficiency in the PCHiC network in all types of origins (Fig. 33). Taken together, these results indicate that origins with high connectivity in the promoter-centered chromatin interaction network are likely to fire with high efficiency, early in S phase.

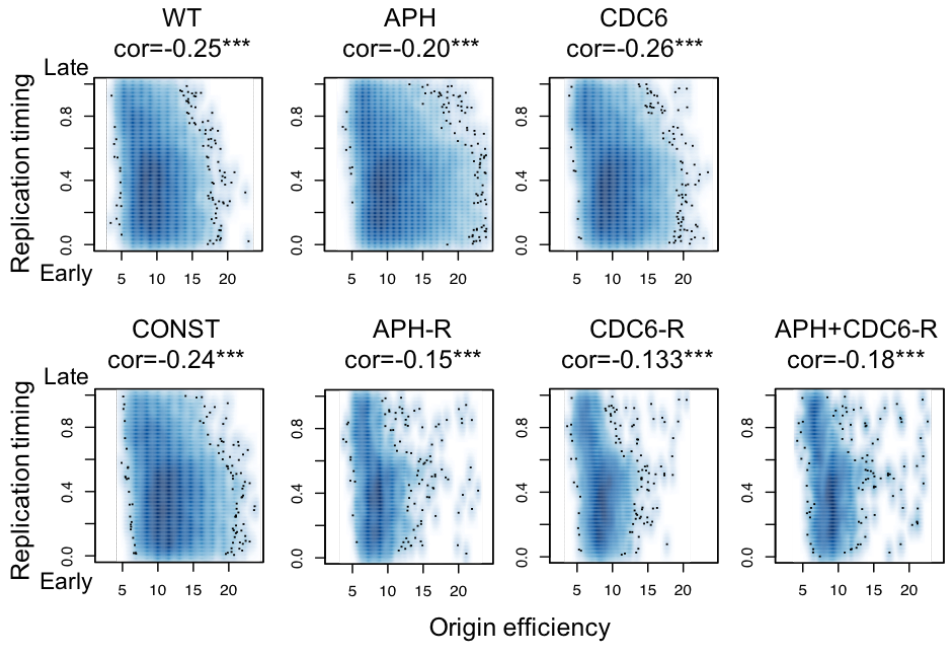


Figure 33. Origins that replicate earlier in S phase tend to be more efficient. Graphs show in cloud form the distribution of origin efficiency in relation to their replication timing in the indicated groups of origins. Origins defined by **stringent** criteria were analysed. For replication timing early=0 and late =1. ***$p<10^{-16}$ in Pearson correlation coefficient.

Origin-origin 3D interactions occur preferentially within TADs

Topologically associated domains (TADs), proposed to be the structural units of chromatin organization (Dixon et al., 2012), are known to correspond to RT domains (RD; Pope et al., 2014). In order to study the relationship between origin-origin interactions in the PCHiC network and TADs, we assigned each chromatin fragment to its corresponding TAD and analysed whether ori-ori contacts are found within TADs (intra-TAD) or whether they span different TADs (inter-TAD). In the WT origin subnetwork (stringent criteria), the majority (63%) of ori-ori contacts were intra-TAD. This is expected, as according to the replication factory concept, interacting origins should display similar RT, which maintained within TADs. We found, however, that approximately one third of ori-ori contacts were inter-TAD. This finding could still be compatible with the replication factory model as long as the TADs with connected origins displayed comparable RT. To address this issue, we calculated the average difference in RT between two contacting origins belonging to intra-TAD, inter-TAD or random ori-ori interactions in the WT origin subnetwork (Fig. 34). The smallest difference in RT was detected for intra-TAD contacts. This difference was slightly higher for inter-TADs interactions, but still clearly below the difference observed in a random ori-ori interaction set. These data suggest that two interacting origins, independently of whether they belong to the same TAD or two separate TADs, tend to replicate synchronously.

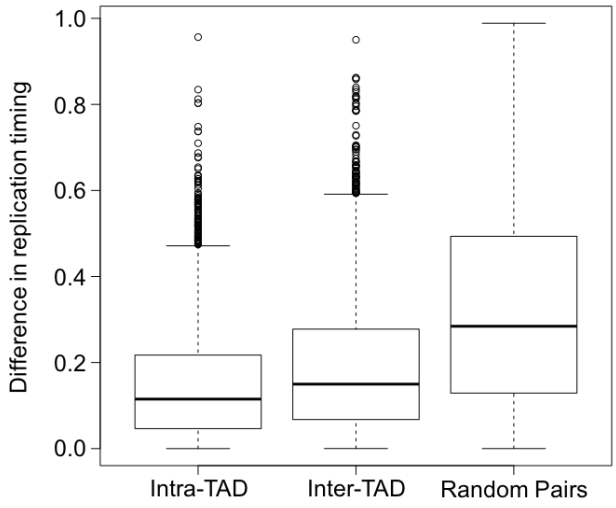


Figure 34. Origins brought together by intra- or inter-TADs interactions display similar replication timing. Box plots show the distribution of differences in replication timing between two interacting PCHiC fragments containing origins. Intra- and inter-TADs contacts were evaluated separately. Origins defined with stringent criteria were analysed. As a control, the distribution of differences in RT between random pairs of fragments is also depicted. Black horizontal lines represent the median difference in RT.

Integration of origin maps with 3D chromatin structure using chromatin assortativity

To further characterize the integration of replication origins in the 3D chromatin structure, we took advantage of a recently developed parameter called **chromatin**

assortativity (ChAs) that enables the integration of linear genomic information with 3D chromatin interaction networks (Pancaldi et al., 2016). The concept of assortativity is derived from social networking studies; when applied to chromatin, it identifies to what extent a given property of any chromatin fragment (e.g. the presence of a specific epigenetic feature) is shared by those fragments that interact with it. ChAs is a correlation coefficient, ranging between -1 and 1, and it is a powerful tool to identify epigenomic features which are associated to 3D chromatin topology (Pancaldi et al., 2016).

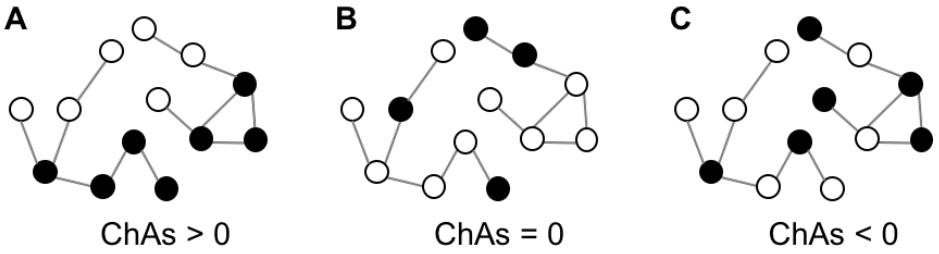


Figure 35. Chromatin Assortativity. (A-C) Schematic illustrating possible outcomes of chromatin assortativity (ChAs) analysis. Black and white nodes represent chromatin fragments that contain, or not, a particular feature, respectively. Edges represent 3D interactions between nodes. Adapted from Pancaldi et al. (2016). See text for details.

For any genome feature, ChAs can fall within three categories: **ChAs>0**, indicating that network fragments containing a particular feature tend to interact with each other, and less with fragments not containing this feature (Fig. 35A); **ChAs=0**, indicating no relationship between the presence of a particular feature in the fragments and its presence in their interacting fragments (Fig. 35B); **ChAs<0,** which would indicate that fragments containing a certain feature interact preferentially with fragments not containing the same feature (Fig. 35C).

Chromatin fragments containing origins tend to interact in 3D

Using assortativity we first determined whether chromatin fragments that contain origins tend to interact with each other. To this aim, **origin assortativity (OriAs**) was measured in the PCHiC network, subdivided into the P-P and P-O subnetworks. The three origin datasets (WT, APH, CDC6) as well as the constitutive origin subset displayed positive assortativity in the P-P subnetwork (Fig. 36A). In contrast, responsive ones were not assortative in this subnetwork. As a control reference, assortativity was not found in multiple sets of randomized origins, represented in the same graph as open circles.

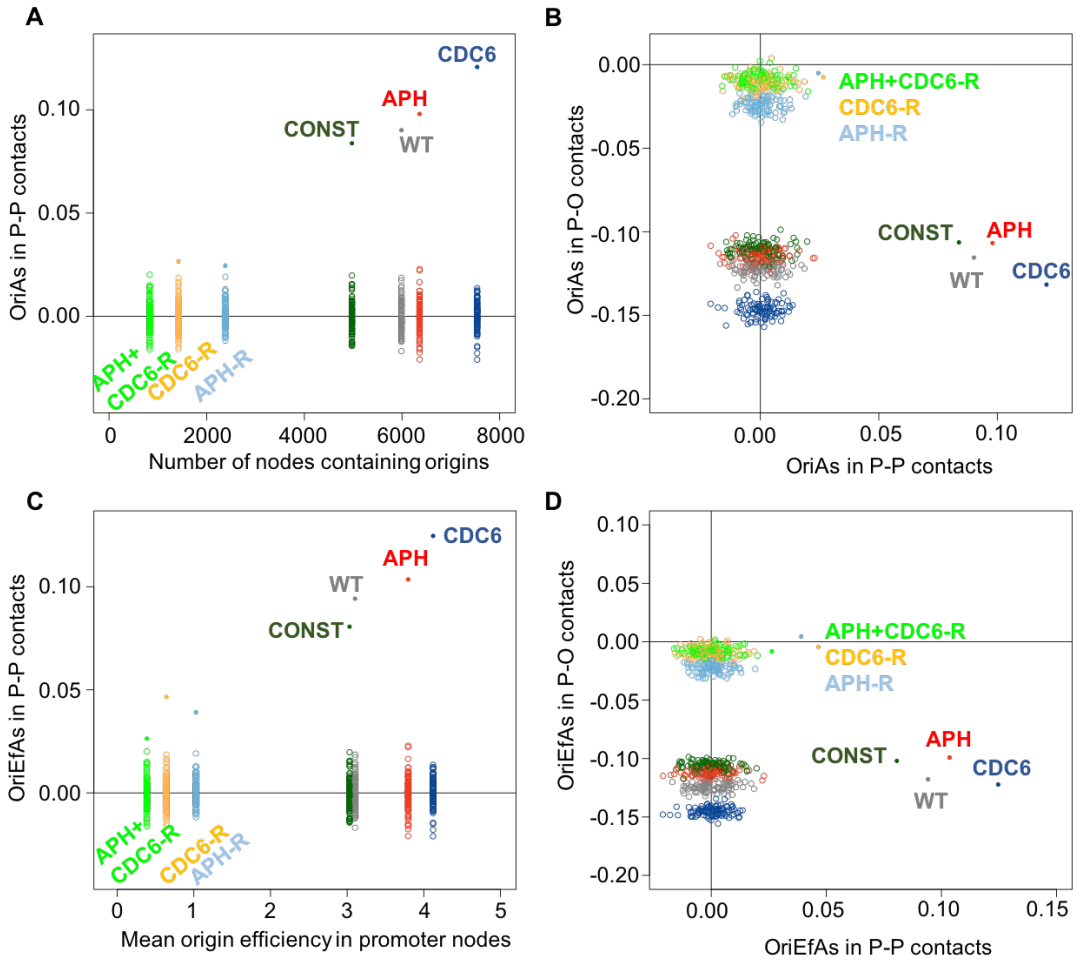


Figure 36. Origins are assortative in the promoter-promoter PCHiC subnetwork. Assortativity was calculated for origins defined by **stringent** criteria**. (A)** OriAs measured for the indicated groups of origins. Filled circle represents the assortativity value. As a control, OriAs for randomized origin groups are also depicted (clouds consisting of open circles). **(B)** Comparative plot of OriAs in P-O vs P-P PCHiC subnetwork. **(C)** OriEfAs of the indicated groups of origins. Filled circle represents the assortativity value. As a control, OriEfAs for randomized origin groups are depicted (clouds consisting of open circles) **(D)** Comparative plot of OriEfAs in P-O vs P-P PCHiC subnetwork.

Next, we examined OriAs between contacts involving promoter and other element fragments. To this aim, OriAs in the P-P and P-O subnetworks were combined in one plot, allowing for comparison of assortativity values (Fig. 36B). As shown in the previous graph, WT, APH, CDC6 and constitutive origins are assortative in the P-P subnetwork, while responsive origins were not. In contrast, assortativity of WT, APH, CDC6 and constitutive origins in the P-O subnetwork was not significantly different to the values presented by sets of randomized origins (Fig. 36B). We concluded that WT, APH, CDC6 and constitutive origins were not assortative in the P-O subnetwork.

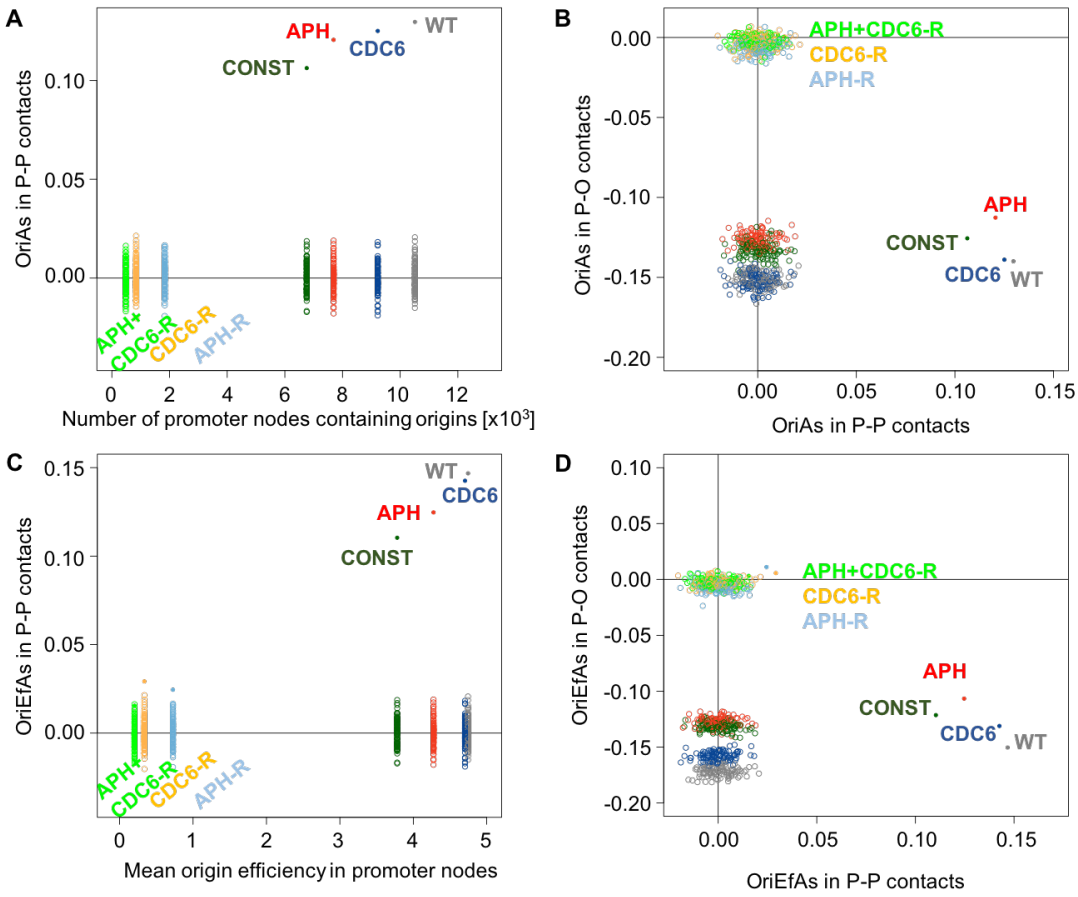


Figure 37. Confirmation of origin assortativity in the promoter-promoter PCHiC subnetwork. Assortativity was calculated for origins defined by **relaxed** criteria**. (A)** OriAs measured for the indicated groups of origins. Filled circle represents the assortativity value. As a control, OriAs for randomized origin groups are also depicted (clouds consisting of open circles). **(B)** Comparative plot of OriAs in P-O vs P-P PCHiC subnetwork. **(C)** OriEfAs of the indicated groups of origins. Filled circle represents the assortativity value. As a control, OriEfAs for randomized origin groups are depicted (clouds consisting of open circles). **(D)** Comparative plot of OriEfAs in P-O vs P-P PCHiC subnetwork.

Origin efficiency is an assortative parameter

To test whether connected origins are likely to fire with similar efficiencies, we calculated the assortativity of origin efficiency **(OriEfAs).** Interestingly, efficiency was assortative in the P-P subnetwork for WT, APH, CDC6 datasets as well as the subset of constitutive origins (Fig. 36C, D). As it was the case with OriAs, OriEfAs was not different from random for the subsets of responsive origins. In the P-O subnetwork, origin efficiency did not show assortativity above the random value in any of the datasets (Fig. 36D). This data strongly suggests that WT, APH, CDC6 and constitutive origins found at promoters that contact each other tend to be activated with similar efficiencies.

The patterns of OriAs and OriEfAs were confirmed with extended datasets created relaxed criteria (Fig. 37).

Origin assortativity in RNAPII and SMC1 networks

To further validate the notion of origin connectivity in the 3D chromatin structure, we took advantage of other chromatin interaction maps available for ESCs such as those mediated by RNA Polymerase II (RNAPII) or cohesin SMC1 subunit. Chromatin connectivity maps for both proteins had been generated using ChIA-PET (chromatin interaction analysis with paired-end tagging; Dowen et al., 2014; Zhang et al., 2013). This assay involves pull-down of chromatin fragments bound to a specific protein of interest and interacting with each other.

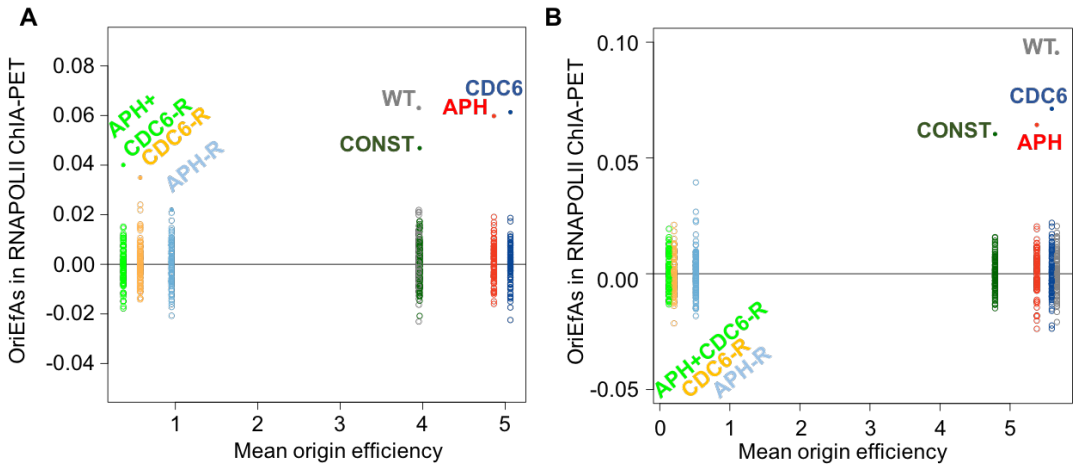


Figure 38. Origin efficiency is assortative in the RNAPOLII ChIA-PET chromatin network. OriEfAs measured for the indicated groups of origins defined by **(A) stringent** and **(B) relaxed** criteria. Filled circle represents the assortativity value. As a control, OriEfAs for randomized origin groups are also depicted (clouds consisting of open circles).

The percentages of origins that localize within the fragments of these networks were lower in comparison to the PCHiC network (Table 8), mostly due to the low genome coverage of the ChIA-PET assays. In the RNAPII network, which is enriched in active promoter interactions, we detected positive values of OriAs (not shown) and OriEfAs in WT, APH, CDC6 and constitutive origins in both the stringent and relaxed datasets (Fig. 38). This indicates that origins are frequently located at chromatin interactions mediated by RNAPII and that contacting origins tend to share similar efficiency.

%	WT	APH	CDC6	CONST	APH-R	CDC6-R	APH+ CDC6-R
RNAPII	21.3	14.4	17.1	29.7	5.09	8.34	5.95
SMC1	10.6	7.91	9.51	14.1	4.12	6.58	4.42
HiCap	25.1	17.5	21.4	34.2	6.45	11.7	7.85

Table 8. Overlap of origins (stringent dataset) with RNAPII, ChIA-PET, SMC1 ChIA-PET and HiCap networks. The percentage of total origins that overlap with chromatin fragments from each network is indicated.

Cohesin participates in the organization of chromatin loops at replication factories (Guillou et al., 2010). It has been proposed that cohesin-stabilized DNA loops bring origins together to be activated simultaneously. Indeed, values of OriAs (not shown) and OriEAs in WT, APH, CDC6 and constitutive origins (both stringent and extended datasets) were higher than expected by random in the SMC1 network (Fig. 39). This result implies that origins also tend to localize at chromatin interactions mediated by cohesin and that origins brought together by these interactions in 3D become activated with similar efficiencies.

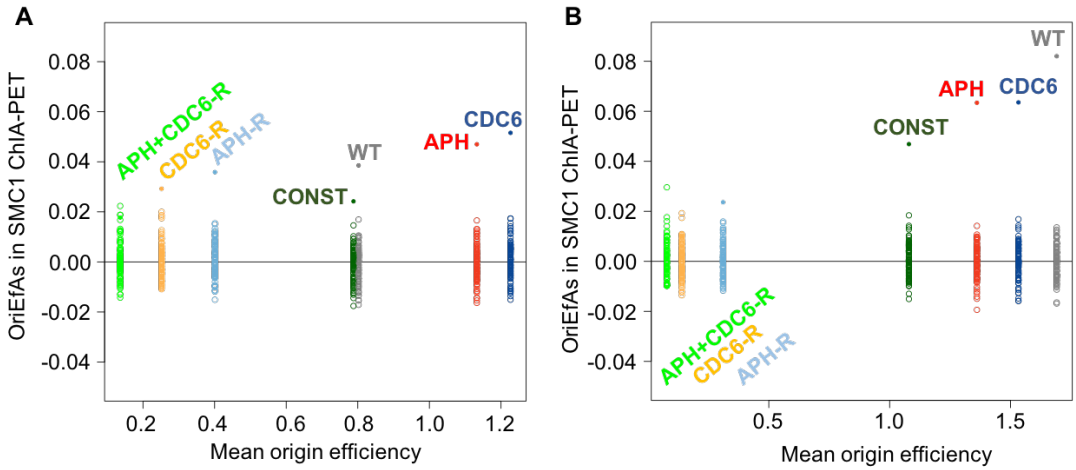


Figure 39. Origin efficiency is assortative in the SMC1 ChIA-PET chromatin network. OriEfAs measured for the indicated groups of origins defined by **(A)** stringent and **(B)** relaxed criteria. Filled circle represents assortativity value. As a control, OriEfAs for randomized origin groups are depicted (clouds consisting of open circles).

Origins connected in 3D tend to be separated by long-range distances

Assortativity was also analysed in a promoter-capture chromatin connectivity map obtained using HiCap (Sahlén et al., 2015). In this approach, a higher resolution 3D map was obtained (fragments have an average ≈0,6kb length in comparison to ≈4kb

in the PCHiC network). This alternative protocol generates shorter interaction fragments, increasing the coverage of short-range interactions but reducing the coverage of long-range ones. Surprisingly, our initial analyses in the HiCap network failed to detect OriAs and OriEfAs above the values expected by random (Fig. 40). We hypothesized that lack of assortativity in this chromatin connectivity map could be due to the fact that long-distance interactions are needed to detect assortativity.

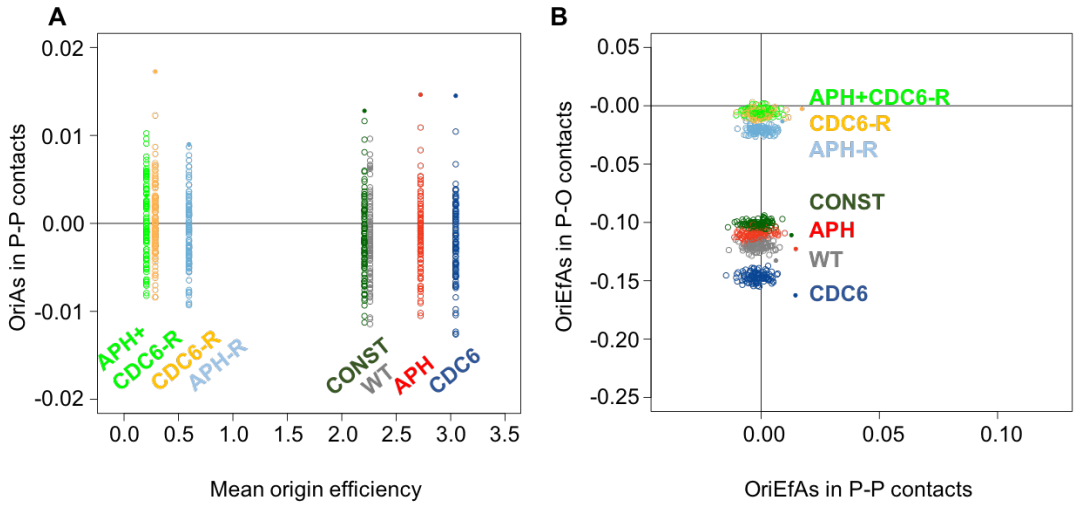


Figure 40. Origin efficiency is not assortative in the promoter-promoter HiCap subnetwork. Assortativity was calculated for origins defined by **stringent** criteria**. (A)** OriEfAs measured for the indicated groups of origins. **(B)** Comparative plot of OriEfAs in P-O vs P-P PCHiC subnetwork. Filled circle represents assortativity value. As a control, OriEfAs for randomized origin groups are depicted (clouds consisting of open circles).

Therefore, OriEfAs values were calculated in the PCHiC network at different ranges of linear separation between contacting fragments by gradually eliminating contacts between genomically “close” regions (Fig. 41A). Interestingly, OriEfAs peaked around 500kb, indicating that this is the most common linear separation between origins that interact in 3D and that origins separated by this distance are most likely to have the same efficiency (Fig. 41A). This distance is significantly higher than the average distance between adjacent origins estimated by other techniques (50-150kb).

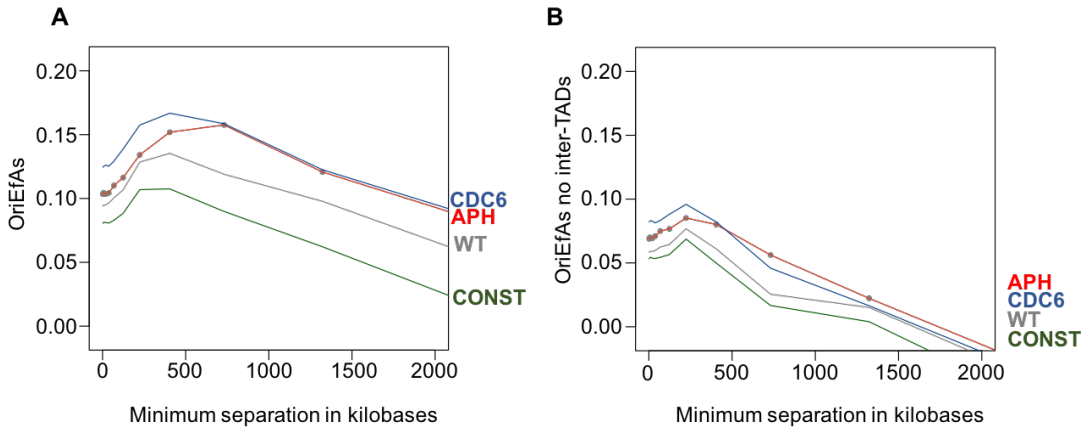


Figure 41. Origins preferentially interact at long-range distances. Contacts which span less than indicated genomic distances were gradually removed from the PCHiC network. Origins defined by **stringent** dataset were used for this analysis. **(A)** OriEfAs values in the PCHiC network at different ranges of linear separation between contacting fragments. **(B)** The same as in (A) but inter-TAD contacts were eliminated from the network.

To better understand the contribution of intra-TAD and inter-TAD contacts for clustering of origins in 3D, a similar analysis was done after elimination of inter-TAD interactions from the network. In this scenario, the peak of OriEfAs shifted towards 250kb (Fig. 41B), indicating that this distance preferentially separates origins within a factory involving exclusively intra-TAD contacts. Notably, removal of inter-TAD interactions leads to a general decrease in OriEfAs values emphasizing their importance in origin clustering in 3D. This result suggests a possible hierarchy of organization of replication factories where individual replication factories formed within TADs are brought together in 3D by inter-TAD interactions (see Discussion).

Chapter 2: Regulation of CDC6 stability by CDC7 kinase

Inhibition of CDC7 causes a reduction in CDC6 protein levels

In the course of our experiments with mouse embryonic fibroblasts (MEFs) derived from the TetO-CDC6 mice strain, described in Chapter 1 (Muñoz et al., 2017), we noticed that CDC6 protein levels decreased when cells were treated with PHA-767491, an inhibitor of CDC7 kinase (Cdc7i; Montagnoli et al., 2008). The drop in CDC6 levels was especially prominent following CDC6 overexpression (Fig. 42A, lanes 3 and 4), but could also be detected in control conditions (Fig. 42A, lanes 1 and 2). The reduction in CDC6 levels upon CDC7 inhibition was conserved in human U2OS cells treated with Cdc7i (Fig. 42B). These results suggest a previously unknown effect of CDC7 on the stability of CDC6.

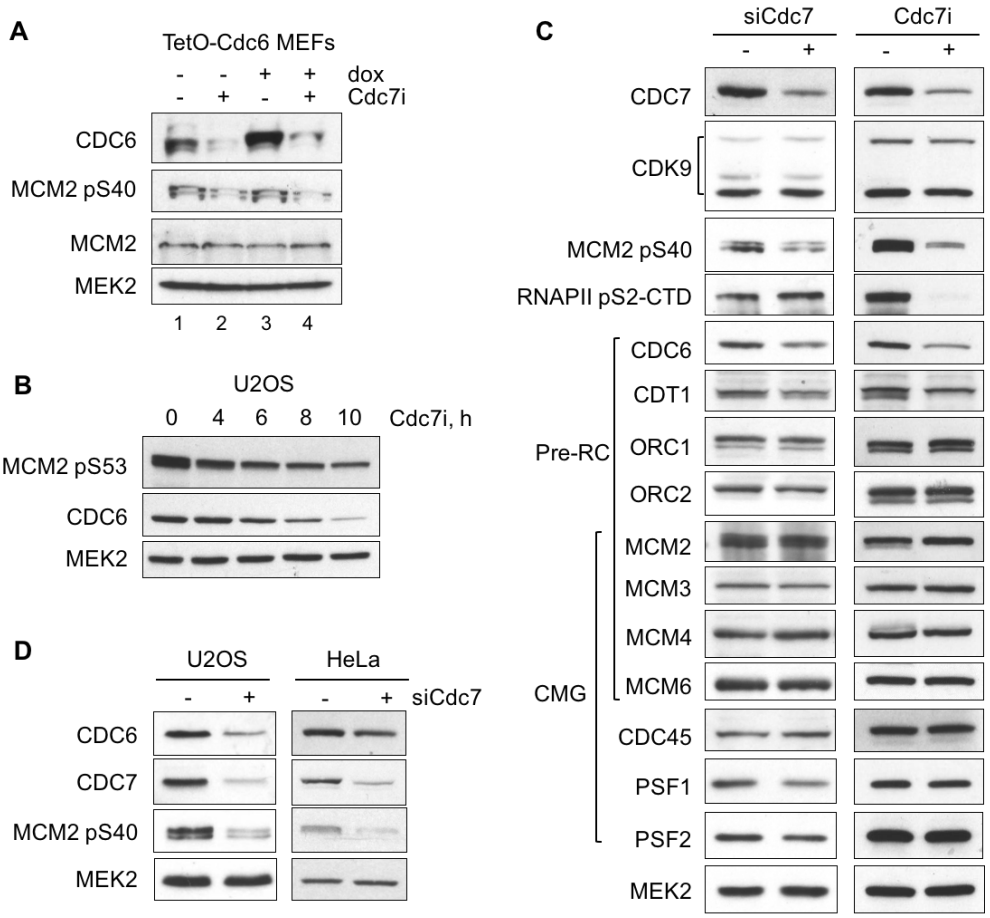


Figure 42. Downregulation or inhibition of CDC7 causes a decrease in CDC6 levels without affecting other components of the pre-RC. (A) Immunoblots from TetO-CDC6 MEFs whole cell extracts (WCE) treated with doxycycline (dox) for 27h and/or Cdc7i for 3h. **(B)** Immunoblots from U2OS WCE treated with Cdc7i for the indicated times. **(C)** Immunoblots from U2OS WCE treated with siCdc7 for 48h or Cdc7i for 6h. Pre-RC, pre-replication complex; CMG, CDC45-MCM-GINS complex. **(D)** Immunoblot from U2OS and HeLa WCE treated with siCdc7 for 48h. MEK2 levels are shown as loading control.

The efficiency of CDC7 inhibition was monitored with specific antibodies that detected the phosphorylation of MCM2 protein at Ser40 and Ser53, which are mediated by CDC7-DBF4 kinase (Fig. 42; Montagnoli et al., 2006). In addition to CDC7, PHA-767489 is also known to inhibit CDK9 kinase (Montagnoli et al., 2008). CDK9 phosphorylates Ser2 at the C-terminal domain (CTD) of RNA polymerase II to stimulate the elongation phase of transcription. Consistent with this, RNAPII pS2-CTD was decreased following Cdc7i treatment (Fig. 42C). To determine whether the reduction in CDC6 levels was actually mediated by DDK, *CDC7* was downregulated using a specific siRNA oligonucleotide (siCdc7). In both U2OS and HeLa cells, siCdc7 led to a reduction in CDC6 levels, supporting the notion that CDC7 contributes to the regulation of CDC6 protein (Fig. 42D).

CDC7 downregulation/inhibition did not affect the levels of other components of the pre-RC and CMG complexes such as ORC1, ORC2, MCM2, MCM3, MCM4, MCM6, CDC45 or PSF2, except for slight reductions in the levels of CDT1 and GINS subunit PSF1. Therefore, CDC7 inhibition does not elicit a global downregulation of DNA replication proteins. As previously reported, Cdc7i also reduced CDC7 protein levels (Fig. 42C; Montagnoli et al., 2008).

CDC6 levels are reduced in all phases of the cell cycle

In human cells, CDC6 levels fluctuate in the cell cycle, being lower during early G1 and higher in the remaining phases (Méndez and Stillman, 2000). For this reason, the drop in CDC6 levels observed after CDC7 inhibition could be due to an indirect effect on cell cycle distribution. However, analyses of DNA content by flow cytometry following siCdc7 or Cdc7i treatments showed only a modest percentual increase in G1 cells, concomitant to a slight reduction in S-phase cells (Fig. 43A,B). To exclude the possibility that the drop in CDC6 levels was caused by this slight accumulation in G1, cells were separated into G1, S and G2/M phases by FACS. The reduction of CDC6 protein caused by siCdc7 was observed in all phases (Fig. 43C). In response to Cdc7i, the decrease in CDC6 levels was apparent in G1 and G2 phases but less clear in S phase (Fig. 43D).

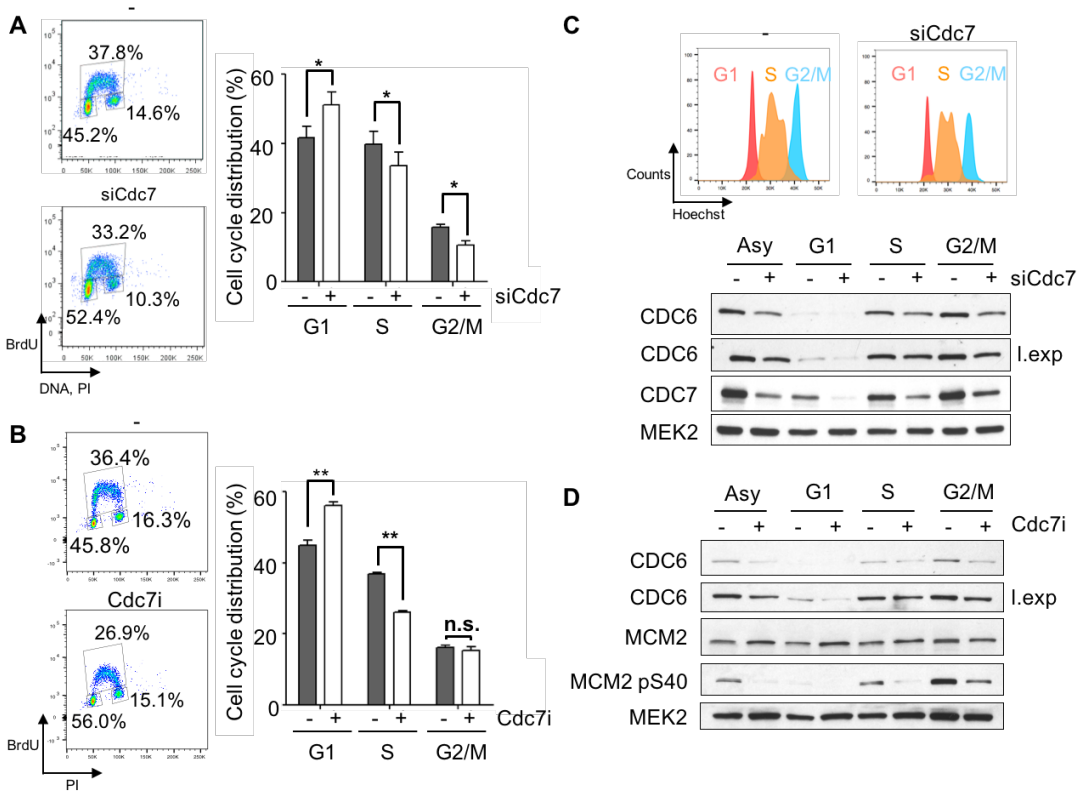


Figure 43. CDC6 levels decrease in all cell cycle phases upon siCdc7. **(A)** Distribution in cell cycle phases in U2OS cells -/+ siCdc7 (48h), as assessed by flow cytometry using propidium iodide (PI) and BrdU staining. Gates corresponding to G1, S and G2/M cells and their percentages relative to the total cell population are indicated. The histogram shows the percentage of cells in each phase (mean value and standard deviation (SD), n=3 assays), *p < 0.1 in Student's t test. **(B)** The same as in (A), but U2OS cells were treated with Cdc7i for 6h; *p < 0.1; **p < 0.01; n.s. = not significant in Student's t test. **(C)** U2OS cells were treated with siCdc7, stained with Hoechst 33342 and sorted by FACS into G1, S and G2/M phases. Flow cytometry graphs indicate the purity of each fraction after cell sorting. Immunoblots of the indicated proteins were performed in WCE from asynchronous (Asy) and sorted cells. s. exp., short exposure; l.exp, long exposure. **(D)** The same as in (C), but U2OS cells were treated with Cdc7i instead of siCdc7 before cell sorting.

Next, the subcellular localization of CDC6 was assessed by biochemical fractionation. Reduced CDC6 levels were detected in both cytosolic and chromatin fractions upon siCdc7 (Fig. 44A) and Cdc7i (Fig. 44B). The fact that the levels of chromatin-bound CDC6 are altered suggests a functional significance for the stabilization of CDC6 protein mediated by CDC7.

The decrease in CDC6 levels is mediated by changes in protein stability

To test the possibility that CDC6 levels could be regulated transcriptionally, *CDC6* mRNA was determined by RT-qPCR following *CDC7* downregulation. No major changes in mRNA levels were detected for *CDC6* or other members of the pre-RC such as *ORC1* and *MCM3* (Fig. 45A). *p21* mRNA, which displays a short half-life,

was also not reduced (it was actually slightly increased), suggesting that transcription is not globally altered (Fig. 45A). Conversely, Cdc7i led to a marked reduction in mRNA levels in all genes tested, consistent with a global transcriptional shut-down probably mediated by the CDK9 inhibition (Fig. 45B).

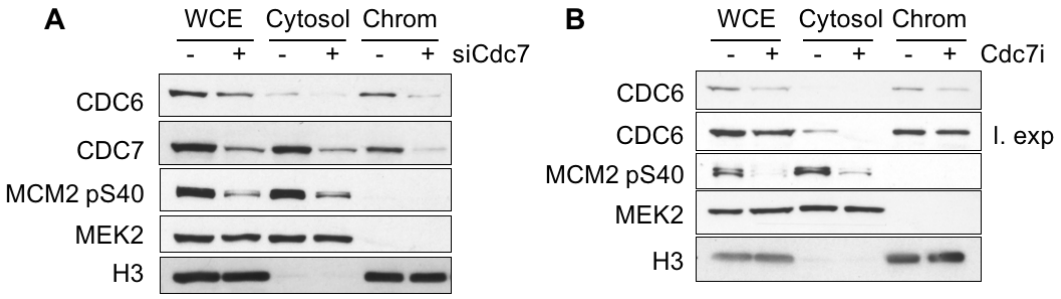


Figure 44. CDC7 downregulation or inhibition affects both cytosolic and chromatin-bound CDC6. U2OS cells were treated with siCdc7 or Cdc7i and subjected to biochemical fractionation. Immunoblots of WCE, cytosolic and chromatin (chrom) fractions for cells treated with siCdc7 **(A)** or Cdc7i **(B)** are shown. MEK2 (cytosolic) and H3 (chromatin-bound) immunoblots serve as fractionation controls.

CDC6 protein is stabilized by proteasome inhibitors such as MG-132 (Méndez and Stillman, 2000). This effect was reproduced in our study in control conditions (Fig. 45C, lanes 1 and 2; Fig. 45D, lanes 1 and 3), and also upon siCdc7 (Fig. 45C, lanes 3 and 4) and Cdc7i (Fig. 45D, lanes 2 and 4). These results support the notion that the reduction in CDC6 levels in response to CDC7 downregulation/inhibition is mediated at least in part by protein destruction via proteasome.

To confirm that the decrease in CDC6 levels was related to protein stability, the half-life of CDC6 was examined upon addition of cycloheximide (CHX), an inhibitor of protein synthesis. CDC6 normal half-life of approximately 2 h was shortened upon Cdc7i treatment (Fig. 45E). A similar effect was observed with CDK inhibitor roscovitine, previously shown to reduce CDC6 stability (Fig. 45F; Mailand and Diffley,2005).

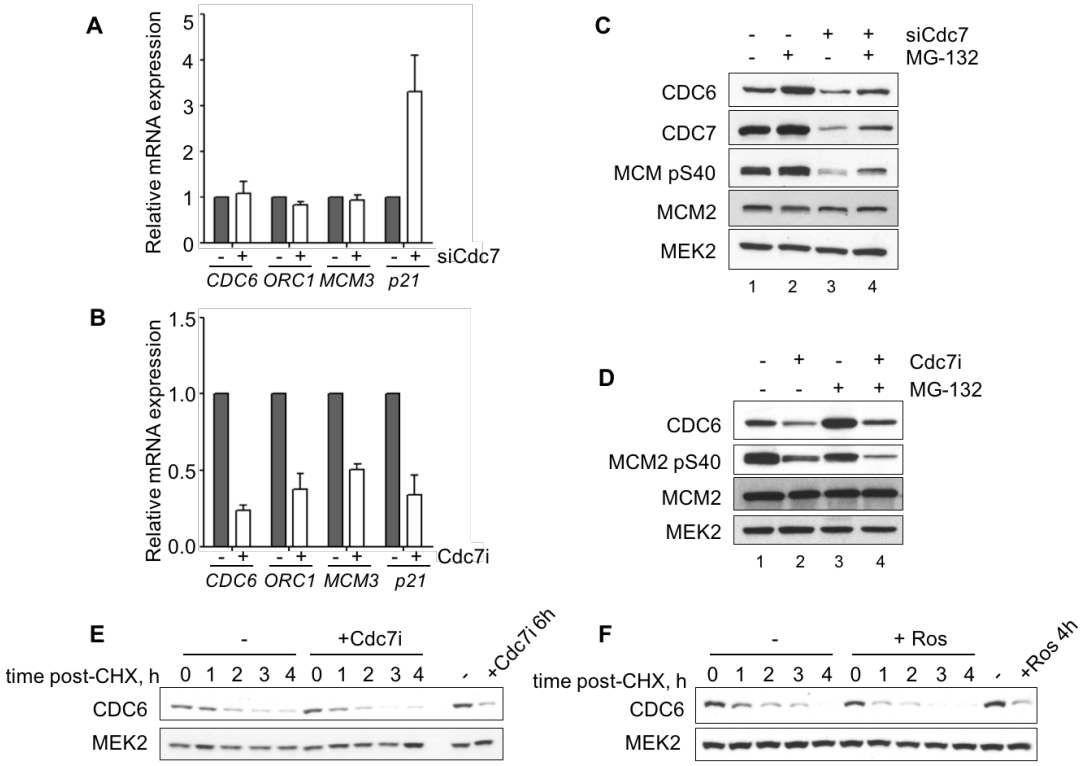


Figure 45. Effect of CDC7 downregulation or inhibition on *CDC6* mRNA and protein stability. RT-qPCR for *CDC6*, *ORC1*, *MCM3* and *p21* mRNA isolated from U2OS cells treated with siCdc7 **(A)** or Cdc7i **(B).** Fold-change (mean and SD) of treated vs control cells is represented; n=3. **(C)** Immunoblots of WCE from U2OS cells treated with siCdc7 and/or proteasome inhibitor MG-132. **(D)** Same as in (C), but cells were treated with Cdc7i instead of siCdc7. **(E, F)** U2OS cells were treated with Cdc7i for 6h (**E**) or roscovitine (Ros) for 4h **(F)** in the presence of cycloheximide (CHX) for the indicated times. Stability of CDC6 was monitored by immunoblotting.

CDC7 downregulation does not affect APC/C activity

In order to elucidate the mechanism responsible for the drop in CDC6 levels, we examined the levels of proteins known to regulate CDC6 stability in the cell cycle such as the APC/C ubiquitin ligase (Petersen et al, 2000). Upon siCdc7 and Cdc7i treatment, APC/C activating subunits CDH1 and CDC20 levels were not upregulated, suggesting that the decrease in CDC6 levels is not due to an enhanced activity of APC/C. Besides, we confirmed that other known targets of this ubiquitin ligase (CycA, CycB, AuroraB; Manchado et al., 2010) were not altered upon CDC7 downregulation or inhibition. A slight reduction in the levels of Geminin, another APC target, was observed (Fig. 46A). We conclude that the activity of APC/C was not largely affected in our experimental conditions. Interestingly, we noticed the decrease in CycE (not a target of APC/C) levels upon CDC7 inhibition (Fig. 46A), in agreement with a previous report (Montagnoli et al., 2008).

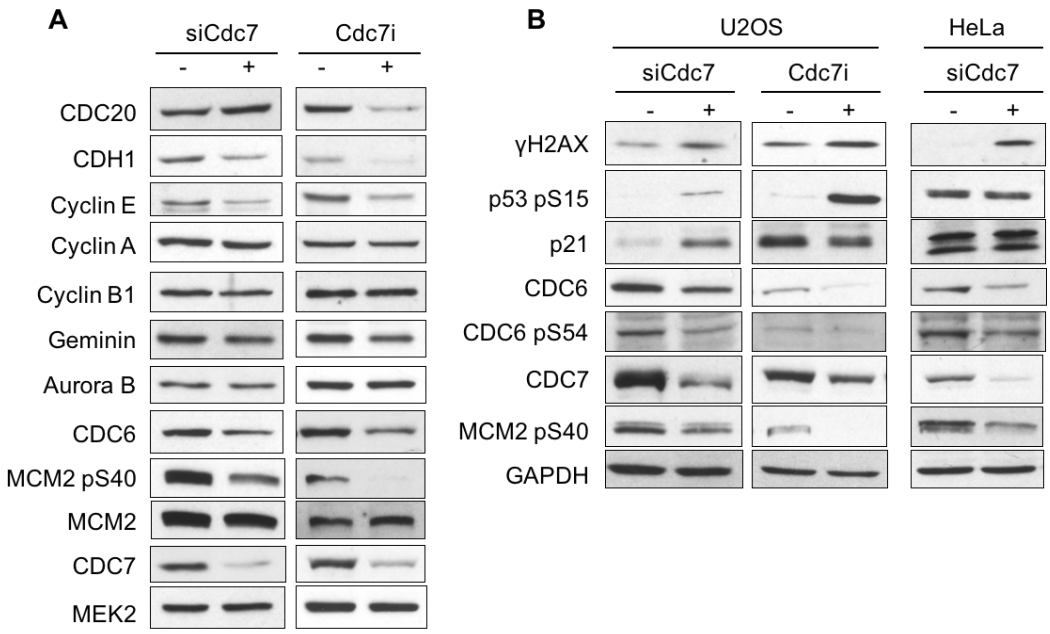


Figure 46. CDC7 downregulation or inhibition does not affect APC/C activity but induces a DNA damage response. (A) Immunoblot detection of the indicated proteins in WCEs of U2OS treated with siCdc7 or Cdc7i. **(B)** Immunoblot detection of the indicated proteins in WCEs from U2OS cells treated with siCdc7 or Cdc7i, or HeLa cells upon siCdc7.

CDC7 downregulation induces p53 and p21

In response to DNA damage, CDC6 is destabilized in a p53-dependent manner. This regulation involves p21, which inhibits a protective phosphorylation event at Ser54 in CDC6 by CDK2-CycE (Duursma and Agami, 2005). We therefore examined whether siCdc7 and Cdc7i induce a p53 response. In both cases, DNA damage marker γH2AX was detected, together with increased levels of p53 phosphorylation at Ser15. siCdc7 also led to increased *p21* mRNA and protein levels (Fig. 45A and 46B). This was not observed after Cdc7i, probably due to the downregulation of *p21* mRNA (Fig. 45B and Fig. 46B). CDC6 phosphorylation at Ser54 was reduced in parallel to the decrease in total CDC6 levels.

These experiments suggest that p53 activation may be responsible for the drop in CDC6 following CDC7 inhibition. However, when CDC7 is inhibited in HeLa cells that carry an inactive form of p53, p21 protein levels did not increase but the levels of CDC6 pS54 were still reduced (Fig 46B). We conclude that CDC6 downregulation upon siCdc7 and Cdc7i involves both p53-dependent and p53-independent mechanisms.

CDC7 phosphorylates CDC6 at multiple sites *in vitro.*

CDC7 kinase phosphorylates preferentially at serine or threonine amino acids that are followed by an acidic residue (S/T-D/E; Charych et al., 2008; Cho et al., 2006;

Montagnoli et al., 2006). Human CDC6 has eight consensus CDC7 phosphorylation sites. Six of them are conserved in mouse and four are maintained between human, mouse and frog CDC6 proteins (Fig. 47). We hypothesized that CDC7 may act directly on CDC6 by phosphorylating it and affecting its stability. In order to address this possibility, GST-CDC6 was purified (Fig. 48A; Méndez and Stillman, 2000). The size of GST-CDC6 is ≈88 kDa (62kDa for CDC6 and 26kDa for GST). Also, CDC7 kinase was purified together with a fragment of DBF4 that is essential for its activation (CDC7/DBF4$^{197-333}$, ≈66 kDa; Fig. 48A). Upon incubation of GST-CDC6 with CDC7/DBF4$^{197-333}$, the GST-tag is cleaved from CDC6 (Fig. 48B, compare lane 2 with lanes 1 and 4), most likely by the action of PreScission protease used for CDC7/DBF4$^{197-333}$ purification. This was unexpected, as the linker sequence between GST and CDC6 contains a target sequence for TEV protease rather than PreScission enzyme. In any case, the TEV target includes the core sequence Phe-Gln-Gly that is also recognized by PreScission. Indeed, we confirmed that purified PreScission excised GST from GST-CDC6 (Fig. 48B, lane 5). The 66 kDa protein species that appears after incubation with CDC7/DBF4$^{197-333}$ (Fig. 48B, lanes 1 and 4) or PreScission protease (Fig. 48B, lane 5) is indeed recognized by CDC6 antibody (Fig. 48B, bottom).

In vitro kinase assays were performed with the purified proteins. CDK2-Cyclin A was used as positive control as it phosphorylates CDC6 at S54, S74 and S106 (Fig. 48C lane 5; Jiang et al., 1999; Petersen et al., 1999). Reactions in the absence of kinase (Fig. 48C, lane 2) or Mg^{2+} metal activator (Fig. 48C lane 2) served as negative controls. In the absence of GST-CDC6 substrate, autophosphorylation of CDC7/DBF4$^{197-333}$ was detected (Fig. 48D, lane 3), in line with previous studies (Cho et al., 2006; Weinreich and Stillman, 1999). The complete *in vitro* kinase reaction revealed the presence of CDC6 phosphorylation (lower band) as well as CDC7 autophosphorylation (upper band; Fig. 48C).

In order to identify specific phosphorylation sites, an *in vitro* kinase assay was followed by mass spectrometry (MS) analysis. Using this approach, five out of the eight CDC7 consensus sites were identified: S30, S134, S421, S439, S504 (Fig. 48D). Additionally, phosphorylation was detected at two non-consensus sites: S6 and S41. CDC7 and DBF4 autophosphorylation sites were also identified (not shown). A parallel *in vitro* kinase assay performed with CDK2-Cyclin A led to the detection of two CDK phosphorylation sites in CDC6: S74 and S106 (Fig.48E). From these experiments we conclude that CDC7 phosphorylates CDC6 *in vitro*.

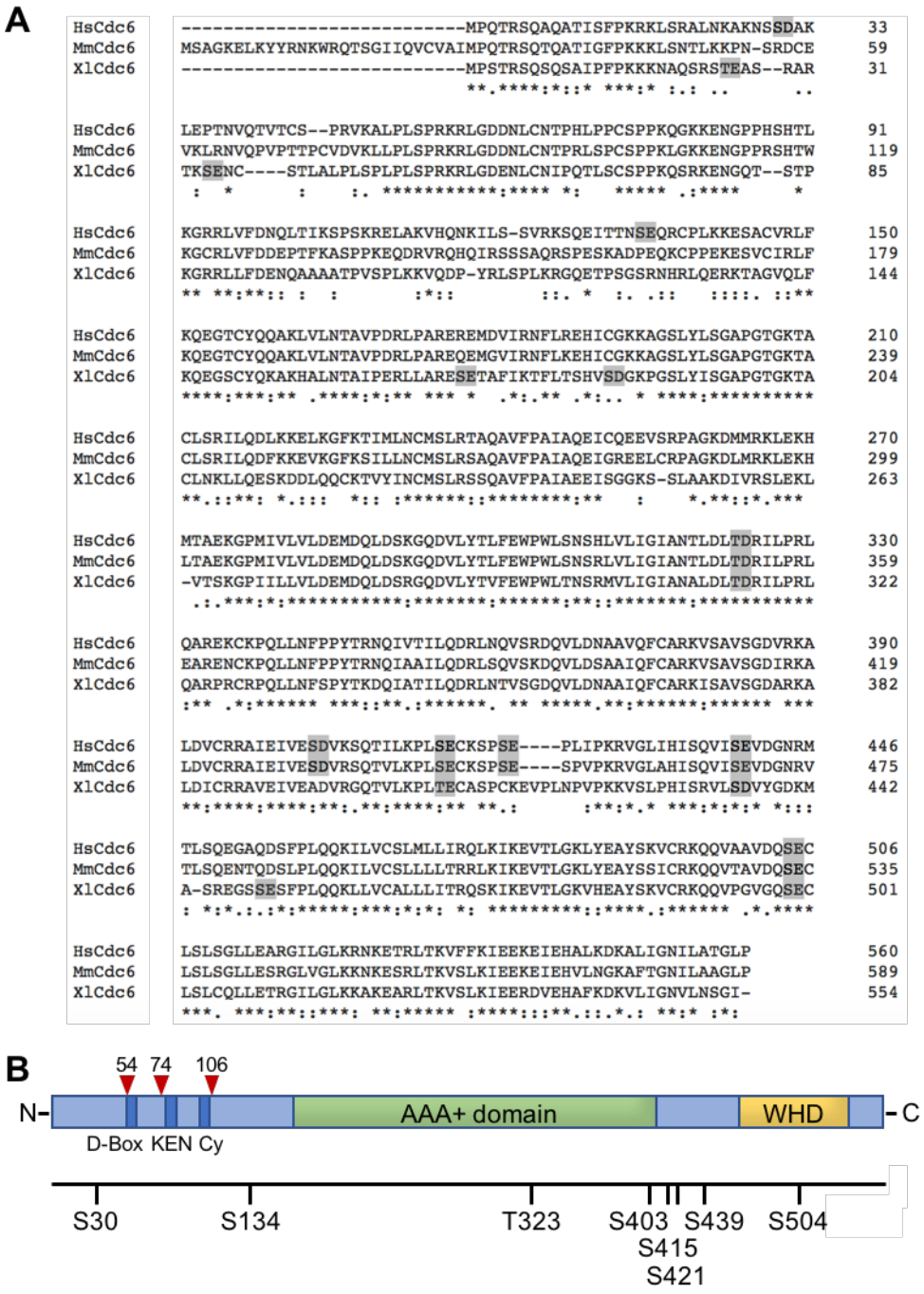
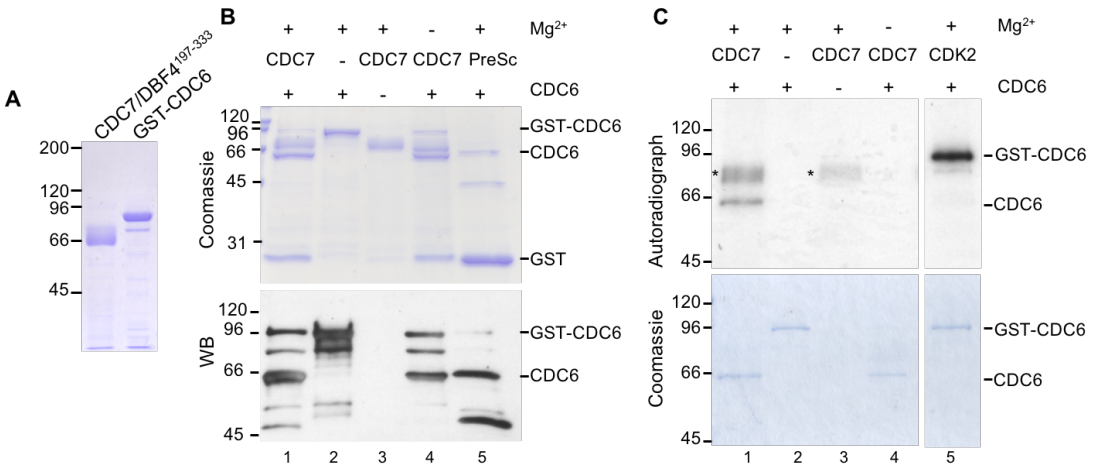


Figure 47. Consensus CDC7 phosphorylation sites in CDC6 protein. (A) Protein sequence alignment of CDC6 protein species from human (Hs), mouse (Mm) and frog (Xl). CDC7 consensus phosphorylation sites are indicated in grey. **(B)** Schematic of the human CDC6. Conserved motifs are indicated. Red arrows mark the positions of three serine residues (S54, S74, S106 that can be phosphorylated by CDK. D-box and KEN motif are conserved destruction boxes recognized by APC/C^{CDH1}. Cy, cyclin binding box. AAA+ domain contains the active site for ATPase activity. WHD, winged-helix domain. Black line at the bottom marks the positions of eight consensus CDC7 phosphorylation sites. Adapted from Borlado and Méndez, (2008).



D

peptide (phospho probabilities)	position	amino acid
S(1)QAQATISFPKR	6	S
NS(0.002)**S**(0.998)**D**AKLEPTNVQTVTCSPR	30	S
LEPTNVQT(0.938)VT(0.061)CS(0.001)PR	41	T
SQEITTN**S**(1)**E**QR	134	S
SQTILKPLS(0.004)ECKS(0.08)P**S**(0.916)**E**PLIPK	421	S
VGLIHISQVI**S**(1)**E**VDGNR	439	S
QQVAAVDQ**S**(0.997)**E**CLS(0.003)LSGLLEAR	504	S

E

peptide (phospho probabilities)	position	amino acid
S(1)QAQATISFPKR	6	S
NSSDAKLEPTNVQTVT(0.002)CS(0.998)PRVK	45	S
LGDDNLCNT(1)PHLPPCS(1)PPKQGK	67	T
LGDDNLCNT(1)PHLPPCS(1)PPKQGK	74	S
LVFDNQLTIKS(0.996)PS(0.004)K	106	S
ILS(0.041)S(0.959)VRK	123	S
SQEITTNS(1)EQR	134	S
SQTILKPLSECKS(0.998)PS(0.002)EPLIPK	419	S
VGLIHISQVIS(1)EVDGNR	439	S

Figure 48. CDC7 phosphorylates CDC6 protein *in vitro*. (A) Coomassie Brilliant Blue staining of purified CDC7/DBF4$^{197-333}$ and GST-CDC6 proteins. **(B)** Incubation of GST-CDC6 with CDC7/DBF4$^{197-333}$ or PreScission (PreSc) protease separates the GST-tag from CDC6. Top, Coomassie staining. Bottom, immunoblot with CDC6 antibody. **(C)** *In vitro* kinase assay followed by autoradiography (upper panel). The lower panel shows a Coomassie staining of the proteins present in the *in vitro* kinase assays. Asterisks mark CDC7 autophosphorylation. **(D)** Table shows the peptides including phosphosites identified by mass spectrometry after *in vitro* kinase reaction similar to the one shown in (C). Consensus phosphorylation sites are shown in grey. **(E)** The same as (D), but the *in vitro* kinase assay was performed using CDK2-Cyclin A and GST-CDC6.

Effect of CDC6 phosphorylation on protein stability

To study the effects of CDC7-mediated CDC6 phosphorylation, human CDC6 mutant versions were generated in which all eight CDC7 consensus sites were modified to substitute the Ser/Thr residues for Ala (S8A) or Asp (S8D), mimicking the unphosphorylated and phosphorylated states of CDC6, respectively. We decided to mutate all consensus sites, because we suspected that the limited sensitivity of the IVK/mass-spec assay could hinder the potential presence of phosphorylation at the remaining positions. In our system, an N-terminal V5 tag was added to CDC6 to differentiate between the endogenous and over-expressed (o/e; V5-tagged) versions of CDC6.

We first tested the stability of mutant CDC6 proteins upon siCdc7 treatment. While endogenous CDC6 and V5-CDC6 dropped upon downregulation of CDC7 kinase,

the levels of phospho-dead (V5-CDC6-S8A) and phospho-mimic (V5-CDC6-S8D) proteins remained stable. We reasoned that because both mutants carry modifications that prevent their phosphorylation by CDC7, they might be resistant to alterations in the levels of this kinase (Fig. 49A). We then evaluated CDC6 stability in the presence of CXH. Whereas WT and S8D versions of CDC6 displayed a similar half-life, the unphosphorylatable S8A CDC6 was clearly less stable (Fig. 49B). These experiments support the notion that CDC7 phosphorylation stabilizes CDC6.

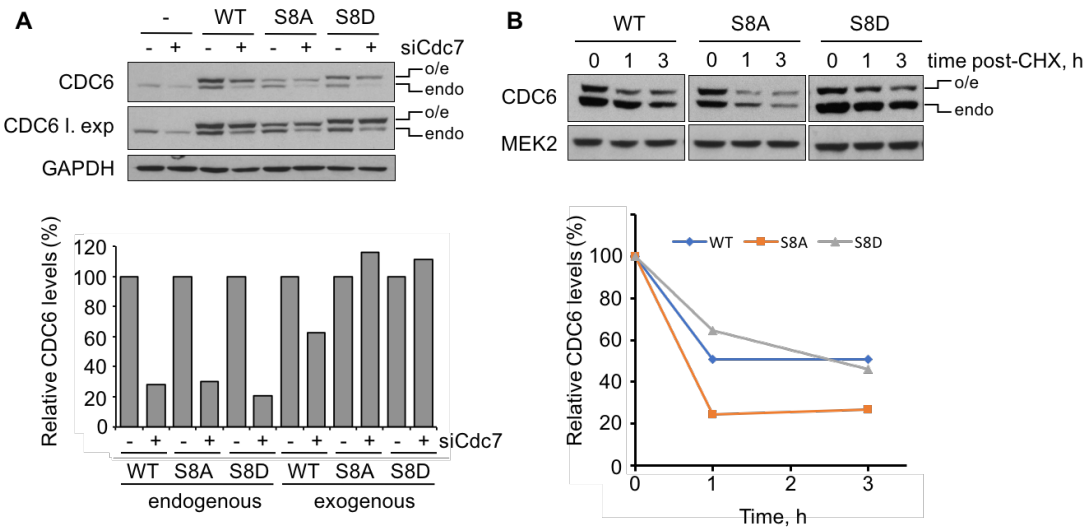


Figure 49. CDC6-S8A and S8D mutants are resistant to CDC7 downregulation. (A) U2OS cells were transfected with siCdc7 and, 24h later, with plasmids carrying different versions of CDC6 (WT, S8A, S8D). The histogram shows the quantification of CDC6 levels as detected in immunoblots, normalized to the siCdc7-untreated condition. **(B) CDC6-S8A mutant displays reduced stability.** WT, S8A or S8D CDC6 mutants were expressed in U2OS cells and their degradation kinetics were estimated by immunoblotting in the presence of CHX. Representative immunoblots are shown. The graph represents quantification of fluctuation of each version of CDC6. Protein levels were normalized to the 0h time point. Quantifications were performed with ImageJ. endo, endogenous CDC6, o/e, over-expressed

Downregulation of CDC6 does not affect MCM helicase loading

To understand the possible function of CDC7-dependent CDC6 phosphorylation, we examined the effect of over-expressing the mutant CDC6 versions. No differences in cell cycle distribution were observed relative to control cells (Fig. 50A, B). As CDC6 is essential for MCM helicase loading onto DNA, we hypothesized that expression of the less stable, phospho-dead version of CDC6 could affect helicase loading. However, neither expression of V5-CDC6-S8A nor V5-CDC6-S8D in asynchronous cell cultures affected the amount of MCM2, MCM3 and MCM4 proteins on chromatin (Fig. 50C).

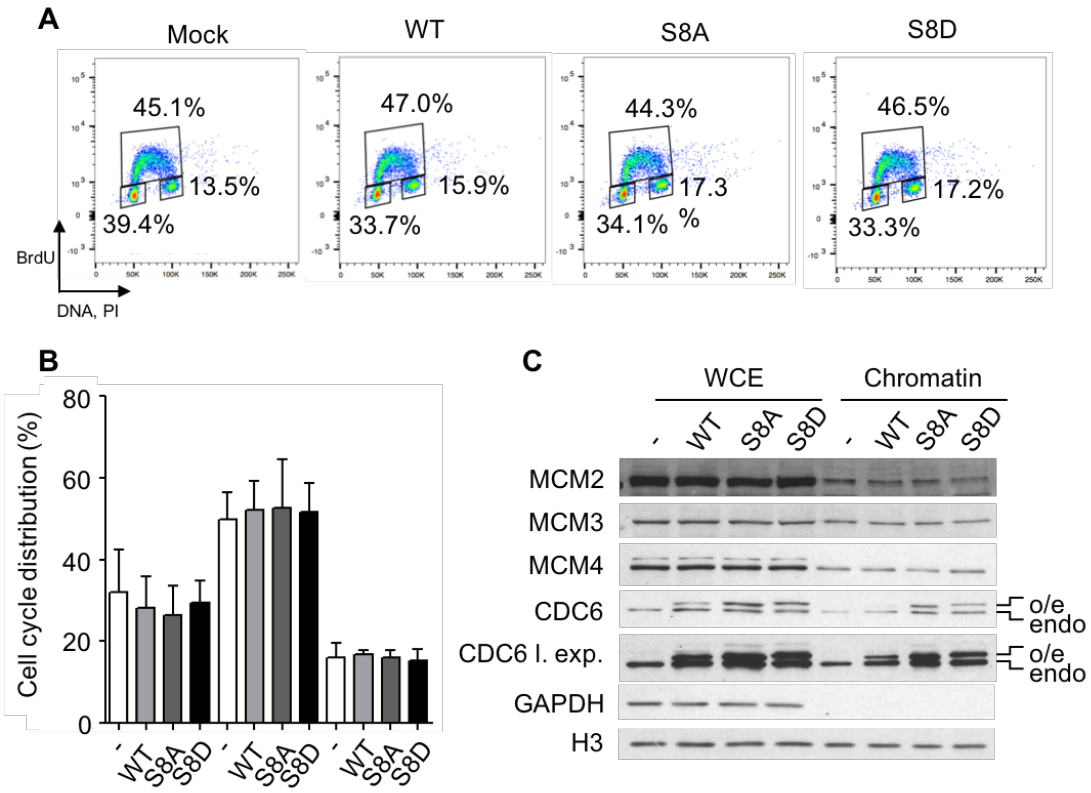


Figure 50. Expression of CDC6-S8A and CDC6-S8D mutants does neither affect cell cycle distribution nor MCM chromatin loading. (A) U2OS were transfected with the indicated versions of CDC6 for 24 h. Cell cycle distribution was assessed by flow cytometry using propidium iodide (PI) and BrdU staining. Gates indicate the percentage of cells in each phase of the cell cycle. **(B)** Histogram showing percentage of cells in G1, S and G2/M phases 24 h after transfection with the different versions of CDC6 (mean value and SD, n=3 assays). **(C)** Cells treated as in (A) were subjected to biochemical fractionation. Immunoblots show the levels of MCM2, MCM3, MCM4 and CDC6 proteins in WCEs and the chromatin fraction. GAPDH (cytosol) and H3 (chromatin-bound) are shown as fractionation controls. Quantifications were performed with ImageJ.

We also monitored whether the reduction in CDC6 levels would affect the licensing reaction specifically at the M-G1 transition. Cells transfected with siCdc7 were synchronized in metaphase with nocodazole and released for different times, followed by chromatin fractionation. Changes in CycB1, Geminin and CDC6 served as controls for the transition from M to G1 (Fig. 51A, B). As expected, the association of CDC6 with chromatin was diminished upon siCdc7 treatment, but MCM helicase loading was not affected, as assessed by the amounts of MCM2, MCM3 and MCM4 subunits (Fig. 51B). It should be noted that endogenous CDC6 was still present in these experiments, which might influence the evaluation of the effect of the CDC6 mutant versions.

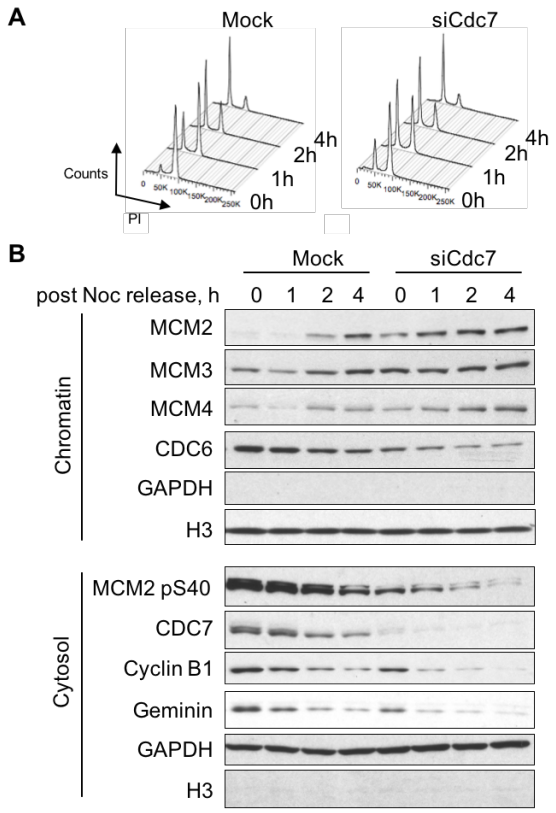


Fig. 51. CDC6 downregulation mediated by siCdc7 does not alter MCM loading in M-G1. (A) U2OS cells -/+ siCdc7 were synchronized with nocodazole for 20h. After block release, cells were collected at different times and subjected to biochemical fractionation. DNA content in each cell population assessed by flow cytometry detection of PI. **(B)** Immunoblots against the indicated proteins in the chromatin or cytosol fractions. GAPDH and H3 are shown as fractionation controls.

In summary, in this chapter we have described a novel regulatory pathway for CDC6 protein stability, mediated by CDC7 kinase. While the biological function of this regulation is still unclear, possible models are presented in the Discussion section.

Discussion

Discussion

Chapter 1: Response of replication origins to stress, and their spatial organization in the nucleus

Genome-wide localization and features of constitutive vs stress-responsive DNA replication origins

We have generated genome-wide maps of replication origins in mouse ESCs in three experimental conditions (WT, APH, CDC6). Before our study, a single genome-wide map and two partial maps of mESCs origins had been published (Cayrou et al., 2011; Sequeira-Mendes et al., 2009). The goal of our analysis was to provide novel insights into the regulation of DNA replication in normal and stress conditions.

The SNS-Seq methodology used for genome-wide mapping of replication origins in mammalian cells has raised controversy, mainly because it has been indicated that insufficient digestion with λ-exonuclease could enrich in G4-quadruplex sequences (Hyrien, 2015; Urban et al., 2015). We have tried to minimize this possible bias in our experiments by performing three rounds of extensive λ-exonuclease digestion in the presence of an exogenous plasmid containing G4 motifs as control. Still we found significant overlap (around 60%) of origins with G4-forming sequences in line with previous reports (Besnard et al., 2012; Cayrou et al., 2012b). However, it should be taken into account, that G4 motifs are very abundant in the genome (nearly 490 000). These structures may therefore contribute to specification of a large subset of origins. Genetic experiments involving the manipulation of G4 structures will be required to determine the precise contribution of these elements to origin definition and activity.

Then, two separate peak-calling algorithms were used to identify replication start sites and restrictive criteria were used to define the final origin datasets. Later on, the results obtained were confirmed using datasets that were defined with more relaxed criteria. Because of these measures we are confident about the quality and reproducibility of the results presented.

Characterization of origin genetic and epigenetic features

We examined the chromatin landscape of replication origins in cells in normal growth conditions (WT), upon replication stress (APH) or overexpression of a key initiator protein (CDC6). In all cases, extensive correlations were established between origin positions and genomic and epigenomic features available for ESCs. Many of the >70

epigenomic marks (Juan et al., 2016) were analysed for the first time in the context of origins, generating one of the most comprehensive characterization of mammalian origin features so far.

In general, our analyses agree with known features of mammalian origins such as their preferred localization at CGI, promoters, intragenic regions, early replicating regions and open chromatin marks (Besnard et al., 2012; Cadoret et al., 2008; Cayrou et al., 2011, 2015; Dellino et al., 2013; Langley et al., 2016; Miotto et al., 2016; Petryk et al., 2016; Picard et al., 2014; Sequeira-Mendes et al., 2009). Because of the exhaustive list of epigenetic marks analysed, we have expanded these observations to include the following considerations:

Connection with promoter elements. It is clear that origins are preferentially located at or around promoter elements, as shown by the strong correlations with the H3K4me3 mark, RNAPII and histone modifying enzymes such as H3K4 methyltransferase MLL and H3K36 demethylases KDM2A/B (Dimitrova et al., 2015; Zhang et al., 2015). Generally, promoters lack the H3K36me2/3 mark and that KDM2 demethylases co-localize with H3K4me3 at CGI promoters to ensure the removal of H3K36 methylation from TSSs (Zhang et al., 2015). In line with these observations, human replication origins preferentially localize at the 5'-end but not the 3' end of expressed genes (Valenzuela et al., 2011).

Interestingly, distinct forms of RNAPII displayed different levels of enrichment: amongst the most frequently present were phosphorylated forms RNAPII-S5 and RNAPII-S7, which have been associated to RNAPII pausing in the proximity of promoters of active genes (Brookes et al., 2012; Core and Lis, 2008). Conversely, the elongating form of RNAPII - characterized by RNAPII-S2 phosphorylated form and presence of H3K36m2/3 marks (Brookes and Pombo, 2009; Dimitrova et al., 2015; Harlen and Churchman, 2017; Kouzarides; Zhang et al., 2015) - are amongst the least enriched features at origins. We have detected a strong correlation of origins with active and poised promoters, but not with transcriptional elongation chromatin states. A lower frequency of origin localization in regions where transcription elongation takes place could help to avoid conflicts between the transcription and replication machineries (García-Muse and Aguilera, 2016). In the future, a more detailed analysis of origin localization in respect to the positioning of different variants of RNAPII could allow for a better understanding of how initiation of DNA replication is coordinated with transcription dynamics.

Polycomb repressive complexes and bivalent promoters. The main role of PcG complexes is to maintain the pluripotent state by silencing developmental regulator genes (Simon and Kingston, 2013). We have found strong correlations of origins with several members of both PcG complexes: PRC1 and PRC2. PRC2 can be recruited to CGI (Ku et al., 2008; Li et al., 2017), where many origins are located. In mESC cells, PRC1 is a master regulator of genome architecture by organizing a network of promoter-promoter contacts (Schoenfelder et al., 2015a). We hypothesize that Polycomb proteins influence the regulation of replication and transcription by controlling the spatial organization of these processes in the nucleus.
Bivalent promoters contain both activator (H3K4me3) and repressive (H3K27me3) marks and are considered to poise expression of developmental genes (Bernstein et al., 2006). Recently, origins in mESCs have been shown to display a preference towards bivalent domains (Cayrou et al., 2015). Indeed, our independent analyses indicate an enrichment at origins of H3K4me3, H3K27me3, EZH2 (methyltransferase of H3K27m3) and the “poised promoters” chromatin state.

Connection with cohesin and mediator complexes. We also found a strong enrichment of the mediator complex and cohesin loading factor NIPBL at origins, as well as moderate association with cohesin subunits SMC1 and SMC3. Presumably, these factors link gene expression and chromatin architecture, as they promote cell type-specific gene activation through enhancer-promoter DNA looping (Kagey et al., 2010). Moreover, mediator interacts with RNAPII and directly regulates transcriptional machinery (Kwak and Lis, 2013). These data further underscore the interconnections between 3D chromatin architecture, replication and transcription.

Connection with DNA methylation. A striking enrichment of TET-associated complexes was seen at origins. In ESCs, TET1 binds within genes and CpG-rich promoters (Voigt et al., 2013; Williams et al., 2011) and is implicated in maintaining their hypomethylated DNA state (Wu et al., 2011). O-linked N-acetylglucosamine (O-GlcNAc) transferase (OGT) is recruited to chromatin through its interaction with TET1 in the proximity of CpG-rich TSS (Vella et al., 2013). OGT also interacts with HCFC1, a component of H3K4me3 methyltransferase complexes SET1A/B and MLL1/2 (Voigt et al., 2013), and this protein network establishes the H3K4me3 mark at CpG rich promoters, positively regulating gene expression (Balasubramani and Rao, 2013; Deplus et al., 2013; Voigt et al., 2013). TET1 can also interact with SIN3 (Vella et al., 2013; Williams et al., 2011), a member of Co-repressor complex, and with

PRC2 (Wu et al., 2011) indicating a potential role in transcriptional repression. Replication origin localization is therefore strongly linked to the modulation of chromatin conformation and gene expression mediated by TET1 associated complexes.

Modulation of origin activity upon stress: new origins and global increase in efficiency

We have uncovered unique information about the localization and characteristics of constitutive and stress-responsive origins in ESCs. When datasets were defined using stringent criteria, the total number of origins upon APH or CDC6 overexpression was increased in comparison to control conditions (Table 6), suggesting the activation of many dormant origins. A marked increase in the number of initiation sites upon exposure to APH has also been reported recently in HCT116 human cancer cells (Utani et al., 2017). However, this effect was not observed when datasets were defined using 'relaxed' criteria. In this case, the relatively low overlap between the two WT replicas resulted in the identification of a very large number of initiation sites in WT cells, which was higher than in APH and CDC6 conditions.

In any case, it is important to remark that very often, an enrichment of SNS-seq reads can be detected in positions where a peak is not recognized by the algorithm in a particular replicate or experimental condition. We interpret these cases as origins that are actually used, although their low efficiency does not reach the threshold of peak detection. These observations strongly imply that **the main replicative response to stress may not consist in the selection of new origin sites but in the modulation of origin activity of pre-existing ones**.

Indeed, APH treatment and CDC6 overexpression increased the global efficiency of origin usage in comparison to control growth conditions (as shown in Fig. 15). A stronger increase in origin efficiency was caused by APH than CDC6, in line with the results obtained with DNA fibres (as shown in Fig. 9 and 10). When origin efficiency was compared only in the subgroup of origins that are common to the WT, APH and CDC6 conditions, we confirmed that the efficiency of these common initiation sites was higher in APH and CDC6 than in the WT group. This supports the notion that stress increments the frequency of origin firing from already existing positions.

We would like to propose that both readouts of SNS experiments to detect extra origin activation (origin firing from new positions and increased efficiency from already existing ones) are caused by the increased efficiency of origin firing upon

stimuli. The vast majority of “dormant” origins detected from apparently new positions would correspond to origins that are active but not efficient enough to reach the threshold of peak detection in control growth conditions. Our observations are in line with the stochastic model of origin firing (Rhind and Gilbert, 2013) that postulates that there are no qualitative differences between different types of origins (early, late, constitutive, flexible, dormant). The only difference between them is their probability of firing/ efficiency with which they become activated.

The relevant question now becomes what determines origin efficiency. Indeed, constitutive origins were much more efficient than responsive ones, even after they had been activated by stress (as shown in Figure 15). This finding parallels the observations that in *Drosophila* and human cells, origins shared by different cell lineages from the same organism (referred to as “developmentally constitutive origins”) display higher efficiency than cell type specific origins (Besnard et al., 2012; Comoglio et al., 2015; Picard et al., 2014).

Different nature of constitutive and stress-responsive origins

Constitutive origins displayed stronger correlations with CGI, promoters, exons and early replicating regions than responsive ones. The latter were more evenly distributed throughout the genome. We conclude that dormant origins do not have strong preference towards any of the genomic features analysed nor to the replication timing regions. Analogously, human “developmentally constitutive origins” tend to replicate early in S-phase while cell-type specific origins do not show such a strong bias in replication timing (Picard et al., 2014).

Constitutive origins are much more frequently associated than responsive origins with the “active promoter” and “poised promoter” chromatin states, as defined by multiple epigenetic marks. Therefore, constitutive start sites show a preference towards open and transcriptionally active chromatin. On the contrary, responsive origins were preferentially associated to heterochromatin regions, which are known to be origin-poor (Cayrou et al., 2011; MacAlpine et al., 2010). A very long distance between two adjacent origins makes it harder for neighbouring forks to rescue an event of fork stalling, and indeed paucity of replication origins contributes to genomic instability at late-replicating common fragile sites (Técher et al., 2017). We postulate that within these zones, a significant increase in origin density in response to stress is necessary to ensure cell survival.

Remarkably, none of the analysed chromatin characteristics was exclusive to constitutive or responsive origins. Rather, quantitative differences were observed in the levels of their overlap/enrichment. Again, this is in line with the stochastic model for origin activation. Open chromatin favours origin positioning and efficiency of firing (Cayrou et al., 2011; MacAlpine et al., 2010). In our view, constitutive origins that strongly correlate with open, transcriptionally active chromatin, display a higher probability of being activated that responsive ones. It is likely that euchromatin provides easier access to limiting replication factors than heterochromatin. Additional factors that could influence the probability of origin activation are the number of licensed origins within a particular genomic region and the number of MCM complexes loaded onto each origin (Rhind and Gilbert, 2013). Easily accessible chromatin would enable the licensing of more origins in a particular genomic region and/or a higher loading of MCM complexes onto initiation sites. The probability of helicase activation would be accordingly increased in euchromatin zones.

Spatial organization of replication origins in the nucleus

Efficient origins are connected in 3D: implications for replication factories

We have integrated for the first time a linear, high-resolution map of replication origins into several 3D chromatin structure networks, providing unprecedented information about the spatial organization of origins within the nucleus.

Using a parameter known as **chromatin assortativity** (Pancaldi et al., 2016), we have found that genomic regions containing origins tend to interact across the genome. This was true for the three origin datasets (WT, CDC6, APH) and the constitutive (CONST) origin subset in several published 3D chromatin maps defined in ESCs using Hi-C, promoter-capture Hi-C (PCHiC), RNAPII ChIA-PET and SMC1 ChIA-PET. Origin-origin interactions were predicted by the replication factory model. According to it, groups of origins are brought together in 3D by chromatin loop formation. In S phase, contacting origins localized at the bases of the loops become activated simultaneously (Fig. 52 and Berezney et al., 2000; Courbet et al., 2008; Jackson and Pombo, 1998).

In general, subsets of responsive origins did not show this tendency. This could be due to the fact that responsive origins become active in the context of existing replication factories, rather than forming new factories consisting only of backup origins. It is conceivable that dormant origins are localized along the DNA loops, at a

distance from the loop base where constitutive origins would be concentrated. Stress conditions might induce changes in loop architecture, bringing dormant origins to the centre of the factory and favouring their activation (Fig. 52; Fragkos et al., 2015). In this regard, it would be interesting to perform a Hi-C experiment to investigate the changes in chromatin structure caused by replication stress, e.g. by aphidicolin treatment.

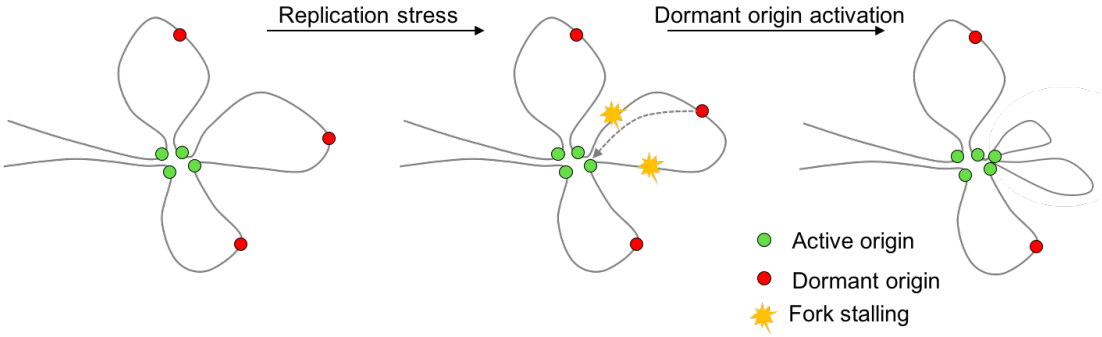


Figure 52. Model of dormant origin activation within a replication factory. Accumulation of stalled forks could trigger loop rearrangements and bring dormant origins to the centre of the factory. See text for details.

Using tools derived from network theory, we also determined that in the promoter-capture network (PCHiC), chromatin fragments containing replication origins display higher connectivity with other genomic regions than the average PCHiC fragment. This was true for all groups and subgroups of replication start sites analysed. Remarkably, constitutive origins presented the highest connectivity from all origin groups analysed (Fig. 30). As constitutive origins are the most efficient, a relationship likely exists between the probability of origin firing and its level of connectivity with other chromatin fragments. Here, we have shown that more connected origins not only tend to fire with higher efficiency, but also have a preference to replicate earlier in S phase. This is in agreement with a previous correlation established between origin efficiency and early replication (Picard et al., 2014).

It would also be expected that origins belonging to interacting chromatin fragments had comparable access to limiting factors for initiation and therefore become activated with similar efficiency. In this regard, we have found that origin *efficiency* is in itself an assortative parameter in the WT, APH, CDC6 datasets and the CONST subset. This result implies that origins contacting in 3D tend to fire with similar efficiencies, further supporting the replication factory model.

Cohesin has been proposed to stabilize DNA loops at replication factories (Guillou et al., 2010). ChIA-PET assays with cohesin component SMC1 capture interactions

mediated by cohesin (Dowen et al., 2014). We find that origins are assortative within the SMC1 ChIA-PET chromatin network, i.e. they are frequently located at chromatin interactions mediated by cohesin. This is another strong argument supporting the model of replication factories. To further validate this result, it would be interesting to calculate origin assortativity in recently reported Hi-C networks determined for cohesin-depleted mESCs (Wutz et al., 2017).

Origins are assortative in the promoter-promoter (P-P) subnetwork of PCHiC as well as the RNAPII ChIA-PET network, indicating a link between the 3D organization of initiation of replication and initiation of transcription. In order to separate the possible influence of 3D chromatin structure and transcription activity on replication origins we could compare the efficiency of activation of those origins localized at transcriptionally active promoters with those localized at poised promoters (transcriptionally inactive, but prepared to be rapidly activated upon developmental cues). If they were similar, this would indicate that origin activation is not dependent on active transcription but rather on 3D chromatin structure, which organizes replication initiation sites spatially.

Replication factories and topologically associated domains (TADs)

Because replication timing domains coincide almost exactly with TADs (Dileep et al., 2015; Pope et al., 2014), one would expect that most origin-origin contacts established within replication factories should belong to the same TAD. However, our results indicate that contacts between chromatin fragments containing origins span both intra- and inter-TADs connections. Interestingly, origins that interact with each other display similar replication timing regardless of the type of contact (inter- or intra-TAD).

Recent reports regarding the spatial organization of chromosomes using Hi-C in single cells have provided unique information about changes in chromatin organization across the cell cycle (Beagrie and Pombo, 2017; Nagano et al., 2017; Stevens et al., 2017). According to these studies, individual cells display a high level of heterogeneity in chromatin folding, and chromosome organization undergoes dynamic changes during interphase (Nagano et al., 2017). Intra-TAD interactions are the strongest during G1 phase. At the start of S phase, the insulation at TAD borders declines, reaching its lowest level in mid-S phase and being low during G2. The partial dissolution of TAD borders in S-phase could explain the substantial number of inter-TAD contacts between fragments containing origins.

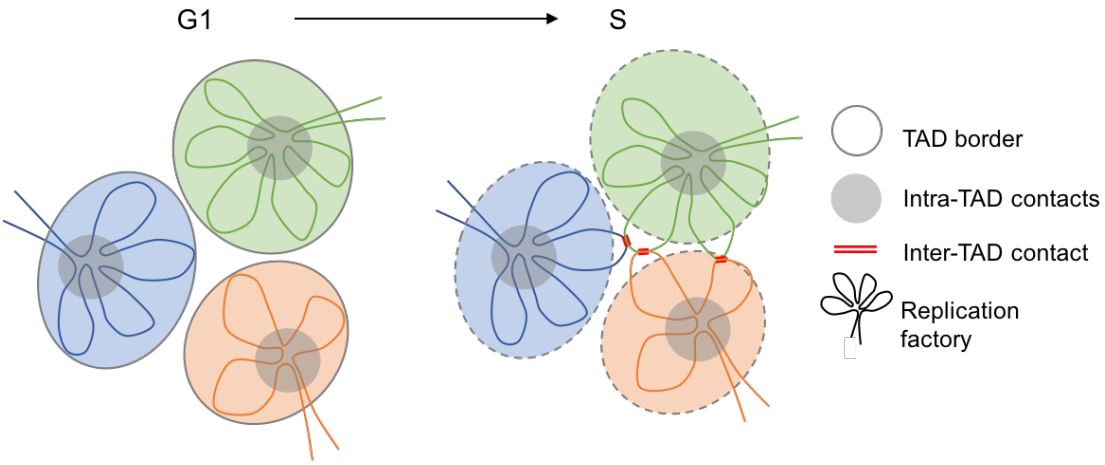


Figure 53. Hierarchical organization of replication origins in the nucleus: a model. At the intra-TAD level, origins are organized within replication factories. Different replication factories could be then connected with each other by inter-TAD interactions. A decrease in the insulation at TAD borders is observed at the beginning of S phase to facilitate inter-TAD contacts to occur. See text for further details.

Multiple studies have indicated that adjacent origins are separated normally by 50-150 kb (Cayrou et al., 2011; Fragkos et al., 2015; Jackson and Pombo, 1998). Our data indicates that the most frequent linear separation between origins that interact in 3D in the PCHiC network is approximately 500 kb. At this distance, interacting origins are most likely to have similar efficiencies of activation. When only intra-TAD origin-origin contacts were considered, the linear separation between origins decreased to approximately 250 kb, closer but still higher than the expected values. Several reasons may contribute to this difference. First, origin-origin interactions may occur between origins that are not adjacent in the linear DNA sequence. It should also be taken into account that our predictions were made in the promoter-centered chromatin interaction map, which includes only a subset of all interactions present in the nucleus. In addition, our analyses consider the presence of one origin per chromatin fragment, but in some cases, there may be more than one origin.

At this point, we would like to propose a model of hierarchical organization of origins within replication factories. The first level of organization would include origin-origin interactions in 3D within TADs. Then, intra-TAD hubs could be further spatially connected through inter-TAD interactions (Fig. 53). Replication factories within TADs would correspond to replication foci that can be microscopically detected by PCNA staining or BrdU incorporation. Two adjacent replication factories brought together by inter-TAD contacts would replicate sequentially in S phase according to a domino-like model of replication foci activation that has been substantiated microscopically

(Maya-Mendoza et al., 2010; Sporbert et al., 2002). In order to obtain additional information about replication factories architecture, we plan to perform similar analysis of distances at which origins preferentially interact using the SMC1 ChIA-PET chromatin network.

In summary, we have aimed at providing new information regarding the characteristics and regulation of mammalian replication origins in control and stress conditions as well as their spatial organization at the genome-wide level. One of the major questions remaining is whether chromatin structure determines the dynamics of replication and transcription, or whether these processes are shaping the spatial organization of the chromatin. A recent study showed that genetic polymorphisms can lead to long-range changes in replication timing and influence gene expression levels. Altered replication timing of a particular genomic region can affect its mutational rate and consequently influence disease and cancer susceptibility (Koren et al., 2014). Another report showed that a single nucleotide polymorphism at a CTCF binding site could alter chromatin topology and transcription, and was also linked to human disease susceptibility (Tang et al., 2015). These observations emphasize the importance of understanding in detail the interplay between 3D chromatin organization, replication and transcription.

Chapter 2: Regulation of CDC6 stability by CDC7 kinase

A CDC7-dependent mechanism of CDC6 protein stabilization

The results shown in Chapter 2 reveal a previously unknown function of CDC7 kinase in the regulation of the stability of CDC6 licensing factor. We have shown that the inhibition or downregulation of CDC7 kinase leads to a decrease in the levels of CDC6 in mouse and human cells. We have determined that the total amount of CDC6 is regulated at the level of protein stability and both the soluble and chromatin-bound fractions of CDC6 are susceptible to changes in CDC7 activity throughout the cell cycle. This effect was not due to a global inhibition of DNA replication in our experimental conditions, as most other factors involved in replication licensing and initiation were not affected. We have not tested the opposite situation, i.e. whether enhanced activity of CDC7, which could be triggered by DBF4 overexpression, results in increased CDC6 levels. It was previously reported that CDC6 is stabilized in a CDK2-CycE-dependent manner during re-entry from G0 into the cell cycle (Mailand and Diffley, 2005) and also in response to genotoxic stress (Duursma and Agami, 2005). Here we propose that CDC7 kinase also has a role in the regulation of CDC6 during the cell cycle in mammalian cells.

CDC7 directly phosphorylates CDC6

We hypothesized that CDC6 could be stabilized by direct CDC7-dependent phosphorylation. Using *in vitro* kinase assays followed by mass spectrometry, phosphorylation events were detected at six out of eight potential CDC7 phosphorylation sites. CDC6 was destabilized *in vivo* when CDC7-dependent phosphorylation was impeded in an unphosphorylatable mutant (S8A). We then monitored how WT and mutant CDC6 proteins respond to a decrease in the levels of CDC7 kinase. As expected, WT-CDC6 levels were reduced upon siCdc7 treatment. In contrast, both phospho-dead and phospho-mimic CDC6 mutants, which cannot be phosphorylated by CDC7, were not affected, further indicating that CDC7 regulates the stability of CDC6 by direct phosphorylation *in vivo*. In *Saccharomyces cerevisiae* Cdc6 is not an efficient substrate for Cdc7 (Weinreich and Stillman, 1999), but the cycle regulation of Cdc6 is very different between yeast and mammalian cells (Arias and Walter, 2007; Diffley, 2004). It is therefore possible that CDC7-dependent phosphorylation of CDC6 creates an additional regulatory mechanism that operates only in higher eukaryotes.

At this stage it is not known how CDC7-dependent phosphorylation could lead to the stabilization of CDC6. In the case of CDK-mediated stabilization of CDC6,

phosphorylated CDC6 fails to interact with APC/C, blocking proteolysis (Duursma and Agami, 2005; Mailand and Diffley, 2005). In contrast, DDK phosphorylation sites in CDC6 are not near APC/C destruction boxes (D-Box and KEN) as it occurs for CDK-dependent phosphorylation sites (Fig. 47 from Results).

Several mechanisms have been described that induce CDC6 proteolysis in S and G2 phases, including PIP-box-mediated proteolysis (Clijsters and Wolthuis, 2014) and $SCF^{CyclinF}$-dependent protein destruction (Walter et al., 2016). These pathways were proposed to serve as mechanisms preventing origin relicensing. Despite this fact, it has been known for a long time that a fraction of CDC6 is protected from these degradation pathways and remains associated to chromatin (Méndez & Stillman, 2000; Coverley et al, 2000). It is possible that this partial protection requires CDC7-mediated phosphorylation. One of the phosphorylation events detected by mass spectrometry lays on S6 (a non-consensus DDK phosphosite), in close proximity to the PIP-box-like degron (Clijsters and Wolthuis, 2014). It is conceivable that phosphorylation of CDC6 at this site could mask this degron, preventing its recognition by $CLR4^{CDT2}$ ubiquitin ligase.

Is CDC6 downregulation in response to limited CDC7 mediated by indirect mechanisms?

CDC6 levels are regulated in a p53-dependent manner. Upon genotoxic stress, p53 leads to p21 upregulation, which in turn inhibits CDK and limits the protective phosphorylation of CDC6 at S54 (Duursma and Agami, 2005). We have observed increased γH2AX levels upon siCdc7 or Cdc7i, suggesting the possible activation of the DNA damage response. At least in U2OS cells, activation of p53 was detected, suggesting that the decrease in CDC6 levels could be in part caused by p53-mediated CDK inhibition. However, it is important to note that in HeLa cells, which are deficient in p53-dependent response, a drop of CDC6 levels was still observed following CDC7 inhibition. Taken together, these results indicate the existence of p53-dependent and p53-independent mechanisms to regulate CDC6 stability upon CDC7 inhibition/downregulation (Fig. 54). To further complicate things, DDK may also be downregulated upon p53 pathway activation, as a mechanism to prevent G1/S transition (Tudzarova et al., 2016). This adds another layer of regulation that could affect CDC6 levels (Fig. 54).

High cyclin E levels are restricted to the G1-S transition. In early S-phase, CDK2-CycA phosphorylates cyclin E targeting it for the SCF^{SKP2} ubiquitin ligase-mediated proteolysis (Depamphilis et al., 2012). We observed that cyclin E levels decrease

upon inhibition of CDC7 activity, in line with previous results (Montagnoli et al., 2008). We speculate that CDC7 could also stabilize cyclin E, potentially by direct phosphorylation (both variants of human cyclin E bear CDC7 consensus phosphorylation sites). Importantly, in some contexts cyclin E contributes to origin licensing by stabilizing CDC6 (Mailand and Diffley, 2005).

Taking all these antecedents into consideration, we propose a model for CDC7-dependent regulation of CDC6 stability. CDC7-mediated phosphorylation acts to stabilize CDC6 protein by a still unknown mechanism and facilitate its accumulation during S, G2 and M phases, when CDC7 is active. Specifically at the G1/S transition, CDC7 could influence CDK2 activity by controlling cyclin E levels. In turn, CDK2-CycE would phosphorylate CDC6 at S54 to stabilize it. In situations of stress, or if CDC7 in inhibited, a p53 response leads to destabilization of CDC6. In addition, p53 activation would contribute to a decrease in CDC7 levels, providing a feedback loop and further destabilizing CDC6 (Fig. 54).

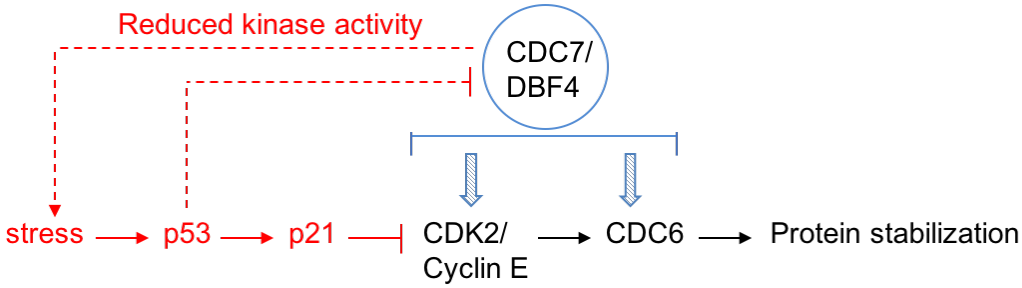


Figure 54. Model of the regulation of CDC6 stability by CDC7 kinase. Details of the model are described in the text.

Potential roles for CDC7 - dependent stabilization of CDC6 during cell cycle

CDC6 performs its origin licensing function in late M/G1 phase. While in yeast Cdc6 degradation at the G1-S transition prevents origin re-licensing in S phase (Diffley, 2004), in mammalian cells a fraction of CDC6 is preserved on chromatin throughout the cell cycle (Coverley et al., 2000; Méndez and Stillman, 2000). CDC6 highest levels are reached in early S, following the rapid increase in DDK activity at the G1/S transition. Thus, cell cycle fluctuations in CDC6 levels resemble the changes in CDC7 activity, strengthening the notion that DDK could have a role in CDC6 stabilization.

We hypothesized that CDC7 could protect CDC6 from APC/C^{CDH1}-dependent degradation by phosphorylating it at the M/G1 transition when licensing step takes place. Intriguingly, the drop in CDC6 levels caused by CDC7 inhibition/ downregulation did not affect the loading of MCM helicase on DNA, at least as determined by biochemical fractionation. This result could be related to the fact that cells with high proliferation rate, such us tumour cells or ESCs, display very high levels of CDC6 (Tatsumi et al., 2006) and origin licensing can occur even after its downregulation.

Overexpression of phosho-dead and phospho-mimic mutants of CDC6 had no impact on MCM chromatin loading or cell cycle progression. The caveat of these experiments is that endogenous CDC6 was still present, making it difficult to assess CDC6-mutant specific effects. To overcome this problem, it would be useful to generate an inducible CDC6 knockout cell line using CRISPR/Cas9 to eliminate the endogenous protein.

CDC7 could protect CDC6 to facilitate some replicative function(s) that are still not well characterized. A provocative hypothesis is that licensing may occur beyond M/G1 in special contexts, such as very origin-poor genomic regions. The fraction of CDC6 bound to chromatin in S and G2 could promote new pre-RC formation or stimulate the activation of already existing pre-RCs at late-replicating domains. Other possibility is that CDC6 protein participates in origin activation. In this regard, chromatin-bound CDC6 in *Xenopus* modulates CDC7-dependent phosphorylation of MCM (Kundu et al., 2010), and CDC6 depletion during S phase in HeLa cells inhibited new origin firing (Lau et al., 2006).

CDC6 may also have functions beyond origin licensing and initiation that justify its stabilization. Human CDC6 localizes at the centrosomes during S, G2 and M phases and may participate in chromosome segregation at mitosis (Kalfalah et al., 2015; Kim et al., 2015). In agreement with this hypothesis, CDC6 depletion in S-phase led to abnormal spindle formation, aberrant chromosome segregation and mitotic cell death in HeLa cells (Lau et al., 2006).

CDC6 was also proposed to participate in checkpoint activation (Clay-Farrace et al., 2003; Oehlmann et al., 2004). In *Xenopus* egg extracts, it is required to activate checkpoint kinase CHK1 in response to aphidicolin (Oehlmann et al., 2004). Overexpression of CDC6 in G2 HeLa cells blocks progression into mitosis but this block is overcome by a CHK1 inhibitor, suggesting that CDC6 prevents premature mitosis via a checkpoint mechanism involving CHK1 (Clay-Farrace et al., 2003).

Recently, it was reported in budding yeast that sequestering of DDK by MCM complexes at unfired origins during S phase prevents cohesin acetyltransferase Eco1 (also a DDK target) from becoming phosphorylated until the late stages of S phase. (Seoane and Morgan, 2017). DDK-dependent phosphorylation of Eco1 targets it for degradation, and the timing of Eco1 degradation is essential for proper chromosome segregation (Lyons and Morgan, 2011). We can also speculate that in mammalian cells, chromatin-bound CDC6 and MCM might sequester CDC7 kinase at non-fired origins to restrict phosphorylation of cohesin acetyltransferase until late S-phase.

In summary, we have found that CDC6 is a previously unknown target of CDC7 kinase. Our results indicate that CDC7 regulates CDC6 stability, by direct and indirect mechanisms. These findings add another layer to the already complex mechanisms that regulate this important replication protein in the cell cycle.

Conclusions

Conclusions

1. Genome-wide maps of replication origins have been generated using SNS-Seq in mouse ESCs in control growth conditions and upon stress induced with aphidicolin or CDC6 overexpression. Origin datasets have been defined using two different algorithms to identify origins at SNS-Seq read peaks.
2. The genomic positions of origins in the WT, APH and CDC6 datasets have been compared to a comprehensive collection of chromatin marks (≈70), many of which were analysed for the first time in the context of origins.
3. The subset of constitutive origins correlates with genomic elements and epigenetic features characteristic of open, transcriptionally active chromatin. Conversely, stress-responsive origins frequently correlate with transcriptionally inactive heterochromatin regions.
4. Constitutive origins display higher efficiency than stress-responsive origins.
5. Replication stress induced by APH or CDC6 overexpression trigger the activation of origins from new positions (´dormant´ origins) and also increase the efficiency of already active ones. We propose that the main response to stress consists in the modulation of the activity of pre-existing origins.
6. Linear maps of replication origins have been integrated into three-dimensional chromatin structure networks. Using the measure of chromatin assortativity, we find that replication origins tend to interact with each other in 3D.
7. Origins that display higher 3D connectivity tend to be activated with higher efficiency and replicate earlier in S phase. Constitutive origins display the highest connectivity from all origin subsets.
8. Origins can be brought together within a same chromatin topologically-associated domain (TAD) and also establish inter-TAD contacts. A hierarchical organization model is proposed in which origins are first organized within intra-TAD replication factories, which then become connected in S phase by inter-TAD origin-origin contacts.
9. CDC7 kinase has a previously undefined role as a regulator of CDC6 protein stability in mammalian cells.
10. The regulation of CDC6 stability by CDC7 kinase likely involves a direct mechanism that includes CDC6 phosphorylation, and an indirect pathway mediated by the regulation of CDK2-CycE activity, which in turn regulates CDC6.

Conclusiones

1. Mediante la técnica de SNS-Seq, se han generado mapas de orígenes de replicación en células ESC murinas, en condiciones normales y en presencia de estrés inducido por afidicolina o sobre-expresión de CDC6. Se han utilizado dos algoritmos independientes para asignar orígenes a los picos de lecturas SNS-Seq.
2. Las posiciones genómicas de los orígenes en células WT, APH y CDC6 se han comparado con una amplia colección de marcadores en cromatina (≈70), muchos de los cuales se analizan por primera vez en el contexto de los orígenes.
3. El subconjunto de orígenes constitutivos correlaciona con elementos genómicos y epigenéticos característicos de cromatina abierta y transcripcionalmente activa. En cambio, los orígenes activados en respuesta a estrés correlacionan principalmente con regiones de heterocromatina transcripcionalmente inactiva.
4. Los orígenes constitutivos se activan con mayor eficiencia que los orígenes de respuesta a estrés.
5. El estrés inducido por APH o CDC6 induce la activación de orígenes "silentes" y también hace aumentar la eficiencia de orígenes activos en condiciones normales. Proponemos que la principal respuesta a estrés está mediada por la modulación de la eficiencia de disparo de orígenes pre-existentes.
6. Los mapas lineales de orígenes de replicación han sido integrados en redes tridimensionales de cromatina. Mediante el parámetro de "asortatividad" empleado en análisis de redes, hemos encontrado que los orígenes tienen tendencia a interaccionar entre ellos en 3D.
7. Los orígenes con mayor grado de conectividad en 3D se activan con mayor eficiencia y replican en la parte más temprana de la fase S. Los orígenes constitutivos presentan el mayor grado de conectividad de todos los subconjuntos de orígenes analizados.
8. Las interacciones entre orígenes se establecen tanto dentro del mismo dominio topológico (intra-TAD) como entre dominios topológicos distintos (intra-TAD). Proponemos un modelo de organización jerárquica en el que los orígenes se disponen en primer lugar en factorías replicativas dentro de un TAD, y éstas a su vez se conectan entre ellas a través de interacciones origen-origen intra-TAD.
9. La quinasa CDC7 tiene una función reguladora de la estabilidad de la proteína CDC6 en células de mamífero, no descrita hasta la fecha.
10. La regulación de la estabilidad de CDC6 por CDC7 implica un mecanismo directo que incluye la fosforilación de CDC6, y un mecanismo indirecto a través de la regulación de CDK2-CycE, cuya actividad controla a su vez los niveles de CDC6

References

References

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

Aladjem, M.I., and Redon, C.E. (2016). Order from clutter: selective interactions at mammalian replication origins. Nat. Rev. Genet. *18*, 101–116.

Alvarez, S., Díaz, M., Flach, J., Rodriguez-Acebes, S., López-Contreras, A.J., Martínez, D., Cañamero, M., Fernández-Capetillo, O., Isern, J., Passegué, E., et al. (2015). Replication stress caused by low MCM expression limits fetal erythropoiesis and hematopoietic stem cell functionality. Nat. Commun. *6*, 8548.

Anglana, M., Apiou, F., Bensimon, A., and Debatisse, M. (2003). Dynamics of DNA replication in mammalian somatic cells: Nucleotide pool modulates origin choice and interorigin spacing. Cell *114*, 385–394.

Antequera, F. (2004). Genomic specification and epigenetic regulation of eukaryotic DNA replication origins. EMBO J. *23*, 4365–4370.

Aparicio, T., Guillou, E., Coloma, J., Montoya, G., and Méndez, J. (2009). The human GINS complex associates with Cdc45 and MCM and is essential for DNA replication. Nucleic Acids Res. *37*, 2087–2095.

Arias, E.E., and Walter, J.C. (2005). Replication-dependent destruction of Cdt1 limits DNA replication to a single round per cell cycle in Xenopus egg extracts. Genes Dev. *19*, 114–126.

Arias, E.E., and Walter, J.C. (2006). PCNA functions as a molecular platform to trigger Cdt1 destruction and prevent re-replication. Nat. Cell Biol. *8*, 84–90.

Arias, E.E., and Walter, J.C. (2007). Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. Genes Dev. *21*, 497–518.

Balasubramani, A., and Rao, A. (2013). O-GlcNAcylation and 5-methylcytosine oxidation: an unexpected association between OGT and TETs. Mol. Cell *49*, 618–619.

Ballabeni, A., Melixetian, M., Zamponi, R., Masiero, L., Marinoni, F., and Helin, K. (2004). Human Geminin promotes pre-RC formation and DNA replication by stabilizing CDT1 in mitosis. EMBO J. *23*, 3122–3132.

Beagrie, R.A., and Pombo, A. (2017). Cell cycle: Continuous chromatin changes. Nature *547*, 34–35.

Berezney, R., Dubey, D.D., and Huberman, J. a (2000). Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. Chromosoma *108*, 471–484.

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.

Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J.-M., and Lemaitre, J.-M. (2012). Unraveling cell type–specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nat. Struct. Mol. Biol. *19*, 837–844.

Bian, Q., and Belmont, A.S. (2012). Revisiting higher-order and large-scale chromatin organization. Curr. Opin. Cell Biol. *24*, 359–366.

Bleichert, F., Botchan, M.R., and Berger, J.M. (2017). Mechanisms for initiating cellular DNA replication. Science *355*, eaah6317.

Blow, J.J., Ge, X.Q., and Jackson, D. a. (2011). How dormant origins promote complete genome replication. Trends Biochem. Sci. *36*, 405–414.

Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nat. Rev. Genet. *17*, 661–678.

Borlado, L.R., and Méndez, J. (2008). CDC6: From DNA replication to cell cycle checkpoints and oncogenesis. Carcinogenesis *29*, 237–243.

Boyer, A.S., Walter, D., and Sørensen, C.S. (2016). DNA replication and cancer: From dysfunctional replication origin activities to therapeutic opportunities. Semin. Cancer Biol. *37–38*.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K.J., Carroll, T., Xie, S.Q., Stock, J.K., Heidemann, M., Eick, D., Nozaki, N., et al. (2012). Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell *10*, 157–170.

Búa, S., Sotiropoulou, P., Sgarlata, C., Borlado, L.R., Domínguez, O., Ortega, S., Malumbres, M., Méndez, J., Búa, S., Sotiropoulou, P., et al. (2015). Deregulated expression of Cdc6 in the skin facilitates papilloma formation and affects the hair growth cycle. Cell Cycle *14*, 3897–3907.

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.

Cairns, J., Freire-Pritchett, P., Wingett, S.W., Várnai, C., Dimond, A., Plagnol, V., Zerbino, D., Schoenfelder, S., Javierre, B.-M., Osborne, C., et al. (2016). CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. Genome Biol. *17*, 127.

Cayrou, C., Coulombe, P., and Méchali, M. (2010). Programming DNA replication origins and chromosome organization. Chromosom. Res. *18*, 137–145.

Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I., Rivals, E., Puy, A., Laurent-Chabalier, S., Desprat, R., et al. (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Res. *21*, 1438–1449.

Cayrou, C., Grégoire, D., Coulombe, P., Danis, E., and Méchali, M. (2012a). Genome-scale identification of active DNA replication origins. Methods *57*, 158–164.

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

Cayrou, C., Ballester, B., Peiffer, I., Fenouil, R., Coulombe, P., Andrau, J., Helden, J. Van, and Méchali, M. (2015). The chromatin environment shapes DNA replication origin

organization and defines origin classes. Genome Res. *25*, 1873–1885.

Chagin, V.O., Stear, J.H., and Cardoso, M.C. (2010). Organization of DNA replication. Cold Spring Harb. Perspect. Biol. *2*.

Chagin, V.O., Casas-Delucchi, C.S., Reinhart, M., Schermelleh, L., Markaki, Y., Maiser, A., Bolius, J.J., Bensimon, A., Fillies, M., Domaing, P., et al. (2016). 4D Visualization of replication foci in mammalian cells corresponding to individual replicons. Nat. Commun. *7*, 11231.

Charych, D.H., Coyne, M., Yabannavar, A., Narberes, J., Chow, S., Wallroth, M., Shafer, C., and Walter, A.O. (2008). Inhibition of Cdc7/Dbf4 kinase activity affects specific phosphorylation sites on MCM2 in cancer cells. J. Cell. Biochem. *104*, 1075–1086.

Cho, W.-H., Lee, Y.-J., Kong, S.-I., Hurwitz, J., and Lee, J.-K. (2006). CDC7 kinase phosphorylates serine residues adjacent to acidic amino acids in the minichromosome maintenance 2 protein. Proc. Natl. Acad. Sci. U. S. A. *103*, 11521–11526.

Clay-Farrace, L., Pelizon, C., Santamaria, D., Pines, J., and Laskey, R.A. (2003). Human replication protein Cdc6 prevents mitosis through a checkpoint mechanism that implicates Chk1. EMBO J. *22*, 704–712.

Clijsters, L., and Wolthuis, R. (2014). PIP-box-mediated degradation prohibits re-accumulation of Cdc6 during S phase. J. Cell Sci. *127*, 1336–1345.

Clijsters, L., Ogink, J., and Wolthuis, R. (2013). The spindle checkpoint, APC/CCdc20, and APC/CCdh1 play distinct roles in connecting mitosis to S phase. J. Cell Biol. *201*, 1013–1026.

Comoglio, F., Schlumpf, T., Schmid, V., Rohs, R., Beisel, C., and Paro, R. (2015). High-Resolution Profiling of Drosophila Replication Start Sites Reveals a DNA Shape and Chromatin Signature of Metazoan Origins. Cell Rep. *11*, 821–834.

Cook, P.R. (1999). The organization of replication and transcription. Science *284*, 1790–1795.

Core, L.J., and Lis, J.T. (2008). Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science *319*, 1791–1792.

Costa, A., Ilves, I., Tamberg, N., Petojevic, T., Nogales, E., Botchan, M.R., and Berger, J.M. (2011). The structural basis for MCM2–7 helicase activation by GINS and Cdc45. Nat. Struct. Mol. Biol. *18*, 471–477.

Costantino, L., Sotiriou, S.K., Rantala, J.K., Magin, S., Mladenov, E., Helleday, T., Haber, J.E., Iliakis, G., Kallioniemi, O.P., and Halazonetis, T.D. (2014). Break-induced replication repair of damaged forks induces genomic duplications in human cells. Science *343*, 88–91.

Courbet, S., Gay, S., Arnoult, N., Wronka, G., Anglana, M., Brison, O., and Debatisse, M. (2008). Replication fork movement sets chromatin loop size and origin choice in mammalian cells. Nature *455*, 557–560.

Coverley, D., Wilkinson, H.R., Madine, M. a, Mills, a D., and Laskey, R. a (1998). Protein kinase inhibition in G2 causes mammalian Mcm proteins to reassociate with chromatin

and restores ability to replicate. Exp. Cell Res. *238*, 63–69.

Coverley, D., Pelizon, C., Trewick, S., and Laskey, R.A. (2000). Chromatin-bound Cdc6 persists in S and G 2 phases in human cells , while soluble Cdc6 is destroyed in a cyclin A-cdk2 dependent process. J. Cell Sci. *113*, 1929–1938.

Deegan, T.D., and Diffley, J.F.X. (2016). MCM: One ring to rule them all. Curr. Opin. Struct. Biol. *37*, 145–151.

Delgado, S., Gómez, M., Bird, A., and Antequera, F. (1998). Initiation of DNA replication at CpG islands in mammalian chromosomes. EMBO J. *17*, 2426–2435.

Dellino, G.I., Cittaro, D., Piccioni, R., Luzi, L., Banfi, S., Segalla, S., Cesaroni, M., Mendoza-Maldonado, R., Giacca, M., and Pelicci, P.G. (2013). Genome-wide mapping of human DNA-replication origins: Levels of transcription at ORC1 sites regulate origin selection and replication timing. Genome Res. *23*, 1–11.

Depamphilis, M.L., Renty, C.M. De, Ullah, Z., and Lee, C.Y. (2012). " The octet ": eight protein kinases that control mammalian DNA replication. Front. Physiol. *3*, 1–20.

Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Méndez, J., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E., et al. (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. EMBO J. *32*, 645–655.

Deursen, F. Van, Sengupta, S., Piccoli, G. De, Sanchez-diaz, A., and Labib, K. (2012). Mcm10 associates with the loaded DNA helicase at replication origins and defines a novel step in its activation. EMBO J. *31*, 2195–2206.

Devbhandari, S., Jiang, J., Kumar, C., Whitehouse, I., and Remus, D. (2017). Chromatin Constrains the Initiation and Elongation of DNA Replication. Mol. Cell *65*, 131–141.

Diffley, J.F.X. (2004). Regulation of Early Events in Chromosome Replication. Curr. Biol. *14*, R778–R786.

Diffley, J.F.X. (2011). Quality control in the initiation of eukaryotic DNA replication. Philos. Trans. R. Soc. B Biol. Sci. *366*, 3545–3553.

Dileep, V., Rivera-Mulia, J.C., Sima, J., and Gilbert, D.M. (2015a). Large-Scale Chromatin Structure–Function Relationships during the Cell Cycle and Development: Insights from Replication Timing. Cold Spring Harb. Symp. Quant. Biol. *80*, 53–63.

Dileep, V., Ay, F., Sima, J., Vera, D.L., Noble, W.S., and Gilbert, D.M. (2015b). Topologically associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication-timing program. Genome Res. *25*, 1104–1113.

Dimitrova, D.S., and Gilbert, D.M. (1998). Regulation of mammalian replication origin usage in Xenopus egg extract. J. Cell Sci. *111 (Pt 19)*, 2989–2998.

Dimitrova, E., Turberfield, A.H., and Klose, R.J. (2015). Histone demethylases in chromatin biology and beyond. EMBO Rep. *16*, 1620–1639.

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.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schujiers, J., Lee, T.I., Zhao, K., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159*, 374–387.

Duursma, A.M., and Agami, R. (2005). CDK-dependent stabilization of Cdc6: Linking growth and stress signals to activation of DNA replication. Cell Cycle *4*, 1725–1728.

Eaton, M.L., Prinz, J. a., MacAlpine, H.K., Tretyakov, G., Kharchenko, P. V., and MacAlpine, D.M. (2011). Chromatin signatures of the Drosophila replication program. Genome Res. *21*, 164–174.

Ekholm-Reed, S., Méndez, J., Tedesco, D., Zetterberg, A., Stillman, B., and Reed, S.I. (2004). Deregulation of cyclin E in human cells interferes with prereplication complex assembly. J. Cell Biol. *165*, 789–800.

Ernst, J., and Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat. Biotechnol. *28*, 817–825.

Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., Li, H., Stillman, B., and Speck, C. (2009). A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. Proc. Natl. Acad. Sci. U. S. A. *106*, 20240–20245.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat. Protoc. *7*, 1728–1740.

Ferreira, M.F., Santocanale, C., Drury, L.S., and Diffley, J.F. (2000). Dbf4p, an essential S phase-promoting factor, is targeted for degradation by the anaphase-promoting complex. Mol. Cell. Biol. *20*, 242–248.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., et al. (2010). Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. Cell *143*, 212–224.

Foulk, M.S., Urban, J.M., Casella, C., and Gerbi, S.A. (2015). Characterizing and controlling intrinsic biases of lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. Genome Res. *25*, 725–735.

Fragkos, M., Ganier, O., Coulombe, P., and Méchali, M. (2015). DNA replication origin activation in space and time. Nat. Rev. Mol. Cell Biol. *16*, 360–374.

Fu, H., Maunakea, A.K., Martin, M.M., Huang, L., Zhang, Y., Ryan, M., Kim, R.G., Lin, C.M., Zhao, K., and Aladjem, M.I. (2013). Methylation of Histone H3 on Lysine 79 Associates with a Group of Replication Origins and Helps Limit DNA Replication Once per Cell Cycle. PLoS Genet. *9*, e1003542.

Fu, H., Martin, M.M., Regairaz, M., Huang, L., You, Y., Lin, C.-M., Ryan, M., Kim, R., Shimura, T., Pommier, Y., et al. (2015). The DNA repair endonuclease Mus81 facilitates fast DNA replication in the absence of exogenous damage. Nat. Commun. *6*, 6746.

García-Muse, T., and Aguilera, A. (2016). Transcription–replication conflicts: how they occur

and how they are resolved. Nat. Rev. Mol. Cell Biol. *17*, 553–563.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. J. Mol. Biol. *196*, 261–282.

Ge, X.Q., and Blow, J.J. (2010). Chk1 inhibits replication factory activation but allows dormant origin firing in existing factories. J. Cell Biol. *191*, 1285–1297.

Ge, X.Q., Jackson, D.A., and Blow, J.J. (2007). Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. Genes Dev. *21*, 3331–3341.

Gilbert, D.M. (2004). In search of the holy replicator. Nat. Rev. Mol. Cell Biol. *5*, 848–855.

Gilbert, D.M. (2007). Replication origin plasticity, Taylor-made: inhibition vs recruitment of origins under conditions of replication stress. Chromosoma *116*, 341–347.

Gilbert, D.M. (2010). Evaluating genome-scale approaches to eukaryotic DNA replication. Nat. Rev. Genet. *11*, 673–684.

Gómez, M., and Antequera, F. (2008). Overreplication of short DNA regions during S phase in human cells. Genes Dev. *22*, 375–385.

Gros, J., Kumar, C., Lynch, G., Yadav, T., Whitehouse, I., and Remus, D. (2015). Post-licensing Specification of Eukaryotic Replication Origins by Facilitated Mcm2-7 Sliding along DNA. Mol. Cell *60*, 797–807.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948–951.

Guillou, E., Ibarra, A., Coulon, V., Casado-Vela, J., Rico, D., Casal, I., Schwob, E., Losada, A., and Méndez, J. (2010). Cohesin organizes chromatin loops at DNA replication factories. Genes Dev. *24*, 2812–2822.

Hayashi, M., Katou, Y., Itoh, T., Tazumi, A., Yamada, Y., Takahashi, T., Nakagawa, T., Shirahige, K., and Masukata, H. (2007). Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. EMBO J. *26*, 1327–1339.

Heichinger, C., Penkett, C.J., Bähler, J., and Nurse, P. (2006). Genome-wide characterization of fission yeast DNA replication origins. EMBO J. *25*, 5171–5179.

Heller, R.C., Kang, S., Lam, W.M., Chen, S., Chan, C.S., and Bell, S.P. (2011). Eukaryotic origin-dependent DNA replication in vitro reveals sequential action of DDK and S-CDK kinases. Cell *146*, 80–91.

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*, 2220–2236.

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.

Hoshina, S., Yura, K., Teranishi, H., Kiyasu, N., Tominaga, A., Kadoma, H., Nakatsuka, A., Kunichika, T., Obuse, C., and Waga, S. (2013). Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. J. Biol. Chem.

288, 30161–30171.

Hozák, P., Hassan, A.B., Jackson, D.A., and Cook, P.R. (1993). Visualization of replication factories attached to a nucleoskeleton. Cell *73*, 361–373.

Huberman, J. a, and Riggs, a D. (1966). Autoradiography of chromosomal DNA fibers from Chinese hamster cells. Proc. Natl. Acad. Sci. U. S. A. *55*, 599–606.

Huppert, J.L., and Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. Nucleic Acids Res. *33*, 2908–2916.

Hyrien, O. (2015). Peaks cloaked in the mist: The landscape of mammalian replication origins. J. Cell Biol. *208*, 147–160.

Ibarra, A., Schwob, E., and Méndez, J. (2008). Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. Proc. Natl. Acad. Sci. U. S. A. *105*, 8956–8961.

Ilves, I., Petojevic, T., Pesavento, J.J., and Botchan, M.R. (2010). Activation of the MCM2-7 Helicase by Association with Cdc45 and GINS Proteins. Mol. Cell *37*, 247–258.

Jackson, D.A., and Pombo, A. (1998). Replicon clusters are stable units of chromosome structure: Evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. J. Cell Biol. *140*, 1285–1295.

Jacob, F., Brenner, S., and Cuzin, F. (1963). On the Regulation of DNA Replication in Bacteria. Cold Spring Harb. Symp. Quant. Biol. *28*, 329–348.

Jiang, W., Wells, N.J., and Hunter, T. (1999). Multistep regulation of DNA replication by Cdk phosphorylation of HsCdc6. Proc. Natl. Acad. Sci. U. S. A. *96*, 6193–6198.

Jin, J., Arias, E.E., Chen, J., Harper, J.W., and Walter, J.C. (2006). A Family of Diverse Cul4-Ddb1-Interacting Proteins Includes Cdt2, which Is Required for S Phase Destruction of the Replication Factor Cdt1. Mol. Cell *23*, 709–721.

Juan, D., Perner, J., Carrillo de Santa Pau, E., Marsili, S., Ochoa, D., Chung, H.R., Vingron, M., Rico, D., and Valencia, A. (2016). Epigenomic Co-localization and Co-evolution Reveal a Key Role for 5hmC as a Communication Hub in the Chromatin Network of ESCs. Cell Rep. *14*, 1246–1257.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kalfalah, F.M., Berg, E., Christensen, M.O., Linka, R.M., Wilhelm, G., Boege, F., Mielke, C., Kalfalah, F.M., Berg, E., Christensen, M.O., et al. (2015). Spatio-temporal regulation of the human licensing factor Cdc6 in replication and mitosis. Cell Cycle *14*, 1704–1715.

Kim, G.S., Kang, J., Bang, S.W., and Hwang, D.S. (2015). Cdc6 localizes to S- and G2-phase centrosomes in a cell cycle-dependent manner. Biochem. Biophys. Res. Commun. *456*, 763–767.

Kitamura, E., Blow, J.J., and Tanaka, T.U. (2006). Live-Cell Imaging Reveals Replication of Individual Replicons in Eukaryotic Replication Factories. Cell *125*, 1297–1308.

Koren, A., Handsaker, R.E., Kamitaki, N., Karlić, R., Ghosh, S., Polak, P., Eggan, K., and

McCarroll, S.A. (2014). Genetic Variation in Human DNA Replication Timing. Cell *159*, 1015–1026.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. PLoS Genet. *4*, e1000242.

Kumagai, H., Sato, N., Yamada, M., Mahony, D., Seghezzi, W., Lees, E., Arai, K.-I., and Masai, H. (1999). A novel growth- and cell cycle-regulated protein, ASK, activates human Cdc7-related kinase and is essential for G1/S transition in mammalian cells. Mol. Cell. Biol. *19*, 5083–5095.

Kundu, L.R., Kumata, Y., Kakusho, N., Watanabe, S., Furukohri, A., Waga, S., Seki, M., Masai, H., Enomoto, T., and Tada, S. (2010). Deregulated Cdc6 inhibits DNA replication and suppresses Cdc7-mediated phosphorylation of Mcm2-7 complex. Nucleic Acids Res. *38*, 5409–5418.

Kunkel, T.A., and Burgers, P.M. (2008). Dividing the workload at a eukaryotic replication fork. Trends Cell Biol. *18*, 521–527.

Kuo, A.J., Song, J., Cheung, P., Ishibe-murakami, S., Yamazoe, S., Chen, J.K., Patel, D.J., and Gozani, O. (2012). The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier–Gorlin syndrome. Nature *484*, 115–119.

Kurat, C.F., Yeeles, J.T.P., Patel, H., Early, A., and Diffley, J.F.X. (2017). Chromatin Controls DNA Replication Origin Selection, Lagging-Strand Synthesis, and Replication Fork Rates. Mol. Cell *65*, 117–130.

Kwak, H., and Lis, J.T. (2013). Control of transcriptional elongation. Annu. Rev. Genet. *47*, 483–508.

Langley, A.R., Gräf, S., Smith, J.C., and Krude, T. (2016). Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq). Nucleic Acids Res. *44*, 10230–10247.

Lau, E., Zhu, C., Abraham, R.T., and Jiang, W. (2006). The functional role of Cdc6 in S-G2/M in mammalian cells. EMBO Rep. *7*, 425–430.

Lee, C., Hong, B., Choi, J.M., Kim, Y., Watanabe, S., Ishimi, Y., Enomoto, T., Tada, S., Kim, Y., and Cho, Y. (2004). Structural basis for inhibition of the replication licensing factor Cdt1 by geminin. Nature *430*, 913–917.

Lee, K.Y., Bang, S.W., Yoon, S.W., Lee, S.H., Yoon, J.B., and Hwang, D.S. (2012). Phosphorylation of ORC2 protein dissociates origin recognition complex from chromatin and replication origins. J. Biol. Chem. *287*, 11891–11898.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, C. -j., Vassilev, A., and DePamphilis, M.L. (2004). Role for Cdk1 (Cdc2)/Cyclin A in Preventing the Mammalian Origin Recognition Complex's Largest Subunit (Orc1) from Binding to Chromatin during Mitosis. Mol. Cell. Biol. *24*, 5875–5886.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, H., Liefke, R., Jiang, J., Kurland, J.V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D.J., Bulyk, M.L., et al. (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. Nature *549*, 287–291.

Li, X., Zhao, Q., Liao, R., Sun, P., and Wu, X. (2003). The SCFSkp2 ubiquitin ligase complex interacts with the human replication licensing factor Cdt1 and regulates Cdt1 degradation. J. Biol. Chem. *278*, 30854–30858.

Löb, D., Lengert, N., Chagin, V.O., Reinhart, M., Casas-Delucchi, C.S., Cardoso, M.C., and Drossel, B. (2016). 3D replicon distributions arise from stochastic initiation and domino-like DNA replication progression. Nat. Commun. *7*, 11207.

Lovejoy, C.A., Lock, K., Yenamandra, A., and Cortez, D. (2006). DDB1 maintains genome integrity through regulation of Cdt1. Mol. Cell. Biol. *26*, 7977–7990.

Lyons, N.A., and Morgan, D.O. (2011). Cdk1-dependent destruction of Eco1 prevents cohesion establishment after S phase. Mol. Cell *42*, 378–389.

MacAlpine, H.K., Gordân, R., Powell, S.K., Hartemink, A.J., and MacAlpine, D.M. (2010). Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. Genome Res. *20*, 201–211.

Mailand, N., and Diffley, J.F.X. (2005). CDKs promote DNA replication origin licensing in human cells by protecting Cdc6 from APC/C-dependent proteolysis. Cell *122*, 915–926.

Manchado, E., Eguren, M., and Malumbres, M. (2010). The anaphase-promoting complex/cyclosome (APC/C): cell-cycle-dependent and -independent functions. Biochem. Soc. Trans. *38*, 65–71.

Mantiero, D., Mackenzie, A., Donaldson, A., and Zegerman, P. (2011). Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. EMBO J. *30*, 4805–4814.

Martin, M.M., Ryan, M., Kim, R., Zakas, A.L., Fu, H., Lin, C.M., Reinhold, W.C., Davis, S.R., Bilke, S., Liu, H., et al. (2011). Genome-wide depletion of replication initiation events in highly transcribed regions. Genome Res. *21*, 1822–1832.

Masai, H., and Arai, K.-I. (2002). Cdc7 kinase complex: a key regulator in the initiation of DNA replication. J. Cell. Physiol. *190*, 287–296.

Masai, H., Matsumoto, S., You, Z., Yoshizawa-Sugata, N., and Oda, M. (2010). Eukaryotic chromosome DNA replication: where, when, and how? Annu. Rev. Biochem. *79*, 89–130.

Maya-Mendoza, A., Petermann, E., Gillespie, D. a F., Caldecott, K.W., and Jackson, D. a (2007). Chk1 regulates the density of active replication origins during the vertebrate S phase. EMBO J. *26*, 2719–2731.

Maya-Mendoza, A., Olivares-Chauvet, P., Shaw, A., and Jackson, D.A. (2010). S Phase Progression in Human Cells Is Dictated by the Genetic Continuity of DNA Foci. PLoS Genet. *6*, e1000900.

McGarry, T.J., and Kirschner, M.W. (1998). Geminin, an inhibitor of DNA replication, is degraded during mitosis. Cell *93*, 1043–1053.

McGuffee, S.R., Smith, D.J., and Whitehouse, I. (2013). Quantitative, Genome-Wide Analysis of Eukaryotic Replication Initiation and Termination. Mol. Cell *50*, 123–135.

McIntosh, D., and Blow, J.J. (2012). Dormant Origins, the Licensing Checkpoint, and the Response to Replicative Stresses. Cold Spring Harb. Perspect. Biol. *4*, a012955–a012955.

Méchali, M. (2010). Eukaryotic DNA replication origins: many choices for appropriate answers. Nat. Rev. Mol. Cell Biol. *11*, 728–738.

Melixetian, M., Ballabeni, A., Masiero, L., Gasparini, P., Zamponi, R., Bartek, J., Lukas, J., and Helin, K. (2004). Loss of Geminin induces rereplication in the presence of functional p53. J. Cell Biol. *165*, 473–482.

Méndez, J., and Stillman, B. (2000). Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. Mol. Cell. Biol. *20*, 8602–8612.

Méndez, J., Zou-Yang, X.H., Kim, S.-Y., Hidaka, M., Tansey, W.P., and Stillman, B. (2002). Human origin recognition complex large subunit is degraded by ubiquitin-mediated proteolysis after initiation of DNA replication. Mol. Cell *9*, 481–491.

Mesner, L.D., Valsakumar, V., Cieslik, M., Pickin, M., Hamlin, J.L., and Bekiranov, S. (2013). Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. Genome Res. *23*, 1774–1788.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Miotto, B., and Struhl, K. (2010). HBO1 Histone Acetylase Activity Is Essential for DNA Replication Licensing and Inhibited by Geminin. Mol. Cell *37*, 57–66.

Miotto, B., Ji, Z., and Struhl, K. (2016). Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. Proc. Natl. Acad. Sci. U. S. A. *113*, E4810–E4819.

Montagnoli, A., Tenca, P., Sola, F., Carpani, D., Brotherton, D., Albanese, C., and Santocanale, C. (2004). Cdc7 inhibition reveals a p53-dependent replication checkpoint that is defective in cancer cells. Cancer Res. *64*, 7110–7116.

Montagnoli, A., Valsasina, B., Brotherton, D., Troiani, S., Rainoldi, S., Tenca, P., Molinari, A., and Santocanale, C. (2006). Identification of Mcm2 phosphorylation sites by S-phase-regulating kinases. J. Biol. Chem. *281*, 10281–10290.

Montagnoli, A., Valsasina, B., Croci, V., Menichincheri, M., Rainoldi, S., Marchesi, V., Tibolla, M., Tenca, P., Brotherton, D., Albanese, C., et al. (2008). A Cdc7 kinase inhibitor restricts initiation of DNA replication and has antitumor activity. Nat. Chem. Biol. *4*, 357–365.

Moyer, S.E., Lewis, P.W., and Botchan, M.R. (2006). Isolation of the Cdc45/Mcm2-7/GINS

(CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. Proc. Natl. Acad. Sci. U. S. A. *103*, 10236–10241.

Mukhopadhyay, R., Lajugie, J., Fourel, N., Selzer, A., Schizas, M., Bartholdy, B., Mar, J., Lin, C.M., Martin, M.M., Ryan, M., et al. (2014). Allele-specific genome-wide profiling in human primary erythroblasts reveal replication program organization. PLoS Genet. *10*, e1004319.

Muñoz, S., and Méndez, J. (2017). DNA replication stress: from molecular mechanisms to human disease. Chromosoma *126*, 1–15.

Muñoz, S., Búa, S., Rodríguez-Acebes, S., Megías, D., Ortega, S., de Martino, A., and Méndez, J. (2017). In Vivo DNA Re-replication Elicits Lethal Tissue Dysplasias. Cell Rep. *19*, 928–938.

Nagano, T., Lubling, Y., Várnai, C., Dudley, C., Leung, W., Baran, Y., Mendelson Cohen, N., Wingett, S., Fraser, P., and Tanay, A. (2017). Cell-cycle dynamics of chromosomal organization at single-cell resolution. Nature *547*, 61–67.

Nieduszynski, C. a, Knox, Y., and Donaldson, A.D. (2006). Genome-wide identification of replication origins in yeast by comparative genomics. Genes Dev. *20*, 1874–1879.

Nishitani, H., Sugimoto, N., Roukos, V., Nakanishi, Y., Saijo, M., Obuse, C., Tsurimoto, T., Nakayama, K.I., Nakayama, K., Fujita, M., et al. (2006). Two E3 ubiquitin ligases, SCF-Skp2 and DDB1-Cul4, target human Cdt1 for proteolysis. EMBO J. *25*, 1126–1136.

Oehlmann, M., Score, A.J., and Blow, J.J. (2004). The role of Cdc6 in ensuring complete genome licensing and S phase checkpoint activation. J. Cell Biol. *165*, 181–190.

Oshiro, G., Owens, J.C., Shellman, Y., Sclafani, R.A., and Li, J.J. (1999). Cell cycle control of Cdc7p kinase activity through regulation of Dbf4p stability. Mol. Cell. Biol. *19*, 4888–4896.

Pancaldi, V., Carrillo-de-santa-pau, E., Javierre, B.M., Juan, D., Fraser, P., Spivakov, M., Valencia, A., and Rico, D. (2016). Integrating epigenomic data and 3D genomic structure with a new measure of chromatin assortativity. Genome Biol. *17*, 152.

Paolinelli, R., Mendoza-Maldonado, R., Cereseto, A., and Giacca, M. (2009). Acetylation by GCN5 regulates CDC6 phosphorylation in the S phase of the cell cycle. Nat Struct Mol Biol *16*, 412–420.

Petersen, B.O., Lukas, J., Sørensen, C.S., Bartek, J., and Helin, K. (1999). Phosphorylation of mammalian CDC6 by cyclin A/CDK2 regulates its subcellular localization. EMBO J. *18*, 396–410.

Petersen, B.O., Wagener, C., Marinoni, F., Kramer, E.R., Melixetian, M., Denchi, E.L., Gieffers, C., Matteucci, C., Peters, J.M., and Helin, K. (2000). Cell cycle- and cell growth-regulated proteolysis of mammalian CDC6 is dependent on APC-CDH1. Genes Dev. *14*, 2330–2343.

Petryk, N., Kahli, M., d'Aubenton-Carafa, Y., Jaszczyszyn, Y., Shen, Y., Silvain, M., Thermes, C., Chen, C.-L., and Hyrien, O. (2016). Replication landscape of the human genome. Nat. Commun. *7*, 10208.

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. Cell *153*, 1281–1295.

Picard, F., Cadoret, J.-C., Audit, B., Arneodo, A., Alberti, A., Battail, C., Duret, L., and Prioleau, M.-N. (2014). The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. PLoS Genet. *10*, e1004282.

Pombo, A., and Dillon, N. (2015). Three-dimensional genome architecture: players and mechanisms. Nat. Rev. Mol. Cell Biol. *16*, 245–257.

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., et al. (2014). Topologically associating domains are stable units of replication-timing regulation. Nature *515*, 402–405.

Powell, S.K., MacAlpine, H.K., Prinz, J.A., Li, Y., Belsky, J.A., and MacAlpine, D.M. (2015). Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. EMBO J. *34*, 531–543.

Prioleau, M.-N., and MacAlpine, D.M. (2016). DNA replication origins—where do we begin? Genes Dev. *30*, 1683–1697.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2004). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. *33*, D501–D504.

Pruitt, K.D., Brown, G.R., Hiatt, S.M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., Farrell, C.M., Hart, J., Landrum, M.J., McGarvey, K.M., et al. (2014). RefSeq: an update on mammalian reference sequences. Nucleic Acids Res. *42*, D756–D763.

Pruitt, S.C., Bailey, K.J., and Freeland, A. (2007). Reduced Mcm2 expression results in severe stem/progenitor cell deficiency and cancer. Stem Cells *25*, 3121–3132.

Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr. Protoc. Bioinforma. *47*, 11.12.1-34.

Quinn, L.M., Herr, A., McGarry, T.J., and Richardson, H. (2001). The Drosophila Geminin homolog: Roles for Geminin in limiting DNA replication, in anaphase and in neurogenesis. Genes Dev. *15*, 2741–2754.

Rainey, M.D., Harhen, B., Wang, G.N., Murphy, P. V., and Santocanale, C. (2013). Cdc7-dependent and -independent phosphorylation of Claspin in the induction of the DNA replication checkpoint. Cell Cycle *12*, 1560–1568.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., and Diffley, J.F.X. (2009). Concerted Loading of Mcm2–7 Double Hexamers around DNA during DNA Replication

Origin Licensing. Cell *139*, 719–730.

Rhind, N., and Gilbert, D.M. (2013). DNA replication timing. Cold Spring Harb. Perspect. Biol. *5*, a010132.

Ricci, M.A., Manzo, C., García-Parajo, M.F., Lakadamyali, M., and Cosma, M.P. (2015). Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. Cell *160*, 1145–1158.

Rivera-Mulia, J.C., and Gilbert, D.M. (2016a). Replicating Large Genomes: Divide and Conquer. Mol. Cell *62*, 756–765.

Rivera-Mulia, J.C., and Gilbert, D.M. (2016b). Replication timing and transcriptional control: Beyond cause and effect - part III. Curr. Opin. Cell Biol. *40*, 168–178.

Rubio-Camarillo, M., López-Fernández, H., Gómez-López, G., Carro, Á., Fernández, J.M., Torre, C.F., Fdez-Riverola, F., and Glez-Peña, D. (2017). RUbioSeq+: A multiplatform application that executes parallelized pipelines to analyse next-generation sequencing data. Comput. Methods Programs Biomed. *138*, 73–81.

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T.C., Robins, A.J., Dalton, S., and Gilbert, D.M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res. *20*, 761–770.

Saha, P., Chen, J., Thome, K.C., Lawlis, S.J., Hou, Z.H., Hendricks, M., Parvin, J.D., and Dutta, A. (1998). Human CDC6/Cdc18 associates with Orc1 and cyclin-cdk and is selectively eliminated from the nucleus at the onset of S phase. Mol Cell Biol *18*, 2758–2767.

Sahlén, P., Abdullayev, I., Ramsköld, D., Matskova, L., Rilakovic, N., Lötstedt, B., Albert, T.J., Lundeberg, J., and Sandberg, R. (2015). Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. Genome Biol. *16*, 156.

Sandhu, K.S., Li, G., Poh, H.M., Quek, Y.L.K., Sia, Y.Y., Peh, S.Q., Mulawadi, F.H., Lim, J., Sikic, M., Menghi, F., et al. (2012). Large-Scale Functional Organization of Long-Range Chromatin Interaction Networks. Cell Rep. *2*, 1207–1219.

Sauvageau, M., and Sauvageau, G. (2010). Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. Cell Stem Cell *7*, 299–313.

Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.-M., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-Cadete, F., Furlan-Magaril, M., et al. (2015a). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. Nat. Genet. *47*, 1179–1186.

Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S.W., et al. (2015b). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res. *25*, 582–597.

Segurado, M., de Luis, A., and Antequera, F. (2003). Genome-wide distribution of DNA replication origins at A+T-rich islands in Schizosaccharomyces pombe. EMBO Rep. *4*,

1048–1053.

Seoane, A.I., and Morgan, D.O. (2017). Firing of Replication Origins Frees Dbf4-Cdc7 to Target Eco1 for Destruction. Curr. Biol. *27*, 2849–2855.e2.

Sequeira-Mendes, J., and Gómez, M. (2012). On the opportunistic nature of transcription and replication initiation in the metazoan genome. BioEssays *34*, 119–125.

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*, e1000446.

Sheu, Y.-J., and Stillman, B. (2006). Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. Mol. Cell *24*, 101–113.

Shima, N., Alcaraz, A., Liachko, I., Buske, T.R., Andrews, C.A., Munroe, R.J., Hartford, S.A., Tye, B.K., and Schimenti, J.C. (2007). A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice. Nat. Genet. *39*, 93–98.

Shimi, T., Butin-Israeli, V., Adam, S.A., and Goldman, R.D. (2010). Nuclear Lamins in Cell Regulation and Disease. Cold Spring Harb. Symp. Quant. Biol. *75*, 525–531.

Siddiqui, K., On, K.F., and Diffley, J.F.X. (2013). Regulating DNA replication in eukarya. Cold Spring Harb. Perspect. Biol. *5*, a012930–a012930.

Simon, J.A., and Kingston, R.E. (2013). Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. Mol. Cell *49*, 808–824.

Smith, D.J., and Whitehouse, I. (2012). Intrinsic coupling of lagging-strand synthesis to chromatin assembly. Nature *483*, 434–438.

Sporbert, A., Gahl, A., Ankerhold, R., Leonhardt, H., and Cardoso, M.C. (2002). DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters. Mol. Cell *10*, 1355–1365.

Stevens, T.J., Lando, D., Basu, S., Atkinson, L.P., Cao, Y., Lee, S.F., Leeb, M., Wohlfahrt, K.J., Boucher, W., O'Shaughnessy-Kirwan, A., et al. (2017). 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature *544*, 59–64.

Sugimoto, N., Tatsumi, Y., Tsurumi, T., Matsukage, A., Kiyono, T., Nishitani, H., and Fujita, M. (2004). Cdt1 Phosphorylation by Cyclin A-dependent Kinases Negatively Regulates Its Function without Affecting Geminin Binding. J. Biol. Chem. *279*, 19691–19697.

Takeda, D.Y., Parvin, J.D., and Dutta, A. (2005). Degradation of Cdt1 during S phase is Skp2-independent and is required for efficient progression of mammalian cells through S phase. J. Biol. Chem. *280*, 23416–23423.

Tanaka, S., and Araki, H. (2013). Helicase activation and establishment of replication forks at chromosomal origins of replication. Cold Spring Harb. Perspect. Biol. *5*, a010371.

Tanaka, S., Umemori, T., Hirai, K., Muramatsu, S., Kamimura, Y., and Araki, H. (2007). CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. Nature *445*, 328–332.

Tanaka, S., Nakato, R., Katou, Y., Shirahige, K., and Araki, H. (2011). Origin association of

Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. Curr. Biol. *21*, 2055–2063.

Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B., et al. (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell *163*, 1611–1627.

Tardat, M., Brustel, J., Kirsh, O., Lefevbre, C., Callanan, M., Sardet, C., and Julien, E. (2010). The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. Nat. Cell Biol. *12*, 1086–1093.

Tatsumi, Y., Ohta, S., Kimura, H., Tsurimoto, T., and Obuse, C. (2003). The ORC1 cycle in human cells: I. cell cycle-regulated oscillation of human ORC1. J. Biol. Chem. *278*, 41528–41534.

Tatsumi, Y., Sugimoto, N., Yugawa, T., Narisawa-Saito, M., Kiyono, T., and Fujita, M. (2006). Deregulation of Cdt1 induces chromosomal damage without rereplication and leads to chromosomal instability. J. Cell Sci. *119*, 3128–3140.

Taylor, J.H. (1977). Increase in DNA replication sites in cells held at the beginning of S phase. Chromosoma *62*, 291–300.

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

Toledo, L.I., Altmeyer, M., Rask, M.B., Lukas, C., Larsen, D.H., Povlsen, L.K., Bekker-Jensen, S., Mailand, N., Bartek, J., and Lukas, J. (2014). ATR prohibits replication catastrophe by preventing global exhaustion of RPA. Cell *156*, 374.

Tudzarova, S., Mulholland, P., Dey, A., Stoeber, K., Okorokov, A.L., and Williams, G.H. (2016). p53 controls CDC7 levels to reinforce G1 cell cycle arrest upon genotoxic stress. Cell Cycle *15*, 2958–2972.

Urban, J.M., Foulk, M.S., Casella, C., and Gerbi, S.A. (2015). The hunt for origins of DNA replication in multicellular eukaryotes. F1000Prime Rep. *7*, 30.

Utani, K., Fu, H., Jang, S.-M., Marks, A.B., Smith, O.K., Zhang, Y., Redon, C.E., Shimizu, N., and Aladjem, M.I. (2017). Phosphorylated SIRT1 associates with replication origins to prevent excess replication initiation and preserve genomic stability. Nucleic Acids Res. *45*, 7807–7824.

Valenzuela, M.S., Chen, Y., Davis, S., Yang, F., Walker, R.L., Bilke, S., Lueders, J., Martin, M.M., Aladjem, M.I., Massion, P.P., et al. (2011). Preferential localization of human origins of DNA replication at the 5'-ends of expressed genes and at evolutionarily conserved DNA sequences. PLoS One *6*, e17308.

Valton, A.L., Hassan-Zadeh, V., Lema, I., Boggetto, N., Alberti, P., Saintomé, C., Riou, J.F., and Prioleau, M.N. (2014). G4 motifs affect origin positioning and efficiency in two vertebrate replicators. EMBO J. *33*, 732–746.

Vashee, S., Cvetic, C., Lu, W., Simancek, P., Kelly, T.J., and Walter, J.C. (2003). Sequence-independent DNA binding and replication initiation by the human origin recognition

complex. Genes Dev. *17*, 1894–1908.

Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K., et al. (2013). Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. Mol. Cell *49*, 645–656.

Voigt, P., Tee, W.-W., and Reinberg, D. (2013). A double take on bivalent promoters. Genes Dev. *27*, 1318–1338.

Walter, D., Hoffmann, S., Komseli, E.-S., Rappsilber, J., Gorgoulis, V., and Sørensen, C.S. (2016). SCF(Cyclin F)-dependent degradation of CDC6 suppresses DNA re-replication. Nat. Commun. *7*, 10530.

Weinreich, M., and Stillman, B. (1999). Cdc7p-Dbf4p kinase binds to chromatin during S phase and is regulated by both the APC and the RAD53 checkpoint pathway. EMBO J. *18*, 5334–5346.

Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *482*, 221–225.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J. V., Cloos, P.A.C., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343–348.

Wilson, B.G., and Roberts, C.W.M. (2011). SWI/SNF nucleosome remodellers and cancer. Nat. Rev. Cancer *11*, 481–492.

Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nat. Methods *6*, 359–362.

Wohlschlegel, J.A., Dwyer, B.T., Dhar, S.K., Cvetic, C., Walter, J.C., and Dutta, A. (2000). Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. Science *290*, 2309–2312.

Wong, P.G., Winter, S.L., Zaika, E., Cao, T. V, Oguz, U., Koomen, J.M., Hamlin, J.L., and Alexandrow, M.G. (2011). Cdc45 limits replicon usage from a low density of preRCs in mammalian cells. PLoS One *6*, e17533.

Woodward, A.M., Göhler, T., Luciani, M.G., Oehlmann, M., Ge, X., Gartner, A., Jackson, D.A., and Blow, J.J. (2006). Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress. J. Cell Biol. *173*, 673–683.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389–393.

Wutz, G., Varnai, C., Nagasaka, K., Cisneros, D.A., Stocsits, R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, J.M., et al. (2017). CTCF, WAPL and PDS5 proteins control the formation of TADs and loops by cohesin. bioRxiv.

Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R. a, Bell, S.P., and Aparicio, O.M. (2001). Genome-wide distribution of ORC and MCM proteins in S.

cerevisiae: high-resolution mapping of replication origins. Science *294*, 2357–2360.

Xiao, S., Xie, D., Cao, X., Yu, P., Xing, X., Chen, C.-C., Musselman, M., Xie, M., West, F.D., Lewin, H.A., et al. (2012). Comparative Epigenomic Annotation of Regulatory DNA. Cell *149*, 1381–1392.

Yamada, M., Sato, N., Taniyama, C., Ohtani, K., Arai, K.I., and Masai, H. (2002). A 63-base pair DNA segment containing an Sp1 site but not a canonical E2F site can confer growth-dependent and E2F-mediated transcriptional stimulation of the human ASK gene encoding the regulatory subunit for human Cdc7-related kinase. J. Biol. Chem. *277*, 27668–27681.

Yang, X.-J., and Seto, E. (2008). The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. Nat. Rev. Mol. Cell Biol. *9*, 206–218.

Yeeles, J.T.P., Deegan, T.D., Janska, A., Early, A., and Diffley, J.F.X. (2015). Regulated eukaryotic DNA replication origin firing with purified proteins. Nature *519*, 431–435.

Zegerman, P., and Diffley, J.F.X. (2007). Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. Nature *445*, 281–285.

Zhang, T., Cooper, S., and Brockdorff, N. (2015). The interplay of histone modifications - writers that read. EMBO Rep. *16*, 1467–1481.

Zhang, Y., Wong, C.-H., Birnbaum, R.Y., Li, G., Favaro, R., Ngan, C.Y., Lim, J., Tai, E., Poh, H.M., Wong, E., et al. (2013). Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature *504*, 306–310.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nat. Rev. Genet. *12*, 7–18.

Zhu, W., and DePamphilis, M.L. (2009). Selective killing of cancer cells by suppression of geminin activity. Cancer Res. *69*, 4870–4877.

Zhu, W., Chen, Y., and Dutta, A. (2004). Rereplication by depletion of geminin is seen regardless of p53 status and activates a G2/M checkpoint. Mol. Cell. Biol. *24*, 7140–7150.

Annex

Annex

During the Doctoral Thesis I also participated in the following study:

Ahuja, A.K., Jodkowska, K., Teloni, F., Bizard, A.H., Zellweger, R., Herrador, R., Ortega, S., Hickson, I.D., Altmeyer, M., Mendez, J., Lopes, M., 2016. A short G1 phase imposes constitutive replication stress and fork remodelling in mouse embryonic stem cells. Nature Communications 7, 10660. doi:10.1038/ncomms10660