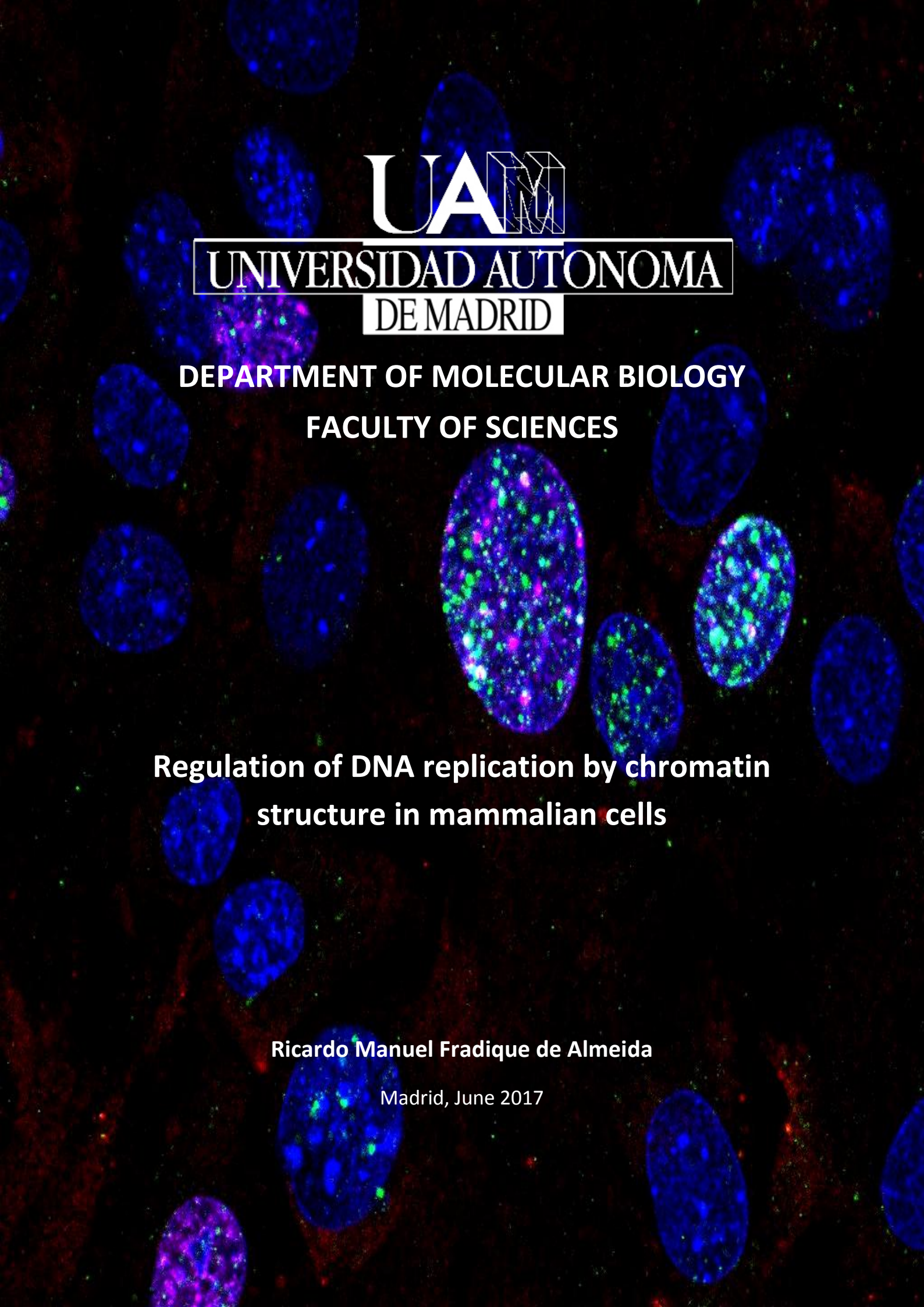


UNIVERSIDAD AUTONOMA DE MADRID

DEPARTMENT OF MOLECULAR BIOLOGY

FACULTY OF SCIENCES

Regulation of DNA replication by chromatin structure in mammalian cells

Ricardo Manuel Fradique de Almeida

Madrid, June 2017

UNIVERSIDAD AUTÓNOMA DE MADRID

DEPARTMENT OF MOLECULAR BIOLOGY

FACULTY OF SCIENCES

Regulation of DNA replication by chromatin structure in mammalian cells

Ricardo Manuel Fradique de Almeida

Graduate in Genetics and Biotechnology

Thesis Director: Dr. María Gómez Vicentefranqueira

Centro de Biología Molecular "Severo Ochoa" (UAM/CSIC)

Madrid, June 2017

This Thesis was supported by a fellowship from the Portuguese Foundation for Science and Technology (FCT - SFRH/BD/81027/2011)

Science knows no country, because knowledge belongs to humanity, and is the torch which illuminates the world.

(Louis Pasteur)

This work would have been impossible without the support and collaboration of many people that were directly or indirectly involved in it. As such I would like to personally thank those who in one way or the other allowed me to reach this point and complete my PhD Thesis:

To María Gómez, director of this thesis, for accepting me not only as a PhD student but also as Master student prior to that. She gave me the opportunity to start a scientific career and her tireless support over the years gave me the motivation necessary to complete my PhD thesis. She always had the door open to discuss any experiment and to solve any doubt and as a PhD student I learned a lot from her. As a mentor she will always be an inspiration for the future.

To Luis Blanco, tutor of this thesis, who was also always available to exchange ideas and give valuable advice.

To all our collaborators over the years, especially Juan Mendez, Sara and Karolina at CNIO how taught me how to perform the DNA fiber technique and in this way contributed immensely to the work that was developed here.

To all the students that passed through the lab over the years, Nicollo, Kenza, Daniel, Marcia and to other friends and colleagues at the CBM and in Madrid, Azmane, Charo, Rafa, Julieta, Pepa, Patricia and Iluminada. Thank you for your companionship.

To Crisanto Gutiérrez and his group, Joana, Bene, Carla, María, Ainhoa, David, Vicky, Sofia Otero, Aitor and all the others, not only for all the labmeetings and scientific discussions that allowed me to have new ideas and improve my work but also for your friendship over this period of more than 5 years.

A very special thanks to all my dear friends, Rodrigo, Gonzalo, José Miguel, Cristina, Laura, Zaida, Clara and Alba. Thank you for all help and the good times that we've spent at the CBM and also outside of it and for being the best "fellowship" that one could ask for in a journey like this. I'm really glad I could come to another country and meat such fantastic people whose friendship I will always cherish.

Quero também agradecer a todo o grupo de amigos, alguns já conhecidos e outro "adquiridos" aqui em Madrid, Fonte, Carol, Vivi, Catarina, Ana Gomes, Diana, Nuna e Bruno pela vossa companhia e amizade que permitiram sair do quotidiano do dia-a-dia e aproveitar Madrid de outra maneira, quer fosse uma simples saída para beber um copo ou durante uma jantarada antes de ir festejar o Europeu para Cibeles ou a vitória na Eurovisão!

A todos os meus amigos de Ovar, principalmente ao Sérgio, Nuno e Leandro pelos bons momentos passados sempre que decidia voltar a “terrinha”.

À Rainha, à Maria Miguel, ao Mirhan, à Fabiana, ao Luís, à Mariana e a pequena Luaninha agradeço a hospitalidade que me concederam ao longo dos anos sempre que decidia ir dar um saltinho até outras paragens.

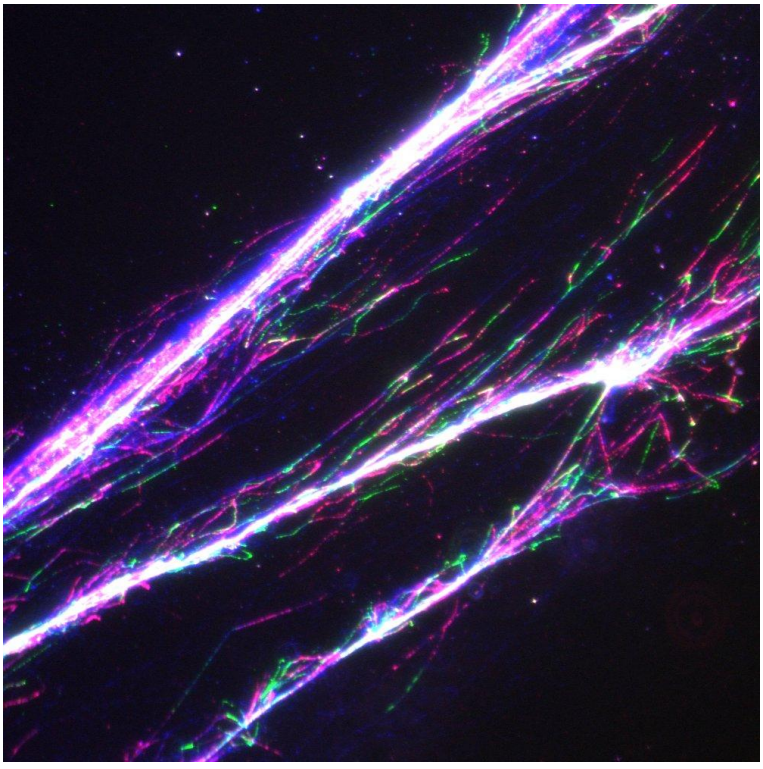
Agradeço também a todos os que por cá passaram a visitar sempre que puderam; Luana, Ruben, Mel, Telomo, Raquel, Sofia Pires, Joaninha, entre outros…

Um agradecimento muito especial a duas das melhores pessoas que conheci aqui em Madrid e ao longo da minha vida, Filipe e Sofia Costa. A vossa grande amizade foi um dos pilares que em muitas ocasiões me manteve de pé e que me permitiu chegar ao final desta tese com alguma sanidade mental. Guardarei sempre boas recordações de todas as nossas viagens juntos, das que fizemos e das que ainda estão para vir… Amigos como vocês só se encontram uma vez na vida e por isso já sabem que podem contar comigo para o que der e vier! P.S.: Obrigado por todos os postais!

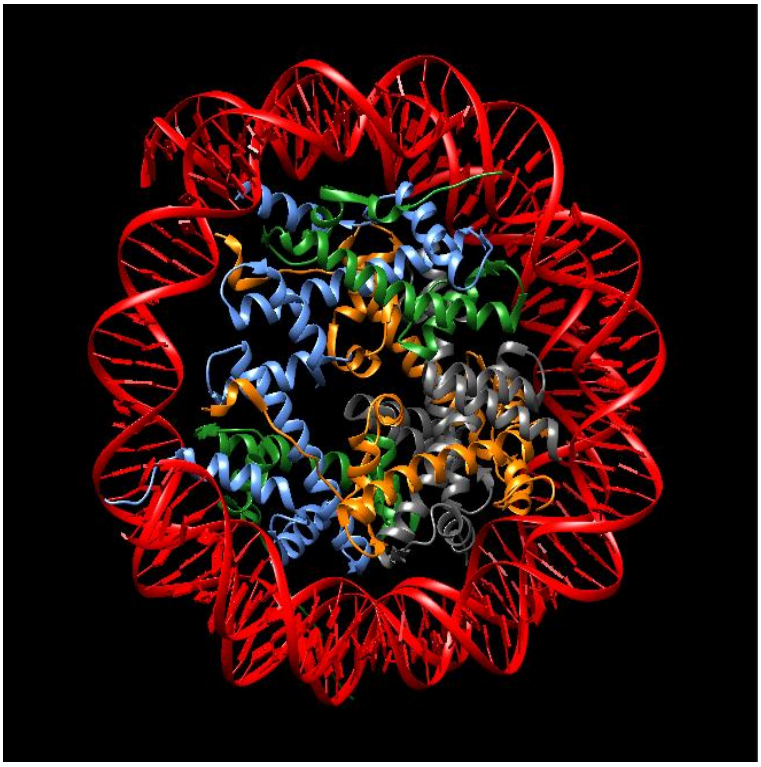
Se a amizade é um dos grandes pilares da minha vida a família é outro. Estas são as pessoas com quem sempre posso contar e que estão sempre lá, independentemente da situação. Ao meu pai Rui e a minha mãe Ilda agradeço o carinho, apoio incondicional e o facto de sempre me transmitirem coragem em todos os momentos da minha na vida e para tentar sempre ser a melhor pessoa que posso. À minha avó Maria por ser uma segunda mãe para mim e de ter estado sempre ao meu lado desde de que era um simples garoto a brincar na sala com legos até hoje. Ao meu tio Carlos por me ensinar que se pode ser sempre uma pessoa bem-disposta independentemente dos obstáculos que a vida nos possa trazer. À Cristina ao Fernando e ao João por serem os melhores sogros e cunhado que eu poderia alguma vez pedir; De igual maneira vocês estiveram sempre aqui para me apoiar em todos os momentos e são já uma parte fundamental e indispensável da minha família com quem fico feliz de compartir a minha vida. Agradeço também o carinho, preocupação e apoio de todos os outros membros da família, avó Ilda, avó Zé (ausente mas não esquecido), e Avó Nair e a toda a família de Mangualde, a qual não posso nomear aqui por inteiro, porque se não já não ia ter espaço para a tese…

Finalmente, nem sei como agradecer a melhor pessoa que conheço e que está sempre do meu lado em tudo o que faço e que é no fundo parte de mim também… A ti Sofia, poderia agradecer mil e uma coisas, o facto de me amares e sempre me apoiares, mesmo quando sou teimoso, o facto de te preocupares comigo a toda a hora, o facto de me teres ajudado ao longo de todos os momentos de stress e de tensão, de sempre estares disposta a sorrir para mim em qualquer momento, mas nem sei bem como posso agradecer tudo isso… Só espero poder ter a sorte de continuar a partilhar a minha vida contigo e de aproveitar todos os bons momentos que ainda estão para vir daqui para a frente… Por isso agradeço-te por seres a minha pessoa, aquela com quem eu posso sempre contar e quero que saibas que tudo isto seria impossível sem ti…

Table of contents



Abbreviations

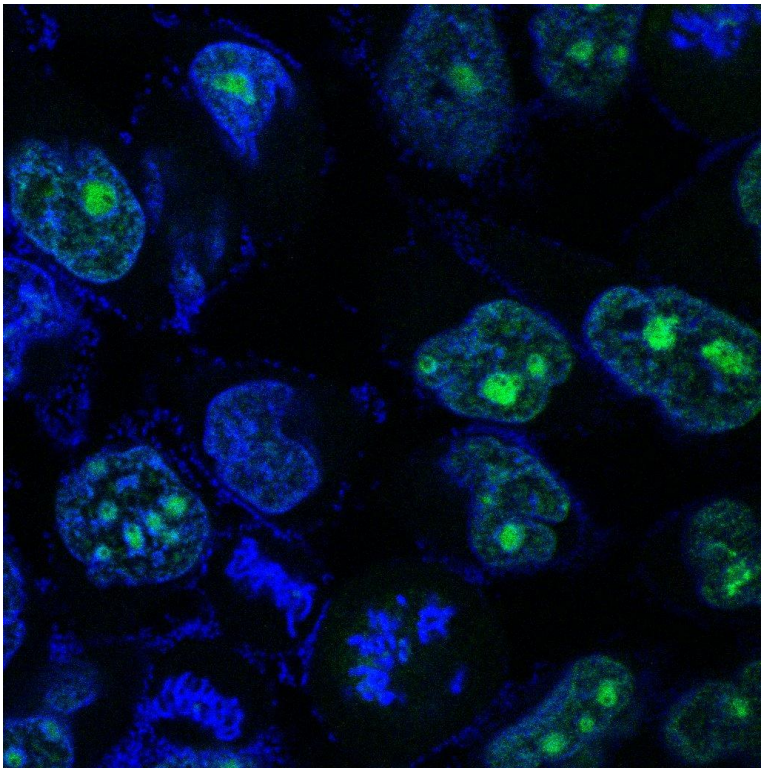


Abbreviation	Description
ACS	ARS Consensus Sequence
ARS	Autonomous Replicating Sequences
BAH	Bromo-Adjacent Homology
BSA	Bovine serum albumin
bp	Base pair
ºC	Degree celsius
Cdc6	Cell division cycle 6
CDK	Cyclin Dependent Kinase
CdT1	Cdc 10 dependent Transcription
CFSs	Common Fragile Sites
CGIs	CpG islands
CHD	Chromodomain-Helicase-DNA binding
ChIP	Chromatin immunoprecipitation
CldU	Chlorodeoxyuridine
CMG	Cdc45/MCM2-7/GINS
DAPI	4',6-diamidino-2-phenylindole
dATP	Deoxyadenosine triphosphate
DDK	Dbf4-dependent protein Kinase
DHFR	Dihydrofolate Reductase
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
dNTPs	Deoxynucleotide triphosphates
Dox	Doxicyclin
DRB	5,6-Dichloro -1-β-D-ribofuranosylbenzimidazole
DSBs	Double Strand Breaks
EDTA	Ethylenediaminetetraacetic acid
EdU	Ethynyl Deoxyuridine
EU	Ethynyl Uridine
FACT	Facilitates Chromatin Transcription
FBS	Fetal Bovine Serum
FRAP	Fluorescence recovery after photobleaching

FriP	Fraction of mapped Reads in Peak regions
FPCM	Fundación Parque Científico de Madrid
G4	G-quadruplexe
HMG	High Mobility Group
HMGA	HMG-AT-hook family
HMGB	HMG Box
HMGN	HMG-nucleosome binding
H3K4me3	Histone H3 Lysine 4 trimethylation
H3K27me3	Histone H3 Lysine 27 trimethylation
H4K12	Histone H4 Lysine 12
H4K20me2	Histone 4 Lysine 20 dimethylation
IdU	Iododeoxyuridine
IgG	Immunoglobulin G
INO80	Inositol requiring 80
IOD	Inter-Origin Distance
ISWI	Imitation Switch
Kb	Kilobase
KO	Knockout
MCM	Minichromosome maintenance
MEFs	Mouse embryonic fibroblasts
mES	Mouse embryonic stem cells
mg	milligram
mL	millilitre
mM	millimolar
MNase	Micrococcal Nuclease
mRNA	messenger RNA
NCP	Nucleosome Core Particle
NDR	Nucleosome Depleted Region
NFR	Nucleosome Free Region
ng	nanogram
NuSA	Nucleosome Scanning Assay
ORC	Origin Recognition Complex
ORIs	Origins of Replication Initiation
PBS	Phosphate-Buffered Saline

PCR	Polymerase Chain Reaction
pH	potential of Hydrogen
PI	Propidium iodide
pmol	picomol
PNK	T4 Polynucleotide Kinase
pre-IC	pre-Initiation Complex
preRC	Pre-Replication Complexe
P-TEFb	Positive Transcription Elongation Factor
PTMs	Post-Translational Modifications
qPCR	quantitative real time polymerase chain reaction
rcf	relative centrifugal force
R-Loops	Hybrid DNA:RNA structures
RNA	Ribonucleic acid
RNAPII	RNA Polymerase II
RT	Room Temperature
RT-qPCR	Reverse-Transcriptase quantitative real time polymerase chain reaction
SD	Standard Deviation
SLBP	Stem Loop binding protein
SNS-seq	SNS sequencing
SNS	Short Nascent Strands
SWI/SNF	Switch/sucrose non-Fermentable
TE	Tris- Ethylenediamine tetraacetic acid
TF	Transcription Factors
TSS	Transcription Start Site
TTS	Transcription termination site
V(D)J	Variable, diverse and joining genes
WT	Wild Type
%	Percentage
µg	microgram
µm	micromolar

Abstract – Resumen



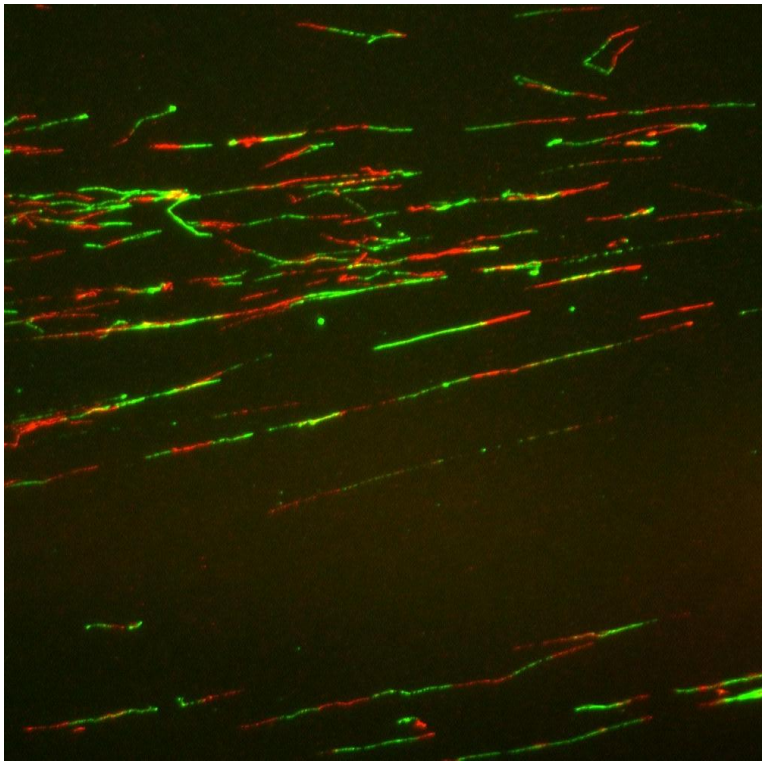
Every time a cell divides the genetic information contained within must be accurately duplicated. This is a feature that universally connects every living organism from the simplest bacteria to the most complex eukaryote, showing its fundamental importance as one of the basic pillars of life on our planet. Since the discovery of the molecular structure of the DNA much has been learned in regards to how different genomes duplicate, especially in what concerns the basic machinery that is involved in this process. In eukaryotic cells, DNA replication occurs during the S-phase of the cell cycle through the activity of hundreds to thousands of replication origins (ORIs) distributed along their large genomes, in a context of a tightly packaged chromatin structure. Even though the molecular mechanism by which ORIs are activated is highly conserved, ORIs do not seem to display any DNA sequence specificity in complex genomes. Nonetheless, they are not randomly distributed, and growing evidence suggests that DNA accessibility could be a major determinant of ORI specification.

The objective of this work was testing the hypothesis that, in mammalian cells, chromatin conformation regulates both the location of the ORIs as well as their activity. To test this hypothesis we first conducted a high-resolution analysis of DNA synthesis start sites and nucleosome architecture at efficient mammalian ORIs. We found that mammalian origins are highly variable in nucleosome conformation and initiation patterns, and that replication initiation profiles mirror nucleosome organization. Second, we performed a functional analysis of the impact of altered chromatin configurations in the definition of the ORIs and in the kinetics of replication elongation. We found that loss of chromatin compaction in H1-depleted cells massively disrupts the replication initiation patterns, triggering the accumulation of stalled forks and DNA damage as a consequence of transcription-replication conflicts. On the contrary, reductions in nucleosome occupancy due to the lack of HMGB1 cause faster fork progression without impacting the initiation landscape or fork stability. Thus, perturbations in the integrity of the chromatin template elicit a range of responses in the dynamics of DNA replication and transcription, with different consequences on replicative stress. These findings have broad implications for our understanding of how defects in chromatin structure, such as those occurring during cellular aging or in some developmental disorders and cancer, contribute to genomic instability.

Siempre que una célula se divide, su información genética debe ser duplicada con precisión. Esta característica une de manera universal a todos los organismos vivos desde la bacteria más sencilla hasta el eucariota más complejo, demostrando su papel fundamental como uno de los pilares básicos de la vida. Desde el descubrimiento de la estructura del ADN se ha aprendido mucho sobre cómo se replican distintos genomas, en especial, sobre la maquinaria básica implicada en este proceso. En células eucariotas la replicación del ADN tiene lugar durante la fase S del ciclo celular, gracias a la actividad de cientos de miles de orígenes de replicación (ORIs) distribuidos a lo largo de sus genomas, en un contexto de estructura de la cromatina altamente compactado. A pesar de que el mecanismo molecular de activación de los ORIs está altamente conservado en eucariotas, los ORIs no muestran especificidad de secuencia en genomas complejos. Sin embargo, no se encuentran distribuidos al azar y, además, numerosas evidencias sugieren que la accesibilidad del ADN puede ser un importante determinante de la especificación de los ORIs.

El objetivo de este trabajo ha sido comprobar la hipótesis de que, en células de mamífero, la conformación de la cromatina regula tanto la especificación de los ORIs como su actividad. Para probar nuestra hipótesis, hemos realizado en primer lugar un análisis de alta resolución de los sitios de inicio de síntesis de ADN y de arquitectura nucleosomal en ORIs eficientes. Hemos comprobado que estos ORIs son altamente variables tanto en su conformación nucleosómica como en los patrones de iniciación de la replicación del DNA y, además, que los perfiles de iniciación de replicación reflejan la organización nucleosómica. A continuación, hemos realizado un análisis funcional del impacto de las alteraciones en la configuración de la cromatina sobre la definición de los ORIs y la cinética de la elongación de la replicación. Hemos descubierto que la pérdida de compactación de la cromatina que ocurre en células deplecionadas en la histona H1 desestabiliza los patrones de iniciación de la replicación, originando la acumulación de horquillas bloqueadas y de daño en el ADN, como consecuencia de conflictos entre la replicación y la transcripción. Por otra parte, la disminución en la ocupación nucleosómica debida a la falta de la proteína HMGB1 tiene como consecuencia un incremento en la velocidad de las horquillas de replicación, sin impacto en su estabilidad y sin alterar los patrones de iniciación de la replicación. Por tanto, hemos encontrado que perturbaciones en la integridad de la cromatina generan un abanico de respuestas en la dinámica tanto de la replicación del ADN como de la transcripción, con distintas consecuencias en estrés replicativo. Estos descubrimientos tienen importantes implicaciones en situaciones fisiológicas o patológicas en las que se producen defectos en la estructura de la cromatina, como ocurre durante el envejecimiento celular o en algunas enfermedades de desarrollo o en el cáncer, en las que se genera inestabilidad genómica.

Introduction



1. Genome Duplication

1.1. The initiation of genome replication in Different Organisms

The accurate duplication of the genetic information is fundamental for the survival and propagation of all living organisms. From the simple bacteria to much more complex higher eukaryotes, faithful inheritance of the genome at each cell division must be achieved to avoid the transmission of mistakes that can lead to potentially dire consequences. The DNA synthesis starts at specific sites called origins of replication initiation (ORIs). The number of ORIs per genome can vary according to the species and cell types and they are recognized through the binding of a heterohexameric complex called the Origin Recognition Complex (ORC), which is well conserved from yeast to humans (Bell, 2002; Bell and Dutta, 2002). Organisms with small genomes like bacteria only need a single ORI in order to replicate their genomes. In *E. coli*, for example, replication takes an average of 30 minutes to be completed after starting from a single specific sequence of DNA with an estimated replication fork speed of 60 kb/min (Jacob *et al*., 1963). Compared with bacteria and other simpler life forms, eukaryotes have significantly larger genomes and therefore need more than one single ORI in order to duplicate their respective DNA content within a few hours. Unicellular eukaryotes, like the yeast *Saccharomyces cerevisiae*, possess several autonomous replicating sequences (ARSs) along their chromosomes, and in those organisms ORC is capable to recognize and bind to these elements. The ARS consensus sequence (ACS) encompasses a T-rich motif which is a pre-requisite for ORC binding (Rao and Stillman, 1995; Xu *et al*., 2006; Chang *et al*., 2008). Although the ACS is capable of functioning as a replicator, upon insertion into a plasmid it was shown that this element by itself is not sufficient to predict or activate ORIs (Stinchcomb *et al*., 1979; Palzkill and Newlon, 1988). *S. cerevisiae* is estimated to have well over 10000 ACS elements but less that 4% of them are actually used as ORIs (Nieduszynski *et al*., 2006). On the other hand, *Schizosaccharomyces pombe* also contains ARS elements but, contrary to its counterpart *S. cerevisiae,* they are not determined by a specific consensus sequence, being comprised of other features like poly-dA/dT tracks and AT-rich islands (Okuno *et al*., 1999; Segurado *et al*., 2003; Dai *et al*., 2005; Heichinger *et al*., 2006). In *S. cerevisiae* there is also is evidence that nucleosome positioning is an important factor for ORI activation. ORIs seem to be flanked by two well positioned histone octamers and ORC is essential to establish this pattern of nucleosomes bordering the ACS elements (Eaton *et al*., 2010).

The human and mouse genomes are several times bigger than those of unicellular eukaryotes, so in order to duplicate their DNA content inside the space of a few hours these organisms need to activate more than just a few hundred ORIs. Those larger

genomes activate several thousand ORIs each S-phase of the cell cycle. Besides ORI number, another fundamental difference is the fact that in higher eukaryotes ORIs do not seem to be defined by DNA sequence and various studies demonstrated that metazoan ORC doesn't display affinity for any consensus sequence *in vitro* (Bell and Dutta, 2002; Vashee *et al*., 2003). This is surprising because the replication machinery, as well as the steps required for ORC assembly and ORI activation, seem to be highly conserved among eukaryotic species (Bell and Dutta, 2002; Vashee *et al*., 2003; Remus *et al*., 2004). Nonetheless, despite the apparent lack of preference for a consensus sequence ORIs in higher eukaryotes aren't distributed in an arbitrary fashion, which is consistent with the notion that these genomes are not randomly organized (Hurst *et al*., 2002; Hurst *et al*., 2004; Batada and Hurst, 2007; Necsulea *et al*., 2009). Different regions in the chromosomes of eukaryotes seem to have a more favorable environment for replication initiation and even in yeast where DNA sequence plays a key role in ORI establishment, telomeric regions seem to replicate late when compared with centromeric proximal regions that replicate in the earlier stages of S phase (MacAlpine and Bell, 2005; Necsulea *et al*., 2009). It seems that factors intimately associated to the transcription process have a strong connection with the temporal and spatial organization of replication. Those factors include local gene density, presence of transcription start sites (TSS) of CpG island promoters and active transcription at certain *loci* (Rocha, 2004; Nieduszynski *et al*., 2006; Cadoret *et al*., 2008; Sequeira-Mendes *et al*., 2009; Cayrou *et al*., 2011; Besnard *et al*., 2012). Additionally, genomic regions that encompass a less condensed chromatin state (euchromatic state), like gene dense regions, might facilitate the accessibility of replication factors to the DNA molecule and in this way promote an easier assembly of the replication complexes (Vashee *et al*., 2003; Schaarschmidt *et al*., 2004; Cadoret *et al*., 2008; Yin *et al.,* 2009; Lubelsky *et al*., 2010). Furthermore, some non-canonical DNA structures, like G-quadruplexes (G4s) have also been shown to be enriched at ORI regions in mammalian species (Cayrou *et al*., 2012; Besnard *et al* 2012), and it is worth mentioning that these structures seem to be preferentially present at specific locations of the genome like telomeric regions and, most importantly, transcription regulatory regions (Rhodes and Lipps, 2015). In addition, ORC1 affinity for G4s is higher when these structures are present on RNA or single-stranded DNA, suggesting that regions with high tendency to accumulate hybrid RNA:DNA structures (R-loops), like active CpG Island promoters, may be more prone to ORC1 binding (Ginno *et al*., 2012, 2013; Hoshina *et al*., 2013).

Together all these data points to a complex scenario where ORI specification in higher eukaryotes seem to be dependent of a combinatorial set of factors. This flexibility

likely allows cells to adapt to different stages of development and to cope with external factors impacting this process, facilitating a faithful duplication of the genomic information (Sequeira-Mendes and Gómez, 2012; Fragkos *et al*., 2015).

1.2. ASSEMBLY AND ACTIVATION OF REPLICATION ORIGINS

The duplication of the DNA molecule must occur only once per S-phase in order to avoid genomic abnormalities that can range from rereplication of small portions of the DNA sequence to whole chromosomal duplications which can in turn lead to much more severe consequences. In order to avoid replication-related genomic instability, the molecular mechanisms involved in controlling this process at the level of ORI establishment and activation are very tightly regulated (Mechali, 2001; Diffley, 2004; Gilbert, 2004; Cvetic and Walter, 2005; Aladjem *et al*, 2006; Arias and Walter, 2007; Rampakakis *et al*., 2009). During the G1 phase of the eukaryotic cell cycle ORC binds to several chromosomal locations and at this stage this complex is capable of recruiting several important factors like Cdt1 and Cdc6 and the MCM helicases, which are the main components of the prereplication complexes (pre-RC) (Bell, 2002; Diffley, 2004; Stillman, 2005). Then, during S phase the pre-RCs are activated through several phosphorilation events, what mediates the recruitment of additional factors promoting ORI licensing and finally leading the MCM complexes to start unwinding the DNA molecule and the replication process to unfold (Mechali, 2010) (Figure 1). From this point onwards the MCM helicases are not able to re-associate with ORIs, ensuring that the licensing process is only triggered once every cell cycle, thus preventing re-replication events (Symeonidou *et al*., 2012). Nevertheless, although licensed ORIs are given a “green light” to start replication it seems that many of them are still not used in an efficient manner and some even stay dormant during S phase (Blow and Ge, 2008; Ibarra *et al*., 2008).

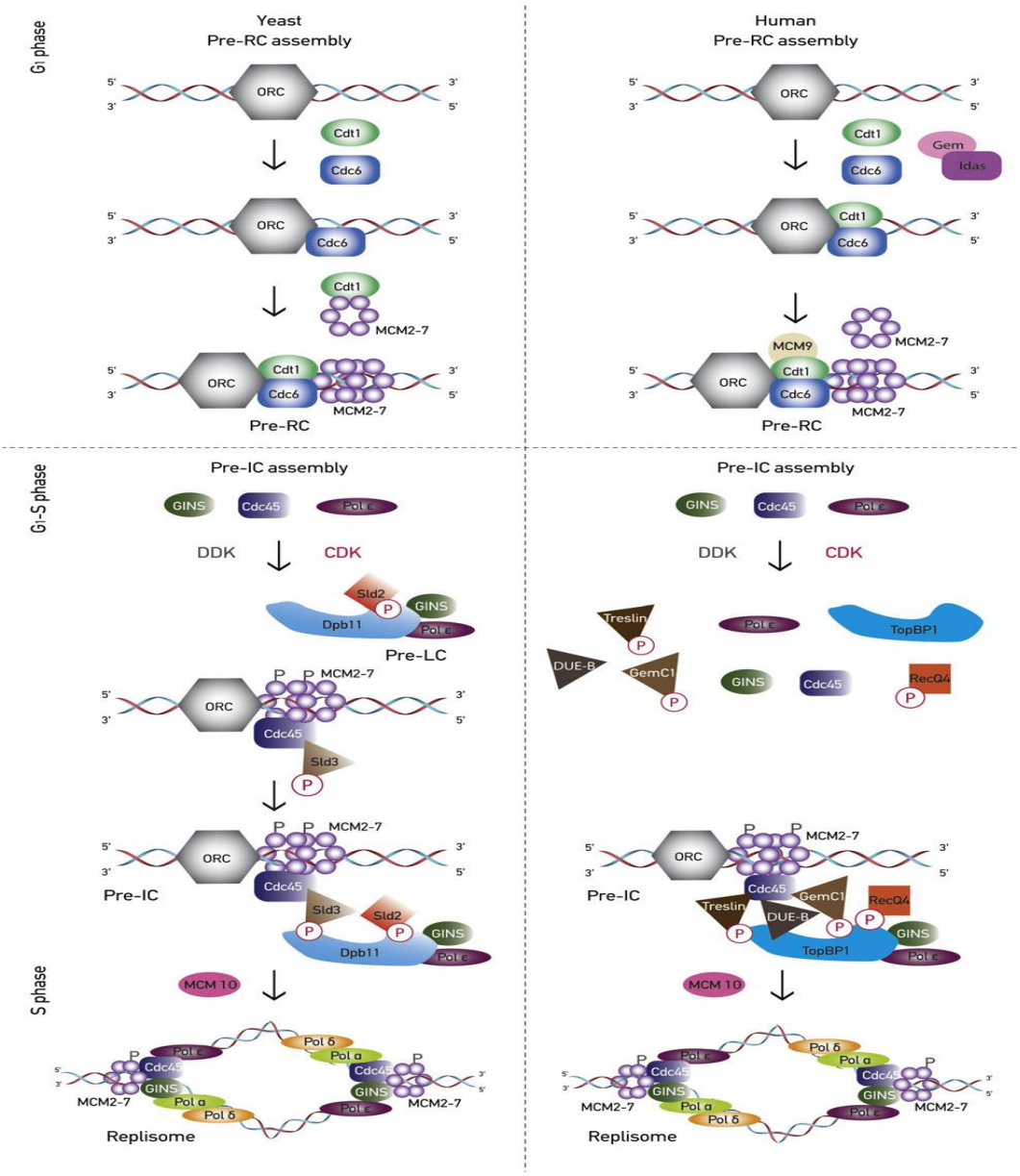


Figure 1: Replisome assembly in budding yeast (left panel) and human cells (right panel). Upon ORC binding, several other factors like Cdt1, Cdc6 and MCM2-7 helicases are recruited to the DNA in order to form the pre-RC. Re-licensing in higher eukaryotes is prevented due to the negative regulation of Geminin. During the G1-S phase transition, CDK and DDK are responsible for the activation of the pre-RCs through the recruitment of additional factors, resulting in the formation of the pre-initiation complexes (pre-ICs). The replisome is finally completed upon the assembly of the active Cdc45/MCM2-7/GINS (CMG) helicase at the pre-ICs and the recruitment of all replicative polymerases (adapted from Symeonidou *et al.*, 2012).

2. The importance of Chromatin for genome regulation and integrity

2.1. Structure of the Chromatin fiber

The large genomes of eukaryotic organisms (encompassing more than 2 meters of DNA in the case of human cells) have to fit inside the space of a small nucleus that isn't more than a few µm in size (Pollard and Earnshaw, 2002). This achievement is possible

because in eukaryotic species the DNA double helix interacts and binds to many proteins to form a chromatin fiber. Chromatin has several levels of organization and it generates the required level of compaction in order to fit the entire genome inside the cell nucleus (Pollard and Earnshaw, 2002) (Figure 2).

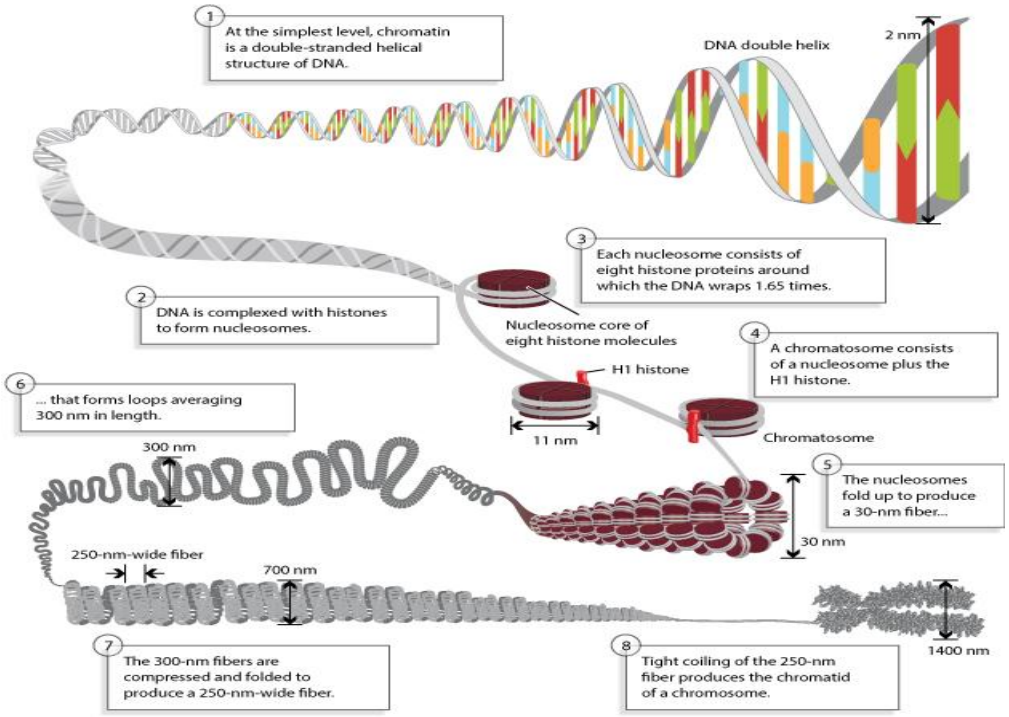


Figure 2: Structure and organization of the chromatin fiber. The DNA molecule interacts with the eight core histones and wraps around them giving raise to the core nucleosome particle. These histone octamers are connected by stretches of DNA known as linker DNA, and the linker histone H1 is frequently present at the DNA entry/exit point of the nucleosome particle. The resulting particle containing H1 is commonly known as the chromatosome. These structures that resemble "beads on a string" are then folded into a fiber-like structure with a 30nm diameter. Further condensation of those fibers occurs to generate other higher-order chromatin structures (adapted from Annunziato, 2008).

The state of chromatin and the global and local levels of condensation of this element are of vital importance for the regulation of all genomic processes like the activation or repression of transcription, splicing events, and DNA replication, among many others. Cell type differentiation and specificity are achieved through modifications in chromatin organization which establishes different levels of accessibility of regulatory elements to the DNA molecule, ultimately changing the cellular transcriptional and replication programs (Vogelauer *et al*., 2002; Donaldson, 2005; Li *et al*., 2005; Knott *et al*., 2009; Hnilicová and Staněk, 2011; Gómez-Acuña *et al*., 2013; Petty and Pillus, 2013). Changes in chromatin conformation are achieved through the action of chromatin remodelers and histone modifying complexes such as histone acetyltransferases, deacetylases, methyltransferases, etc (DesJarlais and Tummino, 2016). Chromatin remodelers can be recruited to gene promoter regions and transform the local structure of

the chromatin by shifting nucleosome positioning and changing the local environment around the DNA molecule thereby modulating the regulation of specific genes (Jiang and Pugh, 2009; Petty and Pillus, 2013).

Nucleosomes constitute the core unit of chromatin and, as such, play a pivotal role in regulating the many mechanisms that require accessibility to the genomic information. The combined action of individual nucleosomes makes it possible to condense the DNA more than 10^4 times due to their highly basic nature which confers them a high affinity for negatively charged molecules, such as nucleic acids (Kornberg, 1974; Allis, 2007). The core nucleosome particle is composed by eight canonical histones H2A, H2B, H3 and H4, which appear in duplicate (Kornberg, 1974) (Figure 3). About 147 bps of DNA are directly bound to the core histone octamer with the DNA molecule completing approximately 1,7 turns around it (Kornberg, 1974; Luger *et al*., 1997). At the dyad axis point, where DNA enters and exits the nucleosome core another histone can be present, the linker histone H1, which interacts with an additional 20 bps of the DNA double helix (Simpson, 1978). In addition to its importance in controlling the chromatin condensation level, histone H1 has a fundamental role in keeping nucleosomes in place and avoiding the sliding of this particle along the DNA (Oudet *et al*., 1975; Li *et al*., 2005).

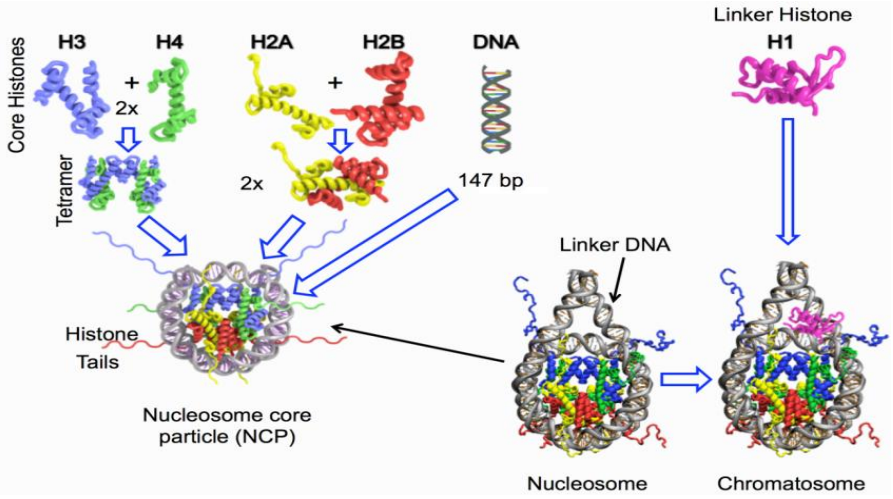


Figure 3: Nucleosome core particle. Two pairs of each histone dimer, H2A-H2B and H3-H4, interact to form the histone octamer core. DNA then wraps around the histones giving raise to the nucleosome core particle (NCP). Histone H1 interacts with the DNA double helix at the dyad axis of the NCP, establishing the chromatosome (adapted from Draizen *et al*., 2016).

Modifications in the structure of the nucleosomes are vital to the architecture of the chromatin and are the basis for the dynamic nature of this element (Jenuwein and Allis, 2001; Margueron *et al*., 2005). Nucleosomes can undergo many changes, mainly on the histone proteins that compose these elements. Histones tails can be subject to acetylation or methylation in different amino acidic residues as well as many other posttranslational

modifications that can ultimately stimulate or repress different chromatin regulatory pathways and lead to changes in the transcriptional program, changing cellular activity and identity (Millar and Grunstein, 2006; Kouzarides, 2007). Histone H3 lysine 4 trimethylation (H3k4me3) is a well known example of an active mark in eukaryotes associated with a general open chromatin structure and very commonly colocalized with promoters of active genes (Santos-Rosa *et al*., 2003; Pray-Grant *et al*., 2005; Sims *et al*., 2005; Wysocka *et al*., 2005). Other types of histone tail methylations, like histone H3 trimethylation of lysine 27 (H3k27me3) are linked with the opposite effect and tend to be found in heterochromatic regions of the genome (Francis *et al*., 2004; Ringrose and Paro, 2004). Domains that carry both the H3k27me3 and the H3k4me3 are known as bivalent and encompass both active and repressive histone marks (Berstein *et al*., 2006). Usually they are associated with CpG island promoters, keeping genes poised for subsequent activation or repression in distinct cell types. This also relates to the fact that half of them coincide with promoters of developmental regulatory genes (Berstein *et al*., 2006; Mikkelsen *et al*., 2007). In addition to posttranslational changes to the properties of nucleosomes, canonical histones can be replaced by histone variants which encompass a slightly different polypeptide sequence. Some particular variants, like H3 variant cenH3, also known as CENP-A in mammals, can be found in specific genomic regions like centromers and plays a major role in kinetochore assembly (Warburton *et al*., 1997; Ouspenski *et al*., 2003; Kamakaka and Biggins, 2005). Others like H2A.Z and H3.3 replace the canonicals H2A and H3 and have a direct effect on the physical properties of nucleosomes, with both variants apparently being implicated in active transcription (Kamakaka and Biggins, 2005; Sarma and Reinberg, 2005; Talbert and Henikoff, 2010). H2A.Z is commonly found in nucleosomes that flank nucleosome free regions (NFRs) at gene transcription start sites (TSSs), in both yeast and human cells, promoting RNA polymerase II (RNAPII) recruitment at those sites (Adam *et al*., 2001; Zlatanova and Thakar, 2008; Hardy *et al*., 2009). Moreover this variant seems to play an important part in processes such as DNA damage signaling and repair, embryonic stem cell differentiation and nucleosome turnover, which might relate to the fact that H2A.Z nucleosomes are more resistant to the binding of the linker histone H1 (Creyghton *et al*., 2008; Zlatanova and Thakar, 2008; Altaf *et al*., 2009; Thakar *et al*., 2009). H3.3 also displays similar features in regards to transcription activation as found in *D. melanogaster* and human cells (Henikoff, 2008). This variant is incorporated into genes upon induction of transcription and is also associated with the elongation phase of this process (Schwartz and Ahmad, 2005; Sutcliffe *et al*., 2009). Furthermore, there is evidence in *D. melanogaster* showing that the turnover rate of H3.3 is faster than that of the canonical

H3, thus contributing to the maintenance of an open and accessible chromatin at transcription initiation sites (Schwartz and Ahmad, 2005).

All these changes are at the very foundation of the dynamic structure that characterizes the chromatin molecule. Remodeling complexes and other factors orchestrate and control the delicate processes responsible for the regulation of the genome and nucleosomes have a direct influence in many of those processes that require accessibility to the DNA molecule and that are of crucial importance for cell fate.

2.2. NUCLEOSOME POSITIONING AND THE MODULATION OF GENOMIC PROCESSES

The 147 bp of DNA that are directly bound to the nucleosome core particle are shielded from most interactions with external enzymes and complexes that have affinity for the double helix molecule. Regions of the genome that are being constantly transcribed encompass a lower level of nucleosome occupancy, essential for allowing the assembly of multiprotein complexes in charge of this processes (Rando and Chang, 2009). It is not simple to decipher and understand nucleosome patterns along the genome of different organisms or even cell types because there are numerous factors that can influence the organization of these nucleoproteins. To help understand this, different concepts were developed over the years such as nucleosome positioning and occupancy. Nucleosome positioning is defined as the probability that a determined nucleosome starts at a given base pair within the genome and nucleosome occupancy refers to the presence or absence of these nucleoproteins over specific genomic sequences (Segal and Widom, 2009; Arya *et al*., 2010). The concept of occupancy differs from positioning in that the former doesn't take in to account where the nucleosome starts as long as the given base pair is covered by nucleosomes (Arya *et al*., 2010). Chromatin remodeler complexes and competition between site-specific DNA-binding proteins such as polymerases and transcription factors (TFs) are very important for modulating nucleosome positioning patterns. The nucleosomes themselves have more affinity for certain DNA sequences, even though the contribution of the sequence seems to vary in the different organisms studied (Ioshikhes *et al*., 1996; Schones *et al*., 2008, González *et al*., 2016). In spite of some specific sequences having more likelihood to bind nucleosomes than others, virtually every DNA sequence can bind and wrap around histone octamers (Sekinger *et al*., 2005; Segal *et al*., 2006; Yuan and Liu, 2008). Therefore, in order to better understand chromatin regulation and dynamics it's essential to study and decipher nucleosomal positioning patterns along the genome of specific cell types or organisms.

2.3. NUCLEOSOME POSITIONING AND TRANSCRIPTION REGULATION

The position of nucleosomes around regulatory elements like TSSs or enhancers is vital for the correct regulation of numerous cellular processes, particularly gene transcription. TFs compete with histone octamers for access to the DNA double helix at promoters and the active RNAPII complexes have the capability to displace nucleosomes ahead of elongation (Lieb, *et al*., 2001; Schwabish and Struhl, 2004; Liu *et al*., 2006). The first maps of nucleosome positioning at a genome-wide scale were developed in the yeast *S. cerevisiae* by hybridizing mononucleosomal DNA resistant to Micrococcal Nuclease (MNase) digestion on microarrays (Bernstein *et al*., 2004; Lee *et al*., 2004). Despite their low resolution these nucleosome maps constituted a significant improvement over prior attempts. One of the main conclusions from these early works was the distinct pattern present at many yeast gene promoters where a NFR was often flanked by two well positioned nucleosomes. Later studies validate those observations, unveiling that the NFR was regularly situated just upstream of the promoters TSS (Lee *et al*., 2007). This seems to be the case in many other eukaryotic species implying that this particular feature is highly significant in regards to transcription regulation (Radman-Livaja and Rando, 2010). When genes are activated or repressed they undergo several epigenetic changes and acquire different nucleosome patterns and this particular placement of nucleosomes around regulatory elements such as TSSs appears to be important to promote the correct binding of key regulatory factors like TFs and promote the accurate assembly of the transcription machinery (Lee *et al*., 2007; Buratowski, 2008). These regions are generally limited by two nucleosomes known as the +1 nucleosome, found upstream, and the -1 nucleosome, found downstream (Peckham *et al*., 2007; Lee *et al*., 2007; Fu *et al*., 2008; Zhang *et al*., 2009; Arya *et al*., 2010) (Figure 4). In *C. elegans* regions located upstream of the TSS show a strict pattern of positioned nucleosomes that is characterized by the presence of two tightly placed histone octamers at the edge of the NFR. Furthermore, the strong positioning of those elements is even more obvious in promoters of genes that are ubiquitously expressed (Valouev *et al*., 2008). In human T cells the +1 nucleosome can be positioned up to 40 base pairs downstream of the TSS in active genes whereas in inactive genes these two elements are only separated by 10bp (Barski *et al*., 2007). Another distinct feature at the TSS of active genes is the strong phasing of the 5 to 10 nucleosomes that are located downstream of the +1 nucleosome (Figure 4) (Arya *et al*., 2010). It seems that RNA polymerase II plays an important role in the establishment of this particular pattern due to its ability to evict or misplace nucleosomes as it moves through the DNA molecule (Izban and Luse, 1992; Studitsky *et al*., 1994; Schwabish and Struhl, 2004; Dion *et al*., 2007; Schones *et al*., 2008; Zhang *et al*., 2009). At the 3' end of a

large number of genes there is also a well-positioned nucleosome (Figure 4). This nucleosome located close to the transcription termination site (TTS) is usually followed by a nucleosome free region (NFR, also known as nucleosome depleted region, NDF) which seems to facilitate the disassembly of the RNA polymerase and all the transcription machinery from the gene (Jiang and Pugh, 2009).

As previously mentioned, nucleosomes can also undergo many structural transformations, such as the replacement of the H2A canonical histone for its variant H2A.Z, or the canonical H3 histone isoform H3.3. This seems to be the case in many regulatory regions like promoters or enhancers, where it was found that the NFR was not in fact free of nucleosomes but rather encompassed a labile nucleosome particle containing both the H2A.Z and H3.3 variants (Jin *et al*, 2009). Besides structural changes, nucleosomes can also suffer positional changes mainly through the action of chromatin remodeler complexes like the SWI/SNIF, ISWI, CHD, INO80 or the High Mobility Group (HMG) family of proteins (Zhang and Wang, 2008; Längst and Manelyte, 2015). The +1 nucleosome for example can be shifted from its original position opening or closing the NFR and thus modulating accessibility to the TSSs (Lomvardas and Thanos, 2001; Koutroubas *et al*., 2008).

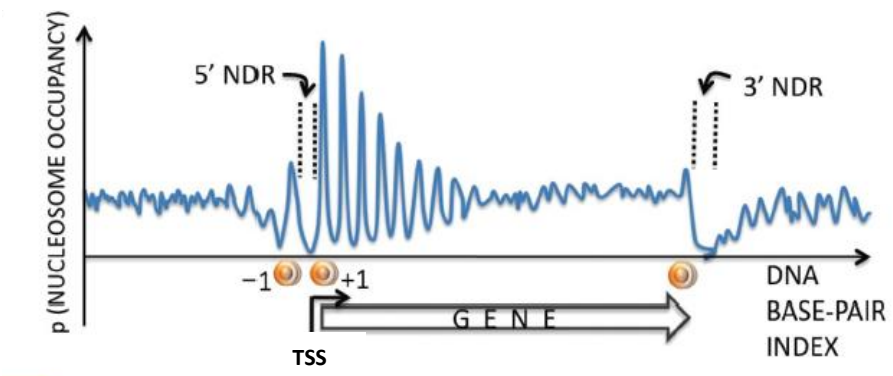


Figure 4: Usual depiction of a nucleosomal occupancy map of actively transcribed genes with the 5' nucleosome depleted region at the promoter and the flanking -1 and +1 nucleosomes. Nucleosomes downstream of the +1 show gradual decrease in their phasing, until the strong positioning patterns are lost. At the 3' transcription termination region there is a well positioned nucleosome, immediately followed by a nucleosome depleted region (adapted from Arya *et al.*, 2010).

2.4. The interplay between transcription, replication and nucleosome positioning

To specify a replication initiation site in the genome ORC must get access the DNA molecule. Origins of replication then require the assembly of a large number of proteins and protein complexes in order to be functional (Figure 1). Hence, the regions to which these complexes bind are expected to be free of nucleosomes or at the very least

encompass nucleosomal particles that are able to shift their positions easily and therefore allow the replication machinery to bind to the DNA (Zhou *et al*., 2005; Field *et al*., 2008; Audit *et al*., 2009). Furthermore, even though ORIs in higher eukaryotes don't show sequence specificity there is evidence that in *Drosophila* ORC has slightly higher affinity for negative supercoiled DNA, implying that processes that generate super-helical tension, such as nucleosome removal, may influence ORI selection (Remus *et al*., 2004). Earlier works in yeast showed that plasmids containing the ARS1 origin needed to have an ACS element free of nucleosomes in order to bind ORC, and that this binding is stabilized by neighboring DNA-protein interactions (Simpson, 1990; Bell *et al*., 1995). Posterior studies focusing on the endogenous ARS1 *locus* supported the theory that the ACS element had to be nucleosome-free for the origin to be functional. However, if the NFR was too large pre-RC formation was inhibited, suggesting that a precise positioning of nucleosomes must take place at the ARS1 *locus* in order to facilitate ORC binding and helicase loading (Lipford and Bell, 2001). In synchronized populations of Chinese hamster cells that contained amplified copies of the dihydrofolate reductase *locus* the ORC and MCM complexes colocalize preferentially at regions of low nucleosome occupancy (Lubelsky *et al.* 2010).

As previously discussed, transcriptional activity has a great influence in chromatin organization, being important in the establishment of a more open and accessible chromatin structure. Over the years many studies have found links between transcription and replication, especially in regards to ORI establishment. In yeast some ORIs contain binding sites for TFs that assist in their activation (Diffley and Stillman, 1989). In other eukaryotic organisms TFs are also capable of influencing the processes of selection and activation of replication origins through the recruitment of histone-modifying and chromatin remodeling complexes and also by interacting with the pre-RC (Cheng *et al*., 1992; Maric *et al*., 2003; Danis *et al*., 2004; Ghosh *et al*., 2004; Minami *et al*., 2006; Sasaki *et al*., 2006; Cayrou *et al*., 2012). One possible explication for these findings involves the changes in transcriptional status occurring during the G1 phase of the cell cycle, which might be able to alter the initial choice of possible origins by influencing the pre-replicative complex assembly. Other possibility is that these changes might also be able to control the activation of the pre-RCs during the S phase. Nonetheless, the specific relationship between those two processes is still no fully understood, especially considering that in order to properly duplicate the genome the replication machinery must avoid at any cost the interference with ongoing transcription, or with any other genomic process that otherwise could impair its progression and consequently generate replication stress, which might eventually lead to genome instability (Helmrich *et al*., 2013). The proper

coordination between those processes is intimately related to the chromatin structure, and changes in the regulation of this element are bound to influence the outcome of both cell division and the patterns of gene expression. The manner in which transcription, replication and chromatin structure interplay is still to this date subject of many studies and analysis, especially due to the complex network of factors that influences them and connects each other. This subject will be the focus of the next section of this introduction chapter.

3. GENOME STABILITY AND ITS RELATIONSHIP WITH CHROMATIN ARCHITECTURE

3.1. IMPORTANCE OF CHROMATIN FOR GENOME INTEGRITY

The epigenetic mechanisms that regulate the state of the chromatin have a fundamental role in maintaining genome stability and ensuring the correct development of an organism. It is therefore not surprising that a large number of diseases are rooted in the epigenetic deregulation of genes due to abnormal changes in the chromatin structure. Those abnormalities can be the result of both acquired and inherited mutations in genes that control modifications in the histone proteins or the DNA molecule (e.g. DNA methylation). Thus, the resulting epigenetic disorders can be of monogenic or multifactorial origin, which can be inherited from the progenitors or acquired by *de novo* mutations in the parental germline or at any stage along the development of the organism (Mirabella *et al*., 2016). The spectrum of epigenetic syndromes is very broad and encompasses numerous neuronal and physical developmental and degenerative disorders, immunodeficiency diseases and many different types of cancer (Cassidy and Schwartz, 1998; Jin and Warren, 2000; Iwase *et al*., 2007; Gibbons *et al*., 2008; Cabianca *et al*., 2012; Tsurusaki *et al*., 2012; Lazo-Gomez *et al*., 2013; Lu *et al*., 2013). In predisposed organisms they can be triggered by many environmental factors like exposure to toxic or harmful compounds, nutrient deficiency and stress (Mirabella *et al*., 2016) (Figure 5). The interaction between chromatin and chromatin binding proteins is of upmost importance to control the dynamics of the different genomic processes and, besides changes in DNA molecule, chromatin remodelers and histones modifications are the main players in the regulation of chromatin stability and epigenetic integrity. So it becomes important to study how specific changes in those factors affect the normal function of the cells and what is the impact for basic genomic processes like transcription, replication and DNA damage repair and recombination.

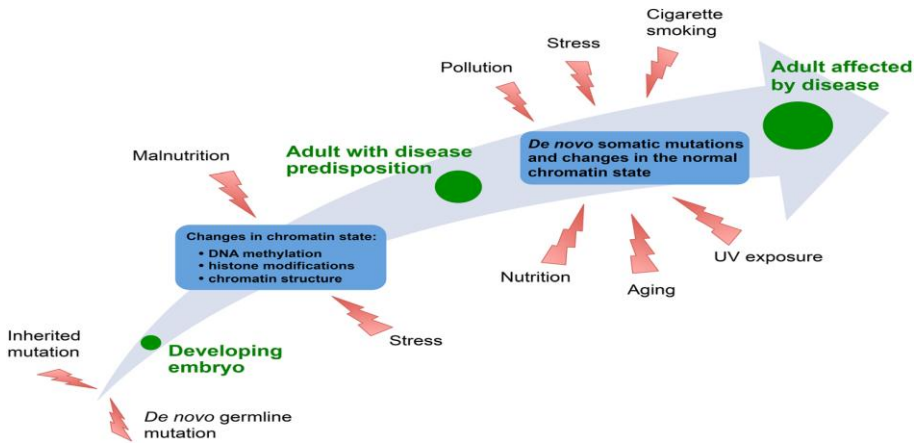


Figure 5: Factors involved in the generation of epigenetic diseases. Diseases caused by chromatin deregulation can arise through environmental stress, either during fetal development or later in life, or by mutations in genes encoding chromatin regulators. These mutations can be heritable or acquired *de novo* leaving the organisms predisposed to disease (adapted from Mirabella *et al.*, 2016)

Two proteins that have become increasingly more studied in the last years due to their importance for chromatin integrity and regulation are the linker histone H1 and the proteins belonging to the High Mobility Group family, mainly the High Mobility Group Box 1 (HMGB1) protein. H1 variants and HMG proteins are expressed in all vertebrates in a ubiquitously manner and are able to bind dynamically to other proteins and factors in the chromatin. They are also the most abundant chromosomal proteins after the core histones themselves (Happel and Doeneck, 2009; Postnikov and Bustin, 2016). H1 and HMG proteins actively affect and change the levels of chromatin condensation and have the ability to regulate many genomic processes in antagonizing ways due to their opposing effect in nucleosome compaction and the fact that they directly compete with each other (Postnikov and Bustin 2016). The specific roles of these proteins are depicted bellow.

3.2. LINKER HISTONE H1 AS A KEY FACTOR FOR CHROMATIN STRUCTURE MAINTENANCE

One of the essential functions of linker histone H1 is to bind to the nucleosome core stabilizing the folding of the nucleofilament into higher order chromatin structures (Figure 2) (Robinson and Rhodes, 2006; Bassett *et al*., 2009; Happel and Doeneck, 2009). It also has a role in regulating transcription by promoting condensation and thereby impairing or at least severely limiting the access of TFs and other complexes to the DNA molecule. However, despite the general notion that H1 acts as a universal repressor of transcription there is increasing evidence that it may regulate transcription at specific levels by interacting with complexes that repress or promote transcriptional activity (Zlatanova *et al*, 2000; Harshman *et al*, 2013). Nonetheless, the specific mechanisms by

which the linker histone H1, its PTMs (post-translational modifications) and related proteins are involved in those processes are still not fully understood mainly due to technical and experimental limitations (Harshman *et al.*, 2013; Roque *et al*., 2016). One factor that also contributes to this apparent complexity in studding and deciphering the full importance and implications of H1 is the fact that this histone is much less conserved between eukaryotes than core histones. Many unicellular species such as yeast encompass only one H1-like gene (Shen *et al.*, 1995; Patterton *et al*., 1998; Ramón *et al*., 2000; Hellauer *et al*., 2001), and in other eukaryotes like *D. melanogaster* there is only one somatic and one germ-line specific H1 variant; dH1 and dBigH1 respectively (Bayona-Feliu *et al*., 2016; Pérez-Montero *et al*., 2016). On the other hand, in mice and humans there are several genes of both somatic and germ-line H1 variants that in many cases might play partially redundant roles (Sirotkin *et al*., 1995; Fan *et al*., 2001; Pérez-Montero *et al*., 2016). Despite this heterogeneity of H1 among different eukaryotes many works in the field of epigenetics underline the extreme importance of this linker histone in the developmental processes of those organisms, as well as the impact on the normal structure and integrity of their genomes. Several years ago it was found that the genome of mouse embryos lacking three of the somatic variants of H1 (H1c, H1d and H1e) encompass only half of the normal levels of this histone, which are not enough to maintain embryonic viability due to the accumulation of a large number of defects (Fan *et al*., 2003). Cell lines derived from those embryos show a reduction in the levels of chromatin compaction and altered expression of a small number of genes, from which imprinted genes represented an enriched category (Fan *et al*., 2005). These cells display also a 75% reduction in H4K12 acetylation, involved in chromatin relaxation, and an average 15 bp reduction in linker DNA size, which might be a mechanism to cope with H1-loss and minimize the effects of genome-wide chromatin decompaction (Fan *et al*., 2005). It seems also that this relaxation in the chromatin structure has an effect on DNA damage signaling. H1 triple knockout (KO) cells appear to be more resistant to DNA damage caused by multiple agents, being able to activate the intra-S and G2/M phase checkpoints more efficiently than normal wild type (WT) mES cells (Murga *et al*., 2007).

All of this evidences points to a scenario where cells have to modify basic parameters in fundamental processes like transcription and replication in order to cope with a drastic change in the structure of the genome and indicates that the linker histone H1 is of upmost importance for the maintenance of this structural integrity, which is so essential to all cellular processes. Furthermore, although there is increasing evidence of the many roles in which H1 is involved, at the moment there´s still not many information on how it directly impacts the kinetics of replication and transcription and what are the

implications of the deregulation of both processes for genome stability in a context of such a drastic change in chromatin structure.

3.3. The role of High Mobility Group Family of proteins in Chromatin dynamics

After the histones, the most prominent components in chromatin are the proteins that belong to the High Mobility Group family. They are divided into three main subtypes, the HMG-AT-hook family (HMGA), the HMG-box family (HMGB) and the HMG-nucleosome binding family (HMGN) (Bustin, 2001). These families are both abundant and ubiquitous along the genome and are also subject to many PTMs that influence their interactions with the DNA molecule and their associated proteins, contributing to modulate the regulation of different genomic processes (Zhang and Wang, 2008). While the HMGN family binds specifically to the nucleosome particle to promote PTMs in the core histones (Lim *et al*., 2004; Postnikov *et al*., 2006; Ueda *et al*., 2006), the other two families, HMGA and HMGB, are characterized as nuclear DNA-binding proteins (Sutrias-Grau *et al*., 1999; Das and Scovell, 2001; Reeves, 2001; Stros *et al*., 2007). Fluorescence recovery after photobleaching (FRAP) experiments in *Arabidopsis* reveal that both HMGA and HMGB proteins are very dynamic inside the nucleus, binding to chromatin only in a transient manner and constantly searching for new target regions (Launholt *et al*., 2006). These two families are known for their role the in regulating transcription activity, with HMGA proteins seeming to be more important in a developmental context as they are more expressed in fast proliferating cells such as embryonic and tumor cells (Fedele *et al*., 1996; Bandiera *et al*., 1998; Chiappetta *et al*., 1998; Reeves, 2001; Sgarra *et al*., 2004), while HMGB proteins operate primarily as architectural remodelers of chromatin structure promoting gene expression mainly due to their opposing role to the linker histone H1 (Sutrias-Grau *et al*., 1999; Das and Scovell, 2001; Stros *et al*., 2007) and their ability to loosen the DNA wrapped around nucleosomes, promoting their sliding (Bonaldi *et al*., 2002; Agresti and Bianchi, 2003; Travers, 2003) (Figure 6). There is also evidence that HMGB proteins further contribute to the modulation of transcription by interacting with certain TFs like P53, p73, and sterol-regulatory element–binding proteins (Agresti and Bianchi, 2003; Bianchi and Agresti, 2005; Najima *et al*., 2005). The HMGB family encompasses three proteins, HMGB1, HMGB2 and HMGB3, with 80% amino acid identity between them and very similar functions, but different spatial and temporal patterns of expression in mammals (Agresti and Bianchi, 2003; Muller *et al*., 2004; Nemeth *et al*., 2005). While embryos are able to express elevated levels of all of them, upon development, HMGB2 and HMGB3 are only significantly present in the testis and lymphoid organs whereas

HMGB1 is almost ubiquitous expressed, being absent only in brain neuron cells (Muller *et al*., 2004; Bianchi and Agresti, 2005). The importance of this last member of the HMGB family has been increasing in recent years as an ever growing number of studies uncover novel functions of this protein, which is extremely well conserved among mammalians, with almost 99% identity between different species (Bustin, 1999; Muller *et al*., 2004).

Besides the above mentioned role in transcription regulation and nucleosome sliding, HMGB1 is involved in other processes such as the regulation of monocyte proinflammatory cytokine synthesis (Andersson *et al*., 2000), and V(D)J (variable, diverse and joining genes) recombination which is the process through which B and T cells assemble diverse gene segments in a random manner to generate distinct receptors (Swanson, 2002). This protein also seems to have a role in the invasive and metastatic properties of cancer (Ellerman *et al*., 2007). Furthermore, there is accumulating evidence that depicts HMGB1 as having a function as a histone chaperone. Yeast mutant cells for the *Hmgb1* orthologous gene *nhp6,* show reduced histone content. Likewise, mammalian cells lacking this element show a reduced number of nucleosomes in their genome due to a significant drop in the levels of all core and many histone variants (Celona *et al*., 2011). Furthermore it was found that HMGB1 enhances nucleosome remodeling and that the addition of HMGB1 in a classical stepwise dialysis nucleosome assembly method increases the assembly of middle-positioned nucleosomes by more than two fold (Bonaldi *et al*., 2002; Osmanov *et al*., 2013). Although this role as a histone chaperone is not fully understood it could be linked to HMGB1s DNA bending properties rather than any direct interaction with histone particles. In order to wrap DNA around histone octamers the nucleic acid molecule must bend several times, a process with a high energy requirement. HMGB1 is able to bend DNA in such a manner and the resulting V shaped structure (Figure 6) might drastically decrease the energy requirements for this process, facilitating the assembly of the prefolded histone octamer. Possible evidence for this comes from yeast where it was found that mutants expressing a defective Nhp6 protein were unable to twist the DNA properly presenting a disruption in the chromatin structure and defects in transcription similar to those observed in *nhp6A/B* deficient cells (Dowell *et al*., 2010).

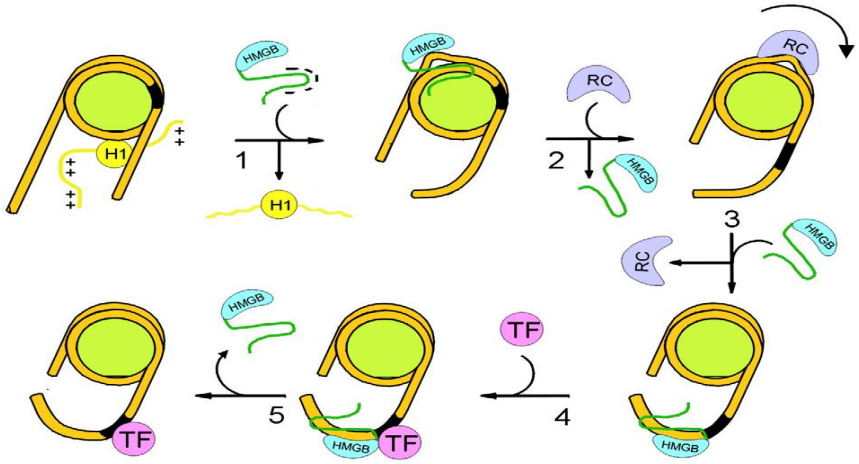


Figure 6: Model of HMGB family function and interactions. HMGB binds to the entry point of nucleosomes and interact with the positively charged flanking sequences of H1 through their negatively charged C-tail, displacing the linker histone and promoting its eviction from chromatin (1). Afterwards a DNA loop is formed at the site where HMGB is bound, which serves as an ideal anchoring point to whom remodeling complexes (RCs) bind. Then HMGB dissociates from the chromatin, followed by the directional propagation of the looped DNA (bound with the RC) around the nucleosome particle (2). This places the TSS (black DNA sequence) out of the histone core. The RC then subsequently dissociates from chromatin and HMGB binds to the DNA molecule in a region adjacent to the TSS bending the DNA in a V shaped structure (3) facilitating the binding of TFs to the TSS (4). Following this HMGB is dissociated once more and the activation of transcription ensues (adapted from Stros, 2010).

The overall decrease in histone levels driven by the lack of HMGB1 leads to a global increase in transcripts in mouse embryonic fibroblasts (MEFs). Loss of Nhp6 also affects the expression of about 10% of yeast genes (Celona *et al*., 2011). HMGB1 also has major impact in normal development. Mouse mutants for this protein are not viable and die shortly after birth (Calogero *et al.*, 1999) and yeast defective for Nhp6 present irregular phenotypes (Paull *et al.,* 1996). Also *Hmgb1*-/- cells are more susceptible to DNA damage by UV and ionizing radiation (Giavara *et al.*, 2005; Celona *et al*., 2011). However, little is known about how HMGB1 and the epigenetic changes caused by its depletion affect DNA replication specifically. A more permissive chromatin sate is a common characteristic between gene promoters and ORIs which might explain the strong correlation found between transcriptional activity and efficient ORIs (Sequeira-Mendes and Gómez, 2012). It seems likely that the overall increase in transcriptional activity is a direct consequence of the lower histone content found in HMGB1 deficient cells, so one might argue that this chromatin environment could favor DNA replication as well, facilitating the assembly of ORC and the activation of pre-RCs. It could also have an impact on the movement and velocity of the replication forks. It has been found recently that human cells lacking the stem-loop binding protein gene (*SLBP*), which controls stability, processing, nuclear export, and translation of canonical histone mRNAs, show a comparable decrease in the levels of canonical histones and some H2A variants (Jimeno-González *et al*., 2015). The resulting changes in the chromatin structure lead to an

increase in RNAPII elongation rates. If a global decrease in histone content can affect the velocity of the transcription polymerases then it also could have an impact on the movement of the replication machinery as well.

Deciphering the full spectrum of consequences that these epigenetic changes caused by lack of HMGB1 have on genome regulation remains an open question, specially due to the fact that there are novel functions of this protein that are still being uncovered. Although much is known about its impact on transcription there is not many data available on how the replication program is affected and what consequences can this have for genome stability and integrity. Answering these questions could help fill the knowledge gap in this particular subject and would contribute to the general understanding of how deregulations in chaperone and histone remodeler activities influence epigenetic stability.

3.4. Chromatin structure and coordination between different genomic processes

Numerous factors have an impact on chromatin stability and the deregulation of any one of them can have different outcomes on the global or local structure of this element and thus modulate many different processes like transcription, replication, repair and recombination in dissimilar ways. Some epigenetic changes may influence replication in a positive manner, like local chromatin decompaction which facilitates ORI firing but also results in conflicts between the replication and transcription machineries (Barlow and Nussenzweig, 2014). Such conflicts between both processes are a great source of genome instability and may have dire consequences for cell viability if left unchecked (García-Muse and Aguilera, 2016; Hamperl and Cimprich, 2016). In Support of this, transcription-replication conflicts are a hallmark of some types of cancer, and mutations in genes involved in preventing and resolving the collisions between both machineries often give raise to many disorders like amyotrophic lateral sclerosis type 4, ataxia-ocular apraxia type 2 and Fanconi anemia (Schwab *et al*., 2015; García-Rubio *et al*., 2015). There are several possible ways in which the two machineries can obstruct each other and, overall, the most common interferences come from head-on or co-directional collisions (García-Muse and Aguilera, 2016) (Figure 7). In normal conditions cells usually deal with the stress caused by these events through the action of specific enzymes such as RNAseH1, which degrades the RNA molecule present in RNA:DNA hybrids, or topoisomerases that resolve the resulting topological stress (Helmrich *et al*., 2013) (Figure 8). Chromatin itself likely plays a crucial role in preventing and regulating possible transcription-replication encounters. For example, both in yeast and humans, it was seen

that cells lacking the FACT (facilitates chromatin transcription) chromatin remodeling complex present a high degree of replication fork stalling, a clear sigh of genomic instability that may be related to abnormal transcription regulation in the absence of this factor. These cells were able to recover from their replication defects by inhibiting transcriptional activity with specific drugs, showing that the abnormal fork staling was mainly due to collisions between both machineries (Herrera-Moyano *et al*., 2014). Furthermore, FACT-depleted cells also show increased levels of R-loops, which are structures composed of three nucleic acid strands forming a RNA:DNA hybrid and leaving one displaced ssDNA strand in a loop (Figure 7c and Figure 8). This is important because although R-loops are described as having regulatory roles in some genomic processes, such as transcription termination (Skourti-Stathaki *et al*., 2014), or immunoglobulin class-switch recombination (Yu *et al*., 2003; Santos-Pereira and Aguilera, 2015), an abnormal accumulation of R-loops in the genome is often related with genomic instability (Tuduri *et al*., 2009; Aguilera and Garcia-Muse, 2012; Skourti-Stathaki *et al*., 2014; Skourti-Stathaki and Proudfoot, 2014). This is mainly due to the fact that these structures actively contribute to the increase in fork stalling frequency (Gan *et al*., 2011) (Figure 7c).

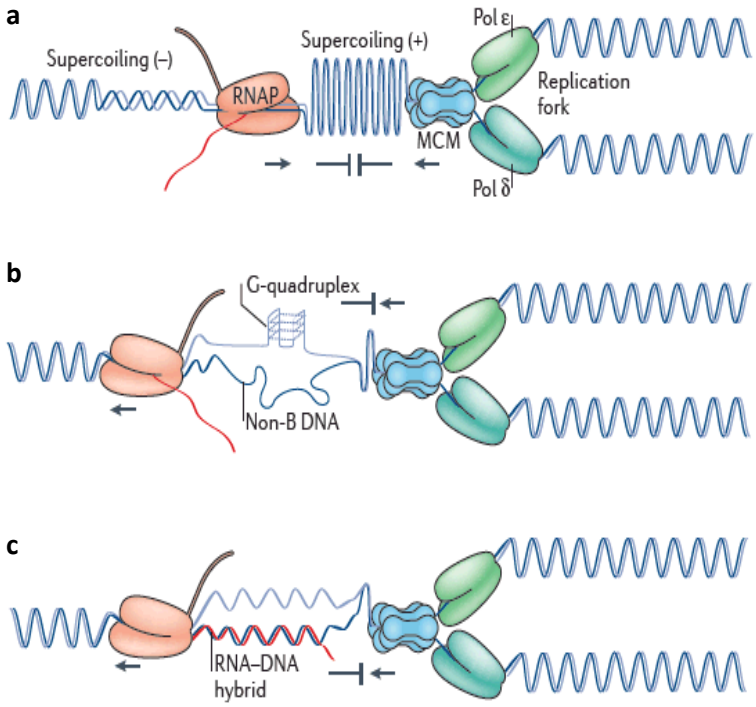


Figure 7: Different types of transcription-replication encounters. **(a)** Head-on collision between transcription and replication machineries. These types of collisions can cause accumulation of positive DNA supercoiling, leading to fork stalling. **(b)** and **(c)** co-directional collisions between both machineries. The negative supercoiling that results from the action of RNA Polymerases unwinds the DNA double helix which leads to the formation of non-B DNA structures, like G-quadruplexes (b) or co-trancriptional R-loops (c), that impair the movement of the incoming replication machinery (adapted from García-Muse and Aguilera, 2016).

Besides being a source for genomic instability, these types of conflicts can fundamentally alter the replication program (Figure 8). There is evidence that in yeast, defects in the transcription machinery related with the termination of this processes can lead to a rearrangement of the MCM2-7 complexes along the genome due to encounters with RNAPII and consequently alter the initiation landscape (Gros *et al*., 2015). Similar studies in Drosophila also demonstrate that the same complexes also shift their initial positions and are relocated to non-transcribed regions of the genome when cells are blocked at the transition from G1 to S phase of the cell cycle (Powell *et al*., 2015). Therefore, it seems that under certain circumstances where the kinetics of the RNA polymerase or the movement of the replication fork is altered, replication initiation might be driven by the elongation of transcription and, has previously mentioned, chromatin structure plays a key role in regulating those processes.

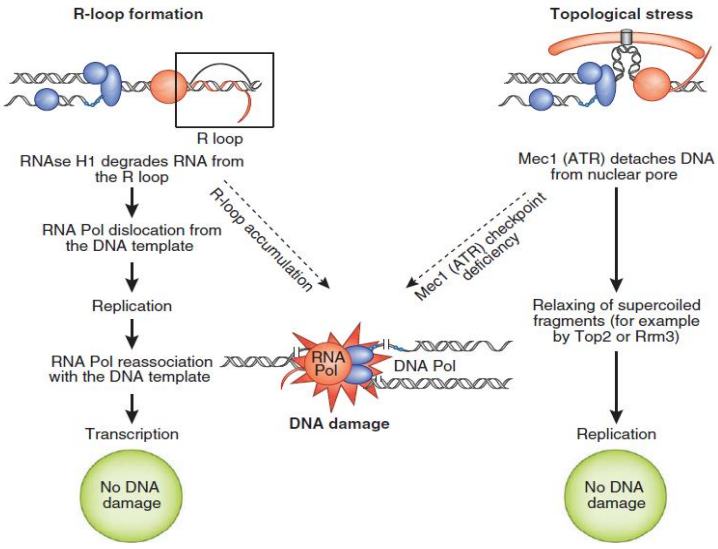
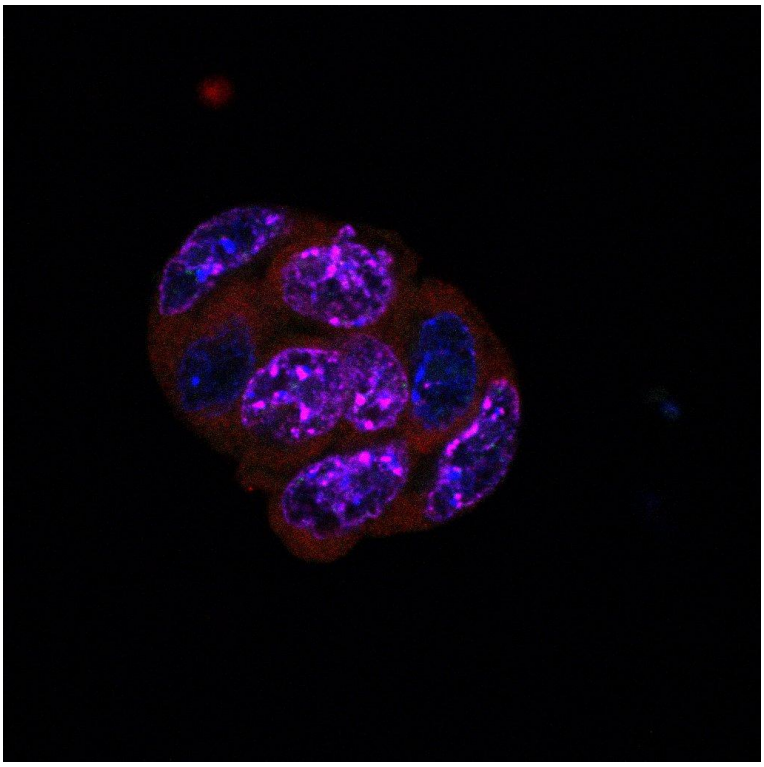


Figure 8: Cellular mechanisms to resolve collisions between active transcription and replication. Two of the most common problems arising from encounters between these two machineries are the generation of RNA:DNA hybrid structures (left) and topological stress that inhibits fork progression (right). If left unchecked genome instability can be generated through the appearance of breaks in the DNA molecule. Usually cells are able to resolve these types of conflicts either through the action of RNaseH1, in the case of R-loop formation, or the activation of the Mec1 (ATR) checkpoint which promotes the recruitment of specific enzymes and topoisomerases to resolve the topological stress and prevent DNA damage (adapted from Helmrich *et al*., 2013).

Genomes with altered chromatin configuration, like cells lacking H1 variants or proteins from the HMG family can be suitable a scenario to study how specific modifications in conformation might alter DNA replication dynamics and what impact would they have on transcription kinetics and replication fork stability as well as which consequences the interplay between those two processes would have for cell viability. This would also contribute to the understanding of the specific nature of the relationship between replication and transcription in the specification of ORIs in mammalians.

Objectives

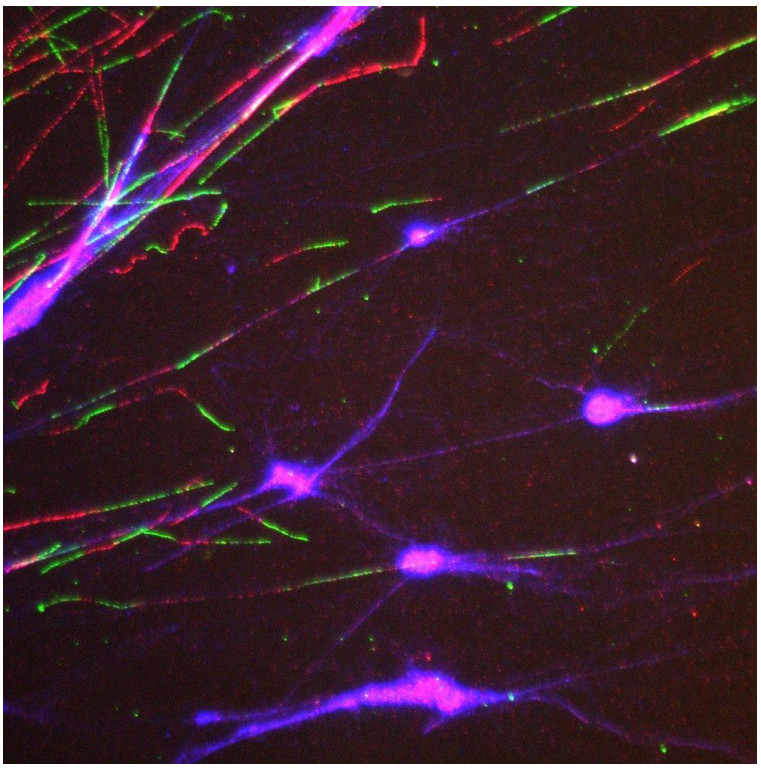


In order to increase our understanding of the role of chromatin in ORI specification and activity, and to determine how specific alterations in this element affect the stability of the replication program, we set the following objectives for this Doctoral Thesis:

- Carry out a high-resolution analysis of nucleosome architecture and DNA synthesis start sites at efficient mammalian ORIs.

- Perform a functional analysis of the impact of altered chromatin configurations in the definition of mammalian ORIs and in the kinetics of replication elongation.

Materials and Methods



1. CELL MANIPULATION

1.1. GROWTH CONDITIONS

All cell types cultured were grown at 37°C in a humid environment with 5% CO_2 inside an incubator (Esco Celculture) and handled in an appropriate P2 cell culture hoods (Faster BH 2004).

Mouse embryonic stem cells (mES) PGK 12.1 (Penny *et al*., 1996) and WW6 (Ioffe *et al*., 1995) were grown in Dulbecco's modified Eagle's medium (DMEM, Invitrogen) supplemented with 10% fetal bovine serum (FBS, Biosera), 1x non-essential aminoacids (Gibco), 2 mM L-glutamine (Gibco), 50 μM β-mercaptoethanol (Gibco), 1 mM sodium piruvate (Gibco), 100 U/mL penicillin plus 100 μg/mL streptomycin (Invitrogen) and 10^3 U/mL LIF (ESGRO). HMGB1 MEFs were derived from 14.5 dpc (Balb/C) embryos (Calogero *et al*., 1999; Celona *et al*., 2011). HeLa cells (ATCC) and both NIH-3T3 (ATCC) and HMGB1 MEFs were grown in DMEM high glucose medium (Lonza) supplemented with 10% FBS (Gibco), 100 U/mL penicillin plus 100 μg/mL streptomycin (Invitrogen), 100 μg/mL streptomycin (Gibco), 2 mM L-glutamine (Gibco), 1x non-essential aminoacids (Gibco) and 50μM β-mercaptoethanol (Gibco). HTC-shSLBP.1 cells were cultured with or without 2 μg/mL Doxycycline (Sigma-Aldrich) for 72 hours as described (Jimeno-Gonzalez *et al*., 2015) in McCoy´s 5A-modified medium (Sigma-Aldrich) supplemented with 10% FBS (Gibco), 100 U/mL penicillin plus 100 μg/mL streptomycin (Invitrogen).

1.2. PASSING CELL CULTURES

Cells at approximately 80% confluence were washed twice with PBS (137mM NaCl, 2.7mM KCl, 10 mM Na_2HPO_4, 1.76 mM KH_2PO_4, pH 7.4) at room temperature and incubated with Trypsin (0,25% with 0,04% EDTA, Invitrogen) for approximately 4 minutes. In the case of mES cells Trypsin was supplemented with 2% chicken serum (Invitrogen). Dissociated cells were resuspended in culture medium to quench trypsin activity and then centrifuged 5 minutes at 200 rcf. Afterwards the cell pellet was resuspended in appropriate medium and divided accordingly to new culture flasks.

1.3. FREEZING AND STORING CELL CULTURES

Cells at approximately 80% confluence were trypsinized as described before and then collected by centrifugation. Afterwards cells were carefully resuspended in freezing medium composed of 90% FBS (Gibco) and 10% DMSO (Merck) at a concentration of

approximately 1x10^6 cells/mL. Freezing was carried out overnight in a Mr. Frosty container to gradually decrease cell temperature until it reached -80^{o}C. The following day frozen cells were transferred to liquid nitrogen for long-term storage.

1.4. Reactivating Cell cultures

Each cell vial containing 1 mL of frozen cells was removed from liquid nitrogen and thawed in a water bath at 37^{o}C and then diluted 1:10 in appropriate, pre-heated, growth medium and mixed gently to dilute the DMSO in the freezing medium. After centrifugation at 200 rcf for 5 minutes the resulting cell pellet was resuspended in culture medium and the cells were transferred into a T25 tissue-culture flask (Thermo Fisher Scientific).

2. Flow Cytometry

In order to analyze cell-cycle profiles, cells were incubated for 20 minutes with 250µM IdU (Sigma-Aldrich) and then fixed overnight in 70% ethanol at -20^{o}C. Cells were then incubated in 2 M HCl (Merck) with 0.5% Triton X-100 (Sigma-Aldrich) for 30 minutes and neutralized with 0,1 M Sodium Tetraborate pH 9.5 (Merck) for 2 minutes before blocking 10 minutes with a solution of 1% BSA (Sigma-Aldrich) and 0.5% Tween20 (Sigma-Aldrich) in PBS. Afterwards cells were incubated for 1 hour at room temperature (RT) with mouse anti-BrdU primary antibody (BD Biosciences) and then with the anti-mouse IgG Alexa-Fluor 647 secondary antibody (Thermo Fisher Scientific) for 30 minutes at RT. Cells were finally stained with PI/RNAse cycle buffer (BD Pharmingen) for another 30 minutes in the dark at RT. All samples were processed in a FACSCanto II (Becton Dockinson) with FACSDiva v6.1.3 analysis software, and then analysed with the FlowJo v10 program.

3. Cell synchronization and immunofluorescence with anti-BrdU Antibodies

A double-thymidine block was performed by growing HeLa cells for 16 hours in culture medium containing 2.5 mM thymidine (Sigma-Aldrich) followed by 9 hours of incubation in fresh medium and further incubation for 16 hours in the presence of 2.5 mM thymidine. Progression through S-phase was checked by FACS analysis at 0 hours, 45 minutes and then every 1.5 hours after block removal. Synchronous entry in S-phase was checked by immunofluorescence on parallel cultures of cells grown on slides labeled

during 15 minutes with 10 mM BrdU (BD Biosciences) prior to cell harvest as described in Leonhardt *et al*, 1992.

4. Purification of nucleic acids

4.1. Extraction of genomic DNA

Genomic DNA was extracted from cells using the standard procedures described in Sambrook, 1989. Trypsinised cells were resuspended in Lysis Buffer consisting of 50 mM Tris pH 8.0, 10 mM NaCl, 10 mM EDTA pH 8.0, 0.5% SDS, 100µg/mL Proteinase K (Roche) and incubated overnight at 37^{o}C. The cell lysate was gently mixed with an equal volume of Tris-saturated Phenol (Sigma-Aldrich) and centrifuged 10 minutes at 1500 rcf. The aqueous phase was then mixed with an equal volume of Phenol:chloroform:isoamylalcohol (25:24:1) saturated with 10 mM Tris pH 8.0 and 1 mM EDTA (Sigma-Aldrich) and centrifuged at 1500 rcf for 5 minutes. DNA was then precipitated with 2 volumes of ice-cold (-20 ^{o}C) 100% ethanol (Merck), washed with 70% ethanol and air-dried before resuspending in TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) supplemented with 0.1 U/µL Units of RNAseOUT^{TM} (Invitrogen). Samples were then kept at 4^{o}C until further processing.

4.2. Purification of small DNA replication intermediates

For each gradient tube total genomic DNA from approximately 10^{9} exponentially growing cells was heat-denatured at 100^{0}C for 10 minutes and then size-fractioned by centrifugation at 78000 rcf in an Ultracentrifuge (Beckman Coulter Optima^{TM} L-100 XP), with a SW-40Ti rotor (Beckman Coulter) and appropriate centrifuge tubes (Beckman Coulter 331374) for 20 hours at 20^{o}C, in a seven-step neutral sucrose gradient (5-20% sucrose, 2.5% steps made in 10 mM Tris pH 8.0, 1 mM EDTA, and 100 mM NaCl), as described in Gómez and Antequera, 2008. After centrifugation, twelve to thirteen 1 mL fractions were collected and the DNA in each fraction was ethanol-precipitated. About 10% of each fraction was analysed in a 1% alkaline agarose gel (50 mM NaOH, 1 mM EDTA) to monitor the fractionation profile (Figure 9a). Sucrose gradient fractions containing replication intermediates ranging between 100-600 nt (usually fraction 3), 300-800 nt (usually fraction 4) and 300-1200 nt (usually fraction 5) of size were treated with polynucleotide kinase (PNK, Thermo Scientific) to phosphorylate 5'-hydroxyl ends and render all DNA molecules in the samples available for λ-exonuclease digestion. This last enzyme degrades fragments contaminating random sheared DNA leaving intact DNA

replication intermediates that are protected by a 5'-RNA-primer. The PNK phosphorylation reaction was performed in the presence of 1 mM dATP (Roche) and 40 Units of RNAseOUT^{TM} (Thermo Scientific) and 100 units of PNK enzyme, for 30 minutes at 37^{o}C. After PNK inactivation with 6.25 µg of protease K (Roche) in the presence of 0.125% sarkosyl and 2.5 µmol of EDTA, the phosphorylated DNA was extracted, precipitated and resuspended in water. Once resuspended, samples were heat-denaturated again for 5 minutes and the digestion with lambda-exonuclease (custom-made, Thermo Scientific) was done with 150 Units of this enzyme plus 1x λ-exonuclease digestion buffer (Thermo Scientific) and 40 Units of RNAseOUT^{TM} at 37°C overnight. Reactions were inactivated during 10 minutes at 75°C, and then extracted with phenol/chloroform, ethanol precipitated and resuspended in water.

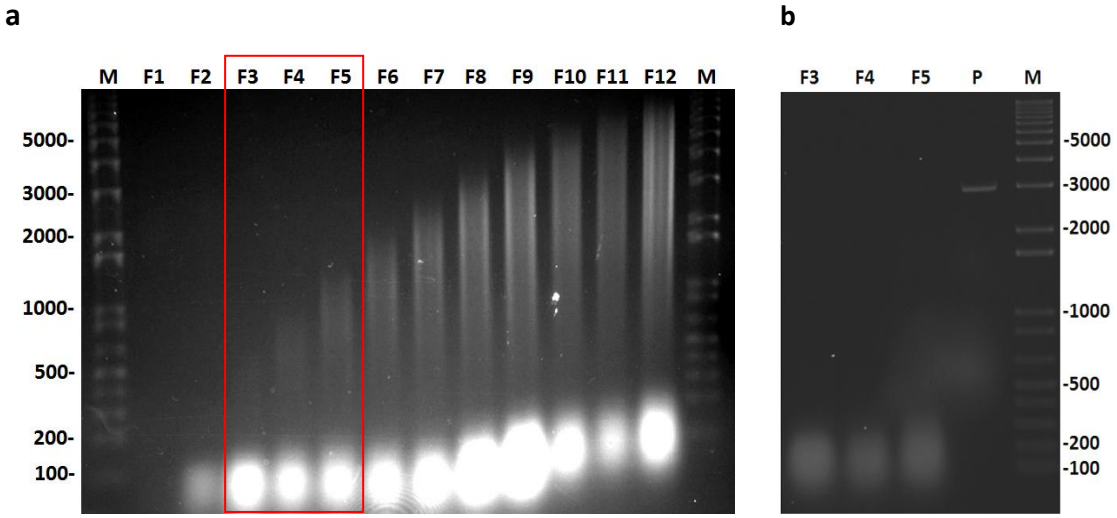


Figure 9: **(a)** Representative example or genomic DNA fractionation in an alkaline 1% agarose gel after a seven-step neutral sucrose gradient. Numbers to the left correspond to the DNA fragment sizes of the molecular marker (1 Kb plus DNA Ladder, Invitrogen) present in the first and last wells (M). Fraction number is indicated above the wells (F1 to F12). Red box indicates the main fractions used for the subsequent experimental procedures present in this work (F3, F4 and F5). **(b)** Representative example λ-exonuclease of digestion control. F3, F4 and F5 indicate the wells were about 5% of each fraction of interest was loaded upon digestion with λ-exonuclease in the presence of a phosphorilated linear 3 Kb plasmid. Undigested control plasmid is indicated as P. M indicates the molecular marker.

The efficiency of the digestion was routinely monitored by adding 20 ng of a linearised 3 Kb-plasmid to a parallel tube containing about 5% of each sample, and incubated in the same conditions. Those controls were then loaded on a 1% agarose gel to confirm plasmid digestion by λ-exonuclease (Figure 9b). DNA from each λ-treated fraction was then precipitated by adding 20 µg/mL glycogen (Roche), 0.3 M sodium acetate pH 5.2, and 2 volumes of ice-cold 100% ethanol. The samples were kept at -20^{o}C for 30 minutes, centrifuged at 20000 rcf for another 30 minutes at 4^{o}C to pellet the DNA and washed with 70% ethanol at room temperature. DNA pellets were air-dried before

resuspending in TE. Final samples were used either for qPCR quantification or further processed for high throughput sequencing.

5. PROCESSING SNS SAMPLES FOR GENOMIC LIBRARY PREPARATION AND SEQUENCING

To prepare replication intermediates for sequencing, SNS were primed with 50 pmol of random hexamer primer-phosphate (Roche) and incubated for 5 minutes at 95°C followed by gradually cooling to 4°C as described in Cadoret *et al.*, 2008. Primer extension was performed by adding 10 mM of dNTPs (Roche) plus 5 Units of exo-Klenow (New England Biolabs), and incubating for 1 hour at 37°C. Reactions were inactivated for 10 minutes at 75°C. Ligation of adjacent fragments of the second-strand synthesis was performed by adding 80 Units of Taq DNA ligase (New England Biolabs), incubating for 30 minutes at 50°C and inactivating the enzyme for 10 minutes at 75°C. Finally, RNA primers were removed by treating samples with 5 Units of RNAse A/T1 Mix (Thermo Scientific) for 30 minutes at 37°C. DNA was extracted with phenol/chloroform (Sigma-Aldrich), ethanol precipitated and resuspended in TE.

DNA libraries were prepared at the Fundación Parque Científico de Madrid (FPCM), with the NBNext kit (New England Biolabs) following manufacturer's instructions, and library fragments were purified from polyacrylamide gels. Each library was sequenced by 2x75 single-end runs on a NS500 system (Illumina) at the FPCM.

6. SNS-SEQ DATA PROCESSING

SNS-seq reads were aligned using the standard BWA-MEM algorithm (Li, 2013), with the parameter -q 1 to avoid multihits. ORI peaks were determined with the scanquantile program (Picard *et al*., 2014), with a p-value threshold of 1E-16. The required genome segmentation used by this peak-calling script was based on replication timing data from mES (Hiratani *et al*., 2011), which accurately matches the read coverage differences between segments. Finally, the peaks separated by less than 200 bp were merged (keeping the lowest p-value), and the peaks with a p-value higher than a threshold were removed; this threshold was fixed depending on the read depth and the background of each track.

Common peaks were obtained with the intersectBed (BedTools), with parameters -wa -f 0.1: nonreciprocal, and with a minimal fraction of 10% of overlap. The pairwise comparison between origins per segment, and the Venn diagrams were generated with

custom R scripts; the dendrogram of the heatmap was calculated with the standard hierarchical clustering implemented on the heatmap.2 function (gplot package).

To account for the observed genomic distribution of replication origins, ORI peaks were compared with the expected proportion calculated from genomic intervals randomly sampled from throughout the genome. The procedure was repeated 1,000 times and the statistical significance of the observed distribution was determined by computing the empirical p-value from the sampling distribution.

7. MONONUCLEOSOMAL DNA EXTRACTION AND PURIFICATION

The method for mapping nucleosomes was adapted from Gong *et al*., 1996. Upon reaching 80% confluence, nuclei were extracted from cells grown under previously described conditions. The medium from each culture flask was removed and the flask was washed twice with PBS. In certain cases cells were previously crosslinked with 1% Formaldehyde (Sigma-Aldrich) for 10 minutes at room temperature, in order to compare NuSA patterns derived from native against crosslinked chromatin. After, cells were washed two times in cold PBS containing protease inhibitors (Sigma-Aldrich) and transferred to cold 50 mL falcon tubes (Falcon). Cell concentration was estimated in a Neubauer chamber. After collecting cells by centrifugation at 200 rcf for 5 minutes at 4^{o}C they were resuspended in cold Homogenization Buffer consisting of 10 mM Tris pH 7.4, 1 mM EDTA, 0.1 mM EGTA, 15 mM NaCl, 50 mM KCl, 0.15 mM spermine (Sigma-Aldrich), 0.5 mM spermidine (Sigma-Aldrich), 0.2% Nonidet P-40 (Sigma-Aldrich), 5% (weight/volume) sucrose (Merck) at a concentration of approximately 1,7x10^{7} cells/mL. The aliquots were then left 3 minutes on ice in order to lysate the cell membrane and obtain intact nuclei. After this, nuclei were centrifuged (20 minutes at 900 rcf at 4^{o}C) through a cushion of Homogenization Buffer containing 10% sucrose. Each pellet was resuspended in 6 mL of cold Wash Buffer (10 mM Tris pH 7.4, 15 mM NaCl, 50 mM KCl, 0.15 mM spermine, 0.5 mM spermidine, 8.5% sucrose, 1 mM of $CaCl_2$) and divided in aliquots of 1 mL. Samples were incubated for 3 or 6 minutes at 25^{o}C with different Units of MNase (Thermo Fisher Scientific), ranging from 0 Units to 2400 Units. Digestions were stopped by complementing the reactions up to 9 mM EDTA and 3.5 mM EGTA. Crosslinks were reversed overnight at 65^{o}C in the presence of 150 mM NaCl. Samples were then incubated for 30 minutes at 37^{o}C with 0.5 µg of RNAse, DNase free (Roche) and then overnight at 45^{o}C in a final concentration of 1% SDS and 0,1mg of proteinase K (Roche). Next day, Nucleosomal DNA was purified by two consecutive phenol-chloroform extractions. All the samples were resuspended in TE and nucleic acid concentration was

measured in a nanodrop spectrophotometer (Nanodrop ND-1000, thermo scientific). Each digestion was analyzed by gel electrophoresis (1.2% agarose gel) (Figure 10a). The DNA fragments corresponding to the mononucleosomal fraction of the samples digested with MNase were sliced from the agarose gels and purified using a Wizard SV gel purification kit (Promega) following manufactures instructions. Samples were then diluted in TE for qPCR analysis using an approach described in Sekinger *et al*, 2005, with amplicons of 60 to 80 bp in size and with a 15 to 25 bp overlap (Supplementary Tables 1 to 5).

Undigested MNase controls were also prepared in parallel in order to perform the qPCR normalization. Figure 10b exemplifies the typical nuclesome positioning pattern inferred by NuSA analysis at a TSS.

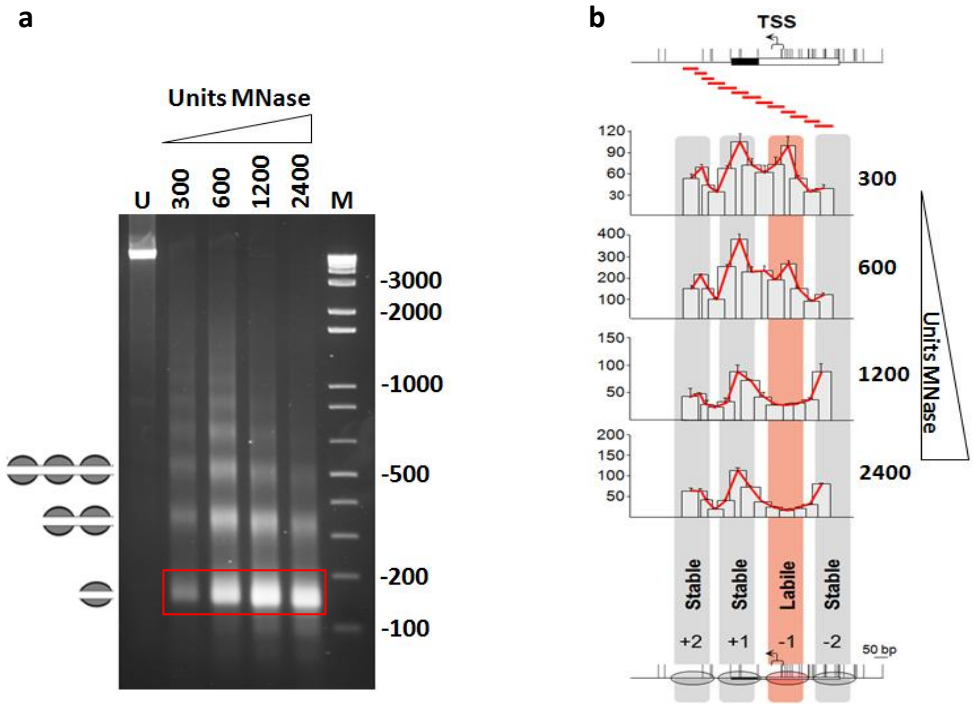


Figure 10: (a) 1,2% agaross gel representing a tipical fragmentation of the chromatin after digestion with increasing MNase concentrations. MNase units are indicated on top of the gel. Undigested control is present on the first well (U). Numbers to the right correspond to the DNA fragment sizes of the molecular marker (M, 1 Kb plus DNA Ladder, Invitrogen). Positions of the bands corresponding to mono-, di-, and tri-nucleosomal DNA fragments are indicated on the left. Bands present inside the red square (mono-nucleosomal DNA fragments) are excised from the gel and purified for NuSA analysis by qPCR. **(b)** Scheme illustrating the nuclesome positioning pattern inferred by NuSA analysis. 4 nucleosomes are detected around the TSS region. The -1 nucleosomes is classified as a labile particle due to the gradual decrease of nucleosomal DNA levels detected in the higher MNase digestions.

8. CHROMATIN IMMUNOPRECIPITATION

Chromatin derived from approximately $1x10^7$ cells was crosslinked with 1% formaldehyde, in PBS, for 10 minutes at room temperature. Crosslinked chromatin was

then digested with 2560 Units of MNase for 6 minutes at 37°C. Digestions were stopped by bringing the samples to 4°C and 9 mM EDTA plus 3.5 mM EGTA. Samples were then homogenized four times through a 20G needle and another four through a 25G needle. 50 mg of digested chromatin was immunoprecipitated using 4 mg of polyclonal anti-histone H3 antibody (Abcam) using either a low-salt buffer (10mM Tris–HCl pH 7.4; 15mM NaCl; 0.2mM EDTA) as previously described in Jin and Felsenfeld, 2007, or in high salt conditions as described in Sequeira-Mendes *et al*, 2009. Results were analyzed by qPCR using the same overlapping amplicons employed for the MNase-treated samples (Supplementary Table 1 to 5).

9. FIBER STRETCHING

Analysis of DNA replication by fiber stretching was adapted from Terret *et al*., 2009. Figure 11 represents a simplified view of this method as well as the most common structures that can be identified.

Exponentially growing cells were pulsed first with 50µM CldU (Sigma-Aldrich) for 20 minutes and then with 250µM IdU (Sigma-Aldrish) for another 20 minutes. Cells were then resuspended at a concentration of $0{,}5x10^6$ cells/mL in cold (4°C) PBS. 2 µL of cell suspension were directly lysed on appropriate microscopy slides (VWR) by adding 10 µL of pre-warmed (30°C) spreading buffer (0.5% SDS, 200 mM Tris pH 7.4, 50 mM EDTA) and incubated for 6 minutes at room temperature in humidity chamber. DNA fibers were stretched by tilting the slide approximately 30°. After air drying, the samples were fixed 2 minutes with cold (-20°C) 3:1 methanol:acetic acid solution. Slides were then incubated in 2.5 M HCl (Merck) for 30 minutes at room temperature, washed three times with PBS, and blocked for 1 hour with a solution of 1% BSA, 0.1% Triton X-100 in PBS. Then the samples were incubated with 1:100 anti-CldU (Abcam), 1:100 anti-IdU (BD) and 1:300 anti-ssDNA (Millipore) antibodies for 1 hour at room temperature in humidity chamber followed by another 30 minutes with the secondary antibodies (all at 1:300 concentration) anti-rat IgG Alexa Fluor 594 (Molecular Probes), anti-mouse IgG1 Alexa Fluor 488 (Molecular Probes) and anti-mouse IgG2a Alexa-Fluor 647 (Molecular Probes). Finally slides were air dried and mounted with Prolong Diamond (Invitrogen). Visual acquisition of the DNA fibers was done in a Axiovert200 Fluorescence Resonance Energy Transfer microscope (Zeiss), using the 40x oil objective, and the images were analyzed with the image processing program ImageJ v1.51a using the conversion factor of 1 µm = 2.59 kb (Jackson and Pombo, 1998). Statistical analysis of all data was performed in Prism v5.0.4 (GraphPad Software) using the non-parametric Mann-Whitney rank sum test as described

in Técher *et al*., 2013 (Supplementary Tables 8, 9, 17 to 19 and 21 to 24). *:p<0.05, **:p<0.01, ***:p<0.001, ****:p<0.0001.

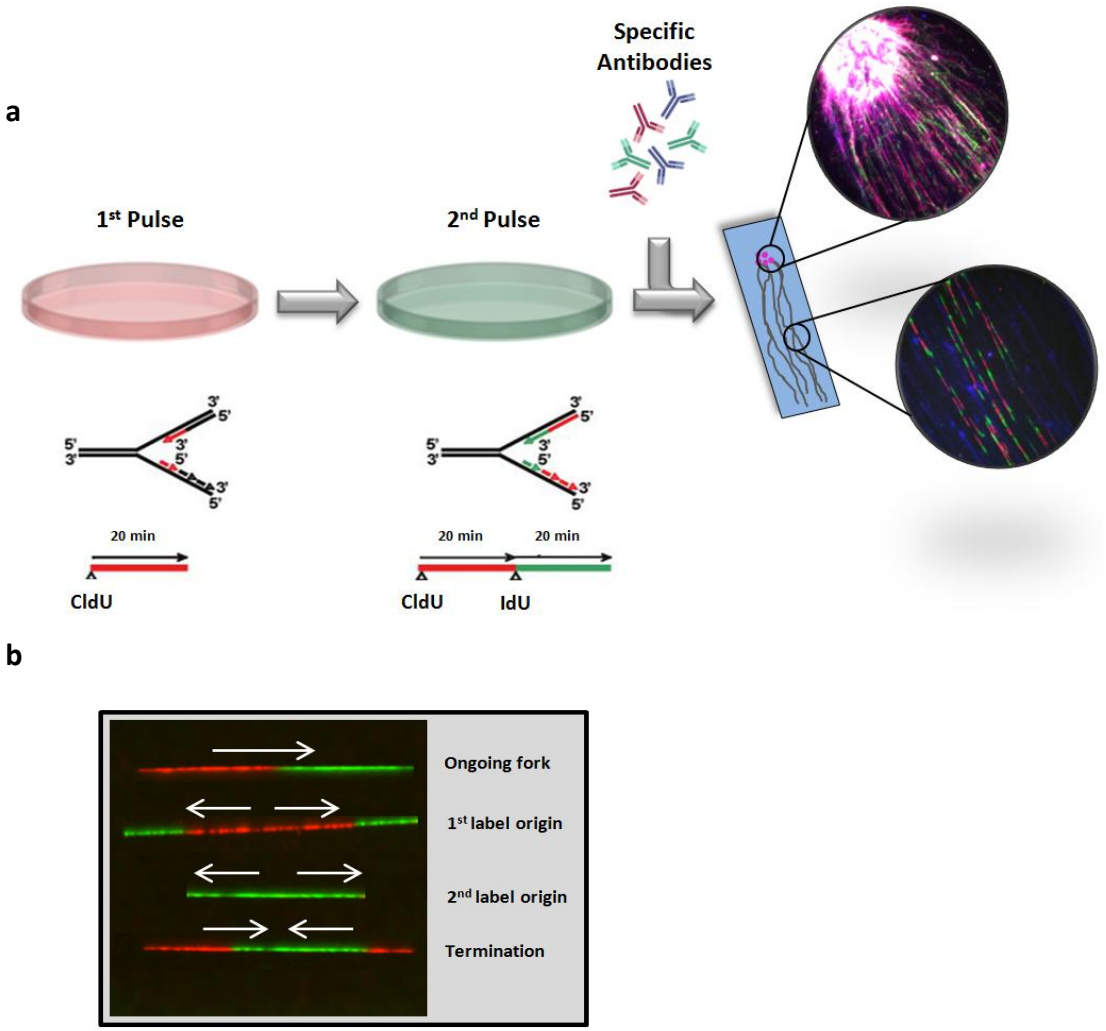


Figure 11: (a) Simplified scheme representing the DNA fiber stretching method. After two consecutive 20 minute pulses with CldU (represented as red tracks) and IdU (represented as green tracks) cells are directly lysed on the slide and fibers are stretched and labeled with specific antibodies. Lysed nuclei can usually be seen on the upper part of the slides whereas stretched fibers are normally present on the middle and lower parts. **(b)** Example of the 4 most prominent structures that can be identified by DNA fiber stretching. White arrows indicate the orientation of the replication forks.

10. Transcription Inhibition

After growing cells to approximately 80% confluence in pre-treated coverslips (VWR), RNAPII-dependent transcription was inhibited by incubating cells in the presence of either 100 µM DRB for 3 hours (Sigma-Aldrich) or 10 µg/mL α-amanitin (Sigma-Aldrich) for 6 hours. RNAPI-dependent transcription was inhibited by incubating cells in the presence of 0.05 µg/mL Actinomycin-D (Sigma-Aldrich) for 1 hour.

11. IMMUNOFLUORESCENCE

11.1. R-LOOP AND γH2AX DETECTION

Cells grown on glass coverslips (VWR) were fixed with 3.7% formaldehyde in PBS for 15 minutes at RT and permeabilized with 0.5% Triton X-100 in PBS for 20 minutes at RT. Samples were blocked with 3% BSA (Sigma Aldrich) in PBS before overnight incubation at 4°C with primary antibodies S9.6 (1:100) (ATCC) and anti-γH2AX (1:250) (Abcam), followed by 1 hour incubation at RT with the respective secondary antibodies anti-mouse Alexa-Fluor 594 and anti-rabbit Alexa-Fluor 488 (Invitrogen) and then 5 minutes staining at RT with 2ng/µl of DAPI (Merck) in PBS. Coverslips were mounted in Prolong Diamond (Life Technologies) and visual acquisition was performed in a LSM510 AxioImager M1 microscope (Zeiss) using either a 63x or a 100x oil objective. Nuclear segmentation was based on DAPI staining. Statistical analyses were performed in Prism v5.0.4 (GraphPad Software) using the non-parametric Mann-Whitney rank sum test (Supplementary Tables 11 to 14, 16 and 20). *:p<0.05, **:p<0.01, ***:p<0.001 , ****:p<0.0001.

11.2. NASCENT DNA LABELING

Cells were incubated in culture medium supplemented with 10 µM EdU for 30 minutes, rinsed in cold medium and washed in PBS before fixation in 3.7% formaldehyde in PBS for 15 minutes at RT. EdU incorporation was revealed with Click-iT® EdU Imaging kits (Invitrogen) using Alexa-Fluor 647 dye according to manufacturer´s instructions. Coverslips were stained at RT with 2ng/µl of DAPI (Merck) in PBS for 5 minutes and then mounted in Prolong Diamond (Life Technologies). Cells were analyzed using the ImageJ v1.51a software and scored as early-, middle-, or late-S phase according to their EdU replication-foci patterns (Nakamura *et al*., 1986; see representative Figure 29a).

11.3. NASCENT RNA LABELING

Cells were incubated in culture medium supplemented with 0.5 mM EU for 1 hour, rinsed in cold medium and in PBS before fixation in 3.7% formaldehyde in PBS for 15 minutes at RT. EU incorporation was revealed with Click-iT® RNA Imaging kits (Invitrogen) using Alexa-Fluor 488 dye according to manufacturer´s instructions. Coverslips were stained at room temperature with 2ng/µl of DAPI (Merck) in PBS for 5 minutes and then mounted in Prolong Diamond (Life Technologies). The mean EU

fluorescence intensity per cell was obtained using ImageJ v1.51a software by averaging the mean grey value of EU signal measured in each nucleus. Nuclear segmentation was based on DAPI staining. Statistical analyses was performed in Prism v5.0.4 (GraphPad Software) using the non-parametric Mann-Whitney rank sum test (Supplementary Tables 10 and 15). *:p<0.05, **:p<0.01, ***:p<0.001 , ****:p<0.0001.

12. RNAPII TRANSCRIPTION ELONGATION

Transient inhibition of RNAPII Ser2 phosphorylation with DRB and release was adapted from Singh and Padget, 2009, following the modifications of Jimeno-Gonzalez *et al*., 2015. Subconfluent cell cultures were incubated in complete medium supplemented with 100 μM DRB for 3 hours, then DRB-medium was washed-off and fresh medium was added to resume transcription elongation. Total RNA from equivalent number of cells was isolated every 10 minutes with RNeasy kit (Quiagen) following the manufacturer´s instructions. One microgram of RNA per time point was treated with DNAse (Roche), and reverse transcription reactions were performed with SuperScript III First Strand Synthesis System (Invitrogen) using random hexamers (invitrogen). Pre-mRNAs at the different time-points were quantified by RT-qPCR with primers spanning different exon-intron junctions (Supplementary Table 25). Pre-mRNA values were normalized to the values of the non-DRB-treated sample, which was set to one.

13. GENE SPECIFIC RT–qPCR

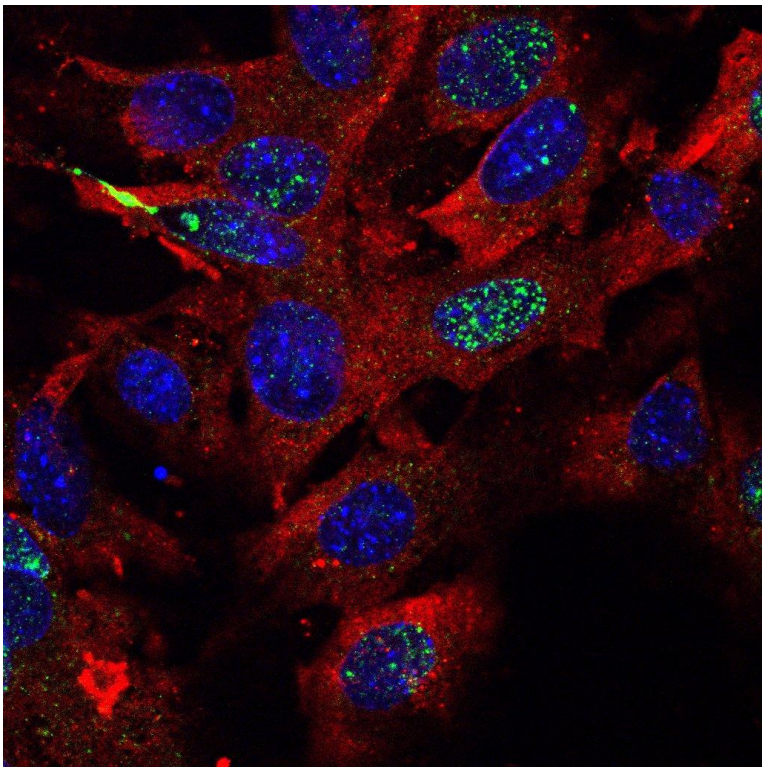
Total RNA was extracted from mES and NIH3T3 cells with a guanidium isothiocyanate solution and cDNA was synthesized using oligo-dT and SuperScript II Reverse Transcriptase (Invitrogen). The amount of cDNA synthesized from 2 μg of total RNA (with or without SuperScript II) was used as a template for each qPCR, using the primer pairs present in Supplementary Table 26. RT–qPCR analysis of *Haus7*, *Vps45* and *Mecp2* mRNA levels was preformed relative to *Hprt* mRNA levels, which was set to one.

14. QUANTITATIVE REAL-TIME PCR

Quantitative real-time PCR (qPCR) was performed in an ABI Prism 7900HT Detection System (Applied Biosystems), using HotStarTaq DNA Polymerase (Qiagen) and following manufacturer's instructions. The primer sequences for all the regions amplified

as well as the qPCR conditions used are indicated in Supplementary Tables 1 to 6, 25 and 26. All qPCR reactions were performed in duplicate. The analyses were carried out using the Applied Biosystems Software SDS v2.4. In the case of absolute quantifications only regions in which the standard amplification curve presented slope value between -3.6 and -3.0 and R2 values greater than 0.990 were considered for analysis.

Results



1. High-resolution analysis of nucleosome architecture and replication initiation at efficient DNA replication origins

1.1. Analysis of nucleosomal profiles at mouse DNA replication origins

The resolution level of ORI genome-wide studies depends greatly on the size of the replication intermediates used for sequencing or for microarray hybridization. Correlating this type of studies with fine chromatin landscape analysis is very difficult as the general resolution of DNA replication studies usually ranges between 1 and 2 kb and nucleosomal particles encompass around 150 bp of DNA. Other major challenge in mapping nucleosomes at such detailed level comes from the fact that specific genomic regions encompass distinct physical properties. Therefore, a meticulous study is required in which digestion with distinct concentrations of MNase and performing ChIP techniques with distinct conditions becomes essential for depicting the dynamics of the histone octamers. Those different experimental conditions, as well as the distinct computational pipelines that can be employed represent a major challenge regarding genome-wide data analysis (Tolstorukov *et al*, 2010).

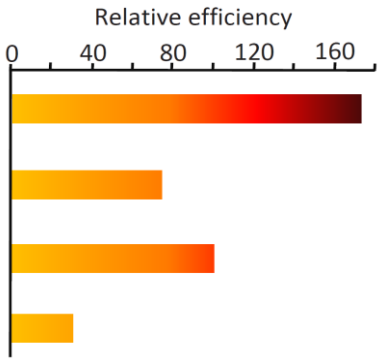
In order to determine the level of influence of local chromatin structure in ORI activity we first decided to conduct a high resolution analysis of nucleosomal architecture at a set of previously described replication initiation sites in the mouse genome. Chosen ORIs were characterized in a previous work from our laboratory (Sequeira-Mendes *et al*, 2009) by hybridizing small replication intermediates, from 300 to 800 bp in size, on tiled microarrays. Table 1 contains the description of the chosen ORIs in terms of genomic location, presence of histone modifications, transcriptional activity, presence or absence of CpG islands (CGIs) and relative ORI firing efficiency. The chosen ORI regions encompass distinct chromatin landscapes in which we performed a detailed nucleosome scanning assay (NuSA) (Sekinger *et al*, 2005; Infante *et al*, 2012).

The NuSA method allowed us to go beyond the resolution of genome-wide studies and characterize the specific nucleosomal landscape at this set of ORIs. By coupling isolation of DNA after MNase digestion with qPCR analysis employing overlapping amplicons (between 60 to 80 bp in size and with 15 to 25 bp overlap) this method is able to generate quantitative high-resolution maps of nucleosome position and occupancy levels (Figure 10b, Material and Methods). When a nucleosome is occupying the same genomic position in all cells of the population analyzed, the amplicons located at the nucleosome-protected DNA will yield higher enrichments when compared with amplicons located at a NDRs or linker DNA regions between histone particles. Due to the shielding

effect that nucleosomes offer, NuSA generates a profile of peaks and valleys that represent the different levels of protection from MNase digestion.

Table 1: Characterization and genomic localization of the mouse origins of DNA replication chosen for detailed NuSA and SNS analysis.

ORI location	Transcriptional activity[b]	Histone modifications[b]	Presence of CpG Island[c]
***Mecp2* promoter**	Moderate	H3K4me3 H3K9, 14 Ac	Yes
***Haus7* promoter**	Moderate	H3K4me3 H3K9, 14 Ac	No
***Vps45* promoter**	Moderate	H3K4me3 H3K9, 14 Ac	No
***Scl7a14* Exon**	NA	NA	No



[a]Previous data from our laboratory generated by SNS hybridization on genomic microarrays (Sequeira-Mendes *et al*., 2009). The location of the ORIs was defined within 800 bp resolution.

[b]Data derived from Sequeira-Mendes *et al,* 2009 and Ku *et al,* 2008.

[c]CGIs were defined by employing the stringent definition from Takai and Jones, 2002.

*NA: Not available

Nucleosome particles can also encompass distinct histone modifications or even distinct histone variants which affect the dynamics of those nucleoproteins, influencing the sensibility to enhanced digestions by nucleases or even ChIP techniques carried out in different conditions (Jin and Felsenfeld, 2007; Jin *et al*, 2009). Taking this in to account we performed purifications of mononucleosomal fractions derived from digestions with increasing amounts of MNase in order to depict the stability of local nucleosome particles (Figure 12a). Thus, nucleosomes that are stably bound after harsh digestions are able to shield the DNA from the action of the nucleases and are consider stable while more labile nucleosome particles aren't able to offer such a strong protection when submitted to strong treatments and thus their signal appears only at mild or soft digestion. As a control we also performed anti-H3 ChIP to detect if the regions classified as nucleosomes by NuSA contained histone particles, and if their stability varied between less (low salt) or more (high salt) stringent conditions (Figure 12b). The length of the scanned was around 600 bp in each case (Supplementary Tables 1-4), and we started by generating maps of nucleosome profiles at promoter-ORIs of actively transcribed genes (Figure 12d). We originally chose those types of genomic regions because nucleosomes located at active promoters are usually distributed in a very characteristic pattern (Figure 4, Introduction; Arya *et al*, 2010), which would validate the NuSA analysis.

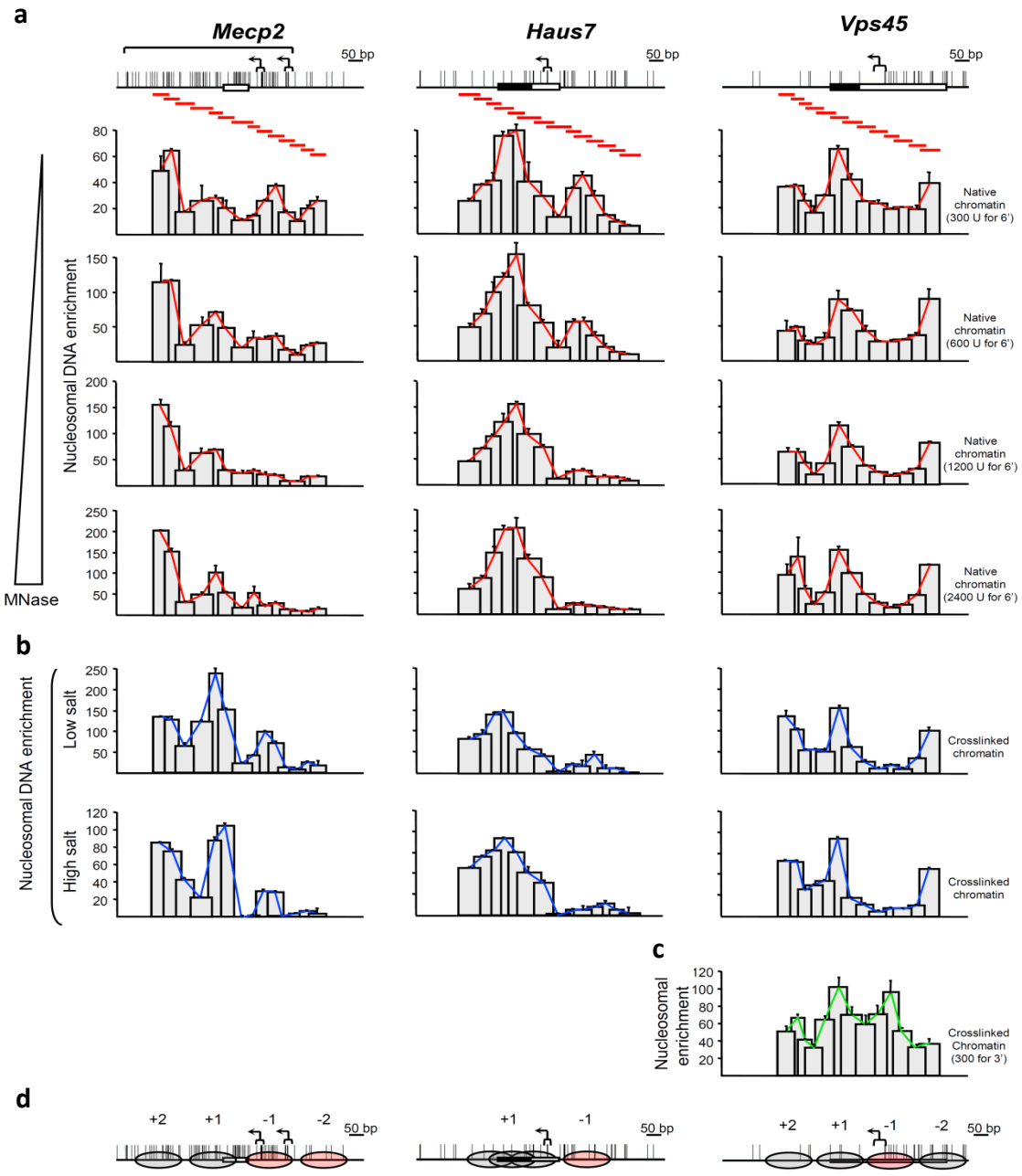


Figure 12: High-resolution mapping of nucleosomal landscapes at ORIs associated with active promoters. **(a)** Nucleosomal profiles obtained from NuSA upon digesting native chromatin with increasing amounts of MNase. The amount (U, units) of MNase used are indicated in each row. Genomic maps of each region are present on the upper part. The bracket above the MeCP2 promoter region indicates the position of the CpG island. Vertical lines on the maps point to the position of CpG dinucleotides. Arrows depict the location of the major TSS inside the promoters as well as the direction of transcription. Red lines below the maps correspond to the location of the overlapping amplicons used to generate the profiles (Supplementary Tables 1 to 3). The level of enrichment obtained for each amplicon is present in the corresponding histograms just below the lines. **(b)** Nucleosomal profiles obtained from the analysis of histone H3 ChIP with the same overlapping amplicons in both stringent (high-salt) and non-stringent (low-salt conditions). **(c)** Profile obtained at the Vps45 promoter region after performing NuSA on mononucleosomal DNA from crosslinked chromatin coupled with low MNase digestion. **(d)** Nucleosome positions, depicted by the NuSA analysis, at the same genomic maps present in (a). The histone particles are represented either as light-grey ovals, considered stable particles, or as red-ovals, considered labile particles. The nucleosomes are numbered according to their position relative to the TSS.

At the MeCP2 promoter region, which associates with a CpG island, we were able to detect four nucleosome particles with distinct characteristics. The two nucleosomes detected downstream of the promoter TSS, identified as +1 and +2, were stable through all the digestions employed with different units of MNase, showing that this particles are very tightly bound to the DNA molecule at those precise positions within the cell population. Furthermore, we also found two labile particles, one occluding the TSS (-1 nucleosome) and other just upstream of this element (-2 nucleosome). These labile nucleosomes could only be detected in H3 ChIPs and at lower MNase digestions, and the levels of nucleosomal protected DNA droped rapidly in harsher digestions with this enzyme (Figure 12, left panels).

At the Haus7 promoter region we also found both types of particles: labile and stable. The -1 nucleosome positioned a few base pairs upstream of the TSS showed a similar pattern as the labile nucleosomes present at MeCP2 region. This particle was also detected on lower MNase digestions and anti-H3 ChIPs. The other histone octamer detected in this region was more resistant to digestion with high amounts of MNase, although the nucleosome protected DNA region detected here was clearly larger than the expected 150 bp (Figure 12, middle panels). This means that this particle does not have a strong positioning in the cell population and its precise location varies some tenths of bp between different cells. This weak positioning was also detected by ChIP.

The genomic region analyzed at the Vps45 promoter revealed three strongly positioned non-labile nucleosomes, two of them located downstream of the TSS (+1 and +2) and one upstream (-2) (Figure 12, right panels). One peculiar feature detected in this first analysis was the 200 bp NFR surrounding the TSS. To make sure this zone was in fact nucleosome free we performed the same NuSA at the Vps45 region using crosslinked instead of native chromatin and digested it for three minutes instead of six. This second analysis revealed that the TSS region was not free of histone particles, unveiling the presence of the -1 nucleosome. Thus, the data obtained at the Vps45 promoter region shows the presence of a highly labile particle at the TSS. The -1 nucleosome was indeed much more labile than any one detected at the other two promoter regions analyzed.

Taken together, the results obtained at those regions are consistent with the notion that labile particles are essential for the accessibility of cis-regulatory elements such as TFs to the DNA molecule (Deal and Henikoff, 2010). The fact that these genomic regions are classified as active promoters points to a scenario were the fast turnover rates of the labile nucleosome particles at the TSS contributes to the maintenance of an active transcription process.

Outside promoter regions, the precise patterns of nucleosome positioning are gradually lost (Figure 4, Introduction; Arya *et al*., 2010; Valouev *et al*, 2011). With this in mind we tried to further validate these results by performing the same NuSA analysis at another previously characterized ORI located inside the intragenic region and several kb far from the promoter region of the *Scl7a14* gene (Table 1; Sequeira-Mendes *et al*., 2009).

Upon analyzing the *Scl7a14* intragenic region we found, as expected, a fuzzy pattern of nucleosome positioning with an almost constant level of histone density inside the region and minimal variations between peaks and valleys (Figure 13). The ORI present at this region, which was previously characterized as less efficient when compared with the other ORIs chosen for NuSA analysis (Table 1), is occupied by stably bound nucleosomes that slightly differ in their position from cell to cell inside the analyzed population. The detected profiles don't seem to vary greatly between different conditions. Nonetheless, histone patterns were more easily depicted from anti-H3 ChIP and higher MNase digestions.

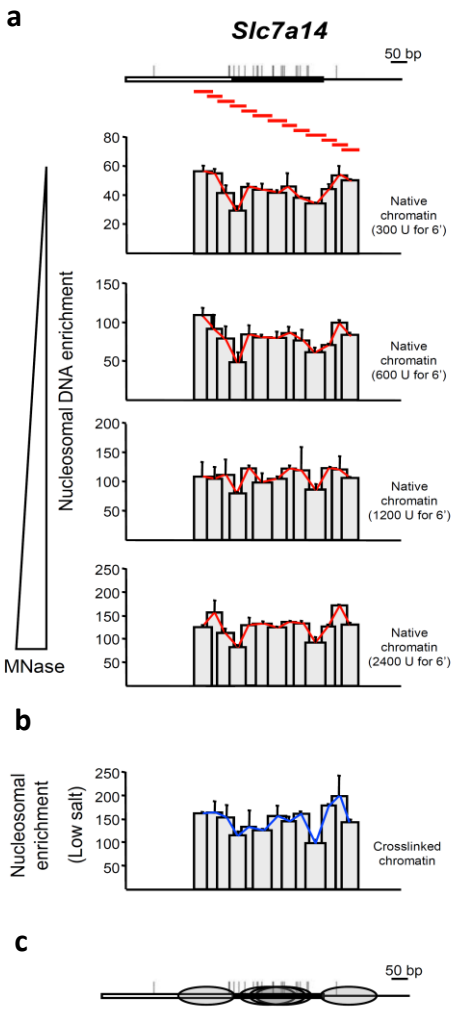


Figure 13: High-resolution mapping of the nucleosomal landscape at the *Scl7a14* intragenic ORI region. Same analysis as in Figure 12 performed at the ORI located within the transcribed region of the *Scl7a14* gene.

The diversity of nucleosomal landscapes found at the different regions analyzed evidenced the resolution power of the NuSA technique. Moreover, the fact that we were able to find nucleosome particles with such dissimilar levels of stability and with distinct positioning patterns inside cell populations further demonstrates that this technique is suitable for in depth analysis of local chromatin landscapes.

1.2. Comparative analysis between nucleosomal profiles and replication initiation sites.

We applied the same scanning method, using overlapping amplicons, to generate a detailed map of SNS profiles at the same ORI regions with the aim of correlating the replication initiation sites with individual nucleosomes. In order to obtain such a detailed high-resolution SNS profile we use three of the smallest fractions of nascent strands obtained after gradient fractionation (Figure 9, Material and Methods). The average size of these fractions ranged between 100 and 600 nt for the smallest one, 300 and 800 nt for the medium sized one and 400-1200 nt for the largest fraction. The level of detail obtained by employing the NuSA scanning method on small size fractions of nascent strands enabled us to reach the level of resolution of 150 bp, which is the average size of nucleosomal protected DNA (Figure 14).

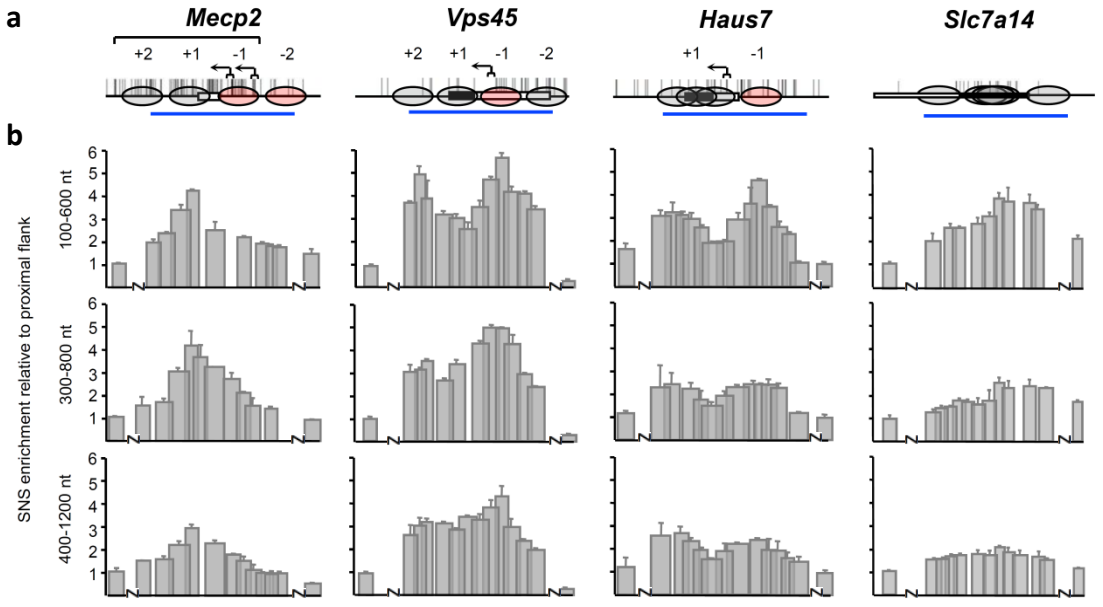


Figure 14: High resolution mapping of SNS enrichment at ORI sites. **(a)** Same genomic maps present in figures 12 and 13 depicting the nucleosome positions inferred by NuSA. **(b)** Histograms representing the enrichment levels of the short replication intermediates isolated from different fractions of the gradients (Figure 9a, Material and Methods) relative to a flanking region. Average sizes of the fractions used are indicated on the left side.

The enrichments found at scanned regions were in general between 2 and 6 times the level of the SNS signal obtained with the flanking amplicons (Supplementary Tables 1 to 5).Those amplicons were located usually 700 to 1200 bp outside each region of interest. Moreover, the enrichment levels are slightly higher in the shorter fractions used compared to the larger ones. This means that as we increase the length of the fragments analyzed we lose resolution. This was expected because the baseline level is adjacently located to the ORI regions from were bidirectional replication elongates. Indeed, the enrichment levels were much higher when we normalized the SNS levels to those obtained at a previously identified control region (Sequeira-Mendes *et al*, 2009; Supplementary Table 6) positioned several kb away from ORI sites (Supplementary Figure 1).

MeCP2 region was previously characterized as an efficient ORI associated with a CpG island (Table 1). The SNS scanning at this region detected a single peak of SNS enrichment. This peak was consistent between the three fractions analyzed and coincided with the location of the +1 nucleosome mapped by NuSA. It is also interesting to point out that this is the location of a non-labile nucleosome, meaning that inside this population ORI firing is preferentially taking place at the position of a stably bound histone particle.

At the other two non CGI promoter-ORIs the SNS profile was slightly different. At both these regions we found two distinct peaks of SNS enrichment separated by approximately 290 and 250 bp, respectively, coinciding both also with regions of positioned nucleosome particles, with a slight distinction from the MeCP2 region. At those regions one of the peaks overlap with a stable nucleosome particle located downstream of the TSS but the other peak is situated at the position marked by the -1 labile histone octamer. At both Vps45 and Haus7 regions the SNS enrichment signal flattens faster as SNS fraction size increases than at MeCP2 promoter-ORI region. This is consistent with previous data that depicts MeCP2 as a much more efficient ORI than either Vps45 or Haus7 (Table 1). Nevertheless, the major peak located at the -1 labile Vps45 nucleosome is still noticeable in the last fraction used for analysis. This is also in agreement with the same preliminary data, which shows that the ORI located at the Vps45 promoter is slightly more efficient than the one located at Haus7 (Table 1). Another interesting point that the analysis of the Haus7 region reveals is that the peak located downstream of the TSS is very broad and doesn't seem to be precisely positioned inside the cell population. The same happens with the nucleosome particle mapped at this exact location, suggesting that the SNS profile is accompanying the nucleosomal profile.

Upon analyzing the non promoter Scl7a14 region we obtained a SNS profile that was not very well defined, indicating that at this ORI region there is a high variability of

replication initiation sites within the cell population. A slight SNS peak at the position of the central nucleosome can be detected at the smaller SNS fractions. However this enrichment is rapidly lost in the larger SNS fractions, which was expected as this region was previously classified as a low efficient ORI. Also, the fact that once more the SNS enrichment patterns accompany the nucleosomal profiles, even at low efficient ORIs, further strengthens the idea of a correlation between the two.

1.3. Correlation between transcription efficiency, nucleosomal architecture and replication initiation sites at promoter-ORIs

The data from the NuSA and SNS profiles at ORI regions suggests a link between nucleosome positioning and ORI firing efficiency, pointing to a scenario were replication initiation and nucleosome positioning are tightly linked. This scenario predicts that shifts in the local chromatin conformation around an ORI would be reflected in the patterns of replication initiation. With this in mind we decided to test if variations in the nucleosomal architecture related to changes in transcription levels at promoter-ORIs could influence the SNS profiles at those sites. We did this by analyzing SNS profiles and nucleosome maps by NuSA in mES cells and NIH3T3 fibroblasts, where the genes *Haus7* and *Vps45* are upregulated (Figure 15a). As expected, changes in the expression levels of *Haus7* and *Vps45* were accompanied by changes in the nucleosomal landscape of their promoter regions (Figure 15a and c, central and right panels). At the *Haus7* promoter region we found that the protected DNA region corresponding to the +1 nucleosome is not as broad in NIH3T3 MEFs as in mES cells. This suggests that in the fibroblast cell population this nucleosome is more precisely positioned than in the other cell line. This shift is important because it leaves the TSS inside a NFR, in agreement with the observed increase in the levels of mRNA expression likely due to higher accessibility of TFs or the transcription initiation machinery to this region. Furthermore, the -1 nucleosome, which was described before as a labile particle in mES cells seems to be even more unstable in fibroblasts, probably contributing to an even larger NFR near the TSS.

A similar scenario was found at the promoter of the *Vps45*, where the labile -1 nucleosome also shows a slightly higher sensitivity to MNase digestion. Moreover, we detected a slight shift in the positions of the +1 and the -1 nucleosomes that leaves the TSS just at the DNA region exposed between those two histone particles. This shift in the positioning, together with the higher instability of the -1 nucleosome is also in agreement with the increase in transcription levels of the *Vps45* gene in NIH3T3 cells. As a control, we analysed the promoter-ORI associated to the housekeeping *MeCP2* gene that is

similarly expressed in both cell types (Figure 15a). Accordingly, NuSA analysis found similar profiles of nucleosome positioning and occupancy in both cell types (Figure 15b and c, left panels). Interestingly, when we compared the nucleosome maps with the SNS profiles, we detected a clear correlation between the chromatin landscape and the replication intermediate profile at each cell type (Figure 15d).

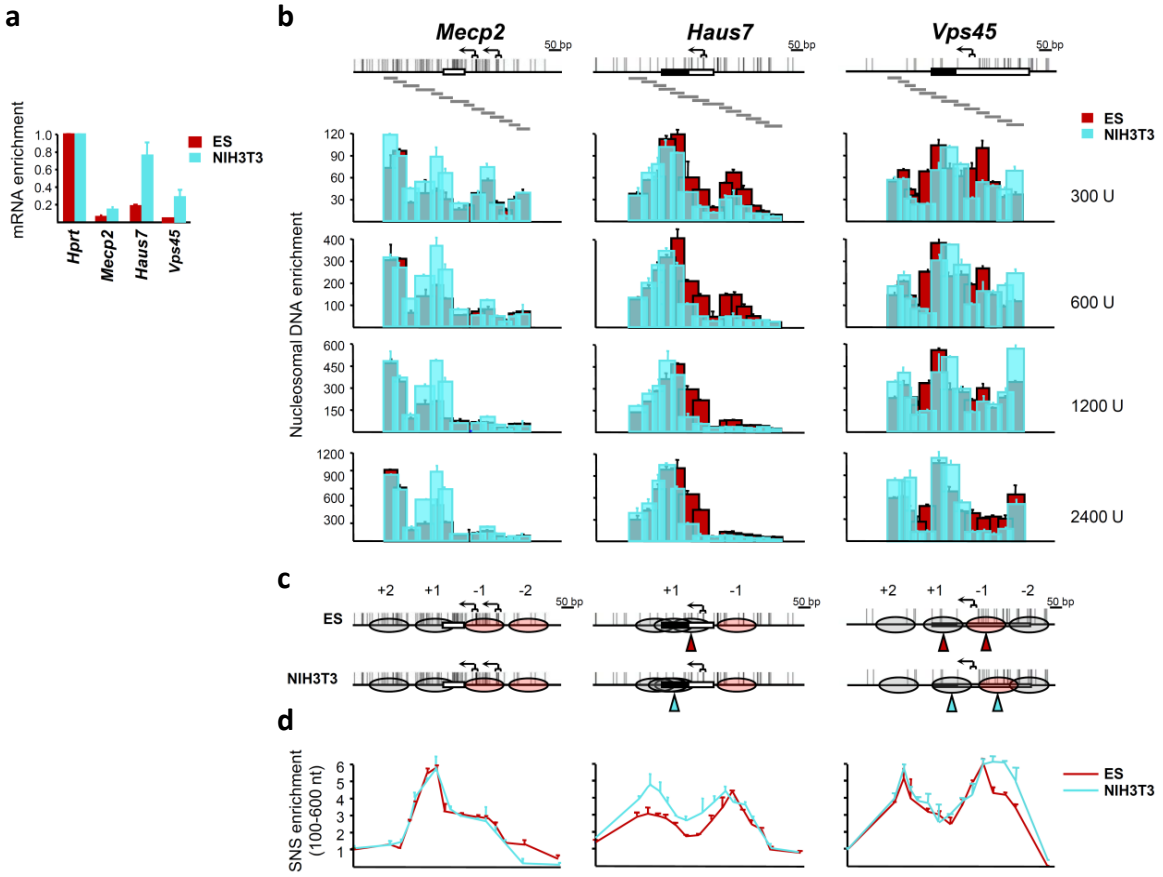


Figure 15: Analysis of mRNA levels, NuSA and SNS enrichment profiles at the three promoter associated ORIs. **(a)** RT-qPCR analysis of mRNA levels (relative to Hprt mRNA) at the three promoter regions. The measurements were performed in both mES cells (dark-red histograms) and NIH3T3 MEFs (light-blue histograms). **(b)** NuSA maps generated for both mouse cell types. Mononucleosomal DNA is derived from either native (*MeCp2* and *Haus7* regions) or crosslinked (*Vps45* region) chromatin. Amount of MNase used in each treatment is indicated on the right. **(c)** Genomic maps depicting nucleosome positions inferred from NuSA analysis. Shifts in nucleosome positions between cell lines are indicated by color-coded triangles. **(d)** SNS enrichment levels relative to flanking control regions. Due to the higher sensibility level, only smaller fractions (100-600 nt) were used to detect differences between the patterns of both cell lines.

In the case of the *Haus7* promoter-ORI region, a sharper peak of SNS enrichment was detected at the location of the +1 nucleosome in NIH3T3 cells, which occupies a narrower region at this cell type, coinciding with a better positioned nucleosome within this cell population. The results obtained at the *Vps45* promoter-ORI region pointed to a similar scenario; the sharp peak of SNS enrichment found in mES cells at the position of the labile -1 nucleosome was much broader in NIH3T3 cells, accompanying the detected

shift in the position of this particle in this cell type. In contrast, we detected no changes in the SNS profile at the initiation site located at the +2 nucleosome coinciding with the invariant position of this nucleosome in both cell lines. Similar results were found at the *MeCP2* promoter-ORI region, in which no significant changes were detected between cell types neither in the nucleosomal landscape, nor in the SNS peaks. Besides reinforcing the notion that local shifts in the nucleosomal landscape at ORI regions are accompanied by parallel changes in the SNS profiles, these findings also indicate that the differences observed were not due to cell type differences but rather to the specific changes that take place in local nucleosome configuration.

1.4. Nucleosomal landscape profiles upon ORI activation

The data so far suggested a correlation between replication patterns and the nucleosome landscape at ORI sites. Peaks of SNS enrichment inside those regions tend to occur at sites occupied by nucleosomes, either labile or more stably bound. It is likely that these histone particles are evicted from chromatin as replication activation unfolds because the pre-RCs need to be loaded onto the DNA in order for ORI firing to occur, a process which demands accessibility to the DNA molecule (Figure 1, Introduction). We sought out to test this by analyzing the nucleosomal landscape at one of the best characterized human replication origins, the LaminB2 ORI (Abdurashidova *et al*, 2000). This ORI has been extensively studied and the region where ORC binds has been precisely mapped by ChIP as well as footprinting protection assays (Ladenburger *et al*, 2002; Abdurashidova *et al*, 2003). It encompasses a 1.2 kb region that has the capability of promoting DNA replication initiation upon insertion into an ectopic site in human cell lines, and the precise site of replication initiation was determined at a single nucleotide resolution (Abdurashidova *et al*, 2000; Paixao *et al*, 2004). Furthermore, it is located near a CpG island associated with the promoter of the neighboring gene *Timm13*. Another advantage of this ORI is that it fires quickly upon entry into S phase (Swarnalatha *et al*, 2012), thus enabling us to study in detail the changes in the chromatin landscape at HeLa cells synchronized in early-S phase of the cell cycle.

The cells were synchronized at the end point of G1 cell cycle phase, just before entry into S phase (Figure 16a). This was performed by a double thymidine block, as described in Materials and Methods. The mononucleosome protected DNA was purified from crosslinked chromatin obtained from the same number of cells before the block release (Figure 16b, dark blue histograms) and 45 minutes after release (Figure 16b, light blue histograms). Additionally we mapped the nucleosomes from exponentially growing

cells in order to compare NuSA profiles from synchronized populations to asynchronic HeLa cells (Figure 16c).

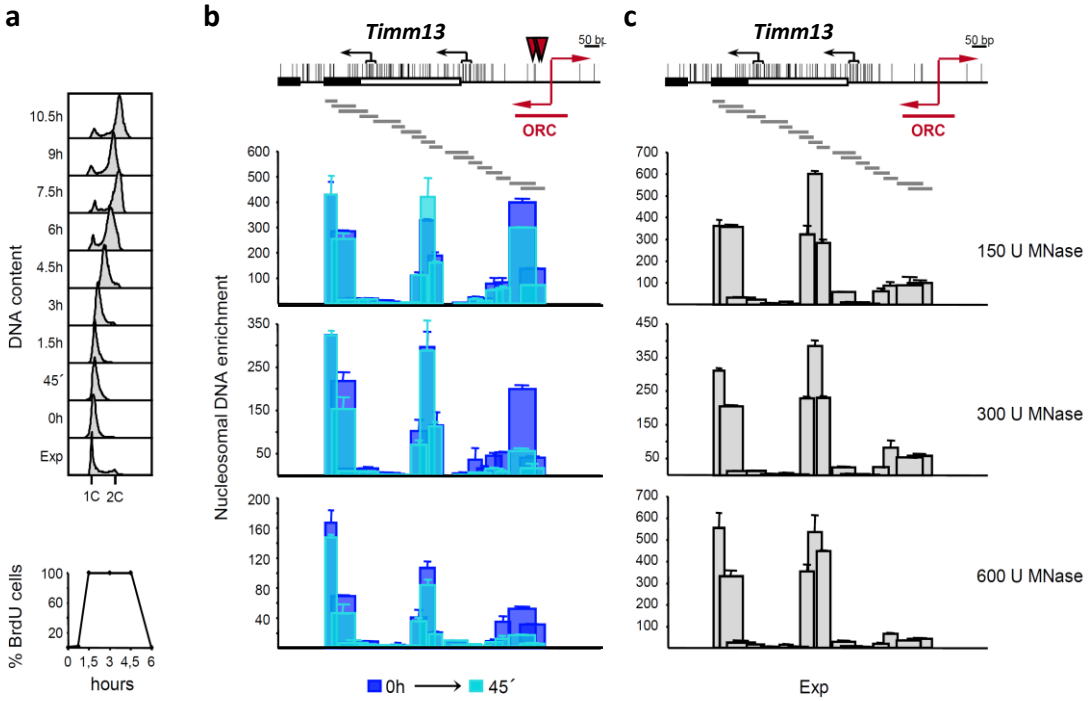


Figure 16: Nucleosomal architecture dynamics at the LaminB2 ORI in G1 and early S-phase of the cell cycle. **(a)** Upper panel depicts HeLa cell cycle progression analyzed by flow cytometry at the different time points (from blocked-0h to 10.5h after release) as well as exponentially growing cells (Exp). Lower panel depicts the level of BrdU incorporation measured at the same time points. **(b)** Nucleosome profiles obtained by performing NuSA on crosslinked chromatin derived from the same number of HeLa cells blocked at G1/S phase (0h, dark-blue histograms) and early-S phase (45' after block release, light-blue histograms). Genomic maps are indicated above and contain the same features previously mentioned for the mouse ORI maps. The red triangles presented above the maps indicate the exact location of ORC2, ORC1 and CDC6 proteins, according to Abdurashidova *et al*, 2003. **(c)** NuSA profiles at the control exponentially growing HeLa cells.

No significant DNA synthesis was detected by flow cytometry analysis or BrdU incorporation analysis at the first two time points of the experiment (Figure 16a, 0h and 45'). NuSA analysis on exponentially growing cells (Exp, grey histograms) revealed three nucleosome particles at this region: two stable particles located near the two major TSSs of *Timm13* gene and one labile nucleosome at the region where ORC was previously mapped. Upon performing the same analysis on synchronized cell populations we found no significant changes in the positions of the histone particles, either at blocked cells (0h time point, dark-blue histograms) or at early S phase cells (45' minutes time point, light-blue histograms). The levels of occupancy also didn't vary between the two time points for the two nucleosomes positioned near the TSSs. The major difference was found at the protected region located at the ORC binding site. Here, NuSA analysis revealed that upon exiting G1 phase, this particle become more labile. This might be linked to the previously mentioned footprinting signal detected at this region in G1/S blocked cells, possibly

related to the recruitment of the replication complex at this precise location (Figure 16b, red triangles; Abdurashidova *et al*, 2003). These findings suggest that, at this ORI site, histone octamers seem to suffer remodeling prior to the onset of replication initiation, supporting a model of dynamic turnover between components of the chromatin structure and the replication machinery.

2. Functional analysis of the cellular responses to alterations in chromatin conformation

Chromatin is the template of all genomic transactions occurring in eukaryotic cells. This element encompasses many different features and its structural integrity, as well as its plasticity and ability for local or global remodelling are key features for safeguarding the balance between all genomic processes such as transcription, replication, recombination and repair. In the first part of this chapter of results we found that slight changes in the local nucleosomal architecture can be linked to variations both in transcription and replication. Furthermore, nucleosome remodelling is also essential to maintain the spatial and temporal balance of these fundamental processes. We therefore decided to increase the scope of our study from a local scenario to a global one, analysing the relationship between chromatin organization, replication and transcription. Specifically, our objective was deciphering how genome-wide alterations in the chromatin conformation affect genomic functions and what consequences those possible changes could have on cell viability.

2.1. DNA replication landscape of cells with reduced nucleosome numbers

In order to characterize the consequences of a global decrease in chromatin nucleosome level on the DNA replication program we first took advantage of a model consisting of a primary MEF cells derived from embryos knockout (KO) for the *Hmgb1* gene (HMGB1-KO). The HMGB1 protein (described in detail in section 3.3 of the Introduction chapter) is a chromatin architectural protein that has the capability of binding to the nucleosomes near the entry/exit point of the DNA in order to promote nucleosome sliding and in this way contributing to the accessibility of TFs to promoter regions (Figure 6; Bonaldi *et al*., 2002; Agresti and Bianchi, 2003; Travers, 2003; Ueda *et al*., 2004; Joshi *et al*., 2012). Interestingly, HMGB1 depletion has been associated with a reduction in nucleosome occupancies, as well as with an overall increase in mRNA transcripts both in the yeast *S. cerevisiae* and in mammalian cells (Celona *et al*., 2011).

We started by isolating small replication intermediates from both WT and HMGB1-KO MEFs as previously described and then sequencing those fragments in high depth (Short Nascent Strands-Sequencing). The outcome of the SNS-seq allowed us to evaluate the genome-wide landscape of replication initiation within the cell population, generating genomic maps in which the peaks of enrichment represent the most likely sites of ORI firing (Prioleau and MacAlpine, 2016). The analysis of the SNS-seq datasets

obtained from all the MEF replicates showed small variations between WT and HMGB1-depleted cells (Figure 17a; additional region also shown in Figure 22).

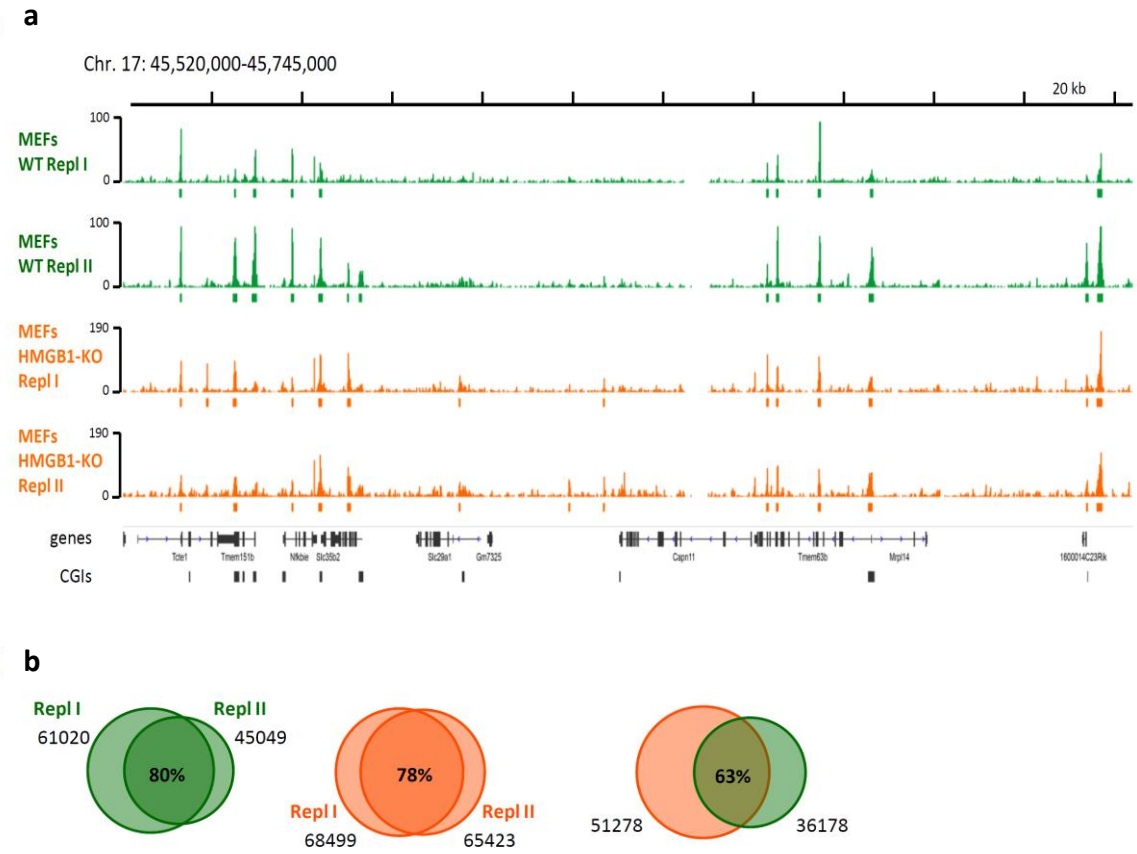


Figure 17: Replication initiation landscape WT and HMGB1-KO MEFs. **(a)** Representative IGV snapshot of the SNS-seq coverage and ORI positions at a region of mouse chromosome 17. Two replicates, indicated as Repl I and Repl II were analyzed separately for each condition. Results from WT MEFs are depicted in green while HMGB1-KO MEFs are presented in orange. SNS samples were derived from sucrose gradient fractions containing replication intermediates ranging between 300-1500 nt in size. Coloured rectangles below each track mark the positions of the ORIs identified in each cell population. Marked below are also the positions of genes and CGIs present at this region. An additional representative region is shown in Figure 22. **(b)** Venn diagrams depicting the overlaps of common ORIs identified between distinct replicates and genotypes (colour code is maintained for WT and HMGB1-KO MEFs). ORIs were defined by applying the SNS-scan algorithm (Picard *et al*., 2014), using as genome segmentation the early and late-timing domains (Hiratani *et al*., 2008) to account for the differences in read coverage between domains. See Supplementary Table 7 and Materials and Methods for details.

The ORI overlap between the two conditions was of 63% (Figure 17b). Taking in to consideration the striking similarities observed between the profiles of WT and HMGB1-KO SNS-Seq, as illustrated in figures 17a and 22, we anticipated this percentage to be slightly higher. Nonetheless, upon calculating the overlap between replicates of the same cell type, we found that this percentage was never higher than 80% (78% between HMGB1-KO replicates and 80% between WT replicates; Figure 17b). This likely reflects the technical limitation of the SNS-seq experiments and computational analysis; some

ORI peaks might not be detected by the peak-calling algorithm when the read-coverage is not sufficient in certain regions giving raise to false negatives. Taking this into account, we concluded that the replication initiation landscape was overall conserved in a context of reduced nucleosome numbers in chromatin.

As an independent approach to validate these results, we preformed single-molecule analysis of replication intermediates by stretching DNA fibers after labelling the cells with consecutive pulses of thymidine analogues. This method allowed us to quantify and characterize specific DNA replication parameters such as inter-origin distances (IOD), fork rates, and fork asymmetry at individual DNA molecules (see Figure 11b for a representation of the most common structures that can be identified using this method). Both scatter plots and frequency distributions showed no significant differences between cell types in terms of IODs, with median values and average frequency distances of around 100 kb among adjacent ORIs (Figure 18a and c). These results were in agreement with those obtained by SNS-seq, which show limited changes in ORI landscapes. As a side note, medians rather than means are represented throughout the different plots in this results chapter because they are less sensitive to presence of outliers and as such more suitable for the analysis of DNA fiber experiments (Técher *et al*., 2013).

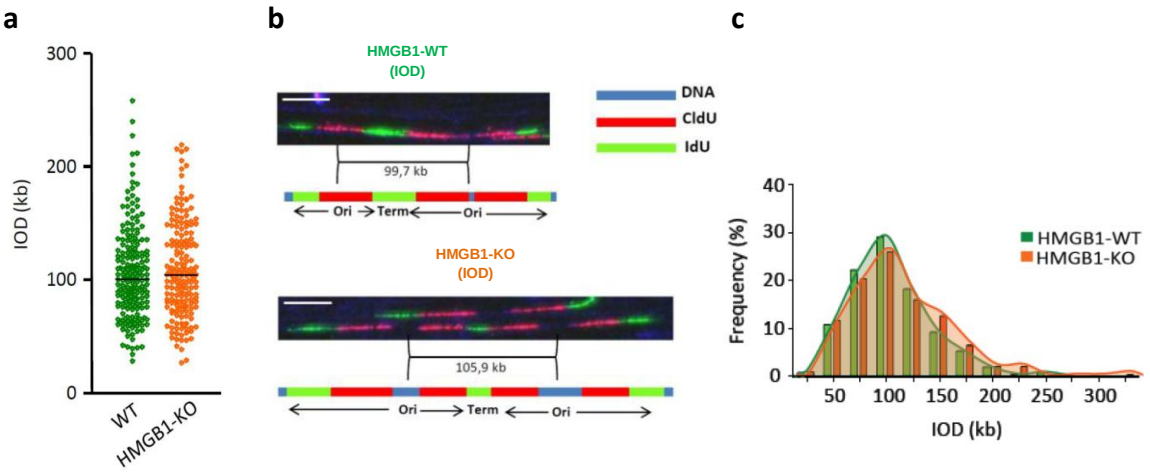


Figure 18: Analysis of distances between replication initiation origins (IODs) in WT and HMGB1-KO MEFs. **(a)** Scatter plot showing the IODs distribution, in kb, of both WT and KO MEFs. Black lines represent the median values. Data are pooled from three replicate experiments. **(b)** Representative example of DNA fibers labelled sequentially for 20 minutes with CldU (red) and IdU (green) used to estimate IODs with the corresponding schematic interpretation below each image. *Ori* corresponds to origin of replication point and *Term* to replication termination point. White scale bar corresponds to 10 µm. **(c)** Frequency distribution of the same IODs present in a. Data are derived from three independent experiments. Numerical values are present in Supplementary Table 8.

When analysing the velocity of the replication forks, however, we noticed a highly significant increase in fork rates ($p<0.0001$) in the cells with reduced histone content

(Figure 19). Replication fork velocities change from 1,49 kb/min in WT to 1,84 kb/min in HMGB1-KO cells, as both the distribution in the scatter plots and the frequencies of individual fork rates indicated (Figure 19a and c).

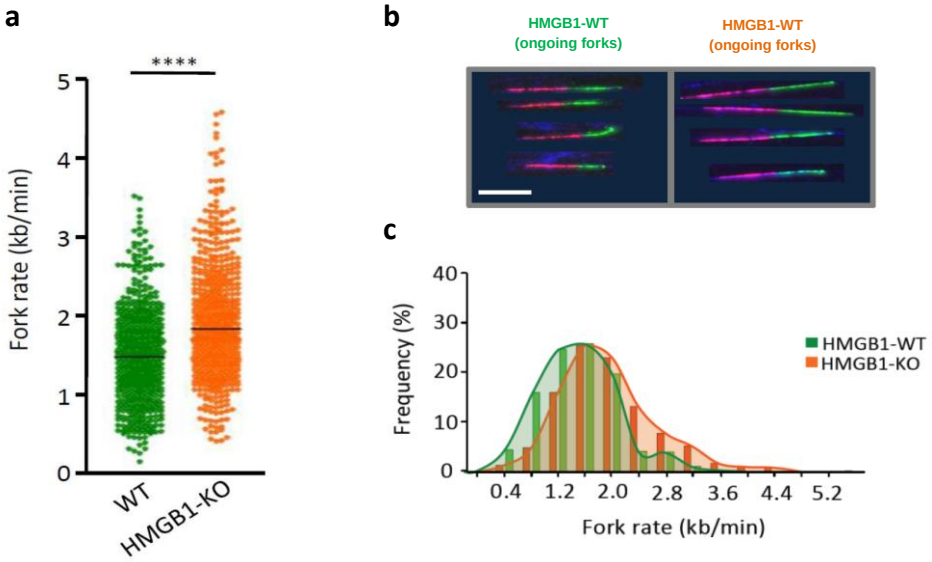


Figure 19: Analysis of replication fork rates in WT and HMGB1-KO MEFs. **(a)** Scatter plot showing the fork rate distribution, in kb/min, of both WT and KO MEFs. Black lines represent the median values. ****p<0.0001. **(b)** Representative example of ongoing forks labelled sequentially for 20 min with CldU (red) and IdU (green) used to estimate fork velocities. White scale bar corresponds to 10 µm. **(c)** Frequency distribution of the same fork rates present in a. Data are derived from three independent experiments. Numerical values are present in Supplementary Table 8.

This result was somewhat unexpected, as usually the changes in fork velocity are accompanied by alterations in the distances between ORIs. We therefore decided to check whether this increase in velocities affect the stability of the sister forks. Thus we measured the global symmetry between left and right green tracks (representing IdU incorporation during the second pulse) present on each side of 1st labelled ORIs detected during DNA fibers analysis (Figure 20b; see also Figure 11 to view an example of 1st labelled ORIs). This approach for measuring fork instability has been previously described and widely used in many studies involving single molecule DNA fiber analysis (Conti *et al*, 2007; Rao *et al*, 2007; Tuduri *et al*, 2009; Fu *et al*, 2014). On stretched DNA molecules, a stable replication bubble labelled with halogenated deoxynucleotides will appear as a symmetric 1st labelled ORI due to the fact that both forks move bidirectionally at approximately the same speed (Figure 20b, inside red lines). Any genomic event that perturbs one of the forks during the second pulse will result in an asymmetric replication bubble (Figure 20b, outside red lines; Rao *et al*, 2007). This enabled us to measure the average differences within the cell population in terms of global asymmetry (Figure 20a). Furthermore, a single ORI signal in which one of the forks is 33% shorter than the other is

usually considered as stalled (Conti *et al*, 2007; Fu *et al*, 2014). Therefore, by plotting the length of left *vs* right fork emanating from the same ORI we could estimate the average percentage of stalled forks in WT and HMGB1-KO cells (Figure 20b and c).

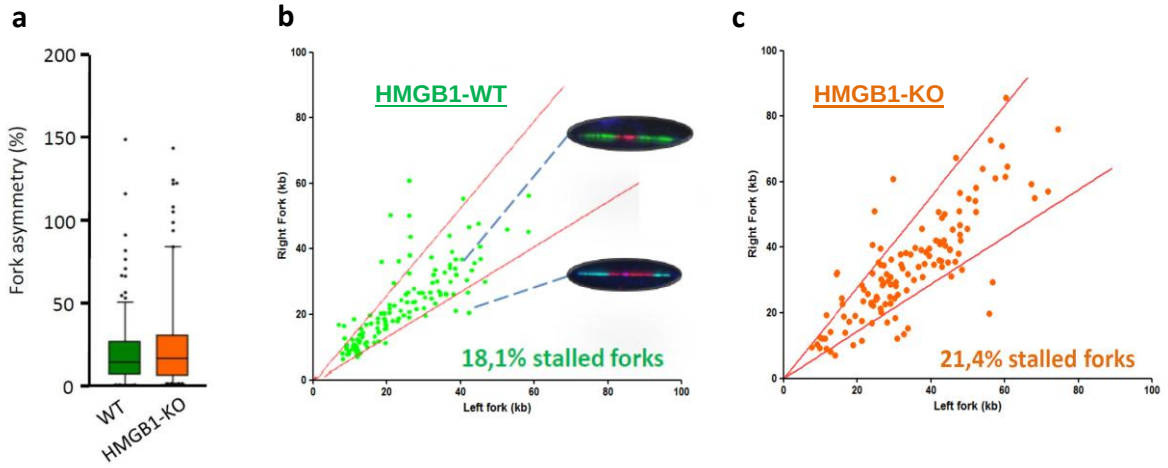


Figure 20: Analysis of replication fork stability in WT and HMGB1-KO MEFs. **(a)** Box plot representing global fork asymmetry levels in both normal and KO MEFs. Fork asymmetry is expressed as the ratio of the longest distance covered to the shortest, for each pair of sister replication forks. Median values are indicated. Data not included between the whiskers are plotted as outliers (black dots). **(b)** and **(c)** represent scatter plots depicting normal and stalled forks detected by DNA fiber analysis at both WT and HMGB1-KO cells. Only those tracks showing differences greater that 33% between the right and the left moving fork (outside the central area delimited by the red lines) were considered as stalled. In b two representative examples of normal (inside the central area) and stalled forks (outside central area) are shown. Data derived from three independent experiments. Numerical values are shown in Supplementary Table 8.

Upon analyzing the global levels of asymmetry we found no significant changes between WT and HMGB1-KO MEFs, with median values of asymmetry between left and right sides of the forks varying around 15-17% for both conditions (Figure 20a). Likewise, the overall percentage of stalled forks was quite similar between both conditions, with WT MEFs encompassing 18,1% of arrested replication forks and HMGB1-KO presenting a value of 21,4%.

To analyse if this reduction in nucleosome occupancy had any influence on the cell cycle we preformed a detailed analysis of the S-phase by first, quantifying the incorporation of thymidine analogues after a short pulse by flow cytometry and, second, by scoring the characteristic patterns of replication factories along S-phase (Nakamura *et al*., 1986) by Immunofluorescence (Figure 21).

Flow cytometry analysis of IdU incorporation showed no major differences in the percentage of cells in S-phase, with 22,4% and 24,8% for WT and KO-HMGB1 MEFs, respectively. Similarly, no differences were found between cell types by scoring early-, middle- and late-S phase patterns of EdU incorporation by Immunofluorecence (Figure 21c). Likewise no differences were detected in the percentage of cells at the other phases

of the cell cycle (Figure 21b, G1 and G2) with either approach. Altogether, the analysis of DNA replication initiation profiles, replication dynamics, and S-phase progression of MEFs lacking HMGB1 indicates that the reduction in nucleosome occupancies caused by the lack of this factor allow faster fork velocities without compromising fork stability or cell cycle advance.

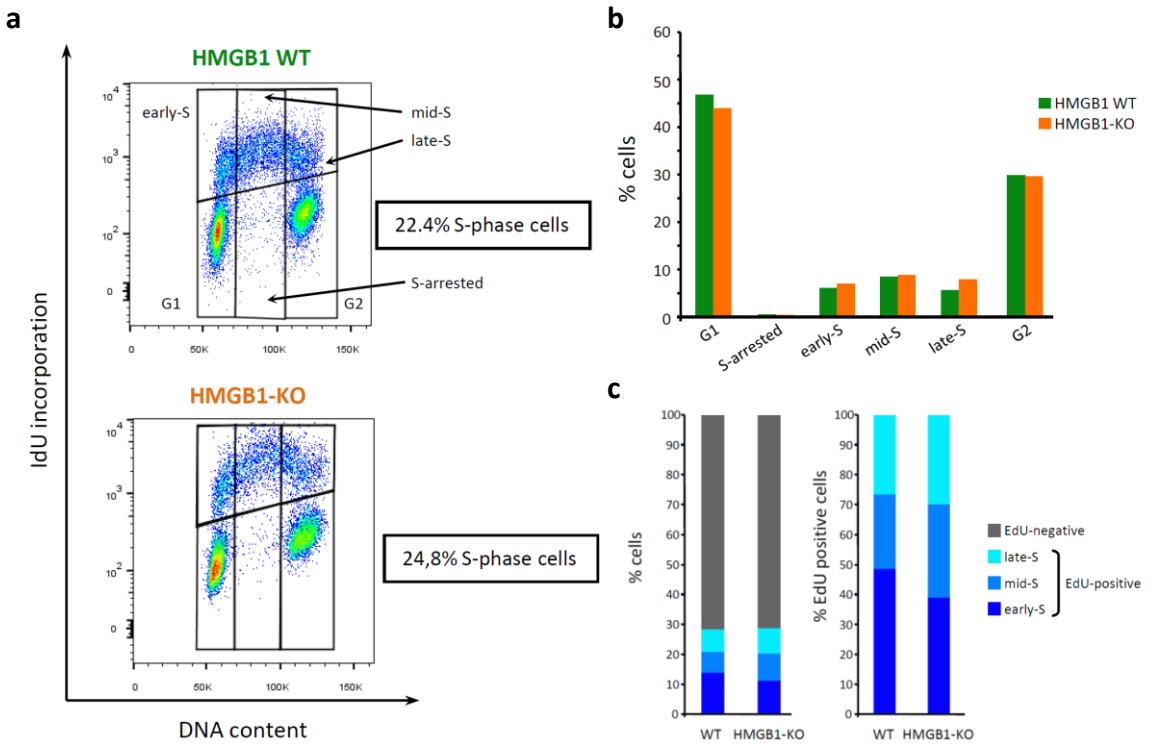


Figure 21: Detailed cell cycle analysis of WT and HMGB1-KO MEFs. **(a)** Cell cycle and IdU incorporation analysis by flow cytometry of both cell types. Plots were divided in six regions, each one representing a different cell cycle phase. G1 is represented in the lower left region as cells with low DNA content and low IdU signal (below baseline). S-phase is represented in the three upper regions where IdU incorporation signal was above the baseline. The division between early-, mid- and late-S was established according to DNA content. S-arrested cells present intermediate levels of DNA content, but no IdU Incorporation signal. G2 is represented in the lower right region as cells with the higher DNA content but no IdU incorporation signal. Total percentages of S-phase cells are indicated. **(b)** Histogram representing the percentages obtained from the analysis shown in a. **(c)** Histograms representing detailed analysis replication factories by scoring EdU patterns. Left histogram shows the percentage of EdU-positive *vs* EdU-negative cells. Right histogram shows the percentage of cells in distinct points of the S-phase according to the EdU replication-foci patterns described by Nakamura *et al*, 1986. Data are representative from two independent experiments.

2.2. DNA REPLICATION LANDSCAPE OF CELLS WITH ALTERED CHROMATIN COMPACTION

We next analysed the impact of alterations in chromatin compaction on the DNA replication program. To accomplish this we made use of another available model consisting of mouse ES cells (mES) lacking three of the somatic variants of the H1 gene (H1-TKO), specifically subtypes H1c, H1d and H1e (Fan *et al*., 2003; Fan *et al*., 2005). H1-TKO mES present a reduction in the levels of H1 in chromatin resulting in the

equivalent of a molar ratio of one histone H1 molecule per four nucleosome particles. Consequently, H1-TKO cells display a globally less compact chromatin structure and a slight reduction of average linker DNA size, although only the expression of a small subset of genes seems to be altered (Fan *et al*., 2003; Fan *et al*., 2005; Murga *et al*., 2007; Geeven *et al*., 2015). SNS-seq analysis of these cells showed considerable differences not only between H1-TKO and normal WT mES cells, but also between these mutant cells and our previously analyzed MEFs lacking HMGB1 (Figure 22).

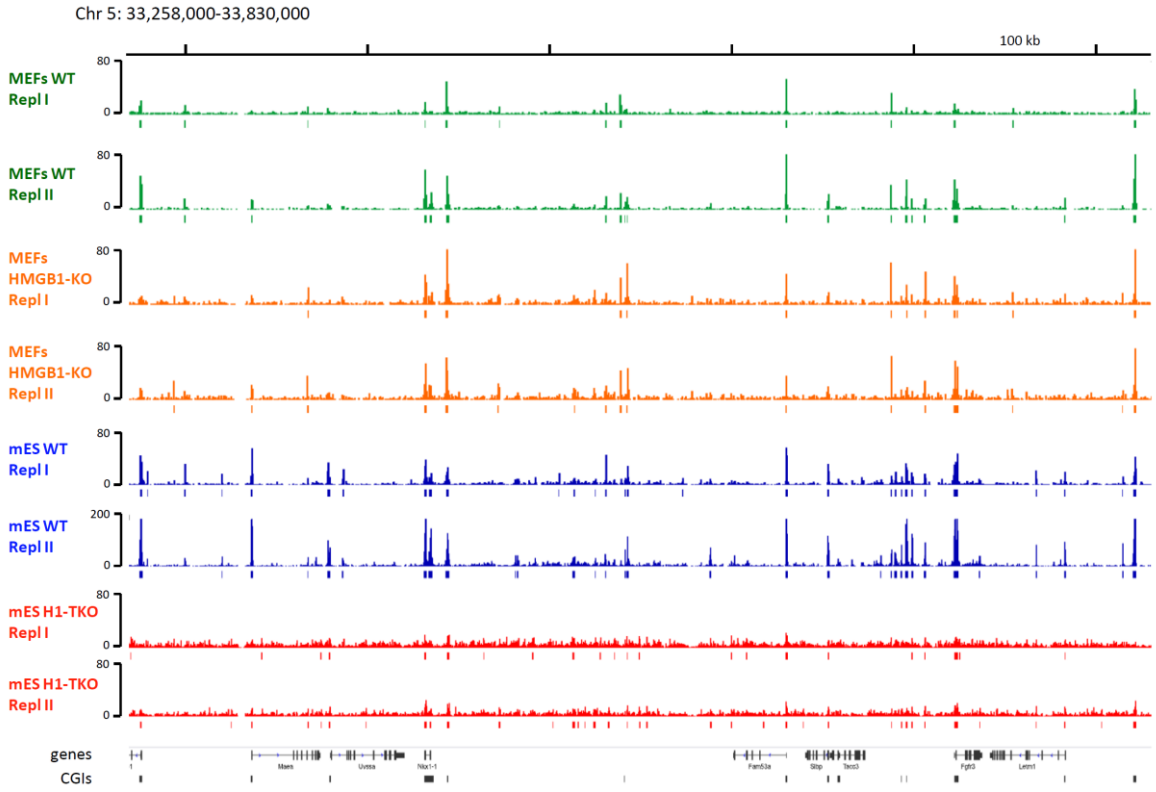


Figure 22: Replication initiation landscape in WT and H1-TKO mES and WT and HMGB1-KO MEFs. Representative IGV snapshot of the SNS-seq coverage and ORI positions at a region of mouse chromosome 5. Two replicates, indicated as Repl I and Repl II were analyzed separately for each condition. Results from WT MEFs are depicted in green, HMGB1-KO MEFs in orange, WT mES cells in blue and H1-TKO mES in red. SNS samples were derived from sucrose gradient fractions containing replication intermediates ranging between 300-1500 nt in size. Coloured rectangles below each track mark the positions of the ORIs identified in each cell population. The other symbols are as in Figure 17.

Although the most prominent SNS enrichments were still detectable in mES cells with reduced amounts of histone H1, the widespread accumulation of short replication intermediates prevent us from applying a reliable peak-calling algorithm to identify preferential sites of replication initiation. Peak quality (signal-to-noise ratio) for all SNS-seq replicates was assessed at our laboratory by performing a Fraction of mapped Reads in Peak regions (FriP) analysis (Supplementary Figure 2).

To confirm that the global DNA replication landscape was indeed altered in H1 mutant cells we carried out the same set of DNA fiber analysis experiments that we employed for MEF cells. The results from SNS-seq indicate a drastic change in replication initiation sites in cells with lower levels of H1, so we expected IOD analysis to reveal this variation as well. Indeed, upon analyzing the average origin distances in WT and H1-TKO cells we detected a significant decrease in IOD distributions when the normal levels of H1 were reduced (Figure 23).

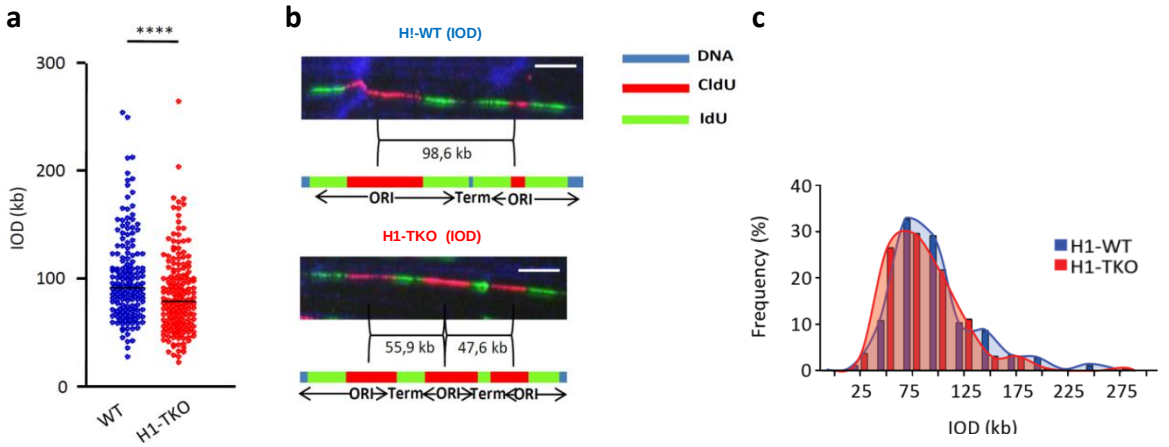


Figure 23: Analysis of inter-origin distances (IODs) in WT and H1-TKO mES cells. **(a)** Scatter plot showing the IODs distribution, in kb, of both WT and KO mES cells. Black lines represent the median values. ****p<0.0001 **(b)** Representative example of DNA fibers labelled sequentially for 20 min with CldU (red) and IdU (green) used to estimate IODs with the corresponding schematic interpretation below each image. *Ori* corresponds to origin of replication point and *Term* to replication termination point. White scale bar corresponds to 10 µm. **(c)** Frequency distribution of the same IODs present in a. Data are derived from three independent experiments. Numerical values are shown in Supplementary Table 9.

The observed decrease in IOD in H1-TKO cells relative to their WT counterparts was of about 15% (Figure 23a and c), indicating that more replication origins are activated when regular levels of linker histones are impaired. This result is in agreement with the genome-wide increase of SNS signal emanating from the genome of H1-TKO cells (Figure 22). This IOD reduction was accompanied by a decrease in average fork velocities of over 30% (Figure 24a and c), indicating that the architectural chromatin defects in H1-TKO cells have a severe impact on the normal progression of the replication machinery. The shorter replication tracks found in H1-TKO (Figure 24b) reflect slower DNA synthesis but can also be the result of fork stalling, which can also be detected as short labelled tracks (Fu *et al*., 2015). We therefore measured the fork asymmetry levels, as well as the percentage of stalled forks, in H1-TKO cells and its WT counterparts (Figure 25).

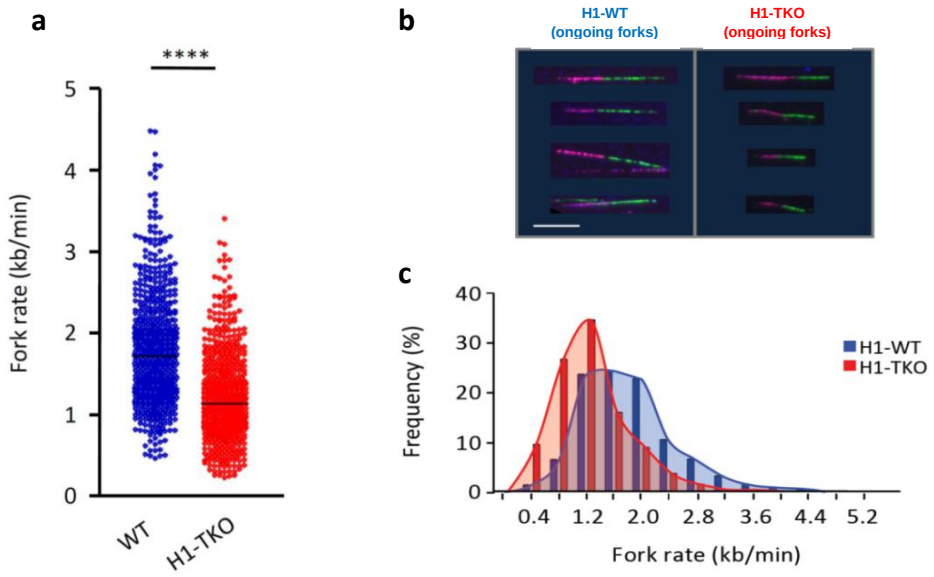


Figure 24: Analysis of replication fork rates in WT and H1-KO mES cells. **(a)** Scatter plot showing the fork rate distribution, in kb/min, in both cell types. Black lines represent the median values. ****p<0.0001. **(b)** Representative example of ongoing forks labelled sequentially for 20 min with CldU (red) and IdU (green) used to estimate fork velocities. White scale bar corresponds to 10 µm. **(c)** Frequency distribution of the same fork rates present in a. Data are derived from three independent experiments. Numerical values are shown in Supplementary Table 9.

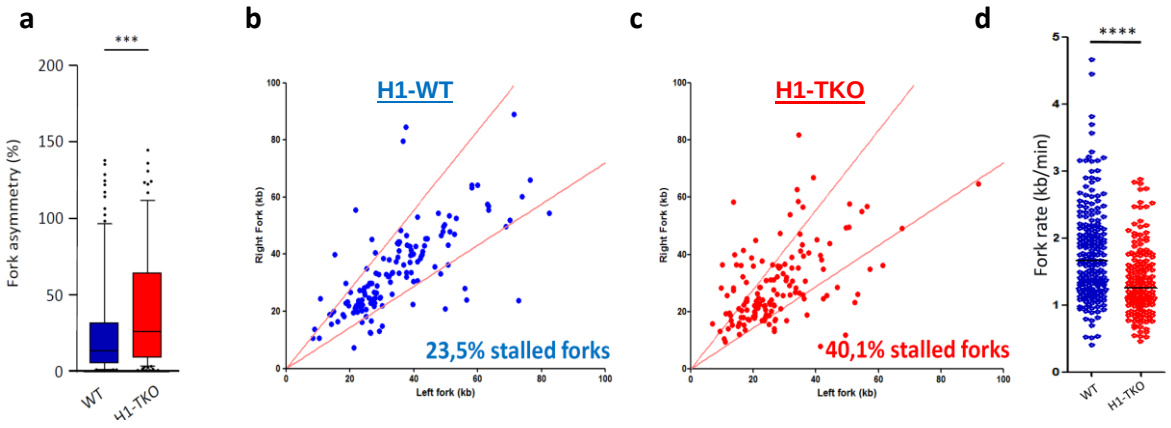


Figure 25: Analysis of replication fork stability in WT and H1-KO mES cells. **(a)** Box plot representing global fork asymmetry levels in both cell types. Fork asymmetry is expressed as the ratio of the longest distance covered to the shortest, for each pair of sister replication forks. Median values are indicated. Data not included between the whiskers are plotted as outliers (black dots). ***p<0.001 **(b)** and **(c)** represent scatter plots depicting normal and stalled forks detected in DNA fiber analysis at both WT and HMGB1-KO cells. Only those tracks showing differences greater that 33% between the right and the left moving fork (outside the central area delimited by the red lines) were considered as stalled. (d) Scatter plot representing the fork rate analysis of the non-stalled forks depicted in b and c. Data derived from three independent experiments. Numerical values are shown in Supplementary Table 9.

We found that, indeed, both fork staling and asymmetry levels were altered in the H1-TKO cells, further demonstrating that the chromatin conformation alterations caused by reduced levels of histone H1 had a deep impact on the replication program. To ensure that the decrease in fork rates that we detected in these cells was not due to the higher

percentage of stalled forks, which might be skewing the first measurement, we analyzed separately the velocities of the forks considered not stalled (Figure 25d, central area inside the red lines). This complementary analysis showed that the significant drop in the average fork velocities was still noticeable ($p<0.0001$) even when we only analyse symmetric forks.

These massive alterations of DNA replication were accompanied by slight alterations in the cell cycle progression of H1-TKO cells (Figure 26).

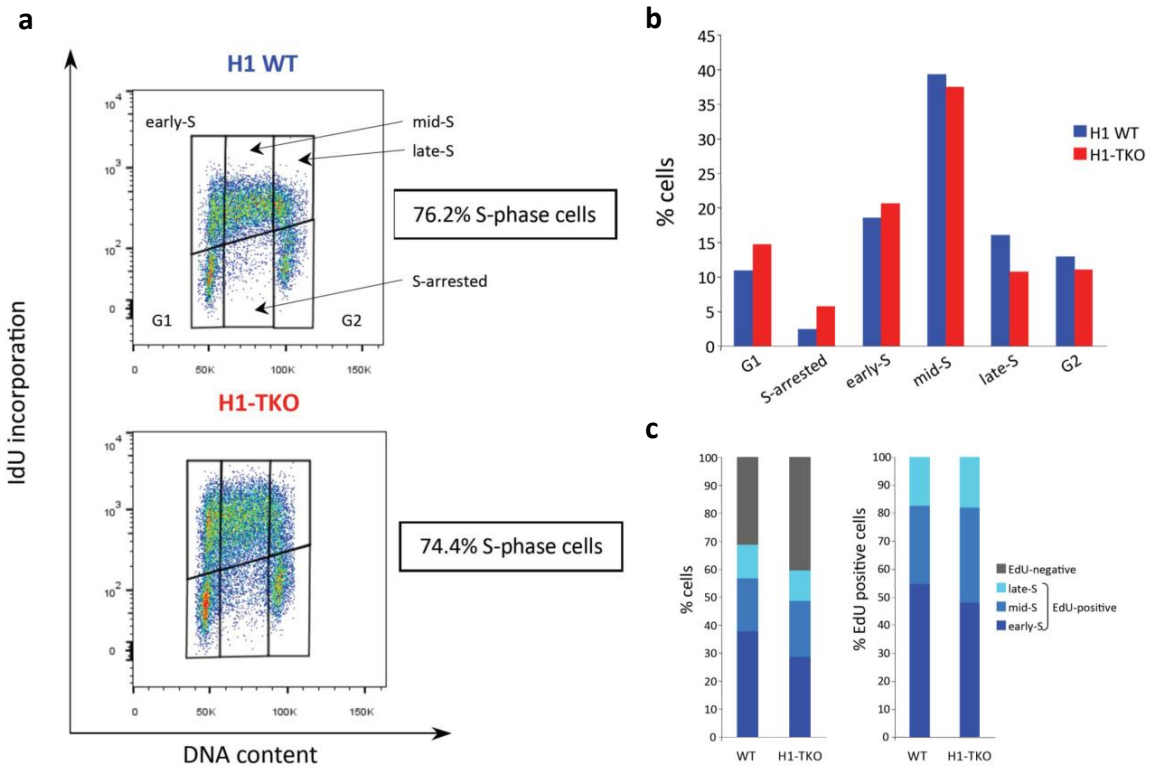


Figure 26: Detailed cell cycle analysis of WT and H1-KO mES cells. **(a)** Cell cycle and IdU incorporation analysis by flow cytometry analysed as described in Figure 21. Total percentages of S-phase cells are indicated. **(b)** Histogram representing the percentages of cells in each cell cycle phase derived from the analysis performed in a. **(c)** Histograms representing the analysis of replication factories by scoring EdU patterns. Left histogram shows the percentage of EdU-positive *vs* EdU-negative cells. Right histogram shows the percentage of cells in distinct points of the S-phase according to the EdU replication-foci patterns described by Nakamura *et al*, 1986 (see Figure 29a for a visual example of the different patterns obtained in mES cells). Data were derived from two independent experiments.

As mES cells have a high rate of replication, a large number of cells were detected in S-phase (76,2% as estimated by flow cytometry and 69% as estimated by EdU incorporation). In H1-TKO cells global S-phase percentages did not vary greatly (74,4% as estimated by flow cytometry and 60% as estimated by EdU incorporation), although some minor variations at G1, G2 and at different points of the S-phase were detected (Figure 26b and c). Nonetheless, cells with reduced H1 levels showed a twofold increase

in cells arrested in S-phase relative to WT cells, from around 2,5% to 6% (Figure 26b). This percentage of S-arrested cells was higher than the one found in both WT and HMGB1-KO MEFs, which was less that 1% (Figure 21b). Quite likely, the cell cycle perturbations of H1-TKO mES cells are associated to the replication fork instability observed, which undoubtedly is impacting the whole replication program, as seen by the altered SNS profiles.

2.3. Analysis of transcription dynamics in histone H1-depleted cells

Given the large range of replication defects in cells with H1-mediated altered chromatin conformation, we next decided to search for the underlying sources of such phenotype. One of the major problems that these cells encompass is the large percentage of replication forks that have halted their progression. Together with the reduced fork velocities observed it seems that the normal movement of the replication forks is being compromised genome-wide. One of the most common obstacles that replication forks have to overcome when the cells are in the process of duplicating their genomic information is active transcription (Figure 7). During S-phase, encounters between the two machineries can occur and, in normal conditions, cells have mechanisms to solve such conflicts (Figure 8; García-Muse and Aguilera, 2016; Hamperl and Cimprich, 2016).

We first sought to understand if the transcription process was altered in H1-TKO cells. We started by analysing the rates of RNAPII elongation at two specific genes, *Med13l* and *Inpp5*α. These genes were previously used to evaluate the dynamics of transcription in mouse cells (Jonkers *et al*., 2014) due to their large size and long distances between exons (Figure 27b, genomic maps). By transiently inhibiting RNAPII activity with 5,6-dichlorobenzimidazole1-β-D-ribofuranoside (DRB) and then follow the synchronous transcription wave from promoter regions, it is possible to estimate the time that a specific exon takes to be expressed by measuring newly synthesized primary transcripts by qPCR spanning different exon-intron junctions (Singh and Padgett, 2009; Jimeno-González *et al*., 2015). DRB specifically inhibits the transition from initiation to elongation phase of transcription by inhibiting CDK9, which is the kinase subunit of the Positive Transcription Elongation Factor (P-TEFb) that is required to promote this specific transition (Bensuade, 2011). Consequently, polymerases that were already in transcription elongation phase will finish in the presence of DRB and dissociate from chromatin upon transcription termination. However, initiating RNAPII will stay in a poised state until DRB removal from the media. In this way, by performing a time course of RNAPII kinetics after

DRB block and release (Figure 27a) we could to estimate the relative movement of RNAPII along the above-mentioned genes in WT and H1-TKO mES cells (Figure 27b).

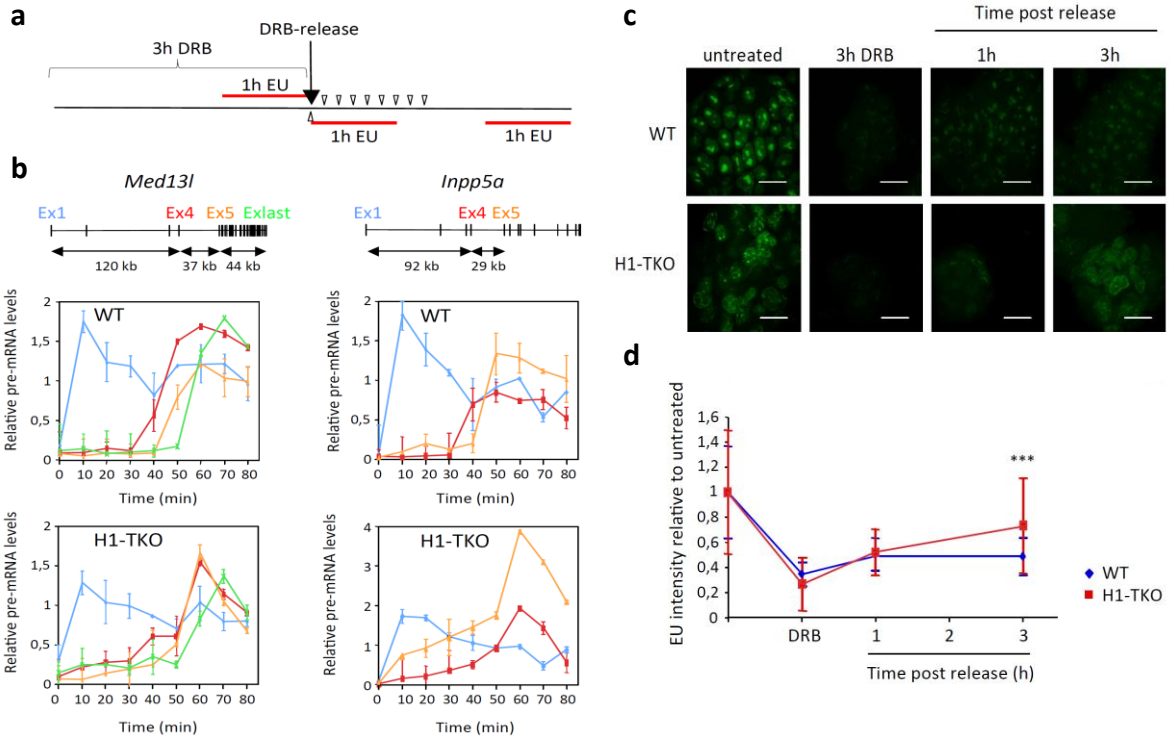


Figure 27: Analysis of transcription dynamics in WT and H1-TKO mES cells. **(a)** Diagram of the experimental design used to measure the rate of transcription elongation through transient inhibition of initiating RNAPII with DRB. Cell samples were incubated with media supplemented with 100 μM DRB for 3 hours. DRB was then washed off and total RNA was extracted from identical number of cells at each time-point (open triangles). Global nascent transcription was analysed by incubating the cells with EU for 1 hour at the time points indicated by the red lines. **(b)** Time course of transcription elongation for the *Med13l* and the *Inpp5a* genes in normal and H1-TKO cells. Genomic maps are present above the histograms with the distance between the primer pairs used to measure transcription levels at the respective exons. Pre-mRNA levels at each time-point were normalized to the values of non-DRB-treated samples. Results are shown as means +/- SD from two independent experiments. The different colours depict the different exons analysed. **(c)** Representative images of nuclear EU staining at the time points indicated in a. White bar corresponds to 20 μm **(d)** Quantification of nuclear EU signal intensity at the same time-points. Values were normalized to those obtained in untreated wt and H1-TKO cells. ***p<0.001. Scatter plots representing the raw values of EU intensity are shown in Supplementary Figure 2. Data derived from two independent experiments. Numerical values are shown in Supplementary Table 10.

Pre-mRNA transcription signals at both genes were detected at exon 1 in both WT and H1-TKO cells at the first 10-minutes time point (Figure 27b, blue lines). However, the normal kinetics of transcription of the other exons changed in H1-TKO cells relative to WT cells, with pre-mRNA expression starting to be detected from earlier time points in mutant cells and, strikingly, expression of exon 5 at both genes (Figure 27b, orange lines) being detected either at similar time points (in *Med13l*), or much earlier (in *Inpp5α*) than expression of exon 4 (Figure 27b, red lines). This result was the first evidence that the dynamics of the transcription process might also be affected in cells lacking the levels

numbers of H1. This prompted us to perform a global quantification of nascent transcription by 5-ethynyl uridine (EU) incorporation before and after DRB block release in order to see if there was any variation in the kinetics of transcription recovery between normal WT and H1-TKO cells (Figure 27a, c and d). After releasing cells from the block, the EU incorporation signals showed a significantly faster recovery rate of RNA synthesis in mutant cells, which is consistent with the increased RNAPII initiation activity previously noticed in the analysis of individual genes. This set of results support the notion that the altered chromatin scenario caused by the lack of normal levels of histone H1 might be contributing to increase DNA replication instability by increasing the number of encounters between the replication and transcription machineries.

As discussed in the Introduction chapter, the deregulation of RNAPII kinetics might also lead to the accumulation of non-canonical structures in the genome, like R-loop structures (Figure 7 and 8). The presence of this RNA:DNA hybrids can be detected *in vivo* by immunofluorescence using the S9.6 antibody (Boguslawski *et al*., 1986). We took advantage of this tool to evaluate if the perturbations in transcription detected in H1-TKO cells were accompanied by the accumulation of RNA:DNA hybrids in chromatin (Figure 28).

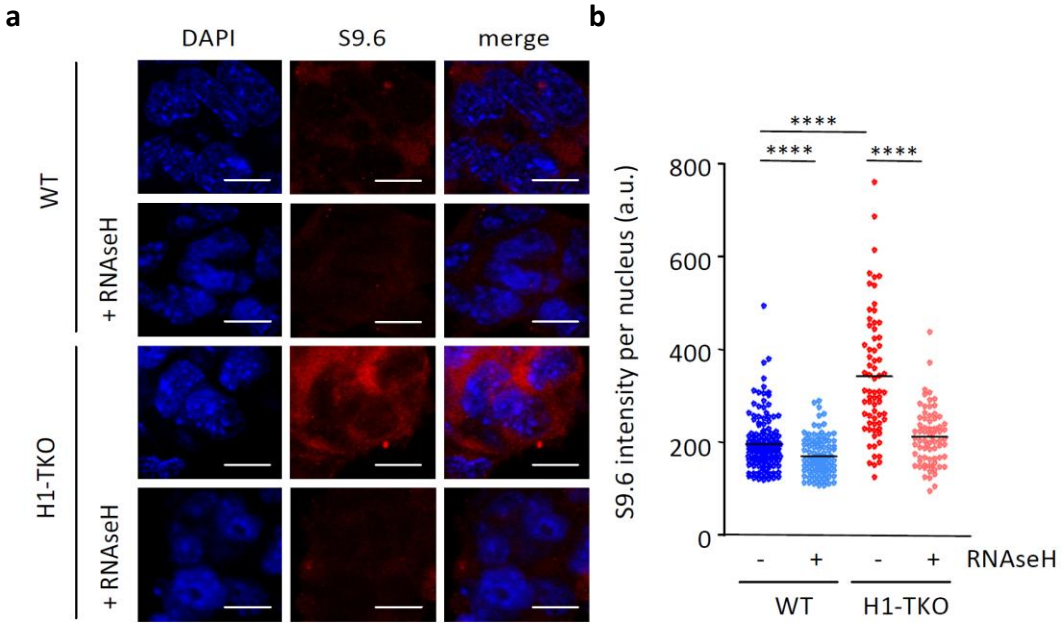


Figure 28: Analysis of RNA:DNA hybrids in mES cells. **(a)** Representative images of S9.6 (R-loop) immunostaining, with or without RNAseH incubation. DAPI staining was used as reference in order to measure nuclear R-loop signal intensities. White scale bar corresponds to 10 μm. **(b)** Scatter plots depicting nuclear R-loop signal in WT and H1-TKO mES cells and the corresponding RNAseH controls. Data are pooled from two independent experiments. Black lines indicate median values. ****p<0.0001. Numerical values are shown in Supplementary Table 11.

We found that H1-TKO cells had a significant increase in nuclear S9.6 signal compared to WT mES cells, which presented fluorescence levels close to the baseline

(Figure 28a and b). Importantly, this signal was sensitive to RNAseH treatment, confirming that the antibody reactivity was due to the presence of R-loops.

We next analyze at which point of the cell cycle, and more specifically in which moment of the S-phase R-loop accumulation was more evident. We did this by co-scoring replication foci-patterns as revealed by EdU incorporation (Figure 29a).

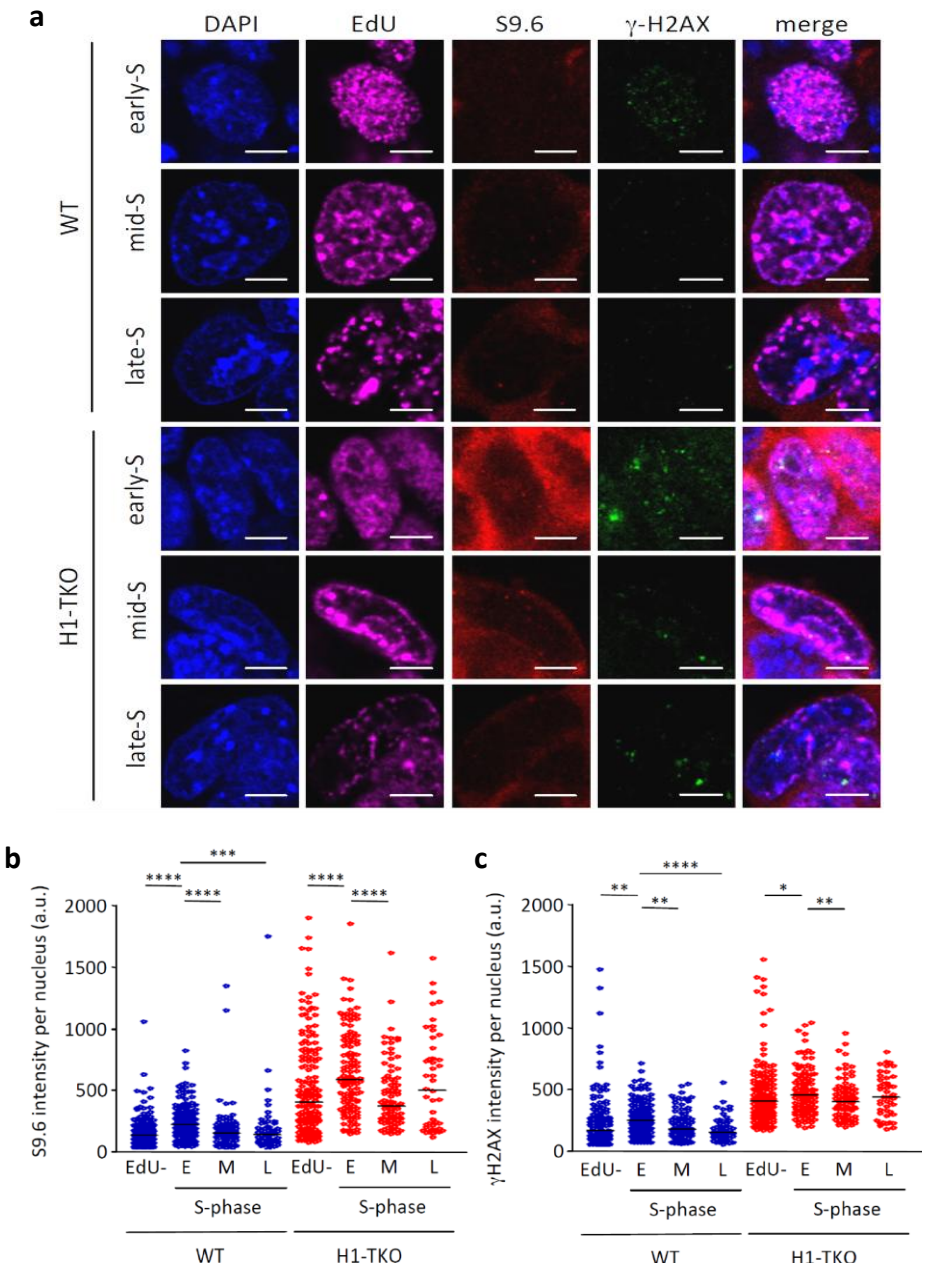


Figure 29: S-phase distribution of nuclear R-loop and γH2AX intensities. **(a)** Representative image of EdU, S9.6 and γH2AX immunostaining. S-phase patterns were classified according to the EdU replication-foci patterns described by Nakamura *et al*, 1986. White bar corresponds to 10 µm **(b)** Scatter plot depicting the distribution of S9.6 (R-loop) nuclear intensity during S-phase. S-phase was divided in three subphases, early (E), middle (M) and late (L). EdU- corresponds to the signal found in cells with no EdU incorporation during the time of the pulse (20 minutes). Values are expressed in arbitrary units (a.u.). Numerical values are shown in Supplementary Table 12 **(c)** Same analysis done in b for nuclear S9.6 γH2AX intensity. Numerical values are shown in Supplementary Table 13. ****p<0.0001; ***p<0.001; **p<0.01; *p<0.05. Data are derived from two independent experiments.

In parallel, we evaluated the possibility of associated genomic instability by analysing the signal of phosphorylated histone H2AX (γH2AX), used as a biomarker for DNA double strand breaks (DBS; Kuo and Yang, 2008). The results showed a significant accumulation of both R-loop structures (Figure 29b) and γH2AX (Figure 29c) in the nucleus of H1-TKO cells at all points of the cell cycle when compared to WT mES cells. Remarkably, the accumulation of both signals was more evident at early stages of S-phase. The fact that these cells presented higher levels of γH2AX is also in agreement with our hypothesis of increased replication/transcription encounters in situations of reduced H1 content, which when left unchecked often result in breaks on the DNA molecule (Aguilera 2002; Li and Manley 2006; Aguilera and Garcia-Muse 2012).

To ensure that this increased accumulation of RNA:DNA hybrids and DSB was a specific defect linked to H1-mediated alterations in the chromatin structure, we checked if HMGB1-KO MEFs, in which a 20% reduction in nucleosome content results in little replication alterations, also showed a similar phenotype (Figure 30).

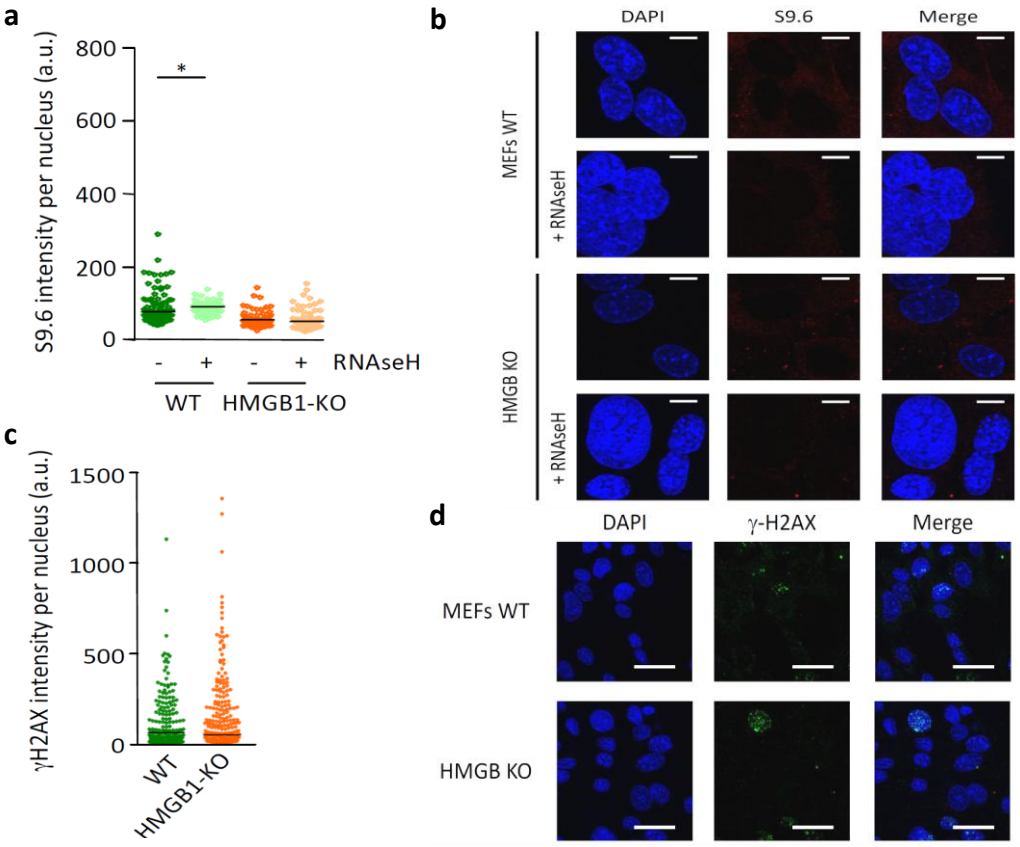


Figure 30: Analysis of RNA:DNA hybrids and γH2AX levels in normal and HMGB1-KO MEFs. **(a)** Scatter plots depicting the nuclear R-loop signal in MEF cells with or without RNAseH treatment. *p<0.05. **(b)** Representative image of R-loop immunostaining in MEF cells with or without RNAseH treatment. White scale bar represents 10 μm **(c)** Scatter plots depicting nuclear γH2AX signal in MEF cells. Data are pooled from two independent experiments. Black lines on the plot indicate median values. See Supplementary Table 14 for numerical values. **(d)** Representative image of γH2AX immunostaining in MEF cells. White scale bar represents 40 μm. Data are derived from two independent experiments. Numerical values are shown in Supplementary Table 14.

In both WT and HMGB1-KO MEF cells we found low levels of R-loops and γH2AX. The levels of both signals did not present significant variations between normal and HMGB1 depleted cells and the only slightly significant variation observed was between WT MEFs and their RNAseH control, regarding R-loop signal (Figure 30a, $p<0.05$). The median value slightly increases in the control sample, which seems counterintuitive, but taking in account that the majority of fluorescence intensity values detected in MEF cells where close to the background levels, as seen in Figure 30b and shown in Supplementary Table 14 (both mean and median values below 100 a.u.), this difference is likely due to variations in background signal. Furthermore, we found no significant difference between R-loop signal intensities in HMGB1-KO and its corresponding RNAseH control, further implying that MEF cells encompass only residual levels of R-loops in chromatin. Since in comparison with H1-TKO cells, HMGB1-KO cells don't show abnormal accumulation of RNA:DNA hybrids, nor significant increased levels of genomic instability as estimated by γH2AX signalling, these evidences suggest that the impairment in global chromatin compaction occurring in H1-TKO cells is critical to maintain the balance between transcription and replication.

2.4. Analysis of replication-transcription conflicts in cells with altered chromatin structure

To further clarify the extent of the interaction between transcription and replication in the context of altered chromatin structure, we decided to perform a detailed study of replication dynamics upon impairing the transcription process. In order to do this we blocked transcription with α-amanitin, which is a stable and irreversible inhibitor of RNAPII and III (RNAPI is insensitive to its action). This cyclic peptide binds near the catalytic centre of the transcription polymerases with high specificity and affinity, trapping them in a conformation that prevents nucleotide incorporation and translocation of the transcripts (reviewed in Bensaude, 2011). Furthermore, α-amanitin promotes the dissociation of RNAPII from chromatin, because it triggers the degradation of its larger subunit, Rpb1 (Bensaude 2011). As the uptake of this drug is slow, usually taking several hours for the effects on transcription to start being noticed we've exposed the mES cells to an appropriate concentration of this drug (10 µg/mL) for six hours. Figure 31a shows a representative picture of the effect of the α-amanitin treatment in global transcription levels estimated by EU incorporation. Notice that, as previously mentioned, α-amanitin has no effect on RNAPI and consequently we could still detect nucleolar signal deriving specifically from RNAPI activity. In order to ensure that α-amanitin was blocking RNAPII

activity in parallel experiments, cells were treated with 0.05 μg/mL of Actinomycin-D for 1 hour, which specifically inhibits RNAPI function (Figure 31a and Supplementary Table 15; Bensaude, 2011). This approximation enabled us to evaluate the effect of α-amanitin on RNAPII transcription. Our expectation was that, after 6h treatment, the H1-TKO mES cells would have time to resolve, at least in part, the accumulation of RNA:DNA hybrid structures in chromatin. This was in fact the case when we analysed R-loop levels upon inhibition of transcription in WT and H1-TKO mES cells (Figure 31b).

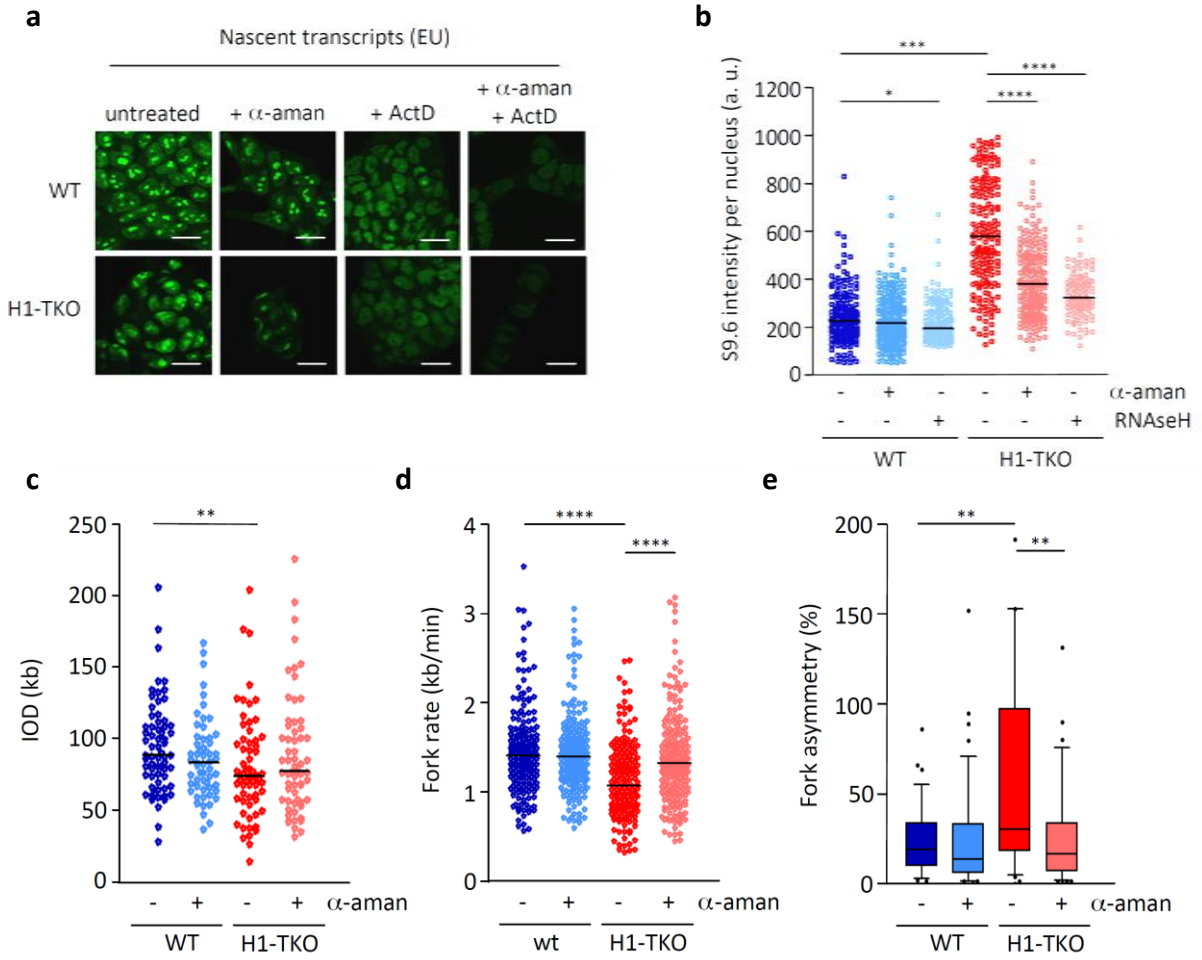


Figure 31: Analysis of RNA:DNA hybrid levels and replication dynamics in mES cells upon transcription inhibition. **(a)** Representative image of EU staining in mES cells. Cells were treated for 6 hours with α-amanitin to specifically block RNAPII. As RNAPI continue to be active upon α-amanitin treatment, cells were treated with 0,05 μg/mL Actinomycin-D to specifically inhibit nucleolar signal and confirm that the extranucleolar EU incorporation dropped in the presence of α-amanitin. Numerical values are shown in Supplementary Table 15. **(b)** Scatter plot depicting the distribution of nuclear R-loop intensities in mES cells upon blocking RNAPII transcription with α-amanitin. Untreated and RNAseH controls are shown. **(c)** Scatter plot depicting IOD distributions in mES cells upon blocking RNAPII transcription with α-amanitin. **(d)** and **(e)** same analysis present in c, for Fork rate and fork asymmetry, respectively. Medians are represented in all plots. ****p<0.0001; ***p<0.001; **p<0.01; *p<0.05. Data are derived from two independent experiments. Numerical values are shown in Supplementary Tables 16 (R-loops), 17 (IOD), 18 (Fork rates) and 19 (Fork asymmetry), respectively.

The drop in R-loop signal detected in H1-TKO cells upon transcription inhibition was highly significant ($p<0.0001$), but not complete, as α-amanitin treated cells didn't reach the low levels found in WT cells or in its RNAseH control. To evaluate if this reduction in the levels of transcriptional R-loops in the genome facilitates the transit of replication forks we measured IODs, fork rates and fork asymmetry levels in WT and H1-TKO mES cells upon α-amanitin exposure (Figure 31c, d and e). IOD analysis didn't reveal any significant change in transcription-blocked H1-TKO cells when compared to their untreated counterpart. However, we found major changes when analyzing the other two parameters (Figure 31d and e, $p<0.0001$ and $p<0.01$). Upon transcription inhibition, both fork velocity and fork asymmetry were recovered to levels similar to the ones found in WT mES cells, indicating that, indeed, the defects in transcription arising from the abnormal chromatin configuration of H1-TKO cells are a major source of the alterations in the replication program. Furthermore, when we analyzed in parallel γH2AX levels upon blocking transcription, we also observed a drastic recovery in H1-TKO cells towards WT levels (Figure 32, $p<0.0001$). The reduction in γH2AX intensity, R-loop levels, and fork asymmetry levels observed in H1-depleted cells treated with α-amanitin seem to connect all these features of replication stress together.

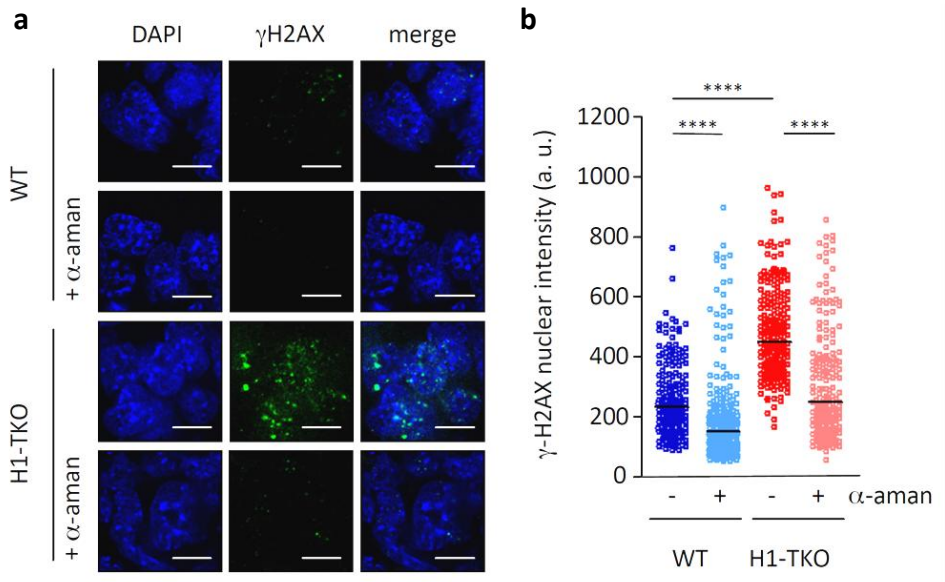


Figure 32: Analysis of γH2AX levels in mES cells upon transcription inhibition. **(a)** Representative images of nuclear γH2AX foci intensity in mES cells untreated or treated with α-amanitin. Treatment times and doses are the same as in Figure 31a. White scale bar corresponds to 10 μm. **(b)** Scatter plot depicting γH2AX intensity per nucleus. Black lines inside the plots indicate median values. Intensity values are indicated in arbitrary units of fluorescence (a.u.). ****$p<0.0001$. Data derived from two independent experiments. Numerical values are shown in Supplementary Table 20.

The set of results obtained until now argue that irregularities in the transcriptional program severely affect the replication dynamics of these cells. To further characterize

this relationship we used the reversible transcription inhibitor DRB (Figure 27a and c) as a "switch" to test whether the observed replication recovery was stable after transcription reactivation (Figure 33).

When analyzing this second transcription inhibition experiment we uncovered several interesting results. Firstly, using a DRB block we found a similar effect on suppressing the replication phenotype of H1-TKO cells as the one described using α-amanitin, where both fork rates and asymmetries were recovered to similar levels to those obtained in WT mES cells (Figure 33c and d, $p<0.0001$ and $p<0.01$). One important difference was that using this transcription block we were also able to observe a recovery in IOD (Figure 33b, $p<0.0001$), which we did not observe when using α-amanitin (Figure 31c). One possibility for this difference could be due to the faster effect of DRB on transcription, which could have bigger impact on cycling cells than the slow action of α-amanitin, giving them more time to normalize their replication program.

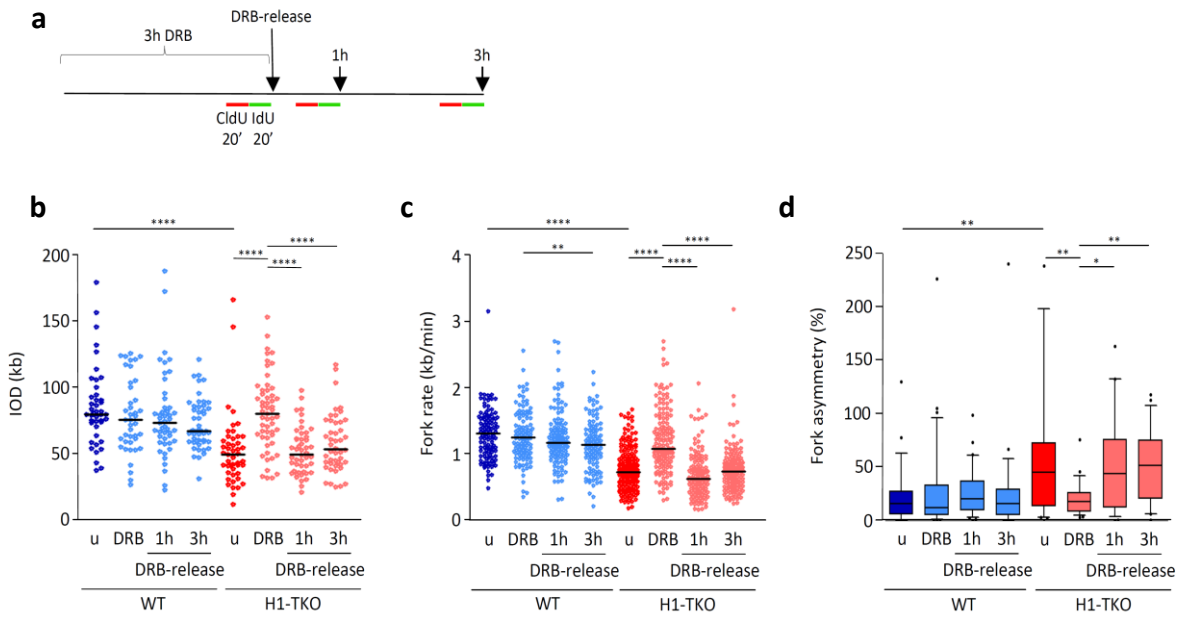


Figure 33: Analysis of DNA replication dynamics in mES cells upon transcription reactivation. **(a)** Diagram of the experimental design used to measure replication dynamics by transient inhibition of transcription with DRB. mES cells were sequentially labelled for 20 minutes with CldU (red line) and IdU (green line) at the indicated time points. The effect of DRB on nascent mRNA transcripts at the same time points is shown in figure 27. Scatter plots depicting IODs (b), fork rates (c) and asymmetry levels (d), respectively, along the experiment. Black line inside plots represents the median. ****p<0.0001; ***p<0.001; **p<0.01; *p<0.05. Data are derived from two independent experiments. Numerical values for all plots are shown in Supplementary Tables 21 (IOD), 22 (Fork rates) and 23 (Fork asymmetry), respectively.

Immediately after releasing the cells from the DRB block when transcription is resumed, all the replication parameters that were suppressed by DRB treatment (DRB) in H1-TKO cells recover to the levels found in untreated (u) cells (Figure 33b, c and d, $p<0.0001$ for IOD and fork rates and $p<0.05$ for fork asymmetries). This result constituted

further proof that, in this context of altered chromatin structure, the balance between transcription and replication is impaired mainly due to conflicts between both machineries.

2.5. Evaluation of RNAPII elongation rates and DNA replication fork rates in cells with reduced histone content

The data so far supported a scenario where the defects in chromatin structure can impact both transcription and replication dynamics, causing replication stress. A prediction from these results is that, in HMGB1-KO MEFs, where replication forks travel faster without detectable increase in fork stalling (Figure 19 and 20), RNAPII transcription elongation rates wouldn't suffer major alterations. To test if this was indeed the case, we performed time-course transcription elongation measurements upon DRB release at the *Med13l* and *Inpp5*α genes, in similar conditions to the one previously performed for mES cells (Figure 27a and b). Transcription analysis at both genes revealed only minor changes between WT and HMGB1-KO cells in the timing of expression of the examined exons (Figure 34).

The lack of alterations in transcription elongation fitted well with previous results showing that these MEF cells encompass low R-loop and γH2AX levels, with no significant changes when HMGB1 is knocked down (Figure 30). Altogether, this data suggested that cells with reduced nucleosome numbers can tolerate fast moving forks without compromising genome stability if transcription is not altered. To further validate this idea we decided to test a complementary scenario and asked if the replication dynamics was altered in situations where a reduction in histone content promotes the increase in RNAPII elongation rates. To accomplish this we made use of a human cell line (HCT116) where the levels of canonical histones can be modulated by knocking down the expression of the stem-loop-binding protein (SLBP) gene, which encodes a histone mRNA regulatory factor (Marzluff *et al*., 2008). A moderate reduction in the levels of SLBP results in a reduction in histone levels, promoting an increase in RNAPII elongation rates and consequent defects in cotranscriptional splicing without leading to cell-cycle defects (Jimeno-Gonzalez *et al*., 2015).

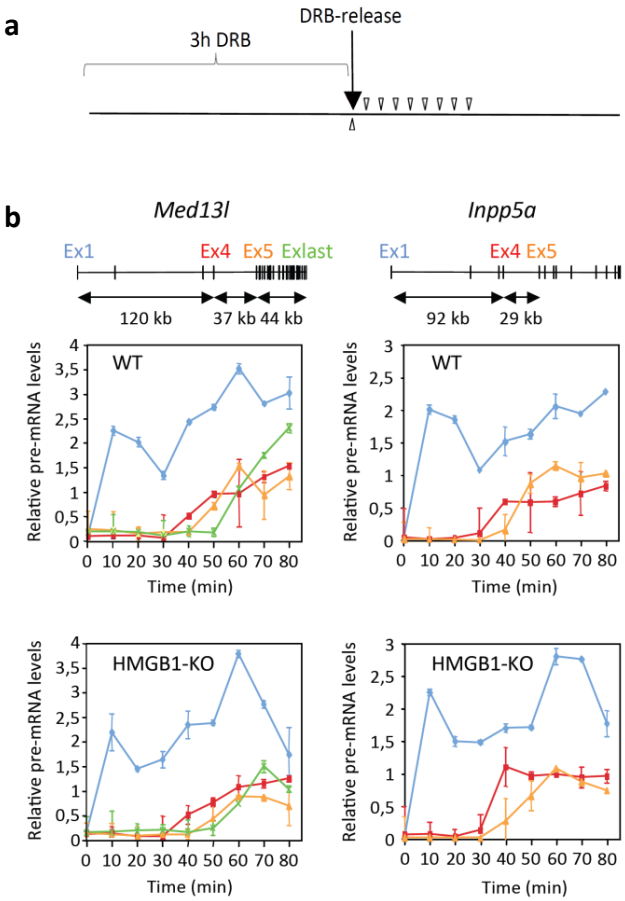


Figure 34: Analysis of transcription elongation rates in WT and HMGB1-KO MEF cells. **(a)** Diagram of the experimental design used to measure the rate of transcription elongation through transient DRB inhibition of initiating RNAPII. **(b)** Time course of transcription elongation for the *Med13l* and the *Inpp5α* genes in both cell types. Pre-mRNA levels at the specified time-points were normalized to the values of non-DRB-treated samples. Results are shown as means +/- SD from two independent experiments. All symbols are as in Figure 27.

Upon performing DNA fiber analysis on Doxycyclin-induced SLBP-knockdown cells we were able to detect a significant increase in replication fork rates (Figure 35b and d, $p<0.0001$), with no variation in IODs nor in fork asymmetry (Figure 35a, c). The average increase in fork velocity was much higher in this nucleosome depleted scenario (40% in average), than in the one observed in HMGB1-KO MEFs (29% in average) (Figure 35d), consistent with the increased RNAPII elongation rates displayed by SLBP-knockdown cells (Jimeno-Gonzalez *et al*., 2015).

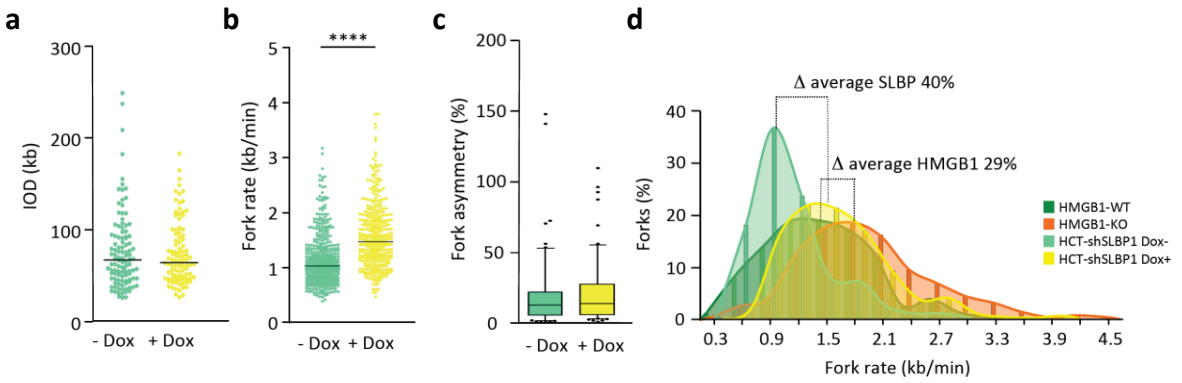
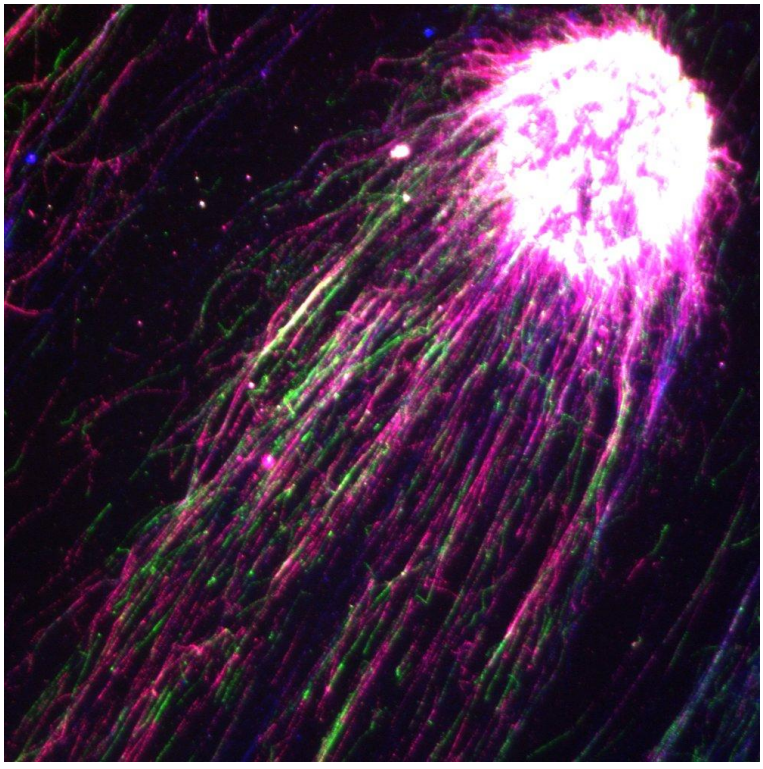


Figure 35: Replication dynamics in cells with reduced histone content.(**a**) IODs, (**b**) fork rates and (**c**) percentage of fork asymmetry calculated from stretched DNA fibers of HCT-shSLBP.1 cells cultured in the absence (light green) or presence (yellow) of Doxicyclin (Dox) for 72 hours. Black lines indicate median values. ****p<0.0001. (**d**) Distribution of replication fork speeds in cells with reduced histone content relative to their respective WT counterparts. Data are derived from two independent experiments. Numerical values are shown in Supplementary Table 24. HMGB1-KO and WT data used here is the same represented in Figure 19, which corresponds to Supplementary Table 8.

Most likely, the specific chromatin changes are occurring in SLBP knock-down cells, such as the drop in the levels of the histone variant H2A.Z and the increase of the variant H3.3 (Jimeno-Gonzalez *et al.*, 2015), seem to have a greater influence on the movement of both the replication and transcription machineries than the ones that occur in HMGB1-depleted cells. The analysis of replication dynamics in this later cell line also reinforces the notion that cells can accommodate fast moving replication forks in contexts of histone depletion without compromising genomic stability, the same not being true in scenarios of altered chromatin structure leading to impaired RNAPII dynamics.

Discussion



Cell viability relies on the correct regulation and fine tuned balance between different genomic processes: replication, transcription, recombination and repair. As we go up in the evolutionary scale and complexity increases, the layers of genomic regulation accumulate in parallel, making it harder to decipher the nature of the interactions between the different processes that are occurring on the same DNA template. One of the most important factors that contribute for the maintenance of genomic stability is chromatin. The correct packaging of the genomic information contained in the DNA molecule into chromatin is indispensable for the regulation of all genomic processes from transcription to DNA damage repair and, most importantly, DNA replication. This is the most fundamental process without which the normal development and correct propagation of all living organism would be impossible. The replication process ensures the accurate duplication of the DNA molecule and any perturbation, either from exogenous or endogenous sources, might interfere with the correct progression and completion of this process, affecting in this way the integrity of the genome, which in turn can have catastrophic consequences for the cell and the organism. The correct coordination between DNA replication and other genomic processes is essential for survival. Thus, deciphering how the structure of chromatin is able to modulate the expression and maintenance of the genetic information encoded in eukaryotic genomes, and how these processes take place within the context of a highly complex and compact environment is of major importance, and will contribute to increase our knowledge in the field of Epigenetics and consequently our better understanding of the many diseases that arise from the abnormal regulation of this fundamental element.

1. Replication start sites and nucleosomal architecture are intimately linked

1.1. The NuSA technique allows detailed profiling of nucleosomal patterns at mammalian ORIs

We've started our work by analysing the specific nucleosomal landscape at a subset of previously characterised promoter-ORIs (Table 1; Sequeira-Mendes *et al*., 2009), aiming to address how it relates to specific genomic processes like transcription or DNA replication efficiency. At those active transcription sites we found the stereotypic nucleosomal configuration described at promoter regions, with well positioned nucleosomes flanking the TSSs. This is consistent with the dynamic process of chromatin remodelling as consequence of an active transcription machinery, that generates an array of well positioned nucleosomes downstream of the promoter region (Figure 4) (Lieb, *et al*., 2001; Schwabish and Struhl, 2004; Liu *et al*., 2006; Arya, 2010). One exception to this

was the promoter region of *Haus7* (Fig 12, central panels). Although at this region we also found a high level of nucleosome occupancy downstream of the TSS, the +1 nucleosome doesn't seem to be well-positioned, indicating that it protects slightly different DNA regions within the mES cell population. Interestingly, when we analyzed the same promoter-ORI region in a cell type in which the expression levels of this gene was six times greater (NIH3T3 mouse fibroblasts, Figure 15), we were able to detect a shift in the location on the +1 nucleosome, which becomes better positioned. This finding illustrates the high sensitivity of the NuSA technique in detecting subtle changes in the local chromatin environment due to alterations in other genomic processes.

By combining diverse conditions of chromatin preparation and MNase digestion with the NuSA assay, we were able to uncover different levels of stability of the various histone octamers. Specifically, by performing MNase digestions on native chromatin under different conditions and by employing distinct stringency treatments for ChIP experiments with antibodies against histone H3 we could differentiate between stable and labile nucleosome particles at the regions analysed (Figure 12). The identification of labile histone octamers at the promoter regions of the active genes *MeCP2*, *Vps45* and *Haus7* is in agreement with the notion that the more dynamic chromatin at those sites is important to allow the transient exposure of the DNA molecule. This likely facilitates the accessibility of cis-regulatory complexes like TFs, chromatin remodelers, initiating transcription complexes and others, that together will contribute to maintain a chromatin domain active for transcription (Deal and Henikoff., 2010). In addition, the fact that the non-promoter region of the *Scl7a14* gene showed an array of poorly positioned nucleosomes with an almost constant level of occupancy values (Figure 13) indicates that the presence of labile particles is a specific feature of the active promoter regions studied.

This set of results are in agreement with the concept that active gene promoters are not in fact nucleosome-free but contain an unstable nucleosome that is usually lost under the usual conditions employed in nucleosomal preparations for genome-wide studies (Jin *et al*, 2009). In fact, the work performed at our lab revealed that when we aligned our detailed nucleosome maps of individual promoter-ORIs with the corresponding maps generated genome-wide from previous work (Tief *et al*, 2012), we are able to notice several missing nucleosome particles in those data-sets (data not shown). This demonstrates that the high-resolution study that we conducted allowed us to unveil the diverse characteristics of the histone particles at those ORI sites. In addition, since all the analysed regions associate with ORIs, our analysis shows that replication initiation can occur in a range of nucleosome configurations.

1.2. DNA SYNTHESIS INITIATES AT SITES OF HIGH NUCLEOSOME OCCUPANCY

We took advantage of the high resolution of the scanning assay of the NuSA technique to couple it to the fine mapping of replication start sites. By employing preparations of replication intermediates from increasing sizes we found that the initiation sites occur at positions of high nucleosome occupancy in all cases (Figure 14 and Supplementary Figure 1). One thing that was noticeable was that the maximum enrichments of SNS detected were comparable between the four regions analyzed, including the ORI associated with *Scl7a14*, which was previously defined as a low efficient ORI (Table 2; Sequeira-Mendes *et al*, 2009). This was due to the fact that SNS enrichments were normalized with values obtained at flanking amplicons that were relatively close to the analysed regions (700 to 1200 bp). This is in line with previously reported low enrichment levels obtained when using similar approaches (Giacca *et al*, 1994; Keller *et al*, 2002; Ladenburger *et al*, 2002; Sequeira-Mendes *et al*, 2009). Nonetheless, the enrichments were much higher and variable between regions when the SNS values were normalized to a control region located several kb and isolated from known ORIs (Supplementary Figure 1), confirming that the SNS levels detected at the regions under analysis corresponded in fact to actual sites of DNA replication initiation.

The resolution of the scanning assay on short SNSs not only allowed us to map in high detail the location of DNA synthesis start sites within the analysed cell population, it also enabled us to correlate the SNS profiles with the nucleosome profiles. As mentioned earlier, we found that peaks of maximum SNS enrichment are located at regions of high nucleosome occupancy. Moreover, we detected a spatial correlation between the shape of the SNS and the nucleosome profiles. At the *MeCP2* promoter-ORI region, for example, we found one single sharp peak of SNSs that coincided with the well-positioned nucleosome +1 (Figure 14, leftmost panels). On the contrary, at the intragenic *Scl7a14* region we detected a less defined peak of SNS enrichment that accompanied the poorly defined nucleosomal array (Figure 14, rightmost panels). Previous evidences suggested that replication initiation points are less precise at ORIs located within sites that encompass less defined nucleosomal patterns (Lubelsky *et al*., 2010). Making use of the DNA fiber technique, the authors analysed the frequency of replication initiation along the DHFR *locus* of Chinese hamster cells, which was previously described as being packaged by poorly positioned nucleosome particles and scattered ORC binding profiles (Dijkwel and Hamlin, 1995; Dijkwel *et al*, 2000). Correspondingly, they found a disparity of replication initiation patterns between fibers broadly distributed along the entire DHFR *locus*, indicating that at this region where nucleosomes didn't show a strict positioning pattern, ORIs didn't had a preferred initiation site. Interestingly, at the other two ORIs

analysed in our work we found a bimodal distribution of SNS peaks (*Vps45* and *Haus7* promoter-ORI regions, Figure 14, middle panels); one of the SNS peaks was located at a position protected by a labile nucleosome particle, and the other peak at a position coinciding with a stable nucleosome. In both cases, however, SNS enrichments were sharper at the locations of well-positioned nucleosomes. Altogether, these results suggest that is nucleosome positioning, and not occupancy, the feature that correlates more strongly with the replication start sites. This bimodal distribution of SNS profiles, which peaks upstream and downstream of the TSSs of both regions has been previously reported in studies that employ hybridization of SNS on microarrays (Cayrou *et al*, 2011), and might depict two distinct sites of preferential initiation at the cell population. Alternatively, there is also the possibility that these bimodal peaks reflect dual initiation events at opposing regions of leading-strand synthesis as previously seen in mouse and *Drosophila* cells (Cayrou *et al*, 2011).

1.3. Replication initiation patterns reflect the nucleosomal architecture

To further test the possible correlation between replication start sites and nucleosome positioning, we take advantage of the expected alterations in the nucleosomal landscapes following promoter activation. When analysing the same promoter-ORI regions in different cell types presenting differences in the transcription levels of those genes, we were able to show that variations in the position of histone octamers were mirrored by shifts in the SNS profiles (Figure 15). Taking in to account that even the most efficient mammalian ORIs likely fire on average in less than 20% of the S-phases (Cayrou *et al*, 2011; Gilbert, 2012), this relationship may be explained by the fact that nucleosome maps were generated in asynchronous cells whereas ORI firing only occurs in a subset of cells during S-phase. Indeed, when analysing nucleosome dynamics upon S-phase entry at the LaminB2 ORI we found that the ORC binding site coincide with the position of a labile nucleosomes that are precisely remodelled before the onset of DNA replication (Figure 16). This result is also in agreement with earlier studies at the genome of the Epstein-Barr virus (Papior *et al*, 2012). The genome of this virus usually persists in infected cells as chromatinized minichromosomes that are replicated by the host machinery, and they exhibit a dynamic MNase pattern displaying increased sensitivity during S-phase, correlated with pre-RC and SNS enrichments. Thus, our findings suggest a relationship between the deposition of histone core particles during the replication process and the synthesis of the new DNA strands in eukaryotic cells. Most interestingly, genome-wide analysis of Okazaki fragment polarity in *S. cerevisiae*

demonstrated that the ligation junctions between adjacent fragments tend to take place at the region coinciding with the nucleosome dyad (Smith and Whitehouse, 2012). When the processivity of the lagging strand or the chromatin assembly was disrupted, both the distribution and the average size of the Okazaki fragments were altered, further implying a strong link between replication and histone repositioning. Whether newly deposited nucleosomes and leading-strand synthesis are coupled in mammalian cells constitutes an interesting open question.

1.4. Nucleosome remodelling and the opportunistic nature of replication initiation

The results obtained at the set of ORI regions analysed here support the notion that efficient ORIs are associated with CpG island-promoter regions, which are genomic sites where the prevalence of histone variants associated with active transcription, such as H3.3 and H2A.Z, contributes to the facilitated remodelling of nucleosomes and the maintenance of an open chromatin structure (Jin *et al*, 2009; Sequeira-Mendes *et al*, 2009; Cayrou *et al*, 2011; Besnard *et al*, 2012). This is also evidenced at the LaminB2 region where we found that the region protected by a labile nucleosome, which coincided exactly with the site of ORC binding, was rapidly exposed upon entry into S-phase (Figure 16). This findings are also reminiscent of work in Drosophila cells, were nucleosome turnover rates were found to be higher at regions that are bound by ORC (Deal *et al*, 2010). At the same time that we performed this analysis, a genome-wide mapping of ORC1 binding sites in HeLa cells was published (Dellino *et al*., 2013). The authors described a strong correlation between ORC1 occupancy and the TSS of both coding and non-coding RNA transcripts, as well as higher transcription levels at mapped ORC sites associated with replication in early S-phase. Altogether, these data point to a scenario where regions with high rates of transcription initiation, which are transiently accessible during the G1-phase of the cell cycle, allow an enhanced recruitment of ORC1 and other pre-RC factors. An attractive possibility could be that ORC interacts directly with certain histone particles, especially labile ones. In this sense, in *S. cerevisiae*, the bromo-adjacent homology domain (BAH) of ORC1 has a major role in facilitating ORC binding to the DNA (Noguchi *et al*, 2006; Muller *et al*., 2010) and most importantly, has the capability of recognizing a mark on the histone H4, H4K20me2, which is specially enriched at ORI sites (Kuo *et al*., 2012; Rivera *et al*., 2014). Therefore, it is possible that ORC recruitment and/or stabilization could be mediated through a direct interplay between the nucleoprotein particles and the ORI recognition complexes themselves (Figure 36a).

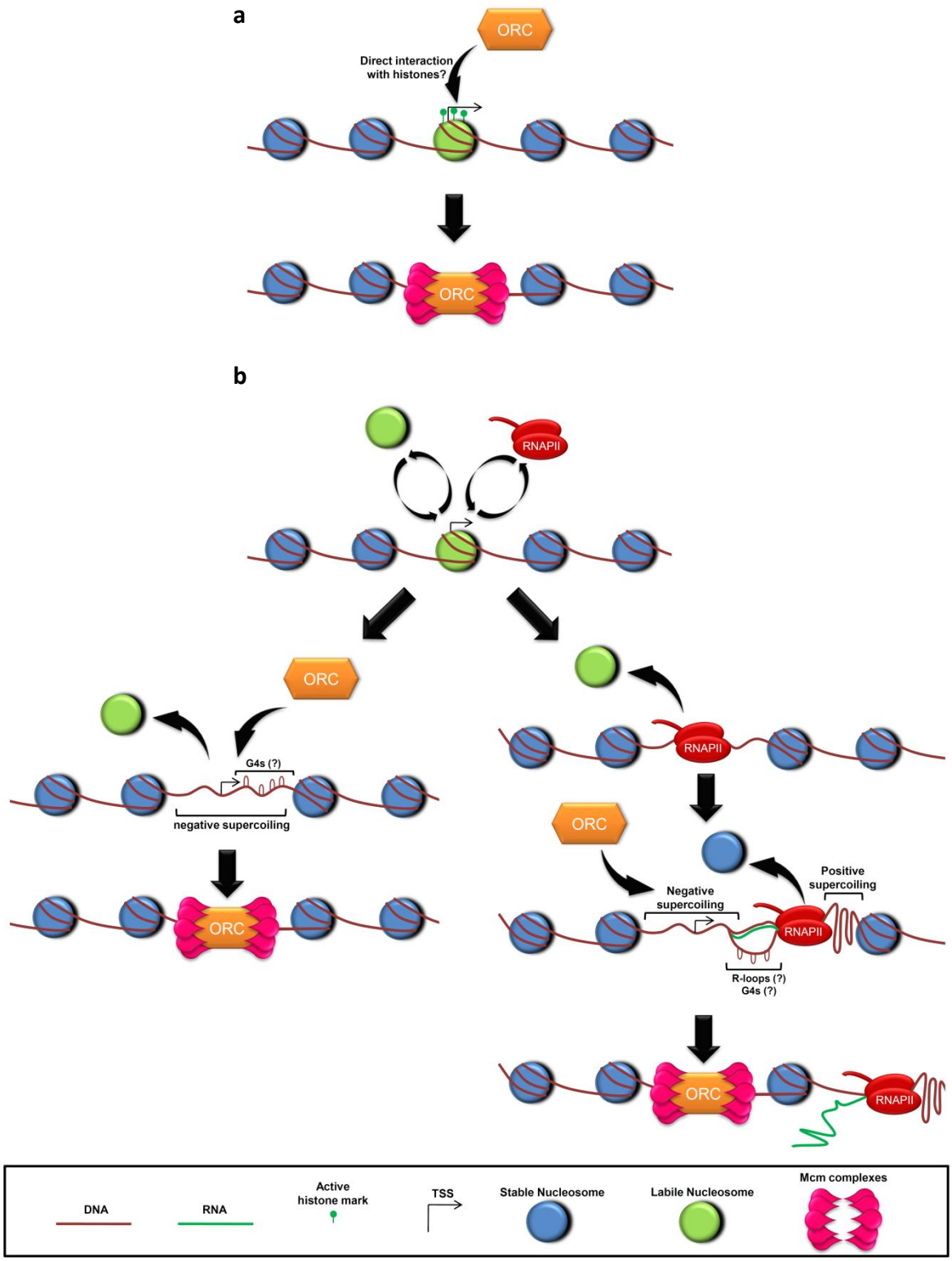


Figure 36: Diagram depicting the opportunistic recruitment of ORC at active promoters. **(a)** Through its BAH domain, ORC might interact directly with histone octamers that encompass specific marks (like H4K20me2; Kuo *et al*., 2012) or maybe with specific histone variants (like H3.3 or H2A.Z) that tend to be more enriched at active promoters. Upon nucleosome eviction ORC binds to the DNA molecule and can recruit additional replication complexes like Mcm2-7 **(b)** The high turnover rates between labile nucleosomes and/or the transcription machinery favor the binding of ORC to the DNA. Nucleosome eviction (left) generates negative helical tension inside the NFR (with possible generation of G4 structures) favoring ORC binding. Similarly RNAPII activity (right) generates negative helical tension behind the transcription machinery (with possible generation of R-loop and associated G4 structures) favoring ORC binding.

Another possible way in which ORC could take advantage of the increased remodelling occurring at promoter sites would be through the process of nucleosome eviction. It has been demonstrated that ORC affinity for DNA increases several fold in regions of the double helix that are being subjected to helical tension, specifically negative supercoiling. One process that can generate this type of tension is nucleosome eviction (Vashee *et al*., 2003; Remus *et al*., 2004; Schaarschmidt *et al*., 2004). This preference is not specific of the ORC complex, as unconstrained supercoiling stores free-energy on the DNA molecule, promoting several protein-DNA interactions (Bates and Maxwell, 2005; Corless and Gilbert, 2016). It can also contribute to conformational changes in DNA, facilitating the formation of non-canonical DNA structures like Z-DNA, cruciform DNA and, most distinguishably, R-loops and/or G-quadruplexes, which seem to be enriched at efficient ORI regions in mammalian species (Cayrou *et al*., 2012; Besnard *et al* 2012; Corless and Gilbert, 2016), and might have a positive role in replication initiation (reviewed in Lombraña *et al*., 2015). The fact that unstable nucleosome particles with high turnover rates are present at active promoter regions can generate such negative tension that, in turn, can increase the probability of ORC binding at those sites during the G1 phase of the cell cycle (Figure 36b, left side). Besides nucleosome removal, another process that can generate supercoiling on the double helix is active transcription (Liu and Wang, 1987; Nelson, 1999; Corless and Gilbert, 2016). This process is capable of generating a high degree of positive supercoiling ahead of the transcription machinery, destabilizing nucleosomes, and, at the same time, generates negative supercoiling behind the transcription machinery (Teves and Henikoff, 2014). This can also be another explanation for the high correlation between ORIs and active transcription units (Figure 36b, right side). Interestingly enough, at most promoters with dispersed TSSs, such as CGI promoters, transcription can initiate in both directions and some reports uncovered two distinct peaks of the active form of RNAPII at those regions (Core *et al*., 2008; He *et al*., 2008; Seila *et al*., 2008; Xu *et al*., 2009; Neil *et al*., 2009). This could fit well with our findings of a bimodal distribution of SNS enrichments at the TSS of promoter-ORIs, further linking both processes. Depending on the directionality of transcription at those promoter ORIs, the ORC protein might take advantage of the negative helical tension generated bind either side, upstream or downstream of the TSS. This would generate distinct peaks of SNS enrichments upon analyzing a cell population, much like the ones that we saw at the promoter regions of *Vps45* and *Haus7* (Figure 14). It would also be in agreement with our recent findings in *Leishmania major*, which suggest that active transcription is a driving force for their nucleosomal organization along the genome and that coupling replication initiation to transcription elongation might be a common solution used for ORI maintenance in this parasite (Lombraña *et al*., 2016).

Combining the reassembly of nucleosomes with the initiation of DNA synthesis at ORIs associated with promoters could facilitate the modulation of chromatin possibly promoting a switch during development and differentiation. Our findings concur with the notion that mammalian replication origins have a high degree of flexibility probably derived from an opportunistic nature of the replication complexes, a property that likely contributes to the robustness of the DNA replication process (Mechali, 2010; Cayrou *et al*, 2011; Sequeira-Mendes and Gómez, 2012; Fragkos *et al*., 2015).

2. Chromatin structure regulates the coordination between the replication and transcription processes

The opportunistic nature of ORC that we hypothesise here constitutes a simple and efficient way to couple nucleosome dynamics to replication initiation, thereby contributing to the establishment and maintenance of the most commonly used (efficient) ORIs. Nonetheless, this opportunistic nature of ORC could put at risk genome integrity, particularly at specific regions which are intrinsically refractory to ORC binding and pre-RC recruitment. In relation with this, some studies suggest that the high sensitivity of common fragile sites (CFSs) to replication stress, such as those studied in human lymphocytes, arises from the combination of late replication and paucity of initiation events at those sites (Letessier *et al*., 2011; Ozeri-Galai., 2011). More recently, other authors reported that genomic regions with poor ORC2 enrichment in human cells often coincide with late replicating domains and large heterochromatic regions. Strikingly, these authors described that those same sites where enriched in CFSs, and regions that are frequently deleted in cancer cells (Miotto *et al*., 2016).

Our findings showing that efficient replication initiation seems to be intimately connected to chromatin structure lead us to ask the question of how alterations in chromatin structure impact the DNA replication programme. To answer this we made use of different genetic systems encompassing a range of chromatin defects and study how those alterations affect the normal replication landscape and replication kinetics *in vivo*.

2.1. Reduced nucleosome occupancies promote increased replication fork rates without compromising the stability of the replication and transcription programs

The rate of replication fork movement is highly dependent on the availability of histones for the correct nucleosome assembly behind the forks (reviewed in Prado and Maya, 2017), and strong inhibition of histone biosynthesis severely affects fork movement, leading to DNA damage and genome instability (Mejlvang *et al*., 2014; Ghule *et al*., 2014). However, in HMGB1-KO cells or SLBP-knockdown cells, which show mild reductions of nucleosome numbers in chromatin, we detected faster replication fork rates (Figure 19 and 35b and d) without major disruption on the global replication initiation landscape or fork stability (Figure 17, 18, 20 and 35a and c). Quite likely, these different results account for the severity of the histone impairment phenotype of previous studies and suggests that cells can tolerate a range of alterations in the histone:DNA ratio. Indeed, we found that HMGB1-KO cells have very similar RNAPII kinetics to their WT counterparts (Figure 34),

while previous work in SLBP knockdown cells detected an increment in the RNAPII elongation rate (Jimeno-González *et al*., 2015). Strikingly, the observed increase in speed of transcription or replication didn't generate any noticeable defects in the division cycle (Figures 17, 18, 21, 22 and 35a and Jimeno-González *et al*., 2015) neither did it increase genome instability in either cell type (Figures 20, 21, 30 and 35c). These results suggest that, in a context of reduced nucleosome occupancies, the molecular machineries move along the DNA molecule at higher rates, as they have a decreased number of natural barriers to overcome along the genome. The fact that we didn't find significant variations in IODs was somewhat unexpected taking into account that usually disturbances in fork progression concomitantly affect the number of activated ORIs in all cellular contexts examined (Anglana *et al*., 2003; Ge *et al*., 2007; Ibarra *et al*., 2008; Courbet *et al*., 2008; Zhong *et al*., 2013; Fu *et al*., 2015). One possibility is that the gentle decrease in nucleosomal occupancy levels occurring in these scenarios is not detectable by the S-phase surveillance mechanisms favouring only the faster processivity of its machinery without compromising stability (Figure 37). However, this does not necessarily mean that reducing nucleosome occupancies has no effect on genomic protection against foreign agents, as there is evidence that HMGB1-KO MEFs are more susceptible to DNA damage by ionizing radiation (Celona *et al*., 2011). This result might indicate that HMGB1-KO cells could have a higher sensitivity for DNA damage detection. Moving forward from here, it would be interesting to perform survival assays and check whether HMGB1-KO cells encompass hypersensitive checkpoints in order to see if they have a more sensitive DNA damage response due to their increased chromatin accessibility, as reported for H1-TKO cells (Murga et al., 2007). It would also be interesting to verify this in a context of induced DNA replication stress to evaluate if this faster response could confer an advantage to the cells when the replication process is challenged.

2.2. Alterations in chromatin structure due to reduced levels of histone H1 severely alter both replication and transcription programs

The chromatin defects caused by a 50% reduction in the amounts of linker histone H1 had a massive impact on the global landscape of replication initiation accompanied by alterations in both the kinetics and the stability of the replication forks, as well as an accumulation of cells arrested in S-phase (Figures 22, 23-26). The widespread distribution of replication intermediates revealed by SNS-seq in H1-TKO cells, together with the observed alterations at fork dynamics, argues in favour of chromatin compaction being of major importance for the DNA replication process. Furthermore, we observed parallel

abnormalities in the transcription cycle in those cells, with defects in RNAPII dynamics and anomalous accumulation of R-loops in chromatin (Figures 27, 28 and 29a and b). According to other reports, there are no significant differences in mRNA abundances between WT and H1-TKO mES cells (Fan *et al*., 2005; Murga *et al*., 2007; Geeven *et al*., 2015). Therefore, our results likely imply that these transcriptional abnormalities are the result of increased non-productive transcription initiation due to the altered chromatin structure of H1-depleted cells. Defects in transcription were also accompanied by replicative stress, as seen by the accumulation of double strand breaks (DSBs), especially at the early stages of S-phase (Figure 29c). Our results partly explain the finding that mES H1-TKO cells encompass a constitutively induced DNA damage checkpoint (Murga *et al*., 2007). Functional studies of transcription inhibition with specific drugs fully recover the replication phenotype of H1-TKO cells (Figures 31 and 33), demonstrating that the movement of the replication machinery is severely challenged by collisions with the transcription machinery or transcription-mediated structures that are left unresolved in the genome. Previous reports described that replicative stress as well as genomic instability are indeed consequences that can arise from the deregulation of transcription (Tuduri *et al*., 2009; Kotsantis *et al*., 2016; Stork *et al*., 2016). Likewise, abnormal accumulation of R-loop structures is able to mediate chromatin compaction at certain genomic regions (Castellano-Pozo *et al*., 2013; Skourti-Stathaki *et al*., 2014), and this might be another factor that can induce the replication instability seen in this cells, further contributing to slow fork velocities and increasing the probability of fork collapse.

Other experiments from our laboratory found that the temporal replication program of H1-TKO cells was globally preserved when compared to their WT counterparts, with only few regions shifting replication timing. The microarrays used in this analysis didn't include heterochromatic regions such as centromeres and telomeres and for this reason we could not analyse if replication timing was specifically affected at those regions. Nonetheless, the lack of significant changes in late-replicating segments of the genome as well as in the S-phase distribution of EdU patterns (Figure 26d) argues against any major changes in heterochromatic regions. In support of our results, the global 3D topological organization of the genome was found to be globally conserved between WT and H1-TKO cells (Geeven *et al*., 2015).

2.3. Chromatin structure impact on genome stability

Histone H1 is commonly described as a general repressor of transcription, achieving this by limiting the accessibility of protein complexes to the DNA molecule. Despite this general idea, there are not many studies addressing in detail the layers of

regulation that different histone H1 variants or their posttranslational modifications are able to impose (Harshman *et al*, 2013). This is mainly due to the fact that linker histones are not so well conserved between eukaryotic species and thus, many of the functionalities of their different variants and PTMs are still not yet fully investigated in spite of the numerous evidences of their importance (Shen *et al.*, 1995; Patterton *et al*., 1998; Ramón *et al*., 2000; Zlatanova *et al*, 2000; Hellauer *et al*., 2001; Harshman *et al*, 2013). For example, histone H1 phosphorylation can promote transcriptional activation of the mouse mammary tumour virus promoter, implying that a rescue of transcription is possible when certain modifications of H1 occur (Lee and Archer, 1998; Bhattacharjee *et al*., 2001; Koop *et al*., 2003). This argues that histone H1 might also be important for the normal regulation of the initial steps of transcription either through direct or indirect interaction with transcription factors. Our findings support this view unveiling yet another regulatory role of histone H1 in preventing accumulation of non-canonical DNA structures deriving from abnormal transcriptional activity. Although this finding needs further investigation, especially in regards to the specific interactions of histone H1 with the transcription machinery and how this contributes to its normal and coordinated kinetics, our results clearly indicate that the action of linker histones is vital to prevent R-loop-induced replication stress.In turn, altered transcription kinetics due to the defective chromatin of H1-depleted cells imposes the slowing or irreversible stop of fork progression, causing the firing of dormant ORIs in order to complete replication. This is a similar response to the one reported upon depletion in the cellular nucleotide pool (Anglana *et al*., 2003; Ge *et al*., 2007; Ibarra *et al*., 2008; Courbet *et al*., 2008), and reminiscent of the frequent fork stalling occurring at AT-rich sequences along the common fragile site FRA16C that, even under normal growth conditions, leads to activation of additional origins to ensure the complete duplication of that region (Letessier *et al*., 2011). Another likely possibility contributing to the altered SNS pattern and the accumulation of stalled replication intermediates found in H1-TKO cells is R-loop formation itself. As mentioned before, there is increasing evidence linking R-loop and G4 structures to ORIs, especially at CpG islands, were the most efficient ORIs tend to locate (reviewed in Lombraña *et al*., 2015). It is possible that the RNA strand of the R-loops might serve as a primer for replication elongation, a similar scenario to that of mitocondrial DNA replication, in which the leading-strand origin is coupled to transcription through the formation of an RNA:DNA hybrid (Chang and Clayton 1985; Chang *et al*., 1985) (Figure 37). It would be interesting to develop genomic maps of R-loop and RNAPII occupancies and integrate those results to the ones obtained by SNS-seq in both WT and H1-TKO mES cells. This could allow us to characterize the relationship between those three features; replication initiation,

transcription initiation and R-loop formation, which might further reveal the opportunistic nature of the replication process in mammalian cells.

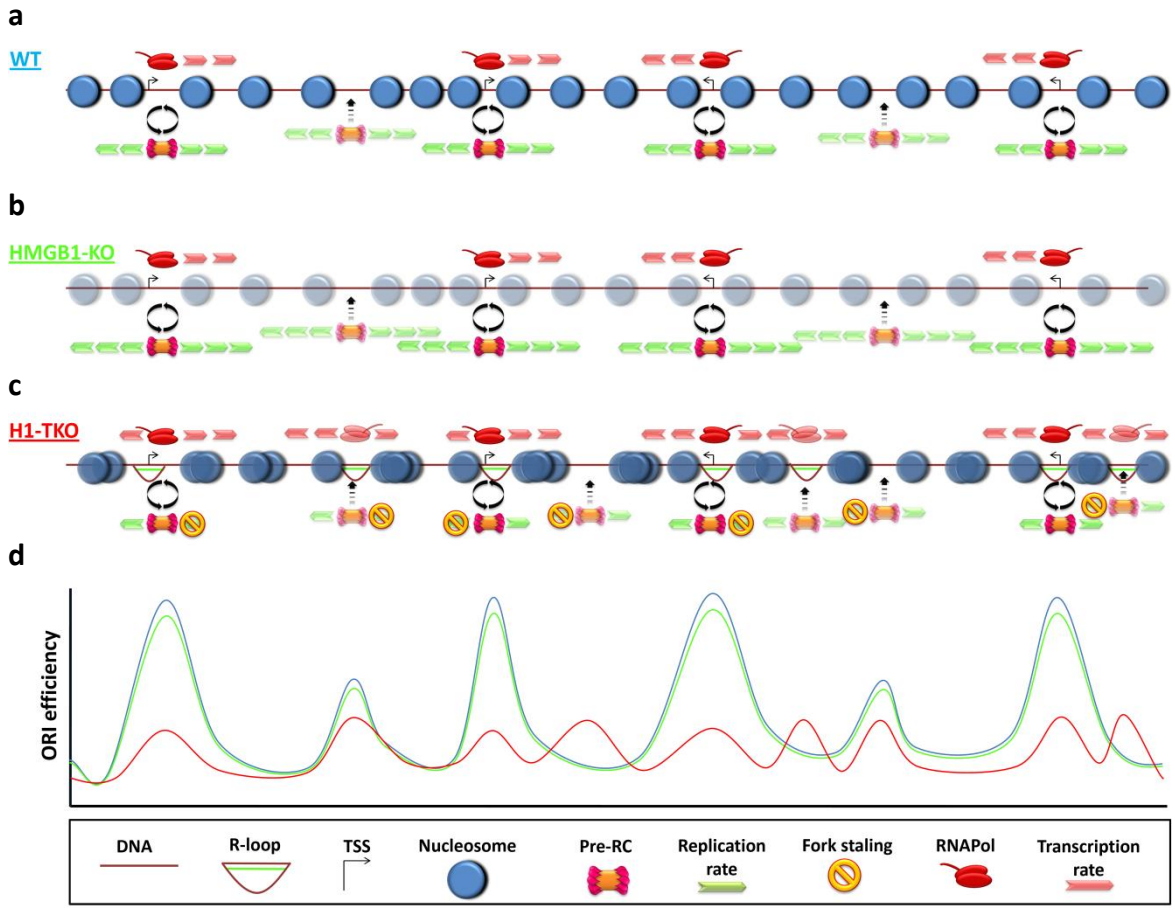
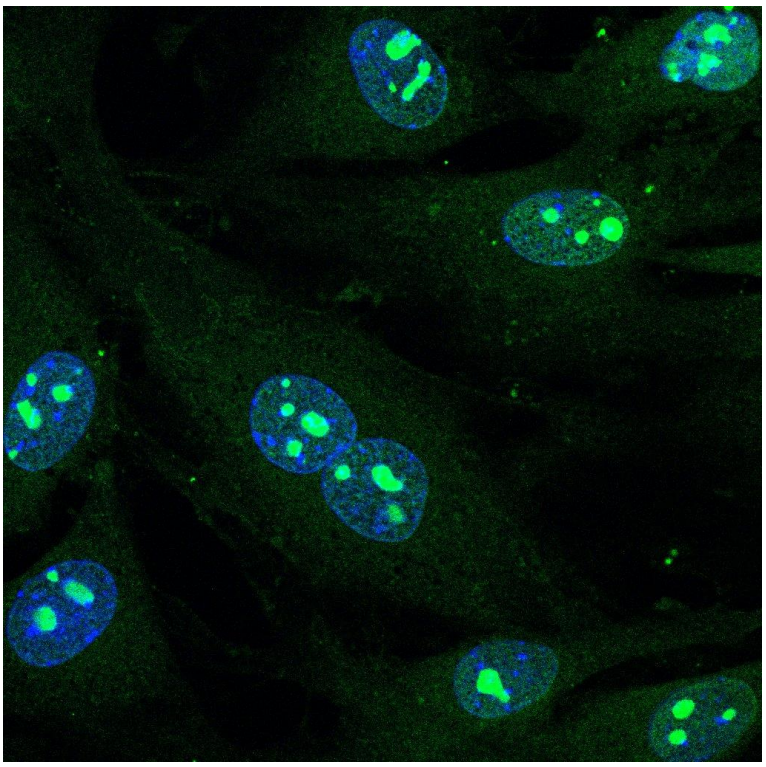


Figure 37: Schematic illustration depicting the chromatin environments of HMGB1-KO and H1-TKO cells. **(a)** In WT cells the replication and transcription machineries progress at similar rates (represented by the number of green and red arrows, respectively) without major conflicts between them. The most efficient ORIs are associated with transcriptional initiation sites, while other low efficiency ORIs (light-couloured pre-RCs) can also be activated to ensure complete DNA duplication. **(b)** When cells lack HMGB1 the levels of nucleosome occupancies drop (light-couloured nucleosomes), without major alterations in nucleosome positioning (Celona *et al*., 2011). The replication initiation sites are largely conserved relative to WT cells, although forks move faster. **(c)** In cells with reduced H1 levels, chromatin-based processes are severely affected: transcription is deregulated and R-loop structures tend to accumulate in the genome. Replication forks slow down, with higher probability of collapsing and additional dormant ORIs are activated. There is also the possibility that R-loop structures serve as a template for non-canonical replication initiation. **(d)** Relative levels of ORI efficiency in the various chromatin scenarios (blue, WT; green, HMGB1-KO; red, H1-TKO).

Collectively, our results unveil the fundamental importance of chromatin structure for regulating, both locally and globally, the genomic processes occurring in the nucleus, as well as maintaining the normal equilibrium between them. On one hand, the strong reduction in histone H1 levels occurring in the cells used in this work causes impaired RNAPII kinetics, miss-regulated R-loop resolution, and increases transcription-replication conflicts. On the other side, moderate reductions in nucleosome numbers, as occurring in

HMGB1-KO and SLBP-knockdown cells, allowed faster progression of replication and transcription machineries to different extents, implying that this co-regulation might act as a mechanism to decrease the probability of encounters between both machineries and thus avoid replicative stress. This work reinforces the notion that advances in the field of Epigenetics are of extreme importance to uncover new knowledge and achieve new breakthroughs in important areas like cellular aging, developmental disorders and cancer research. This is evidenced by the fact that numerous neurodevelopmental and degenerative disorders, immunodeficiency diseases and many different types of cancer are intimately related to some form of chromatin deregulation (Cassidy and Schwartz, 1998; Jin and Warren, 2000; Iwase *et al*., 2007; Gibbons *et al*., 2008; Cabianca *et al*., 2012; Tsurusaki *et al*., 2012; Lazo-Gomez *et al*., 2013; Lu *et al*., 2013). Indeed, today an increasing number of epigenetic regulator inhibitors are already in clinical use or under development (Mirabella *et al*., 2016), reinforcing the therapeutically importance of studying systems with different chromatin alterations, like the ones presented in this work.

Conclusions - Conclusiones



The conclusions that can be drawn from the work developed in this Doctoral Thesis are the following:

1. Mammalian ORIs can encompass different nucleosome configurations. However, replication start sites occur at positions of high-nucleosome occupancy in all cases.

2. Upon S-phase entry, nucleosomes positioned at the initiation site are immediately remodeled, as exemplified by the LaminB2 ORI.

3. Replication initiation patterns reflect nucleosomal architecture: higher origin efficiencies correlate with stronger nucleosome positioning.

4. Decreased nucleosome numbers allow faster replication fork progression without altering the initiation landscape or compromising replication stability.

5. Alterations in chromatin compaction caused by reduced levels of histone H1 severely alter the replication initiation landscape, triggering the accumulation of stalled forks and DNA damage as a consequence of enhanced transcription-replication conflicts.

6. In the absence of the correct amounts of histone H1 in chromatin the dynamics of RNAPII is defective, causing non-productive transcription initiation and R-loop accumulation.

Las conclusiones que hemos obtenido del trabajo desarrollado en esta tesis doctoral son las siguientes:

1. Los ORIs en células de mamíferos pueden contener distintos patrones de arquitectura nucleosomal. Sin embargo, la síntesis del DNA ocurre en posiciones de alta ocupación nucleosomal en todos los casos.

2. Al comienzo de la fase S, los nucleosomas posicionados en los sitios de iniciación son inmediatamente remodelados, como demuestra el análisis del ORI humano LaminB2.

3. Los patrones de iniciación de la replicación reflejan la arquitectura nucleosomal: existe una fuerte correlación entre la eficiencia de los ORIs y el grado de posicionamiento de los nucleosomas.

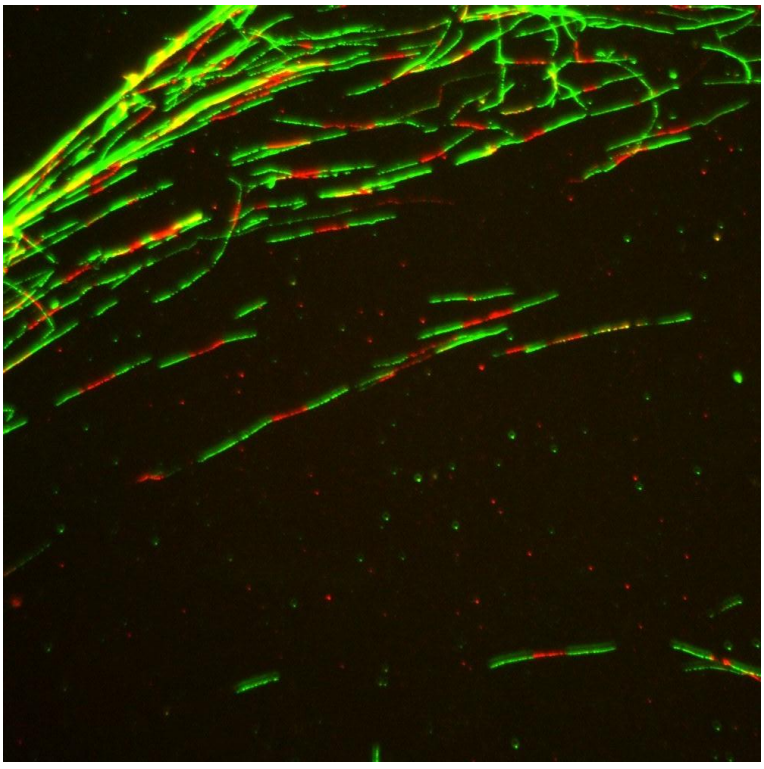
4. Reducciones en el número de nucleosomas en el genoma favorecen el incremento en la velocidad de avance de las horquillas de replicación, sin generar alteraciones en el paisaje global de iniciación de la replicación ni comprometer la estabilidad de las horquillas.

5. Alteraciones en la compactación de la cromatina derivados de la reducción en los niveles de histona H1 alteran drásticamente el paisaje replicativo, generando paradas en las horquillas de replicación y daño en el DNA como consecuencia de conflictos entre los procesos de replicación y transcripción.

6. La disminución en los niveles de histona H1 alteran la dinámica de transcripción de la RNAPII, causando iniciaciones de transcripción no productivas y acumulación de R-loops.

.

Bibliography



Abdurashidova G, Deganuto M, Klima R, Riva S, Biamonti G, Giacca M, Falaschi A (2000). Start sites of bidirectional DNA synthesis at the human lamin B2 origin. Science 287:2023–2026

Abdurashidova G, Danailov MB, Ochem A, Triolo G, Djeliova V, Radulescu S, Vindigni A, Riva S, Falaschi A (2003). Localization of proteins bound to a replication origin of human DNA along the cell cycle. EMBO J 22: 4294–4303.

Adam M, Robert F, Larochelle M, Gaudreau L (2001). H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. Mol Cell Biol 21:6270-9.

Agresti A, Bianchi ME (2003). HMGB proteins and gene expression. Curr Opin Genet Dev 13:170-8.

Aguilera A (2002). The connection between transcription and genomic instability. EMBO J 21:195-201.

Aguilera A, Garcia-Muse T (2012). R loops: from transcription by products to threats to genome stability. Molecular cell 46:115–24.

Aladjem M I, Falaschi A, Kowalski D (2006). Eukaryotic DNA replication origins. Cold Spring Harbor 31–61.

Altaf M, Auger A, Covic M, Cote J (2009) Connection between histone H2A variants and chromatin remodeling complexes. Biochem Cell Biol 87:35–50.

Allis CD, Jenuwein T, Reinberg D, Caparros ML (2007). Epigenetics, 1st ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Andersson U, Wang H, Palmblad K, Aveberger AC, Bloom O, Erlandsson-Harris H, Janson A, Kokkola R, Zhang M, Yang H, Tracey KJ (2000). High mobility group 1 protein (HMG-1) stimulates proinflammatory cytokine synthesis in human monocytes. J Exp Med 192:565–570.

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

Annunziato, A. (2008) DNA Packaging: Nucleosomes and Chromatin. Nature Education 1:26

Arias EE, Walter JC (2007). Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. Genes Dev 21:497-518.

Arya G, Maitra A, Grigoryev SA (2010). A structural perspective on the where, how, why, and what of nucleosome positioning. J Biomol Struct Dyn 27:803-20.

Audit B, Zaghloul L, Vaillant C, Chevereau G, d´Aubenton-Carafa Y, Thermes C, Arneodo A (2009). Open chromatin encoded in DNA sequence is the signature of master replication origins in human cells. Nucleic Acids Res 37:6064–6075.

Bandiera A, Bonifacio D, Manfioletti G, Mantovani F, Rustighi A, Zanconati F, Fusco A, Di Bonito L, Giancotti V (1998). Expression of HMGI(Y) proteins in squamous intraepithelial and invasive lesions of the uterine cervix. Cancer Res 58:426-431

Barlow JH, Nussenzweig, A (2014). Replication initiation and genome instability: a crossroads for DNA and RNA synthesis. Cell Mol Life Sci 71:4545–4559.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007). High-resolution profiling of histone methylations in the human genome. Cell 129:823–837.

Bassett A, Cooper S, Wu C, Travers A (2009). The folding and unfolding of eukaryotic chromatin. Curr Opin Genet Dev 19:159-65.

Batada NN, Hurst LD (2007). Evolution of chromosome organization driven by selection for reduced gene expression noise. Nat. Genet. 39: 945–949.

Bates AD and Maxwell A (2005). DNA topology. Oxford University Press, New York

Bayona-Feliu A, Casas-Lamesa A, Carbonell A, Climent-Cantó P, Tatarski M, Pérez-Montero S, Azorín F, Bernués J (2016). Histone H1: Lessons from *Drosophila*. Biochim Biophys Acta 1859:526-32.

Bell SP (2002). The origin recognition complex: From simple origins to complex functions. Genes Dev 16:659–672.

Bell SP, Dutta A (2002). DNA replication in eukaryotic cells. Annu Rev Biochem 71:333–374.

Bell SP, Mitchell J, Leber J, Kobayashi R, Stillman B (1995). The multi domain structure of Orc1p reveals similarity to regulators Of DNA replication and transcriptional silencing. Cell 83:563-568.

Bensaude O (2011). Inhibiting eukaryotic transcription Which compound to choose? How to evaluate its activity? Transcription 2:103–108.

Bernstein BE, Liu CL, Humphrey EL, Perlstein EO, Schreiber SL (2004). Global nucleosome occupancy in yeast. Genome Biol. 5:R62.

Berstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Weirnig M, Plath K (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 JM, Lemaitre JM (2012). Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nat Struct Mol Biol. 19:837-844.

Bhattacharjee RN, Banks GC, Trotter KW, Lee HL, Archer TK (2001). Histone H1 phosphorylation by Cdk2 selectively modulates mouse mammary tumor virus transcription through chromatin remodeling. Mol Cell Biol 21:5417-25.

Bianchi ME and Agresti A (2005). HMG proteins: dynamic players in gene regulation and differentiation. Curr Opin Genet Dev 15:496-506.

Blow JJ, Ge XQ (2008). Replication forks, chromatin loops and dormant replication origins. Genome Biol 9:244.

Boguslawski SJ, Smith DE, Michalak MA, Mickelson KE, Yehle CO, Patterson WL and Carrico RJ (1986). Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. J Immunol Methods 89:123-130.

Bonaldi T, Längst G, Strohner R, Becker PB, Bianchi ME (2002). The DNA chaperone HMGB1 facilitates ACF/CHRAC-dependent nucleosome sliding. EMBO J 21:6865-73.

Buratowski S (2008). Transcription: Gene expression - where to start? Science 322:1804-1805.

Bustin M (1999). Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. Mol Cell Biol 19:5237–46.

Bustin M (2001). Revised nomenclature for high mobility group (HMG) chromosomal proteins. Trends Biochem Sci 26:152-3.

Cabianca DS, Casa V, Bodega B, XynosA, Ginelli E, TanakaY, Gabellini D (2012). A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. Cell 149:819–831.

Cadoret JC, Meisch F, Hassan-Zadeh V, Luyten I, Guillet C, Duret L, Quesneville H, Prioleau MN (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc. Natl. Acad. Sci. USA 105:15837–15842.

Calogero S, Grassi F, Aguzzi A, Voigtländer T, Ferrier P, Ferrari S, Bianchi ME (1999). The lack of chromosomal protein Hmg1 does not disrupt cell growth, but causes lethal hypoglycaemia in newborn mice. Nature Genet 22:276-280.

Castellano-Pozo M, Santos-Pereira JM, Rondón AG, Barroso S, Andújar E, Pérez-Alegre M, García-Muse T and Aguilera A (2013). R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. Mol Cell 52: 583-590.

Cassidy SB, Schwartz S (1998) Prader-Willi and Angelman syndromes. Disord Genomic Imprinting Med 77:140–151

Cayrou C, Coulombe P, Puy A, Rialle S, Kaplan N, Segal E, Méchali M (2012). New insights into replication origin characteristics in metazoans. Cell Cycle 11:658-67

Cayrou C, Coulombe P, Vigneron A, Stanojcic S, Ganier O, Peiffer I, Rivals E, Puy A, Laurent-Chabalier S, Desprat R, Méchali M (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Res 21:1438-49.

Celona B, Weiner A, Di Felice F, Mancuso FM, Cesarini E, Rossi RL, Gregory L, Baban D, Rossetti G, Grianti P, Pagani M, Bonaldi T, Ragoussis J, Friedman N, Camilloni G, Bianchi ME, Agresti A (2011). Substantial histone reduction modulates genomewide nucleosomal occupancy and global transcriptional output. PLoS Biol 9:e1001086.

Chang DD and Clayton DA (1985). Priming of human mitochondrial DNA replication occurs at the light-strand promoter. Proc Natl Acad Sci USA 82:351–355.

Chang DD, Hauswirth WW and Clayton DA (1985). Replication priming and transcription initiate from precisely the same site in mouse mitochondrial DNA. EMBOJ. 4, 1559–1567.

Cassidy SB, Schwartz S (1998). Prader-Willi and Angelman syndromes. Disord Genomic Imprinting Med 77:140–151.

Chang F, Theis JF, Miller J, Nieduszynski CA, Newlon CS, Weinreich M (2008). Analysis of chromosome III replicators reveals an unusual structure for the ARS318 silencer origin and a conserved WTW sequence within the origin recognition complex binding site. Mol Cell Biol 28:5071–5081.

Cheng LZ, Workman JL, Kingston RE, Kelly TJ (1992). Regulation of DNA replication in vitro by the transcriptional activation domain of GAL4-VP16. Proc Natl Acad Sci USA 89:589–593.

Chiappetta G, Tallini G, De Biasio MC, Manfioletti G, Martinez-Tello FJ, Pentimalli F, de Nigris F, Mastro A, Botti G, Fedele M, Berger N, Santoro M, Giancotti V, Fusco A (1998). Detection of high mobility group I HMGI(Y) protein in the diagnosis of thyroid tumors: HMGI(Y) expression represents a potential diagnostic indicator of carcinoma. Cancer Res 58:4193-8.

Conti C, Saccà B, Herrick J, Lalou C, Pommier Y, Bensimon A (2007). Replication fork velocities at adjacent replication origins are coordinately modified during DNA replication in human cells. Mol Biol Cell 18: 3059-67.

Core LJ, Waterfall JJ, Lis JT (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322: 1845–8.

Corless S and Gilbert N (2016). Effects of DNA supercoiling on chromatin architecture. Biophys Rev 8:245–258

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.

Creyghton MP, Markoulaki S, Levine SS, Hanna J, Lodato MA, Sha K, Young RA, Jaenisch R, Boyer LA (2008). H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. Cell 135:649-61

Cvetic C, Walter JC (2005). Eukaryotic origins of DNA replication: could you please be more specific? Seminars in Cell & Developmental Biology 16:343–353.

Dai J, Chuang RY, Kelly TJ (2005). DNA replication origins in the Schizosaccharomyces pombe genome. Proc. Natl Acad. Sci. USA 102:337–342.

Danis E, Brodolin K, Menut S, Maiorano D, Girard-Reydet C (2004). Specification of a DNA replication origin by a transcription complex. Nat Cell Biol 6:721–730.

Das D, Scovell WM (2001). The binding interaction of HMG-1 with the TATA-binding protein/TATA complex. J Biol Chem 276:32597-605.

Deal RB, Henikoff JG, Henikoff S (2010). Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. Science 328: 1161–1164

Deal RB, Henikoff S (2010). Capturing the dynamic epigenome. Genome Biol 11:218–225

Dellino GI, Cittaro D, Piccioni R, Luzi L, Banfi S, Segalla S, Cesaroni M, Mendoza-Maldonado R, Giacca M, Pelici PG (2013). Genomewide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. Genome Res 23:1–11

DesJarlais R, Tummino PJ (2016). Role of Histone-Modifying Enzymes and Their Complexes in Regulation of Chromatin Biology. Biochemistry 55:1584-99.

Diffley JF (2004). Regulation of early events in chromosome duplication. Curr. Biol, 14:R778-R786.

Dijkwel PA, Mesner LD, Levenson VV, d'Anna J, Hamlin JL (2000). Dispersive initiation of replication in the Chinese hamster rhodopsin locus. Exp Cell Res 256:150–157

Diffley JF, Stillman B (1989). Similarity between the transcriptional silencer binding proteins ABF1 and RAP1. Science 246:1034–1038.

Dijkwel P, Hamlin JL (1995). The Chinese hamster dihydrofolate reductase origin consists of multiple potential nascent-strand start sites. Mol Cell Biol 15:3023–3031

Dion MF, Kaplan T, Kim M, Buratowski S, Friedman N, Rando OJ (2007). Dynamics of replication-independent histone turnover in budding yeast. Science 315:1405–1408.

Donaldson AD (2005). Shaping time: Chromatin structure and the DNA replication programme. Trends Genet 21:444–449.

Dowell NL, Sperling AS, Mason MJ, Johnson RC (2010). Chromatin-dependent binding of the S. cerevisiae HMGB protein Nhp6A affects nucleosome dynamics and transcription. Genes Dev 24:2031–2042.

Draizen EJ, Shaytan AK, Marino-Ramirez L, Talbert PB, Landsman D, Panchenko AR (2016). HistoneDB 2.0: a histone database with variants—an integrated resource to explore histones and their variants. PMID: 26989147.

Eaton ML, Galani K, Kang S, Bell SP, MacAlpine DM (2010). Conserved nucleosome positioning defines replication origins. Genes Dev. 24:748–753.

Ellerman JE, Brown CK, de Vera M, Zeh HJ, Billiar T, Rubartelli A, Lotze MT (2007). Masquerader: High mobility group box-1 and cancer. Clin Cancer Res 13:2836–2848.

Fan Y, Sirotkin A, Russell RG, Ayala J, Skoultchi AI (2001). Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the H1(0) replacement subtype. Mol Cell Biol 21:7933-43.

Fan Y, Nikitina T, Morin-Kensicki EM, Zhao J, Magnuson TR, Woodcock CL, Skoultchi AI (2003). H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. Mol Cell Biol 23:4559-72.

Fan Y, Nikitina T, Zhao J, Fleury TJ, Bhattacharyya R, Bouhassira EE, Stein A, Woodcock CL, Skoultchi AI (2005). Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. Cell 123:1199-212.

Fedele M, Bandiera A, Chiappetta G, Battista S, Viglietto G, Manfioletti G, Casamassimi A, Santoro M, Giancotti V, Fusco A (1996). Human colorectal carcinomas express high levels of high mobility group HMGI(Y) proteins. Cancer Res 56:1896–1901.

Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, Lubling Y, Widom J, Segal E (2008). Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. PLoS Comput Biol 4:e1000216.

Fragkos M, Ganier O, Coulombe P, Méchali M (2015). DNA replication origin activation in space and time. Nat Rev Mol Cell Biol 16:360-74.

Francis NJ, Kingston RE, Woodcock CL (2004). Chromatin compaction by a polycomb group protein complex. Science 306:1574-7.

Fu Y, Sinha M, Peterson CL, Weng Z (2008). The Insulator Binding Protein CTCF Positions 20 Nucleosomes around Its Binding Sites across the Human Genome. PLoS Genet 4:e1000138.

Fu H, Martin MM, Regairaz M, Huang L, You Y, Lin CM, Ryan M, Kim R, Shimura T, Pommier Y and Aladjem MI (2015). The DNA repair endonuclease Mus81 facilitates fast DNA replicationin the absence of exogenous damage. Nat Commun 6: 6746.

Gan W, Guan Z, Liu J, Gui T, Shen K, Manley JL, Li X. (2011). R-loop-mediated genomic instability is caused by impairment of replication fork progression. Genes Dev 25:2041–2056.

García-Muse T, Aguilera A (2016). Transcription-replication conflicts: how they occur and how they are resolved. Nat Rev Mol Cell Biol 17:553-63.

García-Rubio ML, Pérez-Calero C, Barroso SI, Tumini E, Herrera-Moyano E, Rosado IV, Aguilera A (2015). The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. PLoS Genet 11:e1005674.

Ge XQ, Jackson DA and Blow JJ (2007). Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. Genes Dev 21:3331-3341

Geeven G, Zhu Y, Kim BJ, Bartholdy BA, Yang SM, Macfarlan TS, Gifford WD, Pfaff SL, Verstegen MJ, Pinto H, Vermunt MW, Creyghton MP, Wijchers PJ, Stamatoyannopoulos JA, Skoultchi AI and de Laat W (2015). Local compartment changes and regulatory landscape alterations in histone H1-depleted cells. Genome Biol 16:289

Ghosh M, Liu G, Randall G, Bevington J, Leffak M (2004). Transcription factor binding and induced transcription alter chromosomal c-myc replicator activity. Mol Cell Biol 24:10193–10207.

Ghule PN, Xie RL, Medina R, Colby JL, Jones SN, Lian JB, Stein JL, van Wijnen AJ and Stein GS (2014). Fidelity of histone gene regulation is obligatory for genome replication and stability. Mol Cell Biol 34:2650-2659

Giacca M, Zentilin L, Norio P, Diviacco S, Dimitrova D, Contreas G, Biamonti G, Perini G, Weighardt F, Riva S, Falaschi A (1994). Fine mapping of a replication origin of human DNA. Proc Natl Acad Sci USA 91:7119–7123

Giavara S, Kosmidou E, Hande MP, Bianchi ME, Morgan A, d'Adda di Fagagna F, Jackson SP (2005). Yeast Nhp6A/B and mammalian Hmgb1 facilitate the maintenance of genome stability. Curr Biol 15:68-72.

Gibbons RJ, Wada T, Fisher CA, Malik N, Mitson MJ, Steensma DP, Fryer A, Goudie DR, Krantz ID, Traeger-Synodinos J (2008). Mutations in the chromatin-associated protein ATRX. Hum Mutat 29:796–802.

Gilbert DM (2004). In search of the holy replicator. Natl Rev 5:848–855.

Gilbert DM (2012). Replication origins run (ultra) deep. Nat Struct Mol Biol 19:740-2.

Ginno PA, Lott PL, Christensen HC, Korf I and Chedin F (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell 45:814-825.

Ginno PA, Lim YW, Lott PL, Korf I and Chedin F (2013). GC skew at the 5´and 3´ends of human cells links R-loop formation to epigenetic regulation and transcription termination. Genome Res 23:1590-1600.

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

Gómez-Acuña LI, Fiszbein A, Alló M, Schor IE, Kornblihtt AR (2013). Connections between chromatin signatures and splicing. Wiley Interdiscip Rev RNA 4:77-91.

Gong QH, McDowell JC, Dean A (1996). Essential role of NF-E2 in remodeling of chromatin structure and transcriptional activation of the epsilon-globin gene in vivo by 5' hypersensitive site 2 of the beta-globin locus control region. Mol Cell Biol 16:6055–6064.

González S, García A, Vázquez E, Serrano R, Sánchez M, Quintales L, Antequera F (2016). Nucleosomal signatures impose nucleosome positioning in coding and noncoding sequences in the genome. Genome Res 26:1532-1543.

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

Hamperl S, Cimprich KA (2016). Conflict Resolution in the Genome: How Transcription and Replication Make It Work. Cell 167:1455-1467.

Happel N, Doenecke D (2009). Histone H1 and its isoforms: contribution to chromatin structure and function. Gene 431:1-12.

Hardy S, Jacques PE, Gévry N, Forest A, Fortin ME, Laflamme L, Gaudreau L, Robert F (2009). The Euchromatic and Heterochromatic Landscapes Are Shaped by Antagonizing Effects of Transcription on H2A.Z Deposition. PLoS Genet 5: e1000687.

Harshman SW, Young NL, Parthun MR, Freitas MA (2013). H1 histones: current perspectives and challenges. Nucleic Acids Res 41:9593-609.

He Y, Vogelstein B, Velculescu VE, Papadopoulos N, et al. (2008). The antisense transcriptomes of human cells. Science 322:1855–7.

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

Hellauer K, Sirard E, Turcotte B (2001). Decreased expression of specific genes in yeast cells lacking histone H1. J Biol Chem 276:13587-92.

Helmrich A, Ballarino M, Nudler E, Tora L (2013). Transcription-replication encounters, consequences and genomic instability. Nat Struct Mol Biol 20:412-8.

Henikoff S (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. Nat Rev Genet 9:15-26.

Herrera-Moyano E, Mergui X, Garcia-Rubio ML, Barroso S, Aguilera A (2014) The yeast and human FACT chromatin-reorganizing complexes solve R‐loop mediated transcription-replication conflicts. Genes Dev. 28:735–748

Hiratani I, Ryba T, Itoh M, Yokochi T, Schwaiger M, Chang CW, Lyou Y, Townes TM, Schübeler D and Gilbert DM (2008). Global reorganization of replication domains during embryonic ítem cell differentiation. PLoS Biol 6: e245.

Hnilicová J and Staněk D (2011). Where splicing joins chromatin. Nucleus 2:182–188.

Hoshina S, Yura K, Teranishi H, Kiyasu N, Tominaga A, Kadoma H, Nakatsuka A, Kunichika T, Obuse C, Waga S (2013). Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. J Biol Chem 288:30161-71.

Hurst LD, Pal C, Lercher MJ (2004). The evolutionary dynamics of eukaryotic gene order, Nat. Rev. Genet. 5:299–310.

Hurst LD, Williams EJ, Pal C (2002). Natural selection promotes the conservation of linkage of co-expressed genes, Trends Genet. 18:604–606.

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 USA 105:8956–8961.

Infante JJ, Law GL, Young ET (2012). Analysis of nucleosome positioning using a nucleosome-scanning assay. Methods Mol Biol 833:63–87.

Ioffe E, Liu Y, Bhaumik M, Poirier F, Factor SM, Stanley P (1995). WW6: An embryonic stem cell line with an inert genetic marker that can be traced in chimeras. Proc Natl Acad Sci 92:7357–7361.

Ioshikhes I, Bolshoy A, Derenshteyn K, Borodovsky M, Trifonov EN (1996). Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences. J. Mol. Biol. 262:129–139.

Iwase S, Lan F, Bayliss P, de la Torre-Ubieta L, Huarte M, Qi HH, Whetstine JR, Bonni A, Roberts TM, Shi Y (2007). The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. Cell 128:1077–1088.

Izban MG, Luse DS (1992). Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. J. Biol. Chem. 267:13647–13655.

Jackson DA 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 J, Cuzin F (1963). On the regulation of DNA replication in bacteria. Cold Spring Harbor Symp Quant Biol 28:329–348.

Jenuwein T, Allis CD (2001). Translating the Histone Code. Science 293:1074-80.

Jiang CZ, Pugh BF (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat. Rev. Genet. **10**:161-172.

Jimeno-González S, Payán-Bravo L, Muñoz-Cabello AM, Guijo M, Gutierrez G, Prado F, Reyes JC (2015). Defective histone supply causes changes in RNA polymerase II elongation rate and cotranscriptional pre-mRNA splicing. Proc Natl Acad Sci U S A 112:14840-5.

Jin C, Felsenfeld G (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev 21:1519–1529.

Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K, Felsenfeld G (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nuclesome-free regions' of active promoters and other regulatory regions. Nat genet 41:941-945.

Jin P, Warren ST (2000). Understanding the molecular basis of fragile X syndrome. Hum Mol Genet 9:901–908

Jonkers I, Kwak H and Lis JT (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife 3:e02407.

Joshi SR, Sarpong YC, Peterson RC and Scovell WM (2012). Nucleosome dynamics: HMGB1 relaxes canonical nucleosome structure to facilitate estrogen receptor binding. Nucleic Acids Res 20:10161-10171.

Kamakaka RT, Biggins S (2005). Histone variants: deviants? Genes Dev. 19:295–310.

Keller C, Ladenburger EM, Kremer M, Knippers R (2002). The origin recognition complex marks a replication origin in the human TOP1 gene promoter. J Biol Chem 277:31430–31440

Knott SR, Viggiani CJ, Tavare S, Aparicio OM (2009). Genome-wide replication profiles indicate an expansive role for Rpd3L in regulating replication initiation timing or efficiency, and reveal genomic loci of Rpd3 function in Saccharomyces cerevisiae. Genes & Dev 23:1077–1090.

Koop R, Di Croce L, Beato M (2003). Histone H1 enhances synergistic activation of the MMTV promoter in chromatin. EMBO J 22:588-99.

Kornberg RD (1974). Chromatin structure: a repeating unit of histones and DNA. Science 184:868–871.

Kotsantis P, Silva LM, Irmscher S, Jones RM, Folkes L, Gromak N and Petermann E (2016). Increased global transcriptional activity as a mechanism of replication stress in cancer. Nat Comm 7:13087.

Koutroubas G, Merika M, Thanos D (2008). Bypassing the requirements for epigenetic modifications in gene transcription by increasing enhancer strength. Mol. Cell. Biol 28:926-938.

Kouzarides T (2007). Chromatin modifications and their function. Cell 128:693–705.

Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, Adli M, Kasif S, Ptaszek LM, Cowan CA, Lander ES, Koseki H, Bernstein BE (2008). Genome-wide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet 4:e1000242.

Kuo LJ, Yang LX (2008). Gamma-H2AX - a novel biomarker for DNA double-strand breaks. In Vivo 22:305-9.

Kuo AJ, Song J, Cheung P, Ishibe-Murakami S, Yamazoe S, Chen JK, Patel DJ, Gozani O (2012). The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier–Gorlin syndrome. Nature 484:115–119

Ladenburger EM, Keller C, Knippers R (2002). Identification of a binding region for human origin recognition complex proteins 1 and 2 that coincides with an origin of DNA replication. Mol Cell Biol 22:1036–1048.

Längst G and Manelyte L (2015). Chromatin Remodelers: From Function to Dysfunction. Genes 6:299-32.

Launholt D, Merkle T, Houben A, Schulz A, Grasser KD (2006). Arabidopsis chromatin-associated HMGA and HMGB use different nuclear targeting signals and display highly dynamic localization within the nucleus. Plant Cell 18:2904-18.

Lazo-Gomez R, Ramirez-Jarquin UN, Tovar YRLB, Tapia R (2013). Histone deacetylases and their role in motor neuron degeneration. Front Cell Neurosci 7:243.

Lee HL, Archer TK (1998). Prolonged glucocorticoid exposure dephosphorylates histone H1 and inactivates the MMTV promoter. EMBO J 17:1454-66.

Lee CK, Shibata Y, Rao B, Strahl BD, Lieb JD (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet 36:900–905.

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C (2007). A highresolution atlas of nucleosome occupancy in yeast, Nat. Genet. 39:1235–1244.

Leonhardt H, Page AW, Weier HU, Bestor TH (1992). A targeting sequence directs DNA methyiltransferase to sites of DNA replication in mammalian nuclei. Cell 71:865–873.

Letessier A, Millot GA, Koundrioukoff S, Lachages AM, et al. (2011). Cell-typespecific replication initiation programs set fragility of the FRA3B fragile site. Nature 470:120–3.

Li G, Levitus M, Bustamante C, Widom J (2005). Rapid spontaneous accessibility of nucleosomal DNA. Nat Struct Mol Biol 12:46-53.

Li H (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1 [q-bio.GN].

Li X, Manley JL (2006). Cotranscriptional processes and their influence on genome stability. Genes Dev 20:1838-47.

Liu LF, Wang JC (1987). Supercoiling of the DNA template during transcription. Proc Natl Acad Sci USA 84:7024–7027

Lieb JD, Liu X, Botstein D, Brown PO (2001). Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat Genet 28:327–334.

Lim JH, Catez F, Birger Y, West KL, Prymakowska-Bosak M, Postnikov YV, Bustin M (2004). Chromosomal protein HMGN1 modulates histone H3 phosphorylation. Mol Cell 15:573-84.

Lipford J, Bell S (2001). Nucleosomes positioned by ORC facilitate the initiation of DNA replication. Mol Cell 7:21–30.

Liu X., Lee CK, Granek JA, Clarke ND, Lieb JD (2006). Whole-genome comparison of Leu3 binding in vitro and in vivo reveals the importance of nucleosome occupancy in target site selection. Genome Res. 16:1517–1528.

Lomvardas S, Thanos D (2001). Nucleosome sliding via TBP DNA binding *in vivo*. 86. Cell 106:685-696.

Lombraña R, Almeida R, Álvarez A and Gómez M (2015). R-loops and initiation of DNA replication in human cells: a missing link? Front. Genet. 6:158.

Lombraña R, Álvarez A, Fernández-Justel JM, Almeida R, Poza-Carrión C, Gomes F, Calzada A, Requena JM and Gómez M (2016). Transcriptionally Driven DNA Replication Program of the Human Parasite Leishmania major Cell Reports 16:1774–1786

Lu H, Liu X, Deng Y, Qing H (2013). DNA methylation, a hand behind neurodegenerative diseases. Front Aging Neurosci 5:85.

Lubelsky Y, Sasaki T, Kuipers MA, Lucas I, Le Beau MM, Carignon S, Debatisse M, Prinz JA, Dennis JH, and Gilbert DM (2010). Pre-replication complex proteins assemble at regions of low nucleosome occupancy within the Chinese hamster dihydrofolate reductase initiation zone. Nucleic Acids Res. 39:3141-55.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389:251–260.

MacAlpine D, Bell SP (2005). A genomic view of eukaryotic DNA replication. Chromosome Res. 13:309–326.

Margueron R, Trojer P, Reinberg D (2005). The key to development: interpreting the histone code? Curr Opin Genet Dev 15:163-76.

Maric C, Benard M, Pierron G (2003). Developmentally regulated usage of Physarum DNA replication origins.EMBO Rep 4:474–478.

Marzluff WF, Wagner EJ and Duronio RJ (2008). Metabolism and regulation of canonical histones: life without a poly(A) tail. Nat Rev Genet 9:843-854.

Mechali M (2001). DNA replication origins: From sequence specificity to epigenetics. Nat Rev Genet 2:640–645.

Mechali M (2010). Eukaryotic DNA replication origins: many choices for appropriate answers. Nat Rev Mol Cell Biol 11:728-38.

Mejlvang J, Feng Y, Alabert C, Neelsen KJ, Jasencakova Z, Zhao X, Lees M, Sandelin A, Pasero P, Lopes M and Groth A (2015). New histone supply regulates replication fork speed and PCNA unloading. J Cell Biol 204:29-43.

Mikkelsen TS, Ku M, Jaffe DB, Isaac B, Lieberman E (2007). Genomewide maps of chromatin state in pluripotent and lineage-commited cells. Nature 448:553-560.

Millar CB, Grunstein M (2006). Genome-wide patterns of histone modifications in yeast. Nat. Rev. Mol. Cell Biol. 7:657–666.

Minami H, Takahashi J, Suto A, Saitoh Y, Tsutsumi K (2006). Binding of AlF-C, an Orc1-binding transcriptional regulator enhances replicator activity of the rat aldolase B origin. Mol Cell Biol 26:8770–8780.

Miotto B, Ji Z, Struhl K *(*2016*).* Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. *Proc Natl Acad Sci* doi:10.1073/pnas.1609060113

Mirabella AC, Foster BM, Bartke T (2016). Chromatin deregulation in disease. Chromosoma 125:75-93.

Muller S, Ronfani L, Bianchi ME (2004). Regulated expression and subcellular localization of HMGB1, a chromatin protein with a cytokine function. J Intern Med 255:332-343.

Muller P, Park S, Shor E, Huebert DJ, et al. (2010). The conserved bromoadjacent homology domain of yeast Orc1 functions in the selection of DNA replication origins within chromatin. Genes Dev 24:1418–33

Murga M, Jaco I, Fan Y, Soria R, Martinez-Pastor B, Cuadrado M, Yang SM, Blasco MA, Skoultchi AI, Fernandez-Capetillo O (2007). Global chromatin compaction limits the strength of the DNA damage response. J Cell Biol 178:1101-8.

Najima Y, Yahagi N, Takeuchi Y, Matsuzaka T, Sekiya M, Nakagawa Y, Amemiya-Kudo M, Okazaki H, Okazaki S, Tamura Y, Iizuka Y, Ohashi K, Harada K, Gotoda T, Nagai R, Kadowaki T, Ishibashi S, Yamada N, Osuga J, Shimano H (2005). High mobility group protein-B1 interacts with sterol regulatory element-binding proteins to enhance their DNA binding. J Biol Chem 280:27523-32.

Nakamura H, Morita T and Sato C (1986). Structural organizations of replicon domains during DNA synthetic phase in the mammalian nucleous. Experimental Cell Research 165:291-297.

Necsulea A, Guillet C, Cadoret JC, Prioleau MN, Duret L (2009). The relationship between DNA replication and human genome organization. Mol Biol Evol 26:729–741.

Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, et al. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature 457:1038–42.

Nelson P (1999). Transport of torsional stress in DNA. Proc Natl Acad Sci USA 96:14342–14347

Nemeth MJ, Cline AP, Anderson SM, Garrett-Beal LJ, Bodine DM (2005) Hmgb3 deficiency deregulates proliferation and differentiation of common lymphoid and myeloid progenitors. Blood 105:627-634.

Nieduszynski CA, Knox Y, Donaldson AD (2006). Genome-wide identification of replication origins in yeast by comparative genomics. Genes Dev 20:1874–1879.

Noguchi K, Vassilev A, Ghosh S, Yates JL, et al. (2006). The BAH domain facilitates the ability of human Orc1 protein to activate replication origins in vivo. Embo J 25:5372–82.

Okuno Y, Satoh H, Sekiguchi M, Masukata H (1999). Clustered adenine/thymine stretches are essential for function of a fission yeast replication origin. Mol Cell Biol 19:6699–6709.

Oudet P, Gross-Bellard M, Chambon P (1975). Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell 4:281–300.

Osmanov T, Ugrinova I, Pasheva E (2013). The chaperone like function of the nonhistone protein HMGB1. Biochem Biophys Res Commun 432:231-5.

Ouspenski II, Van Hooser AA and Brinkley BR (2003). Relevance of histone acetylation and replication timing for deposition of centromeric histone CENP-A. Exp Cell Res 285:175-88.

Ozeri-Galai E, Lebofsky R, Rahat A, Bester AC, et al. (2011). Failure of origin activation in response to fork stalling leads to chromosomal instability at fragile sites. Mol Cell 43:122–31.

Paixao S, Colaluca IN, Cubells M, Peverali FA, Destro A, Giadrossi S, Giacca M, Falaschi A, Riva S, Biamonti G (2004). Modular structure of the human lamin B2 replicator. Mol Cell Biol 24:2958–2967.

Palzkill TG, Newlon CS (1988). A yeast replication origin consists of multiple copies of a small conserved sequence. Cell:53:441–50.

Papior P, Arteaga-Salas JM, Gunther T, Grundhoff A, Schepers A (2012). Open chromatin structures regulate the efficiencies of Pre- RC formation and replication initiation in Epstein-Barr virus. J Cell Biol 198:509–528

Patterton HG, Landel CC, Landsman D, Peterson CL, Simpson RT (1998). The biochemical and phenotypic characterization of Hho1p, the putative linker histone H1 of Saccharomyces cerevisiae. J Biol Chem 273:7268-76.

Paull TT, Carey M, Johnson RC (1996). Yeast HMG proteins NHP6A/B potentiate promoter-specific transcriptional activation in vivo and assembly of preinitation complexes in vitro. Genes Dev 10:2769-2781.

Peckham HE, Thurman RE, Fu Y, Stamatoyannopoulos JA, Noble WS, Struhl K, Weng Z (2007). **Nucleosome positioning signals in genomic DNA.** *Genome Res*, **17:**1170-1177.

Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N (1996). Requirement for Xist in X chromosome inactivation. Nature 379:131–137.

Pérez-Montero S, Carbonell A, Azorín F (2016). Germline-specific H1 variants: the "sexy" linker histones. Chromosoma 125:1-13.

Petty E and Pillus L (2013). Balancing chromatin remodeling and histone modifications in transcription. Trends Genet 29:621–629.

Picard F, Cadoret JC, Audit B, Arneodo A, Alberti A, Battail C, Duret L and Prioleau MN (2014). The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. PLoS Genet. 10: e1004283 (2014).

Pollard TD, Earnshaw WC (2002). Cell biology. Saunders, Philadelphia, PA.

Postnikov YV, Belova GI, Lim JH, Bustin M (2006). Chromosomal protein HMGN1 modulates the phosphorylation of serine 1 in histone H2A. Biochemistry 45:15092-9.

Postnikov YV, Bustin M (2016). Functional interplay between histone H1 and HMG proteins in chromatin. Biochim Biophys Acta 1859:462-7.

Powell SK, MacAlpine HK, Prinz JA, Li Y, Belsky JA, MacAlpine DM (2015). Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. EMBO J 34:531–543.

Pray-Grant MG, Daniel JA, Schieltz D, Yates JR, Grant PA (2005). Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. Nature 433:434-8.

Prado, F and Maya, D (2017). Regulation of replication fork advance and stability by nucleosome assembly. Genes. pii E49

Prioleau MN and MacAlpine DM (2016). DNA replication origins-where do we begin? Genes Dev 30:1683-1697.

Radman-Livaja M, Rando OJ (2010). Nucleosome positioning: how is it established, and why does it matter? Dev Biol 339:258-66.

Ramón A, Muro-Pastor MI, Scazzocchio C, Gonzalez R (2000). Deletion of the unique gene encoding a typical histone H1 has no apparent phenotype in Aspergillus nidulans. Mol Microbiol 35:223-33.

Rampakakis E, Arvanitis DN, Di Paola D, Zannis-Hadjopoulos M (2009). Metazoan origins of DNA replication: Regulation through dynamic chromatin structure. J Cell Biochem 106:512–520.

Rando OJ, Chang HY (2009). Genome-wide views of chromatin structure. Annu Rev Biochem 78:245–271.

Rao H, Stillman B. (1995). The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators. Proc Natl Acad Sci 92:2224–2228.

Rao VA, Conti C, Guirouilh-Barbat J, Nakamura A, Miao ZH, Davies SL, Saccá B, Hickson ID, Bensimon A, Pommier Y (2007). Endogenous gamma-H2AX-ATM-Chk2 checkpoint activation in Bloom's syndrome helicase deficient cells is related to DNA replication arrested forks. Mol Cancer Res. 5:713-24.

Reeves R (2001). Molecular biology of HMGA proteins: hubs of nuclear function. Gene 277:63-81.

Remus D, Beall EL, Botchan MR (2004). DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. The EMBO journal 23:897-907.

Rhodes D, Lipps HJ (2015). G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res 43:8627-37.

Ringrose L and Paro R (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu Rev Genet 38:413-43.

Rivera C, Gurard-Levin ZA, Almouzni G, Loyola A (2014). Histone lysine methylation and chromatin replication. Biochim Biophys Acta 1839:1433-9.

Rocha EP (2004). The replication-related organization of bacterial genomes, Microbiology 150:1609–1627.

Robinson PJ, Rhodes D (2006). Structure of the '30 nm' chromatin fibre: a key role for the linker histone. Curr Opin Struct Biol 16:336-43.

Roque A, Ponte I, Suau P (2016). Interplay between histone H1 structure and function. Biochim Biophys Acta 1859:444-54.

Sambrook J, Fritschi EF and Maniatis T (1989). Molecular cloning: a laboratorymanual, Cold Spring Harbor Laboratory Press, New York.

Santos-Pereira JM, Aguilera A (2015). R-loops: new modulators of genome dynamics and function. Nat Rev Genet 16:583-97.

Santos-Rosa H, Schneider R, Bernstein BE, Karabetsou N, Morillon A, Weise C, Schreiber SL, Mellor J, Kouzarides T (2003). Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. Mol Cell 2:1325-32.

Sarma K, Reinberg D (2005). Histone variants meet their match. Nat. Rev. Mol. Cell Biol. 6:139–149.

Sasaki T, Ramanathan S, Okuno Y, Kumagai C, Shaikh SS, Gilbert DM (2006). The Chinese hamster dihydrofolate reductase replication origin decision point follows activation of transcription and suppresses initiation of replication within transcription units. Mol Cell Biol 26:1051–1062.

Schaarschmidt D, Baltin J, Stehle IM, Lipps HJ, Knippers R (2004). An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. EMBO J 23:191–201

Shen X, Yu L, Weir JW, Gorovsky MA (1995). Linker histones are not essential and affect chromatin condensation *in vivo*. Cell 82:47-56.

Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K (2008). **Dynamic regulation of nucleosome positioning in the human genome.** *Cell* **132:**887-898.

Schwab RA, Nieminuszczy J, Shah F, Langton J, Lopez Martinez D, Liang CC, Cohn MA, Gibbons RJ, Deans AJ, Niedzwiedz W (2015). The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. Mol Cell 60:351-61.

Schwabish MA, Struhl K (2004). Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. Mol Cell Biol 24:10111–10117.

Schwartz BE, Ahmad K (2005). Transcriptional activation triggers deposition and removal of the histone variant H3.3. Genes Dev 19:804-14.

Segal E, Fondufe-Mittendorf Y, Chen L, Thåström A, Field Y, Moore IK, Wang JZ, Widom J (2006). A Genomic Code for Nucleosome Positioning, *Nature* 442:772-778.

Segal E, Widom J (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. Curr Opin Struct Biol 19:65-71.

Segurado M, de Luis A, Antequera F (2003). Genome-wide distribution of DNA replication origins at A+T-rich islands in Schizosaccharomyces pombe. EMBO Rep 4:1048–1053.

Seila AC, Calabrese JM, Levine SS, Yeo GW, et al. (2008). Divergent transcription from active promoters. Science 322:1849–51.

Sekinger EA, Moqtaderi Z, Struhl K (2005). Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. Mol. Cell 18:735–748.

Sequeira-Mendes J, Diaz-Uriarte R, Apedaile A, Huntley D, Brockdorff N, Gomez M (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells. PLoS Genet 5:e1000446.

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

Sgarra R, Rustighi A, Tessari MA, Di Bernardo J, Altamura S, Fusco A, Manfioletti G, Giancotti V (2004). Nuclear phosphoproteins HMGA and their relationship with chromatin structure and cancer. FEBS Lett 574:1-8.

Shen X, Yu L, Weir JW, Gorovsky MA (1995). Linker histones are not essential and affect chromatin condensation *in vivo*. Cell 82:47-56.

Simpson, RT (1978). Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. Biochemistry 17:5524–5531

Simpson RT (1990). Nucleosome positioning can affect the function of a cis-acting DNA element in vivo. Nature 343:387–389.

Sims RJ, Chen CF, Santos-Rosa H, Kouzarides T, Patel SS, Reinberg D (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. J Biol Chem 280:41789-92.

Singh J and Padgett RA (2009). Rates of in situ transcription and splicing in large human genes. Nat. Stru. Mol. Biol. 16:1128-1133.

Sirotkin AM, Edelmann W, Cheng G, Klein-Szanto A, Kucherlapati R, Skoultchi AI (1995). Mice develop normally without the H1(0) linker histone. Proc Natl Acad Sci U S A 92:6434–6438.

Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature 516:436–9

Skourti-Stathaki K, Proudfoot NJ (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes & development 28:1384–96

Smith DJ and Whitehouse I (2012). Intrinsic coupling of lagging-strand synthesis to chromatin assembly. Nature 483:434–438

Stillman B (2005). Origin recognition and the chromosome cycle. FEBS Lett 579:877–884.

Stinchcomb DT, Struhl K, Davis RW (1979). Isolation and characterisation of a yeast chromosomal replicator. Nature 282:39–43.

Stork CT, Bocek M, Crossley MP, Sollier J, Sanz LA, Chédin F, Swigut T and Cimprich KA (2016). Co-transcriptional R-loops are the main cause of strogen-induced DNA damage. Elife 5:e17548.

Stros M (2010). HMGB proteins: interactions with DNA and chromatin. Biochim Biophys Acta 1799:101-13.

Stros M, Launholt D, Grasser KD (2007). The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. Cell Mol Life Sci 64:2590-606.

Studitsky VM, Clark DJ, Felsenfeld G (1994). A histone octamer can step around a transcribing polymerase without leaving the template. Cell 76:371–382.

Sutcliffe EL, Parish IA, He YQ, Juelich T, Tierney ML, Rangasamy D, Milburn PJ, Parish CR, Tremethick DJ, Rao S (2009). Dynamic histone variant exchange accompanies gene induction in T cells. Mol Cell Biol 29:1972-86.

Sutrias-Grau M, Bianchi ME, Bernués J (1999). High mobility group protein 1 interacts specifically with the core domain of human TATA box-binding protein and interferes with transcription factor IIB within the pre-initiation complex. J Biol Chem 274:1628-34.

Swanson PC (2002). Fine structure and activity of discrete RAG-HMG complexes on V(D)J recombination signals. Mol Cell Biol 22:1340–1351.

Swarnalatha M, Singh AK and Kumar V (2012). The epigenetic control of E-box and Myc-dependent chromatin modifications regulate the licensing of lamin B2 origin during cell cycle. Nucleic Acids Res 40:9021–9035.

Symeonidou IE, Taraviras S, Lygerou Z (2012). Control over DNA replication in time and space. FEBS Lett 586:2803-12

Takai D, Jones PA (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci USA 99:3740–3745.

Talbert PB, Henikoff S (2010). Histone variants--ancient wrap artists of the epigenome. Nat. Rev. Mol. Cell Biol. 11:264–275.

Teif VB, Vainshtein Y, Caudron-Herger M, MallmJP, Marth C, Hofer T, Rippe K (2012). Genome-wide nucleosome positioning during embryonic stem cell development. Nat Struct Mol Biol 19:1185–1192.

Técher H, Koundrioukoff S, Azar D, Wilhelm T, Carignon S, Brison O, Debatisse M, Le Tallec B (2013). Replication dynamics: biases and robustness of DNA fiber analysis. J Mol Biol 425:4845-55.

Terret ME, Sherwood R, Rahman S, Qin J and Jallepalli PV (2009). Cohesin acetylation speeds the replication fork. Nature 462:321-234.

Teves SS and Henikoff S (2014). Transcription-generated torsional stress destabilizes nucleosomes. doi:10.1038/nsmb.2723

Thakar A, Gupta P, Ishibashi T, Finn R, Silva-Moreno B, Uchiyama S, Fukui K, Tomschik M, Ausio J, Zlatanova J (2009). H2A.Z and H3.3 histone variants affect nucleosome structure: biochemical and biophysical studies. Biochemistry 48:10852-7.

Tolstorukov MY, Kharchenko PV, Park PJ (2010). Analysis of primary structure of chromatin with next-generation sequencing. Epigenomics 2:187–197

Travers AA (2003). Priming the nucleosome: a role for HMGB proteins? EMBO Rep 4:131-6.

Tsurusaki Y, Okamoto N, Ohashi H, Kosho T, Imai Y, Hibi-Ko Y, Kaname T, Naritomi K, Kawame H, Wakui K, Fukushima Y, Homma T, Kato M, Hiraki Y, Yamagata T, Yano S, Mizuno S, Sakazume S, Ishii T, Nagai T, Shiina M, Ogata K, Ohta T, Niikawa N, Miyatake S, Okada I, Mizuguchi T, Doi H, Saitsu H, Miyake N, Matsumoto N (2012). Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. Nat Genet 44:376–378.

Tuduri S, Crabbé L, Conti C, Tourrière H, Holtgreve-Grez H, Jauch A, Pantesco V, De Vos, J, Thomas A, Theillet C, Pommier Y, Tazi J, Coquelle A and Pasero P (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nat Cell Biol 11:1315-1324.

Ueda T, Chou H, Kawase T, Shirakawa H and Yoshida M (2004). Acidic C-tail of HMGB1 is required for its target binding to nucleosome linker DNA and transcription stimulation. Biochemistry 43:9901-9908.

Ueda T, Postnikov YV, Bustin M (2006). Distinct domains in high mobility group N variants modulate specific chromatin modifications. J Biol Chem 281:10182-7.

Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K, Sidow A, Fire A, Johnson SM (2008). A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res 18:1051–1063.

Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sido A (2011). Determinants of nucleosome organization in primary human cells. Nature 474:516–520

Vashee S, Cvetic C, Lu W, Simancek P, Kelly TJ, Walter JC (2003). Sequence-independent DNA binding and replication initiation by the human origin recognition complex. Genes dev 17:1894-1908.

Vogelauer M, Rubbi L, Lucas I, Brewer B, Grunstein M (2002). Histone acetylation regulates the time of replication origin firing. Mol Cell 10:1223–1233.

Warburton PE[1], Cooke CA, Bourassa S, Vafa O, Sullivan BA, Stetten G, Gimelli G, Warburton D, Tyler-Smith C, Sullivan KF, Poirier GG, Earnshaw WC (1997). Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Curr Biol 7:901-4.

Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, Roeder RG, Brivanlou AH, Allis CD (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell 121:859-72.

Xu W, Aparicio JG, Aparicio OM, Tavare S (2006). Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae. BMC Genomics 7: 276. doi: 10.1186/1471-2164-7-276

Xu Z, WeiW, Gagneur J, Perocchi F, et al. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature 457:1033–7.

Yin S, Deng W, Hu L, Kong X (2009). The impact of nucleosome positioning on the organization of replication origins in eukaryotes. Biochem Biophys Res Commun 385:363-8.

Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol 4:442–451.

Yuan GC, Liu JS (2008). Genomic sequence is highly predictive of local nucleosome depletion. PLoS Comput Biol 4:e13.

Zhang Y, Moqtaderi Z, Rattner PB, Euskirchen G, Snyder M, Kadonaga TJ, Liu XS, Struhl K (2009). **Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo.** *Nat struct mol biol*, **16:**847-853.

Zhang Q, Wang Y (2008). High mobility group proteins and their post-translational modifications. Biochim Biophys Acta 1784:1159-66.

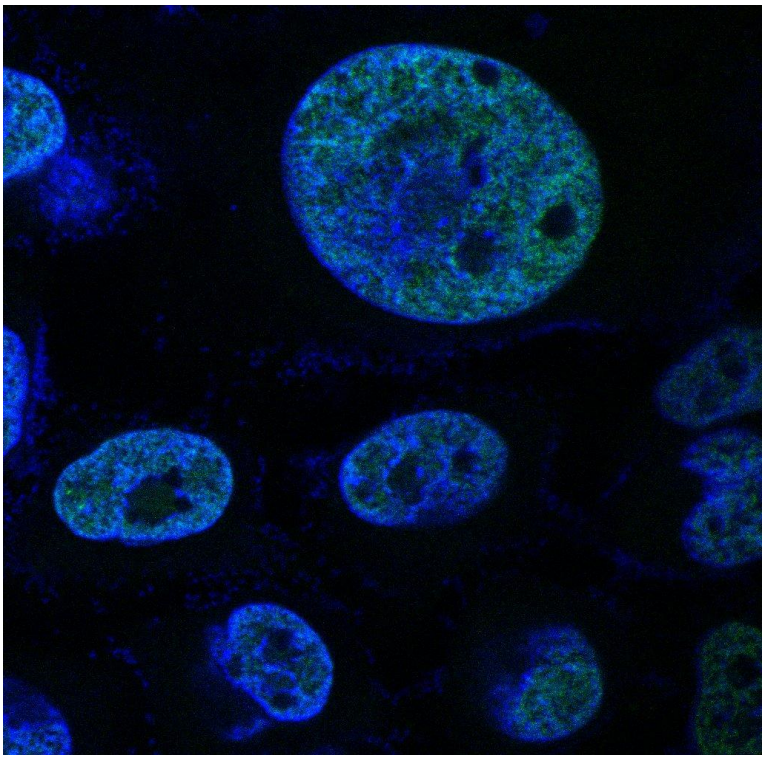
Zhong Y, Nellimoottil T, Peace JM, Knott SR, Villwock SK, Yee JM, Jancusa JM, Rege S, Tecklenburg M, Sclafani RA, Tavaré S and Aparicio OM (2013). The level of origin firing inversely affects the rate of replication fork progression. J Cell Biol 201:373-383.

Zhou J, Chau CM, Deng Z, Shiekhattar R, Spindler MP, Schepers A, Lieberman PM (2005). Cell cycle regulation of chromatin at an origin of DNA replication. Embo J 24:1406–1417

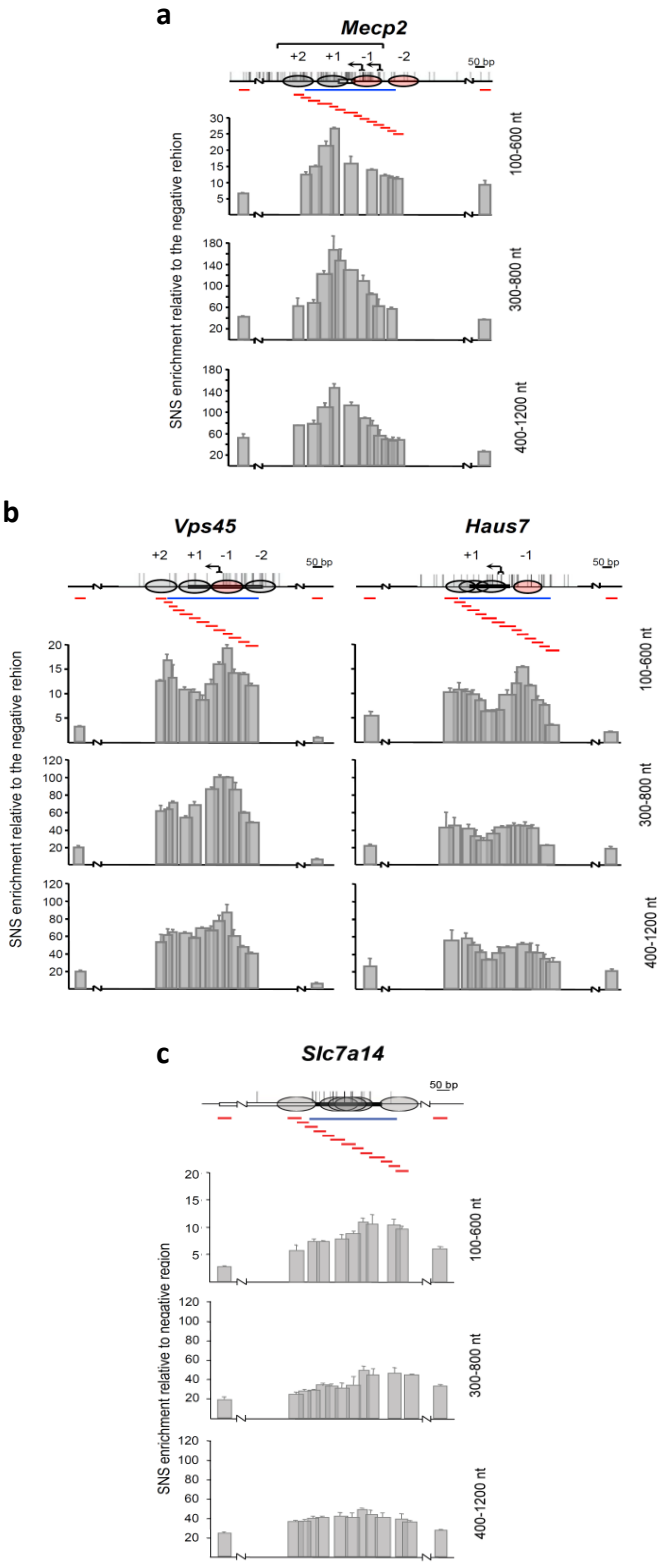
Zlatanova J, Caiafa P, Van Holde K (2000). Linker histone binding and displacement: versatile mechanism for transcriptional regulation. FASEB J 14:1697-704.

Zlatanova J, Thakar A (2008). H2A.Z: View from the Top. Structure16:166-79.

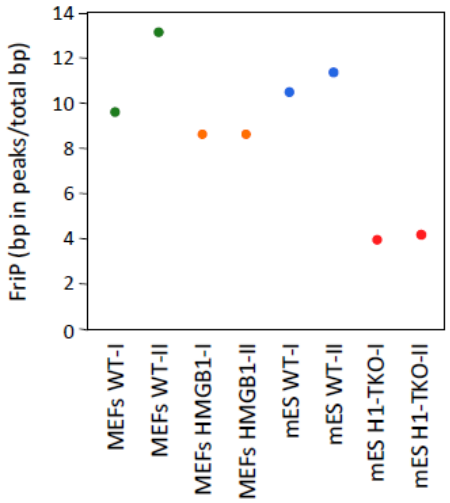
Annex I – Supplementary Information



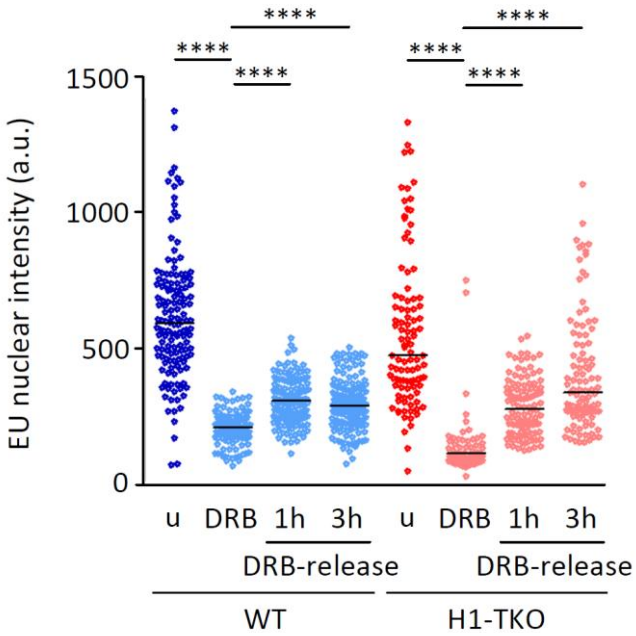
1. Supplementary Figures



Supplementary Figure 1: High-resolution analysis of SNS abundance at promoter-ORIs normalized to a negative region. Maps with the nucleosome patterns and histograms of SNS enrichments at the mouse ORIs present at the promoter regions of *Mecp2* **(a)**, *Vps45*, *Haus7* **(b)** and the intragenic region of *Slc7a14* **(c)**. SNS enrichments were normalized to those obtained at a non-ORI region (Supplementary Tables 1, 2, 3 and 4).



Supplementary Figure 2: Assessment of SNS-seq peak quality by Fraction of Reads in Peaks (FriP) analysis. Dot-plot depicting FriP ratio generated by the number of base pairs in peak regions divided by the total number of mapped base pairs, normalized against the number of peaks. All SNS-seq replicates form MEF and mES cells are represented. Higher ratios indicate more robust and reliable peaks.



Supplementary Figure 3: Quantification of nuclear EU signal intensity upon transcription synchronization in mES cells. Scatter plots represent the absolute values, in arbitrary units (a.u.) of fluorescence, used to obtain the graphic represent in Figure 27d, which depicts the normalized variation of signal intensity throughout the time course of the experiment outlined in figure 27a. u corresponds to untreated sample. Data derived from two independent experiments. ****p<0.0001. Numerical values are shown in Supplementary Table 10. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

2. SUPPLEMENTARY TABLES

Supplementary Table 1: Primer pairs used in the NuSA and SNS mapping experiments for the region surrounding mouse *Mecp2* promoter. Amplicon sizes and annealing temperatures are indicated.

Region	Primer	Sequence (5' to 3')	Anneal. Tre	Amplicon Size
Mecp2 **(572 bp)**	Mecp 1	GAGGGCGTCATTCGAAGGTG	64ºC	59 bp
	Mecp 2	AATCGCGAGCGACGGTTCTC		
	Mecp 3	GCTGAGAACCGTCGCTCGC	66ºC	56 bp
	Mecp 4	CGCCGGTGGTGGCTTTCTC		
	Mecp 5	GAGAAAGCCACCACCGGCG	66ºC	71 bp
	Mecp 6	CGACACGGCTGGCGGATG		
	Mecp 7	CATCCGCCAGCCGTGTCG	60ºC	75 bp
	Mecp 8	CTTGCTGGGGGGCGGGTAG		
	Mecp 9	CTACCCGCCCCCCAGCAAG	64ºC	53 bp
	Mecp 10	GTGAGTGGGACCGCCAAGG		
	Mecp 11	CCTTGGCGGTCCCACTCAC	66ºC	58 bp
	Mecp 12	GCCGAGCGGAGGAGGAGG		
	Mecp 13	CTCCTCCGCTCGGCGCG	60ºC	76 bp
	Mecp 14	GCTGTGGTAAAACCCGTCCGG		
	Mecp 15	TTTTCCGGACGGGTTTTACC	64ºC	48 bp
	Mecp 16	CGCTCCCTCCTCTCGGAG		
	Mecp 17	CTCCGAGAGGAGGGAGCG	62ºC	56 bp
	Mecp 18	GACGTCTGCCGTGCGGGGT		
	Mecp 19	ACCCCGCACGGCAGACGTC	68ºC	57 bp
	Mecp 20	GTGCAGCAGCACACAGGCTG		
	Mecp 21	GACCAGCCTGTGTGCTGCTG	66ºC	60 bp
	Mecp 22	CTCGACAAAGAGCAAGGGGTG		
	Mecp 23	CACCCCTTGCTCTTTGTCGAG	66ºC	54 bp
	Mecp 24	CCAGCCTGGGCTCCACAAC		
	Mecp 25	GTTGTGGAGCCCAGGCTGG	66ºC	50 bp
	Mecp 26	CAATTGAGGGCGTCACCGCT		
	Mecp 27	AGCGGTGACGCCCTCAATTG	58ºC	56 bp
	Mecp 28	CCTCTTTTCCCTGCCTAAAC		
	Mecp flk1	TTGTGGCGCACTCTCCCAAC	66ºC	55 bp
	Mecp flk2	CACACAGACTGGCGCGCGTG		
	Mecp flk3	GCATCCAATGCTCTTTGTGC	60ºC	43 bp
	Mecp flk4	GTCTCTTGTTGAGCATTTGT		

Supplementary Table 2: Primer pairs used in the NuSA and SNS mapping experiments for the region surrounding mouse *Haus7* promoter. Amplicon sizes and annealing temperatures are indicated.

Region	Primer	Sequence (5' to 3')	Anneal. Tre	Amplicon Size
Haus7 **(589 bp)**	Haus 1	AAGCTTTGCCAGGGCCTG	64ºC	70 bp
	Haus 2	GCCATTTGGAGCTTCCTGTG		
	Haus 3	TCCACAGGAAGCTCCAAATG	58ºC	58 bp
	Haus 4	TCCTTGATAGGGGACCAGG		
	Haus 5	TGGTCCCCTATCAAGGAG	56ºC	50 bp
	Haus 6	GAAGCTGAAGGTACTGTGAG		
	Haus 7	CTGGCTCACAGTACCTTCAG	60ºC	64 bp
	Haus 8	GTGATGACTGCGTGGTCAAG		
	Haus 9	CTTGACCACGCAGTCATCAC	62ºC	55 bp
	Haus 10	TGCCGGGTCTGATGACAG		
	Haus 11	TAGTAGCTGTCATCAGACCC	56ºC	76 bp
	Haus 12	GCGTGAACCAGTTGTAGTTC		
	Haus 13	CTGAACTACAACTGGTTCACG	56ºC	65 bp
	Haus 14	GGTGAGCTTTTAGCAGTGTG		
	Haus 15	CGCCGCCACACTGCTAAAAG	64ºC	82 bp
	Haus 16	GATGGTTGGCTTCGCTCCTG		
	Haus 17	GAGCGAAGCCAACCATCG	64ºC	50 bp
	Haus 18	GAAAAGAGGGTGGGCCTTTG		
	Haus 19	CGTCACAAAGGCCCACCCTC	66ºC	57 bp
	Haus 20	ACAGAGCGGCAGCCCAATGG		
	Haus 21	TTGGGCTGCCGCTCTGTC	66ºC	57 bp
	Haus 22	CTTCTCTGCACCCTGCTCCC		
	Haus 23	GGGAGCAGGGTGCAGAGAAG	64ºC	58 bp
	Haus 24	GCCTTTAGCCCGCCTCTG		
	Haus 25	ACAGAGGCGGGCTAAAGGC	62ºC	51 bp
	Haus 26	CCTCCTCCTCTCTGCCTCTC		
	Haus 27	AGGCAGAGAGGAGGAGGC	58ºC	67 bp
	Haus 28	CTGCCTCTAAAAGGAGCTACTC		
	Haus flk1	AAACTGCAAGGAAAAACTCC	56ºC	89 bp
	Haus flk2	TTTCTTAGTCCATCCTGAGG		
	Haus flk3	TATTTCTATCCTCCACAAGG	53ºC	84 bp
	Haus flk4	GAAAACAAGCAACCAAAAGC		

Supplementary Table 3: Primer pairs used in the NuSA and SNS mapping experiments for the region surrounding mouse *Vps45* promoter. Amplicon sizes and annealing temperatures are indicated.

Region	Primer	Sequence (5' to 3')	Anneal. Tre	Amplicon Size
Vps45 **(541 bp)**	Vps 1	AGTTAGAGCTGGTTAGACTG	56ºC	56 bp
	Vps 2	CTGGGAAAGTTGTAGAATG		
	Vps 3	TCTACAACTTTCCCAGTGCC	56ºC	45 bp
	Vps 4	CTGATGACGATGATCAAGTC		
	Vps 5	GACTTGATCATCGTCATCAG	56ºC	46 bp
	Vps 6	GTTCCGCTGTAGAACTTAGC		
	Vps 7	AATGCTAAGTTCTACAGCGG	56ºC	62 bp
	Vps 8	ATAAAGGAAGCTCTCCCTTC		
	Vps 9	TAAAGGAAGGGAGAGCTTCC	58ºC	67 bp
	Vps 10	TCATGGATAAAGAAACGGTGAG		
	Vps 11	CCGTTTCTTTATCCATGAGAAG	58ºC	57 bp
	Vps 12	AAATGATAGAGGACAGCGGG		
	Vps 13	CGCTGTCCTCTATCATTTTG	60ºC	68 bp
	Vps 14	TCAATTCGCCACCATGAATG		
	Vps 15	ATTCATGGTGGCGAATTGAC	60ºC	63 bp
	Vps 16	TAATTCAGCCAGGAAAGTGG		
	Vps 17	CCACTTTCCTGGCTGAATTAACC	64ºC	61 bp
	Vps 18	TTTGGCGACCGGAAGCAG		
	Vps 19	TTCCGGTCGCCAAAGCCTC	64ºC	53 bp
	Vps 20	CCCAGTATCGGAGCTACCCG		
	Vps 21	CGGGTAGCTCCGATACTGG	60ºC	65 bp
	Vps 22	TAGCTGCTGAGTCTGAGTCCC		
	Vps 23	TCAGACTCAGCAGCTAAGCG	60ºC	58 bp
	Vps 24	TTCCACTCCCTACCGAGAAG		
	Vps 25	TTCTCGGTAGGGAGTGGAAG	60ºC	69 bp
	Vps 26	TACTCAGGACCAGAAGCCAG		
	Vps flk1	TACAGGACGAGAATTGGAAC	60ºC	101 bp
	Vps flk2	GACCACTGGGATTAACGGAA		
	Vps flk3	AAGGGAATGAATACAAGGAG	56ºC	74 bp
	Vps flk4	CCAGGGAGCTTTAGGAAC		

Supplementary Table 4: Primer pairs used in the NuSA and SNS mapping experiments for the region surrounding mouse *Slca14* promoter. Amplicon sizes and annealing temperatures are indicated.

Region	Primer	Sequence (5' to 3')	Anneal. Tre	Amplicon Size
Slc7a14 **(582 bp)**	Slc 1	AGGTTATGCCTGGGGATG	58ºC	72 bp
	Slc 2	AGGTGGAAACTTCTTCCAACTC		
	Slc 3	TGAGAGTTGGAAGAAGTTTC	56ºC	56 bp
	Slc 4	GGTTTCTTAGTACATCAGCC		
	Slc 5	GGCTGATGTACTAAGAAACC	56ºC	62 bp
	Slc 6	CTCTCCGGAGTAAGACTAAC		
	Slc 7	GCGCGGTTAGTCTTACTCCG	64ºC	63 bp
	Slc 8	AACAGAACTCGGAGGCCCTG		
	Slc 9	ATCAGGGCCTCCGAGTTCTG	64ºC	58 bp
	Slc 10	GGACGAGTAGCAAAGCGAAAAG		
	Slc 11	TTCGCTTTGCTACTCGTCCG	62ºC	73 bp
	Slc 12	CCGAGGACAAAGGCTTCTATTAC		
	Slc 13	TAGAAGCCTTTGTCCTCGGC	62ºC	69 bp
	Slc 14	GTTCTCCTACGCCACCGAG		
	Slc 15	CTCGGTGGCGTAGGAGAAC	62ºC	56 bp
	Slc 16	GCTACGATGTGGATGATCCC		
	Slc 17	CATCCACATCGTAGCGCTGG	66ºC	59 bp
	Slc 18	AATCAGCGCCCGAGAGCAAG		
	Slc 19	CTCTCGGGCGCTGATTTC	62ºC	74 bp
	Slc 20	CCCTCCCAACAGGTATGCTC		
	Slc 21	AGCATACCTGTTGGGAGGGG	64ºC	56 bp
	Slc 22	CGGGCCCCAAGATAGCTG		
	Slc 23	CAGCTATCTTGGGGCCCG	64ºC	57 bp
	Slc 24	ACCCCAGCCCTAGTCAGGTG		
	Slc 25	AGATCACACCTGACTAGGGC	58ºC	66 bp
	Slc 26	CAAGTCTTTCCAAAATGTGC		
	Slc flk1	CATGTCTGCCAATGCTTTTC	60ºC	88 bp
	Slc flk2	ACCCACCATCAGAGCTTGAG		
	Slc flk3	AAGGCAGAATGCTCGTACTC	60ºC	57 bp
	Slc flk4	AAAACCTGACCACCACTTCC		

Supplementary Table 5: Primer pairs used in the NuSA and SNS mapping experiments for the region surrounding mouse *Slca14* promoter. Amplicon sizes and annealing temperatures are indicated.

Region	Primer	Sequence (5' to 3')	Anneal. Tre	Amplicon Size
LaminB2 **(826 bp)**	Lam 1	GAGGCTTTAGCCGCGACGTC	64ºC	119 bp
	Lam 2	GCTAACATTGTCGGAACAGC		
	Lam 3	CCAGGGGTGGCCCTGTC	64ºC	95 bp
	Lam 4	GGTGAAAGTGCAGATCGCCG		
	Lam 5	CTCCATTATGAGCCCTGGGTC	64ºC	100 bp
	Lam 6	GGCCTCGGTCCGGTTGACT		
	Lam 7	CCCGAAATCGGAGCCG	64ºC	58 bp
	Lam 8	GGCCTCGGTCCGGTTGACT		
	Lam 9	ACCGAGGCCGCGTGCG	66ºC	76 bp
	Lam 10	GGACTCCGTTTCCCGTGGT		
	Lam 11	ACCACGGGAAACGGAGTCC	66ºC	61 bp
	Lam 12	GCGCGGTCGTGTGGGA		
	Lam 13	TAGCTCGTGTAGGTAACGGC	62ºC	72 bp
	Lam 14	GGGCCATTCAAGGTCGCGCG		
	Lam 15	GCTCATGCGGAGGCCTGG	64ºC	51 bp
	Lam 16	CAACCACGGGTAGCTCGTGTAGG		
	Lam 17	CCTACACGAGCTACCCGTGGTTG	64ºC	49 bp
	Lam 18	CAGGGCCTCCCTCTTCCCG		
	Lam 19	CGGGAAGAGGGAGGCCCTG	64ºC	44 bp
	Lam 20	GTAGCCCCCCTACTCCCCGG		
	Lam 21	CCGGGGAGTAGGGGGGCTAC	64ºC	51 bp
	Lam 22	GAGTACAAAGTGATCGGCCTCGG		
	Lam 23	CCGAGGCCGATCACTTTGTACTC	64ºC	77 bp
	Lam 24	GGTGCCTCGTGCGCATG		
	Lam 25	TGCCTCCAGCTCGTCCCG	64ºC	71 bp
	Lam 26	CAACACGCTGTATAGACGCGCC		
	Lam 27	GGCGCGTCTATACAGCGTGTTG	66ºC	47 bp
	Lam 28	GATGCGACCGGGCTCCG		
	Lam 29	CGGAGCCCGGTCGCATC	68ºC	83 bp
	Lam 30	TGGGACCCTGCCCTTTTTTTTC		
	Lam 31	TCAGCTTGTGCAACAGCGTC	66ºC	82 bp
	Lam 32	GCTAGTGTAAACAGGACCCAGGCG		
	Lam 33	GAGGCCCCGGCTCGAG	64ºC	48 bp
	Lam 34	GGTTCTGCCTCTGAGTTTATTCC		

Supplementary Table 6: Primer pairs used for the negative SNS control. Annealing temperature is indicated.

Region	Primer	Sequence (5' to 3')	Anneal. T^{re}
Mouse chrX	mNR 1	CTCACCACCGATGTCTCAAC	60
	mNR 2	CACAAGTAACACAAAGGAAAAGGG	

Supplementary Table 7: Number of reads and ORI peaks obtained in all the SNS-seq experiments. Values correspond to both replicates of WT and HMGB1-KO MEFs (Figure 17 and 22) and WT and H1-TKO mES cells (Figure 22).

SNS sample	Read number	ORI number
MEFs WT-I	55469980	45049
MEFs WT-II	54157881	61020
MEFs HMGB1-I	107801285	65423
MEFs HMGB1-II	119915739	68499
mES WT-I	73122129	94590
mES WT-II	164720257	106891
mES H1-TKO-I	121941851	56615
mES H1-TKO-II	141877552	91273

Supplementary Table 8: Numerical values and statistic parameters of IOD, Fork rate and Fork asymmetry in WT and HMGB1-KO MEFs. IODs correspond to scatter plots and frequency distribution histograms present in Figure 18a and c. Fork rates correspond to scatter plots and frequency distribution histograms present in Figure 19a and c. Fork asymmetry corresponds to box plot present in figure 20a, and the same number of values was used for the scatter plots present in Figure 20b and c. Data derived from three independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	IODs		Fork speed		Fork asymmetry	
	WT MEFs	HMGB1-KO MEFs	WT MEFs	HMGB1-KO MEFs	WT MEFs	HMGB1-KO MEFs
Number of values	202	180	627	596	133	126
Median	101,1	104,8	1,491	1,842	15,09	17,05
Mean	106,1	110,7	1,505	1,944	23,2	29,83
Std. deviation	39,92	43,87	0,5861	0,7341	28,16	40,4
P-value	-	0,3774	-	<0.0001	-	0,4398

Supplementary Table 9: Numerical values and statistic parameters of IOD, Fork rate and Fork asymmetry in WT and H1-KO mES cells. IODs correspond to scatter plots and frequency distribution histograms present in Figure 23a and c. Fork rates correspond to scatter plots and frequency distribution histograms present in Figure 24a and c. Fork asymmetry corresponds to box plot present in figure 25a, and the same values were used for the scatter plots present in Figure 25b and c. The last two columns encompass the values of figure 25f, which represent the velocities of the non-stalled forks represented in 25b and c. Data derived from three independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	IODs		Fork speed		Fork asymmetry		Fork rate (non-stalled forks only)	
	WT mES	H1-TKO mES	WT mES	H1-TKO mES	WT mES	H1-TKO mES	WT mES	H1-TKO mES
Number of values	175	189	679	684	140	144	210	170
Median	91,64	78,87	1,725	1,153	13,79	26,55	1,671	1,265
Mean	100,4	84,49	1,801	1,231	28,78	46,86	1,799	1,358
Std. deviation	38,86	35,63	0,6676	0,5234	39,93	63,15	0,7417	0,5107
P-value	-	<0.0001	-	<0.0001	-	0,0006	-	< 0.0001

Supplementary table 10: Nuclear EU fluorescence intensity values from transcription recovery experiments in mES cells. Numerical values correspond to the ones presented in Supplementary Figure 2. These values were then normalized to the median intensity value of their correspondent untreated samples to generate the graphic present in figure 27d. Data derived from two independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	EU							
	wt mES ASS	**wt mES 3hDRB**	**wt mES 3hDRB+1h release**	**wt mES 3hDRB+3h release**	**H1-TKO mES ASS**	**H1-TKO mES 3hDRB**	**H1-TKO mES 3hDRB+1h release**	**H1-TKO mES 3hDRB+3h release**
Number of values	138	114	149	145	100	61	112	103
Median	595,5	213,5	312	292	476,9	117	282,5	342
Mean	619,1	210,2	311,2	297,9	559	146,1	290,4	409,2
Std. deviation	229,8	59,35	80,2	93,18	277,7	120,2	103	213,1
P-value	<0.0001	-	<0.0001	<0.0001	<0.0001	-	<0.0001	<0.0001

Supplementary Table 11: Nuclear S9.6 (R-loop) fluorescence intensity values in mES cells. Numerical values correspond to the ones presented in Figure 28b. Data derived from two independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	S9.6 fluorescence (a.u.)			
	WT mES RNAse H-	**WT mES RNAse H+**	**H1-TKO mES RNAse H-**	**H1-TKO mES RNAse H+**
Number of values	147	115	69	76
Median	186	161,3	310	209,2
Mean	196,7	168,8	342,9	211,7
Std. deviation	57,59	40	134,5	60,23
P-value	-	<0.0001	-	<0.0001

Supplementary Table 12: Nuclear R-loop fluorescence intensity values in mES cells at different points of the cell cycle. Numerical values correspond to the ones presented in Figure 29b. Data derived from two independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	S9.6 fluorescence (a.u.)							
	WT mES				H1-TKO			
	G1/G2/M	Early S	Middle S	Late S	G1/G2/M	Early S	Middle S	Late S
Number of values	141	162	87	54	173	122	86	46
Median	138,2	226,7	157,9	146,2	415,3	601,0	382,1	513,0
Mean	165,4	245,1	190,1	200,3	523,2	673,3	466,1	586,3
Std. deviation	138,4	144,9	187,1	251,1	393,2	378,2	277,8	404,6
P-value	<0.0001	-	<0.0001	0,0003	<0.0001	-	<0.0001	0,1014

Supplementary Table 13: Nuclear γH2AX fluorescence intensity values in mES cells at different points of the cell cycle. Numerical values correspond to the ones presented in Figure 29c. Data derived from two independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	γ-H2AX fluorescence (a.u.)							
	WT mES				H1-TKO			
	G1/G2/M	Early S	Middle S	Late S	G1/G2/M	Early S	Middle S	Late S
Number of values	141	162	87	54	173	122	86	46
Median	152,3	232,1	159,5	133,9	391,0	442,0	386,2	425,2
Mean	222,6	240,2	196,3	161,1	461,7	468,1	403,1	433,4
Std. deviation	225,1	135,5	125,1	99,3	308,1	184,3	160,6	171,6
P-value	0,0012	-	0,0081	<0.0001	0,0309	-	0,0085	0,4062

Supplementary Table 14: Nuclear R-loop and γH2AX fluorescence intensity values in WT and HMGB1-KO MEF cells. Numerical values correspond to the ones presented in Figure 30a and c. Data derived from two independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	S9.6 fluorescence (a.u.)				γH2AX fluorescence (a.u.)	
	MEFs WT RNAseH-	MEFs WT RNAseH+	MEFs HMGB1-KO RNAse-	MEFs HMGB1-KO RNAse+	MEFs WT	MEFs HMGB1-KO
Number of values	91	88	60	67	189	264
Median	80,54	94,14	57,88	53,22	73,86	62,29
Mean	94,47	92,42	62,37	60,76	136,7	153,8
Std. deviation	43,94	15,88	22,92	27,72	154,7	207,2
P-value	-	0,0226	-	0,2377	-	0,5268

Supplementary Table 15: Nuclear EU fluorescence intensity values in mES cells untreated or treated with transcription inhibitors. Numerical values correspond to all the populations analysed upon exposure to different combinations of transcription blocking compounds represented in Figure 31a). Data derived from two independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	EU							
	ActD-				ActD+			
	WT mES α-amanitin-	WT mES α-amanitin+	H1-TKO mES α-amanitin-	H1-TKO mES α-amanitin+	WT mES α-amanitin-	WT mES α-amanitin+	H1-TKO mES α-amanitin-	H1-TKO mES α-amanitin+
Number of values	87	65	90	88	71	40	118	72
Median	1272	941,6	932,6	421,5	1047	392,9	614,6	225,4
Mean	1285	959,8	997,2	410,3	1007	437,6	656,8	252,6
Std. deviation	376,8	231	377,4	169,4	284,3	242,9	252,4	191,8
P-value	-	<0.0001	-	<0.0001	-	<0.0001	-	<0.0001

Supplementary Table 16: Nuclear R-loop fluorescence intensity values in mES cells untreated or treated with transcription inhibitors. RNAseH controls values are also shown. Numerical values correspond to the ones presented in Figure 31b. Data derived from two independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	S9.6 fluorescence (a.u.)					
	wt mES α-amanitin- RNAse H-	**wt mES α-amanitin+ RNAse H-**	**wt mES α-amanitin- RNAse H+**	**H1-TKO mES α-amanitin- RNAse H-**	**H1-TKO mES α-amanitin+ RNAse H-**	**H1-TKO mES α-amanitin- RNAse H+**
Number of values	182	233	171	175	200	105
Median	217,5	210,7	185,4	572,1	373,4	315,1
Mean	235,9	223,4	210,7	582,3	387,2	327,7
Std. deviation	111,2	107,4	81,19	227,5	148,3	91,46
P-value	-	0,2604	0,0118	-	<0.0001	<0.0001

Supplementary Table 17: IOD values in mES cells untreated or treated with transcription inhibitors. Numerical values correspond to the ones presented in Figure 31c. Data derived from two independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	IODs			
	WT mES α-amanitin-	**WT mES α-amanitin+**	**H1-TKO mES α-amanitin-**	**H1-TKO mES α-amanitin+**
Number of values	66	57	63	61
Median	88,41	83,19	73,42	76,84
Mean	94,17	86,06	79,49	88,6
Std. deviation	32,01	28,04	37,1	41,32
P-value	-	0,1379	-	0,2672

Supplementary Table 18: Same as in Supplementary Table 17 for fork rate measurements. Numerical values correspond to the ones presented in Figure 31d. Data derived from two independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	Fork rate			
	WT mES α-amanitin-	**WT mES α-amanitin+**	**H1-TKO mES α-amanitin-**	**H1-TKO mES α-amanitin+**
Number of values	211	232	205	234
Median	1,405	1,395	1,074	1,316
Mean	1,483	1,426	1,13	1,366
Std. deviation	0,6007	0,4129	0,4119	0,4863
P-value	-	0,5298	-	<0.0001

Supplementary Table 19: Same as in Supplementary Table 17 for fork asymmetry measurements. Numerical values correspond to the ones presented in Figure 31e. Data derived from two independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	Fork asymmetry			
	WT mES α-amanitin-	**WT mES α-amanitin+**	**H1-TKO mES α-amanitin-**	**H1-TKO mES α-amanitin+**
Number of values	39	45	37	45
Median	19,31	14,25	30,83	17,31
Mean	25,01	25,92	58,65	31,55
Std. deviation	20,18	30,63	56,83	43,57
P-value	-	0,3698	-	0,0084

Supplementary Table 20: Nuclear γH2AX fluorescence intensity values in mES cells untreated or treated with transcription inhibitors. Numerical values correspond to the ones presented in Figure 32b. Data derived from two independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	γ-H2AX fluorescence (a.u.)			
	WT mES α-amanitin-	**WT mES α-amanitin+**	**H1-TKO mES α-amanitin-**	**H1-TKO mES α-amanitin+**
Number of values	181	233	213	198
Median	237,3	154,9	453,2	252
Mean	266,8	204,1	484,5	313
Std. deviation	121,8	155,5	151,7	182,9
P-value	-	<0.0001	-	<0.0001

Supplementary Table 21: IOD values in transcription recovery experiments in mES cells. Numerical values correspond to the ones presented in Figure 33b. Data derived from two independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	IODs							
	WT mES ASS	**WT mES 3hDRB**	**WT mES 3hDRB + 1h release**	**WT mES 3hDRB + 3h release**	**H1-TKO mES ASS**	**H1-TKO mES 3hDRB**	**H1-TKO mES 3hDRB + 1h release**	**H1-TKO mES 3hDRB + 3h release**
Number of values	43	41	47	50	47	55	48	46
Median	80,72	76,83	74,25	68,03	50,3	81,17	50,77	54,32
Mean	86,14	79,22	79,42	73,31	53,32	81,55	53,23	59,06
Std. deviation	29,76	27,56	31,89	18,44	27,27	27,99	17,48	21,94
P-value	0,2789	-	0,8344	0,451	<0.0001	-	<0.0001	<0.0001

Supplementary Table 22: Fork rate values in transcription recovery experiments in mES cells. Numerical values correspond to the ones presented in Figure 33c. Data derived from two independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	Fork rate							
	WT mES ASS	WT mES 3hDRB	WT mES 3hDRB + 1h release	WT mES 3hDRB + 3h release	H1-TKO mES ASS	H1-TKO mES 3hDRB	H1-TKO mES 3hDRB + 1h release	H1-TKO mES 3hDRB + 3h release
Number of values	113	114	131	106	194	159	148	171
Median	1,318	1,249	1,175	1,139	0,7327	1,085	0,6314	0,7378
Mean	1,305	1,282	1,247	1,138	0,7711	1,192	0,6847	0,7514
Std. deviation	0,3758	0,3856	0,4087	0,3876	0,3144	0,4695	0,3316	0,3252
P-value	0,614	-	0,2991	0,0087	<0.0001	-	<0.0001	<0.0001

Supplementary Table 23: Fork asymmetry values in transcription recovery experiments in mES cells. Numerical values correspond to the ones presented in Figure 33d. Data derived from two independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	Fork asymmetry							
	WT mES ASS	WT mES 3hDRB	WT mES 3hDRB + 1h release	WT mES 3hDRB + 3h release	H1-TKO mES ASS	H1-TKO mES 3hDRB	H1-TKO mES 3hDRB + 1h release	H1-TKO mES 3hDRB + 3h release
Number of values	22	33	30	25	32	24	29	28
Median	16,53	12,61	20,77	16,4	45,36	18,33	44,05	51,99
Mean	22,77	28,78	26,25	28,45	63,27	20,78	59,71	50,98
Std. deviation	28,87	44,78	22,91	47,19	78,6	16,39	78,27	35,28
P-value	0,8975	-	0,208	0,8016	0,0096	-	0,0188	0,0012

Supplementary Table 24: Numerical values from replication dynamics analysis in HCT-shSLBP cells. Values of IODs, and fork asymmetries correspond to Figures 35a and c, respectively. Fork rate values correspond to Figures 35b and d. Data derived from two independent experiments. Statistical significance was calculated using the Mann-Whitney rank sum test as described in Material and Methods.

	IODs		Fork speed		Fork asymmetry	
	HCT-shSLBP dox-	HCT-shSLBP dox+	HCT-shSLBP dox-	HCT-shSLBP dox+	HCT-shSLBP dox-	HCT-shSLBP dox+
Number of values	109	101	505	414	68	61
Median	65,83	63,1	1,003	1,441	12,56	13,44
Mean	75,52	73,04	1,104	1,546	20,3	22,4
Std. deviation	42,14	33,87	0,4474	0,5757	27,13	24,69
P-value	-	0,9991	-	<0.0001	-	0,477

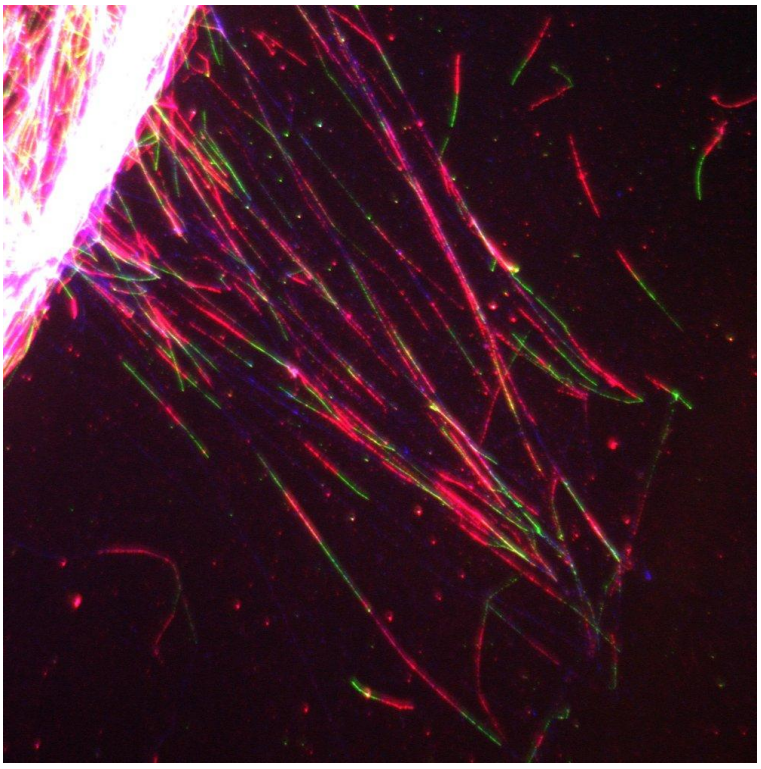
Supplementary Table 25: Primer pairs used in RNAPII transcription elongation measurements. Forward (F) and Reverse (R) primers, as well as annealing temperatures are indicated for each amplicon. qPCR conditions are present in Material and Methods.

Primer	Sequence (5' to 3')	Annealing T^{re} (ºC)
Med13l-Ex1 (F)	CTGGAGGATTGTCACTCCAACC	62
Med13l-In1 (R)	TCCGGGAGGAGAAAGTTGCG	
Med13l-Ex4 (F)	TGTGCGGCCCTATGACAAGG	64
Med13l-In4 (R)	CAGATAACAGATACGCCAGCCC	
Med13l-Ex5 (F)	AGTGTGGAGATAGCTCAGCACC	64
Med13l-In5 (R)	TGCACGCAGTTACGCTGGTG	
Med13l-In-last (F)	AGGTGGCCATGCTGGTGTGC	64
Med13l-Ex-last (R)	CTGGATTGCACGTGAGCCAG	
Inpp5a-Ex1 (F)	ACCGCGGTCCTGCTGGTCAC	64
Inpp5a-In1 (R)	GAAAATGGGGATGTCAGGGTCC	
Inpp5a-Ex4 (F)	AGAATACAACAGGGCGCGTGTC	64
Inpp5a-In4 (R)	GCATGCGTGCCGACTTAGTAC	
Inpp5a-Ex5 (F)	GGAAGCTTTTATTTTCTTCACGAATCC	64
Inpp5a-In5 (R)	GACAACAGAGCTAGAGGGACC	

Supplementary Table 26: Primer sequences used for each gene analyzed by RT-qPCR in mES cells and NIH3T3 MEFs.

Gene	Primer Name	Sequence (5´to 3')
Hprt	HprtA	GCCTAAGATGAGCGCAAGTT
	HprtB	GTGGGAAAATACAGCCAACACT
Haus	HausA	ACATCCATTCCAGTATGTCC
	HausB	GCGTGAACCAGTTGTAGTTC
Vps	VpsA	GGTCGAAGGAAACAAATTGC
	VpsB	TCAATTCGCCACCATGAATG
Mecp	MecpA	ATCATTAGGGTCCAAGGAGG
	MecpB	CTGAAGGTTGGACACGAAAG

Annex II – Publication List



Lombraña R[1], **Almeida R[1]**, Revuelta I[1], Madeira S[1], Herranz G, Bastolla U and Gómez M (2013). High-resolution analysis of DNA synthesis start sites and nucleosome architecture at efficient mammalian replication origins. *EMBO J* 32: 2631-2644. **[1]CO-FIRST AUTHORS**

Lombraña R, **Almeida R**, Álvarez A and Gómez M (2015). R-loops and initiation of DNA replication in human cells: a missing link? *Front Genet* doi:10.3389/fgene.2015.00158

Lombraña R, Álvarez A, Fernández-Justel JM, **Almeida R**, Poza-Carrión C, Gomes F, Calzada A, Requena JM and Gómez M (2016). Transcriptionally driven DNA replication programme of the human parasite *Leishmania major*. *Cell Rep* 16: 1774-17786