Analysis of the contribution of linker histone H1 to the dynamics of RNA:DNA hybrids

Anna Casas Lamesa

Departament de Genètica, Microbiologia i Estadística
Programa de Doctorat de Genètica
Facultat de Biologia
Universitat de Barcelona

ANALYSIS OF THE CONTRIBUTION OF LINKER HISTONE H1 TO THE DYNAMICS OF RNA:DNA HYBRIDS

Memòria presentada per **Anna Casas Lamesa**
per optar al grau de Doctora per la Universitat de Barcelona.

Realitzada sota la direcció dels Drs. Jordi Bernués Martínez i Ferran Azorín Marín a l'Institut de Recerca Biomèdica de Barcelona/Institut de Biologia Molecular de Barcelona, CSIC.

Directors de la tesi		Tutora de la tesi
Dr. Jordi Bernués Martínez	Dr. Ferran Azorín Marín	Dra. Montserrat Corominas Guiu

Doctoranda

Anna Casas Lamesa

Barcelona – 2018

Només en la incertesa
existeix la possibilitat de l'inesperat.

Andrea Köhler
El temps regalat

CONTENTS

CONTENTS

SUMMARY/RESUM

SUMMARY

Histone H1 is a key determinant of higher-order chromatin structures and for a long time it has been considered a main transcriptional repressor. So consistent with its biochemical properties, the absence of histone H1 was expected to have extended consequences in gene expression. On the contrary, upon depletion of histone H1 in *Drosophila melanogaster* (dH1), only a subset of genes was significantly affected [1–3]. These differentially expressed genes were mainly heterochromatic inactive genes that got transcribed [2]. Further studies showed that apart from affecting silencing of heterochromatin, dH1 depletion also induced double-strand breaks (DSBs) on the same heterochromatic sequences that were being up-regulated, suggesting that abnormal transcription of these sequences was resulting in DNA damage [4].

Results reported here suggested that the heterochromatic transcripts up-regulated upon depletion of dH1 induced genomic defects because they were not correctly metabolized and, consequently, formed RNA:DNA hybrids (R-loops). Thus, histone H1 is important in preventing abnormal R-loop accumulation in heterochromatin. On this direction, our approach showed that it was the absence of histone H1 and not simply the relieve of silencing that induced accumulation of R-loops in heterochromatin, as depletion of HP1a (another important heterochromatic element) up-regulated heterochromatin transcription as well, but it neither induced R-loops or DNA damage. Then, further experiments showed that histone H1 was necessary for the binding of hnRNP36 and hnRP48 to heterochromatin. These two proteins are involved in mRNA quality control and, therefore, in the prevention of R-loop formation. Thus, in conditions where histone H1 was absent, the hnRNP36/48 could not bind to the newly synthesized transcript, favoring the formation of abnormal R-loops. Further analyses of the genome-wide distribution of R-loops in *D. melanogaster* showed that histone H1 is important for the regulation of R-loop dynamics beyond heterochromatin, as depletion of histone H1 not only produced a strong accumulation of R-loops in heterochromatin, but also induced important changes in the distribution of physiological R-loops.

Together, we provide evidence of a novel and essential function of linker histones H1 to the regulation of RNA:DNA hybrids in heterochromatin and, thus, to the maintenance of genome integrity and stability. However, further studies are needed to understand the biological importance of the euchromatic changes in R-loop distribution in dH1-depleted cells compared to that of physiological conditions.

RESUM

La histona H1 és un element clau en l'empaquetament d'ordre superior de la cromatina i durant molt de temps se li ha atribuït el paper de repressor transcripcional. Així doncs, d'acord amb les seves propietats bioquímiques, s'esperava que davant la supressió de la histona H1 es produïssin canvis importants d'expressió gènica. De manera contraria però, davant la depleció de la histona H1 (dH1) a *Drosophila melanogaster*, només un petit grup de gens es va veure afectat significativament [1–3]. Aquest grup de gens diferencialment expressats eren principalment gens heterocromàtics silenciats que estaven sent activats [2]. Posteriors estudis van demostrar que a part d'afectar en el silenciament de les seqüències heterocromàtiques, l'absència d'histona H1 també produïda trencaments de cadena doble del DNA en les mateixes seqüències heterocromàtiques que estaven sent activades, suggerint que la transcripció anormal d'aquestes seqüències estava produint dany en el DNA [4].

Els resultats presentats en aquesta tesi proposen que els transcrits heterocromàtics que s'activen en deplecionar la histona dH1 acaben produint defectes genòmics perquè no poden ser correctament metabolitzats i, conseqüentment, formen híbrids d'RNA:DNA (R-loops). Així doncs, la histona H1 és important per prevenir l'acumulació anormal d'R-loops en l'heterocromatina. En aquesta direcció, els nostres resultats demostren que és l'absència de dH1 i no només l'atenuació del silenciament el que provoca l'acumulació d'R-loops en l'heterocromatina, ja que la depleció d'HP1a (un altre component essencial de l'heterocromatina) també activa la transcripció heterocromàtica, però en canvi, no provoca ni R-loops ni dany en el DNA. Després, altres experiments han demostrat que la histona H1 és necessària per la unió d'hnRNP36 i hnRP48 en l'heterocromatina. Aquestes dues proteïnes participen en el control de qualitat de l'mRNA i, per

tant, en prevenir la formació d'R-loops. Així doncs, en condicions en què no hi hagués histona H1, les hnRNP36/48 no podrien unir-se en els nous transcrits, afavorint la formació anormal d'R-loops. Altres anàlisis en la distribució dels R-loops al llarg del genoma de *D. melanogaster* demostren que la histona H1 és important en la regulació de la dinàmica dels R-loops més enllà de l'heterocromatina, ja que la seva absència no només provoca l'acumulació d'R-loops en l'heterocromatina, sinó que també produeix canvis importants en la distribució fisiològica dels R-loops.

En conjunt, en aquest treball s'aporten evidències d'una nova i essencial funció de la histona H1 en la regulació dels híbrids d'RNA:DNA en l'heterocromatina i, per tant, en el manteniment de la integritat i estabilitat del genoma. Tanmateix, més estudis són necessaris per entendre la importància biològica dels canvis en la distribució dels R-loops en les cèl·lules deplecionades de dH1 comparat amb la distribució dels R-loops en condicions fisiològiques.

ABBREVIATIONS

ABBREVIATIONS

AID	Activation-induced cytidine deaminase
CDS	Coding sequence
CGIs	CpG islands
ChIP(-seq)	Chromatin Immunoprecipitation (followed by sequencing)
CRIPR	Clustered regularly interspaced short palindromic repeats
CSR	Class-switch recombination
CTD	C-terminal domain/tail
DNMT	DNA methyltransferase
DRIP	DNA:RNA immunoprecipitation
DSB	Double-strand break
dsRNA	Double-stranded RNA
eccrDNA	Extra-chromosomal rDNA circles
ESCs	Embryonic stem cells
GFP	Green fluorescent protein
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HMT	Histone methyltransferase
HP1:	Heterochromatin Protein 1
HR	Homologous recombination
L3	Third instar larvae
LTR	Long terminal repeats
ncRNA	Non-coding RNA
NHEJ	Non-homologous end joining
NRL	Nucleosomal repeat length
NTD	N- terminal domain/tail
ORF	Open reading frame
PcG	Polycomb group
PCR	Polymerase chain reaction
PEV	Position effect variegation
PTM	Post-translational modification
RNAi	RNA interference
RNAP	RNA polymerase
RNase A	Ribonuclease A
RNH	RNase H

SETX	Senataxin
siRNA	Small interfering RNAs
ssDNA	Single-stranded DNA
TBP	TATA-binding protein
TE	Transposable element
TRF2	TBP-related factor 2
TKO	Triple H1 knock-out
TSS	Transcription start site
TXE	Termination start site
UAS	Upstream activating sequence
UTR	Untranslated regions
WB	Western blot

INTRODUCTION

INTRODUCTION

1. CHROMATIN

1.1. Chromatin structure and the nucleosome core particle

The genome of all eukaryotic organisms is stored inside nucleus in the shape of a compact and folded nucleoprotein complex called chromatin, the structure of which controls essentially all nuclear processes involving DNA; including transcription, DNA replication and DNA repair.

The packaging of DNA into chromatin is primarily controlled by two major types of small, highly basic proteins: the core histones and the linker histones. The first level of chromatin compaction involves the association of DNA with the core histones and the formation of the nucleosome core particle, the smallest packaging unit of the chromatin fiber. The nucleosome core particle consists of a histone octamer comprising two copies each of H2A, H2B, H3 and H4, around which 147 bp of DNA are wrapped in a left-handed 1.65 superhelical turns [5]. These core particles are repeatedly connected by a short stretch of linker DNA, forming a structure resembling beads-on-a-string. A further packaging of DNA into higher order structures involves the linker histone H1, that associates with the nucleosome near the site at which DNA enters and exits the core particle [6]. The complex containing the nucleosome and a linker histone is referred as the chromatosome (**Figure 1**).

Nucleosomes are formed into regularly spaced arrays along DNA and can be mobilized by different ATP-dependent remodeling complexes, such as SWI/SNF or RSC, or ATP-independent ones, like FACT complex [7,8]. Thus, nucleosome positioning is a dynamic process where the average distance between two neighboring nucleosomes changes among species and cell types. While most genomic DNA is occupied by nucleosomes, many functional regions have low occupancy. Thus, nucleosomes occupy preferred positions throughout the genome. For example, in several organisms, promoter regions of transcribed genes are characterized by a nucleosome free region just upstream the transcription start site (TSS), flanked by peaks of high nucleosome occupancy [7]. Then, the nucleosomes are numbered

sequentially relative to TSS; the downstream nucleosome is the +1, and the upstream nucleosome is the −1.

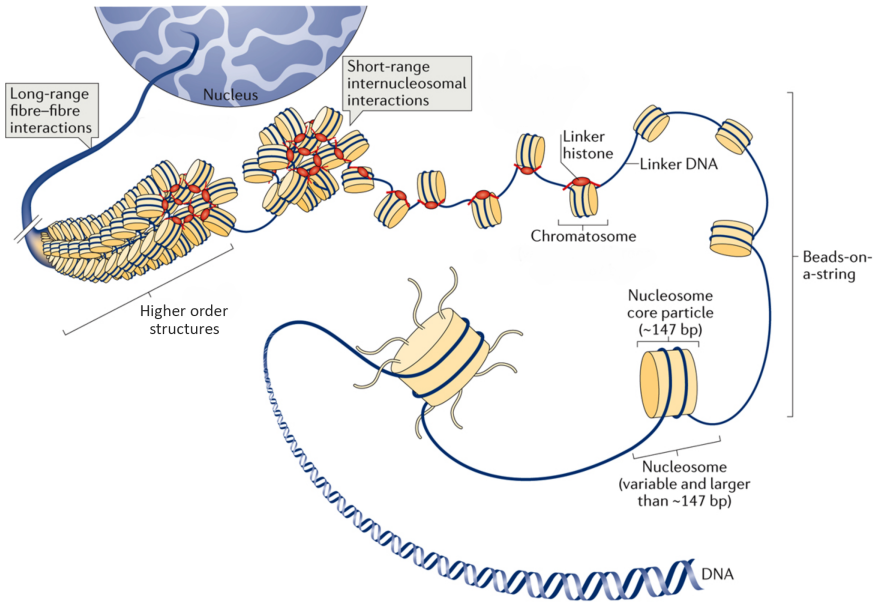


Figure 1. Different levels of DNA compaction and chromatin structure. Histone proteins compact the nucleosomal DNA to fit into the nucleus of the eukaryotic cells. The first level of compaction comes with the core histone octamer, resulting in a DNA-protein complex called nucleosome. Then, the linker histone binds where the internucleosomal DNA enters and exits the nucleosome to form the chromatosome and helps the DNA to fold into higher-order structures. Adapted from [9].

Nucleosome positioning is not determined by a single factor, but rather by the combination of nucleosome remodelers, DNA sequence, transcription factors and other DNA-binding proteins. In respect to the DNA sequence, the histone octamer does not have a binding motif, but considering the constraints in means of the energy required to bend a given genomic sequence, they have preference for those sequences that favor bending around the histone octamer [10]. The strongest effect of DNA sequence on nucleosome depletion is that of poly(dA:dT) sequences, which due to their particular intrinsic bending properties, strongly disfavor nucleosome formation at most promoter sequences [11].

Another level in the complexity of chromatin organization is its spatial organization inside the nucleus. It has been found that the genome is

arranged in a non-random, cell-specific manner, where different regions of intra (cis)- and inter (trans)-chromosomal DNA are specifically looped together in long-range interactions [12].

1.2. Epigenetic regulation

The different cell types comprising an organism have an identical genomic sequence. However, they are not structurally and functionally the same. Thus, the distinct cellular phenotypes are governed by epigenetic mechanisms that manipulate the genome activity and accurately execute particular programs of gene expression, cell-cycle progression and DNA replication in a potentially heritable way. The term "epigenetics" was first used by Conrad Waddington to describe the molecular mechanisms that convert any genetic information into different observable traits or phenotypes [13]. Some of these mechanisms include covalent modifications of cytosine bases, histone composition and histone post-translational modifications, changes in the positioning of nucleosomes and DNA silencing by non-coding RNAs (**Figure 2**).

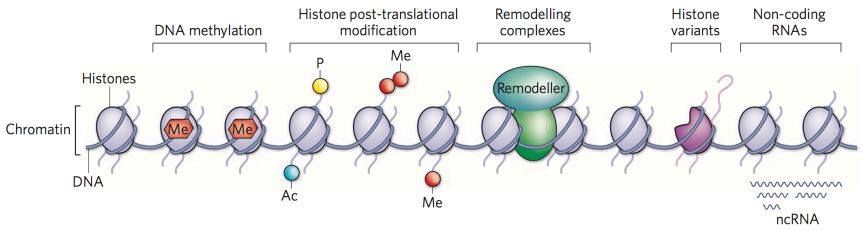


Figure 2. Types of epigenetic modifications. Epigenetic marks are catalyzed by different interrelated mechanisms known to affect chromatin structure: DNA methylation, histone post-translational modifications, nucleosome remodeling by chromatin-remodeling complexes, insertion of histone variants and the effect of non-coding RNAs (ncRNAs). Ac, acetyl; Me, methyl; P, phosphate. Image from [14].

Such epigenetic mechanisms operate coordinately to dynamically regulate chromatin functions and require a substrate that carries this epigenetic information and propagate it information over the time. Thus, histone proteins are the perfect candidates for maintaining and dynamically regulate the epigenetic landscape of eukaryotic genomes. On one hand, histone proteins are subject to a broad range of post-translational

modifications (PTMs), which are thought to contribute to the regulation of genome activity. On the other hand, histones split to daughter cells during each division and transfer this information [15].

1.2.1. Core histone variants

Histone variants are non-canonical variants of histones that are usually expressed at very low levels compared with their conventional counterparts. They are non-allelic variants, representing one or a few amino acid differences, which replace the canonical histones to confer specific expression and novel structural and functional features on the nucleosome, affecting chromatin remodeling and histone PTMs. Also, while the 5' end of canonical histone mRNA consist of a stem-loop, the mRNAs of non-canonical histones are polyadenylated. Differences have also been observed with their transcription regulation. Whereas expression of most canonical histones are synthesized at S phase for a rapid deposition behind replication forks to fill in gaps resulting from the distribution of pre-existing histones, synthesis of histone variants is mainly replication-independent regulated [16].

Among the core histones, H2A has the largest number of variants, including H2A.Z, MacroH2A, H2A-Bbd, and H2A.X. While some are found in most eukaryotic lineages (H2A.Z and H2A.X), others are restricted to vertebrates (MacroH2A and H2A-Bbd). H2A.X and H2A.Z are constitutively expressed and localize throughout the genome, although H2A.Z shows some enrichment in intergenic regions. The defining feature of H2A.X is a C-terminal motif that is rapidly phosphorylated in response to, and at the site of, DNA double-strand breaks (DSBs) and provides signals to concentrate repair proteins [17]. Therefore, although H2A.X may be incorporated at a diluted level throughout the genome, it may be preferentially deposited at the disrupted nucleosomal arrays in response to DSBs.

Drosophila melanogaster has a single H2A variant, H2Av, which combines specific features of the two mammalian homologs, H2A.Z and H2A.X [18]. H2Av is phosphorylated at S137 in response of either DNA damage or to promote transcription. On one side, H2Av phosphorylation in response to DNA damage is important for recruitment of repair and checkpoint proteins, independent of the repair mechanism involved. This modification becomes detectable within minutes of the DSB induction and it is propagated bi-directionally from the break [18]. On the other side, Jil1 phosphorylates H2Av

at promoters, which stimulates PARP-1 enzymatic activity in the surrounding chromatin, leading to further modification of histones and chromatin loosening, to finally promote transcription [19].

Histone H3 has two major variants found in most eukaryotic lineages, H3.3 and centromeric H3 (CENP-A). H3.3 is the major histone H3 that is available for deposition outside of S phase and differs from canonical H3 by only a few amino acids. Nevertheless, the role of its incorporation at specific regions of the genome is still unclear. Whereas the N-terminal tails of the canonical H3 and H3.3 variants are nearly identical, CENP-A proteins have no sequence similarity. CENP-A is a conserved essential protein highly enriched at centromeres and mainly responsible for kinetochore assembly [20,21]

Histone H2B and histone H4 are markedly deficient in variants.

1.2.2. Post-translational modifications of core histones

In addition to histone variants, PTMs of histone proteins have emerged as key players in chromatin structure and regulation. These modifications, which are reversible and mostly occur on the N-terminal histone tail, act as signaling platforms for regulatory proteins, which incorporate, recognize and remove these PTMs, called writers, readers and erasers, respectively, and also for several core histone chaperones, which facilitate deposition or removal of core histones from chromatin [22,23].

PTMs include, among others, phosphorylation at serines (S) and threonines (T), methylation at lysines (K) and arginines (R), and acetylation, ubiquitylation, sumoylation and ADP-ribosylation at lysines (**Table 1**). Furthermore, each lysine residue can accept one, two or three methyl groups (K-me, K-me2, K-me3, respectively), and arginines can be either mono- or di-methylated (R-me, R-me2, respectively; asymmetrically or symmetrically) [22–24].

CHROMATIN MODIFICATIONS	RESIDUES MODIFIED	FUNCTIONS REGULATED
Acetylation	K-ac	Transcription, Repair, Replication, Condensation
Methylation (K)	K-me1, K-me2, K-me3	Transcription, Repair
Methylation (R)	R-me1, R-me2a. R-me2s	Transcription
Phosphorylation	S-ph, T-ph	Transcription, Repair, Condensation
Ubiquitylation	K-ub	Transcription, Repair
Sumoylation	K-su	Transcription
ADP ribosylation	K-ar	Transcription, Repair, Condensation

Table 1. Overview of different classes of PTM identified on histones. The name of the modification, the residue affected and the functions associated are shown. me2a and me2s refers to methylations occurring asymmetrically and symmetrically, respectively. Adapted from [23].

Histones are modified at many sites. However, depending on the nature and position of the modification, PTMs of histones are linked to distinct chromatin structures; regions of open euchromatin or condensed heterochromatin, and specific combinations of them are associated with specific functions and genomic elements (enhancers, TSS, or gene bodies). Therefore, histones can help regulating transcription through these modifications. For instance, H3K27me3 and H4K20me3 are found at inactive promoters, while H3K4me3, H3 and H4 acetylation at active ones [23]. To increase complexity, some opposite marks coexist on adjacent histones within a promoter. This is the case of H4K4me (active transcription) and H3K27me3 (transcription repression) at developmental gene promoters (bivalent promoters) in ESCs [25].

Thus, the interpretation of the core histone PTMs depends on the cellular context and gene studied, but their possibilities are immense.

1.2.3. DNA methylation

DNA modifications, much like histone variants and histone PTMs, play an active role in chromatin regulation, as they can regulate the interaction and subsequent functions of factors that bind DNA. DNA methylation in eukaryotes is a heritable chromatin modification that primarily occurs at cytosine residues (5mC). The DNA methylation pattern in any given cell is the outcome of independent but dynamic methylation and demethylation events, catalyzed by a family of DNA methyltransferases (DNMTs) that

transfer a methyl group from S-adenyl methionine and DNA demethylases, respectively. The majority of DNA methylation occurs on cytosines that precede a guanine in the context of CpG dinucleotides [26]. The CpG dinucleotide distribution across the human genome is irregular, but they concentrate in dense regions called CpG islands (CGIs). These CGIs are found in more than a half of human genes and when they associate with promoters, they are typically hypomethylated [27].

Methylation of the DNA has classically been shown to silence genes by blocking the recruitment of transcription factors and by serving as a binding site for methyl-binding domain factors often associated with chromatin remodelers [28] and transcriptional co-repressor complexes [26]. However, recent studies revealed that a number of DNA-binding transcription factors bind specific DNA sequences in a methylation dependent manner [29]. Hence, methylated DNA sequences can create binding sites for either activators or repressors of transcription, increasing the complexity and diversity of gene regulation. In this regard, several advances in the field have proposed 5mC as a 'fifth base', in addition to adenine, cytosine, guanine and thymine. Besides of 5mC, the discovery of Tet enzymes that mediate the oxidation of 5mC to 5-hidroxymethylcytosine (5hmC) have not only identified an active demethylation pathway, but also a new epigenetic mark candidate [30,31].

1.3. Types of chromatin

The different levels of compaction in interphase nucleus observed by microscopy experiments in 1928 by Heitz [32], forms the basis of the classical categorization of eukaryotic genomes into two main categories: euchromatin, which is characterized by a less condensed, gene-rich and transcriptionally active chromatin, and heterochromatin, which stands out for being a highly condensed, gene-poor and transcriptionally repressed chromatin.

Newer evidences provided by Filion et al. suggested the need of categorize chromatin into a more accurate manner (**Figure 3**). They identified five major chromatin states out of the ChIP-seq profiles of 53 proteins from *Drosophila melanogaster* Kc167 cell line, which are characterized by distinctive protein combinations and histone modifications and are named in five different colors (GREEN, BLUE, BLACK, RED and YELLOW) [33].

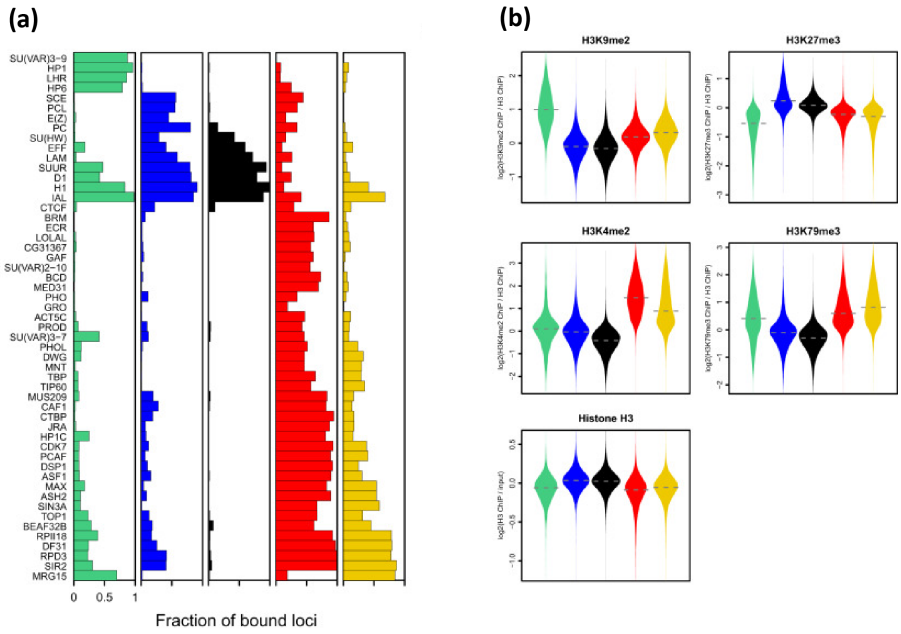


Figure 3. The five chromatin types identified by Filion et al. in *Drosophila*. Distinct protein combinations **(a)** and levels of histone H3 PTMs and histone H3 **(b)** are presented for each of the five-chromatin types. Figure taken from [33].

GREEN and BLUE chromatins correspond to the known heterochromatin types. GREEN chromatin is marked by high content of Su(var)3-9, heterochromatin protein 1 (HP1), and the HP1-interacting proteins LHR and HP6. BLUE matches with Polycomb group (PcG) repressed chromatin.

BLACK is the most abundant chromatin, but it is depleted of PcG, Su(var)3-9, HP1 and related proteins, stressing its difference from GREEN or BLUE heterochromatin types. It is gene-poor and particularly late replicating, and constitutes a strongly repressive environment, as essentially all genes exhibit low levels of transcription. Nevertheless, some genes can become active at restricted tissues.

RED and YELLOW are the euchromatin types where most genes are active. Levels of H3K4me3 and H3K79me3 are typically high and H3K9me2 and H3K27me3 are generally low. Nevertheless, they differ in their molecular organization, regulate a different subset of genes and differ in their H3K36me3 content (RED chromatin genes lack H3K36me3).

Other groups have proposed different chromatin state annotations. For example, Kharchenko et al. profiled a set of eighteen histone modifications and identified nine prevalent combinatorial patterns in *D. melanogaster* S2 cells, which allowed an identification of a larger number of active chromatin states compared to the distribution proposed by Fillion et al. [34]. Also, Ernst et al. defined fifteen chromatin states using nine chromatin marks (core histone PTMs and CTCF) across nine different human cell types. They characterized cell-type-specific patterns of chromatin states focusing on differences in regulatory elements and functional interactions [35].

1.4. Heterochromatin

Heterochromatin is a well-defined form of chromatin with important roles in genome organization, genome stability, chromosome inheritance and gene regulation. In general, formation of heterochromatin is a multi-step process associated with chromatin compaction, silencing and reorganization of the nuclear domains, and can be divided into two types: facultative heterochromatin and constitutive heterochromatin.

The accessibility of facultative heterochromatin is regulated in order to control gene expression, as it often forms at developmentally regulated genes where the level of compaction changes depending on development and/or environmental signals [36,37]. Perhaps, the best-studied example of facultative heterochromatin is the inactive X chromosome (Xi). CGIs of silenced X-linked genes are highly methylated, although the total levels of methylation on Xi appear to be reduced relative to the active X chromosome (Xa) [38]. In addition, the chromatin of Xi is enriched in histone H3 methylated at lysine 27 [39] and macroH2A [40].

In contrast, constitutive heterochromatin preferentially assembles at repetitive elements and maintains high compaction levels all the time. It is basically concentrated in centromeres, therefore implicated in proper chromosome segregation during cell division, and pericentromeric regions. Constitutive heterochromatin contains a distinct molecular signature consisting of heavy methylation of CpG dinucleotides (in higher eukaryotes) and chromatin enrichment in CENP-A, H3K9me and concomitant association with HP1 [41]. Pericentromeric heterochromatin occupies about 30% of fly and human genomes, and is composed by large contiguous stretches of repeated DNA [42]. Accordingly, heterochromatin

has a crucial role in preventing the deleterious consequences that can derive from unscheduled transcription of these repeated DNA. These repetitive elements are satellite DNA sequences and transposable elements (TE). On one side, satellite DNAs are the main constituents of heterochromatin. They are tandemly repeated sequences that are present as long uninterrupted arrays in genetically silent heterochromatic regions. On the other side, TEs can move into new locations upon activation. TEs are mostly located in heterochromatin and proximal heterochromatin-euchromatin transition zones.

Heterochromatin assembly can be divided into three distinct steps: establishment, spreading and maintenance. Heterochromatin is established at nucleation centers through the targeting of histone modifying activities by transcription factors or non-coding RNAs. In higher eukaryotes it starts with the addition of one, two or three methyl groups to histone H3 lysine 9 (H3K9) by a histone methyltransferase, which serves as a binding site for HP1. HP1 is proposed to mediate chromatin condensation via multimerization of HP1 molecules on nearby nucleosomes [43]. HP1 multimers also serve as a scaffold to recruit other interacting factors, including H3K9 methyltransferases and histone deacetylases. This association promotes compaction of the chromatin structure and is thought to prevent or limit access of proteins to DNA to achieve the repressive functions of heterochromatin [36,37]. Subsequently, the combined actions of readers and writers lead to a mechanism of self-propagation along a large domain of chromatin in a sequence-independent manner [36,37,44]. Position effect variegation (PEV) in *D. melanogaster* is the classical example of this positive feedback loop that extends heterochromatin domains (**Figure 4**). PEV describes the mosaic expression of euchromatic genes when they are inserted in or near heterochromatin. Then, after relocation, the repressive heterochromatin components will spread and induce transcriptional silencing of that gene. After an initial choice (active/inactive), the cell will maintain the repressive state in all daughter cells, resulting in patches of cells expressing and not expressing the gene [37].

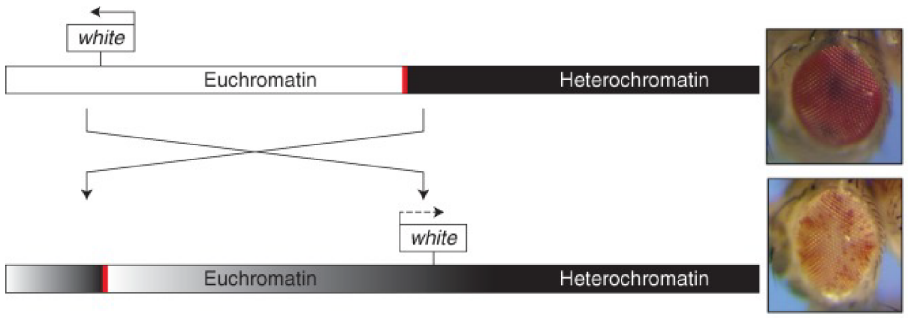


Figure 4. Schematic illustration of *white* position effect variegation. Normally, the *white* gene is located in euchromatin (white bar) and expressed in every cell of the *Drosophila* eye, resulting in a red-eye phenotype (upper-right image). However, translocation of the gene into the pericentric region (black bar; bottom) induces spreading of the heterochromatin. Therefore, the loss of silencing in some cells during differentiation results in a variegated phenotype (red-white colored-eye; bottom-right image). Adapted from [37].

However, because chromatin can spread and induce erroneous silencing, mechanisms to restrict its expansion are necessary. These barriers can be formed through association of chromatin with nuclear structures to form chromatin loops, through association of proteins that promote nucleosome-free regions, nucleosome turn-over to prevent spreading of histone modifications [45] and also binding of anti-silencing factors, such as histone-modifying enzymes to actively counteract heterochromatin associated histone modifications [46]. Furthermore, euchromatin is marked by a variety of modifications that antagonize heterochromatin assembly, so protecting them is also critical for establishing a heterochromatin boundary (H2A.Z and active histone PTMs) [22,47,48] (**Figure 5**).

Nevertheless, in some systems, establishment of heterochromatin is dependent on transcription of heterochromatic repetitive DNAs and non-coding RNAs. In fission yeast, for example, the processing of repetitive DNA transcripts is mediated by the RNAi machinery. In fact, heterochromatin transcripts are converted to double-stranded RNAs by the RNA-dependent RNA polymerase complex and are then processed by the ribonuclease Dicer into small interfering RNAs (siRNAs). The siRNAs guide the RNA-induced transcriptional silencing complex back to nascent transcripts originating from repeats. Then the RNA-induced transcriptional silencing complex associates

with Clr4 methyltransferase, which initiates H3K9 methylation and results in recruitment of chromodomain proteins (such as Chp1, Chp2 and Swi6) to form a self-sustaining loop of heterochromatin assembly [46].

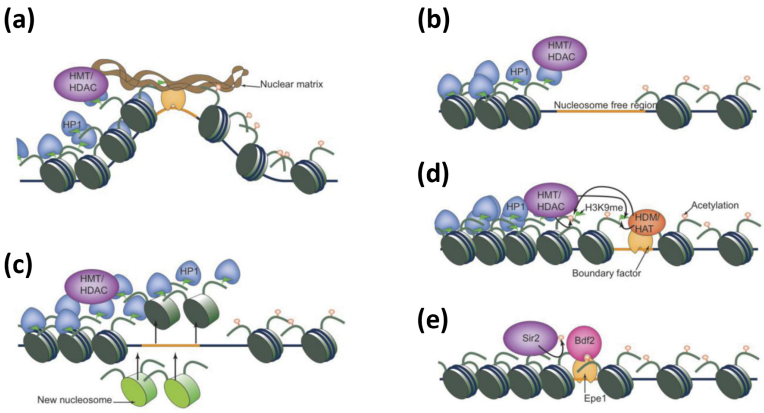


Figure 5. Mechanisms for heterochromatin barrier formation. In general, heterochromatin is flanked by boundary elements, which limit its spreading into surrounding regions in order to maintain correct gene expression profiles. These boundary elements **(a)** can cluster and associate with nuclear structures to form chromatin loops, **(b)** form nucleosome free regions and **(c)** present a high rate of histone turnover to prevent the spreading of heterochromatin modifications, and **(d)** can recruit histone modifying enzymes to counteract heterochromatin histone modifications or **(e)** bind proteins that protect euchromatin modifications. Adapted from [48].

It is important to notice that heterochromatic marks are not excluded from euchromatin. For example, in *D. melanogaster*, HP1a has been shown to bind multiple euchromatic sites and be associated with the expression of some developmental and heat shock–induced puffs [49]. Furthermore, this interaction is independent of H3K9me [50]. Similar observations were done in mammalian chromatin, where HP1 and H3K9me have been found in the transcribed region of active genes [51]. On the same direction, active transcription of protein coding genes is not excluded from heterochromatin. The first heterochromatic genes were discovered in *D. melanogaster* [52], but they have also been reported in other species (mouse, humans, plants).

In addition to histone methylation, the DNA within heterochromatin regions is highly methylated in many higher eukaryotes such as mammals and plants. Thus, heterochromatic silencing in these systems relies on a

crosstalk between DNA methylation and histone methylation. Methylation of DNA in mammals, recruits methyl-binding domain proteins, which, analogously to HP1, serve as platforms for chromatin-modifying factors [36].

2. LINKER HISTONE H1

2.1. Histone H1 structure and family

In metazoans, linker histones H1 are relatively small proteins (~200 amino acids) with a conserved tripartite structure comprising a short, flexible N-terminal tail (NTD), a central globular domain and a long basic C-terminal tail (CTD) (**Figure 6 a**). The globular domain is highly conserved among all H1 subtypes, consists of a winged helix fold and interacts with the nucleosome protecting a 20-bp of linker DNA. The NTDs are not observed in the crystal structures of the nucleosome and little is known about this region. However, free NTD peptides are disordered, and in nucleosomes, the NTD seems to participate in the internucleosomal interactions that drive chromatin fiber condensation. The CTD appears to be the primary determinant of H1 binding to chromatin and is also involved in protein-protein interactions [53,54].

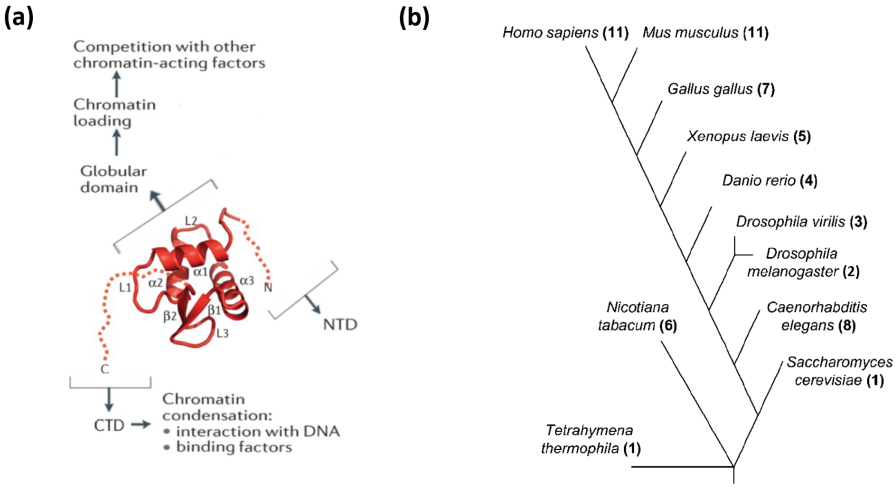


Figure 6. Histone H1 structure and its variants among different species. (a) Crystal structure of chicken erythrocyte linker histone H5. Linker histones in metazoans have a conserved tripartite structure, consisting on the globular domain and the intrinsically disordered N- and C- terminal tails (NTD and CTD, respectively). The globular domain and the CTD mediate the multiple biochemical activities of H1. **(b)** This phylogenetic tree offers an overview of the number of histone H1 variants in different selected organisms. The number of variants is shown in brackets. Adapted from [9,55].

Both core and linker histones mainly use positively charged arginine and lysine residues to interact with the backbone phosphates of DNA through electrostatic interactions in the nucleosome and chromatosome core particles. For example, the CTD of histone H1, which accounts for more than a half of the linker histone H1 sequence, is composed of ~40% of lysine residues [53,54].

In contrast to the extremely evolutionary conserved core histones, linker histones display much higher sequence variability between different species [56]. For example, the level of sequence homology between the histone H1-like protein in *Saccharomyces cerevisiae* (Hho1) and the human H1 is 31% identical and 44% similar, whereas for H4, it is 92% identical and 96% similar. Aside from sequence variability, the different histone H1 proteins are also very variable in structure. For instance, Hho1p contains two globular domains, while *Tetrahymena thermophila* completely lacks a globular domain. Eukaryotes also differ in the number of histone H1 variants [53] (**Figure 6 b**).

2.2. Histone H1 functions

2.2.1. Contribution to chromatin structure

Almost all the nucleosomes in the chromatin contain one molecule of the linker histone H1 (**Figure 7 a**). However, given the abundance of histone H1, the existence of multiple variants or subtypes in many metazoans has further limited progress in understanding their functions, mainly due to lack of specific antibodies recognizing each isoform. Consequently, for a long time histone H1 was thought to be a general chromatin architectural protein. Most recently, these difficulties are being overcome and an increasing number of studies have changed the classical view of histone H1 as a merely structural component of chromatin, providing new insights into the importance of linker histones in epigenetic regulation of transcription and DNA replication, DNA repair and genome stability [2,3,57].

Histone H1 binding to chromatin influences the nucleosomal repeat length, controls the accessibility of linker DNA between two neighbor nucleosomes to DNA-binding proteins, and regulate the condensed states of chromatin (**Figure 7 b**). Notably, a depletion in histone H1 correlates linearly

with a reduction of nucleosomal spacing [58]. Hence, linker histones are required to fold chromatin into higher-order structures, including the so-called 30-nm fiber [6]. It has been observed that nucleosomes are assembled in discrete domains of varying sizes along the chromatin fiber, termed nucleosome clutches, suggesting that *in vivo* 30 nm chromatin structures only exists as short fragments rather than as continuously folded fibers [59].

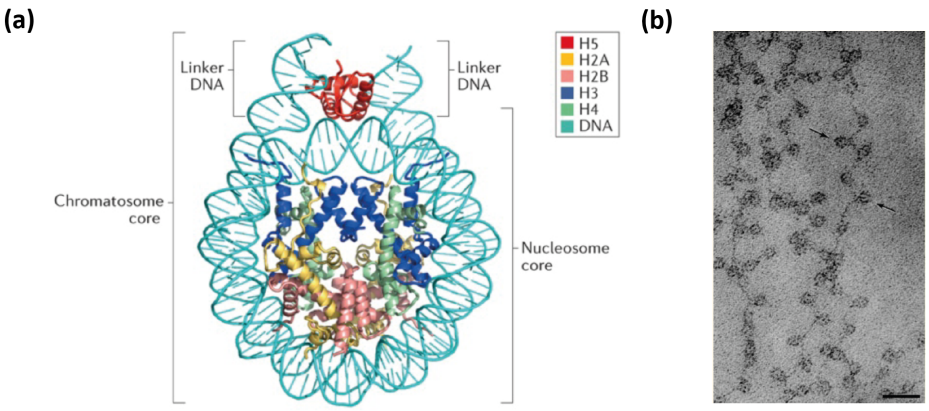


Figure 7. Histone H1 contribution to chromatin structure. (a) In this crystal structure, the chicken linker histone H5 (from **(Figure 6 a)**) sits on the dyad of the nucleosome and interacts with linker DNA. The binding of linker histone confers to chromatin a higher level of compaction. **(b)** Electron microscope image of chromatin showing the beads-on-a-string structure. Scale bar corresponds to 30 nm and nucleosomes are indicated by arrows. Adapted from [9,60].

The association of H1 to chromatin is highly dynamic with an average residence time of minutes [61]. In contrast, core histones have residency times on a timescale of hours [62]. Histone H1 binds to nucleosomes without any known DNA sequence specificity and the mode of binding of histone H1 to the linker DNA is unclear, with several models competing to explain this interaction. The first crystal structure at near-atomic resolution was published in 2015 for the globular domain of chicken erythrocyte histone H5 [63] (**Figure 7**), and since then several advances have made to understand the interactions between the linker histones and DNA. These studies suggest that a small number of residues within the globular domain of the linker histone determine its orientation and binding mode (on-dyad versus off-dyad).

Considering the connection between histone H1 binding and chromatin compaction, which limits access of proteins to DNA molecule, depletion of

H1 was expected to influence global DNA function, having important effects on nuclear structure and gene expression. Surprisingly, unexpected results came from knockout studies in different eukaryotes (see chapter 2.2.3).

2.2.2. Contribution to heterochromatin formation

Probably, one of the best-understood connections between H1 and heterochromatin assembly comes from studies in *D. melanogaster*. Loss of *D. melanogaster* linker histone H1 protein (dH1) caused derepression of more than 50% of TEs, but only about 10% of protein-coding genes, highlighting the essential role of dH1 in silencing heterochromatic elements [1]. dH1 was also observed to participate in the formation of heterochromatin through the interaction with H3K9me2/3, the main PTM associated with heterochromatin. In fact, dH1 tethers Su(var)3-9 to heterochromatin, facilitating methylation of H3K9 and providing a binding site for HP1a (**Figure 8a**) [3]. Although the majority of HP1 proteins are localized in heterochromatin, some isoforms have diverged from heterochromatic functions. In *D. melanogaster*, HP1a predominantly localizes to heterochromatin, whereas HP1c is found in euchromatin and HP1b is found both at euchromatic and heterochromatic domains [64]. Another study in *D. melanogaster* revealed the existence of another pathway of heterochromatin formation involving histone H1, but independent of H3K9me. In this case, the unphosphorylated STAT (Signal Transducer and Activator of Transcription) binds HP1a and both interact with histone H1 to participate in the establishment or maintenance of a certain kind of heterochromatin structure (**Figure 8b**) [65]. This finding indicates that the JAK-STAT pathway, which mediates cytokine signals to activate gene expression, also controls the epigenetic status of cells. So given the evolutionary conservation of the canonical JAK-STAT pathway among different species, this non-canonical mode of JAK-STAT responsible of heterochromatin formation may also be present in vertebrates.

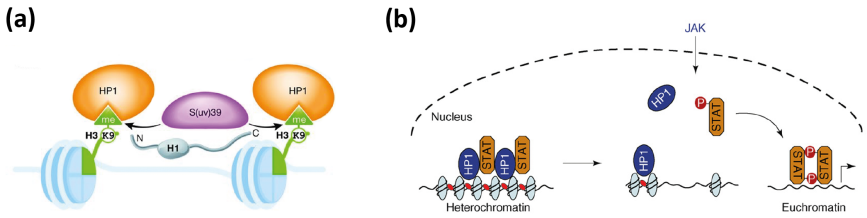


Figure 8. Histone H1 participates in heterochromatin assembly through two independent pathways. One possibility suggests that **(a)** Su(var)3-9, a methyltransferase protein recruited by histone H1 to chromatin, mediates the methylation of H3K9, giving rise to H3K9me2/3 that serves as a binding platform for HP1a (left); another possibility proposes that **(b)** histone H1 (red circle) tethers STAT and HP1 to chromatin (right). Adapted from [3,66]

Several works have also described links between H1 and heterochromatin components in mammalian cells. For example, methylation of lysine K26 in the NTD of H1.4 (H1.4K26me) provides a binding platform for HP1 and is thus linked to transcriptional repression and heterochromatin formation [67]. H1.4K26me is the most abundant methylation occurring on a human linker histone [68]. A similar PTM is found in *D. melanogaster* (K27me2), suggesting a conserved function for this modification [69]. Chromatin immunoprecipitation with antibodies specific for individual human H1 subtypes also pointed at H1.4 to be associated with inactive and condensed chromatin [70].

2.2.3. Regulating the epigenetic state of chromatin

In all organisms analyzed, the absence of histone H1 did not result in an increase of basal transcription, as it was expected for a global transcriptional repressor. Instead, it affected a specific set of genes. Depletion of H1 in *T. thermophila* revealed that H1 is not essential in this organism and that only a specific group of genes is up- or down-regulated [71]. In mouse, elimination of only one H1 variant, and even some pairs of variants did not cause defects in mouse development [72,73]. The absence of any marked phenotype appeared to be due to up-regulation of the remaining subtypes, resulting in maintenance of a normal H1-to-nucleosome stoichiometry. However, the knockout of three H1 isoforms (*H1c*, *H1d*, and *H1e*) simultaneously, led to a 50 % reduction in H1:core particle ratio and the embryos died around mid-gestation, demonstrating the essential role of linker histones in mammals and the importance of a correct stoichiometry of histone H1 deposition on chromatin [74]. Embryonic stem cells (ESCs) derived from these triple H1

knock-out (TKO) embryos also contained about 50% of the normal H1 amount leading to a global reduction in nucleosomal repeat length and less compact packaging of chromatin with selective changes in gene expression [57,75].

The distribution of histone H1 is not uniform along the genome, but rather it is modulated depending on the genomic context. For instance, histone H1 is depleted in TSS of actively transcribed genes that are enriched for active histone marks (H3K4me3), whereas H1 occupancy is increased within silenced chromatin domains enriched for repressive histone marks (H3K9 and H3K27). Also, histone H1 potentiates the repressed state of chromatin by preventing binding of histone methyltransferases (HMTs) that establish positive methylation marks [76,77]. A strong correlation has also been found with histone H1 and hypoacetylation of core histones. It seems that H1 can repress histone acetylation by negatively regulating histone acetyltransferases (HATs) [78] and by interacting with histone deacetylases (HDACs) [79]. However, these are only some of the possibilities for histone H1-mediated transcriptional regulation, as histone H1 is not only linked to gene repression. For example, it has been shown that H1 proteins are necessary for hormone-induced transcription of the mouse mammary tumor virus (MMTV), proving its contribution to transcription activation. In this context, histone H1 binding to the MMTV promoter induces a different conformation of chromatin, which is needed for the recruitment of the hormone receptor and transcription factors [80].

Histone H1 also seems to regulate DNA methylation at specific loci. Yang et al. found in mouse ESCs that some histone H1 subtypes specifically interact with the DNA methyltransferases DNMT1 and DNMT3B and help to recruit these enzymes to the control regions of the h19 and Gtl2 imprinted genes [77].

Moreover, high-throughput chromatin conformation capture applied to TKO cells found that the frequency of inter-domain interactions increased, suggesting that structural changes within topologically associating domains (TADs) occurred when the H1-to-nucleosome stoichiometry was reduced. Interestingly, this increase in inter-TAD interactions correlated with massive epigenetic changes, involving changes in activating histone marks and a

decrease in the sites of DNA methylation [75] and suggesting that linker histones participate in the modulation of the 3D landscape of the genome.

2.3. Linker histone H1 variants

The linker histone H1 family includes multiple variants. The positively charged residues in the globular domains are important for nucleosome binding and are well conserved between species and individual variants (**Figure 9**). In contrast, the linker histone tails among the variants are much less conserved.

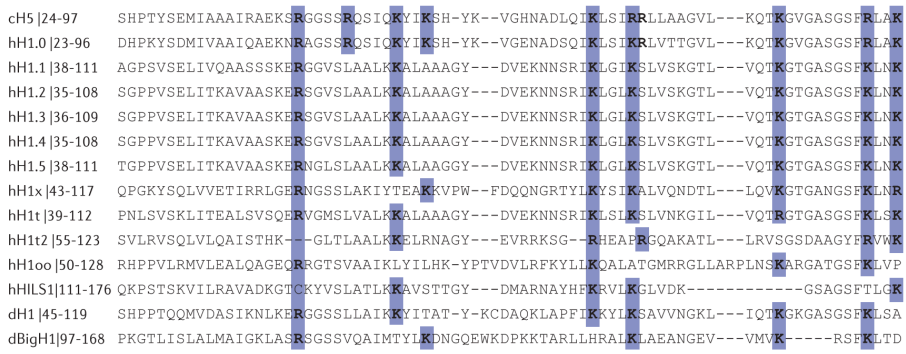


Figure 9. Sequence alignment of the globular domain of different histone H1 variants. The globular domain of the linker histones is highly conserved, whereas the N- and C- terminal domains (not shown) are more variable. The c refers to chicken, h refers to human and d refers to fly isoforms. In blue are highlighted the positively charged residues, which are important for the binding to the nucleosome. Image from [9].

Depending on the organism, different number of histone variants can be found with different degree of similarities. For example, *D. melanogaster* has two histone H1 variants (dH1 and BigH1), whereas in mice and human, eleven genes have been described: seven somatic (H1.1, H1.2, H1.3, H1.4, H1.5, H1.0 and H1.X; **Table 2**) and four germ line specific variants (H1oo in oocytes; H1t, H1T2 and HILS1 in spermatocytes).

The H1.1-H1.5 somatic variants are usually referred as the main histone H1 subtypes. They are replication-dependent genes clustered together in chromosome 6 (HIST1), they are intronless, have short 5'- and 3' non-coding sequences and their mRNA ends with the characteristic histone-mRNA hairpin loop structure that is involved in the replication-dependent regulation of these genes. Instead, H1.0 and H1.X are expressed in a cell-

cycle independent manner and are located on human chromosome 22 and 3, respectively. Their genes are also intronless, but rather than ending with a stem-loop sequence, their mRNAs are polyadenylated [81–84].

NAME	GENE	LOCUS	EXPRESSION
Cell cycle-dependent			
H1.1 (H1a)	HIST1H1A	6p21.3-22	Tissue-specific
H1.2 (H1c)	HIST1H1C		Ubiquitous
H1.3 (H1d)	HIST1H1D		
H1.4 (H1e)	HIST1H1E		
H1.5 (H1b)	HIST1H1B		
Cell cycle-independent			
H1.0 (H1(0))	H1F0	22q13.1	Differentiated cells

Table 2. Linker histone H1 somatic variants in humans, gene and expression pattern. Mouse H1 variant names are shown in brackets. Adapted from [85].

Then, the possibilities for the histone H1 subtypes are immense; they can play both specific and redundant functions, and functions can also be specific to cell type. For that, there are regions on the genome where the H1 subtypes might be interchangeable (e.g. global chromatin organization) and specific regions in the genome where a distinct H1 variant might be needed for a specific function (e.g. precise regulation of genes). For example, it has been observed that the genomic distribution of human histone H1.2 strongly correlates with low gene expression, as it is less abundant than other variants at the TSS of inactive genes and promoters enriched in H1.2 tend to be repressed [86]. Enrichment of different isoforms at certain chromatin types has also been observed. For example, H1.2 and H1.4 are relative depleted in actively transcribed chromatin, whereas heterochromatin contained the four somatic H1.2-H1.4 subtypes [87].

Moreover, H1 composition varies through development and differentiation, as well as between cell types and during disease associated processes. For example, in *D. melanogaster*, the germ line specific variant BigH1 is present in germ cells during the first hours of embryogenesis, but it is replaced by the somatic dH1 variant at cellularization [88]. Another example of replacement can be observed for H1.0, which has been found to be low expressed in pluripotent cells and accumulate on the regulatory regions of differentiation and pluripotency genes during differentiation, becoming the predominant H1 variant in differentiated cells [89]. H1.0 has also been implicated in changes in chromatin structure and function accompanying malignant transformation [90].

2.4. Post-translational modifications of histone H1

H1 resembles the situation in core histones where individual residues can be post-translational modified both at conserved and unique residues. Even though H1 PTMs are much less understood than those of core histones, it is known that they have important roles in regulating chromatin structure and function, as they modulate the interactions with increasing number of proteins.

Post-translational modifications in histone H1 are mainly found on the N-terminal domain and primarily include phosphorylation, acetylation, methylation and ubiquitylation. However, phosphorylation is considered the most frequent and important modification, which depends on different kinases and phosphatases [91,92]. In mammals, histone H1 phosphorylation leads to both chromatin condensation and decondensation depending on the site of phosphorylation. Phosphorylation of histone H1 is often regulated in a cell cycle-dependent manner, as many of the phosphorylation sites have found to happen in consensus CDK-phosphorylation residues and, furthermore, the abundance of phosphorylated forms changes during cell cycle progression. In general, it is preferentially serine-phosphorylated in interphase and mitosis and threonine residues are phosphorylated during mitosis [91,93]. In *D. melanogaster*, however, none of the phosphorylation sites correspond to consensus CDK-sites, as dH1 contains no CDK sites at all [69]. Interestingly, phosphorylation of histone H1 at non-CDK sites have also been found within the N-terminal tail in different species [92,94] and a non-CDK site

has been mapped in the central globular domain of mouse H1.1 and H1.4, which may regulate the interaction of linker histones with DNA [91].

Hence, the relative levels of H1 phosphorylation respond to different stimuli and in addition to functions on cell-cycle, H1 phosphorylation has also been linked to chromatin compaction, DNA damage, apoptosis, DNA ligation, cell differentiation, aging, and cancer [95].

2.5. Linker histone H1 in *Drosophila melanogaster*

The fruit fly *Drosophila melanogaster* has only one somatic histone H1 variant, the product of the *His1* gene. *His1* is present in about 100 copies, located in chromosome II within the tandemly repeated units of the histone cluster (**Figure 10**). Each repeat contains all regulatory sequences to allow expression of each histone genes. Recently, it has been shown that only twelve histone repeats are needed to recapitulate normal histone expression levels [96].

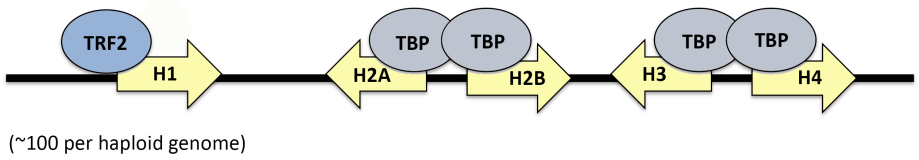


Figure 10. Genomic organization and regulation of Drosophila melanogaster histone H1 locus Representation of one repeat unit of the histone gene locus, which is tandemly repeated up to ~100 copies. The arrows point the direction of transcription and the use of either TBP or TRF2 transcription factors for histone H1 and core histones is indicated.

Despite the insertion of histone H1 locus into the histone repeats, regulation of histone H1 expression is rather independent of core histones. *D. melanogaster* histone H1 genes use TRF2 (TBP-related factor 2) instead of TBP (TATA-binding protein) that it is used for the regulation of core histone genes. Besides, TRF2 is detected on histone H1 genes all along the cell cycle, whereas for core histones, RNA pol II and TBP accumulated at S-phase, suggesting that histone H1 and core histones are regulated differently [97]. Indeed, histone H1 is transcribed throughout S phase in contraposition to core histones, which are only transcribed in a short but intense pulse during early S phase. Another difference is found with the mRNAs half-life, which is much shorter for H1 transcripts than for core histones transcripts [98].

2.6. Histone H1 profiles in cancer

The first link between epigenetics and cancer came from studies in DNA methylation and gene expression, which highlighted general alterations of DNA methylation in cancer. Although less investigated, histone proteins are emerging as targets of important alterations in cancer, as transcriptional and genetic alterations of histone H1 have been found in a variety of cancers.

Examination of various tumor types analyzed by The Cancer Genome Atlas (TCGA) reveals that H1 mRNA levels are often perturbed in cancer and also variant-specific changes can be observed in different tumor types. Aberrant expression patterns for H1 isoforms in cancer are highly heterogeneous, showing variability both inter- and intra-tumor. Furthermore, a positive or negative correlation between the levels of histone H1 and the histopathological grade of the tumors has been established in some cases, a well as the assessment of the aggressiveness of the tumor depending on the heterogeneity in histone H1 levels [99]. For example, increased protein expression levels of H1.5 replication-dependent variant positively correlates with more aggressive prostate tumors [100]. On the contrary, H1.0 replication-independent variant is normally down-regulated in various cancers. However, H1.0 alterations are highly heterogeneous and correlate with tumor grade; grade III-IV gliomas, which are poorly differentiated aggressive cancers, show overall lower levels of H1.0 compared to grade II-lowly aggressive tumors [90].

Also, new insights have reported some recurrent mutations in histone H1 that might disrupt the function of normal histone H1 proteins in the cell. These are the case of mutations in follicular lymphoma or in chronic lymphocytic leukemia [101].

3. RNA:DNA HYBRIDS

3.1. R-loops (RNA:DNA hybrids) in normal conditions

RNA:DNA hybrids, extensively known as R-loops, are three-stranded nucleic acid structures composed of a RNA:DNA hybrid and the displaced single-stranded DNA (ssDNA), complementary to the RNA molecule. R-loops were firstly identified in bacteria, but nowadays they have been described in many organisms. They are generated by a transcription event where the RNA re-anneals with the DNA naturally, leaving the non-template strand alone,

and are key intermediates in specific biological processes; such as replication (**Figure 11a, b**), recombination (**Figure 11c**), transcription (**Figure 11d**) and DNA repair (**Figure 11e**). Beyond these specific roles, R-loops are considered rare transcriptional byproducts that might compromise the integrity of the genome.

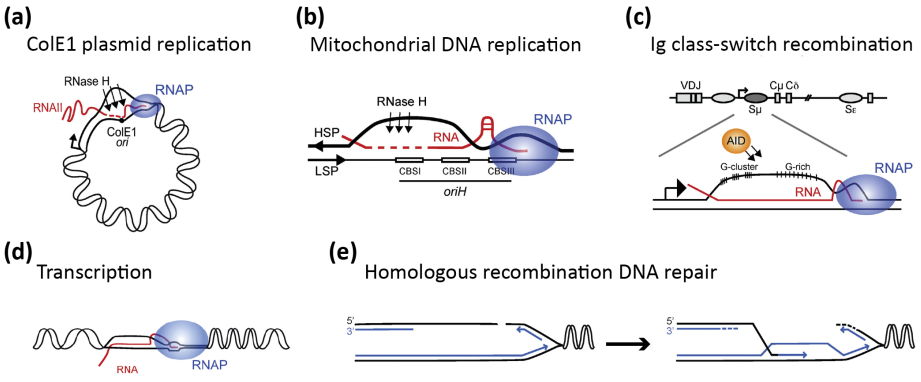


Figure 11. RNA:DNA hybrids are natural intermediates in several biological processes (a) ColE1 initiation of replication is primed with an RNA that forms a stable RNA:DNA hybrid. Then, the RNA is processed by the RNase H1 to generate a 3' end that is extended by the DNA polymerase I. **(b)** Mitochondrial DNA replication at the leading-strand origin is also coupled to transcription through the formation of an R-loop. **(c)** R loops form during transcription of Ig class-switch recombination. Gs are indicated as thin vertical lines. **(d)** The nascent RNA during a transcription event can re-anneal with the DNA template and, therefore, form an R-loop. **(e)** During DNA repair through homology recombination, an R-loop is formed. Adapted from [102].

In vitro characterization of RNA:DNA hybrids suggested that their formation depends on a number of features, among which is important to remark the G·C content, the status of DNA supercoiling, and the presence of DNA cleavages [103].

(a) DNA replication

R-loops can be formed during replication in prokaryotes. For example, the initiation replication of ColE1-type plasmids in *E. coli* relies on the formation of an RNA polymerase (RNAP)-driven sequence, which forms an stable RNA:DNA hybrid with the leading-strand DNA template. Then, the RNA is

processed by the RNase H1 to generate a 3' end that is extended by the DNA polymerase I [104].

Another long-known example of DNA synthesis mediated by an R-loop occurs during bacteriophage T4 infection [105].

R-loops are also a prerequisite to the initiation of transcription of mitochondrial DNA. This replication mechanism reminds the ColE1 replication, where the leading strand is being primed by an RNA molecule produced by the mitochondrial RNA polymerase [106].

(b) Recombination

In order to generate diverse antibody isotypes, vertebrate B cells perform the Ig class-switch recombination (CSR). It consists on a recombination event between two switch (S) regions, which are located downstream of a promoter, are highly repetitive and the non-transcribed strand has a high G-content, thereby generating a G-rich transcript. Then, R-loops act as key intermediates during CSR, in which they form during transcription throughout S regions due to the G-rich nature of their non-transcribed strand [103].

R-loops are also formed in CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 system (**Figure 12**), which provides immunity for prokaryotes against bacteriophages and other genetic elements, and became popular when adapted as a genome-editing tool. Short DNA fragments from infecting phage genomes are incorporated into the host genome within the CRISPR locus. Then, transcription of CRISPR gives rise to an RNA that is processed into short RNAs. These small RNAs grouped with Cas9 act as guides to target homologous DNA loci, generating a trans R-loop. Cas9 then cuts the target locus on each strand, generating a DSB, and ultimately silences the target DNA.

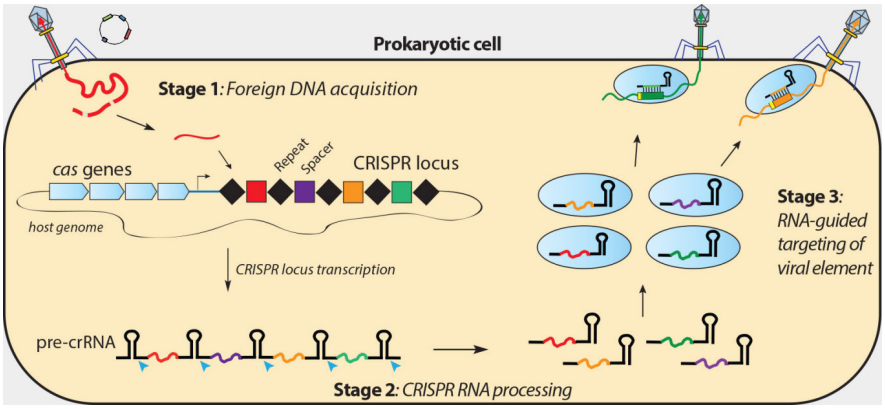


Figure 12. R-loops as key intermediates in a prokaryotic adaptative immunity. CRISPR-mediated immunization occurs through the uptake of DNA from invasive genetic elements such as plasmids and viruses, followed by its integration into CRISPR loci. These loci are subsequently transcribed and processed into small interfering RNAs that guide Cas nucleases to a complementary DNA sequence, resulting in (trans) R-loop intermediates.

(c) Transcription

The most accepted mechanism for R-loop formation during transcription suggests that they are formed as the nascent transcript emerges from the transcription machinery and hybridizes with the complementary DNA template. R-loops formed during transcription are dynamically formed and resolved, and have a half-life of 10 to 20 minutes [107], suggesting that they might be compatible with the normal dynamics of the transcription cycle, as paused promoter-proximal RNAPII half-life is around 7 minutes [108].

During transcription, the RNA synthesis is accompanied by the generation of positive and negative DNA supercoiling in front of and behind the transcription machinery, respectively. Therefore, unwinding of the DNA double helix by negative supercoiling may allow the RNA to re-anneal to its DNA template. Also, DNA hybrid-facilitating DNA sequences, such as G-quadruplex structures or nicks in the non-template DNA strand, could potentiate R-loop accumulation during transcription.

(d) DNA repair

Among the different types of DNA damage, DSBs are the most deleterious defects, as they affect both DNA strands and an inappropriate repair can give

rise to serious genomic and chromosomal defects. Therefore, when DSBs occur, the cell needs to rapidly and efficiently implement a solution. Two possible pathways have been described to repair DSBs. On one side, there is the rapid but error-prone non-homologous end joining (NHEJ) pathway, and, on the other side, the more accurate homologous recombination (HR) pathway, where the sister chromatid is used as the preferential repair template.

However, yeast cells lack of a homologous template for repair. Thus, when DSBs occur, they use a transcript RNA as a template for free-error DSB repair, as the transcript can hybridize with its complementary DNA [109]. Ohle et al. showed that RNA:DNA hybrids form as part of the HR-mediated DSB repair process and that RNase H enzymes are essential for their degradation and efficient completion of the repair mechanism [110], suggesting a role for RNA:DNA hybrids in favoring genomic stability. Similar results have observed in human cells, where R-loops also participate in DSB resolution by HR. More precisely, R-loops can be formed at the resected DNA ends on a DSB generated during S/G2-phase and contribute to the recruitment of members of the HR repair mechanism. At the same time, HR is in complex with RNaseH2, which controls R-loop levels at DSBs, since an excess of R-loops could have detrimental effects [111].

3.2. RNA:DNA hybrids in genome dynamics

Recent studies discovered that R-loop formation over specific, conserved, hotspots occurs at thousands of genes in mammalian genomes and represents an important and dynamic feature of chromatin patterning and gene expression. RNA:DNA hybrids are implicated in all stages of transcription, from initiation to termination, becoming an important regulator of expression. The majority of R-loop forming sequences enriched at the 5' and 3' end of genes are characterized by an asymmetric distribution of guanines and cytosines (GC-skew), probably because the positive GC-skew could favor the thermodynamic stability of R-loop formation [112].

The location of an R-loop within the promoter could influence transcription by disrupting or enhancing transcription-factor binding to any of these sites. For example, Ginno et al. proved that R-loops formed at CGI containing promoters negatively correlate with DNA methylation and are associated with activation of gene expression, suggesting that RNA:DNA

hybrids prevent methylation and transcriptional silencing at CGI promoters. Regulation of the antagonistic relationship between R-loops and CpG promoter methylation is important because its deregulation can lead to defects such as the Aicardi-Goutières syndrome, which is characterized by elevated levels of R-loops within abnormal hypomethylated DNA regions [113]. Interestingly, the R-loops formed at CGI promoters have been proposed to also mediate initiation of DNA replication at these regions [114].

By contrast, R-loops mapped at transcription termination sites have been reported to be critical for transcription termination and RNA Pol II release. Indeed, Senataxin (SETX), a human helicase protein that unwinds the R-loop structures formed downstream the polyadenylated sites, allows the 5'-3' exonucleolytic degradation of the RNA attached to the RNA polymerase II [115].

During transcription elongation, the negatively supercoiled DNA behind the elongating RNA polymerase is susceptible to invasion by the mRNA transcript to form R-loops, reducing the efficiency of transcription. Contrarily, R-loops formed in the gene bodies of yeast, in particular, at the second exon of open reading frames (ORF), were found to have beneficial effects as they facilitate splicing [116].

These structures are also involved in regulating histone modifications. Castellano-Pozo et al. showed that R-loop-accumulating *S. cerevisiae* mutants were enriched in H3S10P, a mark of chromatin condensation and essential for maintaining genome stability, and that this H3S10P accumulation was dependent of R-loops. The same group studied this relationship of R-loops and chromatin condensation marks in several other eukaryotes and found the same positive correlation [117].

Finally, trans R-loops can also occur during gene silencing of non-coding RNAs (ncRNAs), for example, in *Arabidopsis Thaliana* for the COOLAIR lncRNA [118] and in the human *Ube3a* antisense transcript [119].

3.3. Activities controlling R-loops

RNA:DNA hybrids are more stable than dsDNAs (Roberts and Crothers, 1992) and the relative stability of these hybrids depend on the length, the content of deoxypyrimidines/deoxypurines and the A·T/U proportion of the oligomer [120]. The RNA:DNA hybrids adopt a special conformation that is in between the B-form of dsDNA and A-form of dsRNA, which might be crucial

for its recognition by the removing elements. Although R-loops are constantly formed in cells, little is known about the factors that assist their formation. However, given the potential deleterious consequences of persistent R-loops, multiple cellular mechanisms operate to regulate their levels. These mechanisms are involved in removing R-loops once formed or preventing their formation (**Figure 13**).

One of the best-characterized factors responsible for the active removal of R-loops is RNase H (RNH) enzymes. RNH specifically degrades the RNA moiety in RNA:DNA hybrids in a sequence-independent manner [121,122]. In most organisms, two classes of RNH are found: type 1 and type 2. RNH1 consists of a single polypeptide and is present in the nucleus and mitochondria and is essential for mitochondrial replication. RNH2 is a heterotrimer and it is believed to be mainly a repair enzyme, removing, for example, misincorporated ribonucleotides into DNA. RNH2, in contrast to RNH1, can cleave a single nucleotide embedded in a DNA duplex. However, both RNH enzymes have some overlapping specificities in R-loop resolution; it was shown recently that RNH1 is mainly responsible for the resolution of R-loops of transcription-associated RNA:DNA hybrids and RNH2 process R-loops generated during DNA replication and repair [123]. Apart from degrading R-loops, RNH2 activity is also involved in removing the Okazaki fragments from the newly synthesized lagging strand during DNA replication and mis-incorporated ribonucleotides during the DNA replication process [122].

In addition, a number of helicases have been shown to unwind the hybrids *in vitro*, such as the yeast Sen1 or homologous human SETX and the human DHX9 [115], which also acts on G4 structures [124].

A connection between supercoiling and formation of R-loops also seems to occur during transcription. Topoisomerase I enzymes resolve the negative torsional stress behind the transcribing RNAP II to prevent annealing of the nascent RNA with the DNA template and avoiding the formation of R-loops. In *S. cerevisiae*, for example, it has been shown that inherent RNA:DNA hybrids in ribosomal DNA are increased in *top1Δtop2Δ* mutants and further enhanced upon depletion of RNase H1 [125].

Moreover, to avoid the formation of R-loops, eukaryotic cells use RNA-processing and RNA-export factors that co-transcriptionally wrap and pack the nascent transcripts into ribonucleoprotein particles (RNPs) and quickly

export them to the cytoplasm, reducing the rehybridization possibilities of the nascent mRNA to the transiently opened DNA strand behind RNAP [126].

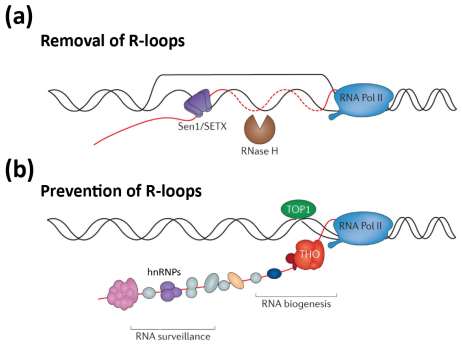


Figure 13. Factors controlling RNA:DNA hybrid levels. Different proteins are involved in **(a)** removing RNA:DNA hybrids, such as RNase H enzymes, helicases (Sen1/SETX) and in **(b)** preventing their formation, such as and topoisomerases (Top1) or mRNA biogenesis and processing factors (THO or hnRNPs). Image adapted from [127]

3.4. R-loops in mutant cells and disease

As mentioned before, R-loops arise naturally in many physiological processes and cells utilize diverse mechanisms to regulate their formation. As the nascent transcript emerges from the RNAPII, it may have two possible outcomes. Either it is precisely controlled, cotranscriptionally packaged into RNPs and exported outside the nucleus, or it may invade the DNA duplex behind the elongating DNA polymerase to form R-loops. Therefore, persistent R-loops due to failure in any of the mechanisms that correctly control R-loop levels are commonly associated with DNA damage and genome instability. Recent studies have implicated R-loops in the pathology of different human diseases [128], further underlining the importance of maintaining the correct R-loop balance.

Different scenarios propose that R-loops can drive genome instability through DNA breakages [129]. Single or double DNA strand breaks are one of the most potent inducers of genome instability as they can lead to mutations, neoplastic transformation or cell death. These breaks can occur either by insults from extrinsic environmental sources or as a result of cellular metabolism. For example, R-loops accumulated due to the absence of several RNA-processing factors or due to the inhibition of topoisomerase I are actively processed to DSBs by the transcription-coupled nucleotide

excision repair endonucleases XPF and XPG (Xeroderma Pigmentosum types F and G [130]). Also, DSBs can arise from transcription-replication conflicts when the replication fork collapses due to R-loop accumulation [131].

Studies in the chromatin area have shown that depletion of histone H1 in *D. melanogaster* induces DNA damage and genome instability. Interestingly, this DNA damage has been presented in the form of DSBs and concomitantly with up-regulation of inactive genes, particularly those genes located in heterochromatin [2,4]. Thus, it could be speculated that from this uncontrolled upregulation of heterochromatic elements, co-transcriptional R-loops were formed, which in turn compromised genome integrity and induced DSBs. Genome instability is a major cause of many developmental disorders and human diseases [129]. Therefore, advances to decipher the molecular mechanisms that contribute to maintain genome stability are of great interest to the biomedical research community.

OBJECTIVES

OBJECTIVES

Histone H1 is an important contributor to chromatin structure and function. Hence, given the importance of chromatin regulation in preserving genome integrity, the main objective of this thesis is to better understand the role of histone H1 to genome stability.

In particular, we aim to decipher the contribution of linker histones H1 to R-loops dynamics in heterochromatin and, thus, the mechanism through which histone H1 proteins prevent R-loop-mediated DNA damage by addressing the following objectives:

1. **Study of the contribution of *Drosophila melanogaster* linker histone H1 (dH1) to the regulation of R-loop dynamics.**
    a. Characterize the specific role of dH1 in preventing R-loop accumulation.
    b. Understand the mechanism by which dH1 prevents abnormal R-loop formation and accumulation in heterochromatin.

2. **Study the biological importance of RNA:DNA hybrids in normal conditions and the relevance of histone H1 in R-loop distribution.**
    a. Map the RNA:DNA hybrids genome-wide in wild type and dH1-depleted conditions.
    b. Analyze the RNA:DNA hybrids formed at genic and intergenic regions.
    c. Assess the particularities of R-loop distribution upon depletion of dH1.

3. **Study the contribution of histone H1 to genome stability in tumor derived cells.**
    a. Characterize the histone H1 and DNA damage levels of different cancer cell lines.
    b. Overexpress H1.4 histone H1 variant in HT29 cancer cell line.
    c. Evaluate the contribution of H1.4 to DNA damage.

RESULTS

RESULTS

1. dH1 IS IMPORTANT FOR THE REGULATION OF R-LOOP DYNAMICS

In previous studies, it was observed that although *Drosophila* histone H1 (dH1) is uniformly distributed along the genome, its depletion affects expression of a relatively small number of genes in a regional manner. Depletion mainly up-regulates inactive genes, induces extra-chromosomal rDNA circles (eccrDNA) accumulation and increases γH2Av in response to DSBs [2]; suggesting that dH1 depletion causes DNA damage and genome instability.

Phosphorylation of the histone H2A variant H2Av has a dual meaning, as it can either bind to active promoters to promote transcription or signal DNA damage upon DSBs. Along the results part and unless indicated, γH2Av will be used as a proxy for DNA damage.

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) experiments against γH2Av, showed that most of the DSBs induced by histone H1 depletion accumulated at repetitive heterochromatic regions, and cell-sorting followed by WB showed that this damage occurred mainly during S-phase; suggesting that instability was linked to problems during replication of heterochromatin [4]. In this regard, we wondered **whether the DNA damage observed in heterochromatin was the result of abnormal R-loop accumulation**, as the unregulated expression of heterochromatic transcripts in dH1-depleted cells could favor their retention in chromatin, facilitating R-loops formation and subsequent DNA damage due to, for example, collisions between the replication fork and the RNAPII.

1.1. Depletion of dH1 induces R-loop accumulation in S2 cells

Following the purpose of assessing if R-loops accumulate in the same heterochromatic regions where DNA damage was produced upon dH1 depletion [4], the distribution of RNA:DNA hybrids was analyzed in conditions of dH1 depletion.

First, in order to obtain an efficient depletion of histone H1 in S2 cells, a double-strand RNA against the whole gene of dH1 was used, as it was described for a different *Drosophila* cell line[132]. Using this dsRNA, S2 cells

were treated during a period of 8 days and important depletion levels for the histone H1 protein were obtained (reduction of 40%). Depletion levels were estimated by western blot (WB) (**Figure 14**). The specificity of dH1 depletion was confirmed, since no changes were observed either in untreated cells or treated with a dsRNA against the bacterial β-galactosidase gene (LacZ).

Next, RNA:DNA immunoprecipitation (DRIP) followed by high-throughput sequencing for these dH1-depleted and untreated cells was performed. DRIP conditions were established after several trials for S2 cells and two independent experiments were carried out. Genomic DNA was first extracted and treated with Ribonuclease A (RNase A) prior to DRIP in order to eliminate any artifacts due to free RNA, which might be present during cell lysis [107]. Genomic DNA was sonicated and the size was checked on a DNA agarose electrophoresis gel. Optimal fragments ranged in size of 200-300 bp. 10% of the genomic extract was separated to be used as an input and the S9.6 antibody, which recognizes RNA:DNA hybrids, was used to immunoprecipitate R-loops.

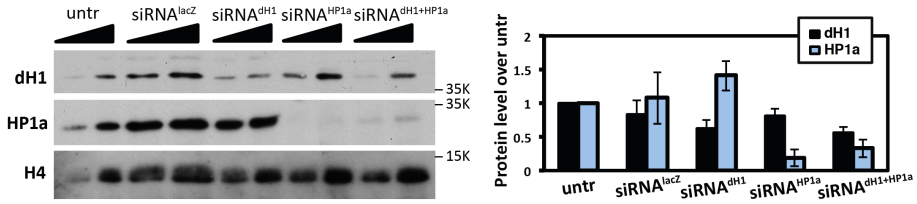


Figure 14. Depletion of dH1 and HP1a. Efficacy of dH1 depletion is tested in WB with specific antibody. Results are the mean of independent depletion experiments (n>4). On the same WB, we also show efficacy of depletion for HP1a and dH1-HP1a (see 1.2-1.5).

To ensure the specificity of the S9.6 antibody when recognizing R-loops, two DRIPs from the same sample were performed in parallel. One was directly immunoprecipitated and the other was treated with RNH just before the immunoprecipitation to remove partially the RNA:DNA hybrids. The resulted immunoprecipitated material and input were sequenced and bioinformatically analyzed using a multi-hit analysis. The coverage of sequencing ranged from 20 to 30-million unique reads per sample. Results showed that S9.6 peaks were distributed along the entire genome in both

dH1-depleted and untreated cells and strongly overlapped (Figure 15 a/b). However, only those regions where there was a >30% reduction (FC < -1.5) on coverage after RNH treatment were considered as R-loop-positive regions. Then, differentially enriched or depleted regions between untreated and dH1-depleted conditions were determined, which detected 189 regions where R-loops abundance specifically increased in dH1-depleted cells, in front of only 14 where R-loops abundance was reduced upon dH1 depletion (Figure 15 c).

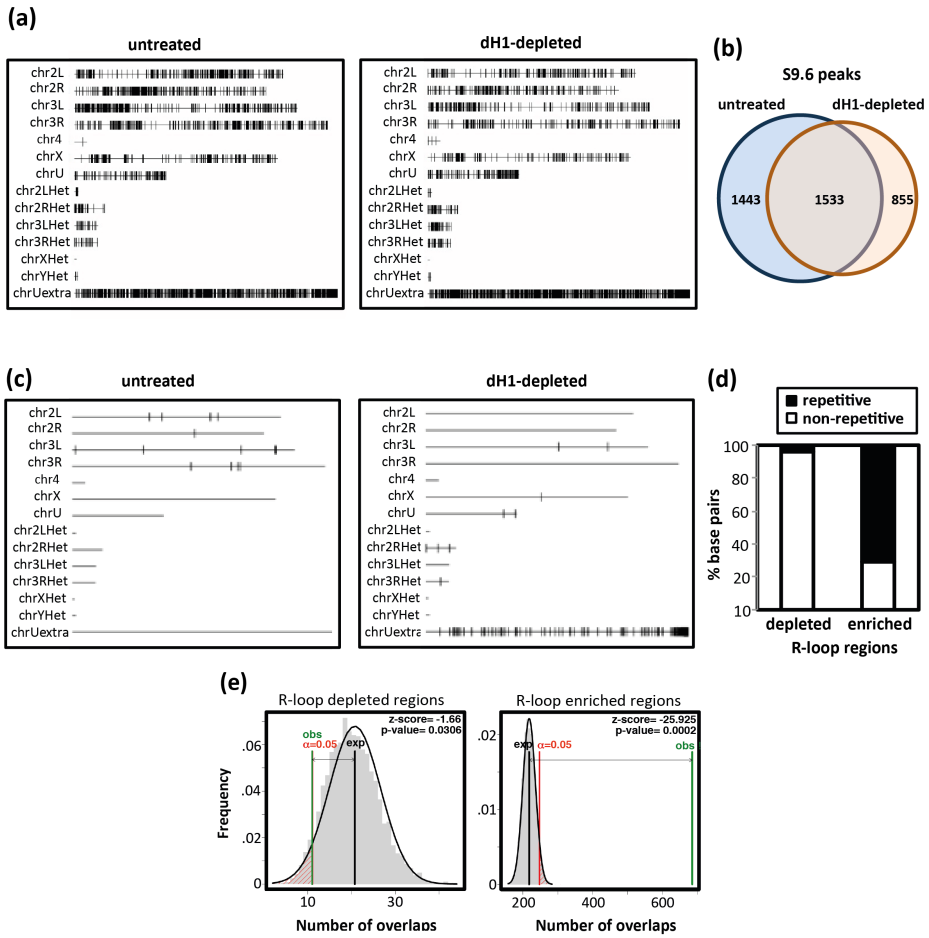


Figure 15. DRIP-seq analysis. (a) Chromosomal distribution of S9.6 peaks detected in control untreated (left) and dH1-depleted cells (right). 2L and 2R, and 3L and 3R correspond to chromosome 2 and 3 left and right arms respect to the position of the centromere, respectively. Chromosome 4 and X are oriented with the centromere to the right. 2LHet, 2RHet, 3LHet, 3RHet, XHet and YHet correspond to partially assembled pericentromeric heterochromatin regions of the indicated chromosomes. U and U Extra correspond to an unordered and not oriented assembly of unplaced

sequence scaffolds. **(b)** Venn diagram showing the intersection between S9.6 peaks in control untreated cells and in dH1-depleted cells. **(c)** Differential enrichment analysis of the regions showing >30% reduction on coverage after RNH treatment. **(d)** The proportion of base pairs (bp) matching to repetitive and non-repetitive elements, as determined by RepeatMasker analysis, are presented for regions showing specific R-loops enrichment and depletion in dH1-depleted cells respect to control untreated cells by DRIPseq. **(e)** Permutation experiments showing statistical significance of the enrichment in repeated DNA sequences of the regions showing specific R-loops enrichment (right) and depletion (left) in dH1-depleted cells respect to control untreated cells. z-scores and permutation p-values of the differences are indicated.

Among the sequences detected to be specifically enriched for R-loops in dH1-depleted cells, around 95% of them accounted for repetitive DNA sequences (**Figure 15 d**), which mainly localized in the U extra chromosome (**Figure 15 c**). In fact, transposons and also some satellite DNAs were principally identified (**Table 3**). Among the pool of transposons detected, there were retrotransposons (LTR and non-LTR types) and DNA transposons. Permutation analyses corroborated the strong enrichment in repetitive DNA elements of the R-loops regions specific of dH1-depleted cells (permutation test p-value <0.0002). The expected frequency of the number of overlaps with repetitive DNA elements, as determined by the regioneR package using the UCSC dm3 RepeatMasker track (March 2017), is presented for 5000 random permutations of the experimentally identified regions and compared with the observed number of overlaps (**Figure 15 e**).

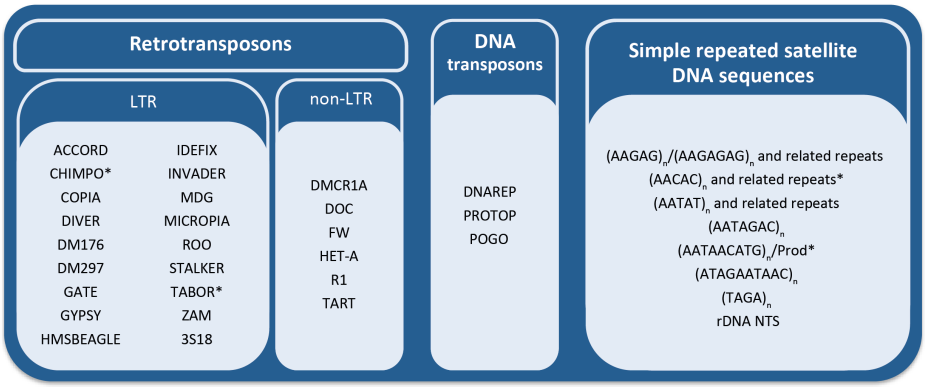


Table 3. Summary of repetitive elements showing specific γH2Av and R-loops enrichment in dH1-depleted cells. *only R-loop enrichment detected.

Further analyses of the genome-wide distribution of RNA:DNA hybrids between control and dH1-depleted cells are presented in Chapter 2.

To confirm DRIP-seq results, two biological replicates of DRIP-qPCR experiments were done. A DRIP followed by qPCR for dH1-depleted cells (siRNA^{dH1}) treated and untreated with RNH prior the immunoprecipitation with the S9.6 antibody was performed. Cells treated with LacZ RNAi (siRNA^{lacZ}) were used as control. Primers amplifying both the open reading frame (ORF) and long terminal repeats (LTR) or untranslated region (UTR) of different heterochromatic sequences were used to check for R-loop enrichment. The percentages of immunoprecipitated material were calculated by the ΔΔCt method and primers for the promoter regions of *mtlh12* and *tubulin* were used as a negative control. In agreement with the results from the DRIP-seq, dH1 depletion significantly increased the proportion of repetitive DNA sequences immunoprecipitated by S9.6 antibodies (both LTR/UTR and ORF) (**Figure 16**). For example, the 3S18, MDG3, GYPSY, ACCORD, INV and HET-A transposons showed a significant enrichment (p-value <0.005) of 3 to 4.5-fold respect to the untreated condition (using the ORF primers). On the contrary, cells treated with LacZ RNAi showed an R-loop enrichment respect to control around 1, suggesting that there are no differences between this condition and the untreated one.

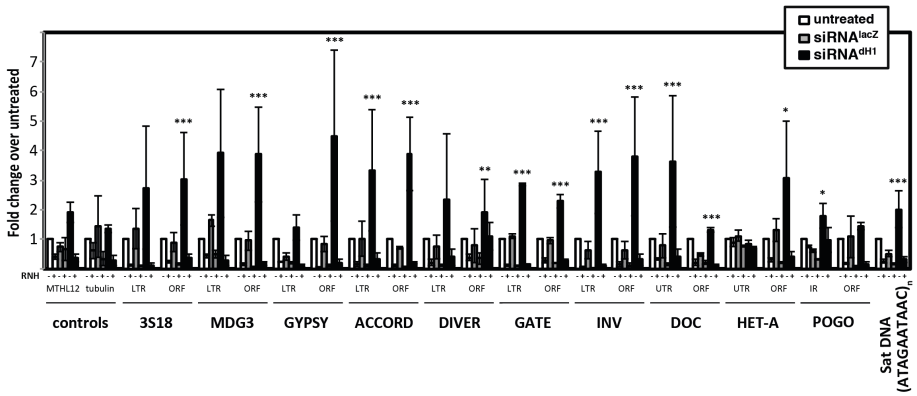


Figure 16. DRIP-qPCR confirmed R-loop accumulation in heterochromatin upon depletion of histone H1. DRIP-qPCR analyses at the indicated regions in siRNA^{dH1} cells, siRNA^{lacZ} and untreated cells. Before immunoprecipitation samples were treated with bacterial RNH (+) or not (-) (N= 2). For each position, the fold-change (FC) respect to the untreated condition at this position is presented. Error bars are s.e.m. The p-values of siRNA^{dH1} with respect to siRNA^{lacZ} are indicated (no asterisk>0.05, *<0.05, **<0.01, ***<0.005; two-tailed Student's t-test).

In this experiment, it could also be observed how the immunoprecipitation of sequences enriched for RNA:DNA hybrids was specific, as treatment with RNH reduced the enrichment values drastically.

1.2. HP1a depletion de-regulates heterochromatin as it happens for dH1 depletion

Whether the accumulation of R-loops detected in heterochromatin of dH1-depleted cells was simply the result of the relief of silencing caused by dH1 depletion or had something to do with a specific function of histone H1 was next addressed. For that, HP1a, which is another essential heterochromatin component also required for silencing, was depleted in S2 cells.

Depletion was done using dsRNA and efficiency was tested by WB (**Figure 14**). After 6 days of treatment, HP1a protein was barely observed for HP1a-depleted cells (siRNA^{HP1a}), suggesting a high degree of depletion.

Total RNA was extracted from HP1a-depleted and untreated S2 cells, and the RNAs were reverse-transcribed into cDNA using random hexamer primers, as the sequences to be analyzed had no polyadenylated tail. Then, the transcription levels of various heterochromatic elements were analyzed by reverse transcription quantitative polymerase chain reaction (RT-qPCR) and the results were compared with the levels found in untreated cells. In order to see the differences respect to dH1 depletion, the results were compared with the up-regulation obtained for dH1-depleted cells [4] (**Figure 17**). Depletion of HP1a induced up-regulation of heterochromatin elements to similar or even higher levels than the ones obtained for dH1-depleted cells.

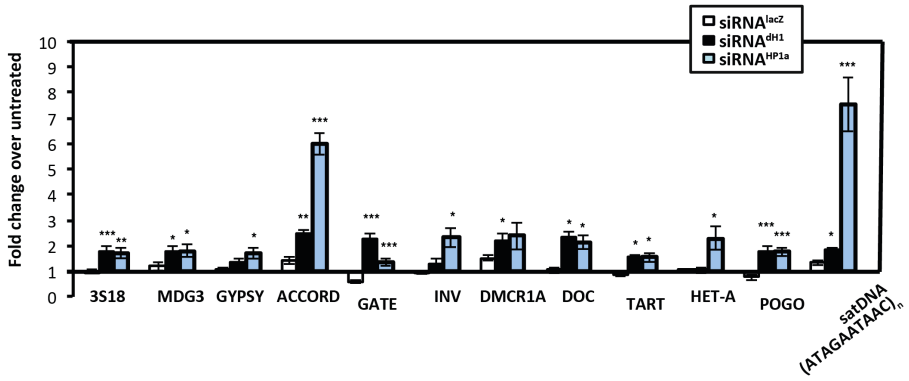


Figure 17. HP1a depletion up-regulates heterochromatin. RT-qPCR analyses of the indicated repetitive elements in siRNA^{HP1a}, siRNA^{LacZ} and siRNA^{dH1} [4]. The fold change expression respect to untreated cells is plotted (N=2). Error bars are s.e.m. The p-values of siRNA^{dH1} and siRNA^{HP1a} respect to siRNA^{LacZ} are indicated (no asterisk>0.05, *<0.05, **<0.01, ***<0.005; two-tailed Student's t-test).

1.3. Depletion of HP1a does neither induce DNA damage nor R-loop accumulation

Once heterochromatin was de-regulated upon depletion of HP1a, we wondered if the transcription of these heterochromatic sequences induced DNA damage as it happened upon depletion of histone H1.

For that, HP1a-depleted and untreated cells were immunostained against γH2Av and the number of foci per cell in both HP1a-depleted and untreated cells was calculated. HP1a-depleted cells did not show any significant increase in DNA damage respect to untreated cells (**Figure 18 a**). These observations were also confirmed by WB (**Figure 18 b**). Both HP1a-depleted and untreated cells showed the same γH2Av levels, which were almost undetectable even at long exposures. Additionally, the levels of γH2Av obtained for HP1a-depleted cells were compared with the ones obtained for dH1-depleted cells [4]. At a difference to dH1-depleted cells, HP1a depletion did not increase γH2Av content.

Next, after having discarded the generation of DNA damage in HP1a-depleted cells as a consequence of heterochromatin up-regulation, it was analyzed whether in this condition R-loops were accumulated in the cells. Immunostaining against S9.6 antibody of S2 cells depleted for HP1a and untreated was performed and analyzed. For the quantification, n>50 cells were used in each condition and the S9.6 reactivity was determined as the

proportion of DAPI area stained with S9.6 antibody (% R-loop/DAPI; **Figure 18 a**). As it happened for γH2Av, HP1a-depleted cells did not show any significant S9.6 increased reactivity. Then, our ratio was compared with the one obtained for depletion of histone H1 [4]. Results showed that whereas depletion of dH1 induced an increase of ~25% of S9.6 area respect to DAPI, depletion of HP1a maintained the same levels of S9.6 staining as the untreated condition, suggesting that depletion of HP1a did not produce accumulation of R-loops.

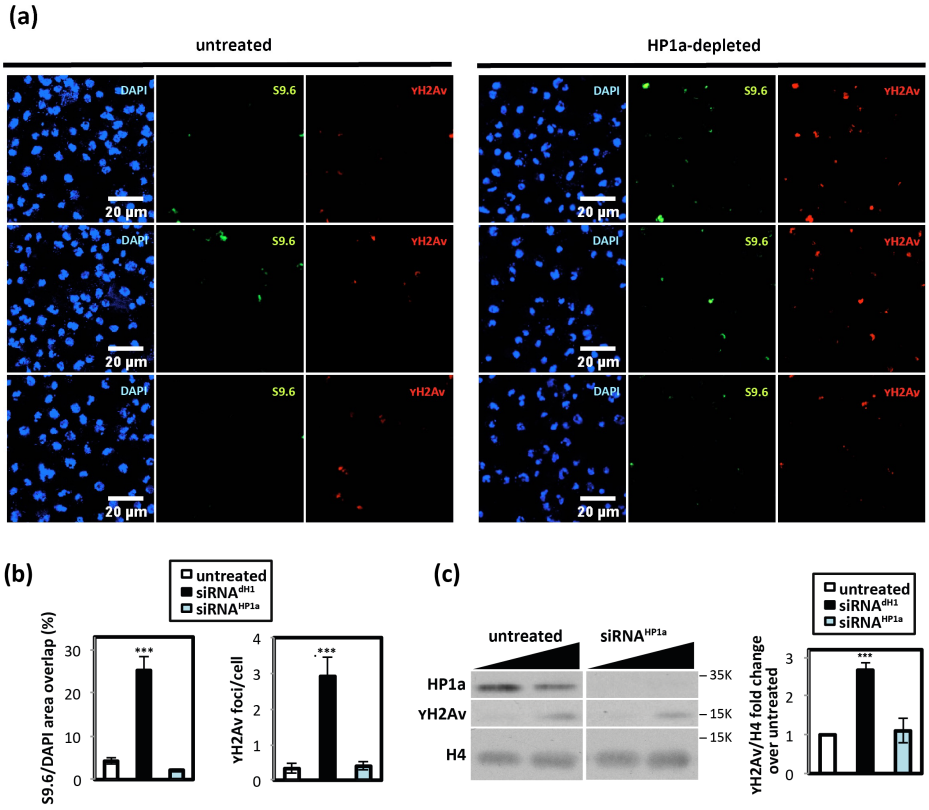


Figure 18. HP1a depletion does not induce DNA damage or R-loop accumulation. **(a)** Immunostainings with γ-H2Av (red) and S9.6 (green) antibodies of siRNA^{HP1a} and control untreated cells. Scale bar corresponds to 20 μm. **(b)** S9.6 (left) and γ-H2Av (right) reactivities are determined as the proportion of DAPI area stained with S9.6 antibodies and the number of foci per cell are presented (n>50 cells for each condition). Data from dH1 depletion is included for comparison [4]. Error bars are s.e.m. The p-values of siRNA^{dH1} and siRNA^{HP1a} respect to untreated are indicated (no asterisk >0.05, ***<0.005; two-tailed Student's t-test). **(c)** WB analyses with αHP1a, αγH2Av and αH4 antibodies of increasing amounts of extracts (lanes 1–2) prepared from siRNA^{HP1a} and untreated cells. The positions corresponding to molecular weight markers are indicated. On the right, quantitative analyses of the results (N=2). The

values obtained for siRNA^{dH1} cells are included for comparison. Error bars are s.e.m. The p-values of siRNA^{dH1} and siRNA^{HP1a} respect to untreated are indicated (no asterisk >0.05, ***<0.005; two-tailed Student's t-test).

1.4. Depletion of dH1 reduces HP1a occupancy at heterochromatin elements

In the light of the results presented above, we sought to analyze how depletion of dH1 or HP1a affected each other's occupancy on heterochromatic sequences. For that, dsRNA against dH1 and HP1a proteins were used to deplete them, separately. Then, 300-500 bp fragmented chromatin was ChIPped followed by qPCR. Two independent experiments were performed for HP1a-, dH1-depleted and untreated cells. We used HP1a antibody to immunoprecipitate all those DNA fragments bound to HP1a. Primers amplifying the ORF of different heterochromatic sequences were used to test HP1a occupancy in heterochromatin upon depletion of HP1a and dH1 proteins. As expected, HP1a was present in the heterochromatin of untreated cells and its abundance was reduced upon depletion of HP1a. Interestingly, upon depletion of dH1, the amount of HP1a-enriched heterochromatic sequences decreased significantly on many of the sequences analyzed, suggesting that dH1 was somehow required for maximal HP1a occupancy at the same heterochromatic elements (Figure 19).

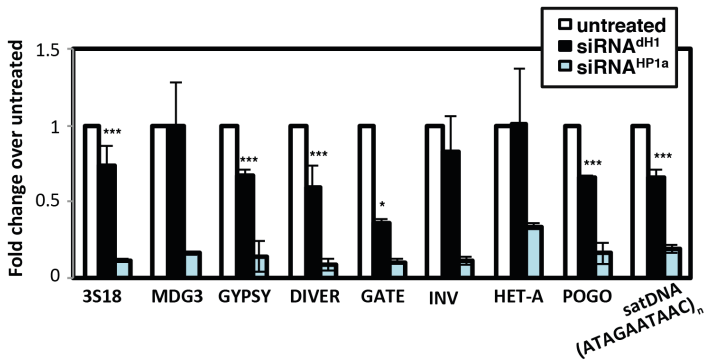


Figure 19. HP1a occupancy in dH1-depleted cells. HP1a ChIP-qPCR analyses of the indicated repetitive elements in siRNA^{HP1a}, siRNA^{dH1} and untreated cells are presented. Primers amplifying the ORF were used. For each repetitive element, the fold change respect to the untreated condition at this repetitive element is presented (N=2). Error bars are s.e.m. The p-values of siRNA^{dH1} respect to untreated are indicated (no asterisk >0.05, ***<0.005; two-tailed Student's t-test).

1.5. Presence of dH1 at heterochromatin is not affected by HP1a depletion

Knowing that HP1a occupancy was affected by the levels of dH1 in heterochromatin, we were interested in evaluating if dH1 occupancy was altered upon depletion of HP1a. So, as done before for HP1a-ChIPs, S2 cells were depleted for dH1 or HP1a separately, ChIPped with dH1 antibody and analysed by qPCR as above. Results showed no significant changes in dH1 binding upon depletion of HP1a (Figure 20). Accumulation of dH1 decreased after treatment with dsRNA against dH1, proving the efficiency of depletion and also the presence of dH1 in those regions in untreated conditions.

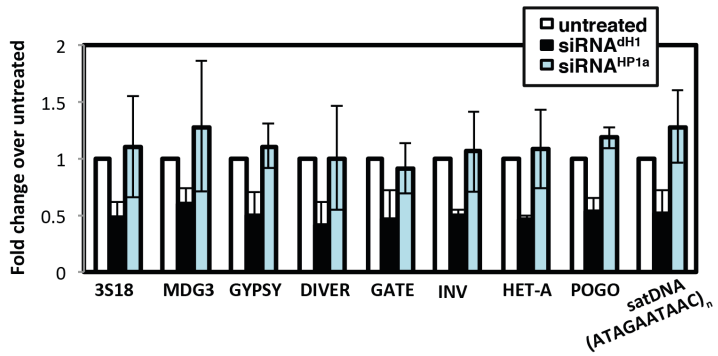


Figure 20. dH1 occupancy in HP1a-depleted cells. dH1 ChIP-qPCR analyses of the indicated repetitive elements in siRNA^{HP1a}, siRNA^{dH1} and untreated cells are presented (N=2). Primers amplifying the ORF were used. For each repetitive element, the fold-change respect to the untreated condition at this repetitive element is presented. Error bars are s.e.m.

1.6. Co-depletion of dH1 and HP1a results in R-loop accumulation in heterochromatin

Then, to assess if the accumulation of R-loops was histone H1-dependent, we decided to co-deplete S2 cells for dH1 and HP1a (siRNA$^{dH1+HP1a}$). For that, we performed two independent depletions. One was done using dsRNA for only HP1a and the other using dsRNA for both dH1 and HP1a. A reduction of 40% for dH1 and more than 60% for HP1a respect to untreated cells was obtained (Figure 14). As it can be observed in the WB, depletion for HP1a was more efficient than the one obtained for dH1. Interestingly, although HP1a occupancy in heterochromatin decreased upon depletion of dH1, total levels of HP1a increased to even higher amounts respect to control. Lu et al.

also obtained increased total cellular HP1 upon depletion of H1 expression in *D. melanogaster* larvae [1].

dsRNA against LacZ was also used to interfere control cells, treating the cells with the maximum amount of dsRNA used in this experiment (the one corresponding to the dsRNA used for depleting dH1 and HP1a, simultaneously). Then, the depletion periods were combined and DRIP was performed for each condition in parallel. The experiment was repeated twice.

As it can be seen in **Figure 21**, untreated cells or cells treated with dsRNA against LacZ and HP1a alone were not enriched for R-loops, as no significant levels of immunoprecipitation using S9.6 were observed. However, when S2 cells were depleted for dH1 in a pool of cells already HP1a-depleted, an increased amount of immunoprecipitated material that was sensitive to RNH treatment was obtained, indicating that dH1 was responsible for the R-loop accumulation in heterochromatic sequences.

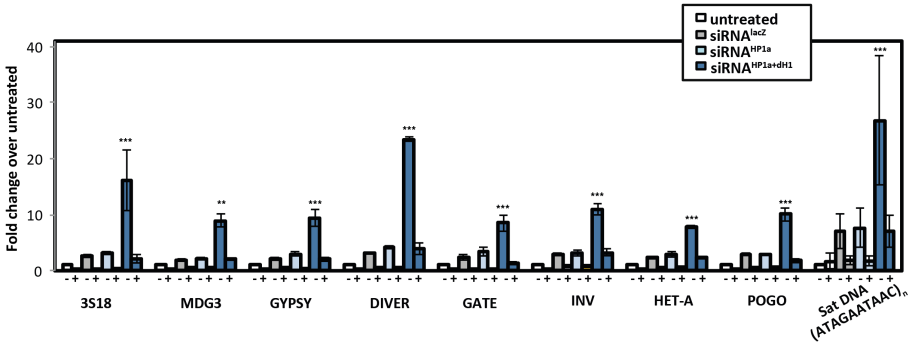


Figure 21. DRIP-qPCR. DRIP-qPCR analyses at the indicated repetitive elements in siRNA^{HP1a}, siRNA$^{HP1a+dH1}$, siRNA^{lacZ} and untreated cells. Before immunoprecipitation samples were treated with bacterial RNH (+) or not (−) (N = 2). For each repetitive element, the fold-change respect to the untreated condition at this repetitive element is presented. Error bars are s.e.m. The p-values of siRNA^{HP1a} and siRNA$^{HP1a+dH1}$ respect to siRNA^{lacZ} are indicated (no asterisk >0.05, **<0.01, ***<0.005; two-tailed Student's t-test).

1.7. The mechanism should involve other co-factors in heterochromatin

At this point, it was clear that absence of dH1 resulted in R-loop accumulation in heterochromatin, suggesting an important role for histone H1 in the regulation of R-loop biology. However, histone H1 proteins lack from enzymatic activity. Consequently, it was hypothesized that the histone H1 specific role in preventing R-loops should involve other proteins with properties that reduce the possibilities of the transcribed RNA to re-anneal to the DNA template or degrade the RNA moiety. For that, histone H1 might be binding some other factors close to heterochromatin that collaborate in the regulation of R-loop dynamics (**Figure 22**).

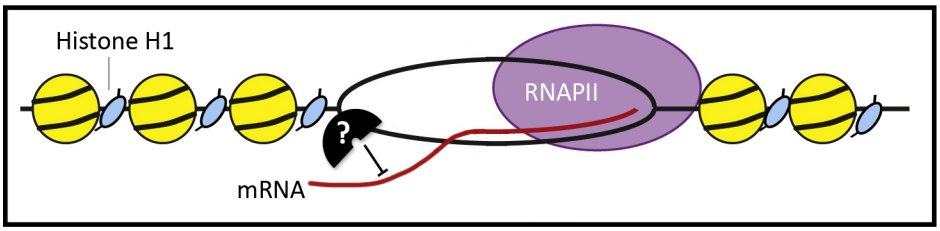


Figure 22. Schematic representation of the proposed mechanism for histone H1 dynamics in R-loop accumulation. During transcription, the transcript needs to be efficiently processed and exported, otherwise it represents a threat to genome stability as it can form an RNA:DNA hybrid. It is proposed that histone H1 recruits a protein with R-loop metabolism activities to prevent their unscheduled formation. The unknown candidate that binds histone H1 and compromises R-loop stability is represented by a question mark.

Hence, with the aim of understanding the exact mechanism by which histone H1 was preventing the accumulation of uncontrolled RNA:DNA hybrids in heterochromatin, we screened for proteins described to interact with histone H1.

1.7.1. Different proteins could contribute to prevent R-loop accumulation in heterochromatin together with dH1

Immunoprecipitation followed by mass spectrometry experiments from nuclear extracts of *Drosophila* cells designed to co-purify partners of dH1 showed that histone H1 interacts with a number of nuclear ribosomal proteins [132]. Also, the *Drosophila Protein Interaction Map* (DPiM) helped us

to identify proteins that pull-down histone H1. The DPiM project indexed protein partners based on the isolation and mass spectrometric analysis of purified protein complexes associated with individually tagged proteins [133]. Although at the moment of the screening the project had not analyzed the histone H1 yet, for all the other proteins already studied, it allowed the identification of proteins that co-immunoprecipitate histone H1.

Then, putting all the data together, 41 possible candidates with activities that could remove R-loops or preserve the RNA from re-annealing to the DNA template and form an RNA:DNA hybrid were selected (**Table 4**). These included several factors associated with RNA and DNA helicase activities, DNA topoisomerases, and factors involved in RNA processing and transport. Finally, the additional requirement for the candidate proteins to be localized at least partially at heterochromatin and that affect heterochromatin silencing were applied, which reduced the list to three candidates: Su(var)2-10 [134], hnRNP36 and hnRNP48 [132] (**Table 4**; in red).

RNA helicases	DNA helicases	DNA topoisomerases	RNA transport	RNA processing
Upf1	rept	scf	thoc5	asf1
Su(var)2-10	Psf3	pho	CG5857	U2af50
l(2)35Df	pont		CG5382	NHP2
Dbp21E2	Mcm5		PCID2	dbe
Hel25E	Mcm7		CG4572	**hnRNP36**
CG9630			CG5053	**hnRNP48**
CG7878			CG8671	CG11266
CG6227				CG16941
CG4901				CG17540
CG3561				CG5454
Dbp73D				CG6227
				CG6995
				CG10909
				CG12301
				CG1542
				CG6712

Table 4. Summary of dH1 interacting factors with activities that could regulate R-loop dynamics. This list highlights proteins that co-IP histone H1 and could remove R-loops (helicases) or prevent their formation (topoisomerases, and RNA transport and processing factors). In red, those related to heterochromatin stability.

1.7.2. Su(var)2-10, hnRNP36 and hnRNP48

Having selected Su(var)2-10, hnRNP36 and hnRNP48 as possible participants in the dynamics of RNA:DNA hybrids together with histone H1, they were next studied in more detail.

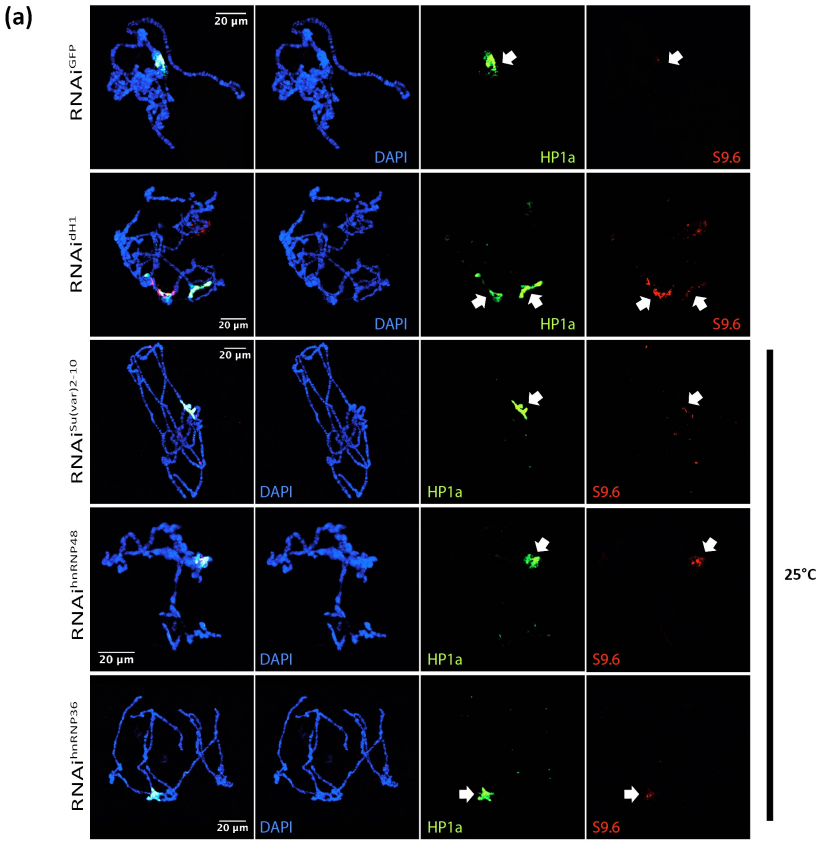
On one hand, Su(var)2-10 (dPIAS) is a dominant suppressor of position-effect variegation and, therefore, affects heterochromatin stability and is required for proper chromosome structure and chromosome inheritance [134]. It is also a SUMO E3 ligase responsible for facilitating the addition of SUMO onto target substrates [135]. While not presenting enzymatic or RNA binding activity by itself, Su(var)2-10 is an RNA helicase binding protein, which could recruit RNA helicases and help histone H1 to remove R-loops and prevent genomic instability by R-loop-mediated DNA damage. On the other hand, hnRNP36 (Hrb87F) and hnRNP48 (Hrb27C) are heterogeneous ribonucleoproteins known to be involved in mRNA biogenesis. They bind to nascent pre-mRNA to form RNP complexes, which become the substrate for subsequent nuclear RNA processing and transport activities [136–139]. Therefore, those hnRNPs could bind to the nascent heterochromatic transcript avoiding the formation of an R-loop and facilitating its export outside the nucleus.

As an initial experiment to analyze the feasibility for these proteins to collaborate with dH1 in preventing R-loop accumulation, the distribution along the chromosomes and the S9.6 reactivity upon depletion of the different candidates was analyzed in polytene chromosomes. Polytene chromosomes of salivary glands are large chromosomes produced by endoreplication, which contain about 1200 DNA strands arranged alongside of each other that facilitate distinctive banding pattern visualization. They have been extensively used to study chromatin structure and proteins bound to chromatin.

Flies carrying constructs designed to knockdown the gene expression of Su(var)2-10, hnRNP36 and hnRNP48 using the RNA interference (RNAi) were obtained. The interference of those genes was regulated through the UAS-GAL4 system, a two-component gene expression system extensively used in *Drosophila*. GAL4 is a yeast transcription factor that binds to a specific UAS (upstream activating sequence, absent in *Drosophila*), allowing the expression of the downstream gene. GAL4 activation is dependent on temperature, producing stronger effects at elevated temperatures [140].

Therefore, placing in one construct a cell-specific promoter followed by the *Gal4* gene, and in a second construct, a hairpin RNA of interest shortly after a UAS, the expression of the double-stranded RNA (dsRNA), which is the precursor of siRNA, is obtained in a controlled manner.

Polytene chromosomes were then prepared from third instar larvae (L3) and were co-stained with HP1a and S9.6 antibodies. HP1a protein was used as a marker of heterochromatin in order to see whether the R-loops accumulated there (**Figure 23 a**). As the specificity of S9.6 signal was already determined in UAS-RNAi^{dH1}; *nub*GAL4, UAS-Dic2 polytene chromosomes [4], we did not find it necessary to overexpress RNH this time. UAS-RNAi^{GFP} flies were used for S9.6 negative control and UAS-RNAi^{dH1} flies were used for S9.6 positive control. Depletion levels of dH1 for UAS-RNAi^{dH1}; *nub*GAL4, UAS-Dic2 are shown in **Figure 23 b**.



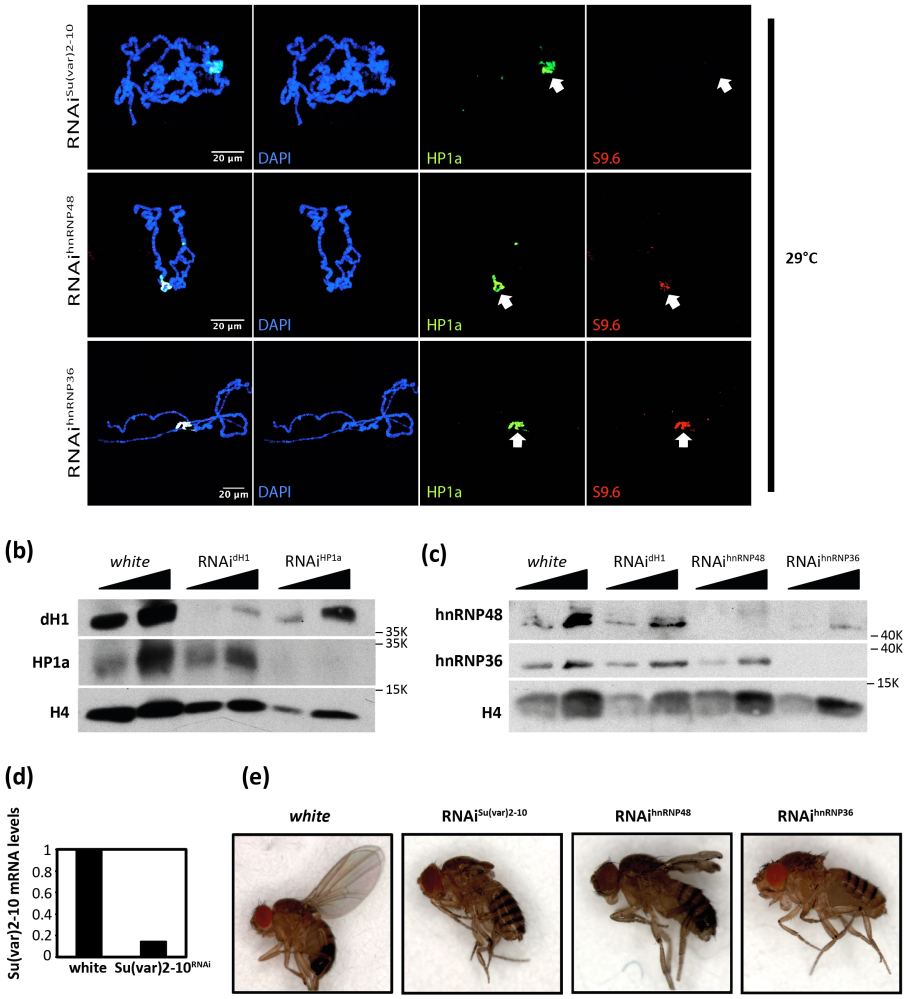


Figure 23. dH1-possible partners in preventing R-loop accumulation. (a) Polytene chromosome immunostainings with S9.6 (red) and HP1a (green) antibodies of siRNA$^{Su(var)2\text{-}10}$; *nub*GAL4, siRNA^{hnRNP48}; *nub*GAL4 and siRNA^{hnRNP36}; *nub*GAL4. Crosses were done at 25°C and 29°C, as indicated. Dapi is in blue and the first image of each row corresponds to the merge of the three channels. HP1a has been used as a marker of heterochromatin (chromocenter of the polytene chromosomes is marked by an arrow). *White* flies and siRNA^{GFP}; *nub*-GAL4 are used for negative control of S9.6 reactivity and siRNA^{dH1}; *nub*-GAL4; Dic2 for positive control. Scale bar corresponds to 20 µm. **(b)** dH1 and HP1a depletion levels in siRNA^{dH1}; *nub*-GAL4; Dic2 and siRNA^{HP1a}; *nub*-GAL4; Dic2 (25°C) are tested in WB with specific antibodies. **(c)** Depletion levels of hnRNP36 and hnRNP48 in siRNA^{hnRNP48}; *nub*GAL4 and siRNA^{hnRNP36}; *nub*GAL4 (29°C) and siRNA^{dH1}; *nub*-GAL4; Dic2 (25°C) are tested in WB with specific antibodies. **(d)** Su(var)2-10 mRNA levels upon S2 cells treated with dsRNA against Su(var)2-10. **(e)** Adult fly wing of siRNA$^{Su(var)2\text{-}10}$; *nub*GAL4, siRNA^{hnRNP48}; *nub*GAL4 and siRNA^{hnRNP36}; *nub*GAL4 (29°C). *White* fly (^{w1118}) wing is used as a control.

a) Su(var)2-10:

*nub*GAL4 virgin females were crossed with homozygous UAS-RNAi$^{Su(var)2\text{-}10}$ males at 25°C, but no S9.6 reactivity or defects on the polytene chromosomes were observed . Then, in order to improve the depletion conditions, crosses were done at 29°C but they did not give any S9.6 reactivity either (**Figure 23 a**). Due to absence of a good antibody to detect Su(var)2-10 protein indirect detection through mRNA levels was performed. Su(var)2-10 levels were significantly reduced at 29°C (reduction of 85%) (**Figure 23 d**; right) and the depletion produced a strong wing phenotype on the adult flies (**Figure 23 e**).

b) hnRNP36 and hnRNP48:

Similar crossings were performed for UAS-RNAi^{hnRNP36} and UAS-RNAi^{hnRNP48}. Although protein down-regulation was already obtained from these crosses at 25°C, crosses at 29°C were done to improve depletion levels, as previously done for Su(var)2-10. Interestingly, depletion of either hnRNP36 or hnRNP48 induced a strong accumulation of S9.6 signal at the chromocenter of polytene chromosomes, especially when depleting the hnRNPs at 29°C (**Figure 23 a**). In these depletion conditions (29°C), protein levels of the hnRNP36 and hnRNP48 were importantly reduced (**Figure 23 c**), and the adult flies presented an important wing phenotype (**Figure 23 e**). Of note, histone H1 levels were maintained (data not shown) and integrity of the chromocenter did not look like affected as indicated by HP1a staining (**Figure 23 a**).

1.8. hnRNP36 and hnRNP48 collaborate with dH1 in the R-loop resolution

Henceforth, after having observed that depletion of hnRNP36 and hnRNP48 strongly increased S9.6 reactivity in heterochromatin, we worked on the hypothesis that those proteins were necessary for histone H1 to prevent R-loops in heterochromatin. Thus, hnRNP36 and hnRNP48 were analyzed in detail.

1.8.1. Profiling of the hnRNP36 and hnRP48 in *D. melanogaster*

The heterogeneous ribonucleoprotein 36 and 48 are described to bind different positions along the chromosomes of *D. melanogaster* [139]. Then, in order to characterize the distribution of these hnRNPs, polytene chromosomes were prepared from *D. melanogaster white* flies (w^{1118}), which will be used as control for all the following polytene chromosome experiments.

Polytene chromosome spreads of *white* L3 larvae and L3 larvae from UAS-RNAi^{hnRNP36} and UAS-RNAi^{hnRNP48} males crossed with *nub*GAL4 virgin females at 29°C were co-immunostained simultaneously with HP1a, hnRNP36 and hnRNP48 antibodies. The staining for both hnRNPs in *white* polytene chromosomes presented a general distribution, but preferentially accumulated at interband regions along the chromosome arms. A strong co-localization with HP1a was also observed, suggesting that the two hnRNPs are localized at the chromocenter (**Figure 24**). More detailed images of hnRNP36/48 and HP1a colocalization are presented in Chapter 1.8.2 (**Figure 26**). Polytene chromosomes depleted for hnRNP36 and immunostained against hnRNP36 antibody did not show any reactivity, validating the depletion of the protein and proving the specificity of the antibody. Also, immunostainings of polytene chromosome spreads against hnRNP48 antibody were also performed in salivary glands depleted for hnRNP36. Remarkably, the distribution of hnRNP48 at the chromocenter was not apparently altered, suggesting that the absence of hnRNP36 did not affect the binding of hnRNP48 in heterochromatin (**Figure 24**; second row). Similar results were observed after performing the same polytene chromosome immunostainings for hnRNP48-depleted flies (**Figure 24**; third row). Likewise, global levels of hnRNP36 were not affected upon depletion of hnRNP48, and vice versa, as it can be observed by the absence of appreciable changes in expression of the other hnRNP in the WB of each particular hnRNP knockdonwn (**Figure 23 c**).

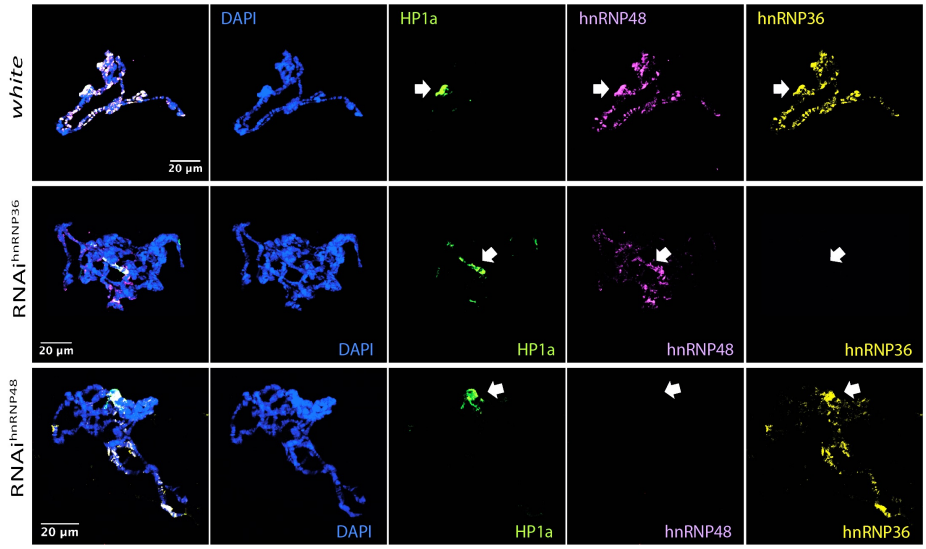


Figure 24. hnRNP48 and hnRNP36 distribution in salivary gland polytene chromosomes. Salivary gland polytene chromosome immunostainings with hnRNP48 (magenta) and hnRNP36 (yellow) antibodies of *white*, siRNA^{hnRNP48};*nub*GAL4 and siRNA^{hnRNP36};*nub*GAL4 flies. Crosses were done at 29°C. Dapi is in blue and HP1a (green) is used as a marker of heterochromatin (chromocenter of the polytene chromosomes is marked by an arrow). The first image of each row represents the merge of the four channels. Scale bar corresponds to 20 µm.

Then, immunoprecipitation experiments against hnRNP36 and hnRNP48 in S2 cells were performed and whether dH1 was pulled-down or not together with the hnRNPs was analyzed by WB. A highly purified dH1 protein was used as a marker in order to identify dH1 band from background, and antibodies against HA and GFP were used to validate the specificity of the immunoprecipitation.

Immunoprecipitation against hnRNP36 immunoprecipitated hnRNP36, but not hnRNP48 (**Figure 25 a**; left). On the same direction, immunoprecipitation against hnRNP48 resulted in two bands corresponding to hnRNP48, and hnRNP36 was not detected (**Figure 25 a**; right). Thus, hnRNP36 and hnRNP48 did not interact and, in agreement with the immunostainings previously shown, they two work independently of each other. In addition, hnRNP36 clearly co-immunoprecipitated with dH1 (**Figure**

25 b; left). This observation confirmed previously published results indicative of a direct interaction between dH1 and hnRNP36 [136].

Co-immunoprecipitation of dH1 with hnRNP48 was not as clear as for hnRNP36. Nevertheless, a slight band that corresponded in size with dH1 could be observed in some of the immunoprecipitations performed (**Figure 25 b**; right).

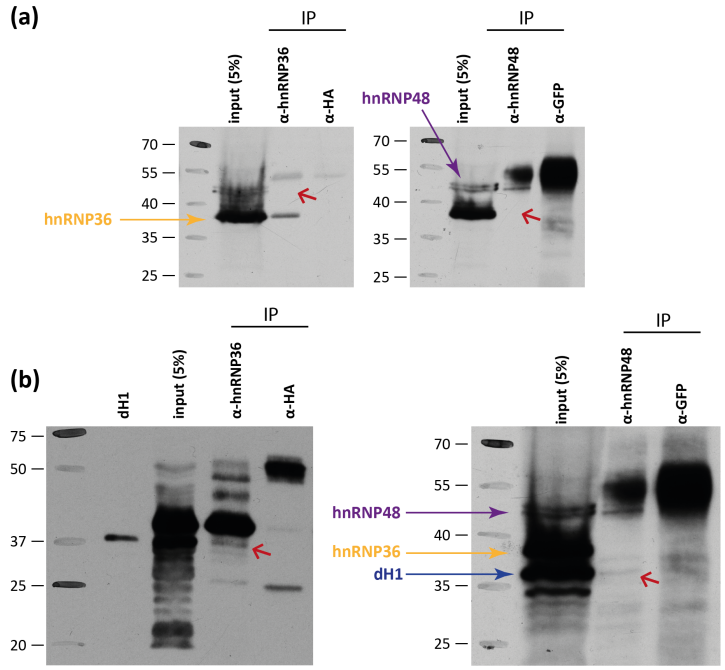


Figure 25. hnRNPs interact with dH1. Immunoprecipitation (IP) assay of S2 cells using antibodies against hnRNP36 and hnRNP48. **(a)** IP against hnRNP36 immunoprecipitated hnRNP36, but not hnRNP48 (red arrow in left gel). Similarly, IP against hnRNP48 immunoprecipitated hnRNP48, but not hnRNP36 (red arrow in right gel). GFP and HA antibodies have been used as a control for specificity of the IP. **(b)** Band corresponding to dH1 after IP with α-hnRNPs is indicated with a red arrow.

1.8.2. dH1 interacts with hnRNP36 and hnRNP48 in heterochromatin

At this stage and consistent with what was already published, it was concluded that hnRNP36 and hnRNP48 proteins, which have mRNA quality control characteristics needed for preventing R-loop accumulation, were found in heterochromatin. Besides, the interaction of hnRNP36 with histone H1, and probably that of hnRNP48 with dH1, was also proved. Finally, and

what was more decisive, depletion of either dH1 or any of the two hnRNPs induced S9.6 reactivity specifically in heterochromatin. Hence, we wanted to answer whether dH1-hnRNP specific interaction was taking place on heterochromatin and whether it was responsible for preventing R-loop accumulation in heterochromatin.

To answer these questions, S2 cells were depleted for hnRNP36 (siRNA^{hnRNP36}) and 48 (siRNA^{hnRNP48}) using specific dsRNA and ChIP-qPCR experiments against dH1 were performed. Protein expression decreased to 45% and 23% respect to untreated for hnRNP36 and hnRNP48, respectively (**Figure 26 a**), and S2 cells depleted for LacZ and untreated were used as controls. For the qPCR, primers amplifying the ORF of different heterochromatic sequences were used. Results showed that upon depletion of any of the hnRNPs in S2 cells, no significant changes for histone H1 accumulation were obtained in these heterochromatic sequences (**Figure 26 b**).

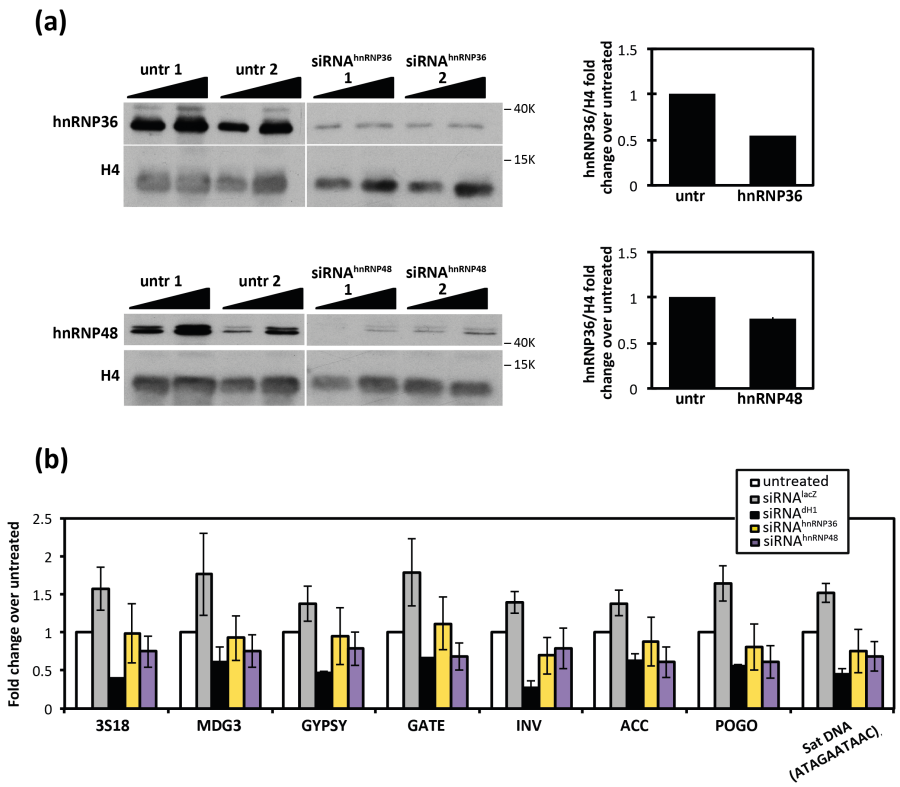


Figure 26. Histone H1 occupancy in heterochromatic sequences is not affected upon depletion of hnRNP36 and hnRNP48. (a) WB analyses showing depletion of

hnRNP36 (upper blot) and hnRNP48 (down blot) in S2 cells. For each sample there are two biological replicates and H4 is included for normalization. The positions corresponding to molecular weight markers are indicated. On the right, quantitative analyses of the results (N=2) are shown. Error bars are s.e.m. **(b)** ChIP-qPCR against dH1 in siRNA^{hnRNP36} (yellow) and siRNA^{hnRNP48} (magenta) S2 cells. Upon depletion of hnRNP36 or hnRNP48, dH1 heterochromatic occupancy diminished compared to untreated and siRNA^{LacZ}. dH1-depleted cells served as control for depletion and dH1-ChIP affinity. ORF primers were used to detect immunoprecipitation of heterochromatic sequences. Error bars are s.e.m. The p-values of hnRNP36-depleted and hnRNP48-depleted cells with respect to untreated cells are >0.05.

An attempt to assess whether hnRNP-binding to heterochromatin was dependent on histone H1 or not was done. Unfortunately, ChIP experiments against hnRNP36 and hnRNP48 in S2 cells depleted for histone H1 did not work at all. Then, in order to overcome those drawbacks and answer this question from other experiments, control and histone H1-depleted polytene chromosomes (UAS-RNAi^{dH1};*nub*GAL4;UAS-Dic2) were immunostained against both hnRNPs and HP1a, simultaneously. Salivary gland polytene chromosomes from RNAi^{dH1} flies had a severely disturbed pattern, as it was previously described [3]. The normal regular band-interband structure was faded out and the chromosome arms were thinner than normal. In addition, DAPI staining of squashed RNAi^{dH1} polytene chromosomes showed the loss of a well-defined chromocenter, where instead of a single focus of pericentric HP1a staining observed for wild-type polytene chromosomes, H1-depleted polytene chromosomes exhibited two or more separated HP1a foci (**Figure 23 a** (second row)**; Figure 27**). That complicated the interpretation of the results. During the preparation of polytene chromosomes, alterations for RNAi^{hnRNP36} chromosomes were also detected (band-interband pattern was difficult to observe in some of the chromosomes and most of them were very fragile).

From the pictures recorded, the intensity of both hnRNPs in HP1a-positive area was quantified. Results showed different scenarios: while some polytene chromosomes missed one or another hnRNPs in a different extent, other chromosomes missed both of them, and/or showed changes in hnRNP intensity. The overall calculation suggested that upon depletion of histone H1, hnRNP36 and hnRNP48 co-localization with HP1a significantly decreased compared to that observed in wild type preparations ($P<0.0001$; Wilcoxon test) (**Figure 27**).

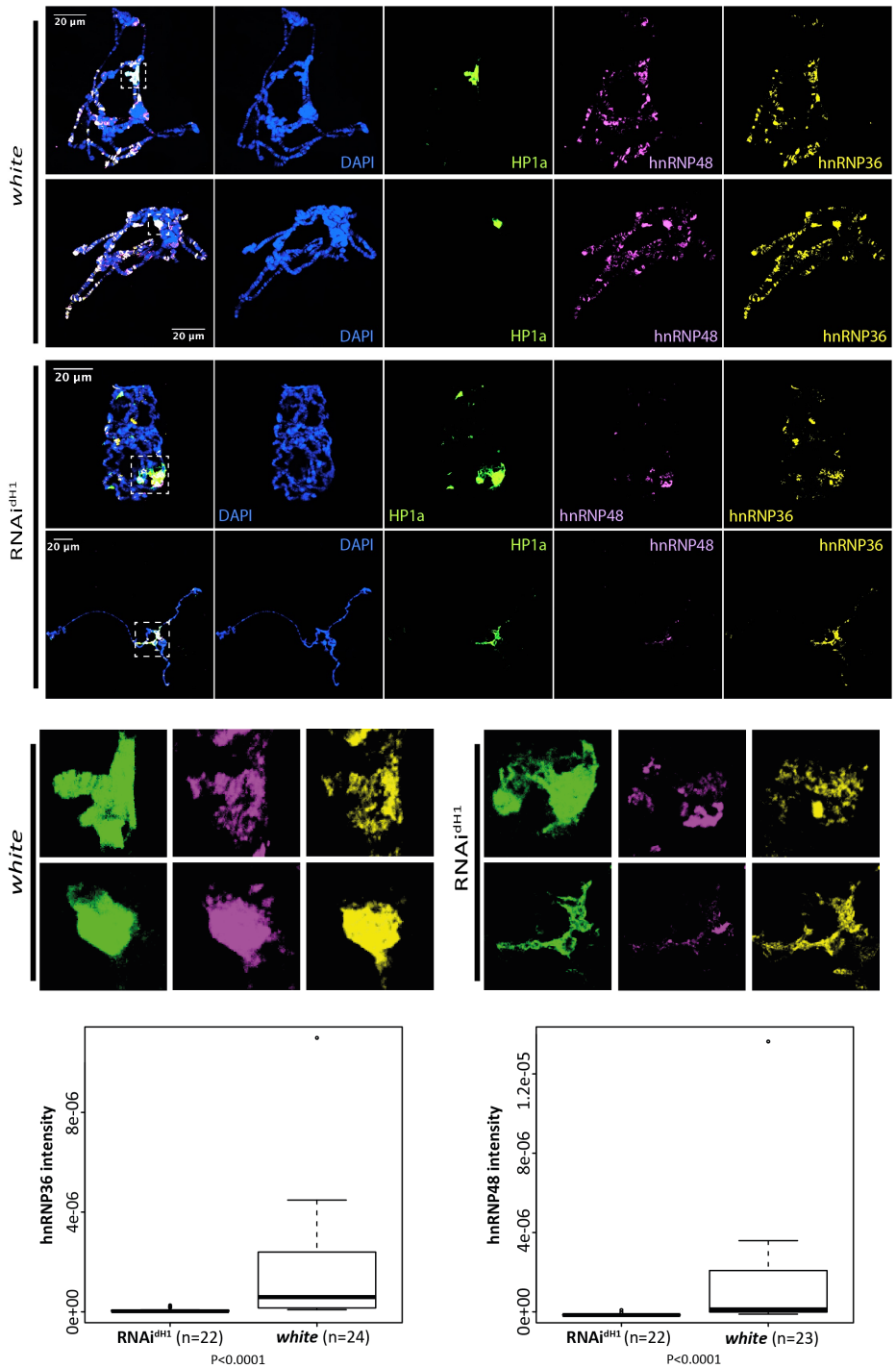
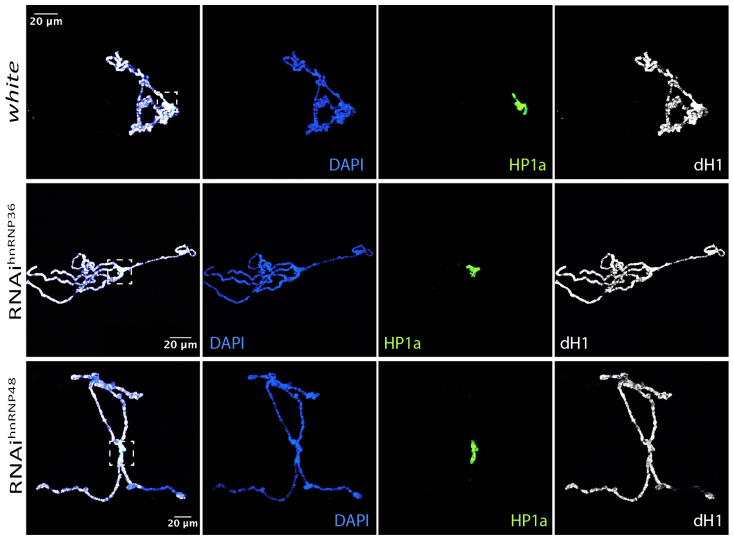


Figure 27. hnRNP36 and hnRNP48 content in heterochromatin decreases upon depletion of dH1. Salivary gland polytene chromosome immunostainings with hnRNP36 (yellow), hnRNP48 (magenta) and HP1a (green) antibodies of *white* and

siRNA^{dH1};*nub*GAL4,Dic2 flies. Crosses performed at 25°C. The amplified chromocenter region is shown for each image. Quantification of the intensity of the hnRNPs in HP1a positive regions normalized to total HP1a area is presented and the p-values of dH1-depleted flies with respect to *white* flies are indicated for each quantification (P<0.0001).

To confirm that immunostainings were consistent enough to sustain our hypothesis, we also immunostained UAS-RNAi^{hnRNP36};*nub*GAL4, UAS-RNAi^{hnRNP48};*nub*GAL4 L3 and *white* polytene chromosomes for dH1. In the same way as described above, dH1 intensity was quantified in these conditions. Results showed a slight decrease in dH1 intensity at the chromocenter upon depletion of hnRNP36 and hnRNP48, respectively (**Figure 28**). However, this decrease in dH1 intensity in HP1a-positive area upon depletion of the correspondent hnRNP was not significant (P=0.451; Kruskal-Wallis test), which goes in agreement with the results obtained by ChIP-qPCR for dH1 (**Figure 26**).



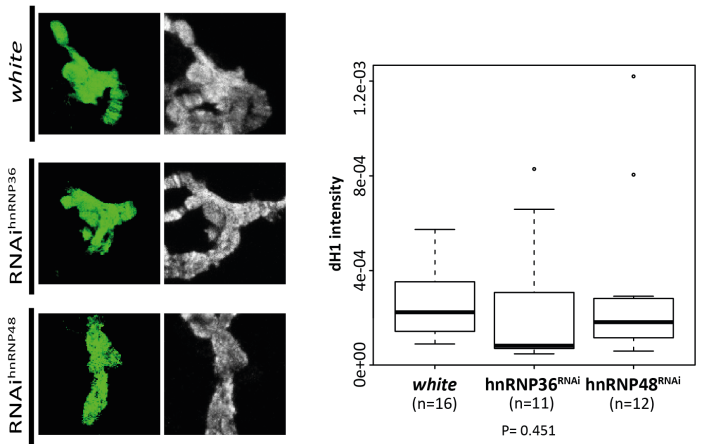


Figure 28. dH1 content in heterochromatin in maintained upon depletion of hnRNP36 and hnRNP48. Salivary gland polytene chromosome immunostainings with dH1 (grey) and HP1a (green) antibodies of siRNA^{hnRNP48};*nub*GAL4 and siRNA^{hnRNP36};*nub*GAL4 flies. Dapi is in blue and the first image of each row corresponds to the merge of the three channels. Crosses are done at 29°C. Scale bar corresponds to 20 µm and the amplified chromocenter region is shown for each image. Quantification of the intensity of the dH1 in HP1a positive regions normalized to total HP1a area is presented and the p-value of hnRNP36-depleted and hnRNP48-depleted flies with respect to *white* flies is indicated (P=0.451).

This experiment also confirmed that the hnRNPs and dH1 proteins colocalize in heterochromatin, as they were all found bound at the chromocenter (**Figure 27**). Due to antibody-based limitations, only co-immunostaining of dH1 and hnRNP36 antibodies could be done in *white* L3. Results confirmed colocalization of both proteins in the chromocenter. However, along the chromosomal arms, they do not colocalize. Whereas histone H1 is found at bands, hnRNP36 is localized at interbands and regions where histone H1 signal is low (**Figure 29**).

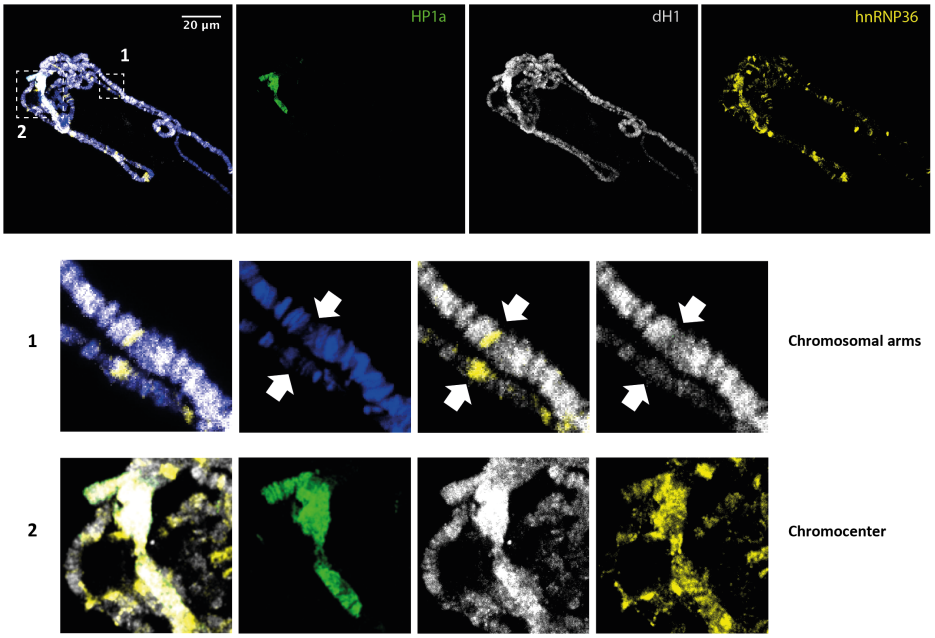


Figure 29. hnRNP36 and dH1 colocalizes in heterochromatin. Salivary gland polytene chromosome immunostainings with dH1 (grey), hnRNP36 (yellow) and HP1a (green; to mark the chromocenter) of *white* flies. Dapi is in blue and the first image of each row corresponds to the merge of the three channels. Scale bar corresponds to 20 µm. Amplified sections of the image are presented: (1) shows the band localization of histone H1 and the interband localization of hnRNP36 along the chromosomal arms. (2) shows the colocalization of hnRNP36 in the chromocenter.

2. THE EFFECT OF dH1 DEPLETION ON THE RNA:DNA HYBRIDS IN *Drosophila melanogaster* EUCHROMATIN

R-loops participate as intermediaries in many biological functions. However, pathological situations may affect their dynamics, inducing their aberrant accumulation and, consequently, becoming a threat to the genome stability. In chapter 1, the importance of histone H1 in R-loop metabolism in heterochromatin was assessed. Thus, to continue our studies, we aimed **to understand how R-loops were normally distributed along the *D. melanogaster* genome and how this distribution was affected upon depletion of histone H1**. For that, the DRIP-seq data was further analyzed.

In the previous DRIP-seq analysis (Chapter 1), as it was used to localize R-loops at heterochromatin upon depletion of histone H1, a multi-hit analysis was used. Heterochromatin is enriched for repetitive sequences. For that,

reads were aligned to all possible sites along the genome. Here, as the aim was to map the normal R-loop distribution genome-wide, the DRIP-seq data was re-processed. This time, only reads aligning to a unique site (single-hit analysis) were considered, whereas those reads aligning to more than one site were eliminated. Therefore, all reads that could be assigned precisely to a single location throughout the genome were mapped.

For each condition (untreated and treated with dH1 or LacZ dsRNA), two biological replicates were performed. As it was done for the previous DRIP-seq analysis, only the common peaks between the two replicates were selected. Then, all peaks presenting sensitivity to RNH treatment were considered R-loops (FC < - 1) and annotated with overlapping and/or closest genes.

2.1. General distribution of RNA:DNA hybrids across the genome

Results revealed a total of 6,039 S9.6 peaks in S2 cells, which occupied 13,931,681 bp and had an average width of 500 bp. Considering an extension of 175 x 10^6 bp of *D. melanogaster* genome, R-loop content corresponded to 7.96%.

As shown in **Figure 30 a**, the vast majority of R-loop signal in control conditions was mapped onto euchromatin regions. They largely localized in both arms of chromosomes II (2L and 2R), III (3L and 3R) and X. Then, the analysis of the distribution of overall S9.6 peaks depending on the feature they overlapped with showed that the largest fraction (>75% of mapped S9.6 peaks) corresponded to genic regions (exons, introns, transcription start and termination sites (TSS, TXE, respectively)). In particular, and as shown in **Figure 30 b/c**, most of the genic R-loops were found within introns or overlapping both intron and exon regions (43% of R-loops mapped over the intersection of exons and introns, whereas 30% of R-loops localized at introns not matching with exons), and there was a considerable enrichment of R-loops in the first intron (11% of R-loops). Interestingly, the R-loop abundance was also remarkably notable in promoters (21%), and only 7% of R-loops localized at exons not coinciding with introns. Intergenic peaks (<25%) corresponded to distal intergenic region (+3 kb from the center of the S9.6 peak to the gene; 19 %).

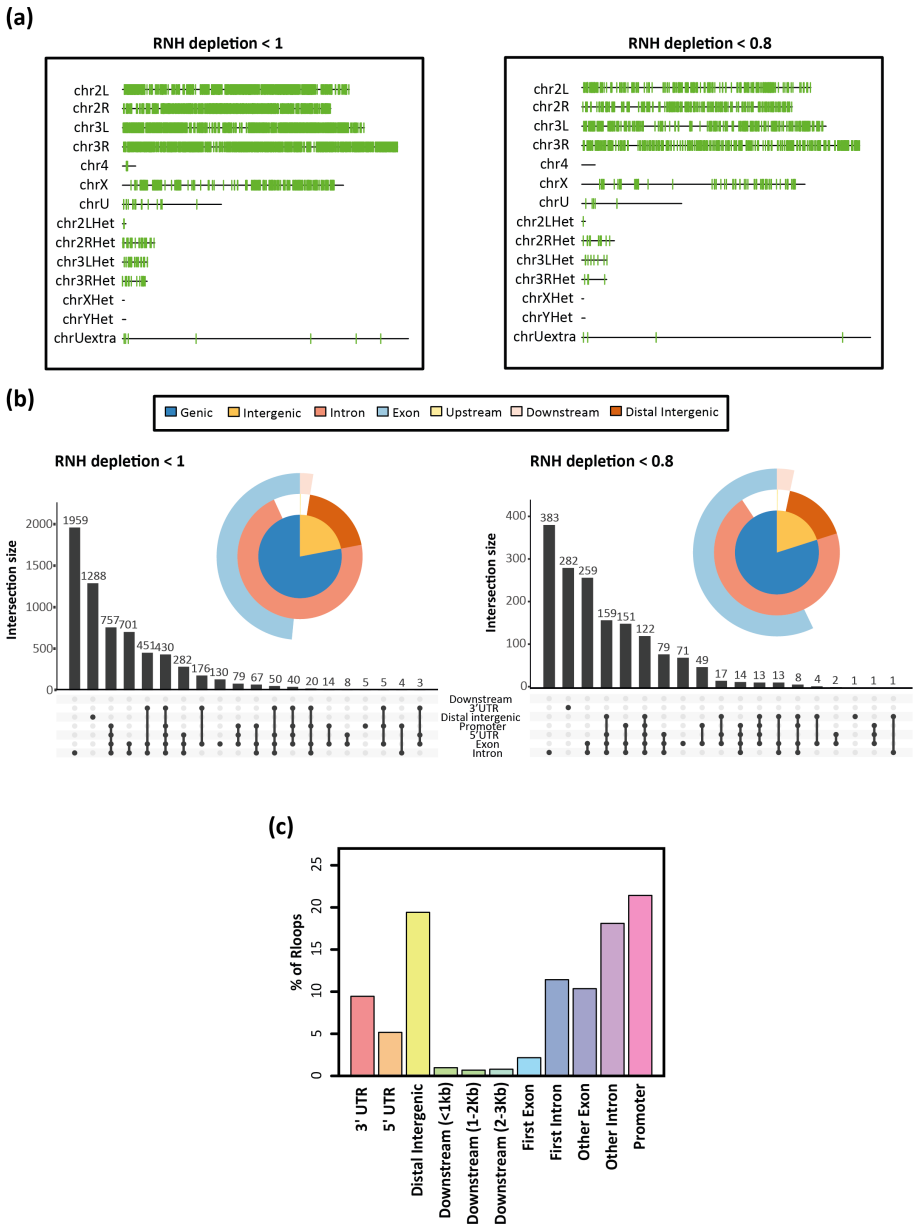


Figure 30. R-loops are distributed along the whole *D. melanogaster* genome. (a) Chromosomal distribution of S9.6 peaks detected in S2 cells for FC <1 (left) and FC <0.8 (right). 2L and 2R, and 3L and 3R correspond to chromosome 2 and 3 left and right arms respect to the position of the centromere, respectively. Chromosome 4 and X are oriented with the centromere to the right. 2LHet, 2RHet, 3LHet, 3RHet and YHet correspond to unassembled highly repetitive chromosome regions. U and U Extra correspond to an unordered and not oriented assembly of unplaced sequence scaffolds. **(b)** S9.6 peaks are classified according to the genomic feature they belong to and are represented in a pie plot. Peaks corresponding to more than one feature

are also presented as a combination of them and the different contributions can be observed in the bar plot. The working FC (FC <1; left) and a higher FC (FC <0.8; right) are presented. **(c)** Bar plot showing the percentage of R-loops along the genome, classified in more detailed genomic features.

The S9.6 peak profile was also analyzed using higher thresholds of RNH sensitivity. Thus, after applying a more stringent FC, the resulting number of R-loops decreased (**Figure 30 a**; right (FC <0.8)). However, the distribution and proportion of peaks for each feature was similar (**Figure 30 b**; right (FC <0.8)). Consequently, to avoid losing of relevant information, the rest of the analyses were performed using a FC of <1.

The classification between genic and intergenic R-loops depended on the annotation of the R package used for each analysis. However, in general, genic consisted on any R-loop overlapping with a gene, distal intergenic consisted on any R-loop localized beyond 3 kb of any gene and the other regions were divided between promoter, 5'-UTR, 3'-UTR and downstream (< 1Kb, 1-2Kb and 2-3kb. Besides, first exon and first intron were considered independently from the others in some figures. Following this division, genic and intergenic R-loops were further analyzed in more detail.

2.1.1. Genic RNA:DNA hybrids

A total of 2,036 genes contained R-loops (aprox. 17,000 genes in *D. melanogaster*). The distribution of R-loops along the genic regions of genes longer than 1kb can be observed in **Figure 31 a**. Interestingly, R-loops mapped within the gene body were found to accumulate near the TSS, whereas the TXE did not show this enrichment. In addition, R-loops accumulated all along the CDS.

Then, trying to answer whether active genes were major hotspots for R-loop formation in the genome, the link between R-loops and transcription was analyzed. Public available expression profiling array data from S2 cells was used for these analyses (GSE 49103). First, genes were classified according to different expression quartiles: [1.81, 3.91], (3.91, 7.6], (7.6, 10.3], (10.3, 13.7]; excluding the non-expressed genes. Thus, as it can be observed, the proportion of genes containing R-loops increased only slightly in those genes showing higher levels of expression (**Figure 31 b**). In fact, all the expression quartiles for S2 cells accumulated a similar proportion of R-

loops, suggesting that the number of genes containing R-loops was independent of their level of expression. The same appreciation was done in **Figure 31 c**, where the different intensities of gene expression accumulated similar levels of R-loops. The bimodal distribution observed on the violin plot is inherent of the technique, as it can also be seen for the whole set of *D. melanogaster* expressed genes (12,607; intensity values >$\log_2$ 2 RMA).

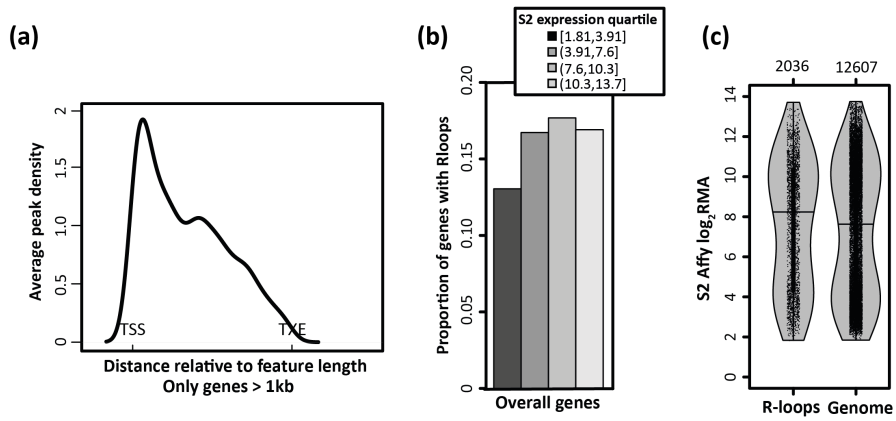


Figure 31. Genes containing R-loops. (a) Location of R-loops along the gene relative to closest/overlapping feature. Only genes > 1 kb are contemplated. **(b)** Bar plot showing the percentage of genes with R-loops. The proportion is divided depending on quartiles of S2 expression. **(c)** The violin plot on the left shows the genes containing R-loops, as distributed depending on their expression level. The violin plot on the right shows the distribution of all *D. melanogaster* genes depending on their level of expression.

Then, knowing that the percentage of the genes accumulating R-loops did not depend on the expression level of those genes, the R-loop content on a gene depending on its expression level was analyzed. As shown in **Figure 32**, no correlation was observed between S9.6-peak intensity (RPKM) and gene expression levels, suggesting that genes containing more R-loops did not necessarily correspond to the higher expressed genes, and vice versa. The absence of correlation was maintained in conditions of validated (FC < 1) and non-validated R-loops (all S9.6 signal), reconfirming that R-loop accumulation did not depend on the expression levels of the genes.

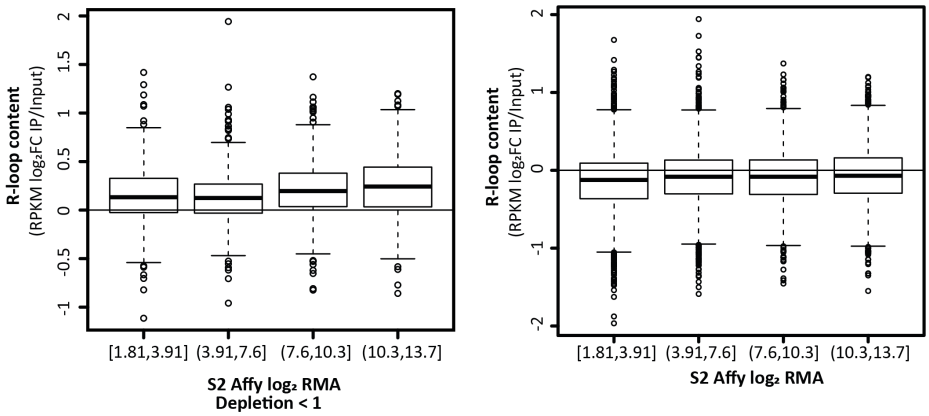


Figure 32. R-loop content in genes depending on their expression levels. The correlation between gene expression (X-Axis) and S9.6 intensity (Y-Axis) is presented for genes classified depending on their expression quartile. On the left, only those genes containing verified R-loops (FC < 1). On the right, all genes presenting S9.6 peaks are plotted.

2.1.2. Intergenic RNA:DNA hybrids

Although R-loops are considered a byproduct of transcription and it was expected that they accumulated along the genic regions of the genome, 25% of them accumulated mainly at distal intergenic regions (**Figure 30 b**). So, to answer whether these distal intergenic RNA:DNA hybrids were an artifact or not, it was first analyzed if these regions contained RNA, as the presence of RNA is needed to form the hybrid.

The result of crossing our DRIP-seq data with high-resolution transcription start site mapping (TSS-seq) [141], suggested that more than 70% of the distal intergenic peaks overlapped with transcription initiation events, which confirmed the presence of RNA in those regions. Once having assessed that distal intergenic peaks were RNA:DNA hybrids, it was consider the possibility that those regions were enhancers and the RNA:DNA hybrid came from the retention of the short ncRNA molecules transcribed from the DNA sequence of these enhancer regions. Thus, from the combination of self-transcribing active regulatory region sequencing (STARR-seq) data [142] with our DRIP-seq, the enhancer contribution of distal intergenic R-loops was evaluated. STARR-seq consists on a genome-wide quantitative enhancer map. Results from the

combination of the STARR-seq and DRIP-seq data showed almost no overlapping. So distal intergenic R-loops could not be attributed to enhancers.

Besides, public available run-on sequencing (GRO-seq; GSE 23543) and RNAPII ChIP-seq (GSE 23542) datasets were briefly studied. To the plain eye, sites of R-loops did not show high levels of transcript accumulation; rather they showed basal transcription levels. However, analysis of the RNAPII phosphorylated on Serine 5 (RNAPIISer5) ChIP-seq data (GSM796331) showed a strong colocalization of R-loops and RNAPII at distal intergenic sites (**Figure 33)**, reconfirming the existence of transcription in those sites.

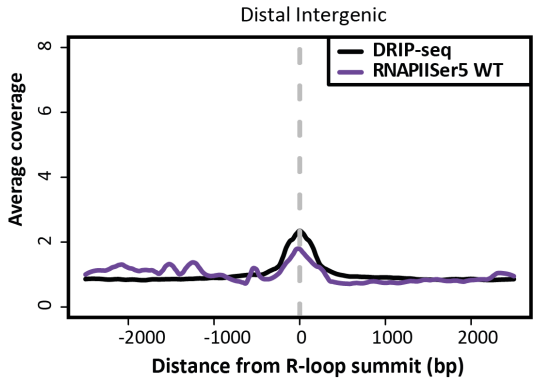


Figure 33. Intergenic R-loops contain RNA. Combination of DRIP-seq (black) and RNAPIISer5-seq (purple) average coverage for the R-loops found at distal intergenic regions. The 0 corresponds to the summit of the R-loop peak.

Then, in order to analyze if those intergenic peaks corresponded to repetitive sequences, the sequence composition was screened with the RepeatMasker program, which looked for interspersed repeats and low complexity DNA sequences. Results indicated that only 7.6% of their base pairs corresponded to repetitive sequences; in particular, most of them corresponded to RNA transposons (6.43%) (**Table 5**).

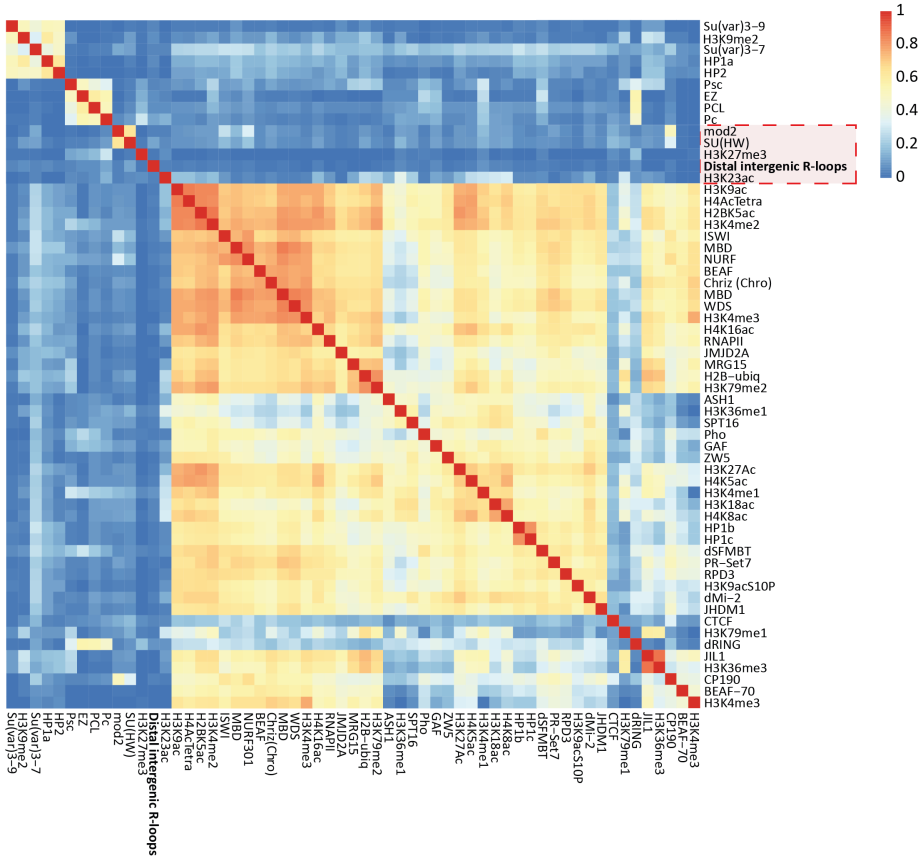
Repetitive DNA elements		number of elements*	length occupied	sequence percentage
RNA transposons:		488	158905 bp	6.43 %
SINEs:		0	0 bp	0.00 %
Penelope-like elements:		0	0 bp	0.00 %
LINEs:		142	35451 bp	1.43 %
	CRE/SLACS	0	0 bp	0.00 %
	L2/CR1/Rex	24	8396 bp	0.34 %
	R1/LOA/Jockey	118	27055 bp	1.09 %
	R2/R4/NeSL	0	0 bp	0.00 %
	RTE/Bov-B	0	0 bp	0.00 %
	L1/CIN4	0	0 bp	0.00 %
LTR elements:		346	123454 bp	5.00 %
	BEL/Pao	35	9056 bp	0.37 %
	Ty1/Copia	8	3389 bp	0.14 %
	Gypsy/DIRS1	272	105711 bp	4.28 %
	Retroviral	0	0 bp	0.00 %
DNA transposons:		66	8096 bp	0.33 %
	hobo-Activator	17	852 bp	0.03 %
	Tc1-IS630-Pogo	9	1546 bp	0.06 %
	En-Spm	0	0 bp	0.00 %
	MuDR-IS905	0	0 bp	0.00 %
	PiggyBac	0	0 bp	0.00 %
	Tourist/Harbinger	4	400 bp	0.02 %
	Other (Mirage, P-element, Transib)	14	1309 bp	0.05 %
Rolling-circles:		0	0 bp	0.00 %
Unclassified transposons:		5	14071 bp	0.57 %
Small RNA:		0	0 bp	0.00 %
Satellites:		2	604 bp	0.02 %
Simple repeats:		83	6449 bp	0.26 %
Low complexity:		0	0 bp	0.00 %

Table 5. Distal intergenic sequences corresponding to transposons and DNA repeats. *Most repeats fragmented by insertions or deletions have been counted as one element.

At this point, we were convinced that distal intergenic S9.6 peaks were RNA:DNA hybrids. However, we were unable to explain their source. Thus, the next step was to identify in which kind of chromatin these distal intergenic R-loops were localized by creating a reference map using the computational approach ChroGPS. ChroGPS integrates genomic distribution data from the modENCODE project and, through multidimensional scaling techniques, allows the visualization in 2D/3D maps of the association between epigenetic factors, or between genetic elements on the basis of their epigenetic state [143].

The ChroGPS factors map integrating intergenic R-loops is shown in **Figure 34**. Intergenic R-loops lie in close vicinity to H3K27me3, suppressor of hairy wing (Su(Hw)) and Modifier of mdg4 (Mod(mdg4)). This is seen both in the heatmap and the 2D map (**Figure 34 a/b**). **Figure 34 c** shows the proportion of distal intergenic R-loops overlapping with the different factors and histone modifications, confirming the strong colocalization with H3K27me3 and the significant colocalization with Su(Hw) and Mod(mdg4).

(a)



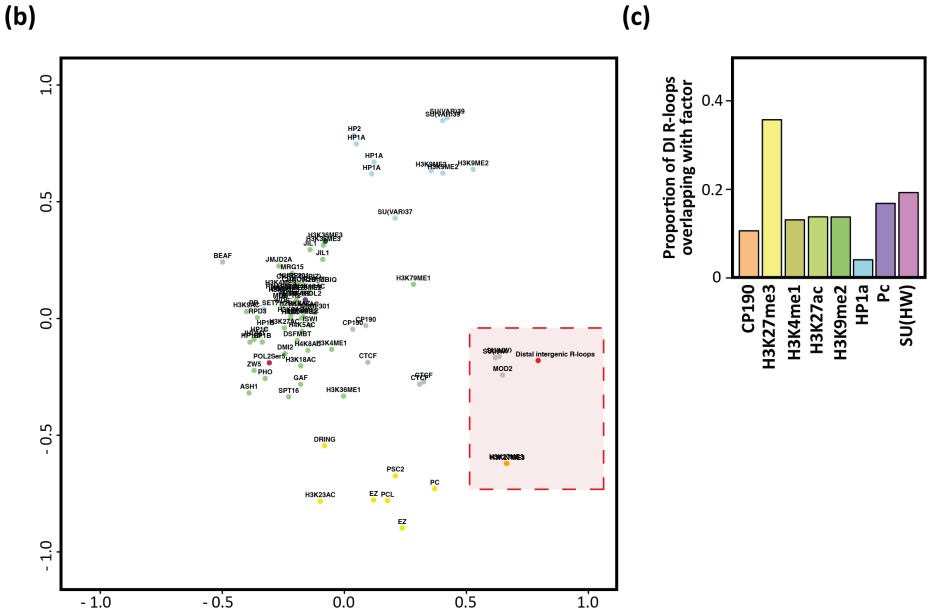


Figure 34. Intergenic R-loop are associated with different epigenetic factors. (a) Heatmap classifying the different ChroGPS elements and the distal intergenic R-loops. Levels of similarity are presented in a color scale from blue (0 similarity) to red (identical). **(b)** 2D map showing the relationship between distal intergenic R-loops and the different chromatin factors. Factors in close vicinity to distal intergenic R-loops are highlighted. **(c)** Proportion of distal intergenic R-loops overlapping with different chromatin factors (CP190, H3K27me3, H3K4me1, H3K27ac, H3K9me2, HP1a, Polycomb (Pc) and Su(Hw)).

H3K27me3 is a repressive and polycomb-associated chromatin mark and both Su(Hw) and Mod(mdg4) are components of the Gypsy insulator. The Gypsy insulator has been related with the nuclear organization and the establishment of higher order chromatin domains, and is thought to bring together several insulator sites in the interphase nucleus [144]. Interestingly, the analysis of the composition of the intergenic R-loop sequences by RepeatMasker showed that 4.28% of distal intergenic DNA corresponded to Gypsy/DIRS1 (**Table 5**).

2.2. R-loop distribution upon depletion of histone H1

Once the physiological distribution of RNA:DNA hybrids in *D. melanogaster* was known, changes produced upon depletion of histone H1 were examined. In contrast to the 7.96% of R-loop content found along the

genome at normal conditions (untreated cells), S2 cells depleted for histone H1 showed a clear reduction: a total of 2,083 peaks were detected, occupying an extension of 3,453,153 bp and, thus, representing 1.97% of the *D. melanogaster* genome.

Therefore, upon dH1-depletion a huge number of R-loops disappeared, others maintained and some appeared in new locations. In **Figure 35 a** (left), a comparison of the genes containing R-loops for untreated and LacZ RNAi-treated cells (control cells) showed that even though the total number of R-loops from control cells was 30% reduced, this reduction was heavily increased for cells treated with dsRNA against dH1, which presented a reduction of 60% in the number of genes with R-loops compared to untreated cells (**Figure 35 a** (right)). In both cases, >75% of the genes with R-loops are also present in the untreated.

Genomic distribution of R-loops for LacZ RNAi-treated cells was similar to the untreated condition (**Figure 35 b**; left). More than 75% of mapped S9.6 peaks corresponded to genic regions. As in untreated, most of the genic R-loops were found within introns or overlapping both intron and exon regions (42% of R-loops mapped over the intersection of exons and introns, whereas 33% of R-loops localized at introns not matching with exons). Also, 23% of R-loops localized in promoters and only 7% of R-loops localized at exons not coinciding with introns. Intergenic peaks (<25%) corresponded basically to distal intergenic regions. Contrary to cells treated with dsRNA against LacZ, the R-loop distribution for cells depleted for histone H1 changed notably (**Figure 35 b**; right and **c**). The intergenic R-loop contribution increased, being more than the 25% of the total number of peaks and, with that, the distal intergenic peaks. Analyzing in detail the different proportions of R-loops between control and dH1-depleted cells, it was observed that the percentage of intron-exon overlapping R-loops was reduced to 27%, whereas the amount of R-loops in introns not matching with exons increased to 41%. Also, that the number of R-loops at exon regions not coinciding with introns maintained similar to the control condition (6%), and that the missing R-loops mostly consisted on peaks localized at the promoter region of genes, which decreased to the half (12%).

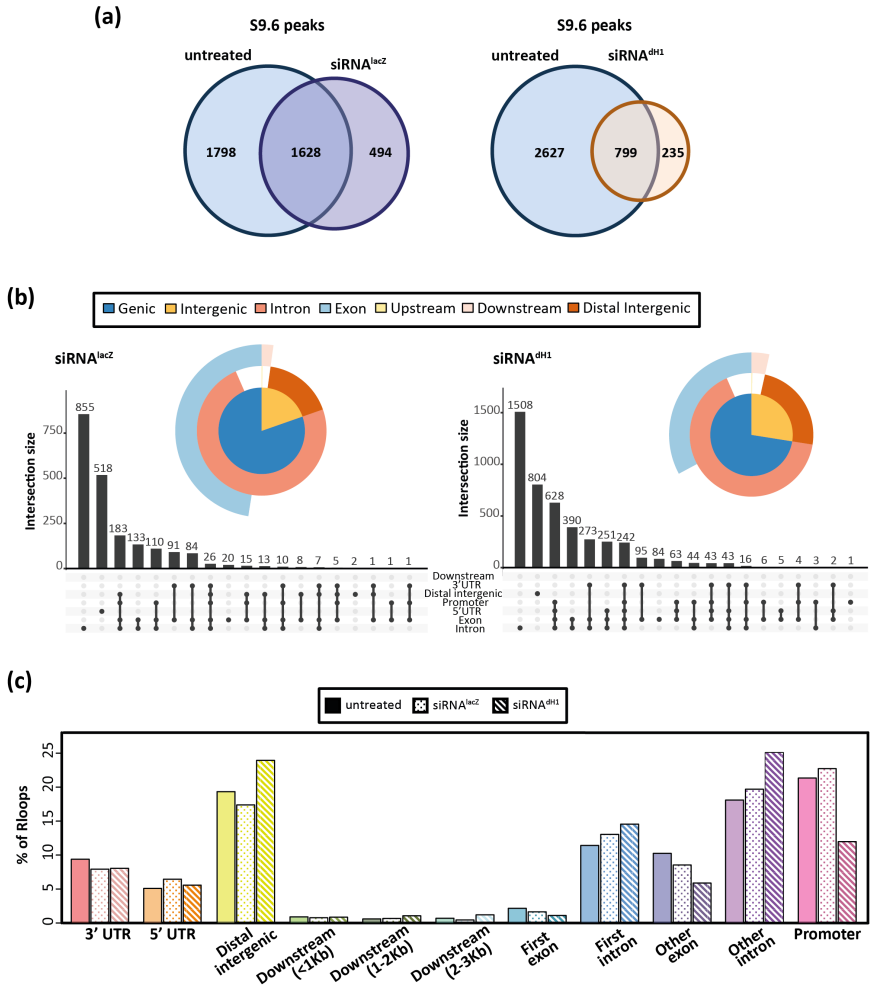


Figure 35. R-loop distribution changes upon depletion of dH1. (a) Venn diagram showing the intersection between genes containing S9.6 peaks in control and untreated cells (left), and in untreated and dH1-depleted cells (right). **(b)** R-loop genomic distribution for siRNA^{lacZ} (control) and dH1-depleted cells. S9.6 peaks are classified according to the genomic feature they belong to and are represented in a pie plot. Peaks corresponding to more than one feature are also presented as a combination of them, and the different contributions can be observed in the bar plot. **(c)** A more detailed representation of the percentage of R-loops depending on genomic feature in control (dotted) and dH1-depleted (striped) cells. Values for untreated condition (plain color) from **Figure 30 c** have been added for comparison.

Then, the R-loop distribution within the gene body was studied. Cells treated with dsRNA against LacZ presented a profile similar to untreated cells (**Figure 36 a** (left) and **Figure 31 a**), where R-loops highly concentrated near the TSS, but also showed accumulation along the whole CDS. Instead, for dH1-depleted cells, an intense peak of R-loops was observed near the TSS, but little R-loop accumulation was distinguished along the rest of the gene (**Figure 36 a** (right)), suggesting that some changes were happening inside the gene body in dH1-depleted cells.

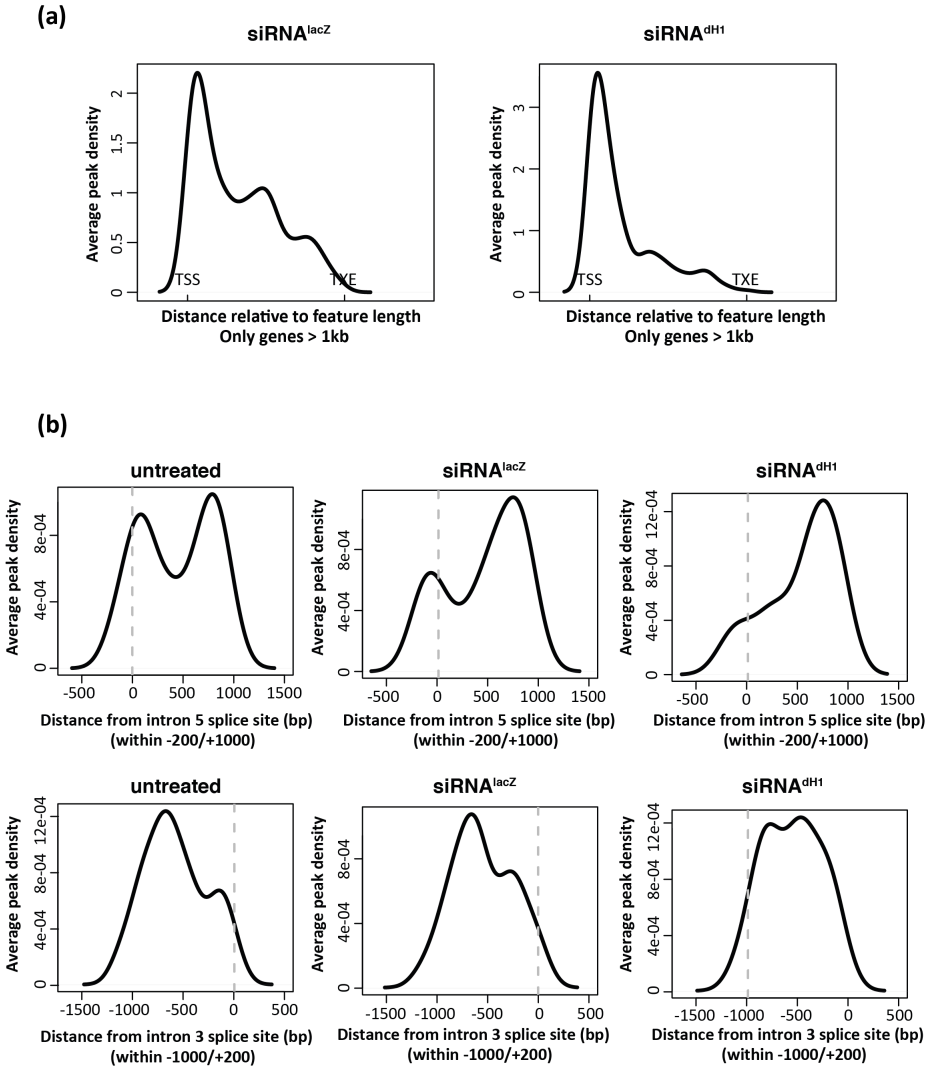


Figure 36. R-loop distribution within the gene body in dH1-depleted cells. (a) Genic R-loop peak location relative to closest/overlapping feature for genes > 1 kb in control (siRNA^{lacZ}) and dH1-depleted cells (siRNA^{dH1}). **(b)** R-loop distribution at a close

distance from 5' splice site (above) and 3' splice site (below) for untreated, control (siRNA^{lacZ}) and dH1-depleted cells (siRNA^{dH1}). 0 indicates the splice site location.

Therefore, in order to decipher the location of these specific changes, the average peak intensity at 5' splice-site and at 3' splice-site for untreated, control and dH1-depleted cells were analyzed (**Figure 36 b**). Results showed that the average peak density profile was different for the dH1-depleted condition, especially in the 5' splice-site.

3. CONTRIBUTION OF HISTONE H1 TO GENOME STABILITY IN TUMOR DERIVED CELLS

Chromatin regulation is critical for differentiation and disease. For example, it is well established that pluripotent stem cells are characterized by an open chromatin conformation with sparse and disorganized heterochromatin. Thus, this particular organization of chromatin confers a general accessibility for transcriptional activation and local silencing of lineage-specific genes [145]. But as cells differentiate, stem cells need to epigenetically repress pluripotent genes, and histone H1 and higher-order chromatin compaction have been shown to be important for the proper differentiation and lineage commitment of pluripotent stem cells [146]. Thus, H1 plays a critical role in pluripotent stem cell differentiation.

Based on these changes in chromatin configuration, it is believed that the same principles may apply during de-differentiation in cancer cells. Considering then the results obtained in *D. melanogaster* upon depletion of histone H1 and hypothesizing that the prevalence of the somatic linker histone variants in cancer cells should be affected in order to display an 'open' chromatin state, **we wondered if histone H1 levels in cancer cells were reduced and if they were relevant for the intrinsic DNA damage that characterize those cells**.

3.1. Colon carcinoma cell lines have lower levels of total histone H1

In order to evaluate the histone H1 content in cancer cells, we analyzed different colon carcinoma cell lines (HCT116, HCT15, LS174T, SW48, SW480, LS180, HT29, Caco2, HCT8) and an endometrium cancer cell line (SKUT-1B).

Regular colon tissue samples were used as a control for normal conditions. These colon cancer cell lines were chosen because they had been characterized extensively and were representative of different stages in tumor progression and malignancy. The endometrium cancer cell line was added to discard specificity of the results to only colon tissue.

Then, in order to obtain the total histone content of each cancer cell line, hydrochloric acid extractions were performed. In addition, perchloric acid extractions were done to isolate only the histone H1 variants. Therefore, for each cell line, a combination of both histone extractions was run on 15% SDS-PAGE gels (**Figure 37 a**), where the perchloric acid extraction was used to identify the bands corresponding to histone H1 in the hydrochloric acid extraction. As it can be observed in **Figure 37 a**, histone H1 proteins displayed three visible bands on the 15% SDS-PAGE gels: the upper one (band 1) corresponded mainly to H1.3, H1.4 and H1.5 variants, the middle band (band 2) was H1.2 and the lowest one (band 3) represented H1.0 [147]. Colorectal and endometrium cells do not express H1.1 [148].

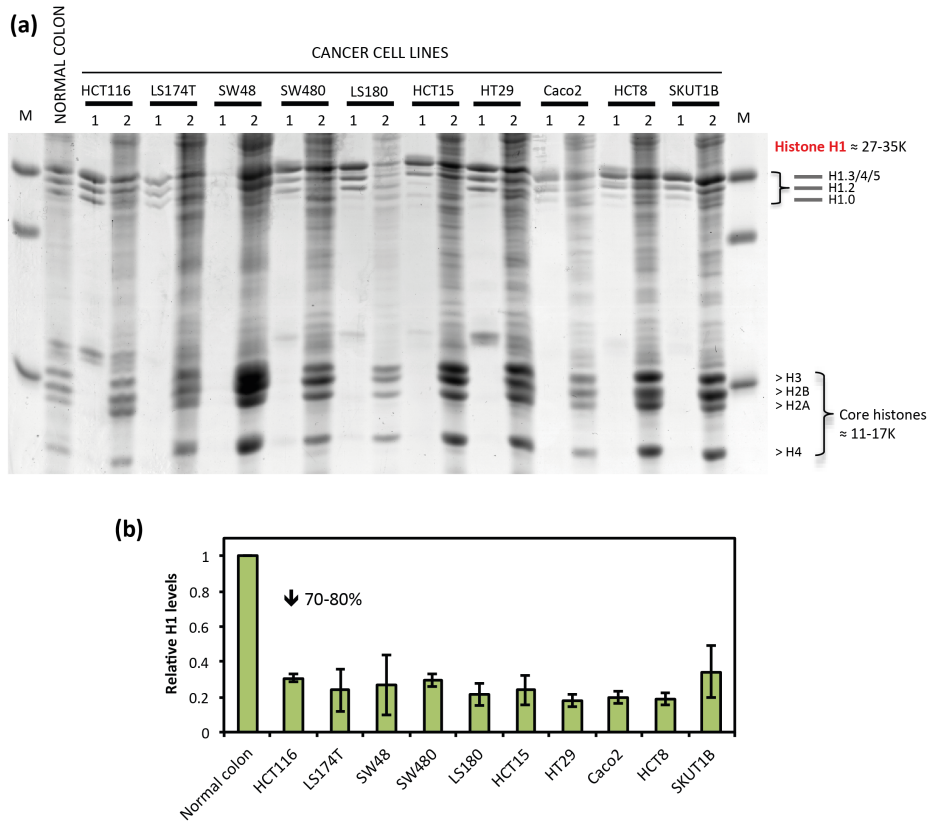


Figure 37. Histone H1 content within different cancer cell lines. (a) Total histone or H1 extracts for different cancer cell lines were obtained from a hydrochloric acid extraction and a perchloric acid extraction, respectively, and run in a 15% SDS-PAGE gel. Then it was stained with coomassie solution. M: molecular weight protein marker; 1 corresponds to perchloric acid extraction, 2 represents the hydrochloric acid extraction. **(b)** Results of histone H1 content compared to H4 for each of the cancer cell line are represented respect to a regular colon sample (N=2). Error bars are s.e.m.

The gels were finally stained with coomassie brilliant blue R250, digitalized using a laser densitometer and quantitated using Fiji software. The amount of total histone H1 was presented respect to H4 for each cell line (H1/H4) and all the ratios were referred to the H1/H4 normal colon ratio. Colon cancer cell lines presented a reduced content of total histone H1 (about 20-30% respect to normal colon tissue, indicative of a 70-80% reduction) (**Figure 37 b**).

3.2. Colon carcinoma cell lines show increased DNA damage compared to a normal tissue

Once proved that colorectal cancer cell lines have reduced histone H1 content compared to a regular colorectal tissue, we wondered how were their levels of DNA damage as cancer cells accumulate an enormous amount of deleterious structures in form of DSBs, chromosomal rearrangements and mutations.

Thus, some of the colon carcinoma cell lines (LS180, HCT15, HT29 and HCT116) were selected to quantify the amount of damage. To get a more exact quantification and avoid technical misinterpretations, an independent WB for each cancer cell line was performed using three biological replicates with two increasing amounts for each sample. In addition, the same sample of histone-extracted proteins from rat liver was loaded onto each gel (**Figure 38 a**). The rat liver sample was selected as a control as we had plenty of it and could be added in each WB. The amount of total γH2A.X was presented respect to H4 for each cell line (γH2A.X/H4) and all the ratios were referred to the γH2A.X/H4 rat liver ratio, to finally obtain a comparative quantification between the different WBs (**Figure 38 b**). γH2A.X has been extensively used as a marker of DSBs [149]. H2A.X is a variant of the H2A protein family, whose phosphorylation represents the first step in recruiting and localizing DNA

repair proteins to sites of DSBs. This phosphorylation is the equivalent of γH2Av in *D. melanogaster*, when γH2Av recognizes and signals DSBs. As expected, γH2A.X was hardly detectable in regular colon tissue samples, while DNA damage levels were clearly detectable and increased in cancer cell lines.

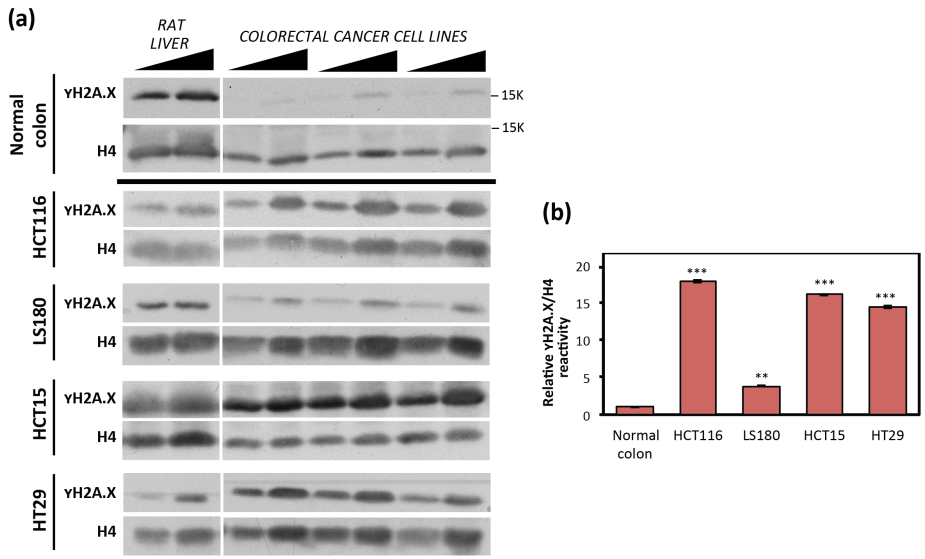


Figure 38. Increased DNA damage in colon carcinoma cell lines (a) WB analyses with αγH2A.X and αH4 antibodies of three independent total histone extracts with its corresponding two increasing amounts prepared from different colon carcinoma cell lines. The positions corresponding to molecular weight are indicated for the normal colon. **(b)** Quantification of the WBs are plotted respect to normal colon (N=3). For each WB, the colorectal cancer γH2A.X/H4 ratio is normalized to the rat liver γH2A.X/H4 ratio. Error bars are s.e.m. The p-values of different colon carcinoma cell line respect to normal colon are indicated (**<0.01, ***<0.005; two-tailed Student's t-test).

3.3. Overexpression of histone H1 variants in HT29 cancer cell line

These different colon cancer cell lines confirmed to be a good model for studying the effects of reduced histone H1 content to DNA damage in a pathological environment, as they naturally contained lower histone H1 levels and an increased amount of γH2A.X. Thus, aiming to establish a

histone H1 content-genome stability correlation, we decided to overexpress specific histone H1 variants to increase the total amount of histone H1 and see whether we restored the normal γH2A.X levels.

Both histone H1.2 and H1.4 variants were selected to be overexpressed into the HT29 cell line, as they are the most expressed variants in all the tissues investigated so far [70]. HT29 derivative cells stably expressing HA-tagged H1.2/H1.4 variants and also a combination of both of them were obtained from lentiviral transductions. For the overexpression of H1.2 and H1.4 variants, a particular pEV833.GFP vector was used (provided by Dr Albert Jordan (IBMB-CSIC)). For each of the variants the specific complete H1 sequence followed by HA was cloned upstream an internal ribosome entry site (IRES)-GFP cassette (**Figure 39**). HEK 293T cells were used to produce the virus particles containing the lentiviral expression vector that was later transfected into the HT29 cells. Also, a pEV833.GFP plasmid without any histone H1 sequence was expressed in HT29 as a control (HT29-pEV833). Because the vectors co-expressed the green fluorescence protein (GFP) marker, infected cells were selected for GFP-positive fluorescence by cell sorting (FACS) and amplified until obtaining a homogeneous population of cells overexpressing H1.2 (HT29-H1.2), H1.4 (HT29-H1.4), or H1.2 and H1.4 (HT29-H1.2/H1.4).

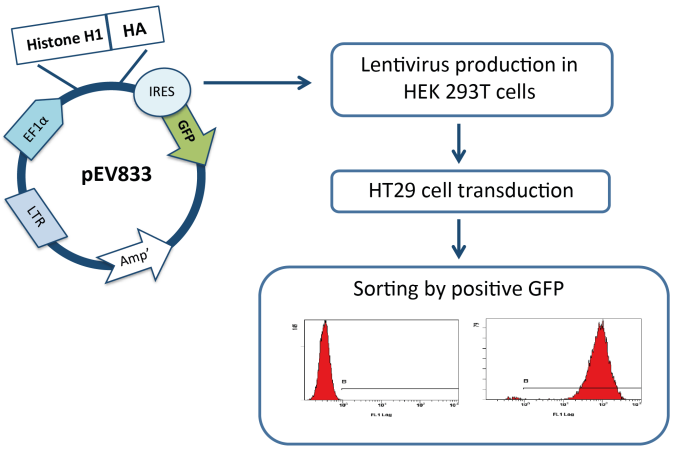


Figure 39. Schematic representation of the HA-tagged H1-variant overexpression in HT29 cells. HEK293T produced virus particles containing H1.4-HA-pEV833.GFP or H1.2-HA-pEV833.GFP vectors. Then, those particles were transfected, where H1.2-HA and H1.4-HA were incorporated into the genome of HT29 cells. HT29 cells stably expressing either H1.2 or H1.4, or a combination of both of them, were selected by FACS. X-axis represent the number of cells and Y-axis the levels of GFP fluorescence.

On the left, the GFP signal corresponds to autofluorescence; on the right, it corresponds to the fluorescence coming from the GFP).

To quantify the efficiency of overexpression, the total histone content and the histone H1s for HT29, HT29-pEV833, HT29-H1.2, HT29-H1.4 and HT29-H1.2/H1.4 were purified, run in a 15% SDS PAGE gel, stained and quantified as mentioned in Chapter 3.1. The histone H1 extraction of HT29-H1.4 cells presented an additional band (band 0) (**Figure 40 a**) that WB analysis confirmed to correspond to H1.4-HA (**Figure 40 b**). Hence, the endogenous H1.4 variant run in the same position as the untreated HT29 cell line (band 1) and the overexpressed form run in this new band (band 0). For HT29-H1.2 cells, the H1.2-HA was almost undetectable and no visible extra band was noticed. Due to the high background after staining the gel, it was not easy to differentiate the bands corresponding to histone H1s in the hydrochloric acid extractions. Therefore, no proper quantification of the total histone H1 content could be performed (**Figure 40 a**).

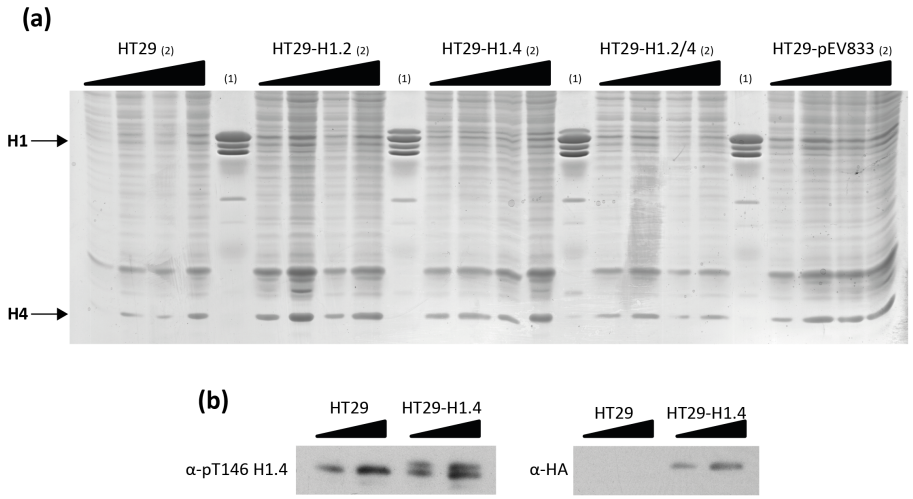


Figure 40. Overexpression of histone H1 variants in HT29 cancer cell line. (a) Total histone or H1 extracts for HT29 cancer cell line overexpressing H1.2, H1.4, H1.2/4 or pEV833 were obtained from a hydrochloric acid extraction and a perchloric acid extraction, respectively, and run in a 15% SDS-PAGE gel. Then it was stained with coomassie solution. 1 corresponds to perchloric acid extraction, 2 represents the hydrochloric acid extraction. **(b)** WB analyses with α-pT147 H1.4 and α-HA antibodies of two increasing amounts of HT29 and HT29-H1.4 are showed. HT29-

H1.4 presented two bands: the upper one corresponding to the exogenous H1.4 (as it is also immunoblotted with HA) and the lower one corresponding to the endogenous H1.4.

Luckily, the perchloric acid extraction SDS-PAGE gel was cleaner and allowed us to analyze the percentage of each band compared to the total amount of histone H1. A total of three rounds of infection-selection were performed to overexpress H1.2 and H1.4 histone proteins. Still, no visible increase was obtained for H1.2 overexpression, whereas a 33% of H1.4 overexpression was obtained for HT29-H1.4 (**Figure 41 a**). In view of the results, HT29-H1.2 and HT29-H1.2/H1.4 cells were discarded. Examining the FACS profiles, a difference in the intensity of GFP was detected between HT29-H1.4 and HT29-pEV833. The GFP intensity was higher in HT29-H1.4 cells, suggesting that a higher amount of plasmid was incorporated into the HT29-H1.4 cells (**Figure 41 b**). The purity of the culture was around 90%, and it was controlled over the time.

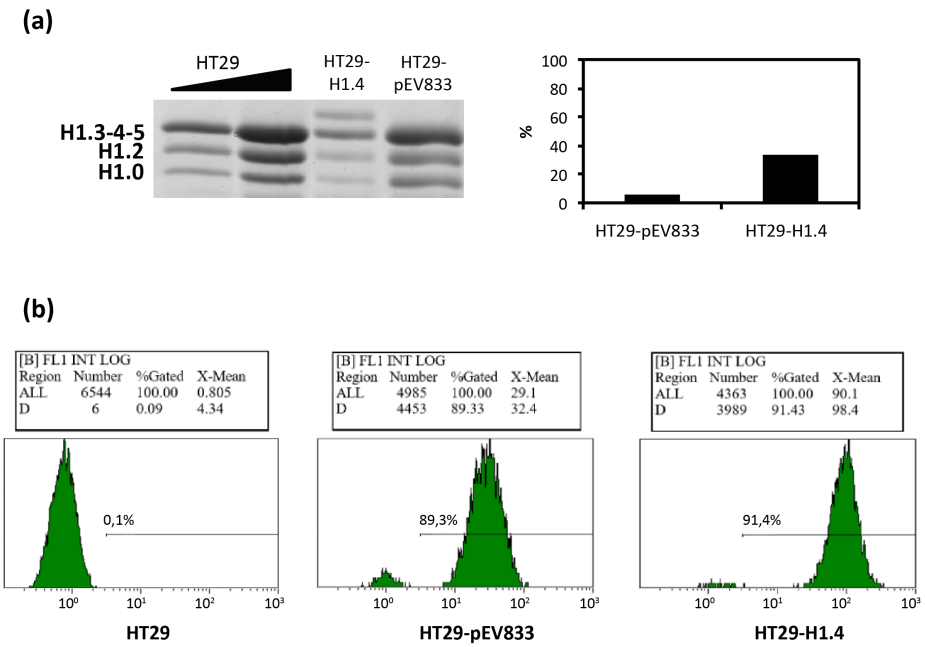


Figure 41. Histone H1.4 overexpression in the HT29 colorectal carcinoma cell line. **(a)** 15% SDS-PAGE gel coomassie staining of a total histone H1 extract from perchloric acid extraction of HT29, HT29-pEV833 and HT29-H1.4 is presented. The

different bands corresponding to the different histone H1 variants are shown and the H1.4 overexpression respect to the untreated cells is plotted. **(b)** FACS profile shows the GFP intensity (X-axis) for each of the different HT29, HT29-pEV833 and HT29-H1.4 cell lines. Number of cells is presented in the Y-axis and % of the cells positive for GFP is indicated.

The levels of histone H1 mRNAs of HT29-H1.4, HT29-pEV and HT29 cells (N=3) were later examined. Briefly, total RNA was reverse-transcribed into cDNA using random hexamer primers, as histone H1s lack from poly(A) tail [81], and analyzed by RT-qPCR. Results were normalized to tubulin and compared to mRNA extracted from a normal colon tissue (**Figure 42**). mRNAs for H1.0, H1.2 and H1.4 were detected, whereas no signal was barely observed for H1.1 (not present in colon), H1.3, H1.5 and H1.X. H1.4 mRNA levels were up-regulated up to 30% in HT29-H1.4 cells respect to the untreated HT29 cells.

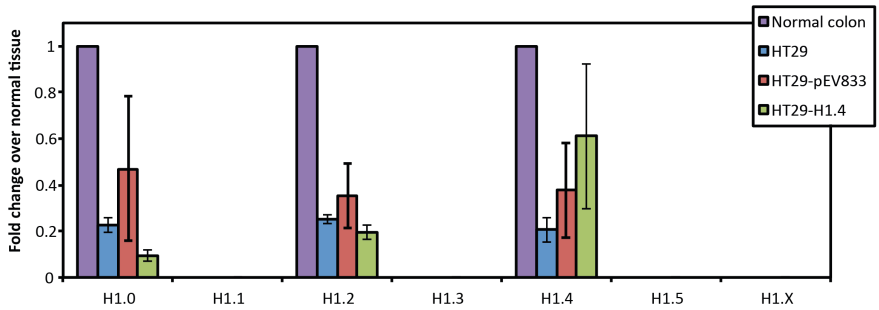


Figure 42. Histone H1 variants profile in HT29-H1.4 overexpressing cells. Real-time quantitative PCR of the indicated histone H1 subtypes in HT29 (blue), HT29-pEV833 (red) and HT29-H1.4 (green) cells. The mRNA levels are expressed as a fold change respect to normal colon (magenta). (N=3) Error bars are s.e.m.

3.4. H1.4-overexpression does not importantly affect the DNA damage in HT29 cells

Having successfully overexpressed H1.4 in HT29 cancer cell line, the next step was to evaluate the impact of the histone H1 overexpression over the DNA damage. For that, levels of γH2A.X were analyzed as it was done for the set of colorectal carcinoma cell lines in Chapter 3.2.

Firstly, the hydrochloric acid extraction samples for HT29, HT29-pEV833 and HT29-H1.4 were used to perform a WB against γH2A.X and the results

were normalized to H4. The averages of several independent WB (N=6) were compared. As shown on the quantifications (**Figure 43 a**), HT29-pEV833 increased the DNA damage respect to HT29-H1.4. However, levels of HT29-H1.4 and the untreated HT29 condition did not show any significant difference.

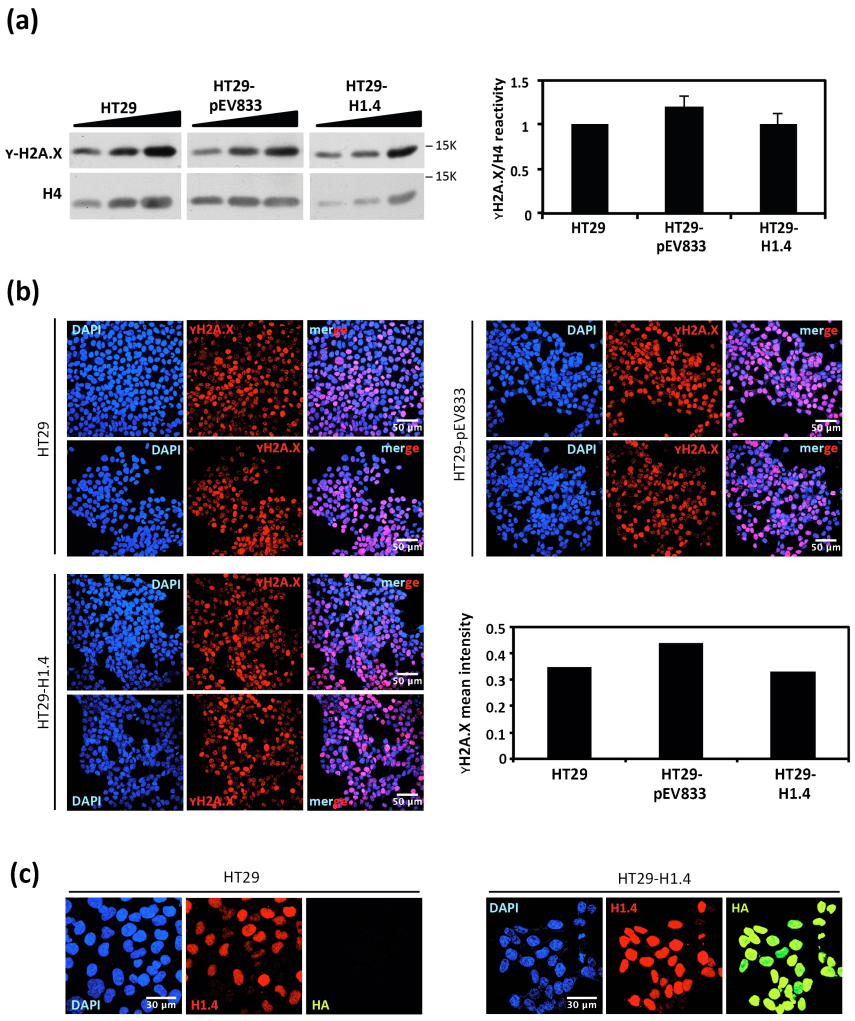


Figure 43. DNA damage in HT29 cells overexpressing H1.4. (a) On the left, a WB analysis of three increasing amounts of a hydrochloric acid extraction sample prepared from HT29, HT29-pEV833 or HT29-H1.4 cells is blotted against γH2A.X and H4 antibodies. On the right, quantification of the γH2A.X/H4 ratio and expressed over HT29 cells (N=6). Error bars are s.e.m. **(b)** Immunostaining of HT29, HT29-pEV833 and HT29-H1.4 cells against γH2A.X (red). Dapi is in blue. The third image of each row represents the merge of the two channels. Scale bar corresponds to 50 µm.

Quantification of the mean γH2A.X intensity of each cell condition and expressed over HT29. **(c)** Immunostaining of HT29 and HT29-H1.4 cells against H1.4 (red) and HA (green). Dapi is in blue. Scale bar corresponds to 30 µm.

Next, the different cell lines were immunostained against γH2A.X and the γH2A.X reactivity was quantified for each condition: HT29 (n=1379 cells), HT29-pEV833 (n=1051) and HT29-H1.4 (n=1583). The Fiji software was used to determine the mean intensity of γH2A.X reactivity in each condition (**Figure 43 b**). Results were similar to those obtained for the WB; suggesting that levels of HT29-H1.4 and the untreated HT29 condition did not importantly differ. In contrast, HT29-pEV33 tended to increase the amount of DNA damage compared to the untreated HT29 cells.

Immunostaining experiments against H1.4 and HA were performed for HT29 and HT29-H1.4 to evaluate the distribution of the H1.4 and H1.4-HA expression. Both H1.4 (corresponding to the endogenous and exogenous H1.4) and H1.4-HA (corresponding to the exogenous H1.4) behaved apparently similar inside the nucleus (**Figure 43 c**).

DISCUSSION

DISCUSSION

1. ABSENCE OF HISTONE H1 INDUCES R-loop-MEDIATED DNA DAMAGE

Histone H1 is an essential component of chromatin that helps to maintain its correct structure and function. Thus, assuming a major role in the chromatin organization, depletion of histone H1 was expected to have important effects on nuclear structure and cell viability. Surprisingly, loss of histone H1 genes in several organisms did not affect viability [71,72,150] and it was not until a 50% reduction of histone H1 content in mouse that it lead to embryonic lethality. And even in those conditions, 50% loss of H1 content was not lethal for stem cells[73].

Drosophila melanogaster is an excellent model organism for studying histone H1 as it has only one somatic isoform, in contrast to vertebrates that have multiple and functionally redundant histone H1 subtypes. Besides, it is very convenient for studying chromatin-related processes and proteins due to the amount of existing possibilities, some of which we have extensively used during the development of this thesis. Among them, we can highlight the high-throughput data available for genetic and genomic studies, the UAS/GAL4 system that provides a high degree of flexibility for expressing transgenes both spatially and temporally, and the clear visualization of chromatin and proteins bound to it through salivary gland polytene chromosomes. For that reason, the main part of the project has been conducted on *D. melanogaster* flies and cells.

In particular, histone H1 is specifically important in heterochromatin, where it has a fundamental role in preserving its highly compact and repressed state. Accordingly, previous studies in *D. melanogaster* have showed that although histone H1 is highly represented along the genome, its depletion affected only a subset of genes [1–3]. These differentially expressed genes were mainly heterochromatic inactive genes that became transcribed [2]. Furthermore, apart from affecting the expression of heterochromatin silencing, its depletion also induced DNA damage and genome instability. In Bayona-Feliu & Casas-Lamesa et al., it was observed that DNA damage (determined by γH2Av reactivity) accumulated at

heterochromatic regions upon depletion of histone H1. Especially, it accumulated at the same heterochromatic sequences that were being up-regulated when levels of histone H1 were decreased (**Figure 44**), meaning that abnormal transcription of these sequences was resulting in DNA lesions. Besides, this DNA damage was generated during S-phase, suggesting that it was dependent on DNA replication [4].

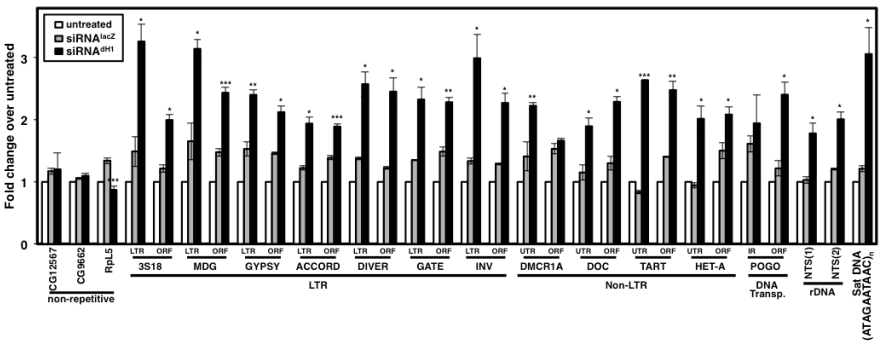


Figure 44. DNA damage induced by dH1 depletion occurs preferentially in heterochromatin. γH2Av ChIP-qPCR analyses at the indicated repetitive and non-repetitive regions in siRNA^{dH1}, siRNA^{lacZ} and untreated cells (N=2). For each position, the fold change respect to the untreated condition at this position is presented. Error bars are s.e.m. the p-values of siRNA^{dH1} respect to siRNA^{lacZ} are indicated (no asterisk >0.5, * <0.01, ***<0.005; two-tailed Student's t-test). Results from [4].

Powerful techniques designed to completely eliminate the genetic function of *D. melanogaster* histone genes have been designed. For example, the genetic system BAC-based platform developed by McKay et al. [96], which allows total deletion of the histone gene cluster. However, as histone H1 protein levels from 20% or below are shown to be lethal in *D. melanogaster* [1], a dsRNA approach to partially deplete dH1 expression was used instead. Moreover, depletion of histone H1 in *D. melanogaster* has proven to be arduous due to the multi-copy nature of the *his1* gene (approximately 100 copies on the haploid genome). Thus, depletion using a dsRNA resulted in an incomplete and heterogeneous down-regulation, where dH1 expression of the knockdown was around 60% of the endogenous levels (**Figure 14**). For that reason, results observed upon depletion of histone H1 might also be partial. To deplete histone H1 in flies, a dsRNA approach driven by the UAS-GAL4 system was also used.

Thus, to increase the depletion efficiency in the system, a combination of the UAS-RNAi^{dH1}*;nub*GAL4 with overexpression of the ribonuclease Dicer was used.

Studies in the field have shown that impairment of the replication fork is a major driver of DNA damage, and a substantial obstacle for its progression is the transcription machinery, in particular, when it is stalled by RNA:DNA hybrids (R-loops) [131,151]. Thus, a possible explanation for the DNA damage observed after relief of normally silenced sequences in heterochromatin was that the heterochromatic transcripts were re-annealing with the DNA template, inducing the formation and accumulation of R-loops.

Most of the current techniques for detecting R-loops are based on the S9.6 antibody. The monoclonal S9.6 antibody recognizes RNA:DNA hybrids with high specificity. However, it can cross-react with other RNA species resulting in false positives [152,153]. Accordingly, the DRIP-seq was performed following an adapted protocol combining treatment with RNase A and RNase H. Pre-treatment with RNase A helped to eliminate artifacts due to free RNA. Halász et al. showed that RNA:DNA hybrids were resistant to RNase A digestion at high ionic strength (300 mM NaCl) [154]. Besides, in order to validate the R-loop signal, a sample in parallel was processed and treated with RNase H prior to the immunoprecipitation. RNase H degrades the RNA moiety of the hybrid and, therefore, removes partially any “real” R-loop signal. The effects of RNase A and H pre-treatment on DRIP experiments were studied by others and proved to be efficient [107,154]. Apart from RNase A and H treatment, in order to reduce inherent biases of the method, an special attention was paid to different parameters along the DRIP workflow; such as avoiding fixation (as formaldehyde fixation induces a conformational change of the DNA, which might abolish R-loops or create ectopic R-loop sites) and shearing the DNA through sonication instead of restriction enzyme digestion to avoid a biased fragmentation [154].

Due to the repetitive nature of heterochromatic DNA, mapping of reads and, consequently, the analysis of the high-throughput sequencing data were complicated. Accordingly, the DRIP-seq data was first studied using a multi-hit approach (Chapter 1), where the reads that map to multiple locations were assigned to all possible sites in the entire genome to scatter them uniformly across all repeat copies. After the differential binding

analysis, an enrichment of validated RNA:DNA hybrids mainly localized in the U Extra chromosome. The U Extra is an unordered artificial chromosome where those unmapped sequences of the *Drosophila* genome are assigned, most of which correspond to heterochromatic elements. So it was not surprising that R-loops concentrated there, as depletion of histone H1 lead to up-regulation of silenced heterochromatic elements [1,2,155]. Comparing the γH2Av- and S9.6-enriched regions, results suggested that R-loops accumulated at the same heterochromatic regions where DNA damage was produced upon depletion of histone H1 (**Table 3**).

2. HISTONE H1 PREVENTS RNA:DNA HYBRID ACCUMULATION IN HETEROCHROMATIN

The above-mentioned observations lead us to question whether R-loops were accumulated in heterochromatin due to a simple relief of silenced sequences or due to a specific role of histone H1. Accordingly, HP1a was depleted. HP1a is also considered an important protein for heterochromatin establishment and maintenance. It is encoded by the *Su(var)2-5* gene and has two important domains; a N-terminal chromo-domain that binds H3K9me2/3, and a C-terminal chromo-shadow domain through which establishes protein-protein interactions, including itself and Su(var)3-9, the histone methyltransferase responsible for the H3K9me mark. Thus, the accepted model proposes that HP1a binds H3K9me2/3 and Su(var)3-9. Concomitantly, Su(var)3-9 methylates nearby H3K9, which in turn recruits HP1a. In this way, HP1a spreading is propagated helping to condense chromatin and promote gene silencing [43,64]. Therefore, as expected, HP1a depletion in S2 cells induced transcription of heterochromatic sequences to similar levels as for depletion of dH1 (**Figure 17**). Nevertheless, its depletion did not induce accumulation of R-loops nor DNA damage (**Figure 18**), suggesting that relief of heterochromatic silencing by itself was not the cause. Importantly, HP1a depletion did not affect the binding of histone H1 to heterochromatin, confirming that histone H1 was distributed along the heterochromatic sequences even when HP1a was missing (**Figure 20**). Then, DRIP-qPCR strengthened these observations, as co-depletion of dH1 in HP1a-depleted cells restored R-loop accumulation in those selected heterochromatic sequences (**Figure 21**).

Zeller et al. found that *Caenorhabditis elegans* H3K9me mutants exhibited increased open chromatin, which induced the expression of repetitive sequences and the formation of R-loops [156]. Nevertheless, the molecular mechanism by which R-loops were formed was not deciphered in this study. *C. elegans* possesses eight different linker histone variants. Previous depletion experiments for all the histone H1 isoforms showed that gene activation was observed mainly with H1.1 (HIS-24), which corresponds to the most abundant histone H1 variant in the roundworm. H1.1 is expressed in both soma and germ line cells, however its biological function is principally associated to the silencing of germ-line-specific chromatin [157]. As it happens for other organisms, H1.1 has also been related to the *C. elegans* HP1 homolog, HPL-1 and HPL-2. However, in this case, H1.1 recruits HPL proteins and together associate with the promoter of genes involved in the transcriptional regulation of many immune response genes [158], instead of participating in the formation of heterochromatin. So, given these differences for histone H1 activities between *D. melanogaster* and *C. elegans*, different implications for histone H1 in the regulation of R-loop dynamics in both organisms cannot be excluded.

Moreover, in spite of the link between H3K9-depletion and transposon up-regulation established in different organisms [1,156,159], our results proved that the up-regulation of heterochromatin per se was not the cause of R-loop-mediated DNA damage, since depletion of HP1a activated heterochromatin but did not compromise the genome integrity as it happened upon histone H1 depletion. Instead, it was histone H1 who specifically participated in the metabolism of heterochromatic R-loops (**Figure 21**). Furthermore, although histone H1 acts upstream Su(var)3-9 in *D. melanogaster* [3], the abrogation of histone H1 does not necessarily imply depletion of H3K9me, minimizing the possibility that the effects observed upon depletion of histone H1 were in fact the consequence of the downstream perturbation of H3K9me. In this regard, although Lu et al. slightly detect H3K9me2 in the chromatin of *D. melanogaster* H1-depleted salivary glands by indirect immunofluorescence, total protein levels in cell lysates were elevated rather than reduced [1]. Accordingly, in our hands, H3K9me2 was also detected in polytene chromosomes depleted for histone H1 (data not shown) and experiments performed by Dr J. Bernués (IBMB, CSIC; data not shown) showed that flies depleted for K9 (K9R)

barely had S9.6 reactivity in the chromocenter, suggesting that the absence of H3K9me neither induce R-loops nor DNA damage as it happens for HP1a. Furthermore, Isawaki et al. observed that H3K9 occupancy among the TEs was not significantly affected after knocking-down histone H1 [155]. So it is conceivable that histone H1 depletion might result in specific redistribution of H3K9me, limiting its density to the most risky sites within pericentric regions.

A recent study agrees with our results, as they observed that histone H1 was important in regulating R-loop accumulation in H1 TKO cells [160]. They showed a deficient RNAPII dynamics as a consequence of histone H1 reduction, which correlated with the accumulation of R-loops. Also, this impaired RNAPII activity favored transcription-replication conflicts that consequently, increased replication fork stalling and DNA damage.

3. hnRNP36 AND hnRNP48 COLLABORATE WITH dH1 IN THE R-loop PREVENTION

Although histone H1 is an important chromatin protein with roles in many biological processes, including heterochromatin formation, gene expression and DNA repair, among its properties, there is none that could actively regulate R-loops. Consequently, we proposed the scenario that histone H1 was acting as a platform to recruit R-loop destabilizing factors (**Figure 22**). Then, from a list of possible candidates known to interact with histone H1 and with enzymatic activities related to R-loop metabolism, hnRNP36, hnRNP48 and Su(var)2-10 were selected because they were described to compromise heterochromatin stability through their suppressor of variegation activities [132,134]. Surprisingly, neither RNH enzymes nor helicases involved in R-loop removal (e.g. senataxin, DHX9, etc.) were detected (**Table 4**), which are some of the main mechanisms described to counteract deleterious R-loops in many organisms [115,124,161–163]. This suggested that histone H1 was binding intermediaries rather than proteins directly responsible for R-loop removal.

Su(var)2-10 was later discarded as its depletion did not induce S9.6 reactivity, in contrast to hnRNP36 and hnRNP48 depletion, which resulted in a strong S9.6 reactivity specifically in the chromocenter (**Figure 24 a**). In agreement with our results, the RNA:DNA hybrid interactome recently published for HeLa cells found that different classes of proteins involved in

RNA metabolism were overrepresented, including hnRNPs [163]. Furthermore, in *white* polytene chromosomes, hnRNP36 and hnRNP48 displayed a strong co-localization with HP1a at the chromocenter, supporting their already described presence at heterochromatin [136,137,139]. Also, both hnRNPs distributed along the chromosomal arms following mainly an interband distribution. The interbands are associated with active chromatin, which agrees with the hnRNPs' described role in mRNA biogenesis during transcription [137]. hnRNP36 and hnRNP48 ChIP-qPCR experiments were performed upon depletion of histone H1 in order to determine if histone H1 was necessary for the recruitment of the hnRNPs to the specific up-regulated heterochromatic sites. Unfortunately, we were unable to obtain any immunoprecipitate. Maybe the hnRNPs were not being detected as they were bound to chromatin through an mRNA, and thus, the mRNA was too unstable to be detected during ChIP experiments. Steps such as sonication and RNase treatment could facilitate the loss of the hnRNP-chromatin interaction through degradation of the mRNA. Besides, it was also possible that the antibodies were not suitable for ChIP experiments. Then, with the aim of determining the binding of hnRNP36 and hnRNP48 to the heterochromatic transcripts, RNA immunoprecipitation (RIP) experiments were performed. However, the approach did not work either.

In line with this, immunostaining of polytene chromosomes upon depletion of histone H1 displayed a significant reduction of hnRNP36 and hnRNP48 at the chromocenter (**Figure 27**). Conversely, immunostaining of polytene chromosomes upon depletion of one or another hnRNP showed no significant changes in histone H1 content at the chromocenter (**Figure 27 b**), and dH1 ChIP-qPCR experiments confirmed these results (**Figure 26 b**). Altogether, it suggested that the hnRNP36 and hnRNP48 depended on histone H1 to bind at heterochromatin and, consequently, absence of histone H1 or of any of the hnRNPs induced R-loop accumulation in heterochromatin. Consistent with that, histone H1 might be needed to stabilize and maintain the chromatin compacted after transcription of heterochromatic sequences, giving time to the hnRNPs to specifically interact with histone H1 and thus, with the nascent transcript. In this regard, depletion of dH1 might complicate and slow down the recruitment of the hnRNPs to the heterochromatic transcripts; increasing the half-life

of RNA close to its DNA template, and consequently, increasing the probability of forming deleterious R-loops.

Additionally, the formation of stable R-loops likely alters local chromatin. This is shown in reconstitution histone-DNA experiments, where data suggest that an RNA in a nucleic acid double strand inhibits the interaction of histones with the hybrid molecule and, consequently, limits its condensation into nucleosomes [164]. This happens because the formation of stable nucleosomes requires certain interactions between the DNA and histones that compose the nucleosome over an extended distance from the octamer's surface. However, it has been observed that these interactions are much less favored with the RNA:DNA hybrids due to their different helical structure, as they adopt a more rigid A form-like conformation [165]. The concurrence of R-loops and an open chromatin conformation has been reported in different other studies [119,166]. Conversely, R-loop destabilization has shown to cause chromatin compaction [166]. Accordingly, this alteration in the chromatin pattern due to R-loop formation may destabilize histone H1 as well. It is possible then that R-loops affect dH1 occupancy on those heterochromatic regions raising a possible self-sustaining mechanism where histone H1 is needed to recruit hnRNP36 and hnRNP48 in heterochromatin and avoid R-loop accumulation, but at the same time, R-loops destabilize the chromatin conformation, complicating histone H1 binding.

Thus, having observed that R-loops are normally spread along the genome and that histone H1 depletion induces global changes on their distribution, why R-loop-mediated DNA damage in dH1-depleted cells was specifically detected in heterochromatin? On one side, heterochromatin poses a general challenge to transcription. So a high proportion of the heterochromatic transcripts expressed upon depletion of dH1 might result aberrant and inefficiently processed and exported, a context that increased the possibilities of forming R-loops. On the other side, dysfunction of factors involved in R-loop dynamics leads to deleterious R-loop accumulation. Although R-loops can be enzymatically removed, they are also prevented by factors involved in the RNA metabolism, such as the hnRNPs. Thus, it seems that histone H1 participates in this "prevention mechanism" by helping hnRNP36 and hnRNP48 to localize in heterochromatin. However, in conditions where histone H1 is not there,

hnRNP36 and hnRNP48 might not be directed to the nascent heterochromatic transcript, and consequently, R-loops could be formed (**Figure 45**).

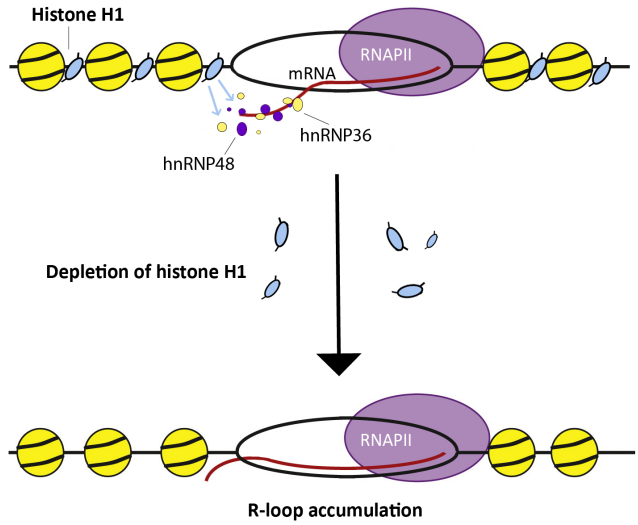


Figure 45. Schematic representation of the proposed model. dH1 specifically participates in the dynamics of RNA:DNA hybrids in heterochromatin by recruiting hnRNP38 and hnRNP48.

As explained before, results showed that hnRNP36 and hnRNP48 also bind to euchromatic regions. However, hnRNPs did not run into dH1 in euchromatin, as shown in white polytene chromosomes stained for hnRNP36 and dH1, where dH1 localized at bands and hnRNP36 tended to localize at interbands (**Figure 29**). Also, upon depletion of these hnRNPs, R-loops specifically formed at the chromocenter and no apparent reactivity was found along the chromosomal arms. Results also showed that histone H1 was still present after depleting one or another hnRNP, confirming that histone H1 by itself cannot prevent heterochromatic R-loops. Because of the condensed nature of heterochromatin, factors responsible to destabilize R-loops may have more difficulties to bind or degrade the transcripts than those happening in euchromatin. Thus, if histone H1 does not mediate hnRNP-binding to the newly transcribed RNA, the probability that this interaction happens by chance is very limited. Contrarily, euchromatin is more easily accessible and many other candidates can

participate in the dynamics of euchromatic R-loops, which might explain why DNA damage occurring after aberrant R-loops is preferentially detected at heterochromatic sites. Blanchette et al. studied the genome-wide distribution of different hnRNPs, including hnRNP36 and hnRNP48, in *D. melanogaster*. In this analysis, only a subset of euchromatic genes were examined, discarding from the analysis all the heterochromatic regions. Interestingly, among the hnRNP euchromatic targets, they described predominantly intronic binding locations for hnRNP36 and largely exonic binding regions for hnRNP48 [137], which happens to meet the sites enriched for genic R-loops (>75% are found within introns and exons). So both hnRNPs could also contribute to prevent euchromatic R-loops independently of histone H1.

Altogether, results presented here conclude that histone H1 is in between the physiological and pathological roles of R-loops; as on the one hand it is responsible for normal distribution of R-loops along the genome and, on the other, it prevents R-loops for remaining too long into heterochromatin and, consequently, DNA damage.

4. HISTONE H1 DEPLETION ALTERS THE NORMAL DISTRIBUTION OF R-loops IN EUCHROMATIN

High-throughput mapping studies have revealed that 5% of the human genome is covered by RNA:DNA hybrids [107], and single-hit analysis of our DRIP-seq data detected that 7.96% of *D. melanogaster* genome contains R-loops. So considering that R-loops a result of transcription, we expected to find more R-loops among the most active genes.

However, genic R-loops did not show a clear correlation with the transcriptional output of those genes (**Figure 31**), suggesting that genic R-loops cannot be explained by the transcription rate. It is likely that the transcription machinery itself attract specific components that help to metabolize the R-loops.

The regulation of RNAPII speed and pausing during transcription are key elements in regulating the level and timing of RNA production. Transcription elongation rates can vary between and within genes[108], so it might be that R-loops accumulate depending on the RNAPII transcription

speed. Therefore, for future analyses, it could be interesting to see whether additional correlations with factors such gene length or transcription speed exist, which would give us some information about the R-loops found along the CDS. On this direction, it could be interesting to study pausing of RNAPII, because it could be possible that at places were the RNAPII pause, the RNA might have more probabilities to re-anneal back to the DNA template and form an R-loop. Interestingly, RNAPII have shown to arrest and accumulate at the promoter-proximal region (some nucleotides downstream the TSS) [167], which could explain the peak of R-loops found at promoters (**Figure 30 c**) and at close distance to the TSS (**Figure 31 a**). These R-loops could be involved in transcription initiation or in controlling timing and rate of transcription, as others have already observed it [168,169]. Likewise, RNAPII pauses after transcription through the poly(A) site and before termination [170]. So it was thought that R-loops would also form in the TXE to facilitate termination by slowing down the advance of the RNAPII and promoting the recruitment of termination factors, however, our results did not show any R-loop enrichment at TXE.

Histone H1 depletion induced important changes in the distribution of R-loops along the genome of *D. melanogaster*. As mentioned before, multi-hit analysis of DRIP-seq revealed an important amount of R-loops accumulating at up-regulated heterochromatic sequences that did not happen for control cells. Besides, single-hit analysis confirmed that R-loops in dH1-depleted cells were also present along euchromatin, however in a different distribution than the one happening in physiological conditions. In fact, the number of R-loops decreased markedly, covering only 1.97% of the genome.

In order to explain the reduction in the R-loop coverage in euchromatic regions of dH1-deplete cells, different possibilities were examined. On one side, it was considered the possibility that this reduction was the result of a bias of the technique. The high-throughput sequencing method distributes the same number of reads in all the samples. Therefore, due to the increased number of R-loops in heterochromatin upon depletion of histone H1, a higher number of reads might have aligned to heterochromatic sequences and, consequently, the number of reads available for euchromatic regions was reduced. However, if it was the case, it should have affected all euchromatic R-loops to the same extent. Thus,

although R-loops in dH1-depleted cells were reduced in many of the genic features, they were more intensively reduced in promoter regions (**Figure 35**). Nevertheless, it could be interesting to perform a DRIP-qPCR against euchromatic regions to prove that there is in fact a reduction in the euchromatic R-loop content. On the other side, depletion of histone H1 could imply defects on pausing and, subsequently, explain why dH1-depleted cells accumulated less R-loops at promoter regions. Besides, depletion of histone H1 might also alter splicing, explaining why R-loop accumulation at splice sites was completely altered (**Figure 36**). To answer whether depletion of histone H1 induces euchromatic R-loop reduction due to RNAPII and splicing defects, further experiments are required.

5. CONTRIBUTION OF HISTONE H1 TO GENOME STABILITY IN TUMOUR DERIVED CELLS

Along this dissertation, it has been explored how deficient histone H1 levels result in detrimental outcomes for the cells. Thus, given the importance of having regulated histone levels and taking advantage of the results obtained after depleting histone H1 in *Drosophila*, we focused the last part of the project in understanding how important the histone H1 levels are to prevent the DNA damage and genome instability in cancer cells.

For that, different colon carcinoma cell lines molecularly and clinically well characterized were used. It is well known that the prevalence of the different linker histone H1 variants varies depending on cell type and differentiation stage. Recent advances have shown how changes in the expression of the different H1 variants modulate self-renewal and differentiation. For example, changes on the levels of H1.0 subtype resulted in a reduced differentiation capacity of ESCs, because H1.0 plays a critical role in regulating differentiation and pluripotency genes [89,146]. Furthermore, less differentiated cells contain a more open chromatin state and, concordantly, lower histone H1 levels [146]. Accordingly, it was hypothesized that the more aggressive a cancer cell line is, the more it divides and resembles a less differentiated cell type state. Hence, besides studying the relationship between the amount of histone H1 and DNA damage, there was an interest in establishing a correlation between H1

levels and aggressiveness of the tumor, as it was thought that the lesser amount of histone H1, the more aggressive the tumor is. A correlation already seen for histone H1.0 variant and patients with malignant gliomas [90].

The analysis of different colon cancer cell lines (HCT116, HCT15, LS174T, SW48, SW480, LS180, HT29, Caco2, HCT8) showed a reduction of 20-40% of total histone H1 levels compared to a normal colon tissue. Furthermore, DNA damage, detected through the phosphorylation of H2A.X, was significantly higher than that of the normal colon tissue. Opposite to expected, further analysis did not give us any detectable correlation between histone H1 levels and malignancy for each of the cell lines (data not shown).

HT29 cell line stably overexpressing H1.4 up to 33% of total histone H1 content (HT29-H1.4) was generated. However, increased total levels of histone H1 could not be determined. This could be explained by the fact that amount of H1.4 copies inserted were too little to show significant changes, as there was no success in overexpressing H1.4 to normal levels. Actually, HT29-H1.4 increased 3-fold the mRNA levels of H1.4 respect to HT29, but still they were 40 % below the levels in the normal colon tissue. Other possibilities such as the cell develop mechanisms to control overall histone H1 amount was also plausible. In this regard and as mentioned above, the mRNA levels of the exogenous H1.4 were 3-fold higher than the endogenous ones in HT29, whereas the protein levels were similar; suggesting that part of the H1.4-HA mRNA was post-transcriptionally regulated and/or that the histone protein turnover was accelerated. This goes in agreement with the observation made in budding yeast, where histone levels are post-translationally regulated by ubiquitylation-dependent proteolysis. Otherwise, excess of core histones resulted in cytotoxic effects via multiple mechanisms [171–173]. Histone H1 levels have also been shown to be tightly regulated in metazoans. For example, mechanisms of negative auto-regulation to avoid excess of histone H1 have been reported. In this case, the overexpression of the fused H1:GFP protein in *Drosophila* was limited and compensated by a decrease of the endogenous H1 form, so that about normal amounts of total H1 were obtained [174]. In this respect, dosage compensatory mechanisms to balance the increased H1.4 levels seemed not to be the case for our system, as the

proportion of each independent band (corresponding to specific H1 variants) did not change after increasing H1.4 levels.

Ariño et al. used the same lentiviral transduction approach to overexpress the histone H1 variants in human breast cancer cell lines. They also obtained HA-tagged H1 variant expression levels lower than or similar to their corresponding endogenous histone [86], confirming the difficulties in overexpressing histone H1 in huge amounts. Another important issue to consider is that the overexpression of H1.4-HA was driven by a constitutive promoter, the human elongation factor-1 alpha (EF1α), whereas the transcription of most of the histone H1 proteins, including H1.4, is subjected to a cell cycle control, increasing the possibilities that part of the H1.4 transcribed out of phase was being degraded.

Hence, one of the situations that could better explain the reason why no significant differences on γH2A.X reactivity were observed between HT29 and HT29-H1.4 cells is the low levels of histone H1 overexpression. However, we cannot discard the possibility that the contribution of histone H1 to the total amount of DNA damage in a cancer cell is little and differences in terms of genome stability and DNA damage are below the threshold to be detected. A part from this, it seemed that γH2A.X reactivity was increased in cells expressing the empty plasmid (HT29-pEV833) compared to HT29-H1.4 and untreated cells, which suggest that the procedure by itself induced damage. This could indicate that part of the recovery in HT29-H1.4 cells was underestimated by the damage produced by the method and that levels of DNA damage after histone overexpression in HT29 cells were lower than presented. Besides, the amount of plasmid introduced in HT29-H1.4 cells was higher than in HT29-pEV833 (as higher amount of GFP is detected by FACS; **Figure 41**), which strengthens the possibility that H1.4 was indeed recovering part of the damage but due to the collateral consequences of the technique, we were unable to detect a significant rescue of the phenotype. Finally, we also need to consider the possibility that the overexpressing approach used in our experiments was not the most appropriate for our objective. We choose a stable histone H1.4 overexpression, which could have allowed cells to develop mechanisms to overcome this overexpression over the time. Instead, a transient overexpression might have induced stronger temporally effects increasing our possibilities to detect them.

In this context and in the absence of any significant decrease on DNA damage, we did not follow the analysis of the contribution of histone H1 to genome stability in cancer cells. However, we believe that improving the technique and obtaining higher overexpression levels of histone H1, we could have been able to appreciate the importance of histone H1 in preventing DNA damage.

CONCLUSIONS

CONCLUSIONS

According to the results obtained in this work, the conclusions generated are the following:

1. *D. melanogaster* linker histone H1 (dH1) participates in the biology of RNA:DNA hybrids (R-loops). It avoids abnormal accumulation of R-loops essentially in heterochromatin.
2. In physiological conditions, R-loops are spread along the genome and do not importantly compromise genome integrity. However, upon depletion of dH1, transcription becomes highly up-regulated and deleterious R-loops accumulate in heterochromatin, which subsequently induce DNA damage observed by reactivity against γ-H2Av. Thus, dH1 is important for prevention of R-loop-mediated genome instability.
3. This dH1 role in R-loop dynamics is specific of this protein and not a mere consequence of heterochromatin deregulation. Depletion of other important heterochromatic proteins, such as HP1a, despite similarly up-regulating transcription of heterochromatic elements, it neither induces R-loop accumulation nor DNA damage.
4. Histone H1 by itself does not have any enzymatic activity able to resolve R-loops. However, it participates in the prevention of R-loops together with hnRNP36 and hnRNP48, two proteins involved in the metabolism of mRNAs.
5. Absence of hnRNP36 and hnRNP48 induce R-loop accumulation in heterochromatin to a similar extent of that obtained for histone H1 depletion. Also, depletion of dH1 reduces the content of hnRNP36 and hnRNP48 in heterochromatin, whereas levels of dH1 do not significantly change upon depletion of any of the hnRNPs. Thus, hnRNP36 and hnRNP48 binding to heterochromatin appears to be dependent on histone H1.
6. In normal conditions, R-loops cover 7.69% of *D. melanogaster*'s genome and are mainly localized at promoters, introns and distal intergenic regions. Upon depletion of histone H1, R-loop content is reduced (1.97%) and its distribution is modified, suggesting that histone H1 is involved somehow in the physiological distribution of RNA:DNA hybrids.

7. Transient R-loops are mainly formed during transcription in physiological conditions. However, genic R-loop accumulation does not correlate with transcription rates. Instead, strong accumulation is observed at promoter regions, suggesting that R-loops might be formed as a consequence of RNAPII pausing.
8. Besides, depletion of histone H1 might affect RNAPII dynamics, as R-loop accumulation in dH1-depleted cells is highly reduced.
9. Distal intergenic R-loops relate with Su(Hw), mdg4 and H3K27m3 and, upon depletion of histone H1, the R-loop content at distal intergenic sites increases.
10. Based on the results obtained for *D. melanogaster*, a correlation between levels of histone H1 and the intrinsic DNA damage that characterizes colorectal carcinoma cell lines was expected. However, after overexpressing H1.4 in HT29 colorectal carcinoma cell line, we were not able to determine any specific change over the genome stability of those cells. Further studies are needed to evaluate more accurately this relationship.

MATERIALS

MATERIALS

1. CELL LINES

To develop this thesis different cell lines have been used.

1. **Drosophila S2 cells** (SL2, Schneider 2; ATCC CRL-1963) The S2 cell line was derived from a primary culture of late stage (20-24 hours old) Drosophila melanogaster embryos [175].
2. **Human cell lines**: different colorectal carcinoma an endometrium cancer cell lines were obtained from Dr. Albert Jordan (IBMB-CSIC). They are listed as follows: HCT116, HCT15, HCT8, SW48, SW480, LS180, LST174T, Caco2 and SKUT1B.
3. **Other cells** also used: HEK 293T.

Samples proceeding from a normal colon tissue were obtained from Dr Eduard Batlle (IRB Barcelona) and a histone H1 rat liver sample from Dra Inma Ponte (UAB).

2. *Drosophila melanogaster* FLY STOCKS

The fly strains used in this study are listed and described in the following table.

NAME OF THE FLY STOCK	CHROMOSOME	REFERENCE
nubb-GAL4	II	From Dr Casali
nubb-GAL4; UAS-Dic2	II;III	From Dr Casali
GFP^{RNAi}	II	[176]
white	X	w^{1118}
dH1^{RNAi}/FM6;sb/TM3$_{SERR}$	X	NIG-FLY 31617R-3
HP1a^{RNAi}	II	VDRC 31995
Su(var)2-10^{RNAi}	III	VDRC 30710
hnRNP48^{RNAi}	II	VDRC 101555
hnRNP36^{RNAi}	II	VDRC 51759

Table 6. Fly strains used during this thesis. VDRC: Vienna Drosophila Resource Center; NIG-FLY: Fly stocks of National Institute of Genetics.

3. OLIGONUCLEOTIDES

The oligonucleotides used in the different experiments are presented in the following table:

NAME	SEQUENCE (5'-3')
Tubulin-F	ACCTGAACCGTCTGATTGGC
Tubulin-R	GCAGAGAGGCGGTAATCGAG
Mthl12-F	TGCAAAGCTATTTAATGCACGA
Mthl12-R	TAAAACTCGACTGGGGTTCC
DivLTR-F	CGCCAAAACTGTGCAGTAGA
DivLTR-R	CAGATAAATGCGTGCGAGAA
DivORF-F	CTCTTACGGTGAGGCTGGAG
DivORF-R	GCTGTCGATTATGCTCGTGA
GATELTR-F	CCGCTCTTCACCTCAGAGTC
GATELTR-R	CCGGGCGTATGTTTATTCAG
GATEORF-F	GCAGACCTGGCAAGTAGAGG
GATEORF-R	GAGCGTTGACCTGAGTAGGC
InvLTR-F	AGATGACAATGTGGCACACG
InvLTR-R	GATCGACGTCAGCAGTCAAA
InvORF-F	GCTTACGCCTTCAAGAAACG
InvORF-R	CAAAAATGGCACATGGTCTG
PogoIR-F	CATCGGCAAGATATCTGCATT
PogoIR-R	CGATGCAGCAAACGTATGAA
PogoORF-F	TACATTTGGTTCGGACAGCA
PogoORF-R	ACGTGCCGGTCAAGAATTAC
3S18LTR-F	CAGCGGAATCAATGTAAGCA
3S18LTR-R	TGGAAAAGTACTGGGCAAGC
3S18ORF-F	GCCATCAATCGCTTCTTCTC
3S18ORF-R	CCAGGAAAGCTTCGTACTGC
Het-AUTR-F	TTCGCTTGCCAAAGACTCTC
Het-AUTR-R	GCTTTTCTTTGCAGCCTGAG
Het-AORF-F	AAACGACGATCTGGACTGCT

Het-AORF-R	CGGAAAAATGCTGGGAGTTA
satDNA-F	AAACACGTCTCCACCCGAAG
satDNA-R	CTATTCTAACATTCGGCATTCCAC
GYPSYLTR-F	GGCTCATTGCCGTTAAACAT
GYPSYLTR-R	GGCGATAGCGATTTGATTGT
GYPSYORF-F	CCTCAGAGCTGTGGTCTTCC
GYPSYORF-R	CAGATGGCAGGTCTTTTGGT
DOCUTR-F	GACATTCGGCATTCCACAGT
DOCUTR-R	ACGTCTCCACCCGAAGACT
DOCORF-F	CGCTGTGCCAGCTGTAAATA
DOCORF-R	ATTGTTGTTGCAAACGGTCA
MDG3LTR-F	TCAGTCGCTGTTGAACCAAG
MDG3LTR-R	TTAGCCGCCGTTTACAGAAG
MDG3ORF-F	AAATGCAAAAAGGCCAAATG
MDG3ORF-R	AGCTAAACGGTTTCGGGTTT
ACCORDLTR-F	TAGGCGACATCAGCAAAGTG
ACCORDLTR-R	ATCGGGTGCAACAGAGTTTC
ACCORDORF-F	CCAACAGCAACAACATGGAC
ACCORDORF-R	AAAAGCCAAAATGTCGGTTG
DMCR1AORF-F	GTTGTGATGCTTGCCTTGTG
DMCR1AORF-R	ATTTCATCTCGTTCGCAACC
TARTORF-F	CCAATGCAACCAAAGCATTA
TARTORF-R	TATGTGTGGGAGGGAGAAGC
T7dH1-F	ATTGTAATACGACTCACTATAGATGTCTGATTCTGCAGTTGC
T7dH1-R	ATTGTAATACGACTCACTATAGTTACTTTTTGGCAGCCGTAG
T7HP1a-F	TAATACGACTCACTATAGGGAGAATCCCGAAACTGAGAACACG
T7HP1a-R	TAATACGACTCACTATAGGGAGATCCGATGCCTTAAGAGTTGG
T7hnRNP36-F	TAATACGACTCACTATAGGGAGACCAACGGAAACTACGACGAT
T7hnRNP36-R	TAATACGACTCACTATAGGGAGATGCTTGGCAATAGCCTTCTT
T7hnRNP48-F	TAATACGACTCACTATAGGGAGAACGAGAGGGGCAAACTTTTT
T7hnRNP36-R	TAATACGACTCACTATAGGGAGAGCGGGACTTCTTCTTCTCCT
T7LacZ-F	TAATACGACTCACTATAGGGATGACCATGATTACGCCAAGC

T7LacZ-R	TAATACGACTCACTATAGGGCAATTTCCATTCGCCATTCAG
Act-F	CGTCCACCATGAAGATTAAGATTGT
Act-R	CAATACTTTTGACTCCCATCCTTTG
Su(var)2-10-F	TGACTGAGGATGCTGACTGC
Su(var)2-10-R	GTGGGCTTACGCTCATTCAT
GAPDH-F	GAGTCAACGGATTTTGGTCGT
GAPDH-R	TTGATTTTGGAGGGATCTCG
H1.0-F	CCTGCGGCCAAGCCCAAGCG
H1.0-R	AACTTGATCTGCGAGTCAGC
H1.1-F	CTCCTCTAAGGAGCGTGGTG
H1.1-R	GAGGACGCCTTCTTGTTGAG
H1.2-F	GGCTGGGGGTACGCCT
H1.2-R	TTAGGTTTGGTTCCGCCC
H1.3-F	CTGCTCCACTTGCTCCTACC
H1.3-R	GCAAGCGCTTTCTTAAGC
H1.4-F	GTCGGGTTCCTTCAAACTCA
H1.4-R	CTTCTTCGCCTTCTTTGGG
H1.5-F	CATTAAGCTGGGCCTCAAGA
H1.5-R	TCACTGCCTTTTTCGCCCC
H1.X-F	TTCCTTCAAGCTCAACCG
H1.X-R	TGCCTTCTTCGCTTTGTG

Table 7. Primers used for qPCR experiments and dsRNA preparation. Sequences are presented in 5'-3' direction. T7 sequence is underlined in primers for dsRNA preparation.

4. ANTIBODIES

Primary antibodies used for the different experiments are presented here:

NAME	SPECIE	CLONALITY	EXPERIMENT	SOURCE
dH1	Rabbit	Polyclonal	WB: 1/10,000 IF Polytene: 1/4,000	[2]

			ChIP: 1 µl	
HP1a	Rat	Polyclonal	WB: 1/10,000 IF Polytene: 1/500 ChIP: 1 µl	[177]
S9.6	Mouse	Monoclonal	DRIP: 4 µl IF Polytene: 1/500 IF S2 cells: 1/300	[178]
hnRNP36	Mouse	Monoclonal	WB: 1/400 IF Polytene: 1/200 IP: 10 µl	From Dr Saumweber [179]
hnRNP48	Rabbit	Polyclonal	WB: 1/1,000 IF Polytene: 1/100 IP: 10 µl	From Dr Johnston[180]
γH2Av	Rabbit	Polyclonal	IF S2 cells: 1/500	Rockland (600-401-914)
γH2A.X	Rabbit		WB: 1/500 IF human cells: 1/200	Santa Cruz (sc101696)
H4	Rabbit	Polyclonal	WB: 1/2,000	Abcam (ab10158)
H1.4pT146	Rabbit	Polyclonal	WB: 1/1,000 IF human cells: 1/500	Abcam (ab3596)
HA	Mouse	Monoclonal	WB: 1/500 IF human cells: 1/100 IP: 4 µl	Roche (11-583-816-001)
GFP	Rabbit	Polyclonal	IP: 4 µl	Roche (11-814-460-001)

Table 8. Name, specie, clonality, dilution for each experiment and source are shown for all the primary antibodies used in this thesis. Abbreviations used: WB, Western Blot; IF, Immunofluorescence; IP, Immunoprecipitation; ChIP, Chromatin Immunoprecipitation; DRIP, DNA:RNA Immunoprecipitation.

Commercial secondary antibodies used for immunofluorescence were either coupled to cyanines (Jackson) or to Alexa fluorophores (ThermoFisher), and to horseradish peroxidase (HRP) conjugated for WB (Jackson).

For ChIP experiments, Protein A Sepharose (GE Healthcare) was used to bind dH1 antibody and Protein G Agarose (Santa Cruz) to bind HP1a antibody.

5. PLASMIDS

In order to produce the lentiviral particles, three plasmids obtained from Dr. Albert Jordan (IBMB-CSIC) were used: pVSV.G, pCMV-dR8.91 and pEV833.GFP.

6. PUBLIC DATA

For the DRIP-seq analysis, several high-throughput data were used:

Expression profiling array: GSE 49103
GRO-seq: GSE 23543
RNAPII (Rpb3) ChIP-seq: GSE 2354
2RNAPII phosphorylated on Serine 5 (RNAPIISer5) ChIP-seq: GSM 796331

METHODS

METHODS

1. MANIPULATION OF CELLS:

1.1. Culturing cells

Human cancer cells were grown at 37°C and 5% CO_2 in Dubecco's modified Eagle's medium (DMEM, Invitrogen) supplemented with 10% fetal bovine serum (FBS, Gibco) and 1% penicillin-streptomycin (Gibco). Every 48-72 h, cells were split at a 1:12 dilution. To dilute them, they were washed with pre-warmed PBS (10 mM Na_2HPO_4, 10 mM NaH_2PO_4, 0.14 M NaCl), spun down (5 min at 1,000 rpm at room temperature) and trypsinized (0.25% Trypsine-0.02% EDTA) for 1 min at 37°C. To stop the reaction, 2 volumes of pre-warmed medium were added. Then, the corresponding number of cells was transferred to a new plate with fresh medium. Human cells were seeded in 100 mm plates (Corning).

S2 cells do not need trypsinization as they grow semi-adherent to the surface of the flask. Therefore, cells were collected by pipetting and transferred to a new flask with conditioned media (Schneider's Insect Medium (Sigma); 10% heat-inactivated FBS, 1% penicillin-streptomycin) in a 1:5 dilution. Cells were cultured at 25°C and passed every 3-4 days. Generally, S2 cells were seeded in T25 and T75 flasks (Corning).

1.2. Freezing and thawing

To keep cell storage and use cells to a maximum of 30 passages, they were continuously frozen and thawed.

To freeze cells, they were washed with PBS and pelleted by centrifuging (1,000 rpm for human cells and 500 rpm for S2 cells) for 5 minutes at 4°C. Cells were resuspended in Freezing Medium (conditioned medium with 10% DMSO) and aliquoted in 1 mL of the cell suspension per vial. Cells were frozen at -20°C for 2 h and at -80°C for 24 h in a polystyrene container, and transferred to liquid nitrogen for long-term storage.

To thaw cells, they were transferred to 15 mL tube with 5 mL conditioned medium. Cells were centrifuged, washed twice and resuspended in conditioned medium.

1.3. Cell counting

To calculate the cell density, 20 µL cells in suspension and 180 µL medium were mixed and poured through the side grooves of the Neubauer chamber previously assembled with the cover slide. Then total cells were counted for 20 squares (4 groups), the numbers were averaged and multiplied by 10^4 (volume in the Neubauer's chamber) and by the dilution factor (10).

1.4. Lentiviral particle production and cell transduction

The protocol for virus production and infection was adapted from [147].

1st day: HEK 293T is a good host cell line as it is easily transfected and supports high-level expression of viral proteins. Thus, HEK 293T cells were cultured at 40% confluence (1.5-3 x 10^6) in a 10 cm plate.

2nd day: For the production of viral particles containing the gene of interest (histone H1 variants), HEK 293T cells were co-transfected with plasmids pVSV.G, pCMV-dR8.91 and pEV833.GFP using calcium phosphate. pVSV.G encodes for the viral envelope and provides a wide range infectivity. pCMV-dR8.91 contains the packaging genes (gag and pol) needed for lentivirus production. pEV833.GFP is a polycistronic vector that encodes for the transgene (H1.2-HA or H1.4-HA) and the GFP reporter gene (**Figure 38**). Hence, to produce the precipitate $CaCl_2$/HBS/DNA, a tube with 250 µL sterile-filtered water, 5 µg pVSV.G, 15 µg pCMV-dR8.91 and 10 µg pEV833.GFP was prepared and 250 µL 0.5 M $CaCl_2$ were added. Then, 500 µL 2x HBS were added at dropwise with sight agitation and incubated for 30 min at room temperature without stirring. The precipitate was added dropwise to whole 30 cm ∅ plate and shake gently to distribute it. It was incubated at 37°C and 5% CO_2 for 6-10 h before washing with PBS and changing medium to continue the incubation.

4th day: On one side, the medium was collected and saved at 4°C and added fresh to the cells. On the other side, HT29 cells were seeded in 6 wells plate.

5th day: The medium was collected and added together with the one from previous day, and filtered with a 0.45 µm syringe filter. Then, it

was carefully added to a ultracentrifuge tube (Beckman Coulter) previously cushioned with 4 mL of 20% sucrose and centrifuged in at 26K for 1.30 h at 4°C. The supernatant was discarded and the pellet, containing the viral particles, was resuspended in 1 mL conditioned medium.

The viral concentration (multiplicity of infection (MOI)) required for an optimal percentage of successfully infected HT29 cells was determined by several infections followed by FACS (Cytomics FC500, Beckman Coulter). Results suggested that the best concentration was a MOI of 800. Thus, HT29 cells were infected (800 µL (MOI 800)) for spinoculation (2h, 1,200 rpm, room temperature). The following days the culture was grown at 37°C and 5% CO_2 in complete DMEM. Once grown enough, the infected cells were selected by sorting (Aria SORP, Becton Dickinson) using the GFP as a selection marker for cells containing the transgene. To increase the number of expressed transgenes, infected cells were re-infected up to three times.

2. MANIPULATION OF FLIES:

The work with flies was done according to the standards [181].

D. melanogaster exhibits a complete metamorphism: egg, larva, pupa and adult. Experiments using *D. melanogaster* flies were done from dissecting 3rd instar larvae (there are 3 stages of larval development; first to third) or analyzing adult fly phenotypes.

Briefly, files were slept by carbon dioxide before manipulating them. To make genetic crosses properly, virgin females were isolated from males. Although virgin females are able to lay eggs, they are not fertilized and do not progress to embryos. Once a female has been crossed with a male, it can lay up to 100 eggs/day. The development of *D. melanogaster* varies with temperature; it takes roughly 10 days to develop and emerge as an adult fly at 25°C, whereas at lower temperatures (18°C), flies take about twice as long to develop. Thus, unless stated otherwise, all the *D. melanogaster* strains were raised at 25°C under standard conditions. Flies were transferred from one vial to another every 14 days to maintain stocks, and were kept at 18°C while not working with them. All flies were kept homozygous or balanced to avoid recombinatorial events. Balancers were linked to dominant markers for easily recognition.

Overexpression and RNA interference (RNAi) experiments were carried out using GAL4/UAS system. The *nubbin*-GAL4 (*nub*GAL4) construct was used to direct the transgene expression to salivary glands and wings.

3. MOLECULAR BIOLOGY METHODS:

3.1. Polymerase chain reaction (PCR)

To perform the reaction, the following master mix was prepared for 20 µL reaction: 2.5 µL Reaction Buffer 10X, 2 µL 2.5 mM dNTPs, 1 µL 10 µM each primer, 1 µL 50 mM $MgCl_2$ solution, 0.5 TAQ, template (200 ng/µL DNA or 15 ng/µL plasmid) and water.

3.2. DNA agarose gel electrophoresis

1% agarose gel electrophoresis was used for staining and visualization of fragmented DNA from ChIP and DRIP (to evaluate sonication efficiency), DNA-T7 (to evaluate DNA amplification) and dsRNA (to evaluate MEGAscript reaction).

Briefly, agarose powder was dissolved in TBE 1X (9 mM Tris base, 2.4 mM EDTA pH 8.0, 0.08 M Boric acid) in the oven, 0.5 µg/µL of ethidium bromide was added and the agarose was poured into a gel tray with the well comb in place. Once solidified, the gel was placed in the box, covered with TBE 1X, and the DNA mixed with Orange G 5X dye (50% Glycerol, 20 mM EDTA, 0.02% Orange G (Sigma)) to a 1X concentration was loaded together with a DNA molecular weight ladder. The gel was run at 80-100 V until the dye was approximately 80% of the way down the gel. To visualize DNA fragments, a UV light device was used.

3.3. Double-strand RNA synthesis

The coding sequence of the interfered gene was first amplified with primers containing the gene sequence plus the sequence of a T7 promoter (for primers used see Materials). dsRNA was then produced using a MEGAscript T7 kit (Ambion) and the RNA was purified with the RNeasy Mini Kit (Qiagen). To set up the MEGAscript reaction (total 20 µl) the following components were mixed in this order: 8 µL NTP's (2 µL of each), 2 µL 10X Buffer, 2 µL Enzyme mix, 4 µL PCR product flanked with T7 and 4 µL H_2O (RNAse free). Then the reaction was incubated at 37°C overnight. 1

µL TURBO DNase was added and incubated for 15 min at 37°C. From here on, RNeasy Mini Kit (Qiagen) was used according the manufacturer's recommended protocol. The volume was increased to 100 µL with RNase free H_2O and 350 µL Buffer RLT and 250 µL of 100% ethanol were added and mixed by pipetting. The sample (700 µL) was transferred to a RNeasy spin column and spun down for 15 s at full-speed. Then, it was washed with 500 µL Buffer RPE and spun down for 15 s at full-speed twice, and centrifuged an extra min at full-speed without any buffer. Finally, the column was placed into a fresh collection tube and 30 µL RNase-free water were added to the column membrane to elute the RNA. The tube was spun down for 1 min at full-speed and it was repeated using the same 30 µL of RNAse-free water to increase the amount of eluted RNA. The RNA concentration was measured using nanodrop and an aliquot was saved to be checked on an agarose gel before keeping the dsRNA at -20°C.

The following table shows the amplification programs used depending on the sequence to be amplified.

T7LZ, T7dH1	**T7HP1a, T7hnRNP36, T7hnRNP48**
Denaturation: 95°C – 15 s	Denaturation: 95°C – 15 s
Annealing: 65°C – 30 s	Annealing: 55°C – 30 s
Extension: 72°C – 45 s	Extension: 72°C – 45 s

3.4. dsRNA treatment of S2 cells

On the first day, $2.5x10^6$ S2 cells were suspended in 2 mL of pre-warmed serum-free medium and seeded into a 5-mL plate. dsRNA (100 µg for $dsRNA^{dH1}$, 30 µg for $dsRNA^{HP1a}$, $dsRNA^{hnRNP36}$ and $dsRNA^{hnRNP48}$, and 30/100/130 µg for $dsRNA^{LacZ}$) was gently added and incubated for 1h at 25°C. After, 3 mL 5X FBS-medium were added and the flasks were gently shacked and incubated for the following days depending on the depletion:

The corresponding doses were added on day 1 and 3 for $siRNA^{lacZ}$, $siRNA^{hnRNP36}$, $siRNA^{hnRNP48}$ and on day 1, 4 and 7 for $siRNA^{dH1}$ and $siRNA^{lacZ}$. On those same days and before adding the dsRNA, the medium was removed and 5mL of complete medium containing the specific amount of dsRNA was added. At day 7, cells were split into two flasks.

3.5. Total RNA extraction and purification

For RNA extraction, a combined protocol using TRIzol® reagent (Invitrogen) and the RNeasy Mini Kit (Qiagen) was used.

1. TRIzol: $1x10^6$ cells were centrifuged at 2000 x g for 5 min, and without washing with PBS, the pellet was dissolved in 500 µL TRIzol by vortexing and incubated for 5 min at room temperature. After addition of 100 µL of chloroform, samples were vigorously shaked and placed on the bechtop for 3 min. Then they were centrifuged for 15 min at 4ºC at maximum speed and the upper aqueous phase was recovered.

2. RNeasy Mini Kit: the manufacturer's guidelines were followed. Briefly, 1 volume of 70% ethanol was added to the aqueous phase and transferred to a RNeasy spin column. It was centrifuged for 15 s at 8,000 x g (if >700 µL, this centrifugation step was repeated). And the column was washed adding 350 µL Buffer RW1 and centrifuged for 15 s at 8,000 x g. A DNase I treatment (10 µL DNase I and 70 µL Buffer RDD) was performed directly to the column membrane for 15 min at room temperature and washed once with Buffer RW1 and twice with RPE. The second RPE wash was centrifuged for 2 min and an additional min at full speed was done to eliminate any carryover. The column was placed into a fresh collection tube and 30 µL RNase-free water were added to the column membrane to elute the RNA. The elution was repeated once again using the same 30 µL of RNAse-free water to increase the amount of eluted RNA. A small aliquot was separated for measuring RNA concentration at nanodrop and the sample was immediately frozen at -80 °C if not directly retrostranscribed (preferred).

3.6. RNA reverse transcription (RT)-qPCR

1. RETROTRANSCRIPTION OF mRNA INTO cDNA: Transcription First Strand cDNA Synthesis Kit (Roche) was used following the manufacturer's recommended protocol. Briefly, the starting reaction mix was set up: 1 µg RNA, either 1 µL oligo-dT (for genes with poly(A) tails) or 2 µL random hexamer primers and RNAse-free H_2O up to 13 µL, and incubated at 65°C for 10 min. The following

components were immediately added to the mix in this order: 4 µL Transcription Reverse Transcriptase Reaction Buffer 5X, 0.5 µL Protector RNase Inhibitor, 2 µL Deoxynucleotide Mix [10mM each], 0.5 µL Transcriptor Reverse Transcriptase. The reaction was mixed gently and incubated with the following program: 50°C for 1 h, 85°C for 5 min, pause at 4°C. Then the retrotranscribed products were diluted 1/10 and stored at -20°C or continued directly with qPCR.

2. QUANTITATIVE (q)-PCR: Given a final reaction of 10 µL per well, the following primer/SYBR Green master mix was prepared for each specific primer: 0.5 µL fw primer [10 µM], 0.5 µL rv primer [10 µM], 5 µL 2X SYBR Green I Master (Roche) and 2 µL water. Then, 2 µL of cDNA and 8 µL of the mix were added in a 96-well plate. The following standard PCR program was used in a Roche LightCycler 480 Instrument: 5 min at 95°C and 45 cycles of 10 s at 95°C, 10 s at 60°C and 10 s at 72°C. Quantitative determination of RNA levels was performed in triplicate and an additional RT reaction was performed as negative control, in which no retrotranscriptase was added, to confirm the absence of genomic DNA contamination. In all cases, values were normalized to the expression of housekeeping genes and relative expression levels were calculated using the standard curve method.

3.7. Phenol:chloroform extraction and ethanol precipitation

DNA was precipitated out of solution for the removal of salts and/or for resuspension in an alternative buffer using a phenol:chloroform extraction followed by an ethanol precipitation in all the cases. Thus, a volume of phenol:chloroform was added to the sample and vigorously vortexed. Then, it was centrifuged to maximum speed for 5 min. The aqueous (upper) phase was carefully removed and transferred to a new tube. To precipitate the DNA, 1/10 of 4M NaCl, 3 volumes of 100% ethanol and 7 µL of 0.25% soluble polyacrylamide were added, vortexed and left precipitating overnight at -20˚C. The pellet was washed in cold 70% ethanol and after a further centrifugation step the ethanol was removed, and the nucleic acid pellet was allowed to dry before being resuspended.

3.8. Sodium Dodecyl Sulfate Polyacrylamide (SDS-PAGE) gel electrophoresis

S2 cells, salivary glands or extract of histone proteins were mixed with PLB 5X (Protein Loading Buffer: 5% SDS, 21.75% Glycerol, 125 mM Tris-HCl pH6.8, 0.25% bromophenol blue) to 1X and β-mercaptoethanol (to a final concentration of 1.5 M), and boiled at 95˚C for 5 min. Samples were loaded in SDS-PAGE along with a molecular weight marker, and the gel was placed inside the electrophorator (Bio-Rad) covered with Laemmli Buffer (0.1% SDS, 200 mM Glycine, 25 mM Tris-HCl pH 8.5) to run until proper protein separation at 25 mA and unlimited voltage. 15% polyacrylamide-SDS gels were used in all the experiments except for the immunoprecipitation, where 10% was used instead.

Separating gel solution

	10%	**15%**
Acrylamide/bis-acrylamide 29:1 (40%)	2.5 mL	3.75 mL
1.5 M Tris-HCl pH8.7	2.7 mL	2.7 mL
10% SDS	0.1 mL	0.1 mL
water	4.075 mL	3.45 mL

Stacking gel solution

Acrylamide/bis-acrylamide 29:1 (40%)	**0.5 ml**
1 M Tris-HCl pH6.8	0.625 mL
10% SDS	0.05 mL
water	3.875 mL

To catalyze the polymerization of acrylamide and bis-acrylamide, 16 µL TEMED and 60 µL ammonium persulfate (APS) were added for the separating gel and 20 µL TEMED and 20 µL APS for the stacking gel.

3.9. Western blot (WB)

Western blots were performed according to standard procedures. After proteins were separated by size, they were transferred from to a nitrocellulose membrane (GE Healthcare Life Sciences) in a wet transfer system with Transfer Buffer (25 mM Tris-HCl, 40 mM Glycine, 0.05% SDS,

20% methanol) at 100 V for 1 h. Then, membranes were blocked with 5% powdered skimmed-milk diluted in PBS-0.1% Tween (PBS-T) at room temperature for 1 h or overnight at 4˚C, indistinctly. Incubation with primary antibodies was performed for either 1 h at room temperature or overnight at 4˚C, indistinctly. The membrane was washed 3 times for 5 min with PBS-T and incubated with peroxidase-conjugated secondary antibody for 1 h at room temperature. The concentration used of the different primary antibodies varied between them (see Materials) and for secondary antibodies was 1:10,000. They were diluted in PBS-T. Finally, the membrane was washed 5 times with PBS-T for 5 min and bound antibodies were detected by the ECL chemiluminescence assay (Amersham) and exposed to autoradiographic films.

For antibodies recognizing a phosphorylation, BSA and TBS-T were used instead of skimmed milk and PBS-T, respectively.

3.10. Coomassie staining of SDS-PAGE gels

After electrophoresis, the SDS-PAGE gel was stained with Coomassie Brilliant Blue R250 (CBB) staining solution (0.1% CBB, 10% acetic acid, 30% methanol) overnight to ensure completely staining. Then, it was destained with 10% acetic acid. Then the gel was digitalized using a GS-800 Calibrated Densitometer (Bio-Rad) laser densitometer to be quantified.

3.11. Genomic DNA purification for DRIP

S2 cells were washed with PBS (10 min at 500 x g at 4˚C), resuspended in 200 µL of Lysis Buffer (Tris-HCl 10 mM pH7.4, EDTA 10 mM, NaCl 10 mM, SDS 0.5%, proteinase K 0.5 µg/µL) and incubated for 1h at 50˚C. Then the DNA was cleaned by consecutive phenol:chloroform extractions until obtain a clean interphase, followed by ethanol precipitation. Then, it was centrifuged at maximum speed for 10 min and the pellet was treated with RNAse A (10 mM Tris pH7.5, 0.5 M NaCl, 6 µg/mL RNase A) for 45min at 37˚C. DNA was extracted with phenol:chloroform and ethanol precipitated. DNA was resuspended in 50 µL of ultrapure water and the concentration was measured by nanodrop. It was kept at -20˚C.

3.12. DNA:RNA hybrid immunoprecipitation (DRIP)

The DRIP method uses the S9.6 antibody to capture RNA:DNA hybrids in their native context, followed by mapping the enriched DNA fragments on a selected number of loci (qPCR) or across the whole genome (deep sequencing). DRIP was adapted from previously described protocols [107,112].

1. SONICATION: 50 µg of genomic DNA were diluted with 200 µL of TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA) and was sonicated for 25 cycles at high power (30 s on-30 s off) using a Bioruptor sonication device (Diagenode). 8 µL of sonicated DNA were run in an agarose gel to check for the fragmented sizes (optimal fragmentation: 300-500 bp).

2. IMMUNOPRECIPITATION: 15 µg of sheared material were equally distributed into 3 different 1.5-mL tubes (5 µg each), called INPUT, RNH- and RNH+. 2U/µL of RNase H (Invitrogen) were added to the RNH+ tube. The two other tubes received TE buffer instead. Then, 5 µL of RNase H 10X buffer (200 mM Tris-HCl pH7.5, 1 M KCl, 100 mM $MgCl_2$, 1mM DTT, 50X sucrose) were added to all three tubes and the corresponding TE to obtain a final volume of 50 µL. The samples were incubated at 37˚C overnight. Next day, INPUT was kept at -20˚C and 500 µL of Binding Buffer (10mM $NaPO_4$, 140 mM NaCl, 0.05% Triton X100) were added to both RNH+ and RNH-, together with 4 µg of S9.6 antibody. Samples were incubated overnight at 4˚C on a rotatory wheel. Before continuing with the immunoprecipitation, 350 µL of Dynabeads-Protein G (Thermo Fisher Scientific) were prepared. They were resuspended in 1 mL of 0.5% BSA-Binding Buffer and incubated for 10 minutes at 4˚C with rotation. This step was repeated 3 times before resuspending them in 300 µL of Binding Buffer. 35 µL of blocked beads were added to the RNH+ and RNH- and incubated for 2h at 4˚C with rotation. Then, beads were washed 3 times with Binding Buffer for 10min at 4˚C with rotation.

3. DNA PURIFICATION AND PRECIPITATION: The beads were resuspended in 100 µL Elution Buffer (Tris-HCl 50mM pH 8.0, EDTA 10mM, SDS 0.5%) and vortexed at maximum for 1 min. The supernatant was separated and the same step was repeated twice

more to obtain 300 µL. To prepare the input, 5 µL of the initial INPUT material (10% input) were added to 295 µL of Binding Buffer to obtain a final volume of 300 µL. The 3 samples (input, RNH+ and RNH-) were treated with 7 µL of proteinase K (10 mg/ml) for 45 min at 55˚C. Samples were purified by phenol:chloroform followed with overnight ethanol precipitation at -20˚C. Cleaned pellets were resuspended in 50 µL of ultrapure water and used for qPCR or sequencing.

For the samples analyzed by qPCR, the percentage of immunoprecipitated material was calculated by the ΔΔCt method. For DRIP-seq analyses, the method used is explained in Chapter 4.1.

3.13. Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation experiments were mainly performed as described in [34]. Approximately, 1 x 10^8 cells were used for each chromatin preparation.

1. CHROMATIN PREPARATION: Cells were cross-linked by adding formaldehyde drop-wise directly to the media for a final concentration of 1.8% for 10 min at room temperature on the shaker. The reaction was stopped by adding Glycine at the final concentration of 0.125 M and incubated for 5 min. Then, the cross-linked cells were transferred into 50 mL tubes and spun at 1,500 g at 4˚C for 3 min. The pellet was resuspeded with 5 mL PBS and transferred into a 15-mL tube. Cells were spun at 1500 x g for 3 min at 4˚C and washed with Buffer A (10 mM Hepes pH 7.9, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X100) and Buffer B (10 mM Hepes pH 7.9, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.01% Triton X100) for 10 min each. Later, the pellet was resuspended in 4.5 mL TE and complemented with 0.5 mL 10% SDS. The tube was inverted 5 times and spun down at 1,500 x g for 3 min at 4˚C. The upper phase was carefully removed and the last step was repeated for two more times with 5 mL TE. Then, TE 1X, SDS 1X and 1mM PMSF were added to a final volume of 4 ml.

2. SONICATION: Chromatin was sheared with a Bioruptor sonication device (Diagenode) to an average fragment size of 300-500 bp (high

power; 25 cycles of 30 s on / 30 s off). The sonicated material was later complemented with 1% Triton X100, 0,1% NaDOC, 140mM NaCl, was incubated for 10 min with rotation at 4˚C and spun down at full-speed at 4˚C for 5min to remove aggregates. Soluble chromatin in the supernatant was divided in aliquots of 500 µL and kept at -80˚C. Sonication efficiency was checked by reverting the crosslinking of 100 µL sheared chromatin with 1% SDS and 0.1 M $NaHCO_3$, and incubated overnight at 65˚C. Then, the DNA was cleaned by phenol:chloroform extraction and ethanol precipitation and was treated with 0.5 µL RNAse A (10 mg/ml) at 37˚C for 20 min before being analyzed on 1 % agarose gel.

3. CHROMATIN IMMUNOPRECIPITATION: Before the chromatin was immunoprecipitated, Protein A sepharose beads (GE Healthcare Life Sciences) or Protein G agarose beads (Santa Cruz) (depending on the species and the Ig subclass of the antibody used; see Materials) were washed with RIPA (140 mM NaCl, 10 mM Tris-HCl pH8.0, 1 mM EDTA, 1 % Triton X100, 0.1% SDS, 0.1% NaDOC) and shortly blocked with RIPA-1% BSA. Then, 30 µL beads as a 50% suspension in RIPA were added to 500 µL of sheared chromatin and incubated at 4˚C for 1 h with rotation. The sample was spun down at 3000 rpm for 2 min at 4˚C and the supernatant (pre-cleared chromatin) was transferred to a fresh tube. The specific antibody was added and incubated overnight at 4˚C with rotation. Next day, 40 µL of overnight blocked and washed beads were added and incubated with rotation for 3 h at 4˚C. Immunocomplexes were then sequentially washed for 5 min with rotation at 4ºC with RIPA (5 times), LiCl ChIP buffer (250 mM LiCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5% NP-40, 0.5% NaDOC) and TE (2 times).

4. ELUTION: The input (50 µL cross-linked chromatin) was included at this stage and processed in the same way. 40 µL TE and 0.5 µL RNAse A (10 mg/ml) were added to the beads and incubated at 37˚C for 30 min. Then, 50 µL 0.2M $NaHCO_3$ and 10 µL 10% SDS were added and vortexed for 30 s. The beads were spun down at full-speed for 30 s and the supernatant was separated. 100 µL Elution Buffer (1% SDS, 100 mM $NaHCO_3$) were added to the beads and vortexed for 30 s. The supernatant was combined with the previous

one and the step was repeated once more to have a final volume of 300 µL. For the input, 200 µL Elution Buffer were added to equal the volume with the other samples. The supernatants were incubated overnight at 65˚C in order to de-crosslink. Then, all three were treated with 3 µL Proteinase K (10 mg/ml) for 3 h at 55˚C and finally extracted with phenol:chloroform and precipitated with ethanol overnight at -20˚C. The pellet was dissolved in 25 µL ultrapure water.

For the qPCR, the percentage of enrichment was calculated by the ΔΔCt method.

3.14. Immunoprecipitation (IP)

1. TOTAL CELL EXTRACT: Total S2 cell extracts for IP were prepared from 2 T75 flasks of S2 cells (80% confluent). Cells were pelleted and washed with PBS three times. Then, the pellet was resuspended in 1 mL IP Buffer (Tris pH8 50mM, NaCl 150mM (150-200mM), EDTA 5mM, NP40 0,5%, PMSF 0,1mM, Protease Inhibitor cocktail (Sigma-Aldrich)) and incubated on ice for 30 min. A Dounce homogenizer with small clearance B pestle was used 25-30 times to disrupt the membranes but maintaining the nuclei intact. Then, NaCl was added to a final concentration of 300 mM and was incubated at 4˚C for 30 min with rotation. The lysate was centrifuged at 14,000 rpm for 15 min at 4˚C. The supernatant (soluble protein extract) was separated in aliquots of 500 µl.

2. IMMUNOPRECIPITATION: 50 Lof protein extract were separated, mixed with PLB-β-mercaptoethanol and kept at -20˚C (input). Then, two extracts of 500 µL were processed in parallel. For the pre-washing, a 30 µL of Protein A sepharose or Protein G agarose beads (depending on the species and the Ig subclass of the antibody used; see Materials) as a 50% suspension in IP Buffer were added and incubated for 1h at 4˚C with rotation. The sample was centrifuged at 3,000 rpm for 2 min and the supernatant was transferred to a new tube. In one tube, the antibody of interest was added, and in the other one, an unrelated antibody. They were incubated overnight at 4˚C with rotation. Protein A sepharose or Protein G agarose beads were blocked with BSA-IP Buffer overnight with rotation. Next day,

beads were washed with RIPA IP Buffer twice and 50 µL of beads as a 50% suspension in IP Buffer were added to the samples and incubated for 2 h at 4˚C with rotation. The beads were washed 5 times with RIPA buffer for 5 min with rotation and then spun down at 2,000 rpm for 2min at 4˚C. The beads were finally resuspended in 20 µL PLB-β-mercaptoethanol. Before loading the sample onto the 10% polyacrylamide gel, it was vortexed and boiled for 5 min at 95˚C, and then spun down as only the supernatant was used. Also 1-5% of input sample was loaded.

3.15. Hydrochloric acid extraction of histones

Cells collected from 80% confluent 100 mm ∅ plate were washed twice with PBS and resuspended in 500 µL of cold Hypotonic Lysis Buffer (10 mM Hepes pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 1.5 mM PMSF, 0.5 mM DTT , 0.05% NP40). Then they were incubated for 10 min on ice and spun for other 10 min at 3,000 rpm at 4˚C. The supernatant was discarded and the pellet was resuspended in 0.2 N HCl in a density of 4 x 10^7 cells/mL and left at 4˚C overnight with rotation. Next day, it was spun at 13,000 rpm for 10 min at 4˚C, and the supernatant was transferred to a 2-mL tube. 6 volumes (750 µL) of ice-cold acetone were added and the tube was inverted 6 times before incubating it on dry ice for 30 min. Then, the sample was centrifuged at 13,000 rpm for 10 min at 4˚C and the supernatant was removed. 1 mL of ice-cold 0.1 N HCl-acetone was added, the tube was inverted 6 times and it was incubated on dry ice for 30 min. Again, it was centrifuged at 13,000 rpm for 10 min at 4˚C and the supernatant was discarded. The pellet was washed with 1 mL of ice-cold acetone, centrifuged at 13,000 rpm for 10 min at 4˚C and the supernatant was discarded. The pellet was vacuum-dried and histones were then carefully solubilized with 200 µL of ultrapure water.

3.16. Perchloric acid extraction of histone H1

Cells collected from 80% confluent 100 mm ∅ were washed twice with PBS and resuspended in 300 µL of ice-cold 5% perchloric acid. Cells were immediately homogenized using a plastic pestle and incubated for 20 min at 4˚C with rotation. Later, they were centrifuged at 13,000 rpm for 10 min at 4˚C and the supernatant was separated. Trichloroacetic acid was added to a final concentration of 18% v/v, was incubated for 15 min at 4˚C with

rotation, and it was centrifuged for 10 min at 13,000 rpm at 4˚C. The pellet was washed with 500 µL of ice-cold 0.1 N HCl/acetone and kept on ice for 15 min. Then it was centrifuged for 10 min at 13,000 rpm at 4˚C. The pellet was resuspended in 1 volume of 0.1 M HCl and 3 volumes of cold acetone were added and proteins were left to precipitate overnight at -20˚C. Next day, it was centrifuged at 13,000 rpm for 10 min at 4˚C and washed with ice-cold acetone twice. The final pellet was vacuum-dried and histone H1s were then carefully solubilized with 50 µL of ultrapure water.

3.17. Preparation and immunostaining of polytene chromosomes

1. PREPARATION OF POLYTENE CHROMOSOME SPREADS: Salivary glands from third instar larvae were dissected in PBS. Approximately 15 larvae were dissected per preparation. Then transferred and incubated for 8 min with Cohen-Gotchell Buffer (10 mM MgCl2, 25 mM sodium glycerol 3-phosphate, 3 mM $CaCl_2$, 10 mM KH_2PO_4, 0.5% NP-40, 30 mM KCl, 160 mM sucrose). The buffer was removed and 100 µL of Fixing Solution 1 (250 µL PBS 10X, 50 µL formaldehyde 37%, 2.2 mL H_2O) was added and incubated for 2 min. During this incubation period, as much fat tissue as possible was removed. Next, the solution was exchanged with Fixing Solution 2 (1.125 mL acetic acid, 50 µL formaldehyde 37%, 1.325 mL H_2O) and incubated for 3 min. Salivary glands were transferred with the minimum amount of Fixing Solution 2 as possible (drops) onto a sylanized coverslip, and the coverslip containing the glands was collected with a slide. Immediately, the coverslip was repeatedly beaten using a needle, while holding it on one side to avoid moving, and it was put downwards and pressed relatively strong using the thumb to squash the cells. The preparation was hold into liquid nitrogen until it was frozen and the coverslip was carefully removed using a razorblade. The preparation was stored in a slide container containing PBS at 4˚C until the immunostaining was performed (up to 4 h).

2. IMMUNOSTAINING: The slide was washed 3 times for 5 min with PBS-0.05% Tween, incubating on the shaker. When immunostained with S9.6 antibody, a 5 min wash with PBS-0.5% SDS was performed instead of the second wash. Then, the preparation was blocked 2 times for 20 min in blocking solution (PBS, 0.05% Tween, 2% BSA).

To incubate with the primary antibodies, a box with wet paper was prepared to keep it humid. 25 µL dilution of the corresponding antibodies in blocking solution were added onto the polytene preparation. Then, it was covered with a coverslip avoiding bubbles and it was incubated in the wet box (first hour at room temperature and then overnight at 4˚C). Next day, the slide was washed 3 times for 5min with PBS-0.05% Tween on the shaker and 25 µL of the corresponding secondary antibody dilution were added (1:400 in PBS-0.05% Tween) and covered with a coverslip as before. It was incubated in the (dark) wet box for 1 h at room temperature. The slide was washed twice for 5min with PBS-0.05% Tween on the shaker and once with PBS. The preparation was mounted with 20 µL Dapi-Mowiol (2 ng/µL of Dapi in Mowiol) and let for 15 min in the dark at room temperature before keeping it at 4˚C. The preparation was later analyzed with a Leica TCS SPE confocal microscopy.

3.18. Cell immunostaining

Two different adherent surfaces were used for immunostaining S2 cells and human cancer cells. Coverslips coated with Concanavalin A were prepared as follows: coverslips were washed with 100% ethanol. Once dried, 50 µL of Concanavalin A (Sigma; 0.5 mg/ml) were added covering the maximum surface as possible and incubated for 15 min. Excess of Concanavalin A was removed and the coverslips were let to dry. To prepare coverslips coated with Poly-L-lysin, the coverslips were washed in 1M HCl at 50˚C for 2 h and washed three times for 5 min with ultrapure water. Then, they were incubated in diluted (1:10) Poly L-Lys solution for 1 h on a shaker, washed again three times with water and dried. Before plating cells, they were UV-treated for 10 min.

1. S2 CELLS: S2 cells were cultured for 30 min at 25˚C on a 12mm-diameter coverslip coated with Concanavalin A. Cells were washed with PBS for 10 min with slight agitation and fixed for 15 min with 4% paraformaldehyde at room temperature without agitation. Cells were washed twice for 10 min with PBS and twice with PBS-0.3% Triton X100-0.2% BSA. In case of immunostaining with S9.6, an extra wash of 0.5% SDS-PBS was performed. Then, diluted primary antibody in PBS-0.3% Triton X100-0.2% BSA was added an incubated

for overnight at 4˚C with agitation. Cells were washed three times with PBS-0.3% Triton X100-0.2% BSA for 10 min and 1:400 diluted secondary antibody in PBS-0.3% Triton X100-0.2% BSA was added and incubated for 1 h at room temperature in agitation. The preparation was mounted with 5 µL Dapi-Mowiol and let for 15 min in the dark at room temperature before keeping it at 4˚C.

2. CANCER CELLS: Cells were incubated overnight on a 12mm-diameter coverslip coated with Poly L-Lys. Then, they were washed twice with PBS in slight agitation for 2 min and were fixed with 4% paraformaldehyde for 10 min at room temperature. Then, cells were washed and incubated with 0,2% Triton X100 in PBS for 10min at room temperature. Two more washes were performed and cells were blocked with PBS-BT (0.1% Triton X100, 3% BSA, 0.05% NaN_3 in PBS) for 1 h. Primary antibody was diluted in PBS-BT and incubated overnight at 4˚C in a wet box. Cells were washed three times for 2 min with PBS-BT and incubated with 1:400 diluted secondary antibody in PBS-BT for 1 h at room temperature in a (dark) wet box. Three 2 min-washes with PBS-BT were performed before the preparation was mounted with 5 µL Dapi-Mowiol. They were let for 15 min in the dark at room temperature before keeping it at 4˚C.

4. ANALYSIS AND VISUALIZATION

4.1. FIJI

Images were recorded on Leica TCS SPE and Leica TCS SP2 AOBS system confocal microscopes, and were imported into Fiji software for measurements and adjustments.

In order to make the images quantifiable, the settings for the different lasers were established differently depending on the experiment and maintained for all the conditions of the same experiment.

Fluorescence confocal images of S2 cells were recorded from bottom to top with 0.5 µm thick stacks and 63X/1.4 oil lense. Stacks were then processed by filtering 8-bit images for γH2Av or S9.6 immunostaining with the corresponding 8-bit DAPI images to subtract signals not colocalizing with DAPI and converted to RGB. For quantification, maximal projections of the RGB stacks were obtained. The total S9.6 area was determined and

expressed as percentage of total DAPI area and the number of γH2Av foci per cell (0 - infinite) was calculated. For γH2A.X quantification in HT29, images were recorded following the above-mentioned procedure and the total γH2A.X reactivity was determined and normalized for the total amount of cells.

Polytene chromosome confocal images were obtained using 40X and 63X/1.4 oil objectives. Images shown are maximum projections of Z stacks sections (1.2 µm). For hnRNP and dH1 signal quantification in positive HP1a area, in addition to DAPI, polytene chromosomes were filtered for HP1a. Then, the mean intensity of this region was determined and normalized respect to the total HP1 area for each polytene chromosome.

4.2. Bioinformatics analyses

The sequencing was performed with the help of the IRB Barcelona Functional Genomics Core Facility and the analyses were done together with the IRB Barcelona Biostatistics/Bioinformatics Core Facility.

Briefly, library construction, cluster generation and sequencing analysis were performed following manufacturer's protocols (www.illumina.com). Reads were aligned against the dm3 UCSC genome 2012 release using Bowtie 0.12.5, allowing for two mismatches and considering all possible alignment sites for each read (for the analysis on Chapter 1) or considering only unique hit alignments (for the analysis on Chapter 2).

To determine location of R-loops across the whole genome, a peak calling using MACS 1.4 in RNH untreated samples was first performed. Then, log2 RPKM immunoprecipitated signal over identified peaks was measured for both RNH untreated and treated sample pairs to measure the RNAH treatment effect, and only those locations showing at least 30% reduction upon RNH1 treatment in both replicates were kept (FC < −1.5) for the analysis on Chapter 1. For the analysis on Chapter 2, all the locations showing a reduction (FC < -1) were kept. Finally, these identified sites were annotated for overlapping and closest genes against the Ensembl genome annotation version 71 (April 2013) using the annotatePeakInBatch function from the ChIPPeakAnno package68. version 2.6.1. FASTA sequences for differentially enriched/depleted sites were retrieved with the getSeq function from the Biostrings package v2.26.3

(using the annotation package BSgenome.Dmelanogaster.UCSC.dm3 v1.3.19) and were analyzed with the RepeatMasker software (http://www.repeatmasker.org, version open-4.0.5) in order to identify repeated elements found among them. Peak location and coverage plots were generated with the htSeqTools package version 1.20.0. Permutation tests to assess enrichment of repeated elements among identified peaks were performed with the regioneR package version 1.2.3, using the UCSC dm3 RepeatMasker track (March 2017).

REFERENCES

REFERENCES

1. Lu, X. *et al.* Linker histone H1 is essential for Drosophila development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure. *Genes Dev.* **23,** 452–65 (2009).

2. Vujatovic, O. *et al.* Drosophila melanogaster linker histone dH1 is required for transposon silencing and to preserve genome integrity. *Nucleic Acids Res.* **40,** 5402–5414 (2012).

3. Lu, X. *et al.* Drosophila H1 regulates the genetic activity of heterochromatin by recruitment of Su(var)3-9. *Science* **340,** 78–81 (2013).

4. Bayona-Feliu, A., Casas-Lamesa, A., Reina, O., Bernués, J. & Azorín, F. Linker histone H1 prevents R-loop accumulation and genome instability in heterochromatin. *Nat. Commun.* **8,** 283 (2017).

5. Luger, K., Mä, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389,** 251–260 (1997).

6. Thoma, F., Koller, T. & Klug, A. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell Biol.* **83,** 403–427 (1979).

7. Hartley, P. D. & Madhani, H. D. Mechanisms that Specify Promoter Nucleosome Location and Identity. *Cell* **137,** 445–458 (2009).

8. Takahata, S., Yu, Y. & Stillman, D. J. FACT and Asf1 Regulate Nucleosome Dynamics and Coactivator Binding at the HO Promoter. *Mol. Cell* **34,** 405–415 (2009).

9. Fyodorov, D. V., Zhou, B. R., Skoultchi, A. I. & Bai, Y. Emerging roles of linker histones in regulating chromatin structure and function. *Nat. Rev. Mol. Cell Biol.* **19,** 192–206 (2017).

10. Thåström, A., Bingham, L. M. & Widom, J. Nucleosomal Locations of Dominant DNA Sequence Motifs for Histone–DNA Interactions and Nucleosome Positioning. *J. Mol. Biol.* **338,** 695–709 (2004).

11. Iyer, V. & Struhl, K. Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J.*

14, 2570–2579 (1995).

12. Cope, N. F., Fraser, P. & Eskiw, C. H. The yin and yang of chromatin spatial organization. *Genome Biol.* **11,** 204–212 (2010).

13. Waddington, C. H. Genetic Assimilation of the Bithorax Phenotype. *Evolution (N. Y).* **10,** 1–13 (1956).

14. Dulac, C. Brain function and chromatin plasticity. *Nature* **465,** 728–735 (2010).

15. Margueron, R. & Reinberg, D. Chromatin structure and the inheritance of epigenetic information. *Nat. Rev. Genet.* **11,** 285–96 (2010).

16. Henikoff, S. & Smith, M. M. Histone variants and epigenetics. *Cold Spring Harb. Perspect. Biol.* **7,** a019364 (2015).

17. Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* **273,** 5858–5868 (1998).

18. Madigan, J. P., Chotkowski, H. L. & Glaser, R. L. DNA double-strand break-induced phosphorylation of Drosophila histone variant H2Av helps prevent radiation-induced apoptosis. *Nucleic Acids Res.* **30,** 3698–3705 (2002).

19. Thomas, C. J. *et al.* Kinase-mediated changes in nucleosome conformation trigger chromatin decondensation via poly(ADP-ribosyl)ation. *Mol. Cell* **53,** 831–842 (2014).

20. Palmer, D. K., Trongt, H. LE, Charbonneau, H. & Margolis, R. L. Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Cell Biol.* **88,** 3734–3738 (1991).

21. Buchwitz, B. J., Ahmad, K., Moore, L. L., Roth, M. B. & Henikoff, S. A histone-H3-like protein in C. elegans. *Nature* **401,** 547–548 (1999).

22. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **2122,** 381–395 (2011).

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

24. Peterson, C. L. & Laniel, M.-A. Histones and histone modifications. *Curr. Biol.* **14,** 546–551 (2004).

25. Voigt, P. *et al.* Asymmetrically Modified Nucleosomes. *Cell* **151,** 181–193 (2012).

26. Klose, R. J. & Bird, A. P. Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **31,** 89–97 (2006).

27. Wang, Y. & Leung, F. C. C. An evaluation of new criteria for CpG islands in the human genome as gene markers. *Bioinformatics* **20,** 1170–1177 (2004).

28. Watt, F. & Molloy, P. L. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.* **2,** 1136–1143 (1988).

29. Hu, S. *et al.* DNA methylation presents distinct binding sites for human transcription factors. *Elife* **2,** e00726 (2013).

30. Kriaucionis, S. & Heintz, N. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science (80-.).* **324,** 929–930 (2009).

31. He, Y.-F. *et al.* Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science (80-.).* **333,** 1303–1307 (2011).

32. Emil, H. Das Heterochromatin der Moose. *Jahrbücher für Wissenschaftliche Bot.* **69,** 762–818 (1928).

33. Filion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* **143,** 212–24 (2010).

34. Kharchenko, P. V *et al.* Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471,** 480–486 (2011).

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

36. Grewal, S. I. S. & Jia, S. Heterochromatin revisited. *Nature Reviews Genetics* **8,** 35–46 (2007).

37. Elgin, S. C. R. & Reuter, G. Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. *Cold Spring Harb. Perspect. Biol.* **5,** a017780 (2013).

38. Viegas-Pequignot, E., Dutrillaux, B. & Thomas, G. Inactive X chromosome has the highest concentration of unmethylated Hha I sites. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 7657–7660 (1988).

39. Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300,** 131–135 (2003).

40. Costanzi, C. & Pehrson, J. R. Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* **393,** 599–601 (1998).

41. Gerstein, M. B. *et al.* Comparative analysis of the transcriptome across distant species. *Nature* **512,** 445–448 (2014).

42. Hoskins, R. a *et al.* Heterochromatic sequences in a Drosophila whole-genome shotgun assembly. *Genome Biol.* **3,** RESEARCH0085.1-16 (2002).

43. Canzio, D. *et al.* Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Mol. Cell* **41,** 67–81 (2011).

44. Hall, I. M. *et al.* Establishment and Maintenance of a Heterochromatin Domain. *Science (80-.).* **297,** 2232–2237 (2002).

45. Garcia, J. F., Dumesic, P. A., Hartley, P. D., El-Samad, H. & Madhani, H. D. Combinatorial, site-specific requirement for heterochromatic silencing factors in the elimination of nucleosome-free regions. *Genes Dev.* **24,** 1758–71 (2010).

46. Zofall, M. & Grewal, S. I. S. RNAi-mediated Heterochromatin Assembly in Fission Yeast. *Cold Spring Harb. Symp. Quant. Biol.* **71,** 487–496 (2006).

47. Donze, D. & Kamakaka, R. T. Braking the silence: How heterochromatic gene repression is stopped in its tracks. *BioEssays* **24,** 344–349 (2002).

48. Wang, J., Lawry, S. T., Cohen, A. L. & Jia, S. Chromosome boundary elements and regulation of heterochromatin spreading. *Cell Molecilar Life Sci.* **71,** 4841–4852 (2014).

49. Piacentini, L., Fanti, L., Berloco, M., Perrini, B. & Pimpinelli, S. Heterochromatin protein 1 (HP1) is associated with induced gene expression in Drosophila euchromatin. *J. Cell Biol.* **161,** 707–714 (2003).

50. Figueiredo, M. L. A., Philip, P., Stenberg, P. & Larsson, J. HP1a Recruitment to Promoters Is Independent of H3K9 Methylation in Drosophila melanogaster. *PLoS Genet.* **8,** e1003061 (2012).

51. Vakoc, C. R., Mandat, S. A., Olenchock, B. A. & Blobel, G. A. Histone H3 Lysine 9 Methylation and HP1γ Are Associated with Transcription Elongation through Mammalian Chromatin. *Mol. Cell* **19,** 381–391 (2005).

52. Hilliker, A. J. Genetic analysis of the centromeric heterochromatin of chromosome 2 of Drosophila melanogaster: deficiency mapping of EMS-induced lethal complementation groups. *Genetics* **83,** 765–782 (1976).

53. Harshman, S. W., Young, N. L., Parthun, M. R. & Freitas, M. A. H1 histones: current perspectives and challenges. *Nucleic Acids Res.* **41,** 9593–9609 (2013).

54. Hansen, J. C., Lu, X., Ross, E. D. & Woody, R. W. Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J. Biol. Chem.* **281,** 1853–1856 (2006).

55. Izzo, A., Kamieniarz, K. & Schneider, R. The histone H1 family: Specific members, specific functions? *Biol. Chem.* **389,** 333–343 (2008).

56. Malik, H. S. & Henikoff, S. Phylogenomics of the nucleosome. *Nat. Struct. Biol.* **10,** 882–891 (2003).

57. Fan, Y. *et al.* Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* **123,** 1199–212 (2005).

58. Hashimoto, H. *et al.* Histone H1 null vertebrate cells exhibit altered nucleosome architecture. *Nucleic Acids Res.* **38,** 3533–3545 (2010).

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

60. Olins, D. E. & Olins, A. L. Chromatin history: Our view from the

bridge. *Nat. Rev. Mol. Cell Biol.* **4,** 809–814 (2003).

61. Misteli, T., Gunjan, A., Hock, R., Bustin, M. & Brown, D. T. Dynamic binding of histone H1 to chromatin in living cells. *Nature* **408,** 877–881 (2000).

62. Kimura, H. & Cook, P. R. Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J. Cell Biol.* **153,** 1341–1353 (2001).

63. Zhou, B.-R. *et al.* Structural Mechanisms of Nucleosome Recognition by Linker Histones. *Mol. Cell* **59,** 628–638 (2015).

64. Vermaak, D. & Malik, H. S. Multiple Roles for Heterochromatin Protein 1 Genes in Drosophila. *Annu. Rev. Genet* **43,** 467–492 (2009).

65. Xu, N., Emelyanov, A. V, Fyodorov, D. V & Skoultchi, A. I. Drosophila linker histone H1 coordinates STAT-dependent organization of heterochromatin and suppresses tumorigenesis caused by hyperactive JAK-STAT signaling. *Epigenetics Chromatin* **7,** 1–13 (2014).

66. Li, W. X. Canonical and non-canonical JAK–STAT signaling. *Trends Cell Biol.* **18,** 545–551 (2008).

67. Daujat, S., Zeissler, U., Waldmann, T., Happel, N. & Schneider, R. HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. *J. Biol. Chem.* **280,** 38090–38095 (2005).

68. Lu, A. *et al.* Mapping of Lysine Monomethylation of Linker Histones in Human Breast and Its Cancer. *J. Proteome Res.* **8,** 4207–4215 (2009).

69. Bonet-Costa, C. *et al.* Combined bottom-up and top-down mass spectrometry analyses of the pattern of post-translational modifications of Drosophila melanogaster linker histone H1. *J. Proteomics* **75,** 4124–4138 (2012).

70. Parseghian, M. H. & Hamkalo, B. a. A compendium of the histone H1 family of somatic subtypes: an elusive cast of characters and their characteristics. *Biochem. Cell Biol.* **79,** 289–304 (2001).

71. Shen, X. & Gorovsky, M. A. Linker histone H1 regulates specific gene expression but not global transcription in vivo. *Cell* **86,** 475–483

(1996).

72. Sirotkin, A. M. *et al.* Mice develop normally without the H1(0) linker histone. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 6434–6438 (1995).

73. Fan, Y., Sirotkin, A., Russell, R. G., Ayala, J. & Skoultchi, A. I. Individual Somatic H1 Subtypes Are Dispensable for Mouse Development Even in Mice Lacking the H1 0 Replacement Subtype. **21,** 7933–7943 (2001).

74. Fan, Y. *et al.* H1 Linker Histones Are Essential for Mouse Development and Affect Nucleosome Spacing In Vivo. *Mol. Cell. Biol.* **23,** 4559–4572 (2003).

75. Geeven, G. *et al.* Local compartment changes and regulatory landscape alterations in histone H1-depleted cells. *Genome Biol.* **16,** 289 (2015).

76. Stützer, A. *et al.* Modulations of DNA Contacts by Linker Histones and Post-translational Modifications Determine the Mobility and Modifiability of Nucleosomal H3 Tails. *Mol. Cell* **61,** 247–259 (2016).

77. Yang, S.-M., Kim, B. J., Norwood Toro, L. & Skoultchi, A. I. H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation. *Proc. Natl. Acad. Sci.* **110,** 1708–1713 (2013).

78. Herrera, J. E., West, K. L., Schiltz, R. L., Nakatani, Y. & Bustin, M. Histone H1 is a specific repressor of core histone acetylation in chromatin. *Mol. Cell. Biol.* **20,** 523–529 (2000).

79. Vaquero, A. *et al.* Human SirT1 Interacts with Histone H1 and Promotes Formation of Facultative Heterochromatin. *Mol. Cell* **16,** 93–105 (2004).

80. Koop, R., Di Croce, L. & Beato, M. Histone H1 enhances synergistic activation of the MMTV promoter in chromatin. *EMBO J.* **22,** 588–599 (2003).

81. Dominski, Z. & Marzluff, W. F. Formation of the 3' end of histone mRNA. *Gene* **239,** 1–14 (1999).

82. Pandey, N. B., Chodchoy, N., Liu, T. J. & Marzluff, W. F. Introns in histone genes alter the distribution of 3' ends. *Nucleic Acids Res.* **18,** 3161–3170 (1990).

83. Albig, W., Kioschis, P., Poustka, A., Meergans, K. & Doenecke, D. Human histone gene organization: Nonregular arrangement within a large cluster. *Genomics* **40,** 314–322 (1997).

84. Marzluff, W. F., Gongidi, P., Woods, K. R., Jin, J. & Maltais, L. J. The human and mouse replication-dependent histone genes. *Genomics* **80,** 487–498 (2002).

85. Happel, N. & Doenecke, D. Histone H1 and its isoforms: Contribution to chromatin structure and function. *Gene* **431,** 1–12 (2008).

86. Millán-Ariño, L. *et al.* Mapping of six somatic linker histone H1 variants in human breast cancer cells uncovers specific features of H1.2. *Nucleic Acids Res.* **42,** 4474–4493 (2014).

87. Parseghian, M. H., Newcomb, R. L. & Hamkalo, B. A. Distribution of somatic H1 subtypes is non-random on active vs. inactive chromatin II: distribution in human adult fibroblasts. *J. Cell. Biochem.* **83,** 643–659 (2001).

88. Pérez-Montero, S., Carbonell, A., Morá, T. S., Vaquero, A. & Azorín, F. The Embryonic Linker Histone H1 Variant of Drosophila, dBigH1, Regulates Zygotic Genome Activation. *Dev. Cell* **26,** 578–590 (2013).

89. Terme, J.-M. *et al.* Histone H1 Variants Are Differentially Expressed and Incorporated into Chromatin during Differentiation and Reprogramming to Pluripotency. *J. Biol. Chem.* **286,** 35347–35357 (2011).

90. Gabrovsky, N., Georgieva, M., Laleva, M., Uzunov, K. & Miloshev, G. Histone H1.0—a potential molecular marker with prognostic value for patients with malignant gliomas. *Acta Neurochir. (Wien).* **155,** 1437–1442 (2013).

91. Wisniewski, J. R., Zougman, A., Krüger, S. & Mann, M. Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. *Mol. Cell. Proteomics* **6,** 72–87 (2007).

92. Snijders, A. P. L. *et al.* Characterization of Post-Translational Modifications of the Linker Histones H1 and H5 from Chicken Erythrocytes Using Mass Spectrometry. *J. Proteome Res.* **7,** 4326–4335 (2008).

93. Talasz, H., Helliger, W., Puschendorf, B. & Lindner, H. In Vivo Phosphorylation of Histone H1 Variants during the Cell Cycle †. *Biochemistry* **35,** 1761–1767 (1996).

94. Garcia, B. a *et al.* Comprehensive phosphoprotein analysis of linker histone H1 from Tetrahymena thermophila. *Mol. Cell. Proteomics* **5,** 1593–1609 (2006).

95. Izzo, A. & Schneider, R. The role of linker histone H1 modifications in the regulation of gene expression and chromatin dynamics. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1859,** 486–495 (2016).

96. Mckay, D. J. *et al.* Interrogating the Function of Metazoan Histones using Engineered Gene Clusters Developmental Cell Resource Interrogating the Function of Metazoan Histones using Engineered Gene Clusters. *Dev. Cell* **32,** 373–386 (2015).

97. Isogai, Y., Keles, S., Prestel, M., Hochheimer, A. & Tjian, R. Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev.* **21,** 2936–2949 (2007).

98. Guglielmi, B., Rochelle, N. La & Tjian, R. Gene-Specific Transcriptional Mechanisms at the Histone Gene Cluster Revealed by Single-Cell Imaging. *Mol. Cell* **51,** 480–492 (2013).

99. Scaffidi, P. Histone H1 alterations in cancer. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1859,** 533–539 (2016).

100. Hechtman, J. F. *et al.* Promyelocytic leukemia zinc finger and histone H1.5 differentially stain low- and high-grade pulmonary neuroendocrine tumors: a pilot immunohistochemical study. *Hum. Pathol.* **44,** 1400–1405 (2013).

101. Li, H. *et al.* Mutations in linker histone genes HIST1H1 B, C, D, and E; OCT2 (POU2F2); IRF8; and ARID1A underlying the pathogenesis of follicular lymphoma. *Blood* **123,** 1487–1498 (2014).

102. Aguilera, A. & García-Muse, T. R Loops: From Transcription Byproducts to Threats to Genome Stability. *Mol. Cell* **46,** 115–124 (2012).

103. Roy, D. & Lieber, M. R. G Clustering Is Important for the Initiation of Transcription-Induced R-Loops In Vitro, whereas High G Density without Clustering Is Sufficient Thereafter. *Mol. Cell. Biol.* **29,** 3124–3133 (2009).

104. Itoh, T. & Tomizawa, J. Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. *Proc. Natl. Acad. Sci. U. S. A.* **77,** 2450–2454 (1980).

105. Nossal, N. G., Dudas, K. C. & Kreuzer, K. N. Bacteriophage T4 proteins replicate plasmids with a preformed R loop at the T4 ori(uvsY) replication origin in vitro. *Mol. Cell* **7,** 31–41 (2001).

106. Xu, B. & Clayton, D. A. RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers. *EMBO J.* **15,** 3135–3143 (1996).

107. Sanz, L. A. *et al.* Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Mol. Cell* **63,** 167–178 (2016).

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

109. Keskin, H. *et al.* Transcript RNA-templated DNA recombination and repair. *Nature* **515,** 436–439 (2014).

110. Ohle, C. *et al.* Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair. *Cell* **167,** 1001–1013 (2016).

111. D'Alessandro, G. *et al.* A role for RNA and DNA:RNA hybrids in the modulation of DNA repair by homologous recombination. *bioRxiv* 255976 (2018).

112. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chédin, F. R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. *Mol. Cell* **45,** 814–825 (2012).

113. Lim, Y. W., Sanz, L. A., Xu, X., Hartono, S. R. & Chédin, F. Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi–Goutières syndrome. *Elife* **4,** (2015).

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

115. Skourti-Stathaki, K., Proudfoot, N. J. & Gromak, N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42,** 794–805

(2011).

116. El Hage, A., Webb, S., Kerr, A. & Tollervey, D. Genome-Wide Distribution of RNA-DNA Hybrids Identifies RNase H Targets in tRNA Genes, Retrotransposons and Mitochondria. *PLoS Genet.* **10,** e1004716 (2014).

117. Castellano-Pozo, M. *et al.* R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol. Cell* **52,** 583–590 (2013).

118. Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J. & Dean, C. R-Loop Stabilization Represses Antisense Transcription at the Arabidopsis FLC Locus. *Science* **340,** 619–621 (2013).

119. Powell, W. T. *et al.* R-loop formation at *Snord116* mediates topotecan inhibition of *Ube3a-antisense* and allele-specific chromatin decondensation. *Proc. Natl. Acad. Sci.* **110,** 13938–13943 (2013).

120. Shaw, N. N. & Arya, D. P. Recognition of the unique structure of DNA:RNA hybrids. *Biochimie* **90,** 1026–1039 (2008).

121. Eder, P. S., Walder, R. Y. & Walder, J. A. Substrate specificity of human RNase H1 and its role in excision repair of ribose residues misincorporated in DNA. *Biochimie* **75,** 123–126 (1993).

122. Cerritelli, S. M. & Crouch, R. J. Ribonuclease H: the enzymes in eukaryotes. *FEBS J.* **276,** 1494–1505 (2009).

123. Chon, H. *et al.* RNase H2 roles in genome integrity revealed by unlinking its activities. *Nucleic Acids Res.* **41,** 3130–3143 (2013).

124. Chakraborty, P. & Grosse, F. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst).* **10,** 654–665 (2011).

125. El Hage, A., French, S. L., Beyer, A. L. & Tollervey, D. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* **24,** 1546–1558 (2010).

126. Huertas, P. & Aguilera, A. S. Cotranscriptionally Formed DNA:RNA Hybrids Mediate Transcription Elongation Impairment and Transcription-Associated Recombination. *Mol. Cell* **12,** 711–721 (2003).

127. Santos-Pereira, J. M. & Aguilera, A. R loops: new modulators of genome dynamics and function. *Nat. Rev. Genet.* **16,** 583–597 (2015).

128. Richard, P. & Manley, J. L. R Loops and Links to Human Disease. *J. Mol. Biol.* **429,** 3168–3180 (2017).

129. Aguilera, A. & Gómez-González, B. Genome instability: a mechanistic view of its causes and consequences. *Nat. Rev. Genet.* **9,** 204–217 (2008).

130. Sollier, J. *et al.* Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. *Mol. Cell* **56,** 777–785 (2014).

131. Gan, W. *et al.* R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* **25,** 2041–2056 (2011).

132. Ni, J.-Q., Liu, L.-P., Hess, D., Rietdorf, J. & Sun, F.-L. Drosophila ribosomal proteins are associated with linker histone H1 and suppress gene transcription. *Genes Dev.* **20,** 1959–1973 (2006).

133. Guruharsha, K. G. *et al.* A protein complex network of Drosophila melanogaster. *Cell* **147,** 690–703 (2011).

134. Hari, K., Cook, K. & Karpen, G. The Drosophila Su(var)2-10 locus regulates chromosome structure and function and encodes a member of the PIAS protein family. *Genes Dev.* **15,** 1334–1348 (2001).

135. Johnson, E. S. Protein Modification by SUMO. *Annu. Rev. Biochem.* **73,** 355–382 (2004).

136. Piacentini, L. *et al.* Heterochromatin Protein 1 (HP1a) Positively Regulates Euchromatic Gene Expression through RNA Transcript Association and Interaction with hnRNPs in Drosophila. *PLoS Genet.* **5,** e1000670 (2009).

137. Blanchette, M. *et al.* Genome-wide Analysis of Alternative Pre-mRNA Splicing and RNA-Binding Specificities of the Drosophila hnRNP A/B Family Members. *Mol. Cell* **33,** 438–449 (2009).

138. Singh, A. K. & Lakhotia, S. C. The hnRNP A1 homolog Hrp36 is essential for normal development, female fecundity, omega speckle formation and stress tolerance in Drosophila melanogaster. *J.*

Biosci. **37,** 659–678 (2012).

139. Matunis, E. L., Matunis, M. J. & Dreyfuss, G. Association of Individual hnRNP Proteins and snRNPs with Nascent Transcripts. *J. Cell Biol.* **121,** 219–228 (1993).

140. Fischer, J. A., Giniger, E., Maniatis, T. & Ptashne, M. GAL4 activates transcription in Drosophila. *Nature* **332,** 853–856 (1988).

141. Henriques, T. *et al.* Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev.* (2018). doi:10.1101/gad.309351.117

142. Arnold, C. D. *et al.* Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science (80-.).* **339,** 1074–1077 (2013).

143. Font-Burgada, J., Reina, O., Rossell, D. & Azorín, F. chroGPS, a global chromatin positioning system for the functional analysis and visualization of the epigenome. *Nucleic Acids Res.* **42,** 2126–2137 (2014).

144. Ghosh, D., Gerasimova, T. I. & Corces, V. G. Interactions between the Su(Hw) and Mod(mdg4) proteins required for gypsy insulator function. *EMBO J.* **20,** 2518–2527 (2001).

145. Efroni, S. *et al.* Cell Stem Cell Article Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell* **2,** 437–447 (2008).

146. Zhang, Y. *et al.* Histone h1 depletion impairs embryonic stem cell differentiation. *PLoS Genet.* **8,** e1002691 (2012).

147. Sancho, M., Diani, E., Beato, M. & Jordan, A. Depletion of Human Histone H1 Variants Uncovers Specific Roles in Gene Expression and Cell Growth. *PLoS Genet.* **4,** e1000227 (2008).

148. Khochbin, S. Histone H1 diversity: bridging regulatory signals to linker histone function. *Gene* **271,** 1–12 (2001).

149. Kuo, L. J. & Yang, L.-X. Gamma-H2AX - a novel biomarker for DNA double-strand breaks. *In Vivo* **22,** 305–9 (2008).

150. Patterton, H. G., Landel, C. C., Landsman, D., Peterson, C. L. & Simpson, R. T. The biochemical and phenotypic characterization of Hho1p, the putative linker histone H1 of Saccharomyces cerevisiae. *J. Biol. Chem.* **273,** 7268–7276 (1998).

151. Wellinger, R. E., Prado, F. & Aguilera, A. Replication Fork Progression Is Impaired by Transcription in Hyperrecombinant Yeast Cells Lacking a Functional THO Complex. *Mol. Cell. Biol.* **26,** 3327–3334 (2006).

152. Zhang, Z. Z., Pannunzio, N. R., Hsieh, C. L., Yu, K. & Lieber, M. R. Complexities due to single-stranded RNA during antibody detection of genomic rna:dna hybrids. *BMC Res. Notes* **8,** 127 (2015).

153. König, F., Schubert, T. & Längst, G. The monoclonal S9.6 antibody exhibits highly variable binding affinities towards different R-loop sequences. *PLoS One* **12,** e0178875 (2017).

154. Halász, L. *et al.* RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome Res.* **27,** 1063–1073 (2017).

155. Iwasaki, Y. W. *et al.* Piwi Modulates Chromatin Accessibility by Regulating Multiple Factors Including Histone H1 to Repress Transposons. *Mol. Cell* **63,** 408–419 (2016).

156. Zeller, P. *et al.* Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability. *Nat. Genet.* **48,** 1385–1395 (2016).

157. Jedrusik, M. A. & Schulze, E. H1.1 in C. elegans chromatin silencing. *Development* **128,** 1069–1080 (2001).

158. Studencka, M. *et al.* Novel roles of Caenorhabditis elegans heterochromatin protein HP1 and linker histone in the regulation of innate immune gene expression. *Mol. Cell. Biol.* **32,** 251–265 (2012).

159. Penke, T. J. R., McKay, D. J., Strahl, B. D., Matera, A. G. & Duronio, R. J. Direct interrogation of the role of H3K9 in metazoan heterochromatin function. *Genes Dev.* **30,** 1866–1880 (2016).

160. Almeida, R. *et al.* Chromatin conformation regulates the coordination between DNA replication and transcription. *Nat. Commun.* **9,** 1590–1604 (2018).

161. Amon, J. D. & Koshland, D. RNase H enables efficient repair of R-loop induced DNA damage. *Elife* **5,** e20533 (2016).

162. Wahba, L., Amon, J. D., Koshland, D. & Vuica-Ross, M. RNase H and

Multiple RNA Biogenesis Factors Cooperate to Prevent RNA:DNA Hybrids from Generating Genome Instability. *Mol. Cell* **44,** 978–988 (2011).

163. Cristini, A., Groh, M., Kristiansen, M. S. & Gromak, N. RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cell Rep.* **23,** 1891–1905 (2018).

164. Dunn, K. & Griffith, J. D. The presence of RNA in a double helix inhibits its interaction with histone protein. *Nucleic Acids Res.* **8,** 555–566 (1980).

165. Noy, A., Pérez, A., Márquez, M., Luque, F. J. & Orozco, M. Structure, Recognition Properties, and Flexibility of the DNA·RNA Hybrid. *J. Am. Chem. Soc.* **127,** 4910–4920 (2005).

166. Boque-Sastre, R. *et al.* Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc. Natl. Acad. Sci.* **112,** 5785–5790 (2015).

167. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13,** 720–731 (2012).

168. Ginno, P. A., Lim, Y. W., Lott, P. L., Korf, I. & Chédin, F. GC skew at the 5′ and 3′ ends of human genes links R-loop formation to epigenetic regulation and transcription termination. TL - 23. *Genome Res.* **23,** 1590–1600 (2013).

169. Chen, L. *et al.* R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Mol. Cell* **68,** 745–757.e5 (2017).

170. Gromak, N., West, S. & Proudfoot, N. J. Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol. Cell. Biol.* **26,** 3986–3996 (2006).

171. Gunjan, A. & Verreault, A. A Rad53 Kinase-Dependent Surveillance Mechanism that Regulates Histone Protein Levels in S. cerevisiae. *Cell* **115,** 537–549 (2003).

172. Singh, R. K., Kabbaj, M. H. M., Paik, J. & Gunjan, A. Histone levels are regulated by phosphorylation and ubiquitylation-dependent proteolysis. *Nat. Cell Biol.* **11,** 925–933 (2009).

173. Singh, R. K., Gonzalez, M., Kabbaj, M. H. & Gunjan, a. Novel E3 ubiquitin ligases that regulate histone protein levels in the budding yeast Saccharomyces cerevisiae. *PLoS One* **7,** e36295 (2012).

174. Siriaco, G., Deuring, R., Mawla, G. D. & Tamkun, J. W. A Novel Approach for Studying Histone H1 Function in Vivo. *Genetics* **200,** 29–33 (2015).

175. Schneider, I. Cell lines derived from late embryonic stages of Drosophila melanogaster. *J. Embryol. Exp. Morphol.* **27,** 353–65 (1972).

176. Font-Burgada, J., Rossell, D., Auer, H. & Azorin, F. Drosophila HP1c isoform interacts with the zinc-finger proteins WOC and Relative-of-WOC to regulate gene expression. *Genes Dev.* **22,** 3007–3023 (2008).

177. Moreno-Moreno, O., Torras-Llort, M. & Azorín, F. Proteolysis restricts localization of CID, the centromere-specific histone H3 variant of Drosophila, to centromeres. *Nucleic Acids Res.* **34,** 6247–6255 (2006).

178. Boguslawski, S. J. *et al.* Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J. Immunol. Methods* **89,** 123–130 (1986).

179. Hovemann, B. T., Reim, I., Werner, S., Katz, S. & Saumweber, H. The protein Hrb57A of Drosophila melanogaster closely related to hnRNP K from vertebrates is present at sites active in transcription and coprecipitates with four RNA-binding proteins. *Gene* **245,** 127–137 (2000).

180. Huynh, J.-R., Munro, T. P., Smith-Litière, K., Lepesant, J.-A. & Johnston, D. S. The Drosophila hnRNPA/B Homolog, Hrp48, Is Specifically Required for a Distinct Step in osk mRNA Localization. *Dev. Cell* **6,** 625–635 (2004).

181. Hales, K. G., Korey, C. A., Larracuente, A. M. & Roberts, D. M. Genetics on the fly: A primer on the drosophila model system. *Genetics* **201,** 815–842 (2015).

ANNEX I

ANNEX II