GENOMIC DISTRIBUTION AND FUNCTIONAL SPECIFICITY OF HUMAN HISTONE H1 SUBTYPES

Lluís Millán Ariño

DOCTORAL THESIS · 2013

Thesis supervisor:

Dr. Albert Jordan Vallès

Molecular Genomics Department

Institut de Biologia Molecular de Barcelona (IBMB) - CSIC

DEPARTMENT OF EXPERIMENTAL AND HEALTH SCIENCES – UPF

a mons pares,

ACKNOWLEDGEMENTS

>This Thesis
AlbertKhademulNúriaElenaAndreaNeusErikRegiJe
anMichelAdéEdurneAlisNachoMónicaVeroJaumeEri
kaCeciliaSilvinaGuilleJofreFanSilvioCristina
IsauraJordiYamGabrielaDonAndrésClaraMariaJor
diJúliaAdamNúriaSilviaAnnieDavidJoanBegoIgor
UViclabPassadísSegonaPlantaHèlixOscarToniVio
letaSergiAllBlacksAllGreysAllOrangeIbekersAm
icsJesusMaggieChecoDídacGerardRaquelCaroleAn
naPedroJaumeAndreaAgustíAlfredLorenaJoseJord
iEstherChanoPaquitaEusebiSoraiaPauJúliaKarin
aMiguelBrunoGermanetMarinaIaiaPadrinaIaioIai
aMarquesAriñoFamíliaDídacCescPapaMamaMíriam

Gràcies!

Abstract

Seven linker histone H1 variants exist in human somatic cells with distinct prevalence among cell types and during differentiation. Despite being key chromatin structural components, it remains elusive how they participate in the regulation of nuclear processes. Moreover, it is not well understood whether the different variants have specific roles or are differentially distributed along the genome. By taking advantage of specific antibodies for H1 variants and HA-tagged recombinant H1s expressed in breast cancer cells, the distribution of somatic variants H1.2 to H1.5, H1.0 and H1X has been investigated by combining ChIP-qPCR, ChIP-chip, and ChIP-seq analysis. All H1 variants bind gene promoters and are depleted from the TSS in active genes, and also from regulatory sites. The extension of H1 depletion at promoters is dependent on the transcriptional status of the gene and differs between variants. Analyses show that histone H1 is not uniformly distributed along the genome and differences among variants exist, being H1.2 the variant showing a more specific pattern and a strongest correlation with gene repression in breast cancer cells. Results suggest that different variants may be present at different chromatin types, and this may depend on the cell type, differentiation state, and whether cells are originated from a neoplastic process. In a second part of the thesis, it is shown that a previously reported H1.4 knock-down cell line presents and off-target effect against lamin B2. Therefore, it has been developed a new inducible knock-down cell line specifically inhibiting H1.4, which resembles previously characterized H1.2 knock-down. Finally, combined depletion of H1.4/lamin B2 and H1.2/H1.4 causes similar effects in T47D breast cancer cell line.

Key words: chromatin, epigenetics, histone H1, ChIP-seq, transcription start site, LADs, CpG islands, lamin B2.

Resum

Fins a set variants de la histona H1 s'han identificat en mamífers, les quals mostres una prevalença diferent entre tipus cel·lulars i durant el procés de diferenciació. Tot i que la histona H1 juga un paper clau en l'estructuració de la cromatina, no s'acaba d'entendre encara com participa exactament en els diferent processos cel·lulars. A més a més, encara no està clar si les diferents variants tenen funcions específiques ni si es distribueixen igual al llarg del genoma. Mitjançant anticossos específics per algunes variants d'H1 i de línies cel·lulars de càncer de mama que expressen H1s recombinant fusionades a un pèptid HA, s'ha estudiat la distribució genòmica de H1.2 a H1.5, H1.0 i H1X, combinant ChIP-qPCR, ChIP-chip i ChIP-seq. Totes les H1s es troben a promotors gènics i empobrides a l'inici de transcripció dels gens actius, i també a les regions reguladores. El grau de disminució d'H1 al promotor depèn de l'estat transcripcional del gen i presenta diferències entre variants. Els anàlisis mostren que la histona H1 no es distribueix uniformement al genoma i que hi ha diferències entre variants, essent H1.2 la variant que presenta un patró més específic i una correlació més forta amb repressió gènica a cèl·lules de càncer de mama. Aquests resultats suggereixen que variants d'H1 diferents es troben presents als diversos tipus de cromatina, i aquest fet podria dependre de la línia cel·lular, l'estat de diferenciació, o de si les cèl·lules s'han originat durant un procés neoplàsic. En una segona part de la tesi, es mostra que una línia cel·lular anteriorment descrita que inhibeix H1.4 presenta un efecte inespecífic contra lamina B2. Així, s'ha desenvolupat una altra línia que inhibeix H1.4 específicament, la qual s'assembla a un mutant anteriorment caracteritzat que inhibeix H1.2. Finalment, la inhibició combinada de H1.4/lamina B2 i H1.2/H1.4 provoca efectes fenotípics semblants a cèl·lules de càncer de mama T47D.

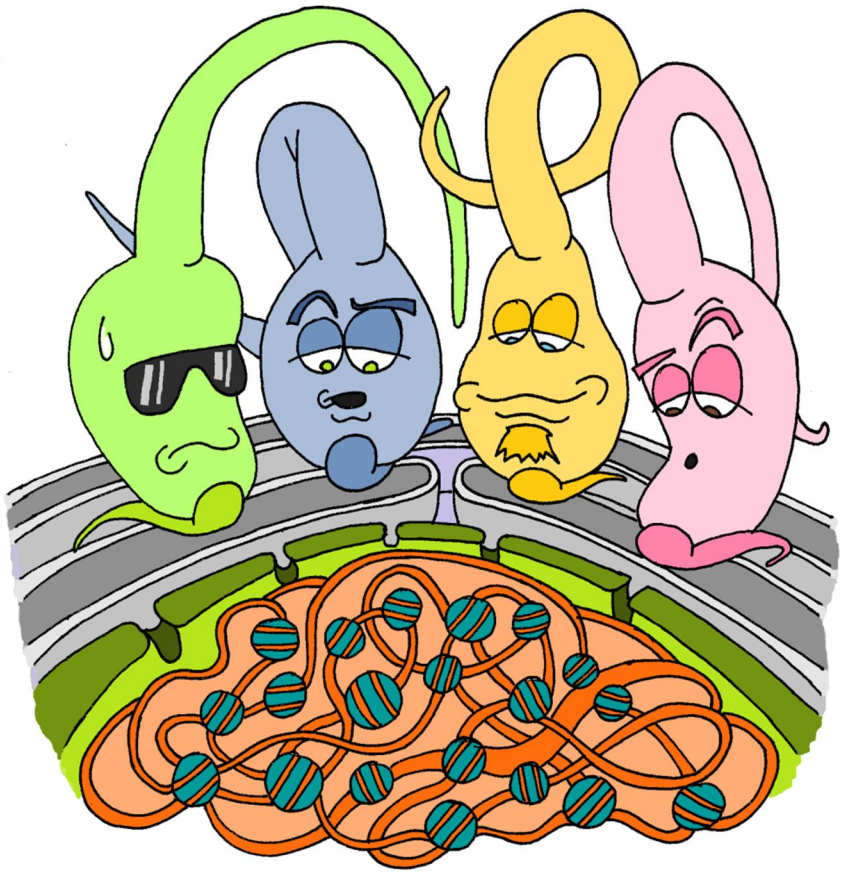
Paraules clau: cromatina, epigenètica, histona H1, ChIP-seq, inici de transcripció, LADs, illes CpG, lamina B2.

Preface

How diploid cells are able to pack about 2 meters of DNA within the cell nucleus is a question that has captivated scientists for many years. The eukaryotic cell has solved this problem by the existence of several proteins that compact the DNA forming a DNA-protein complex called chromatin. The main proteins contributing to this process are histone proteins, which, thanks to their positive electrostatic charge and structure, are able to wrap DNA around a core particle formed by two copies of each of the H2A, H2B, H3, and H4 histone proteins. The resulting complex is called nucleosome, and it is considered the fundamental repeat unit of the chromatin. Additionally, another histone protein, the linker histone H1, sits at the base of the nucleosome and promotes the stabilization and organization of nucleosomes into more complex structures, allowing the DNA to be finally packed and organized in the nucleus. This kind of organization raises other fascinating questions: into such compacted environment, how is chromatin regulated to allow the access of transcriptional and replication machineries at a given place of the genome at the time they are needed? How is genome regulated in order to lead to the formation of different cell types that possess different gene expression profiles? How do genes change their expression in response to environmental stimuli? And, how do all this processes affect diseases like cancer? The answers for those questions are more complex and involve many proteins apart from histones that, alone or by associating with others, are able to modulate the plasticity of the chromatin. Moreover, chromatin epigenetic modifications also contribute to the regulation of gene expression, replication, and DNA damage response. These include DNA methylation, histone post-translational modifications, incorporation of histone variants, and expression of non-coding RNAs. The involvement of core histones in those processes has been studied since a long time ago. However, less attention has been paid to the role of linker histone H1 in chromatin structure and function. Because of its role in the formation of higher order chromatin structures, linker histone H1 has been classically seen as an structural component related with chromatin compaction and, hence, as an obstacle for transcription. However, thanks to the development of knock-out and knock-down studies in several organisms and, more recently, to gene expression microarray technology, it has been demonstrated that H1 plays both a positive and negative role in gene expression. Another controversial issue regarding H1 deals on the existence of multiple H1 variants or subtypes in many organisms. For instance, 11 variants are described in humans, which possess characteristic expression patterns and other specific features. At the moment, it is not still well understood why so many variants exist and, although they seem to share

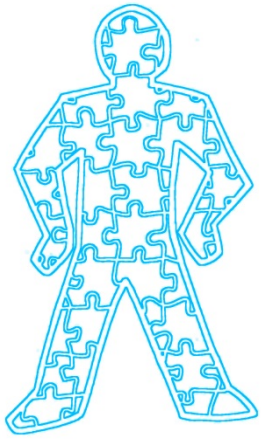
redundant functions, several reports point to the idea that they are also differently involved in specific cellular processes. An increasing effort has been done during the last years to elucidate divergent features among H1 variants, regarding their structure, expression pattern, chromatin dynamics, transcriptional regulation, post-translational modifications, and genome-wide distribution. The aim in our laboratory is to address H1 heterogeneity in breast cancer cell lines by different approaches using recently developed tools. Specific antibodies for most of the somatic H1 variants were achieved. Moreover, we have created inducible T47D breast cancer cell lines specifically inhibiting each of the H1 variants, and also cell lines over-expressing recombinant H1s fused to an HA tag. Therefore, we have recently reported that specific inhibition of each of the H1 variants cause specific phenotypes in breast cancer cell lines. Specifically, H1.2 depletion caused in T47D cells defects in proliferation, cellular arrest in the G1 phase of the cell cycle, and a reduction in the nucleosomal spacing. H1.4 knock-down also affected cell proliferation and an increase in cellular death. Finally, gene expression microarray analyses on those cells showed that different H1 variants regulate a specific subset of genes. One part of this thesis goes in depth into these phenotypic consequences after H1 variant knock-down. On the other hand, the main part deals about the genome-wide distribution of somatic H1 variants in breast cancer cells. These types of studies have been delayed respect to similar ones in core histones due to the lack of specific ChIP-grade antibodies for the H1 variants. However, during the last year, the first reports comparing the distribution of several histone H1 variants in the genome have started to arise. By taking advantage of two specific commercial antibodies for two of the variants and HA-tagged recombinant H1s expressed in breast cancer cells, we have studied the distribution of these variants in the genome, combining ChIP-qPCR, ChIP-chip, and ChIP-seq. Thus, the elucidation of the precise localization of linker histone H1 variants in different cell types related with different chromatin sates will help to understand the heterogeneity of this family of histones and the role that they play in chromatin organization, gene regulation, and other cellular processes.

Santa Bàrbara, September 2013



OK GUYS··· HOW DO
WE DISTRIBUTE IT?

CONTENTS



INTRODUCTION

1. CHROMATIN STRUCTURE

1.1. The double helix and the nucleosome

The DNA (Deoxyribonucleic acid) is the molecule that encodes the genetic instructions determining the makeup of all living cells and many viruses. It was first identified and isolated by Friedrich Miescher [1] and its double helix structure was first discovered by James Watson and Francis Crick [2]. The DNA molecule is a right-handed double helix structure composed of two anti-parallel strands. Each strand is the succession of four nucleotides: Adenine (A), Thymine (T), Guanine (G) and Cytosine (C), joined by phosphodiester bonds. The two strands are held together by hydrogen bonds between base pairs. Adenine pairs only with Thymine (A=T) and Cytosine with Guanine (C≡G). This forms an asymmetric molecule with two grooves of different size (Figure I.1A). The major groove, which is wider than the minor, is where most of sequence specific DNA-binding proteins bind to their cognate DNA through structural domains that recognize specific DNA sequences. On the other hand, the minor groove is the place where non-sequence-specific DNA ligands bind, such as most of the contacts with core and linker histones.

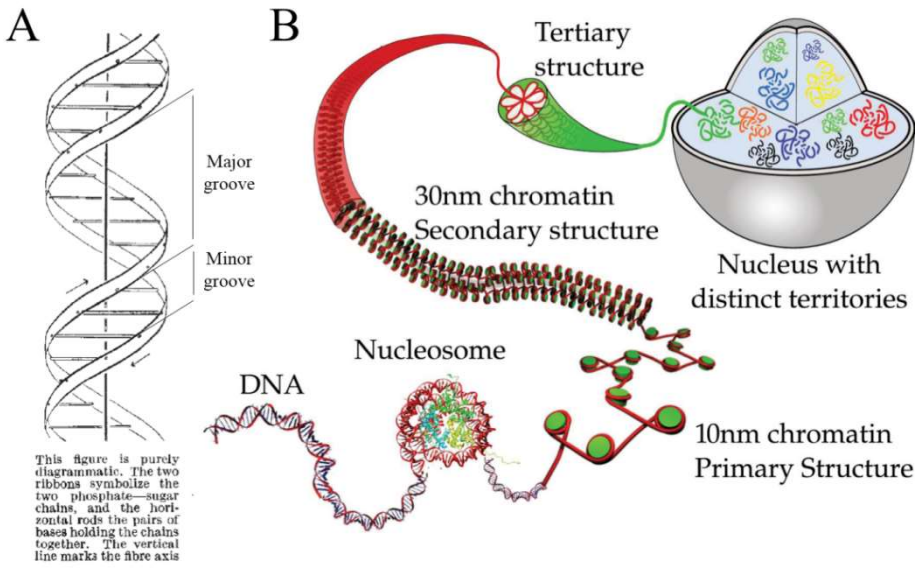


Figure I.1. Representation of the different levels of DNA compaction. (A) *First double-helix model, described by Watson and Crick in 1953.* ***(B)*** *Double-stranded DNA hierarchical organization.*

The eukaryotic DNA is packaged into the nucleus of the cell through its association with histone proteins, forming the chromatin. The fundamental repeat unit of the chromatin is the nucleosome, which consists of 146 bp of DNA wrapped around a histone octamer formed by two copies of the core histone proteins H2A, H2B, H3 and H4 [3, 4] (Figure I.2). Nucleosomes are connected by short DNA segments, termed “linker DNA”, into nucleosomal arrays. Linker

DNA length varies among species, tissues, and even within a single cellular genome. This beads-on-a-string organization of the nucleosome constitutes de “primary structure” of the chromatin [5, 6]. Apart from core histone proteins, another histone protein, the linker histone H1, binds to the nucleosome, siting at the base near the entry and exit sites. Histone H1 promotes the stabilization and organization of nucleosomes into higher-order “secondary” (such as the 30 nm chromatin fiber) or “tertiary” structures, which allows DNA to be organized in the cell nucleus [7-10] (Figure I.1B). Additionally, more recent evidences point to a spatial compartmentalization of the chromatin within the interphase nucleus, which enables the interaction of nucleosomes that may be distant with respect to their primary structure. This 3D organization provides another level of complexity to the hierarchical chromatin organization [11]. Finally, the highest compacted state of the chromatin occurs during cellular division, when metaphase chromosome structures are formed.

At those levels of DNA compaction, various mechanisms are involved in making genome accessible for readout by the complex machineries involved in gene transcription, DNA replication and DNA repair: the histone tail post-translational modifications, the existence of multiple histone variants, and the ATP-dependent chromatin remodelers. All this mechanisms collaborate together to control the repressive nature of chromatin, allowing access to the nucleosomal DNA needed at a certain point. Moreover, other architectural chromatin proteins (ACPs) and nucleosome-binding proteins, like histone chaperones, MeCP2, HP1, HMGs or PARP1, can also affect chromatin structure and dynamics [12].

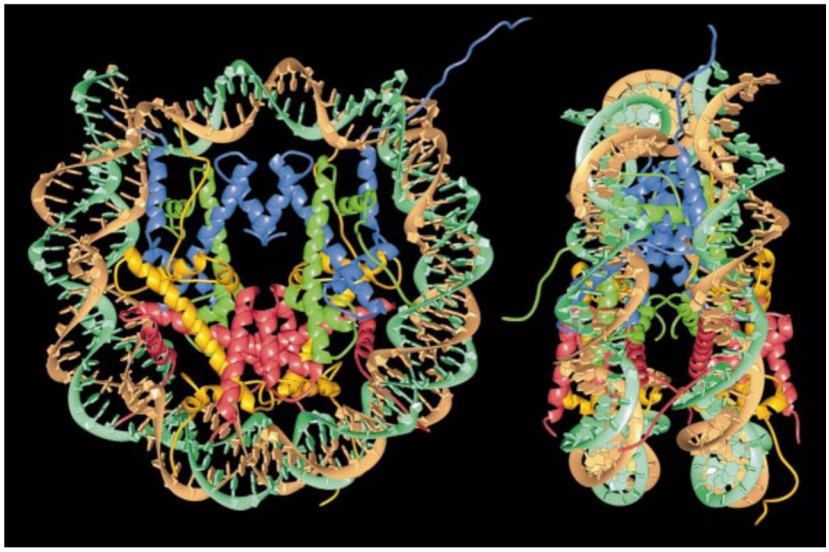


Figure I.2. Structural model of the nucleosome core particle, achieved by Luger, K. et al in 1997. *Ribbon traces for the 146-bp DNA phosphodiester backbones (brown and turquoise) and eight histone protein main chains (blue: H3; green: H4; yellow: H2A; red: H2B). The views are down the DNA superhelix axis for the left particle and perpendicular to it for the right particle. For both particles, the pseudo-twofold axis is aligned vertically with the DNA centre at the top.*

1.2. Core histones

Core histones are the responsible of the nucleosomal organization of the DNA. The histone octamer is composed of a central heterotetramer of histones H3 and H4, flanked by two heterodimers of histones H2A and H2B. The core histone family is a family of small (11-16.2 kDa) basic proteins which are highly conserved in length and amino acid sequence through evolution. They are composed by a globular domain and histone tails. The globular domain is responsible of mediating histone-histone interactions within the octamer and organizing the two wraps of nuclesosomal DNA. The N-terminal domain is a highly basic (rich in Lys and Arg) 20-35 residue segment that extends from the surface of the nucleosome. Finally, only histone H2A has a long C-terminal tail with a large interface with the histone H3-H4 core domains. This histone tails are not responsible of the individual structure of the nucleosome neither of its stability, but they play an essential role in controlling the folding of nucelosomal arrays into higher order structures, due to their role as DNA-binding modules, but also because they promote inter-nucleosomal histone-histone interactions [13]. Furthermore, histone tails are flexible structures that protrude from the nucleosome, and are ideal surfaces for post-translational covalent modifications.

The complexity of the core histone function increases due to the existence of multiple histone variants and the presence of different and specific histone post-translational modifications (PTMs). Due to the increasing number of histone variants and histone PTMs that are being identified nowadays, and considering that each nucleosome contains two copies of each core histone, the number of theoretically possible combinations and variations in the chromatin primary structure is astronomical. This variability affects the compaction of the chromatin and the interaction of other proteins with nucleosomes, shifting the equilibrium between different chromatin states upon cell transcriptional requirements or cellular states.

2. EPIGENETICS

“An epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alteration in the DNA sequence” [14]. Epigenetic modifications include DNA methylation, histone post-translational modifications, incorporation of histone variants, and expression of non-coding RNAs. These traits are mainly acquired under the influence of certain environmental factors. Epigenetic effects have been linked with progression and treatment of cancer, regulation of development and function of the nervous system, gene regulation, cellular stress, nutrigenomics, aging, and DNA repair.

2.1. DNA METHYLATION

DNA methylation of cytosine residues occurs by the addition of a methyl group to the carbon 5 position of cytosine (5-methylcytosine, 5mC) [15]. Most of DNA methylation occurs within CpG dinucleotides, although it has been recently shown that it can also happens outside the CpG context [16]. CpG islands are regions of DNA greater or equal to 500 bp with a G+C content ≥ 0.55 and an observed/expected presence of CpG ≥ 0.6 [17]. They are mainly found at the 5' regulatory regions of around 60% of the genes, but also in gene bodies, intergenic regions or regulatory regions. DNA methylation is catalyzed by DNA methyltransferases (DNMTs). DNMT1 is responsible for the maintenance of DNA methylation, while DNMT3A and DNMT3B are responsible of *de novo* methylation and are also required for methylation maintenance. Demethylation process remains still elusive, but it is now more widely accepted that demethylases exist. Active demethylation can be achieved in the context of cell division or DNA repair through the excision of the methylated base and the insertion of an unmethylated one. Moreover, different enzymes such as ten-eleven translocation (TET) methylcytosine dioxygenases, activation-induced cytidine deaminase (AID) and thymine DNA glycosylase (TDG) also catalyze DNA demethylation (Figure I.3) [18, 19].

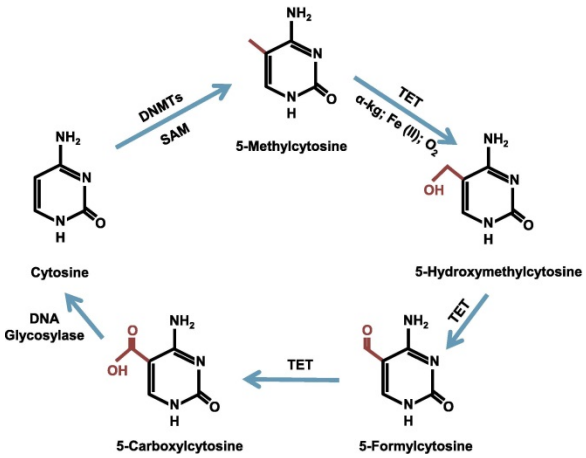


Figure I.3. Active cytosine demethylation through hydroxymethylation by TET. *5-hydroxymethylcytosine (5hmC) has been recently proposed as a player in the removal of methyl groups from cytosine bases. TET enzymes convert 5mC to 5hmC and other intermediates that are finally excised by a DNA glycosilase, returning the cytosine to its unmodified form [19].*

DNA methylation is often associated with transcriptional silencing but recent evidences suggest that the function of DNA methylation seem to vary within context and the genomic localization. Most CpG islands located at the TSS are not methylated in somatic cells and methylation at promoters only occurs when genes should be kept in a repressed state for long-term, like in the case of imprinted genes, X-chromosome inactivation and specific germ cell genes. Methylation in the gene body stimulates transcriptional elongation and has a role in splicing events. Methylation in repetitive regions, such as centromeres, is important for chromosomal stability and suppressing transposon expression. CpG methylation at enhancers

is variable but a link between methylation and reduced activity of the enhancer has been established. At insulators (elements that block the interaction between an enhancer and a promoter), DNA methylation seem to block the binding of CTCF, affecting insulator activity [18]. The precise mechanism by which CpG islands regulate their function is still controversial. However, it is suggested that specific histone modifying enzymes, methyl-CpG binding proteins or other chromatin proteins could be recruited in CpGs by the recognition of either methylated or un-methylated DNA, leading to different chromatin states with varying propensities for gene activation or repression. For example, unmethylated CpG islands recruit H3K36-specific lysine demethylase enzyme (KDM2A), which removes H3K36 methylation, creating a CpG island that is uniquely depleted of this modification [20, 21].

Abnormal methylation has been mainly related with cancer disease in retinoblastoma, colon, breast, lung and ovarian cancer, but also in neurological, metabolic, cardiovascular and autoimmune disorders. Aberrant transcriptional silencing caused by local hypermethylation at promoters of certain tumor-suppressor genes (TP53, RB1, MLH1, p16 or BRCA1), as well as induced genomic instability by global genome hypomethylation, contributes to cell transformation. Moreover, DNA methylation alteration causes deregulation of non-coding RNAs (miRNAs) that may also play a role in tumor suppression. Specific altered DNA methylation profiles in certain tumors raise the possibility to use this epigenetic alteration for diagnosis, prognosis and prediction of response to therapies [22, 23].

2.2. HISTONE POST-TRANSLATIONAL MODIFICATIONS

Histones can be post-translational modified by phosphorylation at serines (Ser) and threonines (Thr), methylation and acetylation at lysines (Lys) and arginines (Arg), and ubiquitylation, sumoylation and ribosylation at lysines (Lys) [24]. This modifications (more than 100 identified nowadays) mainly occur in the N-terminal region but also in the C-terminal or the globular domain [25-27] (Figure I.4). The combination of these histone PTMs, according to the “histone code” hypothesis, have the potential to regulate chromatin function and transcription by recruiting or repelling specific transcription factors and other regulatory chromatin binding proteins [28].

With the recent application of genome-wide studies based on microarray analysis or high-throughput sequencing, mapping of increasing numbers of histone PTMs is achieved in many organisms and, although our understanding of such modifications is increasing, there are still many open questions regarding their precise role or how they are regulated.

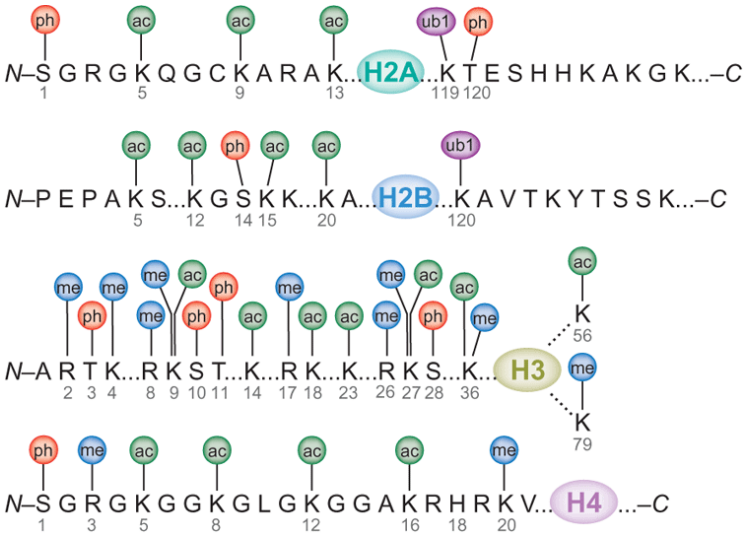


Figure I.4. Post-translational modifications of human nucleosomal histones. *Acetylation (ac), methylation (me), phosphorylation (ph) and ubiquitination (ub1). Globular domains of core histones are represented as colored ovals. Image taken from [29].*

2.2.1. Histone acetylation

Histone acetylation occurs at lysines and has been related with active transcription since early after it was discovered [30]. For instance, H3K9ac is enriched in active promoters and enhancers [31]. It has been proposed that acetylation neutralizes the positive charge of lysine residues, weakening charge-dependent interactions between histones and DNA or other histones [32]. This, increases DNA accessibility to the transcriptional machinery, and is also important in DNA replication and DNA repair processes. Moreover, histone acetylation has also been positively related with pluripotency maintenance (H3K9ac is higher in stem cells than in differentiated cells) and it has also been associated with tumor development and progression.

Histone acetylation is controlled by histone acetyltransferases (HATs) and histone deacetylases (HDACs), which add or remove, respectively, acetyl groups to the histones. The different families of these enzymes are summarized in Table I.1 [33].

Histone AcetylTransferases (HATs):
GNAT
Myst
p300/CBP
Histone DeACetylases (HDACs):
Class I: HDAC1-2
Class II: HDAC4-7 and 9-10
Class III: Sirtuins (1-7)
Class IV: HDAC11

Table I.1. Summary of the major families of HATs and HDACs.

Acetylated lysine residues are recognized by different chromatin associated proteins and chromatin remodeling proteins through a specific protein domain: the bromodomain [34].

Histone	Site	Histone-modifying Enzymes	Proposed Function
H2A	Lys4 (S. cerevisiae)	Esa1	transcriptional activation
	Lys5 (mammals)	Tip60, p300/CBP	transcriptional activation
	Lys7 (S. cerevisiae)	Hat1	unknown
		Esa1	transcriptional activation
H2B	Lys5	p300, ATF2	transcriptional activation
	Lys11 (S. cerevisiae)	Gcn5	transcriptional activation
	Lys12 (mammals)	p300/CBP, ATF2	transcriptional activation
	Lys15 (mammals)	p300/CBP, ATF2	transcriptional activation
	Lys16 (S. cerevisiae)	Gcn5, Esa1	transcriptional activation
	Lys20	p300	transcriptional activation
H3	Lys4 (S. cerevisiae)	Esa1	transcriptional activation
		Hpa2	unknown
	Lys9	unknown	histone deposition
		Gcn5, SRC-1	transcriptional activation
	Lys14	unknown	histone deposition
		Gcn5, PCAF	transcriptional activation
		Esal, Tip60	transcriptional activation
			DNA repair
		SRC-1	transcriptional activation
		Elp3	transcriptional activation (elongation)
		Hpa2	unknown
		hTFIIIC90	RNA polymerase III transcription
		TAF1	RNA polymerase II transcription
		Sas2	euchromatin
		Sas3	transcriptional activation (elongation)
		p300	transcriptional activation
	Lys18	Gcn5	transcriptional activation, DNA repair
		p300/CBP	DNA replication, transcriptional activation
	Lys23	unknown	histone deposition
		Gcn5	transcriptional activation, DNA repair
		Sas3	transcriptional activation (elongation)
		p300/CBP	transcriptional activation
	Lys27	Gcn5	transcriptional activation
	Lys56 (S. cerevisiae)	Spt10	transcriptional activation
			DNA repair
H4	Lys5	Hat1	histone deposition
		Esal, Tip60	transcriptional activation
			DNA repair
		ATF2	transcriptional activation
		Hpa2	unknown
		p300	transcriptional activation
	Lys8	Gcn5, PCAF	transcriptional activation
		Esal, Tip60	transcriptional activation
			DNA repair
		ATF2	transcriptional activation
		Elp3	transcriptional activation (elongation)
		p300	transcriptional activation
	Lys12	Hat1	histone deposition
			telomeric silencing
		Esal, Tip60	transcriptional activation
			DNA repair
		Hpa2	unknown
		p300	transcriptional activation
	Lys16	Gcn5	transcriptional activation
		MOF (D. melanogaster)	transcriptional activation
		Esal, Tip60	transcriptional activation
			DNA repair
		ATF2	transcriptional activation
		Sas2	euchromatin
	Lys91 (S. cerevisiae)	Hat1/Hat2	chromatin assembly

Table I.2. *Summary of histone* ***acetylated*** *residues, the associated modifying enzymes and proposed functions (adapted from Cell Signalling webpage).*

In addition to acetylation, other lysine coenzyme A-dependent acylations have been recently identified: crotonylation, formylation, succinylation, malonylation, propionylation and butyrylation. Although these modifications seem to play a similar role as acetylation because they also neutralize the positive charge of lysine, they must be explored in further detail [35].

2.2.2. Histone methylation

Methylation of histones can occur on lysine or arginine residues, which can be mono- or dimethylated. Moreover, lysines can also be trimethylated. As methylation does not affect the amino acid charge as acetylation, the effect on nucleosome dynamics in transcription, DNA damage and nuclear architecture depends on the binding of regulatory complexes in the methylated residues by the recognition of these modifications through chromo, Tudor, PWWP, MBT or PHD domains [36, 37].

Histone methylation is regulated by histone metyltransferases (HMTs) and histone demetylases (HDMs). Depending if they exert their function on lysine or arginine residues we can distinguish between protein arginine methyltransferases (PRMT) or histone lysine methyltransferases (HKMT). HKMTs, characterized by a catalytic SET domain (except KMT4), present a high degree of enzymatic specificity for the lysine within the substrate, and for the degree of methylation. SUV39H1, Polycomb-repressice complex (PRC) and mixed-linage leukemia (MLL) family proteins are among the most well studied HKMTs [38]. On the other hand, HDMs comprise LSD1 (HDM1A), which specifically demethylates H3K4me2 and H3K4me1 [39], and a broad family of JmjC-containing proteins. PHF8 (KDM7B), a PHD and JmjC domain-containing protein, has been related with H4K20me1 demethylation [40] and JMJD6 has been identified as the first arginine demethylase [41]. So, HDMs also exert their function in a specific manner, depending on the residue and the degree of methylation (Figure I.5).

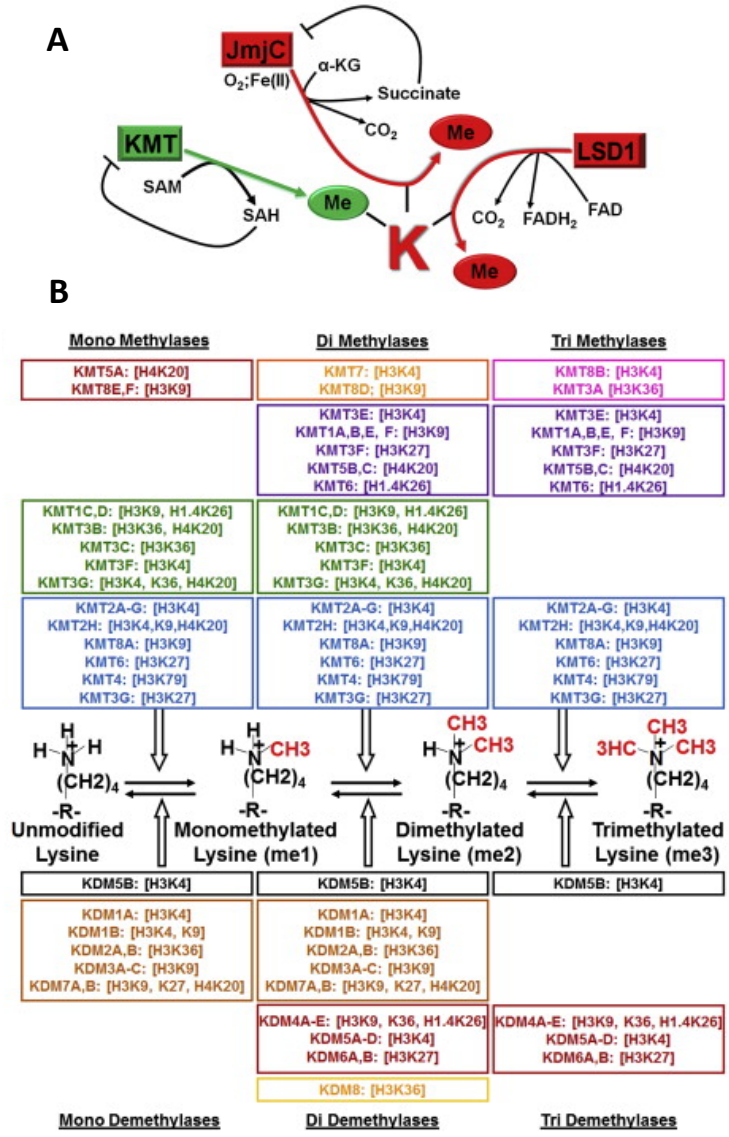


Figure I.5. Histone methylation *[38].*

(A) *General reaction mechanisms of KMTs and KDMs. The catalytic SET domain of KMTs uses S-adenosyl-L-methionine (SAM) as the methyl group donor to methylate histone lysines. LSD1 (KDM1A) histone demthylase contains a flavin adenine dinucleotide (FAD)-dependent amine oxidase domain that demethylates H3K4me2 and H3K4me1. Another class of KDMs uses the JmjC domain to catalyze dimethylation through the oxidation of methyl groups.*

(B) *HKMTs and HKDMs have a high degree of specificity for particular lysine residues and the degree of methylation.*

Thus, the biological function of each of these modifications depends on the residue that is modified, the degree of methylation and the position in the genome (Table I.3). Regarding transcription, histone methylation can be associated with activation (i.e. H3K4me and H3K36me) or repression (H3K9me and H3K27me) in a residue-depending manner [37]. H3K4me3 is enriched at transcriptional start sites of promoters (TSS) and H3K36me2/3 is enriched over gene bodies. As the reduction of these modifications by impairment of their methyltransferases does not critically affect transcription, it has been proposed that H3K4me3 and H3K36me3 may function as regulatory modules in some contexts by inhibiting histone repressive marks such as H3K27me3. Moreover, H3K36me3 has also another role in stabilizing nucleosomes by suppressing nucleosome turnover to prevent aberrant transcription initiation in gene bodies [42]. H3K4me3 is not only present in the promoter regions of active genes. It is also present, together with H3K27me3, in genes "poised" to become activated during

differentiation, forming broad "bivalent domains" [43, 44]. Another well studied modification, H3K4me1, mark distal regulatory regions (enhancers) together with H3K27ac and p300 binding. On the other hand, for H3K9me3 and H3K27me3 repressive marks, associated with heterochromatin formation and Polycomb silencing, respectively, methylation increases the affinity of certain proteins for histone tails promoting nucleosome stability and a closer chromatin state [38]. The characterization of an increasing number of histone methylation marks shows that they are associated with different genomic locations and related with different chromatin states. Furthermore, the regulation of histone methylation by HMTs and HDMs is shown to be important in cancer and neurological diseases [45].

Histone	Site	Histone-modifying Enzymes	Proposed Function
H1	Lys26	Ezh2	transcriptional silencing
H3	Lys4	Set1 (S. cerevisiae)	permissive euchromatin (di-Me)
		Set7/9 (vertebrates)	transcriptional activation (tri-Me)
		MLL, ALL-1	transcriptional activation
		Ash1 (D. melanogaster)	transcriptional activation
	Arg8	PRMT5	transcriptional repression
	Lys9	Suv39h,Clr4	transcriptional silencing (tri-Me)
		G9a	transcriptional repression genomic imprinting
		SETDB1	transcriptional repression (tri-Me)
		Dim-5 (N.crassa), Kryptonite (A. thaliana)	DNA methylation (tri-Me)
		Ash1 (D. melanogaster)	transcriptional activation
	Arg17	CARM1	transcriptional activation
	Lys27	Ezh2	transcriptional silencing
			X inactivation (tri-Me)
		G9a	transcriptional silencing
	Lys36	Set2	transcriptional activation (elongation)
	Lys79	Dot1	euchromatin
			transcriptional activation (elongation)
			checkpoint response
H4	Arg3	PRMT1	transcriptional activation
		PRMT5	transcriptional repression
	Lys20	PR-Set7	transcriptional silencing (mono-Me)
		Suv4-20h	heterochromatin (tri-Me)
		Ash1 (D. melanogaster)	transcriptional activation
		Set9 (S. pombe)	checkpoint response
	Lys59	unknown	transcriptional silencing

Table I.3. *Summary of histone* ***methylated*** *residues, the associated modifying enzymes and proposed functions (adapted from Cell Signalling webpage).*

2.2.3. Histone phosphorylation

As acetylation, phosphorylation brings a negative charge to its modified residue, creating charge repulsion between histones and the DNA backbone, which is also negatively charged [46]. Phosphorylation of histones at serine and threonine has been related with cellular response to DNA double-strand breaks (by γH2A.X) [47], development [48], transcriptional activation (H3S10p is associated with transcriptional activation of induced immediate-early genes [49]), and cell division by altering the binding of chromatin-binding proteins to other modified residues. This is the case of H3S10p, which impairs the binding of HP1 to H3K9me3, releasing it from mitotic chromosomes [50, 51]. (Table I.4)

Histone	Site	Histone-modifying Enzymes	Proposed Function
H1	Ser27	unknown	transcriptional activation, chromatin decondensation
H2A	Ser1	unknown	mitosis, chromatin assembly
		MSK1	transcriptional repression
	Thr119 (D. melanogaster)	NHK1	mitosis
	Ser122 (S. cerevisiae)	unknown	DNA repair
	Ser129 (S. cerevisiae)	Mec1, Tel1	DNA repair
	Ser139 (mammalian H2A.X)	ATR, ATM, DNA-PK	DNA repair
H2B	Ser10 (S. cerevisiae)	Ste20	apoptosis
	Ser14 (vertebrates)	Mst1	apoptosis
		unknown	DNA repair
	Ser33 (D. melanogaster)	TAF1	transcriptional activation
H3	Thr3	Haspin/Gsg2	mitosis
	Ser10	Aurora-B kinase	mitosis, meiosis
		MSK1, MSK2	immediate-early gene activation
		IKK-α	transcriptional activation
		Snf1	transcriptional activation
	Thr11 (mammals)	Dlk/Zip	mitosis
	Ser28 (mammals)	Aurora-B kinase	mitosis
		MSK1, MSK2	immediate-early activation
H4	Ser1	unknown	mitosis, chromatin assembly
		CK2	DNA repair

Table I.4. *Summary of histone* ***phosphorylated*** *residues, the associated modifying enzymes and proposed functions (adapted from Cell Signalling webpage).*

2.2.4. Other histone modifications

Histones can also be modified by ubiquitylation, sumoylation, biotinylation, ADP ribosylation or glycosylation. These modifications are not so well studied as previous ones, but there is increasing data about their role in many different biological processes. While ADP ribosylation is suggested to relax chromatin structure, glycosylation seem to be more related with transcriptional repression. Large modifications such as ubiquitylation or sumoylation impact nucleosome dynamics in a context-dependent manner [42].

Ubiquitylation

Histone	Site	Histone-modifying Enzymes	Proposed Function
H2A	Lys119 (mammals)	Ring2	spermatogenesis
H2B	Lys120 (mammals)	UbcH6	meiosis
	Lys123 (S. cerevisiae)	Rad6	transcriptional activation
			euchromatin

Sumoylation

Histone	Site	Histone-modifying Enzymes	Proposed Function
H2A	Lys126 (S. cerevisiae)	Ubc9	transcriptional repression
H2B	Lys6 or Lys7 (S. cerevisiae)	Ubc9	transcriptional repression
H4	N-terminal tail (S. cerevisiae)	Ubc9	transcriptional repression

Biotinylation

Histone	Site	Histone-modifying Enzymes	Proposed Function
H2A	Lys9	biotinidase	unknown
	Lys13	biotinidase	unknown
H3	Lys4	biotinidase	gene expression
	Lys9	biotinidase	gene expression
	Lys18	biotinidase	gene expression
H4	Lys12	biotinidase	DNA damage response

Table I.4. *Summary of histone* ***ubiquitylated, sumoylated and biotinylated*** *residues, the associated modifying enzymes and proposed functions (adapted from Cell Signalling webpage).*

The combination of all those histone PTMs establishes specific patterns that are related with different chromatin states. For example, particular modifications including H3K4me2/3, H3K36me2/3 and H3K9ac are associated in actively transcribed regions, together with specific transcription factors and RNApolII. On the other hand, other histone PTMs such as H3K27me3 and H4K20me3, are tightly mapped in repressed regions, also enriched in repressive-associated chromatin proteins (Figure I.6).

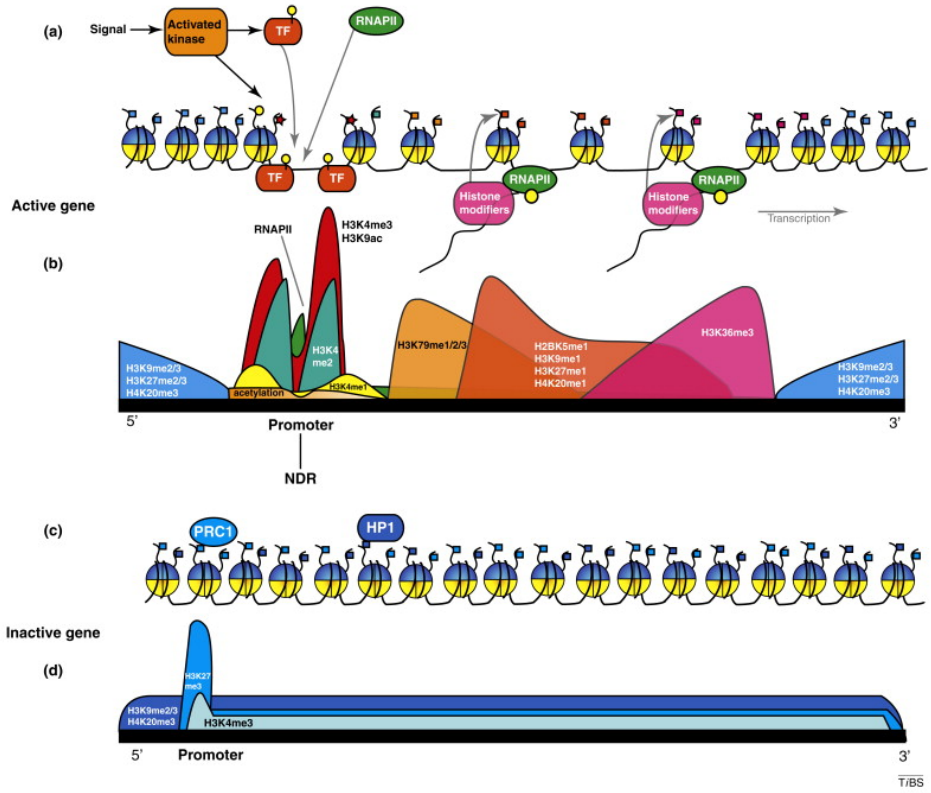


Figure I.6. *Schematic distribution of histone modifications on active and silenced genes, displayed as view on nucleosomes [**(A)** and **(C)**] or over the gene [**(B)** and **(D)**] [52].*

2.3. HISTONE VARIANTS

Apart from the major histone types (H2A, H2B, H3 and H4), in mammals, all of them have diversified into a wide range of histone variants that differ in sequence and structure. This heterogeneity provides another layer of complexity to the variation of chromatin states. Unlike "canonical" histones, which are deposited in a replication-coupled manner during S phase and are expressed from tandem gene arrays, histone variants are incorporated into chromatin throughout the cell cycle in a replication-independent manner and are found in single or low copy number. Histone variants replace "canonical histones" in localized genomic regions, regulating transcription, chromatin structure, DNA damage and repair, epigenetic silencing, and pluripotency and ESC differentiation.

Histone variants are deposited by dedicated chaperones that ensure the correct histone transfer to DNA shielding their positive charge and avoiding unwanted interactions with DNA or other proteins [53-55].

Here, we will focus in the best studied histone variants, members of the H2A and H3 families (summarized in Table I.5). H2B variants are poorly characterized and H4 has few members.

Variant	Function	Chaperones and regulatory factors	Distribution patterns	Deposition	Mutant phenotype
H2A variant					
H2A.Z (H2A.Z.1, H2A.Z.2 and H2A.Z.2.2)	Transcriptional regulation (mostly activation) Genome stability and chromosome segregation Telomere stability (*S. cerevisiae*) Others	SRCAP complex p400 complex (including Tip60) INO80 (negative regulator, *S. cerevisiae*)	TSSs, enhancers, insulators Pericentric chromatin Subtelomeric regions (*S. cerevisiae*)	Replication-independent	*T.thermophila*,*D. melanogaster*,*X. Laevis*, *M. musculus*: lethal (H2A.Z.1) *S. cerevisiae, S. pombe*: transcriptional defects, chromosmome loss (H2A.Z.1)
H2A.X	DSB repair factor recruitment[a] Meiotic gene silencing[a] (*M. musculus*)	Tip60*(D.melanogaster)* FACT Nucleolin	Genome-wide DSB sites[a]	Replication-independent DSB-induced (IRIF)[a]	*M. musculus*: chromosomal aberrations
mH2A (mH2A.1 and mH2A.2)	Transcriptional regulation (mostly repression)	ATRX (negative regulator)	Upstream and downstream of TSS Telomeres SAHFs Xi	Replication-independent	*D. rerio*: brain developmental defects (mH2A.2) *M. musculus*: female liver steatosis (mH2A.1)
H3 variant					
CENP-A	Centromere structure and function Kinetochore assembly (chromosome segregation)	HJURP Mis18 complex	Centromeres	Replication-independent (M/G1)	*M. musculus*: lethal
H3.3	Gene activation Telomere homeostasis Epigenetic memory (*X. laevis*) Spermatogenesis (*D. melanogaster*)	ATRX-Daxx HIRA DEK	Euchromatin (active genes) Promoters and regulatory sequences Telomeric repeats Pericentric heterochromatin	Replication-independent	*D. melanogaster*,*T. thermophila*: sterility *M. musculus*(hypomorphic allele): lethality or fertility defects

Table I.5. *Summary of histone variants. Data refers to mammalian histone variants, unless otherwise specified.* [a] *γ-H2A.X Adapted from [54].*

2.3.1. H2A variants

The H2A family is the largest family of variants among the core histones and is the most structurally diverse, resulting in multiple biological functions. H2A.Z, H2A.X, macroH2A, H2A.B and H2A.J are members of this family.

H2A.Z comprises, in vertebrates, three different isoforms: H2A.Z.1 and H2A.Z.2, arising from different non-allelic genes, and H2A.Z.2.2., a recently uncovered splice variant of H2A.Z.2. Most of the studies regard H2A.Z.1, but not all them necessarily distinguish between isoforms, particularly those based on antibody approaches. H2A.Z participates in different cellular processes, such as transcription regulation, epigenetic memory, heterochromatin boundaries establishment, genome stability and chromosome segregation, and integrity of telomeres. It is enriched at gene promoters and regulatory regions (enhancers and insulators), mainly having a positive role on transcription, probably due to N-terminal tail hyperacetylation. H2A.Z-containing nucleosomes preferentially occupy positions flanking the TSS [56]. However, it is also reported a negative effect in certain genes, may be due to other PTMs [54].

H2A.Z is overexpressed in colorectal, breast, lung and bladder cancers and is suggested to act as an oncogene. An increased chromatin incorporation of H2A.Z due to transcriptional up-regulation or altered chromatin deposition by chaperones could alter gene expression that is relevant to disease progression, as well as to affect other cellular mechanisms regulated by H2A.Z (DSBs or maintenance of telomere integrity).

H2A.X has a longer C-terminal region than the canonical H2A. H2A.X presents various PTMs such as phosphorylation, acetylation and ubiquitylation. Phosphorylation at S139 in the longer C-terminal region (refered as γH2A.X) is carried out by members of the Phospho-Inositide 3-Kinase-related protein Kinase (PIKK) family (ATM, ATR, and DNA-PK). H2A.X S139ph occurs upon DNA damage, spreading over a 2-Mb domain surrounding the DSB, where repair proteins, histone modifying enzymes, and chromatin remodeling complexes are accumulated in subnuclear foci called IRradiation-Induced Foci (IRIF) [57, 58]. Due to its function in response to DSBs, H2A.X is seen as genome caretaker and a tumor suppressor in certain genetic contexts.

MacroH2A or mH2A is the most structurally distinct histone variant due to the presence of a large macro domain of ~30 kDa at its C-terminal end that protrudes from the nucleosome [59]. There are three isoforms reported in mammals: mH2A.1.1 and mH2A.1.2, splicing variants encoded by the same gene, and mH2A.2. Since its identification, it has been suggested a

repressive role for mH2A, as it contributes to silencing of the X chromosome (Xi), is associated with Senescence-Associated Heterochromatic Foci (SAHFs), and is involved in the inactivation of imprinted genes and also autosomal genes [54]. However, it has also a positive effect on transcription in a context-dependent manner [60]. Particular mH2A PTMs or chromatin binding factors are thought to determine the positive or negative role of mH2A in transcription. For instance, S137ph is suggested to have a positive role in a cell cycle-dependent context [61]. Furthermore, mH2A also seems to be involved in differentiation. Knock-down analyses show that differentiation is impaired through a defect in the silencing of pluripotency-related genes [62, 63]. Due to its reported role in regulating cellular proliferation, mH2A is also starting to be considered as a tumor suppressor gene, with reduced expression in several tumor types.

2.3.2. H3 variants

There are reported eight histone H3 variants in humans (H3.1, H3.2, H3.3, H3t (H3.4), H3.5, H3.X, H3.Y, and CENP-A). H3.1 and H3.2 are considered “canonical” H3, some are tissue specific (H3t and H3.5 are expressed in testis), and others primate-specific (H3.X and H3.Y). H3 variants are also post-translational modified and differently located in the genome by specific mechanisms.

H3.3 differs from canonical H3.1 and H3.2 by five and four aminoacids, respectively, in a region important in mediating the interaction with regulatory factors and histone chaperones. Indeed, H3.3 is differently post-translational modified than canonical H3, as it is the case for S31, specifically phosphorylated in mitosis [64]. H3.3-containing nucleosomes are less stable than canonical ones, and even less if they are combined with H2A.Z. Thus, H3.3/H2A.Z nucleosomes are found in promoters and enhancers of highly active genes, mediating an open chromatin state accessible to TFs [65]. Consistent with this fact, H3.3 has been described as a replacement histone, associated with transcriptionally active chromatin, deposited in promoters and gene bodies of active genes, as well as in regulatory sites of both active and inactive genes [66-68]. HIRA (HIstone cell cycle Regulation-defective homolog A) is the responsible chaperone to deposit H3.3 at promoters throughout the cell cycle. Furthermore, H3.3 is also located in telomers and pericentric heterochromatic regions by another chaperone, ATRX and its co-factor Daxx [67].

H3.3 and its chaperones have been related to cancer. H3.3 is mutated and overexpressed in certain human tumors.

CENP-A is a highly specialized variant that is located in centromeres of all eukaryotic cells. It is much less abundant in the cell than canonical H3, but has a crucial role in both centromere formation and maintenance, and kinetochore assembling during chromosome segregation in mitosis and meiosis. Structurally, two N-terminal regions of CENP-A differ from canonical H3, contributing to the more compacted unique centromeric chromatin structure, and to the specific binding of other centromeric proteins such as CENP-B [54]. In humans, CENP-A is incorporated at centromeres by a specific chaperone, HJURP (Holliday Junction Recognition Protein), during G1 phase of the cell cycle [69, 70] in a process that involves a three-step mechanism: (1) recognition and licensing of centromeres, (2) loading of newly synthesized CENP-A with the help of chaperone proteins, and (3) maintenance of newly incorporated CENP-A (reviewed in [71]).

Given the role of CENP-A in centromeres and kinetochore structures, it is not surprising that its deregulation could lead to chromosomal instability and cancer. In fact, many tumors present CENP-A, but also HJURP, overexpression. It is speculated this CENP-A overexpression could lead to miss-localization of the protein along the chromosome arms, altering gene expression and/or other cellular processes.

As in the case of histone PTMs, the specific location of histone variants, driven by dedicated chaperones in a replication-independent manner, establishes different chromatin states leading to the regulation of specific functions and the participation in different cellular processes (Figure I.7).

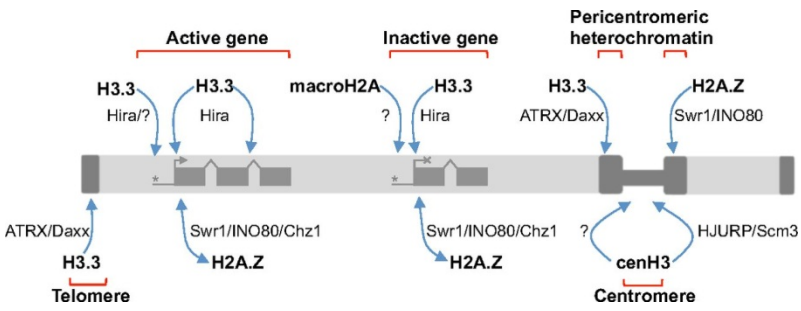


Figure I.7. Summary of histone variants genomic localizations by their dedicated histone chaperones and chromatin remodelers. *Asterisks mark enhancers and question marks indicate unknown deposition pathways.*

2.4. EXPRESSION OF ncRNAs

The central dogma on molecular biology “DNA makes RNA makes PROTEIN” in a sequential non-reversible flow transmission of genetic information has been contradicted during the last years. It is nowadays accepted that up to 70% of the genome is transcribed into RNA that does not serve as template for proteins [72, 73]. RNA molecules that do not encode for proteins are named non-coding RNAs (ncRNAs) and can be divided in several classes depending on their function in protein biosynthesis (tRNAs, rRNAs, snRNAs, snoRNAs), but also in their role in regulating gene transcription and other cellular processes. For instance, microRNAs (miRNAs), molecules about 22 nucleotides long, are able to post-transcriptionally regulate gene expression by recognizing and degrading target mRNA sequences [74], and siRNAs have been related with DNA methylation and histone modifications. A particular class of ncRNAs are the long intergenic non-coding RNAs (lincRNAs), which are modestly conserved transcripts generated through similar pathways than those of protein-coding genes [75, 76].

Non-coding RNA function is extensively studied, and these transcripts are demonstrated to regulate a myriad of cellular processes. They regulate transcription in *cis* but also in *trans*, are involved in dosage compensation and genomic imprinting, are linked to heterochromatin formation via an RNA interference (RNAi) pathway, regulate higher-level nuclear organization and chromosome dynamics, are implicated in cellular differentiation and development, etc. Furthermore, aberrant expression of ncRNAs is linked with pathological processes like cancer [77].

3. CHROMATIN REMODELING FACTORS

Another layer of controlling chromatin compaction relies on the function of chromatin remodelers. These are ATP-dependent multisubunit protein complexes that catalyze the insertion, displacement, restructuration or eviction of nucleosomes, and the replacement of histones, regulating the accessibility of transcription factors to specific DNA sequences depending on the transcriptional requirements, or in replication and DNA repair processes (Figure I.8A). Chromatin remodeling complexes are divided into four groups depending on the ATPase catalytic subunits and domain structure: ISWI, Chd1, Snf2 and Swr1 (Figure I.8B). These catalytic subunits are evolutionary related to the SF2 superfamily of DNA helicases, so it is believed that they use DNA-translocase activity to “pump” DNA around the histone octamer, resulting in nucleosome mobilization [78, 79].

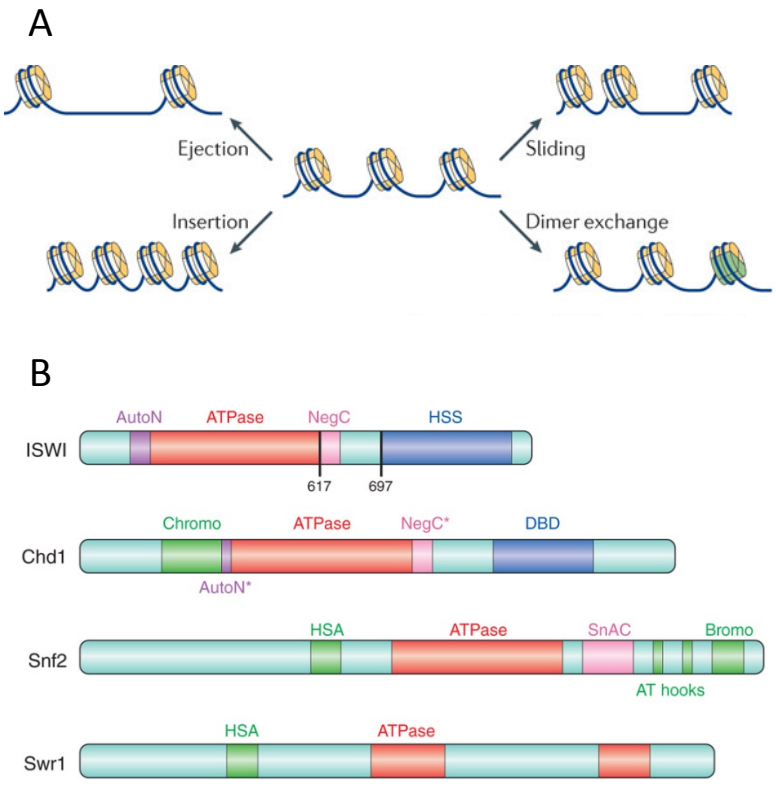


Figure I.8. (A) Actions of chromatin remodeling factors *[80].* ***(B) Families of chromatin-remodeling ATPases with the different domains in different colors.*** *These families are classified according to the different catalytic subunits (in red), but also to the presence of characteristic sequence motifs such us the Chromo and Bromo domains, which mediate the recognition of specific histone PTMs* [79].

The **SWI/SNF** complex, first identified in yeast, (BAF or PBAF in humans) consists in 8 or more subunits. The catalytic subunits Swi2/Snf2 (Sth1 in yeast) and BRM (BRG1 in humans) are accompanied by other auxiliary subunits that modulate targeting, assembly and regulation of specific functions of those complexes. BAF, together with PCAF histone acetyl transferase, participate in the regulation of the mouse mammary tumor virus (MMTV) promoter upon progesterone stimulation by mediating the displacement of H2A/H2B from the nucleosome B [81].

The **ISWI** catalytic subunit (Snf2H and Snf2L in mammals) also interacts with other auxiliary subunits forming different remodeling complexes with different properties. For example, NURF complex, through its histone methyltransferase activity, has been also implicated in nucleosome remodeling upon progesterone stimulation in the MMTV by promoting phosphorylated H1 ejection from regulatory nucleosome B [82]. The action of ISWI remodelers depends on the interaction with the N-t tail of histone H4, through the recently identified AutoN domain [79, 83].

The **CHD** (Chromodomain-Helicase-DNA binding) family is defined by the existence of two chromodomains in the N-terminal of the ATPase domain. They present heterogeneous functions. For instance, the NURD (NUcleosome Remodeling and Deacetylase) complex

contains Chd3/Chd4, histone deacetylases (HDAC1/2), and methyl CpG-binding domain (MBD) proteins, and is implicated in the transcriptional repression of certain genes during development [84]. Furthermore, together with Isw1, it also maintains nucleosomal spacing in yeast [85].

The **SWR1** or **INO80** family contains more than 10 subunits and is characterized by the split of its ATPase domain. They have diverse functions in transcriptional activation, DNA repair, recombination and replication. It is worth noting the ability of these complexes in replacing canonical H2A-H2B dimers with H2A.Z-H2B dimers by the insertion of H2A.Z [86].

4. NUCLEOSOME POSITIONING AND DYNAMICS

Nucleosomes are organized in the genome into nonrandom regularly spaced arrays. Since the recently development of nucleosome mapping techniques such as MNase-seq, the understanding on the genome-wide nucleosome pattern positioning is increasing. Many studies have been performed in simple eukaryotes, but data in mammals and humans is rapidly arising. Nucleosome positioning can vary from perfect positioning (when the nucleosome is located at the same place in all cells of a population) to random positioning (when every cell in the population presents a different location at a given time). In yeast, it is generally accepted that many genomic features are depleted of nucleosomes: promoters, enhancers and terminator regions. Moreover, nucleosomes occupy preferred positions in genes but also in non-gene regions. Active promoters present a nucleosome-depleted or nucleosome-free region (NDR or NFR) immediately upstream the TSS. This region is flanked by two well-positioned nucleosomes (-1 and +1) and the positioning of subsequent nucleosomes upstream or downstream those two progressively decrease (Figure I.9B). The determinants of nucleosome positioning are still controversial but it is demonstrated that it is not achieved by any single factor, rather by the combination of several factors: DNA sequence, DNA-binding proteins, nucleosome remodelers and the RNApolII transcriptional machinery [87, 88].

DNA sequence dictates in part nucleosome positioning. In the rotational positioning of a nucleosome along the DNA helix, more bendable sequences (AT and TA) are in contact with histones and show ~10-bp periodicity, while less bendable sequences (GC) are solvent-exposed, determining the rotational positioning of the nucleosome [89] (Figure I.9A). Regarding translational positioning, Poly(dA:dT) sequences disfavor nucleosome positioning and are typical of linker DNA sequences and promoters in yeast, while G/C-rich sequences present higher nucleosome occupancy [90-92].

However, despite DNA sequence is important in establishing nucleosome depleted regions in promoters and this can be reproduced in *in vitro* experiments with purified histones and DNA, other aspects of the *in vivo* organization of the nucleosome cannot be reconstituted in vitro. Thus, strong positioning of the +1 nucleosome is not observed, nor the formation of highly positioned arrays around it [93]. This supports the idea that other factors beyond the DNA sequence are involved in determining nucleosome positioning *in vivo*. ATP-dependent chromatin remodelers are reported to have a crucial role in dictating nucleosome positioning. Zhang et al. *in vitro* experiments showed that assembled nucleosomes on yeast DNA in the presence of cell extracts as a source of chromatin associated proteins generated positioned nucleosomes only when ATP was added in the extract [94], indicating the importance of ATP-dependent chromatin remodelers. However, the exact position of those nucleosomes, as well as the extension of the downstream nucleosomal array did not completely resembled the *in vivo* conditions. Thus, RNApolII transcription is also involved in establishing nucleosome positioning. In the TSS, the preinitiation complex (PIC), together with associated factors, is demonstrated to have a role in fine tuning the position of the +1 nucleosome, while in the coding region, the elongating PolII machinery is involved in the downstream formation of the nucleosomal arrays. Finally, nucleosome spacing is also controlled by nucleosome remodeling factors. In yeast, strains lacking several remodelers (Isw2) presented drastic alteration of nucleosome positioning patterns. Moreover, when introducing exogenous genomic DNA from one yeast species into another one, the nucleosome spacing is characteristic of the host's, not of the donor species, suggesting that specific remodelers of the host species are involved in the nucleosome positioning of the foreign DNA [88].

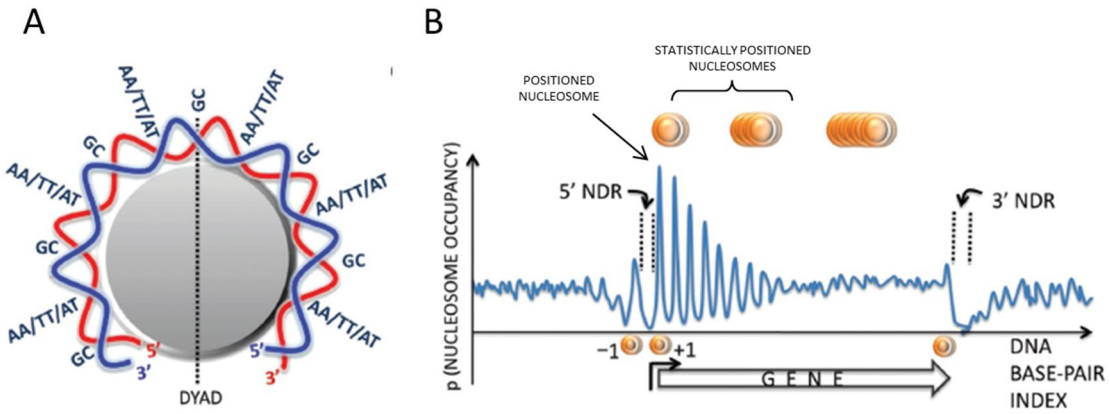


Figure I.9. Stereotypical nucleosome positioning, mainly based on yeast studies. (A) *Nucleosome rotational positioning showing repeated alternating motifs of AA/TT/AT and GC dinucleotides.* ***(B)*** *Typical nucleosomal positioning in genes with the presence of the 5' and 3' nucleosome depleted regions (NDR). The 5' NDR is flanked by tightly positioned nucleosomes with other phased nucleosomes around. Figure adapted form [95].*

Nucleosome positioning studies have been mainly performed in simple eukaryotes. In mammals, genome is more complex and, like in the case of yeast, DNA sequence alone cannot explain nucleosome positioning, especially if we consider that within a same organism, different cell lines sharing the same genome present different average nucleosome spacing. This variability in linker lengths is thought to be related to the required transcriptional activity of a given tissue or cell type. Thus, higher transcriptional active cells present reduced nucleosome-repeat length (NRL), while inactive genomes have long linkers [87]. This linker DNA length variability has been related to differences in linker histone abundance and expression of its variants (see below) [96, 97]. Interestingly, within a same cell, a relation between different chromatin states and internucleosomal spacing is also evident. Comparing different primary human cell lines, Valouev et al. reported that active histone modifications (H3K4me1 and H3K27ac) are associated with shorter internucleosomal linker DNA (~30 bp), while repressive marks (H3K9me3 and H3K27me3) show large spacing (~58 bp) [92]. In the same report, authors estimate that < 20% of nucleosomes are highly positioned *in vivo* and that sequences at the center of highly positioned nucleosomes are enriched in G/C nucleotides, whereas flanking repelling sequences are A/T rich. On the other hand, they show that CpG tetranucleotides and CpG islands are relatively nucleosome free, arguing for a sequence-independent function of CpG in promoting nucleosome depletion at promoters. Finally, binding sites for CTCF insulator and NRSF/REST repressor are also nucleosome depleted and flanked by highly positioned nucleosomes.

Although positioned arrays of nucleosomes have been reported by different groups in human promoters and binding sites for CTCF or other proteins (see above), the importance of DNA sequence in establishing nucleosome positioning is still controversial due to the limited resolution of nucleosome maps generated by single-end MNase-seq [92, 98]. To overcome this issue, paired-end MNase-seq on seven human lymphoblastoid cell lines (LCLs) was recently achieved, obtaining highly precise nucleosome maps [99]. This study shows that, although the translational positioning of nucleosomes is weak, most of them (84%) are significantly more positioned than expected by chance, and some moderately or strongly positioned. According to this data, almost half of the genome is organized in nucleosome arrays, mainly in heterochromatic regions, but also in promoters, active insulators and enhancers, and also flanking transcription factor binding sites. Finally, it was also found, like in other organisms, a clear 10 base pair periodicity in the rotational positioning of nucleosomes *in vivo*.

Furthermore, other recently developed techniques are starting to provide more information about nucleosome positioning and its relation with several chromatin structures or chromatin-

related processes. NOMe-seq (from nucleosome occupancy and methylome sequencing) obtains nucleosome positioning and DNA methylation data from the same DNA molecule, achieving valuable information regarding gene expression regulation. In a recent report using human IMR90 fibroblasts [100], Kelly TK et al. used this method to show distinct nucleosome/methylation configurations associated with different genomic regions. They showed that positioned nucleosomes in CTCF sites are unmethylated while the linker regions between them are methylated. At the level of promoters, they reproduced the NDR upstream the TSS, as well as the downstream strong-positioned nucleosomes for the higher active promoters, claiming that these features are predictors of gene expression, for both CpG and non-CpG containing promoters. Their data also suggests that the NDR upstream the TSS is large enough to accommodates multiple nucleosomes.

Altogether, nucleosome occupancy and positioning are potential mechanisms for controlling transcription and other DNA-templated processes. The understanding of this organization combined with other epigenetic data (DNA methylation, histone PTMs, histone variants replacement, etc.) will improve our knowledge in how are genomes regulated.

5. CHROMATIN SPATIAL ORGANIZATION

The combination of all the different features presented until now enables DNA to be tightly packed into the cell nucleus into such organized way that the genomic information encrypted is made accessible by different complex mechanisms at the proper time to regulate different processes (gene transcription, DNA replication and DNA repair). This organization is therefore responsible for the specification and transition of different cell lineages within multicellular organisms, as well as for the occurrence of many diseases like cancer.

Classically, two different types of chromatin states were considered: *euchromatin* and *heterochromatin* (first termed by Emil Heitz in 1928) (Figure I.10A). Euchromatin is defined as the most active and less compacted portion of the genome (“open” chromatin), presenting higher gene concentration, higher transcriptional activity, and associated with active histone PTMs. On the other hand, heterochromatin represents the more compacted fraction of the genome (“close” chromatin). It is characterized by the presence of centromeric and telomeric repetitive sequences (satellite DNA) and a repressive transcriptional state, marked by repressive histone modifications and the presence of HP1 (Heterochromatin Protein 1). This

type of chromatin can be subdivided within *facultative heterochromatin* (silenced chromatin that can be reversely converted to an "open" state to enable transcription when needed, during cell cycle, differentiation, development, etc.) and *constitutive heterochromatin* (permanently silenced chromatin, typical of centromeres and telomeres). In the nuclei, heterochromatin is placed in the nuclear periphery, while euchromatin is more localized towards the center [101, 102].

However, this first classification of chromatin has been reassessed, and it has been concluded that chromosomes are linearly segmented into hundreds of different domains. Thanks to de development of high-resolution genome-wide methods (ChIP or DamID, combined with sequencing) we know that these chromosome domains present different protein and histone modification composition. Sometimes, this segmentation is the result of DNA sequence composition if we consider that genes are not randomly distributed along the genome. Thus, clustered genes in defined domains tend to be more active than isolated genes in gene-poor regions [103].

The existence of different chromatin regions defined by particular histone PTMs and regulatory protein composition is broadly reported. In *Drosophila* and mammals, proteins from the Polycomb group (PcG) form the multisubunit Polycomb-repressive complexes (PRC1 and PRC2). PRC2 catalyzes H3K27me3 and PRC1 bind to this histone mark through Polycomb (Pc). Both PcG proteins and H3K27me3 are found in scattered domains along the genome, most of which are transcriptionally inactive and related with development. On the other hand, H3K9me2/3 associates with HP1 proteins in pericentric and telomeric regions, where they have structural roles. These domains are referred as LOCKs (Large Organized Chromatin K9 modifications) [104]. Other chromatin domains are formed by the association of genomic regions with relatively fixed nuclear structures. Lamina-associated domains are large regions of about 0.1-10 Mb that interact with the nuclear lamina (NL). They cover near 40% of the genome and are gene-poor, transcriptionally inactive, associated with histone repressive marks (H3K9me2 and H3K9me3) and flanked by CTFC and CpG islands, thus representing a strongly repressive chromatin environment [105] (Figure I.10C). Many LADs are highly conserved between different species, but some of them seem to be cell-type specific [106]. How these domains are established and directed to the NL is still poorly understood, but the combination of DNA sequence (A/T-rich DNA) and chromatin modifications recognized by NL-associated proteins could mediate this organization. LAD structure and composition resembles to another type of chromatin domain associated with the nucleolus (nucleolus-associated

domain – NAD) [107]. In fact, it is proposed that the same type of repressive chromatin could distribute between the NL and the nucleoli in a random manner.

Thanks to the combination of genome-wide chromatin maps and computational approaches, we can nowadays elucidate whether the genome is segmented in a limited number of chromatin states and which is their composition and function. A pioneer study in *Drosophila* Kc167 cells gave light to this issue recently [108]. By analyzing genome-wide DamID data of 53 chromatin components, five chromatin types composed of unique combinations of proteins were defined: HP1 ("green") and Polycomb ("blue") heterochromatin types, two classes of transcriptionally active chromatin ("yellow" and "red"), and a "black" strongly repressive chromatin not associated with HP1 or Polycomb, that belongs to 48% of the genome (Figure I.10B). Similar studies have been carried in humans after this first attempt, defining as well different chromatin environments composed by the combination of several chromatin regulators [109, 110]. Like in the case of LADs, the establishment and maintenance of these chromatin domains is still poorly understood. However, apart from their role in regulating promoter activity by a distal enhancer element, insulator-binding proteins such as CTCF seem to delimit chromatin domains and contribute to the boundaries of some topological domains.

The two-dimensional chromatin fiber is packed in the nucleus within a 3-dimensional organization, providing another layer of complexity for chromatin-based processes. Nuclear architecture and 3D structure of chromosomes is starting to be resolved after the development of microscopy approaches like FISH (fluorescence *in situ* hybridization) and, more recently, crosslinking and intermolecular ligation assays ("3C"-based approaches) [111]. These techniques allow investigating spatial associations between chromatin regions that seem to be far away regarding the linear genome, but that are in close contact within the spatial distribution in the genome. Since the first microscopy observations of chromatin, it was clear that chromosomes were organized in distinguishable segments in interphase, called "chromosome territories" [112] (Figure I.10D). Although, it is considered that each chromosome occupy an individual territory (Figure I.10E) and that long-range interactions are mainly between regions of the same chromosome (*in cis*) and restricted to the same chromosome arm, robust associations *in trans* between different chromosomes are also reported for gene-dense regions. Moreover, FISH and 3C approaches show that active gene-rich domains associate to each other in the nuclear space. This is also true for inactive loci but in much lesser extent [104, 113, 114]. With the last improvement of Hi-C and 5C technologies, high-coverange and resolution chromatin interaction maps have been generated, identifying large megabase-size local chromatin interaction domains, termed as "topologically associating

domains" (TADs) [114-117] (Figure I.10F). In mammalian IMR90 cells, the boundaries of the topological domains are enriched in CTCF, housekeeping genes, transfer RNAs and short interspersed element (SINE) retrotransposons, suggesting that they have a role in establishing such domain interactions within the genome. Interestingly, by comparing human and mouse, it has been seen that these domains are stable between cell-types and highly conserved across species [115].

Altogether, the understanding of nuclear architecture and chromatin structure will give light to resolve how is genome organized and regulated. Importantly, it is recognized that alterations in large-scale chromatin folding can occur in cancer cells, so efforts in elucidating the mechanism by which nuclear structure is established could help in the way to treat cancer [118].

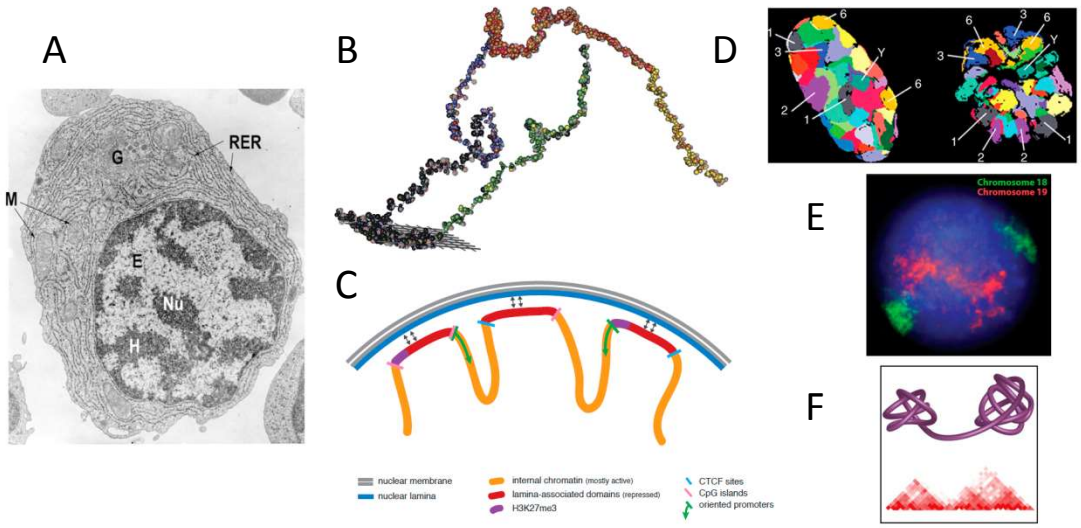


Figure I.10. Evidences for different levels of chromatin organization. (A) *Electron microscopy image of a cell. Heterochromatin (H) appears as electron dense in contrast to euchromatin (E). (Nu) Nucleus; (M) Mitochondria; (G) Golgi complex; (RER) Rough endoplasmic reticulum.* ***(B)*** *Representation in different colors of the five chromatin states identified in Drosophila cells [108]. "Black" chromatin has preference to be associated with the NL (grey lattice) [119].* ***(C)*** *Cartoon representation of lamina associated domains (LADs) and their characteristic features [105].* ***(D)*** *Chromosome territories. Simultaneous delineation of all chromosomes in a human fibroblast nucleus (left) and a prometaphase rosette (right) by multi-color FISH [112].* ***(E)*** *Evidence for the radial organization of chromosomes in the nucleus. FISH image were gene-rich chromosome 19 localized towards the center of the nucleus, while gene-poor chromosome 18 sits close to the periphery [101].* ***(F)*** *Schematic illustration of topological domains (top) and Hi-C data represented as frequency of chromatin interactions (bottom).*

6. CHROMATIN HISTORY

Since the first observation of the cellular nucleus by Robert Brown in 1831 to the current conception we have nowadays about chromatin, many discoveries and landmarks occurred. Figure I.11 summarizes this long, exiting and unfinished trip.

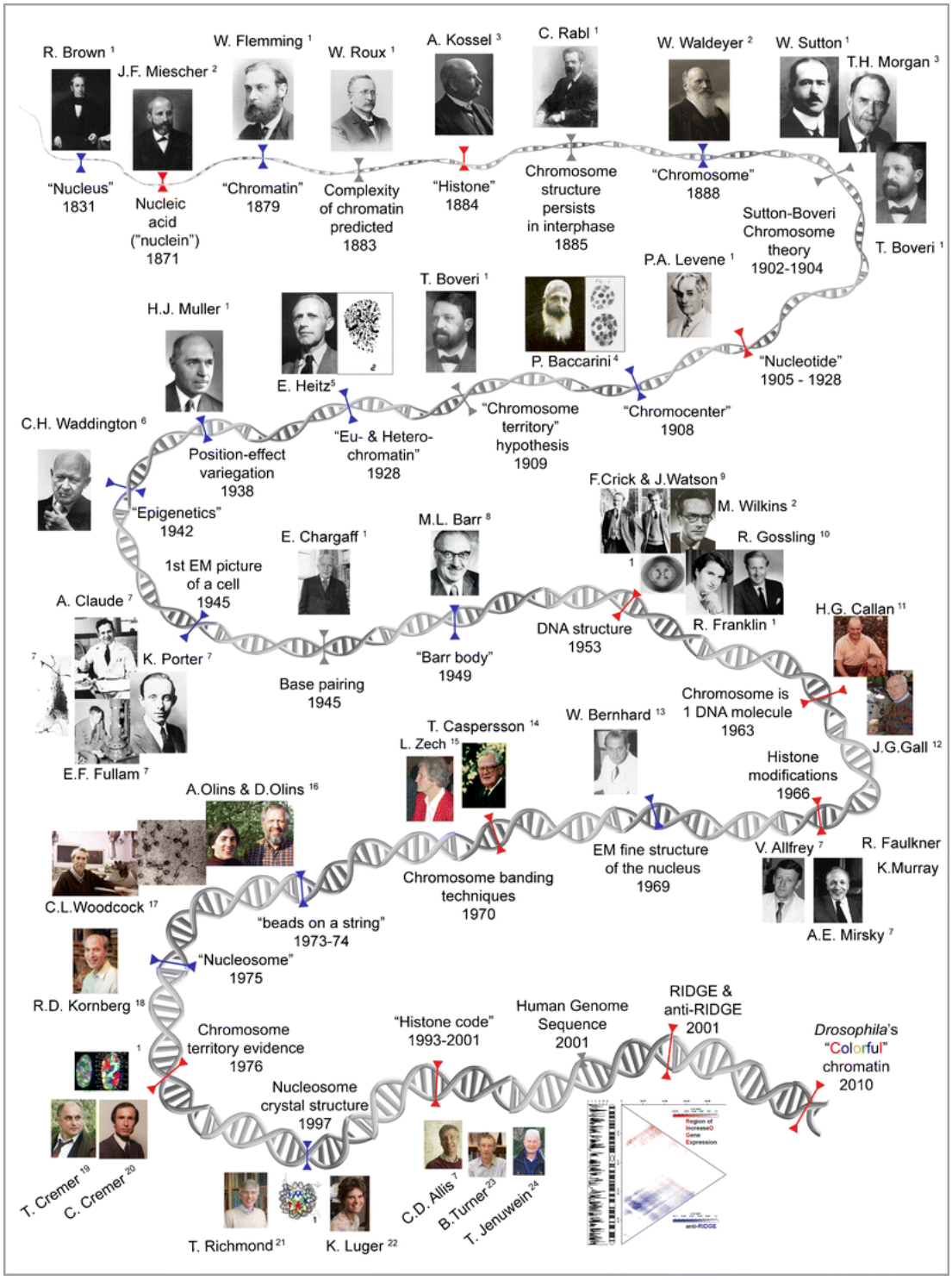


Figure I.11. Timeline compilation of landmark discoveries and concepts (gray) on molecular (red) and cellular (blue) aspects of chromatin. *Taken from [120].*

7. LINKER HISTONE H1

As it has been previously described, the linker histone H1 sits at the base of the nucleosome near the entry and exit sites and is involved in the folding and stabilization of the 30 nm chromatin fiber, allowing a higher degree of DNA compaction. Since many years, most of the work in chromatin structure and function has been focused on the nucleosomal core, composed by the core histones H2A, H2B, H3 and H4, and less attention has been paid to linker histone H1, regarding its structure, function, post-translational modifications, etc.

7.1. LINKER HISTONE H1 FAMILY

Compared with core histones, which are highly conserved in evolution, linker histone H1 family is more divergent [121] and many *subtypes* or *variants* exist in several organisms (note that the "variant" term could also refer to products of polymorphic alleles), from one variant in simple eukaryotes to eleven variants in humans or mouse (Figure I.12). The diversity of histones and the identification of an increasing number of variants have led to confusion in naming since their discovery. Many attempts to unify nomenclature have been done since many years ago. Recently, a unified phylogeny-based nomenclature has been proposed for histone variants [122]. Throughout this thesis I will use the new nomenclature, except for H1x variant, to which I will refer as H1X instead as H1.10 as it is proposed (Table I.6).

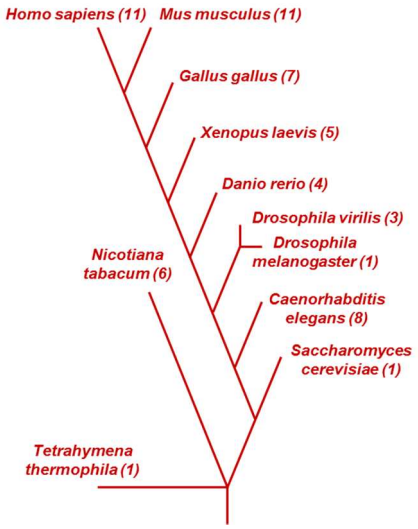


Figure I.12. Evolutionary tree showing the number of H1 variants in different species *[123].*

Humans present eleven histone H1 variants. They can be classified according to different criteria (expression, cell-cycle dependency, and gene location). Thus, while seven of them are somatic subtypes (H1.1 to H1.5, H1.0 and H1X), others are restricted to germ cells, with three testis-specific variants (H1t, H1T2 and HILS1) and one oocyte-restricted variant (H1oo). Among the somatic histone H1 variants, H1.1 to H1.5 are expressed in a replication-dependent manner, whereas H1.0 and H1X are replication-independent. H1.2 to H1.5 and H1X are ubiquitously expressed, H1.1 is restricted to certain tissues, and H1.0 accumulates in terminally differentiated cells. Regarding gene location, H1.1 to H1.5-encoding genes are clustered in a region of chromosome 6, together with the core histone genes, whereas H1X and H1.0 are located on chromosome 3 and 22, respectively. [53, 124, 125].

Genes located in major clusters of chromosome 6 (6p21-p22) are encoded by individual intronless genes, with short 5' and 3' ends. Transcripts lack polyA tails but contain a 3' stem-loop sequence that allows for rapid translation during DNA replication. On the other hand, isolated genes such as H1.0 and H1X are also intronless, but their mRNA is polyadenylated. It is interesting to note that, although clustered genes share same chromosome location and gene structure, they are not expressed equally, as H1.1 and H1t ((TS) H1.6) present tissue-specificity and expression of other subtypes fluctuate different across the cell cycle [126]. Thus, different H1 variant transcription might be tightly regulated in order to obtain proper expression of those variants in different tissues or cells, but also during cell-cycle and differentiation.

Little is known about histone H1 transcriptional regulation, but it is reported that specific sequences in their promoters modulate binding of transcription factors and chromatin proteins, such as H1TF1 and H1TF2 or HIRA. H1 promoters have a CCAAT box upstream of a TATA box, and a GC-rich motif of varied sequence and position upstream of the CCAAT box. They also contain an H1-specific element, the H1 box (AAACACA), which binds the transcription factor H1TF2. This H1 box is missing in the H1.0 promoter [127-129]. More recently, a computational analysis based on genome-wide data suggests E2f1 and E2f4 as master regulators of all histone genes, including H1. Moreover, Zfx is proposed to negatively regulate H1. Other factors such as Smad1, Smad2 and YY1 are found in H1 promoters in a more cell-type specific manner and with some differences between variants [130].

Histone Gene Cluster 1 (Replication-dependent)						
	Human		**Other Mammals**			
gene symbol	**Albig and Doenecke [131]**	**Ohe and Iwai [132]**	**mouse gene symbol**	**Parsegian and Hamkalo [133]**	**Seyedin and Kistler, Lennox and Cohen [134, 135]**	**NEW UNIFIED [122]**
HIST1H1A	H1.1		Hist1h1a	H1a	H1a	H1.1
HIST1H1B	H1.5	H1a	Hist1h1b	$H1^{s}$-3	H1b	H1.5
HIST1H1C	H1.2	H1d	Hist1h1c	$H1^{s}$-1	H1c	H1.2
HIST1H1D	H1.3	H1c	Hist1h1d	$H1^{s}$-2	H1d	H1.3
HIST1H1E	H1.4	H1b	Hist1h1e	$H1^{s}$-4	H1e	H1.4
HIST1H1T	H1t		Hist1h1t		H1t	(TS) H1.6

Orphan Genes (Replication-independent)					
		Human	**Mouse**		
gene symbol	**Alias**	**full name**	**gene symbol**	**alias**	**NEW UNIFIED [122]**
H1F0	H1.0, $H1^{0}$	H1 histone family, member 0	H1f0	H1(0)	H1.0
H1FNT	H1T2	H1 histone family, member N, testis-specific	H1fnt	H1t2	(TS) H1.7
H1FOO	H1oo	H1 histone family, member O, oocyte-specific	H1foo	H1oo	(OO) H1.8
HILS1		Histone H1-like protein in spermatids 1	Hils1	TISP64	(TS) H1.9
H1FX	H1x	H1 histone family, member X	H1fx	H1X	H1.10

Table I.6. Nomenclature for linker H1 histone variants.

7.2. HISTONE H1 STRUCTURE AND ITS ROLE IN CHROMATIN ORGANIZATION

Structurally, histone H1 is a family of lysine-rich proteins that present three distinct domains (Figure I.13): (i) a short ~45 amino acid N-terminal domain, enriched in basic amino acids; (ii) a highly conserved central globular domain, composed of around 75 amino acids; and (iii) a long ~100 amino acid C-terminal domain, positively-charged due to the relative enrichment in lysines, serines and prolines. Like in core histones, these tails are post-translational modified (Table I.7). During chromatin condensation, linker histone H1 binds nucleosomes, binds and deforms the linker DNA, and stabilizes chromatin fibers. These functions are directed by different properties of its structural domains [136, 137].

The central *globular domain* is folded in a 3-helix "winged helix" fold and is sufficient for H1 binding to the nucleosome, although the precise binding site is still controversial [138]. The disordered *N-terminal domain* presents two distinct subregions and the most proximal one to the globular domain, which is highly basic, may contribute to the binding stability of H1 in chromatin [139, 140]. Finally, the *C-terminal domain* shows significant variability among H1 subtypes and species and is required for stabilizing higher order chromatin structures by neutralizing the negative charge of linker DNA. The CTD is unstructured in solution, but it is partially ordered and structured upon interaction with DNA or other protein partners (Figure I.14). This conformational change in the CTD is also responsible for maintaining the DNA interaction and, hence, stabilizing folded chromatin structures. Moreover, it is also involved in modulating the affinity of H1 to chromatin *in vivo*, into much more extend than the NTD [136, 141-143].

Thus, the linker histone H1 domain organization is important for the function of H1 in stabilizing nucleosome structure and condensing higher-order chromatin structures. Linker histone mediates de transition of chromatin primary structure ("beads on a string" conformation) to secondary structures (30-nm chromatin fiber) at increasing salt concentrations. Two models are proposed for the 30-nm fiber: (i) the two-start helix model, consisting in repeating units of nucleosomes folded into a zig-zag arrangement [8], and (ii) the solenoid fiber model [7], where nucleosomes do not adopt a zig-zag orientation and nucleosomes are positioned consecutively in a hand-to-hand orientation.

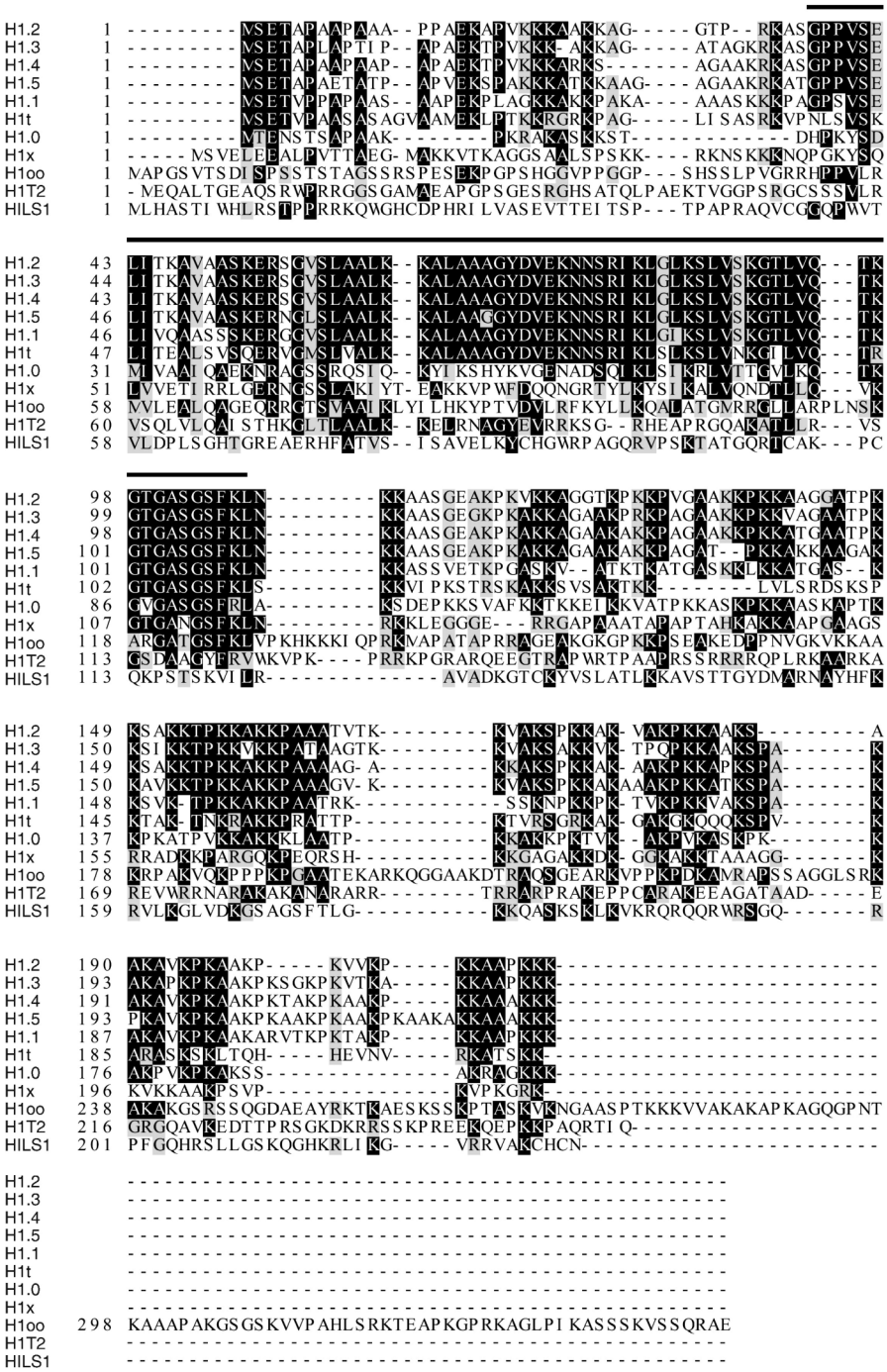


Figure I.13. Amino acid sequence alignment of human histone H1 variants. *Identical residues are in black, and similar in grey. The black bar represents the extension of the globular domain [125].*

Linker histone H1 is also important in determining the nucleosomal repeat length (NRL); it means the distance between consecutive nucleosomes. The paradigm of one histone H1 molecule per nucleosome is no longer believed. It is established that different cell types or cellular states differ on the histone H1/nucleosome ratio, and that the total level of H1 affects

the length of the linker DNA. Thus, high H1-to-nucleosome ratio is related with longer NRL. On the other hand, low amount of H1 per nucleosome is associated with short spacing between nucleosomes. This is characteristic of active chromatin domains or rapidly growing cells, such as ES cells (ratio ~0.5), in contrast to mature cells where chromatin is more compacted (ratio ~0.8) [96, 144-146]. The effect of the different H1 variants on nucleosomal spacing has been challenged by cytosolic microinjection of chicken and human H1 subtype mRNAs in *Xenopus* oocytes. It was demonstrated that H1 subtyes differ in their effect on nucleosomal spacing *in vivo*, suggesting that they have different roles in the organization of the chromatin fiber [147]. Moreover, in breast cancer cells, only H1.2 depletion, and no other variants, causes decrease in the NRL, pointing to specificities of the variants in chromatin organization [97].

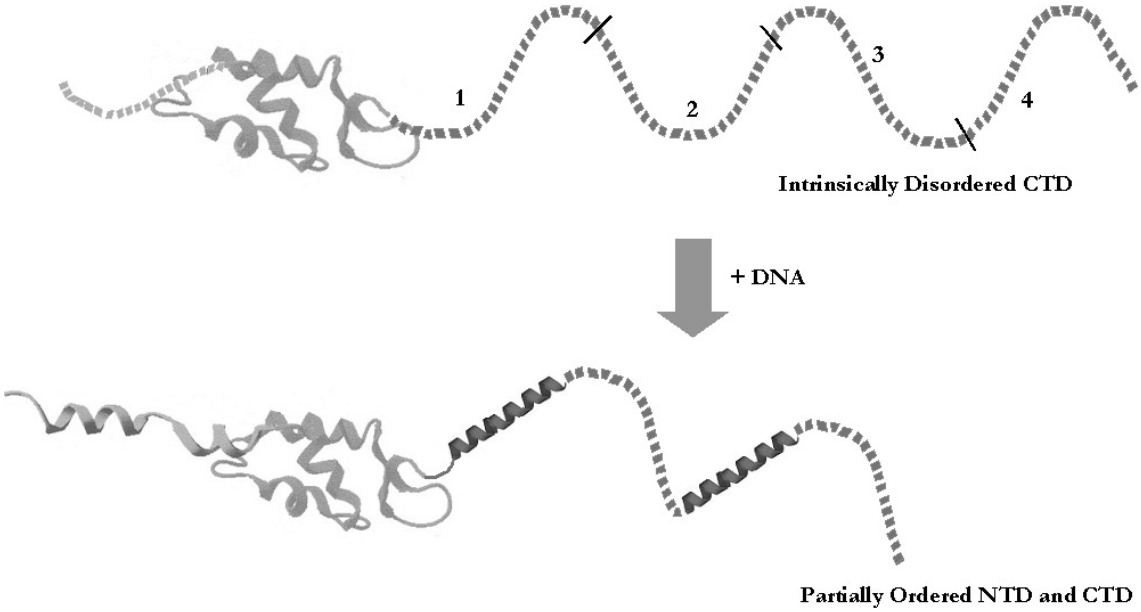


Figure I.14. Model of linker histone H1 domain rearrangement upon interaction with DNA. *The H1 CTD of histone H1 is disordered in the absence of interacting partners. Upon interaction with DNA or other targets in chromatin, the CTD domain adopts a complementary structure, by rearrangement of regions 1 and 3. Interaction with other partners may induce alternative structures. Figure taken from [142], based in the work [148] .*

7.3. HISTONE H1 POST-TRANSLATIONAL MODIFICATIONS

Like core histones, linker histone H1 is post-translational modified. According to the "histone code" hypothesis, histone post-translational modifications provide tools for altering chromatin structure and it is now accepted that H1 PTMs are just as important. With the development of mass spectrometry methods, some attempts in distinct species have been done to map H1 PTMs, showing that H1 is modified mainly by phosphorylation, but also by acetylation and methylation [149-151]. More recent studies show they can also be ubiquitinated and

formylated [152-154]. Although many modifications have been identified by mass spectrometry (Table I.7), only few have been further studied *in vivo*, and some of them are demonstrated to play a role in different processes through the association with different partners. Interestingly, different H1 subtypes present unique modifications, which may be responsible of their specific function in several cellular processes.

H1 variant	Length	Phosphorylation sites	Acetylation sites	Methylation sites	Ubiquitination sites	Formylation sites
H1.2	213	S2, T4, **T31**, S36, **T146**, **T154**, T165, **S173**	S2[a], K17, K34, K46, K52, K63, K64, K85, K90, K97, K169, K192	K34, K52, K64, K97, K106, K119, K168, K187	K46, K64, K75, K85, K90, K97, K106	K17, K34, K46, K63, K64, K75, K85, K90, K97, K160
H1.3	221	T4, **T18**, S37, **T147**, **T155**, T180, **S189**	S2[a], K17, K34, K46, K52, K63, K64, K85, K90, K97, K169	K52, K64, K97, K106, K169	K47, K65, K76, K86, K91, K98, K107	K34, K46, K63, K64, K75, K85, K90, K97, K141, K160
H1.4	219	S2, T4, **T18**, S27, S36, S41, T142, **T146**, **T154**, **S172**, **S187**	S2[a], K17, K26, K34, K46, K52, K63, K64, K85, K90, K97, K169	K26, K52, K64, K97, K106, K119, K148, K169	K17, K21, K34, K46, K64, K75, K85, K90, K97, K106	K17, K34, K46, K63, K64, K75, K85, K90, K97, K110, K140, K160
H1.5	226	S2, T4, T11, **S18**, T39, S44, S107, **T138**, **T155**, **S173**, **T187**, S189	S2[a], K17, K49, K88, K93, K109, K168, K209	K27, K168, K169		K67, K85, K88

Table I.7. Summary of post-translational modifications on the most common human histone H1 variants. *Phosphorylation sites in bold are consensus CDK sites (S/T-P-X-K, where X is any amino acid).* [a] *Denotes N-α-acetylation of the N-terminal residue after methionine removal.*

Phosphorylation is the most well characterized modification of H1 histone and is tightly coupled to cell cycle. H1 phosphorylation progressively increases during S phase, reaches its maximum levels at late G2 and mitosis, and decreases rapidly in telophase. Serine and threonine residues in both N- and C-terminal tails are phosphorylated by cyclin-dependent kinases (CDKs) through the recognition of the (S/T)-P-X-(K/R) motif [149, 155]. H1 phosphorylation in S phase promotes the DNA decondensation needed for replication, while phosphorylation in G2/M is involved in mitotic chromosome condensation. H1 is also phosphorylated by other kinases than CDKs. For instance, H1.4-S27, H1.4-S35 and H1.5-T10 phosphorylations are catalyzed by Aurora B kinase, protein kinase A (PKA), and glycogen synthase kinase-3 (GSK-3), respectively [156-158]. Furthermore, by immunofluorescence analysis it was also demonstrated that different phosphorylations of H1.5 occurred at different cell cycle stages [159]. Finally, and increase in H1 phosphorylation has been seen in cancer. A significant increase in phosphorylated H1 has been observed in high-grade invasive bladder cancer cells compared with normal human bladder epithelial cells, in accordance with the increased proliferation capacity of cancer cells [160].

In agreement with the presence of H1 phosphorylation in interphasic nucleus, H1 phosphorylation is also involved in gene regulation. For instance, H1 is phosphorylated before being displaced from the MMTV upon hormone induction [82, 161, 162]. Moreover, site-specific histone H1 phosphorylation facilitates transcription by RNA polymerase I and II and is implicated in ribosome biogenesis and control of cell growth [163].

H1 phosphorylation, together with methylation and acetylation, is involved in heterochromatinization by regulating HP1 binding. Lysine residue K26 on H1.4 is known to be both methylated and acetylated. G9a histone lysine methyltransferase mediates mono- and di-methylation, whereas JMJD2 removes this modification [164]. Thus, HP1 binds specifically to methylated H1.4-K26, but the interaction is regulated by a "phosphor-switch" on the adjacent H1.4-S27 phosphorylation, which prevents HP1 binding [165]. Moreover, histone deacetylase SirT1 interacts with H1 and deacetylates H1.4 at K26, regulating the formation of repressive heterochromatin [166]. Interestingly, G9a and Glp1 are also able to methylate H1.2-K187 *in vivo* and *in vitro*, but this methylation, which is stable during cell cycle, cannot recruit HP1 and it is not reversed by JMJD2D [167]. These observations favor the view that specific modifications in linker H1 subtypes lead to specific functions in different processes.

Like in the case of core histones, genome-wide mapping of linker histone post-translational modifications will be useful to better characterize their function. At the moment, only one report focusing in a specific acetylation on H1.4 has raised [168]. Kamieniarz et al. showed that H1.4-K34 acetylation by GCN5 is associated with promoters of active genes and regulatory regions. This modification seems to positively regulate transcription, both by increasing H1 mobility and by recruiting transcription factors (TAF1). They also showed that H1.4-K34 acetylation presents a dynamic behavior during spermatogenesis and marks human seminomas.

Linker histone H1 is also PARylated as a result of PARP-1 (poly ADP-ribose polymerase) activity. It has been known for a long time that PARylation of native polynucleosomes promote decondensation, a similar effect than H1 depletion [169]. However, it has been recently proposed that H1 PARylation by PARP-1 mediates H1 displacement from promoters, leading to chromatin remodeling and transcription activation. In fact, PARP-1 and H1 are mutually exclusive in active promoters and depletion of PARP-1 increases the binding of H1 at many target regions [170, 171]. Moreover, activation of PARP-1 by CDK2 contributes to the displacement of histone H1 from progesterone responsive promoters in breast cancer cells [172]. PARP-1 also acts as a coactivator for GATA3 in breast cancer cells to regulate CCND1

transcription by ejecting H1 from the promoter [173]. Additionally, it has been recently described another evidence for histone modification cross-talk, between H3 and H1 [174]. Kassner et al showed that ADP-ribosylation of H3 by ARTD1 prevents H3 methylation by SET7/9, while PARylation of H3 allowed methylation of H1.4 by SET7/9.

7.4. NEW VIEW FOR H1 FUNCTION

Because of its role in the formation of higher order chromatin structures, H1 has been classically seen as a structural component related to chromatin compaction and inaccessibility to transcription factors, RNA polymerase, and chromatin remodeling enzymes [175, 176]. Many studies support this view, as H1 presence in promoter regions impairs transcription of the associated gene. For instance, CHD8 negatively regulates β-catenin function by recruiting histone H1 to the promoters of Wnt target genes [177], and both CHD8 and H1 negatively regulate p53-mediated apoptosis [178]. Also, *Rhox* homeobox gene cluster is demonstrated to be a major target of H1-mediated repression in ES cells [179].

However, in the last years, the view that H1 can play a more dynamic and gene-specific role in regulating gene expression is gaining strength in the field. In fact, overexpression and knock-out or knock-down studies in several organisms revealed than only few genes change in expression upon H1 content alteration and they are either up- or down-regulated, pointing to a complex positive or negative gene-specific function of H1, rather than a general repressive function. The study of H1 function has been addressed by manipulating the amount of H1 in different organisms (Table I.8).

For instance, in *Xenopus laevis* embryos, overexpression and incorporation into chromatin of somatic H1 variants repressed oocyte- but not somatic-type rRNA genes or other Pol III transcripts [180, 181]. This correlates with deletion of H1 in *Tetrahymena*, where it was observed specific gene activation, but not a pronounced effect on gene transcription by Pol II or Pol III [182]. In the same direction, depletion of the unique somatic *Drosophila melanogaster* H1 histone (dH1) affected expression of a limited number of genes [183], and H1-null DT 40 chicken cells caused mainly down-regulation of gene expression, suggesting a positive role for H1 [144]. Finally, global gene analysis in human breast cancer cells after individual depletion of somatic variants (H1.0 and H1.2-H1.5) also affected a reduced number of genes (~6% of the transcriptome), and, interestingly, sometimes different for each H1 variant knock-down, supporting the idea of specific functions for the different H1 variants (see below) [97]. In conclusion, H1 depletion caused in many organisms down-regulation of genes more than up-regulation, contrary to the notion of H1 as a general repressor.

It is also important to note that H1, besides from its role in modulating transcription, has been related with an increasing number of functions related with other DNA-templated processes. For instance, H1 seem to be involved in DNA damage response and apoptosis [183-185], and in DNA replication, as shown by *in vitro* experiments [186]. It is also interesting to note the observation that H1.2 dissociates form chromatin and accumulates into the cytoplasm, where it triggers permeabilization of the mitochondrial membrane and release of apoptotic molecules, upon double strand brakes. These results and others indicate that H1 is implicated in transmitting apoptotic signals from the nucleus to the mitochondria in response to DNA damage [187, 188]. Finally, H1.4, through specific PTMs, has also been related with heterochromatin establishment, as it has already been discussed above [166, 189, 190].

Another controversial issue in the linker histone H1 field deals on the function of its variants. It is still not well understood why so many H1 variants exist and many efforts have been done recently to elucidate if they play specific roles or if they play redundant functions between them. In order to resolve this, single or double H1 variant knock-out studies in mice were generated and, interestingly, they presented no apparent phenotype [191]. This was due to the compensatory up-regulation of other subtypes that maintained a normal H1-to-nucleosome stoichiometry, favoring the view that H1 variants are redundant. However, triple somatic H1 knock-out (H1c, H1d and H1e) mice died at early-embryonic stages (E11.5) because the up-regulation of the remaining subtypes could not fully compensate the lack of total H1. So, correct H1 levels are critical for proper mammalian development [145]. Furthermore, mouse embryonic stem cells derived from triple KO mice presenting a ~50% reduction of total H1 showed a shorter nucleosomal repeat length (NRL) and reduced chromatin compaction. Moreover, these cells presented gene expression alteration in a small number of genes, mainly regulated by DNA methylation (imprinted or X chromosome genes), pointing to a role of H1 in maintaining or establishing DNA methylation patterns [96].

Organism	Variant	#	Phenotype	References
S. cerevisiae	Hho1	KO	Effect on expression of specific genes (up or down)	[192]
		KO	Inhibits DNA repair, effect on life span	[193]
T. thermophila		KO	Effect on expression of specific genes (up or down)	[182]
A. immersus		KO	Effect on life span	[194]
C. elegans	H1.1	RE	Defect in development and transgene silencing	[195]
	H1.1	KO	Globally increased H3K4 and decreased H3K9 methylation	[196]
	H1.X	RE	Uncoordinated and egg laying defective worms	[197]
N. tabacum	H1A, H1B	RE	Defects in flower development and male gametogenesis	[198]
A. thaliana		RE	Aberrant development, DNA hypomethilation	[199]
D. melanogaster	dH1	RE	Compromises fly viability, lethality at larval stage Effect on expression of specific genes in a regional manner (transposons), genome instability and proliferation deffects	[183, 200]
G. gallus (DT 40 cells)	01H1 02H1 03H1 10H1	KO	Changed protein patterns Enhanced expression of remaining H1s	[201]
	H1L, H1R	KO	Enhanced expression of remaining H1s	
	H1R	KO	Accumulation of IR-induced chromosomal aberrations	[185]
	H1-null	KO	Decreased NRL, expanded nuclear volumes, increased chromosome aberration rates, transcription alteration (mostly downregulation)	[144]
X. laevis	H1A	RE/OE	Activation/repression of oocytes 5S gene Defects in the control of mesoderm differentiation	[180, 181, 202]
M. musculus	H1.0, H1c	OE	Increased NRL, differential cell cycle alterations and transcription levels	[203, 204]
	H1a, H1b, H1c, H1d, H1e, H1.0	Single KO	Mice develop normally Positive and negative effects on gene expression	[191, 205]
	H1.0	KO	Mice develop normally Defect in immune system	[206, 207]
	H1t	KO	No phenotype in spermatogenesis Expression of H1.1, H1.2 and H1.4 increased Effect on expression of specific genes	[208-211]
	H1T2	KO	Reduced male fertility	[212, 213]
	H1c+H1d+H1e	Triple KO	50% reduced H1 to nucleosome ratio, embryonic lethal Shorter NRL, genes up- and down-regulated, in particular genes regulated by methylation	[96, 145]
			Hyperresistance to DNA damage	[214]
H. sapiens	H1.0, H1.2, H1.3, H1.4, H1.5	RE	H1.2 and H1.4 caused defects in proliferation. H1.2 KD caused G1 arrest, decreased NRL, variant specific changes in global gene expression	[97]

Table I.8. *Overview of phenotypes observed after manipulation of H1 expression in different organisms. # KO=knock out; RE=reduced expression; OE=overexpression. Adapted from [123].*

7.5. EVIDENCES FOR H1 SUBTYPE SPECIALIZATION

Despite previous observations, increasing evidences support the hypothesis of specific functions for histone H1 variants. Observations in their expression preferences or their different affinity for chromatin, as well as their specific role in various cellular processes, support the idea that heterogeneity of H1 histone family in several organisms is linked to the specialization of these variants in certain cellular processes (Figure I.15). Several evidences

supporting specific functions for H1 variants are described below. Other examples have been already discussed before.

7.5.1. H1 variant sequence conservation

H1 variants are paralog genes, as they were originated by gene duplication events. On the other hand, the corresponding variants within two species are orthologs, because they share a common ancestor before the event of speciation. Ortholog genes are much more conserved than paralog genes; it means that the primary sequence of a given H1 variant is more conserved compared with the corresponding variant of another species than compared with other variants from the same species. This effort of evolution to conserve the sequence of a given H1 subtype indicates that H1 variants specialized in their function after they were originated from a common ancestor. Moreover, an estimation of the rates of nucleotide substitution also supports functional differentiation between H1 variants [215, 216].

7.5.2. Differential expression patterns

H1 subtypes present cell type and tissue-specific expression patterns and their expression is regulated along differentiation and development. It is well established that some histone H1 variants are ubiquitously expressed, while others are tissue restricted. For instance, some H1 variants are only expressed in germ cells (H1t, H1T2, HILS1 and H1oo) while H1.1 is restricted to some tissues (liver, kidney, lung, lymphocytes from thymus and spleen, neurons and germ cells). Moreover, the relative ratio of a given variant in a cell type can be different than in others [134, 217-219], and only H1.2 and H1.4 are expressed in all investigated cells [220]. Histone H1 variant expression pattern is also different during cell cycle, and while expression of some histone variants is couplet to replication, other variants such as H1.0 and H1X are expressed independently of replication in non-proliferating cells [126, 221, 222].

It should also be noted that H1 variant differential expression is related with differentiation. In fact, H1.0 was considered as a replacement histone gene, because its accumulation in terminally differentiated cells that have stopped dividing [222, 223]. A recent report based on *in vitro* differentiation experiments on human ESC and on keratinocyte reprogramming to induce pluripotent cells (iPS) confirms that H1 variants are differentially expressed within these stages [224]. Pluripotent cells have decreased levels of H1.0 and increased levels of H1.1, H1.3 and H1.5, compared with differentiated cells, where H1.0 represents ~80% of the H1 transcripts. It is worth noting that in these systems H1.0 knock-down impaired differentiation. The importance of linker histone H1 in differentiation has also been reported in mice H1 triple

KO cells, where the differentiation capacity of ESCs is severely impaired [225]. ESCs are characterized by an "open" chromatin state compared with differentiated cells, which correlates with reduced levels of H1. Thus, appropriate levels of H1 are needed in order to keep a dynamic chromatin state in stem cells, but also to efficiently repress the expression of pluripotency factors and maintain the epigenetic marks necessary during differentiation. In this direction, a recent report on *Xenopus* embryonic development suggests that proper chromatin H3/H1 ratio is required for correct mesodermal competence, as it keeps chromatin in a malleable state that facilitates differentiation [226]. A similar conclusion arises from another recent report on mouse development, where specific knock-down of H3.3 causes developmental arrest at the morula stage. This effect is rescued by exogenous overexpression of H3.3 but also, partially, by H1 interference [227].

Finally, H1 expression has also been related with cancer processes, and some specific variants seem to be more related than others in particular tumors [160, 228-230]. For instance, in ovarian cancer, H1.3 expression is increased, while expression of H1.0, H1.1, H1.4 and H1X is reduced in malignant adenocarcinomas compared with benign adenomas [229]. Moreover, H1.5 expression correlates with grade of pulmonary neuroendocrine tumors [231]. Like this, H1 expression could be used as a marker or therapeutic target in cancer disease.

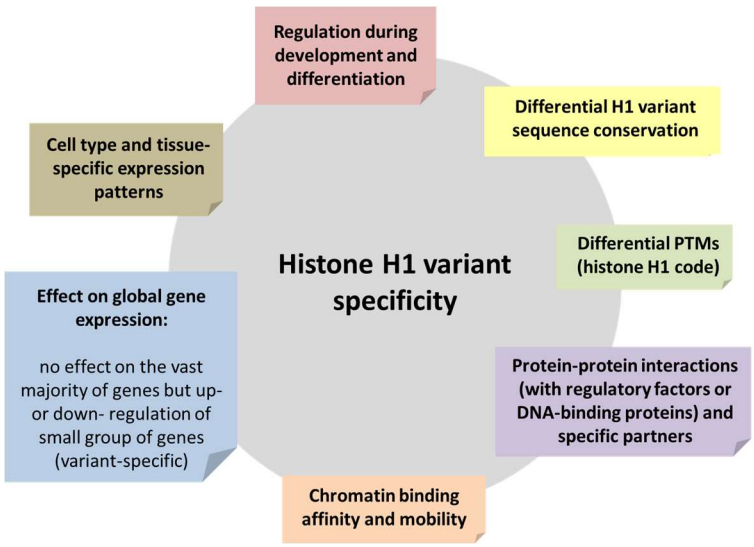


Figure I.15. *Overview of evidences pointing to specific non-redundant roles of linker histone H1 variants.*

7.5.3. Chromatin binding affinity

H1 was considered to be constantly associated to chromatin because of its structural role in chromatin formation. However, this view changed after the development of *in vivo* approaches to study H1 binding to chromatin. Fluorescence recovery after photobleaching (FRAP) with

recombinant H1s fused to GFP showed that they are highly mobile in comparison with core histones, but less than their competitors in chromatin binding: high mobility group (HMG) proteins [143, 232].

Further FRAP experiments and biochemical studies *in vitro* showed that the chromatin binding affinity and the residence time on chromatin was different between H1 subtypes due to differences in the C-t tail, but also in the N-t tail [233]. FRAP experiments by Th'ng et al [234], showed that H1.4 and H1.5 are the variants with higher affinity to chromatin, followed by H1.3 and H1.0, and with H1.1 and H1.2 presenting the highest mobility. Moreover, a report using chromatin assembled in *Drosophila* embryo extracts showed that the H1 variants differ in their ability to reconstitute nucleosomal arrays in vitro - [(H1.5, H1.4)>(H1.3, H1.2, H1.0)>H1.1>H1X]- [235]. Finally, in an *in vitro* competitive assay, in which the binding of the H1 subtypes to long chromatin fragments and to scaffold-associated regions (SARs) was determined, H1.1 was the subtype with the lowest affinity, H1.2 and H1.5 showed intermediate, and H1.3, H1.4 and H1.0 had high affinity to both the SAR and to the non-SAR DNA [236]. Interestingly, H1 mobility is also dependent on specific post-translational modifications, as mutations in the residues associated to a particular PTM show altered FRAP recovery kinetics [158, 168].

7.5.4. Differential histone H1 PTMs and interaction with specific partners as a potential mechanism for specific functions

As it has been previously discussed, H1 subtypes are post-translational modified. These modifications are sometimes specific for a given H1 variant and it is reported that specific PTMs modulate the interaction of H1 variants with different partners. We also discussed how variant-specific histone H1 modifications are catalyzed by different enzymes. This complexity in the "histone H1 code" could explain some reported specific functions for some H1 variants.

Although some groups have started to perform proteomic analysis in H1 [237], the complete interactome for the different H1 variants is still not resolved. However, several reports reveal that H1 variants interact with different partners, and these different associations are responsible for H1 variant specific functions. For instance, in mice, the transcription factor Msx1 interacts with H1b to inhibit *MyoD* transcription and muscle differentiation [238]. H1.5 also interacts with FoxP3 via the leucine zipper (LZ) domain to alter its FoxP3 binding to target genes, modulating gene expression and programming Treg function [239]. Furthermore, H1.2 has been proposed to be part of a complex that acts as a repressor of p53-mediated transcription, through modulation of chromatin remodeling. The association of H1.2 with p53

is disrupted upon DNA damage by H1.2 phosphorylation in T146 by DNA-PK, followed by p300-mediated p53 acetylation, and resulting in increased transcription [240, 241]. Finally, it has been already discussed the importance of H1.4 modifications in K26 and S27 for HP1 binding and heterochromatin formation [165-167].

7.5.5. Effect on global gene expression

Linker histone H1 is not merely a transcriptional repressor as it has also been reported a positive role in gene expression. Moreover, global gene expression analyses in various cell types revealed that histone H1 variants control the expression of different subsets of gens, pointing to a specific role of H1 variants in gene regulation.

Overproduced H1c and $H1^0$ in mice, combined with cell cycle synchronization and expression microarray analysis, showed that H1 acts as specific rather than general regulators [242]. Moreover, many of the genes are uniquely targeted either by H1c or $H1^0$, indicating H1 variant specificity. Interestingly, $H1^0$ repressed more genes than H1c, supporting the idea that $H1^0$ is a strong repressor or transcription.

In human breast cancer cells, individual inducible knock-down of each of the H1 subtypes causes deregulation of a limited fraction of the genome [97]. A portion of genes are affected by the depletion of more than one variant, suggesting a redundant role of the variants in those promoters. However, most of the genes are affected only by one variant, pointing to specific functions on those ones. Moreover, the proportion of genes down- versus up-regulated is also different for the variants, going form 1 for H1.5 to 2.7 for H1.2. Finally, further analysis on expression microarray data reveals that the different H1 variants may be regulating specific functions in the cell, as the deregulated genes are enriched in different cellular processes, pathways or cellular compartments (data not published).

Regarding this issue it would be interesting to determine how H1 variants regulate specific promoters. The exact mechanisms by which H1 regulates the expression of genes, either negatively or positively, is still unknown. It could be that H1 variants are locally enriched in particular promoters or regulatory regions in order to control the expression of the associated gene. But it could also be plausible that specific post-translational modifications in H1 variants, and not their abundance *per se*, is what dictates if one gene must be activated or not.

7.6. H1 NUCLEAR LOCALIZATION

To fully understand the function of histone H1 and its variants, several groups pursued to elucidate the genomic distribution of H1 *in vivo*. However, due to the lack of specific ChIP-grade antibodies for most of the H1 variants, the precise mapping of H1 variants into the genome has been challenging. Initial biochemical and microscope approaches point to a non-uniform distribution of H1 in the cell nucleus and reported differences between variants.

A first analysis by indirect immunofluorescence approaches with specific polyclonal antibodies showed H1.5 localizing at the periphery of the nucleus, where chromatin is more compacted [133]. In a subsequent study, the same group showed that H1.2 was distributed in parallel with DNA concentration, and H1.3 and H1.4 presented a punctuate pattern [243]. Sometime later, chromatin immunoprecipitation (ChIP) studies and analysis by PCR on selected loci in human fetal fibroblasts confirmed previous observations, suggesting that active chromatin is depleted of total histone H1. Moreover, authors claimed differences in the distribution of the somatic H1s on active versus inactive genes and heterochromatic regions [244, 245]. More specifically, active and poised chromatin was characterized by H1.3 and H1.4 depletion, while all variants were present in inactive genes and heterochromatin. Unfortunately, antibodies used in the studies described above were no longer available and other strategies have been recently implemented. Thus, recently, another report using fluorescence microscopy showed that H1 variants fused to GFP in their N-terminal region were associated differently in euchromatin and heterochromatin in human neuroblastoma cells. While H1.0, H1.1, H1.2 and H1.3 were more associated to euchromatin, H1.4 and H1.5 were preferentially located in heterochromatin [234].

Despite all these observations supporting a differential distribution of H1 variants in the genome, further analysis were needed at that time because of the low resolution of the data and the lack of consistency of the results. Therefore, extensive genome-wide localization of the distribution of H1 variants would provide high resolution and the possibility to interrogate all the genome instead of particular loci.

7.6.1. Genome-wide analysis of H1 variant distribution

During the last years, the explosion of high-throughput sequencing technologies in the study of chromatin has provided valuable information about the distribution in the genome of core histones and their post-translational modifications, as well as of transcription factors and other DNA-binding proteins. However, the study of H1 distribution in the genome has been more

problematic because of the heterogeneity of this family and the lack of chip-grade specific antibodies. Moreover, as histone H1 is expected to distribute in almost all genomic loci, in contrast to the discrete and defined position for transcription factors and histone PTMs, the computational demand to analyze these H1 data makes the mission even more challenging. Despite of this, the first total H1 and H1 variants maps started to arise recently.

Genome-wide studies with histone H1 started with ChIP-chip experiments in breast cancer MCF7 cells using an antibody for total H1 [170]. Apart of proving a reciprocal binding of PARP-1 and H1 at transcriptionally active promoters (already discussed), these ChIP-chip experiments on custom promoter-containing arrays showed a clear depletion of H1 ("H1 valley") near the transcription start site (TSS) of active genes, which is not reproduced in repressed promoters. Furthermore, as microarrays also contained ENCODE regions, other significant troughs and peaks of H1 could be observed in more upstream or intergenic regions.

H1 binding to chromatin was also achieved for the unique somatic H1 in *Drosophila* [246]. By performing DamID (DNA adenine methyltransferase identification) couplet to microarray hibridization in Kc167 cells, it was shown that H1 was bound throughout the genome without significant differences between euchromatin and heterochromatin. However, like in the previous report on MCF7 cells, H1 was excluded from active promoters and, also, from other intergenic regulatory regions. Interestingly, they also showed that H3.3 binding in promoters is inversely correlated with H1 presence, suggesting that H3.3 may contribute to H1 exclusion from promoters in order to maintain chromatin in an open state when transcription must take place.

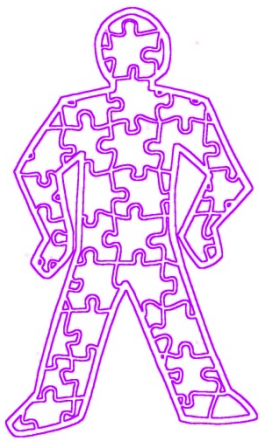
After these firsts' approaches to study H1 distribution in the genome, some groups succeeded recently to obtain the firsts H1 variants maps into the genome. Thus, genome-wide distribution of human H1.5 in IMR90 fibroblasts revealed that this variant presents blocks of enrichment in genic and intergenic regions of differentiated human cells, but not in embryonic stem cells, suggesting that H1.5 pattern establishment depends on cellular differentiation state [247]. Moreover, H1.5 target genes are enriched in gene families, are clustered together in the genome, and present a transcriptionally repressed state. Moreover, gene repression is associated with H1.5 binding, and this H1 variant is necessary for SIRT1 binding, H3K9me2 enrichment, and chromatin compaction.

But it has not been since last year that the genome-wide distribution between different H1 variants was compared. The first attempt focused in mouse H1c and H1d variants [248]. ChIP-seq analysis of N-terminal tagged (with Myc or FLAG) H1c and H1d variants in knock-in mouse

ESC showed depletion of these variants from GC- and gene rich regions, and also in active promoters, presenting the characteristic H1 valley around the TSS. Both H1s presented a positive and negative correlation with H3K9me3 and H3K4me3 respectively, and they were overrepresented in major satellites. The H1 enrichment in major satellites seemed to contribute to the increased NRL observed in pericentromeric regions compared with bulk chromatin. Differences between variants were restricted to cross-comparisons of genome attributes between H1 variants and histone marks enriched regions. In this kind of analysis, when unique peaks for H1c and H1d were compared, H1d was more related to GC-rich sequences and LINES, whereas H1c associated better with AT-rich sequences, Giemsa positive regions and satellite DNA. Finally, genomic distribution of overexpressed FLAG-tagged $H1^0$ mostly resembled that of H1c and H1d, although, besides overrepresentation in major satellites, it was also enriched in minor satellites and LINE L1 elements. This suggests quite differential binding preferences for this variant. However, as $H1^0$ is present at low levels in undifferentiated WT ESCs, it would be necessary to extend these observations to differentiated cells.

Finally, the last reported study of H1 variant distribution until now, has taken advantage of DamID technology to map human H1.1 to H1.5 variants in human lung IMR90 fibroblasts [249]. Authors showed that H1.2 to H1.5 were similarly distributed and were depleted form CpG-dense regions and active regulatory regions. All these variants formed deeps at promoters, enhancers, and CTCF binding sites. Furthermore, H1 abundance was negatively correlated with “active” histone marks and positively correlated with “repressive” ones. Interestingly, the observation that H1 is overrepresented at LADs and that different combinations of H1 variants associated with functionally distinct topological domains, pointed to a possible role of H1 in the three-dimensional organization of the genome, with different H1 variants contributing to the establishment of particular chromatin states. Finally, it is worth noting that H1.1 showed a distinct binding profile from all other variants (H1.2 to H1.5), pointing to a special role of this subtype into chromatin function. This variant is more abundant at promoters and CpGs than others, is not depleted from regulatory regions, and is higher enriched in intergenic regions. Moreover, it is not associated with LADs, but it enriched in polycomb-type chromatin domains.

Altogether, the precise mapping of histone H1 subtypes in different cell lines and conditions will definitively help to understand the function of this histone and its variants, and will bring more light into our understanding of chromatin organization and regulation. So, it will be important in the future to extend the analysis of H1 variant distribution to other cell types, because H1 variant relative abundance in a given cell line or the cellular state (i.e. undifferentiated vs differentiated vs tumorigenic) could influence in the distribution of H1 variants into the genome.



OBJECTIVES

This thesis has been divided in two chapters, corresponding to two different objectives regarding the specificity of human histone H1 variants:

1. CHAPTER I: GENOMIC DISTRIBUTION OF HUMAN HISTONE H1 SUBTYPES.

In this part of the thesis the genome-wide distribution of most of the somatic variants of human histone H1 has been addressed by combining chromatin immunoprecipitation with quantitative PCR, tiling promoter arrays, and high-resolution sequencing. The specific aims in this part were:

a) To validate useful specific ChIP-grade antibodies for the study of H1 variant distribution in breast cancer cells.
b) To set up the conditions to perform ChIP in H1-HA tagged expressing cell lines.
c) To establish a relation between H1 variant occupancy at promoters and expression of the associated genes.
d) To study the H1 variant distribution at promoter regions.
e) To study the genome-wide distribution of H1 variants along the genome and their association with distinct genomic features.

2. CHAPTER II: FUNCTIONAL SPECIFICITY OF HUMAN HISTONE H1 SUBTYPES.

In this second part of the thesis we pretended to further analyze the phenotypic consequences upon H1 variant knock-down in T47D breast cancer cells line. The specific goals of this part were:

a) To further characterize a previously reported inducible H1.4 knock-down cell line.

After reporting an off-target effect of the shRNA in that H1.4 knock-down cell line, we pretended:

b) To develop and characterize a new specific inducible H1.4 knock-down cell line.
c) To develop and characterize an inducible H1 knock-down cell line inhibiting several H1 variants.
d) To further understand the phenotype of all previous cell lines.