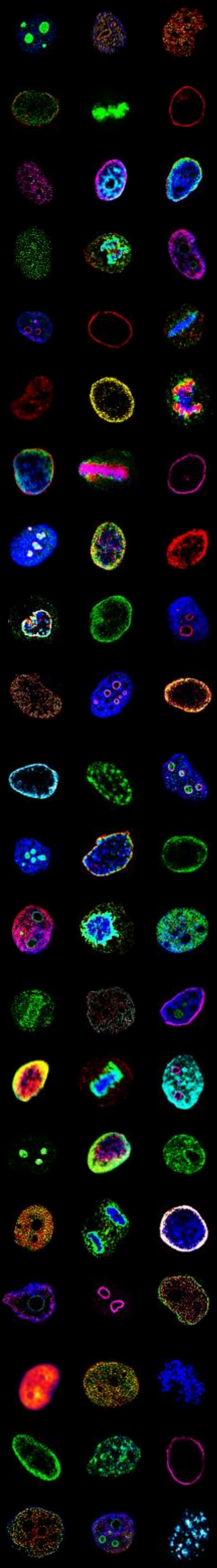
UNIVERSITAT DE
BARCELONA

Systematic characterization of the genome-wide and nuclear distribution of six linker histone H1 variants in human cancer cells

Mónica Salinas Pena

Systematic characterization of the genome-wide and nuclear distribution of six linker histone H1 variants in human cancer cells



Mónica Salinas Pena

Doctoral Thesis · 2023

Doctoral program in Genetics
Faculty of Biology, Universitat de Barcelona

Research performed at the
Institute of Molecular Biology of Barcelona, IBMB-CSIC

Systematic characterization of the genome-wide and nuclear distribution of six linker histone H1 variants in human cancer cells

---

A thesis submitted by Mónica Salinas Pena to obtain the doctoral degree by the Universitat de Barcelona

Author:	*Thesis Director:*	*Tutor:*
Mónica Salinas Pena	Albert Jordan Vallès	Francesc Mestres Naval

Mónica Salina

Barcelona, 2023

Abstract

The histone H1 family comprises up to seven members in human somatic cells. However, H1 studies have been limited by the lack of specific ChIP-grade antibodies. Here, we have mapped six endogenous H1 variants in T47D breast cancer cells, which represent the whole somatic H1 complement in this cell line. ChIP-Seq experiments indicate that H1 variants are categorized into two large groups depending on the local GC content: H1.0, H1.2, H1.3 and H1.5 are enriched at low-GC regions while H1.4 and H1X are more abundant at high-GC genomic regions. Data also uncovers common features for H1 variants, highlighting the existing balance between redundancy and specificity.

Examination of H1 variants abundance within repetitive elements classes denoted that H1.0, H1.2, H1.3 and H1.5 are enriched within Satellite, LINE, LTR or DNA classes while H1.4 and H1X are enriched within SINE or 'Other' classes. This last category comprises SVA retrotransposons, which emerged along hominoid evolution. Interestingly, we have determined that H1X abundance gradually increases from older SVA_A to human-restricted SVA_F families. This unprecedented association of H1 variants abundance and transposable elements (TEs) evolutionary age is also observed across different TE classes. Both H1X and H1.4 are enriched within the most recently evolved TEs along primate evolution, including not only SVAs but also younger Alu, LINE-L1 or LTR repeats. Conversely, H1.2, H1.3, H1.5, H1.0 are enriched in older TEs.

Confocal and super-resolution microscopy experiments further confirm the differential distribution of H1 variants and their distinct contribution to chromatin structure. H1.2, H1.3 and H1.5 are enriched towards the nuclear periphery or lamina-associated domains. H1.4 and H1X show a punctuated pattern throughout the nucleus, with H1X being particularly enriched at nucleoli. H1.0 forms enrichment territories that tend to be peripheral. Moreover, depletion of H1.2, either alone or in combination with H1.4, leads to a general chromatin decompaction that is not observed upon single H1.4 or H1X depletion.

Extensive analysis of multiple cell lines revealed the existence of certain universal distribution features despite variations in the H1 somatic repertoire. Specifically, H1.2, H1.3 and H1.5 consistently show enrichments towards the nuclear periphery in all cell lines examined, suggesting their universal role as components of lamina-associated domains. H1X, whose nucleolar presence is evident in all cell lines, is enriched at high-GC regions and younger SVA and Alu elements, as observed in T47D cells.

In conclusion, we provide the first systematic comparison of six endogenous H1 variants within a mammalian cell type, while also addressing their differential distribution among multiple cell lines. Altogether, our results support H1 variants heterogeneity and highlight their significance as key organizers and regulators of chromatin.

List of contents

Introduction

Materials and Methods

Results Chapter 1. Differential distribution of six endogenous histone H1 variants in T47D breast cancer cells

Results Chapter 2. Heterogeneity of H1 variants content, regulation and distribution across cell lines

Discussion

List of Figures and Tables

Figures-Introduction

Figures-Materials and Methods

Figures-Results Chapter 1

Figures-Results Chapter 2

Figures-Discussion

Tables-Introduction

Tables-Materials and Methods

Tables-Discussion

Introduction

1. Chromatin

Over the past two centuries, significant progress has been made in unraveling the composition and dynamics of the nucleus, since its initial observation by R. Brown in 1831. The term 'chromatin' was later coined by W. Fleming in 1879, followed by the discovery of histones by A. Kossel in 1884. Since then, notable advances have been done in understanding chromatin structure, leading to a complex biological scenario responsible for genome regulation. Our perception of such regulation has constantly progressed, being DNA sequence insufficient to fully understand the genetic program. Due to the complex and dynamic nature of chromatin, our understanding of genome functionality remains incomplete, raising unanswered biological questions that need to be addressed.

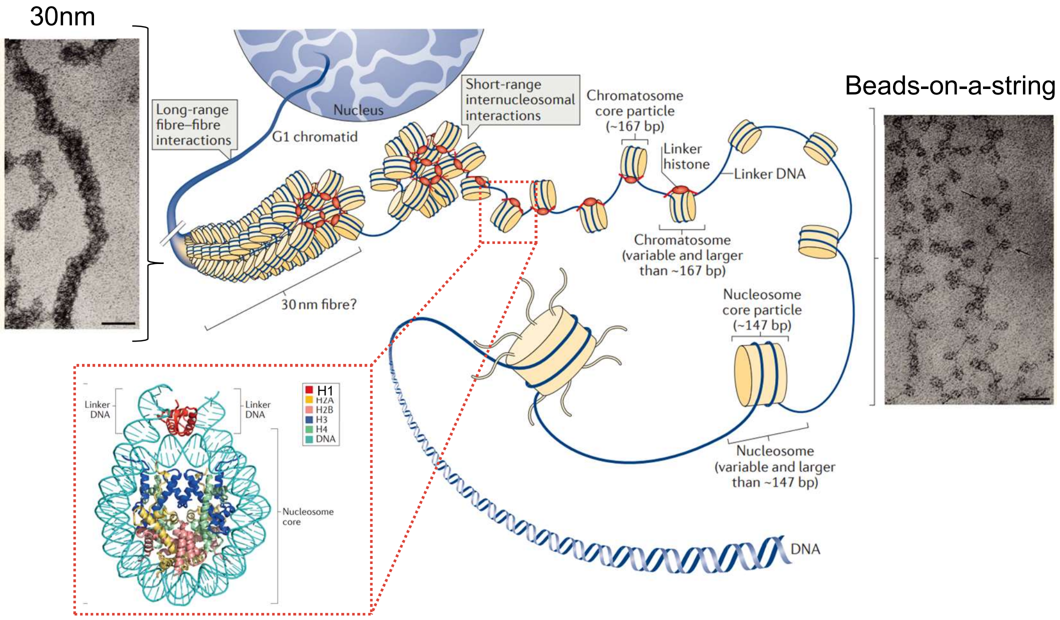


Figure I. 1. Different levels of chromatin folding. Hierarchical levels of DNA compaction in the interphase nucleus: nucleosome core particle, strings of nucleosomes (bead-on-a-string model), the chromatosome core particle, 30nm fibres (the existence of which is debatable *in vivo*) and the association of individual fibres. The crystal structure of the chromatosome core containing the globular domain of chicken H5 (H1.0; shown in red) and fold regions of core histones is shown. Electron micrographs of chromatin spreads at different conformations are shown. Figure adapted from (1,2).

1.1. The nucleosome and chromatin structure

Chromatin is the state in which DNA is packaged within the cell, seen as a DNA-protein polymer in which histone proteins guide DNA compaction. Histones present a highly basic nature, which confers them the ability to strongly bind the negatively charged DNA. The repeating structural and functional unit of chromatin is the nucleosome. Each nucleosome is composed of an octamer of core histones (consisting of two copies of H2A, H2B, H3 and H4) around which ~147 base pairs (bp) of DNA are wrapped. Thus, the basic structure of chromatin has been referred as a 'beads-on-a-string' model or 11nm fiber, in which nucleosomes are spaced by ≈80bp 'linker DNA' throughout the genome **(Figure I.1)**. Chromatin folding into a 30 nm fiber and higher-order structures has been mainly attributed to the addition of histone H1. Histone H1 binds to the nucleosome with ~10 bp of DNA at both the entry and the exit sites of the nucleosome core particle, forming the chromatosome. However, the existence of this hierarchical folding has been a matter of debate due to *in vivo* studies failed to detect stable 30-nm fibers (3–5). Super-resolution microscopy experiments in intact cells suggest that nucleosomes are grouped in discrete domains of variable sizes (termed 'nucleosome clutches') along the chromatin fiber, interspersed with nucleosome-depleted regions **(Figure I.2)** (6).

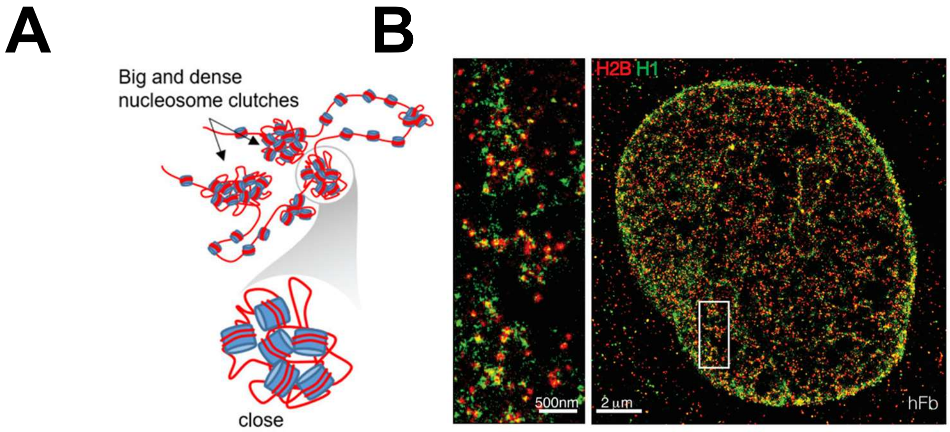


Figure I. 2. Chromatin fibers are formed by heterogeneous groups of nucleosomes *in vivo*. A) Scheme depicting 'nucleosome clutches' of varying sizes. **B)** Super-resolution STORM image of a human fibroblast nuclei stained for core histone H2B (red) and total H1 (green). Figure adapted from (6,7).

1.2. Epigenetic determinants of chromatin

Initial electron microscopy experiments led to the classical view of a bipartite chromatin composition **(Figure I.3)**. Euchromatin is defined as active and open chromatin located at the nuclear center while heterochromatin corresponds to the more compact chromatin fraction, generally positioned at the nuclear periphery and surrounding nucleoli (8–10). This compartmentalization is tightly regulated by multiple epigenetic layers, which ultimately modulate chromatin structure and genome function.

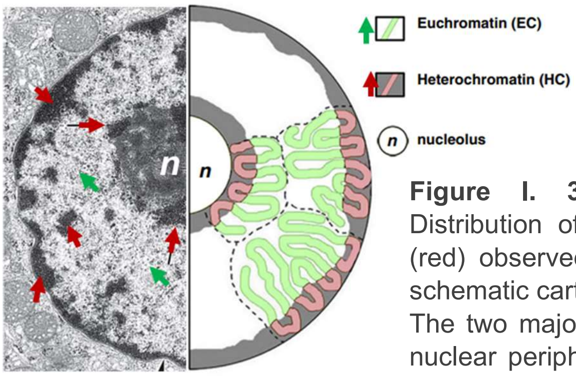


Figure I. 3. Euchromatin and Heterochromatin. Distribution of euchromatin (green) and heterochromatin (red) observed in an electron micrograph (left part) or a schematic cartoon (right part). Nucleolus is also highlighted. The two major heterochromatic regions correspond to the nuclear periphery (lamina-associated-domains, LADs) and the perinucleolar space (nucleolus-associated-domains, NADs). Figure adapted from (8).

1.2.1. DNA methylation

DNA methylation refers to the transfer of a methyl group to the carbon 5 of cytosine residues. This covalent modification is catalyzed by DNA methyltransferases. While DNMT3a and DNMT3B establish *de novo* methylation during development, DNMT1 functions during DNA replication, ensuring the maintenance of DNA methylated patterns. On the contrary, ten-eleven translocation enzymes (TET1, TET2 and TET3) remove the methyl group. Alternatively, DNA demethylation can also occur passively, through sequential cell divisions without DNA methylation maintenance (11).

In mammals, DNA methylation occurs almost exclusively within CpG dinucleotides. CpG sites are not randomly distributed in the genome and they tend to cluster in CpG islands. Although CpG islands are often found at promoters, they are also located within intergenic regions or gene bodies, among others (12).

DNA methylation is in general associated with chromatin silencing, both by direct impediment of transcription factors binding to DNA and through the recruitment of transcriptional repressors. It has a fundamental role in maintaining genome stability and repetitive elements in a repressed state (13). Nonetheless, DNA methylation at gene bodies is positively correlated with gene expression and represents a common feature of ubiquitously expressed genes. It has been suggested that it might be crucial for maintaining transcription efficiency. Indeed, aberrant DNA methylation leads to spurious RNA polymerase II (RNA pol II) entry and cryptic transcription initiation (14).

Aberrant methylation profiles are a hallmark of cancer. Cancer cells exhibit a global loss of DNA methylation, mainly within intergenic and intragenic regions containing repetitive elements. On the other hand, local hypermethylation of promoters from tumor suppressor genes, cell cycle regulators and DNA repair genes is found in multiple cancers. Indeed, several DNA-methylation-based approaches have been used in cancer diagnosis and treatment (15). This is the case of DNMT inhibitors, such as azacytidine (aza), which exhibits antineoplastic effects mediated by DNA hypomethylation and cytotoxicity in tumor cells.

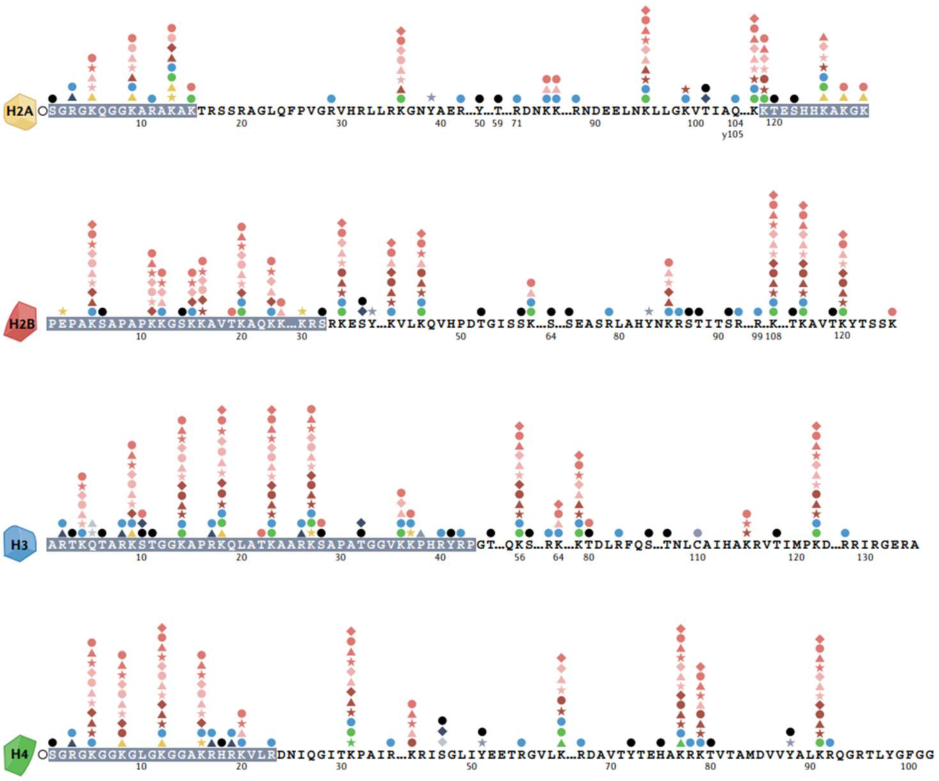
1.2.2. Core histones

Core histone proteins are highly conserved though evolution. In the genome, they are present as multicopy and intronless genes located in the histone clusters. They are synthesized in a replication-dependent manner and their mRNAs are not polyadenylated. Histones present a tripartite structure formed by a central globular domain and N-terminal and C-terminal disordered tails.

Core histones can be decorated with a wide repertoire of post-translational modifications (PTMs) that can affect inter-nucleosome interactions and specifically regulate chromatin processes, constituting the proposed ‘histone code’ (16). These modifications are tightly regulated by multiple enzymes which catalyze their formation (writers), recognize specific PTMs (readers) and remove them (erasers). Moreover, an additional layer of complexity is represented by core histone variants.

1.2.2.1. Core histones post-translational modifications

Histone PTMs are present in both the terminal tails of histones and globular core domains, although N-terminal PTMs are more predominantly studied. PTMs include methylation, acetylation, ubiquitination, ADP-ribosylation and sumoylation of lysine (K) residues; phosphorylation of serine (S) and threonine (T) residues and methylation of arginine (R) residues (**Figure I.4**). Some concrete methylations and acetylations of histone H3 N-terminal tail will be briefly commented.



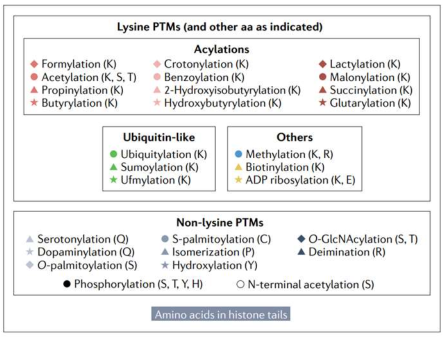


Figure I. 4. Core histone post-translational modifications. The amino acid sequence for histones H2A, H2B, H3 and H4 is shown. Gaps in the sequence are indicated by ellipses. Amino acids within the histone tails are indicated by grey background shading. The most common PTMs are methylation, phosphorylation, acetylation and ubiquitin-like. Figure from (17).

Lysines can be mono-, di- or tri-methylated, while arginines may be mono- or di-methylated. H3K9me3 is a classical marker of constitutive heterochromatin, enriched in repressed genomic regions, including centromeres, telomeres, repetitive sequences but also silent genes. Heterochromatin-protein 1 (HP1) proteins are H3K9me3 readers, promoting chromatin heterochromatinization. On its part, H3K9me2 has been preferentially associated to lamina-associated domains (LADs), a well-known repressive environment. H3K27me3 has been intrinsically linked to facultative heterochromatin and silencing of polycomb target genes (17,18).

Histone methylation is also associated to transcription (19). H3K4me3 enrichment at transcription start site (TSS) of active genes is an evolutionary-conserved and widely studied example. H3K4me3 peak at TSS positively correlates with transcription, enabling the recruitment of the transcriptional machinery. Despite this strong association, functional studies suggest that this mark is not necessary for transcription (20). Indeed, several evidence support that H3K4me3 present could be more related to DNA hypomethylation rather than transcription (21). H3K36me3 is present at gene bodies and associated to transcription elongation but similarly to H3K4me3, several reports suggest that its presence it not required for transcription elongation. Its functional role in the transcriptional process has been proven in preventing intragenic transcription initiation (22), regulating splicing (23) and allowing for proper co-transcriptional modification of RNA (24).

Histone acetylation is generally associated with a more open chromatin state, due to the neutralizing positive charge of lysine residues, which weakens histone-DNA interaction. H3K27ac is considered a hallmark of active enhancers but functional data showed its dispensability for enhancer function (25) and it is also found at active gene promoters. Interestingly, a recent study showed that histone H2B N terminus multisite lysine acetylation (H2BNTac) is a better predictor of active enhancers compared to H3K27ac (26). Overall, although core histone PTMs represent a widely studied epigenetic mechanism that has a direct and indirect impact of chromatin, evidence also supports their dispensability for nuclear processes, highlighting the importance of exploring other determinants, including histone variants.

1.2.2.2. Core histones variants

In contrast to canonical core histones, core histone variants genes are located outside histone clusters and may contain introns. They are synthesized in a replication-independent manner and some variants are not conserved among species.

Conceptually, histone variants functional role is supported by three key aspects. Firstly, their sequence divergence from canonical histones could modulate physical properties of nucleosomes. Secondly, histone variants are deposited by specific chaperones and present unique domains that allow for interaction with concrete partners at specific genomic regions. Thirdly, they can hold PTMs different from their canonical counterparts, which in turn can provide a scaffold for chromatin recruitment of variant-specific interacting factors (27). While the function of some core histone variants will be mentioned in this section, the complete core histone variants diversity is depicted in **Figure I.5.**

The H3 variants illustrate how minor sequence differences can significantly impact the molecular outcomes when they are integrated into chromatin (28). H3.3 is enriched within active chromatin regions, including promoters, enhancers and gene bodies. H3.3 deposition is mediated by the histone regulator A (HIRA) chaperone complex. It has been named ‘replacement variant’ due to its role in histone substitution at active promoters, when elongating RNApol II displaces nucleosomes. In addition, H3.3 is found at repetitive regions, including but not limited to telomeres or pericentromeric chromatin. In this case, its deposition is mediated by ATRX chromatin remodeler and histone chaperone DAXX (ATRX-DAXX) complex. These regions may form abnormal DNA structures difficult to replicate, such us G-quadruplexes (G4). In this context, the ATRX-DAXX-H3.3 axis has been postulated to prevent G-quadruplex DNA-mediated replicative stress (29). On the other hand, CENP-A is a centromere-specific H3 variant, crucial for maintaining centromere identity and proper chromosome segregation during cell division (30).

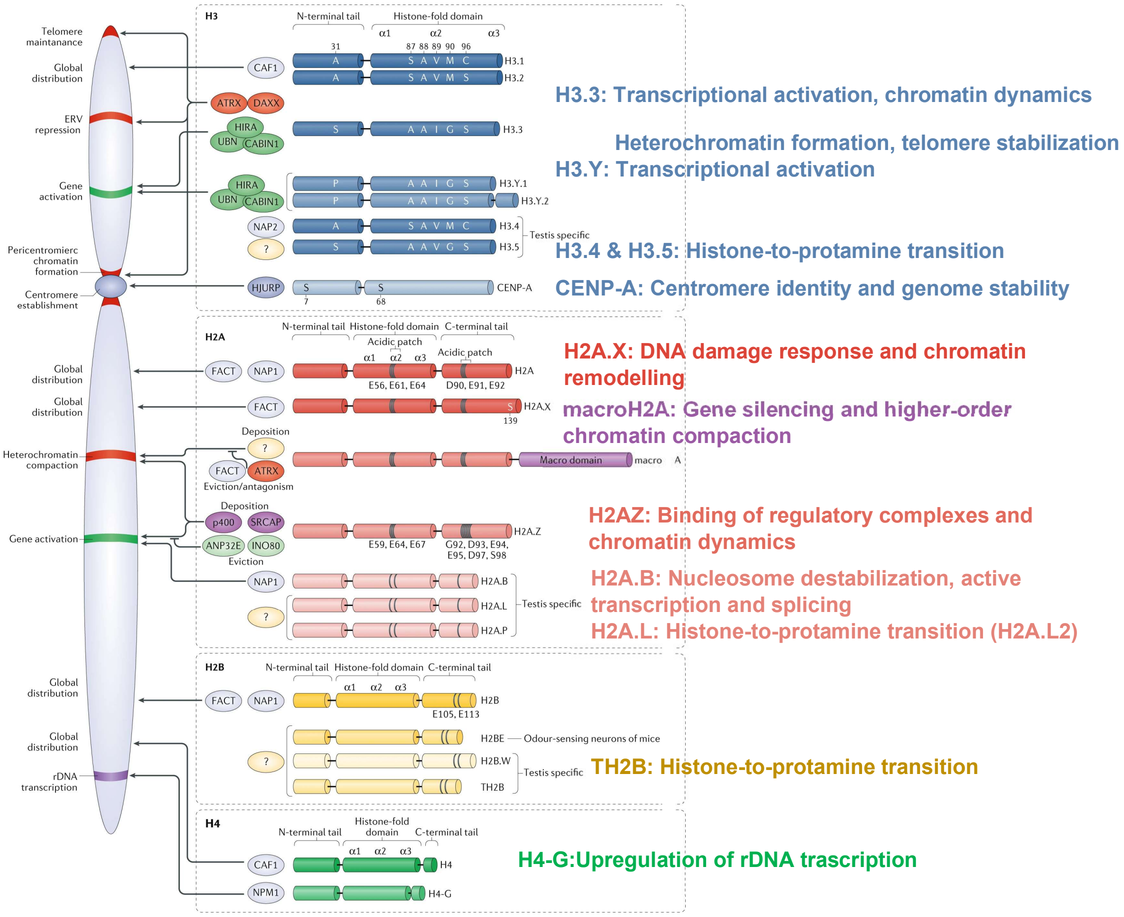


Figure I. 5. Core histone variants. Figure depicts the diversity of core histone variants, their associated chaperone/remodeler machineries and associated functions on chromatin. Specific amino acid residues are illustrated at either key differences among members of a common histone family or at well-established histone variant-specific PTMs. Different shades of color in the structures are used to indicate differences within the domains compared with the canonical histones. Figure adapted from (27).

Regarding H2A variants, H2A.X only differs in four additional amino acids of its C-terminal domain compared to canonical H2A. However, its unique phosphorylation at Serine 139 (referred as ɣH2A.X) is a well-accepted DNA damage marker. ɣH2A.X spreads around the double-strand DNA break (DSBs). Consequently, it forms the so-called ‘ɣH2A.X foci’, supporting the recruitment of the repair machinery to chromatin (31). On the contrary, macroH2A variants present a more divergent structure in comparison to canonical H2A. They are characterized by an additional large C-terminal globular macrodomain bound to the histone fold by a disordered linker region. In mammals, three macroH2A isoforms exist, macroH1A.1.1 and macroH2A.1.2 splice variants and macroH2A.2, encoded by a different gene. In general, macroH2A variants have been associated with chromatin compaction and more particularly to polycomb-repressed regions (32,33). Interestingly, its role in chromatin compaction has been attributed to the linker region, possibly promoting phase separation of heterochromatin (34).

A recently characterized H4 variant, named H4G, is restricted to the Hominidae lineage. It is enriched at nucleoli of multiple human cell lines and it is involved in ribosomal DNA (rDNA) transcription (35,36).

1.2.3. Nucleosome dynamics

As the structural basis of chromatin, nucleosome positions are crucial determinants in gene regulation. Indeed, nucleosomes are highly dynamic structures. Two main parameters could be considered: nucleosome occupancy and positioning **(Figure I.6A)**. Nucleosome occupancy refers to the presence or absence of nucleosomes within a genomic region in a cell population. Nucleosome positioning is defined as the probability of a nucleosome reference point (for example, a dyad) being at a specific genomic coordinate (37).

Gene promoters exhibit concrete nucleosome dynamics **(Figure I.6B)**. Active genes are characterized by a nucleosome-free region at their promoters, allowing for the accessibly of the transcription machinery. Nucleosome +1 (first nucleosome downstream TSS) represents a well-positioned nucleosome.

Nucleosome dynamics can be mapped using micrococcal nuclease (MNase) to cleave the accessible linker DNA between two nucleosomes. By using single-cell micrococcal nuclease sequencing (scMNase-seq), two principles of nucleosome organization were found (38). Nucleosomes in heterochromatin regions and around silent TSSs exhibit cell-to-cell variation in positioning but are uniformly spaced within the nucleosome array. Conversely, nucleosomes surrounding active TSSs and DNase I hypersensitive sites display minimal variation in positioning across cells but show heterogeneous spacing within the nucleosome array **(Figure I.6C)**.

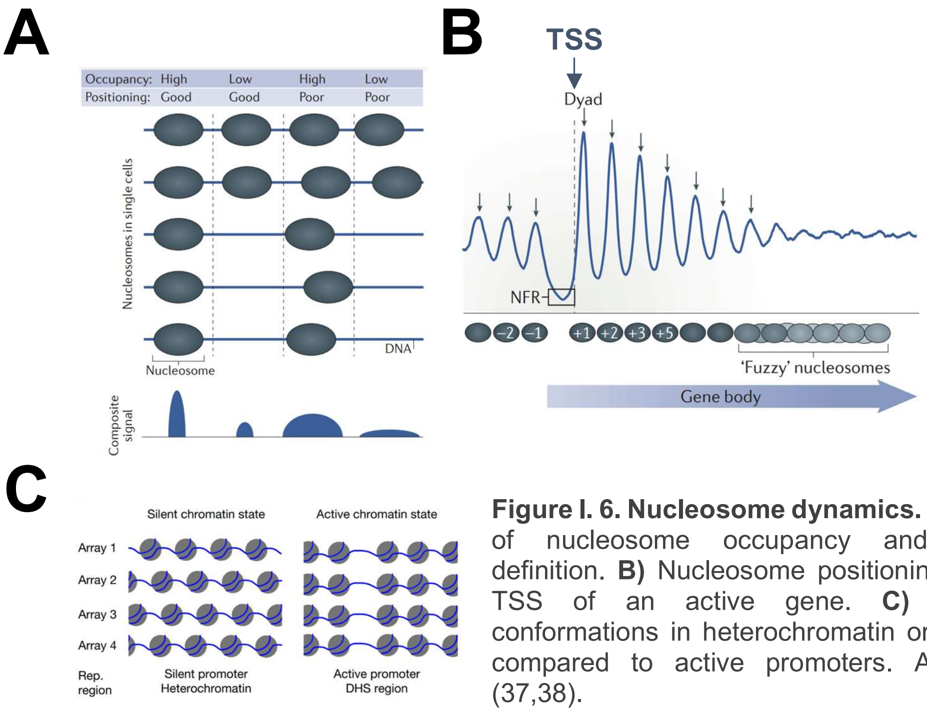


Figure I. 6. Nucleosome dynamics. A) Illustration of nucleosome occupancy and positioning definition. **B)** Nucleosome positioning around the TSS of an active gene. **C)** Nucleosome conformations in heterochromatin or silent genes compared to active promoters. Adapted from (37,38).

1.2.4. Non-coding RNAs

Over the last decades, extensive research has demonstrated that RNAs have diverse regulatory roles in chromatin, extending beyond the traditional messenger RNA (mRNA) functionality (39). Notably, a significant portion of these regulatory functions is attributed to non-coding RNAs (ncRNAs). Multiple types of ncRNAs exist but, particularly, a subset of long-non-coding RNAs (lncRNAs) reside in chromatin.

Chromatin-associated RNAs or chromatin-enriched RNAs (Che-RNAs) are a novel sub-class of lncRNAs tethered to chromatin by RNApol II **(Figure I.7)**. Che-RNAs are mainly cell-type specific and their presence is correlated with proximal gene expression. Despite their similar functionality, Che-RNAs are distinct from enhancer RNAs (eRNAs), which are short bi-directional transcripts that emerge from canonical active enhancers (40,41).

Furthermore, ncRNAs derived from repetitive sequences also cause an impact on chromatin regulation. Although most repetitive elements are silenced, they can become expressed under specific contexts. Indeed, 40% of all lncRNAs sequences are derived from repetitive elements and 80% of human lncRNAs contain some repeat (42,43). A focus on repetitive elements is provided in next section 1.3.

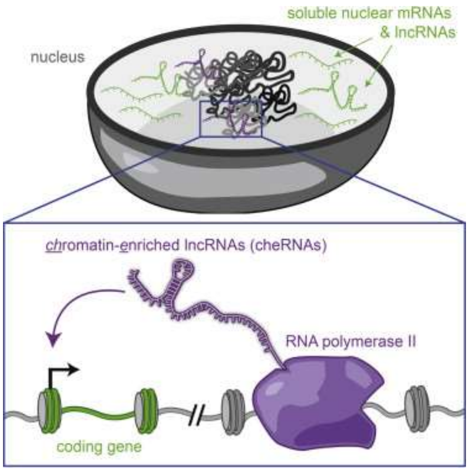


Figure I. 7. Chromatin-enriched RNAs (Che-RNAs). Che-RNAs are a novel subclass of lncRNAs tethered to chromatin by RNApol II that can impact proximal gene expression. Figure from (40).

1.3. Repetitive elements

Repetitive elements can be found in almost all eukaryotic genomes. In humans, over half of the genome is composed of repetitive elements, which mainly include tandem repeats and interspersed transposable elements (TEs) **(Figure I.8)**.

1.3.1. Tandem repeats

Satellites consist of tandemly repeated arrays that are primarily present at centromeres, pericentromeric regions and subtelomeric regions. Depending on the size of the repeating unit they can be classified as microsatellites (also referred as simple of low-complexity repeats), minisatellites or macrosatellites. Macrosatellites can spread up to a few kilobases (kb). Some human macrosatellites examples include D4Z4 (3.3kb microsatellite at subtelomeric regions of chromosomes 4q35 and 10q26), DXZ4 (3 kb CpG-rich macrosatellite present in 12–100 tandem copies on chromosome Xq23) and

NBL2 (from SST1 family, a 1.4kb microsatellite found on the short arm of acrocentric chromosomes 13, 14, 15 and 21) (44).

Primates centromeres are constituted by α-satellites (SATa), which in humans represent over half of the total satellite DNA. On the contrary, mammalian pericentric regions contain different satellite families, including but not limited to satellites I, II and III. Human telomeres consist of thousands of repeats of 5′-TTAGGG-3′ ending in a G-rich 3′-overhang 30–300 nucleotides in length (45). Notably, a recent genome assembly by T2T (telomere-to-telomere) consortium has provided a full map of previously gapped satellite-rich genomic regions. Thus, satellite repeats constitute 6.2% of the T2T genome assembly, with SATa representing the single largest component (2.8% of the genome) (46).

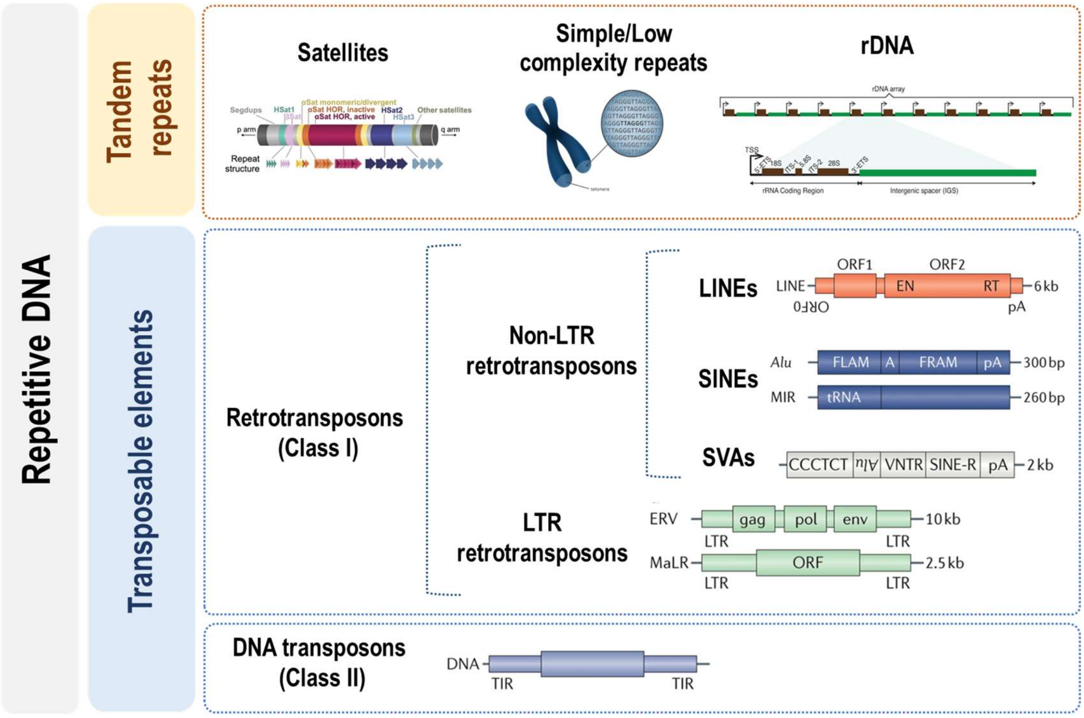


Figure I. 8. Repetitive elements classification. Repetitive elements are classified in tandem repeats or interspersed transposable elements (TEs). Different types of satellites repeats comprise centromeric, pericentromeric, subtelomeric and telomeric regions. A schematic of a generalized human peri/centromeric region is shown (from (46)). Ribosomal DNA (rDNA) is not classified as satellite but it is present as tandem copies of a rDNA repeat (from (47)). TEs are divided into retrotransposons or DNA retrotransposons, according to their transposition mechanism. The structure of characteristic TE families is shown (from (48)).

Although the exact role of satellites remains unclear, they have been associated to multiple processes, including heterochromatin formation. In mammals, satellite transcripts are reported to form RNA:DNA hybrids and thus form a scaffold for the retention of Suv39H, a H3K9me3 methyltransferase (49). Moreover, satellite RNA also participates in the recruitment to scaffold attachment factor B (SAFB) to pericentric regions, promoting heterochromatin stabilization through phase separation (50). Similarly, telomeric and subtelomeric regions can be transcribed into UUAGGG-repeat-containing RNAs (TERRAs), which also have been proposed to control telomere structure (51). These results suggest that centromeric, pericentromeric, telomeric and subtelomeric blocks of constitutive heterochromatin can be transcribed and their satellite-derived RNAs may act as a structural scaffold to maintain the native condensed state of heterochromatin.

Ribosomal DNA (rDNA) can also be considered as an additional tandem repeat structure. In the human genome, they are present as tandem copies of a rDNA repeat that encode for the different rRNAs. rDNA copies are located over the nucleolar organizer regions (NORs) located in the short arms of the acrocentric chromosomes. rDNA is transcribed by RNApol I.

1.3.2. Transposable elements

Unlike tandem repeats, transposable elements (TEs) are interspaced over the genome. Different classes of TEs exist, exhibiting a variable transposition mechanism and structure (**Figure I.8**). Retrotransposons (or Class I TEs) move throughout the genome via RNA intermediates through a 'copy and paste' mechanism. On the contrary, DNA transposons (or Class II TEs) move autonomously as DNA segments via 'cut and paste' mechanism. Retrotransposons include long interspersed elements (LINEs) and short interspersed nuclear elements (SINEs), which are the two major classes of repeats in the human genome. Other classes of retrotransposons are represented by SINE-VNTR-Alu elements (SVAs) and long-terminal-repeats (LTR) elements (52–54).

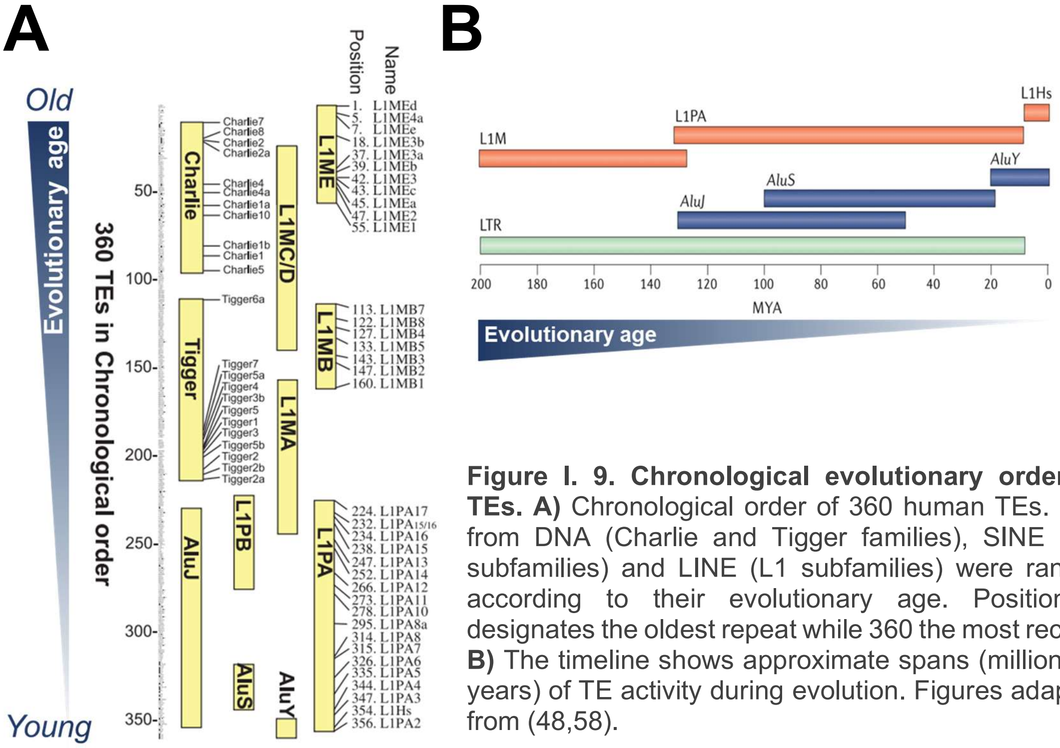


Figure I. 9. Chronological evolutionary order of TEs. A) Chronological order of 360 human TEs. TEs from DNA (Charlie and Tigger families), SINE (Alu subfamilies) and LINE (L1 subfamilies) were ranked according to their evolutionary age. Position 1 designates the oldest repeat while 360 the most recent. **B)** The timeline shows approximate spans (millions of years) of TE activity during evolution. Figures adapted from (48,58).

LINEs make up ≈20% of mammalian genomes. Different LINE families evolved before the mammalian radiation, but the most common mammalian LINEs are represented by LINE-1 (LINE-L1) family. A mammalian full-length L1 sequence includes a 5′ untranslated region (5′UTR) harboring an RNApol II promoter, two open reading frame proteins (ORF1p and ORF2p) and a 3′UTR that contains a poly A tail. LINE-L1 can propagate autonomously in the genome. However, most of LINE-L1 repeats are inactive in the human genome and LINE-L1 activity is restricted to certain copies of the most recently evolved L1 repeats, including L1PA1, L1PA2 or L1HS. Other active TEs arise from Alu and SVAs insertions, non-autonomous retrotransposons that rely on L1 machinery (55,56).

Besides LINEs, SINEs represent the other most prevalent repeat class in the genome. SINEs are transcribed by RNApol III and do not encode for proteins. In humans, SINE repeats are most represented by Alu elements, which have successfully expanded along the primates lineage. Alu are preferentially located within gene-rich regions and include more than one million copies that comprise ≈11% of the human genome (57). Other frequent SINE family are mammalian-wide interspersed repeats (MIRs), although they are a more ancient family and no longer mobile. Regarding Alu elements, Alu evolution is dominated by the amplification of new insertions that accumulate sequence variations. Thus, along primates evolution, different Alu subfamilies arised **(Figure I.9)**. The most ancient one is the AluJ subfamily, followed by AluS subfamilies. The dominant S subfamilies included Sx, Sq, Sp and Sc. More recently, AluY subfamily appeared, with AluYa5 and AluYb8 being the predominating Alu elements in humans (58). Indeed, many copies of AluS and specially AluY subfamilies remain active (59).

SVAs are retrotransposons with a composite structure. SVAs contain a hexameric repeat of the sequence CCCTCT, an antisense Alu-like element, a variable number of tandem repeats (VNTR) region, and a SINE region (SINE-R). Roughly 2700 SVA copies exist in the human genome, which are preferentially located within high-GC regions. SVAs are restricted to hominids and six different families exist in humans. These families are named SVA_A to SVA_F, according to their decreasing evolutionary age **(Figure I.10)**. While older families (SVA_A to SVA_D) are found in other apes species, SVA_E and SVA_F are human-specific (60,61).

Collectively, LINE, SINE and SVA are non-LTR retrotransposons. On the contrary, endogenous retroviruses (ERVs) contain LTRs. Full-length ERVs are composed of three genes, *gag, pol,* and *env,* flanked by LTRs. ERVs make up the ≈8% of the human genome and are the remnants of exogenous retroviruses that integrated into the germline millions of years ago. LTR retrotransposons also include shorter mammalian apparent LTR retrotransposon (MaLR) sequences (62).

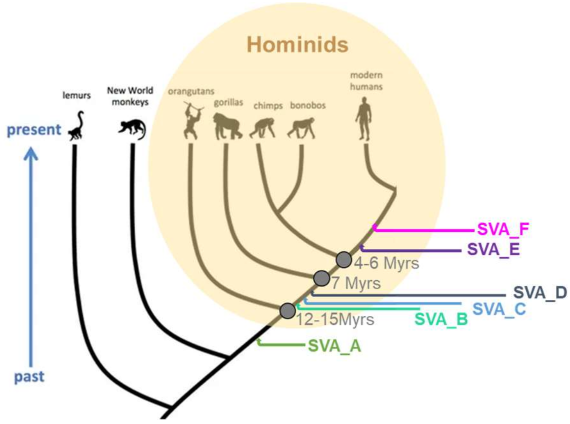


Figure I. 10. Evolution of SVA families. SVAs emerged in the common ancestor of apes but successfully proliferated only in great apes or hominids comprising human, chimpanzee, gorilla, and orangutan. SVA_A (13.56 million years ago (Myrs)) may have expanded contemporary to the divergence of the orangutan and the great apes (human, chimpanzee and gorilla) (12–15 Myrs). The expansion of families SVA_B (11.56 Myrs), SVA_C (10.88 Myrs) and SVA_D (9.55 Myrs) may have predated the human, chimpanzee and gorilla divergence (~7 Mya). Youngest subfamilies SVA_E (3.46 Myrs) and SVA_F (3.18 Myrs) may have expanded after the human and chimpanzee divergence (~4–6 Mya), thus being restricted to humans. Age estimation from (60).

1.3.2.1. Epigenetic regulation of transposable elements

As stated, TEs insertions represent waves of invasion of mobile elements that propagate in the genome up to the loss of their retrotransposition capacity (**Figure I.9**). Thus, the host genome has evolved different epigenetic mechanisms to mediate TE silencing. These mechanisms include DNA methylation (63–68), repressive histone modifications (67,69) and the recruitment of transcriptional repressors complexes involving HP1alpha, KAP1 or KRAB-KZNF proteins (70–74), among others. On the other hand, TEs have also co-opted in the host gene-regulatory networks by acting as cis-regulatory elements, challenging the traditional an outdated notion of 'junk DNA' (75–79).

DNA methylation is a major repressive mechanism of several classes of TEs in both developing and differentiated mammalian cells. Indeed, DNA hypomethylation causes upregulation of ERVs. Importantly, their derived ncRNAs are sensed in the cytosol, activating an anti-viral response and interferon pathway (63,64). This immune-triggering mechanism was termed 'viral mimicry' and highlights the impacts of a dysbalanced TEs regulation. DNA methylation has been more associated to younger TEs, especially ERVs. On the contrary, older ERVs are preferentially silenced by repressive histone modifications. Nonetheless, both mechanisms showed synergistic effects in TE repression (67). In fact, several studies report the existent crosstalk of multiple mechanisms in the regulation of TEs (68,71,80,81)

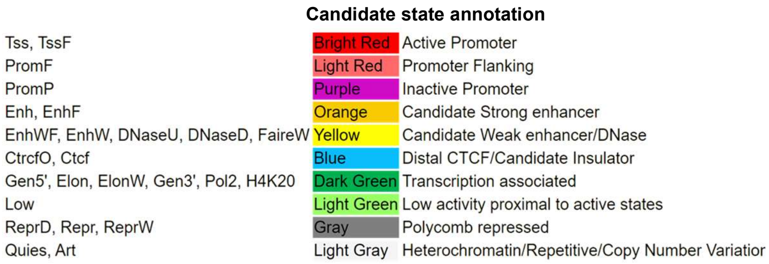
Although Alu elements tend to be highly methylated, it has been reported that DNA hypomethylation does not trigger Alu expression. In contrast, histone methylation may play a dominant role in their silencing (69). These repressive histone modifications are deposited by histone methyltransferases that concurrently can interact with transcriptional repressors. For instance, H3K9me3-methyltransferase SETDB1 is required for ERVs silencing through their interaction with KAP1 and KRAB-ZFP proteins (70,82). Further investigation showed that in mouse embryonic stem cells (mESCs), recruitment of DAXX, H3.3 and KAP1 to ERVs is co-dependent and occurs upstream of SETDB1 (83) . This serves as an example of how histone variants may also act as master regulators in TEs expression. For its part, KAP1-mediated repression has been involved in ERVs silencing in both embryonic and differentiated cells (70–72,82,84,85) but also associated with primate-specific retrotransposons of SVA and L1PA lineages (71,73,86).

KAP1 is recruited by KRAB-ZFP proteins, a large protein family whose evolution have allowed genomes to respond to retrotransposons invasion (87–89). In fact, concrete members repress specific TEs. For instance, ZNF91 is associated to SVA families repression while ZNF93 repressed certain TEs from the L1PA lineage (73,74).

Overall, existing data suggest that TEs are subjected to a specific regulation in which several mechanisms participate, complicating our understanding of TE regulation. Moreover, this epigenetic regulation seems to be context-specific but also dependent on the different classes and evolutionary age of the TEs. The determination of the chromatin landscape within repetitive elements, which can provide meaningful insights into their functional heterogeneity and regulation, remains limited. Thus, the study of additional epigenetic mechanisms emerges as a necessary matter of study.

1.4. Chromatin states or ‘colors’

The growing availability of genome-wide sequencing data has enabled a better understanding of the complex chromatin landscape. From the classical view of heterochromatin and euchromatin (**Figure I.3**), the mapping of all the already-discussed epigenetic determinants has enabled more complex segmentations. These segmentations originate from the computational analysis based on the combinatorial recurrence or absence of histone PTMs, transcription factors, DNA accessibility status, presence of repetitive elements, among others. The first approach was done in Drosophila and authors identified five different types of chromatin, which were referred as chromatin states or ‘chromatin colors’ (90). In human T cells, up to 51 distinct chromatin states were initially defined (91), which could be later grouped according to their predicted functionality. Analysis of multiple ChIP-Seq datasets in nine different human cell lines, identified 15 chromatin states and identify cell-type specific patterns at regulatory regions (92). Similar approaches have been done for multiple groups and consortiums in different cell models, allowing for a functional segmentation of the genome, useful for chromatin studies. **Figure I.11** depicts the 10 groups of chromatin states differentiated by ENCODE using ChromHMM segmentation (93), which will be used in the Results Section.



	Candidate state annotation	
Tss, TssF	Bright Red	Active Promoter
PromF	Light Red	Promoter Flanking
PromP	Purple	Inactive Promoter
Enh, EnhF	Orange	Candidate Strong enhancer
EnhWF, EnhW, DNaseU, DNaseD, FaireW	Yellow	Candidate Weak enhancer/DNase
CtrcfO, Ctcf	Blue	Distal CTCF/Candidate Insulator
Gen5', Elon, ElonW, Gen3', Pol2, H4K20	Dark Green	Transcription associated
Low	Light Green	Low activity proximal to active states
ReprD, Repr, ReprW	Gray	Polycomb repressed
Quies, Art	Light Gray	Heterochromatin/Repetitive/Copy Number Variation

Figure I. 11. Chromatin states. A common set of states were learned by computationally integrating ENCODE ChIP-seq, DNase-seq, and FAIRE-seq data using ChromHMM. Twenty-five states were used to segment the genome, and these states were then grouped (n=10) and colored to highlight predicted functional elements.

1.5. Three-dimensional organization of chromatin

While linear aspects of chromatin have been introduced, a proper chromatin three-dimensional structure is crucial to develop nuclear processes. Chromatin is spatially arranged in the nucleus and different levels of topological organization have been evidenced thanks to the development of chromosome conformation capture techniques and other approaches (**Figure I.12**).

1.5.1. Hierarchical chromatin organization

Chromatin packing involves several hierarchical levels of organization. First, chromosomes do not occupy random positions in the nucleus and form specific ‘chromosome territories’. Gene-poor chromosomes are placed at the heterochromatic nuclear periphery while gene-rich chromosomes were preferentially located in the nuclear interior (10).

Chromosome capture techniques, such as Hi-C, have provided genome-wide maps of how chromatin is three-dimensionally structured. At megabase scale, genome can be segregated into the so-called A and B compartments. The A compartment represents active, accessible chromatin with a tendency to occupy a more central position in the nucleus. The B compartment corresponds to heterochromatin and gene deserts enriched at the nuclear periphery (94). Topological associating domains (TADs) are sub-megabase structures that interact more frequently within themselves than with the rest of the genome (95–97). TADs are conserved across species and cell types and show a coordinated transcriptional status (98,99). These domains are formed by assemblies of chromatin loops with physical properties that, ultimately, depend on the histone composition and modifications of its resident nucleosomes. Loops domain formation has a direct functional outcome in chromatin, for instance permitting enhancer-promoter contacts. CTCF and cohesin proteins are present at TAD boundaries and according to the 'loop extrusion model', cohesin extrude DNA to generate chromatin loops until founding CTCF in a convergent orientation (100).

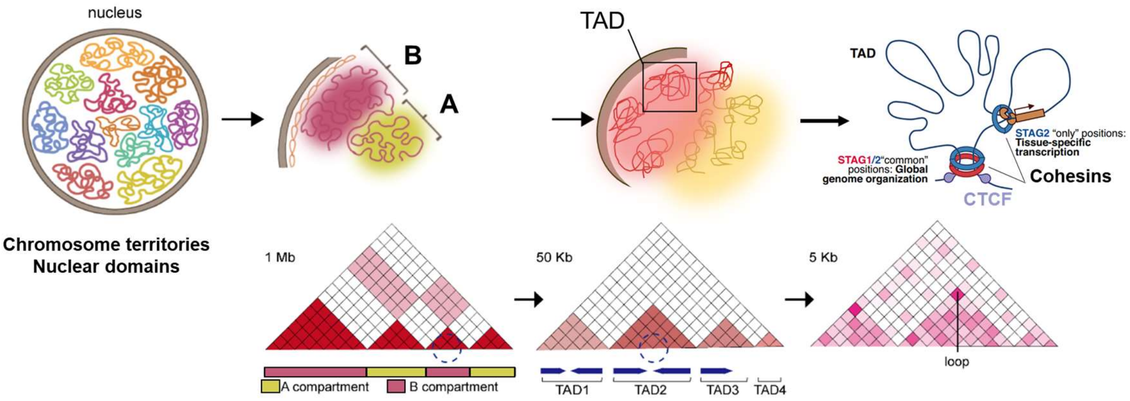


Figure I. 12. 3D genome organization. The topological structures of the 3D genome are organized at four levels, comprising chromosome territories, A/B compartments, TADs and chromatin loops. Contact matrices from Hi-C experiments are illustrated. Adapted from (101,102).

1.5.2. Nuclear domains

Other chromatin domains have been studied, which are crucial to tether chromatin within the nuclear space. Early microscopy experiments identified a condensed chromatin layer adjacent to the nuclear lamina. Further genome-wide experiments termed these regions as lamina-associated domains (LADs). Other example are nucleolus-associated domains (NADs), which constitute perinucleolar heterochromatin.

1.5.2.1. Lamina-associated domains

Chromatin anchoring to the nuclear lamina is a key determinant for the establishment of interphase genome spatial organization. Broadly speaking, LADs are defined as genomic regions that contact with the nuclear lamina. In human cells, they were first identified genome-wide by DamID technology, a proximity DNA labeling method using Dam-LaminB1 fusion protein (103). Human cells have 1000-1500 LADs that cover more than one-third of the genome. Importantly, while constitutive LADs (cLADs) are conserved among different cell types, the lamina association of facultative LADs (fLADs) is cell-type-variable (104,105).

LADs represent a well-known repressive environment, characterized by low gene density, low gene expression and a great overlap with B compartment. Regarding core histone modifications, LADs are marked by H3K9me2/3 histone marks. Nevertheless, while H3K9me3 is also a hallmark of non-LAD heterochromatin, H3K9me2 is considered a conserved chromatin mark more exclusively associated to LADs **(Figure I.13)**. Moreover, the H3K9me2 reader CEC-4 is necessary for LADs anchoring to the nuclear membrane in worms (106). However, in mammalian cells, inhibition of G9a (a H3K9me2 methyltransferase), only reduces but not ablates LADs interaction with the lamina (107,108). These observations suggest that, although having a conserved role in peripheral chromatin anchoring, H3K9me2 is unlikely a sufficient signal in mammals and other anchoring mechanisms exist.

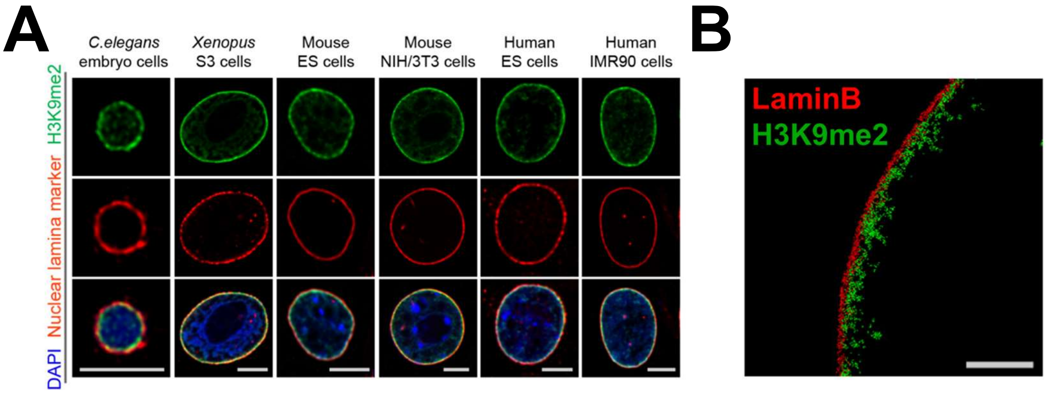


Figure I. 13. H3K9me2-marked chromatin at the nuclear periphery. A) H3K9me2 localization at nuclear periphery is evolutionary conserved. Confocal immunofluorescence images in different species are shown. **B)** Super-resolution microscopy (STORM) image from a mouse cell nuclei. Figure adapted from (106).

1.5.2.2. Nucleolus-associated domains

The nucleolus is a membraneless structure where ribosome biogenesis occurs. In humans, nucleoli assemble around tandem repeats of rDNA, located on the short arms of acrocentric chromosomes. In addition to ribosome biogenesis and regulation, nucleoli also act as central chromatin organizers. Genomic regions positioned close to nucleolus are referred as NADs. Initial genome-wide studies of human NADs were performed in HeLa (109) and later in IMR-90 cells (110), using a biochemical purification of nucleoli. Almost 40% of the genome was identified as NADs, which consist of mainly heterochromatic regions, characterized by gene deserts, low gene expression and repressive histone marks. Moreover, an important overlap with LADs was found. Accordingly, some LADs have been found to stochastically reshuffle after mitosis and associate with nucleoli (107). These observations suggest that the lamina and the nucleolus could act as interchangeable scaffold for heterochromatin positioning.

More recently, the inclusion of HiC-based approaches has provided more accurate NADs genome-wide maps, while clarifying their interconnexion with LADs (111,112). Concretely, in mESCs, NADs showed a different chromatin landscape from LADs (111). Anyway, a considerable overlap with LADs was also found, in accordance with previous evidence. NADs that overlap with LADs were found enriched in H3K9me2 and depleted in H3K27me3, compared to regions exclusively identified as LADs. In addition, a recent work performed in human cells used nucleolar-HiC to identify 264 NADs (112). These

NADs cover 24% of the genome and clustered in three subgroups based on their *trans* interaction frequencies. The three subgroups show differential chromatin properties, including distance to centromere regions and overlapping with LADs.

1.5.3. Giemsa-bands

The interphasic chromatin organization is highly remodeled during mitosis, where chromatin acquires a highly condensed state. Indeed, metaphase chromosomes represent another layer of chromosome architecture that has been studied for years. In 1970s, several staining methods of metaphase chromosomes arised, that is, Giemsa staining (113). Although the precise molecular basis of Giemsa has remained unknown for decades, it is widely accepted that staining correlates with AT-rich sequences and chromatin compaction (114,115). Nevertheless, although staining of G-positive bands (Gpos) negatively correlates with GC content, unstained bands (Gneg) are highly GC heterogeneous.

Giemsa bands (G-bands) have been useful in cytogenetics, allowing detection of chromosomes rearrangements in diseased cells. However, they have not been much explored in relation to functional genomics. During these years, we have *in silico* characterized G-bands and found them useful as epigenetic units to investigate the differential distribution of linker histones ((116) and Appendix-publication list).

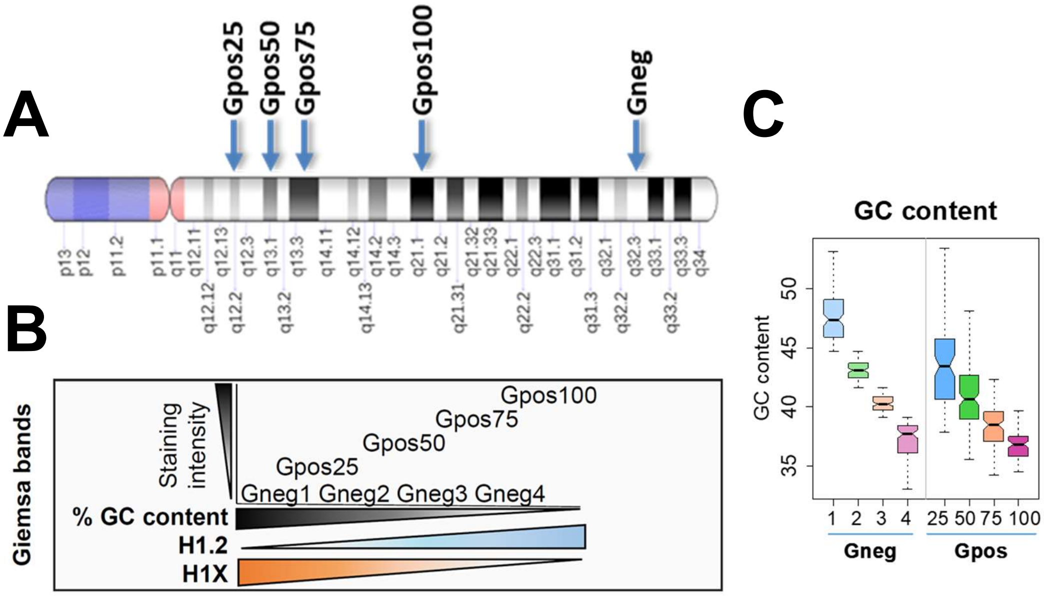
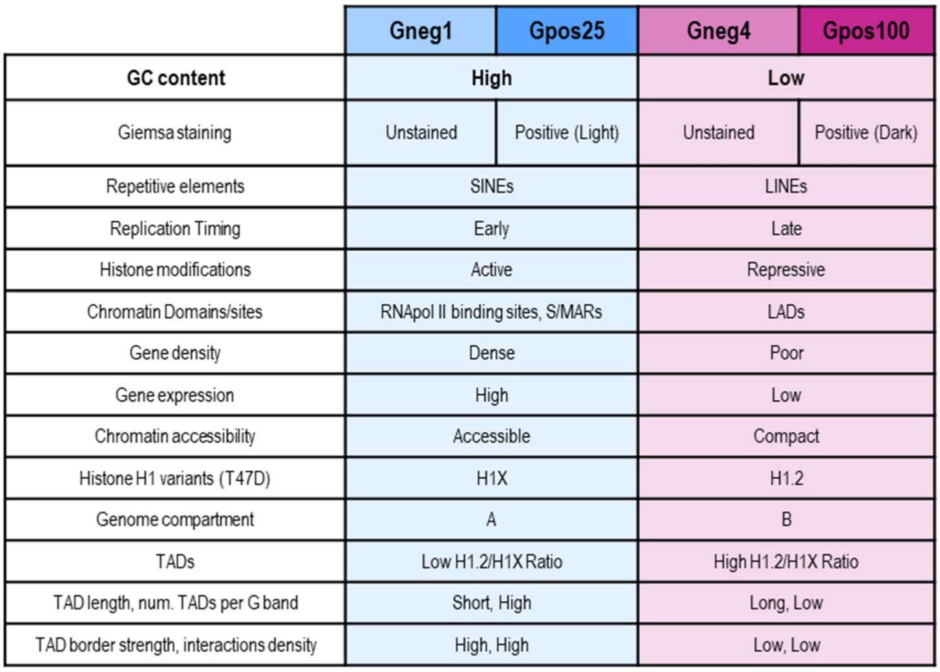


Figure I. 14. Giemsa bands classification. A) Idiogram of a human metaphase chromosome showing banding after Giemsa staining. **B)** G positive bands are classified in 4 types (Gpos25 to Gpos100) according to increasing staining intensity. Unstained bands or interbands are called G negative (Gneg). Gneg bands were classified in four equal groups according to their GC content. **C)** Boxplot shows the percentage of GC content of each G-band group.

Briefly, Gpos bands can be classified in four groups according to its relative Giemsa staining (i.e., Gpos100, Gpos75, Gpos50, Gpos25; being Gpos100 the most stained bands). We then created four equal groups of Gneg bands, based on their GC content (i.e. Gneg4, Gneg3, Gneg2, Gneg1; being Gneg4 the group with a lower GC content) **(Figure I.14)**. Integrating ChIP-Seq, ATAC-Seq and Hi-C data, we performed an epigenetic characterization of these eight groups of G-bands. In general, low-GC bands correlated with compact chromatin while high-GC bands show a strong correlation with active chromatin features (see **Table I.1**).

Overall, G-bands can recapitulate genomic properties of interphasic genome organization. G-bands-based segmentation is useful to compare the genome-wide distribution of chromatin-associated proteins, including histone H1 variants, as detailed in Section 2.3.5.

Table I. 1. Summary of chromatin and topology features of high and low-GC cytobands.



	Gneg1	Gpos25	Gneg4	Gpos100
GC content	**High**		**Low**	
Giemsa staining	Unstained	Positive (Light)	Unstained	Positive (Dark)
Repetitive elements	SINEs		LINEs	
Replication Timing	Early		Late	
Histone modifications	Active		Repressive	
Chromatin Domains/sites	RNApol II binding sites, S/MARs		LADs	
Gene density	Dense		Poor	
Gene expression	High		Low	
Chromatin accessibility	Accessible		Compact	
Histone H1 variants (T47D)	H1X		H1.2	
Genome compartment	A		B	
TADs	Low H1.2/H1X Ratio		High H1.2/H1X Ratio	
TAD length, num. TADs per G band	Short, High		Long, Low	
TAD border strength, interactions density	High, High		Low, Low	

2. Linker histone H1

As previously mentioned, linker histone H1 binds to the entry/exit sites of DNA at the nucleosome and it has been involved in the folding of higher-order chromatin structures. However, histone H1 is a diverse family and growing evidence supports that H1 variants play a more complex role in chromatin regulation. Our knowledge of H1 variants functionality is very limited in comparison to core histone variants or other chromatin-associated proteins, despite being among the most abundant proteins in the nucleus.

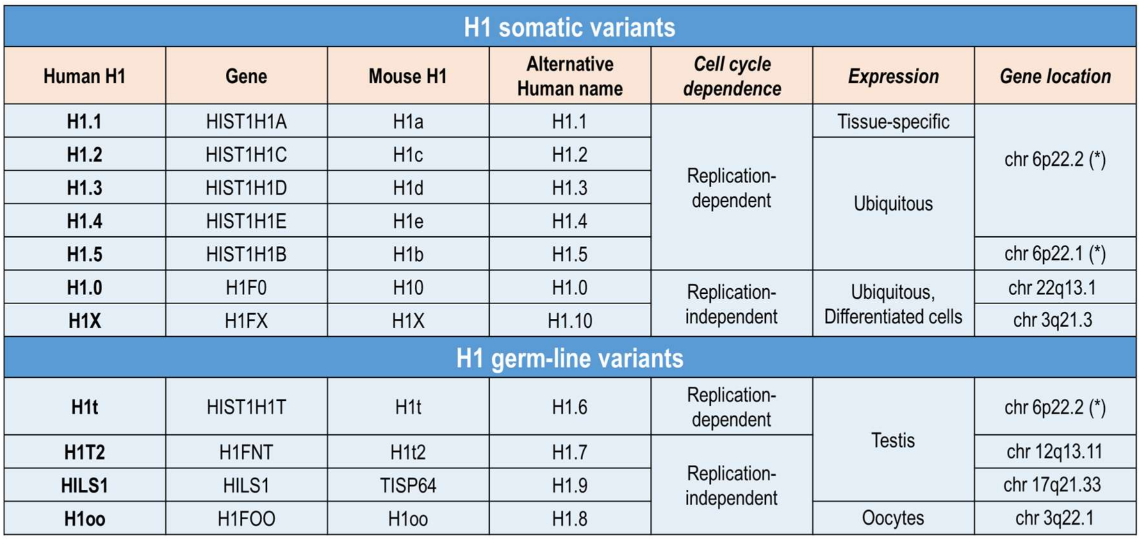
2.1. Histone H1 family

Core histones are highly conserved through evolution but H1 proteins are evolutionary diverse. This diversity ranges from a single H1 protein in yeast or two H1 variants in Drosophila (somatic dH1 and embryonic dbigH1) up to eleven H1 subtypes or variants present in mammals. The human histone H1 family includes seven somatic variants

(H1.1 to H1.5, H1.0, and H1X), three testis-specific (H1t, H1T2, and HILS1) and one oocyte-specific variant (H1oo) **(Table I.2)** (117–119).

Among H1 somatic variants, H1.1–H1.5 are expressed in a replication-dependent manner while H1.0 and H1X are replication-independent. Regarding their patterns of expression, H1.2 to H1.5 and H1X can be ubiquitously expressed, H1.1 is restricted to certain tissues, and H1.0 accumulates in terminally differentiated cells. H1X expression is also reported to be higher in differentiated cells (120). H1.1 to H1.5-encoding genes are located within the histone cluster (*HIST1* locus) of chromosome 6 (6p21-p22), whereas replication-independent variant genes are in different chromosomes outside the histone cluster. Replication-dependent H1 genes are intronless with short 5′ and 3′ ends. Transcripts are not polyadenylated but contain a 3′ stem-loop that allows for a rapid post-transcriptional processing during S phase. On the other hand, H1.0 and H1X genes also lack introns but their mRNAs are polyadenylated.

Table I. 2. Human histone H1 family. Human cells may contain up to eleven histone H1 variants. Asterisks (*) indicate that genes are located in the histone cluster of chr6. Mouse ortholog H1 variants are indicated and defined with a different nomenclature. Alternative human nomenclature refers to the one used in (121).



H1 somatic variants						
Human H1	**Gene**	**Mouse H1**	**Alternative Human name**	***Cell cycle dependence***	***Expression***	***Gene location***
H1.1	HIST1H1A	H1a	H1.1	Replication-dependent	Tissue-specific	chr 6p22.2 (*)
H1.2	HIST1H1C	H1c	H1.2		Ubiquitous	
H1.3	HIST1H1D	H1d	H1.3			
H1.4	HIST1H1E	H1e	H1.4			
H1.5	HIST1H1B	H1b	H1.5			chr 6p22.1 (*)
H1.0	H1F0	H10	H1.0	Replication-independent	Ubiquitous, Differentiated cells	chr 22q13.1
H1X	H1FX	H1X	H1.10			chr 3q21.3
H1 germ-line variants						
H1t	HIST1H1T	H1t	H1.6	Replication-dependent	Testis	chr 6p22.2 (*)
H1T2	H1FNT	H1t2	H1.7	Replication-independent		chr 12q13.11
HILS1	HILS1	TISP64	H1.9			chr 17q21.33
H1oo	H1FOO	H1oo	H1.8		Oocytes	chr 3q22.1

Of note, certain H1 genes located within the same histone cluster show great differences on their expression patterns, as occurs with H1.1 (tissue-specific) or H1t (testis-restricted). This observation suggests that expression of H1 variants is specifically regulated at the transcriptional level. However, the exact molecular mechanisms underlying this H1-variant-specific expression are poorly understood.

Like core histones, H1 variants are lysine-rich proteins that present a tripartite structure composed of two N-terminal and C-terminal tails and a central globular domain. The globular domain is highly conserved and represents the DNA binding region to the dyad axis of the nucleosome. On the contrary, the terminal tails are intrinsically disordered and more divergent between variants, especially the C-terminal.

2.2. Functional role of histone H1. Models for H1 depletion

2.2.1. Histone H1 and chromatin structure

Classically, histone H1 has been seen as a general structural component of chromatin, participating in the formation of higher-order structures. Early *in vitro* studies determined that loss of H1 leads to a more decondensed chromatin, prompting to a more beads-on-a-string or open state (122). Accordingly, H1 binding induces a more compact and rigid nucleosome conformation, which in turn could facilitate the assembly of higher-order structures (123). Moreover, the precise mode of H1 binding can regulate the chromatin structure (124). The 'on-dyad' H1 binding is suggested to result in poor compaction compared to the 'off-dyad' binding, which is proposed to allow for chromatin fiber condensation **(Figure I.15)**. Super-resolution microscopy experiments that support nucleosome arrangement in discrete groups or 'clutches' revealed that histone H1 is enriched within larger and denser clutches that formed closed heterochromatin (6).

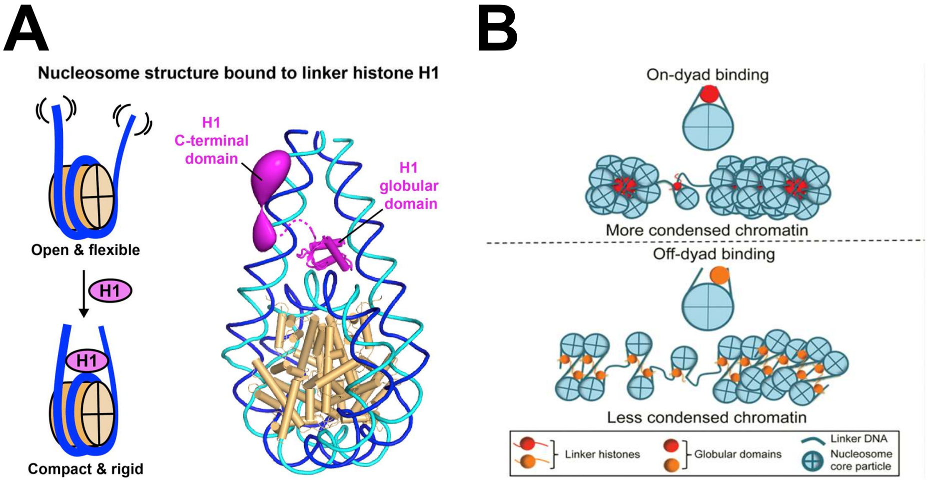


Figure I. 15. Histone H1 and nucleosome structure. A) Histone H1 shifts the conformational landscape of the nucleosome by drawing the two linkers together and reducing their flexibility. Figure from (123). **B)** Histone H1 can recognize the nucleosome by on- or off-dyad binding mode. These two binding modes lead to distinct higher-order structures of chromatin. Figure adapted from (124).

Histone H1 stoichiometry is directly related to nucleosome repeat length (NRL), being a high H1-to-nucleosome ratio associated with longer NRLs. Structural studies suggest that restricted H1 binding in nucleosome arrays with short NRLs could facilitate transcription whereas long NRLs enable H1 binding and thus can form heterochromatin regions (125). Functional studies support that H1 loss is associated with an increase in NRL (122,126–128).

Recent studies have pointed to phase-separation forces as a key determinant in chromatin organization and compaction, allowing for the maintenance of membraneless chromatin compartments. In this context, histone H1 is reported to promote phase separation (129–131), in agreement with its structural role as a chromatin condenser.

Histone H1 can shape other layers of chromatin organization, as denoted by Hi-C experiments performed on triple H1 knock-out (TKO) mESCs (132). These cells harbor a depletion of H1c, H1d, H1e (mouse orthologs of H1.2, H1.4, H1.5), which results in a 50% reduction of the total H1 content (127). Although TAD organization is largely unaltered, inter-TAD interactions increase in TKO cells and interaction changes correlate with epigenetic alterations within TADs. In particular, H1 loss results in a preferential gain of H3K4me1 and H3K4me3 within the most gene-dense TADs, which also showed a loss in DNA methylation **(Figure I.16)**.

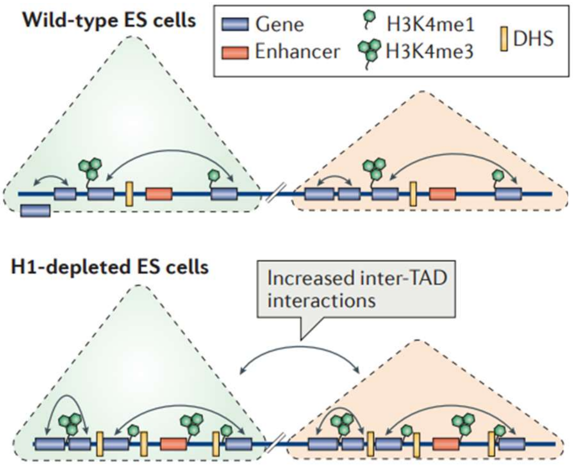


Figure I. 16. TAD chromatin organization upon H1 loss in mESCs. In H1 TKO mESCs the overall genome partition into TADs is not majorly affected. However, within gene-dense TADs, long-range inter-TAD interactions increase and new DNAse hypersensitive sites (DHSs) and sites of H3K4me1 and H3K4me3 are established. Adapted from (2).

2.2.2. Other functions of histone H1

Apart from its role in chromatin structure, histone H1 has been involved in multiple cellular processes, including transcription, replication, genome stability, DNA damage, splicing, apoptosis or heterochromatin maintenance, among others (126,133–140).

In H1 TKO mESCs, H1 reduction alters replication initiation landscape, generating massive replication fork stalling and DNA damage as a result from transcription-replication conflicts (133). In agreement, functional studies in *D.melanogaster* revealed that histone H1 (dH1) prevents R-loop accumulation, thus avoiding increased DNA damage and genome stability in heterochromatin (138).

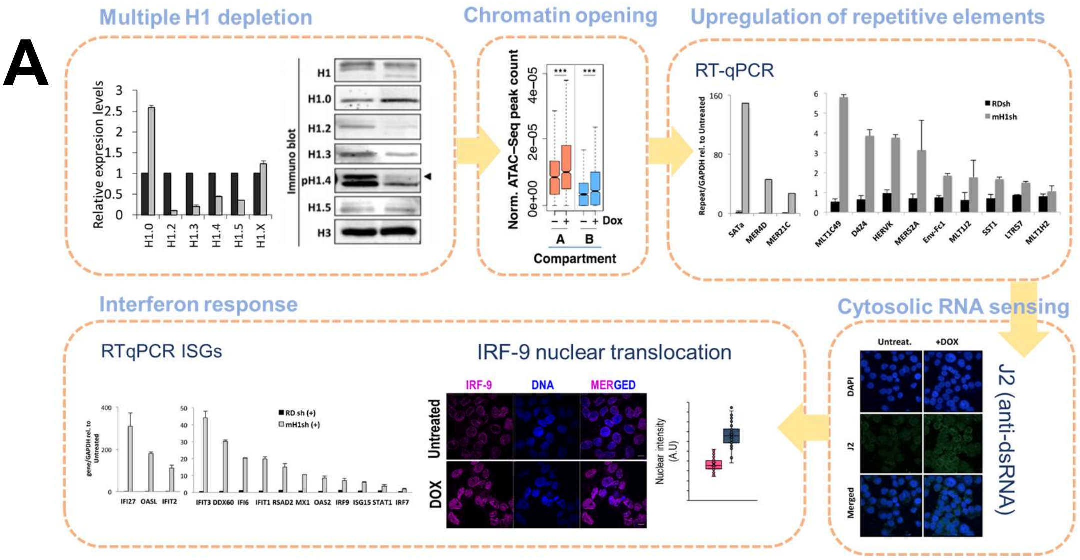
Histone H1 depletion in different cell models has shown that a limited number of genes are either up- or down-regulated, suggesting its role in gene expression regulation rather than functioning as a general transcriptional repressor. Moreover, depletion of different H1 variants deregulates a different subset of genes, pointing to H1-variants-specificity, as mentioned later. Transcriptional alterations upon H1 deficiency are not restricted to coding genes, though. In H1 TKO mESCs, de-repression of major satellites occurs (127) and recent work denoted an aberrant turnover of non-coding transcripts (141). Upon H1 loss, increased transcription of Che-RNAs (see section 1.2.4) occurs, together with reduced levels of RNA-m6a modification, leading to their accumulation on chromatin.

Multiple H1 depletion has also been associated with transcription and replication defects in human cells. Due to these cells are further used in this thesis, consequences of multiple H1 depletion in human cells are detailed in next section 2.2.3.

2.2.3. Consequences of multiple H1 depletion in human cells

In a previous work of the lab, consequences of the simultaneous depletion of multiple H1 variants were evaluated in T47D breast cancer cells (126). Specifically, an inducible short-hairpin-RNA (shRNA) system affecting the expression of several H1 genes (hereafter multiH1 KD) was used. At mRNA level, expression of all replication-dependent H1.2-H1.5 H1 variants is reduced but only H1.2 and H1.4 protein levels decrease. Because of the total H1 reduction, H1.0 becomes upregulated at both mRNA and protein levels. Overall, total H1 content is reduced ≈30%.

MultiH1 KD drastically impairs cell growth and induces a strong interferon (IFN) response (**Figure I.17**). This immune response is evidenced by the activation of IFN signaling transducers, participation of cytosolic nucleic acids receptors, IFN synthesis, up-regulation of IFN-stimulated genes (ISGs) and IRF-9 nuclear translocation. Mechanistically, the IFN response is mediated by the sensing of cytosolic RNAs derived from aberrantly transcribed endogenous repetitive elements (i.e. 'viral mimicry'). H1 reduction causes chromatin opening genome-wide, triggering an increased transcription of satellites and ERVs. Importantly, this phenotype occurs without global changes in histone marks abundance nor at the activated ISGs and repeats. Nevertheless, histone marks status was only evaluated at concrete gene promoters and putative changes could occur genome-wide. Another unsolved question is whether the remaining H1 variants redistribute to compensate H1 reduction, which was technically limited by the lack of specific ChIP-grade antibodies.



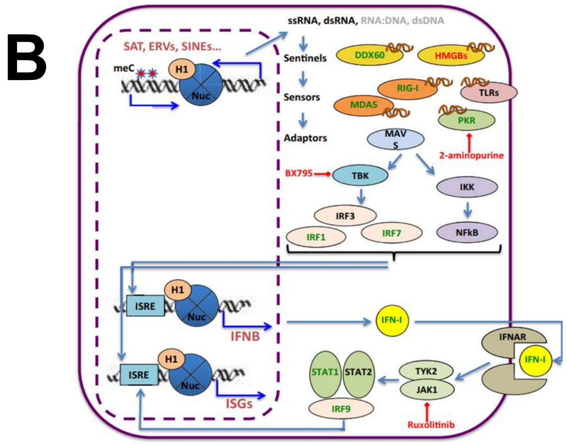


Figure I. 17. Combined H1 depletion in T47D triggers an interferon response via activation of repeats. A) Selected evidence that support the mechanism by which multiple H1 depletion triggers an immune response. Briefly, multiple H1 depletion leads to genome-wide chromatin opening, causing an upregulation of repeats. Their derived-RNAs are exported to the cytosol and recognized by sensors, triggering an anti-viral response. **B)** Illustration depicting the model. Most of the experiments and the cartoon model are taken from (126).

During these last years, we further addressed how this multiH1 depletion impact chromatin structure (142) and non-coding transcription (141). This concrete data is not included in the Results section, but I will briefly highlight here some of the conclusions obtained. Both papers are also included in the Appendix-Publication List.

Upon multiH1 KD, TAD organization is largely unaltered but TADs showed increased internal interactions, a more decompacted structure and an increase in TAD border strength. Besides, the structural changes correlated with a coordinated gene expression response within TADs (142).

In a work led by M.Gómez group (141), we observed that multiH1 KD produces the aberrant chromatin retention of Che-RNAs, similar to H1 TKO mESCs **(Figure I.18A)**. Che-RNAs genomic regions were marked by H3K4me1, as previously reported. Besides, upon multiH1KD their levels of H3K4me3 increased **(Figure I.18B)**. Importantly, as H1 TKO mESCs, multiH1 KD also exhibit a transcription-dependent replicative stress.

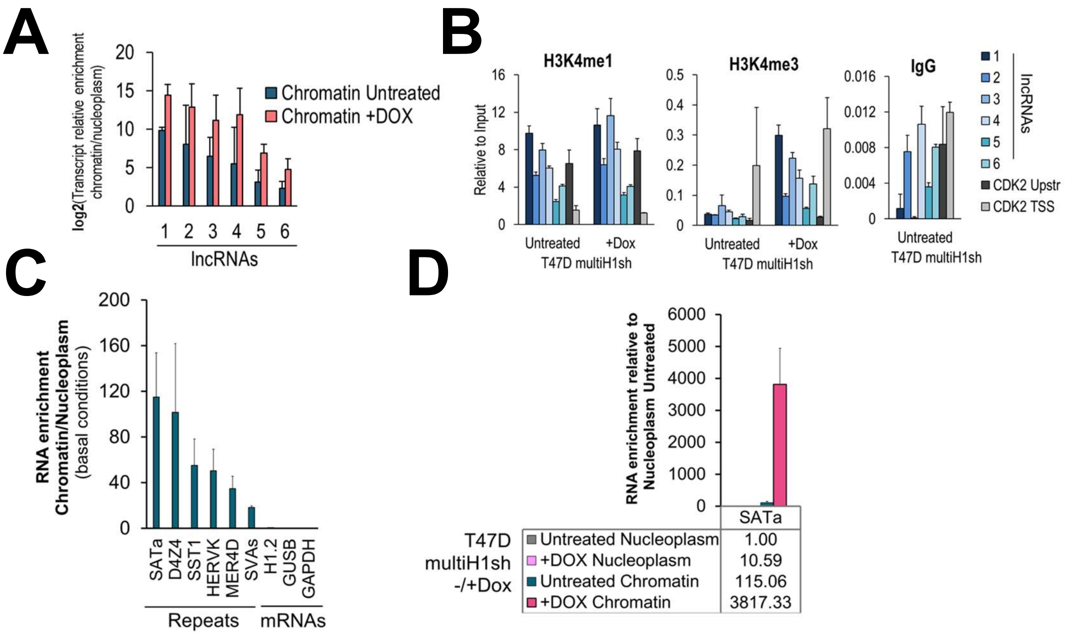


Figure I. 18. Chromatin retention of non-coding transcripts upon multiH1 KD in T47D cells. A) Che-RNA enrichment of lncRNAs in T47D multiH1sh -/+Dox. RT-qPCR data represent mean and SD of four independent biological replicates. Data is expressed relative to Nucleoplasm (Untreated) and in log2 scale. **B)** ChIP-qPCR of H3K4me1 and H3K4me3 in T47D multiH1sh -/+Dox. Data is expressed relative to Input sample. A representative experiment is shown. IgG is used as non-specific control. *CDK2* is used as a control of active gene to show H3K4me3 peak at TSS compared to Distal promoter regions (3Kb Upstream from TSS). Primers used for lncRNAs are <1kb upstream from their predicted TSS. **C)** Chromatin enrichment of repetitive elements under basal conditions. Data refers to the transcript chromatin enrichment compared to Nucleoplasm. **D)** Chromatin retention of SATa upon multiH1 KD depletion. Transcript enrichments in the mentioned fractions are expressed relative to Nucleoplasm under Untreated (basal) conditions. C,D) Data is expressed as mean and SEM from six independent biological replicates.

Notably, we observed that repetitive elements transcripts reside in chromatin under basal conditions (**Figure I.18C**). Interestingly, upon multiH1 KD, an increased SATa retention in chromatin occurs (**Figure I.18D**). This suggests that apart from their cytosolic exportation, SATa-derived-RNAs aberrantly accumulated in chromatin could also have a deleterious impact on heterochromatin maintenance, which could partially mediate the structural changes and phenotype observed. In mice, major satellite repeat transcripts also remain chromatin-associated and are crucial for maintaining a proper heterochromatin structure (49).

2.3. Evidence for H1 variants specificities

Whether human H1 variants are redundant or display specific biological functions is a matter of debate. The studies mentioned in previous Section 2.2. evaluate a scenario in which several H1 variants have been depleted, so consequences on chromatin cannot be directly attributed to a particular H1 variant, but rather to the overall reduction in H1 content. However, other combinations of double and triple H1 KDs in T47D breast cancer cells, which suppose an equivalent reduction in H1 content compared to multiH1 KD (i.e. H1.2+H1.4 KD) do not recapitulate the strong IFN response observed in multiH1 KD cells (126). These observations are an example of how both redundance and specificity may play a role in defining H1 variants functionality.

The initial observation of single or double KO mice without an apparent phenotype led to think that H1 variants could compensate for the loss of other variants (143). Nevertheless, compensation was not observed in a triple somatic H1 KO, which causes embryonic lethality (128). mESCs derived from these mice present an altered chromatin functionality, as has been explained in Section 2.2. Although these early studies support H1 variants redundancy and an importance of the total H1 pool over the specific variants, growing evidence points to a more complex H1 variants-specific dynamics.

2.3.1. Sequence conservation

The first piece of evidence comes from evolution, as H1 ortholog genes are much more conserved than paralog genes. In other words, the amino acid sequence of a given H1 variant is more similar across species compared to other H1 variants from the same species (144).

2.3.2. H1 variants expression patterns

H1 variants are subjected to a specific regulation of their expression patterns. This regulation may be cell-type specific, as the relative ratio of H1 variants contribution to total H1 content is variable between cell types, being H1.2 and H1.4 the only H1 variants found in all cells investigated. Although the exact mechanisms of H1 variants expression regulation are not understood, it is clear that they cannot be explained by the coordinated regulation of their gene loci and a variant-specific regulation exists.

In addition, H1 variants expression patterns change along differentiation or cancer progression. Pluripotent cells present a wider repertoire of H1 variants, with lower levels of replication-independent H1 variants and higher levels of H1.1, H1.3 and H1.5, in comparison with differentiated cells (120). Indeed, H1.0 progressively accumulates along differentiation (120,145,146). H1X has also reported to accumulate along differentiation of human embryonic stem cells and upon retinoic-acid induced differentiation of human NT2 cells (120).

Abnormal H1 variants expression has been associated to certain cancers and specific H1 variants have been even proposed as biomarkers (reviewed in (147)). However, their functional implications as main drivers in the oncogenic process are unclear and many of the observations rely on correlations between H1 variants expression and tumor type or malignancy. The best characterized association with tumorigenicity has been described for H1.0. H1.0 levels are intra-tumor heterogeneous, with reduced H1.0 levels in cells characterized by long-term self-renewal capacities and tumorigenic potential (148). H1.0 silencing is required to maintain self-renewal of the tumor-maintaining cells and re-establishment of H1.0 levels restricts the proliferative potential. Importantly, H1.0 has been reported to act as the major mediator of Quisinostat anti-tumor effect, a second-generation histone-deacetylase inhibitor used for cancer therapy (149).

2.3.3. Differential gene expression regulation

In human T47D breast cancer cells, inducible KD of each of the somatic H1 variants caused deregulation of a limited subset of genes, as denoted by microarray and RNA-Seq experiments (126,150,151). Deregulated genes were affected in both directions (upregulated or downregulated) and in a variant-specific manner. Notably, T47D H1.2 KD caused cell cycle arrest in G1 and downregulation of cell-cycle related genes (150). Moreover, a decrease in NRL was observed in H1.2 KD but not for the other single H1 KDs, highlighting again the existence of functional specificity. Downregulation of cell-cycle genes also occurs in H1.4 KD and multiH1 KD (126).

However, in transcriptomic analyses is difficult to distinguish between direct and indirect effects. Because H1 variants can influence a wide repertoire of nuclear processes, many of the deregulated genes may not be directly regulated by the given H1 variant which is being depleted. In fact, analysis of H1.4 KD and multiH1 KD revealed that upregulated genes present a low basal gene expression (126). Those low-expressed genes are characterized by high H1 levels at their promoter of all the variants (126,152), not specifically enriched within a particular H1 variant.

2.3.4. H1 code: Histone H1 post-translational modifications

H1 variants can be specifically post-translationally modified **(Figure I.19)**, which can modulate protein functionality or promote their interaction with specific partners, as occurs with core histones. Thus, PTMs add an additional level of complexity in H1 variants heterogeneity.

Phosphorylation is probably the best characterized H1 modification. H1 phosphorylation dynamics are related to cell cycle progression and H1 proteins are highly phosphorylated in mitosis (153–157). However, certain H1 phosphorylations are also found in interphase. H1.2-pS173, and H1.4-pS187 were found in interphasic nuclei of HeLa cells, while H1.4-pT154 was restricted to mitosis (154). Regarding their nuclear patterns, H1.2-pS173 and H1.4-pS187 are enriched in nucleoli, where H1.4-pS186 is preferentially associated with active rDNA and hormone-responsive promoters. While interphasic H1 phosphorylation has been primarily associated to transcriptional activation (154,158–161), mitotic hyperphosphorylation is thought to promote chromosome condensation (162,163).

Interestingly, H1.4 interaction with HP1 proteins is controlled by a ‘phospho-switch’ mechanism (164). HP1 binds to H1.4-K26me but phosphorylation on adjacent S27 prevents HP1 from binding, thus revealing a crosstalk between H1 PTMs and heterochromatin formation.

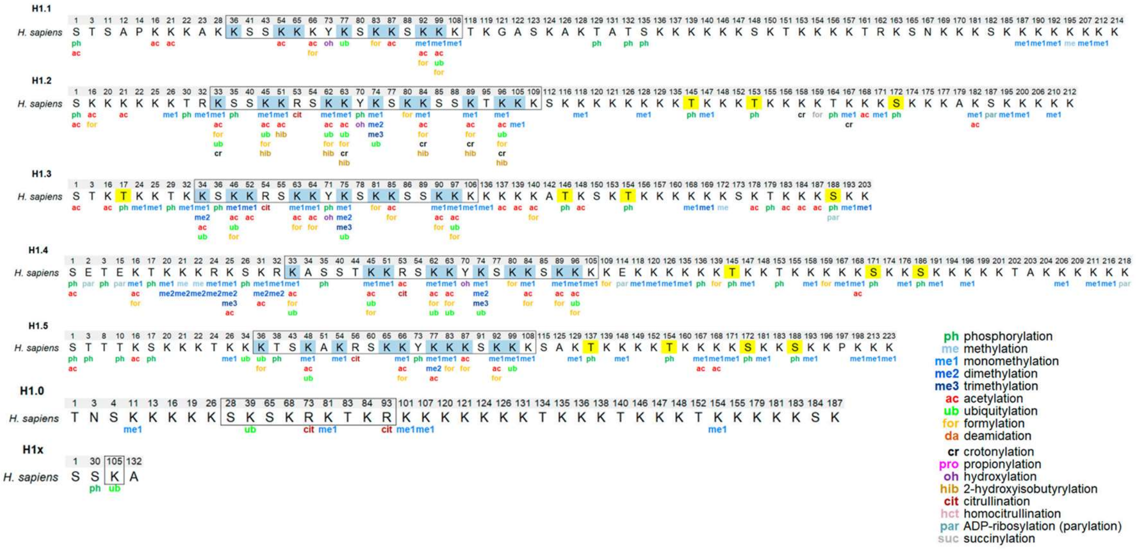


Figure I. 19. H1 variants post-translational modifications. Human somatic H1 variants sequence and their PTMs are shown. Gray boxes indicate the residues in the globular domain. In yellow, phosphorylated residues located at CDK-consensus motifs. In blue, PTM-hotspots in the globular domain. The positions refer to the mature protein, which lacks the initial methionine. Figure adapted from (165).

2.3.5. Differential genomic distribution of H1 variants

Genomic distribution of a protein is tightly linked to their biological function. In this context, although genome-wide maps of H1 variants distribution would be crucial to fully understand H1 variants functional specificity, studies have been limited by the lack of specific ChIP-grade antibodies. To overcome this, several overexpression strategies have been used, although caution must be taken when interpreting results from exogenous H1 variants. Moreover, another drawback of H1 distribution studies has been their limitation to a single variant, hindering a direct comparison between multiple reports.

Initial genome-wide studies used a total H1 antibody to perform ChIP-chip in MCF-7 cells. H1 is not uniformly located thought the genome and a clear depletion of H1 is seen in active TSS, which was referred to as ‘H1 valley’ (166). Genome-wide analysis of *Drosophila* somatic H1 also showed its exclusion from active promoters (167).

Later studies performed the genome-wide mapping of concrete H1 variants in both human and mice. In mESCs, a knock-in system was used to perform ChIP-Seq of N-terminal flagged H1c, H1d (H1.2 and H1.3 orthologs) and overexpressed H1.0 (168). Both H1c and H1d are enriched within AT-rich regions, positively correlate with H3K9me3 and are abundant at major satellites. Notably, the aforementioned ‘H1 valley’ around TSS of active genes was also reported (**Figure I.20A**). Although substantial overlap, these H1 variants show preferential association with some repetitive elements classes. H1c was more associated to satellites, H1d to LINEs and H1.0 was more enriched within minor satellites and LINE-L1.

In human cells, two different reports performed a systematic study of somatic H1 variants, although overexpression strategies were used (152,169). By DamID technology, genome-wide maps of somatic replication-dependent H1.1 to H1.5 were generated (169). H1.2–H1.5 are depleted from CpG-dense, regulatory regions, promoters **(Figure I.20B)** and associated to repressive histone marks. On the other hand, H1.1 shows a more distinct profile, being present at promoters (in comparison with the other H1 variants) **(Figure I.20B)** and associated to polycomb regions. In breast cancer cells, ChIP-Seq of endogenous H1.2 and H1X was performed. In addition, hemagglutinin (HA)-tagged recombinant H1 variants (H1.2-H1.5 and H1.0) were also profiled by ChIP-Seq or ChIP-chip (152). H1.2 correlates with low-gene expression and is abundant at low-GC content regions and LADs, compared to the rest of the variants. Again, H1 variants were found to be depleted at active promoters **(Figure I.20C)**. Thus, H1 variants presence at TSS is inversely correlated with the transcriptional status of the gene, although differences between variants exist **(Figure I.20)**. Subsequent analysis of the data revealed that H1X is enriched within active chromatin, RNA pol II-enriched regions, coding regions and hypomethylated CpG islands (151). Moreover, H1.0 (i.e. H1.0-HA) is enriched at NADs ribosomal DNA, and acrocentric and telomeric satellites (151).

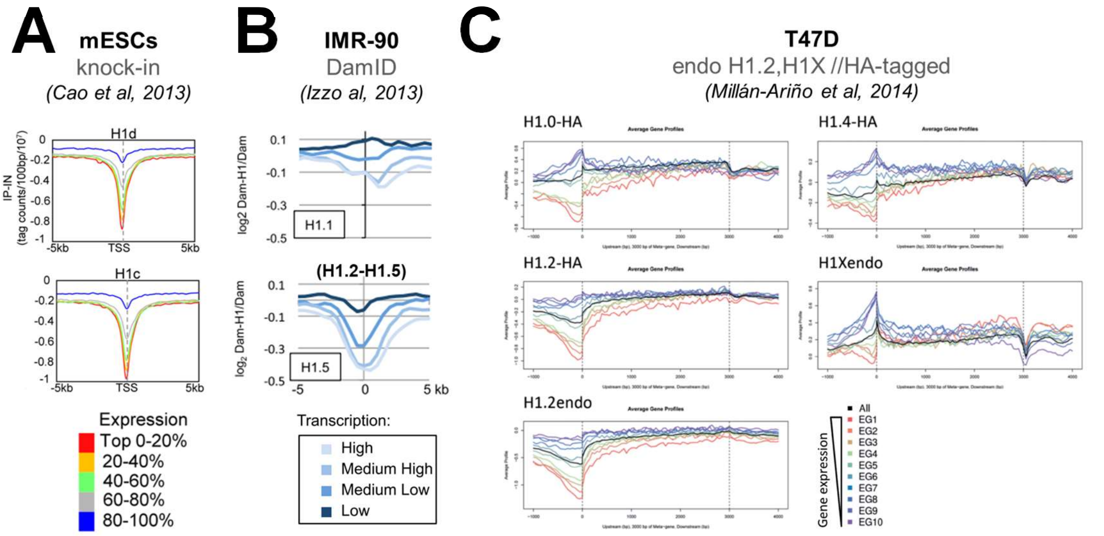


Figure I. 20. H1 variants are depleted from promoters of active genes. The so-called 'H1 valley' was first evidenced by using a total H1 antibody (166) and later confirmed by analysis of multiple H1 variants in different cell models. **A)** Average profile of H1d (H1.3) and H1c (H1.2) around TSS in mESCs (from (168)). **B)** Average profile of H1.1 and H1.5 around TSS in IMR-90. H1.2-H1.4 showed an analogous profile to H1.5, while H1.1 profile was more divergent (from (169)). **C)** H1 variants profile in meta-genes in T47D cells (from (152)). A,B,C) Genes are ranked by their increasing basal expression and a limited number of groups were formed in each case.

As mentioned, the referenced studies used tagged recombinant forms of H1 variants, through different approaches. It is important to highlight that differences with their endogenous counterparts were reported, as evidenced by comparing endogenous H1.2 mapping in T47D cells to H1.2-HA results (152). Apart from endogenous H1.2 and H1X in breast cancer cells, only a few publications have performed a genome-wide profiling of endogenous H1 variants, but they were limited to a single H1 variant.

In IMR90, endogenous H1.5 was associated with gene repression and binding of SIRT1 and H3K9me2 (170). Concretely, it forms blocks of enrichment in locus enriched for gene family clusters. In human skin fibroblasts, ChIP-Seq mapping of endogenous H1.0 revealed a strong correlation with GC content, its abundance at gene-rich chromosomes and a positive correlation with H3K27me3 binding (148).

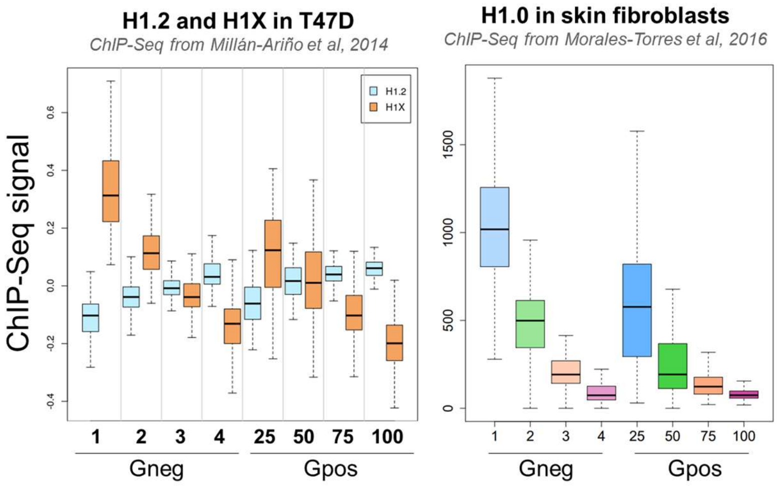


Figure I. 21. Giemsa bands are useful epigenetic units to compare H1 variants genome-wide profiles. G-bands genomic segmentation allows for a direct comparison between different H1 variants within a cell line and also to compare a given H1 variant profile in multiple cell lines. As described in section 1.5.3, G-bands groups correlate with GC content and other epigenetic and topological features of the genome. ChIP-Seq data from endogenous H1.2 and H1X in T47D (from (152)) and endogenous H1.0 in transformed skin fibroblasts (from(148)) are shown. Endogenous H1.2 in T47D is enriched within low-GC G-bands, while both H1X in T47D and H1.0 in skin fibroblasts are more abundant within high-GC bands.

Overall, our understanding of how H1 variants are distributed in the genome and whether their distribution is cell-type-dependent is limited by the lack of specific ChIP-grade antibodies. Moreover, the direct comparison of existing reports can be biased as both experimental and analytical approaches differ between studies. To facilitate comparison of H1 variants profiles we found useful G-bands segmentation, which as highlighted in section 1.5.3, correlates with epigenetic and topological genome features **(Table I.1)**. Thus, different H1 variants within the same cell line can be compared, as a given H1 variant in multiple cell lines **(Figure I.21)**.

Objectives

The present PhD thesis aimed to deepen our understanding of the heterogeneity of somatic variants of human histone H1. Specifically, it pursued to investigate their genomic distribution in a breast cancer cell line and compare H1 variants distribution across multiple human cellular models. To achieve this, the following objectives were proposed, divided into two main sections.

Chapter 1. Differential distribution of six endogenous histone H1 variants in T47D breast cancer cells

- To validate specific ChIP-grade antibodies to map endogenous H1 variants.
- To analyze the differential genomic distribution of somatic H1 variants, with a particular focus on repetitive elements.
- To analyze the H1 variants nuclear distribution and their relative abundance within chromatin compartments.
- To evaluate the putative compensatory mechanisms, in terms of H1 variants redistribution, occurred upon multiple H1 depletion.

Chapter 2. Heterogeneity of H1 variants content, regulation and distribution across cell lines

- To study the variable H1 complement across human cell lines and possible mechanisms of H1 variants expression regulation.
- To analyze the nuclear distribution of H1 variants within a subset of human cell lines.
- To focus on the differential genome-wide distribution of H1X across multiple cell lines.

Materials and Methods

1. Experimental Methods

1.1. Oligonucleotides

Primers were designed with Primer3 and their amplification was firstly checked *in-silico* by using In Silico-PCR tool from UCSC Genome Browser. *UCSC Genes* option was used for primers that amplify cDNA (i.e., designed in different exons for RT-qPCR) and *Genome Assembly* option was used for primers that amplify genomic DNA. Primer efficiency was experimentally checked by calculation of a qPCR amplification standard curve and amplified DNA was run in an agarose gel. Amplicon size was compared to *in-silico* amplification results.

Primer sequences used for qPCR are listed in **Table M.1**.

1.1.1. Considerations about primer design of young transposable elements

Primer design on repetitive sequences is difficult due to their repetitive nature and the sequence similarity between related repeats. To design specific primers for young repeat groups or families and avoid cross-reactivity with other similar repeats we used *an in-silico* approach. Consensus sequences were extracted from Dfam database (171), and multiple sequence alignment of similar repeats or families of the same class was performed with ClustalW. Designed primers were specific according to this multiple sequence alignment and according to all matches returned by In-Silico PCR UCSC Genome Browser tool. Predicted In-silico PCR amplification is indicated in **Table M.1**.

1.2. Antibodies

Primary antibodies used for ChIP/ChIP-Seq, immunoblot and immunofluorescence experiments are listed in **Table M.2**. Importantly, performance and specificity of H1 variants ChIP-grade antibodies were extensively assayed. This validation is included in **Figures R1-R5** in Results section.

Of note, the following H1 variants antibodies were tested for ChIP performance with negative results: H1.3 (abcam, ab24175), H1.0 (abcam, ab11079), H1.0 (Invitrogen, PA5-38570), H1.2-pT165 (Millipore, 06-1370), H1.3/H1.4-pT17 (abcam, ab250946). H1.5 (abcam, ab24175), which has been used for ChIP-Seq elsewhere (170), showed only limited ChIP performance in our hands with poor reproducibility.

Table M.1 Primers. Table includes primer sequences used for qPCR. For repeats primers, In Silico-PCR matches are listed.

			PRIMER NAME	SENSE	SEQUENCE (from 5' to 3')	REGION/COMMENTS
Genes	H1 variants	**H1.0**	H1.0_F	F	CCTGCGGCCAAGCCCAAGCG	
			H1.0_R	R	AACTTGATCTGCGAGTCAGC	
		H1.1	H1.1_F	F	CTCCTCTAAGGAGCGTGGTG	
			H1.1_R	R	GAGGACGCCTTCTTGTTGAG	
		H1.2	H1.2_F	F	GGCTGGGGGTACGCCT	
			H1.2_R	R	TTAGGTTTGGTTCCGCCC	
		H1.3	H1.3_F	F	CTGCTCCACTTGCTCCTACC	
			H1.3_R	R	GCAAGCGCTTTCTTAAGC	
		H1.4	H1.4_F	F	GTCGGGTTCCTTCAAACTCA	
			H1.4_R	R	CTTCTTCGCCTTCTTTGGG	
		H1.5	H1.5_F	F	CATTAAGCTGGGCCTCAAGA	
			H1.5_R	R	TCACTGCCTTTTTCGCCCC	
		H1X	H1X_S	F	CCCAACGATGTAGCGTTTTT	
			H1X_AS	R	AAGGCCGAGAGCCAATAGA	
	ISGs	**IFi27**	IFi27_F	F	TGCTCTCACCTCATCAGCAGT	
			IFi27_R	R	CACAACTCCTCCAATCACAACT	
		OASL	OASL_F	F	GGGACAGAGATGGCACTGAT	
			OASL_R	R	AAATGCTCCTGCCTCAGAAA	
		IFIT2	IFIT2_F	F	ACGGTATGCTTGGAACGATTG	
			IFIT2_R	R	AACCCAGAGTGTGGCTGATG	
		IFIT3	IFIT3_F	F	CGGAACAGCAGAGACACAGA	
			IFIT3_R	R	ATGGCATTTCAGCTGTGGA	
		DDX60	DDX60_F	F	AAGGTGTTCCTTGATGATCTCC	
			DDX60_R	R	TGACAATGGGAGTTGATATTCC	
		IFi6	IFi6_F	F	CTGTGCCCATCTATCAGCAG	
			IFi6_R	R	GGGCTCCGTCACTAGACCTT	
	Others	**GAPDH**	GAPDH_F	F	GAGTCAACGGATTTTGGTCGT	
			GAPDH_R	R	TTGATTTTGGAGGGATCTCG	
		ZNF91	ZNF91 F	F	CCAGACCTGATTACTTATCTGG	ZNF91 gene, exon3-exon4
			ZNF91 R	R	ACATTTTTCATATTTTCTCAGTAATAC	
"H1 valley"		**CDK2 Distal**	3kb cdk2 prom fw	F	CAGCGAGGAAAGTCACATCA	-3kb from CDK2 TSS
			3kb cdk2 prom rv	R	TGGGGTGAGGGTAGTTTCTG	
		CDK2 TSS	cdk2 prom fw	F	GCGGCAACATTGTTTCAAGT	
			cdk2 prom rv	R	GTCGGGATGGAACGCAGTAT	
		NANOG Distal	nanog prom dist F	F	GACAGGGTTTCACCATGTTGGT	-2kb from NANOG TSS
			nanog prom dist R	R	CCGAGCCAGGTGCATCAT	
		NANOG TSS	nanog prom prox F	F	CGGTTTTCTAGTTCCCCACCTA	
			nanog prom prox R	R	CCAAGGCCATTGTAATGCAA	
Repetitive elements	LTR	**HERVK**	HERVK F	F	AGAGGAAGGAATGCCTCTTGCAG	
			HERVK R	R	TTACAAAGCAGTATTGCTGCCCGC	
		HERVE-int	envE F	F	ACTGGCCTTTTCCTAGGTGATAC	envE(RepeatMasker family HERVE-int ERV1). Amplifies several copies of HERVE-int
			envE R	R	TACTATTAATGGCTGCACAAGCA	
		LTR12C	LTR12C F	F	GTCTCGCTGGCTCAGGAGTG	LTR12C (Hominidae LTR-ERV1). In Silico-PCR: 457 LTR12C + 6 LTR12E
			LTR12C R	R	TGAGCTGTAACACTCACCGC	
		LTR5_Hs	LTR5_Hs #1 F	F	AAAGGGTCTGTGCTGAGGAG	LTR5_Hs (human specific LTR-ERVK). In Silico-PCR: 376 LTR5_Hs + 2LTR5 + 2 LTR5B
			LTR5_Hs #1 R	R	AGACATTCCATTGCCCAGGG	
		LTR5_HS	LTR5_Hs #2 F	F	TCTCAAGTACCCAGGGACACA	LTR5_Hs human specific LTR-ERVK
			LTR5_Hs #2 R	R	GTCTTTCCCTTCCCACGAGG	
	LINE-L1	**L1HS_5end**	L1HS_5end #1 F	F	TCCATCTGAGGTACCGGGTT	Young LINEs-L1. In Silico-PCR: 275 L1PA2 + 236 L1HS + 19 L1PA3 + 1 L1PA5
			L1HS_5end #1 R	R	CGATTTTCCAGGTGCGTCC	
		L1HS_5end	L1HS_5end #3 F	F	GGCACACTGACACCTCACA	Young LINEs-L1. In Silico-PCR: 205 L1PA2 + 10 L1PA3 + 9 L1HS
			L1HS_5end #3 R	R	AGATGGGTTTTCGGTGTGGA	
		L1HS_3end	L1HS_3end #2 F	F	GGGCGAAGGACATGAACAGA	Young LINEs-L1. In Silico PCR: 3683 matches, all of them within L1PA lineage (1170 L1PA2 + 1702 L1PA3 + 488 L1HS + others)
			L1HS_3end #1 R	R	CCTCTCCAGCACCTGTTGTT	
		L1PA3	L1HS_3end #2 F	F	GGGCGAAGGACATGAACAGA	L1PA3 (Hominoidea). In Silico-PCR: 1573 matches, all of them within L1PA lineage (1360 L1PA3 + 137 L1PA2 + others)
			L1PA3_3end #1 R	R	GCCAGTGATGGTGAGCATTTT	
		L1PA4	L1PA4 F	F	CATTTGCGGTTCACCAATATC	L1PA4 (Hominoidea). In Silico-PCR: 131 matches, 112 in L1PA4 + rest in other L1PA repeats
			L1PA4 R	R	GCTAGAGGTCCACTCCAGAC	
	SINE-Alu	**AluYa5**	AluYa5 #1 F	F	AGGAGATCGAGACCATCCCG	AluYa5 (Hominidae). In Silico-PCR: 522 AluYa5 + 4 AluY + 2 AluYb8 + 1AluSc8
			AluYa5 #1 R	R	CCACTACGCCCGGCTAATTT	
		AluYb8	AluYb8 #2 F	F	CGAGGCGGGTGGATCATGAGGT	AluYb8 (Hominidae). In Silico-PCR: 1117 AluYb8 + 154 AluYb9 +4 AluY+ 2 AluYk4
			AluYb8 #2 R	R	TCTGTCGCCCAGGCCGGACT	
	SVAs	**SVA_A**	SVA_A #1 F	F	TCTCGCTCACTCAATGCTCA	SVA_A family (Hominoidea).In Silico-PCR:24 SVA_A + 1 SVA_B
			SVA_A #1 R	R	CTGGGAGGTGGAGGTTGTAG	
		SVA_A	SVA_A #2 F	F	CTCGCTCACTCAATGCTCAA	SVA_A family (Hominoidea). In Silico-PCR:32 SVA_A + 1 SVA_D +1 SVA_C+2SVA_B+ 1 SVA_F
			SVA_A #2 R	R	GAGGTGGAGGTTGTAGCGA	
		SVA_E	SVA_E #1 F	F	AATAGAAAGGCGGGAAGGGTG	SVA_E family (Human-specific). In Silico-PCR 97 SVA_E
			SVA_E #1 R	R	CTTCTATCCACACAGACCCGG	
		SVA_F	SVA_F #4 F	F	GACTGGTTTTGGTGGAGACG	SVA_F family (Human-specific). In Silico-PCR: 23 SVA_F
			SVA_F #4 R	R	GGAGGTGTAGGTTGTAGCGA	
		SVA_F	SVA_F #5 F	F	CCTGACTGGTTTTGGTGGAG	SVA_F retrotransposon family (Human-specific). In Silico-PCR: 118 SVA_F
			SVA_F #5 R	R	GCTGGGAGGTGTAGGTTGTA	
		SVA_F	SVA_F #8 F	F	TGTGGAATAGAAAGGCGGGA	SVA_F family (Human-specific). In silico-PCR: 168 SVA_F
			SVA_F #8 R	R	AGGATCCCAAGGCAGAGGA	
		SVAs	SVAs #1_F	F	GTGTACCCAACAGCTCATTG	Not family-specific. In Silico PCR: 680 SVA_D + 257 SVA_F + 1 SVA_C
			SVAs #1_R	R	CACGGCAACCATCCGATTTC	
		SVAs	SVAs #2_F	F	GCCTTGGGATCCTGTTGATC	Not family-specific. In Silico-PCR:683 SVA_D + 289 SVA_F + 107 SVA_E + 10 SVA_C + 6 SVA_B + 1 SVA_A
			SVAs #2_R	R	CTTAACGAGCATGCTGCCTTC	
	Satellites	**SST1**	SST1 F	F	AACCACTGTGACGGGAGAAA	
			SST1 R	R	CTGGGACAGGACGAGACAC	
		SATa	SATa F	F	AAGGTCAATGGCAGAAAAGAA	
			SATa R	R	CAACGAAGGCCACAAGATGTC	

Table M. 2. Primary antibodies. Table includes primary antibodies used for chromatin immunoprecipitation (ChIP), Western-Blot (WB) and/or immunofluorescence (IF).

TARGET	HOST	COMPANY AND REFERENCE	APPLICATION
H1.0	Mouse	Millipore (05-629I clone-3H9I)	ChIP, WB, IF
H1.0	Rabbit	abcam (ab11079)	WB
H1.2	Rabbit	abcam (ab4086)	ChIP, WB, IF
H1.3	Rabbit	abcam (ab203948 clone EPR12683)	ChIP, WB, IF
H1.3	Rabbit	abcam (ab24174)	WB
H1.4	Rabbit	Thermofisher/Invitrogen (702876)	ChIP, WB, IF
H1.5	Rabbit	Thermofisher/Invitrogen (711912)	ChIP, WB, IF
H1.5	Rabbit	abcam (ab24175)	WB
H1X	Rabbit	abcam (ab31972)	ChIP, WB, IF
H1.2-phosphoT165	Rabbit	Millipore (06-1370)	WB, IF
H1.4-phosphoT146	Rabbit	abcam (ab3596)	WB, IF
H1.3/H1.4-phosphoT17	Rabbit	abcam (ab250946 clone EPR18087)	WB, IF
H3K4me1	Rabbit	abcam (ab8895-100)	ChIP
H3K4me3	Rabbit	abcam (ab8580)	ChIP, WB
H3K9me2	Mouse	abcam (ab1220; monoclonal)	IF
H3K9me3	Rabbit	abcam (ab8898)	ChIP, WB, IF
H3K27me3	Rabbit	Millipore (07-449)	ChIP, WB, IF
H3K27ac	Rabbit	abcam (ab4729)	ChIP
H3	Rabbit	abcam (ab1791)	WB
H4	Rabbit	abcam (ab10158)	WB
Lamin A	Mouse	abcam (ab8980 clone 133A2)	IF
Nucleophosmin (NPM1)	Mouse	abcam (ab10530 FC82291)	IF

1.3. Cell lines and culturing conditions

As a principal cell model for H1 studies we used breast cancer T47D-MTVL cell line (or derivative H1 *knock-downs*), which carry one stably integrated copy of luciferase reported gene driven by the MMTV promoter (3.17 clone). However, we also studied H1 variants distribution in a range of tumoral and non-tumoral cell lines. **Table M.3** includes all the human cell lines used in this thesis and their culturing requirements. All cell lines were grown at 37°C with 5% CO_2.

BRAF mutation status of melanoma cell lines is indicated in **Figure R.59**. Overall, BRAF mutation is a key determinant of the tumor response to treatment. BRAF mutated melanoma cases present high response rates to new targeted therapies (combinations of BRAF and MEK inhibitors).

Table M. 3. Human cell lines used. Human cancer and non-tumoral cell lines used in this thesis, their cell or cancer type as their culturing conditions are indicated. Absence of H1 variants at protein level are also listed. Abbreviations: FBS (Fetal bovine serum), P/S (penicillin/Streptomycin), Gln (L-Glutamine), N.E.A.A.S (non-essential aminoacids). Medium supplements are indicated in percentage, but equivalent final concentrations are: 10mM HEPES + 2mM L-Glutamine, 100 U/ml penicillin, and 100 µg/ml streptomycin.

Cell line	Cell/Cancer type	Growing conditions	Notes
T47D (T47D-MTVL)	**breast cancer**	RPMI 1640 medium containing 10%FBS, 1% P/S and 1%Gln	
MCF-7		MEM medium containing 10% FBS, 1% P/S, 1% N.E.A.A.S, 1% sodium pyruvate and 1% Gln	
MDA-MB-231		DMEM/F-12 medium containing 10% FBS, 1% P/S and 1% Gln	NO H1.3+H1.5
SK-MEL-147	**melanoma**	DMEM-GlutaMax medium containing 10% FBS and 1% P/S	
SK-MEL-173			NO H1.3+H1.5
UACC-257			
IGR-39		DMEM-GlutaMax medium containing 10% FBS, 1% P/S and 1% HEPES	NO H1.3+H1.5
SK-MEL-28			
WM266-4			
HCT-116	**colorectal carcinoma**	DMEM-GlutaMax medium containing 10% FBS, 1% P/S and 1% HEPES	NO H1.3+H1.5
HT-29			NO H1.3+H1.5
CaCo-2			
SK-N-SH	**neuroblastoma**		
HeLa	**cervix carcinoma**		NO H1.3+H1.0
HepG2	**hepatocellular carcinoma**		NO H1.3
HEK-293T	**embryonic kidney**		NO H1.5
NTERA-2 (NT2-D1)	**testicular embryonal carcinoma**		
IMR-90	**normal lung fibroblasts**		

1.4. Histone H1 inducible knock-down

To study consequences of single or multiple H1 depletion we used a Doxycycline (Dox)-inducible knock-down (KD) system. The T47D H1 KDs cell lines used in this thesis (**Table M.4**) were already available in the lab and validated in previous works (126,150,151). Although not included in the manuscript, we have used the same strategy to construct H1 KDs in other cell lines. Details of KD construction are briefly explained below.

Table M. 4. Derivative T47D KD cell lines used. Table includes the single or multiple H1 KDs and the control KD (i.e., Randomsh) used in this thesis. Target gene, shRNA lab reference (which indicates sequence position in the corresponding H1 variant mRNA and in the case of 225sh it refers to position in H1.4 mRNA) and target sequence are listed. All H1 KDs were used in previous lab publications (Reference column).

T47D KD	Target gene	shRNA lab ref	Target sequence	Reference
H1.0sh	H1.0	190sh	CGCTGACTCGCAGATCAAG	Sancho et al, 2008
H1.2sh	H1.2	156sh	AGAGCGTAGCGGAGTTTCT	Sancho et al, 2008
H1.3sh	H1.3	555sh	CTGCCAAGAGTCCAGCTAA	Sancho et al, 2008
H1.4sh	H1.4	120sh	GTCCGAGCTCATTACTAAA	Izquierdo-Bouldstridge et al, 2017
H1.5sh	H1.5	75sh	GGCAACTAAGAAGGCTGCC	Sancho et al, 2008
H1Xsh	H1X	87sh	CAACGGTTCCTTCAAGCTCAA	Mayor et al, 2015
multiH1sh	multiple-H1s	225sh	GAACAACAGCCGCATCAAG	Izquierdo-Bouldstridge et al, 2017
Randomsh	None	-	ACGTAGGCTAAGAGAAGCA	Sancho et al, 2008

To produce viral particles containing the HIV-derived vectors, 1.5x10^6 HEK 293T cells were seeded in 100mm plates (≈20% confluency) to be transfected with plasmids ptTR-KRAB-Red or pLVTHM-shH1.n (10 µg), pCMV-R8.91 (15 µg) and pVSVG (5 µg) using calcium phosphate. Medium containing viral particles was collected every 24 hours for 2 days. Medium was centrifuged (1200rpm, room-temperature (RT), 5min), filtered (0.45µm) and ultracentrifuged (26000rpm, 1h 30min, 4°C) in a sucrose gradient (by adding 0.22µm filtered 20% sucrose at the bottom of the ultracentrifugation tube prior centrifugation) to concentrate viral particles. Pellet containing the viruses was resuspended in culture medium and used for infection. Cells were infected by spinoculation, through centrifugation (1200rpm, 2h, RT).

First, a cell line expressing the Dox-responsive KRAB repressor (ptTR-KRAB-Red) was generated. After sorting of RedFP positive cells, this cell line was infected with viruses for expression of the corresponding shRNA (pLVTHM). After addition of Dox (3 days), sorting of double RFP and GFP positive cells was performed. Sorted cells were amplified in the absence of Dox to stablish the cell line.

1.5. Drug treatments

shRNA expression was induced with 6-days treatment of Doxycycline (Dox), in which cells were passaged on day 3. Dox (Sigma) was added at 2.5µg/ml.

5-aza-2´-deoxycytidine (aza) was used to evaluate transcriptional changes upon global DNA hypomethylation. Aza was added at 5 µM for 3 days, in which medium was replaced at day 2 by fresh aza-containing medium. Aza treatment was performed on different cell lines under the same conditions.

Actinomycin D (ActD) treatment leads to inhibition of RNA polymerase I (RNApol I), thus producing the inhibition of ribosomal DNA (rDNA) transcription. ActD used in immunofluorescence experiments, was added at 50 ng/ml for 24h.

To study the distribution of H1 variants in mitosis by immunofluorescence in T47D, we performed Thymidine-Nocodazole synchronization, in order to increase the percentage of mitotic cells in the sample. Notably, Thymidine inhibits DNA synthesis (thus synchronizes cells in G1/S entry phase) while nocodazole inhibits microtubule polymerization, arresting cells in G2-prometaphase. Specifically, cells were plated in 24 mm plates and Thymidine was added at 2mM. After 24h, the medium was PBS-washed and replaced by fresh medium during 3h. Nocodazole was then added at 100 ng/ml for 12h. After nocodazole treatment, medium was PBS-washed and replaced by fresh medium. Immunofluorescence protocol was started after 30min (sample enriched in prophase-metaphase) and after 60 min (enriched in anaphase-telophase) incubations.

To avoid cell cycle differences on the cell population that could potentially affect ChIP-Seq results, some of the H1X ChIP-Seq experiments in different cancer cell lines were performed on G1/S synchronized cells using double thymidine block. Thymidine was added at 2mM for 18h. Cells were then washed with PBS and were incubated with fresh medium for 9h. A second addition of thymidine was repeated under the same conditions and samples were processed for ChIP protocol.

However, after comparison of ChIP-Seq experiments performed on an asynchronized culture and G1/S synchronized cells, we concluded that cell cycle was not a significant source of variability between experiments. For ChIP-Seq experiments in T47D, cells were not synchronized. Nevertheless, under basal conditions, the majority (>70%) of T47D cells are in G1 phase (150).

1.6. RNA extraction and reverse transcriptase quantitative PCR (RT-qPCR)

Total RNA was extracted using the High Pure RNA Isolation Kit (Roche). RNA was eluted in RNase-free water in a final volume of 25µL. RNA was quality-checked (A260/A280≈2; A260/230 >2) and quantified using a Nanodrop or Biodrop. Then, cDNA was generated from 100 ng of RNA using the Superscript First Strand Synthesis System (Invitrogen). Notably, the system uses random hexamers. The reactions were prepared as follows: 4µl 5x VILO Reaction Mix + 2µl 10x Superscript Enzyme Mix + 100 ng of RNA + Nuclease-free water up to 20 µl. Samples were incubated as follows: 25ºC for 10min, 42ºC for 60min, 85ºC for 5min. Resulted cDNA was diluted 1:10 to be used on RT-qPCR.

RT-qPCR was performed using a SYBR Green Master Mix (Invitrogen) and specific oligonucleotides **(Table M.1)** in a QuantStudio 5 machine (Applied Biosystems). Data was analyzed using $2^{-\Delta\Delta Ct}$ method and normalized by GAPDH. If GAPDH resulted an inappropriate housekeeping, other commonly used housekeeping genes (GUSB, TBP, etc.) were used.

In experiments with different treatment conditions, expression was expressed relative to Untreated condition. To determine the H1 variants expression contribution to total mRNA H1, data was normalized by GAPDH and by the corresponding cell line genomic DNA amplification of each primer pair.

1.6.1. Library preparation of Total RNA samples

Concretely, total RNA-Seq of T47D H1Xsh-/+Dox and T47D H1.2sh-/+Dox was performed. Library preparation (library type: DNBSEQ LncRNA-Seq) and sequencing were performed by BGI Genomics (Hong-Kong). For additional quality checking of RNA samples, RIN value and 28S/18S ratio were measured to evaluate RNA integrity.

According to the manufacturer's instructions, the first step involves the removal of ribosomal RNA (rRNA) using target-specific oligos and RNase H reagents to deplete both cytoplasmic (5S rRNA, 5.8S rRNA, 18S rRNA and 28S rRNA) and mitochondrial ribosomal RNA (12S rRNA and 16S rRNA) from total RNA preparations. Following SPRI beads purification, the RNA is fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments are copied into first strand cDNA using reverse transcriptase and random primers, followed by second strand cDNA synthesis using DNA Polymerase I and RNase H. This process removes the RNA template and synthesizes a replacement strand, incorporating dUTP in place of dTTP to generate ds cDNA. These cDNA fragments then have the addition of a single 'A' base and subsequent ligation of the adapter. After UDG treatment, the incorporation of dUTP quenches the second strand during amplification. The products are enriched with PCR to create the final cDNA library. The libraries were assessed quality and quantity in two methods: check the distribution of the fragments size using the Agilent 2100 bioanalyzer and quantify the library using real-time quantitative PCR (TaqMan Probe).

1.7. Genomic DNA extraction

Genomic DNA extraction was performed by Phenol/chloroform purification. Concretely, we extracted DNA from fragmented chromatin samples, in order to quantify them and check sonication. Samples were incubated with Proteinase K overnight at 65ºC. Phenol/chloroform was added and samples was centrifuged (Vmax, 5min, RT). Sodium acetate and ethanol 100% were added to the supernatant (upper phase, carefully avoiding pipetting of the protein interphase) and mixed sample was incubated >30min at -80ºC. After precipitation, sample was centrifuged (Vmax, 40min, 4ºC) and ethanol 70% was added to the pellet. After a final centrifugation (Vmax, 10min, 4ºC), dry pellet was dissolved in nuclease-free water. DNA was quantified using a Nanodrop or Biodrop.

1.8. Histones extraction

For isolation of total histones, cell pellets were resuspended in 1 ml of hypotonic solution [10 mM Tris–HCl (pH 8.0), 1 mM KCl, 1.5 mM MgCl2, 1 mM PMSF, 1 mM DTT] and incubated on ice for 30 min. The nuclei were pelleted at 10000 × g for 10 min at 4ºC. Sulfuric acid (0.2 M) was added to the pellet to extract the histones on ice for 30 min. The solution was centrifuged at 16000 × g for 10 min at 4ºC. Trichloroacetic acid (TCA) was added to the supernatant in order to precipitate histones. After >1h ice-incubation precipitate was centrifuged (16000xg 10 min at 4ºC). Precipitate was washed with acetone and finally resuspended in water. Protein concentration was determined by Micro BCA protein assay (Thermo Scientific), which is based on the use of a standard curve of albumin and a colorimetric assay.

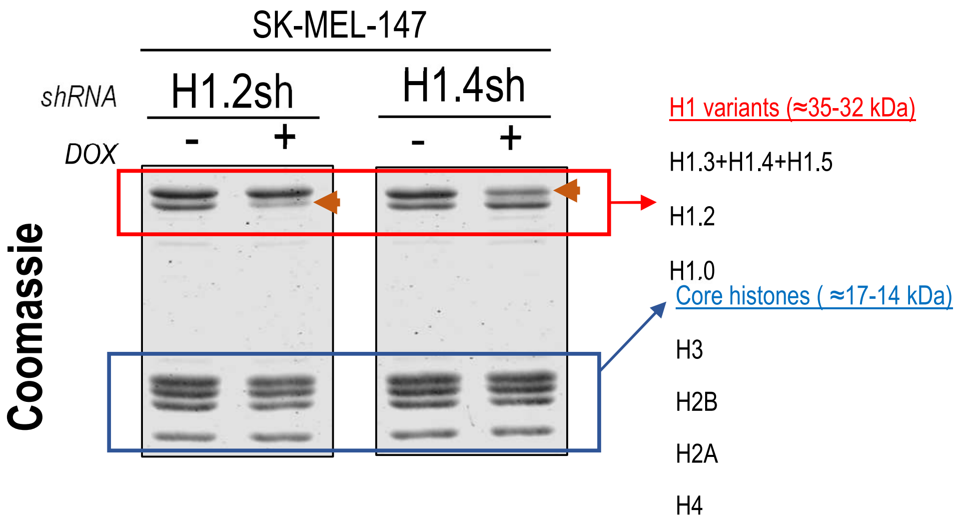


Figure M. 1. Coomassie staining of histone extracts. Coomassie staining of histones is added as a loading control in multiple immunoblot Figures in the Results section. Figure shows as an example the Coomassie staining (14% acrylamide) of SK-MEL-147 H1.2sh and H1.4sh -/+Dox cells. H1 variants appeared as three bands (red): the upper band corresponds to H1.3+H1.4+H1.5, the middle band corresponds to H1.2 and the lower and less-abundant band corresponds to H1.0. Core histones (blue) present lower molecular weight than H1 proteins. H3, H2B and H2A appear in three consecutive bands while the lower band corresponds to H4. In cell lines expressing all H1 variants, the upper H1 band (H1.3+H1.4+H1.5) is more prominent than H1.2 one, but in cell lines lacking H1.3+H1.5 (in which the upper band thus corresponds only to H1.4), both H1.4 and H1.2 bands are stained equally (see Figures R.58-R.59). Orange arrows indicate the corresponding H1 depletion in each single KD.

1.9. Immunoblot

Total histones or chromatin samples (usually 1-10µg) were mixed with LB Buffer (containing β-mercaptoethanol) and denaturalized 5min at 95ºC.

were exposed to SDS-PAGE (14%), transferred to a PVDF membrane, blocked with Odyssey blocking buffer (LI-COR Biosciences) for 1 h at RT. Alternatively, blocking was performed with 5% non-fat milk diluted in PBS-T (PBS- Tween 0.1%). Primary antibody incubation was done at 4ºC overnight. After three PBS-T washes (10min RT), membrane was incubated with secondary antibodies conjugated to fluorescence (IRDye 680 goat anti-rabbit IgG or IRDye 800 goat anti-mouse, Li-Cor) for 1 h at RT (diluted 1:5000 in PBS-T). Three final PBS-T washes (10min RT) were performed. Bands were visualized in an Odyssey Infrared Imaging System (Li-Cor). Coomassie staining, histone H3 and/or histone H4 immunoblotting were used as loading controls. ImageJ software was used for immunoblot quantification and Coomassie profiling. **Figure M.1** shows the characteristic Coomassie staining profile of a histones extract.

1.10. H1 Immunoprecipitation (IP) from total histones

Total histones from T47D H1.4sh Untreated or 6-days Dox-treated cells were extracted as previously described. For IP reaction, 60 µg of histones per condition were incubated overnight at 4ºC in RIPA buffer (150mM NaCl, 50mM Tris pH 7.5, 1%NP-40, 0.5% Sodium Deoxycholate, 0.1%SDS, 1mM EDTA) with protease inhibitors, 5µL of anti-H1.4 ChIP-grade antibody (Invitrogen ref. 702876) and 20µL of Protein A magnetic beads. Unbound histones were washed with RIPA buffer. After 3 washing steps, immunocomplexes were directly eluted from magnetic beads using Loading Buffer (5min 95ºC). Immunoblot analysis were performed using 5 µg of input samples (histones) and 1/6 of total eluted IPed material.

1.11. Chromatin immunoprecipitation (ChIP)

Cultured cells (2-3 150 mm plates) were crosslinked with 1% methanol-free formaldehyde and *crosslinking* reaction was stopped with 0.1M Glycine and incubated 5min RT. Cells were washed twice with pre-chilled PBS, scrapped in PBS-containing protease inhibitors and centrifuged (4000 rpm, 5min, 4ºC). Cell pellets were then resuspended in *Lysis Buffer I*, incubated 10min on ice and pelleted as before (4000 rpm, 5min, 4ºC). Pellet was resuspended in *Lysis Buffer II* and maintained on ice. Chromatin was sonicated using a Diagenode Bioruptor to produce fragments of an average size of ≈200bp. Chromatin can be maintained on ice up to one-week prior immunoprecipitation.

An aliquot of the chromatin sample was used to perform two independents phenol-chloroform DNA extractions to quantify DNA. 1 µg of DNA was loaded in a 1.2% agarose gel to check sonication.

To perform ChIP, 30 µg of fragmented chromatin were diluted in *IP buffer* and immunoprecipitated with the indicated antibody of interest (or rabbit IgG as a non-specific control) overnight at 4ºC in rotation. 20µl of Protein A magnetic beads (Millipore) were added to the incubation to later recover immunocomplexes. Notably, ≈2-3µg of antibody were used, depending on the antibody of interest. However, we have detected interbatch efficiency differences for some antibodies and amount of antibody was adjusted for each batch if necessary. We also noticed that an appropriate chromatin/antibody ratio is crucial to properly detect H1 depletion in H1 KDs by ChIP experiments.

After immunoprecipitation, beads bound to immunocomplexes (DNA-protein-antibody) were sequentially washed (*Wash Buffer I*, *Wash Buffer II*, *Wash Buffer III* and 2 washes of TE 1x; each wash 5min 4ºC on rotation). Reverse-crosslinking was performed using IPure Kit v2 (Diagenode). Input was prepared with 10% of the chromatin material used for ChIP. Final DNA was eluted in a total volume of 50 µl (two elution rounds of 25µl each one). Composition of *ChIP buffers* is indicated in **Table M. 5**.

Table M. 5. ChIP buffers composition. Asterisks (*) indicate that protease inhibitors were added prior use.

ChIP solution	Composition
Crosslinking solution	50mM HEPES pH 8; 0.1M NaCl; 1mM EDTA pH 8; 0.5 mM EGTA pH 8
Lysis buffer I (*)	5mM PIPES pH 8; 85mM KCl; 0.5% NP-40
Lysis buffer II (*)	1% SDS; 10mM EDTA pH 8; 50mM Tris-HCl pH 8.1
IP buffer (*)	0.01% SDS; 1.1% Triton X-100;1.2mM EDTA pH 8; 16.7 mM Tris-HCl pH 8.1; 167 mM NaCl
Wash buffer I	0.1%SDS; 1% Triton X-100; 2mM EDTA pH 8; 20 mM Tris-HCl pH 8.1; 150 mM NaCl
Wash buffer II	0.1%SDS; 1% Triton X-100; 2mM EDTA pH 8; 20 mM Tris-HCl pH 8.1; 500 mM NaCl
Wash buffer III	0.25 LiCl;1% NP-40;1% Sodium Deoxicholate;1mM EDTA pH 8; 10 mM Tris-HCl pH 8.1
Protease inhibitors	1mM Phenylmehtylsulfonyfluoride (PMSF); 1 μg/ml Aprotinin; 1 μg/ml Pepstatin A; 1mM Sodium ortovanadate; 20mM β-Glycerophosphate; 1x Protease Inhibitors Cocktail (Roche)

1.11.1. ChIP-qPCR

H1 binding to concrete genome regions was evaluated by ChIP-qPCR. qPCR was performed under the same conditions described above for RT-qPCR. Data was corrected by Input amplification and compared to IgG negative control. For each H1 ChIP sample, *NANOG* (inactive gene) and *CDK2* (active gene) upstream (3kb from TSS) and TSS regions were checked. H1 has been shown to be depleted at promoters of active genes, so by qPCR we can detect the H1 depletion at TSS of expressed genes compared to upstream regions, forming the so-called ‘H1 valley’ (152). On the contrary, this does not occur in inactive genes, whereby H1 variants are similarly abundant in both TSS and upstream regions.

1.11.2. Library preparation of ChIP samples

Quality checking and library construction of ChIP samples was performed by BGI Genomics (Hong-Kong). Libraries were constructed with 10ng of DNA. Qualified ChIP and Input samples were subjected to end-repair and then 3’ adenylated. Adaptors were ligated to the ends of these 3’ adenylated fragments. Fragments were PCR-amplified and PCR products were purified and selected with the Agencourt AMPure XP-Medium kit. The double stranded PCR products were heat denatured and circularized by the splint oligo sequence. The single strand circle DNA (ssCir DNA) were formatted as the final library and then quality-checked. The library was amplified to make DNA nanoball (DNB) which had more than 300 copies of one molecular.

1.12. Immunofluorescence

For immunofluorescence, cells were directly grown on glass coverslips (0.17mm thickness, 1.5H high-performance; Marienfeld Superior) placed in 24-well plates.

Cells were fixed with 4% paraformaldehyde (20 min; RT), permeabilized with Methanol (10 min RT) and blocked with 5% bovine serum albumin (5% BSA diluted in PBS-Triton 0.1%). Primary antibodies of interest were incubated overnight at 4ºC in a humid chamber (diluted in 5% BSA-PBS-T at optimized concentrations). After a PBS-Triton 0.1% (PBS-T) wash, secondary antibodies conjugated to Alexa fluorophores were incubated 1h RT in the dark (diluted 1:1000 in 5% BSA diluted in PBS-Triton 0.1%). The following conjugated secondary antibodies (Invitrogen) were used: goat anti-rabbit IgG H+L (Alexa 488 or 647); donkey anti-mouse IgG H+L (Alexa 555, 561, 633 or 647). After incubation, samples were washed with PBS-T (x3) and nuclei were stained with Hoechst (25 µg/ml diluted in 5%BSA-PBS-T; 1h RT in the dark). 5 PBS-T washes and a final MiliQ water were performed. Coverslips were mounted using Pro-long glass (Invitrogen). Preparations were maintained 24-48h in the dark at RT and then storage at 4ºC up to image acquisition.

1.12.1. Considerations about immunofluorescence protocol: Optimization and limitations

Super-resolution radial fluctuations (SRRF) was used as super-resolution imaging approach. Importantly, SRRF is an analytical approach. So, independently of upstream image acquisition procedure (confocal or SRRF) the experimental sample preparation workflow followed was the one described above. However, we needed to optimize the protocol to improve quality image and enable SRRF imaging. The protocol already described corresponds to the final and already optimized one, but some key points of the optimization process will be highlighted here.

First, accurate 0.17 mm coverslip thickness is critical for super-resolution imaging. Variations of thickness result in image defects, that is why we used high-precision coverslips. Second, Pro-long glass (whose refractive index is 1.52) considerably improve the inter-channel mismatch observed when using Mowiol (refractive index 1.49). This occurs because immersion oil has a refractive index of 1.51-1.52, which is more similar to Pro-long glass one. Moreover, we ensured Pro-long glass curation (24-48h RT incubation in the dark) prior imaging. This incubation allows Pro-long glass to reach its refractive index (1.52) and improve its antifade effectiveness. Third, to optimize super-resolution imaging of DNA and better visualize DNA structure and also DNA co-localization with H1 variants, we test different conditions to perform DNA staining. Notably, Hoechst gave much better results compared to DAPI. We tested several Hoechst treatments and, in general, a higher Hoechst concentration gave more signal but also increased noise signal. In addition, Hoechst diluent also affected to visualization. We determined that 25 µg/ml diluted in 5%BSA-PBS-T; incubated 1h RT give the best signal-to-noise ratio (**Figure M.2**). Besides, co-incubation of Hoechst and secondary antibodies interfered, so Hoechst incubation was performed sequentially, after secondary antibodies incubation.

Trying to improve the resolution from the experimental perspective, we tested Nanobodies as secondary antibodies, which have a spatially reduced structure compared to standard conjugated antibodies. However, H1 variants intensity signal was reduced compared to the one obtained using conventional secondary antibodies and images resulted inadequate for SRRF imaging. Another possibility that substantially would reduce experimental size of the signal would be the use of conjugated primary antibodies. However, the H1 variants antibodies we use for ChIP are not commercially available in a fluorescent-conjugated format. Although conventional secondary antibodies are larger, their use (indirect immunofluorescence) also amplifies the signal, which is beneficial for SRRF imaging.

Noteworthy, the Dox-inducible H1 KDs express RFP under basal or Untreated conditions while upon Dox treatment the cells are besides GFP+ (see section 1.4. Histone H1 inducible knock-down). For that reason, green and red channels cannot be used in H1 KDs immunofluorescence. As blue channel is used for nuclear staining, we only can use far-red channel (Alexa 647 or 633) without any interference and no co-immunostaining experiments can be performed on our H1 KDs.

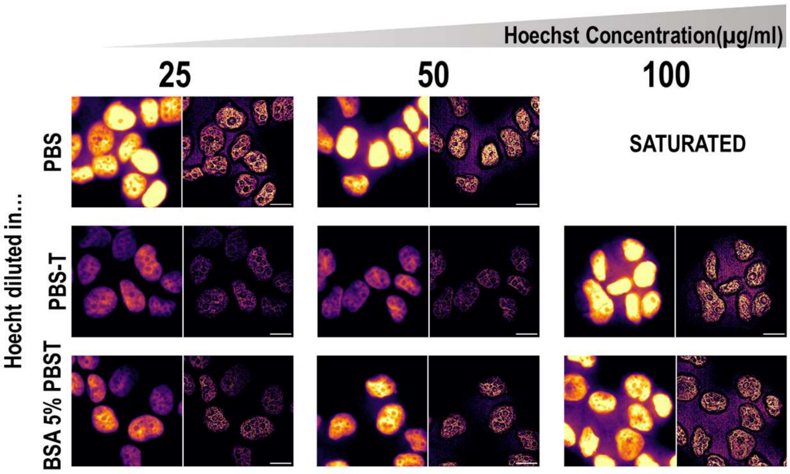


Figure M. 2. DNA staining optimization. Preliminary experiments revealed that Hoechst gave better SRRF results for DNA staining compared to DAPI. We then tested different Hoechst treatment conditions, as indicated. We concluded that Hoechst 25ug/ml diluted in BSA 5% PBS-T, 1h RT post-AB secondary incubation gave the optimal results. For each condition: Wide-field reference (left) and SRRF (right) images are shown.

1.12.2. Confocal image acquisition

All images were acquired in a *Dragonfly 505* multimodal spinning-disk confocal microscope (Andor Technologies, Inc), using a 100x/NA-1.4 oil immersion objective and Fusion analysis software. Camera magnification 2x and pinhole diameter 40 µm were used for confocal imaging. Laser excitation was done sequentially from shorter to longer wavelengths. Alexa-633 and Alexa-647 were excited by 647-nm laser, Alexa-555 and Alexa-561 were excited by 561-nm laser, Alexa-488 was excited by 488-nm laser and Hoechst was excited by 405-nm laser. Exposition time and laser intensity were adapted in each case, ensuring the absence of saturating pixels. 16-bit images were acquired. Confocal 3D images were taken as Z-stacks with 0.11 µm intervals.

Graphics processing units accelerated deconvolution (ClearView-GPU™) was applied in the acquisition protocol (16 iterations). Deconvolution is a mathematical correction to prevent out-of-focus light and optical distortions.

Unless indicated in the Figure legend, deconvolved images are shown and representative confocal images show a single focal plane. Biological conditions that are directly compared were experimentally performed in parallel, as well as acquired in parallel under the same conditions and acquisition settings.

1.12.3. Super-resolution radial fluctuations (SRRF) acquisition

1.12.3.1. SRRF algorithm

SRRF is an analytical super-resolution approach based on observing fluctuations in radial symmetry throughout the image frames (172,173). SRRF assumes the image is formed of point sources convolved with a point spread function (PSF) that displays a higher degree of local radial symmetry than the background (radiality). Thus, radiality refers to the local radial gradient convergence to the center of the fluorophore signal. SRRF algorithm magnifies each image pixel into sub-pixels, and each subpixel is assigned a non-binary value related to the probability that it contains a fluorophore, through radiality calculation. Thus, radiality acts as a proxy for the positions of molecules. Radiality is also tracked over time, to allow de-noising and enhance resolution. This temporal analysis allows to discern specific signal from noise, due to noise is not correlated over time but fluorophores are. Moreover, the highest degree of temporal correlation is located at the centre of the radiality peaks generated by fluorophore signal. **Figure M.3** depicts the principles of SRRF imaging.

1.12.3.2. SRRF acquisition

We implemented SRRF through the SRRF-Stream+ module (Andor) available in the microscope. SRRF was performed for 1) co-localization experiments (H1 variants or H1 variants-DNA) and 2) evaluate chromatin structure at the super-resolution level.

For SRRF co-localization experiments a single Z confocal plane 1024x1024 was imaged. When evaluating H1 variants co-localization with DNA, H1 variants were labelled with Alexa-488 (except for H1.0, which was labelled with Alexa 561) and Hoechst was used for DNA staining. For H1 variants co-localization with H1.0, H1 variants were labelled with Alexa-488 and H1.0 with Alexa-561. In dual color SRRF images, laser excitation was performed sequentially in the following order: 488-nm, 561-nm, 405-nm. Images frames were also acquired sequentially for each channel. The following parameters were used for all conditions: 1x ring radius, 6x radiality magnification (i.e. each pixel is magnified in an array of 6x6 sub-pixels), 500 frames (i.e. temporal analysis). Exposition time 250ms and 18% laser intensity were used for 405-nm channel (Hoechst imaging). Exposition time 180-200 ms and 8-12% laser intensity were used for 488-nm or 561-nm channels. Under these conditions, voxel size corresponds to 8.5 nm (in x,y).

For evaluating DNA structure upon H1 KDs, SRRF imaging of DNA (Hoechst) was used. SRRF was performed on 1024x1024 wide-field images with the following parameters: exposition time 150ms, 10% 405-nm laser intensity, 1x ring radius, 6x radiality magnification, 1000 frames. Under these conditions, voxel size corresponds to 8.5 nm (in x,y).

A higher number of images frames in general correlates with a higher resolution but also with a considerable increase of the acquisition time. However, we found that upon > 500 frames, more instable fluorophores burnt-out, which causes an increment of the noise and artifacts. Hoechst signal is more stable and after 1000 images frames, signal was still detectable, but we found no considerable improvement with >1000 frames.

Importantly, we used Tetraspeck beads (Thermo Fisher Scientific) to acquire confocal and SRRF images and check chromatic aberration. No significant inter-channel mismatch was found.

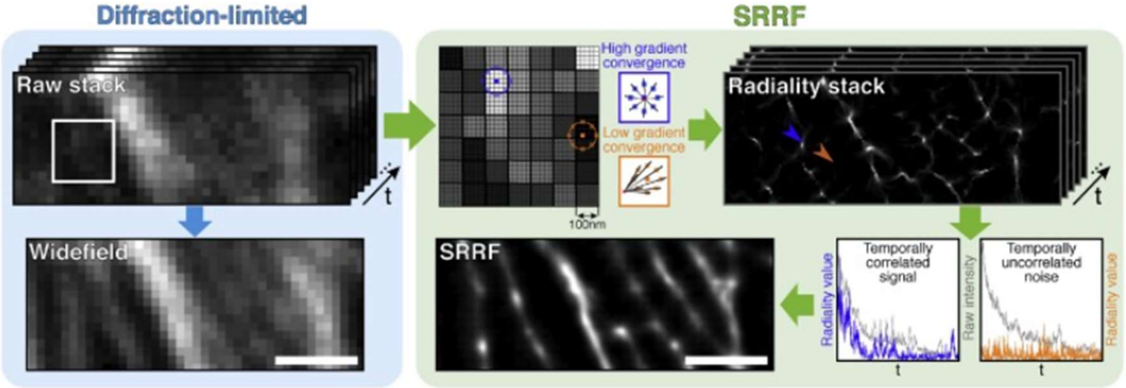


Figure M. 3. SRRF algorithm. Confocal or wide-field raw data is split into subpixels (in the example 5x5 array of subpixels). The blue (high radiality) and orange (low radiality) arrowheads in the radiality stack indicate the location of these subpixels. Temporal analysis along image frames is done. Radiality variation over time is measured to distinguish signal (blue) from noise (orange). Adapted from (173).

1.12.4. Image analysis

Image analysis was performed in ImageJ software. Data was post-processed and plotted in R or Excel.

In general, fluorescence intensity quantification was done by generating masks for each nucleus using Hoechst signal as reference and computing the mean intensity of the proteins of interest. Alternatively, Corrected Total Cell Fluorescence formula was used and calculated as: Integrated Density - Area of selected nucleus x Mean fluorescence of background readings.

Line signal intensity profile plots were created using Plot Profile tool.

Circularity was defined as 4π*area/perimeter^2. Circularity ranges from 0-1, where Circularity=1 corresponds to a perfect circle. It was calculated from Z stack maximum projections considering Hoechst signal.

Analysis of ring intensity distribution was done with a macro available at https://github.com/MolecularImagingPlatformIBMB with minor modifications. Every nucleus is partitioned into four concentric rings with equal areas that converge towards the center of the nucleus. Subsequently, the signal intensity density of the specific interest is assessed for each ring and adjusted to the total intensity density of the nucleus. This approach enables the comparison of the intensity distribution of a target protein among nuclei varying in shape and size. The analysis was done in n=30 nuclei/condition and represented in Violin Plots using *vioplot* R function. The four concentric sections were termed: A1-A4, being A1 the most peripheral and A4 the most central. A 'peripheral index' was calculated for H1.0 and H1.4 immunofluorescences in order to enable a direct comparison between different cell lines. Peripheral Index was defined as (Mean A1 intensity)/ (Mean A4 intensity) ratio.

Pearson´s correlation coefficient (r) was calculated using JaCoP Plugin. Calculation was done from a unique Z central plane in single nuclei and after channel thresholding. r was calculated per each nucleus and then distribution of values in a cell population were represented in R as violin plots (*vioplot* function) or density plots (*ggridges* and *ggplot2* packages).

Co-localization in SRRF images was calculated with an in-house macro. Briefly, after pre-processing steps, it creates a mask of each channel and calculates intersection between both masks of interest (in % of intersected pixels). This process was done in n=20 nuclei and distribution was represented in violin plots (*vioplot* R function).

To objectively compare chromatin structure through DNA super-resolution in different H1 KD conditions, we used DNA-free areas method. DNA-free areas quantification has been used elsewhere to assess DNA compaction (174,175). For the quantification, we used an in-house macro constructed by Elena Rebollo. Briefly, in each nucleus a random sampling of small 200x200pixel areas was manually done, covering all nuclear area and excluding nucleolus. Per each 200x200 bin, auto-local thresholding using the Phansalkar filtering algorithm was applied to Hoechst channel **(Figure M.4)**. So, based on this filtering each bin presents a given % of DNA-covered area and the remaining % of DNA-free areas. In Boxplots corresponding to individual cells **(Figure R.49C)**, distribution of %DNA-free areas in each bin were represented. In **Figure R.49B**, average values of %DNA-free areas were calculated per each nucleus and boxplots represent the distribution of n=20 average nuclear values.

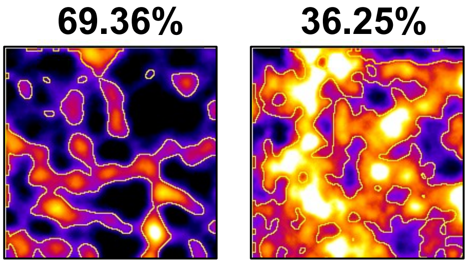


Figure M. 4. Percentage of DNA-free areas calculation macro. Each image shows a 200x200pixel bin from SRRF images of Hoechst staining in T47DmultiH1sh Untreated (left) and +Dox (right) nuclei, after applying auto-local thresholding as indicated in the text. %DNA-free areas is indicated.

Regarding statistics, Kolmogorov-Smirnov test was used to check Normality. If data followed a normal distribution, differences were assessed by paired-t-test. When comparing multiple groups, ANOVA multiple comparison test was performed to evaluate the mean differences between groups and subsequent Tukey multiple comparison test was performed to evaluate the differences between each condition pair.

2. Computational methods

2.1. RNA-Seq processing

The qualified libraries were sequenced pair end on the BGISEQ-500/ MGISEQ-2000 System (BGI).

RNA-Seq reads were mapped to the human reference genome (GRCh37/hg19) using HISAT2 v2.2.1 with default parameters and specifying strand-specific information *(--rna-strandness RF)*. SAMtools v1.11 was used to sort BAM files and filter for properly paired-end reads *(-f 2)*. Aligned reads were mapped to Ensembl GRCh37.87 gene annotation with TEtranscripts v2.1.4 *(--sortByPos --mode multi --stranded reverse)*. DESeq2 v1.26.0 was used to identify differentially expressed genes. Gene expression changes were considered significantly different if the absolute value of the log2(FC) was higher than 1.4 and the adjusted p-value was lower than 0.05.

2.2. ChIP-Seq processing

ChIP-Seq libraries were sequencing by BGI. The DNBs were loaded into the patterned nanoarray and single-end 50 or 100 bases reads were generated in the way of sequenced by combinatorial Probe-Anchor Synthesis (cPAS).

Single-end reads were quality-checked via FastQC (v0.11.9) and aligned to the human GRCh37/hg19 reference genome using Bowtie2 (v2.3.5.1) (176) with default options. SAMtools (v1.9) (177) utilities were used to filter out the low-quality reads with the flag 3844. Input and ChIP samples genome coverage was calculated and normalized by reads per million with BEDTools (v2.28.0) (178), and regions with zero coverage were also reported in the ChIP-Seq annotation *(genomecov -ibam -bga -scale)*. MACS2 (v2.1.2) (179) was used to subtract input coverage from ChIP values to generate continuous signal tracks *(bdgcmp -m subtract)*.

2.3. Operations on genomic intervals

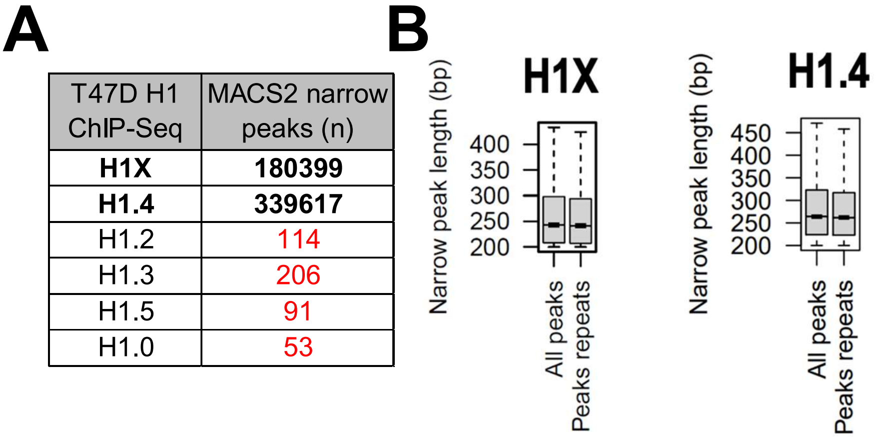
Evaluation of ChIP-seq abundance in genomic regions of interest was done with BEDTools (*bedtools map -o mean*), using as input files a BED file containing the coordinates of the segments and a bedgraph file with the Input-subtracted ChIP-Seq signal.

Intersection or overlap between two sets of genomic regions was calculated with BEDtools (*bedtools intersect*) using BED file coordinates as input files. Restriction parameters *(-f -F)* were applied when indicated.

Other BEDTools functions *(bedtools map -o collapse, bedtools map -o sum)* were also used to calculate percentage of overlapping base-pairs or correspondence between different genomic intervals datasets.

2.4. H1 variants peak calling

H1 variants narrow peak calling was performed with MACS2 *(callpeak --no model –ext size 200)*. Notably, narrow peaks can only be computed for H1.4 and H1X variants in T47D **(Figure M.5A)**. H1 narrow peaks length distribution is shown in **Figure M.5B**. **Table M.6** collects the number of H1 narrow peaks within repetitive elements families.



T47D H1 ChIP-Seq	MACS2 narrow peaks (n)
H1X	**180399**
H1.4	**339617**
H1.2	114
H1.3	206
H1.5	91
H1.0	53

Figure M. 5. H1 peak calling. A) Table contains the number of narrow peaks for H1 variants ChIP-Seq. In T47D, narrow peaks can only be efficiently called for H1X and H1.4. **B)** Boxplots show length distribution of H1X and H1.4 narrow peaks and of those narrow peaks that overlap with repeats.

2.5. ChIP-Seq average profile

ChIP-Seq average profiles around the center of gene transcription start site (TSS) were constructed with CEAS (Cis-regulatory Element Annotation System, v0.9.9.7) (180). Expressed genes were classified in 10 equal groups depending on their basal expression. Group 1 corresponds to the 10% genes with lower expression while group 10 contains the 10% of top-expressed genes. Additionally, group 0 contains non-expressed genes. Gene expression was extracted from RNA-Seq experiments performed in T47D cells.

ChIP-Seq average profiles around the center of other regions of interest (i.e. enhancers) were constructed using Sitepro script from CEAS. Plotting was done in R.

2.6. Heatmap construction

Heatmaps were performed by using the R package *pheatmap*. The euclidean distance measure and the complete cluster method were used in clustering rows and columns.

2.7. ChIP-Seq profile at meta-repeats

H1 variants ChIP-Seq profiles at meta-repeats were constructed using deepTools (*computematrix, plotHeatmap, plotProfile*) (181). All regions were scaled to the same size (*scale-regions* mode). bigwig containing H1 variants ChIP-Seq abundance and bed files containing the coordinates of the repetitive elements of interest were used as input files. Notably, in figures, adjacent plots do not present the same row order. Each plot is ordered by its corresponding ChIP-Seq profile.

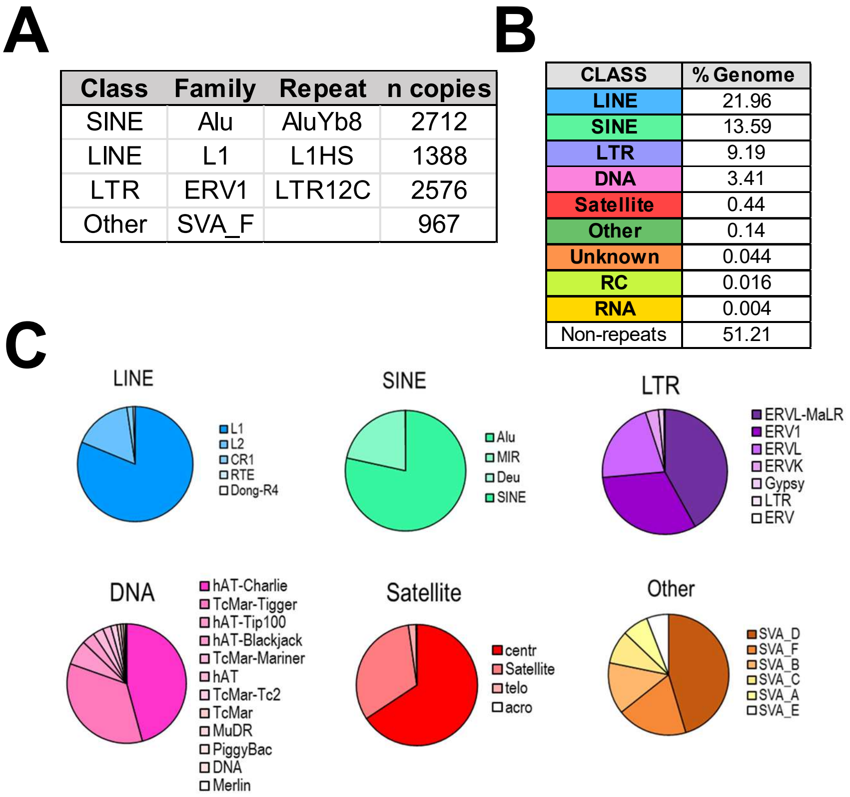
2.8. Genomic data visualization

Genome browser captures were done by loading the corresponding files (bed or tdf formats) to IGV (Integrated Genomics Viewer, Broad Institute).

2.9. Repetitive elements annotation and classification

2.9.1. Repetitive elements annotation

TEtranscripts (v2.1.4) repetitive elements annotation was used (182). In TEtranscripts, GTF files of transposable element annotation were generated from the RepeatMasker tables obtained from UCSC genome database. The TEtranscripts annotation tables were parsed to filter out low complexity and simple repeats, rRNA, scRNA, snRNA, srpRNA and tRNA (which are annotated in RepeatMasker). For that, annotation includes nine different classes of repeats, including: LINE, SINE, LTR, DNA, Satellite, Other, Unknown, RC (Rolling-Circle) and RNA. Repeats with unsure classification named with a '?' at the end of the family or class name (for example, SINE?) were excluded from the analysis.



A

Class	Family	Repeat	n copies
SINE	Alu	AluYb8	2712
LINE	L1	L1HS	1388
LTR	ERV1	LTR12C	2576
Other	SVA_F		967

B

CLASS	% Genome
LINE	21.96
SINE	13.59
LTR	9.19
DNA	3.41
Satellite	0.44
Other	0.14
Unknown	0.044
RC	0.016
RNA	0.004
Non-repeats	51.21

Figure M. 6. Repetitive elements annotation. A) Table containing some examples of repeats classification. Number of copies correspond to those located in chr1-22 and chrX and that not overlap with ENCODE BlackList (hg19). **B)** Table containing the genome occupancy (% length) of each class. **C)** Genome occupancy (total length in bp) of each family of the class is represented in a pie chart and expressed as percentage. Families are ordered based on their decreasing genome occupancy within their respective class.

Repetitive elements can be hierarchically classified at different levels, including Class, Family and Repeat (**Figure M.6**). Finally, each family or repeat present multiple insertions/copies in the genome.

Repeats overlapping problematic regions defined in ENCODE BlackList (183) were excluded from the analysis. Satellite was the class most affected by this removal, due to almost 50% of the satellite repeats overlap with the BlackList.

All analysis including in this thesis are done at class-family-repeat level, without considering concrete individual copies. For this reason, reads that align to multiple positions (multi-reads) are also used for repeats quantification. However, it must be considered that for analyzing multiple genome features (or correlation between different genomic datasets) within a single copy, multi-reads have to be discarded.

Table M. 6. H1X and H1.4 narrow peaks within repetitive element classes. Table includes the number of H1 narrow peaks per family, considering total number of peaks within the family or restricting one peak per repeat copy. Data is also expressed as percentage of peaks per family or percentage of repeats exhibiting at least one peak. In longer repeats, one single copy could harbor more than one H1 peak, so in order to properly calculate the percentage of repeats exhibiting narrow peaks, a maximum of one peak per repeat is considered (so, repeats could have one or more H1 narrow peaks). Number of repeat copies include those located in chr1-22 and chrX and excluded from ENCODE BlackList.

REPEAT CLASSIFICATION			H1X narrow peaks				H1.4 narrow peaks			
Class	**Family**	**copies (n)**	**H1X narrow peaks (n)**	**H1X narrow peaks (1 per repeat)**	**% H1X narrow peaks**	**% family with H1X narrow peak**	**H1.4 narrow peaks (n)**	**H1.4 narrow peaks (1 per repeat)**	**% H1.4 narrow peaks**	**% family with H1.4 narrow peak**
DNA	DNA	2701								
	hAT	15177								
	hAT-Blackjack	19382								
	hAT-Charlie	247696	13	13	0.008	0.005	9	9	0.003	0.004
	hAT-Tip100	29864	2	2	0.001	0.007	1	1	0.000	0.003
	PiggyBac	2295								
	TcMar	5305								
	TcMar-Mariner	15995								
	TcMar-Tc2	7966								
	TcMar-Tigger	100712	8	8	0.005	0.008	10	10	0.003	0.010
	Merlin	54								
	MuDR	1928								
	12	**449075**								
LINE	CR1	59938	2	2	0.001	0.003				
	Dong-R4	551								
	L1	911361	26445	19135	16.376	2.100	43351	29134	13.667	3.197
	L2	456087	35	35	0.022	0.008	8	8	0.003	0.002
	RTE	17416	1	1	0.001	0.006				
	RTE-BovB	651								
	6	**1446004**								
LTR	ERV	553								
	ERV1	164965	1165	1114	0.721	0.675	525	514	0.166	0.312
	ERVK	9500	1	1	0.001	0.011				
	ERVL	156220	21	21	0.013	0.013	21	21	0.007	0.013
	ERVL-MaLR	335366	1499	1496	0.928	0.446	2487	2487	0.784	0.742
	Gypsy	18370	1	1	0.001	0.005				
	LTR	2158								
	7	**687132**								
Satellite	Satellite	2074	1	1	0.001	0.048			0.000	0.000
	acro	6			0.000	0.000				
	centr	1168	51	43	0.032	3.682	94	74	0.030	6.336
	telo	212	1	1	0.001	0.472				
	4	**3460**								
SINE	Alu	1144739	129513	129510	80.200	11.313	269121	269115	84.846	23.509
	Deu	1259	1	1	0.001	0.079				
	MIR	583523	26	26	0.016	0.004	6	6	0.002	0.001
	SINE	951	1	1	0.001	0.105	1	1	0.000	0.105
	4	**1730472**								
Other	SVA_A	247	184	146	0.114	59.109	130	108	0.041	43.725
	SVA_B	461	388	323	0.240	70.065	246	203	0.078	44.035
	SVA_C	277	275	227	0.170	81.949	162	141	0.051	50.903
	SVA_D	1341	1319	1064	0.817	79.344	701	575	0.221	42.878
	SVA_E	233	140	122	0.087	52.361	60	54	0.019	23.176
	SVA_F	967	394	346	0.244	35.781	254	237	0.080	24.509
	6	3526								
RC	**Helitron**	2193								
RNA	**RNA**	708								
			161487	**153640**	**100**		**317187**	**302698**	**100**	

2.9.2. Classification of repetitive elements according to evolution

Repetitive elements were classified from an evolutionary perspective, based on the taxonomic clades or species in which the repeat is known to be present. This information was extracted from Dfam database (171). Clades classification is done according to NCBI taxonomy database.

All repeats present both in Primates and Non-primates species were classified as 'Non-primate'. For repeats restricted to primate species, intermediate clades classification was maintained. Of note, according Dfam database, this classification refers to the oldest clade in which copies of the family/repeat have been found at orthologous positions, meaning that the transposable element was active before the first speciation of extant species in this clade.

B

Class	Family	Repeat	Clades	Notes
LINE	L1	**L1M2**	Non-primates	Present in all primates but also in non-primates species
SINE	Alu	**AluYc3**	Primates	Present in all primates species but not in non-primates
Other	**SVA_D**		Hominidae	Present in Hominidae species and descendants (H.sapiens) but not in older ancestors (not in Hominoidea, Catarrhini, etc.)
Other	**SVA_F**		Homo sapiens	Human-specific

Figure M. 7. Taxonomic classification of repetitive elements. A) Illustration of the intermediate clades of primates. Representative species are depicted. Clades in bold correspond to the ones considered in the analysis. Clades are ordered from older to most-recent (from Primates to humans). **B)** Table containing some repeat examples to clarify taxonomic classification criteria.

The intermediate primates clades present in Dfam database are the following ones (from older to younger): *Primates, Haplorrhini, Simiiformes, Catarrhini, Hominoidea, Hominidae, Homininae, Homo, Homo sapiens* (**Figure M.7A**). Clades with a very low number of repeats (*Haplorrhini, Homininae, Homo*) were re-classified with their more proximal ancestor group. In the case of horizontal transfer within primates and other clades, repeats were re-classified as *Primates* (but these cases were minority, n=6). That is, n=7 primates clades were considered (i.e. *Primates, Simiiformes, Catarrhini, Hominoidea, Hominidae, Homo sapiens*). Some examples of the classification are given in **Figure M.7B**. **Table M.7** shows the clades classification per repeat family. In **Table M.8**, *Hominoidea* repeats are listed.

Regarding plots where H1 variants abundance within different repeats clades where evaluated: in boxplots, mean Input-subtracted ChIP-Seq signal in each copy was plotted; in heatmaps, median abundance per repeat was calculated and then median abundance per family and clade was plotted on the final heatmap.

Table M. 7. Repetitive elements classified according to taxonomic clades. Table includes clade classification of repeats within each family. Number and percentage of repeats belonging to each clade (n=7) is indicated. Clade classification information is extracted from Dfam database (171). Unknown class (n=29 families) is not included in the table but all Unknown repeats are classified as 'Non-Primate'. Notably, TE-transcripts repeat annotation includes n=41 Alu families, but AluYf4, AluYc5 and AluYf5 families are not listed in Dfam database, so n=38 Alu families are included from evolutionary analysis.

REPEAT CLASSIFICATION			CLADES (n= 7)													
Class (n=9)	Family (n=72)	Repeat (n=963)	Non-Primates		Primates		Simiiformes		Catarrhini		Hominoidea		Hominidae		H.sapiens	
			n	%	n	%	n	%	n	%	n	%	n	%	n	%
DNA	DNA	12	11	91.7			1	8.33								
	hAT	9	9	100												
	hAT-Blackjack	8	8	100												
	hAT-Charlie	68	62	91.2	1	1.47	5	7.35								
	hAT-Tip100	24	24	100												
	PiggyBac	6	2	33.3			4	66.67								
	TcMar	6	6	100												
	TcMar-Mariner	5	2	40	1	20	2	40								
	TcMar-Tc2	8	8	100												
	TcMar-Tigger	52	24	46.2	28	53.85										
	Merlin	1		0	1	100										
	MuDR	5		0	5	100										
	12	**204**														
LINE	CR1	15	15	100												
	Dong-R4	1	1	100												
	L1	122	82	67.2	31	25.41	1	0.82	2	1.64	4	3.28	2	1.64		
	L2	4	4	100												
	RTE	3	3	100												
	RTE-BovB	1	1	100												
	6	**146**														
LTR	ERV	3	3	100												
	ERV1	260	48	18.5	94	36.15	67	25.77	36	13.85	8	3.08	7	2.69		
	ERVK	37	0	0			2	5.41	24	64.86	10	27.03			1	2.70
	ERVL	102	85	83.3	10	9.80	7	6.86								
	ERVL-MaLR	80	58	72.5	9	11.25	12	15	1	1.25						
	Gypsy	19	19	100												
	LTR	3	3	100												
	7	**504**														
Satellite	Satellite	15	3	20	12	80										
	acro	1		0	1	100										
	centr	6		0	6	100										
	telo	2		0	2	100										
	4	**24**														
SINE	Alu	38	1	2.6	27	71.05							10	26.32		
	Deu	2	2	100												
	MIR	4	4	100												
	SINE	1	1	100												
	4	**45**														
Other	SVA_A	1		0							1	100				
	SVA_B	1		0									1	100		
	SVA_C	1		0									1	100		
	SVA_D	1		0									1	100		
	SVA_E	1		0											1	100
	SVA_F	1		0											1	100
	6	**6**														
RC	Helitron	4	4	100												
RNA	RNA	1	1	100												

Table M. 8. Hominoidea repeats. Table shows the repeat name of n=48 repeats found in *Hominoidea* and descendant clades (from Hominoidea to H.sapiens). Repetitive element class and family are indicated.

Class	Family	TAXA CLASSIFICATION			
		Hominoidea	Hominidae	Homininae	H. sapiens
LINE	**L1**	L1P1, L1PA3, L1PA4, L1PA5	L1PA2	L1HS	
LTR	**ERV1**	LTR12C, LTR12D, LTR12E, LTR2B, LTR2C,LTR6B, LTR7, HERVE-int	HERV-Fc1_LTR1, HERV-Fc1_LTR2, HERV-Fc1_LTR3, HERV-Fc2_LTR, HERV-Fc2-int	LTR7Y, HERV-Fc1-int	
	ERVK	LTR13, LTR5, LTR5A, MER11B, MER11C, MER11D, MER9a1, MER9a2, HERVK11D-int, HERVK13-int			LTR5_Hs
SINE	**Alu**		AluYa5,AluYa8,AluYb8,AluYb9, AluYd8, AluYg6, AluYh9, AluYk11, AluYk12, AluYk4		
Other	**SVA_A**	SVA_A			
	SVA_B		SVA_B		
	SVA_C		SVA_C		
	SVA_D		SVA_D		
	SVA_E				
	SVA_F				SVA_E,SVA_F

2.10. Genome segmentations and databases

Genome-wide GC content, Giemsa bands (G-bands) coordinates at 850 bands per haploid sequence (bphs) resolution and chromosomes coordinates were obtained from the UCSC human genome database (184,185). G-bands were classified in eight groups as detailed in (116): G-positive (Gpos25 to Gpos100, according to its intensity in Giemsa staining), and G-negative (unstained), which were further divided into four groups according to their GC content (Gneg1 to Gneg4, from high to low-GC content).

HeLa-S3 genome segmentation by ChromHMM (ENCODE) was obtained from UCSC human genome database and 10 groups of chromatin states were evaluated.

T47D enhancers were extracted from EnhancerAtlas 2.0 database (186). Importantly, we confirmed that these regions exhibit enhancer-like features **(Figure M.8)**.

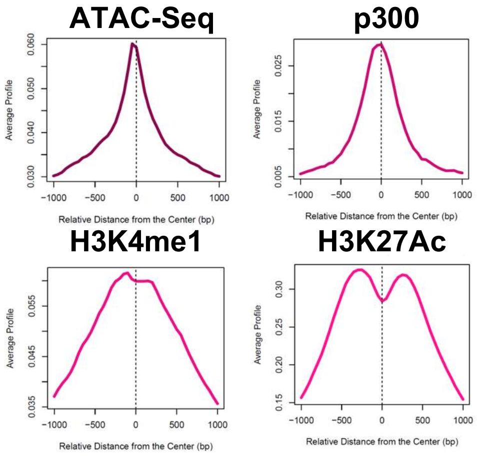


Figure M. 8. Validation of T47D Enhancer Regions from EnhancerAtlas. Input-subtracted ChIP-Seq or ATAC-Seq average profile of enhancer-associated chromatin features. All data used were originated in T47D cells. GEO accession numbers for ChIP-Seq data used here: p300 (GSM1669024, r2); H3K4me1 (SRR891282); H3K27Ac(SRR9886584).

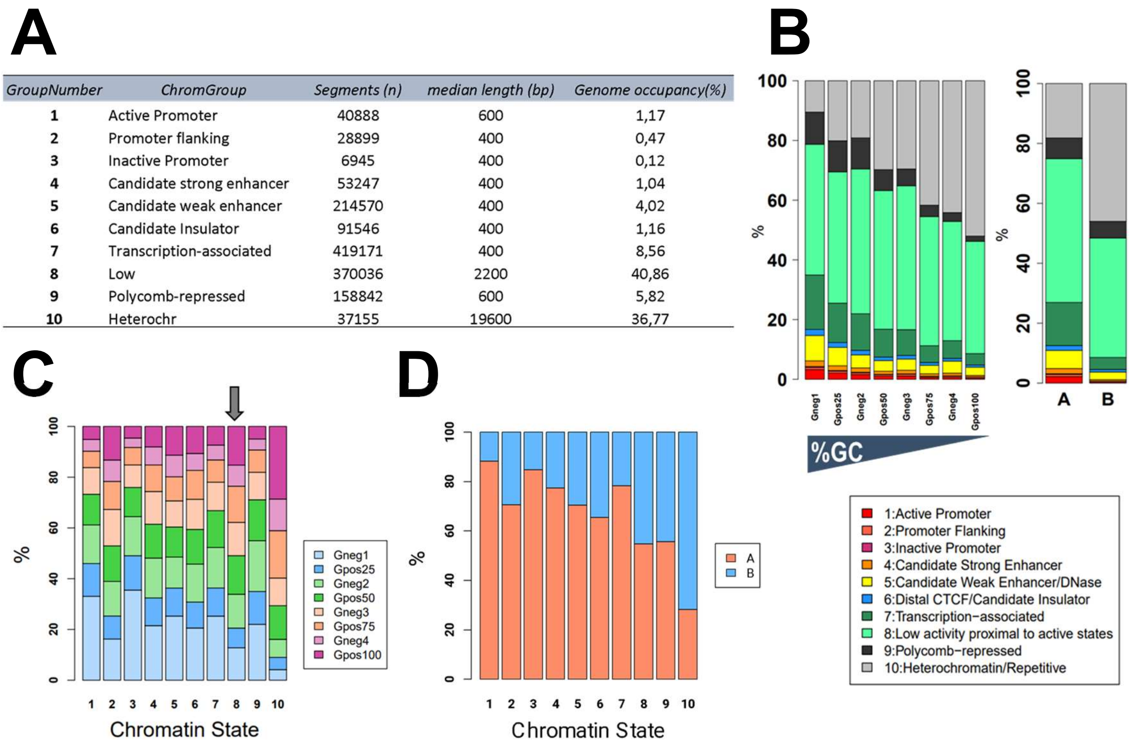
Lamina-associated-domains (LADs) were originally identified in Tig3 fibroblasts through LaminB1 Dam-ID (103). Coordinates were extracted from UCSC database.

Nucleolus-associated-domains (NADs) were originally identified by Nucleolar-HiC in HeLa cells (112).

A, B compartments and TAD coordinates were extracted from Hi-C experiments performed in T47D multiH1 KD cells. TADs were classified in four equal groups depending on their H1.2 and H1X relative ChIP-Seq abundance (H1.2/H1X ratio). In this thesis, the use of Hi-C-derived segmentations is virtually limited to evaluate H1 variants distribution in T47D. For that reason, coordinates under basal conditions (T47D multiH1sh Untreated) were directly used for analysis. For further details regarding Hi-C analysis, see (142) (Appendix I, publication list).

ATAC-Seq data in T47D multiH1 KD cells (GSE100762) was previously generated in the laboratory (126). Differential ATAC-Seq (+/-Dox) signal was calculated in 100-kb bins.

Relationship between different genomic annotations used for evaluation of H1 variants genome-wide distribution, including G-bands, chromatin states and A/B compartments is shown in **Figure M.9**.



GroupNumber	ChromGroup	Segments (n)	median length (bp)	Genome occupancy(%)
1	Active Promoter	40888	600	1,17
2	Promoter flanking	28899	400	0,47
3	Inactive Promoter	6945	400	0,12
4	Candidate strong enhancer	53247	400	1,04
5	Candidate weak enhancer	214570	400	4,02
6	Candidate Insulator	91546	400	1,16
7	Transcription-associated	419171	400	8,56
8	Low	370036	2200	40,86
9	Polycomb-repressed	158842	600	5,82
10	Heterochr	37155	19600	36,77

Figure M. 9. Chromatin states characterization. A) Table indicating the number of regions, their median length (bp) and genome occupancy (percentage) of each chromatin state. **B)** Chromatin state composition of Gpos and Gneg G-bands (left) or A/B compartments (right). **C)** G-Bands composition of each chromatin state. **D)** Compartment composition of each chromatin state. In (B), (C), (D) the base pair overlap between two genomic segmentations of interest is calculated and represented as proportion to the total (Percentage).

2.11. Public data on H1 variants expression and H1 gene methylation

H1 variants expression data and gene methylation in the NCI-60 cell lines panel was available at CellMiner (https://discover.nci.nih.gov/cellminer/) (187). Expression values correspond to gene composite transcript levels (fragments per kilobase per million reads, FPKM) as measured by RNA-seq. In the figures, RNA-Seq signal is expressed as log2(FPKM + 1). Gene methylation data from Illumina 450K methylation BeadChip was expressed as β-values normalized to a value between 0 (unmethylated) and 1 (methylated). β -values from all identifiers (i.e. different probes) corresponding to the same H1 gene were considered and an average β-value was calculated per each H1 variant.

2.12. Genomic datasets generated

Table M.9 shows the ChIP-Seq and RNA-Seq experiments generated in this thesis.

Exp	Cell line	Condition, IP	Sample code	GEO
ChIP-Seq	T47D-MTVL WT	H1.0	AJV20	GSE166645
		H1.2	AJV16	
		H1.3	AJV123K	-
		H1.4	AJV17	GSE166645
		H1.5	AJV21A	
		H1X	AJV19	
	T47D-MTVL multiH1sh (225sh)	H1.0 (Untr)	AJV32L	GSE156036
		H1.2(Untr)	AJV26F	
		H1.3(Untr)	AJV124L	-
		H1.4(Untr)	AJV3	GSE156036
		H1.5(Untr)	AJV30J	
		H1X(Untr)	AJV28H	
		H3K9me3(Untr)	AJV7	
		H3K9me3(Untr)	AJV84X	
		H3K27me3(Untr)	AJV86Z	-
		H3K27me3(Untr)	AJV104R	
		H1.0 (+Dox)	AJV33M	GSE156036
		H1.2(+Dox)	AJV27G	
		H1.4(+Dox)	AJV10	
		H1.5(+Dox)	AJV31K	
		H1X(+Dox)	AJV29I	
		H3K9me3(+Dox)	AJV14	
		H3K9me3(+Dox)	AJV85Y	
		H3K27me3(+Dox)	AJV87A	-
		H3K27me3(+Dox)	AJV111Y	
	SK-MEL-147	H1X	AJV119G	-
	MCF-7		AJ93	
	MCF-7 (*)		AJV38E	
	HeLa		AJ24-25	
	HeLa		AJ52-53	
	SK-N-SH (*)		AJV56X	
	HCT-116 (*)		AJV49P	
RNA-Seq	T47D-MTVL H1.2sh (156sh)	Untr	RNAJV5E	GSE190158
		Untr	RNAJV6F	
		+Dox	RNAJV7G	
		+Dox	RNAJV8H	
	T47D-MTVL H1Xsh (87sh)	Untr	RNAJV16P	-
		Untr	RNAJV17Q	
		+Dox	RNAJV18R	
		+Dox	RNAJV19S	

Table M.9. ChIP-Seq and RNA-Seq data generated. Sample code refers to in-house identification. Asterisks (*) indicate that cells were synchronized in G1/S. AJ93, AJ24-25, AJ52-53 were sequenced in an Illumina Platform at Center of Genomic Regulation (CRG). Gene Expression Omnibus (GEO) accession numbers are provided.

Results

Chapter 1. Differential distribution of six endogenous histone H1 variants in T47D breast cancer cells

1. Differential distribution of six endogenous histone H1 variants in T47D breast cancer cells

Human histone H1 variants distribution studies have been technically limited due to the lack of specific ChIP-grade antibodies. Here, we have explored the differential genome-wide profiling of six endogenous H1 variants in T47D cell line, what represents the whole somatic H1 repertoire in these breast cancer cells. Firstly, we assessed the performance and specificity of H1 antibodies. Secondly, ChIP-Seq data was generated in order to study their genome-wide distribution, at both coding and non-coding genome. Thirdly, using an alternative approach based on imaging experiments, differential H1 variants distribution at certain chromatin domains was evaluated. Lastly, both genomic mapping of endogenous H1s and microscopy approaches were used to answer whether H1 variants exhibit compensatory mechanisms upon multiple H1 depletion.

1.1. Tools for studying endogenous H1 variants: ChIP-grade antibodies validation

1.1.1. Validation of ChIP-grade antibodies performance and specificity to map endogenous H1.0, H1.2, H1.3, H1.4, H1.5 and H1X linker histone variants

During the last few years, we were able to validate the variant-specific performance of six antibodies against different somatic H1 variants in various experiments, including ChIP/ChIP-Seq. All six antibodies are commercially available (see **Table M.2** in Methods). Of note, H1.2 (ab4086), H1X (ab31972) antibodies were previously used in ChIP-Seq experiments in our lab (152). H1.0 (05-629l) was also used for ChIP-Seq in a previous publication, although their variant-specific performance was not fully addressed (148). On the other hand, H1.3 (ab203948), H1.4 (702876) and H1.5 (711912) antibodies had not been used for ChIP experiments before. In fact, to our knowledge, endogenous histone H1.3 had never been mapped in any human cell. Notably, validation of ChIP-grade anti-H1.3 was performed separately **(Figures R4-R5)** since it was the last antibody to become available in the lab. Thus, prior immunoblot experiments **(Figures R1-R3)** were performed with anti-H1.3 (ab24174), which is variant-specific at immunoblot level but is not efficient in ChIP experiments (*data not shown*).

To begin with, we checked variant specificity of the aforementioned antibodies by Western-Blot (WB), using recombinant purified histones **(Figure R.1A)** or total histones extract from the different single H1 variant knock-downs (KDs) in T47D cells **(Figure R.1B)**. Importantly, efficacy of single-KDs was previously confirmed (126,150,151). In all cases, the antibodies resulted to be specific, as the only H1 variant signal decrease observed was the one being depleted in each KD (see diagonal in **Figure R.1B**). Remarkably, H1.0 upregulation upon H1.2 or H1.4 KD was detected, as previously reported (126,150). As expected, no depletion of any H1 variant was observed when using a cell line with a control shRNA (ie. Random shRNA).

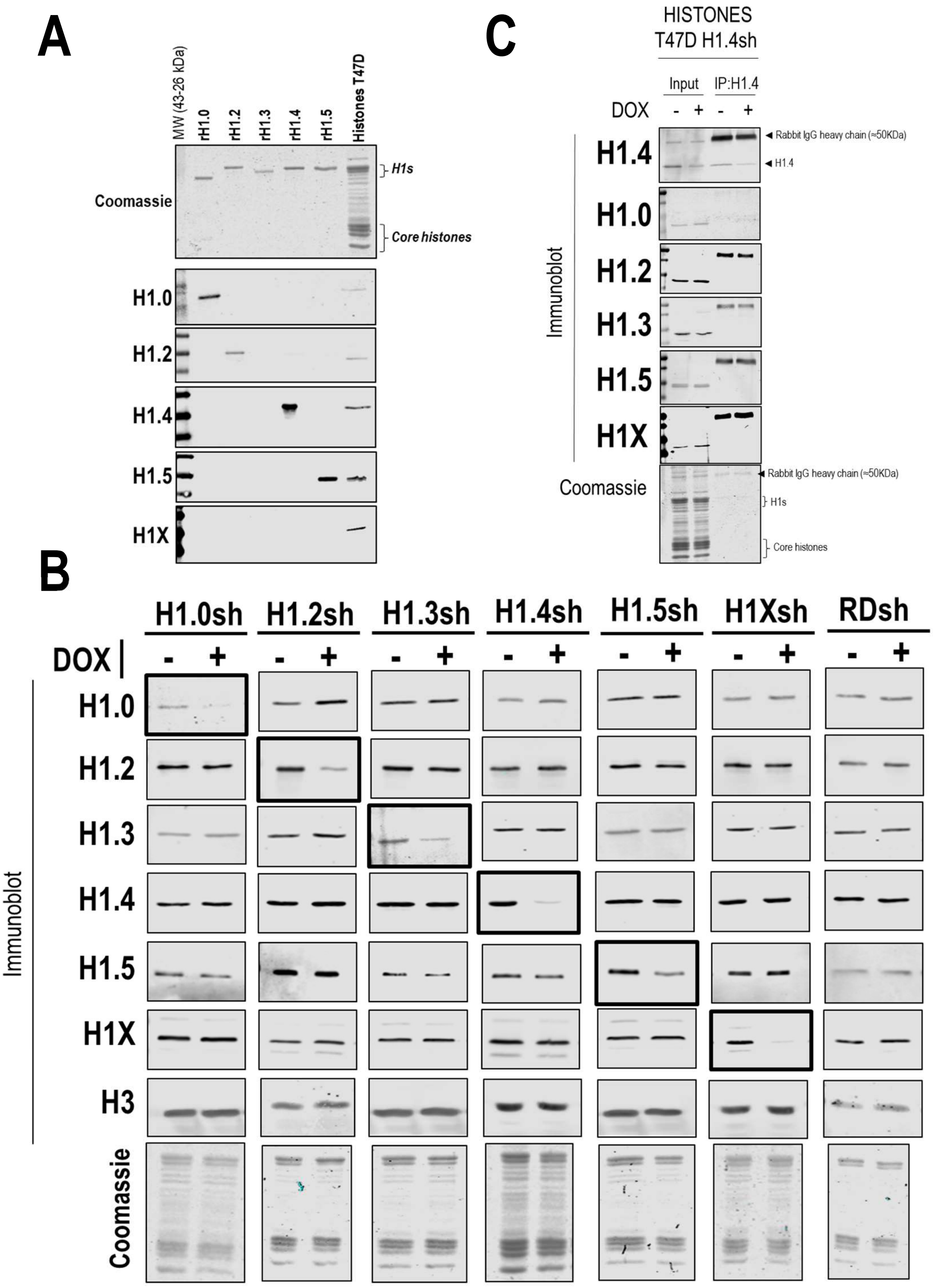


Figure R. 1. H1 variants antibodies specificity on recombinant histones and histone extracts. A) H1 antibodies recognize its specific recombinant variant (rH1). Recombinant human histone H1 variants (125 ng) were immunoblotted with the indicated antibodies. A total histones extract from T47D cells (10µg) was added as a control. All rH1s are native except for 6xHis-HA-H1.2 and 6xHis-H1.4. **B)** Total histones from T47D H1.4sh cells Untreated or 6-days Dox-treated were immunoprecipitated with the H1.4 antibody. Immunoblot analysis with the indicated histone H1 antibodies was performed on input and immunoprecipitated material. **C)** Immunoblot analysis of H1 variants in histone extracts of inducible single-H1 KDs or Randomsh in T47D. In all cases, Coomassie staining is shown as a loading control.

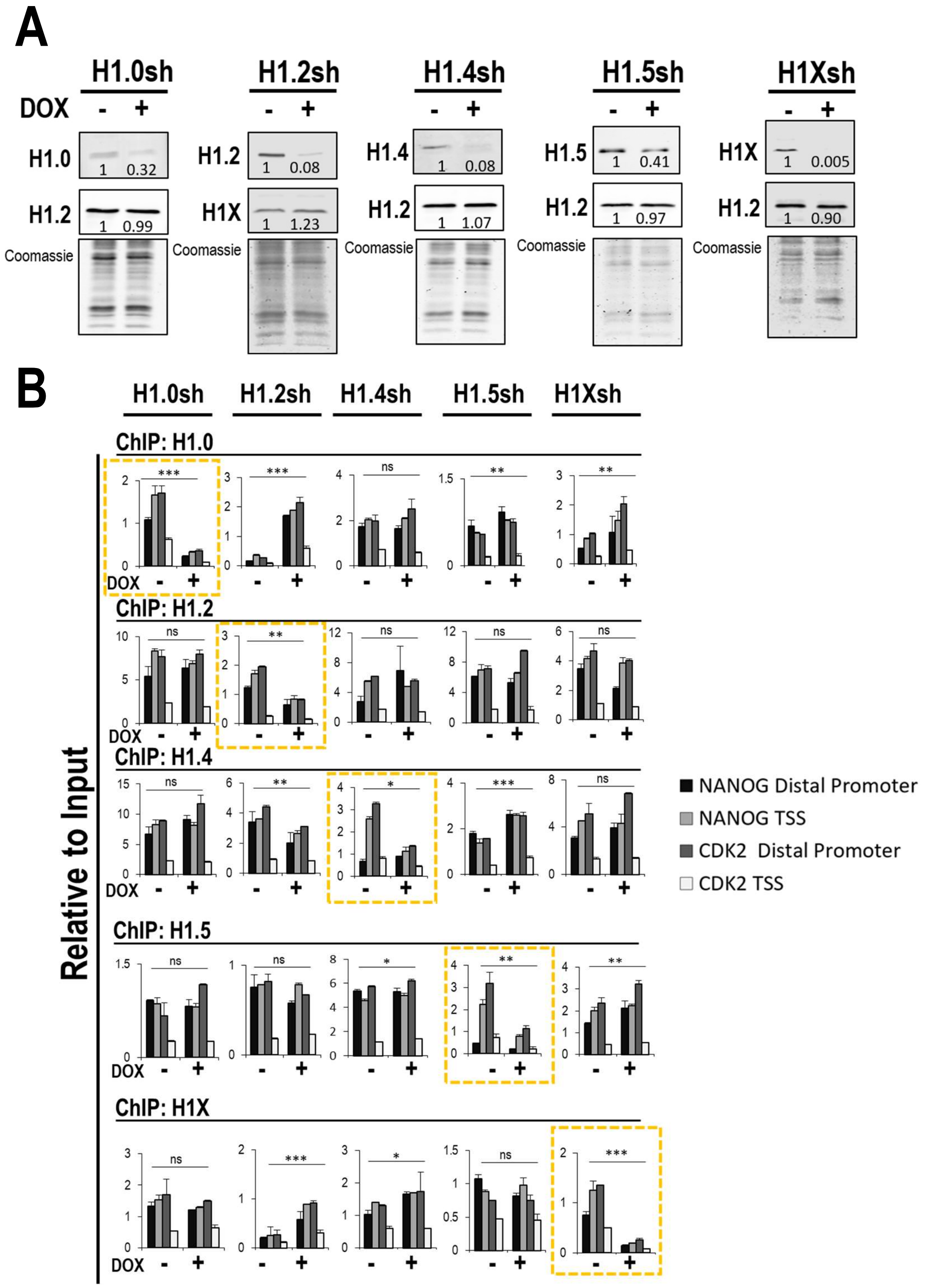


Figure R. 2. H1 variants antibodies work for ChIP. A) Immunoblot analysis of H1 variants in chromatin extracts from inducible single-H1 KDs. **B)** ChIP-qPCR of H1 variants in inducible knockdown T47D cells for different H1 variants. Chromatin samples shown in (A) were used to perform ChIP with antibodies against H1 variants. Resulting DNA was amplified by qPCR. ChIP amplification is shown relative to input DNA amplification. A representative experiment is shown. Statistical differences between Untreated (-Dox) and +Dox immunoprecipitated DNA for each H1 variant are supported by paired-t-test. (***) $p < 0.001$; (**) $p < 0.01$; (*) $p < 0.05$; (ns/non-significant) $p>0.05$.

Next, we performed WB on chromatin samples **(Figure R.2A)** and again observed a strong depletion of the variant which was being depleted in each case. We used those chromatins to perform ChIP of H1 variants in the different H1 KDs and evaluated *NANOG* and *CDK2* Distal promoters and Transcription Start Site (TSS) regions by ChIP-qPCR **(Figure R.2B)**. In all cases, an H1 valley was detected in *CDK2* promoter but not in *NANOG*, as expected from an active and inactive gene, respectively (see Methods section 1.11.1). Importantly, immunoprecipitated (IPed) DNA material was drastically reduced when the H1 variant IPed was depleted, demonstrating variant-specificity.

We used cell lines lacking some of the variants to further check specificity of H1.5 and H1.0 antibodies by WB and ChIP experiments **(Figure R.3)**. As before, the antibodies were specific. H1.0 IPed material in HeLa cells, which lack this variant, was comparable to non-specific IgG control. A similar scenario was observed when performing H1.5 ChIP in cells lacking H1.5 (i.e.HCT-116 cell line) **(Figure R.3B)**.

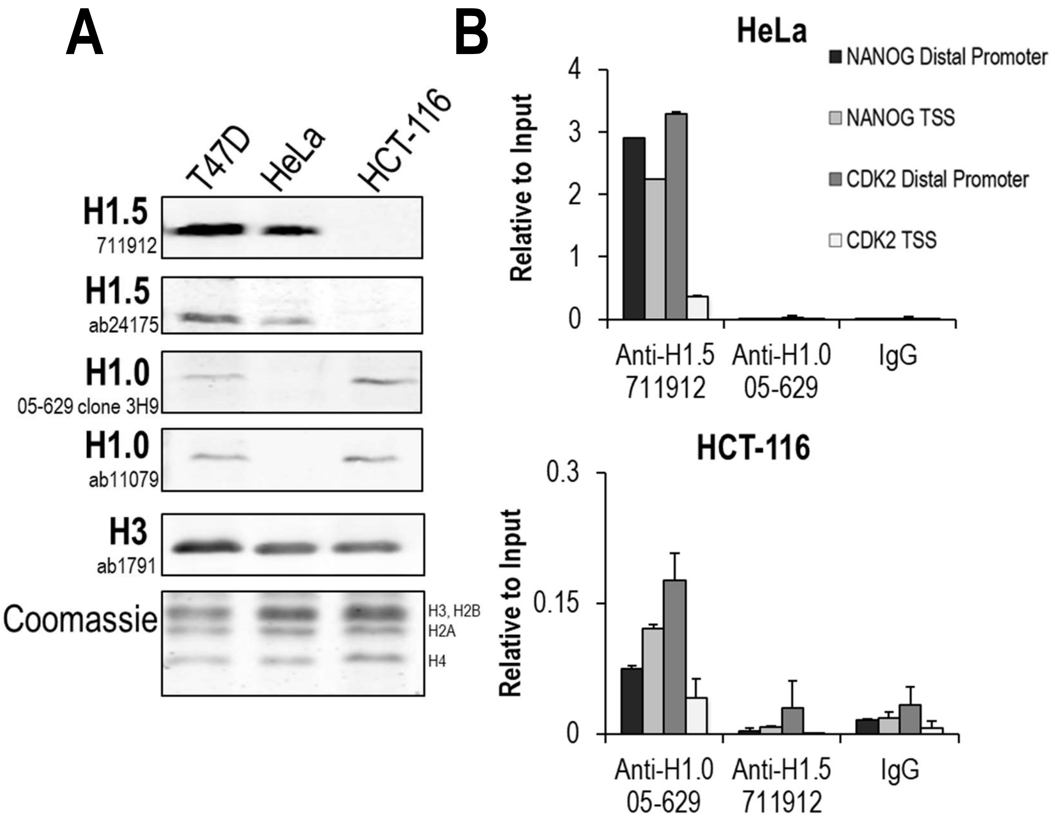


Figure R. 3. Analysis of H1 variants antibodies specificity within cell lines. A) Chromatin extracts (10 µg of protein) from T47D, HeLa and HCT-116 cells were run in SDS/PAGE and immunoblotted with the indicated antibodies against histone H1 variants or histone H3 as a loading control. **B)** ChIP-qPCR of H1 variants in HeLa and HCT-116 cells. Chromatin was used for ChIP with antibodies against H1 variants and unrelated IgG as a control. Resulting DNA was amplified by qPCR. ChIP amplification is shown relative to input DNA amplification. A representative experiment is shown.

Additional validation of H1.4 antibody was performed by immunoprecipitation (IP) of histones extract **(Figure R.1C)**. Upon H1.4 IP, only H1.4 protein was detected. Importantly, upon H1.4 depletion, no detection of other H1 variant was observed, indicating that the antibody did not gain additional cross-reactivities when its primary immunogen was absent.

An analogous characterization was performed with H1.3 (ab203948) antibody. WB of total histones in the different single-KDs showed that H1.3 depletion was only detected upon H1.3 KD **(Figure R.4A)**. WB on recombinant histones also support antibody specificity **(Figure R.4B)**. We also confirmed this depletion on chromatin samples **(Figure R.4C)** and proceed to perform ChIP. ChIP-qPCR experiments in T47D H1.3 KD showed that under basal conditions, H1.3 antibody immunoprecipitated a comparable amount of DNA material compared to the ChIP-grade H1.2 antibody **(Figure R.4D)**. Furthermore, the characteristic H1 valley was observed at *CDK2* TSS compared to distal regions but not at *NANOG*, as expected. Upon H1.3 depletion, H1.3-IPed DNA decreased by >50% but it was still higher than the unrelated IgG control, as we are using a KD partially depleting H1.3 and remaining H1.3 could be IPed.

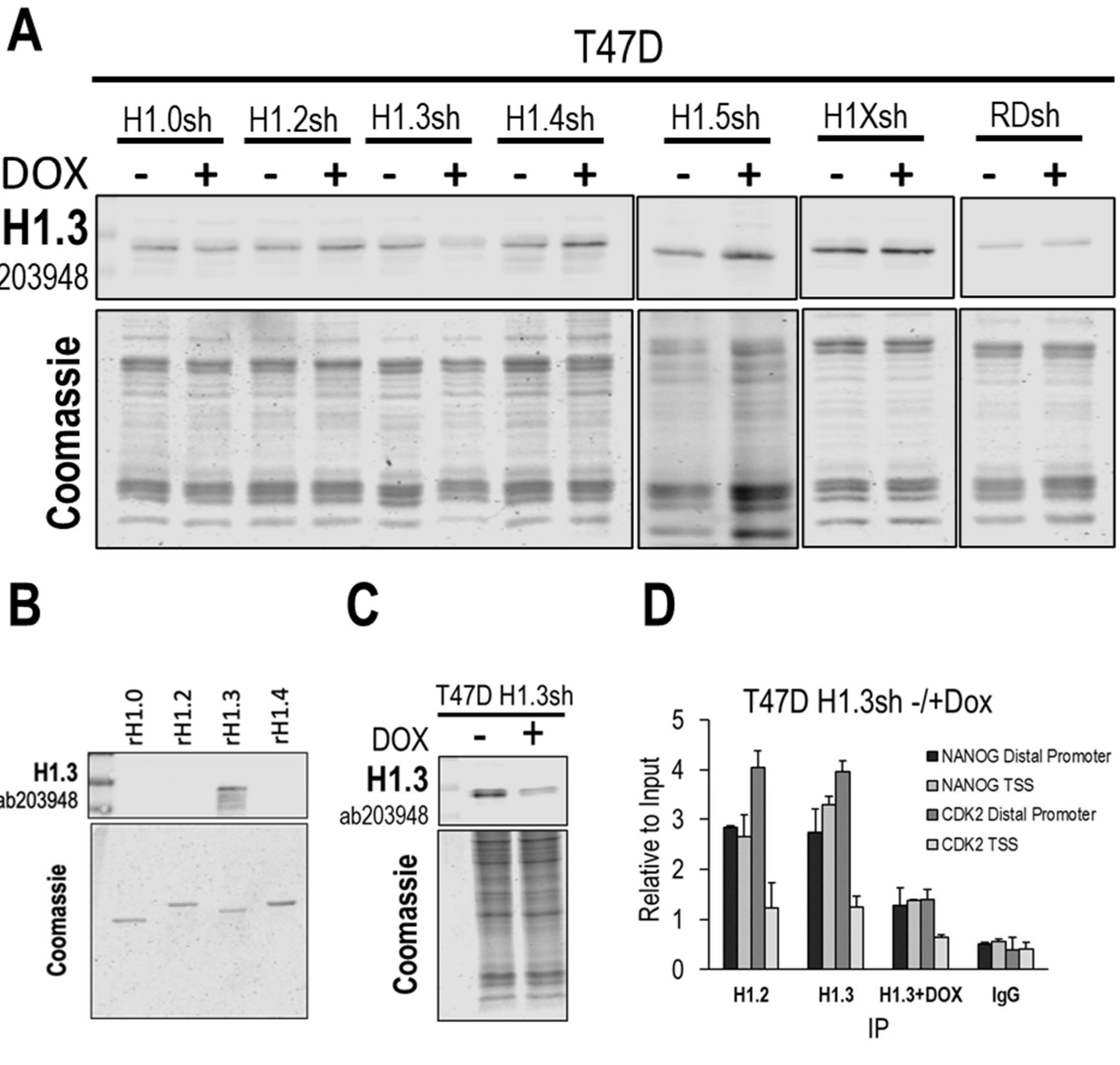


Figure R. 4. Analysis of H1.3 antibody specificity and ChIP performance. A) Immunoblot analysis of H1 variants in histone extracts from inducible single-H1 KDs or Randomsh in T47D. **B)** Recombinant human histone H1 variants (125 ng) were immunoblotted with the indicated antibodies. All recombinant H1 variants are native except for 6xHis-HA-H1.2 and 6xHis-H1.4. **C)** Immunoblot analysis of inducible single H1.3 KD chromatin extracts. Coomassie staining is added as immunoblot loading control. **D)** ChIP-qPCR of H1.3 in inducible H1.3 KD. H1.2 was added as a positive control and non-specific IgG as a negative control.

Then, we used HeLa cells, which lack H1.3 and H1.0, and MDA-MB-231, which do not express H1.3 and H1.5 proteins. H1.3 was not detected in any of these cell lines by WB **(Figure R.5A)** or by immunofluorescence **(Figure R.5B)**.

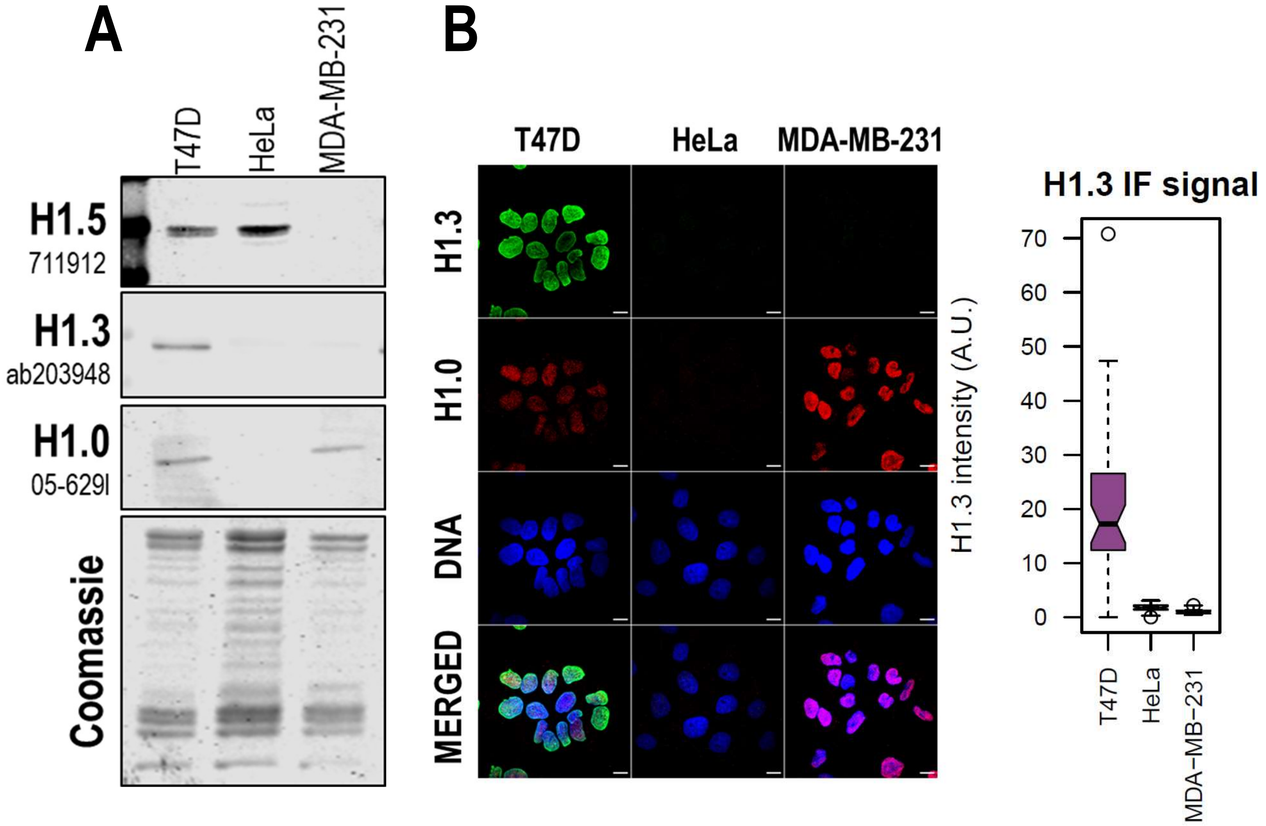


Figure R. 5. Analysis of H1.3 antibody specificity within cell lines. A) Immunoblot analysis of the indicated H1 variants in total histone extracts of T47D, HeLa and MDA-MB-231 cell lines. Coomassie staining is shown as a loading control. **B)** Immunofluorescence analysis of H1.3 and H1.0 in T47D, HeLa and MDA-MB-231. Immunofluorescence quantification of H1.3 signal in the three different cell lines is shown. Scale bar 10 µm.

Taken together, these experiments demonstrate the ChIP-grade performance and specificity of the antibodies tested, allowing for the performance of ChIP-Seq experiments on the six somatic H1 variants in human cells, thus enabling us to decode their differential distribution.

1.2. Genome-wide profiling of six endogenous H1 variants

1.2.1. H1 variants genomic distribution clusters in two different groups depending on the local GC content

We generated ChIP-Seq experiments of six endogenous H1 variants in T47D breast cancer cells to develop a systematic study of their genome-wide distribution profiles. Notably, we are studying the complete H1 repertoire, as the remaining somatic H1 variant, H1.1, is not expressed in these cells.

Giemsa bands (G-bands) were utilized as epigenetic units to compartmentalize the genome and evaluate H1 differential distribution, as mentioned in the introduction. We computed H1 variants ChIP-Seq abundance within G-bands and found that, as expected (116), H1.2 and H1X displayed opposing patterns **(Figure R.6A)**. Specifically, H1.2 was enriched at low-GC G-bands (i.e. Gpos100 and Gneg4 bands) while H1X was more abundant at high-GC G-bands (i.e. Gpos25 and Gneg1). Further analysis of all H1 variants revealed that they co-exist in two clusters based on the GC content **(Figure R.6B)**. H1.4 belonged to the 'H1X-cluster' and was enriched at high-GC bands, while H1.0, H1.3 and H1.5 clustered with H1.2 and were more enriched towards low-GC bands **(Figure R.6B)**.

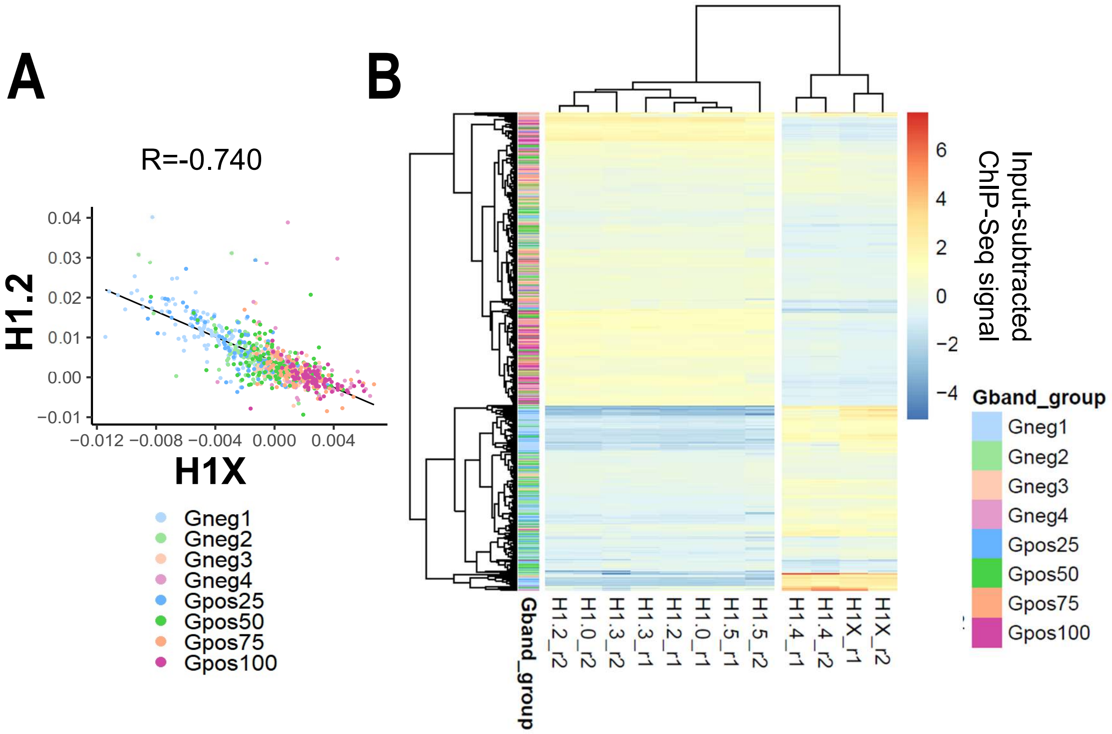


Figure R. 6. H1 variants are differentially distributed within G-bands. A) Scatterplot of H1.2 and H1X Input-subtracted ChIP-Seq abundance within G-bands. Pearson´s correlation coefficient is shown (pvalue < 0.001). Each dot represents a single G-band from the groups indicated in the legend. **B)** Heatmap and cluster analysis of the Input-subtracted ChIP-Seq abundance (scaled) of H1 variants within G-bands. Y-axis annotation shows to which group the G-band belongs according to the legend. Two ChIP-Seq replicates are shown (r1, r2).

The two differentiated clusters were also evident when evaluating H1 variants ChIP-Seq abundance at A/B compartments, which denote active and repressive genome compartments based on Hi-C experiments (**Figure R.7**). H1X and H1.4 were more enriched in the A compartment while H1.0/H1.2/H1.3/H1.5 were more abundant in the B compartment.

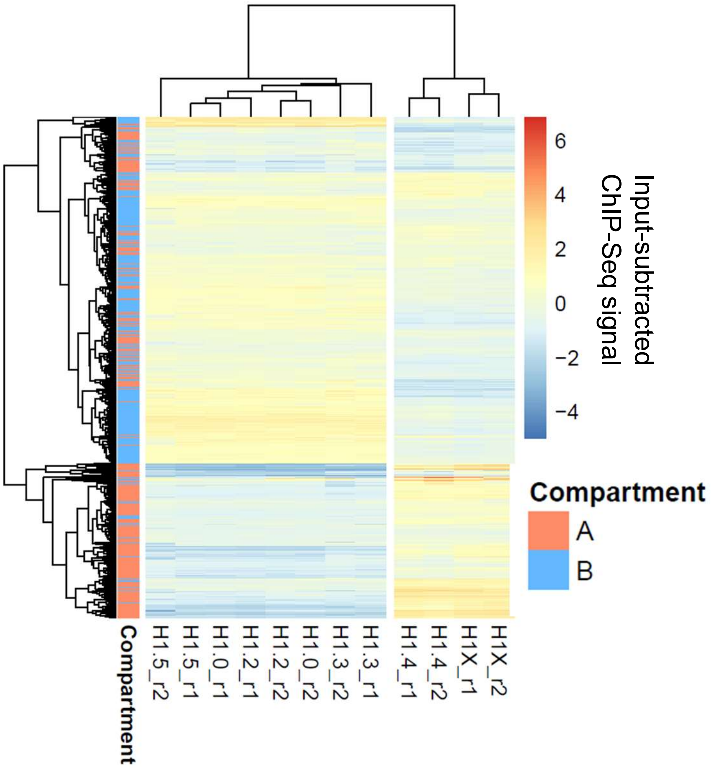


Figure R. 7. H1 variants are differentially distributed within A/B compartments. Heatmap and cluster analysis of the Input-subtracted ChIP-Seq abundance (scaled) of H1 variants within A/B compartments. Y-axis annotation shows compartment identity according to the legend. Two ChIP-Seq replicates are shown (r1, r2). A/B compartments were identified from T47D Hi-C experiments.

We used genome segmentation in 100-kb bins to investigate the correlation between different H1 variants (**Figure R.8**). Correlation of each H1 variant against H1.2 or H1X, representative prototypes of each H1 cluster, was calculated. H1.0/H1.3/H1.5 showed a strong positive correlation with H1.2 and a negative one with H1X, and they were more abundant at bins with low %GC. Conversely, H1.4 and H1X presented a negative correlation with H1.2 within 100-kb bins, with anti-correlation being stronger in the case of H1X. Related to this, although H1.4 and H1X abundance correlated positively, it was evident that higher-GC bins were preferentially enriched in H1X *versus* H1.4 (last scatterplot in **Figure R.8**). This analysis supports the division of H1 variants into two clusters according to GC content and also indicates that H1X better correlates with %GC compared to H1.4.

To simplify the terminology used, we might refer here and along the manuscript to H1.0/H1.2/H1.3/H1.5 as ‘low-GC’ H1s, while H1.4/H1X may be referred to as ‘high-GC’ H1s. This terminology is based on ChIP-Seq data clustering in two groups.

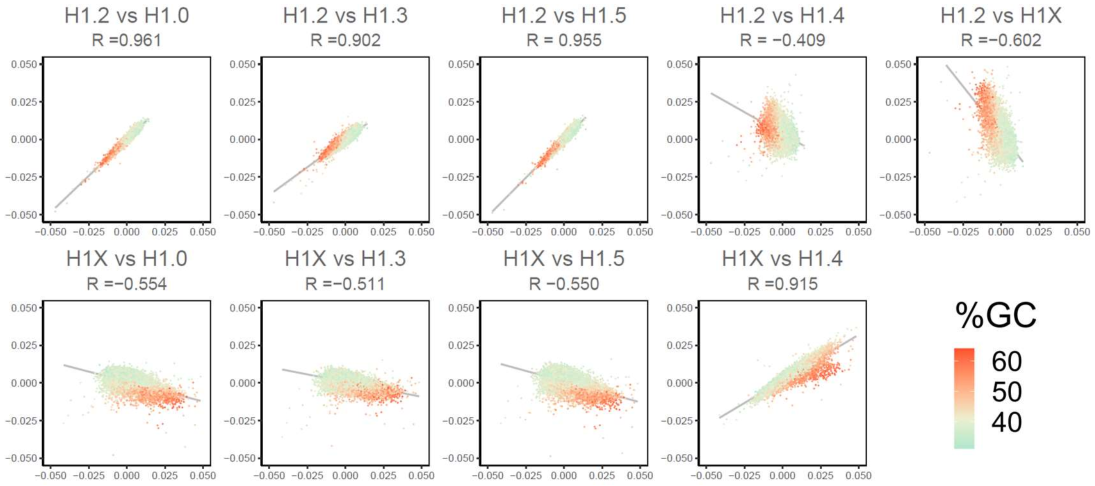
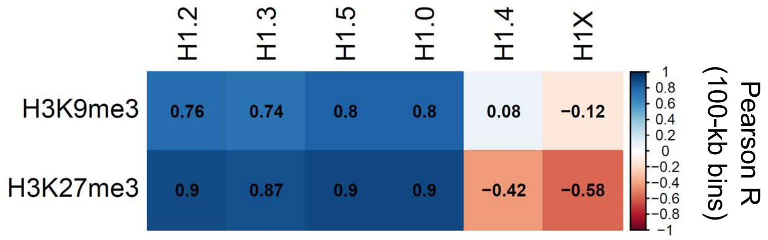


Figure R. 8. H1.2 and H1X differently correlate with the rest H1 variants. Scatterplots of the indicated H1 variant pairs Input-subtracted ChIP-Seq abundance within 100-kb genomic bins. Each dot is a 100-kb bin color-coded by its corresponding GC content. Pearson's correlation coefficient is shown (p-value < 0.001).

We then used the same 100-kb segmentation approach to evaluate the correlation between H1 variants and repressive histone marks H3K9me3 or H3K27me3 (**Figure R.9**). H1 variants belonging to the 'low-GC' cluster showed a strong positive correlation with both H3K9me3 and H3K27me3, with an even stronger correlation found with H3K27me3. In contrast, the anti-correlation of 'high-GC' H1s with these histone marks was not as high as the positive correlation denoted by 'low-GC' H1s. Indeed, H1X and specially H1.4 show almost no correlation with H3K9me3. A negative correlation of 'high-GC' H1s and H3K27me3 was found, especially for H1X.



	H1.2	H1.3	H1.5	H1.0	H1.4	H1X
H3K9me3	0.76	0.74	0.8	0.8	0.08	−0.12
H3K27me3	0.9	0.87	0.9	0.9	−0.42	−0.58

Figure R. 9. H1 variants correlation with repressive histone marks. Correlation plot showing Pearson correlation of H1 variants and H3K9me3 or H3K27me3 Input-subtracted ChIP-Seq abundance within 100-kb genomic bins. Pearson's correlation coefficient (R) is shown and color-coded (p-value < 0.001 in all cases).

1.2.2. Functional annotation of the genome denotes both specific and common features of H1 variants

We aimed to evaluate H1 variants abundance using a segmentation more associated to genome functionality, not as directly related to %GC content as previous approaches. We used as a proxy chromatin states segmentation public data on HeLa cells. The relationship between G-bands, A/B compartments and chromatin states is denoted in **Figure M.9** (Methods).

We computed ChIP-Seq H1 variants´ abundance within chromatin states and represent data as a heatmap **(Figure R.10A)**. H1 variants segregated into the same two differential clusters previously seen. However, data also revealed some common characteristic for all the variants. Certainly, all H1 variants were enriched in *heterochromatin* and *low-activity* chromatin states **(Figure R.10A)**, which represent the majority of the genome **(Figure R.10B)**. H1 variants were also particularly abundant at *polycomb-repressed* state, especially for ‘low-GC’ H1 variants, in concordance with their strong correlation with H3K27me3 **(Figure R.9)**. Most H1 variants were underrepresented from promoters, being H1X the only found to be enriched, especially at *inactive promoters*, compared to the other H1 variants.

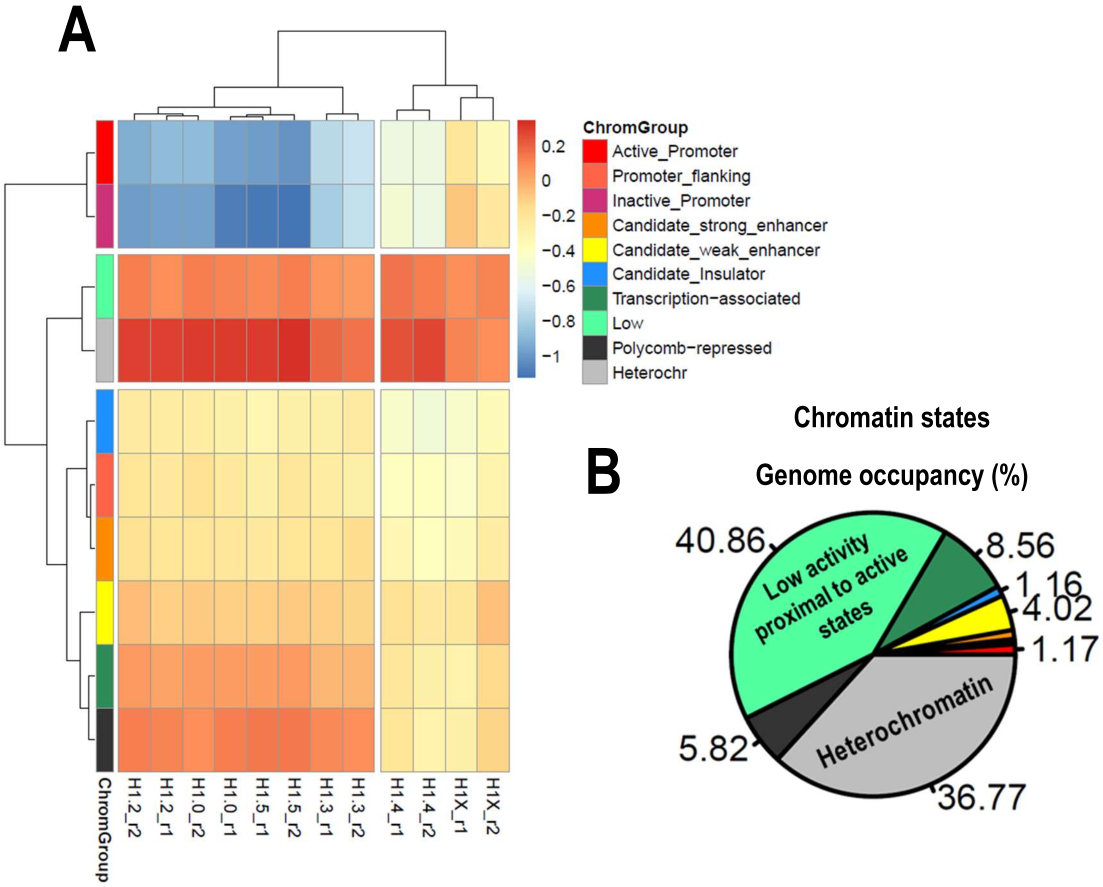


Figure R. 10. Analysis of H1 variants abundance within chromatin states reveals some redundant distribution features for all H1 variants. A) Heatmap and cluster analysis of the median Input-subtracted ChIP-Seq abundance (scaled) of H1 variants within 10 Chromatin states. Each row corresponds to a different chromatin state, according to the Y-axis annotation and the legend. Two ChIP-Seq replicates are shown (r1, r2). **B)** Contribution of each chromatin state to the genome, calculated as percentage of genome occupancy. For more details about Chromatin states annotation, see Materials and Methods **Figure M.9**.

Average H1 profiles around TSS showed that all variants, except for H1X, were depleted from gene promoters depending on their transcriptional status, being more depleted at TSS of highly expressed genes **(Figure R.11)**. This pattern was already reported for endogenous H1.2 in a previous study from the lab (152). Notably, ‘high-GC’ H1s showed different abundance profiles around TSS between them.

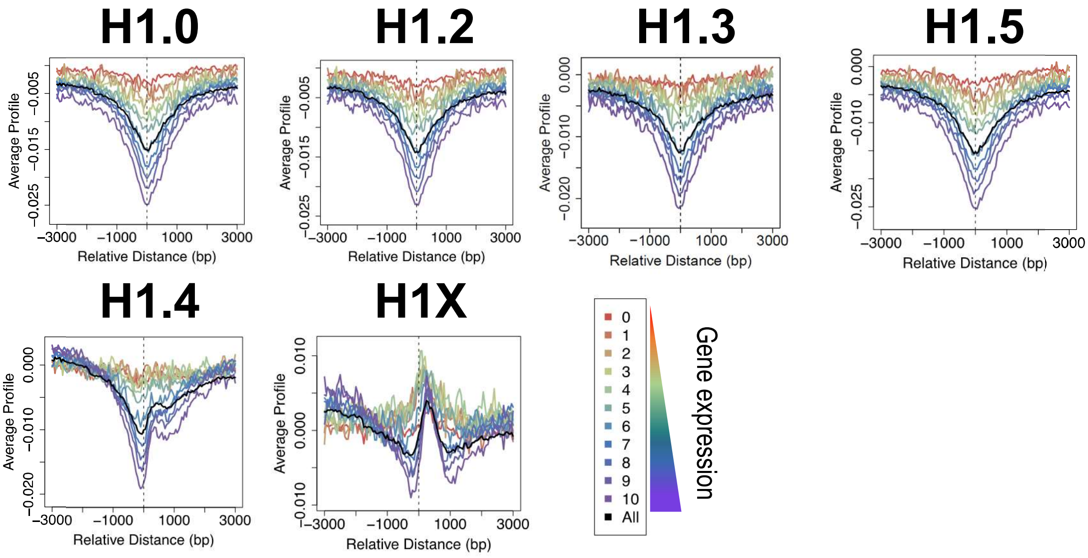


Figure R. 11. H1 variants are in general depleted from promoters depending on their transcriptional status. H1 variants Input-subtracted ChIP-Seq average profile around gene transcriptional start site (TSS). Expressed genes are divided in 10 equal groups (Groups 1-10, each containing 10% of total expressed genes) according to their basal gene expression on T47D RNA-Seq experiments. Group 0 includes non-expressed genes. Average H1 profile for all genes is shown in Black.

As chromatin states used were originated from data on HeLa cells, and many enhancers are known to be predominantly cell-type specific (188), we further analyzed H1 variants abundance within enhancers using T47D enhancer data from EnhancerAtlas 2.0 (186). Firstly, we checked whether the regions exhibited enhancer-like features. As expected, T47D enhancers were highly accessible and showed local enrichment of p300, H3K4me1 and H3K27Ac (**Figure M.8**; methods). The average profile of H1 variants denoted that all H1 proteins, without any exception, were highly depleted from enhancers **(Figure R.12)**.

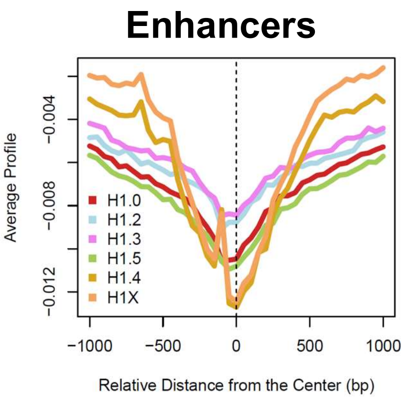


Figure R. 12. H1 variants are depleted from Enhancers. H1 variants Input-subtracted ChIP-Seq average profile around T47D Enhancer regions. T47D Enhancers were extracted from EnhancerAtlas (186).

All things considered, ChIP-Seq data on six endogenous H1 variants in T47D cells support the segregation of the variants into two large genomic clusters depending on the local GC content. However, data also supports common features observed for all the variants studied here, highlighting the co-existent balance between redundancy and specificity.

1.3. H1 variants differential abundance within repetitive elements

1.3.1. H1 variants are differently enriched within particular repetitive element classes

As previously shown, ChIP-Seq data analysis revealed that H1.0, H1.2, H1.3 and H1.5 were enriched in the B compartment while H1.4 and H1X were more abundant in the A compartment (**Figure R.7**).

Nearly half of our genome consists of repetitive sequences, which can be found at both A and B compartments (**Figure R.13A**). Therefore, we next explored H1 variants ChIP-Seq abundance at repetitive or non-repetitive fractions of both compartments, depending on their overlap with some annotated repetitive element (**Figure R.13B**). Interestingly, H1.4 and H1X were generally enriched at repetitive DNA, particularly within repeats located in the A compartment. In contrast, 'low-GC' H1 variants were more enriched within the non-repetitive genomic fraction.

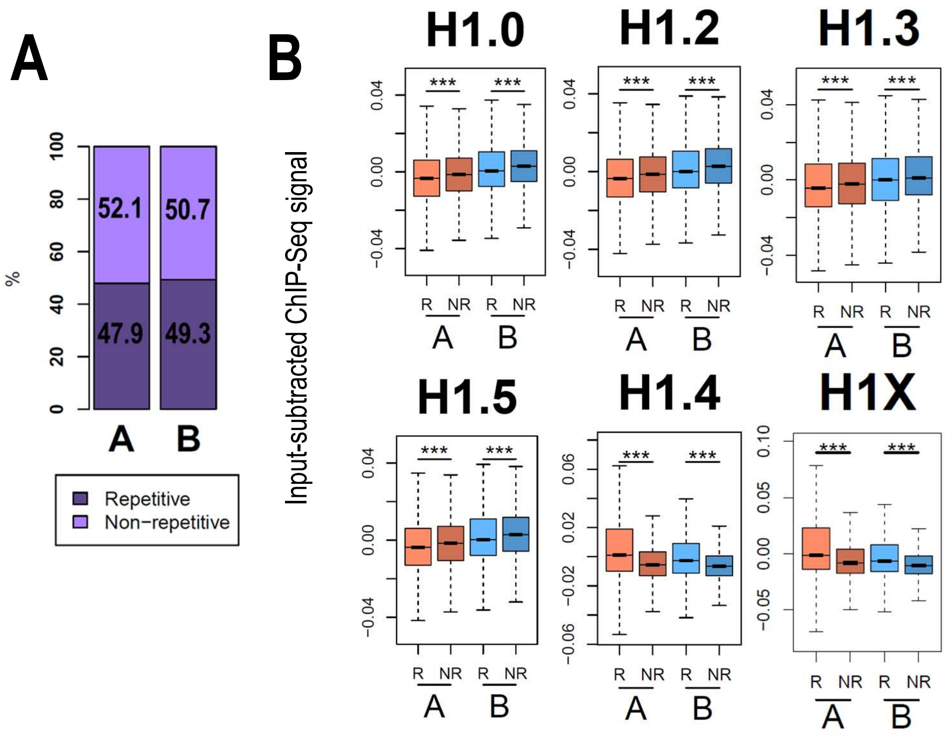
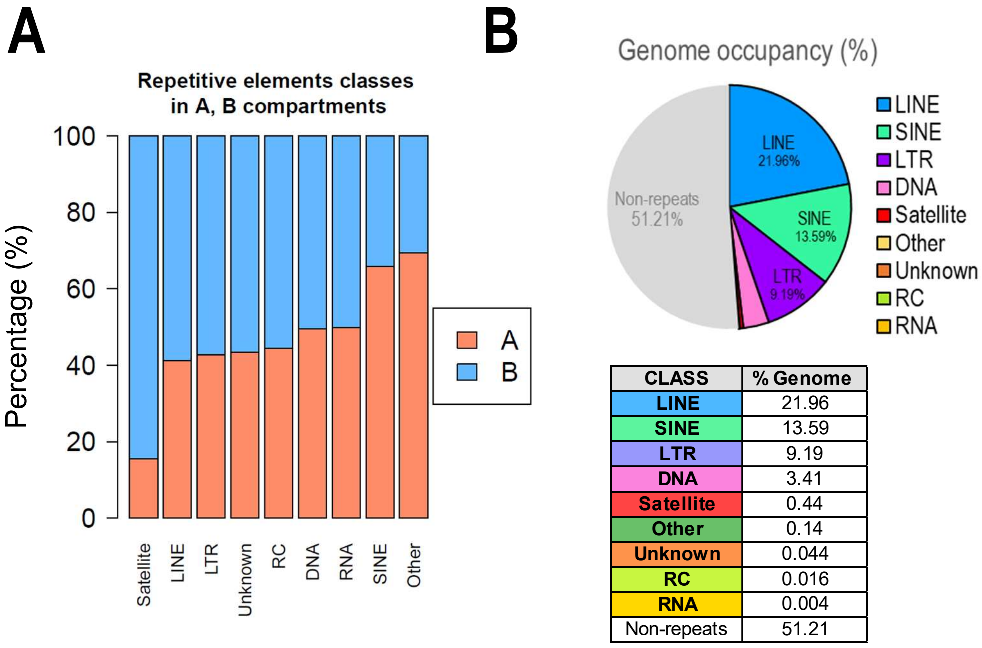


Figure R. 13. H1 variants abundance at the repetitive genome fraction. A) Genome occupancy of Repetitive or Non-repetitive genomic fractions within each compartment type. Overlap (in bp) between Repetitive Elements annotation and each compartment was calculated (i.e. Repetitive fraction) and expressed as percentage. **B)** Boxplot of H1 variants Input-subtracted ChIP-Seq abundance within repetitive (R) or non-repetitive (NR) genome fractions. Intersection between A or B compartment segments and Repetitive element annotation was calculated and each intersected region was classified as R or NR depending on their overlap or not with any type of repeat.

Transposable elements (TEs) classes are differently distributed within A and B compartments (**Figure R.14A**). For example, Satellite, LINE and LTR classes are preferentially located in the B compartment while SINE or 'Other' classes are mostly found at A compartment. 'Other' class comprises SINE-VNTR-Alus (SVA) retrotransposons. The 9 repeat classes also occupy different fractions of our genome, being LINE and SINE elements the predominant ones (**Figure R.14B**). Each class can be further subclassified into different families (see **Figure M.6** and section 2.9.1. in Methods). Due to the differential genome-wide distribution of repetitive elements classes, we aimed to explore the H1 variants abundance at different classes and/or families of TEs.



CLASS	% Genome
LINE	21.96
SINE	13.59
LTR	9.19
DNA	3.41
Satellite	0.44
Other	0.14
Unknown	0.044
RC	0.016
RNA	0.004
Non-repeats	51.21

Figure R. 14. Repetitive element classes are differentially distributed throughout the genome. A) Distribution of each repetitive element class within A or B compartments. Total length (bp) occupied for each class and their overlap with A or B compartment was calculated and expressed as percentage. **B)** Genome occupancy (total bp contribution to the genome) of repetitive elements classes, expressed in percentage.

Analysis of ChIP-Seq H1 variants abundance within repeat classes revealed that H1.4 and H1X were highly enriched at SINE and 'Other' classes (**Figure R.15**). Among both H1 variants, H1X was preferentially enriched at SVA elements (i.e. 'Other' class) while H1.4 relative enrichment at SINE class was higher. On the other hand, 'low-GC' H1 variants were more abundant within LINE, LTR, DNA, Satellite, RNA, RC and 'Unknown' repeats.

Notably, all H1 variants were relatively enriched within Satellite class, although this relative enrichment was more pronounced for 'low-GC' H1 variants. However, as highlighted before (see section 2.9.1 in Methods), almost 50% of satellite repeats are excluded from the analysis, due to their overlapping with poor-mappable regions.

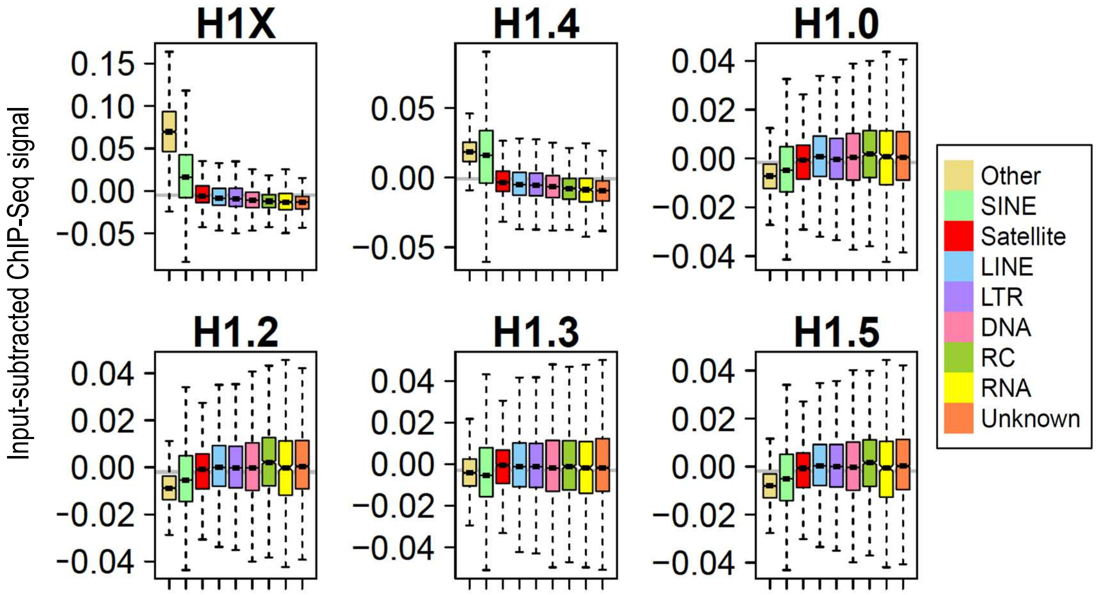


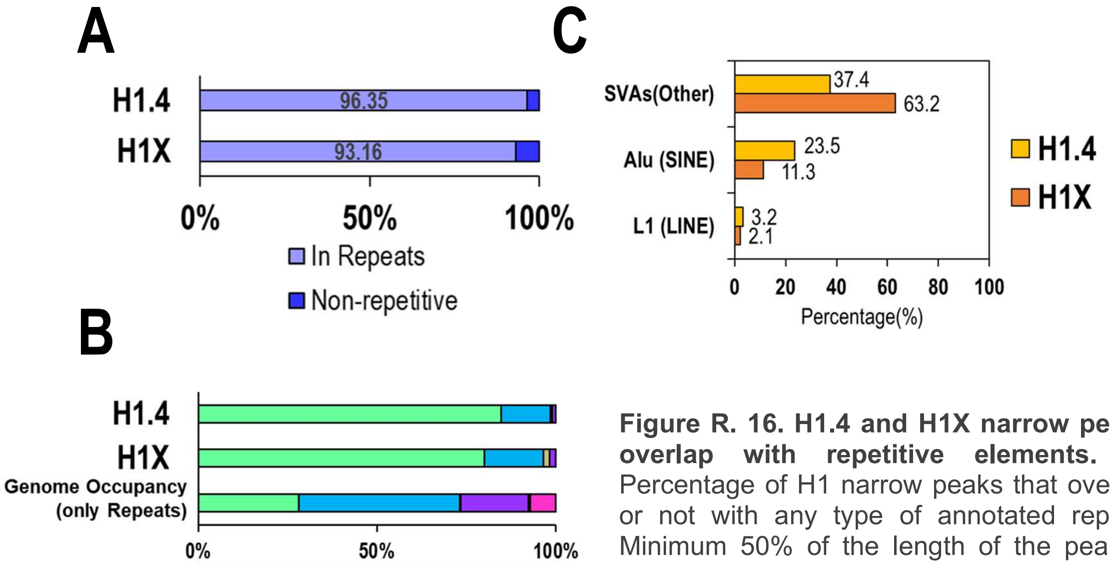
Figure R. 15. H1 variants are differently distributed within repetitive element classes. Boxplot of H1 variants Input-subtracted ChIP-Seq abundance within repetitive element classes (n=9). In each boxplot, gray horizontal line corresponds to the median abundance of the corresponding H1 variant within all repeats.

1.3.1.1. H1X and H1.4 peaks accumulate within some repetitive element classes

To further analyze H1.4 and H1X abundance within repeats, we computed narrow peak calling. Of note, this method can only be applied with these two variants, as the 'low-GC' H1 variants form large genomic domains and peaks cannot be efficiently called (**Figure M.5A**, methods). Surprisingly, almost all H1.4 (96.34%) and H1X (93.26%) overlapped with some repetitive element (**Figure R.16A**). Of these peaks in repeats, the majority of them overlapped SINE elements but also included LINE and, to a lesser extent, SVAs and LTR classes (**Figure R.16B**). Notably, SINE repeats with H1.4 or H1X peaks mostly belonged to the Alu family while LINE were almost entirely represented by L1 family. For their part, peaks within LTR class accumulated in ERV1 and ERVL-MaLR families (**Table M.6**, methods).

However, the calculation of the percentage of H1 peaks within repetitive classes is influenced by genome occupancy of each class (see **Figure R.14**). Therefore, we also calculated the percentage of SVAs, Alu, LINE-L1, LTR-ERV1 and LTR-ERVL-MaLR elements marked with some H1 peak (**Figure R.16C**). Interestingly, 63% of SVA repeats contain (at least) a peak of H1X, while 37% of them showed an H1.4 peak. On the contrary, Alu elements were more associated to H1.4 compared to H1X. While 23.5% of Alu elements showed an H1.4 peak, 11.3% of them exhibited an H1X peak. It is important to note that, due to the high number of Alu copies in the human genome (see **Table M.6**), this difference between the variants implies a huge number of Alu marked with H1.4 but not with H1X. In relation to LINEs, only 2-3% of L1 family showed some H1 peak. Lastly, less than 1% of ERV1 and ERVL-MaLR LTR families contain some H1 peak (**Table M.6**).

Each of these repetitive element classes/families was further analyzed in detail.



	Genome Occupancy (only Repeats)	H1X	H1.4
🟩 SINE	28.07	80.22	84.85
🟦 LINE	45.10	16.40	13.67
🟫 Other	0.29	1.67	0.49
🟪 LTR	18.91	1.66	0.96
🟥 Satellite	0.43	0.03	0.03
■ DNA	7.07	0.01	0.01

Figure R. 16. H1.4 and H1X narrow peaks overlap with repetitive elements. A) Percentage of H1 narrow peaks that overlap or not with any type of annotated repeat. Minimum 50% of the length of the peak is required to overlap a repeat to be considered. **B)** Number of peaks in repeats per repetitive element class, calculated as number of peaks in each class divided by number of peaks in repeats, and expressed as percentage. Genome occupancy of repeats, considering as 100% the 48.78% of the genome which is repetitive (see **Figure R.14**), is shown as a reference. **C)** Percentage of repeats within each indicated class or family that contains at least one H1 peak.

1.3.1.2. H1X shows a close association to SVA retrotransposons

To begin with, SVA retrotransposons can be divided into six different families (from SVA_A to SVA_F). These SVA families were incorporated into our genome during the evolution of hominids, being SVA_A the oldest family and SVA_F the most recently incorporated one. In fact, SVA_E and SVA_F families are human-specific (60). Analysis of H1.4 and H1X peaks within SVA families revealed that the six families were preferentially marked by H1X compared to H1.4 **(Figure R.17A)**. Among them, SVA_C and SVA_D families showed the highest percentage of repeats (≈80%) with an H1X peak.

However, analysis of ChIP-Seq abundance profile at meta-SVAs showed a gradual H1X increase from SVA_A to SVA_F **(Figure R.17B,C)**. Moreover, H1X also expanded to both 5´and 3´flanking regions of the SVA_F repeats. H1.4 local enrichment at SVA was minor compared to H1X and did not significantly increase in an evolutionary-related trend as observed for H1X **(Figure R.17B,C)**. However, H1.4 expansion to flanking regions of SVA_F elements was still observed. Notably, SVA repeats are frequently found in clusters or tandems of multiple SVA elements. This is especially characteristic in SVA_F family. We observed that the H1 expansion observed within the flanking regions of youngest SVA families is due to the H1 variants abundance within adjacent SVAs *(data not shown)*.

It is worth mentioning that differences between narrow peak analysis **(Figure R.17A)** and abundance profiles of H1X **(Figure R.17B)** are due to the differential H1X local enrichment pattern within SVA families. SVA_C and SVA_D exhibited the most sharpened H1X profiles and thus, H1X narrow peaks could be called more efficiently compared to the youngest families. Indeed, the more recent SVA_E and SVA_F families showed a higher H1X abundance but with broader local enrichments.

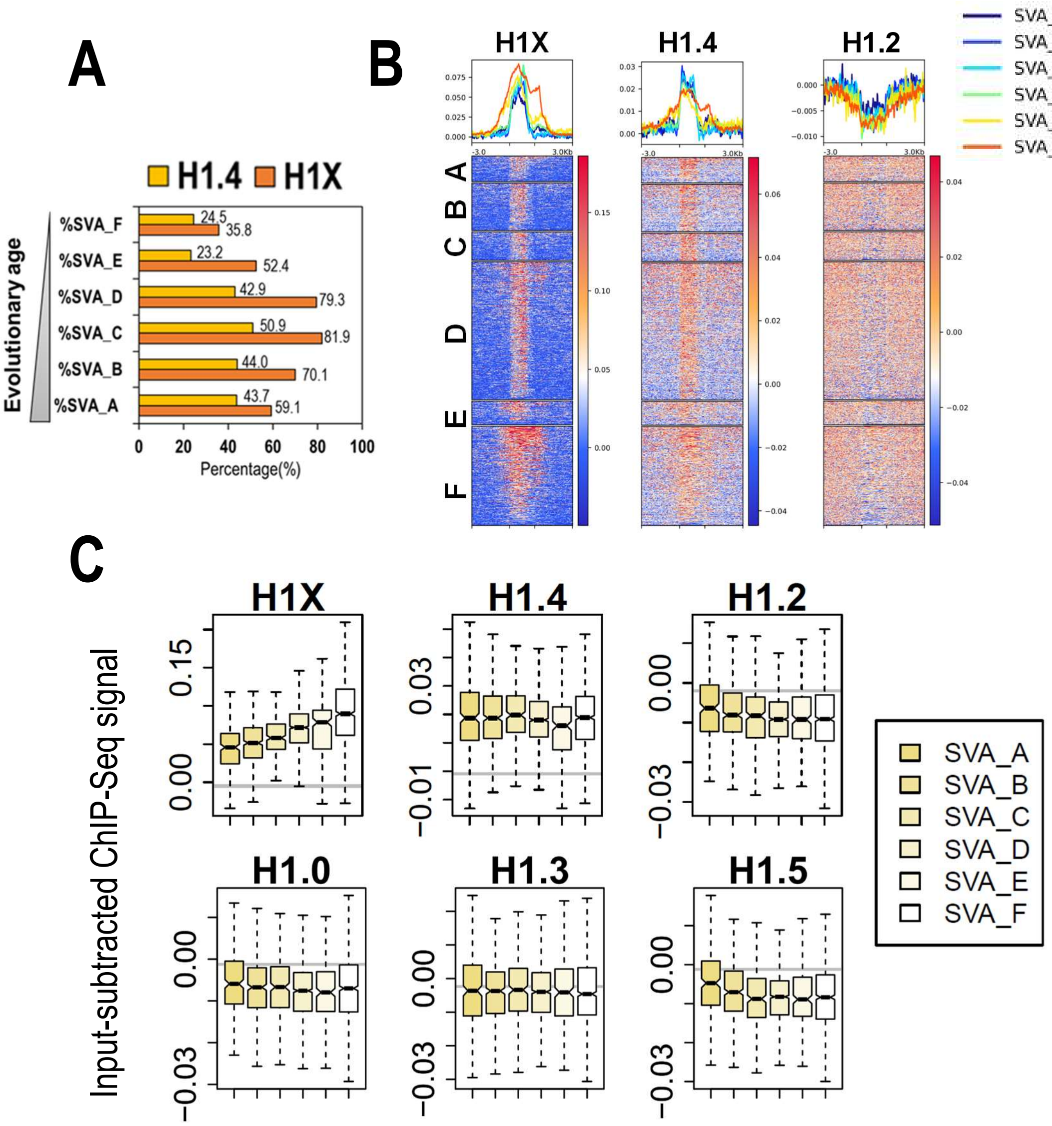


Figure R. 17. H1X is highly enriched at SVA retrotransposon families. A) Percentage of SVA repeats in each family that contain at least one H1 narrow peak. **B)** Meta-repeat profile of H1 variants Input-subtracted ChIP-Seq abundance at SVA repeats and their 3kb flanking regions. In the heatmaps, each row represents a SVA repeat of the indicated family and ordered by the corresponding H1 profile in each case (i.e., heatmaps show an independent order for each H1). Average profile of H1 variants per family is also shown in the upper line graphs. **C)** Boxplot of H1 variants Input-subtracted ChIP-Seq abundance within the six families of the 'Other' class (SVA_A to SVA_F, ordered by evolutionary age). In each boxplot, gray horizontal line corresponds to the median abundance of the corresponding H1 variant within all repetitive element classes (n=9).

We also evaluated the ChIP-Seq abundance of ‘low-GC’ H1 variants, which, in general were depleted within SVA families (i.e.‘Other class), compared to other repeat classes **(Figure R.15)**. H1.2, H1.0 and H1.5 abundance gradually decreased from older SVA_A to younger SVA_F, thus presenting an opposite relative pattern to H1X **(Figure R.17C)**. Importantly, H1.3 showed a distinct profile and was found equally abundant within all SVA families. This pattern is similar to H1.4 one, although it is very important to highlight that, while H1.4 is in general enriched within SVAs, H1.3 is relatively depleted. On the whole, analysis of H1 variants abundance within SVAs denoted differences between H1 variants belonging to the same ‘GC cluster’.

Additional SVA subfamilies have been identified in humans through network analysis based on sequence similarity between different copies (61). Applying different thresholds in their computational analysis, Levy et al. identified n=11 or n=17 SVA subfamilies, which primarily consists of subdivisions of SVA_D family. Of note, SVA_D represents the largest SVA family, in terms of number of repeats and genome occupancy (see **Figure M.6** and **Table M.6** in Methods). We computed H1X abundance within these SVA subfamilies **(Figure R.18)** and observed a general progressive H1X enrichment from older to younger subfamilies, confirming the evolutionary trend observed in previous analyses.

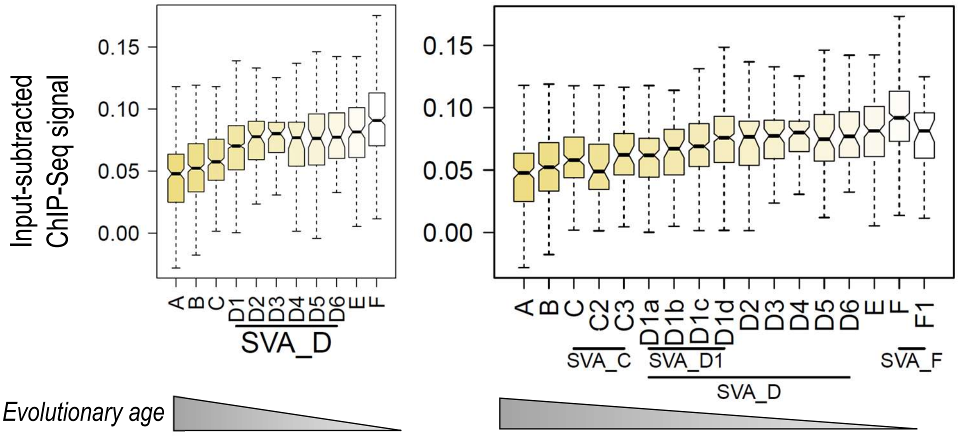


Figure R. 18. H1X abundance gradually increases from older to more recently evolved SVA subfamilies. Boxplot of H1X Input-subtracted ChIP-Seq abundance within the SVA subfamilies stablished by Levy et al. (61). Classification is done applying two different threshold parameters as defined in the original paper, which results in n= 11 (left) or n=17 (right) SVA sub-families. In general, original SVA_D, the majority SVA family, is first subdivided into 6 sub-families. SVA_D1, the majority sub-family of SVA_D can be further sub-divided into D1a-D1d sub-families. All sub-families are ordered by their evolutionary age.

1.3.1.3. Certain Alu elements are predominantly marked by H1.4

SINE repeats are classified in four families: Alu, MIR, SINE and Deu. Alu family is the most abundant family and represents almost 11% of the genome. It consists of 41 different repeats with more than one million copies in the genome. Both H1X and H1.4 were highly enriched within Alu family compared to the rest of SINE families **(Figure R.19A)**. Thus, H1X and H1.4 enrichment within SINE class **(Figure R.15)** is attributed to a particular family. On the contrary, H1.2 is relatively depleted from Alu family, in comparison to MIR, SINE and Deu families **(Figure R.19A)**. H1.0, H1.3 and H1.5 showed the same abundance profile than H1.2 *(data not shown).*

A deeper analysis of Alu family showed that the percentage of copies marked with H1.4 or H1X narrow peaks varied among different Alu elements **(Figure R.19B)**. As previously observed **(Figure R.16C)**, Alu repeats were preferentially marked by H1.4 over H1X. Interestingly, the more recently incorporated Alu elements in our genome (57,58) showed a higher percentage of repeat copies marked with H1.4. For example, almost 50% of AluYa5 repeats were marked by H1.4, while ≈20% of them showed an H1X peak. Similarly, when classifying Alu repeats into subfamilies based on their evolutionary age (58,189), the AluY subfamily, the youngest one, presented the highest H1.4 (and also H1X) peak enrichment **(Figure R.19C)**.

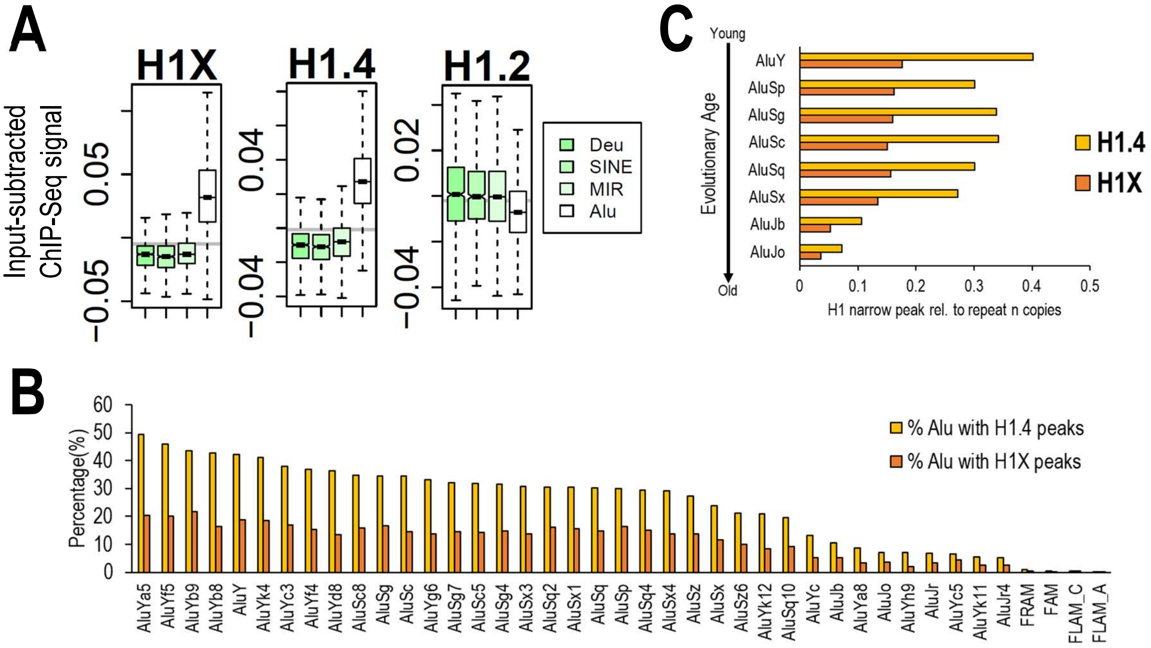
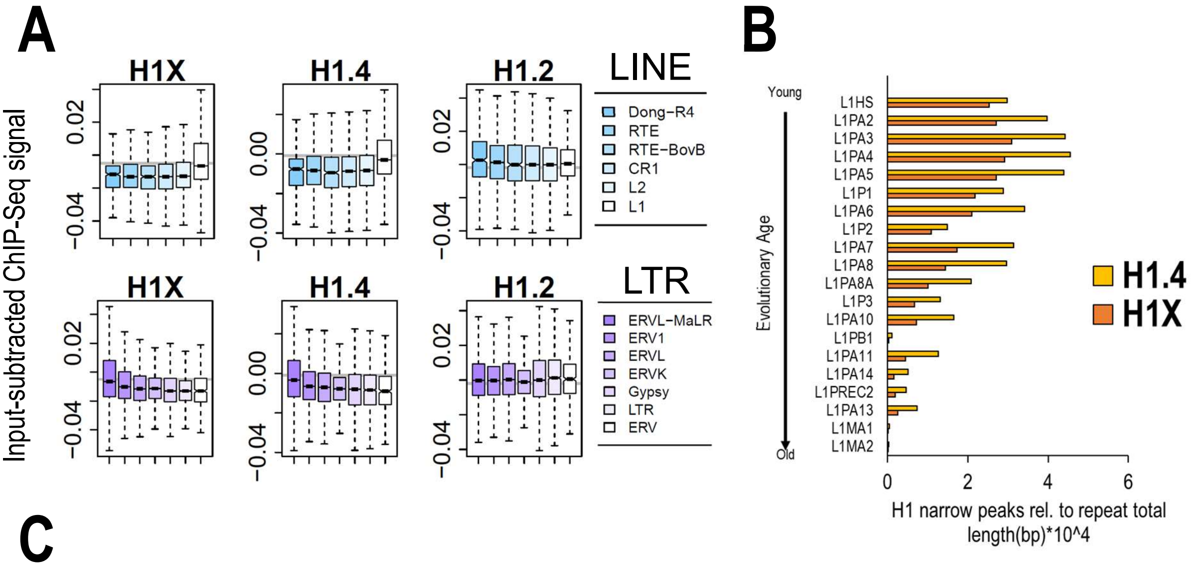


Figure R. 19. H1.4 and H1X are enriched within recently incorporated Alu elements. A) Boxplot of H1 variants Input-subtracted ChIP-Seq abundance within the SINE families (n=4). **B)** Percentage of Alu repeats that contain at least one H1 narrow peak. Alu repeats are ordered by its decreasing percentage of repeats marked by H1.4. **C)** Alu repeats classifies into sub-families (58,189) that contain at least one H1 narrow peak, represented as number of H1 narrow peaks relative to total number of repeat copies per sub-family. Alu sub-families are ordered by their evolutionary age.

1.3.1.4. H1X and H1.4 peaks within a small subset of LINE and LTR repeats

Regarding LINE elements, H1.4 and H1X were not generally enriched within LINE class **(Figure R.15)**. However, H1.4 and H1X were more abundant within LINE-L1 family compared to the rest of the LINEs families **(Figure R.20A)**, suggesting that these variants could be enriched within a subset of LINE-L1 repeats. Indeed, 2-3% of LINE-L1 repeats contained at least one H1 peak **(Figure R.16C)**. Specifically, 98.18% of H1.4 peaks and 98.76% of H1X peaks within LINE class were accumulated within the top 20 most recent L1 among the 122 L1 repeats that comprise the family (*data not shown*). Additionally, ranking of these 20 L1s according to their evolutionary age (58,189), revealed that accumulation of H1.4 and H1X peaks occurred at the most recently-evolved LINE-L1, gradually decreasing towards older L1 elements **(Figure R.20B)**.

Concerning LTRs, as occurred with LINEs, H1.4 and H1X were depleted within LTR class **(Figure R.15)**. Among the different LTR families, the higher H1X and H1.4 abundance was found at ERVL-MaLR family, followed by ERV1 family **(Figure R.20A)**. Less than 1% of all ERV1 and ERVL-MaLR families contain an H1 peak **(Table M.6)** but interestingly, these peaks accumulated within very few LTR repeats **(Figure R.20C)**. Comparing both variants, H1X was more associated to ERV1 family while H1.4 was more related to ERVL-MALR **(Figure R.20C)**. The LTR repeats more extensively marked by both H1.4 and H1X were LTR12C, LTR12E and LTR12D. Concretely, ≈37% of LTR12C repeats contain some H1X peak whereas ≈16% of them exhibited H1.4 peaks.



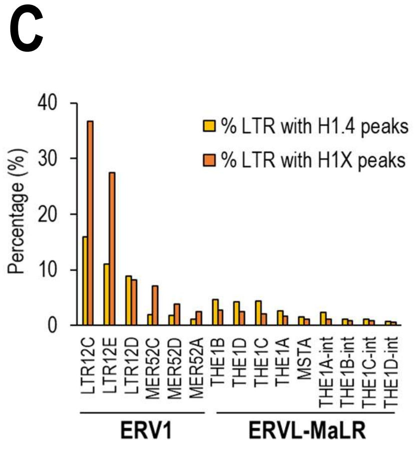


Figure R. 20. H1.4 and H1X are enriched within a small subset of LINE-L1 and LTR repeats. A) Boxplot of H1 variants Input-subtracted ChIP-Seq abundance within the LINE families (n=6) and LTR families (n=7). **B)** H1 narrow peaks within the top-20 most evolutionary recent repeats of the LINE-L1 family. Data is expressed as total H1 narrow peaks corrected by total base-pairs occupied per each LINE repeat. **C)** Percentage of LTR repeats of the ERV1 and ERVL-MaLR families that contain at least one H1 narrow peak. LTR repeats more extensively marked by H1 are shown. Repeats are ordered by decreasing percentage of elements containing H1X peaks.

In general, further analysis of H1 variants ChIP-Seq data denote their differential presence within concrete repeat classes. H1.4 and H1X are highly enriched within SINE and ‘Other’ classes. Exploration of these variants allow us to conclude that narrow enrichment peaks are concentrated within repeats and more precisely, within groups of ‘young’ repeats, evolutionary speaking. However, peaks cannot be computed for other H1 variants. For those reasons, we decided to apply an alternative approach, in order to 1) compare H1.4 and H1X with ‘low-GC’ variants and 2) confirm the association between repeats evolutionary age and certain H1 variants abundance, as suggested by the previous analyses.

1.3.2. Analysis of H1 variants abundance within transposable elements from an evolutionary perspective

1.3.2.1. *H1X and H1.4 are enriched within recently-incorporated transposable element families*

We took advantage of DFam database (171) to classify TEs based on their evolutionary age. The presence or absence of a TE in different taxa indicates when it was integrated into the genome during evolution. In other words, we can classify each repetitive element according to the oldest taxonomic clade or species in which copies have been found (see Methods section 2.9.2). We initially stratified repeats exclusively present in primates (n=440) *versus* older repeats (Non-primates; n= 523). Subsequently, 6 intermediate clades along primates evolution were used. From primates to humans the 6 clades are as follows: *Primates, Simiiformes, Catarrhini, Hominoidea, Hominidae and Homo sapiens.* Percentage of repeats included in each category was variable among different repeat classes and families **(Figure R.21)**. Thus, *Homo sapiens* repeats include only human-specific repeats, which represent the most recently incorporated ones. For example, *Hominoidea* category include those repeats present in *Hominoidea*, *Hominidae* and *H.sapiens* clades (i.e. from small apes to humans), but not in older ancestors (such as *Catarrhini* species). **Table M.7** and **Table M.8** in Methods show the number of repeats per family in each clade and a list of *Hominoidea* clade repeats, respectively.

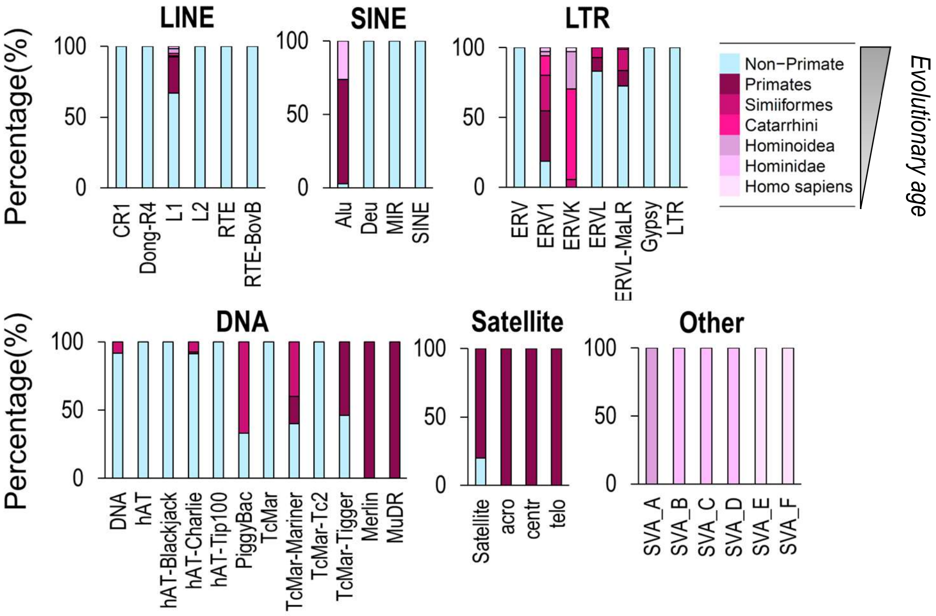


Figure R. 21. Repetitive elements have been incorporated in the genome prior and along primates evolution. Barplots show the percentage of repeats in each family that are found in each clade. Percentage is calculated considering number of repeat types (for example n=122 in LINE-L1 family), not repeat copies.

It should be noted that several analyses in this section include comparison of H1X, H1.4 and H1.2 ChIP-Seq data. H1.2 was selected as representative for the 'low-GC' cluster but figures do not include analysis with all variants to facilitate reading of the section.

Stratification of Primates repeats into the afore-mentioned six intermediate clades denoted that H1X was gradually increasing towards human-specific repeats while the opposite occurred for H1.2 **(Figure R.22)**. H1.4 followed a similar pattern to H1X, except for being less abundant in *H.sapiens* repeats compared to *Hominoidea/Hominidae* ones. Notably, both H1X and H1.4 were more enriched in *Primates* repeats compared to consecutive descendants. This was due to the presence of Alu repeats in *Primates* category but not in *Simiiformes* and *Catarrihini*. As previously observed, both H1X and H1.4 are enriched within Alu family, which represent almost 11% of the genome **(Figure R.19A)**. Members of this family are classified within *Primates* or *Hominidae* taxonomic categories (see **Table M.7** and **Table M.8** in Methods).

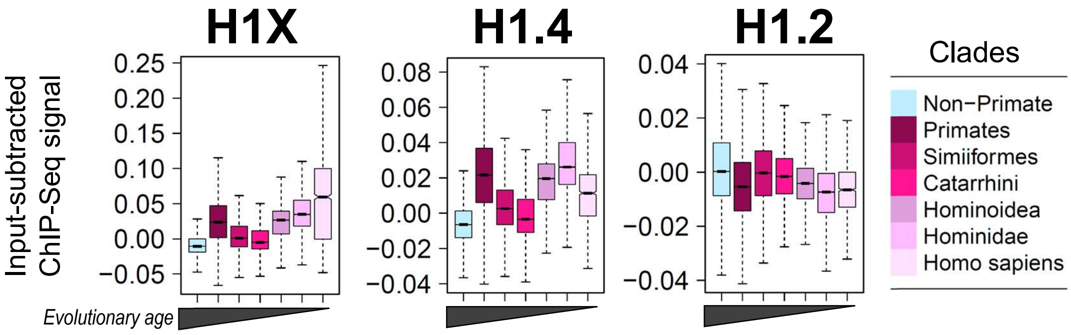


Figure R. 22. H1 variants abundance within repeats incorporated along primate evolution. Boxplot analysis of H1 variants Input-subtracted ChIP-Seq abundance within repetitive elements classified into different taxonomic clades according to Dfam database, ordered by their evolutionary age. The classification of repeats corresponds to the taxonomic clade to which the oldest ancestor in which the repeat has been found, belongs.

Indeed, analysis of repeats classified by family and clades and clustered by H1 variants abundance, resulted in three clusters that were mostly represented by recent, intermediate and older clades, respectively **(Figure R.23A)**. Interestingly, some families segregated along different clusters depending on their evolutionary age (i.e. LINE-L1 family), showing different relative H1 variants abundance. Notably, heatmap analysis also pointed out differences between H1 variants belonging to the same 'GC cluster'. In the case of 'low-GC' H1s, particular differences were found for H1.3, which for example was relatively enriched at DNA-hAT Charlie and LINE-Dong-R4 families, compared to H1.0/H1.2/H1.5.

To further analyze the relationship between H1 variants abundance and evolutionary age of repetitive elements, we next explored H1 variants abundance in each family whose members were represented along different clades separately.

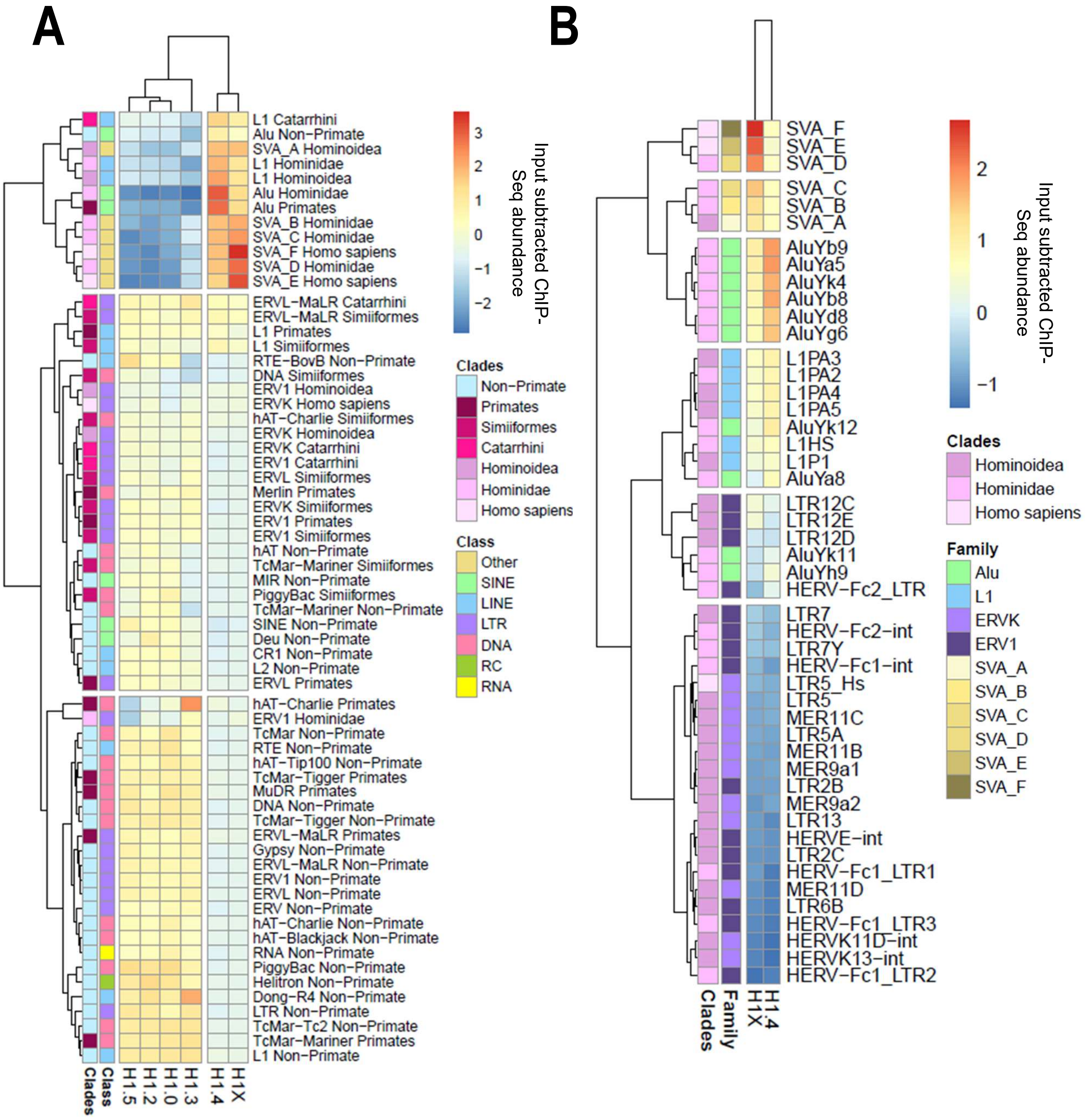


Figure R. 23. H1 variants abundance within repetitive element families from an evolutionary perspective. A) Heatmap and clustering of H1 variants Input-subtracted ChIP-Seq median abundance (scaled) at repetitive elements classified per family and taxonomic clades (n=64; Satellite and Unknown classes are excluded). Y-axis annotation indicates Repetitive element Class and Clades. First, median abundance per repeat was calculated and then median abundance per family and clade was plotted on the final heatmap. **B)** Heatmap and clustering of H1 variants Input-subtracted ChIP-Seq median abundance (scaled) at n=48 repeats from *Hominoidea* and descendant clades. Y-axis annotation indicates repetitive element family and taxonomic clades.

1.3.2.2. *H1X is highly enriched at human-specific SVA families and H1.4 at most-recent Alu elements*

Starting with hominid-specific SVA repeats, and according to previous analyses (**Figure R.17**), H1X abundance gradually increased along the evolution of SVA elements, while H1.4 abundance remained relatively constant (**Figure R.24A**). On the contrary, H1.2 abundance decreased gradually from older towards more recently incorporated SVA elements.

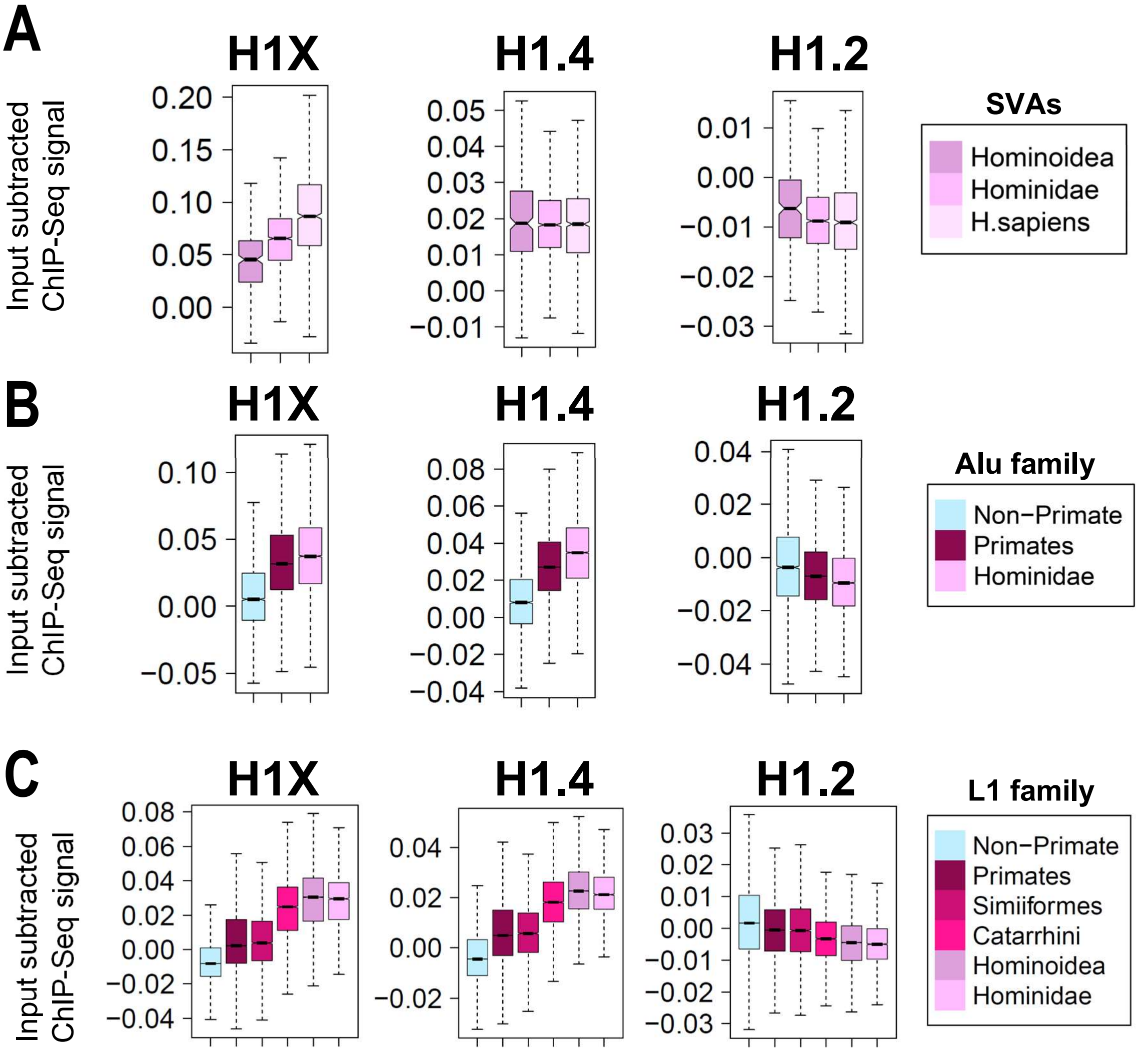


Figure R. 24. H1 variants abundance along SVA, Alu and L1 evolution. Boxplots indicate the H1 variants Input-subtracted ChIP-Seq abundance within repeats of **A)** Other Class (i.e. SVA families); **B)** SINE-Alu family and **C)** LINE-L1 family, classified according taxonomic clades.

A similar scenario was observed when analyzing Alu repeats (**Figure R.24B**). Both H1X and H1.4 were enriched in younger Alu elements but the relative enrichment in more evolutionary young Alu elements (i.e., those found in hominids) was higher for H1.4 than for H1X. On the other hand, H1.2 abundance gradually decreased towards recent Alu. Notably, *Hominidae* Alu include many Alu repeats belonging to the AluY subfamily (see **Table M.8**), like AluYa5, AluYb8 or AluYb9 repeats, which were found to be highly marked by H1X and specially H1.4 peaks in our previous analyses (**Figure R.19 B,C**).

It is worth-mentioning that although Alu family is *per se* primate-specific, fossil Alu monomer (FAM) repeats are included in this family and classified as *Non-primate*. FAM are the evolutionary ancestors from which dimeric Alu subfamilies evolved (57).

Despite both H1.4 and H1X being enriched in SVA and Alu repeats compared to other TE classes, these analyses support the closest association of H1.4 to Alu elements evolution while H1X gradually increases along SVA retrotransposons lineage.

1.3.2.3. H1X and H1.4 are abundant in a small subset of recently-evolved LINE-L1 and LTR elements

Regarding LINE-L1 transposable elements, the family comprises 122 repeats (and almost 1 million copies) that have been incorporating into the genome prior and after primate radiation. Interestingly, H1X and H1.4 abundances increased along L1 evolution, from older to recent L1 elements **(Figure R.24C)**. The opposite trend was found for H1.2, which was more abundant in those repeats also found in non-primates species, which indeed represent the majority of LINE-L1 repeats (**Table M.7** and **Figure R.21**). Notably, some differences were observed within 'low-GC' H1 abundances at non-primate LINE families. Compared to other 'low-GC' H1s, H1.3 was relatively enriched within Dong-R4 family while H1.5 was more enriched within RTE-BovB repeats, where H1.3 was relatively less abundant **(Figure R.23A)**. Moreover, those hominid-restricted LINEs (i.e. *Hominoidea* and *Hominidae*) that exhibit more H1X and H1.4 abundance were also the ones accumulating the H1 narrow peaks within LINE class **(Figure R.20B, R.23)**.

LTR repeats are classified into 7 families (ordered by decreasing genome occupancy as follows: ERVL-MaLR, ERVL, LTR, ERV1, ERVK, Gypsy and ERV). We explored H1 abundance within LTR families represented along primates taxa categories **(Figure R.25)**. H1X and H1.4 had very similar abundance patterns within LTR repeats, being enriched at younger repeats, especially in families ERV1, ERVK and ERVL-MaLR. On the other hand, H1.2 abundance decreased at more recent LTR compared to ancestral clades. Interestingly, although previous analysis showed that H1X or H1.4 did not accumulate within ERVK repeats **(Figure R.20C)**, we confirmed that correlation of evolutionary age and H1X and H1.4 abundances also exists in this LTR family **(Figure R.25)**.

Overall, although both LINE and LTR classes are enriched in 'low-GC' H1 variants abundance compared to H1X/H1.4, analysis of LINE and LTR families according to clades classification showed that H1X and H1.4 are enriched within a small subset of LINE-L1 and LTR repeats, which coincided with the elements most recently incorporated in our genome.

1.3.2.4. H1X and H1.4 within hominoidea-specific repeats

After analyzing repeat families separately, we have shown that both H1X and H1.4 abundances vary with the evolutionary age of repeats from different classes/families. The highest H1X/H1.4 abundances occurred at *Hominoidea/Hominidae/H.sapiens* restricted repeats, when compared to older members of their corresponding family. However, to allow a direct comparison of H1X and H1.4 ChIP-Seq abundances within these elements (n=48; see **Table M.8**), we performed heatmap analysis of the 48 repeats present in *Hominoidea* and descendant clades **(Figure R.23B)**.

Among these *Hominoidea* repeats, the lowest H1X (and H1.4) abundance was found at LTR repeats from ERVK and ERV1 families, except for H1X being relatively enriched at LTR12C/E elements. Interestingly, H1X peaks within LTR repeats also accumulated in these elements, as previously shown **(Figure R.20C)**. Regarding H1X, the highest H1X abundance was found at the six families of SVA elements (from SVA_F to SVA_A as previously analyzed). After SVAs, H1X was more enriched within AluY repeats and also young LINE-L1s elements. For its part, the highest H1.4 abundance was found at most recently evolved AluY repeats (especially at AluYa5 and AluYb8/9). Secondly, and contrary to H1X, H1.4 was more abundance at young LINE-L1s compared to SVA families.

Thus, direct comparison of repeats that showed that highest H1X and H1.4 abundance in each family confirmed that H1X and H1.4 show particular relative abundances within different repeats present in *Hominoidea* species. The analysis confirmed that the highest H1X abundance occurred at SVA families while the highest H1.4 abundance occurred at top-young Alu repeats. Additionally, both H1X and H1.4 are more enriched at recent L1 compared to recent LTR. Among *Hominoidea*-LTRs, H1X was especially enriched at LTR12C/E.

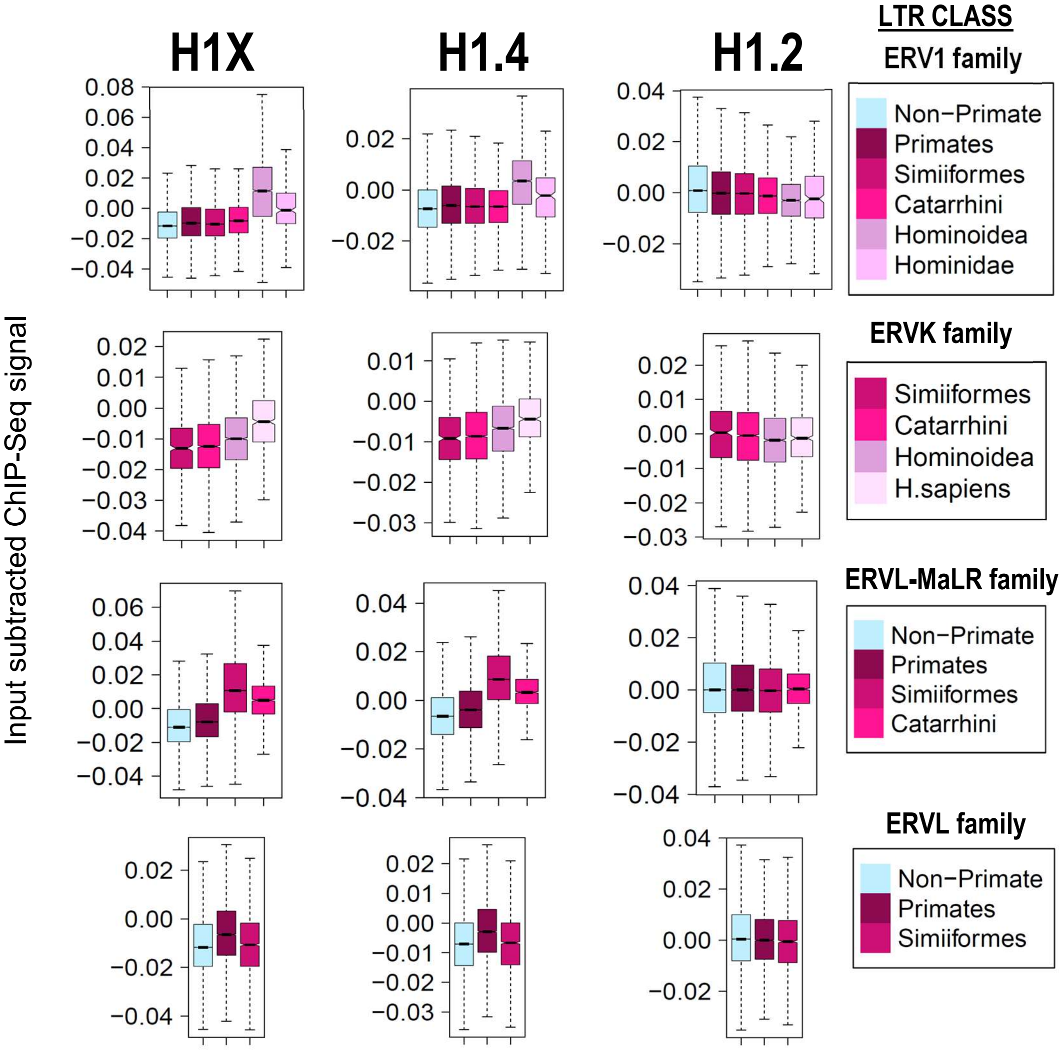


Figure R. 25. H1 variants abundance along LTR families evolution. Boxplots indicate the H1 variants Input-subtracted ChIP-Seq abundance within repeats of LTR families, classified according to taxonomic clades. LTR families ERV, Gypsy and LTR are not included because all repeats within these families are classified as ‘Non-primate’.

1.3.2.5. H1 variants abundance and DNA repeats

Lastly, we focused on DNA class of TEs. DNA repeats are divided into 12 different families, being hAT-Charlie and TcMar-Tigger families the main contributors to DNA class occupancy, which in total represent 3.41% of the human genome **(Figure R.14B, M.6C)**. While most DNA repeats are also found in Non-primates species, some of them have been incorporated through primates evolution **(Figure R.21, Table M.7)**. In general, DNA repeats are more enriched in 'low-GC' H1 variants compared to 'high-GC' ones, as observed in previous Figures **(Figure R.15)**. Interestingly, some differences between the 'low-GC' variants were observed when analyzing hAT-Charlie, as H1.3 was highly enriched within *Primates* hAT-Charlie repeats **(Figure R.23A)**. Indeed, hAT-Charlie Primates repeats was the group showing a highest H1.3 enrichment when classifying repeats according to their family and clade **(Figure R.23A)**.

In summary, evaluation of H1 variants ChIP-Seq abundance within TE families classified according to their evolutionary age confirmed that H1X and H1.4 are enriched within TEs recently-incorporated in our genome. On the contrary, H1.0/H1.2/H1.3/H1.5 are more abundant within older elements of the different families, although analysis also showed some differences between these variants. Both H1X and H1.4 are abundant at recent SVA and Alu retrotransposons. However, top-H1X enrichment is found at SVA families while young AluY elements present the highest H1.4 abundance. Although LINE, LTR and DNA repeats are in general more enriched in H1.2 and the rest of 'low-GC' H1 variants, H1X and H1.4 are also abundant at the most-evolutionary recent LINE-L1 elements and, to a lesser extent, in young LTR repeats.

1.3.3. H1X depletion effects on gene and repeats expression

In general, ChIP-Seq analysis showed a more divergent genome-wide distribution of H1X compared to the rest of the variants (Sections 1.2-1.3). For this reason, we aimed to explore how H1X depletion impact gene and TEs expression.

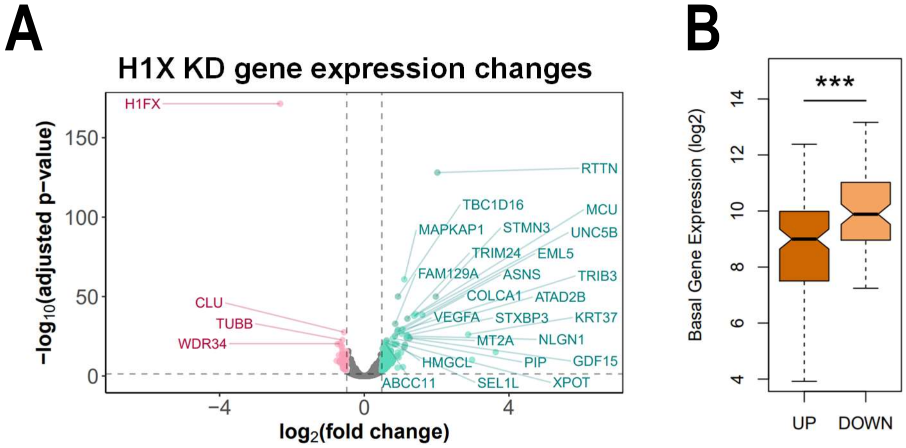


Figure R. 26. Deregulated genes upon H1X depletion. A) Volcano plot shows the deregulated genes upon H1X depletion. Considering FC >1.4 and p-adj > 0.05), 126 genes were upregulated (colored in green) and 81 downregulated (colored in red). **B)** Boxplot shows the basal gene expression of genes that became up- or down-regulated upon H1X KD.

Previous work of the lab studied gene deregulation upon H1X depletion, through a microarray platform (151). To expand previous results, we performed RNA-Seq experiments in T47D H1Xsh-/+Dox cells.

Differential expression analysis denoted that 126 genes were found up-regulated and 81 down-regulated **(Figure R.26A)**, confirming that H1 depletion impact gene regulation in both senses, as previously observed for other variants (126,150,151). We next wondered whether these genes were also found deregulated under other H1 KD conditions. Concretely, H1.4 KD, H1.2 KD and multiH1 KD (i.e. combined H1.2+H1.4 KD, see Introduction, section 2.2.3) were compared. Analysis showed that each KD affected a different subset of genes, as most of the genes were up- or down-regulated exclusively under a specific KD condition **(Figure R.27)**. The higher overlap was found between single H1.2 and multiH1 KD (both UP or DOWN genes), which altered a higher percentage of the gene transcriptome, in comparison to H1.4 or H1X KDs. Regarding deregulated genes upon H1 KD, ≈67% of the upregulated genes (n=85) were not deregulated in other H1 KDs. ≈38% of the downregulated genes (n=31) were exclusively deregulated in H1X KD, while the rest were also found downregulated in other H1 KD conditions. In general, upregulated genes were more specific compared to downregulated genes.

Moreover, we observed that upregulated genes upon H1X depletion presented a lower basal expression than downregulated genes **(Figure R.26B)**, pattern that was previously found when analyzing other H1 KD conditions (126).

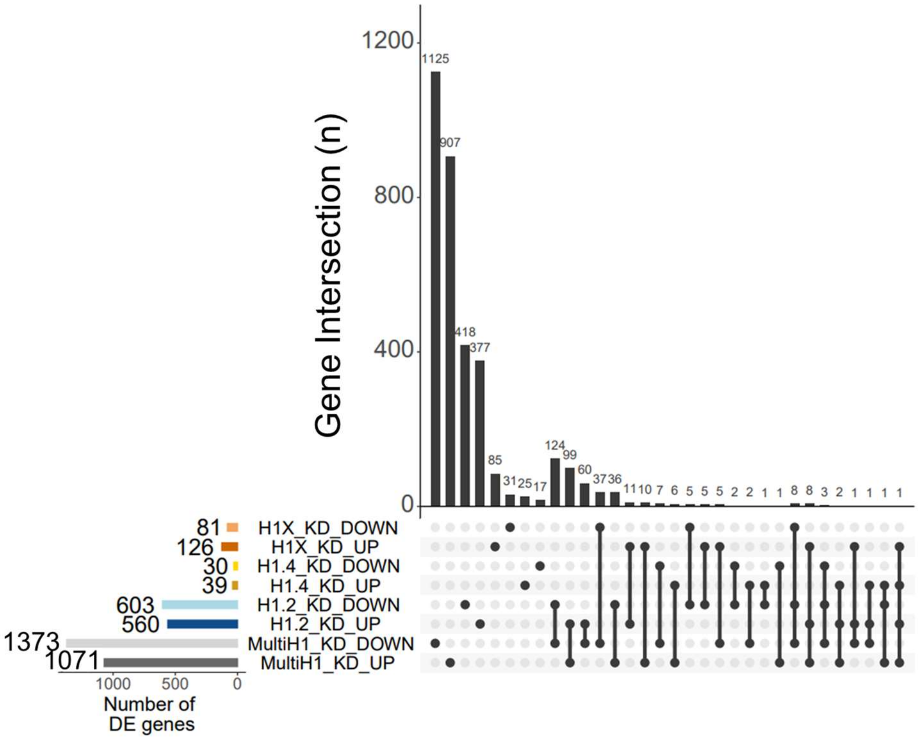


Figure R. 27. Depletion of different H1 variants affect a specific subset of genes. The left barplot shows the number of upregulated (UP) or downregulated (DOWN) genes upon different H1 KD conditions: MultiH1 (i.e. H1.2+H1.4), H1.2, H1.4 or H1X KDs. The main graph represents the intersection between the different datasets, in terms of number of genes found in a single dataset or simultaneously found in multiple datasets. In each case, datasets being intersected are indicated by the bottom-colored gray cells.

We recently realized that the most upregulated genes upon H1X KD did not show a homogeneous transcriptional increase from the first exon. That is, the upregulation was found from non-TSS regions, including intragenic **(Figure R.28A,B)** but also upstream regions **(Figure R.28C)**. In the case of KNG1 (the most upregulated gene), which was not expressed at basal conditions, upregulation upon H1X KD was only detected starting from an intermediate exon **(Figure R.28A)**. A similar scenario was found for RTTN gene **(Figure R.28B)**. In other cases, abnormal expression was not detected intragenically, but found to initiate upstream from TSS (see KRT37 example in **Figure R.28C**). These observations suggest that H1X could prevent both intragenic and intergenic cryptic transcription.

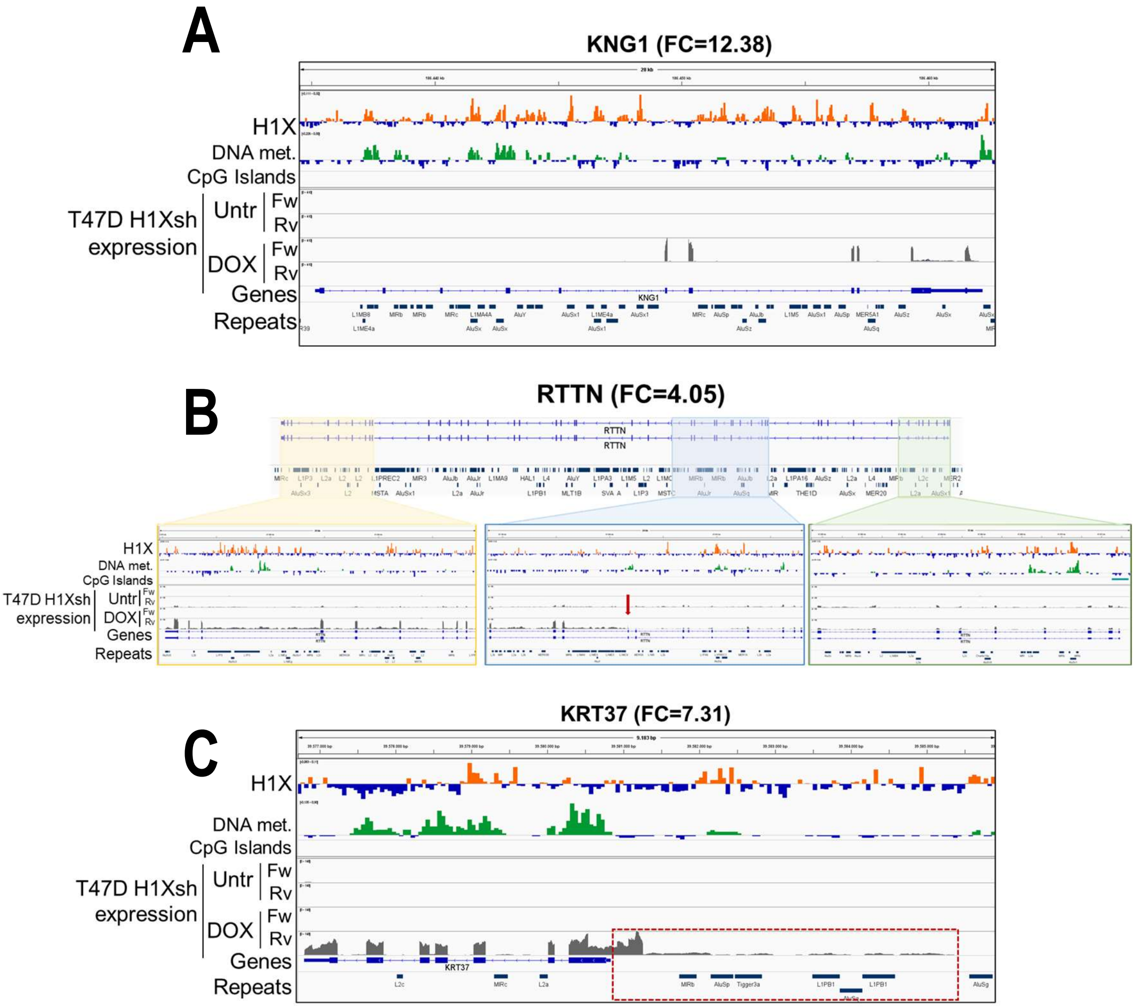


Figure R. 28. Upregulated genes upon H1X depletion are transcribed from non-promoter regions. A, B, C) Browser captures of genes KNG1 (B), RTTN (C), KRT37 (D), which are among the top upregulated genes upon H1X depletion. In the three captures the following data is represented: H1X ChIP-Seq abundance, DNA methylation abundance, CpG islands coordinates, gene expression in T47D H1Xsh-/+Dox as measured by RNA-Seq (coverage) and separated by strand (fw or rv), RefSeq genes and repetitive elements. In (B), three different zoom-ins of the gene are shown. Red arrow indicates the apparent increase in transcription occurred in +Dox condition within an intron. In (C), dashed red box highlights that transcription upregulation occurs upstream KRT37 TSS.

To confirm the genome browser observations, we plotted the exon expression of some upregulated genes at basal and H1X KD conditions **(Figure R.29A)**. The analysis illustrates how upregulation occurred from intermediate exons but was not detected within the first exons of the transcripts. Importantly, we checked whether the first upregulated exon coincides with a first exon of an annotated alternative transcript. However, after manual checking of multiple upregulated genes, this was not observed, discarding that the transcriptional defect was due to upregulation of shorter annotated transcripts isoforms *(data not shown)*.

A global analysis of all upregulated genes showed that intermediate exons are more transcribed in comparison to the first exon in H1 KD, supporting the idea of cryptic or spurious intragenic transcription initiation **(Figure R.29B)**.

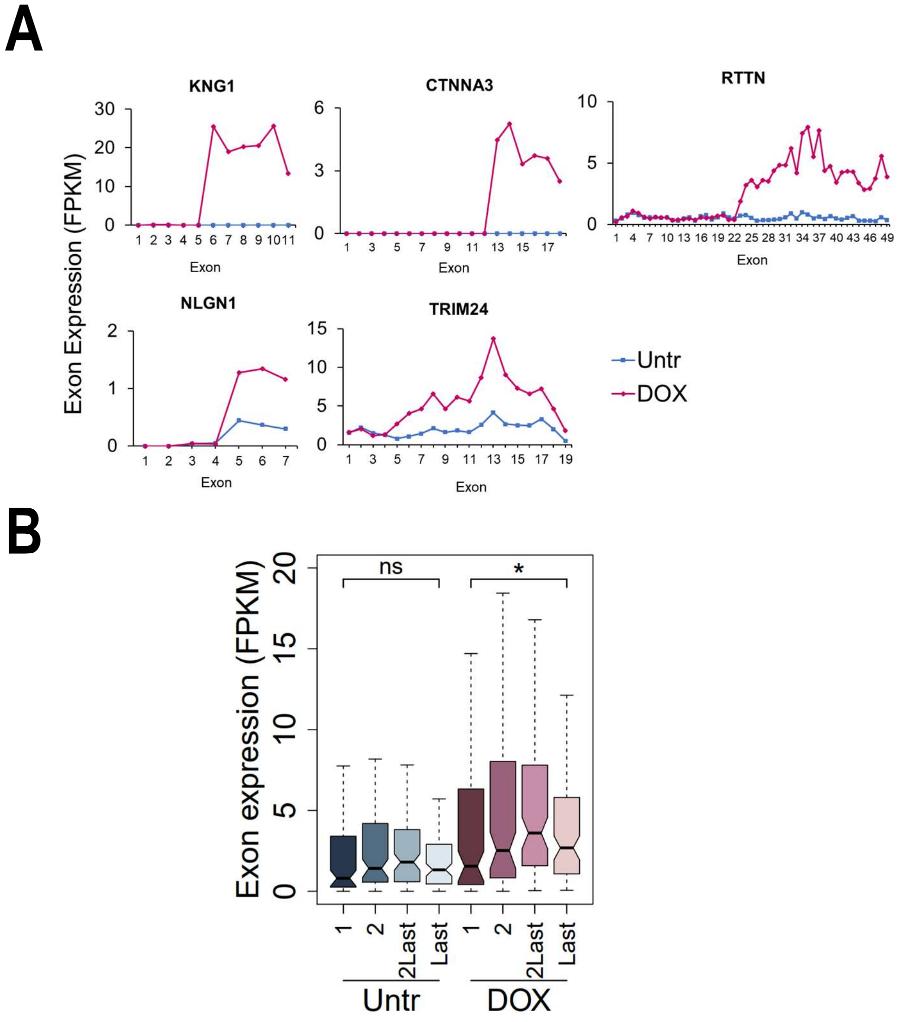


Figure R. 29. H1X depletion causes cryptic intragenic transcription. A) Graphs represent the expression of the different exons in the indicated genes as measured by RNA-Seq experiments in T47D H1Xsh -/+Dox. The five genes shown are among the most upregulated genes upon H1Xsh. **B)** Boxplot show the RNA-Seq expression (expressed in FPKM) of the first exon (1), second (2), penultimate (2last) or last exons of upregulated genes upon H1X depletion, in Untreated or Dox-treated T47D H1Xsh cells. For the analysis, upregulated genes with four or more exons were selected. For each gene, the longest transcript was considered (n=102). One-sample Wilcoxon signed-rank test was used to compare expression of first and last exon in each treatment condition (*) p-value < 0.05; (ns; non-significant) p-value > 0.05.

Overall, these initial analyses point to a role of H1X in transcriptional regulation. Although we still do not know the exact mechanism by which H1X safeguards a proper transcriptional initiation, several possibilities arise, as further discussed later (Discussion section 1.2.5).

On the other hand, we aimed to explore whether H1X depletion have an impact on repetitive elements expression. Concretely, we focused on recently-incorporated TEs which were highly enriched in H1X, as denoted by ChIP-Seq data. We designed specific primers (see Methods section 1.1.1) for these young TEs to analyze by RT-qPCR changes in expression upon H1X depletion. As H1.4 was also enriched within more recent TEs, H1.4 KD was also analyzed.

Both H1X and H1.4 KDs led to moderate upregulation of multiple repetitive elements from different classes (**Figure R.30**). In general, a higher upregulation was found upon H1X depletion. Apart from young TEs of LTR, LINE, SINE and SVAs classes, H1X depletion also caused upregulation of SATa and SST1. SST1 was also upregulated upon H1.4 KD.

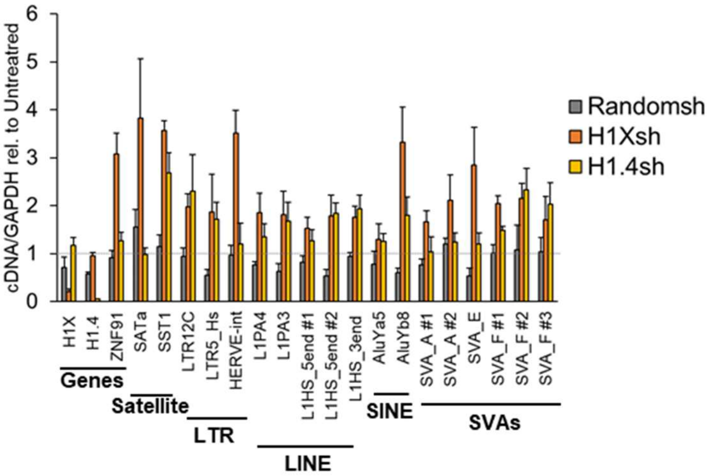


Figure R. 30. H1X and H1.4 depletion lead to moderate transcriptional activation of recently-evolved transposable elements in which these H1 variants are enriched. RT-qPCR performed in T47D H1X or H1.4 KD denoted upregulation of multiple repetitive elements. T47D Randomsh were also analyzed as control. Apart from SATa and SST1 satellites, the rest of repeats analyzed are restricted to Hominoidea species (see Table M.8) and enriched in H1X and H1.4, as denoted by ChIP-Seq analyses. Hash (#) indicates different primer pairs. Expression was corrected by GAPDH and expressed relative to Untreated condition in each case.

To gain insight into the interplay between H1 variants and other repressive mechanisms of TEs, we next analyzed how DNA hypomethylation affected TEs expression. To do so, we performed aza treatment in Dox-inducible H1Xsh cells. RT-qPCR analysis showed that DNA hypomethylation mediated by aza led to a moderate upregulation of repeats, similar or higher than H1X KD (**Figure R.31**) while effects of combined Dox/aza treatment were in general additive in the induction of young TEs from different families.

Interestingly, ZNF91 gene was found upregulated upon H1X KD (**Figure R.30-R.31**, RNA-Seq FC=1.45). ZNF91 is a KRAB-ZNF that represses SVA elements, through recruiting of KAP1/SETDB1 machinery (73,74). ZNF91 upregulation upon H1X KD could indicate that a crosstalk between both mechanisms exist for maintaining SVA elements silent.

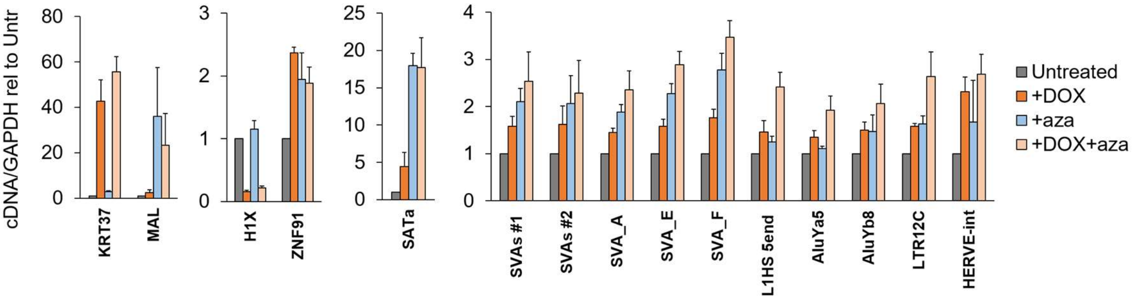


Figure R. 31. H1X depletion and aza treatment have an additive effect in activating young transposable elements. Combined Dox (H1X KD induction) and aza (DNMT inhibitor) treatment was performed and expression was analyzed by RT-qPCR. Figure shows RT-qPCR expression of different genes or repetitive elements. KRT37 gene is among the top upregulated genes upon H1X depletion. MAL is an aza-responding gene. Expression was corrected by GAPDH and expressed relative to Untreated condition.

In summary, these results suggest that H1X and H1.4 act as transcriptional repressors of TEs, including those in which they were found specially enriched. However, H1X or H1.4 KD did not produce a huge upregulation of TEs, indicating that they are an additional silencing mechanism and other repressive mechanisms, including DNA methylation, also play a role.

1.4. Differential distribution of H1 variants within the nuclei and at distinct chromatin domains

To further explore the differential distribution of H1 variants in T47D breast cancer cells analyzed by ChIP-Seq, we took advantage of imaging experiments, including both confocal and super-resolution microscopy. With this complementary approach we aimed to study whether H1 variants show unique nuclear patterns, association with specific chromatin domains and how H1 depletion affects those environments or chromatin structure.

1.4.1. Nuclear distribution of six endogenous H1 variants in T47D breast cancer cells

Immunofluorescence analysis demonstrated that different H1 variants exhibit unique nuclear patterns **(Figure R.32)**. H1.2, H1.3, and H1.5 were observed to be enriched at the nuclear periphery, while H1.0 was distributed throughout the nucleus, with certain territories tending to be located peripherally. On the other hand, H1.4 and H1X were found to be homogeneously distributed throughout the nucleus, with the difference that H1X was particularly abundant in the nucleoli.

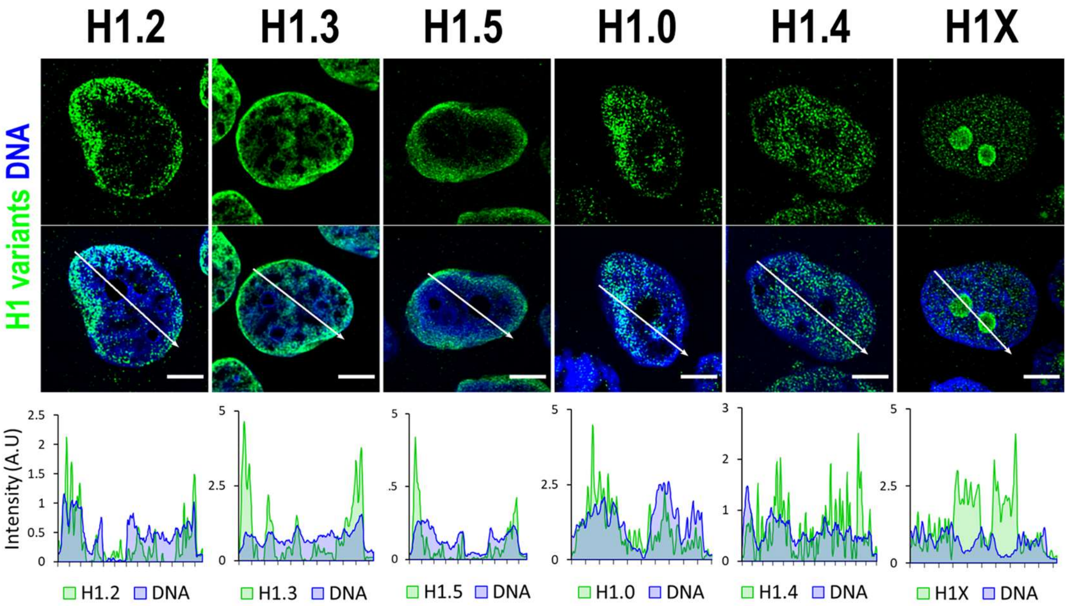


Figure R. 32. H1 variants show differential nuclear profiles. Upper panel: Confocal immunofluorescence of H1 variants (green) and DNA staining (blue). Bottom panel: Intensity profiles of H1 variants and DNA and along the arrows depicted in the upper panel. Scale bar: 5µm

The intensity profiles depicted in **Figure R.32** aim to illustrate how H1 variants belonging to the 'low-GC' cluster identified by ChIP-Seq (i.e. H1.2, H1.3, H1.5 and H1.0) show a coincident pattern with DNA staining one, suggesting an enrichment at more condensed-DNA nuclear areas, including but not limited to the nuclear periphery. H1.4 profile only partially mimicked DNA pattern while H1X profile was opposite to that of DNA. Overall, these results are compatible with the classification of H1 variants into two differential clusters, as suggested by the ChIP-Seq analysis.

To further examine H1 variants nuclear distribution, we performed an analysis of the H1 radial intensity distribution. Each nucleus was automatically divided into four sections of equal area (as exemplified in **Figure R.33A**) and the percentage of H1 intensity present in each area was quantified (**Figure R.33B**). H1.2, H1.3, and H1.5 showed a clear relationship with radiality, becoming increasingly abundant from the nuclear center to the periphery. Importantly, H1.0 was found to be most abundant at the two most peripheral percentiles. A different distribution was observed for H1.4, which was more equally distributed along A1-A4 sections, being more abundant at intermediate A2-A3 percentiles. H1X was gradually increasing towards the nucleus center. It is important to note that, in part, this gradual profile is due to the nucleolar H1X fraction, as nucleoli tend to be located at central nuclear positions and we are not excluding nucleoli from the analysis.

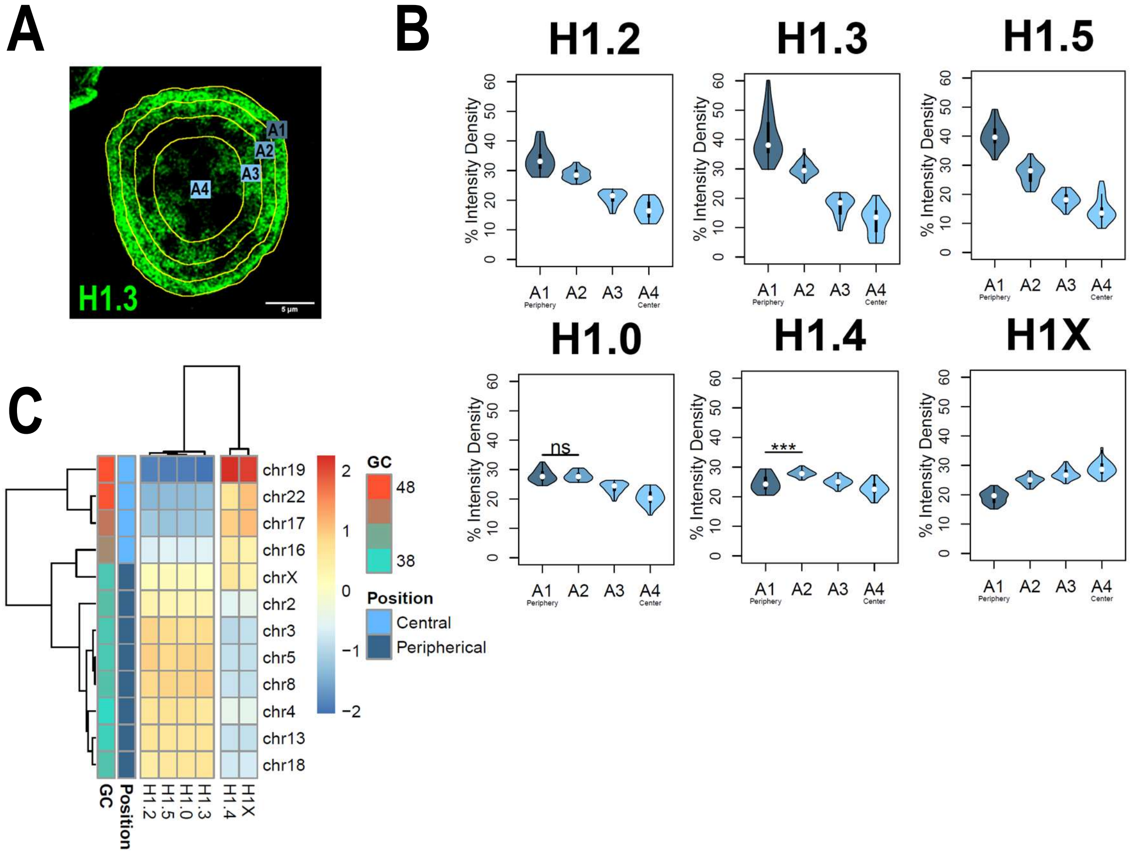
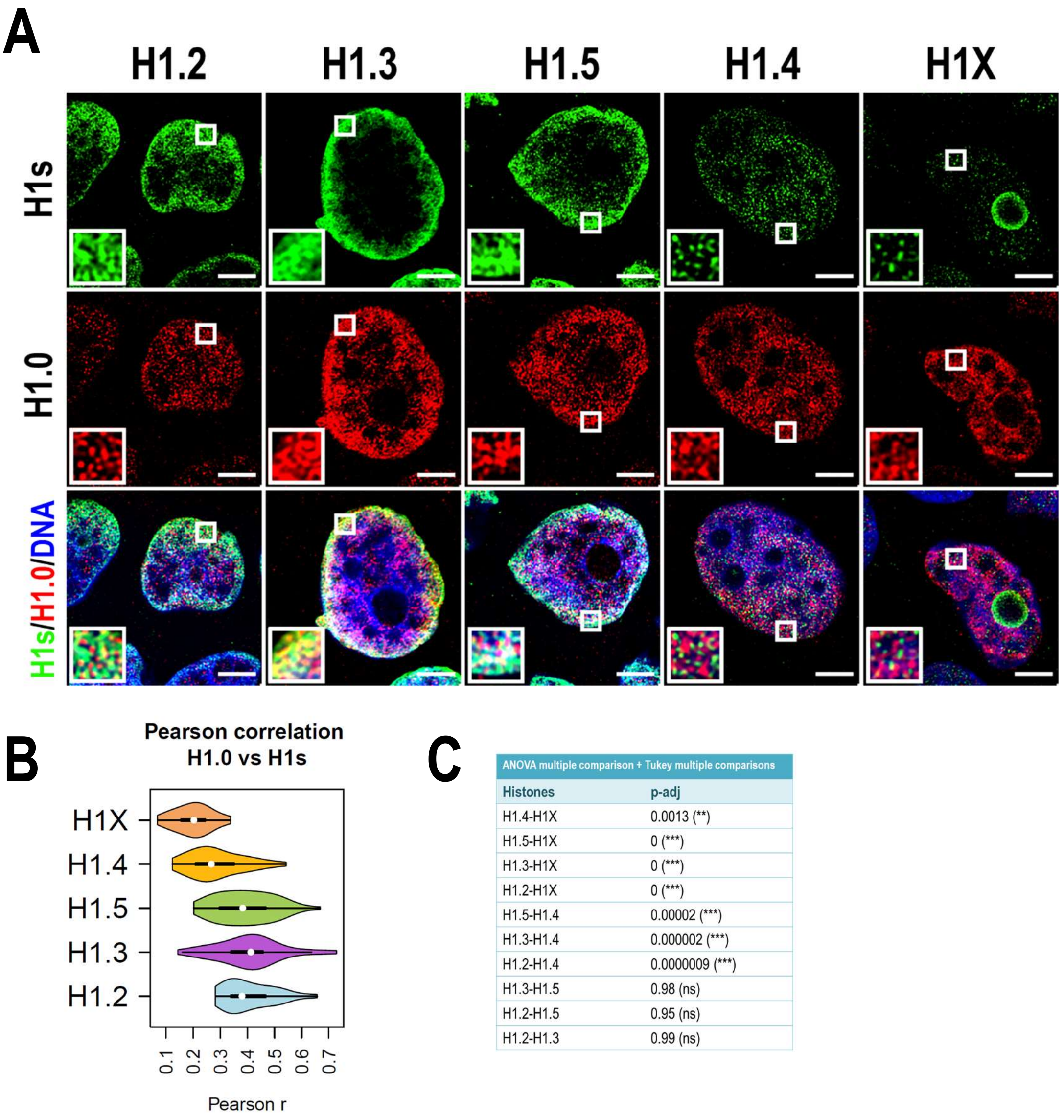


Figure R. 33. H1 variants and chromatin radiality. A) Example of one cell stained with H1.3 antibody in which four sections of an equivalent area and convergent to the nuclear center are shown. Sections are named A1 to A4, from the more peripheral section to the more central one. H1 variants immunofluorescence intensity is measured in each area and expressed as percentage. **B)** Quantifications of H1 variants using this method, where n=30 cells/condition were quantified, and data was represented in violin plots. Statistical differences between A1-A2 for H1,0 and H1.4 are supported by paired t-test (***) p-value<0,001; (ns/non-significant) p-value>0.05. **C)** H1 variants Input-subtracted ChIP-Seq median abundance per chromosome. Y-axis annotation indicated median %GC content per chromosome and their nuclear positions according to (190,191).

Differential H1 'radiality' is related to the spatial organization of chromatin in the nucleus of mammalian cells and the concept of chromosomes territories. Chromosomes are not randomly positioned in the nucleus; gene-poor chromosomes are located at peripheral positions while gene-rich chromosomes tend to occupy central regions (190). We computed H1 variants ChIP-Seq abundance at chromosomes reported to occupy different radial territories **(Figure R.33C)** and, supporting immunofluorescence quantification, we found that H1.2, H1.3, H1.5 and H1.0 were enriched at peripheral chromosomes over central ones. On the contrary, H1.4 and H1X were more abundant at chromosomes located in central positions. Notably, chromosome X showed an increased H1.4 and H1X abundance compared to what would be expected based on its radial position and average %GC content.

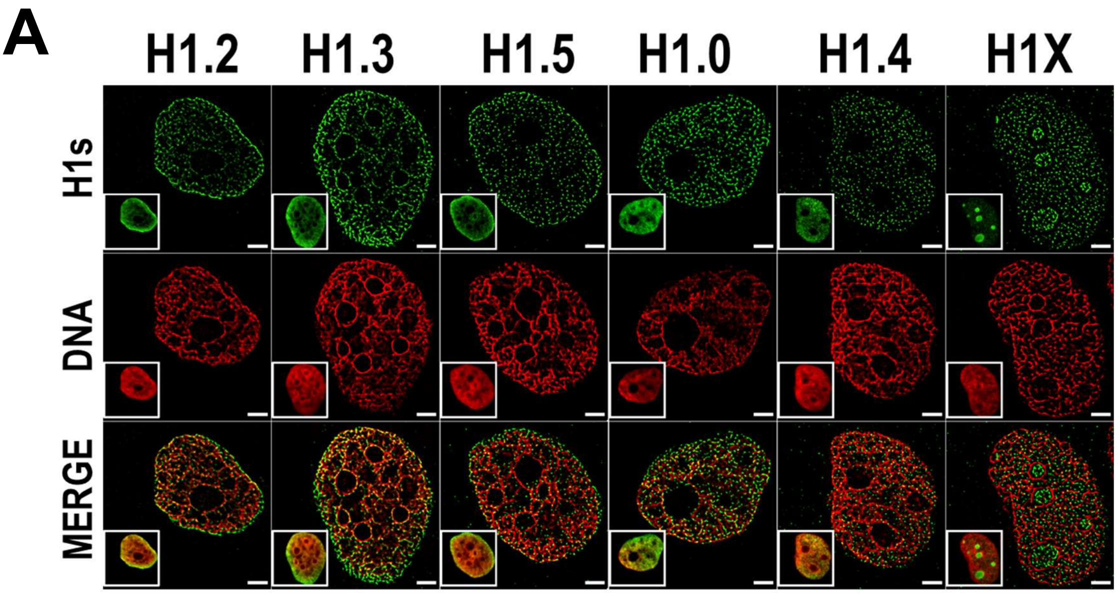


ANOVA multiple comparison + Tukey multiple comparisons	
Histones	**p-adj**
H1.4-H1X	0.0013 (**)
H1.5-H1X	0 (***)
H1.3-H1X	0 (***)
H1.2-H1X	0 (***)
H1.5-H1.4	0.00002 (***)
H1.3-H1.4	0.000002 (***)
H1.2-H1.4	0.0000009 (***)
H1.3-H1.5	0.98 (ns)
H1.2-H1.5	0.95 (ns)
H1.2-H1.3	0.99 (ns)

Figure R. 34. H1 variants co-localization with H1.0 by confocal immunofluorescence. A) Confocal immunofluorescence of the indicated H1 variants (green) with H1.0 (red) and DNA staining (blue). Zoom-in insets highlight H1.0-peripheral enrichment territories. Scale bar: 5mm. **B)** Violin Plots showing the Pearson correlation coefficient (r) distribution of H1 variants with H1.0 in n=40 cells/condition. **C)** Statistical comparison of results shown in B). ANOVA multiple comparison test revealed that significant differences exist between groups. Tukey multiple comparison test was used to compare the H1s-H1.0 r values distribution between different H1s. p-adjusted values are shown (***) p-adj< 0.001; (**) p-adj<0.01; (ns/non-significant) p-adj >0.05).

We next performed co-localization experiments of H1.0 with the rest of H1 variants **(Figure R.34)**. It is worth mentioned that H1.0 antibody is raised in mouse while the rest of H1 variants antibodies are raised in rabbit. So, indirect immunofluorescence experiments could not be performed between other H1 variants combinations. Zoom-in insets in **Figure R.34A** highlight the characteristic peripherical H1.0-enrichment-territories, where H1.0 co-localization was observed with H1.2, H1.3 and H1.5 but not with H1.4 and H1X. Accordingly, Pearson correlation calculation in the whole nuclei revealed that H1.0 intensity better correlated with H1.2/H1.3/H1.5 variants compared to H1.4 and especially H1X, which showed the lowest correlation with H1.0 **(Figure R.34B,C)**.

After analyzing the spatial co-occurrence of H1 variants with DNA signal and the existing co-localization between different variants, we sought to extend these analyses to the super-resolution level. Super-resolution imaging techniques surpass the diffraction limit, enabling visualization of subcellular organization beyond conventional light microscopy resolution (≈250nm). Specifically, we used super-resolution radial fluctuations (SRRF) technique (172,173). SRRF is an analytical approach based on the analysis of radial fluctuations of emitted light from fluorophores, allowing for the reconstruction of super-resolution images (see Methods section 1.12.3 for details).



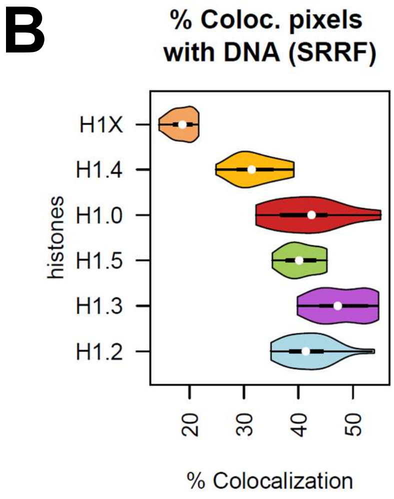
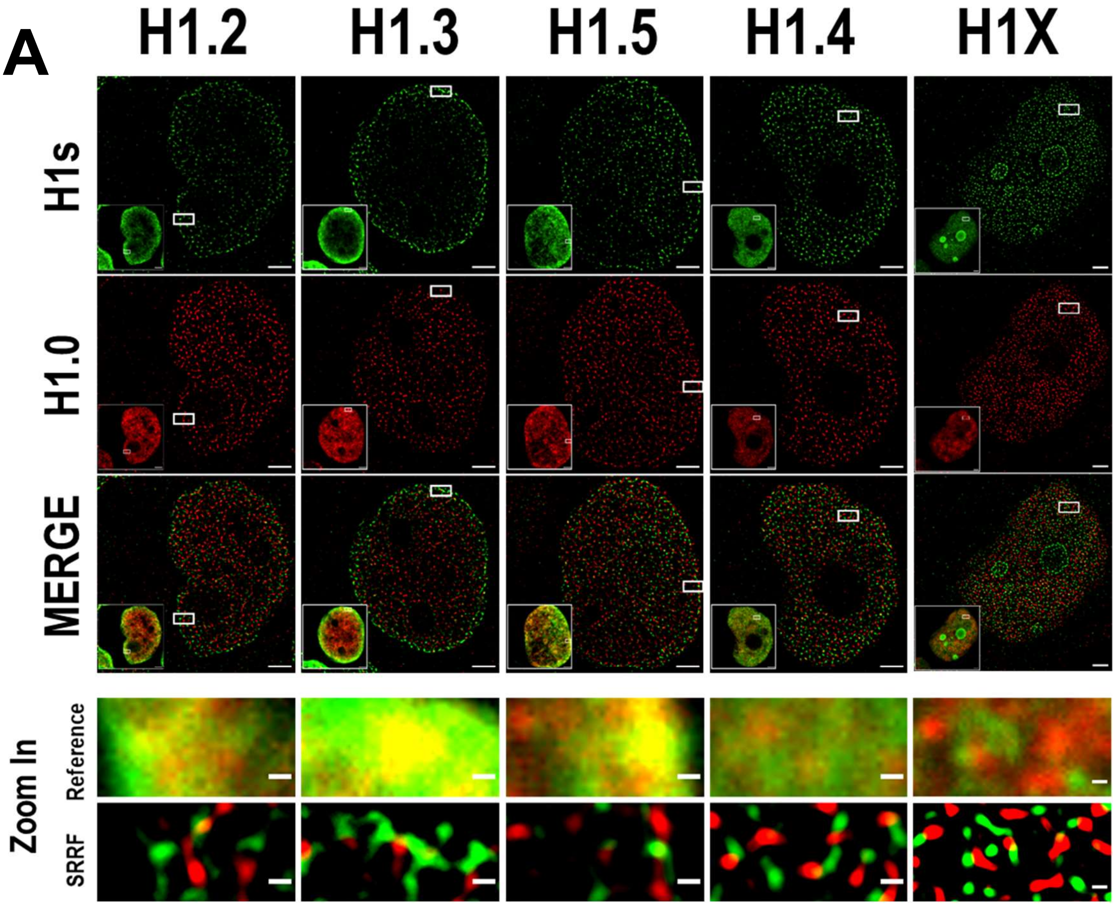


Figure R. 35. H1 variants nuclear distribution by super-resolution microscopy. A) SRRF images of H1 variants (green) and DNA (red). Bottom-left Insets show the Reference confocal image in each case. Scale bar: 2µm. **B)** Percentage of co-localized pixels between H1 variants and DNA by SRRF imaging. n=20 cells/condition were quantified and values distribution were represented as violin plots.

We acquired SRRF images of H1 variants and DNA **(Figure R.35A)**. DNA super-resolution imaging detect DNA in densely packed areas and areas with little or no DNA signal, compared to the typically blurred DNA signal resolved by confocal resolution. Consequently, the DNA we detected through SRRF imaging represents chromatin in a more condensed or closed state, relative to surrounding regions. On the other hand, super-resolution imaging of H1 variants reinforced the different nuclear patterns already seen through confocal microscopy. H1.2, H1.3, H1.5 and to a lesser extent, H1.0 were specially detected at the nuclear periphery, but SRRF imaging emphasized their presence throughout the entire nucleus. SRRF images can be interpreted as a map of probability of a fluorophore (in our case, histone H1 or DNA detection) being located at a specific position, not as linear intensity values. Therefore, even though high intensity values were observed at the nuclear periphery through confocal microscopy, SRRF imaging of H1.2/H1.3/H1.5 highlighted that these variants are not limited to those regions. For its part, H1.4 discrete signals were found throughout the whole nucleus, excluding nucleoli. Conversely, nucleolar H1X was detected and ‘ring’ appearance (see also **Figure R.32, R.34A**) was emphasized. Again, in the case of H1X, super-resolution accentuated the presence of the non-nucleolar H1X-fraction, which was observed throughout the entire nucleus. This information could be misinterpreted at confocal resolution, as high intensity values of nucleolar H1X could lead to an underestimation of its non-nucleolar presence.

The percentage of co-localization of histone H1 variants and DNA signals detected by SRRF was calculated **(Figure R.35B)**. It is worth mentioning that the percentage numbers in **Figure R.35B** must be interpreted comparatively, not as exact values. Both DNA and histone signals are the ones we are detecting under our experimental conditions and specific acquisition settings. If those conditions change, the exact percentage values could vary, but the relative ratio between them should not (if the conditions are the same for all variants). For that, the correct interpretation would be that one cluster of variants co-localizes more with the DNA fraction we are detecting (as previously mentioned, it represents the more condensed DNA) compared to other variants. Taking this into consideration, H1.2, H1.3, H1.5 and H1.0 showed a higher degree of co-localization with DNA compared to H1.4 and H1X, with H1X showing the least colocalization. In summary, our super-resolution co-localization studies of H1 variants with DNA reaffirmed and reinforced the conclusions obtained at a conventional resolution.

We also extended the study of H1.0 co-localization with the rest of H1 variants at super-resolution level **(Figure R.36)**. The zoom-in insets in **Figure R.36A** show some peripheral enrichment-territories of H1.0 at both confocal and SRRF resolutions. Importantly, H1.0 co-localization with H1.2/H1.3/H1.5 observed by confocal microscopy was drastically lost when resolution bypassed the diffraction limit. This became evident by the calculation of the percentage of co-localization **(Figure R.36B)**. Previously observed differences **(Figure R.34B)** were lost and all H1 variants showed ≈25% of co-localization with H1.0. This observation could be underlying the principles of nucleosome composition and 3D chromatin organization. Considering the dissimilarities observed between the standard and super-resolution approaches, results suggest that, in single-cells, at least in more heterochromatic regions where H1.0/H1.2/H1.3/H1.5 are more abundant, H1 variants are not occupying random positions in nearby nucleosomes. If that were the case, co-localization between the different ‘low-GC’ variants (*versus* ‘high-GC’ ones) would not be lost when improving resolution. On the contrary, a compatible model with our results is that, in single cells, heterochromatic 3D-nanodomains tend to be consistently marked by a certain H1 variant.



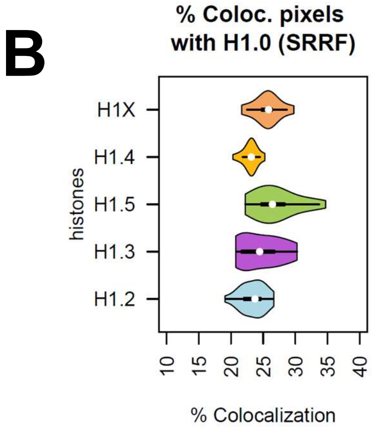


Figure R. 36. H1 variants co-localization with H1.0 by super-resolution microscopy. A) SRRF images of H1 variants (green) and H1.0 (red). Bottom-left Insets show the Reference confocal image in each case. In the bottom panel, the highlighted zoom-in insets at confocal (reference) or SRRF resolutions are shown. Scale bar: 2µm, scale bar in zoom-in insets: 200nm. **B)** Percentage of colocalized pixels of H1 variants with H1.0 by SRRF imaging. n=20 cells/condition were quantified and values distribution were represented as violin plots.

1.4.2. H1 variants distribution patterns along mitosis

As H1 variants showed different nuclear patterns in interphasic cells, we wonder whether these differential patterns were also observed through mitosis.

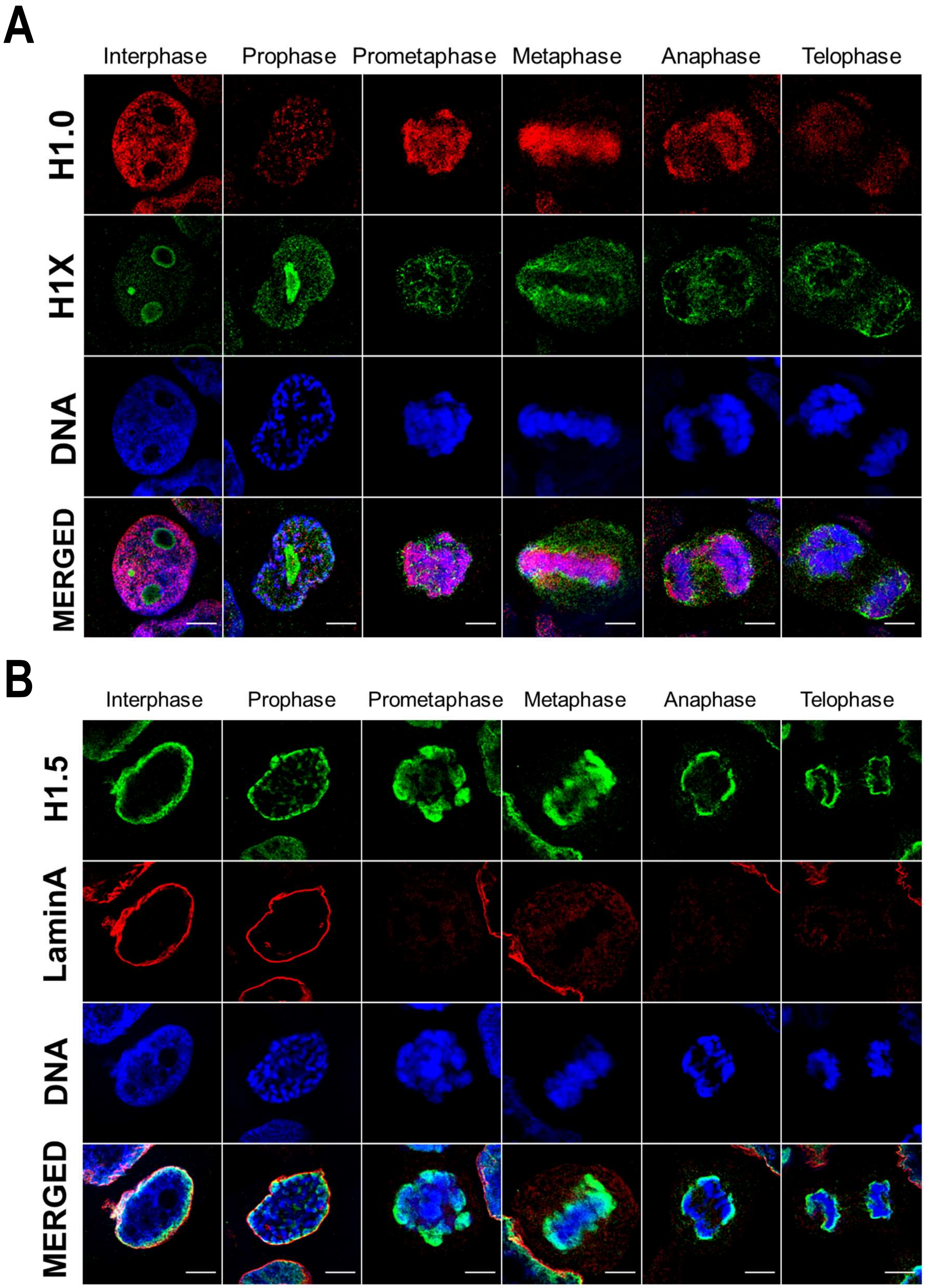
Co-immunostaining of replication-independent H1 variants was performed and their distribution through consecutive phases of mitosis was monitored **(Figure R.37A)**. H1.0 and H1X exhibited completely different distribution patterns. While H1.0 was anchored to mitotic chromosomes, H1X was not recruited to mitotic chromosomes and it accumulated to the perichromosomal region.

H1 variants highly enriched at nuclear periphery during interphase were co-examined with LaminA **(Figure R.37B-D)**. Since nuclear lamina is disassembled during mitosis, we wondered whether the positional information of H1-marked chromatin was maintained through mitosis. To begin with, H1.3 and H1.5 showed similar distribution profiles throughout mitosis **(Figure R.37B,C)**. Both H1 variants persisted at mitotic chromatin, being specially enriched at the periphery of condensed chromosomes. Importantly, H1.3 and H1.5 layers re-associate with the forming nuclear lamina before mitotic exit. Moreover, H1.2 immunofluorescence revealed that H1.2 signal was dispersed after prophase **(Figure R.37D)**. H1.2 was re-detected at anaphase, when its peripheral enrichment was re-acquired, parallel to lamina re-assembly. These results suggest that radial position of H1.2, H1.3 and H1.5-marked chromatin is inherited through mitosis. These genomic regions are re-localized to the nuclear periphery following mitotic division and the nuclear lamina reassembles around H1.2, H1.3 or H1.5-associated chromatin.

The apparent absence of H1.2 at intermediate mitotic stages is striking. Interestingly, a similar scenario was observed for H1.4 *(data not shown)*, as it was found to be depleted at mitotic cells. These observations led us to consider two possible hypotheses: 1) Immunofluorescence signal is not detected because the antibodies are not able to reach metaphasic chromosomes due to extreme chromatin condensation. 2) These H1 variants are not present in mitosis. 3)These H1 variants become highly post-translationally modified during mitosis, and the antibodies do not recognize the H1-modified fraction.

As previously mentioned, other H1s were found anchored to mitotic chromatin **(Figure R.37A)**, so the first and second scenarios related to sterical inaccessibility or protein absence did not seem the most plausible. Considering that H1 proteins are known to be highly phosphorylated during mitosis (153–157,165) and taking advantage of antibodies against concrete H1.2 or H1.4 phosphorylations, we explored the third scenario raised above.

H1.2 Phosphorylation of Threonine 165 (H1.2-pT165) and H1.4 phosphorylation of Threonine 146 (H1.4-pT146) were found to be highly increased in mitosis compared to interphase **(Figure R.38A)**. Thus, this confirms that H1.2 and H1.4 proteins are not absent at certain mitotic phases, but antibodies are unable to recognize their post-translationally modified state. Although these modifications are most prevalent in mitosis, they are also detected in interphasic cells, being specially enriched at nucleoli **(Figure R.38B,C)**, as further analyzed in next sections (see section 1.4.3.2). Early-mitotic H1.2pT165 was found to be associated with condensed chromosomes, with maximum levels occurring at metaphase and drastically dropping down at later mitotic phases **(Figure R.38B)**. These temporal dynamics coincide with the previously discussed lack of H1.2 detection at metaphase and the re-appearance of H1.2 signal at anaphase/telophase **(Figure R.37D)**. In other words, H1.2-pT165 maximum increase was detected at the same time point when H1.2 became undetectable, and the H1.2-pT165 signal decreased when H1.2 signal reappeared. Furthermore, H1.4-pT146 was also more enriched at early mitotic stages, but excluded from metaphasic chromosomes **(Figure R.38C)**. Otherwise, H1.4-pT146 was accumulated adjacent to chromosomes, in the perichromosomal layer, similar to what was observed for H1X **(Figure R.37A)**.


A
Interphase
Prophase
Prometaphase
Metaphase
Anaphase
Telophase
H1.0
H1X
DNA
MERGED
B
Interphase
Prophase
Prometaphase
Metaphase
Anaphase
Telophase
H1.5
LaminA
DNA
MERGED

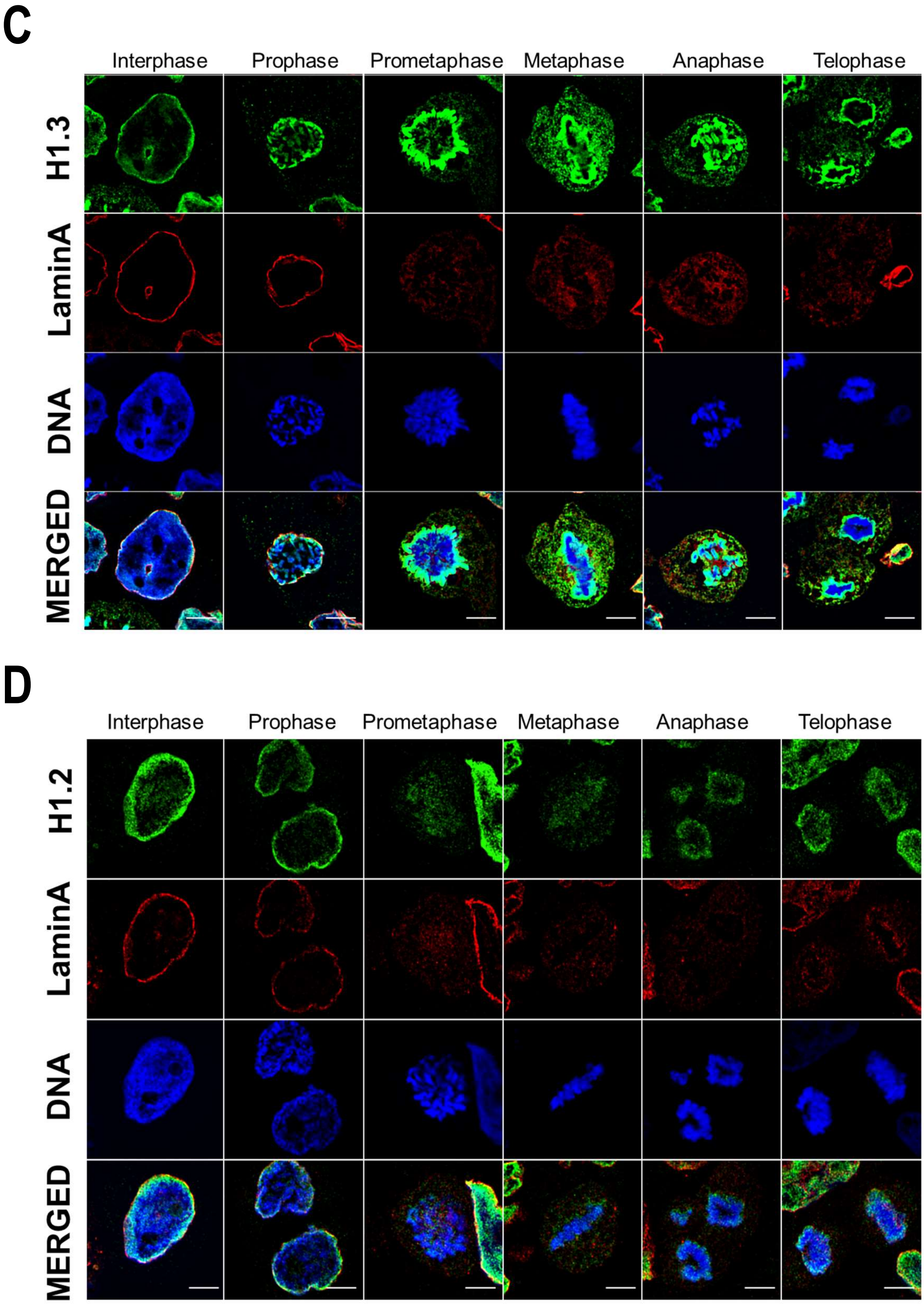
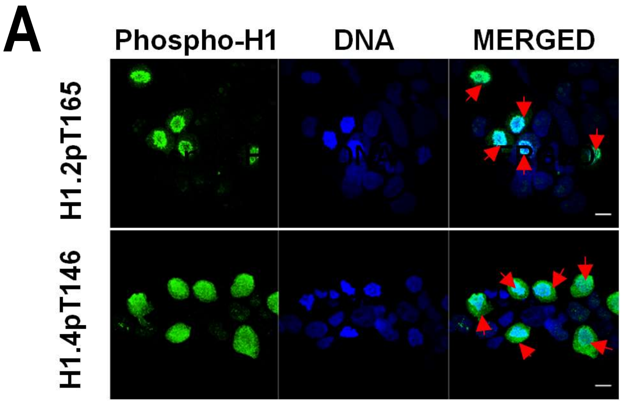


Figure R. 37. H1 variants show differential distribution patterns along mitosis. Immunofluorescence of H1 variants and LaminA along the distinct mitotic phases. In All cases, DNA staining is also shown. **A)** Replication-independent H1.0 and H1X. **B)** H1.5 and LaminA. **C)** H1.3 and LaminA. **D)** H1.2 and LaminA. Scale bar: 5µm



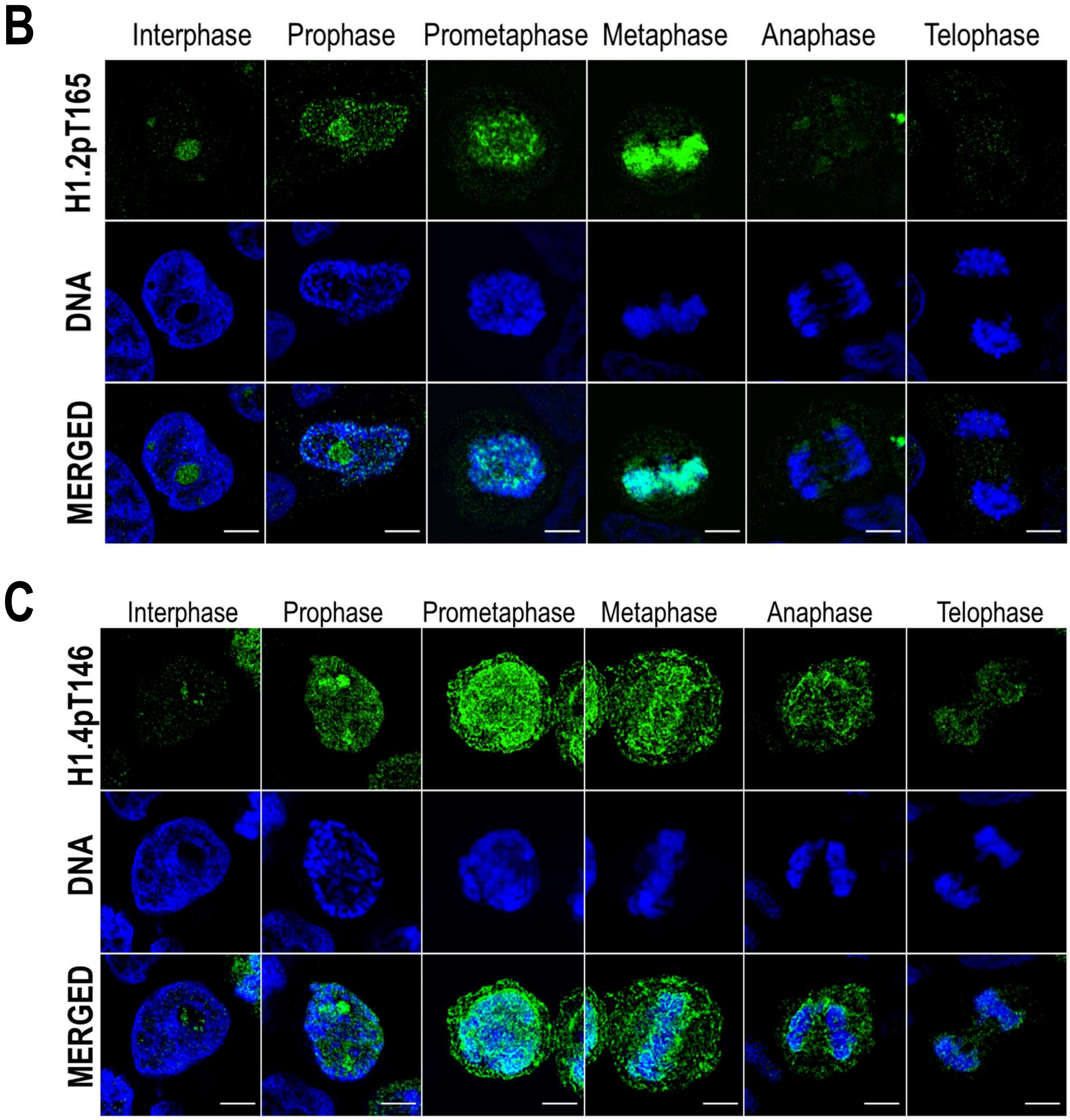


Figure R. 38. H1 is highly phosphorylated in mitosis. A) Immunofluorescence of H1.2-pT165 and H1.4pT146. A Z-stack maximum projection is shown. Mitotic cells are marked by a red arrow. **B)** and **C)** Immunofluorescence of phosphorylated H1.2-pT165 (B) and H1.4-pT146 (C) along the distinct mitotic phases. DNA staining is also shown. Scale bar: 5µm

Overall, analysis of H1 variants during mitosis indicates that the two H1 clusters defined by ChIP-Seq analysis present distinct localization patterns through mitosis. While 'low-GC' H1s (whether phosphorylated or not) are associated with mitotic chromosomes, 'high-GC' variants (phosphorylated or not) are excluded from mitotic chromosomes and accumulate to the perichromosomal region. In addition, H1.3 and H1.5 are enriched towards the peripheral chromosome regions, in comparison to H1.0 or H1.2-pT165. In conclusion, imaging experiments support the differential distribution of H1 variants not only during interphase but also in mitotic cells.

1.4.3. H1 variants distribution within chromatin domains

Euchromatin and heterochromatin are spatially segregated within the nucleus. Genome conformation is also regulated by the tethering of chromatin to scaffold structures, such as the nuclear lamina or nucleolus. In addition to the A/B compartments identified by Hi-C experiments, other domains have been implicated in chromatin organization, such us lamina-associated domains (LADs) or nucleolus-associated domains (NADs). Proper chromatin organization is crucial for genome functionality, so we further explored H1 variants differential distribution within these particular chromatin domains.

1.4.3.1. H1 variants and lamina-associated domains

As H1.2, H1.3 and H1.5 were highly enriched at nuclear periphery (**Figure R.32**) and re-associated to lamina before mitotic exit (**Figure R.37B-D**), we aimed to explore their association with lamina through super-resolution microscopy (**Figure R.39**). Interestingly, in all three cases, the peripheral H1 enrichment seen by confocal microscopy was perfectly resolved by SRRF imaging as an H1 layer adjacent to LaminA layer (see zoom-in insets in **Figure R.39**). These chromatin regions positioned at the nuclear periphery and interacting with nuclear lamin are known as LADs, and they represent a very well-known repressive environment.

To further confirm the enrichment of H1.2/H1.3/H1.5 at LADs, we performed co-immunostaining with H3K9me2, an evolutionarily conserved specific mark of LADs (106). As expected, H3K9me2 was found enriched but not limited to nuclear periphery (**Figure R.40A**). Indeed, H3K9me2 nuclear pattern resembled the distributions of H1.2/H1.3 and H1.5 and a high co-localization was observed at both confocal (**Figure R.40A**) and SRRF resolutions (**Figure R.41**). We next calculated Pearson correlation between H3K9me2 and H1 variants (**Figure R.40B**). H1.2/H1.3 and H1.5 exhibited a strong correlation with H3K9me2. On the contrary, H1.4 and H1X correlation was significantly lower.

These H1 variants and H3K9me2 co-exist within LADs, forming a layer at the nuclear periphery. Notably, co-localization was also observed in other nuclear regions, including perinucleolar chromatin, which represents another heterochromatic environment. This coexistence at nucleolus-associated domains (NADs) was observed both at confocal (**Figure R.40A**) and super-resolution levels (**Figure R.41**). H1 variants abundance at NADs will be analyzed in the next section.

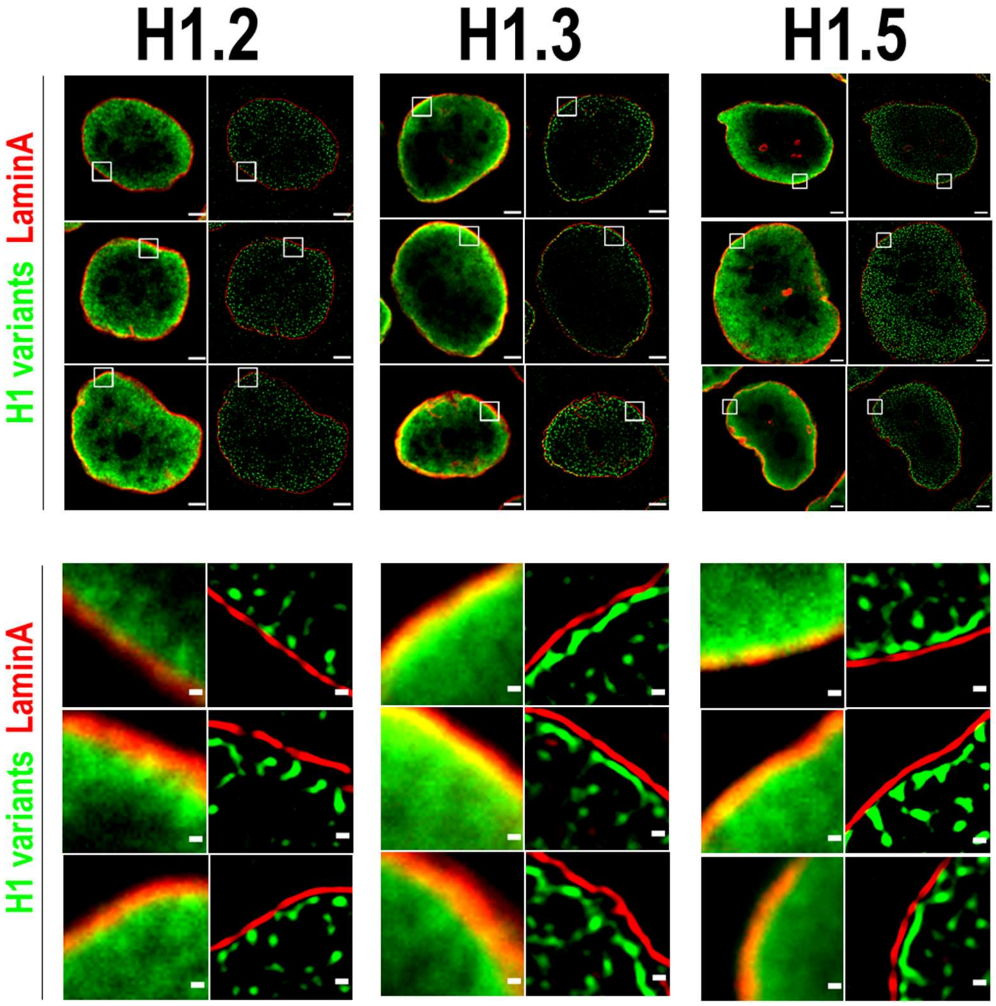
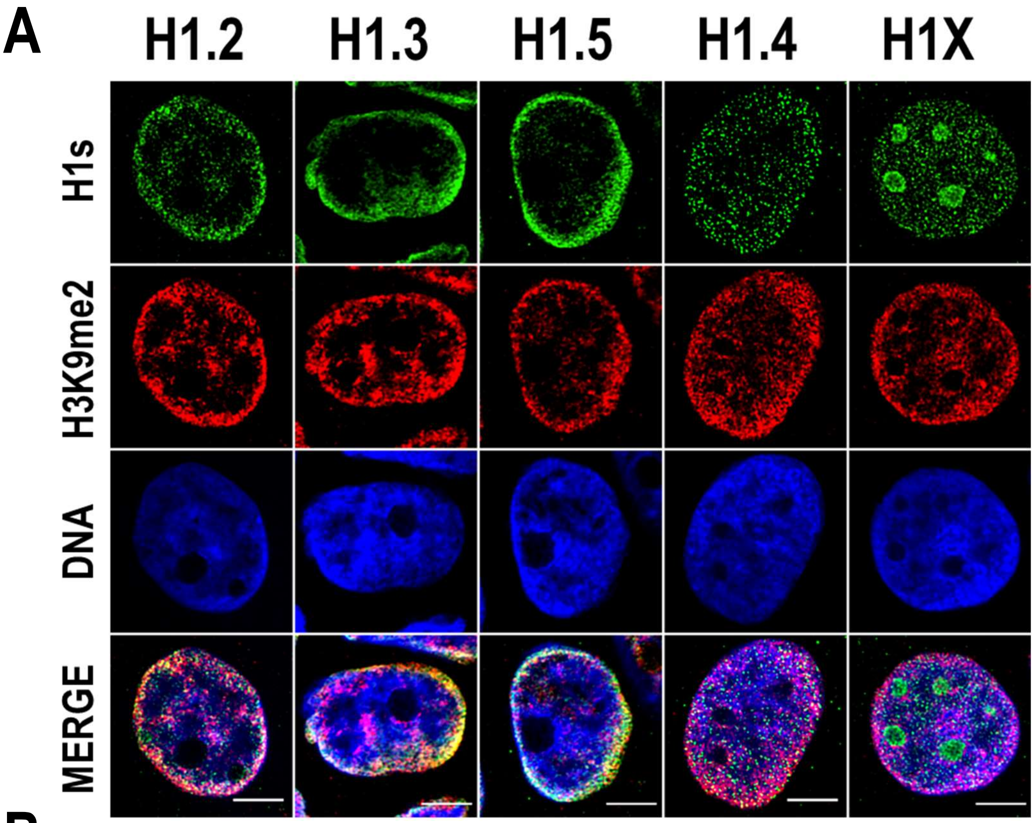


Figure R. 39. H1.2, H1.3 and H1.5 are enriched within LADs. Representative confocal (left) and super-resolution (right) images of a T47D cells stained for H1.2, H1.3 or H1.5 (in green) and Lamin A (in red) obtained using SRRF. Full nuclei (upper panel) and zoomed views of nuclear periphery (bottom panel) are shown. Scale bars: 2 µm (upper panel) and 200nm (bottom panel).

As aforementioned, H1.0 was also found to be partially enriched at nuclear periphery (**Figure R.32-R.33**). We could not perform co-immunostaining of H1.0 with LaminA or H3K9me2 because these antibodies are raised in the same species. Therefore, we studied H1 variants abundance within LADs using ChIP-Seq data. LADs comprise ≈35-40% of the mammalian genome and can be identified using genome-wide approaches. Taking advantage of publicly available LADs coordinates, we computed H1 variants ChIP-Seq abundance in LADs (**Figure R.47**). 'Low-GC' H1 variants were more abundant at LADs compared to regions in A compartment. On the other hand, both H1.4 and H1X were more enriched in the active A compartment regions.

On the whole, we demonstrated that H1.2/H1.3/H1.5, and also H1.0, are constituents of lamina-associated chromatin, as supported by microscopic and ChIP-Seq experiments.



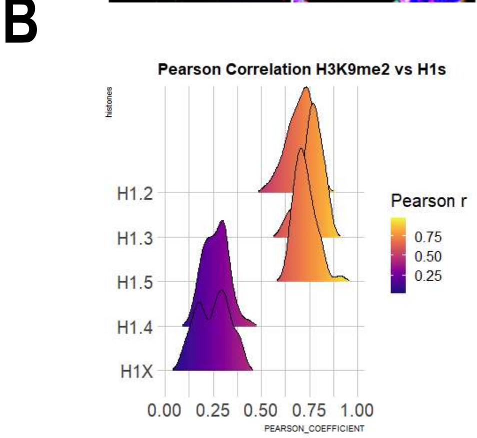


Figure R. 40. H1.2, H1.3 and H1.5 highly co-localize with H3K9me2. Confocal immunofluorescence of H1 variants (green), H3K9me2 (red) and DNA (blue). B) Pearson correlation coefficient (r) of H1 variants and H3K9me2. r values distribution in n=50 cells/condition are shown. Scale bar: 5μm

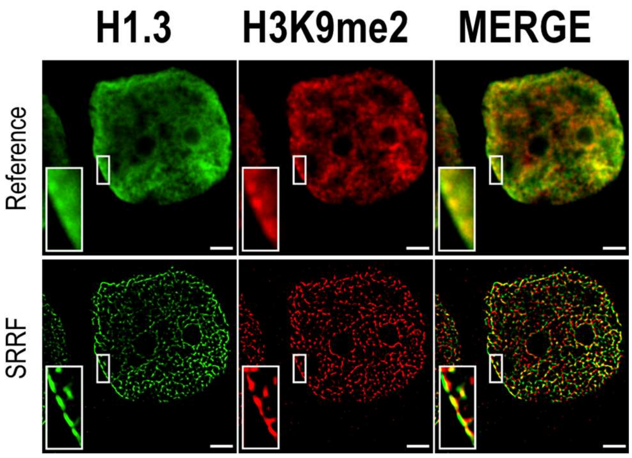


Figure R. 41. H1.3 and H3K9me2 co-localize by super-resolution imaging. H1.3 and H3K9me2 immunofluorescence at confocal (reference) and super-resolution (SRRF) level. A zoon-in inset of the peripheral layer formed by both H1.3 and H3K9me3 is shown. Scaler bar: 2μm

1.4.3.2. H1 variants, nucleolus and nucleolus-associated domains

Nucleolus is a membraneless sub-nuclear compartment where ribosome biogenesis occurs. Together with the nuclear lamina, it represents a hub for heterochromatin organization. Specifically, chromatin regions positioned in close proximity to the nucleolus are known as nucleolus-associated domains (NADs).

We previously identified H1X enriched at nucleolus, using both confocal and super-resolution microscopy **(Figure R.32, R.35)**. To begin with, nucleolar H1X enrichment was found using alternative permeabilization methods *(data not shown)* and both nucleolar and non-nucleolar signals were drastically reduced upon specific H1X depletion **(Figure R.42)**. These data confirm that nucleolar H1X enrichment is not an artifact.

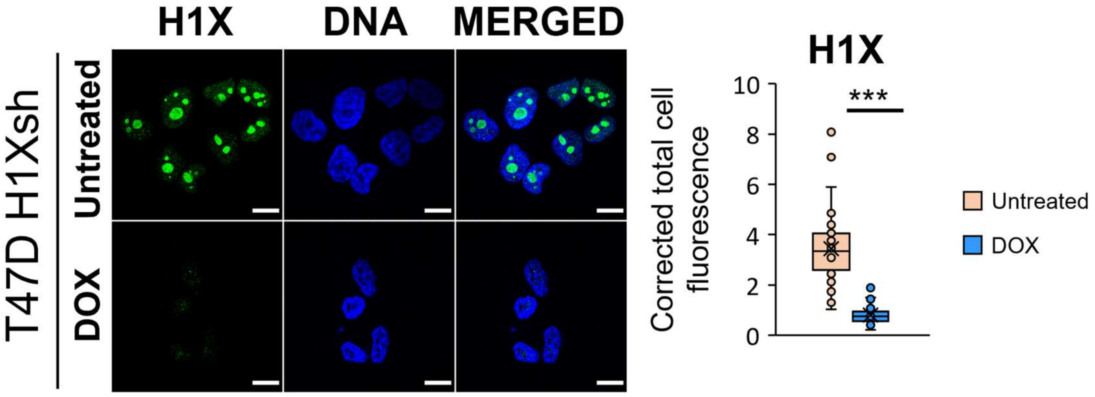


Figure R. 42. H1X immunofluorescence signal is dramatically reduced upon H1X depletion. Immunofluorescence of H1X and DNA in T47D H1Xsh -/+Dox. A Z-stack of five consecutive Z planes is shown. Scale bar: 10mm. H1X immunofluorescence signal quantification (n=41 cells/condition) is shown and supported by paired-t-test. (***) p-value < 0.001

Co-immunostaining of H1X and the nucleolar marker nucleophosmin (NPM1) confirmed that H1X was located inside nucleoli, with a tendency to form a ring-like pattern close to NPM1 **(Figure R.44)**. Indeed, this enrichment near nucleolar rim was better resolved by SRRF, where it became evident that nucleolar H1X formed an adjacent layer to NPM1 on the inner side of nucleoli **(Figure R.43)**.

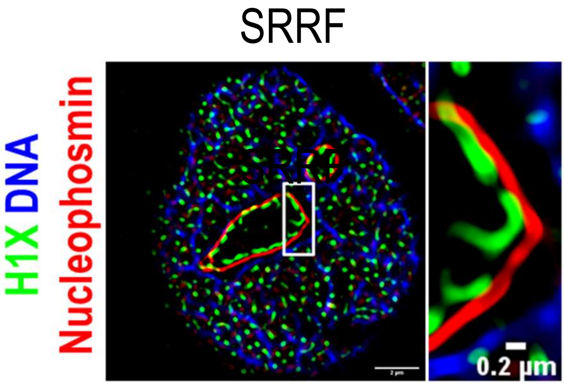


Figure R. 43. H1X forms a nucleolar adjacent layer to Nucleophosmin one. SRRF image of H1X, NPM1 and DNA. Zoom-in highlights the H1X nucleolar layer. Scale bar: 2µm. Scale bar in zoom-in: 0.2µm

Nonetheless, H1X was not the only H1 variant present at nucleoli, as immunofluorescence of H1.2-pT165 and H1.4-pT146 revealed their presence at nucleoli of interphase cells (**Figure R.38B,C**). It is worth mentioning that although these modified histones are predominantly associated to mitosis, they are still present at interphase, exhibiting specific nuclear patterns. Importantly, not all phosphorylated H1 variants are present in the interphasic nucleoli. Phosphorylation in T17 (present in both H1.3 and H1.4) was not found at nucleoli, but enriched to the nuclear periphery *(data not shown)*.

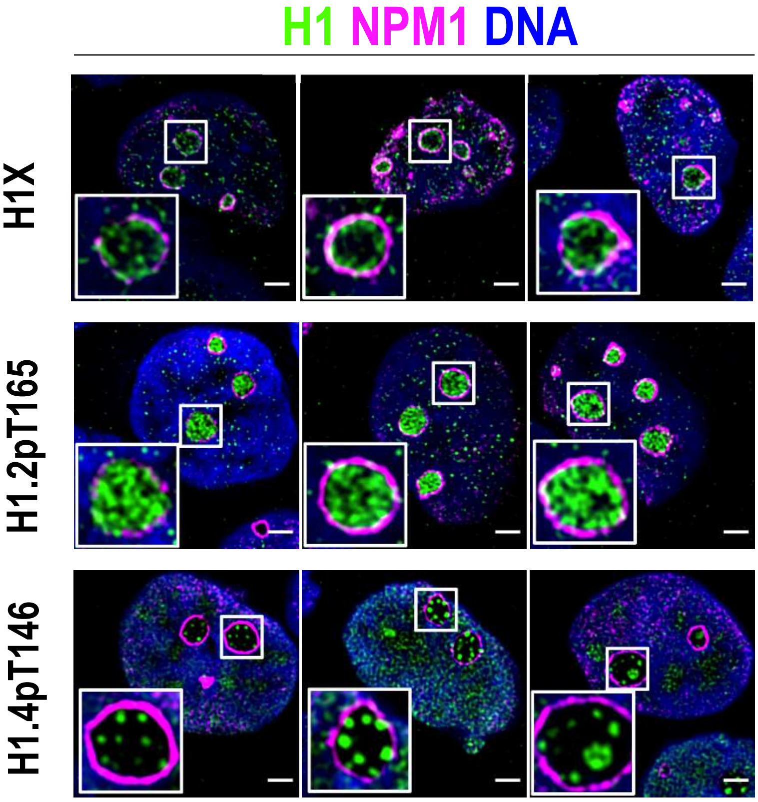


Figure R. 44. H1X is not the only H1 enriched in nucleoli. Immunofluorescence of H1X, H1.2-pT165 or H1.4-pT146, Nucleophosmin (NPM1) and DNA. Insets show a zoom-in of a single nucleolus. Scale bar: 2μm.

To gain insight into their spatial relationship with nucleoli, we performed co-immunostaining of H1.2-pT165 or H1.4-pT146 with NPM1 in asynchronous cells (**Figure R.44**). H1.2-pT165 was highly enriched at the whole nucleoli of most interphase cells. Notably, in some cells, relative enrichment at nucleolar rim, was also detected, similar to H1X. Remarkably, many cells displayed H1.4-pT146 clusters of punctate nucleolar staining, although H1.4-pT146 was not restricted to nucleoli. It was also detected within the rest of the nucleus, specially coinciding with DNA-free staining regions. These characteristic nucleolar dots seem to represent active ribosomal DNA (rDNA) transcription, although to fully confirm this, co-localization with active rDNA transcriptional machinery factors (ie: UBF, RNApol I) would be needed. Importantly, published UBF and RNA pol I immunofluorescences showed the same nucleolar pattern as observed for H1.4-pT146 (192). A second interphasic H1.4-pT146 pattern was observed (**Figure R.45**), with no nucleolar enrichment but speckled enrichment territories along the nucleus that overlap with regions with less DNA staining, i.e., less condensed chromatin. This speckled staining could represent the localization of transcriptionally active chromatin near RNA splicing factories, as has been proposed before for H1.4-pS187, which display an analogous interphasic staining (154).

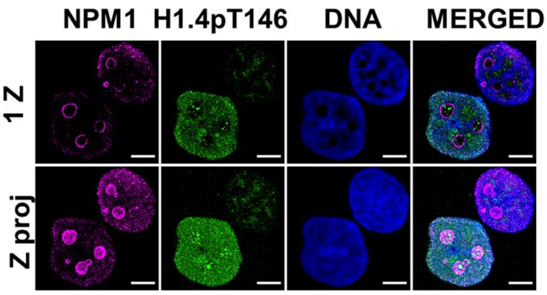


Figure R. 45. Two interphasic patterns of H1.4pT146. Immunofluorescence of NPM1, H1.4-pT146 and DNA in T47D cells. A unique central confocal Z plane (1Z) or the Z maximum projection (Zproj) are shown. Scale bar: 5µm

Next, we tested whether the nucleolar localization of H1 variants depends on nucleolar integrity by rRNA transcription inhibition with Actinomycin D (ActD) treatment (**Figure R.46**). ActD treatment triggers large-scale structural reorganization of the nucleoli, with the migration of some nucleolar markers to the nucleolar-remnant periphery, forming the so-called nucleolar caps while other nucleolar proteins are translocated to nucleoplasm (193). Each nucleolar cap represents UBF-loaded rDNA repeats from a single nucleolar organizer region (NOR). As expected, ActD treatment triggered a total translocation of NPM1 to nucleoplasm and alterations in DNA distribution were also evident. H1.4-pT146 was redistributed to the nucleolar caps, as it would be expected from active rRNA transcription machinery components. Nucleolar enrichment of H1.2-pT165 was completely lost, and it was not detected at nucleolar caps. However, the H1X characteristic nucleolar ring was still found in the remnant nucleoli of a considerable fraction of cells. Thus, these observations might suggest a more structural role of H1X in nucleoli rather than a more functional or regulatory one. On the contrary, H1.2-pT165 and H1.4-pT146 seem to execute a functional role, as the nucleolar distribution of these post-translationally-modified H1 variants depends on functional nucleoli, with H1.4-pT146 being presumably associated to RNA pol I active transcription.

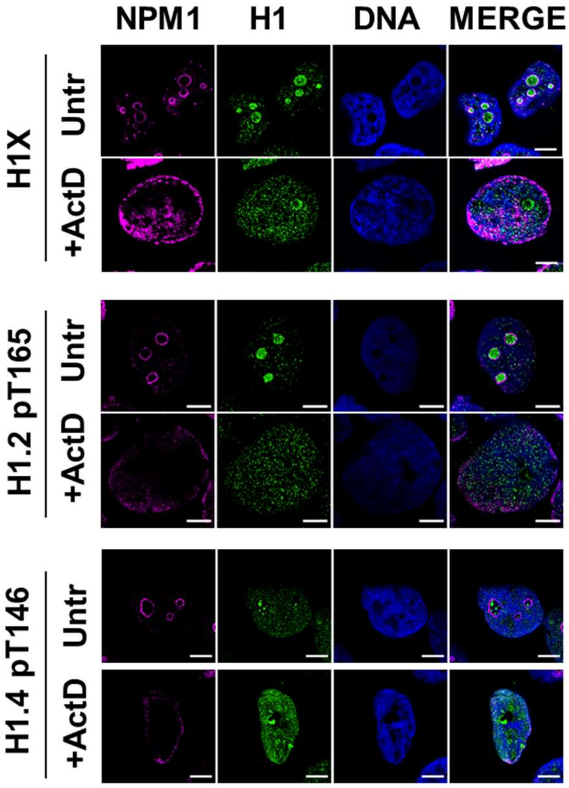


Figure R. 46. Nucleolar H1 variants behavior upon Actinomycin D treatment. Immunofluorescence of H1X, H1.2p-T165 or H1.4-pT146 (green) co-immunostained with Nucleophosmin (magenta) and DNA (blue) under Untreated or Act-D-treated conditions. Scale bar: 5µm

We further explored H1 variants relationship with nucleolar organization by analyzing ChIP-Seq H1 variants abundance within NADs. We used a recently-published NADs mapping performed in HeLa cells, which identified 264 NADs (112). As NADs and LADs show a substantial overlap, we analyzed separately those regions defined exclusively as NAD or LAD and those that overlap (NAD/LAD), similar to the analysis performed in (111). 'Low-GC' H1 variants were enriched within both NADs and LADs, being the NAD/LAD group the one presenting a higher abundance of these variants. On the contrary, 'high-GC' H1 variants are depleted from both NADs and LADs repressive domains, compared to the A compartment **(Figure R.47)**.

Overall, 'low-GC' H1 variants are abundant within NADs, highlighting their presence within multiple repressive compartments in the nucleus and their close association to heterochromatin. Moreover, H1X and phosphorylated-H1.2 or H1.4 exhibit characteristic nucleolar patterns in interphase that could reflect a different functional role in nucleolar dynamics.

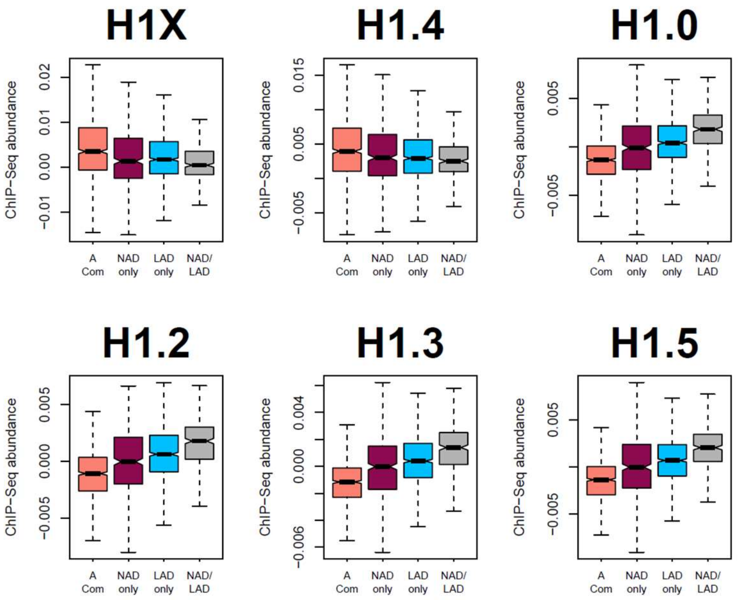


Figure R. 47. H1 variants ChIP-Seq abundance within LADs and NADs. Boxplots show the Input-subtracted H1 variants ChIP-Seq abundance within regions exclusively defined as NADs (NAD only) or LADs (LAD only) and those genomic segments defined as both NADs and LADs (NAD/LAD). A compartment regions are included as a reference. NADs coordinates were extracted from (112).

1.4.4. Chromatin structural changes upon H1 depletion

After investigating the differential nuclear distribution of H1 variants, we examined how H1 depletion affects chromatin organization. We analyzed various H1 depletion conditions, including combined depletion of H1.2 and H1.4 (i.e., multiH1 cells, described in the Introduction), and single depletion of H1.2, H1.4 or H1X. These knock-down (KD) conditions were selected for several reasons. Firstly, multiH1 cells have a severely compromised H1 content, as both H1.2 and H1.4 proteins are depleted. To analyze the individual contribution of each variant, single KDs of H1.2 and H1.4 were also included in the study. Additionally, H1X KD was also analyzed, as it is the least abundant H1 variant in T47D and it shows a more unique distribution pattern. With these KDs, our aim was to analyze the effects of depleting differently genome-wide distributed H1 variants and to compare scenarios with varying percentages of total H1 content reduction.

Using super-resolution imaging of DNA we were able to visualize chromatin structure **(Figure R.48)**. We performed super-resolution imaging of DNA under the previously-mentioned H1 KD conditions to evaluate changes in chromatin structure **(Figure R.49A)**. MultiH1 KD led to a general disruption of chromatin organization compared to control conditions. This disruption was also evident, albeit to a lesser extent, upon single depletion of H1.2. In contrast, single depletion of H1.4 or H1X did not appear to produce changes in chromatin architecture at the level studied.

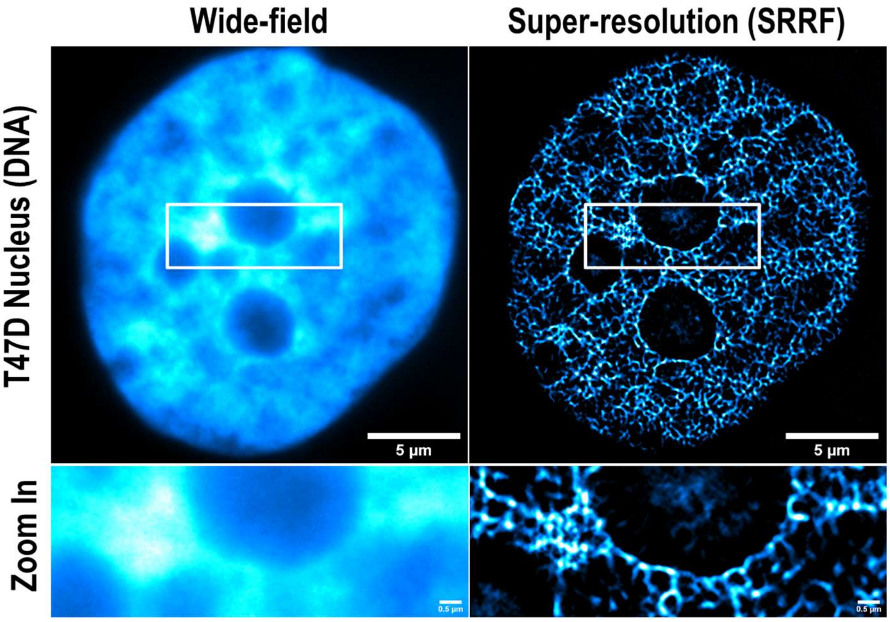


Figure R. 48. SRRF imaging is useful to evaluate chromatin structure. Representative images of DNA staining visualized with wide-field or super-resolution (SRRF) microscopy in a T47D nucleus. Bottom panels show a zoom-in of the perinucleolar region. Scale bar: 5µm (full nucleus) or 500nm (Zoom-in insets).

To quantify chromatin those structural changes upon H1 depletion, we used DNA-free areas analysis, as reported elsewhere (174,175). Methodology details are explained in the Material and Methods section. Briefly, under standard or control conditions, DNA signal becomes redistributed and accumulates in densely packed areas. This leads to the appearance of areas with no DNA signal or low-density DNA signal (referred to as DNA-free areas). This is evident in **Figure R.48** and Untreated conditions in **(Figure R.49A)**. Considering this, upon DNA decompaction, a decrease in the percentage of DNA-free areas is expected. We quantified DNA-free areas under the different H1 KD conditions **(Figure R.49B)**. Thus, the percentage of DNA-free areas was strongly reduced upon multiH1 KD, supporting the apparent DNA decompaction seen in the images **(Figure R.49B)**. Similarly, H1.2 depletion also led to a decreased percentage of DNA-free areas, although the reduction was minor compared to multiH1 KD. Depletion of H1.4 or H1X did not lead to significant changes in % DNA-free areas. Furthermore, **Figure R.49C** shows DNA-free areas quantification per cell.

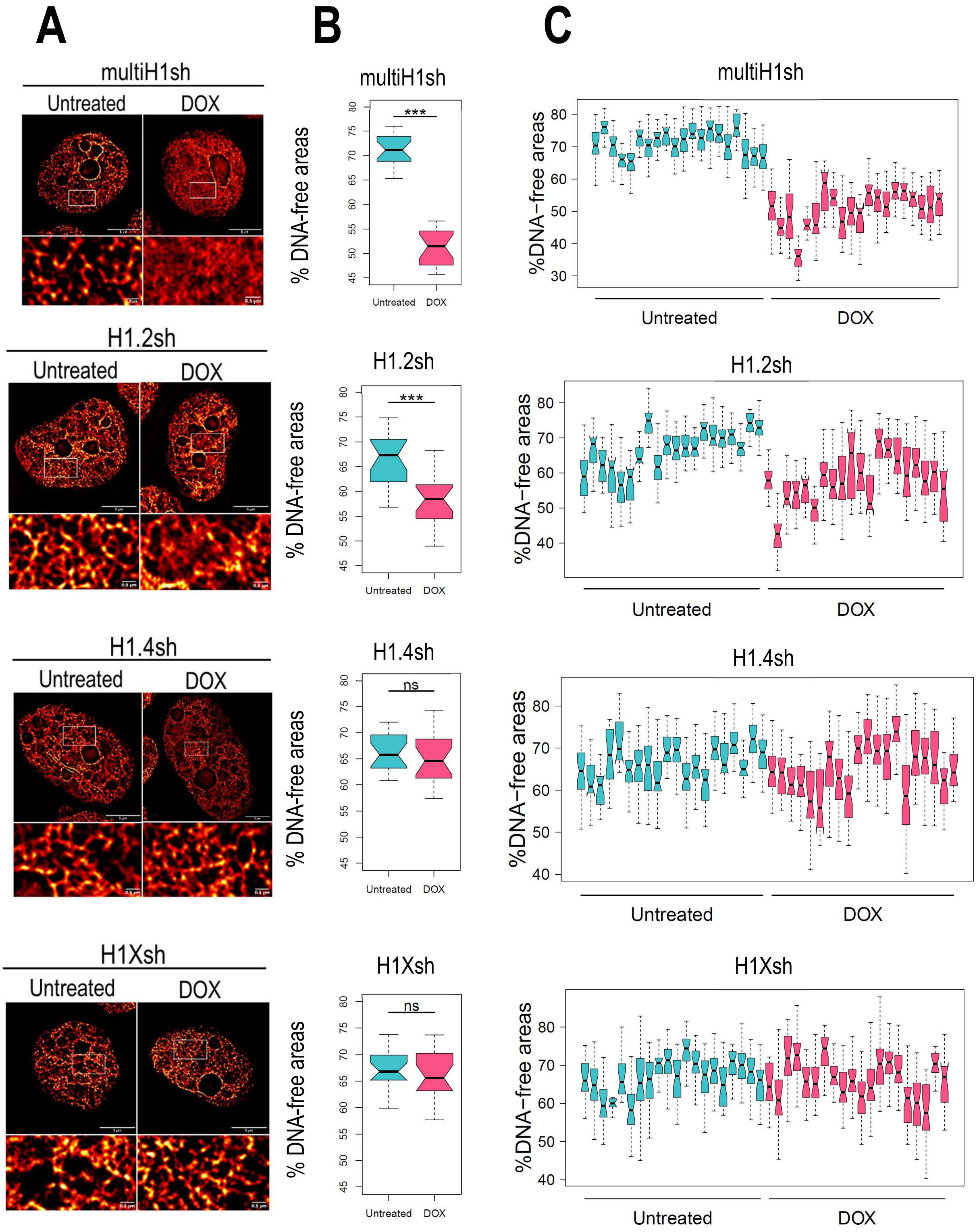


Figure R.49. Multiple H1 depletion and single H1.2 depletion lead to chromatin decompaction. A) Representative SRRF images of DNA staining in the different H1 KD conditions. In the bottom pannels, a zoom-in inset is shown to appreciate DNA pattern in both Untreated and Dox conditions. Scale bar: 5µm (full nucleus) and 500nm (zoom-in). **B)** DNA-free areas percentage quantification in the different H1 KDs. n=20 cells/condition were quantified and the boxplot were constructed with the 20 average values in each condition. **C)** Quantification of % DNA-free areas per cell. % DNA-free areas is calculated in small regions all over the cell (see Methods section 1.12.4), that is why in (C) a distribution of values per cell is represented. Statistical differences between Untreated and Dox-treated conditions are supported by paired-t-test. (***) p-value < 0.001; (ns/non-significant) p-value > 0.05.

In conclusion, combined depletion of H1.2 and H1.4 but also single depletion of H1.2 have an impact on chromatin structure, leading to a general chromatin decompaction that, on the contrary, is not observed upon H1.4 or H1X depletion.

Additionally, we performed immunofluorescence of LaminA in different H1 KDs **(Figure R.50A)**. Interestingly, multi H1 KD cells showed an altered Lamin A pattern **(Figure R.50A)** and also an irregular nuclear morphology **(Figure R.50B)**. Moreover, lamin filaments structure was also disturbed upon H1.2 or H1.4 single depletion, but remain intact upon depletion of H1X and also in control cells (i.e. Randomsh). Although these results could also suggest that multiH1, H1.2 and H1.4 KD present defects on microtubule polymerization or cytoskeleton structure, we have not explored this possibility.

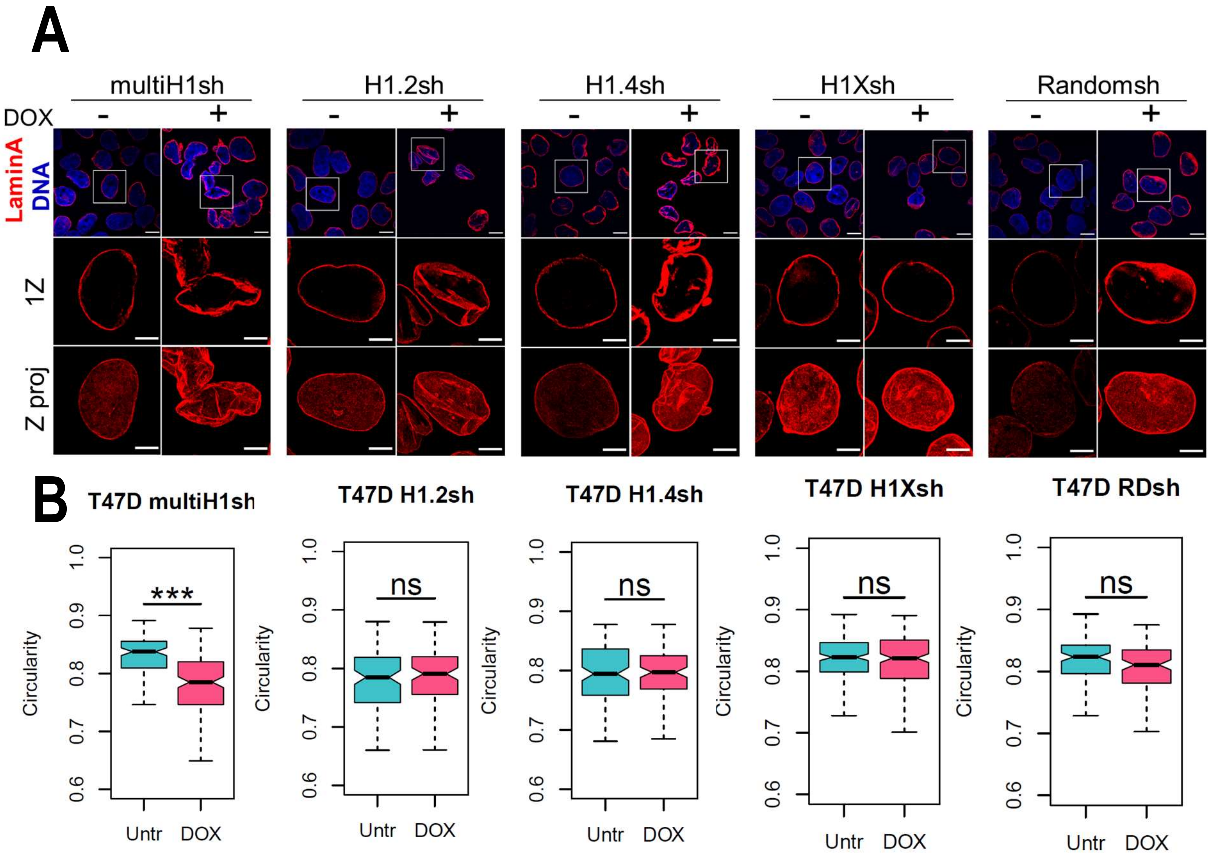


Figure R. 50. Multiple H1 and single H1.2 or H1.4 depletion lead to defects in Lamin structure. A) Immunofluorescence of LaminA in different H1 KD conditions. Per each condition several cells are shown and a single cell is zoom-in, shown as one single confocal Z plane or as the maximum projection oz consecutive Z planes (Zproj), to better appreciate lamin net structure. Scale bar in the upper images showing several cells: 10 µm. Scale bar in single-cell images: 5µm **B)** Boxplots show the nuclei circularity upon the different H1 KDs. Circularity values range from 0-1, being circularity=1 a perfect circle. Circularity was evaluated in n=100-200 nuclei in each condition.

1.5. H1 variants re-distribution upon multiple H1 depletion

Combined depletion of H1.2 and H1.4 (i.e. multiH1 KD cells) was reported to trigger an innate immune response via activation of diverse repetitive elements (126), as detailed in section 2.2.3 in the Introduction. However, it was an unsolved question whether remaining H1 variants redistributed or not to compensate for H1.2 and H1.4 losses due to the lack of specific ChIP-grade antibodies at that time. For this reason, we explored these putative compensatory mechanisms performing ChIP-Seq of the endogenous variants in multiH1 KD cells. In parallel, H3K9me3 was also profiled. Of note, endogenous H1.3 mapping in multiH1 KD was not performed, due to the unavailability of the antibody in our lab when this part of the project was developed.

First, efficacy of multiH1 KD was tested on chromatin samples by WB **(Figure R.51A)**. As expected, multiH1 KD led to H1.2 and H1.4 depletion at the protein level. As a consequence, replication-independent H1 variants (i.e. H1.0 and H1X), especially H1.0, became upregulated, attempting to compensate H1 global reduction. We then perform ChIP with antibodies against endogenous H1.0, H1.2, H1.4, H1.5 and H1X. ChIP-qPCR was performed at Distal and TSS regions of *NANOG* and *CDK2* genes **(Figure R.51B)**. All ChIPs worked efficiently compared to unspecific IgG control. As expected, an H1 valley at the active gene *CDK2* TSS (compared to distal regions) was observed. Upon multiH1 KD condition, IPed DNA with H1.2 and H1.4 antibodies significantly decreased, while H1.0 IPed material increased.

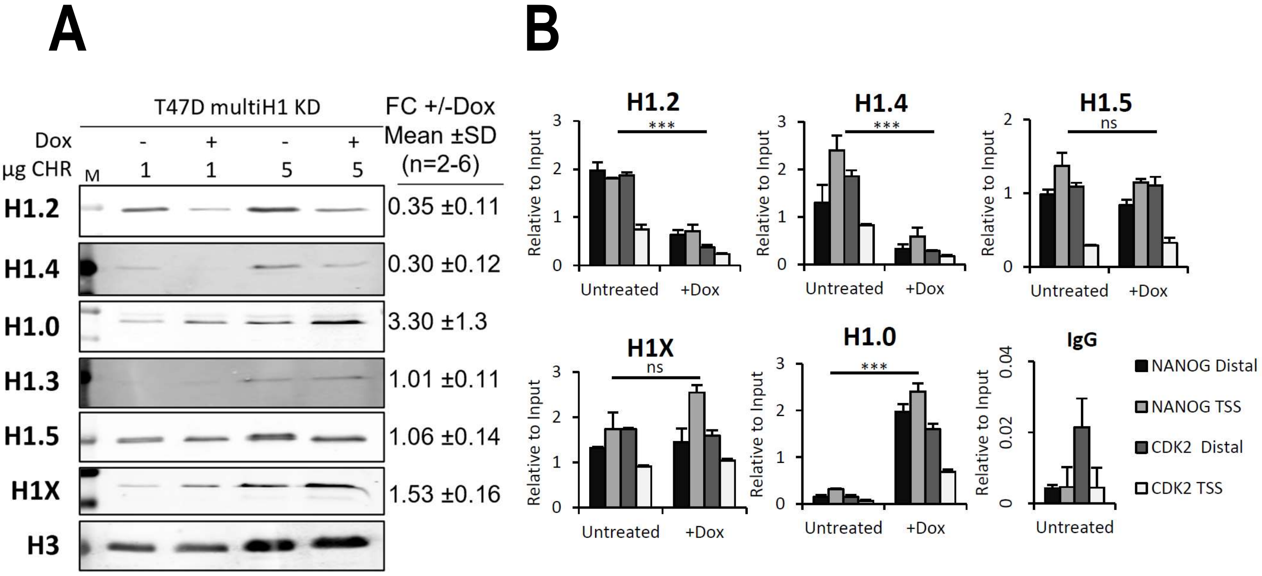


Figure R. 51. H1 variants protein changes in T47D multiH1 KD. A) Immunoblot analysis of H1 variants in chromatin extracts (1 or 5µg) from T47D multiH1sh Untreated or Dox-treated. Histone H3 was added as a loading control. ImageJ immunoblot quantification of multiple experiments is indicated as mean (ratio Dox/untreated) and SD. Number of biological replicates used for quantification were: n = 6 (H1.2, H1.4 and H1.0), n = 4 (H1X), n = 2 (H1.3, H1.5). **B)** ChIP-qPCR of endogenous H1 variants in T47D multiH1sh -/+Dox cells. Unspecific IgG was added as a negative control. ChIP amplification is shown relative to Input DNA amplification. Statistical differences between Untreated and +Dox immunoprecipitated DNA for each H1 variant are supported by paired-t-test. (***) $p < 0.001$; (ns/non-significant) $p > 0.05$.

We proceeded with high-throughput sequencing to analyze putative genome-wide distribution changes upon multiH1 KD. Consideration of how protein levels change in response to multiH1 KD **(Figure R.51)** is essential for interpreting the H1 variants ChIP-Seq data obtained in multiH1 KD versus basal conditions (i.e. multiH1sh-/+Dox). For H1.0, whose protein levels increase upon multiH1 KD, ChIP-Seq data would reveal where it is being incorporated into chromatin. ChIP-Seq data of H1.5 and H1X, whose protein levels do not change or increase slightly, respectively, will allow us to decipher whether these variants re-distribute upon H1 content reduction. It is worth mentioning that, although H1.2 and H1.4 are strongly depleted, they can still be immunoprecipitated. For this reason, we also performed ChIP-Seq of H1.2 and H1.4 in Dox-treated-cells, in order to evaluate the genomic distribution of the remaining H1.2 and H1.4 proteins. Noteworthy, two scenarios are possible: they can be homogenously depleted all over the genome or, on the contrary, being selectively depleted from certain regions.

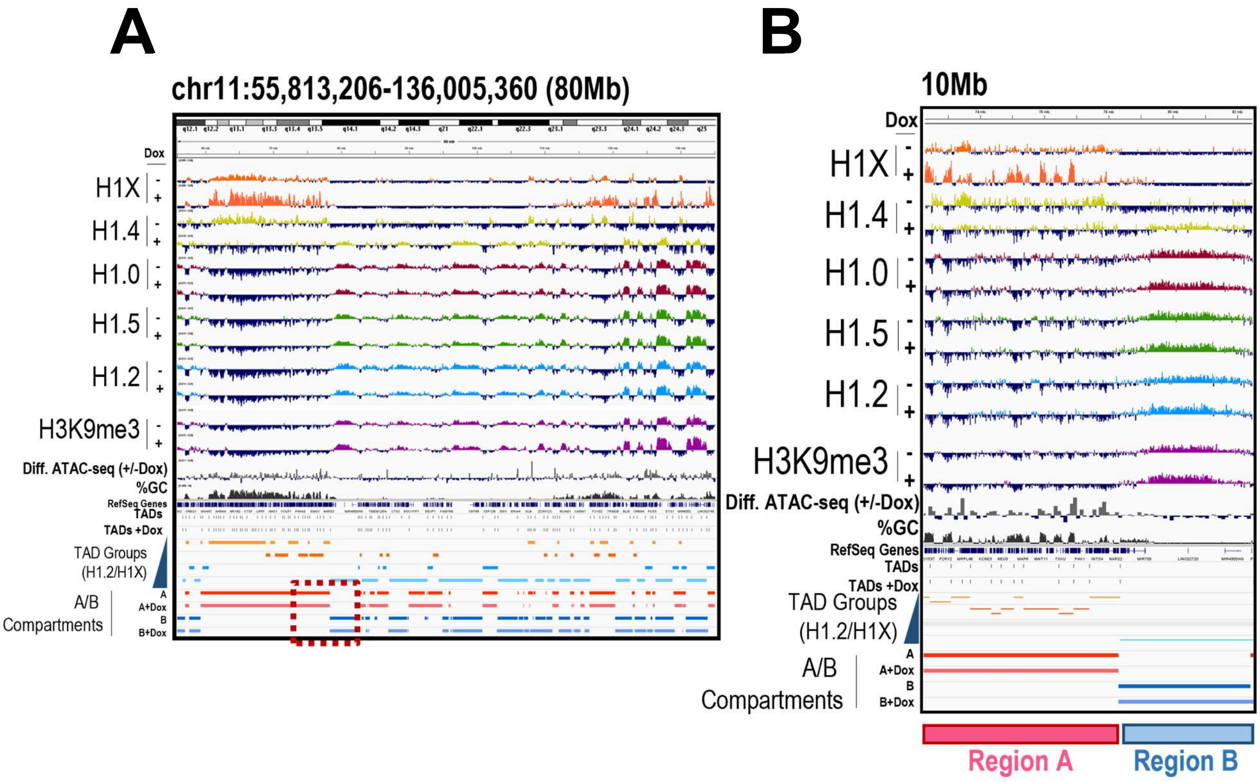


Figure R. 52. Chromatin landscape in T47D multiH1 KD. A) IGV browser screenshots of human chromosome 11 (partial) showing input-subtracted ChIP-seq abundance of H1 variants and H3K9me3 in T47D breast cancer cells under basal (-Dox) or multiple H1 depletion (+Dox). Increased chromatin accessibility upon H1 depletion is shown by differential ATAC-Seq signal (+/- Dox). TADs (TAD borders and TADs classified in four groups according to their H1.2/H1X ratio) and A/B compartments derived from Hi-C experiments under the same conditions are also shown. **B)** Zoom-in of the 10-Mb region highlighted in red in (A).

ChIP-Seq data inspection in a genome browser denoted that H1.2, H1.5 and H1.0 did not show significant distribution changes upon multiH1 KD at the megabase level **(Figure R.52A,B)**. On the other hand, H1X increased at high-GC and A compartment regions. Remaining H1.4 in multiH1 KD was enriched at low-GC and B compartment, showing an opposite profile to basal H1.4 in T47D.

We evaluated H1 variants ChIP-Seq abundance within 100-kb genomic bins (**Figure R.53**). Upon multiH1 KD, H1.2, H1.0 and H1.5 did not significantly change their relative genomic distribution, whereas H1X increased at high-GC regions. H1.4 was relatively enriched towards low-GC regions in Dox-treated cells. These results indicate that while H1.2 is being homogeneously depleted throughout the genome, H1.4 is preferentially depleted from high-GC regions, and remaining H1.4 stays at low-GC regions. H1.0, whose expression and protein levels increased, is incorporated at low-GC regions that already occupied under basal conditions.

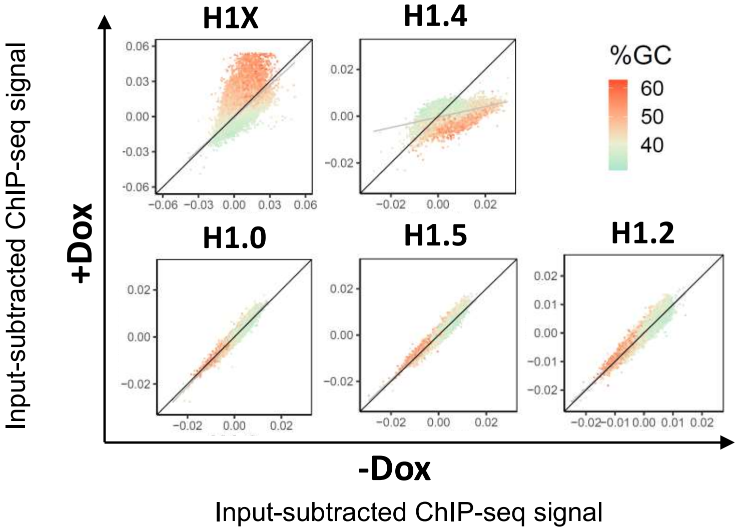


Figure R. 53. H1 variants redistribution upon multiH1 KD in GC bins. Scatter plots of H1 variants input-subtracted ChIP-seq abundance within 100-kb genomic bins in multiH1 KD cells treated or not with Doxycycline. The GC content at each bin is color-coded.

H1 variants re-distribution was also evaluated at the level of chromatin states (**Figure R.54**). Notably, **Figure R.54A** shows the median ChIP-Seq abundance at each chromatin state. **Figure R.54B** shows a random selection of 50 groups, each composed by 20 random segments per each chromatin state. This analysis was performed to confirm re-distributions upon multiH1 KD, as each chromatin state is composed by a variable number of segments (see **Figure M.9A** in Methods).

Thereby, the analysis confirmed that H1.2, H1.5 and H1.0 did not show significant re-distribution changes upon multiH1 KD. On the contrary, H1.4 clustered with 'low-GC' H1 variants in the multiH1 KD condition, confirming their preferential depletion from high-GC regions observed before. Upon multiH1 KD, H1X profile became even more divergent from the other H1 variants, exhibiting a relative decrease in *heterochromatin* state and an increase in almost the rest of chromatin states, but particularly at *polycomb-repressed* and promoters. The highest H1X abundance increase was found at *inactive promoters*. H1X profile around TSS of genes grouped by their basal expression confirmed that, upon multiH1 KD, H1X generally increased at TSS, specially of low-expressed genes (**Figure R.55**).

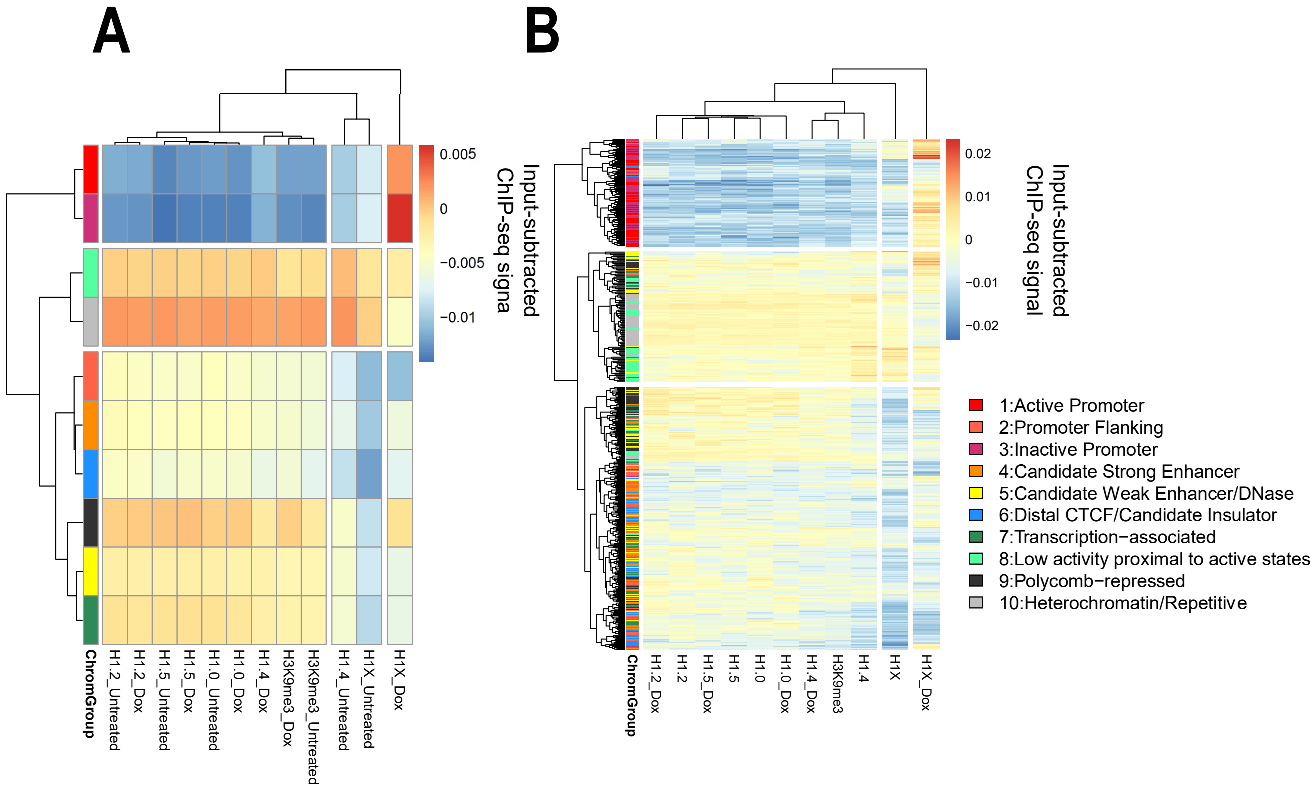


Figure R. 54. H1 variants redistribution upon multiH1 KD in chromatin states. Heatmap and cluster analysis of the Input-subtracted ChIP-Seq abundance of H1 variants within Chromatin states in T47D multiH1 KD cells. **A)** median abundance per chromatin state. **B)** Random sample of genome fragments belonging to the 10 chromatin states. For each chromatin state 1000 fragments were randomly picked and 50 groups of 20 fragments were randomly generated. Each lane of the heat map represents the median input-subtracted ChIP-seq abundance of H1 variants in a group.

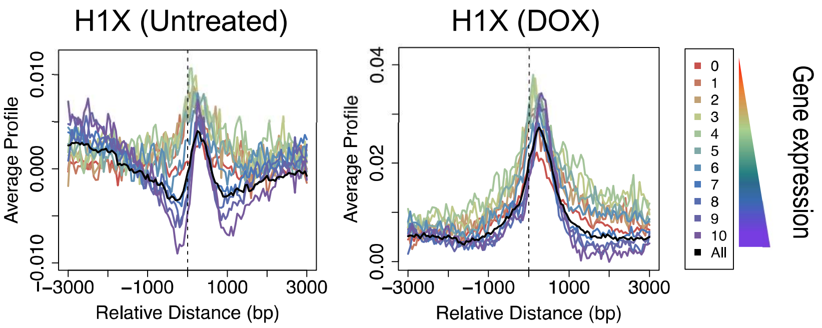


Figure R. 55. H1X increases at promoters upon multiH1 KD. H1X Input-subtracted ChIP-Seq average profile around gene transcriptional start site (TSS) in T47D multiH1sh Untreated and Dox-treated cells. Expressed genes are divided in 10 equal groups (Groups 1-10, each containing 10% of total expressed genes) according to their basal gene expression in T47D RNA-Seq experiments. Group 0 includes non-expressed genes. Average H1 profile for all genes is shown in black.

Furthermore, we explored H1 variants re-distribution upon multi H1 KD by immunofluorescence **(Figure R.56)**. As expected, both H1.2 and H1.4 signals drastically decreased. For that reason, detection of their remaining fractions is not possible with this technique. H1.0 but also H1X signals increased but no relative re-distribution was observed at the nuclear level. Notably, H1X was still present within nucleoli upon Dox-treatment. H1.5, whose levels did not significantly change, did not show any re-distribution. Overall, imaging experiments support ChIP-Seq data, as no drastic re-distributions are detected. Re-distribution of H1X towards higher-GC regions can only be observed at ChIP-Seq level and does not imply a significant alteration of the H1X nuclear pattern seen by immunofluorescence.

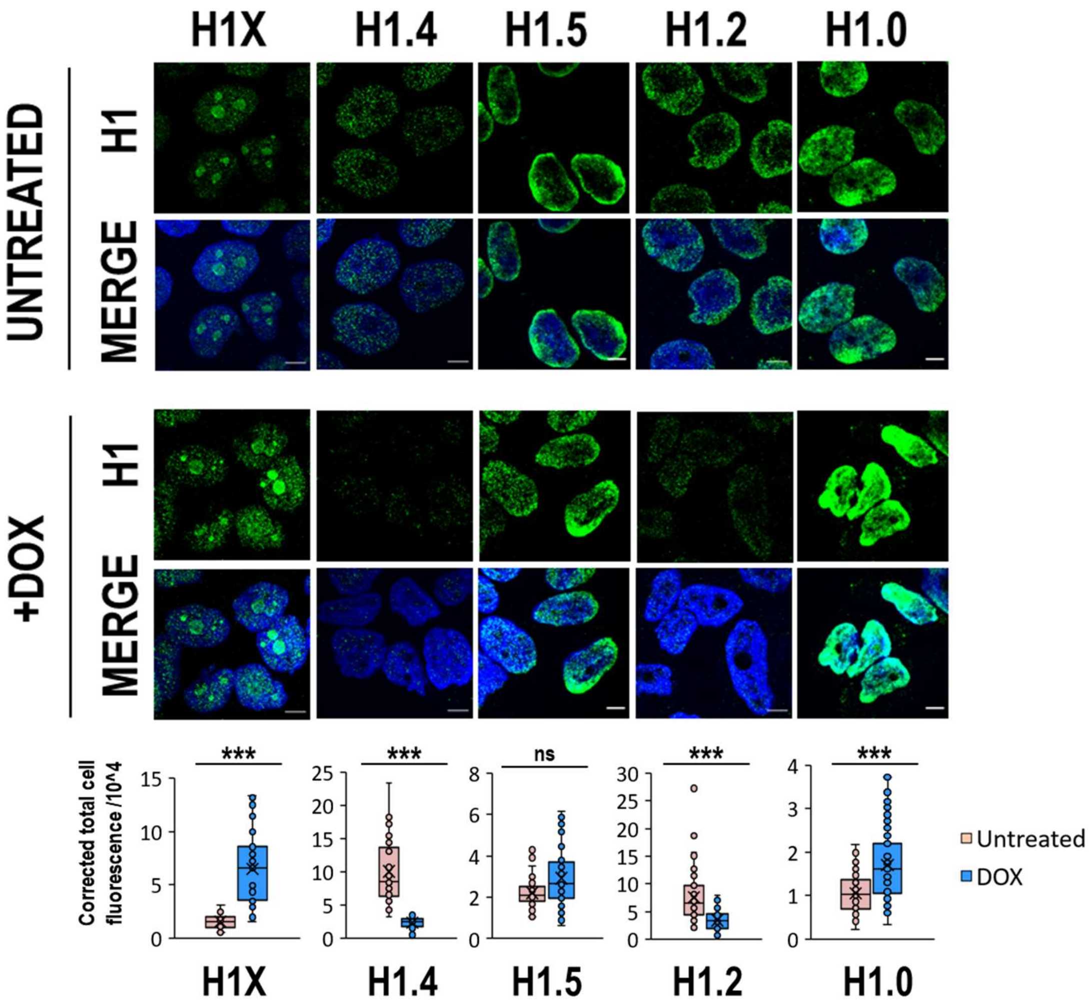


Figure R. 56. H1 variants nuclear distribution remains unaltered upon multiH1 KD. Immunofluorescence of H1 variants (green) and DNA staining (blue signal in MERGE) in T47D multiH1sh Untreated and Dox-treated cells. Bottom boxplot panels represent the H1 variants immunofluorescence signal in both conditions. Number of cells used for quantification (-/+Dox): n=35 (H1X), n=30 (H1.4), n=55(H1.5), n=50 (H1.2), n=70(H1.0). Statistical differences are supported by paired-t-test: (***) $p < 0.001$; (ns/non-significant) $p > 0.05$.

Although H3K9me3 global levels remain unchanged upon multiH1 KD (126) we performed H3K9me3 ChIP-Seq to evaluate a possible genome-wide re-distribution. At the megabase level, significant re-distribution of large H3K9me3 repressive domains were not observed **(Figure R.52)** However, analysis within 100-kb genomic bins revealed that some re-distribution occurred, even though H3K9me3 ChIP-Seq abundance positively correlated in untreated and Dox-treated conditions **(Figure R.57A,B)**. As expected, basal H3K9me3 was enriched towards low-GC and B compartment bins. This was also true upon multiH1 KD condition, but H3K9me3 was more relatively depleted from high-GC bins corresponding to the A compartment. Only slight H3K9me3 re-distribution changes were observed when exploring chromatin states **(Figure R.54)**. We found that the highest H3K9me3 abundance was at *heterochromatin* in both basal and Dox-treated conditions. However, relative increases were detected at *candidate insulators* and *polycomb-repressed* states upon multiH1 KD. Both H1X and H3K9me3 increased at *polycomb* regions in multiH1 KD, maybe trying to maintain these regions silent upon depletion of H1.2, which was enriched at *polycomb* under basal conditions.

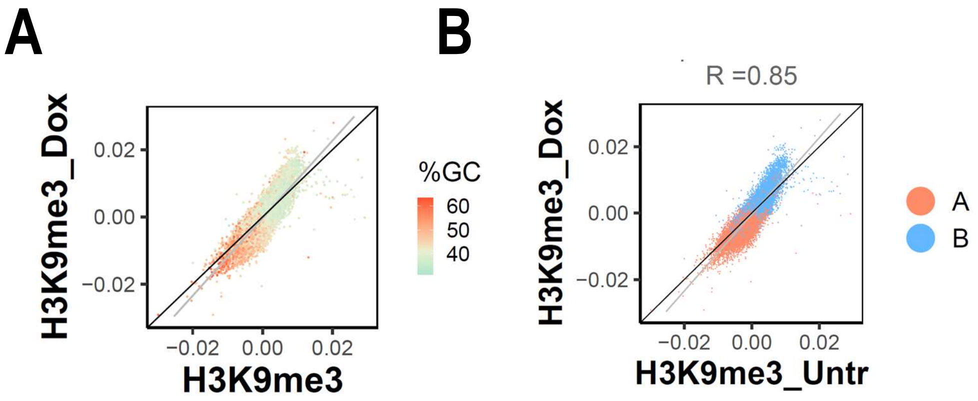


Figure R. 57 H3K9me3 is slightly re-distributed upon mutiH1 KD. Scatterplot of Input-subtracted H3K9me3 ChIP-Seq abundance in T47D multiH1sh Untreated (X-axis) or Dox-treated (Y-axis). Each dot is a 100kb-bin colored by **A)** % GC or **B)** compartment. Pearson correlation coefficient between H3K9me3 Untreated and +Dox samples is shown in (B); R=0.85, p-value <0.001.

On the whole, chromatin homeostasis is highly compromised upon multiH1 KD, as general chromatin decompaction occurs **(Figure R.49)**, accompanied by repetitive elements expression and an interferon response (126). Nevertheless, these alterations take place with only small changes in H1 variants or H3K9me3 distribution, as analyzed here.

Results

Chapter 2. Heterogeneity of H1 variants content, regulation and distribution across cell lines

2. Heterogeneity of H1 variants content, regulation and distribution across cell lines

Total H1 content and the contribution of H1 variants to total H1 are known to vary among cell types, and the genomic distribution of H1 variants seems to differ from one cell type to another (119). However, no systematic studies addressing these questions have been performed to date. Therefore, after investigating the genomic and nuclear distribution of H1 variants in T47D cells, we aimed to extend the study to multiple cell lines, including determination of H1 content and putative regulatory mechanisms of H1 variants.

2.1. H1 variants content and regulation across cell lines

2.1.1. Determination of the H1 repertoire in different cell lines

We first investigated the protein content of H1 variants in different cell lines, most of which had a tumoral origin (see cell line origins in Methods), using Western blot (WB) **(Figure R.58)**. A panel of six melanoma cell lines was also included in the study **(Figure R.59A)**. As previously reported, H1.2 and H1.4 were present in all tested cell lines (194–197). Notably, H1X was also universally-expressed. While H1.0 was only absent in HeLa cells, which also lacked H1.3, H1.3 and H1.5 proteins were not expressed in several cell lines. Specifically, we consistently found a simultaneous lack of both H1.3 and H1.5. Concretely, this was evident in colon cancer cell lines HCT-116, HT-29 **(Figure R.58)** and melanoma cell lines IGR-39 and SK-MEL-173 **(Figure R.59A)**, as well as in breast cancer MDA-MB-231 **(Figure R.5A)**. Moreover, HepG2 cells show no protein expression of H1.3 (and although detectable, very low levels of H1.5) while H1.5 was absent in 293T.

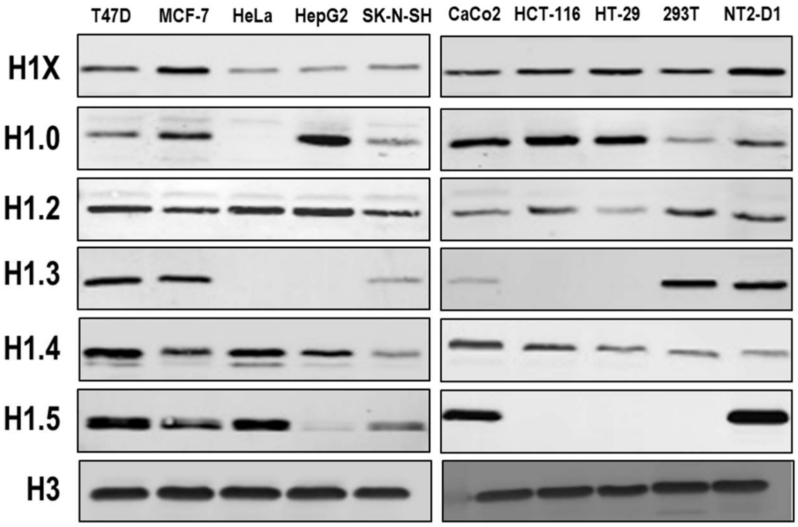


Figure R. 58. H1 somatic repertoire in different cell lines. Immunoblot analysis of H1 variants in histones extracts from different tumoral and non-tumoral cell lines. Histone H3 is added as a nuclear control. For details of the cell lines, see **Table M.3** in Methods Section.

H1 variants are detected in three consecutive bands in the Coomassie staining of histones, which correspond to H1.3+H1.4+H1.5, H1.2 and H1.0 (from higher to lower molecular weight). Quantification of Coomassie staining revealed that melanoma cell lines lacking H1.3 and H1.5 (i.e. IGR-39 and SK-MEL173) had a similar amount of H1.2 and H1.4 and also had more H1.0 compared to other melanoma cell lines tested (**Figure R.59B**). The higher H1.0 content was also evident in other non-melanoma cell lines lacking H1.3 and H1.5 (see HepG2, HCT-116 and HT-29 in **Figure R.58**). It should be noted that H1 variants' contribution to total H1 content, evaluated by WB or Coomassie staining, represents just a proxy and can only be fully addressed by proteomic approaches.

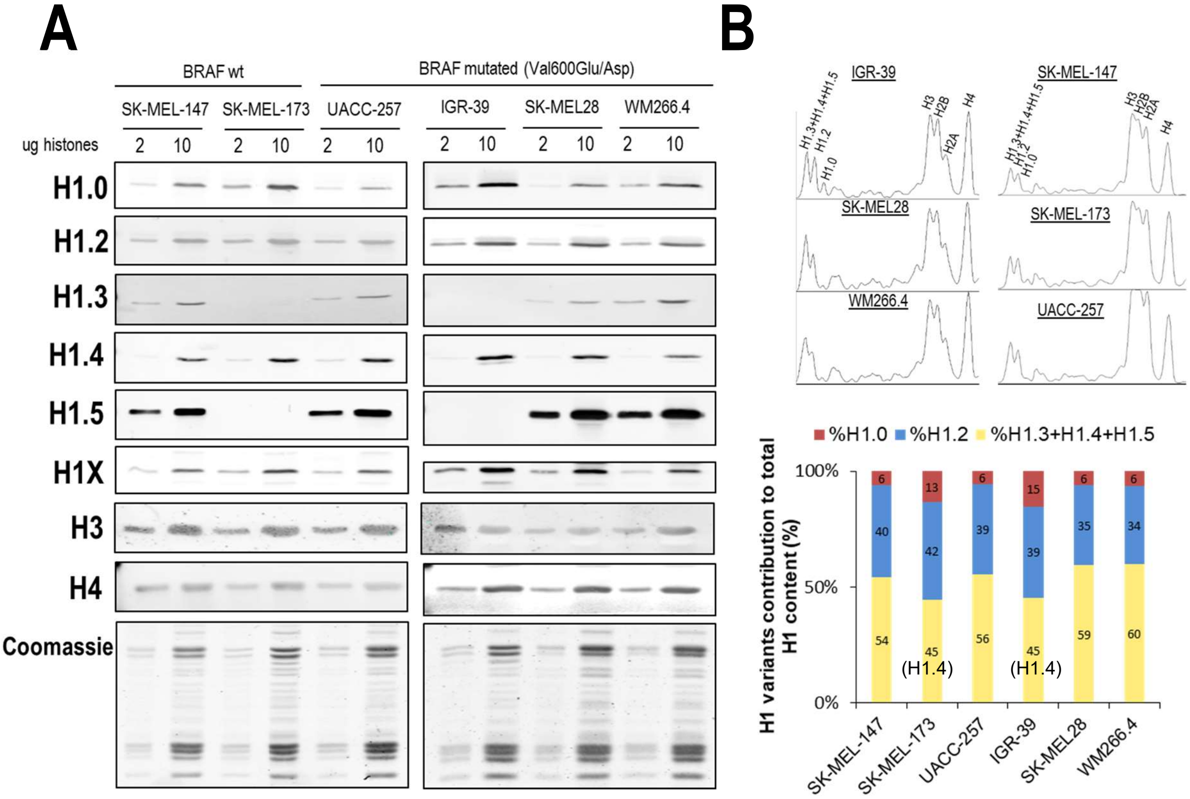
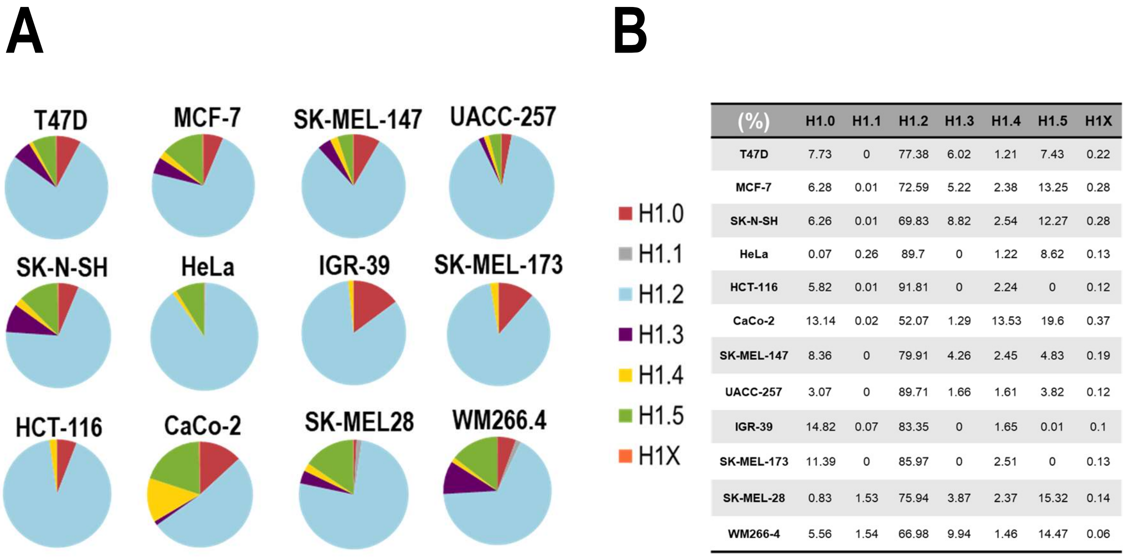


Figure R.59. H1 somatic repertoire in melanoma cell lines. A) Immunoblot of H1 variants in histones extracts (2 or 10 µg) from six melanoma cell lines. Histone H3 and H4 are added as nuclear controls and Coomassie staining is shown. BRAF mutation status is indicated (see Methods section 1.3). **B)** Top panels show the ImageJ profiling of histones Coomassie shown in (A). The bottom panel represents the contribution of the three H1 Coomassie bands to total H1 content. Data is normalized by H4 band and represented in percentage. In cell lines lacking H1.3 and H1.5, the upper H1 Coomassie band corresponds to H1.4.

The correlation between mRNA and protein levels of H1 variants may not always be direct, and the concrete post-transcriptional and post-translational regulatory mechanisms of H1 variants are not well understood. Therefore, we also evaluated H1 variants mRNA expression levels by reverse-transcriptase-quantitative PCR (RT-qPCR) (**Figure R.60A,B**). H1.2 was the most expressed at mRNA level in all cell lines. On the contrary, although proteomic experiments revealed that H1.4 protein tends to represent a considerable percentage of total H1 (A. Roque personal communication), their mRNA levels were low. Indeed, in cell lines lacking H1.3 and H1.5, upper H1 band in Coomassie staining of histones correspond to H1.4. Thus, in SK-MEL-173 and IGR-39 melanoma cells, each H1.2 and H1.4 represent ≈40% of the total H1 protein. On the contrary, at mRNA level, H1.2 represents >80% of the H1 transcripts in these two cell lines, while H1.4 only 1.6-2.5%. Simultaneous absence of H1.3 and H1.5 seen by WB was also detected by RT-qPCR (IGR-39, SK-MEL-173, HCT-116), indicating that their repression mechanism occurs at the transcriptional level. In HeLa cells, which lack H1.3 and H1.0 at protein level **(Figure R.58)**, no H1.3 and almost no H1.0 gene expression was detected **(Figure R.60B)**. In those cell lines where we observed a higher H1.0 protein content, H1.0 mRNA levels were in general also high. Particularly, CaCo-2, IGR-39 and SK-MEL-173 were the cell lines where H1.0 represented the higher percentage of total H1 gene expression **(Figure R.60A,B)**. In conclusion, our analysis supports H1 variants variable content among cell types and suggests that regulatory mechanisms differ between variants.



(%)	H1.0	H1.1	H1.2	H1.3	H1.4	H1.5	H1X
T47D	7.73	0	77.38	6.02	1.21	7.43	0.22
MCF-7	6.28	0.01	72.59	5.22	2.38	13.25	0.28
SK-N-SH	6.26	0.01	69.83	8.82	2.54	12.27	0.28
HeLa	0.07	0.26	89.7	0	1.22	8.62	0.13
HCT-116	5.82	0.01	91.81	0	2.24	0	0.12
CaCo-2	13.14	0.02	52.07	1.29	13.53	19.6	0.37
SK-MEL-147	8.36	0	79.91	4.26	2.45	4.83	0.19
UACC-257	3.07	0	89.71	1.66	1.61	3.82	0.12
IGR-39	14.82	0.07	83.35	0	1.65	0.01	0.1
SK-MEL-173	11.39	0	85.97	0	2.51	0	0.13
SK-MEL-28	0.83	1.53	75.94	3.87	2.37	15.32	0.14
WM266-4	5.56	1.54	66.98	9.94	1.46	14.47	0.06

Figure R. 60. H1 variants expression levels in cancer cell lines. Gene expression levels of H1 variants in different cancer cell lines were analyzed by RT-qPCR. Data is corrected by GAPDH and normalized by the corresponding genomic DNA amplification. Corrected expression data from all H1 variants is summed to calculate total H1 expression and represent values as percentage. Data is represented as pie charts (A) or numerically collected in a table (B).

2.1.2. H1 variants regulation by DNA methylation

The concomitant absence of H1.3 and H1.5 appears to occur at the transcriptional level, which is compatible with repression by DNA methylation. To explore this possibility, we analyzed public data from NCBI-60 cell lines panel and found that H1.2, H1.4 and H1X expression showed no correlation with methylation status of the corresponding genes **(Figure R.61)**. In contrast, H1.0, H1.1, H1.5, H1.3 expression levels showed a positive correlation with gene methylation status and they were not expressed in all cell lines. This observation indicates that in some cell lines, expression of these variants in repressed by DNA methylation. Moreover, gene methylation data from cancer patients (TCGA) revealed that methylation of H1 variant genes varied between cancers originating from different tissues **(Figure R.62)**. H1.2, H1.4 and H1X genes were unmethylated in the three datasets analyzed, while the gene methylation of the other H1 variants was variable. Overall, analysis of public data supports the universal expression of H1.2/H1.4/ H1X observed by WB and a correlation between the expression of the remaining variants and their gene methylation status.

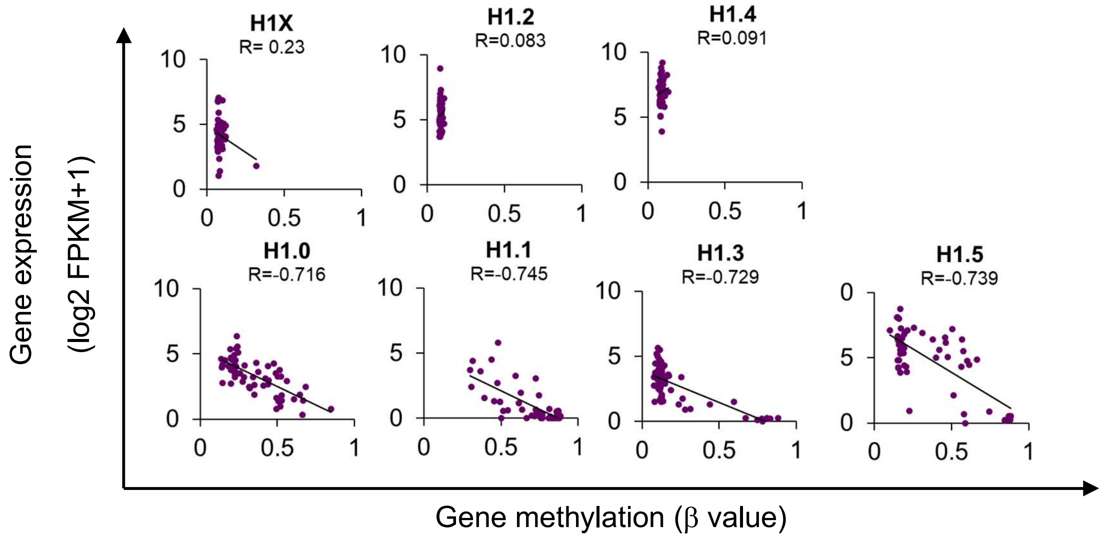


Figure R. 61. H1 variants expression and gene methylation levels across cell lines. Scatterplots between H1 variants expression (Y-axis) and H1 gene methylation (X-axis) from NCI-60 public data. Gene expression from total RNA-Seq data is expressed in log2 (FPKM+1) while gene methylation is expressed as b value (b =0 is totally unmethylated and b=1 is totally methylated). Each dot represents a cell line from the NCI-60 panel. Pearson correlation coefficients (R) are shown.

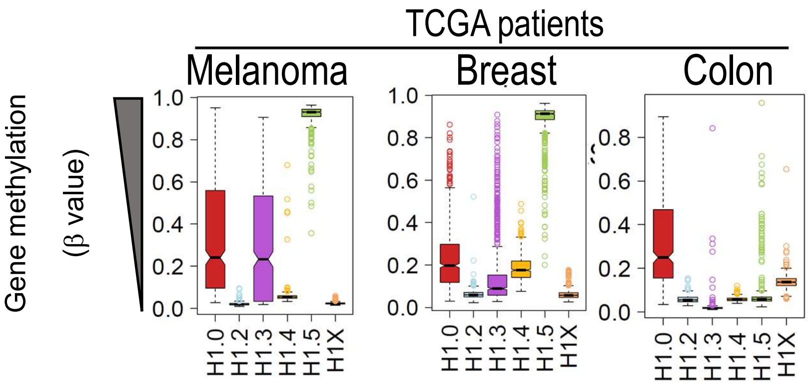


Figure R.62. H1 variants are differentially methylated in cancer patients. Boxplots show the DNA methylation of H1 genes in different cancer datasets from TCGA project. Gene methylation is expressed as β value (β =0 is totally unmethylated and β =1 is totally methylated).

We next explored whether transcriptional repression of H1 variants was reversed by inhibition of DNA methylation. To do so, we treated some of the cell lines that lack H1 variants at the protein level **(Figure R.58, R.59)** with 5-aza-2'-deoxycytidine (aza). In those cell lines lacking H1.3+H1.5, a huge mRNA upregulation of these variants occurred upon aza treatment **(Figure R.63)**. H1.0 expression was also upregulated in HeLa cells (which lack H1.0 protein) but to a lesser extent. Moreover, H1.1 expression was also upregulated upon aza treatment in all cell lines. Notably, H1.1 is not expressed basally in most cell lines **(Figure R.60B)**. Differences in the magnitude of upregulation upon DNA hypomethylation may underlie H1 variant-specific regulatory mechanisms of mRNA content. Thus, this data supports the idea of H1.3 and H1.5 being repressed by DNA-methylation in a subset of cell lines.

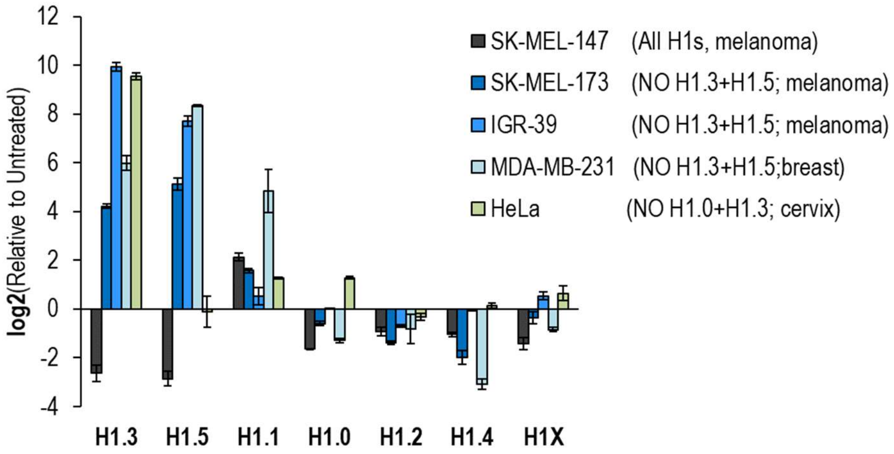


Figure R. 63. DNA hypomethylation leads to upregulation of H1 variants that are naturally absent in cell lines. H1 variants expression levels in different cell lines under Untreated and aza-treated conditions were analyzed by RT-qPCR. Barplot shows relative expression of H1 variants upon aza treatment compared to Untreated condition, corrected by GAPDH and expressed as log2. SK-MEL-147 is added as a control cell line which expresses all H1 variants (without considering H1.1).

2.2. H1 variants distribution differences across cell lines

2.2.1. H1 variants nuclear distribution across different cell lines

To evaluate differential and common distribution patterns of H1 variants among different cell lines, we performed immunofluorescence of six endogenous H1 variants in some of the cell lines previously mentioned in this section **(Figure R.64)**. Importantly, H1.2/H1.3/H1.5 were universally enriched at the nuclear periphery, as observed in T47D. H1.0 and H1.4 were distributed throughout the nucleus, making it difficult to directly compare their distribution patterns by immunofluorescence experiments. Lastly, H1X was also distributed throughout the entire nucleus, but the intensity of its nucleolar enrichment was variable between cell lines. However, as further studied later, it was still present at nucleolus in all cell lines tested, contrary to the other H1 variants, which are depleted at nucleoli.

As both H1.3 and H1.5 seemed to be universally enriched at LADs or peripheral chromatin, we wonder whether in cell lines lacking these two variants, re-distribution of the remaining H1 proteins to the nuclear periphery occurs. To do so, we performed H1 variants immunofluorescence in a subset of cell lines lacking H1.3 and H1.5 **(Figure R.65)**. H1.2 was also enriched at the nuclear periphery, as before. H1.4 and H1.0 appeared to have a more peripheral distribution compared to cell lines expressing all H1 variants evaluated here. On the contrary, H1X was not found enriched at the nuclear periphery in any cell line tested. In all cases, H1X was distributed throughout the nucleus in a punctuated pattern, with diverse nucleolar enrichment between cell lines, as previously observed.

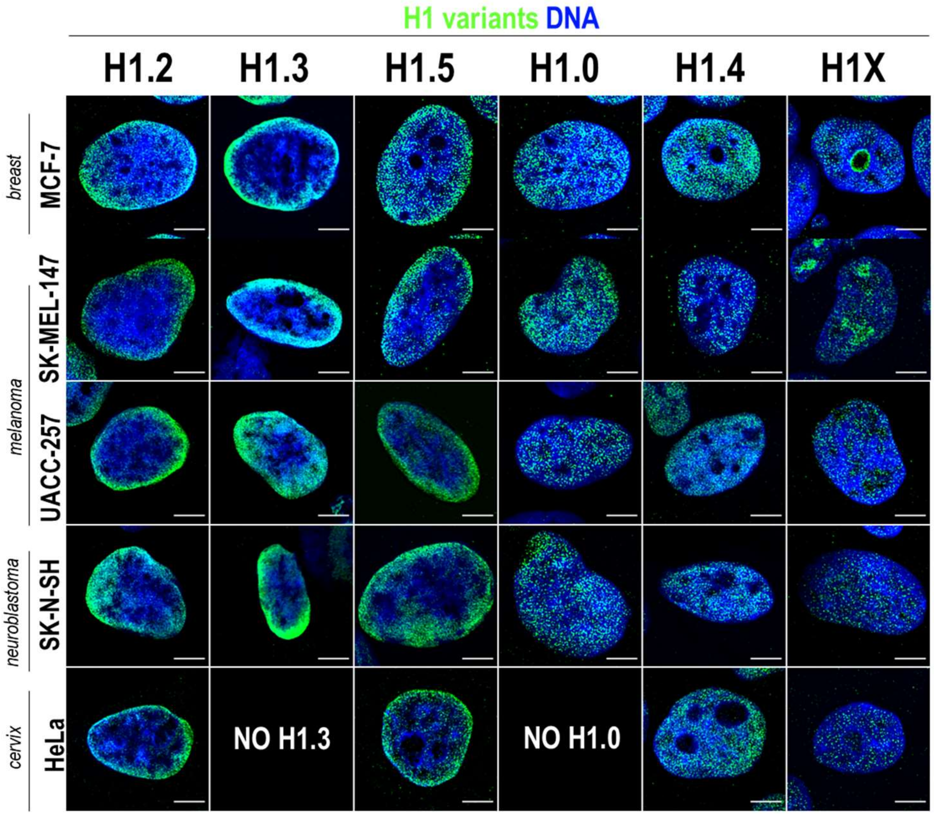


Figure R. 64. **Nuclear distribution of H1 variants in different cell lines.** Immunofluorescence analysis of H1 variants (green) with DNA staining (blue). Merged images are shown. H1.3 and H1.0 grids in HeLa cells are empty, as HeLa cells do not express these variants (Figure R.58). Tumoral origin of the cell lines is indicated. Scale bar: 5µm

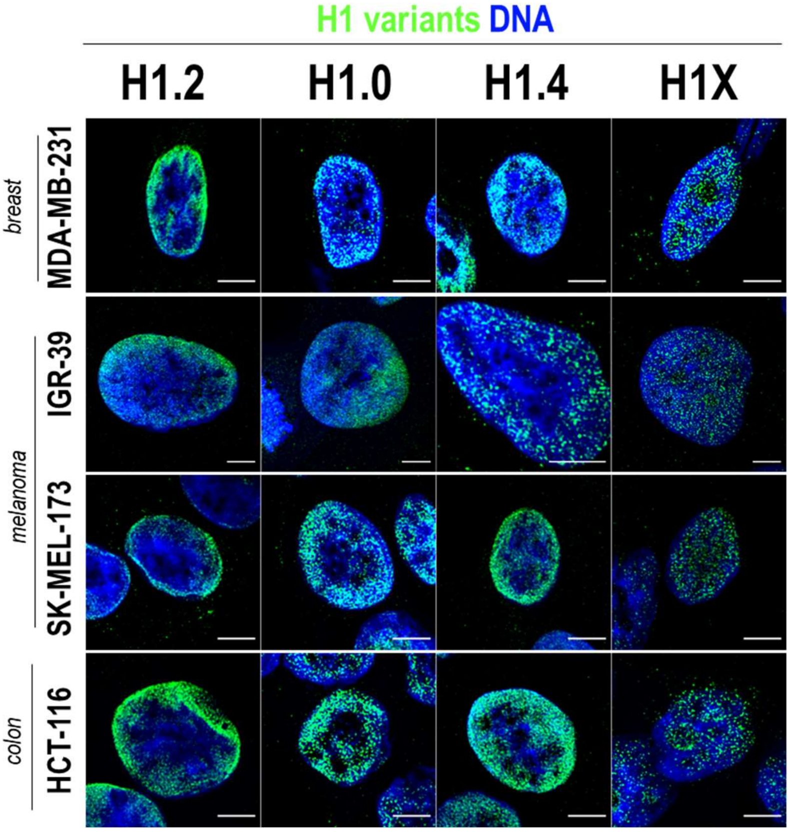


Figure R. 65. Nuclear distribution of H1 variants in cell lines lacking H1.3 and H1.5 Immunofluorescence analysis of H1 variants (green) with DNA staining (blue). Merged images are shown. Tumoral origin of the cell lines is indicated. Scale bar: 5µm

To quantify the apparent H1.0 and H1.4 enrichments at the nuclear periphery in cells lacking H1.3/H1.5 compared to the remaining cell lines, we performed radial intensity analysis **(Figure R.66)**, as previously done in T47D **(Figure R.33)**. Regarding H1.4 radial distribution, quantification in cell lines expressing the six H1 variants analyzed revealed that radial intensity distribution was similar to H1.4 in T47D **(Figure R.66A)**. In most cell lines, H1.4 was more enriched in A2 *versus* the most peripheral quartile (i.e. A1). In addition, intensity decreasing towards the nucleus center was not too accentuated, as also observed in T47D. On the other hand, in cell lines lacking H1.3/H1.5, H1.4 was equivalently enriched in both A1-A2 quartiles or even more enriched in A1 compared to A2. Besides, gradual decreasing towards the nucleus center was more accentuated in these cell lines. Importantly, H1.4 radial quantification in HeLa cells, which lack H1.3/H1.0, also denoted a peripheral distribution of H1.4. Analogous results were obtained for H1.0 **(Figure 66B)**. H1.0 in MCF-7 breast cancer cell line followed a similar radial distribution to T47D. However, it was more enriched at A2 compared to A1 quartile in other cell lines evaluated. This might suggest a tissue-specific variable distribution of H1.0. In cell lines lacking H1.3/H1.5, H1.0 also showed a more peripherical radial intensity profile.

To make a more direct comparison between different cell lines, we calculated a 'Peripheral index' for H1.4 and H1.0 nuclear distribution. Peripheral Index was defined as the ratio between mean intensity in A1 section and mean intensity in A4. That is, the higher Peripheral Index, the higher enrichment towards the nuclear periphery. Calculation of H1.4 and H1.0 Peripheral Index confirmed that those cell lines with a compromised H1 repertoire presented a higher Peripheral Index for both H1 variants, compared to those cell lines expressing all H1 variants **(Figure R.67)**.

Altogether, analysis of radial intensity of H1.4 and H1.0 in multiple cell lines provides quantitative support for their more peripheral distribution in cell lines with a compromised H1 repertoire.

As all cell lines evaluated were cancer cell lines, we next performed immunofluorescence of H1 variants in a non-tumoral cell line. Specifically, we used IMR90 normal fibroblasts cells **(Figure R.68)**. Firstly, we observed that chromatin was more organized in IMR90 cells compared to cancer cell lines, as heterochromatic domains were well differentiated by DNA staining. In this way, H1.2/H1.3/H1.5 distribution mimicked the DNA pattern, including peripheral condensed DNA regions. H1.0 was distributed throughout the nucleus but certain enrichment at peripheral heterochromatic foci was also observed. On the other hand, H1.4 was also distributed along the nucleus but tended to present an opposite profile to DNA. The same was observed for H1X, which additionally was highly enriched at nucleoli. We could conclude that, albeit chromatin organization may differ from cell line to cell line (specially from cancer to non-tumoral ones), H1.2/H1.3/H1.5/H1.0 are more related to heterochromatic nuclear regions, including the nuclear periphery. In contrast, H1.4 and H1X are more abundant in less condensed chromatin domains. Thus, the nuclear distribution panorama of H1 variants somehow resembles the one observed in T47D and other cells lines, especially for H1.2/H1.3/H1.5 and H1X.

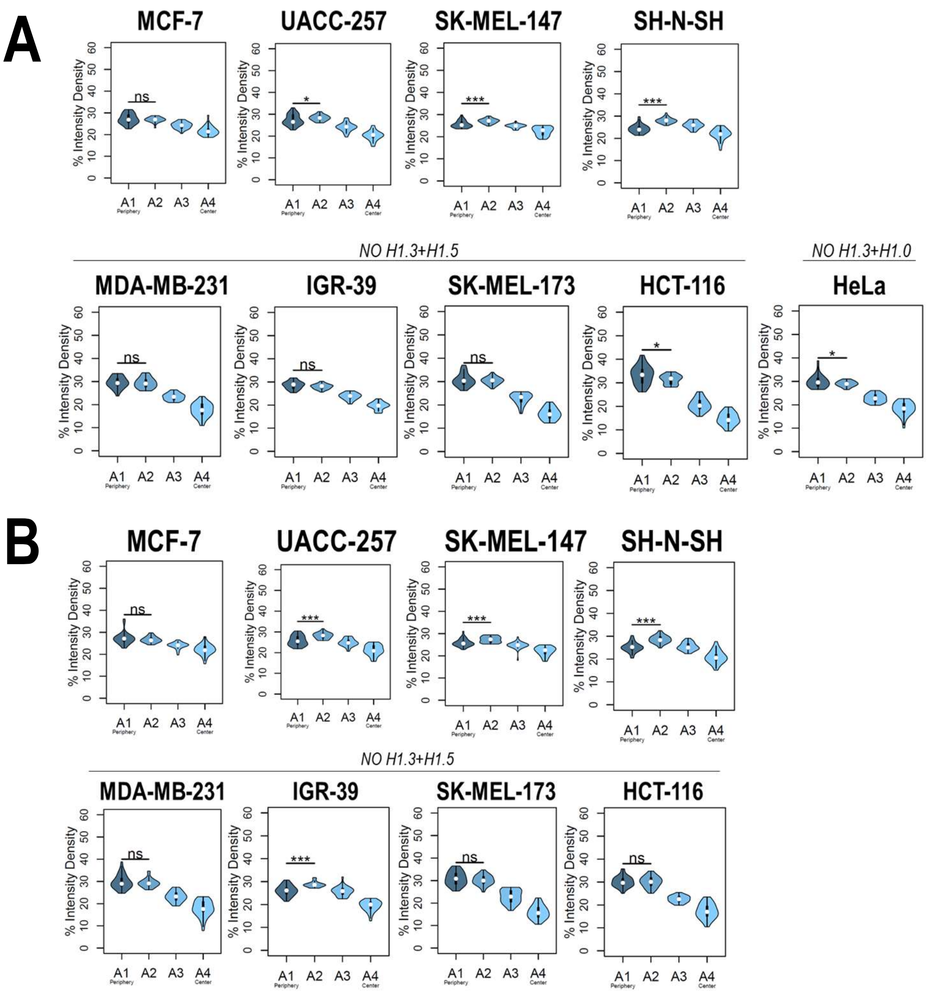
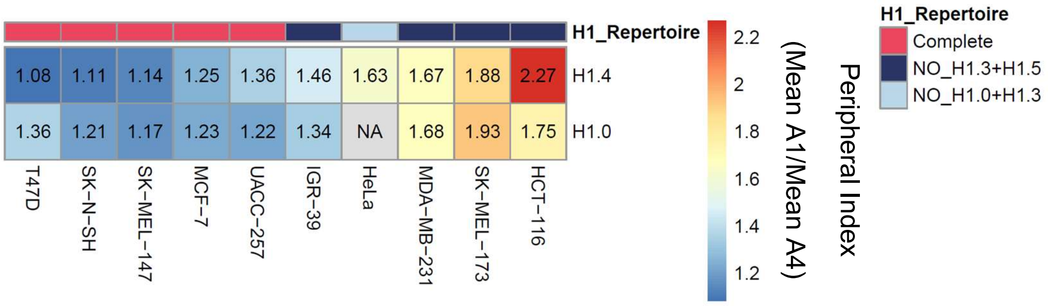


Figure R. 66. H1.4 and H1.0 radiality quantification in multiple cancer cell lines. Graphs show immunofluorescence quantification signal of **A)** H1.4 and **B)** H1.0 in different cell lines. As illustrated in **Figure R.33**, four sections of an equivalent area and convergent to the nuclear center are created per each cell. Sections are named A1 to A4, from the more peripheral section to the more central one. H1 variants immunofluorescence intensity is measured in each area and expressed as percentage. Cell lines with a compromised H1 somatic repertoire are indicated. n=30 cells/cell line were quantified, and data was represented in violin plots. Statistical differences between A1-A2 are supported by paired t-test (***) p-value<0,001; (*) p-value < 0.05; (ns/non-significant) p-value>0.05.



	T47D	SK-N-SH	SK-MEL-147	MCF-7	UACC-257	IGR-39	HeLa	MDA-MB-231	SK-MEL-173	HCT-116
H1.4	1.08	1.11	1.14	1.25	1.36	1.46	1.63	1.67	1.88	2.27
H1.0	1.36	1.21	1.17	1.23	1.22	1.34	NA	1.68	1.93	1.75

Figure R. 67. H1.4 and H1.0 show a more peripheral distribution in cell lines with a compromised H1 repertoire. Numbers correspond to Peripheral Index value in each cell line and color-coded as indicated. Peripheral index was calculated with data from Figure R.66 and defined as the ratio between Average value in A1 peripheral section and A4 central section. So, a higher Peripheral Index is related to a more peripheral nuclear distribution.

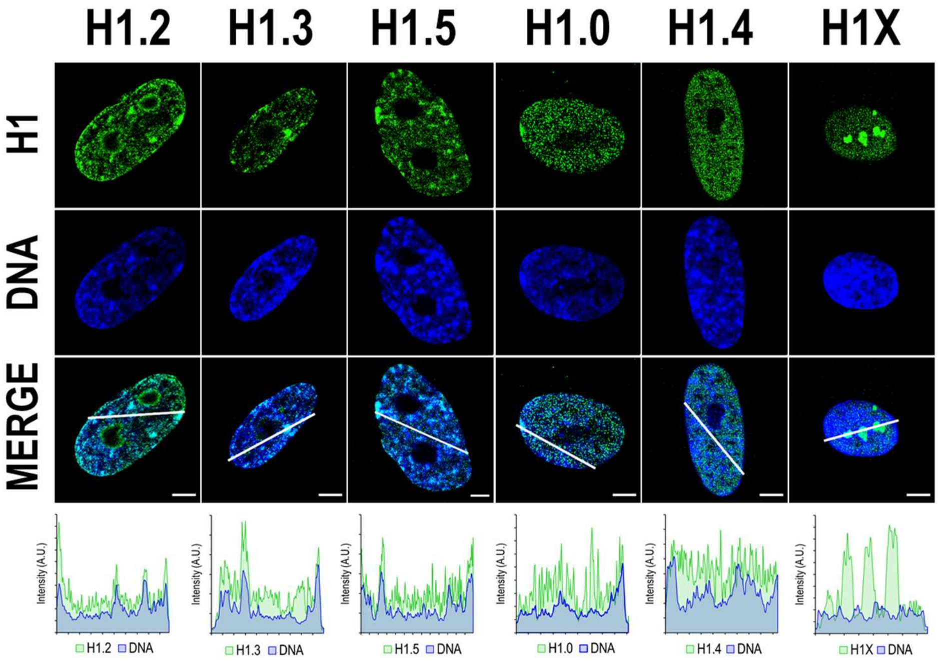


Figure R. 68. H1 variants nuclear distribution in IMR-90 cells. Immunofluorescence of H1 variants (green) and DNA staining (blue) in non-tumoral IMR-90 cells. Bottom graphs show Intensity profiles along the lines drawn in the Merge channel. Scale bar: 5µm.

2.2.2. Nucleolar enrichment and genomic distribution of histone H1X in different cell lines

We next focused our research on histone H1X. As depicted in **Figures R64-65**, H1X was distributed throughout the whole nucleus in a punctuated pattern, with a variable nucleolar enrichment between different cell lines. Co-immunostaining of H1X with the nucleolar marker NPM1 confirmed H1X nucleolar enrichment (**Figure R.69**).

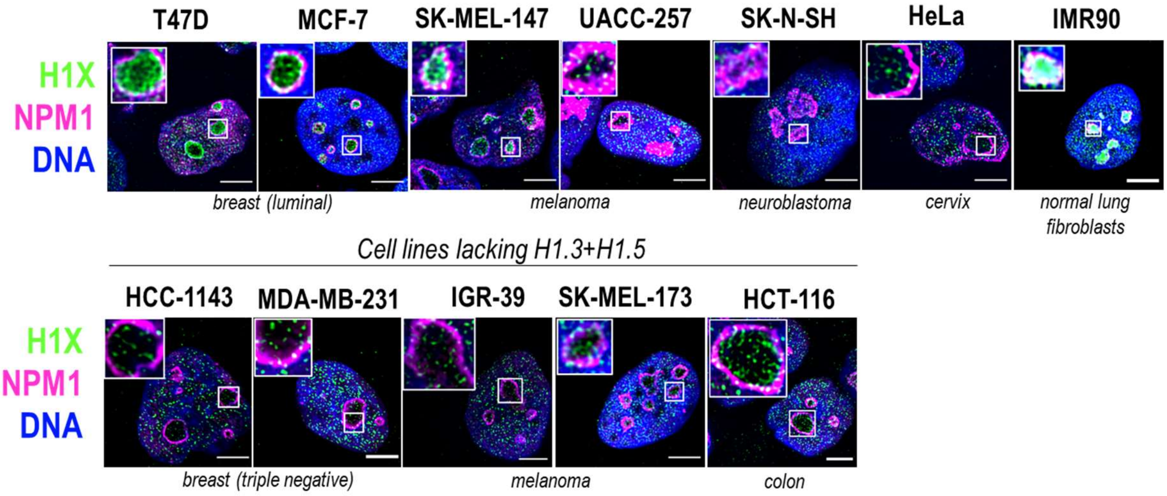


Figure R. 69. H1X is detected within nucleoli of all cell lines. Immunofluorescence of H1X (green), nucleolar marker NPM1 or Nucleophosmin (magenta) and DNA staining (blue). Merge images are shown. Insets show a zoom-in of a single nucleolus. Bottom panel includes cell lines lacking H1.3 and H1.5. Cell line origin is indicated. Scale bar: 5µm.

Among the cell lines tested, T47D, MCF-7, SK-MEL-147, HCT-116 and IMR-90 showed the most prominent H1X nucleolar enrichment. Notably, in MCF-7 breast cancer cells, H1X formed a layer at the nucleolar rim, adjacent to NPM1, similar to what was observed in T47D cells. Nevertheless, it is essential to note that H1X was still present at nucleoli in all cell lines tested, making it the H1 variant most associated with nucleoli, where other H1 variants are underrepresented.

To further study H1X distribution, we performed H1X ChIP-Seq in several of the cell lines analyzed (specifically in MCF-7, SK-MEL-147, SK-N-SH, HeLa and HCT-116). We used G-bands segmentation to compare the H1X ChIP-Seq abundance in the mentioned cell lines. H1X showed a strong correlation with %GC content in all cell lines, being highly enriched at high-GC G-bands **(Figure R.70)**, similar to what was observed in T47D **(Figure R.6)**.

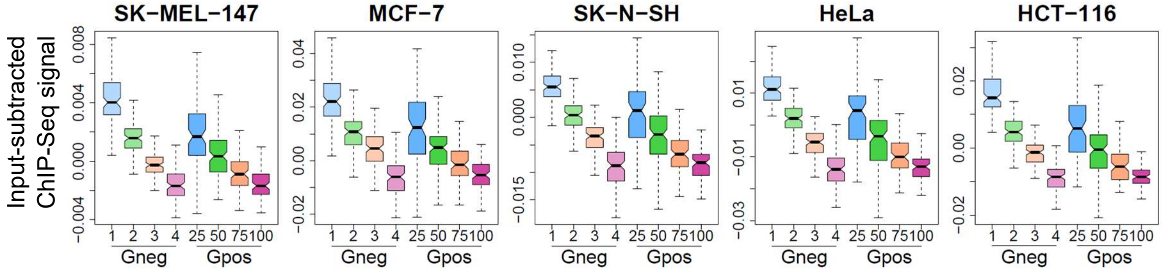


Figure R.70. H1X is enriched at high-GC bands in multiple cell lines. Boxplots show the H1X Input-subtracted ChIP-Seq signal at eight groups of G-bands in five different cancer cell lines.

Although being highly enriched at high-GC regions, analysis in T47D cells revealed that H1X was particularly associated with some repetitive elements. Specifically, it was highly enriched at SVA retrotransposons and other repeats recently-incorporated in our genome (see Results Chapter 1, section 1.3). As H1X ChIP-Seq analysis in different cell lines showed a universal correlation with GC content **(Figure R.70)**, we wondered whether H1X distribution within repetitive elements was also universal or exhibited cell-line specific profiles.

Firstly, a comparative study of H1X abundance within repeat classes confirmed that the repeat class showing the highest H1X enrichment was SVA retrotransposons (i.e. Other Class) in all the cell lines tested **(Figure R.71)**. In general, H1X was also preferentially enriched within SINE class, although relative enrichment compared to 'Other' Class was variable. Similarly, relative enrichment between other repeat classes also exhibited cell-line variability. For instance, in SK-MEL-147 and HeLa cells, H1X was more enriched within Satellite compared to SINE, but the contrary was observed in MCF-7, SK-N-SH and HCT-116.

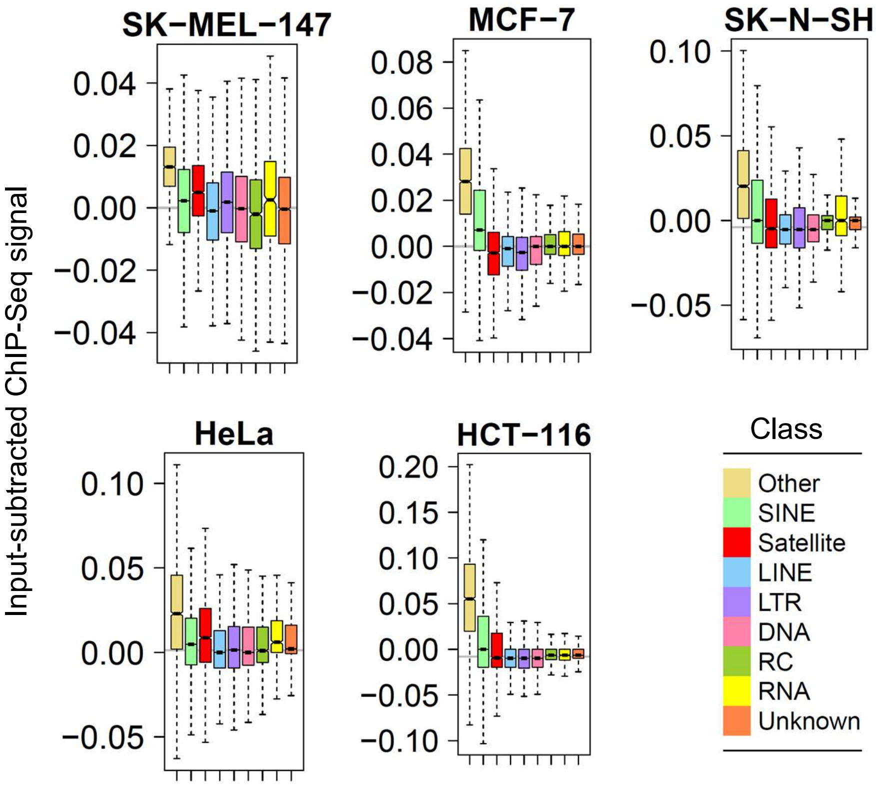


Figure R. 71. H1X is enriched within Other and SINE classes in cancer cell lines. Boxplot of H1X Input-subtracted ChIP-Seq signal within repetitive element classes in five cancer cell lines.

Given the enrichment of H1X in SVAs compared to other classes of repetitive elements, a more detailed analysis was conducted to examine the abundance of H1X within SVA retrotransposons. Profiling of H1X at meta-SVA repeats revealed H1X was enriched at SVAs compared to their surrounding regions in all cell lines analyzed (**Figure R.72**). Besides, a gradual increase from SVA_A to SVA_F families was found, as previously observed in T47D cells. Notably, this gradual increase was not only restricted to the repeat itself, but also expanded to the flanking 5´and, especially, 3´regions. As commented in the Chapter 1, this is due to SVA repeats tend to be in clusters, especially for SVA_F repeats. Thus, a characteristic H1X local peak enrichment was found at 3´ of SVA_F family repeats, the most-recently incorporated SVA in the human genome. On the contrary, for the older SVA families, a local H1X decrease was observed just *upstream* of the indicated 3´local SVA_F peak. Importantly, this valley was observed in all cell lines, gradually disappearing from SVA_A to SVA_F, but with cell-line-specific relative magnitudes. For example, SK-MEL-147 cells exhibited the least prominent H1X 3´ valley, similar to T47D profile (**Figure R.72** SK-MEL-147 vs **Figure R.17B** T47D).

As the relationship between H1X abundance and SVA families' evolutionary age was evident in all cell lines, we next explored whether this association is maintained for other repeat classes, as occurred in T47D. Stratification of repeats into different clades denoted that H1X was enriched within most evolutionary recent repeats in the five cell lines (**Figure R.73**). Concretely, for each cell line, H1X was enriched within Hominidae and Homo sapiens repeats.

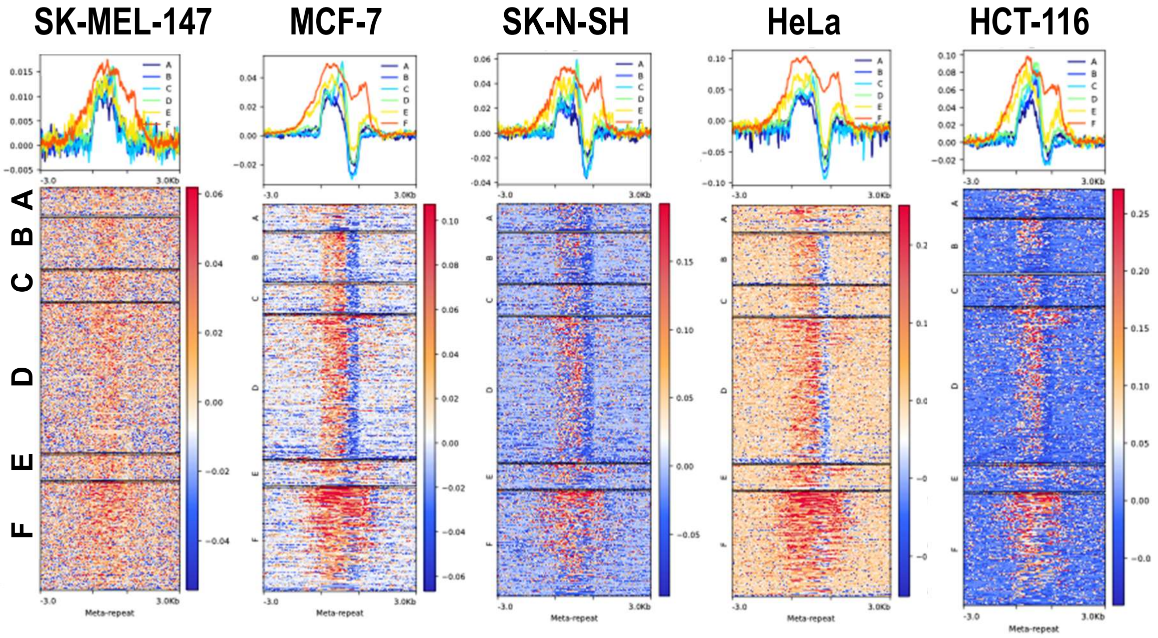


Figure R. 72. H1X is enriched within SVA families in cancer cell lines. Meta-repeat profile of H1X Input-subtracted ChIP-Seq abundance in five cancer cell lines at SVA repeats and their 3kb flanking regions. In the heatmaps, each row represents a SVA repeat of the indicated family and ordered by the corresponding H1 profile in each case (i.e., heatmaps show an independent order for each cell line). Average profile of H1X per family is also shown in the upper line graphs.

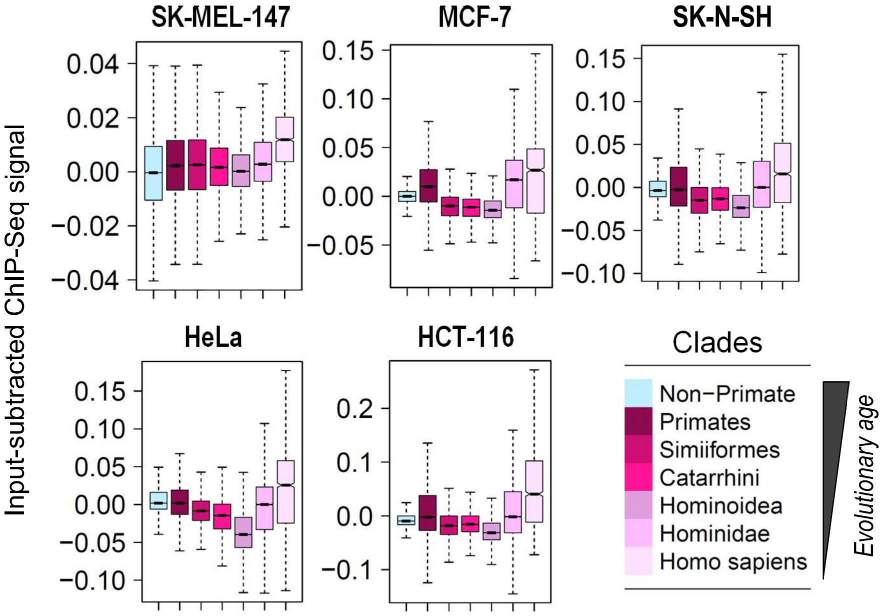


Figure R. 73. H1X abundance increases within repeats recently incorporated along primate evolution in different cell lines. Boxplot analysis of H1X Input-subtracted ChIP-Seq abundance in five cancer cell lines within repetitive elements classified into different taxonomic clades according to Dfam database, ordered by their evolutionary age. The classification of repeats corresponds to the taxonomic clade to which the oldest ancestor in which the repeat has been found, belongs.

We next considered the different repeat families classified by clades. Clustering analysis of H1X abundance in different cell lines confirmed that H1X distribution within TEs in SK-MEL-147 was more similar to the distribution observed in T47D, compared to the other cell lines evaluated **(Figure R.74)**. Moreover, it became evident that in all cases the TE families showing the highest H1X abundance were SVA_D, SVA_E and SVA_F, while the rest of SVA families and Alu repeats also showed a high H1X abundance in all cell lines. However, this analysis revealed notable differences for multiple LINE and LTR families. In HeLa, HCT-116, MCF-7 and SK-N-SH, H1X was relatively depleted in the younger clades of L1 and ERVL-MaLR repeats, in comparison to *Non-Primates* repeats of these classes. To further analyze these differences between multiple cell lines, we next evaluated the different repeat classes separately.

The universal relationship of H1X abundance and SVA evolutionary age was further confirmed upon SVA families' taxonomic classification **(Figure R.75A)**. H1X was found to be universally enriched at more-recently incorporated SVA elements. The highest enrichment was found at human-specific SVA families (i.e. SVA_E and SVA_F).

Analogous examination of Alu family also showed a gradual increase of H1X along Alu elements evolution, which was evident in all cell lines analyzed **(Figure R.75B)**.

Conversely, this evolutionary trend was not universally observed in LINE-L1 or LTR families **(Figure R.75C, R76)**. Regarding LINE-L1 family, the only cell line in which H1X was gradually increasing from older to more recent L1 elements was SK-MEL-147 **(Figure R.75C)**, although the tendency was not as accentuated as previously observed in T47D. On the contrary, in the remaining cell lines H1X was more abundant at older L1 elements compared to recently-evolved ones. A similar scenario was observed when evaluating the different LTR families **(Figure R.76)**. Again, the cell line in which H1X showed a more similar behavior compared to T47D was SK-MEL-147. Indeed, in this cell line H1X was enriched at more recent LTR repeats, including all the four LTR families. On the other hand, in the rest of cell lines analyzed, H1X abundance was gradually decreasing from older to more recent LTR repeats. This tendency was evident along ERVK, ERV1 and ERVL-MaLR repeat families' evolution. However, H1X abundance at the youngest LTR-ERVL repeats (i.e. *Simiiformes* ERVL) was higher compared to the evolutionary-prior group (i.e. ERVL *Primates* repeats).

We next focused on the most recent repeats restricted to *Hominoidea* species. As denoted by previous analysis, the younger families of SVA repeats showed the highest H1X abundance in all cell lines, followed by the rest of the SVA families and Alu repeats **(Figure R.74B)**. Heatmap analysis showed that young LINEs from the L1PA lineage clustered together and were the *Hominoidea* repeats presenting lower H1X abundance in almost all cell lines except for T47D. On the other hand, H1X relative abundance between *Hominoidea* LTR repeats from ERV1 and ERVK families were variable between different cell lines. Despite this variability, *Hominidae* ERV1 repeats clustered together and tended to present higher H1X levels than other LTR repeats. Interestingly, H1X enrichment within LTR12C and LTR12E was universally found in all cell lines.

Altogether, the comparative H1X ChIP-Seq analysis in different cell lines revealed that H1X is universally associated to SVA retrotransposons, especially at more recently-incorporated ones. Additionally, H1X is also abundant at youngest Alu elements in all cell lines tested. Nevertheless, H1X exhibits a non-universal pattern at other repeats, including LINE and LTR classes.

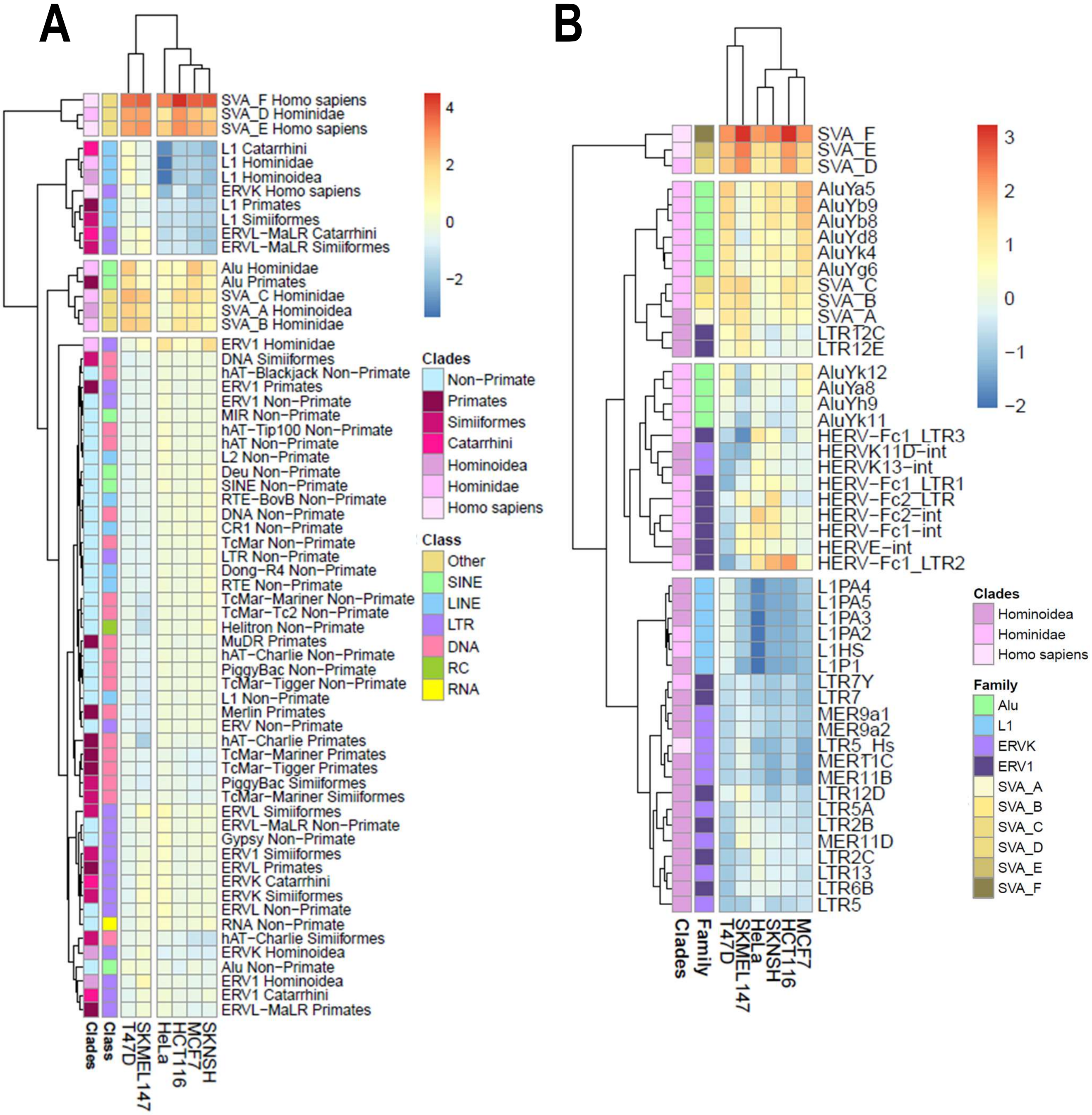


Figure R. 74. H1X abundance within repetitive element families from an evolutionary perspective in different cell lines. A) Heatmap and clustering of H1X Input-subtracted ChIP-Seq median abundance (scaled) in six cancer cell lines at repetitive elements classified per family and taxonomic clades (n=64; Satellite and Unknown classes are excluded). Y-axis annotation indicates Repetitive element Class and Clades. First, median abundance per repeat was calculated and then median abundance per family and clade was plotted on the final heatmap**. B)** Heatmap and clustering of H1X Input-subtracted ChIP-Seq median abundance (scaled) in six different cancer cell lines at n=48 repeats from *Hominoidea* and descendant clades. Y-axis annotation indicates repetitive element family and taxonomic clades.

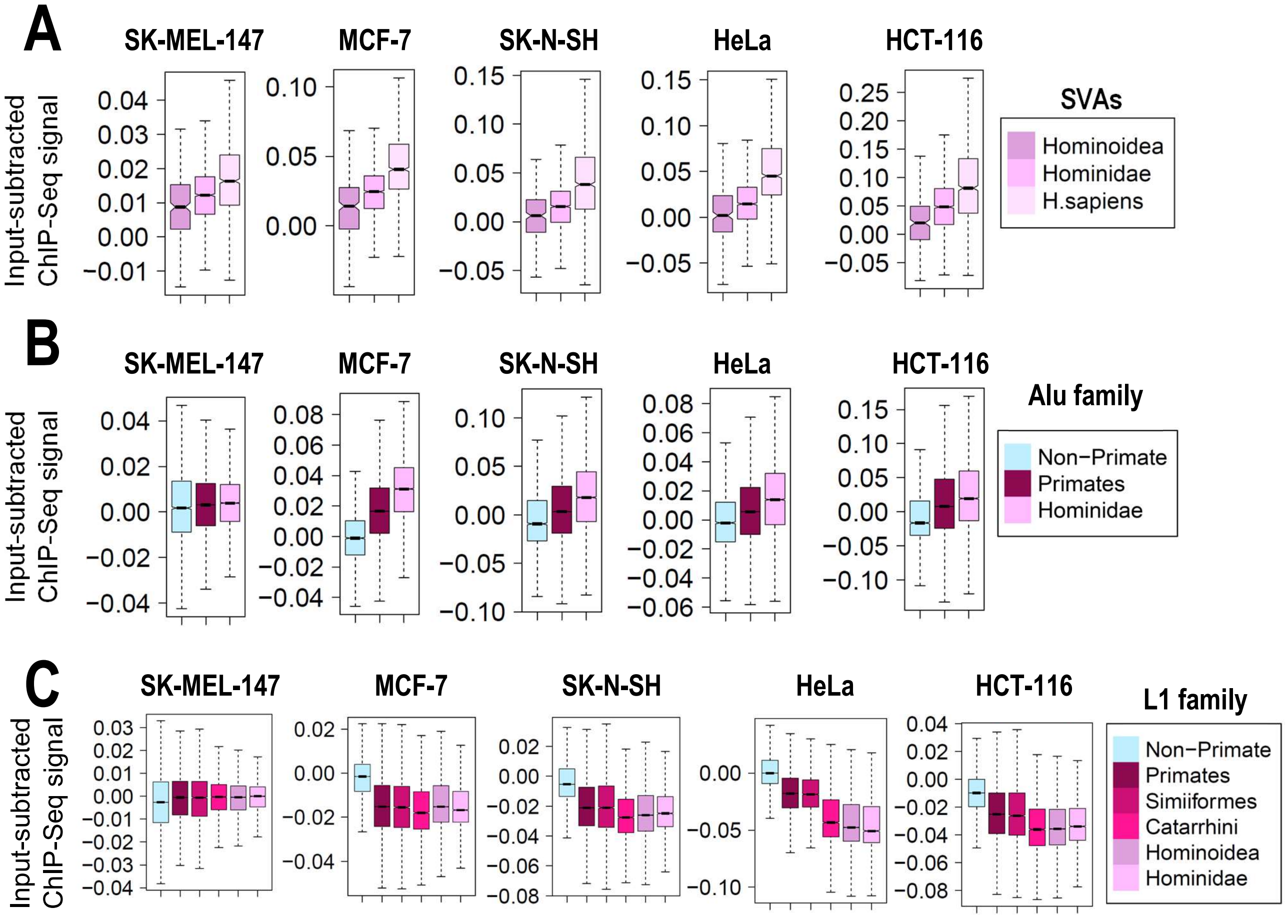


Figure R. 75. H1X abundance along SVAs, Alu and L1 evolution in different cell lines. Boxplots indicate the H1X Input-subtracted ChIP-Seq abundance in five cancer cell lines within repeats of **A)** Other Class (i.e. SVA families); **B)** SINE-Alu family and **C)** LINE-L1 family, classified according taxonomic clades.

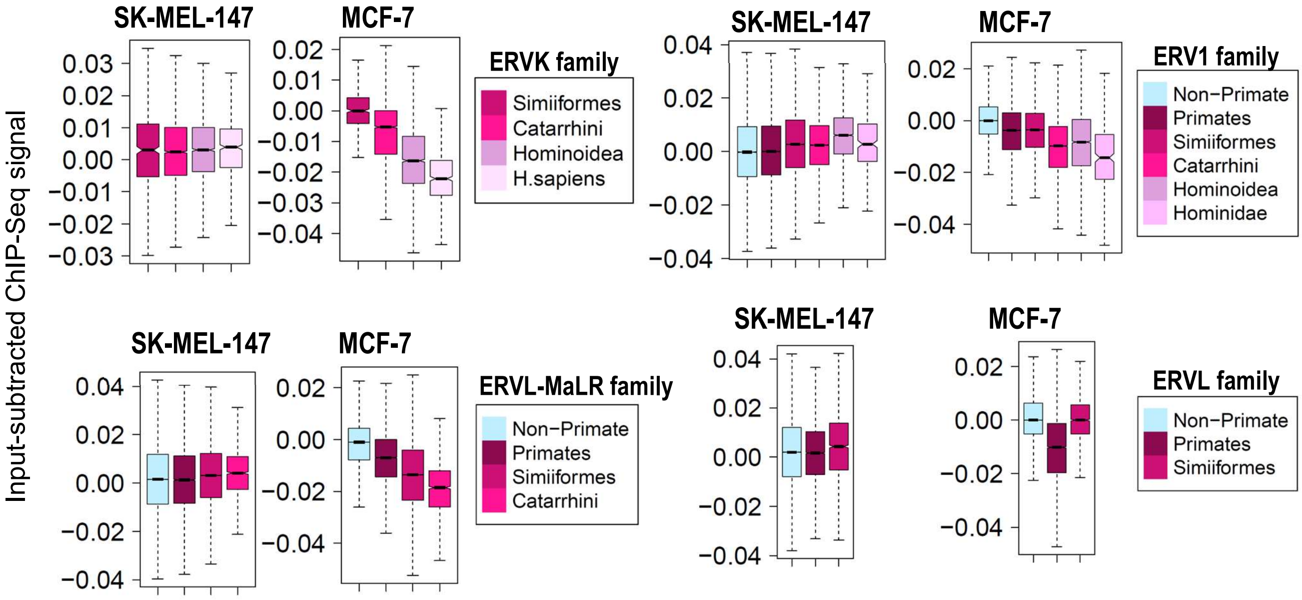


Figure R. 76. H1X abundance along LTR families evolution in different cell lines. Boxplots indicate the H1X Input-subtracted ChIP-Seq abundance within repeats of LTR families, classified according to taxonomic clades. H1X profiles from SK-MEL-147 and MCF-7 cell lines are shown. Notably, SK-N-SH, HeLa and HCT-116 cell lines show an analogous profile to MCF-7. LTR families ERV, Gypsy and LTR are not included because all repeats within these families are classified as *Non-primate*.

Discussion

In this section of the manuscript, the results obtained will be discussed in light of the current literature. Different perspectives or potential future experiments of interest will also be commented. The section will be divided into two major parts, according to the order presented in the Results section. However, determining whether the findings are extrapolatable or not to other cell models is important when comparing results with existing literature. Therefore, certain findings from Results Chapter 2 will also be integrated into the first section. Finally, a third section including a global view of the results is included.

1. Discussion about the differential distribution of six endogenous histone H1 variants in T47D breast cancer cells

1.1. Drawing the genome-wide profile of six endogenous somatic H1 variants

Study of H1 variants genomic distribution have been technically limited by the lack of specific ChIP-grade antibodies. Here, we have extensively assayed specificity and ChIP performance of different commercial H1 variants antibodies **(Figures R.1-R.5)**. Antibodies against endogenous H1.0, H1.2 and H1X were successfully used in previous ChIP-Seq studies (148,152). On the other hand, the selected antibodies against endogenous H1.3, H1.4 and H1.5 have never been used for ChIP-Seq before, to our knowledge. These antibodies now represent a reliable tool to overcome the everlasting technical limitation of H1 studies, allowing for the mapping of almost all H1 somatic variants, with exception of H1.1, which expression is restricted to certain tissues.

Using the mentioned antibodies, we performed the first genome-wide mapping of six endogenous H1 variants within a mammalian cell model. Of note, this represents the complete H1 somatic complement in T47D breast cancer cells, due to H1.1 is not expressed in this cell line. Previous studies by the lab demonstrated that endogenous H1.2 and H1X have different genomic patterns distribution (151,152). Now, we have observed that apart from H1.2; H1.0, H1.3 and H1.5 are also enriched at low-GC regions. On the other hand, H1.4 and H1X are more abundant at high-GC regions, although H1X correlation with GC content is higher. Thus, H1 variants in T47D can be segregated in two differentiated groups based on the local GC content **(Figures R.6-8)**.

Up to date, previous studies addressing differential distribution of multiple H1 variants within a cell line used overexpressed H1 proteins, both in mouse (168) and human cells (152,169). Regarding human studies, H1.1-H1.5 were profiled through DamID technique in IMR90 cells (169). In addition, previously mentioned publications from the lab used endogenous H1.2 and H1X mapping but also HA-tagged recombinant H1 variants (152). It is important to note that previous profiling of stably expressed H1.0 and H1.4 fussion proteins in T47D (fussed to an HA-tag at C-terminus) revealed their enrichment at high-GC regions (152). Both the drastic differential results from exogenous and endogenous H1.0 mapping, as well as the direct comparison between endogenous H1.2 and H1.2-HA profiles (152), point out the relevance of profiling endogenous linker histone variants.

1.1.1. The redundant nature of H1 variants

H1 content in T47D is estimated to be 9% for H1.0, 23% for H1.2, 13% for H1.3, 24% for H1.4 and 31% for H1.5, while H1X protein levels are minority (150). So, based on our ChIP-Seq results, most H1 protein is preferentially enriched at low-GC regions, which supports its role as a heterochromatic protein. It is worth mentioning that, although H1 variants are preferentially enriched at certain genomic regions, they are still present throughout the whole genome. Indeed, functional annotation also revealed common features for H1 variants. All H1 variants are abundant at heterochromatin and low-activity regions, which represent most of the genome (**Figure R.10**), supporting again their role as general repressors. In addition, all H1 variants are depleted from enhancers (**Figure R.12**) and, with exception of H1X, from promoters depending on their transcriptional status (**Figure R.10, R.11**). Compared to the rest of H1 variants, H1X is present at TSS of genes, particularly for the inactive ones. These observations point out the existent balance between H1 variants redundancy and specificity and highlight H1X as the most divergent linker histone, in terms of genomic distribution.

1.1.2. Considering different resolution levels to fully address H1 variants ChIP-Seq comparison

As said, although showing differential enrichments, H1 variants cover all over the genome. Because of that, conventional peak calling algorithms do not recapitulate the complete information of H1 abundance. For that reason, instead of determination of H1-enriched regions, the best approach to directly and systematically compare multiple H1 variants is the evaluation of their ChIP-Seq abundance in different genome segments. We used different genomic segmentations that underlie diverse biochemical information, such as 3D conformation (A/B compartments) or chromatin functionality (chromatin states or 'colors'), among others. However, it is also important to consider the length or resolution we are evaluating in each case (**Figure D.1**).

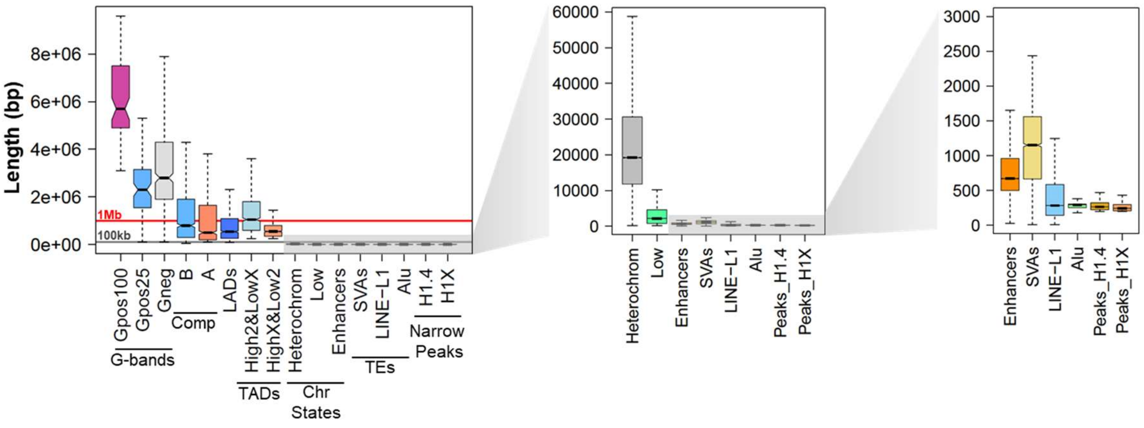


Figure D. 1. Length distribution of genomic segments used for evaluation of H1 variants ChIP-Seq abundance. Giemsa bands Gpos100 (low-GC), Gpos25 (high-GC), Gneg (Unstained, GC heterogeneous); Compartments and topologically associated domains (TADs) detected in our Hi-C experiments in T47D cells. TADs groups indicate: 25% TADs with the highest H1.2/H1X ratio (High2&LowX) and 25% TADs with the lowest H1.2/H1X ratio (Low2&highX); Enhancer regions are T47D enhancers from EnhancerAtlas. Bin-size segmentation of 100kb, which has also been used in this thesis, is indicated with a grey line. 1Mb-size is indicated with a red line.

H1.4 was thought to be associated with heterochromatic regions due to its reported binding to HP1alpha (but also other HP1 isoforms) through the methylation of its K26 residue (164). This is in apparent contradiction of H1.4 being enriched within the active compartment or more euchromatic genomic regions, as suggested by our analysis. However, it is important to note that this information is extracted from ChIP-Seq data inspection at the mega-base level, as denoted by G-bands **(Figure R.6)** or A/B compartments **(Figure R.7)** segmentations. Indeed, further analysis of the data revealed that, although being abundant within the active compartment, H1X and H1.4 were in fact highly enriched within repetitive elements located in the A compartment **(Figure R.13)**. This represents an example of how important is to consider different levels of resolution to fully compare H1 variants ChIP-Seq data **(Figure D.2)**.

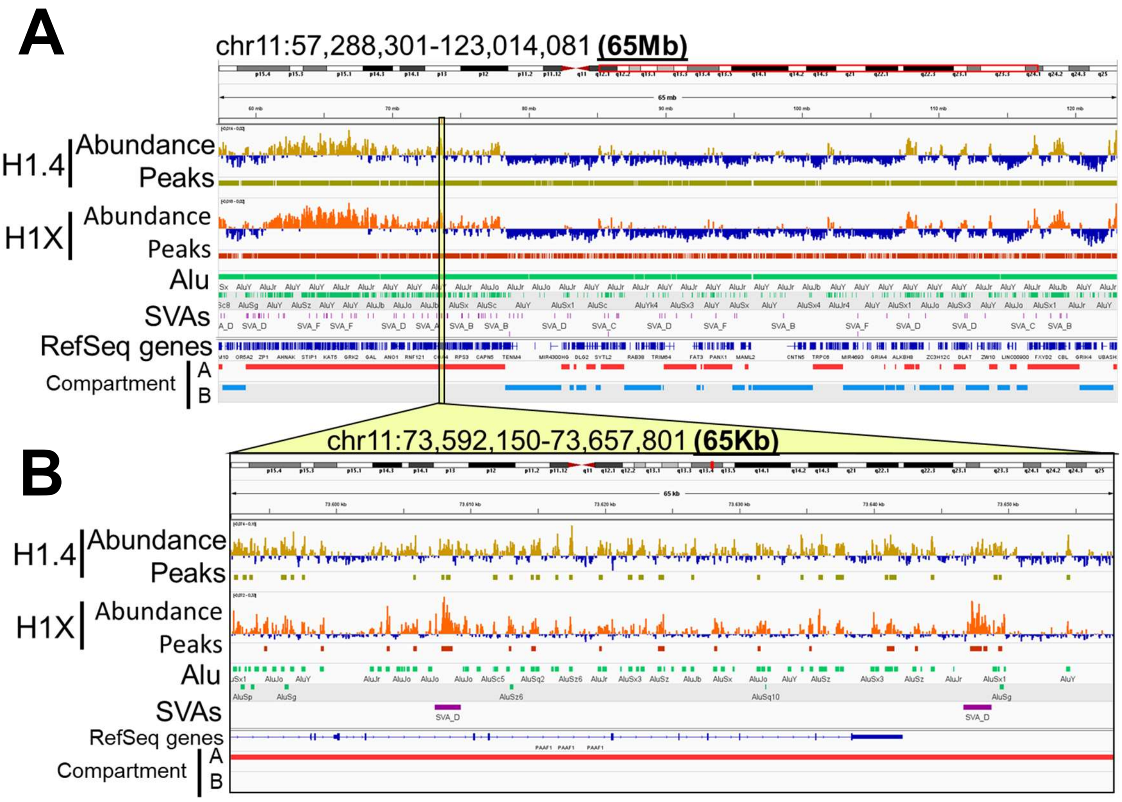


Figure D. 2. Browser capture of H1.4 and H1X ChIP-Seq profiles evaluated at different resolutions. **A)** H1.4 and H1X are enriched within active, A compartment regions. **B)** Further exploration reveals that both H1 variants are indeed highly enriched within repetitive elements, especially Alu and SVAs elements, which are preferentially located within A compartment **(Figure R.14)**. Input-subtracted ChIP-seq abundance and Narrow Peaks are shown for H1 variants.

Thus, huge mega-base domains of H1.4 and repressive HP1alpha are not expected to coincide. In agreement, immunofluorescence data revealed that HP1 isoforms foci do not overlap with H1.4 **(Figure D.3)**. On the contrary, we could hypothesize that H1.4 could bind HP1alpha and participate in local repression of transposable elements in T47D. Indeed, HP1alpha has been associated to transposon silencing through its KAP1-mediated recruitment (70,72). Importantly, HP1alpha recruitment has also been associated with repression of several evolutionary young transposable elements (198). Those repeats, including younger AluY elements, present high H1.4 abundance in T47D.

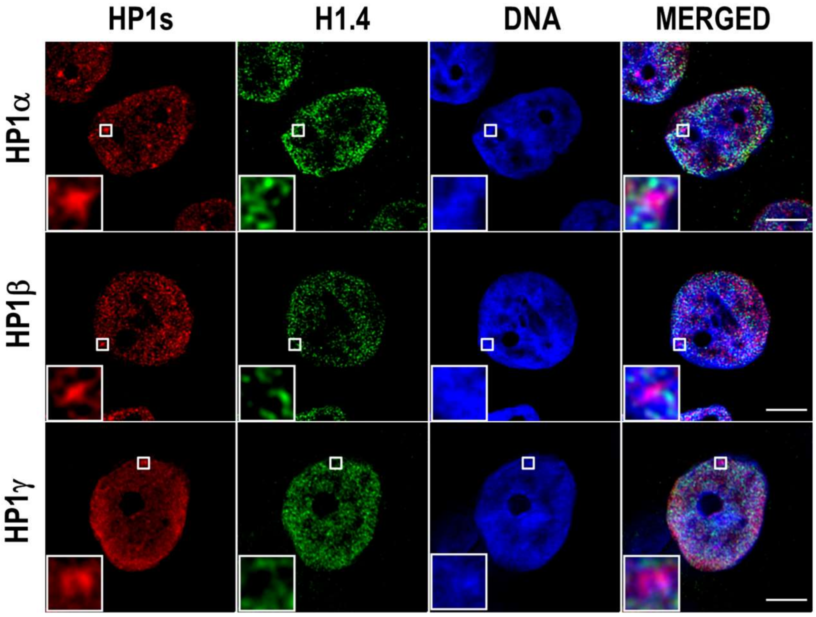


Figure D. 3. H1.4 does not co-localize with HP1 heterochromatic foci. Immunofluorescence of H1.4 and HP1 isoforms was performed in T47D cells. Zoom-in insets show HP1 foci, coincident with DNA condensation but not with H1.4 signal. Scale bar: 5µm.

1.2. H1 variants profiling within repetitive elements

To gain more insights into the binding profile of H1 variants, in Results section 1.3 (Chapter 1) we evaluated H1 variants ChIP-Seq abundance within repetitive elements. We found that H1 variants are differentially distributed within repetitive elements classes. While H1X and H1.4 are enriched in SVAs and SINEs, 'low-GC' H1 variants are more abundant at LINE, LTR, Satellite or DNA classes **(Figure R.15)**.

1.2.1. About H1 variants and nucleosome positioning

Narrow peaks can be called for some of the H1 variants (H1.4 and H1X in T47D, see **Figure M.5A**). Due to the fact of mapping linker histones and considering the short length of those narrow peaks (≈250bp) **(Figure M.5B)**, H1 peak calling is intrinsically biased by nucleosome positioning **(Figure D.4)**. Nucleosome positioning refers to the positions of the nucleosome core particles relative to the DNA sequence in a cell population. A well-positioned nucleosome occupies similar positions in all cells. In contrast, a non-positioned nucleosome occupies different positions across the cell population (199,200). That is, two factors are needed to obtain a narrow H1 peak: 1) nucleosome is well-positioned in the cell population and 2) that nucleosome is consistently bound by H1.4 or H1X. So, by narrow peak calling we are not analyzing the complete H1.4/H1X enriched regions, but those where a very 'sharp' ChIP-Seq enrichment profile is observed.

Interestingly, it has been reported that ~80% of strong nucleosomes overlap with repeats (189). Moreover, these strong nucleosomes are preferentially located within evolutionary young Alu and LINE-L1 elements. Accordingly, >90% of H1.4 or H1X narrow peaks rely on repetitive elements **(Figure R.16A)** and they are biased towards more recent repeats of the Alu and L1 families but also SVA retrotransposons **(Figures R16B,C)**. Strong nucleosome positioning has been postulated as another layer of transcriptional repression, masking the access to the transcription machinery. Now, we know that these strong nucleosomes are in addition loaded with H1.4 and/or H1X and this linker histone presence could contribute to the cooperative silencing of TEs.

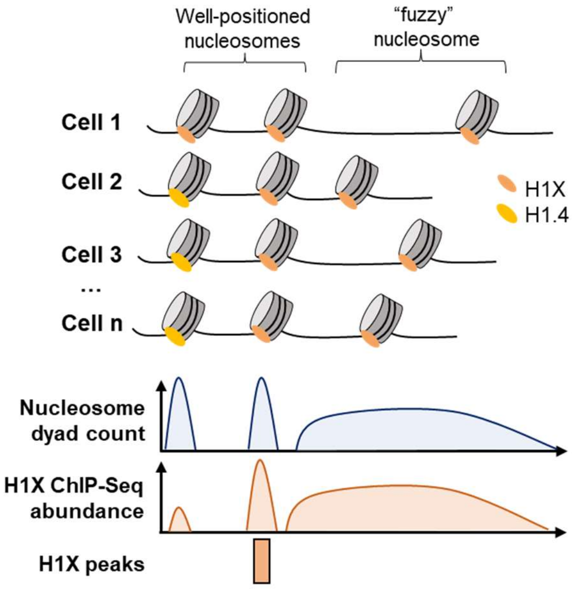


Figure D. 4. Nucleosome positioning and H1 variants peak calling. In T47D, narrow peaks can only be called for H1.4 and H1X but not for the rest of the variants. As we are mapping linker histones and narrow peaks have a length of ≈200-300bp **(Figure M.5B)**, peak calling is biased by nucleosome positioning in the cell population. H1 narrow peaks cannot recapitulate the complete H1 variants enrichment profile but provide a meaningful information of precise H1 binding.

1.2.2. H1.4 and H1X preferred TEs: a question of numbers?

The presence of H1.4 or H1X in those well-positioned nucleosomes is repeat class-dependent. Thus, H1X is preferentially associated to SVA retrotransposons while H1.4 is more related to Alu elements **(Figure R.16B,C, R.17A, R.19B,C)**.

H1.2 and H1.4 are universally expressed in all cells tested (194–197). Moreover, we have added H1X as a universally expressed H1 variant **(Figures R.58-R.59)**, but H1X protein levels are much lower compared to H1.2 or H1.4 (*data not shown*). Related to this, H1.4 is associated to Alu elements, which successfully expanded through primates' evolution and represent almost 11% of the human genome, with more than one million Alu copies (57). Indeed, Alu and LINE-L1 families are the largest families of TEs in the genome **(Figure R.14B, M.6C, Table M.6)**. On the other hand, H1X (a minor contributor to the total H1 content) is enriched within a minority class of repeats. SVA repeats (≈3600 copies) accounts only for the 0.14% of our genome **(Figure R.14B)**.

Worth mentioning, H1X is not only universally expressed within human cell lines, but also is universally enriched within SVAs **(Figures R.71-R.72)**, as further discussed later. Of note, H1X is also universally enriched within Alu repeats, as observed for T47D, although its highest enrichment is constantly found at SVAs, specially at most recent ones. In this sense, whether H1.4 is universally enriched within Alu elements or other H1 variant may exert this role is still unknown.

1.2.3. About Histone H1 variants genomic distribution and TEs evolution

We have shown an unprecedented relationship between H1 variants abundance and evolutionary age of TEs (**Figures R.22-R.25, R. 73-R.76**). Thus, in T47D, H1X and H1.4 are enriched towards most recent TEs of the human genome, including SVAs, Alu, but also at youngest repeats of LTR and LINE-L1 classes/families (**Figures R.22-R.25**). On the other hand, H1.2/H1.3/H1.5/H1.0 are more enriched at oldest repeats.

Throughout evolution, waves of retrotransposon insertions have invaded mammalian genomes. For each invasion, the host genome finds a repressive mechanism to prevent the retrotransposon transcription. A clear example relies on the evolutionary coincidence between expansion of KRAB-KZNF gene family and TE insertions (87–89). However, this so-called 'arms race model' (73) does not fully explain the co-option between inserted TEs and the host genome and their integration in the regulatory networks (201).

In the case of H1 proteins, H1 subtypes divergence occur prior mammalian radiation. That is, the seven somatic H1 variants present in humans are also present in other mammal species (144). For that reason, H1.4 or H1X proteins did not represent an example of 'arm race mechanism', as the proteins did not emerge themselves to repress primate-invading Alu elements or hominid-specific SVAs. Although these H1 variants were already functional in the ancestral genomes, TEs invasion could modulate their functionality or genomic distribution in a lineage or species-specific manner. Indeed, this scenario may also occur with KZNF proteins, as there are many examples of TEs being silenced by KZNF proteins emerged before their invasion (201).

ChIP-Seq experiments of H1 variants in different species would shed light to the interplay between functional adaptation of H1 variants and TE repression through genome evolution. Those experiments would explore whether T47D 'low-GC' H1 variants are enriched in the same repetitive elements, as older LINEs, LTRs or DNA repeats, in species with a common ancestor. Moreover, they would point out whether H1X and H1.4 abundance universally increase towards species-specific TEs across clades.

Examination of other hominoid genomes is intriguing. SVA retrotransposons evolved through hominid lineage, being SVA_E and SVA_F restricted to humans (60,61). Analogously, other species-specific SVAs exist in other hominids (61,202–205). Thus, histone H1 relationship with evolution would be reinforced if H1 variants abundance patterns in other SVA lineages resemble those observed in humans (**Figure D.5**). This is especially interesting for H1X, whose enrichment and increase towards more recent SVA families seems to be universal among all human cell lines examined (**Figure R.71-R.72**). Other SVA-related-retrotransposons exist in *Hominoidea* species and are also hypothetical candidates to present high H1X abundance. This is the case of LAVA repeats (L1-AluS-VNTR-Alu), which are found in gibbons, where SVA elements have not expanded as successfully as in other hominoid species (203–205).

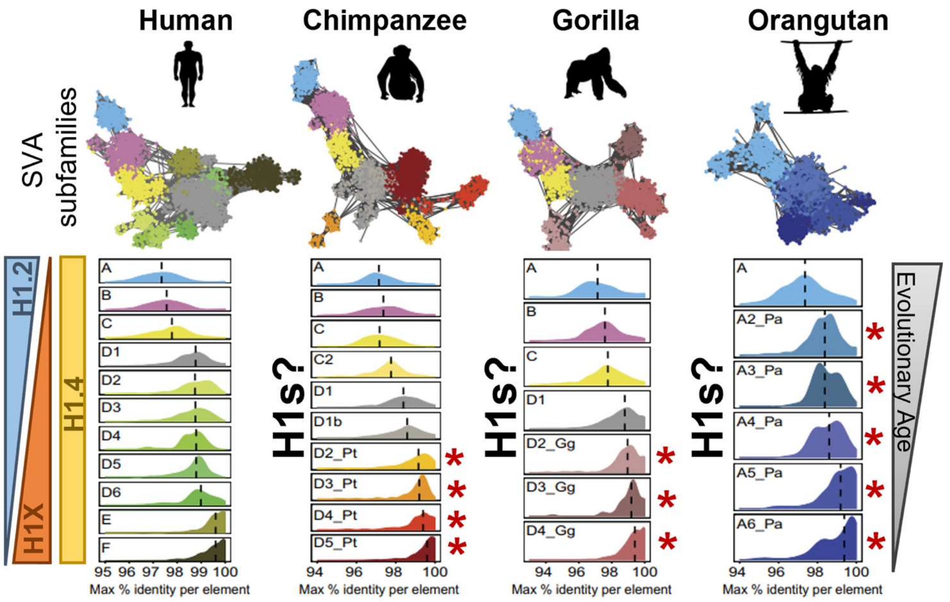


Figure D. 5. SVA subfamilies in hominoid genomes. Putative H1 variants ChIP-Seq studies would elucidate whether H1 abundance patterns observed in human cells analogously evolved in other hominoid species. SVAs evolutionary lineage in each genome is shown. In the bottom panels, SVA subfamilies are ranked from older to more recent subfamilies in each genome, based on divergency levels of the copies. Red asterisks mark species-specific SVA subfamilies in non-human primates. T47D H1X abundance mapped on these 11 SVA subfamilies is evaluated in **Figure R. 18**. Figure adapted from Levy et al. (61).

Moreover, study of H1 variants genomic distribution in non-primate backgrounds is potentially interesting. In mouse embryonic stem cells (mESCs), ChIP-Seq of tagged H1.0 and knock-in of H1c and H1d (mouse orthologous of H1.0, H1.2 and H1.3) have been performed (168). Regarding repetitive elements, the three H1 variants were found enriched within satellites and LINEs. Although limitations of exogenous H1 mapping have already been commented, those results somehow could be in concordance with those observed in T47D cells. In T47D, H1.0/H1.2/H1.3 are also enriched in satellite and LINE classes **(Figure R.15)**.

SVA and Alu elements do not exist in mouse, so it is unknown in which repetitive elements H1X and H1e (orthologous of H1.4) would be highly enriched in mouse. In mouse, B1 repeats represent the analogous to human Alu elements and constitute ≈7% of the genome. Alu and B1 are thought to have the same ancestral origin and later independently evolved after the primate-rodent evolutionary split (57). Related to this, H1X but especially H1.4 abundances, although increasing towards more recent Alu, they are enriched within the whole Alu family **(Figure R.19A, R.24B)**. This enrichment is also observed in ancestral Alu monomers that gave rise to dimeric Alu subfamilies (see H1X/H1.4 enrichment in Alu Non-Primate in **Figure R.23A**). Based on these results, we could hypothesize that in rodents, H1 variants genomic distribution within B1 repeats could have evolved similarly. Worth mentioning, ChIP-grade antibodies used here fail to recognize mouse H1 variants **(Figure D.6)**, so new antibodies would be needed to perform comparative genomics across different species. Whether these antibodies recognize H1 variants in more evolutionary proximal species, as other primate species, is still unknown.

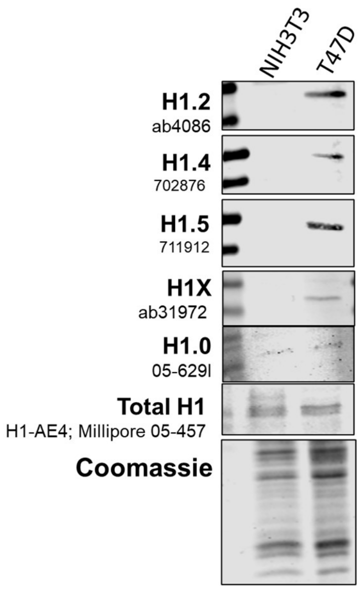


Figure D. 6. Immunoblot of H1 variants in mouse cells. Total histones were extracted from NIH3T3 mouse fibroblasts and T47D human breast cancer cells. ChIP-grade H1 variants specific antibodies did not recognize mouse H1 variants. Immunoblot of anti-H1.0 should be repeated to verify antibody performance. We have not explored H1 variants expression in NIH3T3. Although H1.2/H1.4 and H1X are in principle expressed in every cell type, it is also plausible that absence of detection of H1.0 or H1.5 is due to no expression of these variants.

Beyond hypotheses, the balance between conserved TE regulatory mechanisms across species and species-specific TE regulation is poorly understood. On the one hand, a comparative study of human and chimpanzees iPSCs revealed that H3K9me3 profiles in orthologous TEs (including SVA, LTR and LINE-L1 elements), were overall conserved between both species (206). This study suggests that limited inter-species differences in TE silencing mechanisms exists in primates. On the other hand, in human differentiated cells, TEs are reported to act as enhancer elements in a cell-type specific manner (75,76,79,207). Thus, the comparative study of H1 variants genomic distribution within TE classes in different species emerges as another epigenetic layer to consider in the question. Similarly, the differential interplay of H1 variants abundance with other described repressive mechanisms such as histone modifications or DNA methylation is also an interesting matter of study.

1.2.4. About satellites repeats and mapping limitations

We have not deeply studied the differential H1 variants distribution at satellite repeats due to technical limitations. Satellite class represent the 0.4% of the human genome (**Figure R.14B**), but almost 50% of satellites are located within problematic regions where read mapping is inaccurate, such as pericentromeric regions. Thus, the removal of these regions is essential to properly interpretate results (183). Notably, the rest of repeat classes are largely unaffected when these problematic regions are excluded from the analysis. Considering this, all H1 variants were somehow abundant within satellite class but, compared to other repeat classes, enrichment was higher in 'low-GC' H1 variants (**Figure R.15**). This agrees with satellites being preferentially located at the B compartment (**Figure R.14A**).

Besides, upon multiple H1 depletion (ie H1.2+H1.4 KD), alpha-satellite (SATa) but also pericentromeric SST1 and subtelomeric D4Z4, become upregulated (126). Single H1X depletion also led to moderate upregulation of SATa and SST1 (**Figure R.30**) while H1.0 depletion trigger SATa expression *(data not shown)*. These results reinforce the notion that different H1 variants could participate in satellites regulation.

Recently, a new assembly of the human genome was released by the Telomere-to-Telomere (T2T) Consortium (208). Thanks to long-read sequencing techniques, T2T assembly completed the previously unaddressed 8% of the human genome. Some of the largest filled gaps include centromeric and pericentromeric satellite arrays **(Figure D.7)**. Indeed, satellite base-pairs have increased a ≈97% relative to prior hg38 assembly. Moreover, transcriptional and epigenetic studies of these new assembled repeat elements were already performed (209). This serves as a proof-of-concept of future long-read sequencing experiments to overcome the mapping limitation of short-read data to TEs, including H1 variants studies.

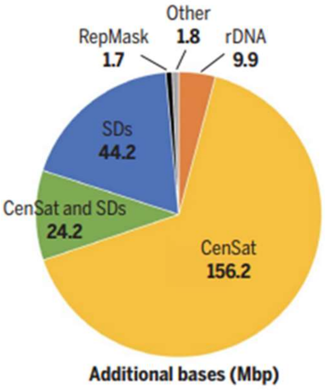


Figure D. 7. Additional bases in the T2T-CHM13 assembly relative to GRh38 (hg38). CenSat (Centromeric Satellite), SDs (Segmental Duplications), RepMask (RepeatMasker), rDNA (Ribosomal DNA). From Nurk et al. (208)

1.2.5. Histone H1 binding serves as an additional factor in the cooperative repression of TEs and as a transcriptional modulator

We have shown that H1X or H1.4 depletion cause a moderate transcriptional activation of young TEs in which these variants are enriched **(Figure R.30)**. Moreover, TEs desrepression was accentuated when Dox and aza treatments were combined in H1Xsh cells **(Figure R.31)**. That is, when cells have concurrently compromised DNA methylation and H1X levels. Accordingly, it has been shown that H1 and DNA methylation cooperatively silence transposons in plants (210). While H1 depletion disperses heterochromatin, only weakly de-repress TEs. However, TEs expression is triggered upon combined H1 depletion and DNA hypomethylation. Moreover, several reports support that repression of TEs is mediated by multiple co-existing mechanisms (67,68,71,81). Thus, we suggest that histone H1 variants are an additional TE repressor participating in this cooperative regulation.

The study of the interplay between H1 variants and other repressive mechanisms, such as histone modifications or KRAB-ZNF-KAP1 axis, would clarify the differential molecular function of H1 variants in repetitive elements regulation. Indeed, we have observed that ZNF91 is upregulated upon H1X KD, but not upon H1.4 KD **(Figure R.30)**. ZNF91 is a master repressor of SVA elements and some young LINEs of the L1PA lineage (73). Moreover, derepression of SVAs is reported to cause upregulation of neighboring ZNF genes (74). It is plausible to propose that H1X can be involved in this feedback loop regulation. Hypothetically, SVA derepression mediated by H1X depletion could incite upregulation of ZNF91 to control SVA activation. This scenario could explain why just mild upregulation of SVAs is shown in H1X KD. Combined H1X and ZNF91 depletion emerges as an interesting background to test whether this feedback occurs.

Upregulation of young TEs upon H1X KD could have different consequences on the nuclear homeostasis. It has been reported that LTR repeats can act as cryptic promoters following treatment with DNMT and HDAC inhibitors (68). Concretely, these induced non-annotated TSSs frequently overlap with LTR12C repeats. Interestingly, these LTR12C-derived transcripts are spliced into protein-coding exons, originated truncated transcripts lacking the 5' regions of the canonical gene mRNA. Our analysis shows that H1X is enriched within LTR12C repeats, not only in T47D but also in other cell lines evaluated (**Figure R.20C, R.74B**). Moreover, we have recently detected that upregulation of deregulated genes upon H1 KD does not occur from the annotated TSS (**Figure R.28-R.29**), suggesting that H1X prevents cryptic transcription. Many TEs are found near the putative cryptic initiation sites (see **Figure R.28**). The possibility that H1X-mediated desrepression of TEs could trigger this cryptic transcription needs to be addressed. Mechanistically, other epigenetic alterations could be involved too. For instance, intragenic DNA hypomethylation also causes cryptic transcription (14). As we have observed certain synergy between H1X depletion and DNA hypomethylation in TEs activation, DNA methylation landscape and its possible association with the intragenic cryptic transcription in H1X KD are worth to be investigated.

Regarding histone H1 and transcription dynamics, RNApol II initiation landscape is impaired in H1 TKO mESCs, linking H1 depletion to transcription initiation from non-promoter positions (133). We need to evaluate whether our findings regarding cryptic transcription upon H1X depletion can be extended to other H1 variants KDs or remain H1X-specific. Based on the observation in H1 TKO mESCs, it could be that depletion of other H1 variants or even a more reduced H1 content (i.e. combined depletion of several H1s) trigger intragenic transcription of different gene subsets in human cells. If that were the case, although exhibiting a similar outcome, the mechanisms by which each H1 KD mediate that cryptic transcription could be different. In addition, it is important to highlight that, in the case of H1X depletion, total H1 content remain almost unaltered, as H1X represent a small fraction of total H1. For this reason, transcriptional consequences occurred upon H1X depletion, cannot be attributed just to a dysbalanced H1 content, but rather point to a role of H1X in chromatin regulation.

1.3. H1 variants occupy differential nuclear regions in interphase cells

To dissect H1 variants differential distribution of H1 variants T47D we also used a microscopy approach (section 1.4 of Results Chapter 1). Imaging data is useful to study spatial distribution patterns within the nucleus of individual cells, which are intrinsically related to genome functionality. On the other hand, through ChIP-Seq experiments we can explore the similarities and differences between H1 variants in a cell population at high resolution genome-wide. Because of that, combination of both approaches gives a more complete information that cannot be obtained by only one type of experimental data.

We found that H1.2, H1.3 and H1.5 and to a lesser extent H1.0 are enriched towards nuclear periphery. On the other hand, H1X and H1.4 are distributed throughout the nucleus with H1X being highly enriched in nucleoli **(Figure R.32)**. Moreover, colocalization of the variants revealed that H1.0 better colocalizes with H1.2/H1.3/H1.5 compared to H1.4 or H1X **(Figure R.34)**. These observations fit with the two differential clusters observed by ChIP-Seq but also add differences between the variants in terms of spatial distribution.

Similar to the importance of considering resolution for drawing interpretations in ChIP-Seq experiments (discussed in section 1.1.2), it is equally crucial in microscopy. Indeed, the term 'co-localization' is intrinsically linked to the optical resolution of the image. Thus, we applied super-resolution techniques to better address nuclear localization of H1 variants. In concrete, we used Super-Resolution Radial Fluctuations (SRRF). Super-resolution imaging of H1 variants revealed that H1.2/H1.3/H1.5/H1.0 coincide more with DNA pattern compared to H1.4 and H1X, confirming their segregation in two groups **(Figure R.35)**. Of note, detected DNA fraction, as emphasized in the results section, represents the DNA in a more compacted state, relative to neighboring regions. This is the underlying reason why a higher colocalization with DNA signal is linked to an enrichment in the inactive nuclear regions.

1.3.1. About three-dimensional distribution of H1 variants and chromatin folding

Emerging evidence from super-resolution microscopy indicates that nucleosomes are grouped in heterogenous nanodomains termed 'clutches' (6). Moreover, topologically associated domains (TADs) represent structural chromatin folding units at the sub-megabase scale (95–97). We have shown that H1 variants form spatially separated nanodomains throughout the nucleus, visualized as a 'punctuate' signal by super-resolution imaging, but with the afore-mentioned differential variant-specific local enrichments **(Figure R.35)**. Similarly, super-resolution imaging of core histone H2B also present this clustered pattern in human fibroblasts (6). It is important to mention that nucleosome clutches were originally defined using STORM technique (6), whose resolution is higher than the one achieved by SRRF. For that reason, we favour the idea that nanodomains formed by H1 variants would be more equivalent to TADs or sub-TADs rather than to nucleosome clutches. However, with our data we cannot establish a direct correlation between both types of 3D structures.

By conventional immunofluorescence, preferential localization of H1.0 with other ‘low-GC’ H1 variants over H1.4/H1X was observed (**Figure R. 34**). However, this preferential colocalization was not evidenced at the super-resolution level and all H1s colocalized similarly with H1.0 (**Figure R. 36**).

This observation may suggest that domains spatially arranged at the 3D level are homogeneously marked by a certain H1 variant and not by random H1 variants. Nevertheless, in a cell population those nanodomains could be marked by different H1 variants. This could explain why at single-cell level we lost preferential co-localization of ‘low-GC’ H1s with H1.0 (compared to H1.4/H1X) while by ChIP-Seq data, mega-base domains of ‘low-GC’ H1 variants coincide (**Figure D.8**). If those nanodomains were homogeneously marked by the same H1 variant in the cell population, we would observe differential enrichments between H1 variants belonging to the same GC cluster, even at the mega-base level. This apparent intra-population ambiguity may be indicating a structural role of ‘low-GC’ H1 variants and emphasizes the existing partial redundancy among certain H1 variants. In fact, shifts on the ChIP-Seq H1 variants distribution tend to coincide with TAD borders and H1 variants are more homogenous within the same TAD than between TADs (see browser captures in **Figure R.52**, **D.8** and (116)). This observation also highlights the relationship between H1 distribution and the structural properties of chromatin.

However, H1 variants co-localization has only been explored in comparison with H1.0, which is enriched at ‘low-GC’ regions. For that, these 3D distribution patterns may be preferentially occurring at more closed chromatin or large heterochromatic domains but not especially at the active compartment. Indeed, differences between H1.4 and H1X ChIP-Seq profiles at the mega-base level are greater compared to those observed between different ‘low-GC’ H1 variants (as an example see **Figure R.6**, **R7**). Related to this, extensive super-resolution imaging co-localization experiments between all the H1 variants are of great interest, but other strategies such as primary antibody fluorescence labelling must be used.

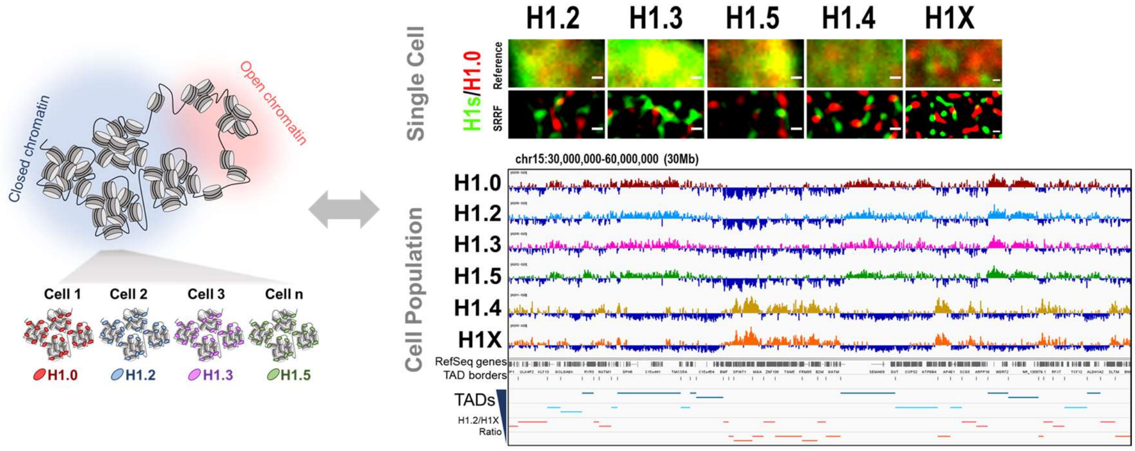


Figure D. 8. Proposed model for H1 variants composition in chromatin structural domains. By confocal immunofluorescence H1.0 better co-localizes with other ‘low-GC’ H1 variants compared to ‘high-GC’ H1s. However, by super-resolution microscopy this preferential localization is lost and all H1 variants show a similar co-localization with H1.0. At cell population level, ChIP-Seq profiles show that large enrichment domains of ‘low-GC’ H1 variants coincide. We thus propose that in single-cell, three-dimensional heterochromatic nanodomains are preferentially marked by a certain ‘low-GC’ H1 variant. This H1-marking may occur stochastically in different cells, explaining the observed ChIP-Seq results.

1.3.2. H1 variants and LADs

H1.2, H1.3 and H1.5 are highly enriched at the nuclear periphery in T47D cells **(Figure R.32)** but also in all cell lines analyzed **(Figure R.64-R.65)**. Super-resolution microscopy revealed that these H1 variants form an adjacent layer to Lamina **(Figure R. 39)** and highly co-localize with H3K9me2 **(Figure R.40-R.41)**.

H3K9me2 is not only a universal component of LADs, but also it is indispensable for peripheral heterochromatin anchoring to the nuclear lamina (106). Due to the also universal H1.2/H1.3/H1.5 enrichment in LADs, these H1 variants could be postulated as potential orchestrating factors for chromatin tethering to the lamina. Thus, future functional studies should address how single or combined depletion of these peripheral H1 variants affects chromatin organization at the nuclear periphery in different cell models.

Proper chromatin-lamina interactions are crucial to maintain chromatin dynamics (211–214). LADs detachment through Lamin B1 KO in human cells led to abnormal segregation of chromosome territories and A/B compartments, as well as global chromatin decompaction (212). Actually, we have shown that H1.2 depletion in T47D cells also led to a global chromatin decompaction **(Figure R.49)**. Furthermore, LADs detachment from the nuclear periphery occurs upon cellular senescence (211,215). While loss of histone H1 has been found in senescence (216), whether H1 depletion is a driver of the lamina-associated structural changes occurring in senescence is unknown. Similarly, whether this relationship between senescence and H1 loss is variant-specific or is preferentially associated to those H1s enriched within LADs, has not been studied yet.

Interactions between the nuclear lamina and LADs are disrupted at early stages of mitosis and re-established upon mitotic exit. In general, mitosis involves large structural reorganization of chromatin (217) that is accompanied by eviction of multiple chromatin factors from DNA (218). On the other hand, factors that persist attached to chromatin, including multiple histone variants and histone modifications (219), are suggested to act as spatial ‘bookmarks’. H3K9me2, through a phospho-methyl switch (H3K9me2S10p), is reported to safeguard positional information of LADs through mitosis (106). Indeed, we have found that interphasic ‘low-GC’ H1 variants persist more attached to chromatin during mitosis, compared to ‘high-GC’ ones **(Figure R.37-R.38)**. Moreover, H1.3 and H1.5, which are highly associated to LADs in interphase, persist in the peripheral layer of mitotic chromosomes **(Figure R.37B,C)**, showing an analogous profile to the one reported for H3K9me2 (106), depicted in **Figure I.13**. On the other hand, H1.2 attachment to mitotic chromatin is regulated by phosphorylation at early mitotic stages, but interestingly, H1.2 layer re-associates to the forming lamina upon mitotic exit **(Figure R.37D, R.38B)**. Due to the strong similarities observed for H1.2, H1.3 and H1.5 with H3K9me2 in both interphase and mitosis, investigation of the possible role of these linker histones as 3D positional ‘bookmarks’ of LADs is intriguing.

1.3.3. Nucleolar H1X and its possible functions

H1X nucleolar enrichment has already been described in previous works (151,220). In addition, other histone variants have been found at nucleoli, including testis-restricted H1T linker histone (221) or certain core histone variants (35,36,222)

Importantly, nucleolar localization of H1X persists after inhibition of RNApol I by Actinomycin D **(Figure R.46)**, which was also reported previously (220). On the contrary, Nucleophosmin is totally redistributed upon ActD treatment while DNA and nucleolus structure are altered. The fact that H1X localization is mostly independent on nucleolus functionality led us to suggest that it could execute a structural role. Indeed, this behavior was previously reported for other nucleolar components, that upon ActD treatment were retained in the central body of the reminiscent nucleolus (223). In the cited publication, authors discussed that this could indicate a function in retaining the structural integrity of the nucleolus.

The nucleolus is a membraneless organelle formed through liquid-liquid phase separation driven by multivalent interactions of its components (224). Several molecular features are known drivers for phase separation, including protein-protein or protein-RNA interactions and highly intrinsically disordered regions (225). Indeed, nucleolar proteins are highly disordered compared to cytosolic proteome. Among nucleolar sub-compartments, proteins localized to the nucleolar rim are the most disordered (226). We found that nucleolar H1X is enriched, although not limited, at the nucleolar rim, adjacent to the inner side of Nucleophosmin layer **(Figure R.43-R.44)**. Histone H1 proteins have a well-known highly disordered structure and have been shown to phase separate *in vitro* (129–131). However, the functional relevance of H1 variants as promotors of phase separation in living cells have not been explored.

It is important to mention that, upon H1X depletion, the characteristic nucleolar ring-like pattern of Nucleophosmin remains unaltered *(data not shown)*. Thus, H1X seems not indispensable for maintaining nucleolar structure, but most probably emerge as an additional contributor to the multiple interactions that formed liquid-liquid scaffolds. Besides, we need to consider that in our model H1X is not being completely depleted. Although H1X KD shows a high efficiency of H1X depletion **(Figure R.1B, R.42)**, the remaining H1X could be sufficient to maintain nucleolar structure.

Another appealing possibility is that H1X participates in the silencing of transcriptionally inactive rDNA. While active rDNAs repeats are located in the nucleolar interior, inactive rDNA repeats are in the nucleolar periphery. Although we have not extensively explored this possibility, preliminary RT-qPCR indicates that a slight upregulation of rRNA occurred upon H1X ablation **(Figure D.9)**. Notably, upregulation was already detected for pre-rRNA transcripts (ITS2 in **Figure D.9**). The concrete mechanism by which H1X lost could trigger the expression of rRNA is unknown. It could be a direct effect of H1X displacement from inactive rDNA repeat chromatin. In fact, in mouse fibroblasts histone H1 has already been implicated in direct repression of inactive ribosomal chromatin (221). However, the cited study evaluated total histone H1 pool not individual contributions of H1 variants. On the other hand, and related to the putative structural role of nucleolar H1X, H1X depletion could partially (but not completely) disrupt nucleolar assembly through loss of certain protein-protein and/or protein-RNA interactions that ultimately affect ribosomal transcription.

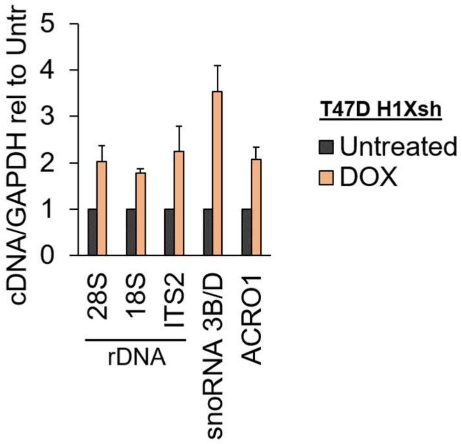


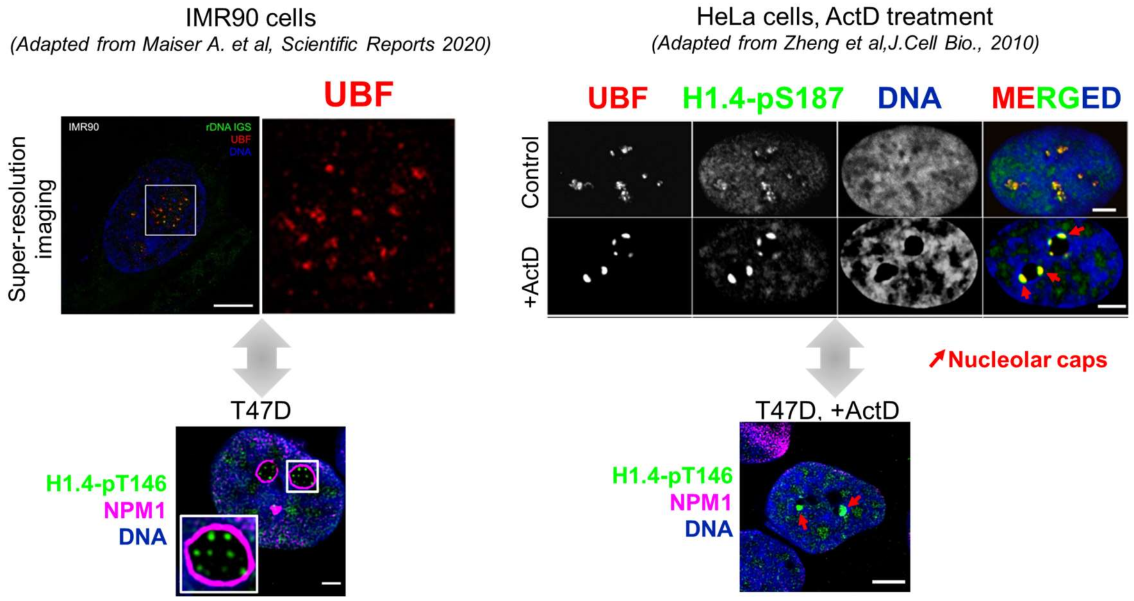
Figure D. 9. H1X could have a role in nucleolar function. Preliminary RT-qPCR analysis showed a mild upregulation of rDNA transcripts and nucleolar-associated RNAs upon H1X KD. Different regions of rDNA repeat were evaluated, including those encoding for 28S and 18S and ITS2. ITS2 region is only present in pre-rRNA 45S, which is later processed. A small-nucleolar-RNA (snoRNA 3B/D) and acrocentric satellite (ACRO1) were also evaluated, as they are related to nucleolar functionality. ACRO1 is found in the short arm of acrocentric chromosomes, where nucleolus organizer regions are located.

1.3.4. H1 variants post-translational modifications represent an additional layer in H1 heterogeneity

In addition to H1X, immunofluorescence analysis revealed the presence of specific H1 post-translational modifications (PTMs) in the nucleoli of interphase cells. Two different phosphorylations in the C-terminal domain of H1.2 and H1.4 were analyzed, with both highly increasing their relative abundance in early mitosis stages **(Figure R.38A)**. However, they are still present in interphase with characteristic patterns **(Figure R.44)**.

In interphase, H1.2-pT165 was almost exclusively restricted to the whole nucleoli, while H1.4-pT146 was present both at nucleoli and the non-nucleolar compartment of interphase cells. H1.4-pT146 nucleolar signals resembled those reported for rDNA transcriptional machinery components, such as UBF (192), suggesting that they represent active rDNA transcription. Upon ActD treatment, H1.4-pT146 seems to mimic the characteristic redistribution pattern of the rRNA transcription machinery whereas H1.2-pT165 relocates to the nucleoplasm **(Figure R.46)**. This result indicates that, nucleolar localization of both phosphorylated H1s is dependent on a proper nucleolus functionality, unlike H1X. Because we are missing direct co-immunostaining experiments with some transcriptional rDNA-associated factor to fully ensure the apparent association of H1.4-pT146 with active rDNA transcription, **Figure D.10** depicts the similarities found between our H1.4-pT146 results and those reported in bibliography for UBF. In general, these findings suggest that specific phosphorylated H1 variants may play a role in regulating nucleolar functions, particularly in rRNA transcription.

These observations agree with the existent evidence about H1 phosphorylation. Firstly, histone H1 variants are highly phosphorylated in mitosis (153–157). Secondly, other site-specific phosphorylated H1.2 and H1.4 have been found at nucleoli of interphasic HeLa cells, being H1.4-pS187 preferentially associated with transcriptionally active rDNA (154). Indeed, later ChIP-Seq analysis of H1.4-pS187 revealed its more global association with gene activation, not only at active rDNA (161). H1.4-pS187 was present around TSS of active genes and enriched at those responsive-genes upon estradiol treatment. This distribution is in total opposition of total H1.4 being depleted from active genes, as expected, and as also observed in **Figure R.11**.


IMR90 cells
(Adapted from Maiser A. et al, Scientific Reports 2020)

HeLa cells, ActD treatment
(Adapted from Zheng et al,J.Cell Bio., 2010)

Figure D. 10. H1.4-pT146 seems to be associated to active rDNA transcription. In T47D, H1.4pT146 nucleolar signal appear as isolated domains that resemble those reported for UBF (Upstream Binding Transcription Factor). UBF is a characteristic protein component of the active rDNA chromatin. H1.4-pT146 pattern after RNApol I inhibition through ActD treatment also resembled the staining reported for UBF. RNApol I inhibition triggers a large-scale structural reorganization of the nucleoli . Active rDNA repeats and their associated transcription machinery translocate to the nucleolar periphery segregating into the so-called ‘nucleolar caps’. For its part, H1.4-pT146 seems to translocate in these structures upon ActD treatment, mimicking the characteristic re-distribution pattern of the rRNA transcription machinery. As shown in adapted panels from Zheng et al., a similar distribution profile has been found for other phosphorylated H1.4 (H1.4-pS187). Upper panels are adapted from (154,192).

Notably, it is worth mentioning that not all phosphorylated H1 proteins may be present at interphasic nucleoli. In fact, it has been reported that some site-specific phosphorylations are restricted to mitosis (154). Besides, phosphorylation of T17 (present both in H1.3 and H1.4) did not show a nucleolar enrichment in our hands (**Figure D.11**). This last result could also suggest that post-translational modifications of N-terminal residues *versus* C-terminal ones may result in a differential functional outcome.

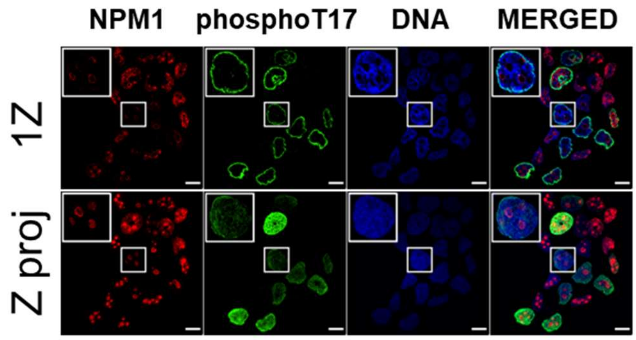


Figure D. 11. Phosphorylated-T17 H1 variants are excluded from interphasic nucleoli. Immunofluorescence of NPM1, phosphoT17-H1 (present in both H1.3 and H1.4) and DNA staining is shown. A single cell is zoomed-in. A single confocal Z plane (1Z) or the maximum projection of consecutive Z planes (Z proj) are shown. Note that this phosphorylation is also enriched at mitosis (see prophase cell in the field). Scale bar:10 µm.

Altogether, these results point out how the examination of phosphorylated, but also other post-translationally modified H1 proteins, represent an additional level of functional diversity of histone H1 variants. In case of core histone proteins, specific PTMs are widely studied and well-known associated to particular chromatin features and functions (reviewed in (17)). In comparison, our knowledge of histone H1 variants site-specific PTMs is still limited (reviewed in(165)).

1.3.5. Limitations related to antibody recognition

The imaging study of H1 PTMs also highlights a secondary lesson to consider in H1 studies. When monitoring H1.2 distribution pattern through mitosis, we found an apparent missing of H1.2 during prophase-metaphase **(Figure R.37D)**. However, immunofluorescence of H1.2-pT165 reveal a peak of enrichment of this modification in the same exact time-lapse **(Figure R.38B)**. That is, the antibody against H1.2 is unable to recognize the phosphorylated state of H1.2. Accordingly, nucleolar interphasic H1.2-pT165 cannot be picked up when evaluating total H1.2. An analogous scenario was found for H1.4 due to H1.4 antibody failed to detect mitotic or nucleolar H1.4 signals observed in H1.4-pT146 immunofluorescence **(Figure R.38C)**. Apart from the putative inability of the antibody to recognize a specific PTM, the observations could have different explanations in mitosis and interphase. In mitosis, H1 variants are highly phosphorylated at multiple residues, so maybe the protein conformation changes a lot and/or the accessible residues for the total H1.2 antibody are being occluded by phosphorylation.

In addition, we do not know what percentage of the total H1 variant carry a concrete modification. For example, if the ratio H1.2-pT165/total H1.2 in interphase is very low, we would not detect the nucleolar signal when evaluating total H1.2 due to a stronger immunostaining outside the nucleoli. To address the question, proteomic studies are needed.

On the other hand, mitotic signal of antibodies against H1.3, H1.5, H1.0 and H1X was detected **(Figure R.37A-C)**. These variants are also modified in mitosis so it seems that the antibody could still recognize at least some of the post-translationally modified H1s. For example, H1.5 is reported to be hyperphosphorylated in mitosis (153). However, we cannot discard that concrete PTMs could not be picked up with the antibodies used.

In conclusion, we have seen that the antibodies used, although being H1 variant-specific, could not be recapitulating the complete panorama of all modified forms of the given H1 variant. Thus, caution is needed to interpret results coming from different sources, including the use of different antibodies.

1.3.6. H1 loss & chromatin decompaction: A balance between H1 variant specificity and total H1 reduction

We have explored how H1 variants depletion affect chromatin structure through super-resolution imaging of DNA **(Figure R.49)**. Moreover, to gain quantitative insight into the chromatin compaction changes upon different H1 KD conditions, we measured the percentage of DNA-free-areas. Notably, super-resolution imaging of DNA has been successfully used to evaluate chromatin structure under different decompacting conditions such as TSA treatment (6,7) or Suv39H1 depletion (227). On its part, DNA-free-areas quantification has been reported elsewhere to assess global chromatin compaction changes (174,175).

In T47D cells, combined depletion of H1.2 and H1.4 (i.e. multiH1 KD) caused a global chromatin decompaction, evidenced by the drastic reduction of the %DNA-free areas **(Figure R.49B,C)**. This is in agreement with previously generated ATAC-Seq experiments in these cells, which pointed to a genome-wide gain of chromatin accessibility. Accordingly, Hi-C data analysis in multiH1 KD cells also showed more de-compacted TAD structures ((142) and Appendix I-Publication list)

In accordance, chromatin structural changes upon simultaneous depletion of multiple H1 variants has also been evidenced by others using H1 triple KO (TKO) mESCs. These cells lack H1c, H1d and H1e (H1.2, H1.3 and H1.5 orthologs) and, consequently, present a 50% of total H1 content (127). This H1 loss translated into several chromatin structure changes, including decreased global nucleosome spacing and reduced local chromatin compaction. Later genome-wide studies confirmed that H1 TKO mESCs showed specific changes in the structural segmentation of chromosomes and an increase of inter-domain interactions (132). More recently, two reports also showed that simultaneous depletion of several H1 variants lead to a global chromatin decompaction in mice (228,229). Concretely, Hi-C and ATAC-Seq experiments were performed in CD8+ T-lymphocytes of conditional H1 TKO mice (228) and germinal centre B cells of H1c$^{-/-}$H1e$^{-/-}$ mice (229).

Although the above-discussed publications study a multiple H1 variants deficiency scenario, the differential contribution of individual H1 variants to chromatin structure has not been explored before.

Interestingly, analysis of T47D single KDs revealed that single H1.2 depletion also led to chromatin decompaction, but not as pronounced as multiH1 KD **(Figure R.49)**. On the contrary, single depletion of H1.4 or H1X did not cause a significant alteration of chromatin structure. Of note, by super-resolution imaging of DNA and % DNA-free-areas quantification we are able to measure global changes in chromatin structure. However, we cannot rule out the possibility of concrete genomic regions being opened or 3D-modified upon H1.4 or H1X depletion, which could be potentially observed by other approaches, such as ATAC-Seq or Hi-C experiments.

The fact that both multiH1 KD and H1.2 KD but not H1.4 or H1X KD caused global chromatin decompaction raises some considerations. Firstly, the structural defects cannot be explained just for the total H1 reduction. Indeed, both H1.2 and H1.4 proteins contribution to total H1 content is the same in T47D, estimated to be 23-24% in each case (150). Besides, both single KDs show a high and similar efficiency of H1 depletion **(Figure R.1, R.2A)**. Thus, H1 variant specific functionality, related to their differential genomic distribution seems to play a role. These results could support the putative structural function of H1.2 (and maybe also for the rest of 'low-GC' variants), as already discussed. Secondly, in multiH1 KD cells, total H1 content is reduced ≈30% and chromatin decompaction is more drastic compared to single H1.2 depletion **(Figure R.49)**. Due to the fact that H1.2 and H1.4 occupy different genomic regions **(Figures R.6-R.9, R.15)**, the more drastic effects on decompaction in multiH1 KD cells seem to be due, at least in part, to the additive depletion of two H1 variants with non-redundant functions. On this point, analogous super-resolution experiments in other single but also other multiple H1 KDs combinations would be interesting.

In T47D, previous work of the lab demonstrated that other combinations of double or triple KDs did not reproduce the satellites induction nor the interferon response seen upon H1.2+H1.4 depletion (126). Nevertheless, it is plausible that these conditions may lead to chromatin structural changes, as the total H1 content was similarly compromised compared to multiH1 KD (i.e.H1.2+H1.4 KD). Several key points may be important to consider when interpreting the different single or combined H1 KD alterations in chromatin compaction. These include: 1) Genomic distribution of the H1 variant(s) being depleted, 2) Total H1 reduction and 3) KD efficiency.

In the whole, super-resolution microscopy of DNA enables us to decipher global compaction changes upon several H1 KDs conditions and revealed that H1 variants have specific roles in shaping genome architecture. Moreover, both the total H1 reduction but also the H1 variant repertoire have an impact on the global chromatin compaction homeostasis.

1.4. Chromatin alterations in multiH1 KD cells occur without drastic H1 variant redistribution changes

We have also addressed whether alteration of the total H1 content and relative abundance of the different variants affect the genomic localization of remaining histones. We performed H1 variants ChIP-Seq in T47D multiH1 KD cells, which showed an already-commented altered chromatin landscape (see Introduction section 2.2.3 and discussion section 1.3.6).

Our results indicate that H1.0, whose protein levels increase trying to compensate for total H1 reduction, maintained their relative enrichment at low-GC regions. On the contrary, a redistribution towards higher GC regions was found for H1X. It is important to mention that H1X protein levels also increase upon multiH1 KD, although to a lesser extent compared to H1.0. For that, ChIP-Seq distribution changes in H1X could be due to both redistribution of H1X and incorporation to newly H1X into chromatin. H1.5 and remaining H1.2 did not show changes in their relative genomic distribution. Lastly, remaining H1.4preferentially remained at low-GC regions **(Figure R.53)**. Thus, while H1.2 is being homogeneously depleted from the whole genome, H1.4 is more selectively depleted from high-GC regions, what can support its role as a heterochromatic protein. An alternative explanation could be that the H1.4 antibody, upon depletion of its specific epitope, gained cross-reactions with other variants located at low-GC regions. To discard this possibility, we performed H1.4 immunoprecipitation in T47D H1.4 KD histones extract, and no detection of any additional H1, apart from remaining H1.4, was found **(Figure R.1C)**.

Profiling within chromatin states showed that upon multiH1 KD, H1.4 clustered with low-GC variants but data also showed particular distribution changes for H1X **(Figure R.54, R.55)**. H1X abundance relatively decreased at heterochromatin and specially increased in promoters and polycomb-repressed chromatin. This last result could suggest that H1X relocates there in attempt to maintain silencing of those polycomb regions.

Importantly, specific distribution changes picked up by ChIP-Seq do not translate into general nuclear distribution changes, as observed by immunofluorescence **(Figure R.56)**.

It is important to mention that, although H3K9me3 levels do not change upon multiH1 KD (126), H3K9me3 profiling revealed its relative depletion from high-GC and A compartment regions **(Figure R.57)**. In concordance with H1X, H3K9me3 abundance at polycomb was also increased upon multiH1 KD **(Figure R.54)**. In parallel with H3K9me3 relative depletion at A compartment, ATAC-Seq signal increase, although evident genome-wide, was higher at A compartment compared to B (**Figure R.52**, (142) and Appendix I, Publication List). Thus, combined depletion of H1.2 and H1.4 do not only have an impact on heterochromatin, but also at other genomic regions. This is probably due to the fact of depleting two H1 variants with opposite genomic distributions.

Existing reports also highlight the role of H1 proteins in maintaining chromatin structure not only at heterochromatin but also at A compartment and regions associated to H3K27me3 and PRC2 signature (228,229).

Overall, H1 variants ChIP-Seq data in multiH1 KD cells showed that upon multiple H1 depletion, remaining H1 variants cannot completely change their genomic distribution to compensate H1 loss. Although data covered some concrete redistribution changes that support H1 proteins role in maintaining homeostasis of both A and B compartments, H1 variants nuclear distribution is overall robust with respect to H1 depletion. This could support H1 variants specific identity over redundancy.

To integrate the new results of multiH1 KD cells with the pre-existing evidence, **Table D.1** summarizes chromatin alterations triggered by multiH1 KD in comparison with another model for H1 depletion, H1 TKO mESCs.

Table D. 1. Chromatin alterations in two different models for multiple H1 depletion. T47D multiH1 KD cells were originally described in (126) while H1 triple KO mouse embryonic stem cells (mESCs) in (127). Differences and similarities in H1 variants content and genomic distribution, chromatin structure, transcriptional output and other epigenetic layers are indicated. Question marks (??) indicate unexplored features. Results from multiH1 KD cells are reported in (126), papers listed in Appendix I and in this thesis. Results from H1 TKO mESCs are reported in (127,132,133,230) as well as in (141) (see Appendix I). NRL = Nucleosome Repeat Length. Che-RNAs = Chromatin-enriched RNAs. (*) Although no repeats expression was found in H1 TKO mESCs, additional CRISPR-Cas9 inactivation of Hist1h1a (H1.1) and Hist1h1b (H1.5), which suppose an 80% total H1 content reduction, led to major satellites, LINE-L1 and ERVs induction (230).

MODEL FOR H1 DEPLETION	**multiH1 KD (T47D human breast cancer)**	**H1 triple KO (mouse ESCs)**
H1 complement		
H1 variants depleted	↓ H1.2, H1.4	↓ H1c,H1d, H1e (= H1.2, H1.3, H1.4)
H1 variants compensation	↑ H1.0 (x3) ↑H1X (x1.5)	↑ H1.0, H1a, H1b (x2-3) (= H1.0, H1.1, H1.5) H1X?
Total H1 content reduction	≈30%	≈50%
H1 variants genomic distribution		
WT	LowGC: H1.0, H1.2, H1.3, H1.5 HighGC: H1.4, H1X	LowGC: H1.0, H1.2, H1.3 (tagged-H1s)
Re-distribution?	H1.0, H1.5, remaining H1.2 → = Remaining H1.4 → LowGC H1X → Higher GC H1.3 ??	??
Chromatin structure		
NRL	Reduced	
Compaction	Global decompaction	
TADs/Compartments	TADs, A/B largely intact	
	↑ TAD border strength	??
Interactions	A: ↑intra.-TAD ↓ inter-TAD	↑ distal interactions
	B: ↑ intra-TAD = inter-TAD	
Transcription		
Repeats	↑ Satellites and ERVs	= (*)
Genes	Interferon response. TADs-coordinated transcriptional response	Few changes. Maintainance of ES cell identity
	Transcriptional changes no correlated with histone marks or DNA methyl. changes	
Histone marks status		
Global levels	No change	↓x2 H3K27me3, ↓ x4 H4K12Ac Rest: No change
Distribution	Relative H3K9me3 reduction in A comp.	Unaltered repressive marks Changes H3K4me1/me3, accumulated in gene-rich TADs
Other		
DNA methylation	Largely unaffected	No change global DNA methyl. Changes biased to hypomethylation, accumulated in gene-rich TADs and enhancers
Replication/DNA damage	??	Maintained Replication timing lanscape ↑ R-loops ↑ RNApol II spurious initiation
	↓ Stalled forks,↑ fork assymetry, ↑ γH2AX	
	Replication conflicts are transcription-dependent	
CheRNAs	Che-RNAs retention	

2. Discussion about H1 variants heterogeneity in different cell lines

In Results Chapter 2, we studied the H1 variants heterogeneity in different cell lines, in terms of content and nuclear distribution. We also addressed the interplay of DNA methylation and H1 variants expression. Lastly, we took a deeper look at the genomic distribution of H1X in different cancer cell lines. In this section, a discussion of these results is provided.

2.1. Heterogeneous H1 complement in human cell lines

We first studied the H1 complement at protein levels in different cell lines **(Figure R.58-R.59)**. H1.2, H1.4 but also H1X were expressed in all cell lines tested. On the other hand, H1.0, H1.3 and H1.5 are completely absent in certain cell lines. Universal expression of H1.2 and H1.4 was already reported (194–197) but these results add H1X as a universally expressed H1 variant and highlights its putative functional relevance despite not being a highly abundant H1.

These particular expression patterns support H1 variants specificity. Importantly, H1.2, H1.4 and H1X universal presence over the rest H1 variants cannot be explained just by their gene location. Replication-independent H1.0 and H1X genes are present as orphan genes in chr22 and chr3, respectively. Replication-dependent H1 variants genes are located in the histone cluster of chromosome 6. Indeed, H1.3 gene is located closer to H1.4 and H1.2 genes compared to H1.5 gene, despite concomitant absence of H1.3 and H1.5 was found repeatedly **(Figure R.58-R.59)**. Thus, each H1 gene seems to confront a variant-specific regulation, independent of the epigenetic environment of the histone locus.

Additionally, replication-independent H1 variants levels could also be regulated by common mechanisms in certain contexts. Both H1.0 and H1X levels increase to compensate H1 depletion in multiH1 KD cells **(Figure R.51)** (126), and H1.0 upregulation occur upon single H1.2 or H1.4 depletion **(Figure R.1)** (126,150). Accordingly, cell lines lacking H1.3 and H1.5, tend to have higher relative levels of replication-independent H1 variants **(Figure R.58-R.59)**. Notably, HeLa cells, which lacks H1.0, do not have increased H1X levels compared to other cell lines. These results suggest that replication-independent variants can respond to replication-dependent H1 deficiency but not to other replication-independent H1s reduction.

Regulation of mRNA and protein H1 levels seems to be controlled in a variant-specific manner. For example, H1.2 is the main H1 transcript in all cell lines evaluated, while H1.4 transcription tend to represent only 2% of total H1 mRNA levels **(Figure R.60)**. On the contrary, both H1.2 and H1.4 contribute similarly to total H1 protein levels. These differences between H1 mRNA and protein repertoire suggests that H1 variants suffer a post-transcriptional and post-translational regulation in a variant-specific manner that apparently may be shared in different cell lines. Transcriptional and translational dynamics of H1 variants are a worthy matter of study, including post-transcriptional modifications that can affect to mRNA stability.

2.2. H1 variants regulation: DNA methylation and other mechanisms

Absence of H1.0, H1.3 and H1.5 occur at both mRNA and protein levels **(Figure R.58-R.60)**. That repression at the already transcriptional level prompted us to investigate whether these variants could be regulated by DNA methylation. In fact, we found that H1.0, H1.1, H1.3 and H1.5 expression correlate with their corresponding encoding-gene DNA methylation **(Figure R.61)**, and that these genes can be methylated in cancer patients **(Figure R.62)**. On the contrary, H1.2, H1.4 and H1X genes are broadly unmethylated in cell lines and in cancer patients, in agreement with their universal expression in all the cell lines tested **(Figure R.58-R.59)**. More importantly, DNA hypomethylation triggered by aza treatment leads to a transcriptional de-repression of H1.0, H1.1 and specially H1.3 and H1.5 variants in cells where they were not expressed **(Figure R.63)**.

The observed H1 silencing pattern could be associated with the tumoral origin of the cell lines analyzed. H1.0 is highly expressed in most adult tissues but downregulation is found in multiple cancers (148). Indeed, H1.0 expression has been reported to be regulated by DNA methylation of a CpG island shore inside H1.0 gene, in agreement with our results (148). However, other non-mutually exclusive regulatory mechanisms may act on H1 variants. For example, in multiH1 KD cells, H1.0 upregulation is mediated by core histone acetylation at H1.0 promoter and without changes in DNA methylation (126). Accordingly, treatment with histone deacetylase inhibitor (HDACi) Trichostatin A (TSA), leads to H1.0 upregulation in T47D (data not shown). Another HDACi, named Quisinostat, has been reported to play an anti-proliferative potential directly driven by H1.0 induced-expression in multiple cancer types (149). In conclusion, H1 variants exhibit a particular regulation driven by diverse mechanisms, that can also be context-specific.

2.3. Consequences of a compromised H1 repertoire

The first unsolved question regarding cell lines with a compromised H1 somatic repertoire, is whether they exhibit a reduced H1 to core histone stoichiometry compared with other cell lines expressing all somatic H1 variants. We could not address that point with the experiments performed in this thesis, due to the limitations of Coomassie staining. Otherwise, proteomics-based approaches are needed.

It is tempting to speculate about the molecular outcome of having a reduced H1 somatic repertoire, especially in those cell lines simultaneously lacking H1.3 and H1.5, as this combination was found recurrent. Notably, to our knowledge this is the first time that concomitant absence of these two H1 variants is found in multiple cancer cells.

Undifferentiated cells also show a limited H1 somatic complement compared to differentiated cells (120). Based on this, we could think that restriction of H1 variety could have an impact on modulating the transcriptional tumoral programs that can ultimately affect proliferation or malignancy. In fact, it has been shown that in tumors, H1.0 silencing promotes self-renewal capacities and thus maintain the long-term proliferative tumor potential (148).

Cell lines without H1.3 and H1.5 expression could resemble the H1 variants situation occurred in multiH1 KD cells, in terms of replication-dependent H1 defect and relative increase of replication-independent H1 variants. As already stated, multiH1 KD cells are characterized by a strong interferon response, triggered by the expression of repetitive elements (Introduction section 2.2.3). Similarly, low expression of replication-dependent H1 variants and higher expression of replication-independent ones was found in pancreatic adenocarcinomas that over-expressed the interferon signature (126). In that dataset, the replication-dependent H1 variants more reduced compared to normal samples were precisely H1.3 and H1.5. An initial RT-qPCR analysis showed that cell lines lacking H1.3 and H1.5 have increased expression of certain interferon-stimulated genes and repeats, in comparison with cell lines from the same tissue origin that have a complete H1 somatic repertoire **(Figure D.12A,B)**. In case of breast cancer cell line MDA-MB-231, which lacks H1.3/H1.5, expression levels were analogous to those seen in T47D multiH1 KD **(Figure D.12B)**. Although these results may be interpreted with caution, as expression patterns of isolated genes or repeats may vary between cell lines, they undoubtedly lay the foundation for further studying the association between the lack of H1 and the immune signature.

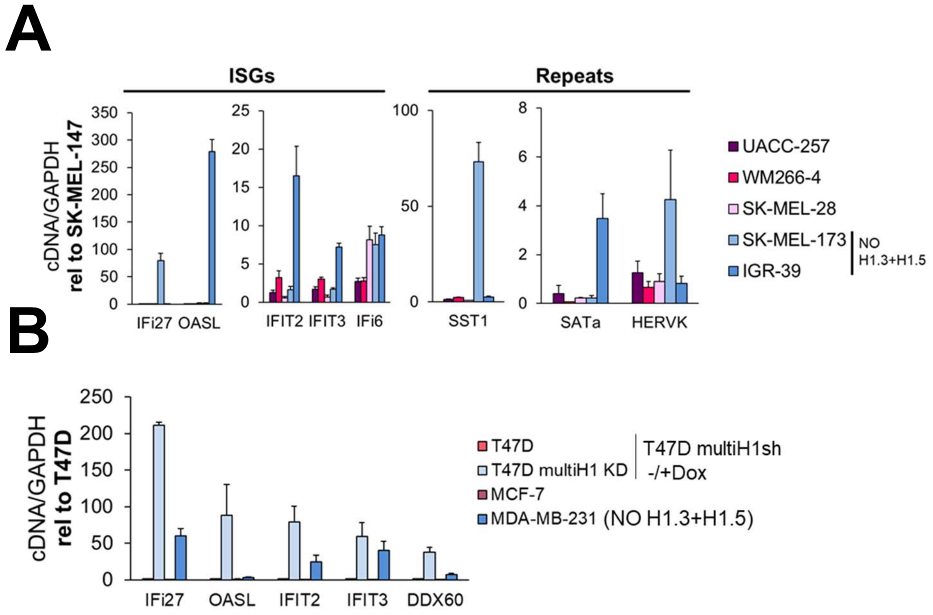


Figure D. 12. Cell lines lacking H1.3 and H1.5 show high basal expression of repetitive elements in comparison with cell lines with a more variable H1 somatic repertoire. A) RT-qPCR of interferon stimulated genes (ISGs) and repetitive elements in different melanoma cell lines. Expression data is corrected by GAPDH and normalized by SK-MEL-147 cell line basal expression, which presents a complete H1 repertoire. SK-MEL-173 and IGR-39 cell lines show concomitant absence of H1.3 and H1.5 **(Figure R.59, R.60)**. **B)** RT-qPCR of ISGs in breast cancer cell lines. MDA-MB-231 cell line does not express H1.3 and H1.5 **(Figure R.5)**. Expression data is corrected by GAPDH and normalized by basal expression in T47D multiH1sh Untreated cells. T47D multiH1sh+Dox cells, which are reported to trigger a high expression of multiple ISGs (126) is also shown.

Chromatin-mediated triggering of an immune response via repeats-derived-RNAs expression has been described upon alteration of other epigenetic regulators, apart from histone H1. These include DNA methylation (63,64), histone-modifying enzymes (231–235), chromatin remodelers (236,237) or transcriptional repressors (238,239), among others (240–243). Noteworthy, several of these reports also provide strong *in vivo* evidence that the final activation of immune signature stimulates anti-tumor immunity and synergizes with immunotherapy treatments (63,64,231–234,237,238,242,243). Although the induction of immune response might be achieved differently (but also in a potential interconnected manner), these mechanisms converge to a final beneficial outcome. Overall, the consequences of a limited H1 variants repertoire in cancer cells needs further investigation. In the same way, future studies should address the potential anti-tumor role of the interferon response mediated through H1 deficiency.

2.4. H1 variants distribution in different cell lines

2.4.1. A general view of H1 variants distribution studies

We systematically analyzed nuclear distribution of six endogenous variants in different cancer cell lines **(Figure R.64-R.65)**. Immunofluorescence experiments revealed that H1.2, H1.3 and H1.5 are universally enriched towards nuclear periphery. H1.0 and H1.4 are distributed throughout the whole nucleus but they show a more peripheral distribution in a subset of cell lines lacking H1.3 and H1.5 **(Figure R.65-R.67)**. H1X is also distributed throughout the whole nucleus in all cell lines tested with a variable relative nucleolar enrichment **(Figure R.69)**. Thus, data suggest universal nuclear patterns for certain H1 variants. However, distribution of H1 variants is accepted to be completely cell-type specific (119) and no universal distribution features have been described for any H1 variant up to date.

The presumption that H1 variants are specifically distributed among different cell lines comes from combining various pieces of evidence from different publications (148,152,168–170). These include studies performed in both human and mouse cells, as detailed in Introduction section 2.3.5. They mostly address the analysis of a single H1 variant in a particular model or cell line. The only two studies that have addressed a systematic analysis of different variants have been performed in a single cell model and through the mapping of exogenous variants. These include the analysis of H1.1-H1.5 using DamID in IMR-90 cells (169), as well as an earlier study from the lab, where ChIP-Seq of endogenous H1.2 and H1X and the exogenous H1.0-HA and H1.4-HA was performed in T47D cells (152). However, limitations of exogenous H1 variants mapping have already been discussed (section 1.1). Moreover, a comparative study of nuclear or genomic distribution of a single H1 variant in different cell models has not been performed so far. In the whole, the direct comparison of these studies is biased by the different origin of the data and the varied methodologies used, which in many cases involve the over-expression of H1 variants. In contrast, we performed the first systematic analysis of six endogenous H1 variants in multiple human cell lines, thus surpassing the commented limitations from previous studies.

Our data unveils, for the first time, universal nuclear patterns exhibited by specific H1 variants. Our findings highlight a potentially more uniform distribution of H1 variants among cell lines than previously anticipated, particularly for certain variants. These results emphasize the significance of accurately mapping endogenous H1 variants and employing standardized experimental and analytical workflows to prevent misinterpretation of the data. Collectively, these findings provide compelling evidence that initiates the exploration of the long-standing enigma regarding the balance between universality and specificity in the distribution of H1 variants within human cells.

2.4.2. H1 variants distribution: Universality *versus* cell-type specificity

2.4.2.1. H1.2, H1.3 and H1.5: universal enrichment at the nuclear periphery

Here, we report for the first time a universal H1.2, H1.3 and H1.5 enrichment towards the nuclear periphery in human cells **(Figures R.32, R.64-R.65, R.68)**. The observed nuclear patterns suggest an enrichment in large low-GC domains and B compartment, similarly to those observed in T47D by ChIP-Seq. Nevertheless, we cannot rule out these variants to show cell-type specific distribution profiles at precise genomic loci, including both LADs and non-LADs chromatin. To address that, ChIP-Seq experiments should be performed. In case of LADs, it is well known that some LADs interact with the nuclear membrane in a cell-type specific manner (i.e., facultative LADs, fLADs) (104). Moreover, other non-LADs peripherical chromatin domains show features of inactive, tissue-specific enhancers. These domains are also marked by H3K9me2 but unlike LADs, they have limited lamina interaction (244). Thus, while H1.2/H1.3/H1.5 may be universally enriched within constitutive LADs, other peripheral domains show cell-type variability and could also exhibit a variable epigenetic landscape, including H1 variants enrichment. Mechanistically, it is unclear whether chromatin positioning influences H1 variants landscape or whether the marking by particular H1 variants determines radial positioning of chromatin. Another unsolved question is whether H1 variants peripheral targeting is mediated by direct interactions with specific chromatin factors.

Universal H1.2, H1.3 and H1.5 enrichment at nuclear periphery directly point to these H1 variants as conserved components of LADs, as has been described for H3K9me2 (106). Besides, mouse and human constitutive LADs are largely conserved (104). Overall, whether H1.2/H1.3/H1.5 enrichment within LADs is conserved through evolution and which mechanisms drive this association are interesting matters of study.

2.4.2.2. H1.0 and H1.4: A more dynamic distribution?

H1.0 and H1.4 are distributed throughout the nucleus with no relevant enrichment within immunofluorescence-recognizable chromatin structures such as the nuclear periphery or the nucleolus. Interestingly, both H1.0 and H1.4 exhibit a more peripheral distribution in cell lines lacking H1.3 and H1.5 **(Figure R.64-R.67)**. As H1.3 and H1.5 are enriched within nuclear periphery when expressed, it seems that H1.0 and H1.4 try to balance H1 content at LADs when H1.3 and H1.5 are absent. This suggests that H1 levels are important to maintain peripheral chromatin. Of note, this apparent re-distribution behavior contrast to the one observed in T47D multiH1 KD. Upon multi-H1 KD, H1 variants distribution is overall robust **(Figures R.52-R.56)**, with no apparent changes at the nuclear level **(Figure R.56)**. These represent two different H1-compromised scenarios. In the first one, H1.3 and H1.5 silencing is intrinsically linked to the cell line identity. On the contrary, multiH1 KD cells represent an inducible H1.2 and H1.4 depletion, abnormal in T47D cells. Our results suggest that compensatory mechanisms between different H1 variants, in terms of distribution, may be limited when perturbing H1 levels but achieved when H1 repertoire is 'naturally' compromised. This particular H1 context could also represent an acquired adaptative mechanism of certain cancer cells. In addition, it is important to consider that H1.2 and H1.4 depletion in multiH1 KD cells was partial, this could also explain why other variants are not being relocated. Another possibility is that the observed differences can be attributed to the absence of different H1 variants in each case (i.e. H1.3 and H1.5 versus H1.2 and H1.4).

Regarding H1.0, it was previously associated to nucleoli. Proteomic studies in four human cell lines demonstrated that almost 1/3 of the candidate H1.0-binding proteins localized to nucleolus and were related to nucleolar functionality (245). Remarkably, the experiments were performed by pulled-down of exogenous, chimeric HaloTag-H1.0 protein. Importantly, direct H1.0 nucleolar localization or rRNA metabolism alterations upon H1.0 depletion were not reported. Our results show that H1.0 is depleted from nucleoli in all cell lines analyzed **(Figures R.32, R.64-R.65)**. We also checked that H1.0 does not redistribute to nucleoli upon H1X depletion in T47D cells (data not shown). Nevertheless, we cannot discard that H1.0 interacts with nucleolus-related proteins, as it can be enriched at perinucleolar heterochromatin or NADs, as observed in T47D cells (see perinucleolar regions in **Figures R.32, R.34, R.35-R.36**).

2.4.2.3. H1X: Nucleolar presence, correlation with GC content and SVAs enrichment

Immunofluorescence analysis revealed that H1X is distributed in a punctuated pattern throughout the whole nucleus with a variable relative nucleolar enrichment between cell lines **(Figure R.64-R.65, R.69)**. Despite the differences, in all cell lines tested nucleolar H1X is detected **(Figure R.69)**, while the rest of H1 variants are strongly depleted from nucleoli. Significantly, one of the cell lines showing higher H1X nucleolar enrichment is non-tumoral cell line IMR-90 **(Figure R.68-R.69)**, thus confirming that H1X accumulation in the nucleolus is not restricted to a tumoral phenotype.

While immunofluorescence analysis is very useful to evaluate differential association with large chromatin domains, we aimed to explore the similarities and differences of H1 variants distribution in different cell lines at a genomic level. For that purpose, we focus on histone H1X and performed H1X ChIP-Seq in five additional cancer cell lines.

First, we used genomic segmentation based on G-bands to directly compare the H1X profiles in the different cell lines, thus reinforcing the utility of G-bands as epigenetic units to compare H1 variants or a given H1 variant in different cell models (see Introduction sections 1.5.3 and 2.3.5). Analysis revealed that H1X is universally enriched at high-GC genomic regions **(Figure R.70)**.

Further analysis of ChIP-Seq abundance within repetitive elements revealed both common and variable features of H1X. The repeat class showing the highest H1X abundance in all cell lines is Other Class. While H1X is also enriched at SINE and frequently at Satellite classes, its relative enrichment in those classes compared to SVAs, is variable **(Figure R.71)**. As observed in T47D, H1X abundance gradually increases towards younger SVA and Alu elements in all cell lines **(Figure R.72, R.75)**. However, H1X profiles at LINE and LTR elements is not universal and for some cell lines even show an opposite tendency to the one observed in T47D, being more depleted at young LINE-L1 or LTR elements **(Figure R.75C, R.76)**. Peculiarly, LTR12C/E elements somehow break the divergences observed in LTR family and shows a universal enrichment of H1X **(Figure R.74B)**.

The recurrent H1X enrichment within SVA retrotransposons but also younger Alu point H1X as a putative master regulator of their expression. It also suggests that regulation of these families or their epigenetic status may be more constitutive between cell lines compared to other TE classes.

TE expression is dysregulated in cancer cells compared to normal tissues in a cancer-type specific manner. LTR TEs exhibit the highest number of overexpressed families across multiple cancer types, followed by DNA and LINE. Of these classes, the most recurrently overexpressed TEs are among the evolutionary youngest in the human genome (246). As we are analyzing cancer cell lines, differences in H1X distribution at LINE and LTR could be reflecting a cell-line specific transcriptional status of these elements. For example, it has been reported that LINE-L1 and more concretely L1HS, which is the youngest elements from LINE-L1 family, are highly expressed in MCF-7 cell line (56,247). We have shown that H1X is depleted from younger LINE-L1 repeats in MCF-7 **(Figure R.75C)**. Whether cell-type specific H1X variable binding to TEs is cause or consequence of their putative differential transcriptional status between cell lines, is still unknown.

As discussed in Section 1.2.3, the universal enrichment of H1X in SVA and Alu families opens door to study their association from an evolutionary perspective in different species. Moreover, it is also important to highlight that we analyzed six different cancer cell lines, so both non-tumoral differentiated or embryonic stem cells represent additional models for evaluating H1X distribution. Altogether, comparative analysis of H1X distribution in all these different epigenetic backgrounds will provide meaningful insights of the functional implication of H1X in TE silencing.

3. Global discussion

All things considered, this doctoral thesis provides the first systematic study involving six endogenous H1 variants within a mammalian cell type, accompanied by a comprehensive comparison between different cell lines. ChIP-Seq and imaging experiments included here demonstrate H1 variants heterogeneity and their non-random genomic distribution (**Table D.2**). Consequently, the present work brings to light the limitations of considering H1 variants as a total H1 pool of mere chromatin compactors. H1 variants represent an insufficiently explored additional determinant of the intricate chromatin landscape, alongside widely-studied core histones variants or core histones PTMs. Moreover, future studies that delve deeper into the roles of H1 variants as both chromatin organizers and regulators (**Figure D.13**) emerge as necessary to fully understand the molecular fundamentals underlying nuclear programs and their implications in disease.

Table D. 2. Summary of the genomic distribution of six endogenous H1 variants in T47D. Table summarizes the results from ChIP-Seq and immunofluorescence experiments from six endogenous H1 variants in T47D breast cancer cells. Regarding repetitive elements, it is indicated whether H1s are generally enriched or depleted in the specified repeat class but also how H1 abundance varies within a class or family in relation to the evolutionary age of those repeats. Notably, blue asterisks (*), highlight the features that have been universally found when analyzing multiple cancer cell lines (see Results Chapter 2), which include: 1) H1.2, H1.3, H1.5 enrichment at LADs; 2) H1X enrichment at SVAs and Alu, with higher abundance at SVAs; 3) H1X gradual enrichment towards younger SVA and Alu elements; 4) H1X enrichment at LTR12C/E.

T47D H1 variants		**H1.0**	**H1.2**	**H1.5**	**H1.3**	**H1.4**	**H1X**
G-bands		Low-GC				High-GC	
Compartment		B				A	
Chromatin states	**Heterochromatin**	Enriched					
	Low activity	Enriched					
	Promoters	Depleted, especially from active TSS					Enriched compared to the rest H1s
	Enhancers	Depleted					
Repetitive elements	**LINE**	Enriched				Depleted	
		Gradual decrease from older to younger LINEs				Enriched in a small subset of young LINEs from L1 family	
	SINE	Depleted, especially from Alu family				Enriched (Alu family) (H1X*). Comparative abundance: H1.4 > H1X	
		Gradual decrease from older to younger Alu				Increase abundance towards younger Alu (AluY) (H1X*)	
	Other (SVAs)	Depleted				Enriched (H1X*). Comparative abundance: H1X > H1.4	
		Gradual decrease from SVA_A to younger SVA_F family			SVA_A-SVA_F similar abundance	SVA_A-SVA_F similar abundance	Gradual increase from SVA_A to younger SVA_F family (*)
	LTR	Enriched				Depleted. Enriched in a small subset of young LTRs	
		Gradual decrease from older to younger LINEs				More associated to youngest ERVL-MaLR	More associated to youngest ERV1 (LTR12C/E) (*)
	DNA	Enriched				Depleted	
	Satellite	Enriched, specially "low-GC" H1 variants					
Nuclear domains	**LADs**	Enriched (H1.2, H1.5, H1.3*)				Depleted	
	NADs	Enriched				Depleted	
	Nucleolus	Depleted (phospho H1.2/H1.4 enriched)					Enriched

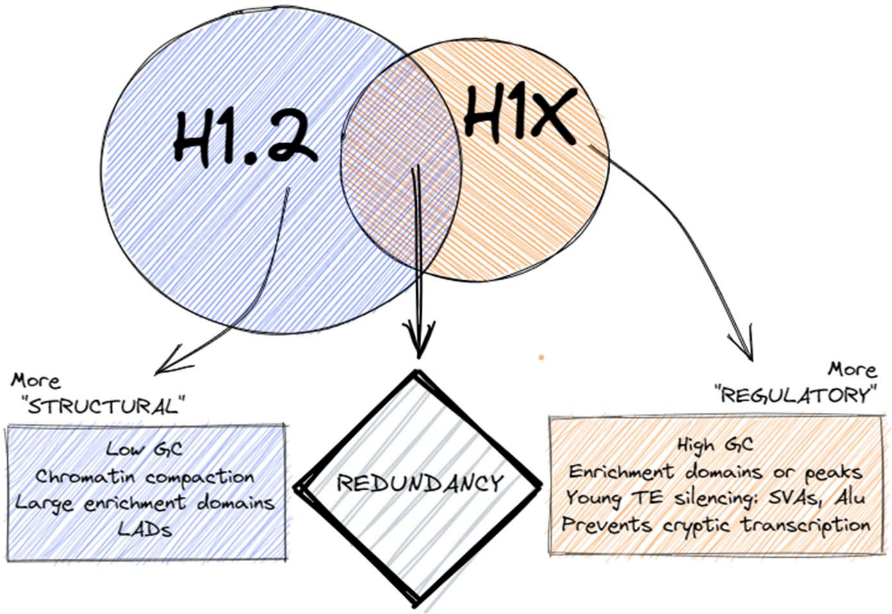


Figure D. 13. **Dual role of histone H1 variants as chromatin structural components and regulators.** Although H1 variants are partially redundant (discussion section 1.1.1), variant-specific distribution and functionality also exist. H1.2 and H1X represents two differentiated H1 groups. Importantly, both H1 variants are universally expressed. H1.2 forms large enrichment domains at low-GC chromatin and LADs. LADs represent a structural chromatin unit that segregates heterochromatin to the nuclear periphery. Due to the universal localization of H1.2 **(Figures R.32, R.64-R.65)**, the chromatin decompaction observed by H1.2 depletion **(Figure R.49)**, and the shared enrichment domains with other 'low-GC' H1 variants at the population level (discussion section 1.3.1, **Figure D.8**), we propose that H1.2 may have a more structural role in chromatin. On the other hand, H1X, which is universally enriched at high-GC regions **(Figure R.6)**, forms smaller enrichment domains but also precise peaks of enrichment **(Figures R.16, D.2)**. These narrow peaks overlap with young TEs **(Figure R.17A, R.19B, R.20B,C)** and H1X participates in their silencing **(Figure R.30, R.31)**. Importantly, enrichment within recently-evolved SVAs and Alu elements in observed in multiple cell lines **(Figure R.71-R.74, R.75A,B)**. Moreover, H1X depletion does not trigger global chromatin decompaction **(Figure R.49)** but produces aberrant cryptic transcription **(Figure R.28-R.29)**. For all these observations, we suggest that H1X may play a more regulatory role in chromatin.

Conclusions

Chapter 1. Differential distribution of six endogenous histone H1 variants in T47D breast cancer cells

1. Selected antibodies against endogenous human H1.0, H1.2, H1.3, H1.4, H1.5 and H1X are variant-specific and show a good performance on ChIP experiments.

2. H1 variants are differentially genome-wide distributed in two large groups: H1.0, H1.2, H1.3 and H1.5 are enriched towards low-GC, B compartment regions, whereas H1.4 and H1X are more abundant at high-GC, A compartment regions.

3. H1 variants also show common distribution features, including enrichment at heterochromatin or low-activity states, as well as depletion from enhancer regions. Besides, H1 variants are in general depleted from promoters depending on their transcriptional status. However, H1X is enriched within inactive promoters, in comparison with the rest of H1 variants.

4. H1 variants are differentially enriched within particular repetitive elements classes: H1.2, H1.3, H1.5, H1.0 are enriched at Satellite, LINE, LTR and DNA repeats, among others, while H1.4 and H1X are enriched at SINE and Other (i.e. SVAs) classes. Concretely, H1X is more associated with SVA retrotransposons while H1.4 is preferentially linked to Alu elements.

5. Enrichment peaks can be computed for H1X and H1.4 but not for the other variants. Notably, almost all these narrow peaks overlap with repetitive elements.

6. Regarding SVAs, H1X abundance gradually increases from SVA_A to SVA_F, while H1.4 shows a similar abundance across all the six families.

7. Regarding SINEs, H1X and H1.4 are highly enriched within Alu family, while H1.2 is more abundant at Deu, SINE and MIR families compared to Alu. The highest H1X and H1.4 enrichment is found at youngest AluY elements.

8. Regarding LINEs, H1X and H1.4 are enriched in L1 family compared to the rest families. On the contrary, H1.2 is abundant in all LINE families. H1X and H1.4 mark a small subset of evolutionary recent LINE-L1 elements.

9. H1X and H1.4 are enriched within transposable elements recently incorporated in the genome along primates evolution, including youngest elements from SVAs, SINE-Alu, LINE-L1 and LTR classes/families. The highest H1X and H1.4 abundance occur within *Hominoidea*-restricted repeats. On the contrary, H1.2/H1.3/H1.5/H1.0 abundance gradually increases with the repeat evolutionary age.

10. H1X and H1.4 depletion cause a moderate upregulation of young transposable elements in which these variants are enriched.

11. H1X depletion causes spurious/cryptic intragenic transcription at particular genes found upregulated.

12. H1 variants occupy differential nuclear domains as evidenced by confocal and super-resolution microscopy. H1.2, H1.3 and H1.5 are enriched towards the nuclear periphery or lamina-associated domains. H1.4 and H1X show a punctuated pattern throughout the whole nucleus, with H1X being enriched at nucleoli. H1.0 forms enrichment territories that tend to be peripheral.

13. H1.0 preferential co-localization with H1.2, H1.3 and H1.5 observed by confocal immunofluorescence, is lost in super-resolution microscopy experiments. This suggests that three-dimensional structural domains are consistently marked by a given H1 variant in single-cells.

14. H1 variants show different distribution patterns along mitosis, showing a variant-specific attachment to mitotic chromosomes.

15. H1.2-pT165 and H1.4-pT146 levels increase during early stages of mitosis, showing particular dynamics and binding distribution patterns. During interphase, both post-translationally modified H1s show nucleolar enrichment.

16. Combined H1.2 and H1.4 depletion (i.e., multiH1 KD cells) but also single H1.2 depletion lead to global chromatin decompaction, while H1.4 or H1X depletion does not.

17. MultiH1 knock-down occurs without drastic re-distribution of remaining H1 variants or H3K9me3. However, ChIP-Seq data revealed that H1X re-distributes towards higher GC regions while remaining H1.4 preferentially reside at low-GC regions. H3K9me3 was more selectively depleted from high-GC regions.

Chapter 2. Heterogeneity of H1 variants content, regulation and distribution across cell lines

1. H1 variants somatic repertoire is variable across different cell lines. H1.2, H1.4 and H1X are universally expressed. Concomitant absence of H1.3 and H1.5 is recurrently found in multiple cell lines.

2. Expression of H1.0, H1.1, H1.3 and H1.5 can be regulated by DNA methylation.

3. H1 variants nuclear distribution show both universal and variable patterns among cell lines. H1.2, H1.3 and H1.5 are enriched towards nuclear periphery in all cell lines analyzed. H1.0 and H1.4 are distributed throughout the whole nucleus but in cell lines lacking H1.3 and H1.5, they show a more peripheral distribution. H1X show a cell-line dependent nucleolar enrichment but its presence within nucleoli is detected in all cell lines analyzed.

4. H1X is universally enriched at high-GC regions as well as within SVA and Alu repeats.

5. H1X abundance gradually increases from older to younger SVA families, as well as towards more recent Alu elements in all cell lines examined.

6. H1X abundance within LINE and LTR classes shows variable patterns and correlation with evolutionary age in different cell lines. Interestingly, H1X is universally enriched at LTR12C/E elements.

References

1. Olins DE, Olins AL. Chromatin history: Our view from the bridge. Nat Rev Mol Cell Biol. 2003;4(10):809–14.
2. Fyodorov D V, Zhou BR, Skoultchi AI, Bai Y. Emerging roles of linker histones in regulating chromatin structure and function. Nat Rev Mol Cell Biol. 2018;19:192–206.
3. Joti Y, Hikima T, Nishino Y, Kamada F, Hihara S, Takata H, et al. Chromosomes without a 30-nm chromatin fiber. Nucleus. 2012;3(5):404–10.
4. Nishino Y, Eltsov M, Joti Y, Ito K, Takata H, Takahashi Y, et al. Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. EMBO Journal. 2012;31(7):1644–53.
5. Fussner E, Strauss M, Djuric U, Li R, Ahmed K, Hart M, et al. Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. EMBO Rep. 2012;13(11):992–6.
6. Ricci MA, Manzo C, García-Parajo MF, Lakadamyali M, Cosma MP. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. Cell. 2015;160(6):1145–58.
7. Otterstrom J, Castells-Garcia A, Vicario C, Gomez-Garcia PA, Cosma MP, Lakadamyali M. Super-resolution microscopy reveals how histone tail acetylation affects DNA compaction within nucleosomes in vivo. Nucleic Acids Res. 2019;47(16):8470–84.
8. Solovei I, Thanisch K, Feodorova Y. How to rule the nucleus: divide et impera. Curr Opin Cell Biol. 2016;40:47–59.
9. Jost KL, Bertulat B, Cardoso MC. Heterochromatin and gene positioning: Inside, outside, any side? Chromosoma. 2012;121(6):555–563.
10. Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet. 2001;2(4):292–301.
11. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology. 2013;38(1):23–38.
12. Jones PA. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13(7):484–92.
13. Pappalardo XG, Barra V. Losing DNA methylation at repetitive elements and breaking bad. Epigenetics Chromatin. 2021;14(1):25.
14. Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, et al. Intragenic DNA methylation prevents spurious transcription initiation. Nature. 2017;543(7643):72–7.
15. Davalos V, Esteller M. Cancer epigenetics in clinical practice. CA Cancer J Clin. 2022;
16. Brian Strahl D, David Allis C. The language of covalent histone modifications. Nature. 2000;403(6765):41–5.
17. Millán-Zambrano G, Burton A, Bannister AJ, Schneider R. Histone post-translational modifications — cause and consequence of genome function. Nat Rev Genet. 2022;23(9):563–80.
18. Kouzarides T. Chromatin Modifications and Their Function. Cell. 2007;128(4):693–705.
19. Li B, Carey M, Workman JL. The Role of Chromatin during Transcription. Cell. 2007;128(4):707–19.
20. Henikoff S, Shilatifard A. Histone modification: Cause or cog? Trends in Genetics. 2011;27(10):389–96.

21. Hughes AL, Kelley JR, Klose RJ. Understanding the interplay between CpG island-associated gene promoters and H3K4 methylation. Biochim Biophys Acta Gene Regul Mech. 2020;1863(8):194567.

22. Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, et al. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell. 2005;123(4):581–92.

23. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. Regulation of alternative splicing by histone modifications. Science (1979). 2010;327(5968):996–1000.

24. Huang H, Weng H, Zhou K, Wu T, Zhao BS, Sun M, et al. Histone H3 trimethylation at lysine 36 guides m6A RNA modification co-transcriptionally. Nature. 2019;567(7748):414–9.

25. Zhang T, Zhang Z, Dong Q, Xiong J, Zhu B. Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. Genome Biol. 2020;21(1):45.

26. Narita T, Higashijima Y, Kilic S. Acetylation of histone H2B marks active enhancers and predicts CBP/p300 target genes. Nat Genet. 2023 Apr;55(4):679–92.

27. Martire S, Banaszynski LA. The roles of histone variants in fine-tuning chromatin organization and function. Nat Rev Mol Cell Biol. 2020;21(9):522–41.

28. Szenker E, Ray-Gallet D, Almouzni G. The double face of the histone variant H3.3. Vol. 21, Cell Research. 2011.

29. Teng YC, Sundaresan A, O'Hara R, Gant VU, Li M, Martire S, et al. ATRX promotes heterochromatin formation to protect cells from G-quadruplex DNA-mediated stress. Nat Commun. 2021;12(1).

30. Gambogi CW, Black BE. The nucleosomes that mark centromere location on chromosomes old and new. Essays Biochem. 2019;63(1):15–27.

31. Hauer MH, Gasser SM. Chromatin and nucleosome dynamics in DNA damage and repair. Genes Dev. 2017;31(22):2204–21.

32. Buschbeck M, Uribesalgo I, Wibowo I, Rué P, Martin D, Gutierrez A, et al. The histone variant macroH2A is an epigenetic regulator of key developmental genes. Nat Struct Mol Biol. 2009;16(10):1074–9.

33. Douet J, Corujo D, Malinverni R, Renauld J, Sansoni V, Marjanović MP, et al. MacroH2A histone variants maintain nuclear organization and heterochromatin architecture. J Cell Sci. 2017;130(9):1570–82.

34. Kozlowski M, Corujo D, Hothorn M, Guberovic I, Mandemaker IK, Blessing C, et al. MacroH2A histone variants limit chromatin plasticity through two distinct mechanisms. EMBO Rep. 2018;19(10):e44445.

35. Long M, Sun X, Shi W, Yanru A, Leung STC, Ding D, et al. A novel histone H4 variant H4G regulates rDNA transcription in breast cancer. Nucleic Acids Res. 2019;47(16):8399–409.

36. Pang MYH, Sun X, Ausió J, Ishibashi T. Histone H4 variant, H4G, drives ribosomal RNA transcription and breast cancer cell proliferation by loosening nucleolar chromatin structure. J Cell Physiol. 2020;235(12):9601–8.

37. Lai WKM, Pugh BF. Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nat Rev Mol Cell Biol. 2017;18(9):548–62.

38. Lai B, Gao W, Cui K, Xie W, Tang Q, Jin W, et al. Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. Nature. 2018;562(7726):281–5.

39. Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021;22(2):96–118.

40. Werner MS, Ruthenburg AJ. Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. Cell Rep. 2015;12(7):1089–98.

41. Werner MS, Sullivan MA, Shah RN, Nadadur RD, Grzybowski AT, Galat V, et al. Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription. Nat Struct Mol Biol. 2017;24(7):596–603.

42. Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. 2012;13(11):R107.

43. Kannan S, Chernikova D, Rogozin IB, Poliakov E, Managadze D, Koonin E V., et al. Transposable element insertions in long intergenic non-coding RNA genes. Front Bioeng Biotechnol. 2015;3:71.

44. Dumbovic G, Forcales S V., Perucho M. Emerging roles of macrosatellite repeats in genome organization and disease development. Epigenetics. 2017;12(7):515–26.

45. Thakur J, Packiaraj J, Henikoff S. Sequence, chromatin and evolution of satellite DNA. Int J Mol Sci. 2021;22(9):4309.

46. Altemose N, Logsdon GA, Bzikadze A V., Sidhwani P, Langley SA, Caldas G V., et al. Complete genomic and epigenetic maps of human centromeres. Science (1979). 2022;376(6588):eabl4178.

47. Agrawal S, Ganley ARD. The conservation landscape of the human ribosomal RNA gene repeats. PLoS One. 2018;13(12):e0207531.

48. Burns K. Transposable elements in cancer. Nat Rev Cancer. 2017;17(7):415–24.

49. Camacho OV, Galan C, Swist-Rosowska K, Ching R, Gamalinda M, Karabiber F, et al. Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. Elife. 2017;6:e25293.

50. Huo X, Ji L, Zhang Y, Lv P, Cao X, Wang Q, et al. The Nuclear Matrix Protein SAFB Cooperates with Major Satellite RNAs to Stabilize Heterochromatin Architecture Partially through Phase Separation. Mol Cell. 2020;77(2):368–83.

51. Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, Lingner J. Telomeric repeat-containing RNA and RNA surveillance factors at mammalian chromosome ends. Science (1979). 2007;318(5851):798–801.

52. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. Genome Biol. 2018;19(1):199.

53. Platt RN, Vandewege MW, Ray DA. Mammalian transposable elements and their impacts on genome evolution. Chromosome Research. 2018;26(1–2):25–43.

54. Levin HL, Moran J V. Dynamic interactions between transposable elements and their hosts. Nat Rev Genet. 2011;12(9):615–27.

55. Kaer K, Speek M. Retroelements in human disease. Gene. 2013;518(2):231–41.

56. Freeman B, White T, Kaul T, Stow EC, Baddoo M, Ungerleider N, et al. Analysis of epigenetic features characteristic of L1 loci expressed in human cells. Nucleic Acids Res. 2022;50(4):1888–907.

57. Deininger P. Alu elements: Know the SINEs. Genome Biol. 2011;12(12):236.

58. Giordano J, Ge Y, Gelfand Y, Abrusán G, Benson G, Warburton PE. Evolutionary history of mammalian transposons determined by genome-wide defragmentation. PLoS Comput Biol. 2007;3(7):e137.

59. Bennett EA, Keller H, Mills RE, Schmidt S, Moran J V., Weichenrieder O, et al. Active Alu retrotransposons in the human genome. Genome Res. 2008;18(12):1875–83.

60. Wang H, Xing J, Grover D, Hedges Kyudong Han DJ, Walker JA, Batzer MA. SVA elements: A hominid-specific retroposon family. J Mol Biol. 2005;354(4):994–1007.

61. Levy O, Knisbacher BA, Levanon EY, Havlin S. Integrating networks and comparative genomics reveals retroelement proliferation dynamics in hominid genomes. Sci Adv. 2017;3(10):e1701256.

62. Smit AFA. Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic Acids Res. 1993;21(8):1863–72.

63. Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, et al. Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. Cell. 2015/09/01. 2015;162(5):974–86.

64. Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, Shen SY, et al. DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. Cell. 2015/09/01. 2015;162(5):961–73.

65. Jansz N. DNA methylation dynamics at transposable elements in mammals. Essays Biochem. 2019;63(6):677–89.

66. Ohtani H, Ørskov AD, Helbo AS, Gillberg L, Liu M, Zhou W, et al. Activation of a subset of evolutionarily young transposable elements and innate immunity are linked to clinical responses to 5-azacytidine. Cancer Res. 2020;80(12):2441–50.

67. Ohtani H, Liu M, Zhou W, Liang G, Jones PA. Switching roles for DNA and histone methylation depend on evolutionary ages of human endogenous retroviruses. Genome Res. 2018;28(8):1147–57.

68. Brocks D, Schmidt CR, Daskalakis M, Jang HS, Shah NM, Li D, et al. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. Nat Genet. 2017;49(7):1052–60.

69. Varshney D, Vavrova-Anderson J, Oler AJ, Cowling VH, Cairns BR, White RJ. SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. Nat Commun. 2015;6:6569.

70. Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, et al. KAP1 controls endogenous retroviruses in embryonic stem cells. Nature. 2010;463(7278):237–40.

71. Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, et al. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. Genome Res. 2014;24(8):1260–70.

72. Tie CH, Fernandes L, Conde L, Robbez-Masson L, Sumner RP, Peacock T, et al. KAP 1 regulates endogenous retroviruses in adult human cells and contributes to innate immune control. EMBO Rep. 2018;19(10):e45000.

73. Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, et al. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature. 2014;516(7530):242–5.

74. Haring NL, van Bree EJ, Jordaan WS, Roels JRE, Sotomayor GC, Hey TM, et al. ZNF91 deletion in human embryonic stem cells leads to ectopic activation of SVA retrotransposons and up-regulation of KRAB zinc finger gene clusters. Genome Res. 2021;31(4):551–63.

75. Su M, Han D, Boyd-Kirkup J, Yu X, Han JDJ. Evolution of Alu Elements toward Enhancers. Cell Rep. 2014;7(2):376–85.

76. Jacques PÉ, Jeyakani J, Bourque G. The Majority of Primate-Specific Regulatory Sequences Are Derived from Transposable Elements. PLoS Genet. 2013;9(5):e1003504.

77. Trizzino M, Park YS, Holsbach-Beltrame M, Aracena K, Mika K, Caliskan M, et al. Transposable elements are the primary source of novelty in primate gene regulation. Genome Res. 2017;27(10):1623–33.

78. Jordan IK, Rogozin IB, Glazko G V., Koonin E V. Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends in Genetics. 2003;19(2):68–72.

79. Simonti CN, Pavličev M, Capra JA. Transposable element exaptation into regulatory regions is rare, influenced by evolutionary age, and subject to pleiotropic constraints. Mol Biol Evol. 2017;34(11):2856–69.

80. Liu M, Thomas SL, DeWitt AK, Zhou W, Madaj ZB, Ohtani H, et al. Dual inhibition of DNA and histone methyltransferases increases viral mimicry in ovarian cancer cells. Cancer Res. 2018;78(20):5754–66.

81. Liu M, Ohtani H, Zhou W, Ørskov AD, Charlet J, Zhang YW, et al. Vitamin C increases viral mimicry induced by 5-aza-2'-deoxycytidine. Proc Natl Acad Sci U S A. 2016;113(37):10238–44.

82. Rowe HM, Friedli M, Offner S, Verp S, Mesnard D, Marquis J, et al. De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. Development (Cambridge). 2013;140(3):519–29.

83. Elsasser SJ, Noh KM, Diaz N, Allis CD, Banaszynski LA. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. Nature. 2015;522(7555):240–4.

84. Ecco G, Cassano M, Kauzlaric A, Duc J, Coluccio A, Offner S, et al. Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. Dev Cell. 2016;36(6):611–23.

85. Brattås PL, Jönsson ME, Fasching L, Nelander Wahlestedt J, Shahsavani M, Falk R, et al. TRIM28 Controls a Gene Regulatory Network Based on Endogenous Retroviruses in Human Neural Progenitor Cells. Cell Rep. 2017;18(1):1–11.

86. Castro-Diaz N, Ecco G, Coluccio A, Kapopoulou A, Yazdanpanah B, Friedli M, et al. Evolutionally dynamic L1 regulation in embryonic stem cells. Genes Dev. 2014;28(13):1397–409.

87. Kapopoulou A, Mathew L, Wong A, Trono D, Jensen JD. The evolution of gene expression and binding specificity of the largest transcription factor family in primates. Evolution (N Y). 2016;70(1):167–80.

88. Thomas JH, Schneider S. Coevolution of retroelements and tandem zinc finger genes. Genome Res. 2011;21(11):1800–12.

89. Huntley S, Baggott DM, Hamilton AT, Tran-Gyamfi M, Yang S, Kim J, et al. A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the

evolutionary history of a large family of transcriptional repressors. Genome Res. 2006;16(5):669–77.

90. Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, et al. Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. Cell. 2010;143:212–24.

91. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat Biotechnol. 2010;28(8):817–25.

92. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473(7345):43–9.

93. Ernst J, Kellis M. ChromHMM: Automating chromatin-state discovery and characterization. Nat Methods. 2012;9(3):215–6.

94. Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science (1979). 2009;326(5950):289–93.

95. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485(7398):376–80.

96. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012;485(7398):381–5.

97. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, et al. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell. 2012;148(3):458–72.

98. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159(7):1665–80.

99. Le Dily FL, Baù D, Pohl A, Vicent GP, Serra F, Soronellas D, et al. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. Genes Dev. 2014;28(19):2151–62.

100. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. 2016;15(9):2038–49.

101. Kojic A, Cuadrado A, De Koninck M, Gimenez-Llorente D, Rodriguez-Corsino M, Gomez-Lopez G, et al. Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. Nat Struct Mol Biol. 2018;25(6):496–504.

102. Sun X, Zhang J, Cao C. CTCF and Its Partners: Shaper of 3D Genome during Development. Genes (Basel). 2022;13(8):1383.

103. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature. 2008;453(7197):948–51.

104. Meuleman W, Peric-Hupkes D, Kind J, Beaudry JB, Pagie L, Kellis M, et al. Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. Genome Res. 2013;23(2):270–80.

105. Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SWM, Solovei I, Brugman W, et al. Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. Mol Cell. 2010;38(4):603–13.

106. Poleshko A, Smith CL, Nguyen SC, Sivaramakrishnan P, Wong KG, Murray JI, et al. H3K9me2 orchestrates inheritance of spatial positioning of peripheral heterochromatin through mitosis. Elife. 2019;8:e49278.

107. Kind J, Pagie L, Ortabozkoyun H, Boyle S, De Vries SS, Janssen H, et al. Single-cell dynamics of genome-nuclear lamina interactions. Cell. 2013;153(1):178–92.

108. Harr JC, Luperchio TR, Wong X, Cohen E, Wheelan SJ, Reddy KL. Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. Journal of Cell Biology. 2015;208(1):33–52.

109. Nemeth A, Conesa A, Santoyo-Lopez J, Medina I, Montaner D, Peterfia B, et al. Initial genomics of the human nucleolus. PLoS Genet. 2010/04/03. 2010;6(3):e1000889.

110. Dillinger S, Straub T, Nemeth A. Nucleolus association of chromosomal domains is largely maintained in cellular senescence despite massive nuclear reorganisation. PLoS One. 2017;12(6):e0178821.

111. Bersaglieri C, Kresoja-Rakic J, Gupta S, Bär D, Kuzyakiv R, Panatta M, et al. Genome-wide maps of nucleolus interactions reveal distinct layers of repressive chromatin domains. Nat Commun. 2022;13(1):1483.

112. Peng T, Hou Y, Meng H, Cao Y, Wang X, Jia L, et al. Mapping nucleolus-associated chromatin interactions using nucleolus Hi-C reveals pattern of heterochromatin interactions. Nat Commun. 2023;14(1).

113. Caspersson, T., Lomakka, G., Zech L. The 24 fluorescence patterns of the human metaphase chromosomes — distinguishing characters and variability. Hereditas. 1971;67(1):89–102.

114. Comings DE. Mechanisms of Chromosome Banding and Implications for Chromosome Structure. Annu Rev Genet. 1978;12:25–46.

115. Holmquist G, Gray M, Porter T, Jordan J. Characterization of Giemsa dark- and light-band DNA. Cell. 1982;31(1):121–9.

116. Serna-Pujol,N., Salinas-Pena,M., Mugianesi,F., Lopez-Anguita,N.,Torrent-Llagostera,F.,Izquierdo-Bouldstridge,A.,Marti-Renom,MA.,Jordan A. TADs enriched in histone H1.2 strongly overlap with the B compartment, inaccessible chromatin and AT-rich Giemsa bands. FEBS J. 2021;288(6):1989–2013.

117. Happel N, Doenecke D. Histone H1 and its isoforms: Contribution to chromatin structure and function. Gene. 2009;431(1–2):1–12.

118. Izzo A, Kamieniarz K, Schneider R. The histone H1 family: specific members, specific functions? Biol Chem. 2008/01/23. 2008;389(4):333–43.

119. Millán-Ariño L, Izquierdo-Bouldstridge A, Jordan A. Specificities and genomic distribution of somatic mammalian histone H1 subtypes. Biochim Biophys Acta Gene Regul Mech. 2016 Mar 1;1859(3):510–9.

120. Terme JM, Sesé B, Millán-Ariño L, Mayor R, Belmonte Izpisúa JC, Barrero MJ, et al. Histone H1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency. Journal of Biological Chemistry. 2011;286(41):35347–35357.

121. Talbert PB, Ahmad K, Almouzni G, Ausiá J, Berger F, Bhalla PL, et al. A unified phylogeny-based nomenclature for histone variants. Epigenetics Chromatin. 2012;5:7.

122. Woodcock CL, Skoultchi AI, Fan Y. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. Chromosome Res. 2006;14(1):17–25.

123. Bednar J, Garcia-Saez I, Boopathi R, Cutter AR, Papai G, Reymer A, et al. Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. Mol Cell. 2017;66(3):384–97.

124. Zhou BR, Jiang J, Feng H, Ghirlando R, Xiao TS, Bai Y. Structural Mechanisms of Nucleosome Recognition by Linker Histones. Mol Cell. 2015;59(4):628–38.

125. Dombrowski M, Engeholm M, Dienemann C, Dodonova S, Cramer P. Histone H1 binding to nucleosome arrays depends on linker DNA length and trajectory. Nat Struct Mol Biol. 2022;29(5):493–501.

126. Izquierdo-Bouldstridge A, Bustillos A, Bonet-Costa C, Aribau-Miralbés P, García-Gomis D, Dabad M, et al. Histone H1 depletion triggers an interferon response in cancer cells via activation of heterochromatic repeats. Nucleic Acids Res. 2017;45(20):11622–11642.

127. Fan Y, Nikitina T, Zhao J, Fleury TJ, Bhattacharyya R, Bouhassira EE, et al. Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. Cell. 2005;123(7):1199–212.

128. Fan Y, Nikitina T, Morin-Kensicki EM, Zhao J, Magnuson TR, Woodcock CL, et al. H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. Mol Cell Biol. 2003;23(13):4559–72.

129. Gibson BA, Doolittle LK, Schneider MWG, Jensen LE, Gamarra N, Henry L, et al. Organization of Chromatin by Intrinsic and Regulated Phase Separation. Cell. 2019;179(2):470–84.

130. Shakya A, Park S, Rana N, King JT. Liquid-Liquid Phase Separation of Histone Proteins in Cells: Role in Chromatin Organization. Biophys J. 2020;118(3):753–64.

131. Turner AL, Watson M, Wilkins OG, Cato L, Travers A, Thomas JO, et al. Highly disordered histone H1−DNA model complexes and their condensates. Proc Natl Acad Sci U S A. 2018;115(47):11964–9.

132. Geeven G, Zhu Y, Kim BJ, Bartholdy BA, Yang SM, Macfarlan TS, et al. Local compartment changes and regulatory landscape alterations in histone H1-depleted cells. Genome Biol. 2015;16:289.

133. Almeida R, Fernández-Justel JM, Santa-María C, Cadoret JC, Cano-Aroca L, Lombraña R, et al. Chromatin conformation regulates the coordination between DNA replication and transcription. Nat Commun. 2018;9(1):1590.

134. Glaich O, Leader Y, Lev Maor G, Ast G. Histone H1.5 binds over splice sites in chromatin and regulates alternative splicing. Nucleic Acids Res. 2019;47(12):6145–6159.

135. Thorslund T, Ripplinger A, Hoffmann S, Wild T, Uckelmann M, Villumsen B, et al. Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage. Nature. 2015;527(7578):389–93.

136. Konishi A, Shimizu S, Hirota J, Takao T, Fan Y, Matsuoka Y, et al. Involvement of histone H1.2 in apoptosis induced by DNA double-strand breaks. Cell. 2003;114(6):673–88.

137. Vujatovic O, Zaragoza K, Vaquero A, Reina O, Bernues J, Azorin F. Drosophila melanogaster linker histone dH1 is required for transposon silencing and to preserve genome integrity. Nucleic Acids Res. 2012;40(12):5402–14.

138. Bayona-Feliu A, Casas-Lamesa A, Reina O, Bernués J, Azorín F. Linker histone H1 prevents R-loop accumulation and genome instability in heterochromatin. Nat Commun. 2017;8(1):283.

139. Laybourn PJ, Kadonaga JT. Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. Science (1979). 1991;254:238–245.

140. Li Z, Li Y, Tang M, Peng B, Lu X, Yang Q, et al. Destabilization of linker histone H1.2 is essential for ATM activation and DNA damage repair. Cell Res. 2018;28(7):756–70.

141. Fernández-Justel JM, Santa-María C, Martín-Vírgala S, Ramesh S, Ferrera-Lagoa A, Salinas-Pena M, et al. Histone H1 regulates non-coding RNA turnover on chromatin in a m6A-dependent manner. Cell Rep. 2022;40(11):111329.

142. Serna-Pujol N, Salinas-Pena M, Mugianesi F, Le Dily F, Marti-Renom MA, Jordan A. Coordinated changes in gene expression, H1 variant distribution and genome 3D conformation in response to H1 depletion. Nucleic Acids Res. 2022;50(7):3892–910.

143. Fan Y, Sirotkin A, Russell RG, Ayala J, Skoultchi AI. Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the H1(0) replacement subtype. Mol Cell Biol. 2001;21(23):7933–43.

144. Ponte I, Vidal-Taboada JM, Suau P. Evolution of the vertebrate H1 histone class: evidence for the functional differentiation of the subtypes. Mol Biol Evol. 1998;15(6):702–8.

145. Helliger W, Lindner H, Grubl-Knosp O, Puschendorf B. Alteration in proportions of histone H1 variants during the differentiation of murine erythroleukaemic cells. Biochem J. 1992;288(Pt 3):747–51.

146. Zlatanova J, Doenecke D. Histone H1 zero: a major player in cell differentiation? The FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 1994;8(15):1260–8.

147. Scaffidi P. Histone H1 alterations in cancer. Biochim Biophys Acta Gene Regul Mech. 2016 Mar 1;1859(3):533–9.

148. Morales-Torres C, Biran A, Burney MJ, Patel H, Henser-Brownhill T, Cohen AHS, et al. The linker histone H1.0 generates epigenetic and functional intratumor heterogeneity. Science (1979). 2016 Sep 30;353(6307):aaf1644.

149. Morales-Torres C, Wu MY, Hobor S, Wainwright EN, Martin MJ, Patel H, et al. Selective inhibition of cancer cell self-renewal through a Quisinostat-histone H1.0 axis. Nat Commun. 2020;11(1):1792.

150. Sancho M, Diani E, Beato M, Jordan A. Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. PLoS Genet. 2008;4(10):e1000227.

151. Mayor R, Izquierdo-Bouldstridge A, Millán-Ariño L, Bustillos A, Sampaio C, Luque N, et al. Genome distribution of replication-independent histone H1 variants shows H1.0 associated with nucleolar domains and H1X associated with RNA polymerase II-enriched regions. Journal of Biological Chemistry. 2015 Mar 20;290(12):7474–91.

152. Millán-Ariño L, Islam ABMMK, Izquierdo-Bouldstridge A, Mayor R, Terme JM, Luque N, et al. Mapping of six somatic linker histone H1 variants in human breast cancer cells uncovers specific features of H1.2. Nucleic Acids Res. 2014;42(7):4474–4493.

153. Talasz H, Sarg B, Lindner HH. Site-specifically phosphorylated forms of H1.5 and H1.2 localized at distinct regions of the nucleus are related to different processes during the cell cycle. Chromosoma. 2009;118(6):693–709.

154. Zheng Y, John S, Pesavento JJ, Schultz-Norton JR, Schiltz RL, Baek S, et al. Histone H1 phosphorylation is associated with transcription by RNA polymerases I and II. Journal of Cell Biology. 2010;189(3):407–15.

155. Sarg B, Helliger W, Talasz H, Forg B, Lindner HH. Histone H1 phosphorylation occurs site-specifically during interphase and mitosis: identification of a novel phosphorylation site on histone H1. J Biol Chem. 2006;281(10):6573–80.

156. Hergeth SP, Dundr M, Tropberger P, Zee BM, Garcia BA, Daujat S, et al. Isoform-specific phosphorylation of human linker histone H1.4 in mitosis by the kinase Aurora B. J Cell Sci. 2011;124(Pt 10):1623–8.

157. Happel N, Stoldt S, Schmidt B, Doenecke D. M phase-specific phosphorylation of histone H1.5 at threonine 10 by GSK-3. J Mol Biol. 2009;386(2):339–50.

158. Vicent GP, Nacht AS, Font-Mateu J, Castellano G, Gaveglia L, Ballare C, et al. Four enzymes cooperate to displace histone H1 during the first minute of hormonal gene activation. Genes Dev. 2011;25(8):845–62.

159. Koop R, DiCroce L, Beato M. Histone H1 enhances synergistic activation of the MMTV promoter in chromatin. EMBO J. 2003;22(3):588–99.

160. Vicent GP, Koop R, Beato M. Complex role of histone H1 in transactivation of MMTV promoter chromatin by progesterone receptor. J Steroid Biochem Mol Biol. 2002;83(1–5):15–23.

161. Saha A, Seward CH, Stubbs L, Mizzen CA. Site-specific phosphorylation of histone h1.4 is associated with transcription activation. Int J Mol Sci. 2020;21(22):8861.

162. Th'ng JPH, Guo XW, Swank RA, Crissman HA, Bradbury EM. Inhibition of histone phosphorylation by staurosporine leads to chromosome decondensation. Journal of Biological Chemistry. 1994;269(13):9568–73.

163. Izzo A, Schneider R. The role of linker histone H1 modifications in the regulation of gene expression and chromatin dynamics. Biochim Biophys Acta. 2016 Mar 1;1859(3):486–95.

164. Daujat S, Zeissler U, Waldmann T, Happel N, Schneider R. HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. Journal of Biological Chemistry. 2005;280(45):38090–5.

165. Andrés M, García-Gomis D, Ponte I, Suau P, Roque A. Histone H1 post-translational modifications: Update and future perspectives. Int J Mol Sci. 2020;21(16):5941.

166. Krishnakumar R, Gamble MJ, Frizzell KM, Berrocal JG, Kininis M, Kraus WL. Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. Science. 2008;319(5864):819–21.

167. Braunschweig U, Hogan GJ, Pagie L, van Steensel B. Histone H1 binding is inhibited by histone variant H3.3. EMBO J. 2009;28(23):3635–45.

168. Cao K, Lailler N, Zhang Y, Kumar A, Uppal K, Liu Z, et al. High-resolution mapping of H1 linker histone variants in embryonic stem cells. PLoS Genet. 2013;9(4):e1003417.

169. Izzo A, Kamieniarz-Gdula K, Ramírez F, Noureen N, Kind J, Manke T, et al. The Genomic Landscape of the Somatic Linker Histone Subtypes H1.1 to H1.5 in Human Cells. Cell Rep. 2013;3(6):2142–54.

170. Li JY, Patterson M, Mikkola HKA, Lowry WE, Kurdistani SK. Dynamic Distribution of Linker Histone H1.5 in Cellular Differentiation. PLoS Genet. 2012;8(8):e1002879.

171. Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob DNA. 2021;12(1):2.

172. Gustafsson N, Culley S, Ashdown G, Owen DM, Pereira PM, Henriques R. Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations. Nat Commun. 2016;7:12471.

173. Culley S, Tosheva KL, Matos Pereira P, Henriques R. SRRF: Universal live-cell super-resolution microscopy. International Journal of Biochemistry and Cell Biology. 2018;101:74–9.

174. Neguembor MV, Martin L, Castells-García Á, Gómez-García PA, Vicario C, Carnevali D, et al. Transcription-mediated supercoiling regulates genome folding and loop formation. Mol Cell. 2021;81(15):3065–81.

175. Martin L, Vicario C, Castells-García Á, Lakadamyali M, Neguembor MV, Cosma MP. A protocol to quantify chromatin compaction with confocal and super-resolution microscopy in cultured cells. STAR Protoc. 2021;2(4):100865.

176. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.

177. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

178. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.

179. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

180. Shin H, Liu T, Manrai AK, Liu SX. CEAS: Cis-regulatory element annotation system. Bioinformatics. 2009;25(19):2605–6.

181. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44(W1):W160-5.

182. Jin Y, Tam OH, Paniagua E, Hammell M. TEtranscripts: A package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics. 2015;31(22):3593–9.

183. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep. 2019;9(1):9354.

184. Donna Karolchik 1, Angela S Hinrichs, Terrence S Furey, Krishna M Roskin, Charles W Sugnet, David Haussler WJK. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 2004 Jan;32(Database issue):D493-6.

185. Navarro Gonzalez J, Zweig AS, Speir ML, Schmelter D, Rosenbloom KR, Raney BJ, et al. The UCSC Genome Browser database: 2021 update. Nucleic Acids Res. 2021 Jan 8;49(D1):D1046–57.

186. Gao T, Qian J. EnhancerAtlas 2.0: An updated resource with enhancer annotation in 586 tissue/cell types across nine species. Nucleic Acids Res. 2020;48(D1):D58–64.

187. Reinhold WC, Varma S, Sunshine M, Elloumi F, Ofori-Atta K, Lee S, et al. RNA sequencing of the NCI-60: Integration into CellMiner and CellMiner CDB. Cancer Res. 2019;79(13):3514–24.

188. Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol. 2015;16(3):144–54.

189. Li C, Luscombe NM. Nucleosome positioning stability is a modulator of germline mutation rate variation across the human genome. Nat Commun. 2020;11(1):1363.

190. Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, Bickmore WA. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. Hum Mol Genet. 2001;10(3):211–9.

191. Girelli G, Custodio J, Kallas T, Agostini F, Wernersson E, Spanjaard B, et al. GPSeq reveals the radial organization of chromatin in the cell nucleus. Nat Biotechnol. 2020;38(10):1184–93.

192. Maiser A, Dillinger S, Längst G, Schermelleh L, Leonhardt H, Németh A. Super-resolution in situ analysis of active ribosomal DNA chromatin organization in the nucleolus. Sci Rep. 2020;10(1):7462.

193. Burger K, Mühl B, Harasim T, Rohrmoser M, Malamoussi A, Orban M, et al. Chemotherapeutic drugs inhibit ribosome biogenesis at various levels. Journal of Biological Chemistry. 2010;285(16):12416–25.

194. Piña B, Suau P. Changes in the proportions of histone H1° subtypes in brain cortical neurons. FEBS Lett. 1987;210(2):161–4.

195. Lennox RW, Cohen LH. The histone H1 complements of dividing and nondividing cells of the mouse. Journal of Biological Chemistry. 1983;258(1):262–8.

196. Meergans T, Albig W, Doenecke D. Varied expression patterns of human H1 histone genes in different cell lines. DNA Cell Biol. 1997;16(9):1041–9.

197. Parseghian MH, Hamkalo BA. A compendium of the histone H1 family of somatic subtypes: An elusive cast of characters and their characteristics. Biochemistry and Cell Biology. 2001;79(3):289–304.

198. Zhang X, Jiang Q, Li J, Zhang S, Cao Y, Xia X, et al. KCNQ1OT1 promotes genome-wide transposon repression by guiding RNA–DNA triplexes and HP1 binding. Nat Cell Biol. 2022;24(11):1617–29.

199. Struhl K, Segal E. Determinants of nucleosome positioning. Nat Struct Mol Biol. 2013;20(3):267–73.

200. Singh AK, Mueller-Planitz F. Nucleosome Positioning and Spacing: From Mechanism to Function. J Mol Biol. 2021;433(6):166847.

201. Imbeault M, Helleboid PY, Trono D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature. 2017;543(7646):550–4.

202. Damert A. Phylogenomic analysis reveals splicing as a mechanism of parallel evolution of non-canonical SVAs in hominine primates. Mob DNA. 2018;9:30.

203. Carbone L, Alan Harris R, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J, et al. Gibbon genome and the fast karyotype evolution of small apes. Nature. 2014;513(7517):195–201.

204. Okhovat M, Nevonen KA, Davis BA, Michener P, Ward S, Milhaven M, et al. Co-option of the lineage-specific LAVA retrotransposon in the gibbon genome. Proc Natl Acad Sci U S A. 2020;117(32):19328–38.

205. Lupan I, Bulzu P, Popescu O, Damert A. Lineage specific evolution of the VNTR composite retrotransposon central domain and its role in retrotransposition of gibbon LAVA elements. BMC Genomics. 2015;16(1):389.

206. Ward MC, Zhao S, Luo K, Pavlovic BJ, Karimi MM, Stephens M, et al. Silencing of transposable elements may not be a major driver of regulatory evolution in primate iPSCs. Elife. 2018;7:e33084.

207. Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, et al. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. Nat Genet. 2013;45(7):836–41.

208. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze A V., Mikheenko A, et al. The complete sequence of a human genome. Science (1979). 2022;376(6588):44–53.

209. Hoyt SJ, Storer JM, Hartley GA, Grady PGS, Gershman A, de Lima LG, et al. From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. Science (1979). 2022;376(6588):eabk3112.

210. Choi J, Lyons DB, Kim MY, Moore JD, Zilberman D. DNA Methylation and Histone H1 Jointly Repress Transposable Elements and Aberrant Intragenic Transcripts. Mol Cell. 2020;77(2):310–23.

211. Chandra T, Ewels PA, Schoenfelder S, Furlan-Magaril M, Wingett SW, Kirschner K, et al. Global Reorganization of the Nuclear Landscape in Senescent Cells. Cell Rep. 2015 Feb;10(4):471–83.

212. Chang L, Li M, Shao S, Li C, Ai S, Xue B, et al. Nuclear peripheral chromatin-lamin B1 interaction is required for global integrity of chromatin architecture and dynamics in human cells. Protein Cell. 2022;13(4):258–80.

213. Zheng X, Hu J, Yue S, Kristiani L, Kim M, Sauria M, et al. Lamins Organize the Global Three-Dimensional Genome from the Nuclear Periphery. Mol Cell. 2018;71(5):802–15.

214. Briand N, Collas P. Lamina-associated domains: Peripheral matters and internal affairs. Genome Biol. 2020;21(1):85.

215. Sadaie M, Salama R, Carroll T, Tomimatsu K, Chandra T, Young ARJ, et al. Redistribution of the Lamin B1 genomic binding profile affects rearrangement of heterochromatic domains and SAHF formation during senescence. Genes Dev. 2013;27(16):1800–8.

216. Funayama R, Saito M, Tanobe H, Ishikawa F. Loss of linker histone H1 in cellular senescence. J Cell Biol. 2006;175(6):869–80.

217. Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods. 2012;9(10):999–1003.

218. Martínez-Balbás MA, Dey A, Rabindran SK, Ozato K, Wu C. Displacement of sequence-specific transcription factors from mitotic chromatin. Cell. 1995;83(1):29–38.

219. Wang F, Higgins JMG. Histone modifications and mitosis: countermarks, landmarks, and bookmarks. Trends Cell Biol. 2013;23(4):175–84.

220. Stoldt S, Wenzel D, Schulze E, Doenecke D, Happel N. G1 phase-dependent nucleolar accumulation of human histone H1x. Biol Cell. 2007;99(10):541–52.

221. Tani R, Hayakawa K, Tanaka S, Shiota K. Linker histone variant H1T targets rDNA repeats. Epigenetics. 2016;11(4):288–302.

222. Jiang X, Wen J, Paver E, Wu Y, Sun G, Bullman A, et al. H2A.B is a cancer/testis factor involved in the activation of ribosome biogenesis in Hodgkin lymphoma. EMBO Rep. 2021;22(8):e52462.

223. Shav-Tal Y, Blechman J, Darzacq X, Montagna C, Dye BT, Patton JG, et al. Dynamic sorting of nuclear components into distinct nucleolar caps during transcriptional inhibition. Mol Biol Cell. 2005;16(5):2395–413.

224. Lafontaine DLJ, Riback JA, Bascetin R, Brangwynne CP. The nucleolus as a multiphase liquid condensate. Nat Rev Mol Cell Biol. 2021;22(3):165–82.

225. Shin Y, Brangwynne CP. Liquid phase condensation in cell physiology and disease. Science (1979). 2017;357(6357):eaaf4382.

226. Stenström L, Mahdessian D, Gnann C, Cesnik AJ, Ouyang W, Leonetti MD, et al. Mapping the nucleolar proteome reveals a spatiotemporal organization related to intrinsic protein disorder. Mol Syst Biol. 2020;16(8):e9469.

227. Xu J, Ma H, Ma H, Jiang W, Mela CA, Duan M, et al. Super-resolution imaging reveals the evolution of higher-order chromatin folding in early carcinogenesis. Nat Commun. 2020;11(1):1899.

228. Willcockson MA, Healton SE, Weiss CN, Bartholdy BA, Botbol Y, Mishra LN, et al. H1 histones control the epigenetic landscape by local chromatin compaction. Nature. 2021;589(7841):293–298.

229. Yusufova N, Kloetgen A, Teater M, Osunsade A, Camarillo JM, Chin CR, et al. Histone H1 loss drives lymphoma by disrupting 3D chromatin architecture. Nature. 2021;589(7841):299–305.

230. Healton SE, Pinto HD, Mishra LN, Hamilton GA, Hamilton GA, Wheat JC, et al. H1 linker histones silence repetitive elements by promoting both histone H3K9 methylation and chromatin compaction. Proc Natl Acad Sci U S A. 2020;117(25):14251–8.

231. Zhang SM, Cai WL, Liu X, Thakral D, Luo J, Chan LH, et al. KDM5B promotes immune evasion by recruiting SETDB1 to silence retroelements. Nature. 2021;598(7882):682–7.

232. Morel KL, Sheahan A V., Burkhart DL, Baca SC, Boufaied N, Liu Y, et al. EZH2 inhibition activates a dsRNA–STING–interferon stress axis that potentiates response to PD-1 checkpoint blockade in prostate cancer. Nat Cancer. 2021;2(4):444–56.

233. Peng D, Kryczek I, Nagarsheth N, Zhao L, Wei S, Wang W, et al. Epigenetic silencing of TH1-type chemokines shapes tumour immunity and immunotherapy. Nature. 2015;527(7577):249–53.

234. Deblois G, Tonekaboni SAM, Grillo G, Martinez C, Kao YI, Tai F, et al. Epigenetic switch–induced viral mimicry evasion in chemotherapy-resistant breast cancer. Cancer Discov. 2020;10(9):1312–29.

235. Avgustinova A, Laudanna C, Pascual-García M, Rovira Q, Djurec M, Castellanos A, et al. Repression of endogenous retroviruses prevents antiviral immune response and is required for mammary gland development. Cell Stem Cell. 2021;28(10):1790–804.

236. Zhou D, Wu Z, Park JG, Fiches GN, Li TW, Ma Q, et al. FACT subunit SUPT16H associates with BRD4 and contributes to silencing of interferon signaling. Nucleic Acids Res. 2022;50(15):8700–18.

237. Pan D, Kobayashi A, Jiang P, De Andrade LF, Tay RE, Luoma AM, et al. A major chromatin regulator determines resistance of tumor cells to T cell-mediated killing. Science (1979). 2018;359(6377):770–5.

238. Shen JZ, Qiu Z, Wu Q, Finlay D, Garcia G, Sun D, et al. FBXO44 promotes DNA replication-coupled repetitive element silencing in cancer cells. Cell. 2021;184(2):352–69.

239. Song Y, Hou G, Diep J, Ooi YS, Akopyants NS, Beverley SM, et al. Inhibitor of growth protein 3 epigenetically silences endogenous retroviral elements and prevents innate immune activation. Nucleic Acids Res. 2021;49(22):12706–15.

240. Srour N, Villarreal OD, Hardikar S, Yu Z, Preston S, Miller WH, et al. PRMT7 ablation stimulates anti-tumor immunity and sensitizes melanoma to immune checkpoint blockade. Cell Rep. 2022;38(13):110582.

241. Bowling EA, Wang JH, Gong F, Wu W, Neill NJ, Kim IS, et al. Spliceosome-targeted therapies trigger an antiviral immune response in triple-negative breast cancer. Cell. 2021;184(2):384–403.

242. Zhou X, Singh M, Santos GS, Guerlavais V, Carvajal LA, Aivado M, et al. Pharmacologic Activation of p53 Triggers Viral Mimicry Response Thereby Abolishing Tumor Immune Evasion and Promoting Antitumor Immunity. Cancer Discov. 2021;11(12):3090–105.

243. Ishizuka JJ, Manguso RT, Cheruiyot CK, Bi K, Panda A, Iracheta-Vellve A, et al. Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade. Nature. 2019;565(7737):43–8.

244. Smith CL, Poleshko A, Epstein JA. The nuclear periphery is a scaffold for tissue-specific enhancers. Nucleic Acids Res. 2021;49(11):6181–95.

245. Kalashnikova AA, Winkler DD, McBryant SJ, Henderson RK, Herman JA, DeLuca JG, et al. Linker histone H1.0 interacts with an extensive network of proteins found in the nucleolus. Nucleic Acids Res. 2013;41(7):4026–35.

246. Kong Y, Rose CM, Cass AA, Williams AG, Darwish M, Lianoglou S, et al. Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. Nat Commun. 2019;10(1):5228.

247. Philippe C, Vargas-Landin DB, Doucet AJ, Van Essen D, Vera-Otarola J, Kuciak M, et al. Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. Elife. 2016;5:e13926.

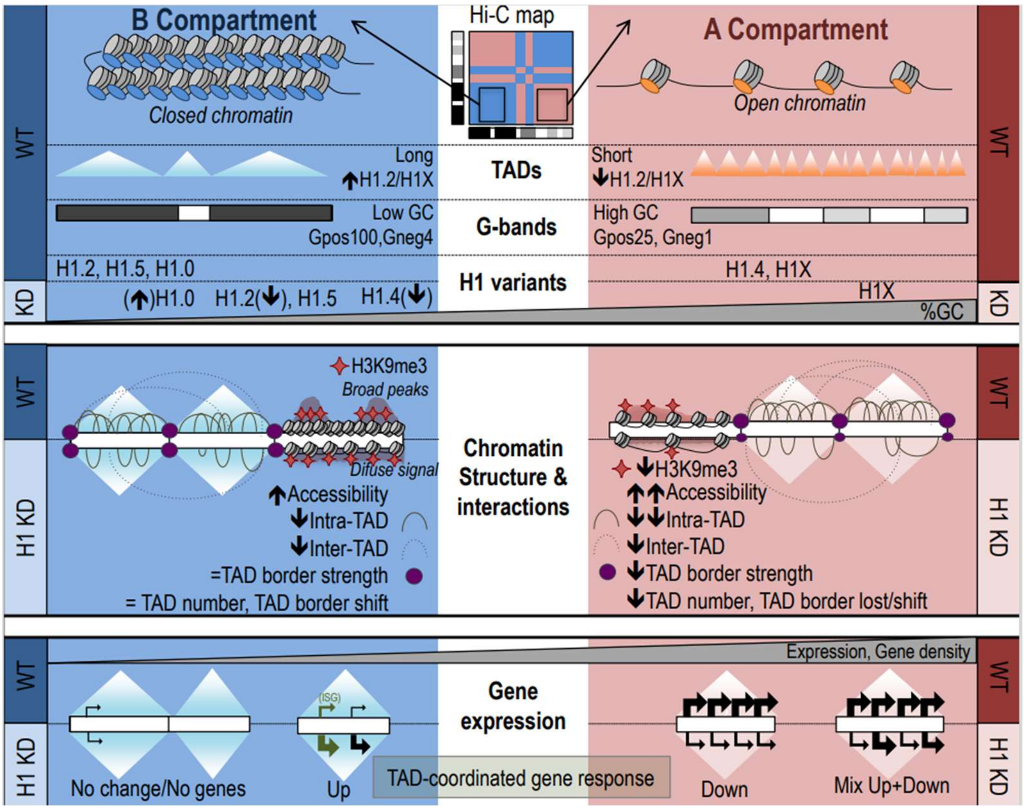
Appendix I.

Publication List

Núria Serna-Pujol*, **Mónica Salinas-Pena***, Francesca Mugianesi*, François Le Dily, Marc A Marti-Renom, Albert Jordan. Coordinated changes in gene expression, H1 variant distribution and genome 3D conformation in response to H1 depletion, *Nucleic Acids Research*, Volume 50, Issue 7, 22 April 2022, Pages 3892 3910, https://doi.org/10.1093/nar/gkac226

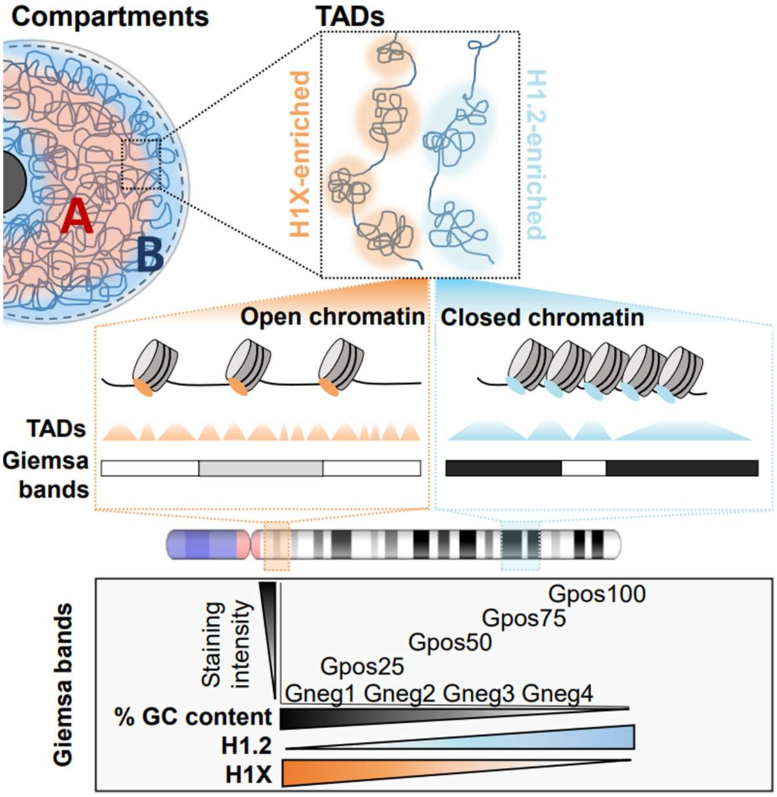
*Co-first authors

Abstract: Up to seven members of the histone H1 family may contribute to chromatin compaction and its regulation in human somatic cells. In breast cancer cells, knock-down of multiple H1 variants deregulates many genes, promotes the appearance of genome-wide accessibility sites and triggers an interferon response via activation of heterochromatic repeats. However, how these changes in the expression profile relate to the re-distribution of H1 variants as well as to genome conformational changes have not been yet studied. Here, we combined ChIP-seq of five endogenous H1 variants with Chromosome Conformation Capture analysis in wild-type and H1.2/H1.4 knock-down T47D cells. The results indicate that H1 variants coexist in the genome in two large groups depending on the local GC content and that their distribution is robust with respect to H1 depletion. Despite the small changes in H1 variants distribution, knock-down of H1 translated into more isolated but de-compacted chromatin structures at the scale of topologically associating domains (TADs). Such changes in TAD structure correlated with a coordinated gene expression response of their resident genes. This is the first report describing simultaneous profiling of five endogenous H1 variants and giving functional evidence of genome topology alterations upon H1 depletion in human cancer cells.



Núria Serna-Pujol, **Mónica Salinas-Pena**, Francesca Mugianesi, Natalia López-Anguita Francesc Torrent-Llagostera, Andrea Izquierdo-Bouldstridge, Marc A Marti-Renom, Albert Jordan, *FEBS J.*, Volume 288, Issue 6, March 2021, Pages 1989-2013, https://doi.org/10.1111/febs.15549

Abstract: Giemsa staining of metaphase chromosomes results in a characteristic banding useful for identification of chromosomes and its alterations. We have investigated in silico whether Giemsa bands (G-bands) correlate with epigenetic and topological features of the interphase genome. Staining of G-positive bands decreases with GC content; nonetheless, G-negative bands are GC heterogeneous. High-GC bands are enriched in active histone marks, RNA polymerase II, and SINEs and associate with gene richness, gene expression, and early replication. Low-GC bands are enriched in repressive marks, lamina-associated domains, and LINEs. Histone H1 variants distribute heterogeneously among G-bands: H1X is enriched at high-GC bands and H1.2 is abundant at low-GC, compacted bands. According to epigenetic features and H1 content, G-bands can be organized in clusters useful to compartmentalize the genome. Indeed, we have obtained Hi-C chromosome interaction maps and compared topologically associating domains (TADs) and A/B compartments to G banding. TADs with high H1.2/H1X ratio strongly overlap with B compartment, late replicating, and inaccessible chromatin and low-GC bands. We propose that GC content is a strong driver of chromatin compaction and 3D genome organization, that Giemsa staining recapitulates this organization denoted by high-throughput techniques, and that H1 variants distribute at distinct chromatin domains.



José Miguel Fernández-Justel, Cristina Santa-María, Sara Martín-Vírgala, Shereya Ramesh, Alberto Ferrera-Lagoa, **Mónica Salinas-Pena**, Javier Isoler-Alcaraz, Magdalena M. Maslon, Albert Jordan, Javier F. Cáceres, María Gómez, *Cell Reports*, Volume 40, Issue 11, 13 September 2022, 111329, https://doi.org/10.1016/j.celrep.2022.111329

Abstract: Linker histones are highly abundant chromatin-associated proteins with well-established structural roles in chromatin and as general transcriptional repressors. In addition, it has been long proposed that histone H1 exerts context-specific effects on gene expression. Here, we identify a function of histone H1 in chromatin structure and transcription using a range of genomic approaches. In the absence of histone H1, there is an increase in the transcription of non-coding RNAs, together with reduced levels of m6A modification leading to their accumulation on chromatin and causing replication-transcription conflicts. This strongly suggests that histone H1 prevents non-coding RNA transcription and regulates non-coding transcript turnover on chromatin. Accordingly, altering the m6A RNA methylation pathway rescues the replicative phenotype of H1 loss. This work unveils unexpected regulatory roles of histone H1 on non-coding RNA turnover and m6A deposition, highlighting the intimate relationship between chromatin conformation, RNA metabolism, and DNA replication to maintain genome performance.