UNIVERSITAT DE BARCELONA

Analysis of the genomic distribution of linker histone H1 variants in human

Anàlisi de la distribució genòmica de les variants d'histona H1 en humans

Andrea Izquierdo Bouldstridge

Tesi Doctoral

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

ANALYSIS OF THE GENOMIC DISTRIBUTION OF LINKER HISTONE H1 VARIANTS IN HUMAN

ANÀLISI DE LA DISTRIBUCIÓ GENÒMICA DE LES VARIANTS D'HISTONA H1 EN HUMANS

Memòria presentada per **Andrea Izquierdo Bouldstridge** per optar al grau de doctora per la Universitat de Barcelona.

Realitzada a l'Institut de Biologia Molecular de Barcelona, CSIC sota la direcció del Dr. Albert Jordan Vallès

Barcelona – 2018

Director de la tesi	Tutor de la tesi	Doctoranda
Albert Jordan Vallès	Francesc Mestres Naval	Andrea Izquierdo Bouldstridge

Para Mater, Gordon, Román y Lein

ACKNOWLEDGEMENTS

Quiero agradecer a todas las personas que me han acompañado en este proyecto y hecho posible este momento. Su ayuda, comprensión, cooperación y motivación han sido pilares indispensables para mí.

A mis compañeros de laboratorio, tanto los que estuvieron largos como cortos períodos de tiempo. Lluís, Neus, Erik, Núria, Carles, Alicia, Dani gracias por compartir vuestro buen humor conmigo y ayudarme a crecer científica, política y también personalmente. También a todas *las chicas de Marian*, Raquel, Stella, Simona y Claudia, por los chistes en los seminarios, las charlas en los pasillos y congresos, me llevo un gran recuerdo. También quiero agradecer a todas las personas que trabajaron y convivieron conmigo durante mi período como estudiante de carrera y master: Francesc, Jordi y Ferran por vuestra supervisión, apoyo, motivación y ayuda. A mis amigos *bioinformáticos* (Laura, Marc y Edu) por nuestras comidas, a los *Azorines*, Antonia, Montse, Raúl, Ismael, Lucia y muchos otros.

A Albert por permitirme realizar la tesis doctoral en su laboratorio, por su paciencia y valiosa opinión científica basada siempre en un gran razonamiento crítico.

Finalmente, y en especial, a toda mi familia y amigos.

Numancia, 2018

CONTENTS

CONTENTS

Seven linker histone H1 variants are present in human somatic cells with distinct prevalence across cell types. Using variant-specific antibodies to H1 and hemagglutinin (HA)-tagged recombinant H1 variants expressed in breast cancer cells, their genomic distribution was assessed. Specifically, ChIP-Seq data was obtained for two replication-dependent (H1.2 and H1.4) and replication-independent H1 variants (H1.0 and H1X) together with core histone H3. Briefly, we have previously reported (Appendix I) that H1.2 is the H1 variant that better correlates with gene repression. It was found enriched at GC-poor, gene-poor and intergenic chromosomal domains in addition to lamin-associated domains (LADs).

We further explored linker histone H1 variant distribution and strikingly, we found that distribution of replication-independent H1 variants (H1.0 and H1X) is distinct (Chapter I). H1.0 was found enriched at nucleolar features such as nucleolus-associated domains (NADs), nucleolus organizer regions (NORs) encoding for the 45S rDNA, specifically at non-transcribed spacers and also in 5S rDNA. Specific repetitive sequences such as SINE-VNTR-Alu (SVA) retrotransposons and telomeric and ACRO1 satellites showed also a specific enrichment of H1.0.

On the other hand, H1X has been associated to actively transcribed chromatin indicated by a colocalization with RNAPII-enriched regions and an enrichment towards the 3' end of active genes. In addition, constitutive exons, included alternatively spliced exons and retained introns are enriched in H1X. Further, specific non-coding RNA (miRNA and snoRNA), mainly found at introns showed a H1X enrichment. Our results point to a potential role of H1X in elongation, splicing or non-coding RNA regulation, which might be prompting gene transcription without changes in core histone PTMs.

Furthermore, depletion of multiple H1 variants (H1.2 and H1.4) triggers an interferon response due to an aberrant transcription of repetitive elements in breast cancer cells (Chapter II). Transcription of repetitive elements was observed by an increase in their RNA levels (RT-qPCR), increase in cytoplasmic dsRNA (immunofluorescence) and transcription of intergenic regions (RNA-Seq). Variants H1.2 and H1.4 seem to be critical in the observed phenotype but rescue experiments showed redundant functions for H1 variants. The molecular mechanism that leads to transcription of repetitive elements upon multiH1 KD, as happens for DE genes upon single or multiple H1 variants KD, is still unsolved. We were able to show an increase in nucleosome accessibility genome-wide (ATAC-Seq) that did not fully correlate with the observed transcriptional changes in multiple H1 depleted cells. Surprisingly, post-translational modifications of core histone remained unchanged.

Existeixen set variants d'histona H1 en cèl·lules somàtiques humanes amb una prevalença diferent segons el tipus cel·lular. Utilitzant anticossos específics contra variants d'H1 i variants d'H1 recombinants etiquetades amb hemaglutinina (HA), vam avaluar la seva distribució genòmica en cèl·lules de càncer de mama. Concretament, vam obtenir dades de ChIP-Seq per a dues variants d'H1 dependents de replicació (H1.2 i H1.4), les dues variants independents de replicació (H1.0 i H1X) i la histona H3. Anteriorment, vam observar (Appendix I) que H1.2 és la variant que millor correlaciona amb la repressió gènica. A més a més, H1.2 es trobava enriquida a dominis cromosòmics pobres en GC, pobres en gens i intergènics i també, als dominis associats a lamin (LADs).

Després, vam explorar amb més profunditat, la distribució de les variants d'histona H1 i, sorprenentment, vam trobar que la distribució dels variants independents de replicació (H1.0 i H1X) és diferent (Chapter I). H1.0 estava enriquida en regions genòmiques associades al nuclèol com els dominis associats al nuclèol (NADs), les regions organitzadores del nuclèol (NORs) que codifiquen per l'ARN ribosomal 45S, específicament en les regions espaiadores no transcrites i, també, en el ADN ribosomal 5S. Elements repetitius com els retrotransposons SINE-VNTR-Alu (SVA) i els satèl·lits telomèrics i ACRO1, també, mostraven un enriquiment específic d'H1.0.

Per altra banda, vam trobar que H1X estava associada a cromatina activa transcripcionalment ja que vam demostrar una col·locació amb les regions enriquides amb RNAPII i un enriquiment cap a l'extrem 3' de gens actius. A més, totes les regions codificants que s'inclouen en el transcrit final (exons constitutius, exons inclosos alternativament i introns retinguts) estan enriquides en H1X. Algunes especies d'ARN no codificant (miRNA i snoRNA), que es troben principalment en introns, mostraven un enriquiment en H1X. Els nostres resultats apunten a que H1X pot tenir un paper en la regulació de l'elongació, splicing o els ARNs no codificants, que podria estar induint la transcripció de gens, sense canvis en les modificacions post-traduccionals d'histones.

La depleció de varies variants d'histona H1 (H1.2 i H1.4) desencadena una resposta d'interferó degut a una transcripció aberrant d'elements repetitius en cèl·lules de càncer de mama (Chapter II). La transcripció d'elements repetitius la vam observar mitjançant un augment dels seus nivells d'ARN (RT-qPCR), un augment dels ARN de doble cadena (dsRNA) al citoplasma (immunofluorescència) i la transcripció de regions intergèniques (RNA-Seq). El mecanisme molecular que condueix a la transcripció d'elements repetitius, tal com succeeix en els gens desregulats en cèl·lules deplecionades d'una sola variant, encara no està resolt. Vam poder mostrar un augment global en l'accessibilitat a la cromatina (ATAC-Seq) que no correlaciona completament amb els canvis transcripcionals observats en deplecionar múltiples variants d'H1. Sorprenentment, les modificacions post-traduccionals d'histones es mantenen intactes.

Existen siete variantes de histona H1 en células somáticas humanas con una prevalencia diferente según el tipo celular. Utilizando anticuerpos específicos contra variantes de H1 y variantes de H1 recombinantes etiquetadas con hemaglutinina (HA), evaluamos su distribución genómica en células de cáncer de mama. Concretamente, obtuvimos datos de ChIP-Seq para dos variantes de H1 dependientes de replicación (H1.2 y H1.4), las dos variantes independientes de replicación (H1.0 y H1X) y la histona H3. Anteriormente, observamos (Appendix I) que H1.2 es la variante que mejor correlaciona con la represión génica. Además, H1.2 estaba enriquecida en dominios cromosómicos pobres en GC, pobres en genes y intergénicos y también, en los dominios asociados a lamin (LADs).

A continuación, exploramos con más profundidad, la distribución de las variantes de histona H1 y, sorprendentemente, encontramos que la distribución de las variantes independientes de replicación (H1.0 y H1X) es diferente (Chapter I). H1.0 estaba enriquecida en regiones genómicas asociadas al nucléolo como los dominios asociados al nucléolo (NADs), las regiones organizadoras del nucléolo (NORs) que codifican para el ARN ribosomal 45S, específicamente en las regiones espaciadoras no transcritas y, también, en el ADN ribosomal 5S. Elementos repetitivos como los retrotransposones SINE-VNTR-Alu (SVA) y los satélites teloméricos y ACRO1, también, mostraron un enriquecimiento específico de H1.0.

Por otro lado, encontramos que H1X estaba asociada a cromatina activa transcripcionalmente ya que demostramos una colocalización con las regiones enriquecidas en RNAPII y un enriquecimiento hacia el extremo 3' de genes activos. Además, todas las regiones codificantes que se incluyen en el transcrito final (exones constitutivos, exones incluidos alternativamente y intrones retenidos) están enriquecidas en H1X. Algunas especies de ARN no codificante (miRNA y snoRNA), que se encuentran principalmente en intrones, mostraban también un enriquecimiento en H1X. Nuestros resultados apuntan a que H1X puede tener un papel en la regulación de la elongación, splicing o los ARNs no codificantes, que podría estar induciendo la transcripción de genes, sin cambios en las modificaciones post-traduccionales de histonas.

La depleción de varias variantes de H1 (H1.2 y H1.4) desencadena una respuesta de interferón debido a una transcripción aberrante de elementos repetitivos en células de cáncer de mama (Chapter II). La transcripción de elementos repetitivos la observamos mediante un aumento de sus niveles de ARN (RT-qPCR), un aumento de los ARN de doble cadena (dsRNA) en el citoplasma (inmunofluorescencia) y la transcripción de regiones intergénicas (RNA-Seq). El mecanismo molecular que induce la transcripción de elementos repetitivos, tal como sucede en los genes desregulados en células deplecionadas de una sola variante, aún no está resuelto. Pudimos mostrar un aumento global en la accesibilidad a la cromatina (ATAC-Seq) que no correlaciona completamente con los cambios transcripcionales observados al deplecionar múltiples variantes de H1. Sorprendentemente, las modificaciones post-traduccionales de histonas se mantienen intactas.

ABBREVIATIONS

ABBREVIATIONS

A	Adenine
A3SS/A5SS	Alternative 3’/5’ splice sites
AFEs	Alternative first exons
ALEs	Alternative last exons
ASEs	Alternatively spliced exons
ATAC-Seq	Assay for transposase-accessible chromatin sequencing
ATRX	α thalassemia-mental retardation, X-linked
aza-dC	5-aza-2’-deoxycytidine
BRG1	Brahma-related gene 1
C	Cytosine
CAF-1	Chromatin assembly factor 1
CBP	C reaction protein
CDKs	Cyclin-dependent kinases
CENP-A	Centromeric protein A
cGAMP	Cyclic GMP-AMP
cGAS	cyclic GMP-AMP synthase
ChIA-PET	Chromatin interaction analysis by paired-end tag
ChIP	Chromatin immunoprecipitation
cLADs	Constitutive lamina-associated domains
CLIP-Seq	Cross-linking immunoprecipitation sequencing
CSCs	Cancer stem cells
CTCF	CCCTC-binding factor
CUL4A	E3 ubiquitin ligase cullin 4A
D	Aspartic acid
Dam	DNA adenine methyltransferase
DamID	DNA adenine methyltransferase identification
DAXX	Death domain-associated protein 6
DDR	DNA damage response
DE	Differentially expressed
DFC	Dense fibrillar component
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DRIP-Seq	DNA:RNA immunoprecipitation sequencing
DSB	Double strand breaks

dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
E	Glutamic acid
ERVs	endogenous retroviruses
ES	Enrichment score
ESCs	Embryonic stem cells
EZH2	Enhancer of zeste homolog 2
FAIRE-Seq	Formaldehyde-assisted isolation of regulatory elements sequencing
FC	Fibrillar center
FDR	False discovery rate
FISH	Fluorescence in situ hybridization
fLADs	Facultative lamina-associated domains
FPKM	Fragments per kb million
FRAP	Fluorescence recovery after photobleaching
G	Guanine
GAS	Gamma interferon activation site
GC	Granular component
GFP	Green fluorescent protein
GO	Gene ontology
GRO-Seq	Global run-on sequencing
GSEA	Gene set enrichment analysis
HA	Hemagglutinin
HDAC	Histone deacetylase
HDACi	Histone deacetylase inhibitor
hESC	human embryonic stem cells
Hi-C	High throughput chromosome conformation capture
HIRA	Histone cell cycle regulation-defective homolog A
HJURP	Holliday junction recognition protein
HMEC	Human mammary epithelial cells
HMG	High-mobility group
HMM	Hidden Markov model
HP1	Heterochromatin protein 1
HP1γ	Heterochromatin protein Cbx3
HSMM	Human skeletal muscle myoblasts
hTERT	human telomerase reverse transcriptase
HUVEC	Human umbilical vein endothelial cells
IAP	intracisternal A particle

IFN	Interferon
IP	Immunoprecipitation
iPSCs	induced pluripotent stem
iRNA	interference RNA
ISGs	Interferon-stimulated genes
ISRE	interferon-stimulated response element
iX	inactivated X chromosome
K	Lysine
KD	Knockdown
KO	Knockout
LADs	Lamin-associated domains
LINE	Long interspersed nuclear elements
LINEs	Long interspersed nuclear elements
lncRNA	Long non-coding RNA
Mb	Megabases
MeCP2	Methyl-CpG-binding protein 2
MEF	Mouse embryonic fibroblasts
miRNA	microRNA
MMTV	Mouse mammary tumour virus
MNase	Micrococcal nuclease
mRNA	messenger RNA
mtrRNA	mitochondrial ribosomal RNA
multiH1	Multiple H1
MXE	Mutually exclusive exons
NADs	Nucleolus associated domains
ncRNA	non-coding RNA
NES	Normalised enrichment score
NETs	Neutrophil extracellular traps
NHEK	Normal human epidermal keratinocytes
NHLF	Normal human lung fibroblasts
NL	Nuclear lamina
NORs	Nucleolus organizer regions
NRL	Nucleosome repeat length
PADIs	Peptidylarginine deiminases
PARP1	Poly ADP-ribose polymerase
PASs	Polyadenylation signals
Pc	Polycomb
PCA	Principal component analysis

PCR	Polymerase chain reaction
PDAC	Pancreatic ductal adenocarcinoma
PIKK	Phosphor-inositide 3-kinase-related protein kinase
pRb	Retinoblastoma tumour suppressor protein
PRC2	Polycomb repressive complex 2
PTB	Polypyrimidine tract binding protein
PTEN	Phosphatase and tensin homolog
PTMs	Post-translational modifications
Q	Glutamine
qPCR	Quantitative polymerase chain reaction
R	Arginine
rDNA	ribosomal DNA
RI	Retained introns
RIP-Seq	RNA immunoprecipitation sequencing
RNA	Ribonucleic acid
RNAPI	RNA polymerase 1
RNAPII	RNA polymerase 2
RPK	Reads per kb
RPKM	Reads per kb million
RPM	Read per million
RRBS	Reduced representation bisulphite sequencing
rRNA	ribosomal RNA
RT	Reverse transcriptase
S	Serine
SARs	Scaffold-associated regions
SE	Skipped exon
SINEs	Short interspersed nuclear elements
siRNA	small interfering RNA
SIRT1	Sirtuin-1
snoRNA	small nucleolar RNA
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
STAT	Signal transducer and activator transcription
SVA	SINE-VNTR-Alu
T	Threonine or Thymidine
TADs	Topologically associated domains
TAF-I	Template activating factor-I (also known as SET)
TERRAs	Telomeric repeat-containing RNAs

TEs	Transposable elements
TKO	Triple knockout
TPase	Transposase
TPM	Transcripts per million
tRNA	transfer RNA
TSA	Trichostatin A
TSS	Transcription start site
WGBS	Whole genome bisulphite sequencing

INTRODUCTION

INTRODUCTION

1. CHROMATIN

1.1. The nucleosome and chromatin structure

Deoxyribonucleic acid (DNA) is a molecule that carries the genetic information, which determines growth, development, functioning and reproduction of all known living organisms and many viruses. DNA is a double-stranded molecule, as first described James Watson and Francis Crick. Two antiparallel strands are held together by weak hydrogen bonds. These strands are strings of nucleotides joined by phosphodiester bonds, each one containing one of four nitrogen-containing bases (cytosine (C), guanine (G), adenine (A) or thymine (T)) a sugar called deoxyribose and a phosphate group. Base pairs naturally only occur between adenine and thymine (A = T) and cytosine and guanine (C ≡ G). These nitrogenous bases are hydrophobic in contrast to the deoxyribose and phosphate molecules, which are hydrophilic. DNA is coiled such that the nitrogenous bases are in the inner portion of the molecule avoiding the water-containing nucleoplasm. In addition, the double helix is skewed as the strands are not directly opposite to each other leading to a major and minor groove (Figure 2). In the major groove the edges of the bases are more accessible allowing a better recognition of specific sequences by transcription factors and other sequence-specific binding proteins. In addition, the double helix structure of DNA allows its replication and transcription.

A human somatic cell (diploid) contains approximately 6 billion base pairs, with roughly 2m of length that are tightly packed in a nucleus, with an average size of 10µm. The proteins in charge of this huge chromosomal DNA compaction are called histones and the resulting DNA-histone complex, chromatin. Histones are a family of small, positively charged proteins termed H1, H2A, H2B, H3 and H4. DNA is negatively charged, due to the phosphate groups in its phosphate-sugar backbone, so histones bind with DNA very tightly. Two copies of the four core histones (H2A, H2B, H3 and H4) form an octamer that constitutes the basic repeating structural (and functional) unit of chromatin, the nucleosome (Figure 1, 2). Approximately, 146bp of DNA are wrapped around a nucleosome. All nucleosomes are connected by stretches of “linker DNA”, which can be up to 80bp. The formed chromatin structure resembles “beads on a string” of DNA under electron microscope [1, 2] (Figure 1B).

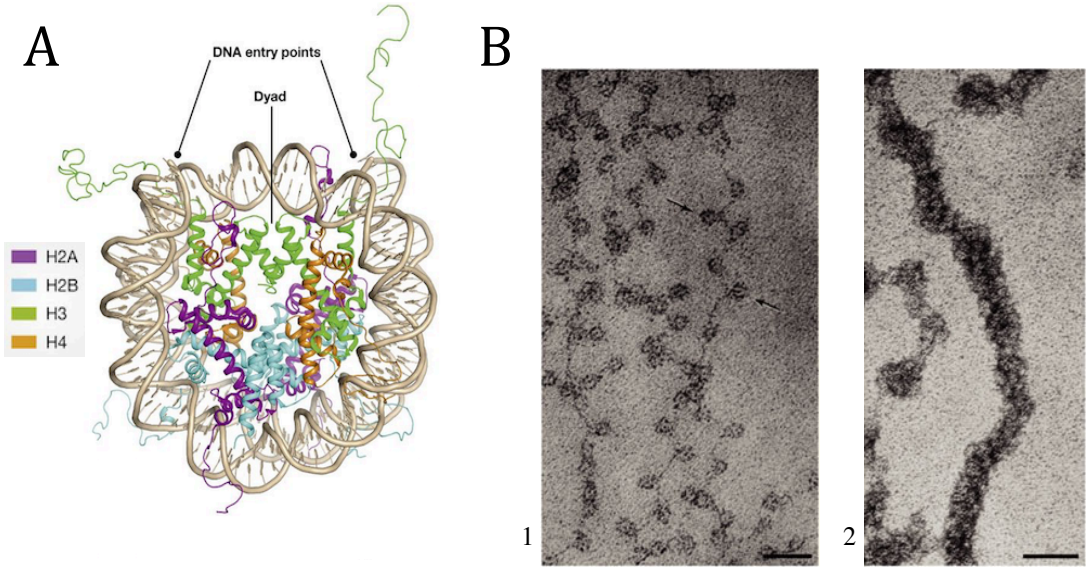


Figure 1. The nucleosome core particle (**A**) Crystal structure of a nucleosome particle. Two copies of each core histone (H2A: purple, H2B: blue, H3: green and H4: orange) and 146bp of DNA wrapped around form the nucleosome. (**B**) Electron micrograph of chromatin showing the "beads on a string" conformation (1) and the 30nm fiber (2). Arrows indicate nucleosomes. Figures adapted from [5, 6].

The remaining histone, histone H1, binds to nucleosomes near the entry/exit sites of linker DNA. The binding of linker histone H1 forms the chromatosome, which confers to chromatin a higher level of compaction allowing the formation of the so-called "30nm fiber" (Figure 1B). So, stabilization of condensed states of chromatin is the most commonly attributed function to linker histone H1 [2, 3]. Finally, chromatin is tightly condensed to form chromosomes, founding the highest level of compaction in the chromatids at metaphase (Figure 2).

Additionally, chromosomes are not randomly distributed along the nucleus but rather form territories where distant nucleosomes from different chromosomes may interact. Chromatin plays also an important role in the 3D organization of the nucleus, which varies along differentiation, tissue, and cell type… [4].

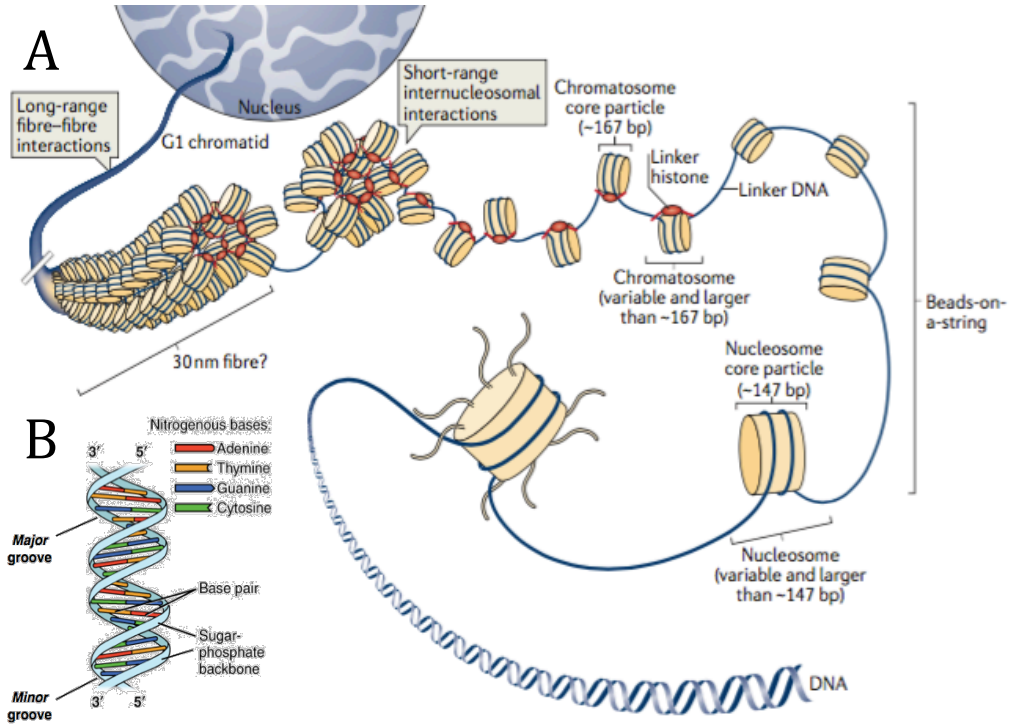


Figure 2. Different levels of DNA compaction and chromatin structure. (**B**) At the simplest level, DNA is a double-stranded helical structure. (**A**) DNA is complexed with histones to form nucleosomes, the basic chromatin unit. With the binding of linker histone H1 to linker DNA, the nucleosomes fold up to produce a "30nm fiber". The "30nm fiber" is tightly coiled to finally produce the highly condensed chromatin of a chromatid at metaphase. Adapted from [2, 4].

1.2. Core histones

Core histones control the packaging and ordering of DNA into units called nucleosomes. They are a family of small (11-16.2 kDa) basic proteins highly conserved across evolution with four major members: H2A, H2B, H3 and H4. Multicopy and intronless genes encode all core histones and their mRNAs are non-polyadenylated. Its structure is composed of a globular domain and tails. Their globular domains mediate the formation of the octamer of two copies of each member and the stabilization of two wraps of DNA. On the other hand, histone tails at the N-terminal domain are highly basic (rich in lysines and arginines) and extend from the surface of the nucleosome. One long tail at the C-terminal domain is only found in H2A. Tails contribute to inter-nucleosomal interactions and folding of chromatin due to their interaction with DNA-binding proteins. In addition, they are post-translational modified allowing a modulation of chromatin compaction by changing the physical properties of individual nucleosomes or by interacting with specific partners. Another level of chromatin modulation is at the level of the less abundant and non-canonical histone variants.

1.2.1. Post-translational modifications

As mentioned, core histones are post-translational modified mainly at their N-terminal tail. They can be phosphorylated at serines (S) and threonines (T), methylated and acetylated at lysines (K) and arginines (R), and ubiquitylated, sumoylated and ribosylated at lysines (Table 1) [7]. Many studies have shown that specific combinations of histones post-translational modifications (PTMs) have a clear correlation with cellular processes and functions (Table 1). For instance, tri-methylation of H3 at K9 has been related to constitutive heterochromatin, the condensed and transcriptionally inactive state of chromatin. H3K9me3 binds heterochromatin protein 1 (HP1) responsible of the transcriptional repression and actual formation and maintenance of heterochromatin [8]. Further, di-methylation at K9 is a characteristic PTM of the inactivated X chromosome (iX) [9] and phosphorylation of H2A at S10 and T119 and H3 at T3, S10 and S28 of condensed mitotic chromatin.

Core histones PTMs at promoter regions have been found to correlate with the transcriptional status of a gene. For example, H3K27me3 and H4K20me3 are found at inactive promoters while H3K4me3, H3 and H4 acetylation at active ones (Figure 3) [10]. Another interesting example is at enhancers, where core histones PTMs are also different depending in its transcriptional status. Active enhancers are bordered by widely spaced nucleosomes carrying modifications such as H3K4me2 and H3K27ac; in contrast to poised enhancers, which have H3K27me3 and reduced chromatin accessibility. Finally, repressed enhancers are at dense nucleosome assemblages with H3K27me3 [11].

All these correlations of core histones PTMs with transcription lead to the controversial hypothesis of a “histone code” proposed by Strahl and Allis, where distinct core histone PTMs are combined to create a read code for other proteins to trigger distinct cascades [12]. One of the arguments against this theory is methodological, as populations of cells have been used and it cannot distinguish whether two PTMs might be from two subpopulations. In addition, some PTMs are also transient. While everybody agrees about the correlation, the interpretation of the combination of core histone PTMs depends on the cellular context and gene studied. An interesting result was found while studying developmentally regulated genes in fly and worm, where those genes are transcribed in the absence of canonical histone activation marks [13]. Their results support a more important regulation of gene expression by transcription factors rather than by histone PTMs.

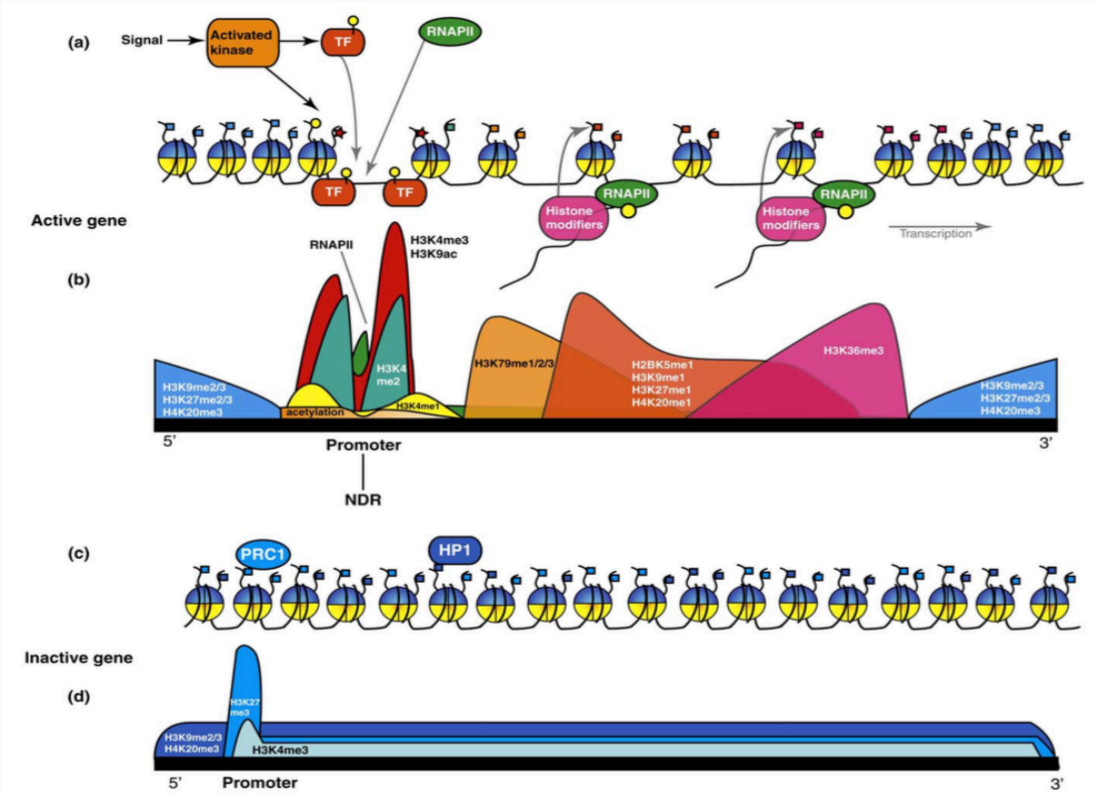


Figure 3. Main core histones post-translational modification found at genes and promoters with different transcriptional status (active and repressed genes). Figure taken from [10].

Table 1. Main histone post-translational modifications. Alternative residue numbers that refer to mammalian histones are shown in red. Continues in next page. Table adapted from [7].

	Position	Enzyme			Function	Recognized by	Reversed by		
		S. cerevisiae	*D. melanogaster*	Mammals			*S. cerevisiae*	*D. melanogaster*	Mammals
Lys Acetylation	H2AK4 H2AK5 H2AK7	Esa1	Tip60	TIP60, P300/CBP, HAT1			Rpd3		
	H2BK11 H2BK12 H2BK16 H2BK15	Gcn5 Gcn5, Esa1		P300/CBP, ATF2			Rpd3, Hda1		
	H3K9	Gcn5, Rtt109	dGcn5	SRC1, PCAF			Rpd3, Set3, Hda1,	dHDAC1	SIRT1/6
	H3K14	Gcn5, Hpa2, Esa1, Elp3, Sas2, Sas3	dGcn5, Taf1, dCBP	P300/CBP, TAF1, hGCN5, PCAF, MOZ, MORF, TIP60, SRC1, HBO1	Transcription activation	Bromo, PHD	Hos2, Hst1		
	H3K18 H3K23	Gcn5 Gcn5, Sas3	dCBP	P300/CBP					
	H3K27	Gcn5, Rtt109	dCBP						
	H3K36	Gcn5							
	H3K56	Rtt109	dCBP	P300/CBP		PH, (Snf5)	Hst3/4	Sir2	SIRT1/2/6
	H4K5	Esa1	Hat1, dCBP	HAT1, TIP60, P300, HBO1		Bromo	Rpd3, Set3, Hos2, Hst1		
	H4K8	Esa1, Elp3, Gcn5	dCBP	TIP60, P300, HBO1		Bromo			
	H4K12	Hat1, Esa1, Hpa2							
	H4K16	Sas2, Esa1	Mof, Atac2	hMOF, TIP60, ATF2		Bromo	Sir2	Sir2	SIRT1/2

Table 1. Main histone post-translational modifications. Alternative residue numbers that refer to mammalian histones are shown in red. Continues in next page. Table adapted from [7].

	Position	Enzyme			Function	Recognized by	Reversed by		
		S. cerevisiae	*D. melanogaster*	Mammals			*S. cerevisiae*	*D. melanogaster*	Mammals
Lys Methylation	H3K4	Set1	Trx, Trr, Ash1, Set1	SET1, NSD2-3, SET7/9, MLL1-4, SMYD3, ASH1L	Transcription activation	PHD, Chromo, WD40, ADD, Tudor, MBT, Zf-CW	Jhd2	Lid, Su(var)3-3	LSD1, AOF1, JARID1A-D
	H3K9		Su(var)3-9, Ash1, G9a	SUV39H1/2, G9a, Eu-HMT1, SETDB1, RIZ1, ASH1	Silencing	PHD, Chromo, Tudor, WD40, Ankyrin	Rph1	dKDM48	LSD1, JHDM2A/B, JMJD2A-D, KIAA1718, PHF8
	H3K23					Chromo			
	H3K27		E(z)	EZH1/2, NSD2-3, G9a	Silencing	Chromo, WD40			
	H3K36	Set2	Set2, Mes4, Ash1	SETD2-2, NSD1-3, SMYD2, ASH1L, SETMAR	Transcription elongation	Chromo, PHD, PWWP	Jhd1, Rph1, Gis1	dKMM4A/B	JHDM1A/B, JMJD2A-C
	H3K79	Dot1	Grappa	DOT1L		Tudor			
	H4K20		Pr-set7, Suv4-20, Ash1	ASH1L, NSD1, Pr-SET7, SUV4-20H, NSD2	Transcription activation /repression	Tudor, MBT, PWWP, WD40			KIAA1718, PHF8
Arg Methylation	H3R2me2a H3R2me2s			PRMT6, PRMT5/7	Transcription repression/ activation				JMJD6
	H3R8			PRMT5	Transcription repression				PAD4
	H3R17			CARM1	Transcription activation				PAD4
	H3R26			CARM1					
	H4R3			PRMT1, PRMT5	Transcription activation				JMJD6, PAD4

Table 1. Main histone post-translational modifications. Alternative residue numbers that refer to mammalian histones are shown in red. Table adapted from [7].

	Position	Enzyme			Function	Recognized by	Reversed by		
		S. cerevisiae	*D. melanogaster*	Mammals			*S. cerevisiae*	*D. melanogaster*	Mammals
Phosphorylation	H2BS10 H2BS14	Ste20		MST1					
	H3S10	Snf1, Ipl1	Jil1, Aurora B	MSK1/2, IKKα, PKB, RSK2, PIM1, Aurora B, JNK	Transcription activation, mitosis, meiosis	(Gcn5)	Glc7	PP2A	
Ubiquitination	H2AK119		Ring1B	Ring1B, 2A-HUB	Transcription repression			Calypso	USP3, USP16, USP21, USP22, A2-DUB
	H2BK123 H2BK120	Rad6, Bre1	Rad6, Bre6	HR6A/B, RNF20/40	Transcription activation	(Cps35)	Ubp8, Ubp10	Nonstop, Scrawny, USP7	
Sumoylation	H2AK126 H2AB6/7 H2ABK16/K17	Ubc9, Siz1, Siz2			Transcription repression				
	H4K5 K8 K12 K16 K20	Ubc9, Siz1, Siz2		UBC9					

1.2.2. Variants

All core histones have diversified into a wide range of variants with different sequence and structure, which have been conserved through evolution. In contrast to canonical core histones, which are encoded in multicopy genes and highly expressed, histone variants have a single or low copy number of genes and a very low expression. They are deposited in replication-independent manner to replace their conventional histone counterpart in specific genomic localizations and through specific processes such as differentiation or maintenance of a chromatin states. Nucleosome containing histone variants have different structural and functional properties, affecting chromatin through histone PTMs and specific interactions. It is worth saying that histone variants are also post-translational modified. Thus, histone variants confer another layer in the complexity of chromatin regulation.

Here, we will focus in the best-studied histone variants, members of H2A and H3 families (Table 2).

Table 2. Histone H3 and H2A variants in human. Table adapted from [14, 15].
[a] *Drosophila melanogaster* has a single H2A variant, H2Av in addition to the major H2A. H2Av is not only member of H2A.Z family; it also contains an SQ motif similar to mammalian H2A.X. It is phosphorylated at S137 and hence it is a functional homologue of H2A.X.

Histone	Variant	Deposition	Functional association
H3	H3.3	Replication- independent	Transcriptionally active regions
	CENP-A	Replication-dependent	Centromeric nucleosomes
H2A	H2A.Z	Replication-independent	Different functions in various organisms: maintenance of pericentric and telomeric heterochromatin, transcriptional activation and viability
	H2A.X	Replication-independent	Sex body in mammals, site of DNA double stranded breaks; condensation and silencing of male sex chromosome.
	MacroH2A	Replication-independent	Inactivation of X-chromosome, interferes with both transcription factor binding and SWI/SNF remodelling
	H2A.Bbd	Replication-independent	Close spacing of nucleosomes

H3.3

H3.3 variant differs from the two canonical H3 variants (H3.1 and H3.2) by five and four aminoacids, respectively, conferring to H3.3 a different interaction network with regulatory factors and histone chaperones.

For instance, H3.3 variant is enriched throughout the gene body of transcribed

genes as well as their promoter regions. In addition to active promoter, H3.3 is also found at inactive promoters accounting for a poised state of these genes [16]. Histone PTMs that mark active chromatin such as acetylation and H3K4 methylation are increased in H3.3. Indeed, H3.3 also has specific PTMs such as phosphorylation of S31 in mitosis. What is the exact function of these histone H3.3 PTMs and how they are established are relevant questions to be answered in the field.

While canonical H3 is incorporated to chromatin by the chaperone complex CAF-1, H3.3 incorporation is mediated by at least two different chaperone complexes: HIRA and DAXX-ATRX. HIRA is responsible for depositing H3.3 at active transcribed regions in contrast to ATRX chaperone and its co-factor DAXX, which deposit H3.3 at telomeres and pericentric heterochromatin in mouse embryonic stem cells (ESC) and mouse embryonic fibroblasts (MEF), respectively. Still, the specific function of H3.3 at these highly repressed regions needs to be further explored to determine if it plays an essential role in cell division and genome stability or if it is linked to the transcriptional status.

CENP-A

CENP-A is the H3 variant highly enriched at centromeres. CENP-A has evolved rapidly, so its protein sequence differs amongst species. Still, CENP-A is thought to be essential for centromere structure and function and kinetochores formation in all eukaryotes. The main structural differences with canonical H3 are found at two regions of the N-terminal tail, which confers to centromeres a more compacted and unique chromatin. Nucleosomes containing CENP-A allow the binding of other centromere-formation proteins such as CENP-B. In human, CENP-A is a replication dependent variant incorporated by a specific chaperone named HJURP.

H2A.X

Histone H2A.X is defined by its SQ[E/D]ϕ motif (where ϕ is a hydrophobic aminoacid) in the C-terminal region, which is not found in canonical H2A. After DNA damage, members of the phosphor-inositide 3-kinase-related protein kinase (PIKK) family phosphorylate this C-terminal tail at S139 (referred as γH2A.X). γH2A.X expands several Mb, in mammals, from double strand DNA breaks sites creating the so-called “γH2A.X foci”. Upon the activation of the DNA damage response, the creation of “γH2A.X foci” is an early event that triggers structural chromatin alterations at those sites to foster the DNA repair.

Moreover, γH2A.X and the DNA repair machinery have been found involved in the process of X chromosome inactivation in mammals.

H2A.Z

Histone H2A.Z is only 60% identical to canonical H2A. It is highly conserved, as the identity is roughly 80% between different species. Three isoforms have been described in humans: H2A.Z.1, H2A.Z.2 and H2A.Z.2.2. Many H2A.Z biological roles have been revealed in transcription regulation, DNA repair, heterochromatin formation, chromosome segregation and mitosis [17]

H2A.Z has been found enriched at gene promoters in several organisms although having both active and repressive roles on transcription [18] by affecting nucleosome mobility and positioning. In addition, H2A.Z is associated with enhancers and insulators as well as heterochromatin.

MacroH2A

MacroH2A is about three times larger than canonical H2A as it harbours about 30 kDa more in its C-terminal tail extending out from the nucleosome [19]. The repressive role of macroH2A in chromosome X inactivation (iX) has been widely studied setting the view of macroH2A as an epigenetic repressor. However, some studies challenge it by reporting a positive role of macroH2A, not yet mechanistically understood, in some target genes.

In addition, macroH2A has also a role in maintaining and stabilizing differentiated epigenomes as knockdown (KD) experiments in frogs showed an inhibition of reprogramming [20].

H2A.Bbd

H2A.Bbd (Barr body deficient) variant has a shorter C-terminal tail than canonical H2A and a part of the docking domain is also missing. These features confer to H2A.Bbd-containing chromatin a more open structure and a less tight wrapped DNA. So, nucleosomes bearing this variant are less resistant to MNase digestion and enriched within actively transcribed genes. As expected, analysis of transcription profiles from cells depleted for H2A.Bbd showed a deregulation of mRNA splicing patterns and a widespread genic down-regulation [21].

H2A.Bbd is not ubiquitously expressed, being strongly expressed in testis and less in brain. Thus, H2A.Bbd might have a tissue-specific function as shown in mouse spermatogenesis [22].

1.3. Types of chromatins

Chromatin has classically been divided in two main types: heterochromatin and euchromatin. Heterochromatin is characterized by densely packed nucleosomes with high levels of DNA methylation, high AT content and gene-poor genomic regions with low or none transcription. It is more frequently localized at the nuclear periphery and surrounding nucleoli as shown by its darker staining in transmission electron microscopy (Figure 4). A major function of heterochromatin is to protect the underlying DNA from being accessed by dedicated machineries such of those of transcription and DNA repair. Heterochromatin has been further subdivided into facultative and constitutive heterochromatin.

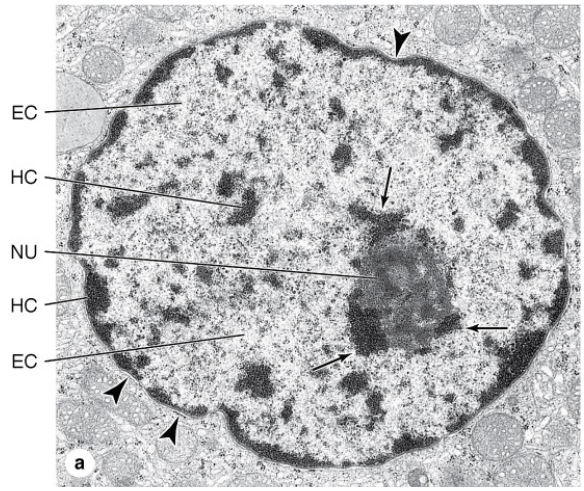


Figure 4. Transmission electron microscope view of a typical cell nucleus clearly shows heterochromatin (HC) and a more diffuse staining of euchromatin (EC). The arrows indicate the nucleolus-associated heterochromatin around the nucleolus (NU). Arrowheads indicate areas where the perinuclear space between the two membranes of the nuclear envelope is clearly seen. Just inside the nuclear envelope is a thin electron-dense region containing the nuclear lamina and more heterochromatin. Figure from [23].

An enrichment of core histones, high levels of H3K9me3 and H4K20me3, and the binding of HP1 proteins characterize constitutive heterochromatin. It is found at pericentromeric, telomeric and ribosomal regions and localized at the same genomic regions in every cell type. All three regions contain, in most organisms, repetitive sequences. Telomeres are constituted by a short highly conserved DNA sequence (5' TTAGGG 3') with specific binding-proteins. In contrast, pericentromeric regions contain the vast majority of constitutive heterochromatin and the repetitive sequences found are less conserved. In humans, centromeres consist mainly of alpha satellites and pericentromeric regions of specific satellites such as satellites I, II and III, among others. An epigenetic dysregulation in the expression of these repetitive sequences in

telomeric, pericentromeric and centromeric regions leads to abnormal processes such as cancer, chromosomal rearrangements… In addition, it is worth saying that constitutive heterochromatin is as well found at specific loci along different genomes.

On the other hand, facultative heterochromatin refers to regions with genes that need to be silent in a developmental, cell-specific manner. H3K27me3 and H2AK119ub are repressive histone marks of facultative heterochromatin. Polycomb proteins play an important role in linage choices during differentiation and development through chromatin modification. PRC2 contains the H3K27 methyltransferase (EZH2) and PRC1 the E3 ubiquitin ligases (RING1B) that mediates H2AK116 ubiquitinilization and also Polycomb that mediates, through its chromodomain, the binding to H3K27me3. It is worth saying that Polycomb complexes have been also found in constitutive H3K9me3-containing pericentromeric chromatin although, under specific processes [24].

In contrast to heterochromatin, euchromatin is associated with activation having high GC content, low levels of DNA methylation and containing gene-rich regions highly transcribed.

With the appearance of massive parallel sequencing another layer of information has been added and the classical heterochromatin/euchromatin classification has been reassessed. Filion et al. identified five major types of chromatin in *Drosophila melanogaster* cell Kc167 [25]. They used ChIP-Seq data from 53 selected chromatin components including a variety of histone-modifying enzymes, specific histone PTMs binding proteins, general members of the transcriptional machinery, nucleosomes remodelers, insulators, heterochromatin proteins, structural components of chromatin and a selections of DNA binding factors. After applying a Principal Component Analysis (PCA), the first three principal components were fitted onto a 5-state Hidden Markov Model (HMM). The identified states were named in five different colours (GREEN, BLUE, BLACK, RED and YELLOW) (Figure 5).

GREEN chromatin corresponds to the well-known heterochromatin having a high content of heterochromatin protein 1 (HP1), a H3K9 methyltransferase (Su(Var)3-9). As expected, a high enrichment of its representative core histone PTM, H3K9me2, is found. BLUE chromatin represents Polycomb repressed chromatin as shown by an extensive binding of Polycomb related proteins (PC, E(Z), PCL and SCE) and a high content of H3K27me3.

BLACK chromatin is the most abundant chromatin covering a 48% of the probed genome and under developmental control. Interestingly, consists of a previously

unknown combination of chromatin-binding proteins contributing to a strongly repressive environment. The lack of Polycomb proteins, HP1 and Su(Var)3-9, active histone PTMs (H3K4me3 and H3K79me2) and an extremely low transcription lead to the assumption that it is a different type of heterochromatin previously unknown.

On the other hand, transcriptionally active euchromatin was subdivided in RED and YELLOW. In contrast to the other identified types, RED and YELLOW chromatin have high levels of H3K4me3 and H3K79me3 and low levels of H3K9me2 and H3K27me3, as expected for actively transcribed chromatin (Figure 5). However, RED and YELLOW chromatin differ in their molecular organization; regulate a different subset of genes and in their H3K36me3 content. Why active genes in RED chromatin lack H3K36me3 remains to be elucidated.

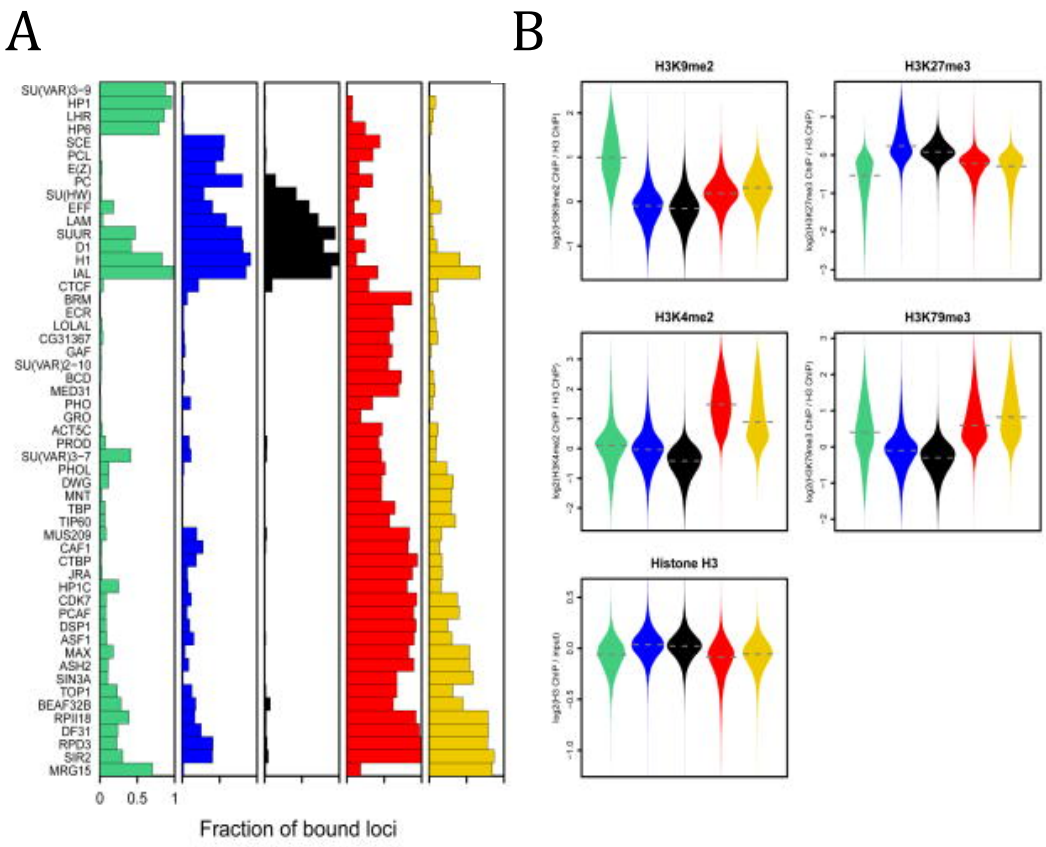


Figure 5. Five chromatin types identified by Filion et al. in *Drosophila melanogaster*. Distinct combinations of chromatin components (**A**) and core histones PTMs (H3K9me2, H3K27me3, H3K4me2, and H3K79me3) and Histone H3 (**B**) in each identified chromatin type (GREEN, BLUE, BLACK, RED and YELLOW). Figure adapted from [25].

In the chromatin classification done by Filion et al. only ChIP-Seq data of chromatin components was analysed. In other words, histone PTMs, nucleosome

accessibility and other datasets with chromatin information were not included to perform the classification. Histone PTMs were rather used to perform enrichment analysis in the already identified chromatin types.

Another interesting study is the one performed by Ernst and Kellis, which applied a similar HMM method using 38 different histone PTMs in human CD4 T-cells, as well as histone variant H2A.Z, RNA polymerase 2 (RNAPII) and CTCF [26]. They were able to identify up to 51 different chromatin states representing promoter, transcription-associated, active intergenic, large-scale repressed and repeat-associated regions in one cell line. The number of identified states (51) shows the enormous complexity of chromatin factors and core histone PTMs combinations found in only one human epigenome.

Additionally, Ernst et al. identified chromatin states in nine human cell lines [27]. They identified 15 chromatin states present in all cells lines studied, see Figure 6 for a summary of each identified chromatin state and its associated genomic region.

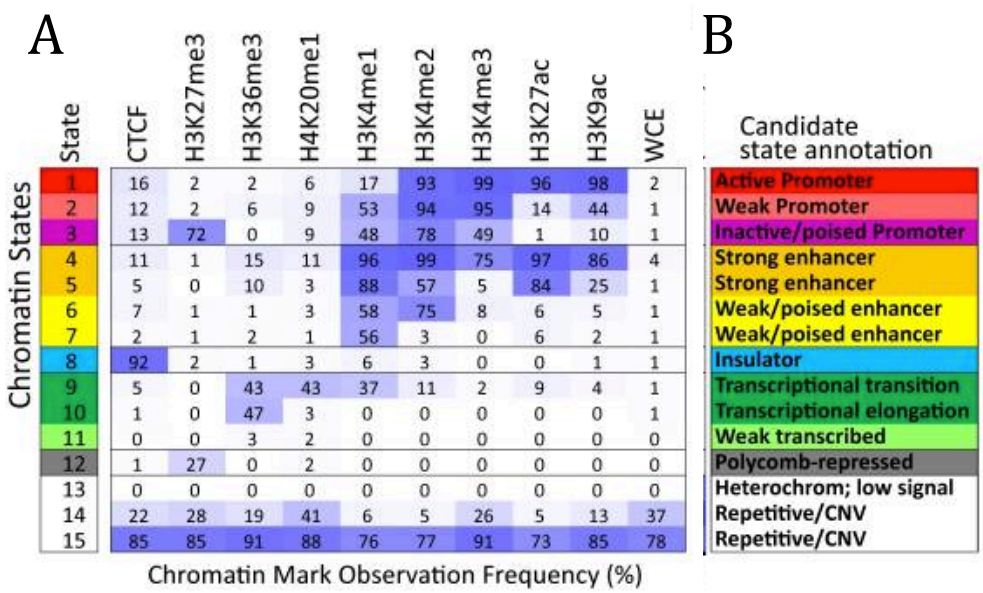


Figure 6. Chromatin states identified by Ernst et al. in nine human cell lines (GM12878, H1-hESC, K562, HepG2, HUVEC, HMEC, HSMM, NHEK and NHLF). Table showing (**A**) the frequency of each chromatin mark (core histones PTMs and CTFC) and (**B**) its candidate annotation. Figure adapted from [27].

They were able to characterize cell-line specific patterns of chromatin states suggesting differences in regulatory elements (enhancers) and functional interactions. Analysing differences in those chromatin states between cell lines new enhancers and gene specific expression patterns could be discovered. As expected, some chromatin states are at same genomic location within all cell-lines studied. Genes expressed in all

cell lines (housekeeping genes, for example) will all have active chromatin states. In contrast, well-known heterochromatic regions such as centromeres, telomeres, and repetitive sequences… will be represented by heterochromatic, repressed states in all cell lines.

Since then, several groups and softwares (such us chromHMM and Segway) have been developed, bringing some light in the complexity of understanding chromatin types/states at a chromosomal level [28, 29].

1.4. Chromatin spatial organization

Another level in the complexity of chromatin organization is its spatial organization in the nucleus. Heterochromatin is preferentially found at the nuclear periphery and surrounding nucleoli (Figure 4), indicating that specific chromatin states might be specifically organized in the nucleus. In addition, DNA fluorescence in situ hybridization (FISH) showed that specific genomic loci are preferentially located within the nucleus.

The study of this specific nuclear organization of heterochromatin lead to the identification of two chromatin domains related to its nuclear spatial organization: lamin-associated domains (LADs) and nucleolus-associated domains (NADs) (Figure 7).

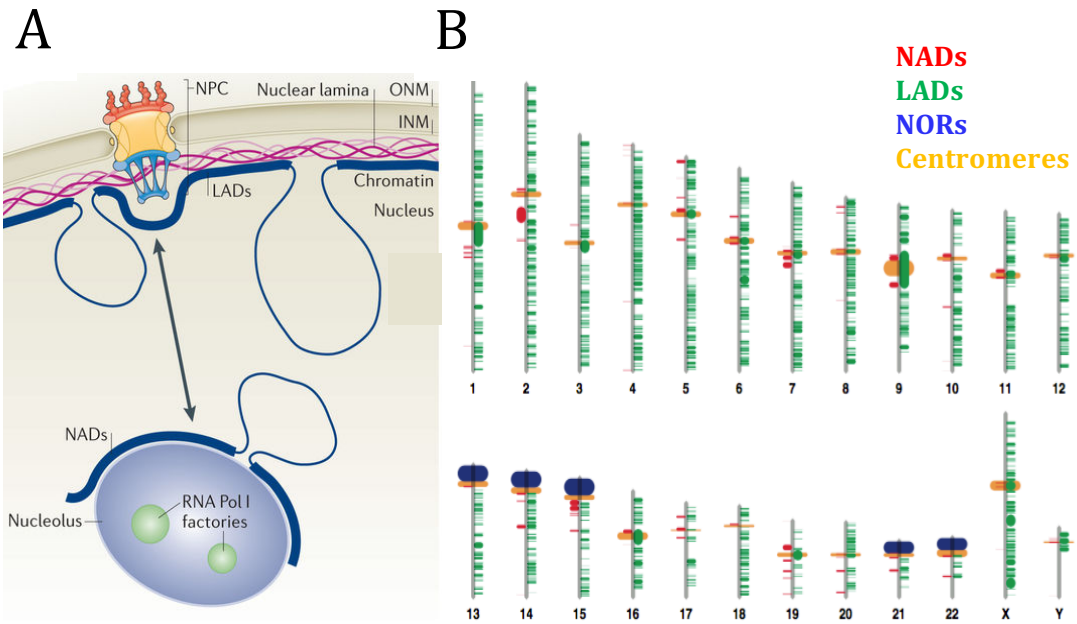


Figure 7. Lamin-associated domains (LADs) and nucleolus associated domains (NADs). (**A**) Spatial nuclear organization. ONM, outer nuclear membrane; INM, inner nuclear membrane; NPC, nuclear pore complex. (**B**) Chromosome location. Figures from [30, 264].

1.4.1. Lamina-associated domains

Nuclear lamina (NL) is a dense fibrillar network of proteins composed of lamins and nuclear lamin-associated membrane proteins. Lamin proteins are implicated in many nuclear processes such as transcription, DNA replication and repair, and cell cycle control. Guelen et al. identified chromatin domains that are in direct contact with lamin B1, named lamina-associated domains (LADs) in human IMR90 fibroblasts [31]. They used they DamID technology in which lamin B1 protein was tagged to the bacterial DNA adenine methyltransferase (Dam), leading to a methylation of adenine in chromatin domains close to lamin B1. Those methylated adenines are then mapped to the reference genome and consequently, DNA-laminB1 contact sites are identified.

LADs cover near 40% of the genome and as expected, they show heterochromatic features such as gene-poor regions, low transcription, high levels of H3K9me2 and H3K9me3, and replicate late during S-phase. They are large domains (0.1-10 Mb) and the insulator protein CTCF and CpG islands demarcate their borders. Interestingly, facultative heterochromatic marks such as H3K27me3 are also found in LADs' borders. LADs are highly abundant domains; mouse and human cells have, on average, 1000-1500 LADs.

As seen for many chromatin features, many LADs are highly conserved between different species but some seem to be cell-type specific. Thus, two types of LADs have been assessed: constitutive LADs (cLADs) and facultative LADs (fLADs). fLADs account for those LADs changing its genomic distribution in different cell types. On the other hand, cLADs are cell-type invariant, A/T rich regions, rich in LINEs (long interspersed nuclear elements) and poor in SINEs (short interspersed nuclear elements) and they are the LADs having the lowest gene content [32, 33].

In fact, long-range hypomethylation (more than 100bp) and focal DNA hypermethylation at CpG islands coincide with LADs in colorectal cancer [34]. Although the association is not yet well understood, it shows the importance of LADs in maintaining chromatin organization and regulating crucial cellular processes.

1.4.2. Nucleolus-associated domains

Nucleolus is the largest structure in the nucleus of all eukaryote cells and where ribosomal DNA (rDNA) transcription, ribosomal RNA (rRNA) and ribosome biogenesis occurs. rDNA is transcribed by RNA polymerase I and encoded in nucleolus organizer regions (NORs) located on the short arms of the acrocentric chromosomes (13, 14, 15, 21 and 22, in humans). These regions are composed of tandem copies of the 45S single transcription unit which will give rise to 18S, 5.8S and 28S rRNA, flanked by a non-transcribed spacer. The repetitiveness of NORs allows the transcription of multiple

copies of rDNA at the same time but they need to be tightly regulated to avoid aberrant recombination events or DNA damage.

Highly transcribed NORs are looped inside nucleoli in contrast to repressed NORs, which stay at the nucleolar periphery silenced in heterochromatic regions. Therefore, as seen in Figure 4, a heterochromatin dense layer is found surrounding nucleoli. In 2010, two different studies in HeLa cervix carcinoma cells and in HT1080 fibrosarcoma cells used high-throughput DNA sequencing and identified those heterochromatic nucleolus-associated domains (NADs) [34, 30]. Importantly, they obtained similar results.

NADs occupy 4% of the human genome (Figure 7B) and contain, as expected, rDNA but they are also enriched in specific gene families and repetitive DNA sequences, such as centromeric and pericentromeric satellites, poorly expressed genes and a high amount of AT-rich sequences. Not surprisingly, NADs are also enriched in repressive core histone PTMs H3K9me3, H4K20me3 and H3K27me3. It is worth mentioning that perinucleolar regions are not exclusively transcriptionally silent, as highly expressed classes of 5S rRNA and transfer RNA (tRNA) are also found in NADs [35].

LADs and NADs share many heterochromatic characteristics such as poor gene-content, low transcription, presence of repetitive DNA sequences, enriched in AT-rich sequence and the same core histone PTMs (mainly H3K9me3 and H3K27me3) and both contain the vast majority of heterochromatic, centromeric and pericentromeric regions (Figure 7B). It is worth saying that constitutive LADs, those having a more extreme heterochromatic state, are more similar to NADs than facultative LADs.

Studies identifying NADs [36, 37] showed an overlap with LADs (Figure 7B) and using a single-cell approach, Kind et al. showed that a considerable proportion of LADs could be found at NADs after mitosis [38, 33]. Thus, DNA can be stochastically found either in LADs or in NADs, showing the potential role of those domains in organizing silent chromatin, although it is not known if a domain-specific silencing mechanism exist.

Importantly, specific DNA sequences preferentially contact each other, thus, another level in global domain organization has been assessed. Since decades, experimental approaches, such as Giemsa staining of interphase nucleus, autoradiographs of a diploid Chinese Cell hamster and immunocytochemical identification of microirradiated DNA and more recently, in situ hybridization experiments evidenced a territorial organization of interphase chromosomes [39].

Consequently, within chromosome territories, contacts happen more frequently than expected by chance and the appearance chromatin conformation capture techniques allowed the identification of those contact genome-wide.

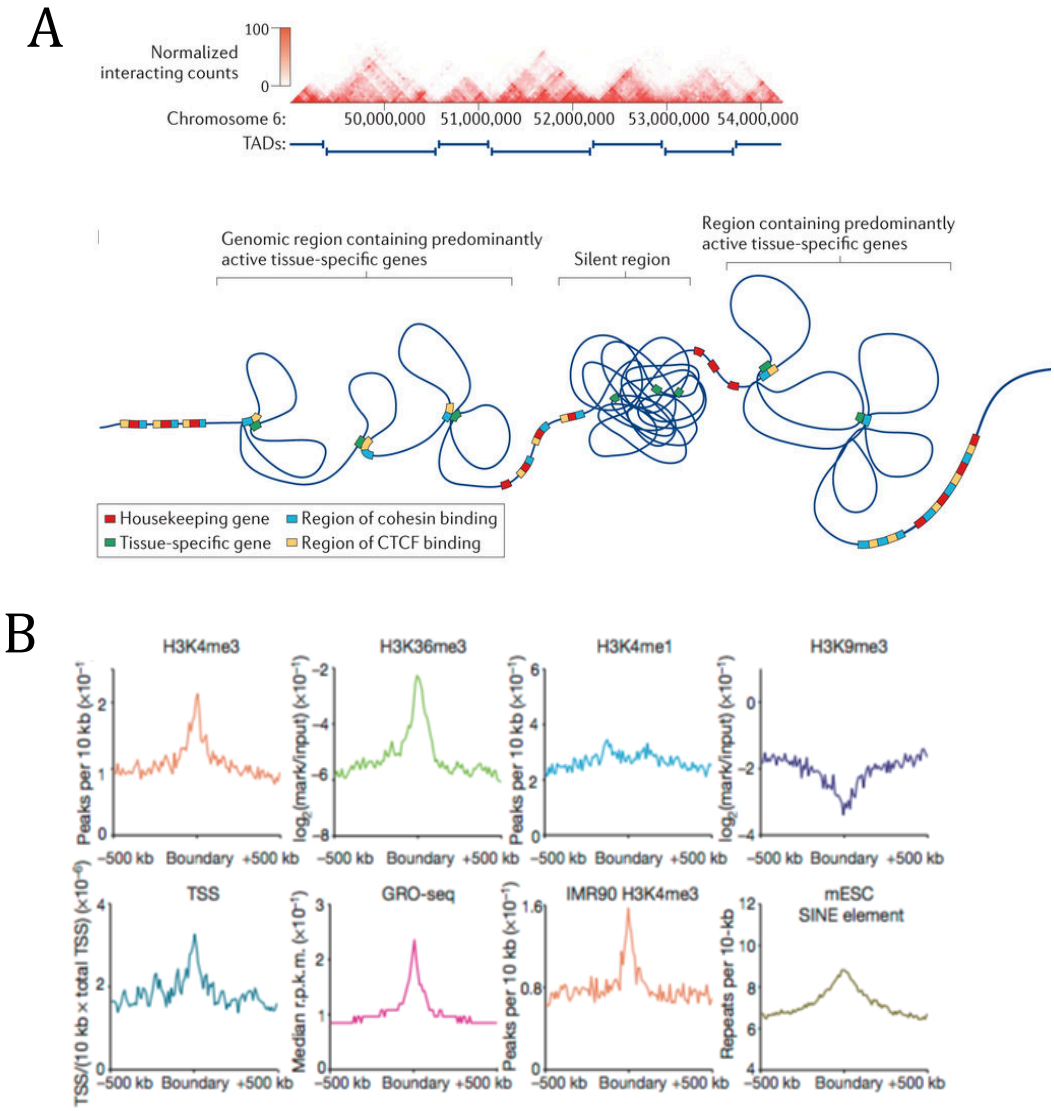


Figure 8. Topologically associated domains (TADs). (**A**) Normalized interacting counts and identified TADs (**B**) TADs boundaries characteristics. Figure adapted from [32, 38].

Dixon et al. firstly identified, using Hi-C, topologically associated domains (TADs) in mouse and human ESC and human IMR90 fibroblasts [38] (Figure 8A). TADs define regions that contact each other (inter- or intra-chromosomal) with the same chromatin environment, well flanked by insulator proteins. Within active TADs active tissue-specific regions are found while constitutively expressed genes (housekeeping genes) are more frequently found in between TADs, in boundaries [25]. TADs boundaries are well defined by CTFC binding and it is observed an enrichment in

repetitive sequences (SINEs) and several active core histone PTMs (H3K4me3, H3K36me3) and depletion of repressive ones (H3K9me3), they also have a high content of transcription start sites (TSS) and accumulation of new mRNA transcripts, assessed by global run-on sequencing (GRO-Seq) (Figure 8B). As expected, repressed TADs are found in the nuclear periphery, seen by a high overlap with LADs.

Enhancers and their regulated promoters are more frequently found within the same TAD and a regulation of TADs' boundaries allow the formation of new enhancer-promoter interactions. Thus, it is not surprising that some TADs are cell type specific and regulated during cellular processes, such as differentiation or stress (reviewed in [40]), while the vast majority are invariant and constitutively found in all cell types. Interestingly, constitutive TADs are also highly conserved across species, as seen between human and mouse [38].

In summary, 3D chromatin organization revealed several domains (LADs, NADs and TADs) that are conserved among species, cell types suggesting the importance of chromatin organization in regulating specific cellular processes such as differentiation, stress… The aberrant genetic program seen in cancer cells might be a consequence of a disorganization of these chromatin domains that need to be tightly regulated, stablished and maintained.

1.5. Chromatin and splicing

As seen, chromatin is involved in regulating important cellular processes such a differentiation, response to stress, DNA damage and repair… Another important process where chromatin has been implicated, and not yet mentioned, is splicing.

It has been found that exonic DNA presents a distinct chromatin landscape than intronic DNA. Genome-wide nucleosome positioning assessed by MNase-Seq showed nucleosomes are not randomly distributed along genes. Instead, exons have a higher nucleosome occupancy than introns, although only when the GC content is higher than its flaking introns, intron-exon junctions also have high nucleosome occupancy and, intriguingly, the average size of a mammalian exons is similar to the 146 bp wrapped around the nucleosome [41]. Therefore, this specific nucleosome positioning may have a protective role in exon definition and splicing regulation. In addition, several core histones PTMs such as H3K36me3, H3K4me3, and H3K27me2… are specifically enriched in exons and some like H3K9me3 depleted [41, 42]. Both, specific core histone PTMs and high nucleosome occupancy at exons do not correlate with the transcriptional status of that gene suggesting a possible splicing regulatory role.

As splicing occurs co-transcriptionally, RNAPII elongation rate has been shown to have an impact on splicing. Fast RNAPII elongation rate leads to a reduction of splicing efficiency and a low exon inclusion in alternative splicing events. Treatment with histone deacetylase (HDAC) inhibitors, which confers to chromatin a less compacted structure, changed alternative splicing events [44].

Therefore, chromatin is also implicated in alternative splicing, which plays critical roles in differentiation, development and disease. Single genes can code for multiple proteins or isoforms in a well-regulated process. Specific tissues, cellular types… have different splicing events depending on cellular needs and this is influenced by the chromatin landscape of that particular alternatively spliced regions. Two non-exclusive models have been proposed (Figure 9). One kinetic model, which states that chromatin structure directly influences RNAPII elongation rate, the final responsible of alternative splicing outcome. Instead, chromatin-adaptor recruitment model states that proteins, which read specifically located core histone PTMs, are the responsible for the different splicing events in a more fine-tuned process (Table 3).

Table 3. Interaction of chromatin and splicing machinery. Table from [43].

Histone modification	Interacting protein	Link to splicing
H3K4me3	Chd1	Chd1 associates with SRp20 (SRSF3) and U2 snRNP (via SF3 subunits) and increases efficiency of pre-mRNA splicing
H3K4me3	Sgf29	Sgf29 interacts with SF3B5 (SF3b10) and SF3B3 (SF3b130) subunits of U2 snRNP
H3K9me3	PTB, hnRNP A1, hnRNP A/B, hnRNP A2/B1, hnRNP K, hnRNP L	PTB and most of the hnRNP proteins are direct regulators of alternative splicing but it is not known whether the association of hnRNP proteins with H3K9me3 affects splicing
H3K9me	HP1 (HP1a)	HP1 binds to Drosophila hnRNP proteins (PEP, DDP1, HRB87F)
H3K36me3	MRG15	MRG15 recruits PTB; tethering of PTB to chromatin changes alternative splicing
H3K79me	TP53BP1	TP53BP1 immunoprecipitates U1 and U2 snRNA (but also other small RNAs)
histone H3 (not phosphorylated at S10)	SRp20 (SRFS3), SF2/ASF (SRSF1)	Both SR proteins participate in constitutive and alternative splicing, but the role of interaction with histone H3 in splicing is not known
DNA methylation	MeCP2	MeCP2 regulates alternative splicing

Several investigations have shown evidences supporting both models. For instance, Luco et al. showed a strong correlation between several core histone PTMs and the splicing factor polypyrimidine tract binding protein (PTB) [46]. Alternatively, spliced exons PTB-dependent were enriched in H3K36me3 and depleted in H3K4me3. The adaptor protein MRG15 specifically binds H3K36me3 and recruits PTB to nascent

mRNA, provoking exon skipping. The explained molecular mechanism does not change RNAPII elongation rate thus, supporting a chromatin-adaptor recruiting model.

However, H3K36me3 positioning also supports the kinetic model. It has been shown that H3K36me3 recruits HDAC1, a histone deacetylase, leading to a more repressive chromatin state and thus, slowing RNAPII elongation rate [76]. Interestingly, the main core histone PTM marking heterochromatin, H3K9me3, is also shown involved in controlling alternative splicing [48]. H3K9me3 reduces RNAPII elongation rate by recruiting heterochromatin protein Cbx3 (HP1γ). In fact, HP1γ is also critical in constitutive splicing of transcripts, whose genes are marked with H3K9me3 [49].

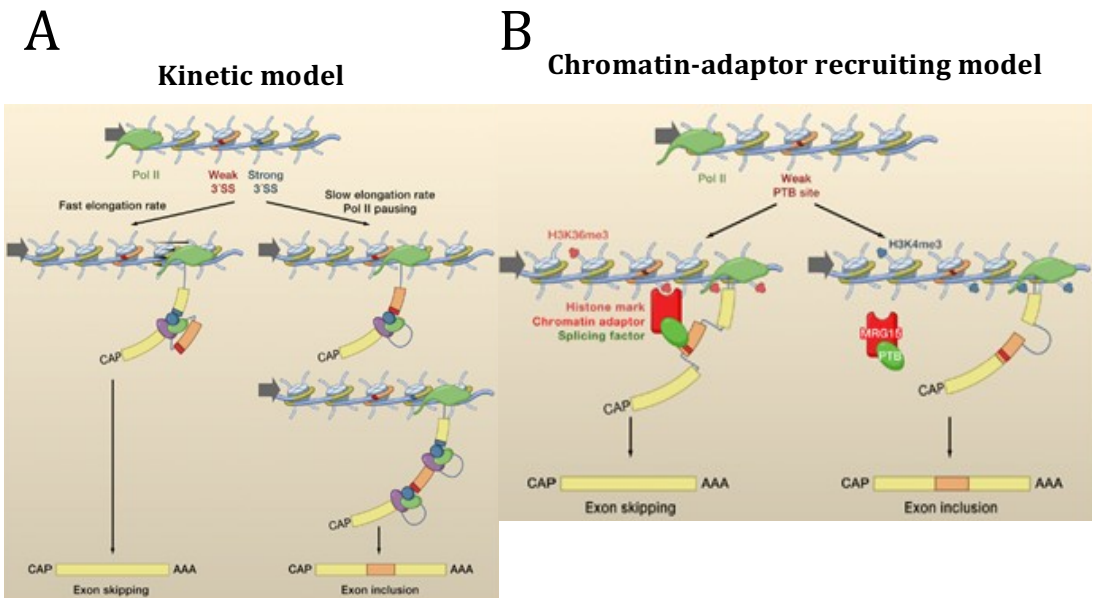


Figure 9. Two non-exclusive models for chromatin role in alternative splicing. (**A**) A kinetic model where RNAPII elongation rate is crucial for splicing decisions and some chromatin landscapes such as nucleosome occupancy directly influence in this process. (**B**) A chromatin-adaptor recruiting model where chromatin binding-proteins are essential for splicing decisions. Figure taken from [45].

As explained, it is clear that chromatin influences splicing outcomes however, some evidences have shown that splicing components are also regulating chromatin structure by establishing or maintaining epigenetics marks such as H3K36me3 [50, 51]. Genome-wide analyses have shown that splicing enhances H3K36 methyltransferase HYPB/Setd2 recruitment to the elongating RNAPII and to H3K36me3-associated proteins [50]. In addition, inhibition of expression of the splicing factor SAP130 leads to a reduction of H3K36me3 levels and less recruitment of HYPB/Setd2.

1.6. Heterochromatin and repetitive elements

The last human genome estimation showed that 60-70% of the genome is repetitive and only a 1,5% is protein-encoding. Repetitive elements are from different types (Table 4), being the simplest microsatellites, dinucleotides. Instead, transposons are the more complex class, remnants of integrated RNA and DNA viruses, which have the ability to copy and paste themselves in a different location; in addition, to encode for non-functional viral proteins or intermediates. Some of them also have a specific chromosome location such as tandem repeats in centromeres or telomeres but, all of them, coincide with heterochromatic regions (Figure 10, 11A) [52].

Table 4. Classification of human repetitive elements. In humans, LTR-retrotransposons are called human ERVs (HERVs). Adapted from [58, 61].

Tandem repeats	Satellite DNA	Centromeric and pericentromeric (200 bp) Minisatellites (20-35 bp, with a conserved core sequence of 10-15 bp) Microsatellites (2-5 bp)	
	Telomeres (6 bp)		
Transposons	Class I **Retrotransposons**	Non-LTR retrotransposons (RNA intermediate)	Long Interspersed Nuclear Elements (**LINEs**) Short Interspersed Nuclear Elements (**SINEs**) As an example: Alu Composite SINEs As an example: SINE-VNTR-Alu (**SVA**), human specific, fusion SINE-R and Alu separated by variable number of tandem repeats (VNTR)
		LTR-retrotransposons (contain envelope proteins)	Endogenous Retroviruses (**ERVs**) Ty1 copia Ty3 gypsy …
	Class II **DNA transposons**		

Figure 10. Human repetitive elements chromosome distribution, which greatly coincides with heterochromatin. CT, Centromere; PCT, pericentromere. Figure adapted from [52].

1.6.1. Tandem repeats

Tandem repeats have a structural role in centromeres and telomeres and a proper control of those elements is crucial to avoid interchromosomal rearrangements. In addition, tandem repeats are highly repetitive, meaning that many copies are found one next to the other. In contrast, transposable elements account for the vast majority of intrachromosomal rearrangements and are less frequently found adjacent [53, 54].

In most metazoans, telomeres are constituted by a short DNA repeat (5'-TTAGGG-3'), which is highly conserved and chromatin at those regions is enriched in H3K9me3. Telomeres are bound by highly conserved protein complexes, which protect chromosomal ends from being recognized by dsDNA breaks repair machineries. The high conservation across evolution of the DNA motif and bound-proteins, suggests that chromatin might play a mild role in this context [52].

In contrast to telomeres, centromeres also contain repetitive elements, although less conserved. In humans, repetitive alpha satellite DNA defines the centromeres regions and they appear to be neither necessary nor sufficient for centromere function. Although surprisingly, centromeric satellites have a monomer length of 170 or 340 bp, pointing to the wrap of one or two nucleosomes, respectively [57]. Besides, a specific centromeric H3 variants (CENP-A) has been identified with important structural and functional centromeric roles, showing that centromeres are epigenetically regulated. In addition, CENP-A containing-nucleosomes are found interspersed with canonical nucleosomes with active PTMs in contrast to chromatin

surrounding centromeres (pericentromere), which contains high levels of repressive core histone PTMs and DNA methylation (Figure 11B) [55, 56].

Pericentromeric heterochromatic regions, contains also a high number of repetitive elements, such satellites I, II and III in human (Figure 11A). Those repetitive sequences are the less poorly conserved, even between chromosomes of the same species. This suggests an epigenetic regulation that do not rely in protein machineries specifically recognizing a sequence motif. Interestingly, maybe because of technical issues when working with so compacted DNA, the importance of pericentromeric regions is unclear in metazoans and their presence may not confer any advantage [52]. Nevertheless, pericentromeric heterochromatin needs to be established and maintained, as a dysregulation leads to genomic instability, seen in various diseases, such as cancer.

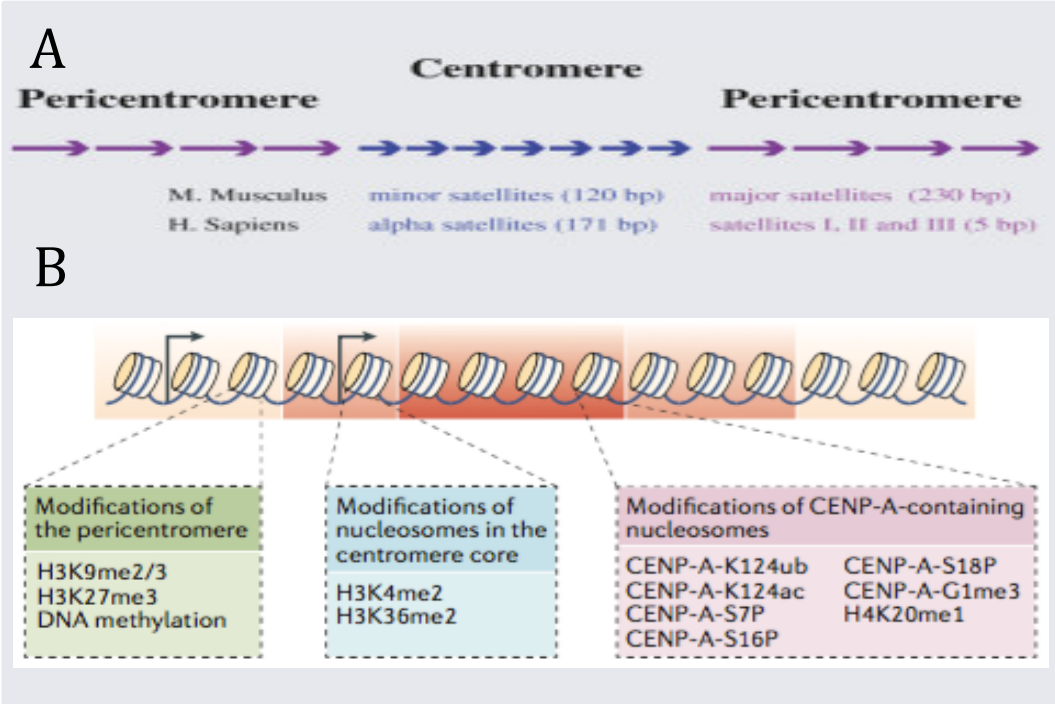


Figure 11. Centromeres and pericentromere regions (**A**) Satellite distribution. (**B**) Chromatin features such as core histone PTMs, histone variants and DNA methylation. Figure adapted from [52, 56].

Interestingly, 97-98% of the human genome is transcriptionally active and produces stable RNA. The majority of those RNAs are non-coding RNA (ncRNA), which have been shown to have a function in heterochromatin and centromere formation, among other process such as gene expression regulation. Centromeric and pericentromeric regions have been found transcriptionally active in several organisms, including human [58]. It was recently assessed that those transcripts play a critical role

in the establishment and function of those regions. Indeed, defective transcription of those repetitive sequences leads to aberrant chromosome segregation during cell divisions in yeast, HeLa cells and human artificial chromosomes [58].

Maida et al. described a non-telomeric function for human telomerase reverse transcriptase (hTERT) at mitosis. hTERT forms a complex together with Brahma-related gene 1 (BRG1) and nucleostemin (NS) (TBN complex), which produces double-stranded RNAs homologous to centromeric alpha satellites and transposons that, when processed with the interference RNA (iRNA) machinery, are targeted to those sites (Figure12A). Binding of the TBN complex promotes heterochromatin assembly at particular sites and allows a proper mitotic progression, by exerting its functions in centromeres. In fact, a rather similar mechanism is also found in yeast [59].

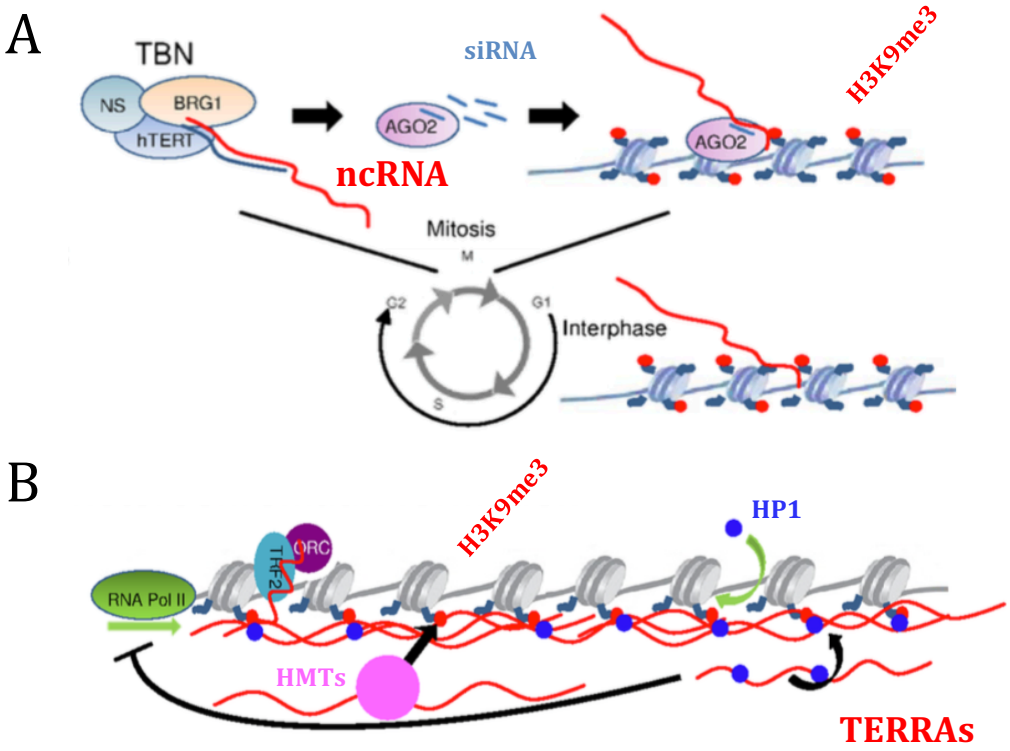


Figure 12. Repetitive element transcription and heterochromatin formation. (**A**) (Peri)centromeric repetitive regions and transposons are being transcribed, by a non-telomeric function of human telomerase reverse transcriptase (hTERT) to non-coding RNA, prompting H3K9 methylation in a AGO1-dependent manner. BRG1: Brahma-related gene 1; NS: nucleostemin. (**B**) Telomere heterochromatin formation through telomeric repeat-containing RNA (TERRAs) transcription. HMTs: histone methyltransferases; siRNA: small interfering RNA. Figure adapted from [58].

In addition to heterochromatin formation, the active transcription found at centromeres has also been related to the CENP-A loading. In fact, in humans, a centromeric 1.3kb lncRNA has been found to directly interact with CENP-A and its

chaperone and is required for its targeting to centromere [60]. Other transcripts encoded in centromeric regions have been found to mediate mitotic Aurora B activity and the localization of kinetochore proteins, such as CENP-C [58].

At telomeres, repetitive elements transcription is also observed *in vivo* and has also a link in heterochromatin formation at those regions, similar to the ncRNA originated at centromere. Specifically, lncRNAs named telomeric repeat-containing RNA (TERRA), transcribed by RNA polymerase II, are an integral component of telomeric heterochromatin. TERRAs are mainly composed of subtelomeric sequences and C-rich telomeric repeats and their length can vary from 100bp up to 9kb. They are conserved across evolution, found from yeast to human. Their expression is cell-cycle dependent, showing at G1-phase the highest level and at late S/G2 the lowest. TERRAs act as a scaffold in recruiting HP1 and other chromatin remodelling factors, leading to high levels of H3K9me3 and heterochromatin formation (Figure12B). TERRAs provoke its own repression, preventing heterochromatin hyperformation and telomere expansion [58].

1.6.2. Transposable elements

Transposable elements (TEs) are ubiquitous in eukaryotic genomes and differ from tandem DNA, having a more dispersed genomic distribution and their copies are less frequently found adjacent. Instead, all repetitive elements share a common heterochromatic silencing (Figure 10).

Transposable elements have the ability to replicate and change its genomic position, independently of the host chromosomes replication but relying in host machinery. Some of this TEs are able to produce viral proteins or viral intermediates but none produce fully viable virions. Besides, TEs can lead to deleterious effect if integrated in a host protein-coding genes. Moreover, their integration leads to chromatin changes in *cis* and *trans* that can induce or repress gene transcription of neighbouring genes (Figure 13A). Indeed, it is known that an uncontrolled transcription of TEs can lead to high rate of mutations and chromosomal rearrangements provoking disruptive and deleterious effects.

There are two major classes of TEs: retrotransposons which generally function via reverse transcriptase (RT) and DNA transposons which encode for protein transposase (TPase). TEs harbour regulatory sequences such as RNAPII and III promoters and polyadenylation signals (PASs) and those regulatory elements and protein-coding regions differ between TEs classes (Figure 13B).

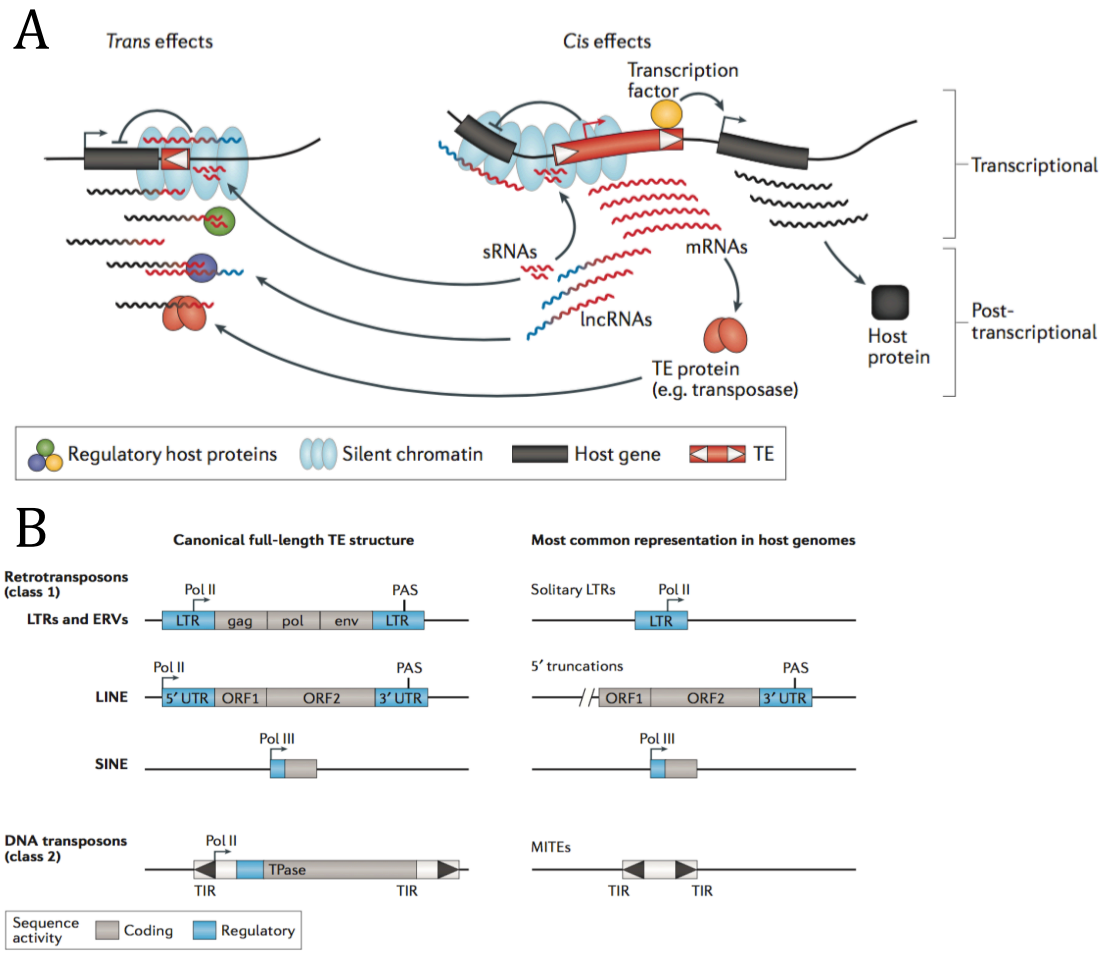


Figure 13. Transposable elements. (**A**) Different types of regulatory activities exerted by TEs. (**B**) Schematic of major TEs classes and their typical genetic organization Figure adapted from [62].

Multiple mechanisms keep those repetitive elements silent, being H3K9me3 the first line of defence, in both somatic and germline. TEs families with potentially high transcription rates, contain a higher H3K9me3 content and increased DNA methylation. In addition, others contain H3K27me3; although, usually found with H3K9me3. In fact, depletion of H3K9me3 methyltransferase (SETDB1) in mammals, allows the transcriptions of repetitive elements in germline and triggers transposition [54]. DNA methylation is extensively modulated during gametogenesis and embryogenesis and most TEs become hypomethylated. Walter et al. showed that H3K9me3 and, most importantly H3K27me3 chromatin can ensure the control of TEs in periods of intense changes of DNA methylation levels [63].

TEs expression needs to be tightly regulated at chromatin level to avoid transposition activity, pathogenic TE-encoded RNA or proteins and nucleic acid toxicity

due to high rate of transcriptions which prompts an innate immune response (See *3. Interferon response and chromatin*).

Despite huge efforts in silencing TEs by DNA methylation and/or heterochromatin establishment and deleterious effect observed when not properly controlled, several evidences are pointing to a regulatory role of TEs [61]. There are cases where TEs function as a tissue-specific promoter modulation transcription of key genes defining that tissue. In humans, erythroid and IFN-inducible enhancers are ERV9 and MER41 TEs, respectively and MER39 promoter is controlling prolactin levels in endometrium [254, 64, 65]. Similar promoter and enhancer function have been attributed to TEs in other organisms and even an insulator function for mice B2 SINE retrotransposon has been described (Introduction, Figure 8) [62]. It is thus nowadays hypothesised a dynamic evolution of TEs regulatory activities.

Interestingly, a role for ERVs in regulating early mammalian development is gaining strength. In two-cell mice embryos, ERVs are from the first sequences to be transcribed during zygotic genome activation [66]. In humans, expression of distinct families of primate-specific ERVs fluctuate through development until cell completely differentiate and they contain OCT4 and NANOG binding sites at their long terminal repeat (LTR) promoter [62].

The activity of primate-specific HERV-H is highly correlated with pluripotent ESCs and depletion of lncRNA derived from HERV-H enhances a rapid cellular differentiation [67, 68, 271]. Additionally, heterochromatin is very distinct in ESCs where a high expression of repetitive elements is observed without deleterious effects [69]. ESCs chromatin is widely more accessible and a distinct morphology has been observed, characterized by fewer but larger heterochromatic domains [69, 70]. In pluripotent ESCs key structural chromatin components such as HP1, linker histone H1 and core histones display a more hyperdynamic binding, which is not seen in lineage-committed but undifferentiated cells. ESCs have unique core histone PTMs patterns known as divalent domains, extensive genome regions containing together H3K27me3 and H3K4me3. This distinct epigenetic landscape and high expression of repetitive elements without deleterious effects in ESCs is not well understood; indeed, reprogramming to pluripotency leads to a loss of transcriptional control of ERVs [71].

In summary, repetitive elements trigger a challenge at cellular level, as they need to be tightly controlled because an aberrant expression can be easily prompted, due to its repetitive nature. Thus, repetitive elements are frequently found at heterochromatin, where the heaviest packaged and repressed DNA is located. Indeed,

a proper epigenetic control of repetitive elements is crucial to avoid severe diseases, such as cancer or immunodeficiencies where high mutations and chromosome rearrangements rates are seen, maybe due to its uncontrolled expression.

However, active transcription of repetitive elements has been observed in normal conditions, which trigger heterochromatin establishment and maintenance, forming a functional and regulated loop mechanism. For instance, its expression and presence also have structural roles in specific genomic regions like centromeres or telomeres. Besides, transposable elements have specific-regulatory roles, define pluripotent ESC state and are regulated through development and differentiation. Therefore, a proper epigenetic control of repetitive elements, mainly through heterochromatin and DNA methylation is crucial to avoid centromere/telomere disorganization and uncontrolled activity of transposable elements both indispensable for a right cell functioning.

2. LINKER HISTONE H1

As mentioned, linker histone H1 binds to the entry/exit site of DNA on the surface of nucleosome allowing the folding and stabilization of nucleosomes, conferring to chromatin a more compacted state. Unlike core histones, studying histone H1 has been challenging mainly due to the lack of specific antibodies recognizing each family member. However, in the last years, the classical view of linker histone H1 as a merely structural component of chromatin has changed and linker histone H1 has emerged as an important regulator of specific processes, as happens with core histones.

2.1. Linker histone H1 family and structure

Unlike core histones, which are highly conserved in evolution, linker histone H1 is more divergent between species and, due to gene duplication events, several variants or subtypes exists, eleven in humans or mice (Table 5).

Table 5. Linker histone H1 variants in human, gene and expression. Mouse H1 variants names are shown in brackets

Somatic variants			**Germ-line variants**		
Name	*Gene*	*Expression*	*Name*	*Gene*	*Expression*
Replication-dependent			H1oo (H1oo)	H1FOO	Oocytes
H1.1 (H1a)	HIST1H1A	Tissue-specific	H1t (H1t)	HIST1H1T	Testis
H1.2 (H1c)	HIST1H1C	Ubiquitous	H1T2 (H1t2)	H1FNT	
H1.3 (H1d)	HIST1H1D		HIL12 (TISP64)	HILS1	
H1.4 (H1e)	HIST1H1E				
H1.5 (H1b)	HIST1H1B				
Replication-independent					
H1.0 (H1(0))	H1F0	Differentiated cells			
H1X (H1X)	H1FX				

In humans, somatic H1 variants are H1.1 to H1.5, H1.0 and H1X, being H1.0 and H1X expressed in a replication-independent manner. Replication-dependent variants (H1.1 to H1.5) have paired expression with DNA replication and core histone expression, thus at S-phase of the cell cycle. Interestingly, H1.1 is restricted to certain tissues and H1.0 accumulates in terminally differentiated cells while the rest are ubiquitously expressed. Germ-line specific variants are H1t, H1T2 and HILS1 in testis and H1oo in oocytes (Table 5) [79-81].

In addition to being expressed in a replication-independent manner, H1X and H1.0 differ from the other variants in its gene location and mRNA. H1.1 to H1.5-encoding genes are clustered in chromosome 6 (6p21-p22), named cluster *HIST1*, together with core histone genes. Their mRNAs tails are not polyadenylated but they are still rapidly translated due to a 3' stem-loop sequence. Instead, single intronless genes located in chromosome 3 and 22 encode H1X and H1.0, respectively. In contrast to somatic replication-dependent H1 variants, their mRNAs are polyadenylated.

It is well established that H1 variants expression pattern differs in cell-types and under specific processes such as differentiation and development. In diseases such as cancer, an abnormal expression of H1 variants is observed thus, a proper H1 stoichiometry is crucial. Interestingly, although H1 levels are tightly controlled, cells try to compensate the lack of one or several variants by up-regulating others (mainly H1.0), indicating that H1 variants also have redundant functions. All this indicates that the *HIST1* gene cluster is tightly regulated but how H1 genes are specifically expressed and regulated is not yet well understood.

All linker histone H1 variants have a shared common structure composed of a well-conserved central globular domain, a long C-terminal tail of approximately 100 aminoacids and a short N-terminal tail.

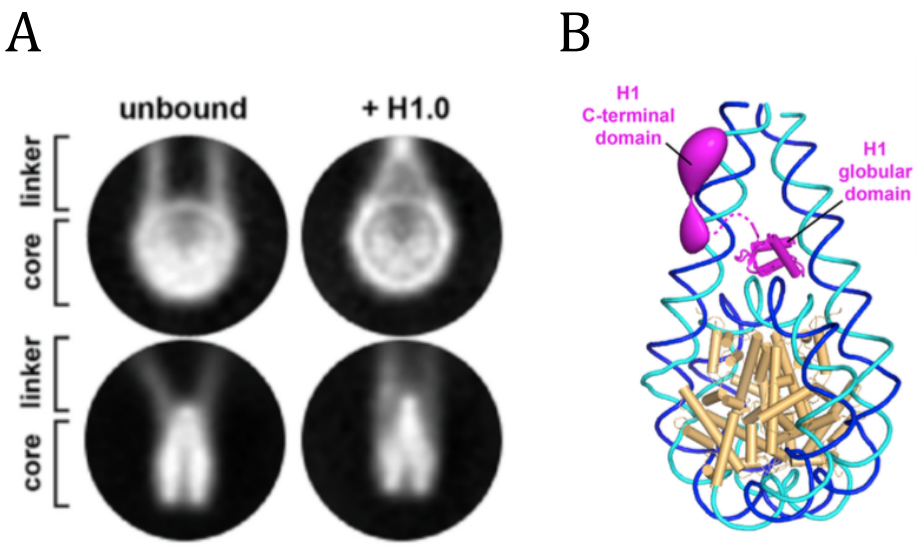


Figure 14. Nucleosome core particle bound to linker histone H1 (**A**) Cryo-electron microscopy images of nucleosomes unbound or bound with linker histone H1.0 (**B**) Structure of linker histone H1 bound to nucleosome. Figure adapted from [48, 50].

The globular domain consists of a 3-helix "winged helix" fold with a classical helix-turn-helix motif; it is responsible of DNA binding and highly conserved. The exact

location of the binding site is unknown however, it is generally accepted that it contacts at least 2 strands of DNA near the nucleosome dyad, stabilizing DNA wrapping and giving a more compact structure (Figure 14) [50].

Both C- and N-terminal tails are enriched in basic aminoacids (mainly lysines) conferring to H1 a high density of positively charged residues, which allows the stabilization of the negatively charged DNA. Both domains are intrinsically disordered, meaning they lack a fixed or ordered structure until they are bound and then, stabilized [51].

It is in the C-terminal domain where a high variability in sequence between variants (Figure 15) and species is seen [51, 52]. It is then, not surprising that it modulates H1 affinity to chromatin and confers to H1 variants different chromatin affinities (See *2.4.3. Chromatin binding affinity of histone H1 variants*). The C-terminal domain protrudes without an ordered structure until it binds DNA or other proteins, which maintains and stabilizes the DNA interaction much more than the N-terminal tail (Figure 14) [53-56] and it might be responsible for the specificities of H1 variants.

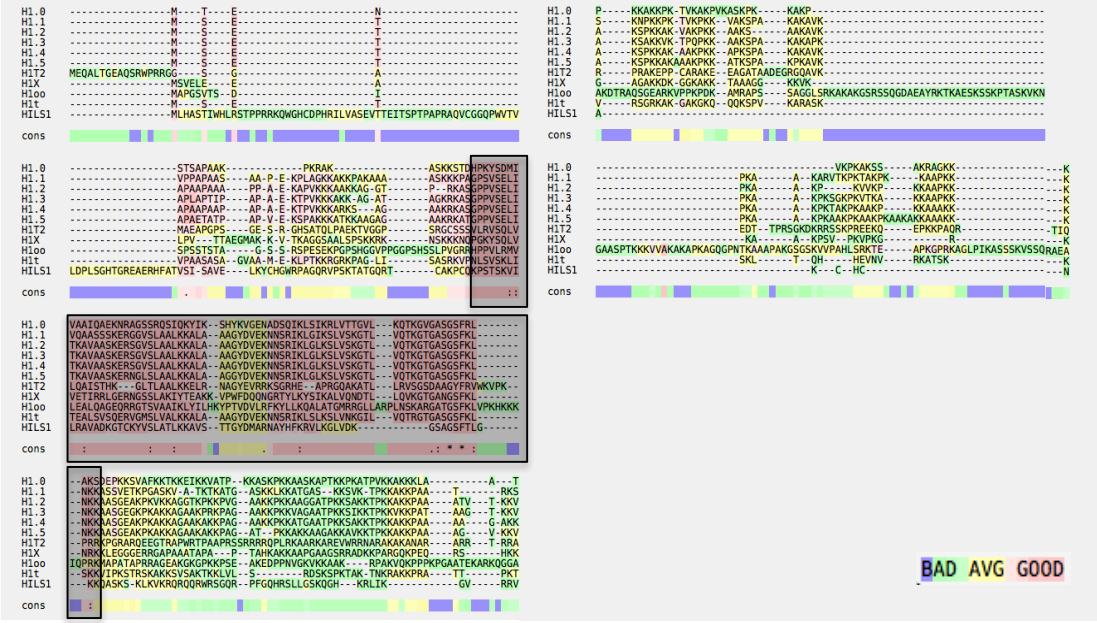


Figure 15. Aminoacid sequence alignment of human linker histone H1 variants. The conserved globular domain is shown in black shadow. Alignment was performed using M-coffee software with default parameters [49].

2.2. Linker histone H1 function

Classically, linker histone H1 has been seen as a mere structural component of chromatin by stabilizing nucleosomes and condensing higher-order structures. Early work comes from *in vitro* systems comparing H1-containing and H1-delpeted

chromatin, knowing that the dissociation of H1 from chromatin at a lower ionic strength is lower than that of core histones. Loss of H1 leads to chromatin decondensation prompting a more “beads on a string” chromatin structure [72] and it also influences nucleosome repeat length (NRL) [73]. In agreement, recent studies combining super-resolution nanoscopy with computer simulations allowed visualizing and counting of nucleosomes along chromatin in a single nucleus [70]. They showed that nucleosomes are assembled in heterogeneous groups of varying sizes (‘clutches’) and as expected, transcriptionally active chromatin associated with RNAPII is found in small ‘clutches’. Largest ‘clutches’ are enriched in linker histone H1 and heterochromatin. As expected, knowing that H1 is a condenser and stabilizer of chromatin, H1 was also found to be a repressor of transcription [74, 75], an inhibitor of nucleosome sliding [76] and a modulator of the remodelling chromatin factor SWI/SNF [77-79]. Further experiments showed that transcriptionally active chromatin is typically depleted in H1 and the presence of H1 in promoter regions impairs genic transcriptional activity [80-86].

However, it is known that histone H1 is highly dynamic and mobile, as fluorescence recovery after photobleaching (FRAP) studies with H1 fused to green fluorescent protein (GFP) revealed residency times of 3-4min, while core histones stay hours [87, 88]. And, some positive roles in transcription have also been observed [89], as depleting H1 leads to an up- but also to a down-regulation of genes in several organisms having a single H1 variant like *Tetrahymena thermophila* and yeast [90, 91] but also in humans, which have eleven H1 variants [92].

In addition to have a direct role in compacting chromatin structure by modulating nucleosome particle structure, linker histone H1 influences transcription and chromatin structure by competing or directly interacting with other chromatin regulators or transcription factors (Figure 16). High-mobility group (HMG) proteins were shown to compete with H1 for chromatin binding sites changing chromatin condensation and access of regulatory factors [93]. Pioneer transcription factors like FoxA, crucial in cell differentiation, change the epigenetic landscape by displacing linker histone H1 from enhancers [94]. Similar competition mechanisms with H1 are described for other transcription factors and enzymes like poly ADP-ribose polymerase 1 (PARP1) or methyl-CpG-binding protein (MeCP2) [2, 95] and even with other core histone variants like H3.3 [96].

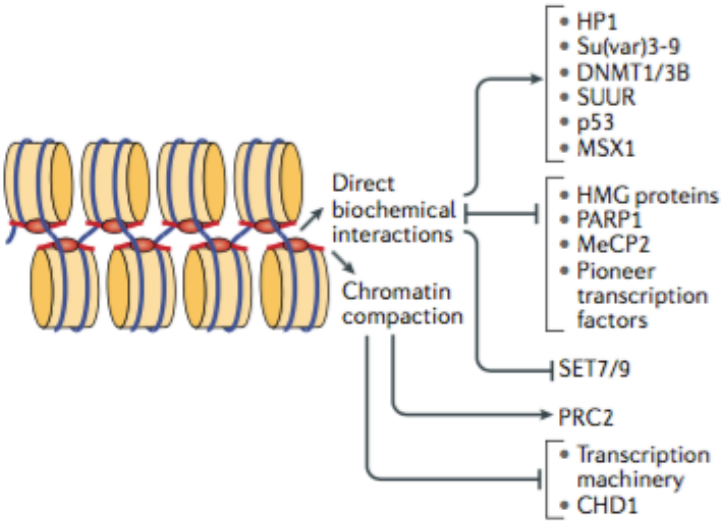


Figure 16: Alternative mechanisms used by linker histone H1 to modulate the activity of chromatin. Figure from [2].

In contrast, other proteins directly interact with linker histone H1 to regulate target gene transcription or chromatin compaction. That is the case of the tumour suppressor enzyme, phosphatase and tensin homolog (PTEN), which displace H1 from target genes activating transcription [97]. The opposite happens with template activating factor-I (TAF-I, also known as SET), which represses interferon-stimulated genes (ISGs), by recruiting H1 to their promoters [98].

In *Drosophila melanogaster*, H1 physically interacts and recruits the H3K9 methyltransferase Su(var)3-9 and the SNF2-like ATPase (SUUR) leading in this case to chromatin condensation in larva polythene chromosomes [99, 100]. In this organism, histone H1 also serves as a reservoir for STAT (signal transducer and activator transcription) by directly interacting with it. As a result, a new regulatory role for *Drosophila* histone H1 in the JAK-STAT signalling has been proposed [101].

Other interactions have been found to be H1 variant-specific such as H1.5 with the myogenesis transcription repressor MSX1 or mice H1 variants with DNA methyltransferases DNMT1 and DNMT3B or even more complex, different variant-specific PTMs like H1.4K26me with HP1 or H1.2 not phosphorylated at T146 with p53, among other (See *2.4. Specificities of histone H1 variants*).

Another important proposed role for H1 in chromatin is modulation of the three-dimensional genome architecture. Geeven et al. performed high-throughput chromatin conformation capture (Hi-C) analysis in H1 triple-knockout (H1c, H1d and H1e; H1.2, H1.3 and H1.4 in human, respectively) mouse ESCs [102]. They found that depletion of H1 changes the epigenetic signature genome-wide but mainly at gene-

dense topological-associated domains (TADs). New DNase hypersensitive sites are established and increased levels of H3K4me1 and H3K4me3 (a mark of potential enhancers and active promoters, respectively) are observed upon loss of histone H1. Interestingly, no changes were found in the "repressive" core histone PTMs specifically, H3K9me3 and H3K27me3. In addition, gene-dense TADs lose DNA methylation at enhancer regions and only few methylation gains are observed genome-wide. Instead, CpG-rich promoter maintain their methylation status in H1 TKO.

Surprisingly, they showed that reduced amounts of H1 does not have a major effect in overall genome organization. However, the frequency of long-range inter-TAD interaction increase, specifically in those TADs more extensively epigenetically modified upon H1 depletion.

Although the vast proportion of linker histone H1 is located inside the nucleus, certain H1 variants have been recently identified in the cytoplasm and on cell membranes, pointing to extranuclear and extracellular functions [103, 104] (Figure 17).

Interestingly, H1.2 variant has been associated with the process of apoptosis. Upon DNA double-stranded breaks (DSBs), in a p53-dependent manner, H1.2 is released to the cytoplasm prompting the secretion of cytochrome c from mitochondria, which eventually triggers apoptosis [105]. This study suggests a new role for linker histone H1.2 in apoptotic signalling from the nucleus to the cytoplasm.

Other reports showed a linker histone H1 role in innate immune response. Granules of macrophages, named murine microbicidal protein (MUMP) have been shown to contain linker histone H1 [106] and acts as a pattern recognition receptor in non-specific cytotoxic cells of catfish [107] (Figure 17).

Macrophages in the liver, named Kupffer cells express linker histone on the cell membrane, which binds to thyroglobulin mediating its clearance [108]. Epithelial cells of intestinal villus also express linker histone in the tip, interacting with *Escherichia coli* and specifically with a type of virus, Norovirus [109-111].

Strikingly, linker histone H1 is also released extracellularly. Upon infection, neutrophils establish the so-called neutrophil extracellular traps (NETs), networks of extracellular fibers, mainly composed of DNA and antimicrobial agents, including linker histone H1, core histones and elastase. Indeed, the major component in NETs is linker histone H1. The main function of NETs is to trap pathogens and to protect surrounding cells from proteins released by neutrophils. Interestingly, fragments of histone H1 interact bacterial membranes, disrupting and increasing its permeability, achieving easily pathogen death [106, 112, 113]. Linker histone H1 also directly interacts with C

reaction protein (CBP) extracellularly, suggesting a role in CBP-dependent microbe clearance [114].

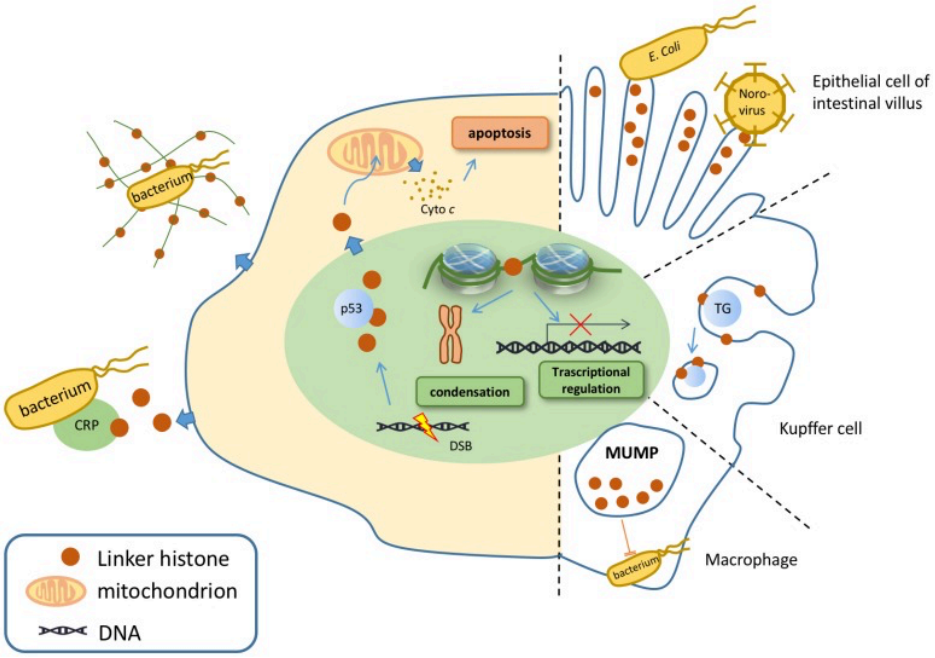


Figure 17. Extranuclear and extracellular functions of linker histone H1. Upon DSBs, H1.2 is secreted to the cytoplasm triggering apoptosis. Macrophages contain linker histones granules that show antimicrobial activity. Extracellular H1 forms neutrophil extracellular traps or binds with c-reaction protein to kill pathogens. On the cell membrane of Kupffer cells, binds to tyroglobulins, internalizing and clearing them. Figure adapted from [104].

What still is a matter of debate in the field is the specificity of H1 variants, as knockout (KO) of single or double H1 variants in mice [115] has no apparent phenotype. Triple KO mice of H1 variants (H1c, H1d and H1e; human H1.2, H1.3 and H1.4, respectively) are embryonically lethal, suggesting that the total H1 content is crucial and H1 variants may have redundant roles. Despite this observation, new evidences have pointed to a specific function and distribution of linker histone H1 variants (See *2.4. Specificities of histone H1 variants*).

2.3. Histone H1 post-translational modifications

As mentioned, histone H1 is also post-translational modified (Table 6). As happens with core histones PTMs, H1 PTMs are thought to regulate chromatin states by modulating its structure and interacting proteins.

Table 6. List of post-translational modifications on the most common histone H1 variants, as identified by mass spectrometry. Phosphorylation sites in bold are consensus CDK sites ((S/T)-P-X-(K/R), where X is any aminoacid).
[a] Denotes N-α-acetylation of the N-terminal residue after methionine removal.
Table adapted from [128].

H1 variant	Phosphorylation	Acetylation	Methylation	Ubiquitination	Formylation
H1.2	S2, T4, **T31**, S36, **T146**, **T154**, T165, **S173**	S2[a], K17, K34, K46, K52, K63, K64, K85, K90, K97, K169, K192	K34, K52, K64, K97, K106, K119, K168, K187	K46, K64, K75, K85, K90, K97, K106	K17, K34, K46, K63, K64, K75, K85, K90, K97, K160
H1.3	T4, **T18**, S37, **T147**, **T155**, T180, **S189**	S2[a], K17, K34, K46, K52, K63, K64, K85, K90, K97, K169	K52, K64, K97, K106, K169	K47, K65, K76, K86, K91, K98, K107	K34, K46, K63, K64, K75, K85, K90, K97, K141, K160
H1.4	S2, T4, **T18**, S27, S36, S41, T142, **T146**, **T154**, **S172**, **S187**	S2a, K17, K26, K34, K46, K52, K63, K64, K85, K90, K97, K169	K26, K52, K64, K97, K106, K119, K148, K169	K17, K21, K34, K46, K64, K75, K85, K90, K97, K106	K17, K34, K46, K63, K64, K75, K85, K90, K97, K110, K140, K160
H1.5	S2, T4, T11, **S18**, T39, S44, S107, **T138**, **T155**, **S173**, **T187**, S189	S2[a], K17, K49, K88, K93, K109, K168, K209	K27, K168, K169		K67, K85, K88

By far, **H1 phosphorylation** is the most well characterized H1 PTM, firstly described in 1972 [129]. Histone H1 tails, especially the C-terminal, contains several (S/T)-P-X-(K/R) motifs, which are recognized by cyclin-dependent kinases (CDK) and then, phosphorylated at serine and threonine residues [130, 131]. Phosphorylation levels are lowest during G1 phase of the cell cycle, increase during S phase and reach a maximum at late G2 and mitosis, which sharply decrease in telophase.

In mitosis, CDK1/CycB is the main responsible of H1 phosphorylation, although other kinases have been found implicated. For instance, H1.2S27, H1.4S35 and H1.5T10 are phosphorylated by Aurora B kinase, protein kinase A (PKA) and glycogen synthase kinase-3, respectively. H1 phosphorylation during G2/M phases of the cell cycle promotes mitotic chromosome condensation [132-134]. In contrast, H1 phosphorylation during S-phase favours chromatin decondensation needed for the progression of the replication fork [135]. H1 phosphorylation has seen increased in bladder cancer, not surprising in cells that lack a proper proliferation control [136].

As mentioned, H1 phosphorylation favours DNA decondensation and consequently, several publications confirmed the link with active transcription. Vincent

et al. showed that the induction of the mouse mammary tumour virus (MMTV) promoter after progesterone treatment needs the removal of H1 [137-139]. This is ultimately controlled by the CDK2/CycA kinase, which phosphorylates H1 leading to its removal and thus, facilitating the access of the transcription machinery to the MMTV promoter. Another link with active transcription was found at rRNA promoters, pointing to a role of H1 also in RNA polymerase I (RNAPI) transcription. Zheng et al. [140] identified H1.2S173, H1.3S172 and H1.4S187 phosphorylation sites enriched in nucleoli in HeLa S3 cells and confirmed by ChIP experiments that H1.4S187p is associated with active rRNA promoters and is induced at hormone response elements. In addition, H1.4S27 phosphorylation inhibits the binding to methylated H1.4K26 of HP1 thus; inhibiting heterochromatin formation [149].

Interestingly, phosphorylation of H1.2 on T146 leads to dissociation of H1.2 from p53, activating p53 target genes, inducing DNA damage response and eventually, apoptosis [133]. It is worth saying, that this study provides a molecular function beyond chromatin regulation for linker histone H1 as H1.2 directly interacts with p53 in the cytoplasm.

It seems contradictory that H1 phosphorylation can condense chromatin during G2/M phases and promote transcription and chromatin decondensation during S phase. A model to solve this issue comes from Roque et al. who analysed H1 secondary structure bound to DNA with partial or full phosphorylation of the C-terminal domain [142]. They showed that different phosphorylation levels lead to different proportions of α-helix, β-structures and unstructured regions and that partial phosphorylation impaired the capacity to compact chromatin. Different site-specific H1 phosphorylation may lead to specific structural changes [142, 145] and thus, affecting chromatin condensation or decondensation in different unknown ways.

Another important PTM is **H1 methylation** (Table 6), which is mostly found within the N-terminal tail. Within, K26 methylation of H1.4 is the most abundant and interestingly, probably conserved in *D. melanogaster* (H1K27me2) [144], which suggests also a conserved function. H1K26 methylation is catalysed by the methyltransferase G9a but also by the Ezh2 in the PRC2 complex. Instead, the lysine demethylase JMJD2/KDM4 removes this modification [145-147]. Methylated H1.4K26 recruits HP1 and L3MBTL, which form heterochromatin and thus, gene silencing. Interestingly, HP1 binding to methylated H1.4K26 is inhibited if the adjacent serine (S27) is phosphorylated [149] (Figure 18). In *D. melanogaster*, H1K27me2 is found accumulated in pericentromeric heterochromatin in metaphase [144], although not proven, it might contribute also to heterochromatin formation and/or maintenance. In

addition, an unknown protein can also acetylate human H1.4K26 and upon deacetylation by SIRT1, H1.4K26 can be methylated and subsequently, facultative heterochromatin is formed [150]. T47D cells only expressing a H1.4K26A mutant display defects in cell-cycle progression and misregulation of specific genes, showing the crucial role of methylated H1.4K26 [153].

H1 acetylation occurs both in tails and globular domains. Within the globular domain, acetylation is at those residues directly involved in DNA binding [151] and, as happens with core histone acetylation, it is thought to decompress chromatin and thus, activate transcription.

In agreement, at the N-terminal tail, the acetyltransferase GCN5 acetylates H1.4K34 and activates transcription by firstly, recruiting TAFI, a subunit of the transcription factor TFIID and secondly, by directly reducing H1 chromatin affinity [152] (Figure 18). And, as mentioned before, H1.4K26 acetylation is related to active transcription [150].

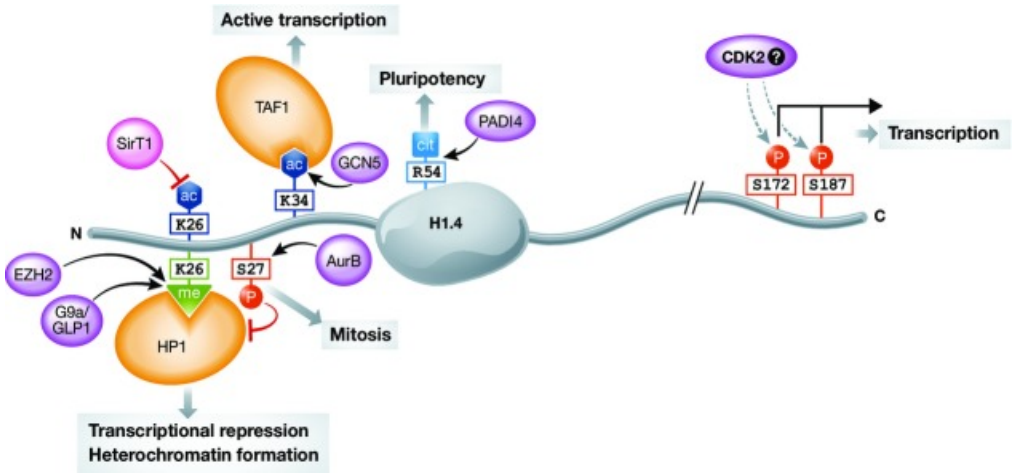


Figure 18. H1.4 modifications with a known cellular function. Figure from [148]

Christophorou et al. described **H1 citrullination** in mice ESCs [154]. Arginine residues are converted to the non-coded amino acid citrulline by peptidylarginine deiminases (PADIs), which leads to the loss of a positive charge. PADI4 is expressed during pluripotency and reprogramming, it binds regulatory elements of key stem cells genes and activates their transcription. PADI4 citrullinates H1 at R54, evicting H1 from chromatin, resulting in global chromatin decondensation and active transcription of pluripotency cell genes (Figure 18).

H1 ubiquitylation has been found to have a role in the DNA damage response [155]. Upon double-strand DNA breaks, their adjacent chromatin regions are ubiquitylated to generate binding sites for the DNA repair machinery. E3 ubiquitin ligase RNF8 ubiquitylates linker histone H1 that recruits RNF168 triggering the amplification of K63-linked ubiquitin chains by the E2 ubiquitin-conjugating enzyme UBC13. Linker histone H1 ubiquitylation plays an important role in initiation and amplification of K63-linked ubiquitin chains upon DSBs.

H1 ADP-ribosylation by poly ADP-ribose polymerase 1 (PARP1) mediates H1 displacement from promoters leading to transcription activation and chromatin remodelling. In fact, H1 and PARP-1 exhibit a reciprocal binding pattern at promoters and depletion of PARP-1 leads to an increase in H1 binding [156, 157]. Furthermore, CDK2 contributes to H1 displacement from progesterone responsive promoters by activating PARP1 [158].

PARP1 physically associates with the transcription factor GATA3, regulator of mammary gland development and a marker of luminal breast cancer. CCND1 gene activation by GATA3 is mainly regulated by PARP1, which triggers linker histone H1 displacement at its promoter [159].

Additionally, H3 ADP-ribosylation done by ARTD1 leads to the inhibition of H3 methylation by SET7/9 [160]. SET7/9 methylates H1.4 while H3 is being ADP-ribosylated, showing how different affinities of H1 and H3 for modifying enzymes are also crucial and play a role in the crosstalk between different chromatin components.

In addition, linker histone H1 is carbonylated, formylated, denitrated, crotonylated and lysine 2-hydroxyisobutyrylated although, their function is still unknown [161].

2.4. Specificities of histone H1 variants

As mentioned, all linker histone H1 variants have a common structural function in chromatin and they can play redundant roles, Fan et al. performed the main experiments supporting this idea [115]. Single or double KO mice are viable with no apparent phenotype; instead, triple KO mice are embryonically lethal indicating that the total H1 content and not the lack of any particular H1 variant is crucial (See *2.2. Linker histone H1 function*). Even though, new evidences are pointing to a more complex and dynamic role of H1 variants in chromatin but also, in specific cellular processes such as DNA damage response, pluripotency, development or cancer.

2.4.1. Sequence conservation

Linker histone H1 variants were originated due to gene duplication events, meaning they are paralog genes; H1 variants genes in other species are ortholog. As expected for paralog and ortholog genes, H1 variants between species have a higher degree of identity than between H1 variants in the same specie [162, 163]. The less conserved terminal domains of H1 variants show significantly higher similarity between species than within species in contrast to the globular domain, which is much conserved across species and between variants (Figure 15) [162]. These observations suggest a positive selection of H1 variants across species through evolution, meaning that H1 variants might have specific and differential functions that need to be conserved.

2.4.2. Expression patterns of histone H1 variants

Another evidence supporting H1 variants functional specificity is their differential expression pattern in different cell types and cellular processes such as development, pluripotency, differentiation or cancer.

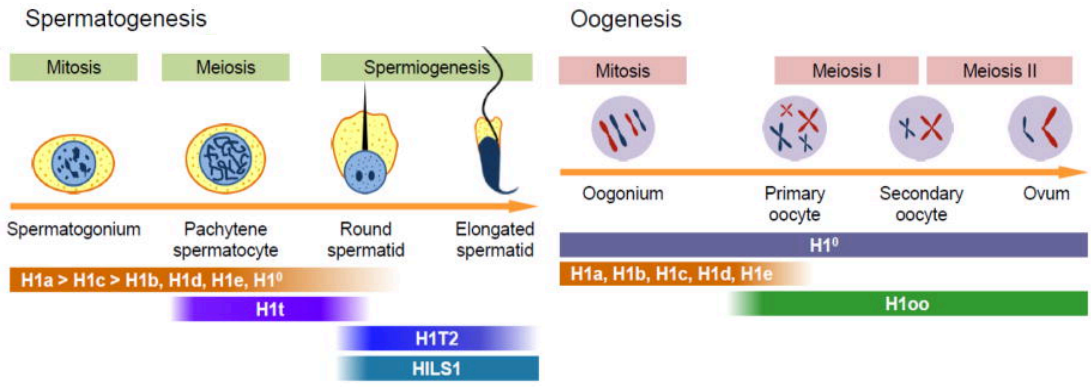


Figure 19. The expression patterns of histone H1 variants during gametogenesis (H1a, H1b, H1c, H1d H1e and H1⁰ in mice; H1.1, H1.5, H1.2, H1.3, H1.4 and H1.0 in humans, respectively). Figure from [175].

H1.1 was only found in some tissues such as thymus, testis, spleen and lymphocytic and neuronal cells [164, 165]. Similarly, some H1 variants are expressed only in germ line cells, being **H1oo** restricted to oocytes and **H1t, H1T2** and **HILS1** to testis. Germ-line specific H1 variants expression varies along gametogenesis in a highly regulated process (Figure 19). During spermatogenesis in mice, somatic H1 were detected in spermatogonia (predominantly H1a and H1c) whose levels decrease in meiotic spermatocytes until complete disappearance in spermatids [166, 167].

Expression of the other testis-specific H1 variants (H1T2 and HILS1) is only found in spermatids [168-171]. Instead, in oogenesis, somatic H1 variants, except $H1^0$, are gradually replaced with H1oo variant [172-174].

Replication-independent **H1.0** was instead, only found in terminally differentiated cells [176, 177]. Interestingly, Terme et al. showed that KD of H1.0 in human ESCs did not affect proliferation but impaired differentiation [178]. For instance, an 80% of H1 transcripts in differentiated cells are from H1.0. In agreement, pluripotent cells have a decreased level of H1.0 but also an increased level of H1.1, H1.3 and H1.5, showing that other histone H1 variants, apart from H1.0, are also differentially incorporated during differentiation and reprogramming to pluripotency. Interestingly, a recent study performed by Morales et al. showed that H1.0 abundance determine which cells within a tumour can sustain the long-term cancer growth [179]. H1.0 levels can explain intratumour heterogeneity between differentiated and self-renewing cancer stem cells. Cells within a tumour containing low levels of H1.0 are cancer stem cells, which have activation of self-renewal genes. In contrast, cells with high levels of H1.0 account for differentiated tumour cells with self-renewal genes repressed by H1.0 in AT-rich domains. They proposed H1.0 as a cancer biomarker with a prognostic value and a potential therapeutic target. Importantly, although authors showed an increase of H3K27me3 and CpG methylation levels in the promoter region of H1.0 gene, the precise regulation of H1.0 expression is still not solved. Other evidences support a dysregulation of H1 variant levels in cancer (Figure 20) [180]. For example, breast invasive carcinoma has an overall increase in H1 transcription in contrast to colorectal cancer, which has major changes in individual H1 mRNA levels.

Furthermore, **H1.5** protein levels positively correlate with high-grade pulmonary neuroendocrine and prostate tumours, which both showed a stronger and more homogenous immunofluorescence staining [181, 182]. H1.5 is a replication-dependent H1 variant whose expression decreases in differentiated and quiescent cells; thus, in cancer cells with high degree of proliferation is expected an increase in H1 replication-dependent variants. Subsequently, H1.5 can serve as a cancer biomarker. Similarly, in malignant adenocarcinomas an increase of H1.3 and a decrease of H1.0, H1.1, H1.4 and H1X levels are observed [183]. Strikingly, H1 expression patterns discriminate adenocarcinomas vs. adenomas with high accuracy.

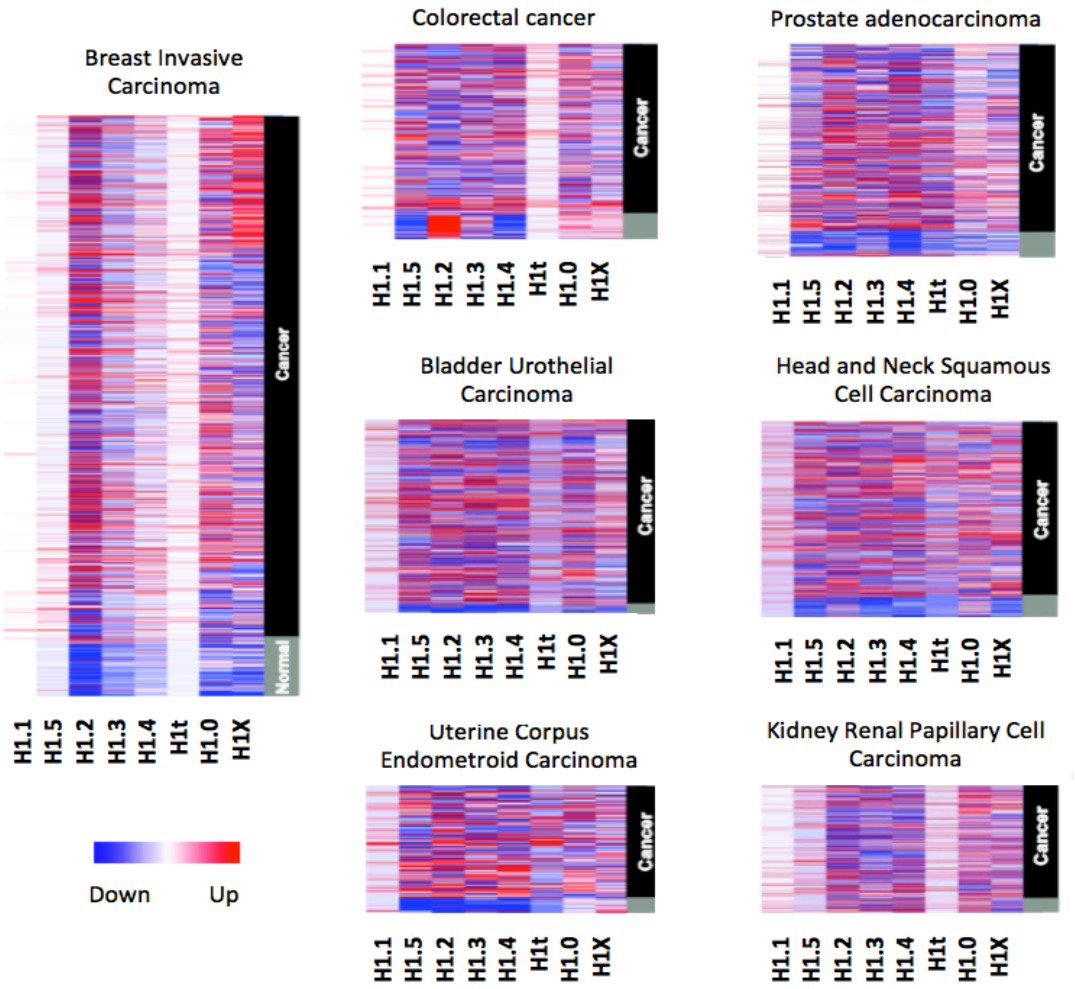


Figure 20. Transcriptional alternation of 8 histone H1 variants genes in cancer. Heatmap representing relative expression levels of H1 genes in the indicated cancers (black samples) and corresponding normal tissues (grey sample). Each row is an individual. Figure adapted from [180].

Altogether point to a specific function of H1 variants or maybe to a specific combination of H1 variants in different tissues that when not controlled could trigger the epigenetic changes seen in cancer cells. Nevertheless, care should be taken, as it is still not known if H1 alterations in cancer are a cause or a consequence. In addition, there is also a high variability between cancer samples with some individuals having opposite H1 expression patterns.

As said, different combinations of H1 variants exist depending on cellular type or process. Although being replication-dependent, **H1.2 and H1.4** transcription is not restricted to S-phase and their levels are constant as cells become quiescent, differentiated or both [188, 189] and only these two variants have been found in all

investigated cells [184-187]. It seems that both subtypes (H1.2 and H1.4) are crucial for cell functioning; indeed, H1.2 KD in breast cancer cells drastically reduced cell proliferation, arresting cells at G1-phase [92]. In contrast, **H1.1, H1.3 and H1.5** drop sharply their expression when cell become quiescent, as expected for replication-dependent variants.

As a replication-independent H1 variant, **H1X** levels remain nearly unchanged during S-phase. Although, Happel et al. showed that H1X accumulates in the nucleoli in G1-phase [190], indicating a possible control of H1X activity by changing nuclear subcompartment localization. In addition, H1X is highly expressed in human neuroendocrine cells and tumours [191] and has been proposed as a biomarker with prognostic value in astrocytic gliomas [192].

Interestingly, similarly to the other replication-independent (H1.0), H1X expression levels significantly increased upon induction of differentiation with retinoic acid in the embryonic carcinoma cell line NT2 [193]. They also showed incorporation of H1X to the promoter region of NANOG, a key stem cell transcription factor, highly repressed in differentiated cells.

In summary, linker histone H1 variant expression and composition is dependent on cellular types and tissues and is highly dynamic in cellular processes such as differentiation, pluripotency or development. H1 variants appear to have specific roles; thus, a tight regulation of H1 variant expression is necessary. In agreement, it has also been described that complex diseases such as cancer, have an altered H1 variant expression.

2.4.3. Chromatin binding affinity of histone H1 variants

Another important point pointing to an H1 variant specialization is their different binding affinities to chromatin. As mentioned before, FRAP studies with H1 fused to GFP showed that linker histone H1 is more mobile than core histones, although less than high mobility group (HMG) proteins [93]. In addition, H1 variants present different nucleosomal binding and affinity determined by differences in both C- and N-terminal tails [194].

Further FRAP experiments performed by Th'ng et al. showed that H1.1 and H1.2 presented the highest mobility followed by H1.0 and H1.3 and the less mobile H1 variants are H1.4 and H1.5 [214]. Partially in agreement, H1 variants differ in their ability to reconstitute nucleosomal arrays *in vitro*. H1.4 and H1.5 are the variants with higher ability followed by (H1.3, H1.2 and H1.0) > H1.1 > H1X [195]. Further *in vitro* competitive assays using H1 variants and two DNA types: long chromatin fragments

and scaffold-associated regions (SARs) showed that, although the absolute affinities of H1 variants for different DNA sequences vary widely, the relative affinities are conserved. Indeed, H1.1 was found to have the lowest affinity, H1.2 and H1.5 intermediate and H1.3, H1.4 and H1.0 the highest [196].

Interestingly, mutations in H1 residues, which are post-translational modified, also altered FRAP recovery kinetics [197, 198]. And, as explained, somatic histones are reversibly phosphorylated in interphase and mitosis, some in a subtype-specific manner, changing H1 affinity for nucleosomes [199, 200]. These studies suggest that H1 variant affinity can be modified by specific H1 post-translational modifications and thus, regulate H1 functions.

2.4.4. Histone H1 variants interacting partners

As explained above, H1 can be post-translational modified and these H1 variant specific PTMs have different interacting proteins and are catalysed by different enzymes (Figure 18). For example, when G9a or EZH2 enzymes methylate H1.4 at K26, HP1 is recruited, triggering the formation of heterochromatin and interestingly, the whole process is inhibited if H1.4S27 is phosphorylated [149]. Instead, acetylated H1.4K34 activates transcription by binding to the transcription factor subunit TAFI [198]. Also, when DNA-PK phosphorylates H1.2 at T146, it leads to the disruption of a complex containing H1.2 and p53, triggering the activation of p53 target genes, some of them inducing apoptosis [129].

In addition to specific H1 PTMs interacting with specific partners, several studies assessed H1 variant-specific partners without studying their specific PTMs. Indeed, variant H1b in mice (H1.5 in humans) has been found to interact with the transcription repressor MSX1 impairing MyoD transcription and myogenesis [202]. H1.5 also interacts with the forkhead box transcription factor FoxP3, altering its binding to target genes, which modulate expression and program CD4+CD25+ regulatory T cell function [203].

Recently, proteomic analysis (LC-MS/MS) revealed that H1.0 interacts (mainly via its C-terminal tail) with an extensive network of nucleolar proteins related to rRNA biogenesis, pre-mRNA splicing, and ribosomal proteins [204, 205]. This study prompts to a new paradigm for linker histone H1.0 in RNA metabolism; however, proteomic analyses of other H1 variants are needed to know if it is a specific H1.0 function. Using HeLa cells expressing flag- and HA-tagged H1.2 and co-purifying interacting complexes, Kim et al. also found H1.2 interacting with four ribosomal proteins in addition to co-repressors proteins and additional factors such as hnRNPK, nucleolin, DNA-PK… [206].

Furthermore, H1.2 occupancy is strongly and specifically stimulated when H3K27 is methylated by EZH2 [207]. The C-terminal tail of H1.2 is critical for the binding to H3K27me3 nucleosomes and inactivation of target genes. Besides, H1.2 was also found to stably interact with the E3 ubiquitin ligase cullin 4A (CUL4A) and PAF1 elongation complexes, which triggers target gene transcription via induction of H4K31 ubiquitylation, H3K4 trimethylation and H3K79 dimethylation [208]. Interestingly, both studies from the same group show that H1.2 can act as an activator or a repressor of transcription, depending on the interacting partners.

Recently, another report showed H1.2 also forms a complex with retinoblastoma tumour suppressor protein (pRb), a master regulator of cell proliferation through the interaction and regulation of the transcription factor E2F [201]. pRb-H1.2 complex was enriched on chromatin of E2F target genes, enhancing transcriptional repression and cell cycle arrest. When pRb gets phosphorylated by cyclin/CDKs, pRb-H1.2 complex gets disrupted and transcription activated. In agreement, Sancho et al. showed that H1.2 depletion leads to cell cycle G1-phase arrest [92]. Further, as mentioned before, H1.2 is released to the cytoplasm in a p53-dependent manner upon DNA double-stranded breaks, triggering apoptosis by prompting the release of cytochrome c from the mitochondria [105].

Moreover, linker histone H1 subtypes in mice differentially recruit DNMTs at H19 and Gtl2 loci, repressing its transcription by DNA methylation. H1c and H1(0) (H1.2 and H1.0 in humans, respectively) bound to DNMT1 and DNMT3B much weakly that the other tested H1 variants (H1a, H1b, H1d and H1e) [210, 211]. This report shows an interesting cross-talking between two epigenetic mechanisms to regulate gene expression, that seems specific for both DNMTs and H1 variants.

2.4.5. Gene expression regulation

Another important aspect of H1 variants function is gene expression regulation. Classically, linker histone H1 has been considered a transcriptional repressor due to its structural function in compacting chromatin. Nevertheless, global gene expression analyses upon histone H1 KD in several cell lines revealed that apart from up-regulated genes, a vast subset of genes are down-regulated pointing to a role of H1 as a transcriptional activator. In addition, H1 variant-specific regulation of gene transcription is observed.

In mice, overexpression of H1c and H1(0) (H1.2 and H1.0 in humans, respectively) in cell cycle synchronized cells and early time point assays, using expression microarrays, showed that H1c and H1(0) act primarily as specific rather than global regulators of gene expression [209]. Many of the genes were uniquely

affected by H1c or H1(0) although some were also affected by both variants suggesting specific but also common functions for H1 variants. In addition, it was found that H1(0) down-regulated a major number of genes than H1c, showing that H1(0) is a stronger repressor of transcription.

Importantly, as explained before, KO of single or double H1 variants in mice has no apparent phenotype in contrast to triple KO (TKO) mice, which are embryonically lethal [115]. Mice ESCs derived from the TKO showed a 50% reduction in H1 content and only a 0.56% of 4500 genes tested changed its expression. Those differentially expressed genes are enriched in imprinted and sex-chromosomes genes, which are normally regulated by promoter methylation at their CpG, pointing to a cross-talk between H1 and DNA methylation. In agreement with Yang et al. who also showed some, but not all, H1 variants interacting with DNMT1 and DNMT3B in mice [210, 211]. Moreover, TKO mice ESC showed decreased levels of H4K12 acetylation and H3K27 methylation in addition to a reduction in nucleosomal repeat length (NRL).

Another study in human breast cancer cells, using inducible shRNA-mediated knockdown systems for single H1 somatic variants, also observed variant-specific regulation of transcription, as the subset of differentially expressed genes varies between variants [92]. Some genes differentially expressed upon single H1 variant KD are dysregulated by several H1 variants, pointing also to redundant roles; nevertheless, most of the genes are only affected by one specific H1 variant and the ratio of down- versus up-regulated genes differs between variants, being 1 for H1.5 and 2.5 for H1.2. Interestingly, H1.2 depletion caused decreased global nucleosome spacing and cell cycle G1-phase arrest by repressing the expression of cell cycle-related genes. Concomitantly, H1.4 depletion caused cell death. Thus, depletion of individual H1 variant in T47D cell line has specific effects on gene expression and consequently, different variant-specific phenotypes are observed.

Regarding variant-specific gene expression regulation it would be interesting to solve the exact mechanism by which a locus is either activated or inhibited in a specific cell type via a specific H1 variant or variant-specific post-translational modification.

2.4.6. Nuclear localization and genomic distribution

Another important point indicating that linker histone H1 variants may have specific functions is its nuclear and genomic distribution, as several reports show a differential distribution of H1 variants in distinct cell types.

Firstly, immunofluorescence studies using polyclonal antibodies against H1.5 variant showed enrichment at the nuclear periphery where, as explained, a more

compacted chromatin is found [212]. In addition, they found that H1.2 distribution positively correlated with DNA concentration and H1.3 and H1.4 had a more punctuate pattern staining [213]. Another report also showed a differential nuclear distribution of H1X variant at nucleoli in G1-phase of the cell cycle [190]. By chromatin immunoprecipitation (ChIP) coupled to polymerase chain reaction (PCR) in selected genes it was also shown that H1 is globally depleted from active chromatin and that this depletion is variant-specific [116]. All H1 variants where found in heterochromatin and inactive genes while in active chromatin, variants H1.3 and H1.4 showed the higher depletion. Th'ng et al. also showed differences in H1 composition between pericentromeric and nucleolar heterochromatin and euchromatin, using green fluorescence protein (GFP) fusion proteins of H1 variants [214]. Specifically, the GFP patterns indicated that H1.1-H1.3 are more commonly found in euchromatin, whereas H1.4 and H1.5 in heterochromatin.

Strikingly, H1 variant composition correlation with chromatin status varies broadly between these studies, which used different cell lines, suggesting a cell-type specific function for H1 variants. Although, what is clear is that H1 variants have a different nuclear localization.

The appearance of high-throughput sequencing, providing high-resolution data, allowed the study of H1 variant localization genome-wide and as expected, differences were found.

In IMR90 fibroblast, H1.5 distribution, assessed by ChIP-Seq, showed that H1.5 forms block of enrichment in differentiated cells but not in human ESCs, associates with gene repression and is required for sirtuin-1 (SIRT1) binding and H3K9me2 enrichment [215]. Remarkably, 37% of H1.5 target genes in differentiated cells belong to gene family clusters, revealing a possible variant-specific regulation in differentiated cells. H1.5 distribution was reported to be variant-specific as it was compared to H1.3, however the specificity of the H1.3 antibody is missing.

In contrast, Cao et al. used a knock-in system with tagged H1c, H1d and H1(0) (H1.2, H1.3 and H1.0 in humans, respectively) variants in mice ESCs to assess their differential distribution genome-wide [216]. Both variants H1c and H1d were depleted from GC- and gene-rich regions and in active promoter (H3K4me3). Around TSS, a typical H1 distribution (the aforementioned "H1 valley") is seen, with dips much deeper at highly active than at silent genes. Instead, they were enriched at high H3K9me3 heterochromatic regions but not at H3K27me3. Moreover, H1c and H1d were also enriched in major satellites, which had higher nucleosome spacing than bulk chromatin. Indeed, the distribution of H1d and H1c were highly correlated through the genome (R=0.7866) but comparing specific peaks for each variant showed some

differences. Specifically, H1d unique peaks were relatively enriched at GC-rich sequences and LINEs, whereas H1c unique peaks at AT-rich sequences, Giemsa positive regions and satellite DNA.

Overexpressed H1(0) variant in mice ESCs was also found depleted from active promoters and enriched at major satellites like H1c and H1d. Other repetitive elements such as minor satellites and long interspersed nuclear elements L1 (LINEs) showed a specific H1(0) enrichment. Although, as H1(0) is low represented in ESCs, its overexpression might not represent true binding sites in differentiated cells, where H1(0) expression is highly induced.

Another interesting study showing a differential genomic distribution in replication-dependent somatic variants (H1.1 to H1.5) was performed in human lung IMR90 fibroblasts, using an integrative DNA adenine methyltransferase identification (DamID) analysis [217]. In agreement with Cao et al. [216], H1.2 to H1.5 had highly correlated DamID binding profiles, forming the expression-dependent "H1 valley". In addition, they were found depleted from CpG-dense regions, active promoters and *cis*-regulatory regions such as enhancer and CTCF-bound insulators. As expected, H1 subtypes were depleted at active core histone PTMs (H3K4me3 and H3K9ac) regions and enriched at repressive (H3K9me3 and H3K27me3) ones. Interestingly, H1.2 to H1.5 were found enriched at lamina-associated domains (LADs) pointing to a role of H1 in establishing and/or maintaining 3D chromatin organization [ref LADs]. Importantly, H1.1 showed a very distinct DamID profile compared to the other H1 variants, showing no enrichment in LADs and a higher enrichment than H1.2 to H1.5 in promoters, CpG and regulatory regions. Interestingly, chromatin states characterized by a high abundance of H1.1 are polycomb-type domains, assessed by H3K27me3 presence.

Instead, using breast cancer cell line T47D, Millán-Ariño et al. showed a distinct distribution for H1.2 compared to other somatic H1 variants (*Appendix I*) [218]. ChIP-chip and ChIP-Seq experiments were performed using available specific antibodies for H1.2 and H1X and antibody against hemagglutinin (HA) for H1.2 to H1.5 and H1.0 tagged to HA. H1.2 was the H1 variant that best correlated with gene repression, low GC content, gene-poor regions and it was the most enriched H1 variant in LADs. In fact, H1 distribution around the TSS indicated a transcriptional dependence for all H1 variants although H1.2 dip was much deeper and wider and even present at 10% most repressed genes. Comparing somatic H1 variants distribution with core histone H3 distribution, it can be seen that "H1 valley" is wider than the TSS nucleosome free region, extended up to 3Kb at promoter regions. Thus, it seems that regulatory factors and transcriptional machinery to perform their function, need a deeper and wider H1 displacement.

Interestingly, apart from showing uncovered features from H1.2 in T47D cell line, the two endogenous variants studied (H1.2 and H1X) were compared in different cancer cell lines and the ratio between H1.2 and H1X at distal promoter regions varies between them. Consistently, although not genome-wide, ChIP-qPCR experiments in three different chicken cell lines also showed a differential binding of H1 variants and H5 variant in selected genes [219].

Furthermore, when they compared endogenous H1.2 and HA-tagged H1.2 genome-wide few differences were observed. Although not shown, it is stated that tagged H1 levels at the protein level are close to or below endogenous levels despite of higher mRNA levels, suggesting a tight post-transcriptional regulation of H1 abundance and relative ratio of H1 variants, differing between cellular types. In agreement, Cao et al. knock-in system with tagged H1 in mice ESC showed that they are functionally equivalent to endogenous H1 [216]. Thus, until better endogenous antibodies specifically recognizing H1 variants appear, tagged H1 variants systems are the best approaches.

In summary, linker histone H1 variants distributions are broadly highly correlated in agreement to a general structural function of histone H1 in stabilizing, regulating and/or maintaining chromatin states. However, when looking deeply, some differences between H1 variant distributions can be found, showing variant-specific features. Although when comparing different cell lines, the relative ratio of H1 variants varies widely and similar specific features have been found for different H1 variants in different cell lines (for example H1.2 in breast cancer cells T47D with H1.2 to H1.5 in IMR90 fibroblasts). Thus, it will be interesting to analyse in parallel the distribution of histone H1 variants in several cell lines taking into account the relative ratio of H1 variants, which varies between differentiated, cancerous and stem cells (See *Appendix II* for a review of the genomic distribution specificities of linker histone H1 variants).

3. INTERFERON RESPONSE AND CHROMATIN

The immune system responds to the presence of pathogens by an innate, non-specific response and by an adaptive, pathogen-specific response. Innate immune system provides immediate and generalized defence and activates the long-lasting adaptive immunity. One of the bacterial features which activate an innate immune response are nucleic acids, sensed as exogenous.

The induction of type I interferon (IFN) is a hallmark of nucleic acids sensing by the innate immune system [220]. DNA located outside from the nucleus is sensed as exogenous DNA by several sensors that recognize double-stranded DNA (dsDNA), including proteins from the AIM2 family, the DDX family, RNA polymerase III and cyclic GMP-AMP synthase (cGAS) (Figure 21A) [221]. Upon sensor activation, STING-dependent signalling pathways are induced, activating the transcription factors NF-κB and IRF3, which promote IFNβ and cytokine production (Figure 21A) [221].

Another nucleic acid sensed by the innate immune system as exogenous is double-stranded RNA (dsRNA), which are replication intermediates for RNA viruses (Figure 21B) [222]. Specific sensors such as MDA5 (melanoma-differentiation-associated gene 5, also known as IFIH1) and RIG-I (retinoic acid inducible protein 1, also known as DDX58) are activated upon dsRNA presence in the cytoplasm also triggering an IFNβ production. Notably, RIG-I preferentially binds to short dsRNA and specifically binds to single-stranded RNA containing 5'-triphosphate, such as viral RNA. Instead, MDA5 recognizes preferentially long dsRNA (>1000bp) with no end specificity. Despite differences in dsRNA species, both activated sensors are finally recruited by the adaptor protein MAVS (also known as IPS-1, CARDIF or VISA) to the outer mitochondrial membrane, leading to IRF3, IRF7 and NF-κB activation, prompting a type I interferon response [223-225].

Upon IFNB gene transcription activation due to exogenous intracellular nucleic acid sensing, IFNβ is secreted from the cell, activating IFN receptors in near cells, leading to a JAK/STAT signalling cascade inducing interferon-stimulated genes (ISGs) transcription. All ISGs contain a sequence motif, named interferon-stimulated response element (ISRE), which is specifically recognized by the transcription factor ISGF3 complex that consists of STAT1, STAT2 and IRF9. In contrast to type I interferons (mainly IFNα and IFNβ), genes specifically induced upon type II interferon (IFNγ) contain a different sequence motif, a gamma interferon activation site (GAS), which is also specifically recognized.

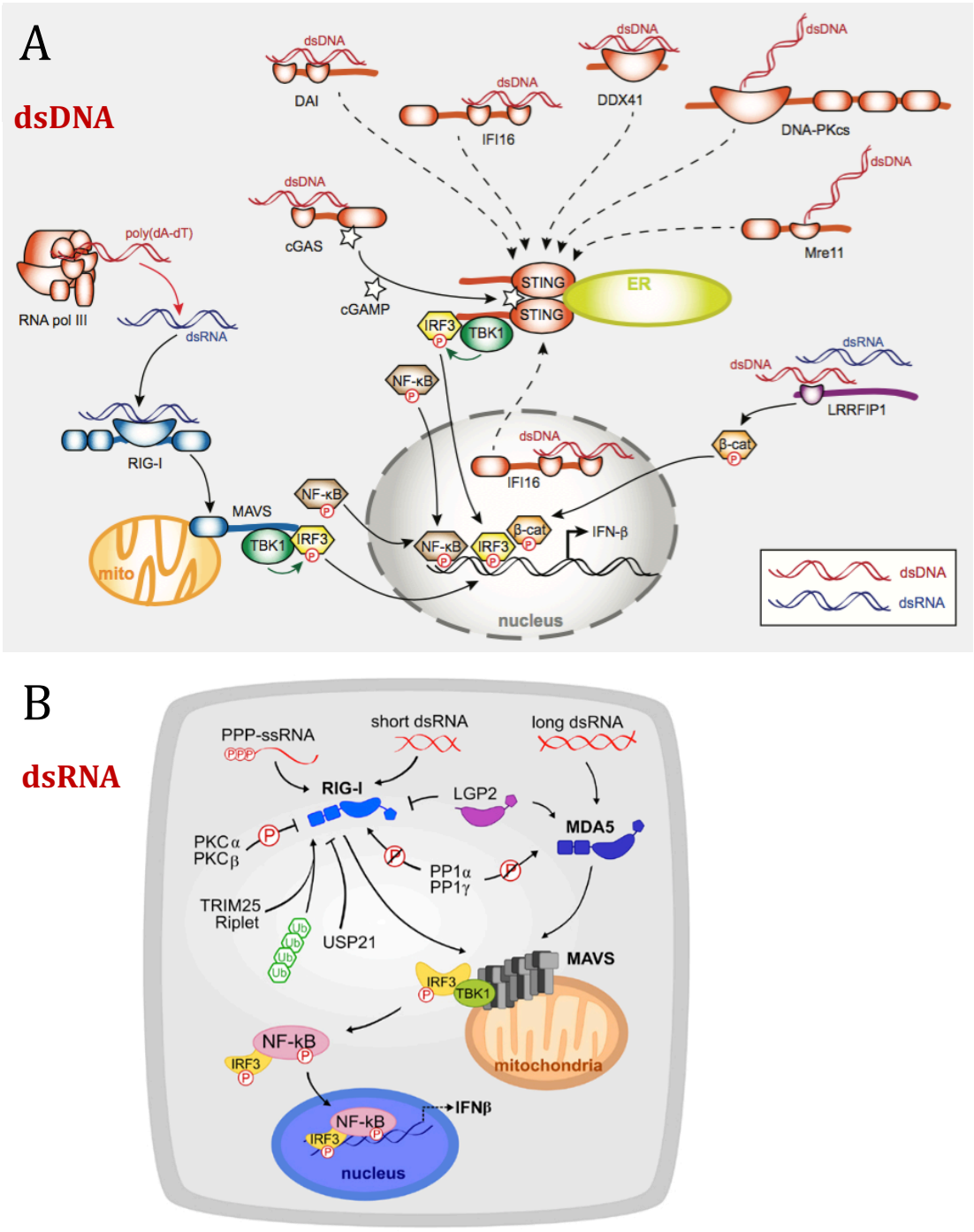


Figure 21. Intracellular receptors sensing nucleic acids involved in the induction of interferon β (IFNβ). (**A**) dsDNA sensors. Multiple DNA sensors have been proposed to activate a STING-dependent signalling pathway culminating in the activation of the transcription factors IRF3 and NF-κB. (**B**) dsRNA sensors. MDA5 and RIG-I recognize different dsRNA leading both an IFN production via the adaptor protein MAVS. Figures from [221, 222].

It is estimated that around an 8% of the human genome comprises endogenous retroviruses (ERVs), the vast majority defective due to deleterious mutations [218]. No single human ERV locus has been found that can produce infectious virions however, some of them are transcribed and translated producing nucleic acids and proteins with viral signatures, which can trigger an innate immune response [227, 228]. In fact, several autoimmune diseases have been related to a dysregulation in ERVs transcription [229, 230]. Thus, ERVs transcriptional silencing is crucial to evade an innate immune response, which leads, when not properly controlled, to cell death. ERVs silencing is mainly mediated through robust and inaccessible heterochromatin establishment and DNA methylation [231, 232].

In addition to repress ERVs, chromatin proteins have been shown to directly play a role in the innate immune system. The high mobility group box (HMGB) proteins are abundant proteins that, when released extracellularly, synergises with cytokines to promote inflammation [233, 234]. And, as explained, a similar extracellular role in innate immune response has been observed for linker histone H1, in both cases helping to promote and induce pathogen death (Figure 17). Intracellularly, HMGB proteins bind with high affinity to nucleic acids and promote activation of TLRs, RIG-I and intracellular DNA sensors, acting as sentinels [235].

In addition to endogenous retroviruses repression and a direct role in innate immune response system, a proper establishment and maintenance of chromatin is crucial to avoid R-loop accumulation. R-loops are nucleic acid structures consisting of an RNA-DNA hybrid and displaced ssDNA, which are ubiquitous in organisms from bacteria to mammals upon transcription. However, a high R-loop accumulation leads to genome instability, a driver mechanism of cancer [237, 238].

RNA-DNA hybrids are also specifically recognized by DNA sensors independently of MAVS, the main protein in the RNA sensing cascade (Figure 21). RNA-DNA hybrids bind to cGAS, which results in the synthesis of cyclic GMP-AMP (cGAMP), triggering the activation of STING in the endoplasmic reticulum (Figure 21A) [236]. STING activation leads to IFNβ production and thus, activation of the innate immune response. Retroviruses could provide another source of intracellular RNA-DNA hybrids, generated upon reverse transcription.

Interestingly, RNA-DNA hybrids are also found in R-loops that when aberrantly accumulated leads to DNA breaks, triggering genome instability and eventually, cell death. *In vivo*, R-loops are prevented and removed by ribonucleases of class H (RNAse H), among other proteins. Mutations in genes encoding for RNAse H (RNASEH2A, RNASEH2B and RNASEH2C) have been observed in chronic inflammatory diseases, such as Aicardi-Goutières syndrome. Thus, it was hypothesized that those mutations may

result in a less efficient removal of R-loops, thereby culminating in RNA-DNA hybrid accumulation, leading to a constant immune response due to cGAS-STING activation [236]. In fact, it has been shown a genome-wide hypomethylation in Aicardi-Goutières syndrome and an accumulation of RNA-DNA hybrids, preferentially found in repetitive elements and intergenic regions [239]. Again, regions tightly controlled by heterochromatin processes (See *1.6. Heterochromatin and repetitive elements*).

Another epigenetic mechanism has been involved in the activation of the interferon response, DNA methylation. Firstly, Leonova et al. showed that DNA hypomethylation in the absence of p53 leads to a transcriptional activation of repetitive elements, such as pericentromeric repeats and ERVs, prompting an interferon type I response, which was named TRAIN (transcription of repeats activates Interferon) [243]. Importantly, high doses of DNA-demethylating agents for long time periods were used, pointing to indirect effects in the p53 involvement. Concomitantly, recently, two reports showed that TRAIN is induced upon treatment with DNA-demethylating agents in a p53-independent manner [241, 242]. And, even the same group showed that curaxin, a small molecule, which disrupts histone/DNA interaction, can induce TRAIN independently of the p53 status of the cell [243]. Thus, p53 does not seem to play a direct role in the interferon response seen upon DNA hypomethylation. Interestingly, it seems that the observed IFN response is due to the accumulation of dsRNA, which activate MDA5-MAVS pathway (Figure 21B) [241-243].

Other chromatin decondensing drugs such as trichostatin A (TSA), a histone deacetylase inhibitor (HDACi), leads to an IFN response in mice embryonic fibroblasts (MEFs) and not a so robust response in HeLa-TI cells [243]. Thus, different chromatin opening agents can induce an interferon response, depending its magnitude on the cell type. In the same study, it is hypothesised that nucleosome opening by different epigenetics mechanisms is sufficient to allow transcription of repetitive elements and thus, induce TRAIN.

In addition to maintain a proper control of repetitive elements, chromatin also modulates the DNA damage response (DDR). There are two types of DNA damage: DNA breaks (single- or double-stranded) and mismatched bases, sensed by several specific proteins (Figure 22). The activation of the immune response is prompted by different sensors, transducer kinases and effector of DNA damage response. In addition to directly interacting with activating immune system proteins, DNA damage releases dsDNA, which activate IFN responses by STING-dependent pathways (Figure 22) [244].

Replication and transcription, processes greatly modulated by chromatin, are great sources of DNA damage. Interestingly, a cross-talk between chromatin and DNA damage responses has been widely observed. Chromatin has been implicated as a mediator of the DDR mainly as an important mark to recruit DDR machinery to DNA sites that were damaged. As an example, core histones and linker histone H1 have been implicated in marking dsDNA breaks sites, phosphorylation of H2A.X variant (named γH2A) is increased and, as mentioned, linker histone H1 poly-ubiquitylation is an important signalling intermediate. Finally, both γH2A and polyUb-H1, mediate DDR machinery recruitment to specific DNA sites [245, 155]. Apart from signalling, chromatin structure is also properly remodelled at those sites where DDR machinery is operating to increase its accessibility and thus, its efficiency [246].

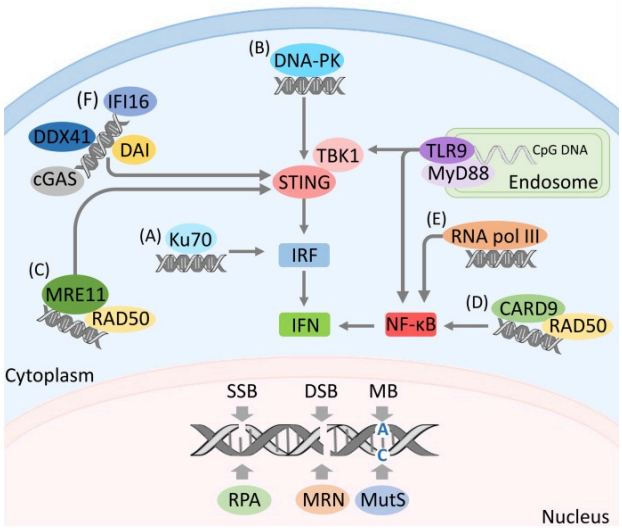


Figure 22. DNA damage response and innate immune system nucleic acid sensing. Diverse nuclear sensors recognize DNA damage including the protein complexes such as replication protein A (RPA) and MRE11-RAD50-NBS1 (MRN) and also MutS proteins. Many DDR proteins directly interact and activate a IFN response in a STING-dependent manner. Figure adapted from [243].

In summary, a proper control of repetitive elements and DNA damage, which is influenced by epigenetic mechanisms (chromatin-related proteins and DNA methylation) is crucial to avoid innate immune responses, which can eventually prompt cell death (Figure 23). These links between chromatin and nucleic acid sensing by innate immune mechanism are of great interest. Indeed, complex human diseases such as cancer or aging, whose epigenetic landscape is completely changed, increased DNA damage and aberrant transcription of repetitive elements has been observed. In addition, the IFN response is activated in many types of tumors [247-250]. It is tempting to speculate that those epigenetic changes leaded to the aforementioned

processes and are, among others, responsible for the IFN response seen in some cancers (Figure 20).

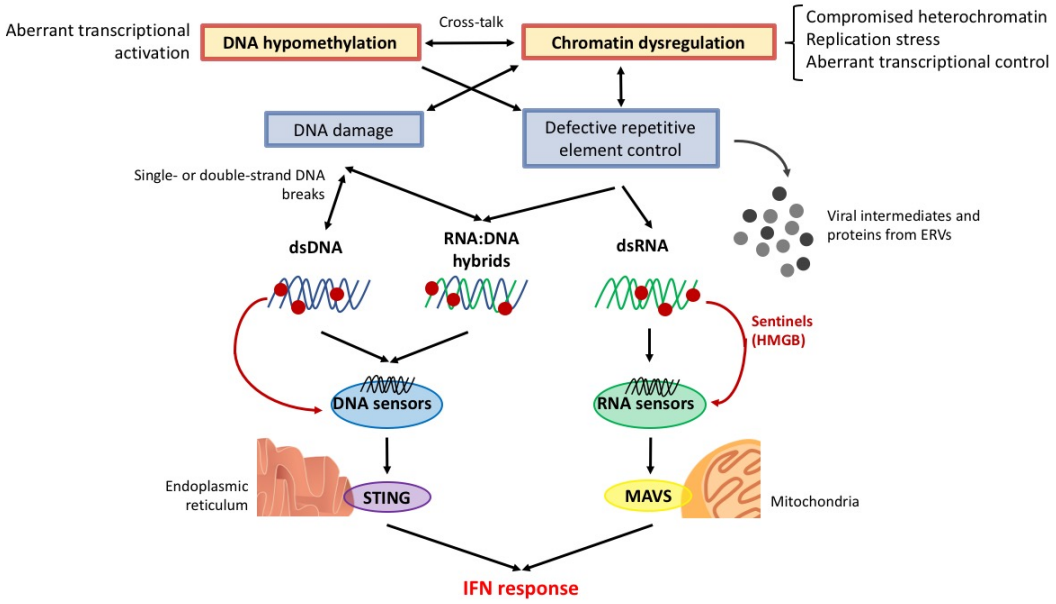


Figure 23. Nucleic acids sensed by the innate immune response triggered by an aberrant epigenetic regulation.

Surprisingly, transcription of repetitive elements has been shown to be modulated naturally in development, without compromising cell viability (See *1.6. Heterochromatin and repetitive elements*). RNA sequencing throughout early mouse embryogenesis revealed a dynamic and stage-specific transcription of repetitive elements, which most of them become repressed before implantation. This repression comes mainly by loss of activating PTMs (H3K4me3) rather than acquisition of repressing ones (H3K9me3) [251]. In this regard, pluripotent embryonic stem cells, which have, as explained before, a unique chromatin landscape characterized by a global chromatin opening, have elevated intergenic and intronic transcription with high transcription of retrotransposons [69, 70, 252]. How ESCs bypass this huge transcription is still not known, as wide-spread chromatin accessibility at repetitive elements linked stem cells and human cancer, leading to complete distinct outcomes [245].

Recently, an even more surprising, it has been observed that ERVs shaped evolution of a transcriptional network, underlying IFN response [254]. Chuong et al. assessed that ERVs constitute a dynamic reservoir of IFN-inducible enhancers. By analysing ChIP-Seq data of IRF1 and STAT1, they found peaks at ERVs near interferon-stimulated genes (ISGs) and described a functional role for ERVs in regulating, rather than triggering, the innate immune response.

OBJECTIVES

OBJECTIVES

The presented doctoral thesis had two main objectives, addressed by experimental and computational methods.

1. Study the specific genome-wide distribution of linker histone H1 variants (ChIP-Seq) in breast cancer cells
    a. Analyse regions not included in the human reference genome such as repetitive elements and ribosomal DNA
    b. Deeply analyse coding regions, specifically exons and introns
    c. Assess H1 distribution at CpG islands and regions with an aberrant DNA methylation in T47D cell line

2. Analyse transcriptomic and genomic changes in breast cancer cells depleted of several H1 variants
    a. Analyse differentially expressed genes (RNA-Seq) upon multiH1 variant knockdown
    b. Assess chromatin accessibility (ATAC-Seq) and changes in core histone post-translational modifications upon multiH1 knockdown.
    c. Analyse expression of repetitive elements by experimental methods such as immunofluorescence or RT-qPCR.

PUBLICATIONS

REPORT OF THE THESIS SUPERVISOR

The publications included in the doctoral thesis of **Andrea Izquierdo Bouldstridge** are two and both have been published in international journals ISI (International Scientific Indexing). Both publications have been done in our laboratory at the Institute of Molecular Biology of Barcelona, CSIC and Andrea Izquierdo-Bouldstridge signs as first co-authors in both. Included in appendix sections, she also was involved in another article as third author and in a review as second author.

The papers in order of appearance in the thesis' chapters are the followings:

CHAPTER I

Mayor R*, **Izquierdo-Bouldstridge A***, Millán-Ariño L, Bustillos A, Sampaio C, Luque N and Jordan A. Genome distribution of replication-independent histone H1 variants shows H1.0 associated with nucleolar domains and H1X associated with RNA polymerase II-enriched regions. **J Biol Chem**. 2015 Mar 20; 290(12):7474-91. doi: 10.1074/jbc.M114.617324 PMID: 25645921

* Both authors contributed equally to this work

Impact Factor: 4.125

In this paper, the genomic distribution of replication-independent linker histone H1 variants was assessed by combining computational methods and experimental procedures. In short, specific features for H1.0 related to the nucleoli and H1X with active chromatin regions were observed.

Andrea Izquierdo was the only bioinformatician involved in this project and she performed all computational analysis with already produced and published ChIP-Seq data of our laboratory combined with published data of other laboratories. Apart from standard bioinformatic procedures such as average signal profile around specific locations or peaks overlapping genomic features, she performed not standard bioinformatic analysis. Specifically, RNA-Seq data was analysed to retrieve alternative splicing events in our cell line or ChIP-Seq data was realigned to a costume reference genome containing ribosomal DNA in addition to look for enrichment in repetitive sequences. Specifically, she produced the results published in Figures 1, 2, 4, 5, 6, 8C and 8D.

The experimental methods performed in this paper were mainly done by Regina Mayor (first co-author with Andrea Izquierdo). Although stated as data not shown, Andrea Izquierdo performed ChIP-qPCR on selected genes looking for changes in core histone post-translational modificiations upon inhibition of H1X variant. Besides, Andrea Izquierdo actively contributed to experimental design, analysis and results discussion.

None of the published results in this paper have been used in another doctoral thesis.

CHAPTER II

Izquierdo-Bouldstridge A*, Bustillos A*, Bonet-Costa C, Aribau-Miralbés P, García-Gomis D, Dabad M, Esteve-Codina A, Pascual-Reguant L, Peiró S, Esteller M, Murtha M, Millán-Ariño L and Jordan A. Histone H1 depletion triggers an interferon response in cancer cells via activation of heterochromatic repeats. **Nucleic Acid Res**. 2017 Nov 16; 45(20): 11622-42. doi: 10.1093/nar/gkx746 PMID: 28977426

* Both authors contributed equally to this work

Impact Factor: 10.162

In this paper, the effect of multiple H1 variants depletion was assessed at the genomic and transcriptomic level again combining computational and experimental procedures. Shortly, upon multiH1 inhibition, a huge interferon (IFN) response is observed due to transcription of heterochromatic repeats that are sensed by the innate immune system as foreign.

Andrea Izquierdo, as happens for the other article, performed all bioinformatic analysis although, for the RNA-Seq data, with some help of other co-authors (Marc Dabad and Anna Esteve-Codina) due to the limited bioinformatic structure in our institute which has a poor computational power. Specifically, she analysed ATAC-Seq, RNA-Seq and ChIP-Seq data in the context of multiH1 variant depletion in addition to microarray data from pancreatic normal and cancerous tissue deposited in public repositories.

In addition to all bioinformatic methods, Andrea Izquierdo also designed and produced all experiments related to the expression of repetitive sequences in heterochromatic regions (RT-qPCR, immunofluorescence against double-strand RNA, ChIP-qPCR and Western Blot of core histone post-translational modifications). The contribution of Andrea Izquierdo was a key point in this article, as she actively

contributed to think, designed, analyse and discuss all experiments performed by her and other members of the laboratory included in this article.

In summary and more specifically, Andrea produced the results published and shown in Figures 2, 5, 6, 7 and 8 (and Supplementary Figures 7 and 8 and Table 2).

None of the bioinformatic results were used in another doctoral thesis. However, some of the experimental results were used for the doctoral thesis of the other first co-author, Alberto Bustillos. Specifically, Figures 3, 4 (and Supplementary Figures 1C, 1D, 3, 4, 5C and 6) which demonstrated firstly the critical role of H1.2 and H1.4 in triggering the interferon response. Secondly, that IFN is secreted in the media and thirdly that depletion of several sensors and adaptor of the IFN pathway by shRNA or chemical inhibitors impaired the triggered IFN response.

Andrea's contribution in the following papers, included in the appendix sections, was minor compared to those included in Chapter I and II. In order of appearance are the following:

APPENDIX I

Millán-Ariño L, Islam AB, **Izquierdo-Bouldstridge A**, Mayor R, Terme JM, Luque N, Sancho M, López-Bigas N and Jordan A. Mapping of six somatic linker histone H1 variants in human breast cancer cells uncover specific features of H1.2. **Nucleic Acid Res**. 2014 Apr; 42(7): 4474-93. doi: 10.1093/nar/gku079 PMID: 24476918

Impact Factor: 10.162

Within this paper, specific features of genomic distribution of linker histone H1.2 variant in breast cancer cell line were observed. Shortly, H1.2 was the H1 variant more related to gene repression, gene-poor regions and to lamin-associated domains (LADs).

Andrea Izquierdo contributed in the bioinformatic analysis shown in Figures 4C, 4D, 5B, 5C and 6, with already mapped ChIP-Seq data produced and analysed by Lluis Millán-Ariño and Abul Islam, respectively. Although she participated in this paper and it is related to her thesis results, she includes it in an appendix because Lluis Millán-Ariño used all the results in his doctoral thesis.

APPENDIX II

Millán-Ariño L, **Izquierdo-Bouldstridge A** and Jordan A. Specificities and genomic distribution of somatic mammalian histone H1 subtypes. **Biochim Biophys Acta**. 2016 Mar; 1859(3): 510-9. doi: 10.1016/j.bbagrm.2015.10.013 PMID: 26477490 Review

Impact Factor: 4.702

Within this review of the specificities and genomic distribution of somatic mammalian histone H1 variants, mainly done by the first author Lluis Millán-Ariño, Andrea Izquierdo contributed in the writing of her already published results shown in Chapter I, where, as said, H1.0 was related to nucleoli and H1X to actively transcribed regions.

Albert Jordan Vallès

Thesis supervisor

CHAPTER I

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 290, NO. 12, pp. 7474–7491, March 20, 2015

Genome Distribution of Replication-independent Histone H1 Variants Shows H1.0 Associated with Nucleolar Domains and H1X Associated with RNA Polymerase II-enriched Regions*

Received for publication, November 5, 2014, and in revised form, December 24, 2014 Published, JBC Papers in Press, February 2, 2015, DOI 10.1074/jbc.M114.617324

Regina Mayor[1,2], **Andrea Izquierdo-Bouldstridge**[1], **Lluís Millán-Ariño**[3], **Alberto Bustillos, Cristina Sampaio, Neus Luque, and Albert Jordan**[4]

From the Institut de Biologia Molecular de Barcelona, Consejo Superior de Investigaciones Científicas, Barcelona, Catalonia 08028 Spain

Background: There are seven histone H1 variants in somatic mammalian cells, two of which are replication-independent, H1.0 and H1X.
Results: In breast cancer cells, H1.0 is enriched at nucleolus-associated domains, whereas H1X is associated with RNA polymerase II-enriched regions.
Conclusion: Most H1 variants show great redundancy across the genome, but there is also some specificity.
Significance: Some H1 variants may have specific functions.

Unlike core histones, the linker histone H1 family is more evolutionarily diverse, and many organisms have multiple H1 variants or subtypes. In mammals, the H1 family includes seven somatic H1 variants; H1.1 to H1.5 are expressed in a replication-dependent manner, whereas H1.0 and H1X are replication-independent. Using ChIP-sequencing data and cell fractionation, we have compared the genomic distribution of H1.0 and H1X in human breast cancer cells, in which we previously observed differential distribution of H1.2 compared with the other subtypes. We have found H1.0 to be enriched at nucleolus-associated DNA repeats and chromatin domains, whereas H1X is associated with coding regions, RNA polymerase II-enriched regions, and hypomethylated CpG islands. Further, H1X accumulates within constitutive or included exons and retained introns and toward the 3′ end of expressed genes. Inducible H1X knockdown does not affect cell proliferation but dysregulates a subset of genes related to cell movement and transport. In H1X-depleted cells, the promoters of up-regulated genes are not occupied specifically by this variant, have a lower than average H1 content, and, unexpectedly, do not form an H1 valley upon induction. We conclude that H1 variants are not distributed evenly across the genome and may participate with some specificity in chromatin domain organization or gene regulation.

There are five major classes of histones that participate in the correct folding of eukaryotic DNA into chromatin: the core histones H2A, H2B, H3, and H4, which form an octamer and constitute the nucleosome core particle, and the linker histone H1, which binds to the nucleosomes near the entry/exit sites of linker DNA. Stabilization of the condensed states of chromatin is the function most commonly attributed to the linker histone (1, 2), in addition to its inhibitory effect *in vitro* on nucleosome mobility (3) and transcription (4).

Histone H1 in humans is a family of closely related, single gene-encoded proteins, including seven somatic subtypes (H1.1 to H1.5, H1.0, and H1X), three testis-specific variants (H1t, H1T2, and HILS1), and one restricted to oocytes (H1oo) (5, 6). Among the somatic histone H1 variants, H1.1 to H1.5 are expressed in a replication-dependent manner, whereas H1.0 and H1X are replication-independent. The H1.1 to H1.5-encoding genes are clustered in a region of chromosome 6 together with the core histone genes, whereas the H1X and H1.0 genes are on chromosomes 3 and 22, respectively. H1.2 to H1.5 and H1X are ubiquitously expressed, H1.1 is restricted to certain tissues, and H1.0 accumulates in terminally differentiated cells. There are few studies characterizing the most recently identified and distantly related human variant, H1X, and its specific function in the cell remains unknown. Like H1.0, it has been suggested that H1X is enriched in a less accessible region of chromatin, but expression of the two variants is regulated differently (7). It has been shown previously that H1X accumulates in nucleoli in G_1 and is distributed across the entire nucleus in the S phase (8). The same year, Takata *et al.* (9) found that H1X was preferentially located at the chromosome periphery in mitosis, and they observed defects in chromosome alignment and segregation after H1X knockdown (KD).[5] Taken

* This work was supported by funding from the Spanish Ministry of Science and Innovation (MICINN), European Regional Development Fund Grant BFU2011-23057, and Generalitat de Catalunya Grant 2009-SGR-1222.
The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession nos. GSE49345 and GSE62766).
[1] Both authors contributed equally to this work.
[2] Recipient of a Técnico de Apoyo contract from Consejo Superior de Investigaciones Científicas-MICINN.
[3] Recipient of a Formación de Personal Universitario predoctoral fellowship from MICINN.
[4] To whom correspondence should be addressed: Institut de Biologia Molecular de Barcelona (IBMB-CSIC), C/Baldiri Reixac 4, Barcelona, Catalonia E-08028, Spain. Tel.: 34-93-402 0487; Fax: 34-93-403 4979; E-mail: albert.jordan@ibmb.csic.es.

[5] The abbreviations used are: KD, knockdown; TSS, transcription start site; H3K4me3, H3K9me3, and H3K36me3, histone H3 Lys-4, -9, and -36 trimethylation, respectively; ChIP-seq, ChIP-sequencing; LAD, lamina-associated domain; NAD, nucleolus-associated chromatin domain; qPCR, quantitative PCR; RNAPII, RNA polymerase II; ASE, alternatively spliced exon.

together, these findings indicate that H1X may have functions that differ from those of the other variants.

Because it participates in the formation of higher order chromatin structures, H1 is seen as a structural component related to chromatin compaction and inaccessibility to transcription factors and to RNA polymerase. Nonetheless, it has also been suggested that histone H1 plays a more dynamic and gene-specific role, participating in the regulation of gene expression. Previous studies on the effect of H1 depletion on global gene expression have found no effect on the vast majority of genes but rather have detected up- or down-regulation of small groups of genes (10–13). It is not clear whether the different variants have specific roles or regulate specific promoters. In mice, single or double H1 variant knockouts have no apparent phenotype due to compensatory up-regulation of other subtypes (14). These reports have favored the view that H1 variants are redundant.

On the other hand, we reported that depletion of single H1 subtypes by inducible RNA interference in breast cancer cells produced a range of phenotypic effects (10), suggesting different functions for the various H1 variants in somatic cells. Furthermore, H1 subtypes can be post-translationally modified, and these modifications modulate their interaction with various other proteins. This could explain some reported specific functions for certain H1 variants (15–24). Moreover, H1 subtypes have cell type- and tissue-specific expression patterns, and their expression is regulated over the course of differentiation and development (25–30). Different H1 subtypes have also been differentially related to cancer processes (31–34).

To fully understand the function of histone H1 and its variants, several studies have explored the genomic distribution of H1 *in vivo*. Initial biochemical and microscopy-based approaches suggested a non-uniform distribution of H1 in the cell nucleus and found differences between variants (35–37). However, due to the lack of specific ChIP-grade antibodies for most H1 variants, it has been challenging to identify the precise mapping of H1 variants in the genome until recently. Two reports, using ChIP of tagged H1 variants in mouse embryonic stem cells and DamID technology in human IMR90 cells, respectively, showed depletion of H1c and H1d from guanine-cytosine (GC)- and gene-rich regions as well as an overrepresentation in major satellites (38) and depletion of H1.2 to H1.5 from CpG-dense and regulatory regions, only H1.1 having a distinct profile (39). Moreover, it has previously been shown that when a gene is transcriptionally active, there is depletion of H1 (an H1 valley) at the TSS of its promoter (40).

Using variant-specific antibodies against H1 and hemagglutinin (HA)-tagged recombinant H1 variants expressed in breast cancer cells, we investigated the distribution of six H1 variants in promoters (ChIP-chip) and genome-wide (ChIP-seq), including H1.0 and H1X, for the first time (41). In short, we reported that histone H1 is not uniformly distributed across the genome, and there are differences between variants, H1.2 showing the most specific pattern and strongest correlation with low gene expression. H1.2 is enriched at chromosomal domains with low GC content and is associated with gene-poor chromosomes, intergenic DNA, and lamina-associated domains (LADs). Meanwhile, other variants are associated with higher GC content, CpG islands, and gene-rich domains. Overall, the distribution of H1.2 along chromosomes differed from that of other variants, including H1.0 and H1X, the two variants most structurally distant within the somatic H1 family.

In this new work, we have further analyzed the distribution of H1 variants in other genomic regions, including repetitive DNA, nucleolus-associated chromatin domains (NADs), and ribosomal DNA (rDNA), and their association with methylated CpG sites and RNA polymerase II-enriched regions. This analysis has revealed that H1.0 and H1X are enriched at particular regions compared with the other variants. H1.0 is the variant that is most abundant at NADs, rDNA, and certain satellite repeats related to nucleolus organizer regions. The association of H1.0 with nucleolar chromatin has been confirmed by immunoblotting on fractionated cellular extracts. In contrast, H1X is associated with RNA polymerase II-enriched sites, coding regions, and hypomethylated CpG islands. Notably, the H1X content at coding regions is higher at active genes, especially toward the 3′ end of genes, and more abundant at exons and intron-exon junctions than within introns themselves. We have also further investigated the functionality of H1X by testing the effect of an inducible KD of this H1 variant on cell proliferation and global gene expression.

EXPERIMENTAL PROCEDURES

Cell Lines and Culture Conditions

T47D-MTVL (carrying one stably integrated copy of luciferase reporter gene driven by the murine mammary tumor virus promoter) (42) and MCF7 breast cancer cells were separately grown at 37 °C with 5% CO_2. T47D-derivative cells were grown in RPMI 1640 medium, supplemented with 10% FBS, 2 mM L-glutamine, 100 units/ml penicillin, and 100 μg/ml streptomycin. MCF7 cells were grown in minimum Eagle's medium containing 10% FBS, 1% penicillin/streptomycin, 1% glutamine, and 1% sodium pyruvate. Doxycycline (Sigma) was added at 2.5 μg/ml when required.

Drug-inducible RNA Interference

H1X KD cell lines were established from T47D-MTVL and MCF7 breast cancer cells. Plasmids for the lentivirus vector-mediated drug-inducible RNA interference system (pLVTHM, ptTR-KRAB-Red, pCMC-R8.91, and pVSVG) were provided by Dr. D. Trono (University of Geneva) (58). After testing five shRNAs against H1X from the MISSION library (Sigma-Aldrich), the 21-mer H1X-specific target sequence 5′-CAACGGTTC-CTTCAAGCTCAA-3′ was chosen to generate the inducible system. The 71-mer oligonucleotides for shRNA cloning into Mlu/ClaI-digested pLVTHM were designed, annealed, and phosphorylated as recommended by Dr. Trono (see the Tronolab Web site). For the production of viral particles containing the lentiviral vector and infections, see Sancho *et al.* (10). The inducible knocked down cell lines were sorted in a FACSCalibur machine (BD Biosciences) for RedFP-positive and GFP-positive fluorescence after 3 days of doxycycline treatment. Then, cells were amplified in the absence of doxycycline until an experiment was performed. Over a 6-day treatment with doxycycline, cells were passaged on day 3. When required, serum-

containing medium was replaced with serum-free medium on day 4 to arrest growth.

Histone H1 Extraction, Gel Electrophoresis, and Immunoblotting

Histone H1 was purified by lysis with 5% perchloric acid for 1 h at 4 °C. Soluble acid proteins were precipitated with 30% trichloroacetic acid overnight at 4 °C, washed twice with 0.5 ml of acetone, and reconstituted in water. Protein concentration was determined with the Micro BCA protein assay (Pierce). Purified histones were exposed to SDS-PAGE (10%), transferred to a PVDF membrane, blocked with Odyssey blocking buffer (LI-COR Biosciences) for 1 h, and incubated with primary antibodies overnight at 4 °C and with secondary antibodies conjugated to fluorescence (IRDye 680 goat anti-rabbit IgG, LI-COR) for 1 h at room temperature. Bands were visualized in an Odyssey infrared imaging system (LI-COR). Polyclonal antibodies specifically recognizing human H1 variants, including those generated in our laboratory (10), are available from Abcam: H1.0 (ab11079), H1.2 (ab17677), H1.3 (ab24174), H1.4-T146p (ab3596), H1.5 (ab24175), and rabbit antiH1X (ab31972). Mouse anti-H1X was obtained from Sigma (SAB1400328). Other antibodies used were β-tubulin (Sigma, nrT4026), nucleophosmin (Abcam, ab15440), nucleolin (Abcam, ab22758), H3K4me3 (Millipore, 07-473), and H3K9me3 (Abcam, ab8898).

Cell Fractionation for Purification of Nucleoli

Cell fractionation was performed as described by Andersen *et al.* (43). Briefly, 30 million cells were resuspended in 1 ml of Buffer A (10 mM HEPES-KOH, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, and protease inhibitors: 1 mM phenylmethylsulfonyl fluoride, 10 μg/ml leupeptin, 0.1 units/ml aprotinin, 1 mM orthovanadate, and 50 mM NaF) and incubated for 10 min on ice. Then the cell pellet was homogenized, by passing the cell suspension through a 23-gauge needle 15 times and through a 25-gauge needle 10 times. From this, we collected the total protein fraction. The homogenized suspension was pelleted at 228 × *g* for 5 min at 4 °C, and the supernatant was taken as the cytoplasmic fraction. The remaining pellet was resuspended in Buffer B (0.25 M sucrose, 10 mM $MgCl_2$, and protease inhibitors) and was homogenized again by passing the suspension through a 23-gauge needle 10 times. Then it was centrifuged at 1,430 × *g* for 5 min at 4 °C on a sucrose cushion (Buffer C: 0.35 M sucrose, 0.5 mM $MgCl_2$, and protease inhibitors). The remaining pellet was resuspended with Buffer C and sonicated for six cycles of 10 s on ice. The sonicated sample was centrifuged at 2,800 × *g* for 10 min at 4 °C on a sucrose pillow (Buffer D: 0.88 M sucrose, 0.5 mM $MgCl_2$, and protease inhibitors). The supernatant was collected as the nucleoplasm fraction. The nucleoli pellet was washed with Buffer C and centrifuged at 200 × *g* for 2 min at 4 °C. Then it was resuspended with lysis buffer (SDS (2%), 67 mM Tris-HCl, pH 6.8). Protein concentration in all fractions was determined with the Micro BCA protein assay kit (Pierce). Fractionated extracts were exposed to SDS-PAGE (10%), transferred to a PVDF membrane, and immunoblotted as described above. Immunoblot band intensities were measured using ImageJ (version 1.48) software and normalized by Coomassie staining.

Immunostaining

Cells were grown over coverslips, washed twice with PBS, and fixed with 4% formaldehyde for 15 min at room temperature. After three washes, they were permeabilized with Triton X-100 for 15 min at room temperature and blocked with bovine serum albumin for 1 h. Then the cells were incubated with primary antibodies diluted with bovine serum albumin for 1 h at room temperature in darkness. After the pertinent washes, the secondary antibodies Alexa-555 and Alexa-647 were added for 1 h at room temperature in darkness. The nucleus was stained with DAPI. The coverslips were mounted on the glass slides using Mowiol mounting medium. The samples were visualized by confocal laser scanning microscopy using a Leica TCS SPE system.

Cell Cycle Analysis

Cells were washed with cold 1× PBS, fixed in 70% ethanol, and stained with analysis solution: 3% ribonuclease A (Sigma) (10 mg/ml) and 3% solution A (38 mM sodium citrate, 500 μg/ml propidium iodide) in 1× PBS. Samples were analyzed with a FACSCalibur machine, using CellQuest Pro Analysis software (both from BD Biosciences) and ModFit LT software (Verity Software House).

Chromatin Immunoprecipitation

Immunoprecipitation of chromatin was performed according to the Upstate (Millipore) standard protocol. Briefly, cells were fixed using 1% formaldehyde for 10 min at 37 °C, harvested, and sonicated to generate chromatin fragments of 200–500 bp. Then 20 μg of sheared chromatin was immunoprecipitated overnight with 2 μg of antibody. Immunocomplexes were recovered using 20 μl of protein A magnetic beads, washed, and eluted. Cross-linking was reversed at 65 °C overnight, and immunoprecipitated DNA was recovered using the PCR purification kit from Qiagen. Genomic regions of interest were identified by real-time quantitative PCR (qPCR) using SYBR Green Master Mix (Invitrogen) and specific oligonucleotides in a Roche Applied Science 480 light cycler machine. Each value was corrected by the corresponding input chromatin sample. Oligonucleotide sequences used for the amplifications are shown in Table 1.

RNA Extraction, Reverse Transcriptase qPCR, and Expression Microarrays

Total RNA was extracted using the High Pure RNA isolation kit (Roche Applied Science). Then cDNA was generated from 100 ng of RNA using the Superscript first strand synthesis system (Invitrogen). Gene products were analyzed by qPCR, again using SYBR Green master mix (Invitrogen) and specific oligonucleotides in a Roche Applied Science 480 light cycler machine. Each value was corrected by human *GAPDH* and represented as relative units. Each experiment was performed in duplicate. Gene-specific oligonucleotide sequences are shown in Table 1. The procedures for microarray hybridization using an Agilent platform (SurePrint G3 Human Gene Expression 8x60K version 2) and data analysis are described elsewhere (41). Gene ontology analysis was performed using the DAVID soft-

TABLE 1
Primer sets used for RT-PCR and ChIP-qPCR

Name of gene	Forward primer (5′–3′)	Backward primer (5′–3′)
ChIP-qPCR		
KNG1 TSS	TCCCAGTTGGCTCTTGATTC	TTTCCTCGGACTGTGATTCC
KNG1 −3 kb	GTGCAGGATGGGTGATTTTT	CCTGTGCTTCAACACCATTC
KRT37 TSS	AATCAAGGCAGGAGGTCAAA	CTTCAGATCAGCTGGGAAGG
KRT37 −3 kb	GGCACTTGTAGTGACCTGGAT	CCGAAGTCCTCAAAGTCCAT
UGT2B10 TSS	AAGGATGGCTCTGAAATGGA	CTGTATTCTGCGGCCCATAC
UGT2B10 −3 kb	GGCATTGGATATTGGCTGTC	TCACCCAGATTTCCCTTTTG
AMTN TSS	CGTGGACCCAAAGGTAACAT	TGTTGAAACTGGCTGGCATA
AMTN −3 kb	TGACATGTGCATTCAATCAGC	GCCCTTTAGTTCCAGGCATT
SPINK9 TSS	CGGACACCAGGTCACTTCTT	TTGCAAGTGTCAGAGCCAAG
SPINK9 −3 kb	TCAAGTTCACCAGGCTTTTTG	CCTCTATGATGAGTCCAGCTC
ALOX15B TSS	TAACCAGGGGCAATAACCAG	CCACGATGCTGACAGACACT
ALOX15B −3 kb	TTGAAAACGTGTGGGTCTTG	CACCTTTGGAGCAATGTCTG
CDK2 TSS	GCGGCAACATTGTTTCAAGT	GTCGGGATGGAACGCAGTAT
CDK2 −3 kb	CAGCGAGGAAAGTCACATCA	TGGGGTGAGGGTAGTTTCTG
FOXB2 TSS	GTGCGGAGAGATTCTGTGGT	AGAGTAGGGCGGTTTTTGGT
FOXB2 −3 kb	CTGTAGCGAGCTCACCCAGT	ACAAATCTTGGGCGCATAAC
TBKBP1 TSS	AGGCCCGAGAGAAGTACACA	CGAAAGCAGGAGTAGGCAGT
TBKBP1 −3 kb	TGCAATGAGATCAGGTCCAG	GTTGGTGGCAAAAGTCCATT
ACTL7B TSS	AGGTGGGGGATCTCATTTCT	CTTGCTCCCCTTCTCACATC
ACTL7B −3 kb	GGTCCCAAGACTGTGTCCAT	AGACAGCTCCTCTCCCTTCC
JUN TSS	GGGTGACATCATGGGCTATT	GCCCGAGCTCAACACTTATC
JUN −10 kb	CCTTTTTGTCCCTCCAAACA	TCTAGGAACTGAGCCCTCCA
RT-PCR		
KNG1	GTGGTGGCTGGATTGAACTT	CGCAAATCTTGGTSGGTGGT
KRT37	TGGGGAGATGATTCTGAAGG	TGCTACCGGTTGATTTAGGG
UGT2B10	GACCTGCTGAATGCACTGAA	ACTGGAACCAGGTGAGGTTG
AMTN	AGCAGGAGGAGCAGGTGTAA	CCAAATTCGAGGCAGCTTAG
ALOX15B	GAAGTGGCTGCCAAAGAGAC	GCTGGCCTTGAACTTCTGAC
SPINK9	GAATGTGCCAAACAGACGAA	GTTTTGCCATCAGATCCACA
H1X	TTCCTTCAAGCTCAACCG	TGCCTTCTTCGCTTTGTG
H1.0	CCTGCGGCCAAGCCCAAGCG	AACTTGATCTGCGAGTCAGC
H1.1	CTCCTCTAAGGAGCGTGGTG	GAGGACGCCTTCTTGTTGAG
H1.2	GGCTGGGGGTACGCCT	TTAGGTTTGGTTCCGCCC
H1.3	CTGCTCCACTTGCTCCTACC	GCAAGCGCTTTCTTAAGC
H1.4	GTCGGGTTCCTTCAAACTCA	CTTCTTCGCCTTCTTTGGG
H1.5	CATTAAGCTGGGCCTCAAGA	TCACTGCCTTTTTCGCCCC

ware (Database for Annotation, Visualization, and Integrated Discovery).

Analysis of ChIP-seq Data

Because there are a limited number of H1 variant-specific ChIP-grade antibodies (only H1.2 and H1X being available to us), we developed T47D-derived cell lines stably expressing hemagglutinin (HA)-tagged versions of each of the five somatic H1 variants expressed in most cell types (H1.0, H1.2, H1.3, H1.4, and H1.5) (10). Therefore, in addition to using H1.2 and H1X antibodies to pull down these variants in parental T47D cells, an anti-HA antibody was used to specifically pull-down H1-associated chromatin fragments in cells expressing H1-HAs. ChIP-chip and ChIP-seq data on the occurrence of H1 variants at promoters and genome-wide in T47D-derivative cells, respectively, were generated in previous research, and the analysis was reported elsewhere (41). Briefly, ChIP-seq libraries were prepared with the ChIP-seq Sample Preparation Kit (Illumina), and sequencing was performed with the Illumina HiSeq 2000 system. Read mapping and peak detection methods have been described before (41). Other types of analysis used were as follows.

Publicly Available Genome-wide Location Data Analysis—Genomic locations of CpG islands and LADs (in hg18) were taken from the University of California Santa Cruz (UCSC) database (44), genomic locations of NADs from Nemeth *et al.* (45), and RNA polymerase II (RNAPII) binding sites from Ballaré *et al.* (46). Further, acromeric satellite 1 (ACRO1) genomic locations (in hg 18) were taken from the UCSC database. Repetitive sequences were taken from RepBase database (47). The mean methylation levels at individual CpG islands were calculated by assessing the overlap between the methylation levels from Vanderkraats *et al.* (48) and the CpG islands using BedTools (49). The genomic locations of hyper- and hypomethylated regions from the T47D cell line were recalculated as in Ref. 50, the source of the raw data.

RNA-sequencing data from the T47D cell line was taken from Vanderkraats *et al.* (48). Reads were mapped to the hg18 genome using the TopHat algorithm (version 2.0.12) (51). Next, we extracted the database of "cassette" exons and retained introns included in the MISO software package (52). The inclusion levels (Ψ) of alternatively spliced exons (ASEs) and retained introns were estimated using the MISO algorithm (52) with default parameters. Exons with inclusion level $\Psi \geq 0.9$ were considered to be included ASEs, and those with $\Psi \leq 0.1$ were considered excluded ASEs. Retained introns with $\Psi \geq 0.9$ were considered to be retained in the T47D cell line.

H1 Occupancy at Genomic Features—Input-subtracted normalized average H1 variant read density was calculated at each location enriched in CpG islands, NADs, hyper- and hypomethylated regions, ACRO1, repetitive elements, exons, introns, ASEs, retained introns, and RNAPII and represented in box plots using in-house R scripts. As a control, a random sample of bulk genomic windows with equal width was used to perform the significance test (Kolmogorov-Smirnov test). In addition, for H1 abundance at ACRO1, a second method was used, namely mapping to sequences in RepBase (47), with the bowtie aligner (53) allowing for multiple positions.

ChIP signals around the center of RNAPII binding sites were calculated using normalized input-subtracted average tag numbers in each 50-bp bin in a set window. Relative distances of each tag from the aforementioned positions and average signals were determined using the Sitepro script from the CEAS package (54) and plotting using R. Continuous ChIP signal profile distribution of reads along the metagene, exons, and introns were performed using CEAS (54). Correlation analysis between NAD content and H1.0 abundance on individual chromosomes was performed using in-house R scripts.

H1 Occupancy at Individual Chromosomes—Occupancy of H1 variants at all human chromosomes is measured in terms of the mean of the input-subtracted ChIP-seq signal in 50-bp windows. LAD and NAD occupancy at all chromosomes was calculated as the number of bases coinciding with LADs or NADs divided by the length of the chromosome. Expression on each chromosome is the mean of the expression of all genes in that particular chromosome. Heat maps and dendrograms were created with in-house R scripts.

H1 Occupancy at rDNA—The abundance of H1 variants on rDNA was assessed as described previously (55). In short, because the rDNA sequence is not included in the reference genome, a custom hg18 assembly was constructed with the bowtie-build tool (53), adding a human rDNA repeat (GenBank™ accession number U13369). Alignment was carried out with the bowtie aligner (53), allowing up to two mismatches, and only unique hits were kept. The input-subtracted ChIP-seq signal in the rDNA sequence was calculated in reads/kilobase/million mapped and plotted using in-house R scripts.

Genomic Distribution of Replication-independent H1 Variants

Overlap Analysis of H1 Islands—The number of enriched and depleted H1 islands that overlapped NADs and RNAPII binding sites was calculated using BedTools (49) and plotted using in-house R scripts. Box plots showing the methylation levels of CpG islands overlapping enriched and depleted islands were calculated in the same way. Features were considered to overlap if the genomic intervals shared at least one base. Chromatin states based on the combined presence of H1 variant-enriched regions were calculated with a multivariate hidden Markov model using the chromHMM software (56).

Human H1 Variant Nomenclature

The correspondence of the nomenclature of the human H1 variants with their gene names is as follows: H1.0, HIF0; H1.1, HIST1H1A; H1.2, HIST1H1C; H1.3, HIST1H1D; H1.4, HIST1H1E; H1.5, HIST1H1B; and H1X, HIFX.

RESULTS

Human H1 Variants Are Differentially Associated with NADs and Repetitive DNA—To further explore whether the distribution of H1 variants is heterogeneous along several genome features or chromatin domains, we used our previously reported ChIP-seq data on endogenous H1.2, H1X, and H3 and HA-tagged H1.0, H1.2, and H1.4 (41). DNA sequences associated with the human nucleolus have recently been identified and used to define NADs by Németh *et al.* (45). Different gene families and certain satellite repeats are the major building blocks of NADs, which constitute about 4% of the genome. Using the input-subtracted ChIP-seq signal, we investigated the occupancy of H1 variants within NADs. H1.0 was significantly enriched at NADs (Fig. 1*A*). Furthermore, H1.0 was the variant that had the largest number of H1-enriched regions overlapping NADs (Fig. 1*B*).

A large part of chromosome 19 is associated with the nucleolus and is reported to be located in central regions of the interphase nucleus, being close to the nucleoli (45). We have previously reported that H1.0 is highly enriched at this chromosome (41). Correlation analysis between H1 variant ChIP-seq signals and NAD content at each chromosome confirmed that H1.0 is the most abundant variant at chromosomes with a higher NAD content, whereas H1.2 signals were negatively correlated with NAD content (Fig. 1, *C* and *D*). As predicted, there was a negative association between the content of NADs and LADs at chromosomes, the former being located within the inner part of the nucleus and the latter at the periphery. We previously reported that H1.2 overlaps with LADs (41).

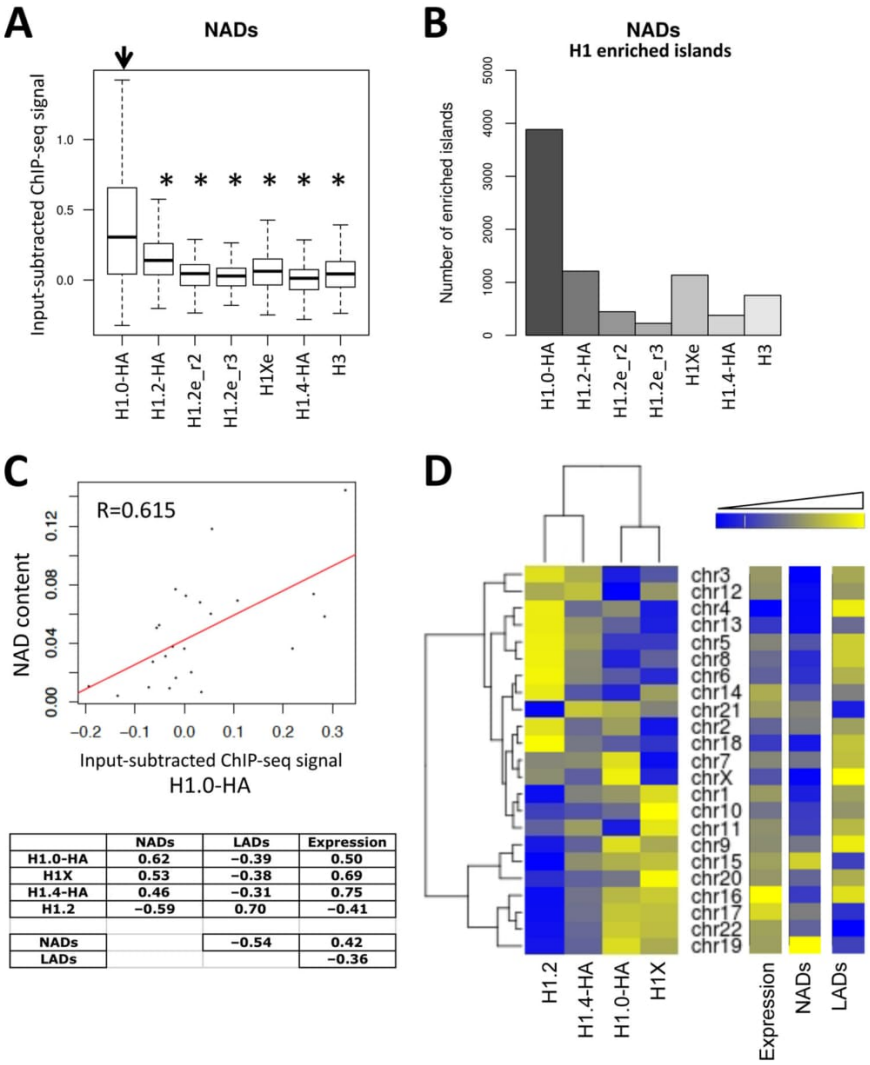
Ribosomal DNA encoding the 45 S single transcription unit giving rise to the 18 S, 5.8 S, and 28 S rRNA and flanked by non-transcribed spacers is present as repetitive DNA at the short arms of acrocentric chromosomes, called nucleolus organizer regions, within NADs. We aligned the input-subtracted H1 variant ChIP-seq signal to the rDNA complete repeating unit and found that H1.0 was enriched in the rDNA, mostly in the non-transcribed spacers (Fig. 2*A*). Instead, at the single transcription unit, H1X was locally enriched. H1.0 was also enriched at the 5 S ribosomal RNA subunit (encoded in tandem arrays, the largest one on chromosome 1), whereas H1X was slightly enriched at microRNAs and small nucleolar RNAs, compared with other variants (Fig. 2*B*).

Next, we aligned the input-subtracted H1 variant ChIP-seq signal to many repetitive DNA categories found in RepBase. One of the few categories that showed differential occupancy was ACRO1 (a 147-bp satellite found in the short arm of acrocentric chromosomes, where nucleolus organizer regions are located), which presented H1.0 enrichment (Fig. 2*C*). H1.0 was also enriched at SINE-VNTR-Alus (SVAs, non-autonomous, hominid-specific non-LTR retrotransposons) and telomeric satellites (Fig. 2*D*).

In summary, H1.0 is found to be enriched at DNA associated with nucleoli, including NADs, rDNAs, and acrocentric and telomeric satellites, suggesting that it could be involved in the stabilization of perinucleolar late-replicating heterochromatin. In contrast, H1X is overrepresented in the coding region of non-coding RNAs, such as 45 S rRNA, miRNA, and small nucleolar RNA, possibly related to the association of this variant with transcribed genes (see below).

H1.0 Is the H1 Variant Most Closely Associated with the Nucleolus—Because previous studies have shown enrichment of H1X and H1.0 in the nucleolus of cells by immunostaining (8, 9, 54), we further explored the localization of H1 variants at nucleoli by cellular fractionation and immunoblotting. Total, cytoplasmic, nucleoplasmic, and nucleolar extracts were prepared from T47D cells. Immunoblots were performed with specific antibodies for the six somatic H1 variants as well as tubulin as the cytoplasmic control and nucleophosmin and nucleolin as nucleolar controls, and the band intensity was quantified (Fig. 3). The cytoplasm was devoid of histone H1, whereas all variants were abundant in the nucleoplasm. At nucleoli, H1.0 was the most enriched variant, compared with total or nucleoplasm extracts. H1X and H1.5 were also enriched to some extent compared with the other variants. Overall, cellular fractionation confirmed our ChIP-seq results, pointing toward a specific association of H1.0 with nucleolar chromatin, although other variants, such H1X and possibly H1.5, are also present. In agreement with this, it has recently been reported that H1.0 interacts with an extensive network of proteins, many of them functioning in RNA metabolism in the nucleolus (55). Additionally, H1.5 was pulled down with H1.0 in one of the cell types used.

H1X Is Highly Associated with RNA Polymerase II-enriched Regions, Exons, Hypomethylated CpG Islands, and Active Transcription—We have described here and elsewhere (41) that different chromosomes have different abundances of H1 variants and that this is related to their gene content, mean gene expression, NAD and LAD content, and ultimately their position within the nucleus. Nonetheless, chromosomes are not uniform, and there may be territories or domains with different forms of chromatin organization. Using our ChIP-seq-derived data on genomic regions enriched in each individual H1 variant (41), we identified chromatin states based on the combined presence of H1-enriched regions, and we analyzed the presence of specific associated features (Fig. 4*A*). Because the number of islands of H1 enrichment identified was limited (ranging between 7,000 for H1.2 and 49,000 for H1.0 (41)), most of the genome was in a chromatin state without H1 islands. The next most abundant states contained islands enriched in H1.0, H1X, or H1.0 and H1X simultaneously. As reported previously, chro-



	NADs	LADs	Expression
H1.0-HA	0.62	−0.39	0.50
H1X	0.53	−0.38	0.69
H1.4-HA	0.46	−0.31	0.75
H1.2	−0.59	0.70	−0.41
NADs		−0.54	0.42
LADs			−0.36

FIGURE 1. **H1 variant abundance at NADs and across human chromosomes.** *A*, box plots showing the occupancy of H1 variants (input-subtracted ChIP-seq signal) within NADs. Significance was assessed using the Kolmogorov-Smirnov test. Significant enrichment of H1.0 compared with other variants is marked with *asterisks* ($p < 0.001$). NAD data determined in HeLa cells were obtained from Németh *et al.* (45). The labels for the different ChIP-seq data sets are consistent with those used elsewhere (41), H1.2e and H1Xe referring to endogenous H1 variants immunoprecipitated with variant-specific antibodies and H1.2_r2 and H1.2_r3 being two independent ChIP-seq replicates. *B*, number of H1 variant-enriched regions overlapping with NADs. Areas were considered to show enrichment of H1 variants if there was a -fold change greater than or equal to 2 compared with inputs derived from ChIP-seq data. *C*, correlation scatter plot between the occupancy of H1.0 at all chromosomes and the NAD content. The table *below* shows Pearson's correlation coefficient (*R*) between the occupancy of H1 variants at all chromosomes and the NAD or LAD content or mean gene expression. Correlation between NAD content, LAD content, and gene expression of all chromosomes is also shown. *D*, heat map and dendrogram of the occupancy of H1 variants (mean input-subtracted ChIP-seq signal over 50-bp genomic windows) at individual chromosomes. Mean gene expression as well as NAD and LAD contents of all chromosomes are shown as heat maps.

matin states containing H1.2 were associated with lamina, as were states containing H1.4. States containing other variants or combinations of variants were associated with genes and CpG islands, especially those containing both H1.0 and H1X.

Because, in this work and elsewhere, we have found that the association with gene promoters and coding regions differs between H1 variants, we explored the overlap of H1 variant-enriched or -depleted regions with RNAPII binding sites, using data from T47D cells obtained by Ballaré *et al.* (46). H1X was the variant showing the greatest overlap of enriched regions and least overlap of depleted regions with RNAPII peaks (Fig. 4*B*). Genes with an RNAPII peak had a higher mean expression level than other genes (data not shown), in agreement with our previous observation that H1X-enriched target genes are highly expressed and H1X-depleted target genes are repressed (see Fig. 7 in Ref. 41). Next, we analyzed the strength of input-subtracted H1 variant ChIP-seq signals within RNAPII peaks, and we again found that H1X was enriched, unlike other variants, which were depleted, compared with random control samples representing regions of the bulk genome (Fig. 4*C*). This was also observed when H1 occupancy was explored around the center of RNAPII binding sites (Fig. 4*D*). Previously, we reported that H1X is the variant most enriched at DNase-hypersensitive sites and FAIRE regions as well as being asso-

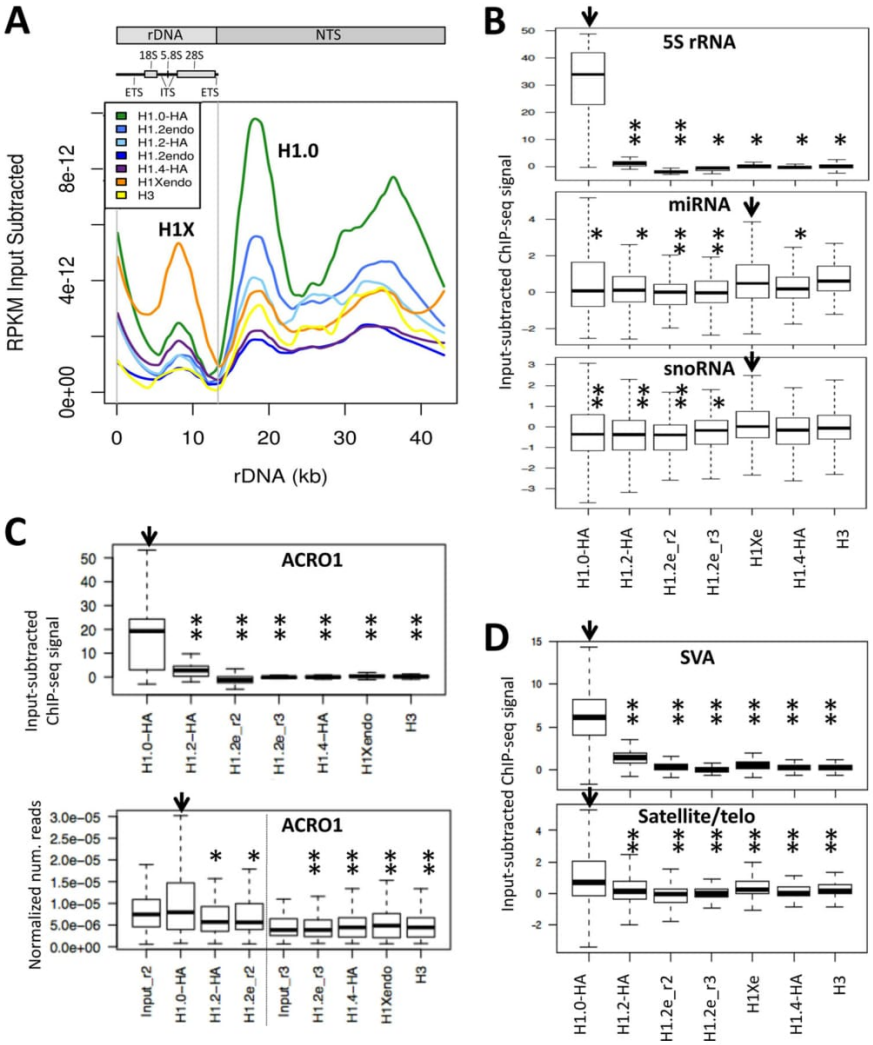


FIGURE 2. **H1 variant abundance at rDNA, non-coding RNA genes, and repetitive DNA.** *A*, alignment of the input-subtracted H1 variant ChIP-seq signal to the human ribosomal DNA complete repeating unit obtained from GenBank™ (U13369.1). To avoid bias in the alignment, the rDNA sequence was added to the human reference genome (hg18), and alignment was only allowed to a single position. A schematic representation of the rDNA repeating unit is shown *above*, with the rRNA transcription unit on the *left* and the non-transcribed spacer (*NTS*) on the *right*. *ITS/ETS*, internal/external transcribed spacer. *B*, box plots showing the occupancy of H1 variants (ChIP-seq signal) within some human non-coding RNA genes: 5 S ribosomal RNA subunit (*rRNA*), microRNA (*miRNA*), and small nucleolar RNA (*snoRNA*). *C*, box plots showing the occupancy of H1 variants (ChIP-seq signal) within human acromeric satellite 1, performing single mapping to the reference genome (*top*) or multiple mapping to a repeat database (RepBase) (*bottom*). *D*, box plots showing the occupancy of H1 variants (ChIP-seq signal) within some human DNA repeats: SINE-VNTR-Alu (*SVA*; non-autonomous, hominid-specific non-LTR retrotransposons) and telomeric satellites. Significance was assessed using the Kolmogorov-Smirnov test. Significant enrichment of H1.0 or H1X (marked with an *arrow*) compared with other variants is marked with *asterisks* (*, $p < 0.05$; **, $p < 0.001$).

ciated with various histone H3 post-translational modifications and the one least depleted at p300 and CTCF sites (41). Additionally, H1X shows the highest correlation with GC content. Taken together, these findings suggest that H1X may have a role in transcriptional regulation of gene expression.

Notably, whereas the H1X content at distal promoters and TSS is lower at active than at inactive genes (an H1 valley), it is higher at coding regions of active genes, especially toward the 3′ end (Fig. 5*A*). This is not true for the other somatic H1 variants (41), the opposite trend being observed for H1.2, and this may be related to the association of H1X with RNAPII described herein. Moreover, H1X is the most abundant variant within exons (Fig. 5*B*). Furthermore, this variant is far more abundant at retained introns than the mean across all introns and is more abundant at included ASEs than excluded exons. Core histone H3 was also found to be more abundant at exons than introns, reflecting the reported higher nucleosome occupancy of exons (56), but this was not seen for all H1 variants (Fig. 5*B*). Overall, H1X is more abundant at exons and intron-exon junctions than within introns themselves and always higher at active genes (Fig. 5*C*). H1X accumulation toward the 3′ end of coding regions and at transcribed exons resembles H3K36me3 distribution (57).

Furthermore, we have reported that H1 variant abundance at CpG islands is heterogeneous across the genome, H1.0 and H1X being clearly overrepresented and H1.2 underrepresented

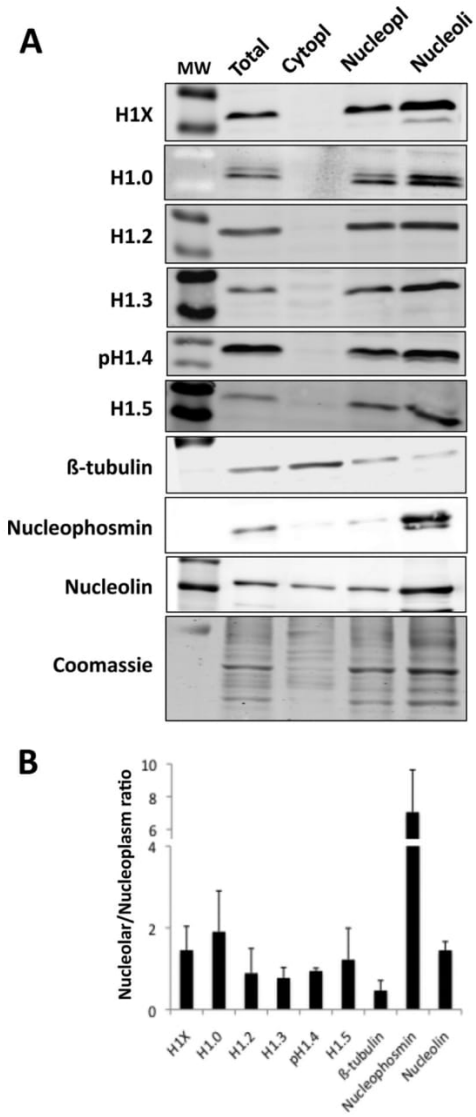


FIGURE 3. **H1 variant distribution within different cell compartments.** *A*, immunoblot of H1 variant abundance in total cell lysate, cytoplasm, nucleoplasm, and nucleolar fractions of T47D cells. 30 μg of protein extract was resolved in SDS-PAGE (10%) and immunoblotted with antibodies specific for H1X, H1.0, H1.2, H1.3, H1.4-T146p, H1.5, β-tubulin, nucleophosmin, and nucleolin. An acrylamide gel was stained with Coomassie solution and used as loading control and for normalization upon quantification of bands. Molecular weight marker (*MW*) bands shown in the H1 blots correspond to 34,000 and 26,000. *B*, representation of the normalized relative units of variants H1X, H1.0, H1.2, H1.3, H1.4, and H1.5, β-tubulin, nucleophosmin, and nucleolin present in the nucleolar fraction divided by the relative units in the nucleoplasm fraction. Immunoblot band intensities were measured using ImageJ version 1.48 software and normalized by Coomassie staining. The means and S.D. values (*error bars*) are shown for three independent fractionation experiments.

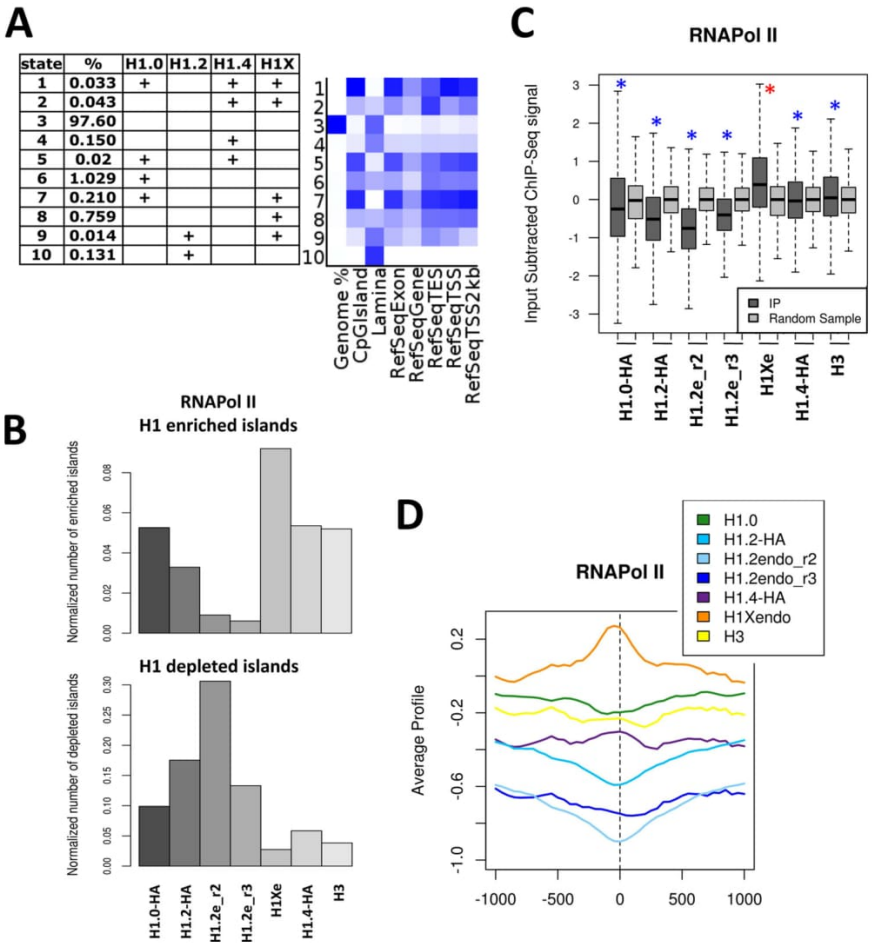
(Fig. 6*A*). Approximately 12% of H1X-enriched regions overlap with CpG sites (41). Next, we investigated the methylation state of CpG islands overlapping H1 variant-enriched or -depleted regions. Although CpG islands overlapping H1-enriched islands are more methylated than those in H1-depleted regions, in agreement with the general consideration that H1-containing chromatin is repressive, H1X-enriched CpGs are less methylated than those overlapping other H1 variants (Fig. 6*B*). In summary, H1X is abundant at CpG islands, but the methylation level of these islands is lower than average, further confirming the relationship of H1X with active transcription. In contrast, H1.2 is disfavored at CpG sites, but the sites at which it is found are highly methylated, related to repressed chromatin.

The pattern of DNA methylation is known to be altered in cancer cells. In general, there is genome-wide intergenic hypomethylation and localized hypermethylation at particular promoters (including tumor-suppressor genes) and CpG-rich and gene-related regions. We have analyzed the occupancy of H1 variants at two subsets of the genome, defined as hyper- or hypomethylated regions in breast cancer cells (T47D) compared with normal human mammary epithelial cells (data obtained from Ruike *et al.* (50)). Hypermethylated regions in cancer cells were enriched in H1.0 and H1X, whereas hypomethylated regions were enriched in H1.2 (Fig. 6*C*). In other words, both sets of analysis confirm the preferential association of H1X with coding regions (hypermethylated in cancer) and active promoters containing hypomethylated CpG islands and the preferential association of H1.2 with intergenic regions (hypomethylated in cancer) and inactive promoters (containing methylated CpGs).

H1X Depletion in Breast Cancer Cells Does Not Alter Proliferation but Does Alter the Expression of Certain Genes—Given the specific association of H1X with RNAPII and active transcription in general, we decided to knock down this variant to explore the effect on cell proliferation and gene expression, in a manner comparable with the inducible depletion of the other five somatic H1 variants that we reported elsewhere (10). We used an inducible shRNA lentiviral expression system based on a Tet-ON strategy (58) to infect T47D and MCF7 breast cancer cells, and stable cell lines were established as described previously (10). Specific H1X depletion upon doxycycline treatment (6 days) was confirmed by immunoblotting, reverse transcription coupled with real-time PCR (RT-qPCR), and immunofluorescence (Fig. 7, *A–C*). No changes in the expression of other H1 variants were detected upon H1X depletion. Moreover, expression of a so-called H1X antisense gene (H1X-AS1) located upstream of the H1X gene in the human genome was not affected either by H1X KD or treatment with trichostatin A, a histone deacetylase inhibitor that induced H1X expression (data not shown), ruling out any functional association between H1X-AS and H1X expression.

Depletion of H1.2 or H1.4 in T47D cells affected cell proliferation and promoted arrest in the G_1 phase of the cell cycle but not the depletion of H1.0, H1.3, or H1.5 (10). H1X-induced depletion slightly slowed down cell proliferation, although the change was not significant (data not shown), and no effect was detected on the cell cycle profile (Fig. 7*D*) in T47D or MCF7 cells compared with untreated cells or control cells expressing random shRNA. Overall, these results suggest that H1X does not play a role in cell proliferation and that this variant is dispensable for the cell lines analyzed under normal growth conditions.

Next, we studied the consequences of H1X depletion on global gene expression using a genome-wide Agilent microarray platform containing ~28,000 mRNAs and 7,400 long noncoding RNAs (see "Experimental Procedures"). T47D cells harboring H1X shRNA were treated or not with doxycycline for 6



state	%	H1.0	H1.2	H1.4	H1X
1	0.033	+		+	+
2	0.043			+	+
3	97.60				
4	0.150			+	
5	0.02	+		+	
6	1.029	+			
7	0.210	+			+
8	0.759				+
9	0.014		+		+
10	0.131		+		

FIGURE 4. **H1X is enriched at RNA polymerase II-enriched regions.** *A*, identification of chromatin states based on the combined presence of H1 variant-enriched regions. The 10 most likely states obtained from a multivariate hidden Markov model are presented in the *table*, listing the frequency of each state. The abundance of different genomic features within each state is shown as a heat map. *Darker blue* corresponds to greater enrichment. *B*, normalized number of H1 variant-enriched and -depleted regions overlapping with RNA polymerase II binding sites (data from T47D cells obtained from Ballaré *et al.* (46)). *C*, box plots showing the occupancy of H1 variants (input-subtracted ChIP-seq signal) within RNA polymerase II binding sites. Significance was assessed using the Kolmogorov-Smirnov test, taking as a control a random sample of bulk genome windows with equal width to the RNA polymerase II sites. Enrichment and depletion are marked with *red* and *blue asterisks*, respectively. *, p value$<$0.001. *D*, mean input-subtracted ChIP-seq signal of H1 variants around the center of RNA polymerase II binding sites.

days and serum-starved for 48 h for synchronization in G_1 prior to RNA extraction to avoid differences between H1 KDs that produce G_1 arrest, as reported elsewhere (10). Applying a -fold change threshold of 1.4 and false discovery rate of $q \leq 0.05$, we found 149 genes to be up-regulated and 45 down-regulated upon H1X KD. The basal expression level, without doxycycline, of up-regulated genes (mean ± S.D. = 8.011 ± 1.99) was lower than the mean expression level across the entire transcriptome (8.723 ± 2.57), whereas that of down-regulated genes was higher than the mean (8.796 ± 1.97) ($p < 0.005$). Some genes, such as *KNG1* and *KRT37* were up-regulated more than 50-fold in H1X KDs (Fig. 8*A*). Gene deregulation by H1X KD was confirmed by RT-qPCR of selected genes in independent samples (Fig. 8*B*). Gene ontology analysis of both up- and down-regulated genes was performed, and interestingly, the most significant functions identified were related to cell movement and transport, and common functions were found between up- and down-regulated genes (data not shown).

Genes Specifically Deregulated by Knockdown of a Particular H1 Variant Are Not Enriched in That Variant at the Promoter—We have previously shown that KD of individual H1 variants deregulates a small subset of genes (≤2%) specific for each variant, including up- or down-regulated genes in similar proportions (10). Here, we have confirmed this for H1X KD. One hypothesis would be that these subsets contain genes specifically targeted by or with prevalence of some of the H1 variants. We explored the occupancy of H1 variant at promoters specifically deregulated by particular variant KDs, and we found no differences in their abundance compared with the mean abundance across all genes (data not shown). For example, the H1X content at distal promoters of genes up- or down-regulated by inducible H1X KD was similar to or at most only slightly lower than the H1X content distribution across all genes (Fig. 8*C*). H1X content was lower at down-regulated genes (high basal expression) than at up-regulated genes (low basal expression), in agreement with the lower H1X content at the distal promoter of

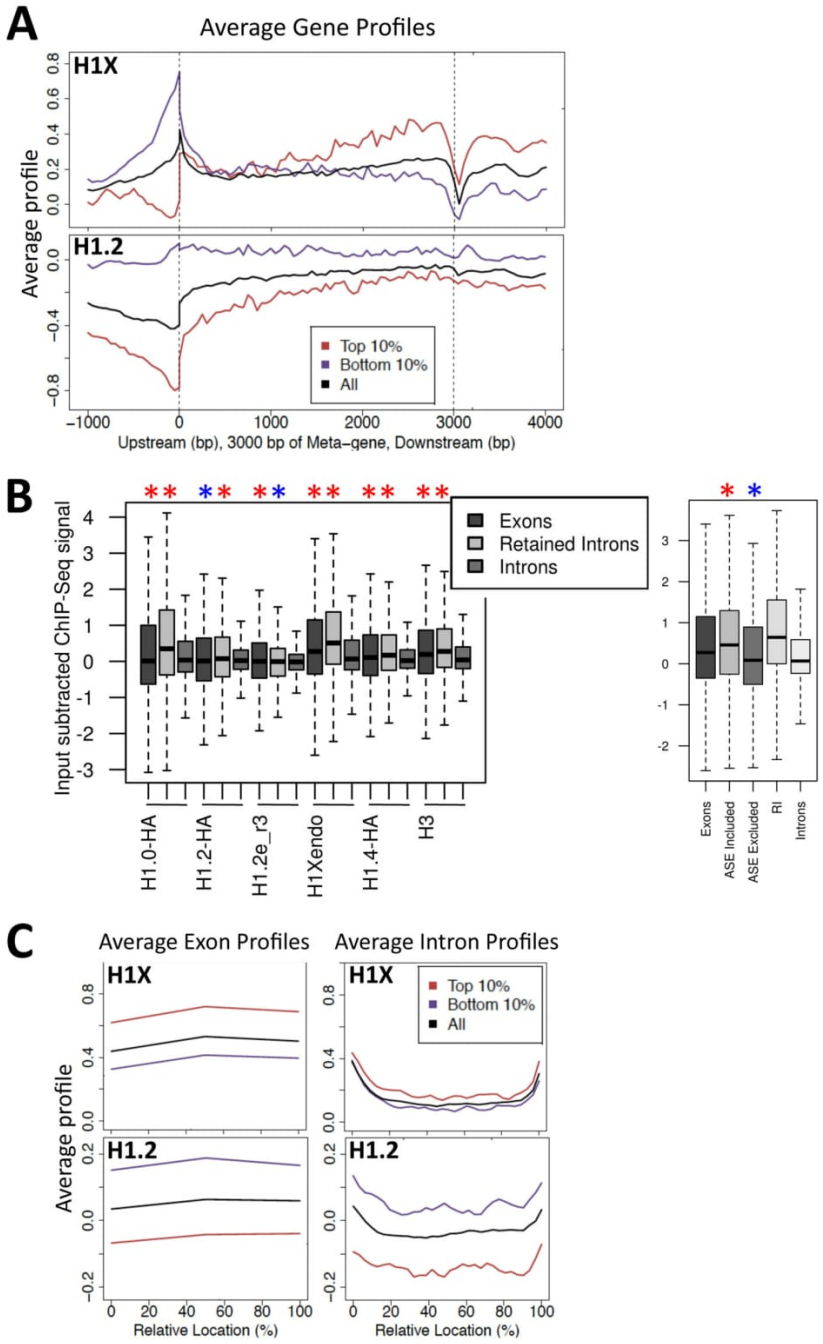


FIGURE 5. **H1X is enriched in constitutive and included alternatively spliced exons within active coding regions.** *A*, H1 abundance (mean input-subtracted ChIP-seq signal) around the body of the top and bottom 10% expressed genes, compared with the mean H1 abundance for all genes (shown in *black*). Gene regions are represented as a 3-kb-long metagene surrounded by a 1-kb region upstream TSS and 1-kb downstream transcription termination site (*TTS*). *B*, box plots showing the occupancy of H1 variants (mean input-subtracted ChIP-seq signal) at exons, retained introns (*RI*), and total introns. In the *right panel*, H1X occupancy is also shown at included and excluded ASEs. Significance was assessed using the Kolmogorov-Smirnov test to compare exons or retained introns with total introns or ASEs with total exons in the *right panel*. Enrichment and depletion is marked with *red* and *blue asterisks*, respectively. *, $p < 0.001$. *C*, H1 abundance as in *A* around exon (*left*) or intron (*right*) profiles.

genes expressed above average (Fig. 5*A*) (41). Up-regulated genes, despite having below average basal expression, do not have above average H1X content (*i.e.* they present less H1X than expected). Moreover, genes dysregulated at H1X KD cells also were observed to have below average H1 content at distal promoters for the other H1 variants (*i.e.* H1.0, H1.2, and H1.4) (Fig. 8*C*).

As mentioned above, whereas the H1X content at the promoter of active genes is lower than at inactive genes, the H1X

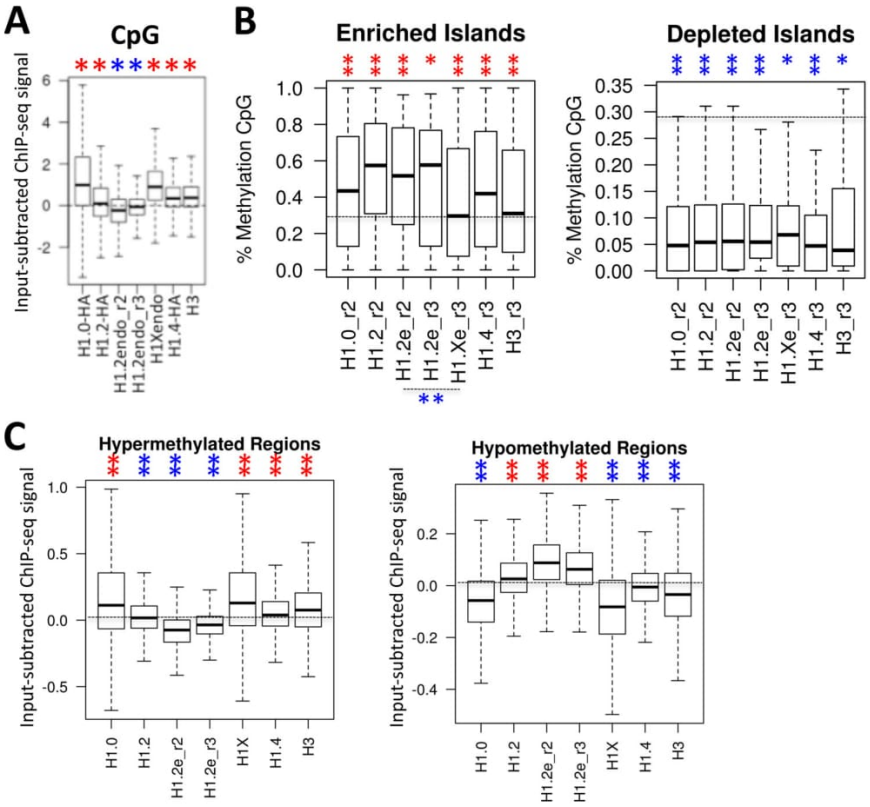


FIGURE 6. **H1 abundance at methylated CpG islands.** *A*, box plots showing the occupancy of H1 variants (mean input-subtracted ChIP-seq signal) at CpG islands. Significance was assessed using the Kolmogorov-Smirnov test taking as a control a random sample of windows with equal width to the CpGs. Enrichment and depletion are marked with *red* and *blue asterisks*, respectively. *, $p < 0.001$. *B*, box plots showing the methylation level of CpG islands overlapping H1 variant-enriched and -depleted regions. Data on the methylation levels at individual CpG islands were calculated by assessing the overlap between the genome methylation levels from Vanderkraats *et al.* (48) and the CpG islands from the UCSC database. Significance was assessed using the Kolmogorov-Smirnov test, taking the methylation level of all CpGs in the genome as a control. Significantly increased or decreased methylation at H1-enriched or -depleted regions is marked with *red* and *blue asterisks*, respectively. Differential methylation between H1X- and H1.2-enriched regions was also tested. *, $p < 0.01$; **, $p < 0.001$. *C*, box plots showing the occupancy of H1 variants (mean input-subtracted ChIP-seq signal) at hypermethylated ($n = 500$) and hypomethylated ($n = 5{,}000$) regions in breast cancer cells (T47D) compared with normal human mammary epithelial cells (data obtained from Ruike *et al.* (50)). Significance was assessed using the Kolmogorov-Smirnov test, taking as a control a random sample of windows with equal width to the hyper- and hypomethylated regions. Enrichment and depletion is marked with *red* and *blue asterisks*, respectively. **, $p < 0.001$.

content at coding regions is higher at active genes. Accordingly, genes down-regulated in H1X KD cells, which present higher basal levels than average or up-regulated genes, showed higher H1X content along their coding region and at collapsed exons and introns (Fig. 8*D*).

Promoters of Up-regulated Genes Become Deprived of H1X upon H1X Knockdown but Do Not Show an H1 Valley or Active Histone Marks—Next, we explored changes in H1 and histone marks at the promoters of six genes up-regulated upon H1X KD by ChIP-qPCR. In all promoters, H1X was removed upon doxycycline treatment of inducible H1X KD cells. In parallel, no significant changes in the promoter occupancy by H1.2 (Fig. 8*E*) or total H1 or H3 (data not shown) were observed. Moreover, these ChIP results confirmed that H1X was not the only H1 variant occupying these H1X-responsive promoters.

We and others have reported elsewhere that there is a valley in H1 occupancy at active promoters compared with that in surrounding regions (40, 41). We compared H1 occupancy at TSS and −3 kb upstream (the distal promoter). Only *UGT2B10* showed a small decrease in H1 at TSS in the absence of doxycycline. This is in agreement with the limited basal expression of these genes, as described above. Nonetheless, these genes were strongly up-regulated upon H1X KD. Hence, an H1 valley was expected under doxycycline treatment, but this was not observed, whereas some of the active genes tested as controls (*CDK2*, *TBKBP1*, and *JUN*) did show an H1 valley at TSS (Fig. 8*E*). Furthermore, enrichment of H3K4me3, a mark of active transcription, was not stronger at TSS of the up-regulated genes under stimulatory conditions, whereas it was present at constitutively active promoters tested as a control (data not shown). In summary, H1X KD up-regulated a limited number of genes in a manner that does not seem to involve regular mechanisms leading to transcription initiation, such as H1 removal and histone H3 Lys-4 methylation at promoters; nor was it linked to removal of a specific H1 variant targeting a promoter for repression.

DISCUSSION

H1.0 Is Enriched at Nucleolus-associated Chromatin—We previously reported that the distribution of H1 variants in the

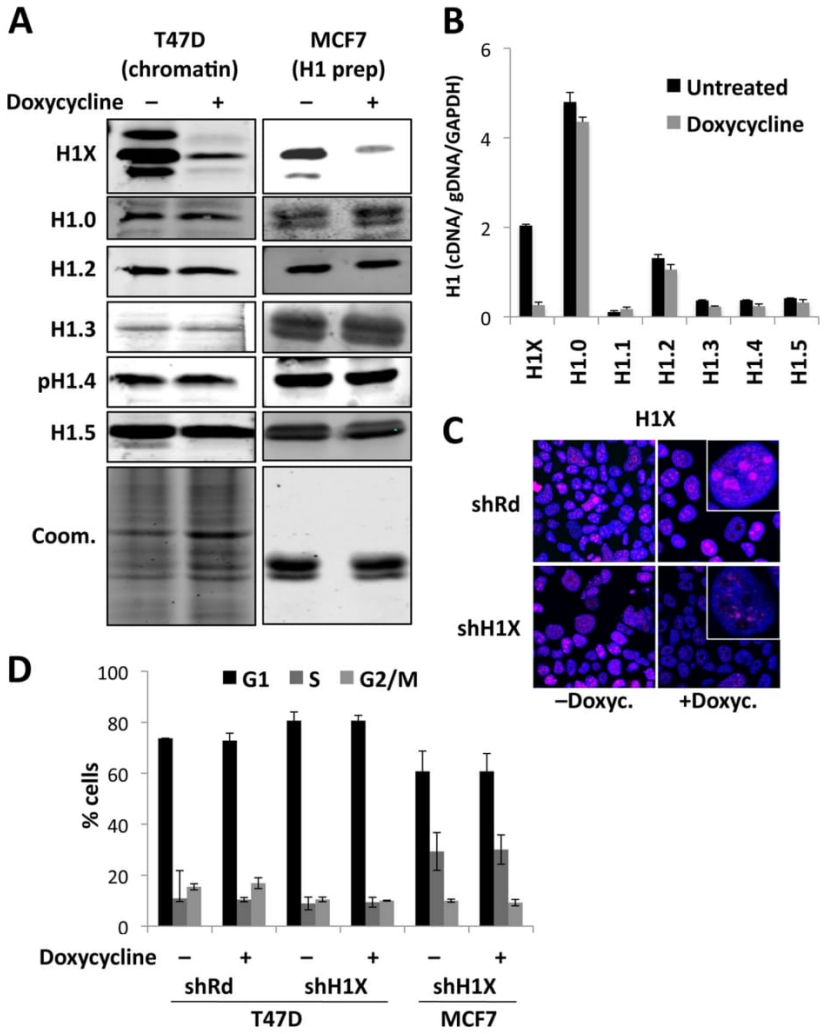
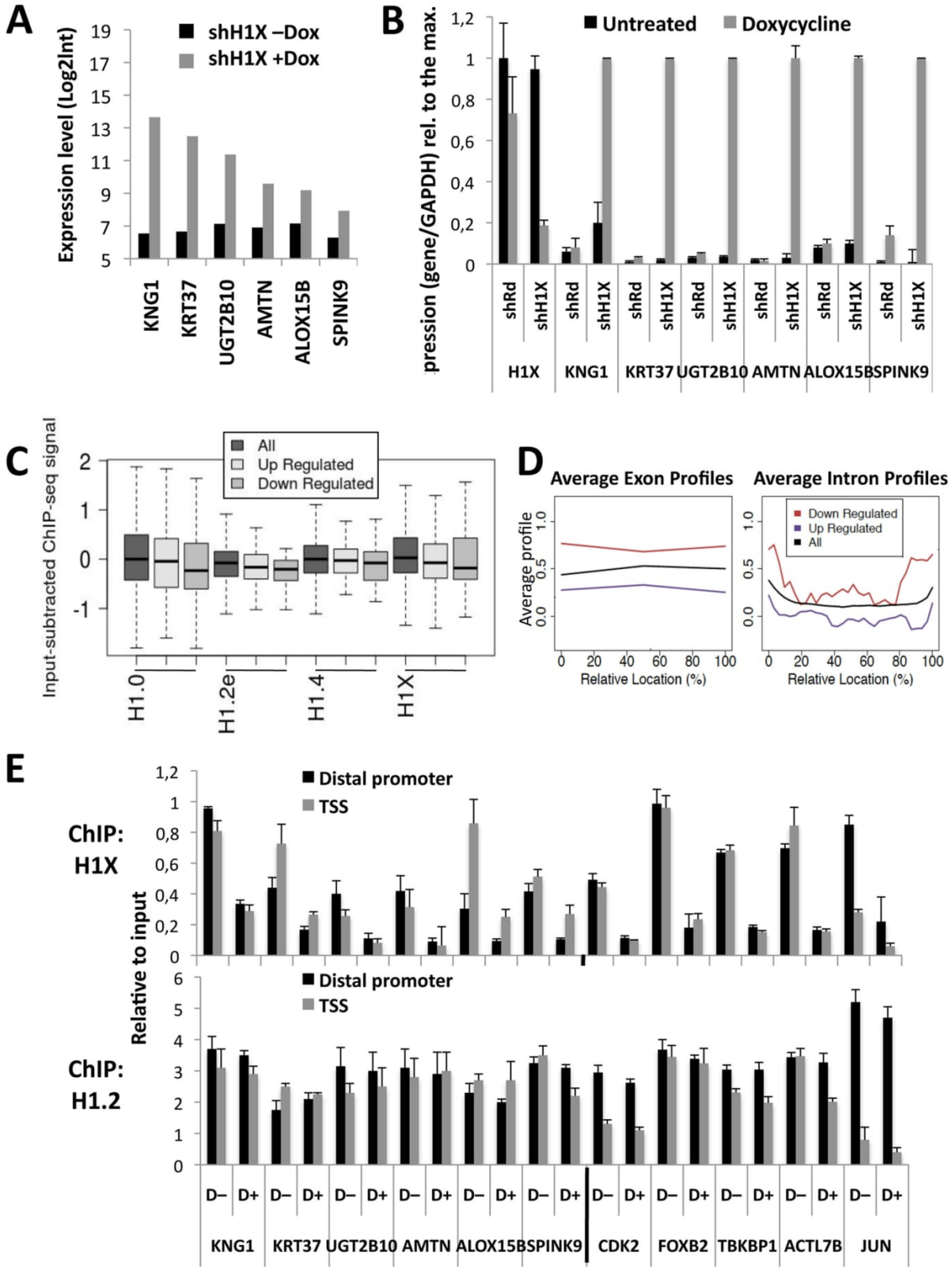


FIGURE 7. **Depletion of histone H1X in breast cancer cells does not alter cell proliferation.** *A*, inducible depletion of H1X in T47D and MCF7 breast cancer cells. T47D- and MCF7-derived cells stably infected with a lentiviral inducible system for the expression of an shRNA against H1X were treated for 6 days with doxycycline or left untreated. Total chromatin or H1 extracts, respectively, were analyzed by immunoblot against H1 variant-specific antibodies or Coomassie-stained as a loading control. Phospho-Thr-146 antibody was used to detect H1.4. *B*, expression of H1 variants after induced depletion of H1X in T47D cells. Reverse transcription real-time PCR of shH1X T47D cells treated or not with doxycycline for 6 days was performed with specific oligonucleotides for H1 variants. Each value was corrected by *GAPDH* and the value of real-time PCR amplification with the same primer set of genomic DNA extracted from the same cell line. *C*, indirect immunofluorescence detection of H1X in T47D cells. T47D cells expressing the shH1X or a random shRNA (*shRd*) as a control, grown in the absence or presence of doxycycline, were fixed and stained with an H1X antibody. DNA was labeled with DAPI. *D*, cell cycle profile after propidium iodide staining of H1X knockdown T47D and MCF7 cells grown for 6 days in the absence or presence of doxycycline. Data are expressed as the percentage of cells in G_1, S, and G_2/M cell cycle phases. The means and S.D. values (*error bars*) for two independent experiments are shown. T47D cells harboring a random shRNA (*shRd*) were used as a control.

breast cancer cell line T47D is not uniform and that H1.2 is the variant that shows the most distinctive pattern. Specifically, H1.2 was found to be non-abundant at genes but enriched at chromosomal domains with low GC content and associated with gene-poor chromosomes, intergenic DNA, and LADs. In contrast, other variants are associated with a relatively high GC content, CpG islands, and gene-rich domains (41). Because not all repetitive DNA is included in the conventional alignment of ChIP-seq data with the human genome, we further investigated the abundance of H1 variants at repetitive features. Strikingly, we found that H1.0 is associated with several repetitive DNA elements related to nucleoli, including ribosomal DNA and acrocentric satellites, as well as NADs. Furthermore, cell fractionation followed by immunoblotting with variant-specific antibodies provided further data suggesting that H1.0 is the variant most enriched at nucleoli.

The nucleolus is the site of ribosome biogenesis and is surrounded by a shell of late replicating condensed heterochromatic DNA (59, 60). DNA associated with this nucleolar chromatin has recently been identified (45, 61) and, in addition to the rDNA repeat units located at human acrocentric chromosomes, includes specific sequences from most chromosomes in a reproducible and heritable manner. NADs have in common a low amount of AT-rich sequence elements, low gene density,

Genomic Distribution of Replication-independent H1 Variants



and enrichment in silent genes. Some of these sequences may also be associated with the nuclear envelope forming the LADs (61). Both the nucleolus-associated and nuclear periphery-associated chromatin domains exhibit common features, being highly condensed and known to replicate preferentially at late stages of the S phase. We have found different H1 variants, H1.0 and H1.2, to be associated with NADs and LADs, respectively, suggesting that these two compartments may have different components, and specific H1 variants may contribute to their organization or stabilization.

H1.0 has been previously reported to be mainly located in chromatin regions that are not affected by micrococcal nuclease digestion, in condensed chromatin, and in perinucleolar regions in certain types of tissue, although it is not fully excluded from active chromatin (54). Recently, a network of proteins interacting with H1.0 in four different cell lines was identified, including splicing factors and proteins involved in rRNA biogenesis, ribosome function/translation, and cellular transport, most of them identified as components of the nucleolus (55). As a consequence, it has been proposed that H1.0 could be a key regulator of nucleolar function and that nucleoli may be the source of the slower exchanging fraction of H1 in the cell. Nonetheless, other H1 variants have also been identified in proteomic profiling of the human nucleolus (62), and phosphorylated H1.2 and H1.4 are associated with RNA polymerase I activity and rRNA biogenesis and have been localized to the nucleolus (63).

It was previously reported that H1X associates with nucleoli in the G_1 phase of the cell cycle (8, 9), although it was suggested that H1X is not located directly at the sites of rDNA transcription but rather at inactive ribosomal genes. Here we also show H1X to be located at nucleoli by immunostaining, but our data indicated that H1.0 is the main nucleolar H1 in the breast cancer cells analyzed. Nonetheless, neither H1.0 nor H1X are restricted to nucleoli, both being found overlapping with somatic H1 variants other than H1.2 throughout the genome of T47D cells (41). H1.0 and, probably, H1X are synthesized independently of DNA replication, and hence, they accumulate when cells stop proliferating and start to differentiate and, consequently, may replace replication-dependent variants. High-throughput mapping of H1.0 and H1X in the genome of differentiated cells has not yet been performed. Additionally, because, for instance, H1.0 content is highly reduced in HeLa cells (data not shown), it would be interesting to analyze whether the association of H1.0 with nucleoli is conserved through different cell types or H1X becomes prominent in other cells as well as whether oncogenic transformation influences the specificity of H1 variant localization.

H1X Associates with RNAPII-enriched Regions, Included Exons, Hypomethylated CpG Islands, and Actively Transcribed Coding Regions—Although histone H1 is often regarded as a basic component of chromatin, growing evidence suggests that particular H1 variants are involved in regulating gene expression at a more specific level. We have found that H1X is the variant that best colocalizes with RNAPII in T47D cells and accumulates at coding regions, mainly exons, of expressed genes. Moreover, we previously reported that expressed genes are devoid of H1, including H1X, at promoters (an H1 valley). Accordingly, the association of H1X with RNAPII might be related to the elongation process, as suggested by the accumulation of H1X toward the 3′ end of coding regions. Besides distinct patterns of expression and localization in different types of tissues and cells, it has also been postulated that H1 associates with regulatory proteins or chromatin components to control their activity. The question of whether H1X is occupying the expected position at nucleosomes at these gene regions or interacts with members of the elongating complex is intriguing. Interestingly, H1X was first found in a two-hybrid screen with the WD40 repeat region of the transcription factor TFIID as the bait (64), although this association was not further explored functionally. In relation to this, it has recently been described that H1.2 functionally interacts with Cul4A E3 ubiquitin ligase, PAF1 elongation complexes, and the serine 2-phosphorylated form of RNAPII that potentiates core histone modifications and targets gene transcriptional elongation in HeLa cells (65). Moreover, in the same study, WDR5, a substrate adaptor for Cul4A E3 ligase, was found to co-purify with six of the somatic H1 variants (H1.0 to H1.5); H1X was not explored. H1X playing a role in T47D cells similar to that of H1.2 in HeLa, while H1.2 is excluded from actively transcribed regions in T47D cells, is an intriguing possibility, compatible with the view that the distinct patterns of expression and localization of H1 variants in different types of tissues and cell types may provide an important regulatory mechanism of gene expression.

H1X accumulation toward the 3′ end of coding regions and at transcribed exons resembles H3K36me3 distribution. H3K36me3 is co-transcriptionally deposited in a splicing-dependent manner and represses internal initiation. Moreover, this histone mark seems to associate with exons included by the splicing machinery more than with ASEs (56, 57, 66, 67). Included exons show higher nucleosome occupancy, either to protect fidelity or to slow down transcription to ensure inclu-

FIGURE 8. **Genes deregulated by H1X KD are not enriched in specific H1 variants.** *A*, expression levels of the top six up-regulated genes in H1X knockdown T47D cells treated or not with doxycycline for 6 days. Expression data were obtained by hybridization with an Agilent microarray in duplicate, and log2 values are represented. *B*, expression of genes up-regulated upon H1X KD measured by real-time PCR in H1X and random shRNA-expressing cells treated or not with doxycycline for 6 days. Expression of H1X was measured to test its inhibition by the inducible shRNA. *GAPDH* was measured for normalization. Expression data are presented relative to the maximal value for each gene. The means and S.D. values (*error bars*) are shown from a representative experiment measured in duplicate. *C*, box plots of H1.0, H1.2, H1.4, and H1X abundance (input-subtracted ChIP-seq signal) at distal promoter regions (−3,200 to −2,000 bp relative to TSS) for the genes up- or down-regulated upon H1X knockdown, compared with the H1 abundance of total genes. *D*, H1X abundance (mean input-subtracted ChIP-seq signal) around exon (*left*) or intron (*right*) profiles of genes up- or down-regulated upon H1X knockdown, compared with the mean H1 abundance for all genes (shown in *black*). *E*, H1X and H1.2 abundance at the up-regulated gene promoters in H1X KD cells. ChIP with specific antibodies for H1X and H1.2 was performed in H1X KD T47D cells treated (+*D*) or not (−*D*) with doxycycline, and the abundance of immunoprecipitated material was quantified by real-time PCR with oligonucleotides for the indicated promoters (−3 kb distal promoter or TSS) and corrected by input DNA amplification with the same primer pair. Genes that did not change their expression in the H1X KD microarray were also analyzed for comparison. The means and S.D. values are shown from a representative experiment measured in duplicate.

sion. It would be interesting to investigate a hypothetical association of H1X with factors that set or recognize H3K36me3 and to investigate whether H1X plays a specific role in marking or protecting exons to be expressed. Alternatively, H1X-mediated compaction could slow RNAPII elongation, favoring alternative exon inclusion. In relation to this, we have found that H1X is more abundant at included ASEs than at excluded exons. Interestingly, H1X is also enriched in retained introns. It has also been reported that DNA methylation is enriched in included exons and that inhibition of DNA methylation, as well as inhibition of histone deacetylase activity, results in aberrant alternative splicing (68). Regarding our observations, H1X could be an additional player in the functional interconnections between chromatin structure, transcriptional elongation, and splicing, raising the intriguing possibility of the existence of an epigenetic memory for splicing patterns that could be inherited.

We have shown that CpG islands overlapping H1-enriched regions are hypermethylated compared with the mean for CpG islands across the genome, whereas those overlapping H1-depleted regions are hypomethylated, further supporting the general notion that H1-containing chromatin is repressive, at least at gene promoters where most of the CpG islands are located. Similar results were reported by Izzo *et al.* (39). Interestingly, there are differences between variants, in agreement with further observations relating H1X to active transcription and H1.2 to repressive chromatin. In other words, H1X is more abundant than H1.2 at CpG islands, and those islands coinciding with H1X enrichment are significantly less methylated than H1.2-occupied islands. It remains to be explored in human cells whether there is a direct interplay between H1 variants and DNA methylation at CpG islands or just co-localization of features related to active or repressed chromatin. In mouse embryonic stem cells, it has been reported that there is interaction of DNA methyltransferases with H1 variants and recruitment to two imprinting control loci for their repression, except for H1c (H1.2) that did not interact (69). Because H1.2 is disfavored at coding regions, according to our data, it is plausible that it is not involved in the regulation of gene expression by associating to DNA methyltransferases, but further investigation is required to clarify this hypothetical interplay between histone H1 and DNA methylation.

Cancer cells have an altered methylation pattern compared with that in healthy cells, namely general hypomethylation and localized hypermethylation at certain promoters containing CpG islands. By analyzing the H1 variant content at regions described to be significantly hypo- or hypermethylated in T47D cells compared with normal mammary epithelial cells, we found hypomethylated DNA to be enriched in H1.2 and found enrichment of H1.0 and H1X at hypermethylated DNA. In turn, regions hypomethylated in cancer have been related to repressive chromatin, transcriptional inactivation, and large genomic structures, such as LADs (70–72), features associated with H1.2 according to our results. Thus, we hypothesize that different H1 variants may be involved in establishing or maintaining altered DNA methylation patterns in the course of cancer. At the same time, we have shown that different H1 variants are associated with different nuclear chromatin structures, such as LADs or NADs, which are developmentally regulated and altered in cancer (73, 74). The view that the origin of cancer may lay in an epigenetic dysregulation that would increase transcriptional noise, variability, and gene expression plasticity is gaining strength (75). Altered expression of H1 variants during the onset of cancer (31, 32) could participate in global chromatin rearrangements by altering the formation of such chromatin structures or chromatin spatial interactions.

H1X Is Dispensable in Breast Cancer Cells but Alters Gene Expression through an Unknown Mechanism—We reported elsewhere that H1.2 and H1.4 KD in T47D cells slowed cell proliferation and, in the H1.2 KD at least, caused the arrest of the cell cycle in G_1 (10). Using the same methodology, no growth phenotype was observed for H1.0, H1.3, or H1.5 KD, as observed here for H1X KD in T47D and MCF7 cells. It is noteworthy that H1X expression was detected not only in T47D and MCF7, but also in HeLa, 293T, and Jurkat cells, all of the cell lines tested to date (data not shown) (7). Elsewhere, H1X was found to be expressed in all tissues examined (64). This is also true for H1.2 and H1.4 but not for the other variants (H1.1 being undetected in most cell types, whereas there is a low level of expression of H1.0 and H1.3 in HeLa and of H1.5 in 293T) and can be considered an indication of the prominent role of H1.2 and H1.4 in human cells. Moreover, regulation of H1X expression differs not only from that of the replication-dependent variants but also from that of replication-independent H1.0 (7). Taken together, these observations suggest that H1X may have a specific and prominent role among H1 functions in human cells, although this has not become apparent upon H1X KD. One possibility could be that shRNA-mediated H1X depletion was not complete or that other H1 variants undertake some of its functions or localization. For this, H1.0 seems to be the stronger candidate, because the two variants show considerable overlap throughout our analysis despite specific enrichment at RNAPII binding regions and nucleolar chromatin, respectively. As yet, no attempt has been made to develop a double H1.0/H1X KD, but such an approach would clarify this possibility.

H1X KD affected expression of a small number of genes, either up- or down-regulated, as observed when some of the other H1 variants were knocked down with the doxycycline-inducible system in the same cells (10). Moreover, small changes in gene expression upon H1 depletion have also been reported in many other systems, including knock-out mice (11–13). Because H1X localization is enriched at RNAPII sites and at gene exons and correlates with high expression within the body of genes, we could expect that H1X KD would have a larger effect on global gene expression than what we observed. Alternatively, H1X could play a role in the coupling of elongation and splicing. Again, insufficient depletion or redundancy between variants could explain the lack of major effects.

Our hypothesis was that genes dysregulated upon H1 KD might have a higher content of that particular H1 variant at the promoter and, consequently, be more exposed to its depletion if H1 plays a regulatory role, but this was not confirmed. Instead, the H1X (and general H1) content under uninduced conditions of genes up-regulated upon H1X KD was lower than expected according to its low basal expression level. Moreover, the con-

tent of H1X along the coding region of these genes was normal, and the abundance of H1X-enriched or -depleted islands within the coding region of genes down-regulated in H1X KD was not significantly different from the abundance of other H1 variants (data not shown). ChIP showed that variants other than H1X were present at the promoter of up-regulated genes, and their abundance did not increase upon H1X depletion.

Surprisingly, genes up-regulated upon H1X KD did not present features of promoter clearance and transcription initiation, such as an H1 valley or increased methylation of histone H3 Lys-4 at TSS, and this cannot be attributed to a lack of nucleosomes because H1 and H3 were clearly observed. Expression of genes in the absence of histone marking has been reported elsewhere (76–78). Maybe H1X clearance is sufficient for these particular promoters to allow recruitment of transcription factors and machinery. Alternatively, gene induction is not caused by increased initiation but rather by a later step that was blocked by H1X, perhaps related to its hypothetical role in elongation or splicing. There is no doubt that further studies are required to understand how particular H1 variants regulate a limited subset of genes and whether this relates to the specific localization of variants across genomic features, chromatin domains, and nuclear territories that we are starting to envisage.

REFERENCES

1. Thoma, F., Koller, T., and Klug, A. (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell Biol.* **83,** 403–427
2. Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J., and Woodcock, C. L. (1998) Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 14173–14178
3. Pennings, S., Meersseman, G., and Bradbury, E. M. (1994) Linker histones H1 and H5 prevent the mobility of positioned nucleosomes. *Proc. Natl. Acad. Sci. U.S.A.* **91,** 10275–10279
4. Laybourn, P. J., and Kadonaga, J. T. (1991) Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. *Science* **254,** 238–245
5. Happel, N., and Doenecke, D. (2009) Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene* **431,** 1–12
6. Izzo, A., Kamieniarz, K., and Schneider, R. (2008) The histone H1 family: specific members, specific functions? *Biol. Chem.* **389,** 333–343
7. Happel, N., Schulze, E., and Doenecke, D. (2005) Characterisation of human histone H1x. *Biol. Chem.* **386,** 541–551
8. Stoldt, S., Wenzel, D., Schulze, E., Doenecke, D., and Happel, N. (2007) G1 phase-dependent nucleolar accumulation of human histone H1x. *Biol. Cell.* **99,** 541–552
9. Takata, H., Matsunaga, S., Morimoto, A., Ono-Maniwa, R., Uchiyama, S., and Fukui, K. (2007) H1.X with different properties from other linker histones is required for mitotic progression. *FEBS Lett.* **581,** 3783–3788
10. Sancho, M., Diani, E., Beato, M., and Jordan, A. (2008) Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. *PLoS Genet.* **4,** e1000227
11. Fan, Y., Nikitina, T., Zhao, J., Fleury, T. J., Bhattacharyya, R., Bouhassira, E. E., Stein, A., Woodcock, C. L., and Skoultchi, A. I. (2005) Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* **123,** 1199–1212
12. Shen, X., and Gorovsky, M. A. (1996) Linker histone H1 regulates specific gene expression but not global transcription *in vivo*. *Cell* **86,** 475–483
13. Lin, Q., Inselman, A., Han, X., Xu, H., Zhang, W., Handel, M. A., and Skoultchi, A. I. (2004) Reductions in linker histone levels are tolerated in developing spermatocytes but cause changes in specific gene expression. *J. Biol. Chem.* **279,** 23525–23535
14. Fan, Y., Sirotkin, A., Russell, R. G., Ayala, J., and Skoultchi, A. I. (2001) Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the H1(0) replacement subtype. *Mol. Cell Biol.* **21,** 7933–7943
15. Daujat, S., Zeissler, U., Waldmann, T., Happel, N., and Schneider, R. (2005) HP1 binds specifically to Lys^{26}-methylated histone H1.4, whereas simultaneous Ser^{27} phosphorylation blocks HP1 binding. *J. Biol. Chem.* **280,** 38090–38095
16. Hale, T. K., Contreras, A., Morrison, A. J., and Herrera, R. E. (2006) Phosphorylation of the linker histone H1 by CDK regulates its binding to HP1α. *Mol. Cell* **22,** 693–699
17. Hergeth, S. P., Dundr, M., Tropberger, P., Zee, B. M., Garcia, B. A., Daujat, S., and Schneider, R. (2011) Isoform-specific phosphorylation of human linker histone H1.4 in mitosis by the kinase Aurora B. *J. Cell Sci.* **124,** 1623–1628
18. Kamieniarz, K., Izzo, A., Dundr, M., Tropberger, P., Ozretic, L., Kirfel, J., Scheer, E., Tropel, P., Wisniewski, J. R., Tora, L., Viville, S., Buettner, R., and Schneider, R. (2012) A dual role of linker histone H1.4 Lys 34 acetylation in transcriptional activation. *Genes Dev.* **26,** 797–802
19. Kim, K., Jeong, K. W., Kim, H., Choi, J., Lu, W., Stallcup, M. R., and An, W. (2012) Functional interplay between p53 acetylation and H1.2 phosphorylation in p53-regulated transcription. *Oncogene* **31,** 4290–4301
20. Kuzmichev, A., Jenuwein, T., Tempst, P., and Reinberg, D. (2004) Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol. Cell* **14,** 183–193
21. Lee, H., Habas, R., and Abate-Shen, C. (2004) MSX1 cooperates with histone H1b for inhibition of transcription and myogenesis. *Science* **304,** 1675–1678
22. Vaquero, A., Scher, M., Lee, D., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2004) Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. *Mol. Cell* **16,** 93–105
23. Weiss, T., Hergeth, S., Zeissler, U., Izzo, A., Tropberger, P., Zee, B. M., Dundr, M., Garcia, B. A., Daujat, S., and Schneider, R. (2010) Histone H1 variant-specific lysine methylation by G9a/KMT1C and Glp1/KMT1D. *Epigenetics Chromatin* **3,** 7
24. Terme, J. M., Millán-Ariño, L., Mayor, R., Luque, N., Izquierdo-Bouldstridge, A., Bustillos, A., Sampaio, C., Canes, J., Font, I., Sima, N., Sancho, M., Torrente, L., Forcales, S., Roque, A., Suau, P., and Jordan, A. (2014) Dynamics and dispensability of variant-specific histone H1 Lys-26/Ser-27 and Thr-165 post-translational modifications. *FEBS Lett.* **588,** 2353–2362
25. Zhang, Y., Cooke, M., Panjwani, S., Cao, K., Krauth, B., Ho, P. Y., Medrzycki, M., Berhe, D. T., Pan, C., McDevitt, T. C., and Fan, Y. (2012) Histone h1 depletion impairs embryonic stem cell differentiation. *PLoS Genet.* **8,** e1002691
26. Lennox, R. W., and Cohen, L. H. (1983) The histone H1 complements of dividing and nondividing cells of the mouse. *J. Biol. Chem.* **258,** 262–268
27. Piña, B., Martínez, P., and Suau, P. (1987) Changes in H1 complement in differentiating rat-brain cortical neurons. *Eur. J. Biochem.* **164,** 71–76
28. Meergans, T., Albig, W., and Doenecke, D. (1997) Varied expression patterns of human H1 histone genes in different cell lines. *DNA Cell Biol.* **16,** 1041–1049
29. Parseghian, M. H., and Hamkalo, B. A. (2001) A compendium of the histone H1 family of somatic subtypes: an elusive cast of characters and their characteristics. *Biochem. Cell Biol.* **79,** 289–304
30. Terme, J. M., Sesé, B., Millán-Ariño, L., Mayor, R., Izpisúa Belmonte, J. C., Barrero, M. J., and Jordan, A. (2011) Histone H1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency. *J. Biol. Chem.* **286,** 35347–35357
31. Sato, S., Takahashi, S., Asamoto, M., Nakanishi, M., Wakita, T., Ogura, Y., Yatabe, Y., and Shirai, T. (2012) Histone H1 expression in human prostate cancer tissues and cell lines. *Pathol. Int.* **62,** 84–92
32. Medrzycki, M., Zhang, Y., McDonald, J. F., and Fan, Y. (2012) Profiling of linker histone variants in ovarian cancer. *Front. Biosci.* **17,** 396–406
33. Sjöblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., Mandelker, D., Leary, R. J., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S. D., Willis, J., Dawson, D., Willson, J. K., Gazdar, A. F., Hartigan, J., Wu, L., Liu, C., Parmigiani, G., Park, B. H.,

Genomic Distribution of Replication-independent H1 Variants

Bachman, K. E., Papadopoulos, N., Vogelstein, B., Kinzler, K. W., and Velculescu, V. E. (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* **314,** 268–274
34. Telu, K. H., Abbaoui, B., Thomas-Ahner, J. M., Zynger, D. L., Clinton, S. K., Freitas, M. A., and Mortazavi, A. (2013) Alterations of histone H1 phosphorylation during bladder carcinogenesis. *J. Proteome Res.* **12,** 3317–3326
35. Orthaus, S., Klement, K., Happel, N., Hoischen, C., and Diekmann, S. (2009) Linker histone H1 is present in centromeric chromatin of living human cells next to inner kinetochore proteins. *Nucleic Acids Res.* **37,** 3391–3406
36. Th'ng, J. P., Sung, R., Ye, M., and Hendzel, M. J. (2005) H1 family histones in the nucleus: control of binding and localization by the C-terminal domain. *J. Biol. Chem.* **280,** 27809–27814
37. Parseghian, M. H., Newcomb, R. L., Winokur, S. T., and Hamkalo, B. A. (2000) The distribution of somatic H1 subtypes is non-random on active *versus* inactive chromatin: distribution in human fetal fibroblasts. *Chromosome Res.* **8,** 405–424
38. Cao, K., Lailler, N., Zhang, Y., Kumar, A., Uppal, K., Liu, Z., Lee, E. K., Wu, H., Medrzycki, M., Pan, C., Ho, P. Y., Cooper, G. P., Jr., Dong, X., Bock, C., Bouhassira, E. E., and Fan, Y. (2013) High-resolution mapping of h1 linker histone variants in embryonic stem cells. *PLoS Genet.* **9,** e1003417
39. Izzo, A., Kamieniarz-Gdula, K., Ramírez, F., Noureen, N., Kind, J., Manke, T., van Steensel, B., and Schneider, R. (2013) The genomic landscape of the somatic linker histone subtypes H1.1 to H1.5 in human cells. *Cell Rep* **3,** 2142–2154
40. Krishnakumar, R., Gamble, M. J., Frizzell, K. M., Berrocal, J. G., Kininis, M., and Kraus, W. L. (2008) Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. *Science* **319,** 819–821
41. Millán-Ariño, L., Islam, A. B., Izquierdo-Bouldstridge, A., Mayor, R., Terme, J. M., Luque, N., Sancho, M., López-Bigas, N., and Jordan, A. (2014) Mapping of six somatic linker histone H1 variants in human breast cancer cells uncovers specific features of H1.2. *Nucleic Acids Res.* **42,** 4474–4493
42. Truss, M., Bartsch, J., Schelbert, A., Haché, R. J., and Beato, M. (1995) Hormone induces binding of receptors and transcription factors to a rearranged nucleosome on the MMTV promoter *in vivo*. *EMBO J.* **14,** 1737–1751
43. Andersen, J. S., Lyon, C. E., Fox, A. H., Leung, A. K., Lam, Y. W., Steen, H., Mann, M., and Lamond, A. I. (2002) Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **12,** 1–11
44. Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., Eussen, B. H., de Klein, A., Wessels, L., de Laat, W., and van Steensel, B. (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453,** 948–951
45. Németh, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Péterfia, B., Solovei, I., Cremer, T., Dopazo, J., and Längst, G. (2010) Initial genomics of the human nucleolus. *PLoS Genet.* **6,** e1000889
46. Ballaré, C., Castellano, G., Gaveglia, L., Althammer, S., González-Vallinas, J., Eyras, E., Le Dily, F., Zaurin, R., Soronellas, D., Vicent, G. P., and Beato, M. (2013) Nucleosome-driven transcription factor binding and gene regulation. *Mol. Cell* **49,** 67–79
47. Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110,** 462–467
48. Vanderkraats, N. D., Hiken, J. F., Decker, K. F., and Edwards, J. R. (2013) Discovering high-resolution patterns of differential DNA methylation that correlate with gene expression changes. *Nucleic Acids Res.* **41,** 6816–6827
49. Quinlan, A. R., and Hall, I. M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842
50. Ruike, Y., Imanaka, Y., Sato, F., Shimizu, K., and Tsujimoto, G. (2010) Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. *BMC Genomics* **11,** 137
51. Zentner, G. E., Saiakhova, A., Manaenkov, P., Adams, M. D., and Scacheri, P. C. (2011) Integrative genomic analysis of human ribosomal DNA. *Nucleic Acids Res.* **39,** 4949–4960
52. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25
53. Ernst, J., and Kellis, M. (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9,** 215–216
54. Gorka, C., Fakan, S., and Lawrence, J. J. (1993) Light and electron microscope immunocytochemical analyses of histone H1(0) distribution in the nucleus of Friend erythroleukemia cells. *Exp. Cell Res.* **205,** 152–158
55. Kalashnikova, A. A., Winkler, D. D., McBryant, S. J., Henderson, R. K., Herman, J. A., DeLuca, J. G., Luger, K., Prenni, J. E., and Hansen, J. C. (2013) Linker histone H1.0 interacts with an extensive network of proteins found in the nucleolus. *Nucleic Acids Res.* **41,** 4026–4035
56. Hnilicová, J., and Staněk, D. (2011) Where splicing joins chromatin. *Nucleus* **2,** 182–188
57. Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X. S., and Ahringer, J. (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* **41,** 376–381
58. Wiznerowicz, M., and Trono, D. (2003) Conditional suppression of cellular genes: lentivirus vector-mediated drug-inducible RNA interference. *J. Virol.* **77,** 8957–8961
59. Sadoni, N., Langer, S., Fauth, C., Bernardi, G., Cremer, T., Turner, B. M., and Zink, D. (1999) Nuclear organization of mammalian genomes: polar chromosome territories build up functionally distinct higher order compartments. *J. Cell Biol.* **146,** 1211–1226
60. Ferreira, J., Paolella, G., Ramos, C., and Lamond, A. I. (1997) Spatial organization of large-scale chromatin domains in the nucleus: a magnified view of single chromosome territories. *J. Cell Biol.* **139,** 1597–1610
61. van Koningsbruggen, S., Gierlinski, M., Schofield, P., Martin, D., Barton, G. J., Ariyurek, Y., den Dunnen, J. T., and Lamond, A. I. (2010) High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Mol. Biol. Cell* **21,** 3735–3748
62. Jarboui, M. A., Wynne, K., Elia, G., Hall, W. W., and Gautier, V. W. (2011) Proteomic profiling of the human T-cell nucleolus. *Mol. Immunol.* **49,** 441–452
63. Zheng, Y., John, S., Pesavento, J. J., Schultz-Norton, J. R., Schiltz, R. L., Baek, S., Nardulli, A. M., Hager, G. L., Kelleher, N. L., and Mizzen, C. A. (2010) Histone H1 phosphorylation is associated with transcription by RNA polymerases I and II. *J. Cell Biol.* **189,** 407–415
64. Yamamoto, T., and Horikoshi, M. (1996) Cloning of the cDNA encoding a novel subtype of histone H1. *Gene* **173,** 281–285
65. Kim, K., Lee, B., Kim, J., Choi, J., Kim, J. M., Xiong, Y., Roeder, R. G., and An, W. (2013) Linker Histone H1.2 cooperates with Cul4A and PAF1 to drive H4K31 ubiquitylation-mediated transactivation. *Cell Rep.* **5,** 1690–1703
66. Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R., and Blencowe, B. J. (2013) Dynamic integration of splicing within gene regulatory pathways. *Cell* **152,** 1252–1269
67. Gómez Acuña, L. I., Fiszbein, A., Alló, M., Schor, I. E., and Kornblihtt, A. R. (2013) Connections between chromatin signatures and splicing. *Wiley Interdiscip Rev. RNA* **4,** 77–91
68. Maunakea, A. K., Chepelev, I., Cui, K., and Zhao, K. (2013) Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* **23,** 1256–1269
69. Yang, S. M., Kim, B. J., Norwood Toro, L., and Skoultchi, A. I. (2013) H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 1708–1713
70. Hon, G. C., Hawkins, R. D., Caballero, O. L., Lo, C., Lister, R., Pelizzola, M., Valsesia, A., Ye, Z., Kuan, S., Edsall, L. E., Camargo, A. A., Stevenson, B. J., Ecker, J. R., Bafna, V., Strausberg, R. L., Simpson, A. J., and Ren, B. (2012) Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* **22,** 246–258
71. Berman, B. P., Weisenberger, D. J., Aman, J. F., Hinoue, T., Ramjan, Z., Liu, Y., Noushmehr, H., Lange, C. P., van Dijk, C. M., Tollenaar, R. A., Van Den Berg, D., and Laird, P. W. (2012) Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.* **44,** 40–46
72. Hansen, K. D., Timp, W., Bravo, H. C., Sabunciyan, S., Langmead, B.,

McDonald, O. G., Wen, B., Wu, H., Liu, Y., Diep, D., Briem, E., Zhang, K., Irizarry, R. A., and Feinberg, A. P. (2011) Increased methylation variation in epigenetic domains across cancer types. *Nat. Genet.* **43,** 768–775

73. Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S. W., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R. M., van Lohuizen, M., Reinders, M., Wessels, L., and van Steensel, B. (2010) Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38,** 603–613
74. McDonald, O. G., Wu, H., Timp, W., Doi, A., and Feinberg, A. P. (2011) Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition. *Nat. Struct. Mol. Biol.* **18,** 867–874
75. Timp, W., and Feinberg, A. P. (2013) Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nat. Rev. Cancer* **13,** 497–510
76. Hodl, M., and Basler, K. (2012) Transcription in the absence of histone H3.2 and H3K4 methylation. *Curr. Biol.* **22,** 2253–2257
77. Nègre, N., Brown, C. D., Ma, L., Bristow, C. A., Miller, S. W., Wagner, U., Kheradpour, P., Eaton, M. L., Loriaux, P., Sealfon, R., Li, Z., Ishii, H., Spokony, R. F., Chen, J., Hwang, L., Cheng, C., Auburn, R. P., Davis, M. B., Domanus, M., Shah, P. K., Morrison, C. A., Zieba, J., Suchy, S., Senderowicz, L., Victorsen, A., Bild, N. A., Grundstad, A. J., Hanley, D., MacAlpine, D. M., Mannervik, M., Venken, K., Bellen, H., White, R., Gerstein, M., Russell, S., Grossman, R. L., Ren, B., Posakony, J. W., Kellis, M., and White, K. P. (2011) A cis-regulatory map of the Drosophila genome. *Nature* **471,** 527–531
78. Zhang, H., Gao, L., Anandhakumar, J., and Gross, D. S. (2014) Uncoupling transcription from covalent histone modification. *PLoS Genet.* **10,** e1004202

CHAPTER II

Nucleic Acids Research, 2017 **1**
doi: 10.1093/nar/gkx746

Histone H1 depletion triggers an interferon response in cancer cells via activation of heterochromatic repeats

Andrea Izquierdo-Bouldstridge[1,†], Alberto Bustillos[1,†], Carles Bonet-Costa[1], Patricia Aribau-Miralbés[1], Daniel García-Gomis[1], Marc Dabad[2,3], Anna Esteve-Codina[2,3], Laura Pascual-Reguant[4], Sandra Peiró[4], Manel Esteller[5,6,7], Matthew Murtha[5], Lluís Millán-Ariño[1] and Albert Jordan[1,*]

[1]Institut de Biologia Molecular de Barcelona (IBMB-CSIC), Barcelona, Catalonia 08028, Spain, [2]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Catalonia 08028, Spain, [3]Universitat Pompeu Fabra (UPF), Barcelona, Catalonia 08003, Spain, [4]Vall d'Hebron Institute of Oncology, Barcelona, Catalonia 08035, Spain, [5]Cancer Epigenetics and Biology Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet, Barcelona, Catalonia 08028, Spain, [6]Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona (UB), Catalonia 08028, Spain and [7]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia 08028, Spain

Received May 26, 2017; Revised August 10, 2017; Editorial Decision August 11, 2017; Accepted August 15, 2017

ABSTRACT

Histone H1 has seven variants in human somatic cells and contributes to chromatin compaction and transcriptional regulation. Knock-down (KD) of each H1 variant in breast cancer cells results in altered gene expression and proliferation differently in a variant specific manner with H1.2 and H1.4 KDs being most deleterious. Here we show combined depletion of H1.2 and H1.4 has a strong deleterious effect resulting in a strong interferon (IFN) response, as evidenced by an up-regulation of many IFN-stimulated genes (ISGs) not seen in individual nor in other combinations of H1 variant KDs. Although H1 participates to repress ISG promoters, IFN activation upon H1.2 and H1.4 KD is mainly generated through the activation of the IFN response by cytosolic nucleic acid receptors and IFN synthesis, and without changes in histone modifications at induced ISG promoters. H1.2 and H1.4 co-KD also promotes the appearance of accessibility sites genome wide and, particularly, at satellites and other repeats. The IFN response may be triggered by the expression of noncoding RNA generated from heterochromatic repeats or endogenous retroviruses upon H1 KD. In conclusion, redundant H1-mediated silencing of heterochromatin is important to maintain cell homeostasis and to avoid an unspecific IFN response.

INTRODUCTION

There are five major classes of histones that participate in the correct folding of eukaryotic DNA into chromatin: the four core histones H2A, H2B, H3 and H4 which form an octamer that constitutes the nucleosome core particle, and the linker histone H1, which binds nucleosomes near the entry/exit sites of linker DNA. Stabilization of the condensed states of chromatin is the function most commonly attributed to linker histone (1), in addition to its inhibitory effect in vitro on nucleosome mobility (2) and transcription (3). Unlike core histones, the H1 histone family is more evolutionary diverse and many organisms have multiple variants or subtypes. In humans, histone H1 is a family of closely related single-gene encoded proteins, including seven somatic subtypes (from H1.1 to H1.5, H1.0 and H1X), three testis-specific variants (H1t, H1T2 and HILS1) and one restricted to oocytes (H1oo) (4–6). Among the somatic histone H1 variants, H1.1 to H1.5 are expressed in a replication-dependent manner, while H1.0 and H1X are replication-independent. H1.2 to H1.5 and H1X are ubiquitously expressed, H1.1 is restricted to certain tissues, and H1.0 accumulates in terminally differentiated cells.

One of the major open questions in the field is whether different somatic H1 subtypes are mainly redundant or have specific functions, functions that are perhaps cell type

*To whom correspondence should be addressed. Tel: +34 402 0487; Fax: +34 403 4979; Email: ajvbmc@ibmb.csic.es
†These two authors contributed equally to this work as first authors.
Present address: Lluís Millán-Ariño, Integrated Cardio Metabolic Centre, Department of Medicine Huddinge, Karolinska Institutet, Stockholm, Sweden.

specific. It is well established that H1 subtype composition varies through development and differentiation as well as between cell types and during disease associated processes such as neoplastic transformation (7–15). H1 subtypes are post-translationally modified, both at conserved and unique residues, and these modifications may modulate their interaction with an increasing number of proteins at the nucleoplasm or in chromatin (16,17). These interactions could explain some reported specific functions for certain H1 variants (18–23). Furthermore, genomic distribution of H1 variants is not absolutely redundant, with enrichment of different variants at certain chromatin types (reviewed in (6)). We have previously investigated the distribution of four H1 variants genome-wide in breast cancer cells, concluding that H1 variants are not distributed uniformly along the genome, H1.2 being the one showing the most specific pattern and strongest correlation with low gene expression (24,25).

Previous studies on the effect of H1 depletion on global gene expression have found no effect on the vast majority of genes, but rather have detected variant specific up- or down-regulation of small subsets of genes (26–29). However, it is not clear whether these effects are driven by variant specific roles to regulate particular promoters or play distinct roles depending on the cell type. For example, H1.2 has been reported to act as a coactivator by bridging between RNA polymerase II and components of the elongation apparatus in 293T cells, and to act as a corepressor by establishing a positive feedback loop with EZH2-mediated H3K27me3 deposition in MCF7 cells (30,31). Other H1 variants have been involved in chromatin compaction or gene silencing, such as H1.4 (21,22,32,33). In general, silent genes contain histone H1 at the promoter, and a 'H1 valley' appears upstream of transcription start site (TSS) upon gene activation (25,34,35).

Altering the expression of H1 variants has been proven useful to study the contribution of individual variants to nuclear processes and to investigate the effects of global H1 decrease. In mice, single or double H1 variant knock-outs have no apparent phenotype due to compensatory up-regulation of other subtypes (36). These reports have favored the view that H1 variants are redundant. Knocking-out additional subtypes cannot be fully compensated by up-regulation of the remaining subtypes, resulting in embryonic lethality and demonstrating that the total amount of H1 is crucial for proper embryonic development (37). Despite embryonic lethality, triple knock-out H1.2–H1.3–H1.4 (TKO) mouse embryonic stem (ES) cells were obtained with a 50% global reduction in total H1 (27). Among the observed effects of TKO, a subset of imprinted genes are up-regulated that correspond to specific CpG regions becoming demethylated. Yang *et al.* demonstrated that some H1 subtypes specifically interact with the DNA methyltransferases DNMT1 and DNMT3B to promote methylation of imprinted genes (38). Major satellite repeats are also de-repressed in TKO (39). More recently, changes in the epigenetic signature of thousands of potential regulatory sites across the genome of TKO ES cells have been described, clustered to gene-dense topologically associating domains (TADs) (40).

We have previously reported that depletion of single H1 subtypes by inducible RNA interference in cancer cells produced a range of phenotypic effects, suggesting non-redundant functions for some of the histone H1 variants in somatic cells (26). Here, we further investigate consequences of depleting H1 variants by simultaneous depletion of several variants in breast cancer cells. We generate two new short-hairpin-RNAs (shRNAs), one for the specific depletion of H1.4, and the other affecting the expression of several H1 genes (multiH1sh), but mainly H1.2 and H1.4 at the protein level. While H1.2 and H1.4 single knock-downs (KDs) decrease proliferation similarly, multiple H1 KD has drastic consequences on cell growth and induces a strong transcriptional interferon (IFN) response. We observed the activation of IFN signaling transducers, participation of cytosolic nucleic acids receptors, IFN synthesis and up-regulation of IFN-stimulated genes (ISGs). In parallel, satellites and endogenous retroviruses are also up-regulated and cytosolic RNA is increased. This is the first report of multiple H1 depletion in human cancer cells, and we show the importance of histone H1 to maintain heterochromatin integrity and to avoid a growth-inhibiting IFN response. We also show that pancreatic carcinomas characterized by a constitutively induced IFN response express low levels of several H1 variants.

MATERIALS AND METHODS

Cell lines, culturing conditions and treatments

Breast cancer T47D-MTVL cells (carrying one stably integrated copy of luciferase reporter gene driven by the MMTV promoter) or derivative cells, were grown at 37°C with 5% CO_2 in RPMI 1640 medium, supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin, and 100 μg/ml streptomycin, as described previously (26). HeLa cell line was grown at 37°C with 5% CO_2 in DMEM GlutaMax medium containing 10% FBS and 1% penicillin/streptomycin. MCF7 cell line was grown at 37°C with 5% CO_2 in MEM medium containing 10% FBS, 1% penicillin/streptomycin, 1% non-essential amino acids, 1% sodium pyruvate and 1% glutamine. Doxycycline (Sigma) was added at 2.5 μg/ml. IFNß (Sigma), Ruxolitinib (Selleckchem), BX795 (Selleckchem) and 2-aminopurine (Sigma) were added at indicated concentrations.

Drug-inducible and constitutive RNA interference

Inducible H1 knock-down cell lines were established from T47D-MTVL cells as described previously (26). Plasmids for the lentivirus vector-mediated drug-inducible RNA interference system (pLVTHM, ptTR-KRAB-Red, pCMC-R8.91 and pVSVG) were provided by Didier Trono (University of Geneva) (41). The 71-mer oligonucleotides for shRNA cloning into Mlu/ClaI-digested pLVTHM were designed, annealed and phosphorylated as recommended by Didier Trono (http://tronolab.epfl.ch/). Target sequences are GTCCGAGCTCATTACTAAA for H1–4sh and GA ACAACAGCCGCATCAAG for multiH1sh. For the production of viral particles containing the lentiviral vector and infections, see Sancho *et al.* (26). The inducible knocked-down cell lines were sorted in a FACSCalibur machine (Becton Dickinson) for RedFP-positive and GFP-positive fluorescence after 3 days of doxycycline (Dox) treat-

ment. For the constitutive depletion of TLR3, STING, MAVS, MDA5 and IFNAR, shRNA-expressing pLKO.1 vectors from the MISSION library (Sigma-Aldrich) were used. Viral particles production and infections were performed as described (26). Cells infected with the shRNA-expressing lentivirus were selected with 2 mg/ml puromycin (Sigma-Aldrich) 24 h after infection.

Stable expression of HA-tagged H1 variants

Generation of T47D derivative cells stably expressing HA-tagged H1 variants using the lentiviral expression vector pEV833.GFP provided by Eric Verdin (Gladstone Institutes) was achieved as described previously (26).

RNA extraction and reverse transcriptase (RT)-qPCR

Total RNA was extracted using the High Pure RNA Isolation Kit (Roche). Then, cDNA was generated from 100 ng of RNA using the Superscript First Strand Synthesis System (Invitrogen). Gene products were analyzed by qPCR, using SYBR Green Master Mix (Invitrogen) and specific oligonucleotides in a Roche 480 Light Cycler machine. Each value was corrected by human GAPDH and represented as relative units. Each experiment was performed in duplicate. Specific qPCR oligonucleotide sequences are available as Supplementary Table S1.

Histone and total protein extraction, gel electrophoresis and immunoblotting

Histone H1 was purified by lysis with 5% perchloric acid for 1 h at 4°C. Soluble acid proteins were precipitated with 30% trichloroacetic acid overnight at 4°C, washed twice with 0.5 ml of acetone and reconstituted in water. 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 $MgCl_2$, 1 mM PMSF, 1 mM DTT] and incubated on ice for 30 min. The nuclei were pelleted at 10 000 × g for 10 min at 4°C. Sulfuric acid (0.2 M) was added to the pellet to extract the histones on ice for 30 min. The solution was centrifuged at 16 000 × g for 10 min at 4°C. TCA was added to the supernatant in order to precipitate histones and the precipitate was washed with acetone and finally resuspended in water. For isolation of total protein, cells were resuspended with Lysis Buffer [25 mM Tris-H–Cl (pH 7.5), 1% SDS, 1 mM EDTA (pH 8), 1 mM EGTA (pH 8), 20 mM B-glycerolphosphate] containing protease and phosphatase inhibitors, and boiled 20 min at 95°C. Cell lysates were obtained by centrifugation and protein concentration was determined by Micro BCA protein assay (Pierce). Purified histones or cell lysates were exposed to SDS-PAGE (12%), transferred to a PVDF membrane, blocked with Odyssey blocking buffer (LI-COR Biosciences) for 1 h, and incubated with primary antibodies overnight at 4°C and with secondary antibodies conjugated to fluorescence (IRDye 680 goat anti-rabbit IgG, Li-Cor) for 1 h at room temperature. Bands were visualized in an Odyssey Infrared Imaging System (Li-Cor). Polyclonal antibodies specifically recognizing human H1 variants, including those generated in our laboratory (26), are available from Abcam: H1.0 (ab11079), H1.2 (ab17677), H1.3 (ab24174), H1-T146p (ab3596), H1.5 (ab24175) and antiH1X (ab31972). Other antibodies used were: H1 AE-4 (Millipore, 05-457), beta-tubulin (Sigma, no. T4026), H3K4me3 (Abcam, ab8580), H3K9me3 (Abcam, ab8898), H3K27me3 (Millipore, 07-449), H4K20me3 (Abcam, ab9053), HP1α (Active Motif, 39295). H1-T146p was used to identify H1.4 by immunoblot, as the immunogen was a synthetic peptide derived from within residues 100–200 of human H1.4, phosphorylated at T146; this antibody could also recognize phospho T146 in H1.2, H1.3 (both 88% sequence identity with immunogen). To our knowledge this is the best available antibody to detect loss of H1.4 despite targeting a uniquely phosphorylated residue. To quantify H1 depletion with Coomassie staining, we load different amounts of total histones and choose the ones that showed the best linear range. After scanning with Image Gauge, individual histone H1 and H4 bands were quantified and H1 was expressed relative to H4, and relative to untreated cells. These were approximate quantifications due to the limitations of Coomassie staining of histone preps.

Immunofluorescence

One milliliter of hypotonic MAC buffer [50 mM glycerol, 5 mM KCl, 10 mM NaCl, 0.8 mM $CaCl_2$ and 10 mM sucrose] was added to 0.5 million cells (resuspended in 1 ml RPMI 1640 medium) and incubated at room temperature for 5 min. 400 μl were spun down for 10 min at 500 rpm in a TermoShandon Cytospin 4 using a single-chamber Cytospin funnel as described in (42). Cells were air-dried for 1h at room tempearature. Then, fixed with 4% paraformaldehyde for 1h at room temperature. After three washes, they were permeabilized with 0.2% Triton X-100 for 15 min at room temperature and blocked with 3% bovine serum albumin for 1 h at room temperature. Then, the cells were incubated with primary antibody recognizing dsRNA (Scicons J2, English and Scientific Consulting Kft, Hungry) diluted in the blocking buffer, over-night at 4°C. After the pertinent washes, the secondary antibody Alexa Fluor 488 donkey anti-mouse (Life Technologies, A21202) was added for 1 h at room temperature in darkness. Nuclei were stained with DAPI, coverslips mounted using Mowiol, and samples were visualized by confocal laser scanning microscopy using a Leica TCS SP5 system.

Cell proliferation analysis

Cells of interest expressing an shRNA (GFP-positive) were mixed 1:1 with parental cells (GFP-negative) and cultured. Every three days, the cells were split and the percentage of cells that were GFP-positive was measured by FACS.

MNase digestion

Cells treated or not with Dox for six days were collected, resuspended and incubated for 10 min at 4°C in Buffer A [10 mM Tris–HCl pH7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.3 M Sacarosa and 0.2 mM PMSF], after that 0.2% of NP40 was added and incubated 10 min more at 4°C. After centrifugation, pellets were resuspended in Buffer A + 10 mM $CaCl_2$ and 0.4 u of MNase per 5 millions of nuclei and incubated

at 37°C. At different incubation times, MNase reaction was stopped by adding 10 mM EDTA pH 8. DNA was purified using EZNA Tissue DNA columns (OMEGA VWR) and run on a 2% agarose gel.

Chromatin immunoprecipitation (ChIP)

Immunoprecipitation of chromatin was performed according to the Upstate (Millipore) standard protocol. Briefly, cells were fixed using 1% formaldehyde for 10 min at 37°C, harvested and sonicated to generate chromatin fragments of 200–500 bp. Then, 30 μg of sheared chromatin was immunoprecipitated overnight with 2 μg of antibody. Immunocomplexes were recovered using 20 μl of protein A magnetic beads, washed and eluted. Cross-linking was reversed at 65°C overnight and immunoprecipitated DNA was recovered using the PCR Purification Kit from Qiagen. Genomic regions of interest were identified by real-time PCR (qPCR) using SYBR Green Master Mix (Invitrogen) and specific oligonucleotides in a Roche 480 Light Cycler machine. Each value was corrected by the corresponding input chromatin sample. Oligonucleotide sequences used for the amplifications are available on request.

mRNA library preparation and sequencing (RNAseq)

The library from total RNA was prepared using the TruSeq® Stranded Total Sample Preparation kit (Illumina Inc.) according to manufacturer's protocol. Briefly, rRNA was depleted from 0.5 ug of total RNA using the Ribo-Zero Gold Kit followed by fragmentation by divalent cations at elevated temperature resulting into fragments of 80–450nt, with the major peak at 160nt. First strand cDNA synthesis by random hexamers and reverse transcriptase was followed by the second strand cDNA synthesis, performed in the presence of dUTP instead of dTTP. Blunt-ended double stranded cDNA was 3′adenylated and the 3′-'T' nucleotide at the Illumina indexed adapters was used for the adapter ligation. The ligation product was amplified with 15 cycles of PCR. Each library was sequenced using TruSeq SBS Kit v3-HS, in paired end mode with the read length 2 × 76 bp. We generated minimally 37 million paired end reads for each sample run in a fraction of a sequencing lane on HiSeq2000 (Illumina, Inc) following the manufacturer's protocol. Images analysis, base calling and quality scoring of the run were processed using the manufacturer's software Real Time Analysis (RTA 1.13.48) and followed by generation of FASTQ sequence files by CASAVA.

RNAseq data analysis

The mapping of the reads against the human reference genome (GRCh38) was done using the STAR program (version 2.5.1b) (43) and the quantification of genes was done with the RSEM program (version 1.2.28) (44). Differential expression analysis was done with DESeq2 (version 1.10.1) (45). Heatmaps, correlation scatter plot and boxplots were done using in-house R scripts. The expression values are those generated by DESeq2 using variance stabilizaing transformations. Genes differentially expressed by IFN treatment were selected from Interferome database (version 2.01) (46) applying a filter of 2 < fold-change < –2 and those differentially expressed in more than one experiment. Enrichment of transcription factor binding DNA motifs was done using Homer software (47). Gene ontology enrichment analysis was performed using the R package 'goseq' (version 1.18.0) (48) and reduced by semantic similarity using REViGO software (49). Pathway enrichment analysis and figures were generated using ReactomeFIPlugin (50) for Cytoscape (51). To perform the enrichment analysis of repetitive sequences the RNA-seq reads were mapped to the Repbase database (version 21.02) (52) using Bowtie aligner (default parameters) (53). We only considered those reads which both paired-end reads were aligned to the same repetitive sequence. To test for significance a Fisher exact test was performed. Significantly enriched repetitive sequence were those with a adjusted *P*-value < 0,05 and not enriched in RDsh samples.

Assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq)

ATAC experiment was performed as described (54). 75 000 cells treated were harvested and treated with transposase Tn5 (Nextera DNA Library Preparation Kit, Illumina). DNA was purified using MinElute PCR Purification Kit (Qiagen). All samples were then amplified by PCR using NEBNextHigh-Fidelity 2× PCR Master Mix (New Englands Labs.) and primers containing a barcode to generate the libraries. DNA was again purified using MinElute PCR Purification Kit and samples were sequenced using Illumina HiSeq 2500 system. Paired-end reads were first trimmed to 30 bp to remove adapter sequences. Then, aligned to the reference genome (hg19) using bowtie software (53). Duplicated reads and those mapping to the mitochondrial chromosome were removed. Accessibility peaks were obtained using MACS2 peak caller software (55). Average accessibility profile around the TSS of selected genes and pie charts representing the annotation of peaks was performed using CEAS software (56). For the repeats accessibility analysis, annotated repeats were taken from the RepeatMasker track from the UCSC and the number of peaks overlapping them computed. For the analysis of peaks overlapping up- or down-regulated genes at the coding regions or promoter, or repeats, a permutation test (10^4 permutations) was performed using the R package regioneR (57).

Expression data accession numbers

RNAseq data is available in the Gene Expression Omnibus (GEO) database under the accession number GSE83277. Microarray data accession numbers are GSE11294 and GSE12299. ATACseq data accession number is GSE100762.

Human H1 variant nomenclature

The correspondence of the nomenclature of the human H1 variants with their gene names is as follows: H1.0, H1F0; H1.1, HIST1H1A; H1.2, HIST1H1C; H1.3, HIST1H1D; H1.4, HIST1H1E; H1.5, HIST1H1B; and H1X, H1FX.

RESULTS

Simultaneous depletion of multiple H1 variants in breast cancer cells impairs cell growth drastically

We have previously reported the effects of the inducible depletion of six somatic histone H1 variants on cell growth and gene expression in breast cancer cells (24,26). H1.2 depletion caused cell-cycle arrest in G1 phase, while H1.4 KD caused cell death in T47D cells. Each variant KD altered the expression of small distinct subsets of genes.

We have designed and assayed new shRNAs targeting H1. One of such shRNAs (sh120; H1.4sh) was specific for H1.4 (Figure 1A and B), while another (sh225; multiH1 or mH1sh from now on) reduced the expression of several H1 transcripts, although only H1.2 and H1.4 proteins were seen depleted consistently when tested by immunoblotting after 6 days of Doxycycline induced shRNA expression (Figure 1A and C, and Supplementary Figure S1). H1.3 was also seen partially depleted in some experiments. The two new shRNAs caused H1.0 up-regulation, both at the mRNA and protein level. This was previously observed for H1.2 KD, but not for the other H1 variants (26). Coomassie staining of histones extracted from multiH1 KD cells showed that the total content of H1 was reduced to ca. 70% despite of H1.0 up-regulation (Figure 1C and D). The intensity of protein bands containing H1.2 and H1.4 were decreased.

In addition, multiH1sh expression caused a significant reduction in global nucleosome spacing as assessed by Micrococcal nuclease digestion of bulk chromatin (Figure 1E and Supplementary Figure S2). This was almost undetectable upon H1.4 KD, but was previously described for H1.2 depletion in breast cancer cells (26).

Expression of both shRNAs impaired cell proliferation (Figure 1F). The effect of inhibiting simultaneously H1.2 and H1.4 in multiH1 KD cells on the growth rate was additive compared to single H1 depletion. H1 depletion increased the number of cells in G1 phase of the cell cycle and reduced cells in S phase, with the effects of multiH1sh being more pronounced than H1.4sh (data not shown).

Transcriptome changes upon histone H1 variants knockdown

We studied the consequences on global gene expression of multiple H1 KD and H1.4-specific depletion compared to the random shRNA using RNAseq. Data confirmed the efficacy of H1.4 and multiH1 shRNAs to inhibit the corresponding H1 genes without significant compensations other than some H1.0 up-regulation (Figure 2A). Results were compared with microarray data obtained before with the KD of the other variants (24,26). As previously described, depletion of individual H1 variants altered the expression of small subsets of genes, considerably specific for each variant, and containing both up- and down-regulated genes in different proportions. The inhibition of multiple H1s affected a larger number of genes, some already observed in single variant KDs (Figure 2B and Supplementary Table S2). In particular, 38% of genes affected in H1.4sh were also affected in multiH1sh (which also targets H1.4), while only 20–28% of the genes affected in the other single variant KDs (H1.0, H1.2, H1.3 or H1.5) were significantly affected in the multiH1 KD.

Among genes up-regulated in the multiH1 KD, there were many interferon-stimulated genes (ISGs) (Figure 2C). Comparison of our data with the Interferome database containing microarray data of IFN-treated cells, showed that 36% of the up-regulated and 18% of the down-regulated genes in the multiH1 KD were known to also respond to short- or long-term INF treatments. This is in comparison to the 6% of the total transcriptome known to be susceptible to IFN treatment (Figure 2C). Almost all the multiH1 KD up-regulated genes with the highest fold-change are ISGs. Moreover, fold-change gene induction upon multiH1 depletion was proportional (positive correlation) to the fold-change induction in response to IFN in multiple experiments deposited in the Interferome database (Figure 2D).

Motif analysis in the promoters of multiH1sh deregulated genes found significant enrichment of interferon-stimulated response elements (ISRE), binding sites for interferon-responsive transcription factors (IRFs), but not in gamma-activated sites (GAS) (Figure 2E). ISRE are found in type I IFN-responsive genes, while GAS elements are present in type II IFN-stimulated genes. Down-regulated genes contained motifs for E2F, Sp1 and NFY transcription factors (q-value<0.0001). Gene-ontology analysis confirmed the enrichment of IFN signaling pathways in the multiH1sh up-regulated genes, while multiH1sh and single H1.4sh down-regulated genes showed GO-terms related to cell-cycle regulation (Figure 2F). No significantly enriched DNA motifs were found in the promoters of H1.4sh affected genes, and no GO-terms were found for the H1.4sh up-regulated genes.

Interestingly, basal gene expression (without doxycycline) of genes up-regulated in H1 KD cells was significantly lower than basal expression of genes down-regulated, despite of the variability within each gene set (Figure 2G). After H1 depletion, overall gene expression of up- and down-regulated genes was more similar. This observation could suggest that histone H1 may play a role in the maintenance of repression in the up-regulated genes. Nevertheless, both up- and down-regulated genes showed an average gene expression above the average of RefSeq genes or even above the expressed transcriptome (Supplementary Figure S7A).

In conclusion, depletion of individual H1 variants effects a small subset of genes, while multiple variant knock-down alters gene expression extensively, larger than the additive effect of individual variants, including an unexpected large proportion of type I interferon-stimulated genes. The fact that there is limited overlap between deregulated genes in H1.2 or H1.4 KDs and multiH1 KD suggests that these genes are not directly regulated by these variants, but the combined depletion of the two variants may trigger an interferon response.

Induction of interferon-stimulated genes upon combined depletion of histone H1 variants

We next sought to understand how the interferon response is induced in H1-depleted breast cancer cells. Either H1 is particularly involved in repressing the promoters of ISGs with little variant specificity, or general H1 depletion from

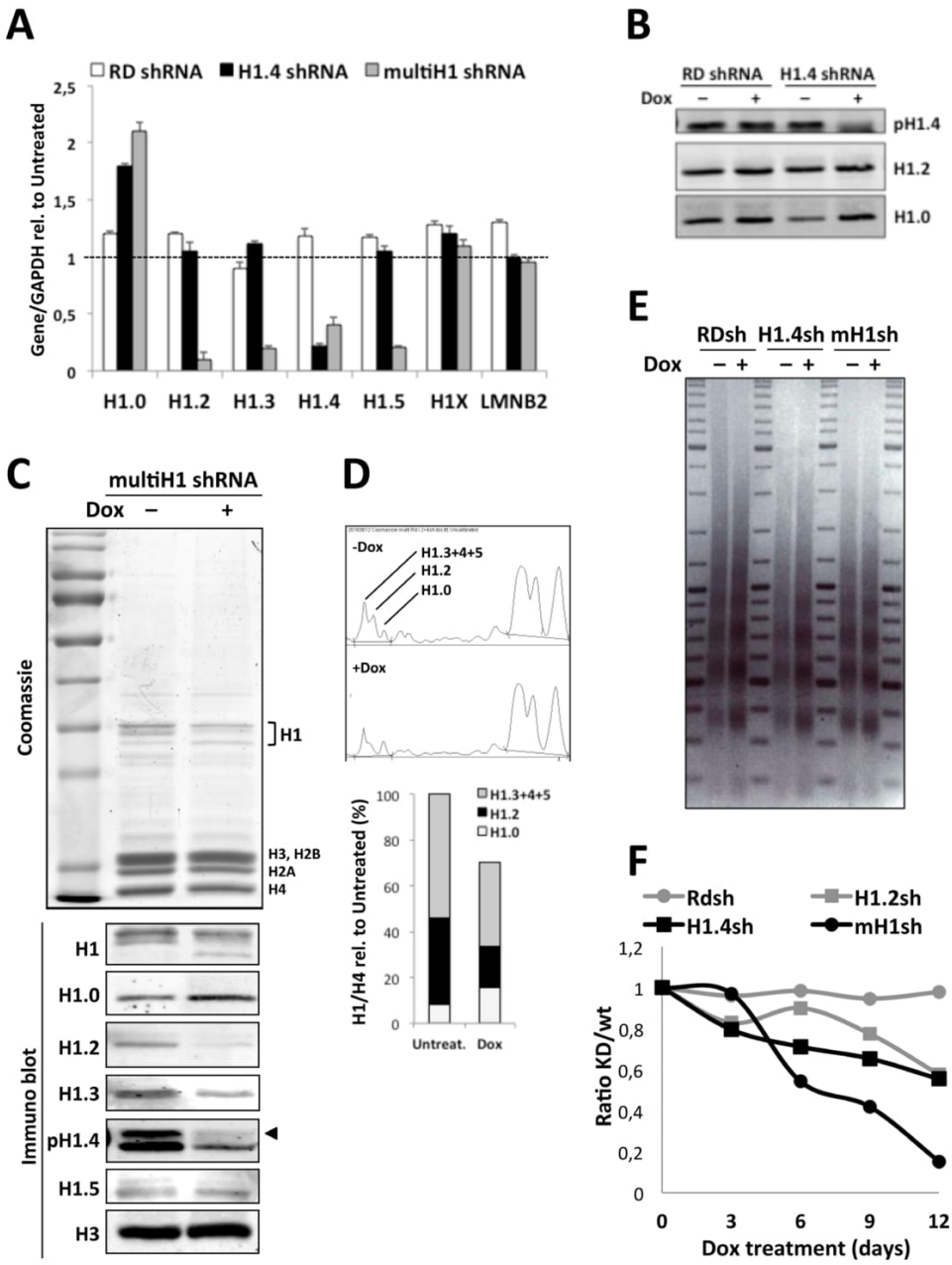
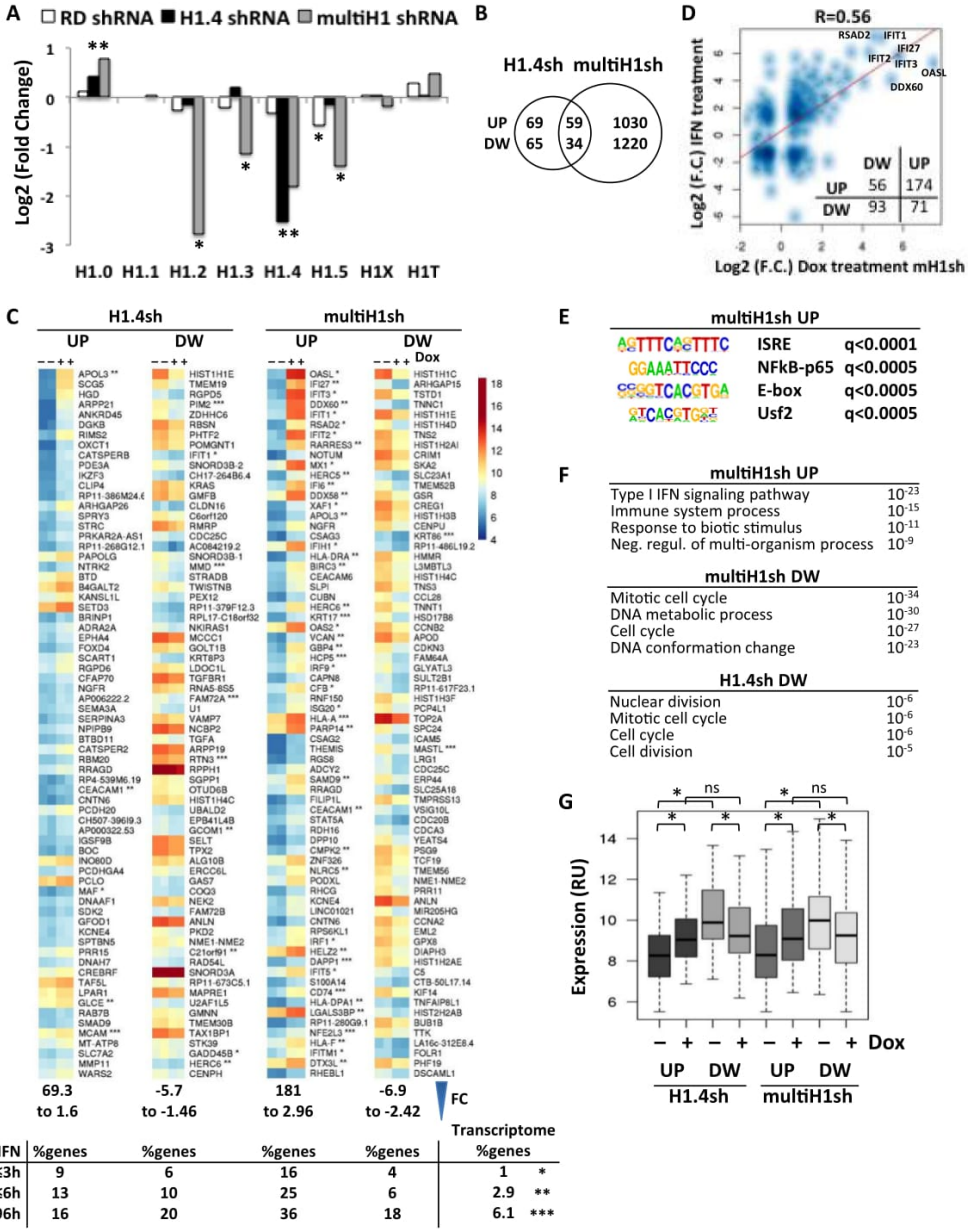


Figure 1. Inducible shRNAs for the depletion of histone H1 variants in human breast cancer cells. (**A**) Expression of H1 variants upon inducible expression of new shRNAs. T47D derivative cells stably infected with inducible lentiviruses for the expression of random, H1.4 or multi-H1 shRNAs were treated or not for 6 days with doxycycline (Dox), RNA was extracted and gene expression was analyzed by RT-qPCR with oligonucleotides for the indicated H1 genes. Gene expression in response to Dox is expressed corrected by GAPDH and relative to untreated cells. Values represent the mean and SD of a representative experiment performed in triplicate. (B and C) H1 depletion was additionally tested by immunoblotting with the H1 variant-specific antibodies indicated. H1 phospho-T146 antibody was used to detect H1.4 (pH1.4) (see Materials and methods). (**B**) H1 from random and H1.4 shRNA cells treated or not with Dox for 6 days were acid-extracted. (**C**) Total histones were extracted from multi-H1 shRNAs cells, run on SDS-PAGE and stained with Coomassie or immunoblotted. (**D**) Coomassie-stained histone bands were scanned and quantified. The graph denotes the decrease of total H1 and variations of each H1-containing band (corrected to histone H4) upon Dox treatment of multiH1 KD cells. (**E**) H1 KD causes a reduction in nucleosome spacing. Nuclei from the KD cells indicated treated or not with Dox for 6 days were treated with MNase and the profile of bulk chromatin was analyzed in gel electrophoresis. A representative experiment is shown. (**F**) Effect of H1 depletion on cell proliferation. In order to measure the effect of H1 depletion on cell proliferation, each of the H1 variant KD cell lines indicated (RFP and GFP-positive) was mixed 1:1 with parental T47D cells (RFP and GFP-negative) and treated with Dox. Every three days, cells were split and the percentage of double-positive cells was measured by FACS. Data is expressed as ratio of positive (KD) versus negative (wild-type) cells along time and corresponds to a representative experiment performed in duplicate.



IFN	%genes	%genes	%genes	%genes	Transcriptome %genes	
≤3h	9	6	16	4	1	*
≤6h	13	10	25	6	2.9	**
≤96h	16	20	36	18	6.1	***

Figure 2. Transcriptome changes upon histone H1 variants knock-down. (**A**) H1 variants expression in the random, H1.4 and multiH1 KD cells upon a 6-day Dox treatment extracted from RNAseq data. Data is shown as Log2 of the fold-change Dox-treated compared to untreated. Significant changes with *P*-value < 0.001 are shown with asterisks. (**B**) Venn diagram showing the number of genes up- or down-regulated in H1.4 and multiH1 KD cells (and not changing in random shRNA ±Dox samples) established by RNAseq (FC ≥ 1.4, adjusted *P*-value ≤ 0.05). (**C**) Genes most highly up- (UP) and down-regulated (DW) in H1.4 and multiH1 KD cells (70 genes of each are shown), not changing in RDsh samples, ordered according to fold-change from top to down. Genes that respond to IFN according to the Interferome data base (in more than one experiment, with FC ≥ 2) are labeled with asterisks: *, 0.5–3 h IFN-treatment; **, 3–6 h IFN; ***, 6–96h IFN. The table below shows the percentage of genes up- or down-regulated in each H1 KD (FC ≥ 1.4, adjusted *P*-value ≤ 0.05), or in the total transcriptome, that respond to IFN. (**D**) Correlation scatter plot between gene expression fold-change in multiH1sh cells upon Dox-treatment (RNAseq data, FC ≥ 1.4) and fold change upon IFN treatment (Interferome data base, FC ≥ 2). R is the Pearson's correlation coefficient, with *P*-value < 0.0001. Insert table: Number of genes deregulated in multiH1sh that are up- or down-regulated in response to IFN. (**E**) Enriched transcription factor binding DNA motifs in multiH1sh up-regulated genes. (**F**) Analysis of enriched gene-ontology terms in H1 KD deregulated genes. GO terms (biological processes) were collapsed by semantic similarity using REVIGO software and the four more significant terms are shown. Enrichment score is shown as adjusted *P*-value. (**G**) Box plots showing the expression profiles of genes up- and down-regulated in H1.4 and multiH1 KD cells, treated or not with Dox. Significance was tested using the Kolmogorov–Smirnov test (**P*-value < 0.05).

the genome somehow triggers the response by promoting IFN synthesis. First, we confirmed by reverse transcription and semi-quantitative PCR (RT-qPCR) the induction of several ISGs in multiH1sh but not control random shRNA (RDsh) cells upon 6-days of doxycycline treatment (Figure 3A). Some of the ISGs showed inductions as high as 100- to 300-fold. A 3-days time curve treatment with doxycycline showed that H1 transcripts, mainly H1.2 and H1.4, were significantly depleted after 12 h (Figure 3B). ISG up-regulation was first observed 48 h after treatment start and was increased at 72 h, both for genes highly induced (IFI27, IFIT2 and IFI6) or less induced (STAT1, IRF1 and IRF7). Interestingly, the gene encoding for IFN-ß (IFNB), which was not found up-regulated in the RNAseq data after 6 days treatment, was induced at early time points (Figure 3B). In a longer H1 KD time curve, ISG expression was progressively increased up to 9 or 12 days of doxycycline treatment (Figure 3C and Supplementary Figure S3). Instead, IFNB, after peaking at 3 days, decreased its expression, explaining why it was not detected in the RNAseq experiment (data not shown). In the same experiment, upon removal of doxycycline at day 3, recovery of H1 variants transcript accumulation was not observed at day 6 but it was at day 9. Upon H1 expression recovery, ISGs were efficiently down-regulated to basal level, indicating that the IFN response induction can be reversed upon H1 recovery.

Induction of ISGs was not observed in single H1 KDs, neither in genome-wide data (microarrays and RNAseq), nor by RT-qPCR. Figure 3D shows the absence of ISGs induction in H1.2 KD constructed in T47D, MCF7 and HeLa cell lines. In contrast, when the H1.2 inducible shRNA was introduced into T47D H1.4 KD cells, to obtain a new multiH1 KD cell line (H1.4/H1.2sh; 80–90% of total H1), ISGs were highly induced upon doxycycline treatment (Figure 3E and Supplementary Figure S4). This indicates that the combined depletion of H1.2 and H1.4 in T47D with variant-specific shRNAs is sufficient to promote the IFN response and validates results obtained with the multiH1sh construct. Induction of ISGs was not unique to T47D cells as it was also observed in MCF7 infected with the multiH1sh, where H1.2 and H1.4 are depleted (data not shown).

We also tested whether other H1 KD combinations were able to reproduce this effect. The H1.5 inducible shRNA was introduced into T47D H1.2 or H1.4 KD cells and all Dox-induced H1 depletions tested by RT-qPCR, Coomassie and immunoblot. Total H1 content was reduced to 60–80% of wild-type (Supplementary Figure S5). Simultaneous depletion of H1.2 and H1.5 caused ISG up-regulation to some extent (Figure 3F). Nonetheless, fold-change induction of ISGs was much higher in H1.4/H1.2sh than in H1.2/H1.5sh cells (Figure 3E versus F). No ISG up-regulation was observed upon combined depletion of H1.4 and H1.5, suggesting that H1.2 depletion is crucial, although not sufficient, to induce the IFN response. Interestingly, the total amount of histone H1 depleted in H1.4/H1.5sh or H1.2/H1.5sh cells was not lower than the achieved with the other KD combinations (multiH1sh or H1.4/H1.2sh), indicating that the induction of the IFN response is variant specific and not likely due to total H1 reduction. Finally, H1.4/H1.5sh and H1.2/H1.5sh cells were further infected with the H1.3 inducible shRNA. While ISGs remained unaltered in H1.4/H1.5/H1.3sh cells, they were further induced in H1.2/H1.5/H1.3sh cells, still to lesser levels than H1.4/H1.2sh (Supplementary Figure S5C).

To test whether ISGs induction in multiH1sh cells could be complemented by overexpressing distinct H1 variants, we infected multiH1sh cells with lentiviruses expressing GFP and hemaglutinin peptide (HA)-tagged H1.2, H1.3 or H1.4. Efficient infection was confirmed by FACS analysis of GFP (80–90% infection) as well as RT-qPCR of H1 expression (Figure 3F). Although the recombinant H1 genes were sensitive to the multiH1 shRNA and expression of HA-tagged H1s was diminished upon Dox treatment, some expression remained above the levels found in parental multiH1sh cells, sufficient to observe partial complementation of the effect of multiH1 KD on ISGs induction (Figure 3F). Immunoblotting confirmed that HA-tagged H1s were still present after Dox treatment (Supplementary Figure S6). It is noteworthy that not only H1.2 or H1.4 overexpression partially complemented the multiH1 KD, but also H1.3, indicating that the IFN response induction by H1 depletion is not fully variant-specific.

Synthesis of interferon and induction of IFN signaling upon multiple H1 depletion

Together these data suggest that H1 depletion mimics or induces the IFN response, for example inducing IFN production. The IFN response is triggered in cells when foreign DNA or RNAs are recognized by cytosolic nucleic acid sensors that trigger a signaling cascade, leading to the activation and nuclear entry of IFN-responsive transcription factors (IRFs). These IRFs then bind and regulate the promoter of IFN-encoding genes and other ISRE-containing target genes. IFN is secreted to the media, where it binds and activates IFN receptors at the surface of responsive cells, which activate the Janus kinase (JAK)-STAT signaling pathway leading to further activation of IRFs and ISRE-containing target genes (by type I IFNs), or to the activation of genes containing gamma-activated sites (GAS) elements (by type II IFN). IFN stimulation classically results in an anti-viral immune response and also has anti-tumor effects inducing ISGs with anti-proliferative and pro-apoptotic functions (58,59).

Induction of ISGs upon H1 depletion could be mediated by IFN synthesis, because, as reported above, the IFNB gene (but not IFNA) was up-regulated at short times after Dox addition to multiH1sh cells (Figure 3B). First, we confirmed that T47D derivative cells were sensitive to commercial IFN-ß, and induction of ISGs by IFN-ß at 0.15 μg/ml was comparable to the upregulation achieved by the combined KD of H1.4 and H1.2 (Figure 4A). ISG induction was not increased by higher concentrations of IFN-ß (1.2 μg/ml), but was synergistic with the effect of H1.4/H1.2 KD (Dox treatment), suggesting that ISG induction by H1 depletion is not only due to IFN synthesis. The effect of IFN-ß was not potentiated by knocking-down individually H1.2 or H1.4 (Figure 4A). We then tested whether IFN was being produced and liberated to the extracellular media. Conditioned growth media of multiH1 KD cells treated with Dox for 3 days was added to wild-

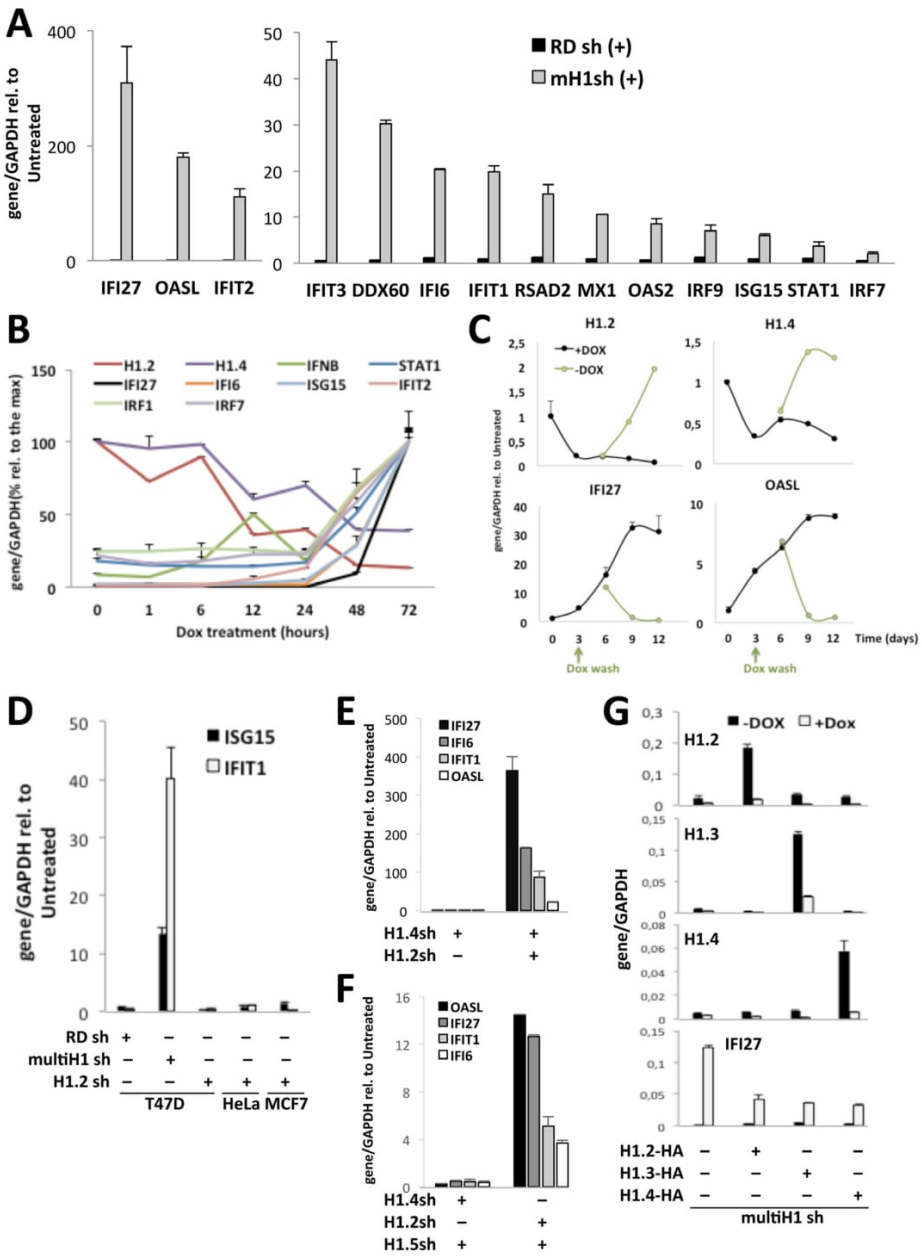


Figure 3. Induction of interferon-stimulated genes upon combined depletion of histone H1 variants. (**A**) Induction of ISGs in multiH1 KD cells. T47D cells containing RDsh or multiH1sh were treated or not with Dox for 6 days and gene expression was tested by RT-qPCR with oligonucleotides for the indicated genes. Gene expression in response to Dox is expressed corrected by GAPDH and relative to untreated cells. (**B**) ISG and H1 variants expression along time after Dox treatment. T47D cells containing multiH1sh were treated with Dox up to 72 h and gene expression was tested as in (A). Gene expression is corrected by GAPDH and relative to the highest data point for each gene. (**C**) H1 expression recovery and reversal of ISG expression after Dox removal. MultiH1 KD cells were treated with Dox for 3 days and split, washed and maintained without Dox or left with Dox up to 12 days. Gene expression was tested by RT-qPCR every 3 days, corrected by GAPDH and shown relative to untreated (time 0). (**D**) ISG induction occurs in multiH1sh but not H1.2sh cells. The indicated cell lines containing RDsh, multiH1sh or H1.2sh were treated or not with Dox for 6 days and ISG15 and IFIT1 gene expression was tested and expressed as in (A). (**E**) ISGs are induced in a double H1.4/H1.2 KD cell line. T47D cells containing H1.4sh were infected or not with the H1.2 shRNA-expressing lentivirus. Resulting cells were treated or not with Dox for 6 days and ISG expression was tested as in (A) and expressed relative to untreated cells. (**F**) ISG induction in a double H1.2/H1.5 KD cell line. T47D cells containing H1.2sh or H1.4sh were infected with an H1.5 shRNA-expressing lentivirus. Resulting cells were treated or not with Dox for 6 days and ISG expression was tested as in (A). (**G**) H1 overexpression in multiH1 KD cells partially blocks ISG up-regulation. T47D cells containing multiH1sh were infected or not with a lentivirus for the expression of HA-tagged H1.2, H1.3 or H1.4 variants. Resulting cells were treated or not with Dox for 6 days and H1 variant and ISG gene expression was tested as in (A). Values represent the mean and SD of representative experiments performed in triplicate.

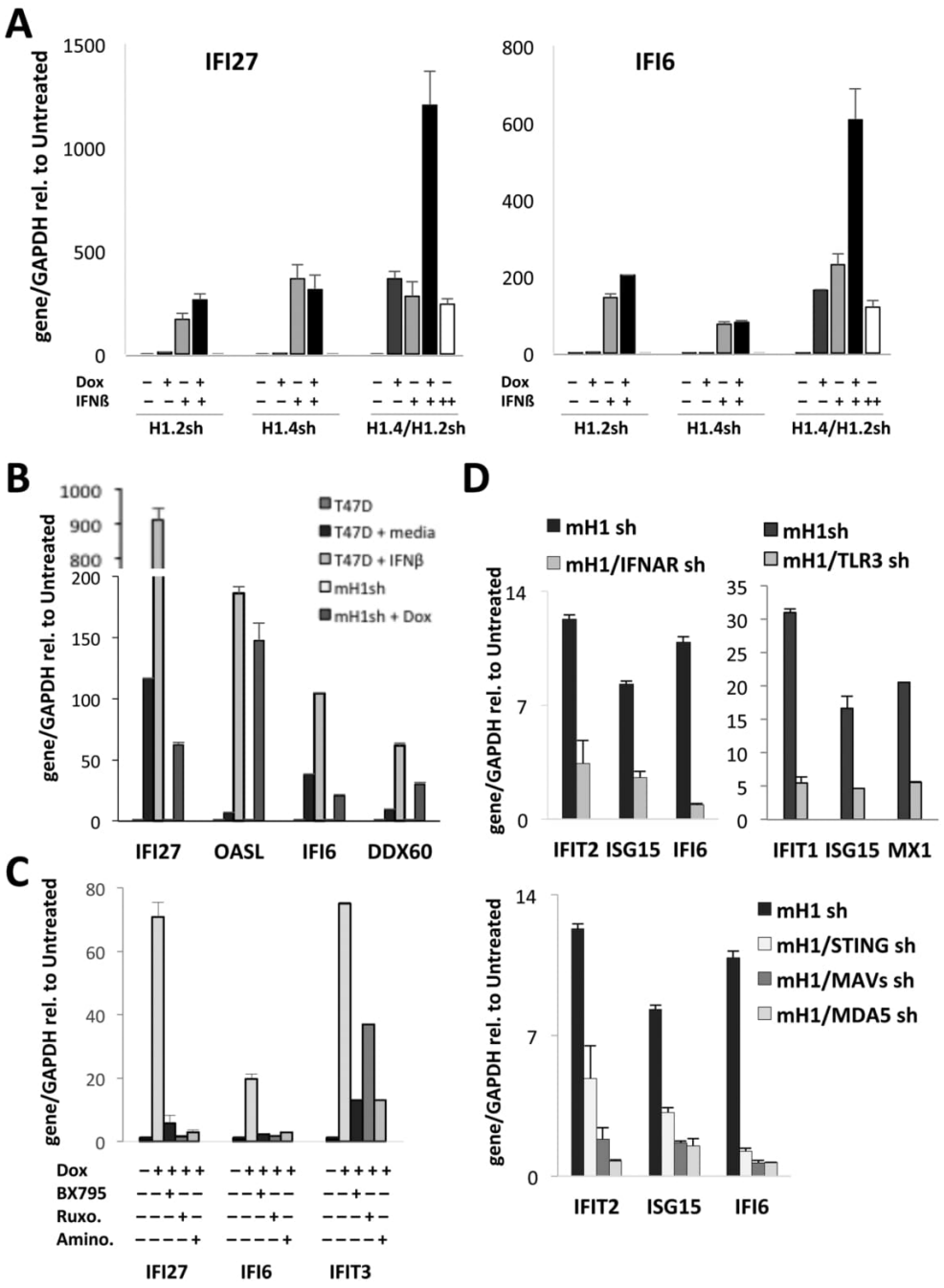


Figure 4. Activation of the interferon response pathway upon depletion of histone H1. (**A**) IFN-ß enhances ISG induction in H1.4/H1.2sh cells. T47D cells containing H1.2sh, H1.4sh or H1.4/H1.2sh were treated or not with Dox for 6 days and IFN-ß at 0,15 (+) or 1,2 (++) μg/ml was added for the last 8 h before analysis of ISG expression, corrected by GAPDH. (**B**) ISG induction by IFN liberated to the media. T47D cells containing multiH1sh were treated with Dox for 3 days, media was collected and added to wild-type T47D cells for 24 h. For comparison, T47D cells were treated with commercial IFN-ß (0.15 μg/ml) for 24 h, and multiH1sh cells treated with Dox for 3 days. ISG expression was tested by RT-qPCR, corrected by GAPDH and expressed relative to untreated cells. (**C**) ISG induction is blocked with IFN pathway inhibitors. T47D multiH1sh cells were treated or not with Dox, 0.5 μM BX795, 1 μM ruxolitinib or 10 mM 2-aminopurine, for 3 days, and ISG expression was tested as in (A). Gene expression is corrected by GAPDH and relative to the highest data point for each gene. (**D**) Blocking ISG induction by knocking-down type-I IFN receptor or IFN pathway sensors and transducers. T47D multiH1sh cells stably expressing shRNAs against IFNAR1, TLR3 (upper panels), STING, MAVS or MDA5 (lower panel) were treated or not with Dox for 6 days and ISG expression was tested as in (A) compared to parental multiH1sh cells. Gene expression is corrected by GAPDH and represented relative to untreated cells. Values represent the mean and SD of representative experiments performed in triplicate.

type T47D cells and ISG expression was tested. The media was able to stimulate expression of ISGs to some extent, although to a lesser extent than was achieved by treating cells with commercial IFN-ß (Figure 4B). Moreover, when a shRNA against the type I IFN receptor IFNAR1 was introduced into multiH1sh cells, ISG stimulation upon Dox treatment was partially affected (Figure 4D). In conclusion, multiH1 depletion has an effect on the induction of ISGs that is partially explained by the production and reentry of IFN.

To further determine the involvement of signaling associated to the IFN response, we analyzed the effect of inhibitors of kinase TBK-1, an intermediate in the sensing cascade, and Janus kinase (JAK) 1 and 2, BX795 and Ruxolitinib, respectively. Both inhibitors impaired the up-regulation of several ISGs (Figure 4C).

DNA and RNA cytosolic sensors trigger the IFN response with different specificity. Examples include MDA5 (IFIH1), RIG-I (DDX58), TLRs, cGAS or IFI16. These receptors activate intermediates such as MAVS (IPS-1) or STING that activate kinase TBK-1, which activates IRFs. We introduced shRNAs to deplete MDA5, TLR3, MAVS and STING from multiH1sh cells and observed a decrease in ISG stimulation upon Dox treatment (Figure 4D). Importantly, the factor that contributed the most to ISG up-regulation in multiH1 KD cells was MDA5, an RNA sensor, and the one that did it the least was STING, an adaptor of the DNA sensor pathway. Treating multiH1sh cells with 2-aminopurine, an inhibitor of IFN-inducible dsRNA-activated protein kinase R (PKR, EIF2AK2), which is involved in sensing dsRNA and activating MAVS, also impaired the activation of several ISGs (Figure 4C).

All together, these data show that upon H1 depletion an IFN response is being induced with participation of cytosolic sensors and signal transducers and features the synthesis of IFN, suggesting that some nucleic acid may be produced at the cytosol as a consequence of H1 depletion. Several genes involved in sensing and responding to foreign nucleic acids were up-regulated in multiH1 KD cells, including DDX60, IFI16, MDA5, RIG-I, LGP2, STAT1, IRF1, IRF7, IRF9, OASL and OAS2. Enrichment pathway analysis denoted that pathways like 'RIG-I/MDA5 (RNA sensing molecules)-mediated induction of IFN-alpha/beta' were significantly enriched, but not 'cytosolic sensors of pathogen-associated DNA', suggesting that the IFN response was being induced by RNA molecules upon multiH1 KD (data not shown).

Induction of repeats and endogenous retroviruses upon histone H1 knock-down

It has been previously reported that DNA methyltransferase inhibitors (DNMTis) induce the IFN response pathway in cancer cells with anti-proliferative effects by inducing synthesis of endogenous dsRNAs from endogenous retroviruses, DNA repeats and non-coding transcripts (60–62). We hypothesized that spurious transcription could be taking place from non-coding RNA genes, DNA repeats, satellites or endogenous retroviruses (ERVs), which are generally located in heterochromatin, and are sensed as foreign thus activating the IFN response.

Expression of several repeats was assessed by RT-qPCR in T47D multiH1sh cells upon Dox treatment, using random shRNA cells as a control (Figure 5). Some repeats were chosen because were found up-regulated in our multiH1sh RNAseq data (MER4D, HERVK and LTR57) (data not shown). Others were selected due to its high abundance of histone H1 variants according to our ChIP-seq data (MLT1J2) (data not shown, (25)). Expression of the endogenous retroviruses (ERVs) MER4D, MER21C, MLT1C49, HERVK, Env-Fc1, LTR57, MER25A and MLT1J2, the subtelomeric repeat D4Z4 and satellite A, are shown to be up-regulated to different extent upon multiH1 inhibition, but not in control treatment (Figure 5A). Up-regulation of these repeats was also observed in H1.4 + H1.2 combined KD cells, but not as much in individual H1 KDs (Figure 5B). Although some repeats are induced to some extent in individual H1 KDs, this is insufficient to trigger the IFN response. Higher induction of a wider number of repeats achieved with the double KD may be the cause of the IFN pathway stimulation observed.

To test that expression of ERVs is the origin of IFN signaling activation, we tested whether inhibition of retrotranscriptase (RT) activity affected the up-regulation of ISGs in multiH1 KD cells. MultiH1sh cells were treated for 3 days with doxycycline and a cocktail of RT inhibitors (Tenofovir 0.5 μM, Nevirapine 17 μM, Emtricitabine 7.3 μM). ISGs were induced similarly in the presence of RT inhibitors (data not shown), indicating that cDNA synthesis is not required and confirming that RNA could be the triggering molecule.

According to our data, a plausible model on how multiple H1 KD induces the IFN response would be that H1 depletion from heterochromatin may induce several non-coding RNAs from endogenous retroviruses, satellite and other repeats, that are sensed in the cytosol as exogenous (ds)RNAs. Immunofluorescence with the specific antibody J2 denoted that dsRNA was accumulated into the cytosol of multiH1sh cells upon doxycycline treatment (Figure 5C). Furthermore, analysis of the distribution of RNAseq reads in multiH1 cells demonstrates that upon H1 inhibition the proportion of intergenic and intronic reads respect to exonic reads was increased significantly (ca. 16% and 30%, respectively) (Figure 5D). This suggests that H1 depletion increases transcriptional noise, which may favor the accumulation of unexpected RNAs.

H1 abundance at genes deregulated by H1 depletion corresponds to its basal expression level

Additionally, H1 could accumulate at the ISG promoters for repression in the absence of IFN signaling, and severe H1 depletion could be enough to induce several of those genes (as reported elsewhere (63)), including IFN-encoding genes, mimicking an IFN-like response. Using our previously generated ChIP-seq data on H1 variants distribution in T47D cells (25), we analyzed H1 abundance at promoter and coding regions of multiH1sh regulated genes compared to total genes. Genes up- or down-regulated in multiH1 KD cells are not enriched in any particular H1 variant compared to the total genes of the genome, although some H1s including H1.2 are more abundant in up- than down-regulated genes, which likely accounts for the fact that H1.2 is the vari-

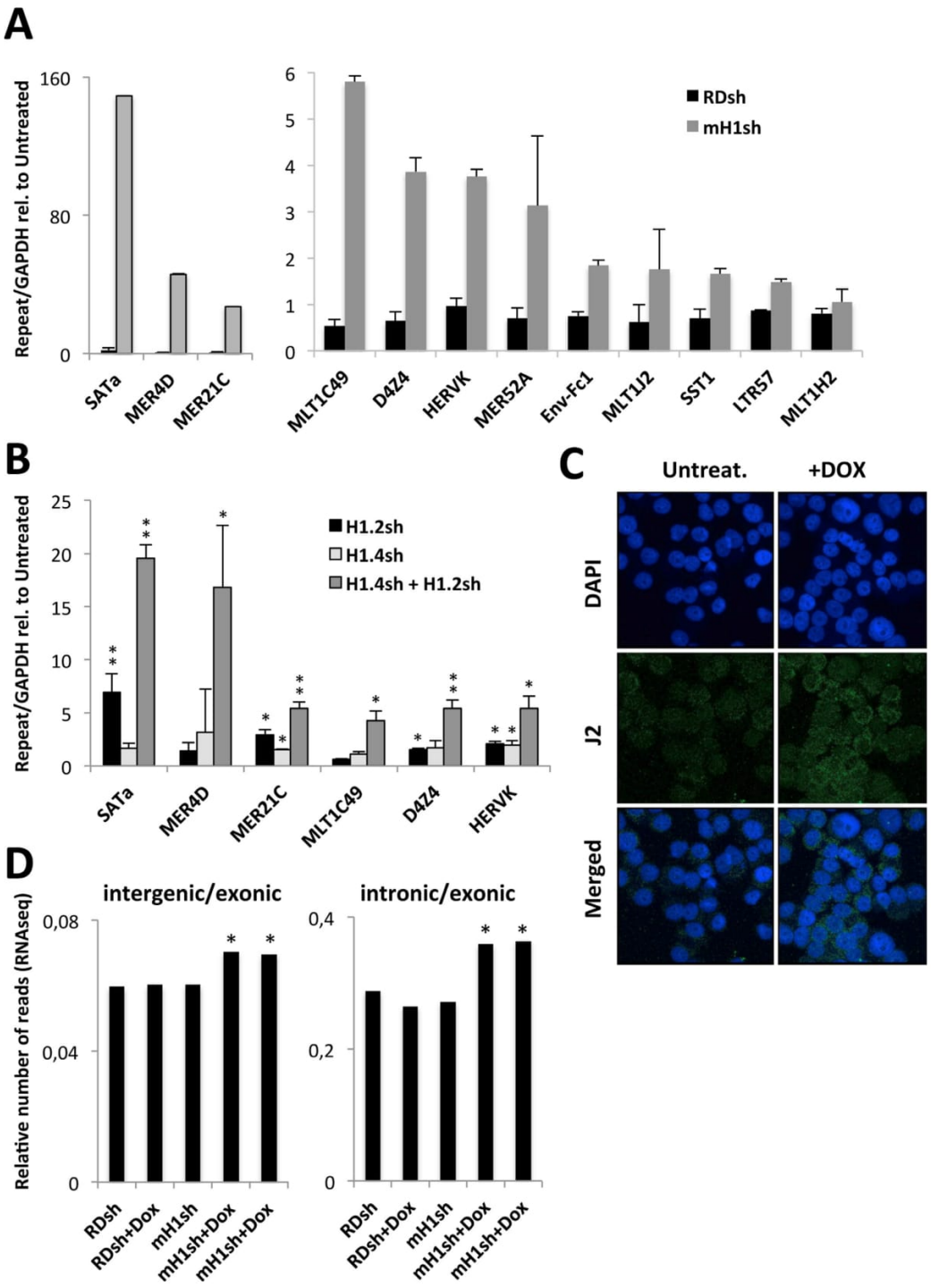


Figure 5. Expression of repetitive DNA upon histone H1 knock-down. (**A**) T47D cells containing RDsh or multiH1sh were treated or not with Dox for 3 days and expression of several repeats and endogenous retroviruses were tested by RT-qPCR. Expression in response to Dox is expressed corrected by GAPDH and relative to untreated cells. (**B**) T47D cells containing H1.2sh, H1.4sh or H1.4/H1.2sh were treated with Dox for 3 days and expression of repeats was tested and represented as in (A). Values represent the mean and SD of representative experiments performed in duplicate. Significance was tested using a *t*-test (**P*-value<0.05; ***P*-value < 0.01). (**C**) Accumulation of dsRNA in multiH1sh cells. Cells were treated or not with Dox for 6 days and submitted to immunofluorescence staining of dsRNA using J2 antibody. DAPI was used to stain the nucleus. (**D**) Relative number of intergenic or intronic RNAseq reads compared to exonic reads in RDsh or multiH1sh treated or not with Dox. Asterisks mark significance ($P < 2e{-}16$) in Dox-treated compared to untreated.

ant that better correlates negatively with gene expression (25), and up-regulated genes present lower basal gene expression as shown in Figure 2G (Supplementary Figure S7B and C). Interestingly, genes within the up-regulated gene set that show the highest responsiveness to IFN according to the Interferome data base, presented a lower basal gene expression average and a higher H1 content around the TSS (data not shown).

H1 variant ChIP-seq signals were similar around TSS of genes (25) up- and down-regulated genes in multiH1sh (Figure 6B and Supplementary Figure S7D). Both sets of genes showed H1.2 depletion around the TSS. Instead, H1.0 and H1.4 showed a local enrichment 1-kb at each side of the TSS of up-regulated genes, while down-regulated genes showed an average depletion of H1 around the TSS. Overall, H1 was more abundant in up-regulated genes, although not as high as in non- or lowly-expressed genes.

Genes highly responsive to multiH1sh and IFN may be tightly repressed with participation of H1.2 or H1.4 (among other variants) and, as a consequence, severe H1 depletion may render these genes open for transcription to take place. An open question is what makes these genes particular for being up-regulated upon H1 depletion. We randomly selected a subset of genes that remained repressed upon H1 reduction and analyzed the content of H1.2 in their proximal promoters. We found that H1 content at up-regulated genes was not significantly different to genes with low expression levels (Supplementary Figure S7E) and indeed H1 content most consistently correlated to expression levels in the absence of doxycycline. In conclusion, an increased H1 content does not seem to be the reason for up-regulation upon H1 depletion.

Gene activation by H1 depletion occurs without canonical marks of active chromatin

We next analyzed H1 variant levels at gene promoters by ChIP-qPCR. We found that H1.2 was present and being removed equally from up-regulated ISGs and uninduced (control) genes upon multiH1 KD, as well as from several DNA repeats and ERVs analyzed (Figure 6A). H1X abundance remained unchanged upon doxycycline treatment, even at induced genes (ISGs and H1.0). As a control, decreased H1s at TSS of active genes (CDK2 and H1.0) compared to an upstream (–3 kb) region, was observed, but not for a repressed gene (NANOG).

In parallel, we analyzed changes in core histone modifications related to gene expression and open chromatin (H3 and H4 pan-acetylation, and H3K4me3), or repression and closed chromatin (H3K9me3, H3K27me3 and H4K20me3). In general, active marks were present at active genes and absent from repeats, as expected. Repressive marks were more abundant at repeats and NANOG. Interestingly, H1.0 activation upon multiH1 KD was only marked by histone acetylation at TSS, but not H3K4me3. In contrast, all histone marks remained unaltered at TSS of all the ISGs analyzed that get activated upon multiH1 KD (Dox treatment) (Figure 6A). This is particularly puzzling for the canonical transcription initiation-associated mark H3K4me3. Differences between ISGs existed, but were not increased with Dox. Repressive marks were also mainly unaltered, and the same happened within repeats. We have previously reported that genes strongly up-regulated upon H1X KD did not present an H1 valley nor H3K4me3 increase at TSS (24), suggesting that gene activation by H1 depletion may occur without canonical marks of active chromatin, as recently reported to occur in certain genes (64). How the ISGs become activated upon H1 depletion without changes in histone marks remains elusive. Activation could be mediated by recruitment of activated IRFs into ISG promoters where active marks are already constitutive and H1 is depleted. It is noteworthy that in an experiment of treatment with IFN-ß, active marks were also not observed to be increased (data not shown).

We also analyzed whether these marks were altered globally by western blot on chromatin extracts. H3K9me3 and H3K27me3 marks were not altered, although HP1α signal was decreased in multiH1 KD cells (Figure 6C). H1.4 with the K26me3 post-translational modification has been proposed to bind the heterochromatin protein HP1α (21,32), suggesting H1.4 depletion could have an effect on chromatin-bound HP1, potentially impacting heterochromatin integrity and repeat repression.

Changes in chromatin accessibility upon H1 depletion

We have also analyzed changes in chromatin accessibility upon multiH1 KD. We assessed chromatin accessibility by ATAC-seq in T47D multiH1sh upon Dox treatment compared to untreated. As a control, accessibility at TSS of genes with unchanged expression was analyzed. Genes highly expressed showed more accessibility than poorly expressed genes (Figure 7A). Genes up and down-regulated in multiH1 KD cells showed slightly different accessibility levels at TSS, and only up-regulated genes showed increased TSS accessibility upon H1 KD (Figure 7B). Next, we searched for peaks of increased accessibility in Dox-treated multiH1sh cells compared to untreated cells within the genome-wide ATAC-seq data. Then, we calculated whether up or down-regulated genes presented an increased number of such newly accessible peaks compared to permutations of randomized genes. Up-regulated genes showed a tendency to have increased number of accessibility peaks both at their promoter and coding regions (P-value = 0.17 and 0.06, respectively), in agreement with its increased transcriptional state in multiH1 KD cells. Conversely, down-regulated genes showed reduced accessibility (P-value = 0.18 and 0.03 for promoter and coding regions, respectively).

Finally, we computed the genomic distribution of accessibility peaks and whether peaks were enriched within some category of repetitive DNA. Upon H1 depletion, accessibility peaks were enriched within introns and distal intergenic regions (Figure 7C), and at satellites, rDNA and simple repeats (Figure 7D). Among satellites where peaks were found enriched, we found SATa and SST1 which were up-regulated upon multiH1 KD (Figure 5A). These results support the notion that H1 depletion generates regions of open chromatin, including regions where repeats were repressed, that might generate synthesis of (ds)RNAs that are sensed at cytoplasm and induce the IFN signaling response.

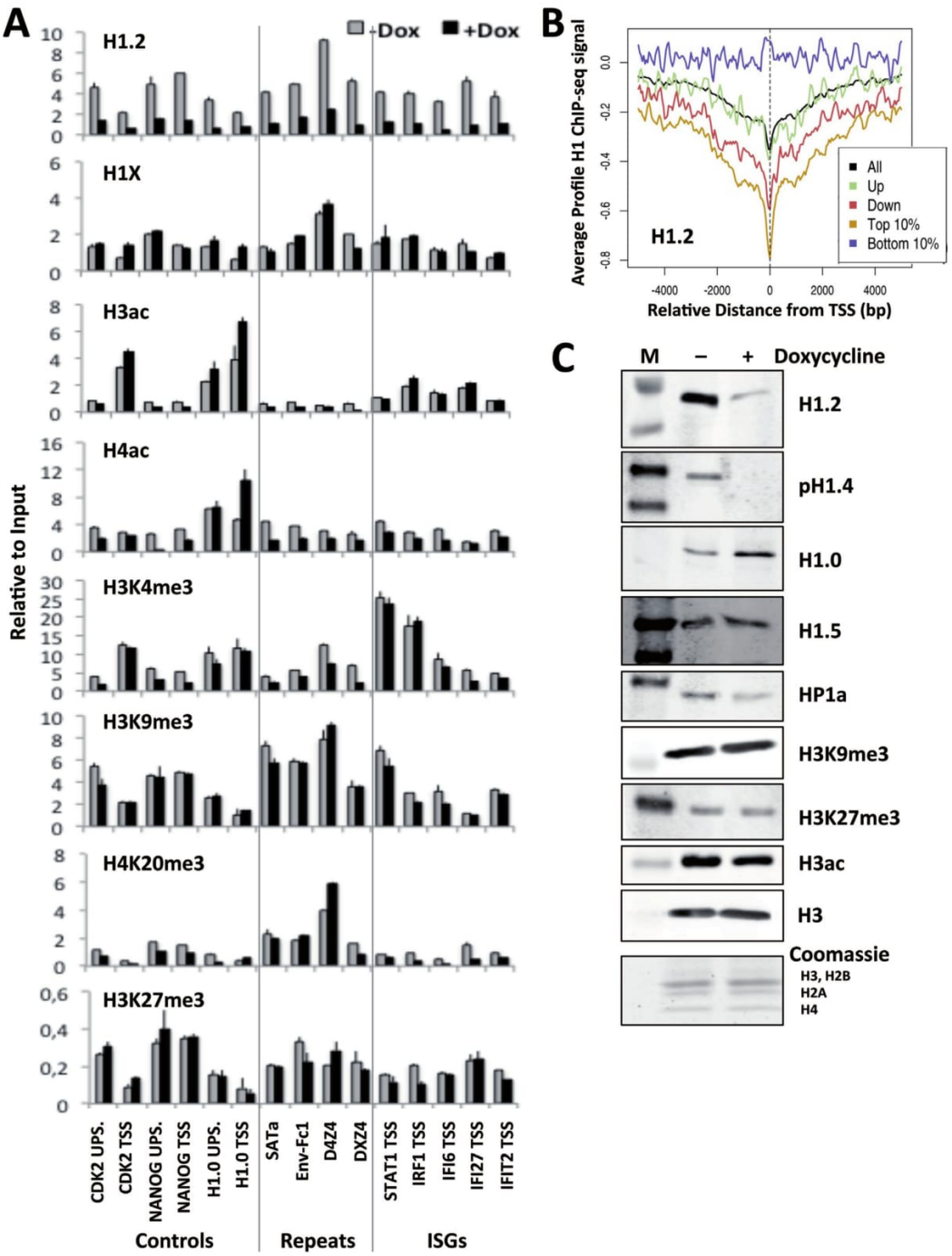
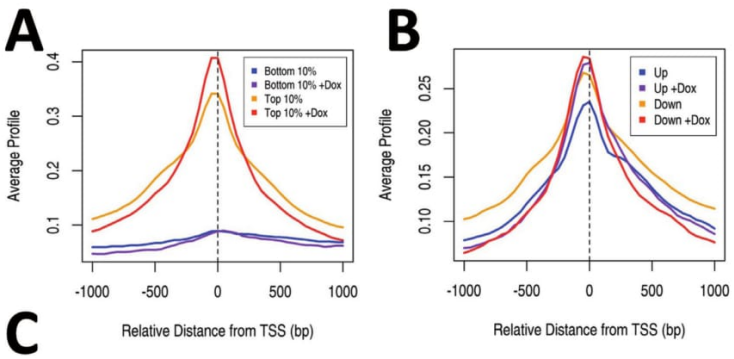
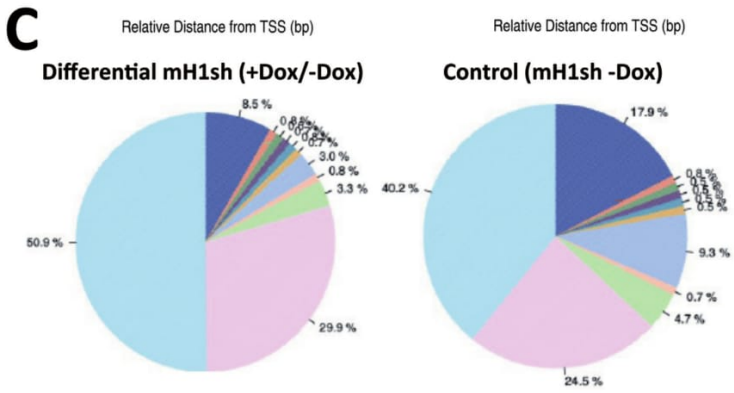


Figure 6. Changes at ISG promoters and chromatin upon H1 depletion. (**A**) Histone modification changes at promoters upon multiH1 KD. T47D cells containing multiH1sh were treated or not with Dox for 6 days and chromatin was extracted and immunoprecipitated (ChIP) with the indicated antibodies. Abundance at specific genomic regions (ISGs, DNA repeats and control genes) was quantified by real-time PCR and expressed relative to input DNA amplification. Oligonucleotides at the indicated promoters correspond to TSS or -3kb distal promoter (UPS). Values represent the mean and SD of a representative experiment performed in duplicate. (**B**) H1.2 abundance at TSS of deregulated genes. Average density profile of H1.2 abundance (input-subtracted ChIP-seq signal) around the TSS of up- or down-regulated genes upon multiH1 KD. RefSeq genes (All), as well as bottom and top 10% expressed genes (groups 1 and 10 in Supplementary Figure S7A), are shown as reference. (**C**) Chromatin-bound HP1α is decreased upon multiH1 KD. T47D cells containing multiH1sh were treated with Dox for 6 days and chromatin was extracted and immunoblotted with the indicated antibodies. Coomassie staining of core histones is shown.





	Differ. +Dox/-Dox	Control -Dox	Genome
Promoter (<1000 bp)	8,5	17,9 *	1,1
Promoter (1000-2000 bp)	0,8	0,8	0,7
Promoter (2000-3000 bpp)	0,6	0,5	0,6
Downstream (<1000 bp)	0,7	0,5	0,9
Downstream (1000-2000 bp)	0,8 *	0,5	0,7
Downstream (2000-3000 bp)	0,7	0,5	0,6
5'UTR	3	9,3 *	0,4
3'UTR	0,8	0,7	1,4
Coding exon	3,3	4,7 *	1,9
Intron	29,9 *	24,5	41,6
Distal intergenic	50,9 *	40,2	50

*p-value<0,001

D

Repeat category	No.of peaks	
DNA	354	
LINE	1258	
Low_complexity	1073	*
LTR	1282	
Other	166	*
RC	1	
RNA	0	
rRNA	25	*
Satellite	3551	*
scRNA	0	
Simple_repeat	3866	*
SINE	2016	
snRNA	4	
srpRNA	4	
tRNA	9	
Unknown	0	

Satellites	No.of peaks
ALR/Alpha	966
(CATTC)n	580
CER	26
D20S16	124
(GAATG)n	1074
GSAT	55
GSATII	159
GSATX	41
HSAT4	20
HSATII	368
MSR1	19
SATR1	41
SST1	24
TAR1	12

Figure 7. Changes on chromatin accessibility upon depletion of histone H1. T47D cells containing multiH1sh were treated or not with Dox for 6 days and submitted to ATAC-seq. (A and B) Average profiles of accessibility around the center of TSS, normalized to reads depth, for the following groups of genes: bottom and top 10% expressed genes (groups 1 and 10 in Supplementary Figure S7A) (**A**), and up- and down-regulated genes in multiH1 KD cells (**B**). (**C**) Genomic annotation of high accessibility regions (narrow peaks) enriched in multiH1sh Dox-treated compared to untreated cells, shown as pie diagram and table (% of total). For comparison, the genomic annotation of accessibility sites in control cells (untreated) is shown. Asterisks mark genomic categories enriched in differential mH1sh accessibility peaks compared to control or vice versa (P-value < 0.001). (**D**) Distribution of high accessibility peaks enriched in multiH1sh cells among the different annotated categories of genomic DNA repeats. Asterisks mark those repeat categories where the number of peaks obtained is significantly higher (P-value < 0.001) than expected in randomized samples (10^4 permutations) mimicking the accessibility peaks file. Within the satellite category, types that were significantly enriched are shown in the right panel.

Low expression of replication-dependent H1 variants in pancreatic adenocarcinomas over-expressing genes of the IFN pathway

Seeing a potential link between H1 variant levels and activation of the IFN response, we looked for other models in which this link may be apparent. Resistance to oncolytic viral therapy in pancreatic adenocarcinomas and multiple myelomas has been associated with overexpression of the anti-viral innate immune response (65). Monsurrò *et al.* identified two molecular phenotypes of pancreatic cancer characterized by a differential expression of genes associated with IFN signaling (66). We have constructed a cluster dendrogram of the expression of IFN signaling related genes in pancreatic cancer and normal samples obtained from databases that has defined three groups of samples, where group 1 contain mainly normal samples and groups 2 and 3 contain most tumor samples (Supplementary Figure S8). Groups 2 and 3 show overall up-regulation of IFN-related genes compared to group 1, while no differences in expression of randomly chosen genes exist among groups (Figure 8). Next we plotted the expression of H1 variants in all the samples of the three groups, and observed that many replication-dependent H1 variants are down-regulated in sample groups presenting ISGs up-regulation, while replication-independent variants are up-regulated, resembling the situation in our T47D multiple 'replication-dependent' H1 KD where ISGs are up-regulated.

However, most cancer types show an up-regulation of H1-encoding genes compared to normal tissue, including breast carcinoma (10). An interesting reported exception is the low H1.2 expression in colorectal cancer compared to normal tissue. ISG expression was not increased in colorectal cancer compared to normal samples despite of the low H1.2 expression in cancer samples, indicating that the correlation in pancreatic cancer does not extend to colorectal cancer where only H1.2 is down-regulated (data not shown). This is consistent with our results, wherein we only register an IFN response in multi-variant down-regulation.

DISCUSSION

This work extends our research on the biological effects of depleting histone H1 variants in cancer cells, by adding shRNA-expressing vectors against H1.4 and simultaneous depletion of several variants with a single shRNA (multiH1), or combining shRNAs. We reported before that individual H1 KDs alter expression of small subsets of genes in T47D cells, with little overlap between variants. Knocking-down individual variants affects proliferation to different degrees. Our experiments show that any H1 variant KD is unable to completely impair cell growth, whether because depletion is incomplete or because H1 variants present some redundancy. Instead, simultaneous depletion of multiple H1 variants has drastic consequences on cell proliferation. We have tested different combinations of H1 KDs, and found co-KD of H1.2 + H1.4 to be the combination with greatest inhibition of cell growth (data not shown). Our results suggest that these two variants are most important for cell homeostasis. Either the two variants play redundant roles, or the observed consequences of multiH1s are due to the additive effects of impairing H1.2 and H1.4 non-redundant functions. Because distribution of H1.2 and H1.4 in the genome is not redundant (25), we favor the second explanation.

Nonetheless, the gene expression profile of multiH1 KD cells is not explained by simple addition of the H1.2 and H1.4 KD profiles. Combination of gene expression profiling and ChIP-seq data of H1s variants fails to show that these genes are specifically targeted and directly regulated by the corresponding variant. Multiple H1 KD present a considerable number of genes deregulated; roughly, 3-times the number of the sum of genes deregulated in H1.2 and H1.4 single KDs. Interestingly, a big proportion of genes, in particular the strongest up-regulated genes, are related to the IFN response. It is worth mentioning that genes up-regulated upon multiH1 KD are genes that are silenced under basal conditions. The opposite is true for the down-regulated genes. As before, we do not support the hypothesis that all these ISGs are maintained repressed particularly by H1.2 and H1.4, as others have suggested (63). Although we have found that H1s are enriched in the promoter and coding regions of genes up-regulated in multiH1 KD compared to other genes, we show that this H1 content is characteristic of genes presenting that low basal expression rates and do not represent especially silenced H1 targets. We have shown before that H1 content correlates positively with gene silencing in T47D, with H1.2 having the highest degree of correlation (24,25).

We believe ISG induction in multiH1 KD may be caused by the production and cytosolic sensing of unusual nucleic acids, and secondarily through the synthesis of IFN. We show here that several intermediate components of the IFN signaling pathway are up-regulated upon doxycycline treatment of multiH1 KD cells, type-I IFN is being expressed and liberated to the media, and ISG stimulation is depending on the presence of several pathogen recognition receptors (MDA5, TLR3) and adaptor molecules (MAVS, STING), on type-I IFN receptors, and on TBK and JAK activation.

Several recent reports have shown that DNMT inhibitors with anti-proliferative effects induce the IFN response pathway in cancer cells by inducing synthesis of dsRNAs from endogenous retroviruses, DNA repeats and non-coding transcripts (60–62). The striking similarity of these reports with our observations led us to investigate whether multiH1 depletion was leading to spurious transcription from non-coding RNA genes, DNA repeats, satellites or endogenous retroviruses (ERVs), generally located in heterochromatin. Enrichment pathway analysis of our multiH1 KD expression data indicated that the IFN response is induced by RNA transcripts, not by DNA, as only genes belonging to the RNA-sensing pathway are up-regulated. Accordingly, we detected expression of satellite repeats and ERVs that might be induced upon heterochromatin relaxation. Fittingly, both variants depleted by our multiH1sh have been associated with maintaining heterochromatin. First, H1.4 has been suggested to associate with heterochromatin through its ability to recruit HP1, L3MBTL1 and polycomb proteins through H1.4K26me (21,22,32,33,67,68). H1.2 may have multiple roles, as it has been associated with active transcription and repression and in complex with different partners (30,31,69). In T47D cells, we have charac-

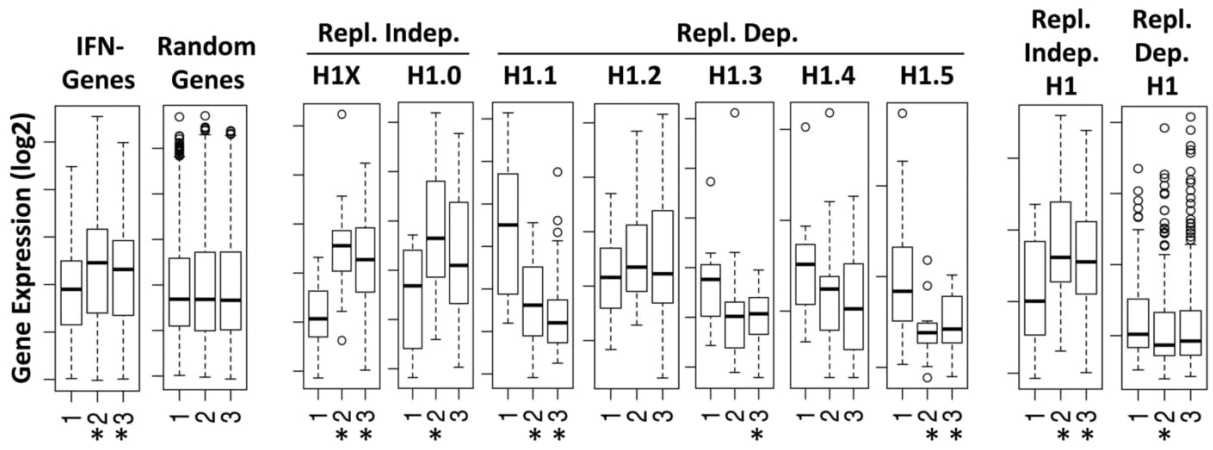


Figure 8. H1 expression in pancreatic normal and cancer samples clustered according to the expression of IFN signaling related genes. Supervised cluster analysis of IFN related genes in 36 pancreatic tumor and 16 normal samples using microarray data deposited in GSE16515, defined 3 groups of samples (see Supplementary Figure S8). Box plots represent the expression of IFN related genes and H1 variants, ether individually or grouped as replication dependent or independent H1s, in the three groups of samples. Expression of a representative set of genes chosen randomly is shown as a control. Significance was assessed using the Kolmogorov-Smirnov test. Significant difference between groups 2 or 3 and group 1 is marked with asterisks ($P < 0.05$). Relative expression units have been omitted for simplicity.

terized H1.2 as enriched within lamin-associated domains (LADs), intergenic regions, repressed chromatin and gene-poor chromosomes (25).

We have tried to establish whether derepression of repeats and induction of the IFN response is due to overall decrease on the H1 content, or to the specific depletion of H1.2 + H1.4. Single KD of these variants does not cause these effects in T47D cells. Other double KD combinations are less effective (H1.2 + H1.5) or ineffective (H1.4 + H1.5). This would suggest that H1.2 is important to maintain repression, but not sufficient alone, and requires H1.4 also participating.

An alternate hypothesis is that total H1 content could be the key determinant of the observed multiH1 KD effects. Indeed, up-regulation of ISGs in T47D multiH1 KD was complemented equally by H1.2, H1.3 or H1.4 overexpression, going against the notion that H1.2 is essential, and favoring redundancy between variants. We reported elsewhere the approximate content of H1 variants in T47D cells, representing H1.2, H1.4 and H1.5, 20–30% of the total H1 content each (26). Then, assuming complete depletion, H1.2 + H1.4 inhibition should not represent a decrease on the total H1 content larger than the inhibitions H1.2 + H1.5, or H1.4 + H1.5. In fact, Coomassie staining of histone extracts has shown that total H1 content was reduced to 70–80% of wild-type in multiH1 or H1.4/H1.2 KD cells. H1 content reduction in the other double KDs (H1.2/H1.5 and H1.4/H1.5) we generated was comparable (down to 60–80%) despite of incomplete H1.5 depletion. Furthermore, simultaneous depletion of H1.3/H1.4/H1.5 rendered ISGs unaltered. In conclusion, the observed effects are better explained by the specificity of H1.2 and H1.4 functions than by the reduction of overall H1 content.

Because total H1 content and the abundance or distribution of the different variants may differ between cell types, KD of single or multiple variants may have different outcomes within different models. Despite of this, we have detected ISG induction also in MCF7 multiH1 KD cells. Interestingly, by analyzing publically available transcriptomic data, we have detected ISG up-regulation in H1.2 (and EZH2) KD experiments performed by another group in MCF7 (31) and 293T cells (30) (data not shown), although induction of the IFN response was not reported in those studies. Both data sets show a better overlap with our T47D multiH1sh up-regulated genes subset than with the H1.2sh up-regulated genes. This indicates that H1 variant depletion has different transcriptional effects in different cell types, and H1.2 depletion is not sufficient to induce the IFN response in T47D cells but induces to some extent several ISGs in HeLa (see below) (63), MCF7 and 293T, further arguing in favor of a special role for H1.2. All together these studies confirm the notion that H1 depletion may trigger this innate immune response in several cell types.

Single and double H1 knock-outs (KO) in mice have limited effects due to compensation and redundancy, but drastic effects were observed in triple KO (TKO) embryos and MEFs due to reduction of H1 total content to 50% (27,36,37). Yang *et al.* demonstrated that some H1 subtypes specifically interact with the DNA methyltransferases DNMT1 and DNMT3B to promote methylation of imprinted genes, which become up-regulated in mouse TKO (27,38). Major satellites are also up-regulated in mouse TKO (39), and because DNMTis induce the IFN response pathway in cancer cells by inducing synthesis of DNA repeats (60–62), we hypothesized that multiple H1 depletion in breast cancer cells could lead to DNMTs displacement and demethylation of heterochromatic repeats leading to IFN signaling. Then, treatment with DNMTis and H1 depletion would have similar effects. Treatment of T47D cells with 5-aza-2′-deoxycytidine (aza-dC) for 3 days did not induce the expression of ISGs, but enhanced the effect of multiH1 KD (Supplementary Figure S9). Response of the DNA

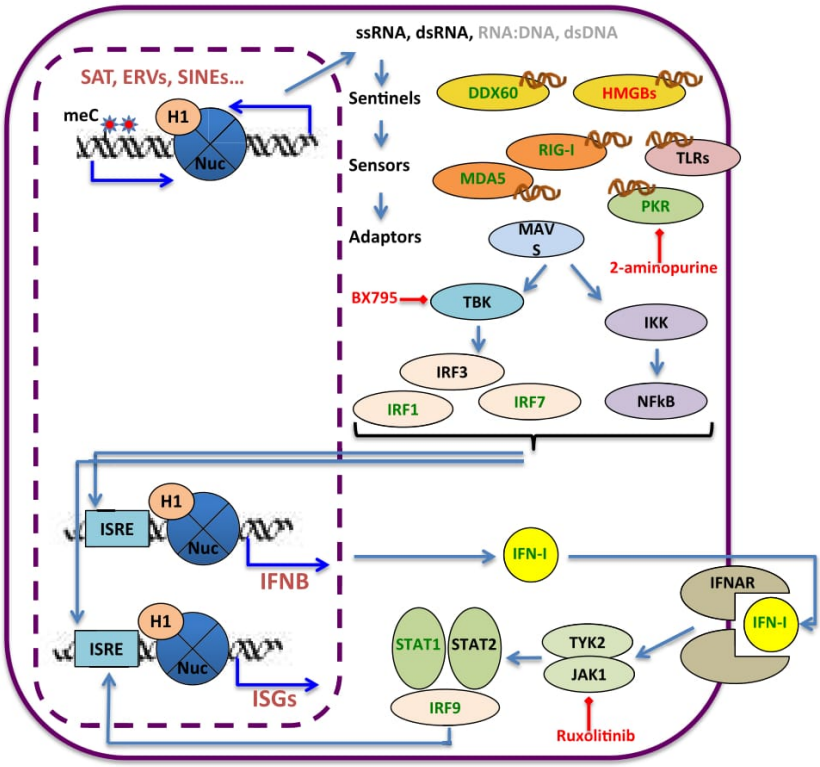


Figure 9. Model of H1 involvement in the repression of the interferon response. Histone H1 participates in the repression of strongly silenced genes such as ISGs under basal conditions, and in the repression of non-coding transcripts derived from repeats such as satellites, transposons and endogenous retroviruses located in heterochromatic regions which may cause, upon induction in H1-depleted cells, activation of the IFN signaling pathway. Transcription of repeats may give raise to ss/dsRNAs (60–62) or non-B DNA structures such as R-loops (70), that could be sensed by universal (HMGBs) or ligand-specific cytosolic sentinels (DDX60) that activate RNA (MDA5/IFIH1, RIG-I, LGP2, OASs) or DNA sensors (cGAS, IFI16), or by membrane-associated Toll-like receptors (TLRs). These proteins transduce the signal to specific adaptors (MAVS/IPS1, STING), leading to the activation of TBK-IRFs and IKK-NFkB. These transcription factors bind specific sites at ISG promoters, including genes for the synthesis of type I IFN that upon interaction with specific receptors lead to the activation of JAK-STAT pathway and additional IRFs. For simplicity, few of the sensor and adaptor proteins are drawn. ISGs up-regulated in multiH1 KD cells include DDX60, MDA5/IFIH1, RIG-I/DDX58, LGP2, PKR, OAS1/IFI4, OAS2, OASL, IRF1, IRF7, IRF9, IFNB, STAT1 (in green). HMGB1, HMGB2 and HMGB3 are down-regulated. Inhibitors of members of the IFN signaling pathway are shown (in red).

repeats to aza-dC was heterogeneous (data not shown). So, in T47D cells, DNMTis do not induce ISGs nor repeats as observed in multiH1 KD, so we do not favor the model in which the effects observed upon H1 depletion are mediated only by DNA demethylation. Nonetheless, it was interesting to address the global methylation state of DNA in multiH1 KD cells. In this direction, we have evaluated changes on global DNA methylation upon multiH1 KD with a genome-wide CpG array that showed virtually no differentially methylated CpGs after six days of doxycycline treatment. Methylation of several ERV elements was also analyzed by bisulphite PCR and changes were not found (Supplementary Figure S9).

Due to all evidences reported here, we favor a model in which histone H1 participates in the repression of genes such as ISGs under basal conditions, and in the repression of non-coding transcripts derived from repeats such as satellites, transposons and endogenous retroviruses located in heterochromatic regions which may cause, upon induction in H1-depleted cells, activation of the IFN response (Figure 9). We have directly observed the increase of cytosolic dsRNA within multi H1 KD cells, increased expression of satellites and repeats, enhanced accumulation of RNAseq reads from intergenic and intronic regions, and increased chromatin accessibility of all these regions. It has been previously shown that tumor cells may contain retroelements-derived DNA in the cytosol forming structures including R-loops, whose levels are increased by genotoxic agents, which activate the IFN response (70). We cannot disregard the possibility that upon expression of repeat elements, R-loops are formed and RNA:DNA or triple stranded fragments are detected in the cytosol by sensors that induce the IFN pathway (71). Overall, our data suggest that activation of ISGs by H1 KD is mediated by activating the IFN signaling pathway, rather than by direct promoter relaxation due to H1 displacement.

ChIP analysis of ISG promoters upon multiH1 KD has shown that active marks are not increased as it would be expected for genes that are induced strongly. It has been recently reported that developmental genes are activated

in the absence of positive marks with the participation of program-specific transcription factors (TFs), allowing rapid activation and deactivation (64). ISGs may fall in this category, where clearance of histone H1, together with binding of IRFs and NFkB to the promoters, may be sufficient for activation. Forced depletion of H1 may be enough to create a permissive chromatin environment to allow transcription without the need of the contribution of active histone marks. Despite of this, H1 clearance may not be sufficient without upstream induction of the IFN signaling pathway that activate NFkB and IRFs to bind to the ISG promoters. This may explain why the control genes analyzed (and the large proportion of the transcriptome), that also show H1.2 clearance, are not induced upon multiH1 KD, favoring the model that IFN signaling is required. Moreover, our ChIP data and data extracted from the UCSC browser shows that many ISGs present significant amounts of H3K4me3 under basal conditions. This is a particularity of genes that respond rapidly to stimulus, and may explain why upon induction these genes do not show increased levels of this mark, neither upon IFN-ß treatment.

A role of H1 in the regulation of ISGs has been suggested before. The H1 chaperone SET/TAF-I has been suggested to participate in the H1-mediated silencing of ISG15, ISG54, ISG56 (IFIT1) and IFITM1 in HeLa cells (63). In the absence of IFN, these genes were discretely up-regulated (two-fold, in average) not only in TAF-I KD but also in H1.2 KD cells. We have confirmed that TAF-I KD causes modest up-regulation of some ISGs in T47D cells (data not shown), but H1.2 KD does not in our hands, until combined with another H1 (H1.4) variant KD. Then, induction fold-changes are much higher (Figure 3A), further suggesting multiH1 KD induces ISGs not only due to its local contribution to promoter chromatin structure, but also inducing the IFN pathway upstream due to derepression of DNA repeats. Whether this IFN signaling is caused by heterochromatin instability and expression of repeats, originated in the altered H1 expression, DNA hypomethylation or other chromatin alterations, are appealing possibilities that need to be explored.

DATA AVAILABILITY

RNAseq data is available in the Gene Expression Omnibus (GEO) database under the accession number GSE83277. ATACseq data accession number is GSE100762.

SUPPLEMENTARY DATA

Supplementary Data are available Online.

ACKNOWLEDGEMENTS

We thank former members of the laboratory Regina Mayor, Jean-Michel Terme and Neus Luque for discussions, Alicia Roque for critical reading of the manuscript, and Marian Martínez-Balbas, Ferran Azorín and Jordi Bernués for sharing reagents and discussions. We thank Javier Martínez-Picado for sharing with us the RT inhibitors. We thank CRG and CNAG sequencing core facilities for the ATAC-seq and RNAseq experiments, respectively.

FUNDING

Spanish Ministry of Economy and Competitiveness (MINECO) and European Regional Development Fund [BFU2014–52237-P]. A.B. received a predoctoral fellowship from SENESCYT from Ecuador. M.D. was recipient of a fellowship from MINECO [PTA2014–09515-I]. A.E.-C. was funded by the RED-BIO project of the Spanish National Bioinformatics Institute (INB) [PT13/0001/0044]. The INB is funded by the Spanish National Health Institute Carlos III (ISCIII) and MINECO. Funding for open access charge: Spanish Ministry of Economy and Competitiveness (MINECO) and European Regional Development Fund [BFU2014-52237-P].
Conflict of interest statement. None declared.

REFERENCES

1. Bednar,J., Horowitz,R.A., Grigoryev,S.A., Carruthers,L.M., Hansen,J.C., Koster,A.J. and Woodcock,C.L. (1998) Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 14173–14178.
2. Pennings,S., Meersseman,G. and Bradbury,E.M. (1994) Linker histones H1 and H5 prevent the mobility of positioned nucleosomes. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 10275–10279.
3. Laybourn,P.J. and Kadonaga,J.T. (1991) Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. *Science*, **254**, 238–245.
4. Happel,N. and Doenecke,D. (2009) Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene*, **431**, 1–12.
5. Izzo,A., Kamieniarz,K. and Schneider,R. (2008) The histone H1 family: specific members, specific functions? *Biol. Chem.*, **389**, 333–343.
6. Millan-Arino,L., Izquierdo-Bouldstridge,A. and Jordan,A. (2016) Specificities and genomic distribution of somatic mammalian histone H1 subtypes. *Biochim. Biophys. Acta*, **1859**, 510–519.
7. Terme,J.M., Sese,B., Millan-Arino,L., Mayor,R., Izpisua Belmonte,J.C., Barrero,M.J. and Jordan,A. (2011) Histone H1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency. *J. Biol. Chem.*, **286**, 35347–35357.
8. Pan,C. and Fan,Y. (2016) Role of H1 linker histones in mammalian development and stem cell differentiation. *Biochim. Biophys. Acta*, **1859**, 496–509.
9. Sato,S., Takahashi,S., Asamoto,M., Nakanishi,M., Wakita,T., Ogura,Y., Yatabe,Y. and Shirai,T. (2012) Histone H1 expression in human prostate cancer tissues and cell lines. *Pathol. Int.*, **62**, 84–92.
10. Scaffidi,P. (2016) Histone H1 alterations in cancer. *Biochim. Biophys. Acta*, **1859**, 533–539.
11. Medrzycki,M., Zhang,Y., McDonald,J.F. and Fan,Y. (2012) Profiling of linker histone variants in ovarian cancer. *Front. Biosci. (Landmark Ed.)*, **17**, 396–406.
12. Meergans,T., Albig,W. and Doenecke,D. (1997) Varied expression patterns of human H1 histone genes in different cell lines. *DNA Cell Biol.*, **16**, 1041–1049.
13. Parseghian,M.H. and Hamkalo,B.A. (2001) A compendium of the histone H1 family of somatic subtypes: an elusive cast of characters and their characteristics. *Biochem. Cell Biol.*, **79**, 289–304.
14. Zhang,Y., Cooke,M., Panjwani,S., Cao,K., Krauth,B., Ho,P.Y., Medrzycki,M., Berhe,D.T., Pan,C., McDevitt,T.C. *et al.* (2012) Histone h1 depletion impairs embryonic stem cell differentiation. *PLoS Genet.*, **8**, e1002691.
15. Torres,C.M., Biran,A., Burney,M.J., Patel,H., Henser-Brownhill,T., Cohen,A.S., Li,Y., Ben-Hamo,R., Nye,E., Spencer-Dene,B. *et al.* (2016) The linker histone H1.0 generates epigenetic and functional intratumor heterogeneity. *Science*, **353**, aaf1644.
16. Kalashnikova,A.A., Rogge,R.A. and Hansen,J.C. (2016) Linker histone H1 and protein-protein interactions. *Biochim. Biophys. Acta*, **1859**, 455–461.

17. Izzo,A. and Schneider,R. (2016) The role of linker histone H1 modifications in the regulation of gene expression and chromatin dynamics. *Biochim. Biophys. Acta*, **1859**, 486–495.
18. Hergeth,S.P., Dundr,M., Tropberger,P., Zee,B.M., Garcia,B.A., Daujat,S. and Schneider,R. (2011) Isoform-specific phosphorylation of human linker histone H1.4 in mitosis by the kinase Aurora B. *J. Cell Sci.*, **124**, 1623–1628.
19. Kamieniarz,K., Izzo,A., Dundr,M., Tropberger,P., Ozretic,L., Kirfel,J., Scheer,E., Tropel,P., Wisniewski,J.R., Tora,L. *et al.* (2012) A dual role of linker histone H1.4 Lys 34 acetylation in transcriptional activation. *Genes Dev.*, **26**, 797–802.
20. Weiss,T., Hergeth,S., Zeissler,U., Izzo,A., Tropberger,P., Zee,B.M., Dundr,M., Garcia,B.A., Daujat,S. and Schneider,R. (2010) Histone H1 variant-specific lysine methylation by G9a/KMT1C and Glp1/KMT1D. *Epigenet. Chromatin*, **3**, 7.
21. Hale,T.K., Contreras,A., Morrison,A.J. and Herrera,R.E. (2006) Phosphorylation of the linker histone H1 by CDK regulates its binding to HP1alpha. *Mol. Cell*, **22**, 693–699.
22. Vaquero,A., Scher,M., Lee,D., Erdjument-Bromage,H., Tempst,P. and Reinberg,D. (2004) Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. *Mol. Cell*, **16**, 93–105.
23. Kim,K., Jeong,K.W., Kim,H., Choi,J., Lu,W., Stallcup,M.R. and An,W. (2012) Functional interplay between p53 acetylation and H1.2 phosphorylation in p53-regulated transcription. *Oncogene*, **31**, 4290–4301.
24. Mayor,R., Izquierdo-Bouldstridge,A., Millan-Arino,L., Bustillos,A., Sampaio,C., Luque,N. and Jordan,A. (2015) Genome distribution of replication-independent histone H1 variants shows H1.0 associated with nucleolar domains and H1X associated with RNA polymerase II-enriched regions. *J. Biol. Chem.*, **290**, 7474–7491.
25. Millan-Arino,L., Islam,A.B., Izquierdo-Bouldstridge,A., Mayor,R., Terme,J.M., Luque,N., Sancho,M., Lopez-Bigas,N. and Jordan,A. (2014) Mapping of six somatic linker histone H1 variants in human breast cancer cells uncovers specific features of H1.2. *Nucleic Acids Res.*, **42**, 4474–4493.
26. Sancho,M., Diani,E., Beato,M. and Jordan,A. (2008) Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. *PLoS Genet.*, **4**, e1000227.
27. Fan,Y., Nikitina,T., Zhao,J., Fleury,T.J., Bhattacharyya,R., Bouhassira,E.E., Stein,A., Woodcock,C.L. and Skoultchi,A.I. (2005) Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell*, **123**, 1199–1212.
28. Shen,X. and Gorovsky,M.A. (1996) Linker histone H1 regulates specific gene expression but not global transcription in vivo. *Cell*, **86**, 475–483.
29. Lin,Q., Inselman,A., Han,X., Xu,H., Zhang,W., Handel,M.A. and Skoultchi,A.I. (2004) Reductions in linker histone levels are tolerated in developing spermatocytes but cause changes in specific gene expression. *J. Biol. Chem.*, **279**, 23525–23535.
30. Kim,K., Lee,B., Kim,J., Choi,J., Kim,J.M., Xiong,Y., Roeder,R.G. and An,W. (2013) Linker Histone H1.2 cooperates with Cul4A and PAF1 to drive H4K31 ubiquitylation-mediated transactivation. *Cell Rep.*, **5**, 1690–1703.
31. Kim,J.M., Kim,K., Punj,V., Liang,G., Ulmer,T.S., Lu,W. and An,W. (2015) Linker histone H1.2 establishes chromatin compaction and gene silencing through recognition of H3K27me3. *Sci. Rep.*, **5**, 16714.
32. Daujat,S., Zeissler,U., Waldmann,T., Happel,N. and Schneider,R. (2005) HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. *J. Biol. Chem.*, **280**, 38090–38095.
33. Kuzmichev,A., Jenuwein,T., Tempst,P. and Reinberg,D. (2004) Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol. Cell*, **14**, 183–193.
34. Krishnakumar,R., Gamble,M.J., Frizzell,K.M., Berrocal,J.G., Kininis,M. and Kraus,W.L. (2008) Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. *Science (New York, N.Y.)*, **319**, 819–821.
35. Vicent,G.P., Nacht,A.S., Font-Mateu,J., Castellano,G., Gaveglia,L., Ballare,C. and Beato,M. (2011) Four enzymes cooperate to displace histone H1 during the first minute of hormonal gene activation. *Genes Dev.*, **25**, 845–862.
36. Fan,Y., Sirotkin,A., Russell,R.G., Ayala,J. and Skoultchi,A.I. (2001) Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the H1(0) replacement subtype. *Mol. Cell. Biol.*, **21**, 7933–7943.
37. Fan,Y., Nikitina,T., Morin-Kensicki,E.M., Zhao,J., Magnuson,T.R., Woodcock,C.L. and Skoultchi,A.I. (2003) H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. *Mol. Cell. Biol.*, **23**, 4559–4572.
38. Yang,S.M., Kim,B.J., Norwood Toro,L. and Skoultchi,A.I. (2013) H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 1708–1713.
39. Cao,K., Lailler,N., Zhang,Y., Kumar,A., Uppal,K., Liu,Z., Lee,E.K., Wu,H., Medrzycki,M., Pan,C. *et al.* (2013) High-resolution mapping of h1 linker histone variants in embryonic stem cells. *PLoS Genet.*, **9**, e1003417.
40. Geeven,G., Zhu,Y., Kim,B.J., Bartholdy,B.A., Yang,S.M., Macfarlan,T.S., Gifford,W.D., Pfaff,S.L., Verstegen,M.J., Pinto,H. *et al.* (2015) Local compartment changes and regulatory landscape alterations in histone H1-depleted cells. *Genome Biol.*, **16**, 289.
41. Wiznerowicz,M. and Trono,D. (2003) Conditional suppression of cellular genes: lentivirus vector-mediated drug-inducible RNA interference. *J. Virol.*, **77**, 8957–8961.
42. Perez Losada,A., Woessner,S., Sole,F., Florensa,L. and Bonet,C. (1994) Chromosomal and in vitro culture studies in a case of primary plasma cell leukemia. *Cancer Genet. Cytogenet.*, **76**, 36–38.
43. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
44. Li,B. and Dewey,C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
45. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
46. Rusinova,I., Forster,S., Yu,S., Kannan,A., Masse,M., Cumming,H., Chapman,R. and Hertzog,P.J. (2013) Interferome v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res.*, **41**, D1040–D1046.
47. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
48. Young,M.D., Wakefield,M.J., Smyth,G.K. and Oshlack,A. (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*, **11**, R14.
49. Supek,F., Bosnjak,M., Skunca,N. and Smuc,T. (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, **6**, e21800.
50. Wu,G., Feng,X. and Stein,L. (2010) A human functional protein interaction network and its application to cancer data analysis. *Genome Biol.*, **11**, R53.
51. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
52. Bao,W., Kojima,K.K. and Kohany,O. (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA*, **6**, 11.
53. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
54. Buenrostro,J.D., Giresi,P.G., Zaba,L.C., Chang,H.Y. and Greenleaf,W.J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*, **10**, 1213–1218.
55. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
56. Shin,H., Liu,T., Manrai,A.K. and Liu,X.S. (2009) CEAS: cis-regulatory element annotation system. *Bioinformatics*, **25**, 2605–2606.
57. Gel,B., Diez-Villanueva,A., Serra,E., Buschbeck,M., Peinado,M.A. and Malinverni,R. (2016) regioneR: an R/Bioconductor package for

the association analysis of genomic regions based on permutation tests. *Bioinformatics*, **32**, 289–291.
58. Chawla-Sarkar,M., Lindner,D.J., Liu,Y.F., Williams,B.R., Sen,G.C., Silverman,R.H. and Borden,E.C. (2003) Apoptosis and interferons: role of interferon-stimulated genes as mediators of apoptosis. *Apoptosis*, **8**, 237–249.
59. Thyrell,L., Erickson,S., Zhivotovsky,B., Pokrovskaja,K., Sangfelt,O., Castro,J., Einhorn,S. and Grander,D. (2002) Mechanisms of Interferon-alpha induced apoptosis in malignant cells. *Oncogene*, **21**, 1251–1262.
60. Chiappinelli,K.B., Strissel,P.L., Desrichard,A., Li,H., Henke,C., Akman,B., Hein,A., Rote,N.S., Cope,L.M., Snyder,A. *et al.* (2015) Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including edogenous Retroviruses. *Cell*, **162**, 974–986.
61. Roulois,D., Loo Yau,H., Singhania,R., Wang,Y., Danesh,A., Shen,S.Y., Han,H., Liang,G., Jones,P.A., Pugh,T.J. *et al.* (2015) DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell*, **162**, 961–973.
62. Leonova,K.I., Brodsky,L., Lipchick,B., Pal,M., Novototskaya,L., Chenchik,A.A., Sen,G.C., Komarova,E.A. and Gudkov,A.V. (2013) p53 cooperates with DNA methylation and a suicidal interferon response to maintain epigenetic silencing of repeats and noncoding RNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, E89–E98.
63. Kadota,S. and Nagata,K. (2014) Silencing of IFN-stimulated gene transcription is regulated by histone H1 and its chaperone TAF-I. *Nucleic Acids Res.*, **42**, 7642–7653.
64. Perez-Lluch,S., Blanco,E., Tilgner,H., Curado,J., Ruiz-Romero,M., Corominas,M. and Guigo,R. (2015) Absence of canonical marks of active chromatin in developmentally regulated genes. *Nat. Genet.*, **47**, 1158–1167.
65. Moerdyk-Schauwecker,M., Shah,N.R., Murphy,A.M., Hastie,E., Mukherjee,P. and Grdzelishvili,V.Z. (2013) Resistance of pancreatic cancer cells to oncolytic vesicular stomatitis virus: role of type I interferon signaling. *Virology*, **436**, 221–234.
66. Monsurro,V., Beghelli,S., Wang,R., Barbi,S., Coin,S., Di Pasquale,G., Bersani,S., Castellucci,M., Sorio,C., Eleuteri,S. *et al.* (2010) Anti-viral state segregates two molecular phenotypes of pancreatic adenocarcinoma: potential relevance for adenoviral gene therapy. *J. Transl. Med.*, **8**, 10.
67. Trojer,P., Zhang,J., Yonezawa,M., Schmidt,A., Zheng,H., Jenuwein,T. and Reinberg,D. (2009) Dynamic histone H1 isotype 4 methylation and demethylation by histone lysine methyltransferase G9a/KMT1C and the Jumonji domain-containing JMJD2/KDM4 proteins. *J. Biol. Chem.*, **284**, 8395–8405.
68. Trojer,P., Li,G., Sims,R.J. 3rd, Vaquero,A., Kalakonda,N., Boccuni,P., Lee,D., Erdjument-Bromage,H., Tempst,P., Nimer,S.D. *et al.* (2007) L3MBTL1, a histone-methylation-dependent chromatin lock. *Cell*, **129**, 915–928.
69. Kim,K., Choi,J., Heo,K., Kim,H., Levens,D., Kohno,K., Johnson,E.M., Brock,H.W. and An,W. (2008) Isolation and characterization of a novel H1.2 complex that acts as a repressor of p53-mediated transcription.*J. Biol. Chem.*, **283**,9113–9126.
70. Shen,Y.J., Le Bert,N., Chitre,A.A., Koo,C.X., Nga,X.H., Ho,S.S., Khatoo,M., Tan,N.Y., Ishii,K.J. and Gasser,S. (2015) Genome-derived cytosolic DNA mediates type I interferon-dependent rejection of B cell lymphoma cells. *Cell Rep.*, **11**, 460–473.
71. Rigby,R.E., Webb,L.M., Mackenzie,K.J., Li,Y., Leitch,A., Reijns,M.A., Lundie,R.J., Revuelta,A., Davidson,D.J., Diebold,S. *et al.* (2014) RNA:DNA hybrids are a novel molecular pattern sensed by TLR9. *EMBO J.*, **33**, 542–558.

Supplementary Material

Histone H1 depletion triggers an interferon response in cancer cells via activation of heterochromatic repeats

Andrea Izquierdo-Bouldstridge[1,#], Alberto Bustillos[1,#], Carles Bonet-Costa[1], Patricia Aribau[1], Daniel García-Gomis[1], Marc Dabad[2,3], Anna Esteve-Codina[2,3], Laura Pascual-Reguant[4], Sandra Peiró[4], Manel Esteller[5,6,7], Matthew Murtha[5], Lluís Millán-Ariño[1,^], Albert Jordan[1,*]

Supplementary Figure Legends

Figure 1. Inducible shRNAs for the depletion of histone H1 variants in human breast cancer cells. (**A**-**B**) T47D derivative cells stably infected with inducible lentiviruses for the expression of H1.4 or multiH1 shRNAs were treated or not for 6 days with doxycycline (Dox), histones were extracted and H1 depletion was tested by immunoblotting with the H1 variant-specific antibodies indicated. **(C)** Total histones were extracted from multi-H1 shRNAs cells treated with Dox up to 12 days, run on SDS-PAGE and stained with Coomassie or immunoblotted with antibodies indicated. **(D)** Coomassie-stained histone bands were scanned and quantified. The graph denotes the decrease of total H1 and variations of each H1-containing band upon Dox treatment of multiH1 KD cells.

Figure 2. H1 KD causes a reduction in nucleosome spacing. Nuclei from random or multiH1 KD cells treated or not with Dox for 6 days were treated with MNase and the profile of bulk chromatin was analyzed in gel electrophoresis (**A**). Two representative experiment are shown. (**B**) Ethidium bromide-stained bands were scanned, quantified and used to extrapolate DNA length of each band by comparison with the DNA ladder. Plot of nucleosome number versus DNA length was used to calculate the corresponding nucleosome repeat length (NRL).

Figure 3. H1 expression recovery and reversal of ISG expression after Dox removal. MultiH1 KD cells were treated with Dox for 3 days and split, washed and maintained without Dox or left with Dox up to 12 days. Gene expression was tested by RT-qPCR every 3 days, corrected by GAPDH and shown relative to untreated (time 0). Data corresponds to the experiment shown in Figure 3C.

Figure 4. ISGs are induced in a double H1.4/H1.2 KD cell line. (**A**) T47D cells containing H1.4sh were infected with the H1.2 shRNA-expressing lentivirus. Resulting cells were treated or not with Dox for 6 days and H1 gene expression was tested by RT-qPCR with oligonucleotides for the indicated genes, corrected by GAPDH and shown relative to untreated cells. (**B**) Total histones were extracted from H1.4/H1.2sh cells treated or not with Dox for 6 days, run on SDS-PAGE and stained with Coomassie or immunoblotted. **(C)** Coomassie-stained histone bands were scanned and quantified. The graph denotes the decrease of total H1 and variations of each H1-containing band upon Dox treatment of H1.4/H1.2 KD cells. (**D**) H1.4/H1.2sh cells were treated with Dox for the time indicated up to 72 hours and gene expression was tested by RT-qPCR. Gene expression is corrected by GAPDH and relative to the highest data point for each gene. (**E**) The indicated cell lines containing H1.2sh, H1.4sh or both together were treated or not with Dox for 6 days and expression of several ISGs was tested and shown corrected by GAPDH and relative to untreated cells.

Figure 5. H1 variants expression in double and triple H1 KD cells. (A) T47D cells containing H1.2sh or H1.4sh were infected with the H1.5 shRNA-expressing lentivirus. Resulting cells were treated or not with Dox for 6 days and H1 gene expression was tested by RT-qPCR with oligonucleotides for the indicated variants, corrected by GAPDH and shown relative to untreated cells. **(B)** Total histones were extracted from H1.2/H1.5 KD and H1.4/H1.5 KD cells treated or not with Dox for 6 days, run on SDS-PAGE and stained with Coomassie or immunoblotted. Coomassie-stained histone bands were scanned and quantified. The graph denotes the decrease of total H1 and variations of each H1-containing band upon Dox treatment. (**C**) H1.2/H1.5 KD and H1.4/H1.5 KD cells were infected with the H1.3 shRNA-expressing lentivirus. Resulting cells were treated or not with Dox for 6 days and gene expression was tested by RT-qPCR, corrected by GAPDH and shown relative to untreated cells.

Figure 6. Expression of HA-tagged H1 variants in multiH1 KD cells. T47D multiH1sh cells infected with lentiviruses for the expression of HA-tagged H1.2, H1.3 or H1.4 variants were treated or not with Dox for 6 days and total histones were extracted for immunoblotting with an anti-HA antibody. A gel stained with Coomassie is sown as a loading control.

Figure 7. H1 abundance at genes responsive to multiH1 KD corresponds to its basal expression level. (A) Box plots showing the expression profiles of genes up- and down-regulated in multiH1 KD cells, treated or not with Dox, analyzed by RNAseq. For comparison, the basal expression profiles of all genes and expressed genes is included. In the left panel, the expression profiles of non-expressed genes (ca. 57% of the total) and expressed genes divided into 10 equal groups are also shown. Outlier genes are not included. Groups 6 and 8, those that resemble the most to up- and down-regulated gene means, are labeled. Significances shown were tested using the Kolmogorov-Smirnov test (*, p-value<0.05). **(B)** H1 is enriched at multiH1sh up-regulated compared to down-regulated genes. Box plots of H1 variants abundance (input-subtracted ChIP-seq signal) at proximal promoter regions (-2,000 to -1 bp relative to TSS) for the genes up- or down-regulated upon multiH1 KD, compared to the mean H1 abundance of total RefSeq genes (All), non-expressed (Not) or expressed genes (10 equal groups of increasing expression). Significance was assessed using the Kolmogorov-Smirnov test. Significant difference between up- or down-regulated genes is marked with asterisks (p<0.0001). Expression and H1 content levels of up- and down-regulated genes is similar to groups 6 and 8 of expressed genes, respectively (labeled in orange). (**C**) Box plots of H1 variants abundance at coding regions for the genes up- or down-regulated upon multiH1 KD, compared to the mean H1 abundance of total genes. Significance was assessed using the Kolmogorov-Smirnov test. Significant difference between up- or down-regulated genes is marked with asterisks (p<0.0001). Three replicates of H1.2 ChIP-seq data are shown. **(D)** Average density profile of H1 variants abundance (input-subtracted ChIP-seq signal) around the TSS of up- or down-regulated genes upon multiH1 KD. RefSeq genes (All), as well as bottom and top 10% expressed genes (groups 1 and 10 in panel **A**), are shown as reference. **(E)** H1 content at up-regulated genes corresponds to genes presenting low expression levels. The H1.2 content at multiH1sh up-regulated genes (green barr) is not significantly different to the H1.2 content distribution in groups of genes presenting similar expression levels randomly chosen.

Figure 8. IFN signaling-related gene expression profiles in pancreatic normal and cancer samples. Supervised cluster analysis of IFN related genes (see (71)) in 36

pancreatic tumor and 16 normal samples using microarray data deposited in GSE16515. Three clusters are formed and have been named as Group 1 to 3. The number of normal and tumor samples in each group is indicated.

Figure 9. H1 knock-down has little impact on genomic DNA methylation. (**A**) Lack of ISG induction by DNA methylation inhibitors. MultiH1sh cells were treated or not with Dox or aza-deoxycitidine 5μM for 3 days and ISG expression was analyzed, corrected by GAPDH and expressed relative to the maximal value for each gene. (**B**) Global changes in DNA CpG methylation were investigated using the Illumina 850K array from multiH1sh cells treated or not with Dox for 6 days. Of the ~850,000 CpGs represented on the array, only 11 residues were observed to have large difference in methylation (change in beta value >0.3). (**C**) The genes associated with these residues are listed. None of these genes was altered in its expression upon multiH1 KD. (**D**) Bisulphite sequencing analysis of three LTR regions of HERV-K(HML-2) family members located on chromosomes 7, 11, and 12. Methyl and un-methylated CpGs present in the LTRs are represented as white and black circles, respectively.

Supplementary Methods

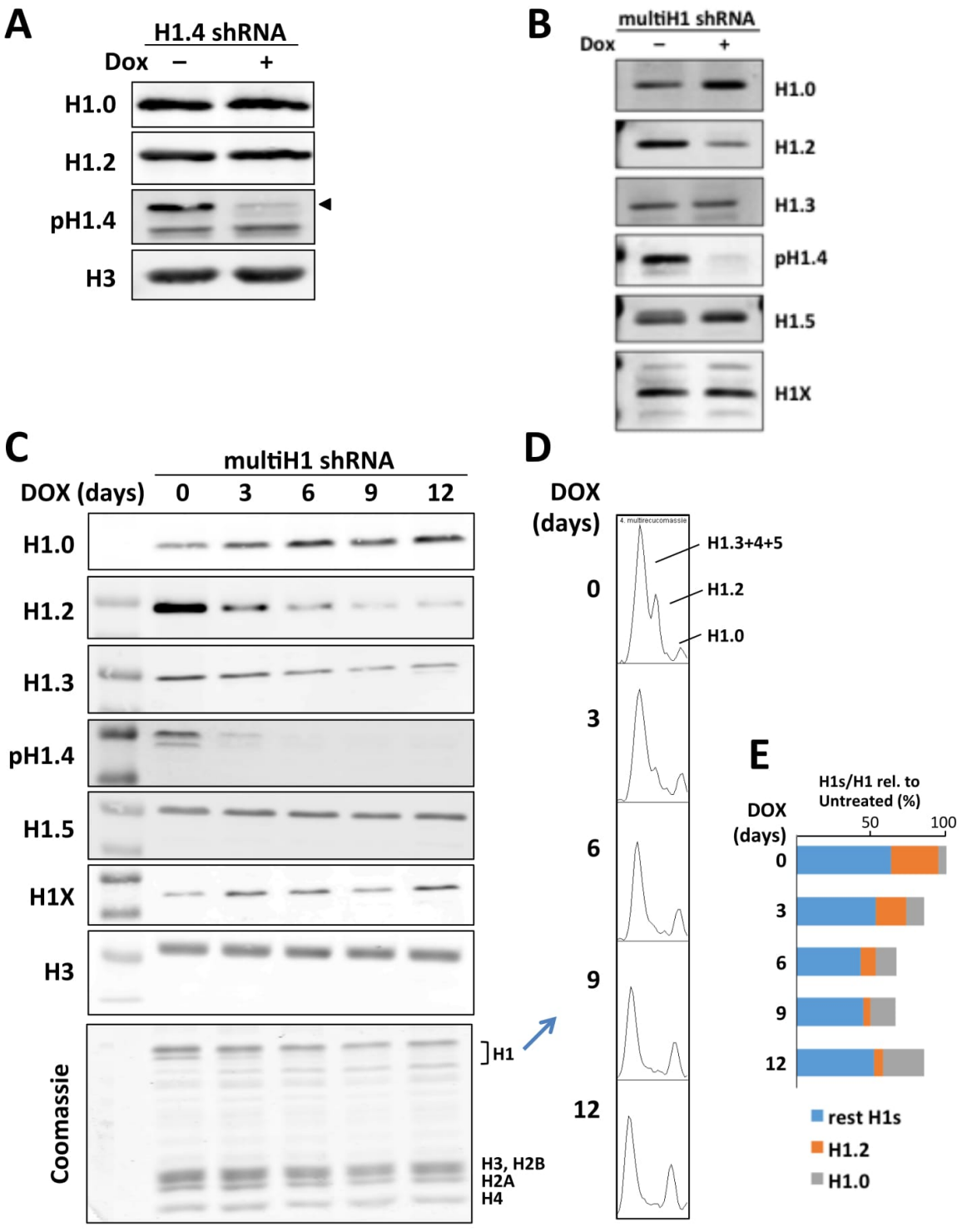
Global DNA Methylation Analysis. DNA bisulfite conversion reactions were carried out with 500ng of genomic DNA using EZ-96 DNA Methylation kit (Zymo Research Corp, Irvine, CA, USA) following the manufacturer's recommendations. Bisulfite-converted DNA (bs-DNA) was and then hybridized on Infinium HumanMethylation850 BeadChip array following the manufacturer's instructions for automated processing of the arrays with a liquid handler (Illumina Infinium HD Methylation Assay Experienced User Card, Automated Protocol - 15019521 v01). We normalized raw intensity data using manufacturer's method (GenomeStudio) implemented on minfi package (preprocessIllumina; minfi v1.12). Unsupervised heatmap representation of the 11 differentially methylated residues was produced in R Studio.

Bisulphite sequencing. Bisulphite treated DNA was also used for detection of methylation changes by sequencing. We amplified, in total, three LTR loci from Human Endogenous Retrovirus Family HERV-K(HML-2) from bisulfite-treated DNA in wildtype and knockdown cells by nested PCR (1). Approximately 150ng of treated gDNA was used as template in the first round of PCR with the following conditions: 2 min at 94°C; 30 cycles, consisting of 15 s at 94°C, 30 s at 55°C, and 40 s at 72°C; and finally 7 min at 72°C. PCR products were purified by the NucleoSpin PCR Clean-Up kit (Macherey-Nagel). The second round of PCR used the eluate as template. PCR products were subsequently cloned into the pGEM-T vector (Promega), and single clones were sequenced with the Applied Biosystems® 3730 DNA Analyzer.

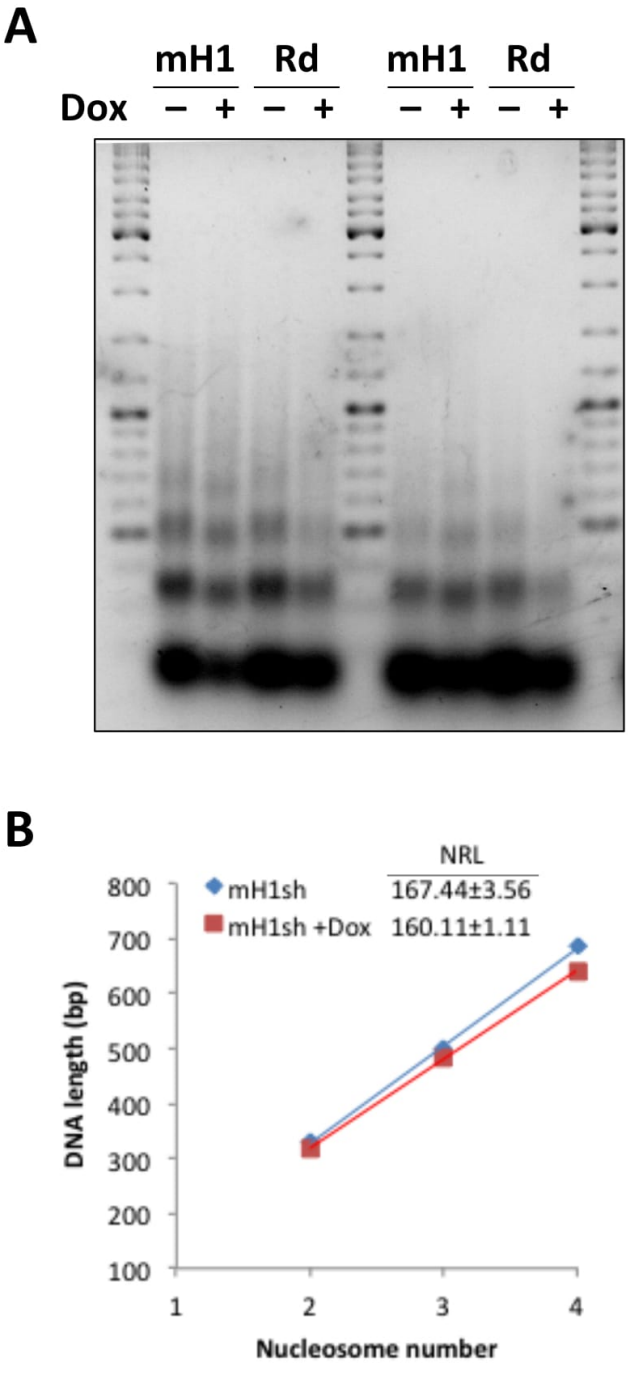
References

1. Lavie, L., Kitova, M., Maldener, E., Meese, E., and Mayer, J. (2005) CpG Methylation Directly Regulates Transcriptional Activity of the Human Endogenous Retrovirus Family HERV-K(HML-2). J. Virol. 79: 2876-883.

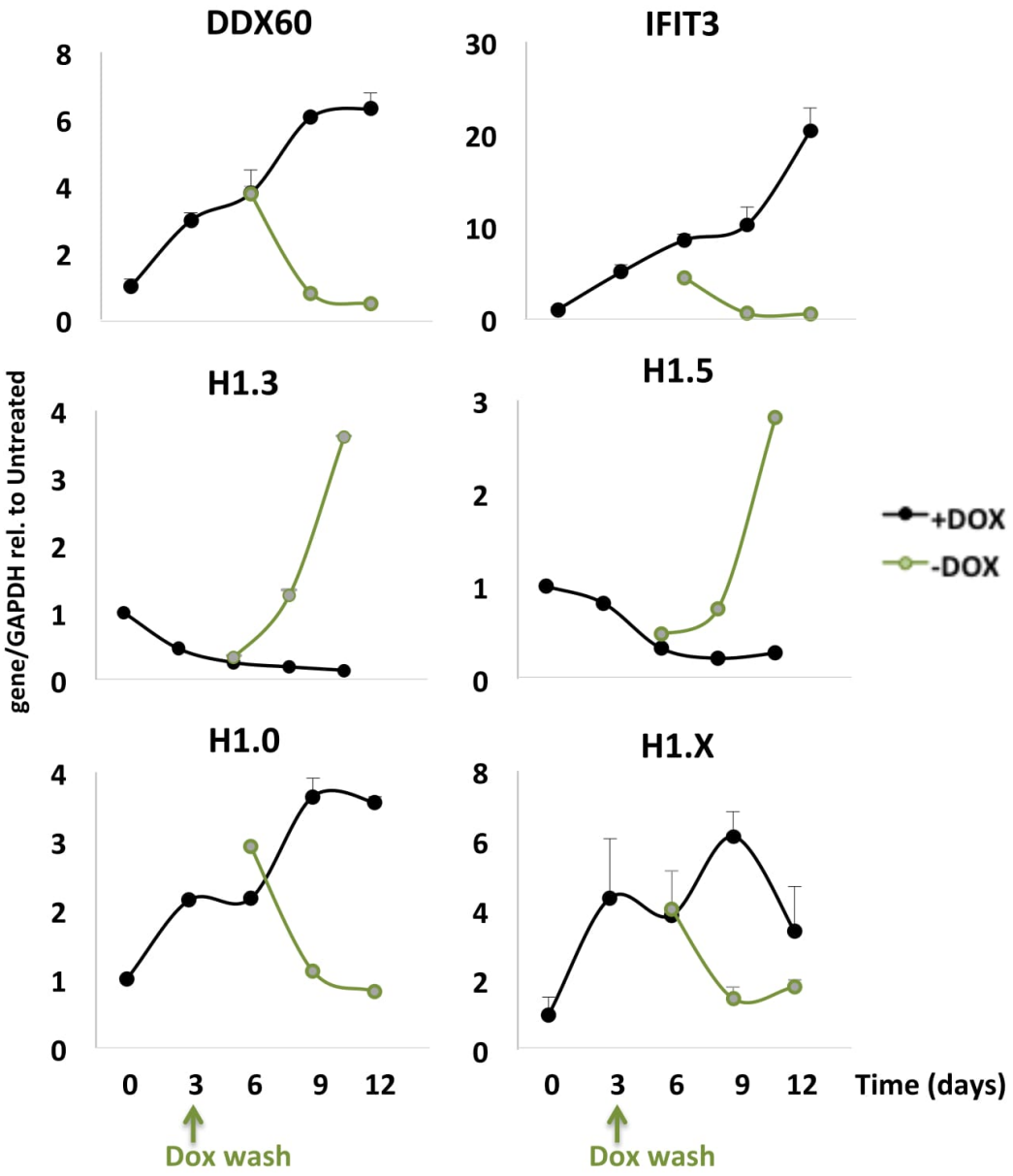
Izquierdo et al. Supplementary Figure 1



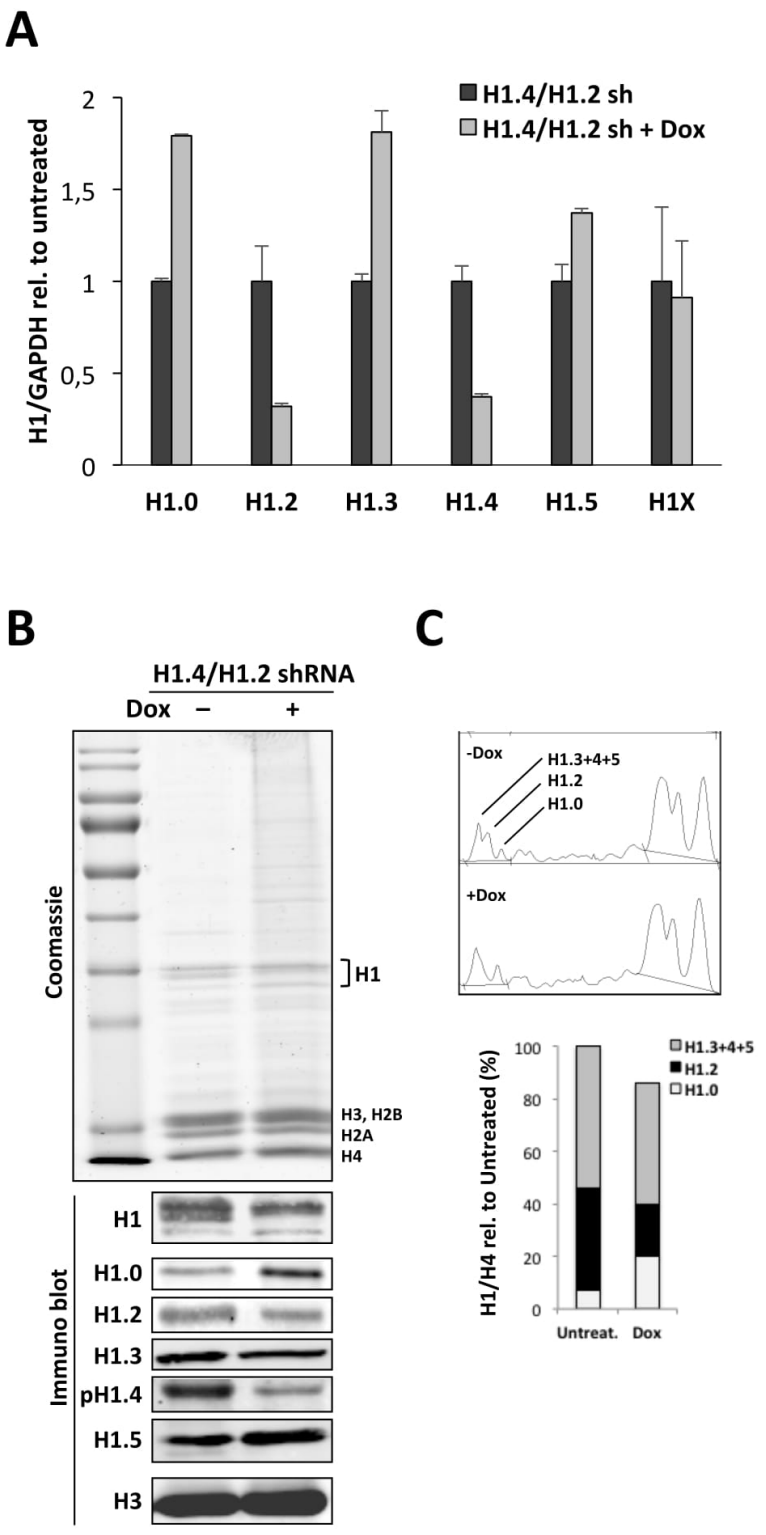
Izquierdo et al. Supplementary Figure 2



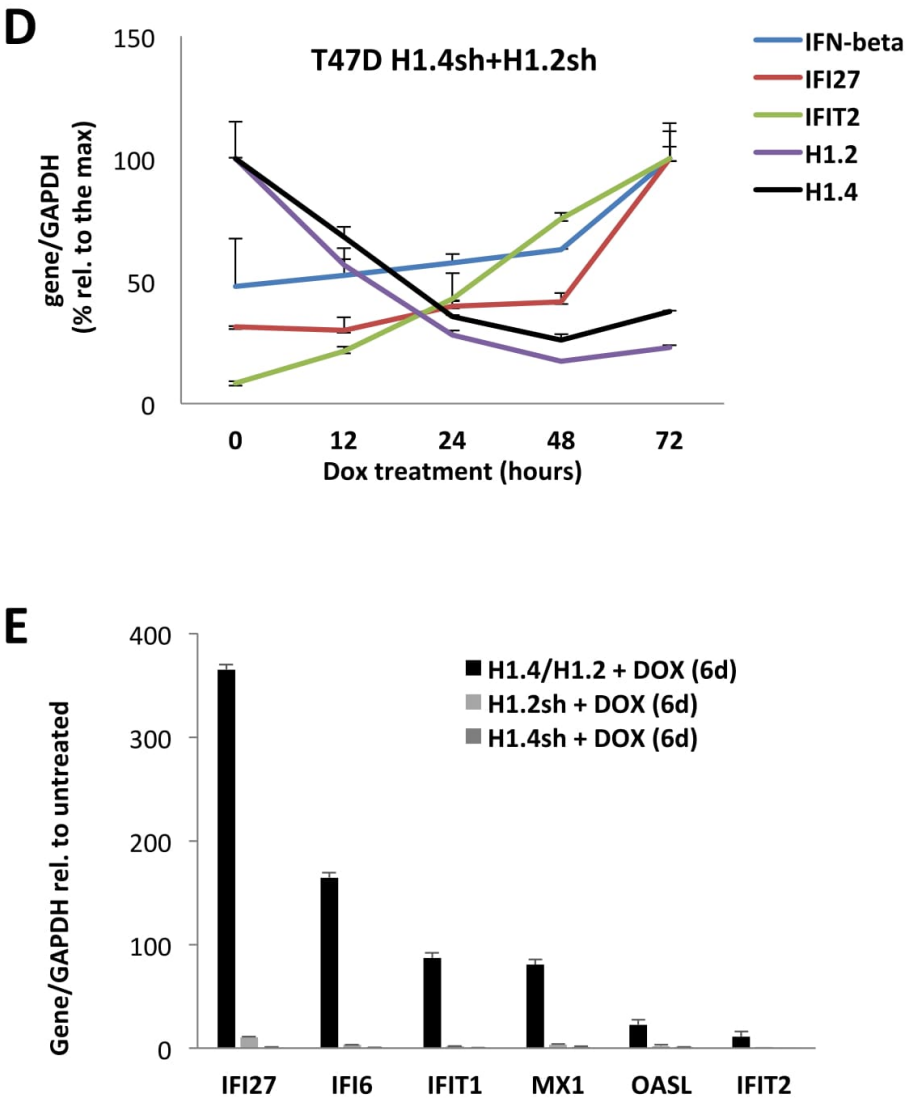
Izquierdo et al. Supplementary Figure 3



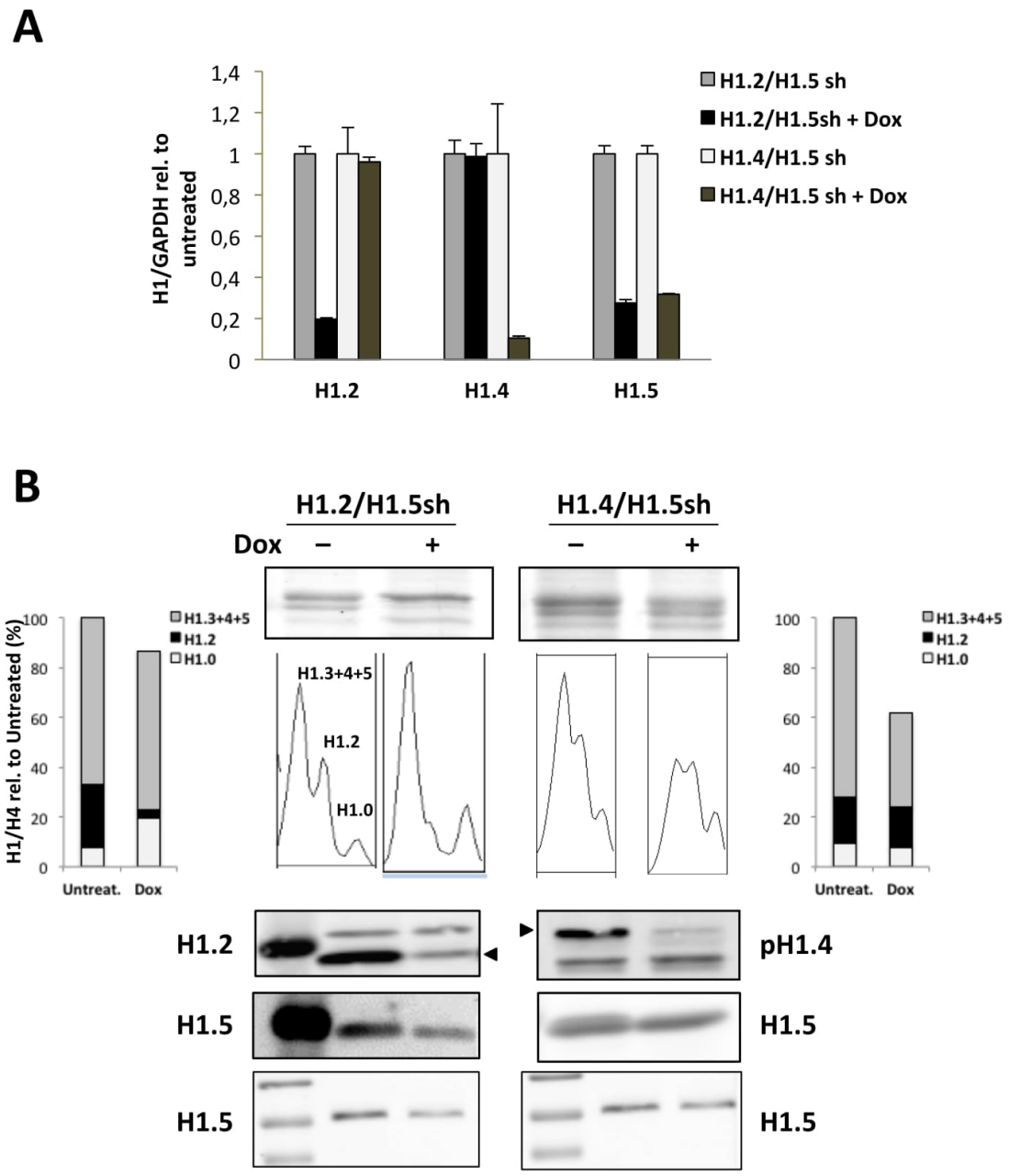
Izquierdo et al. Supplementary Figure 4



Izquierdo et al. Supplementary Figure 4

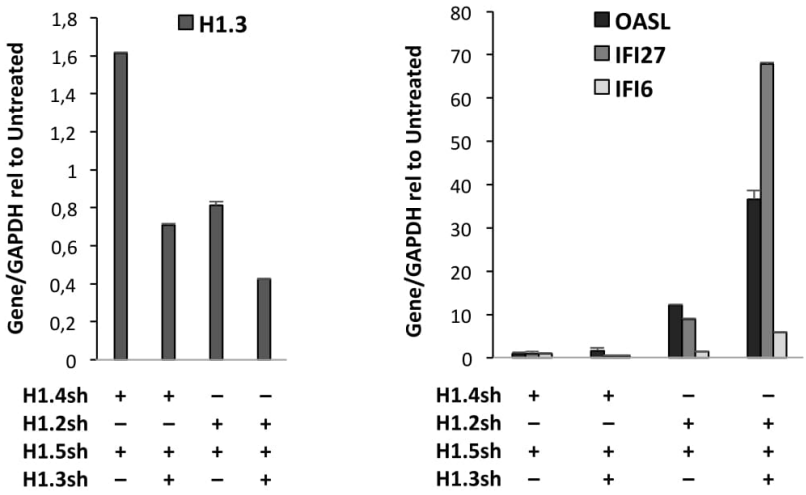


Izquierdo et al. Supplementary Figure 5

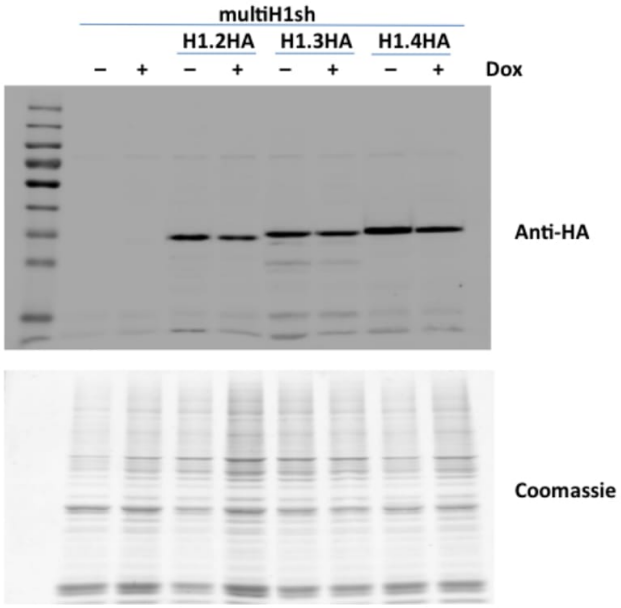


Izquierdo et al. Supplementary Figure 5

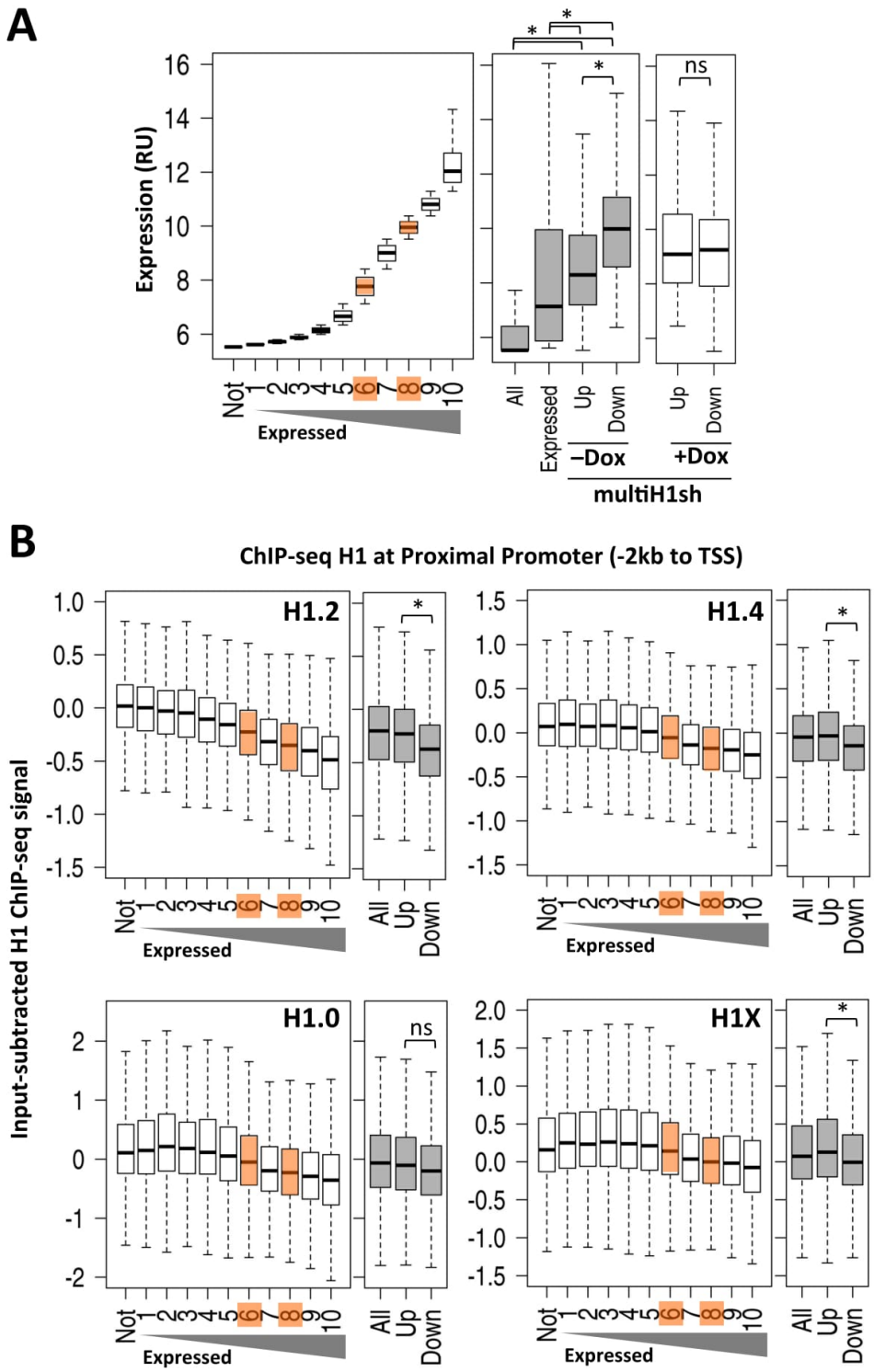
C



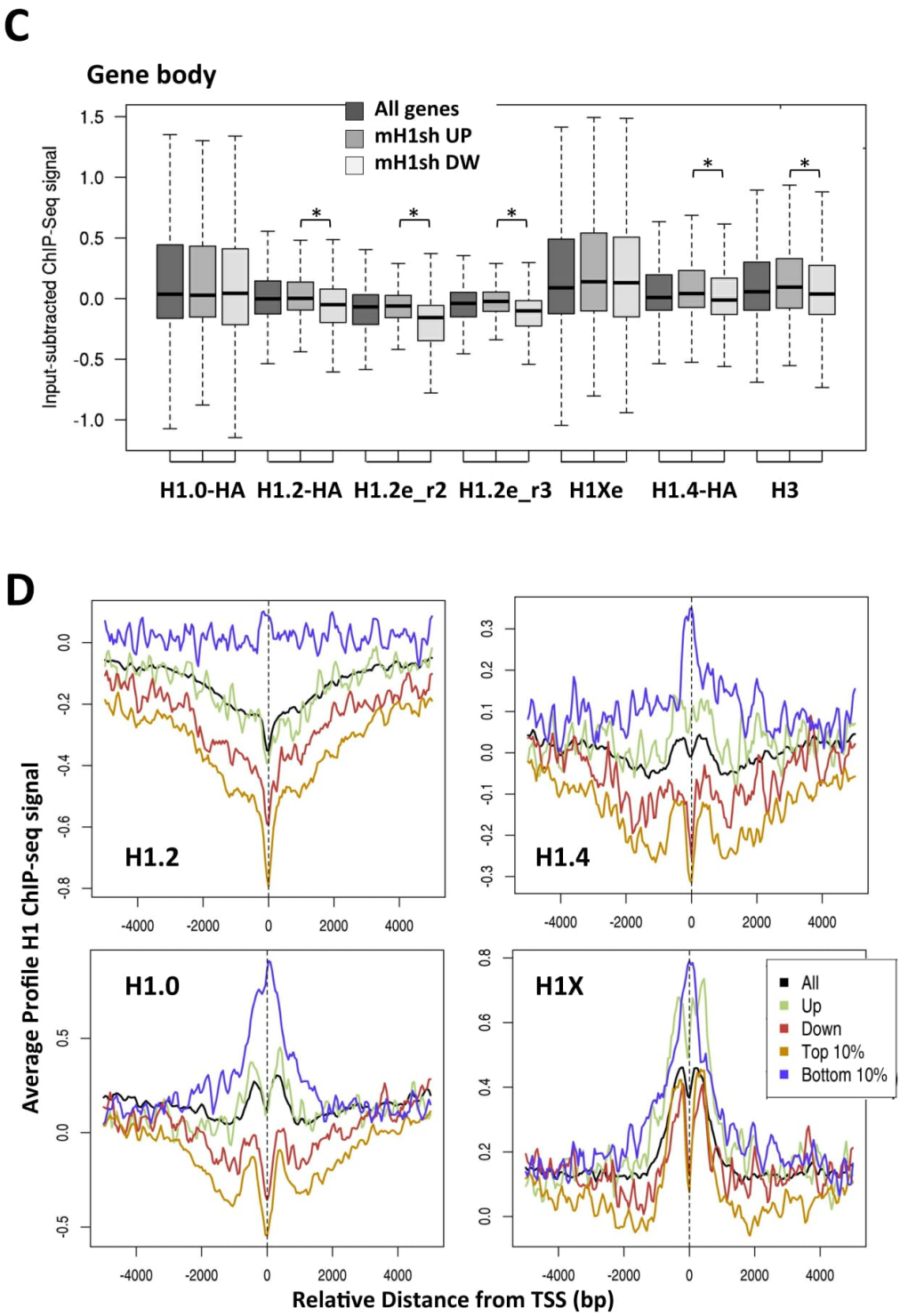
Izquierdo et al. Supplementary Figure 6



Izquierdo et al. Supplementary Figure 7

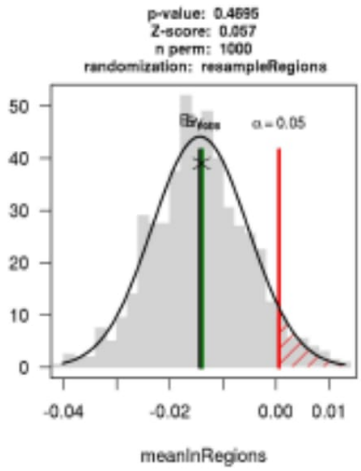


Izquierdo et al. Supplementary Figure 7

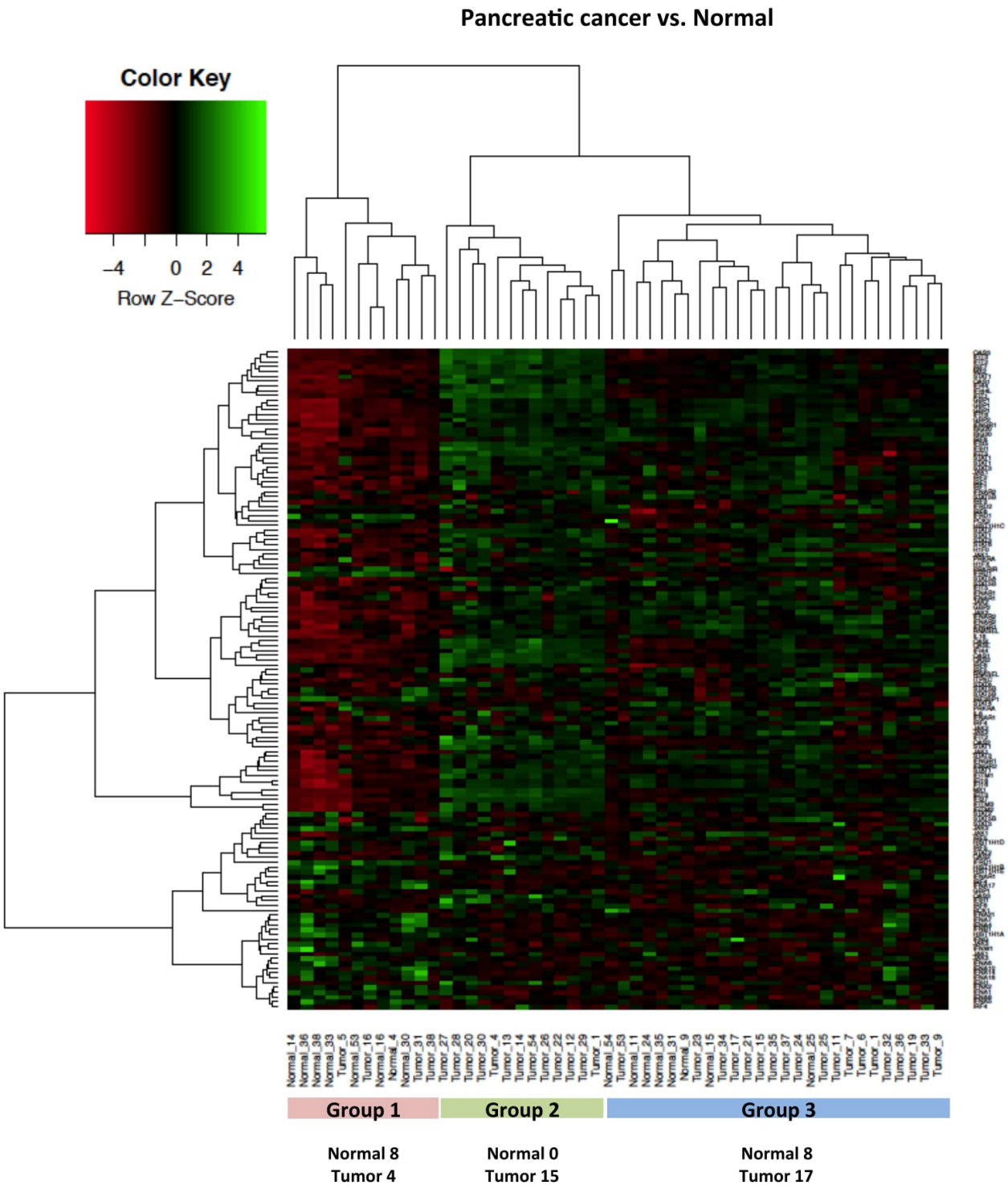


Izquierdo et al. Supplementary Figure 7

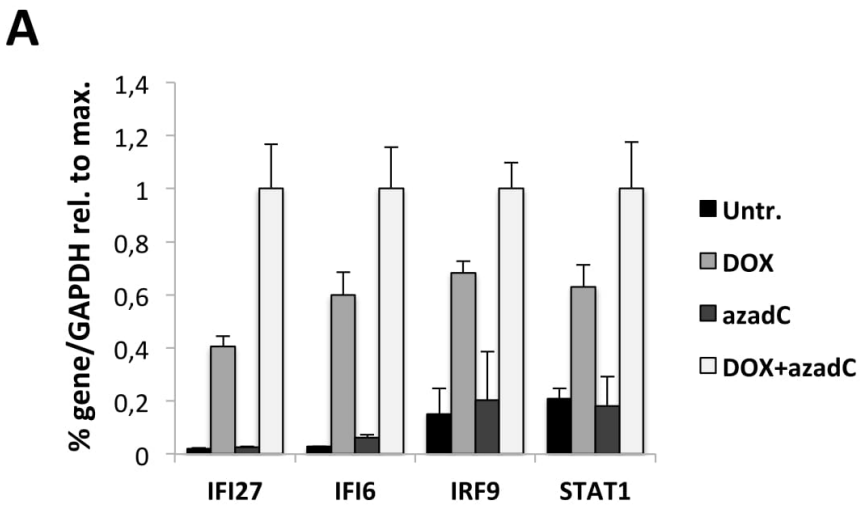
E

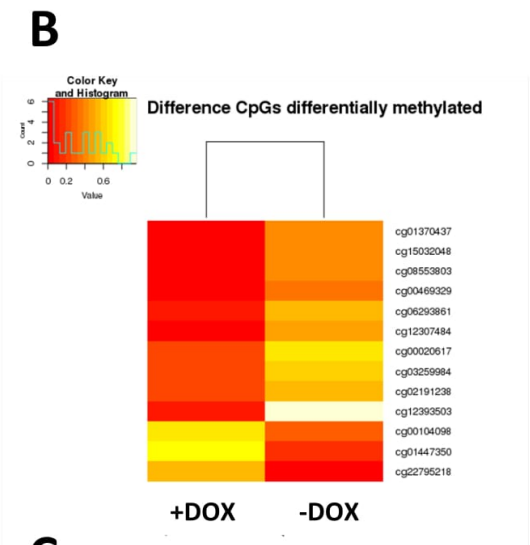


Izquierdo et al. Supplementary Figure 8



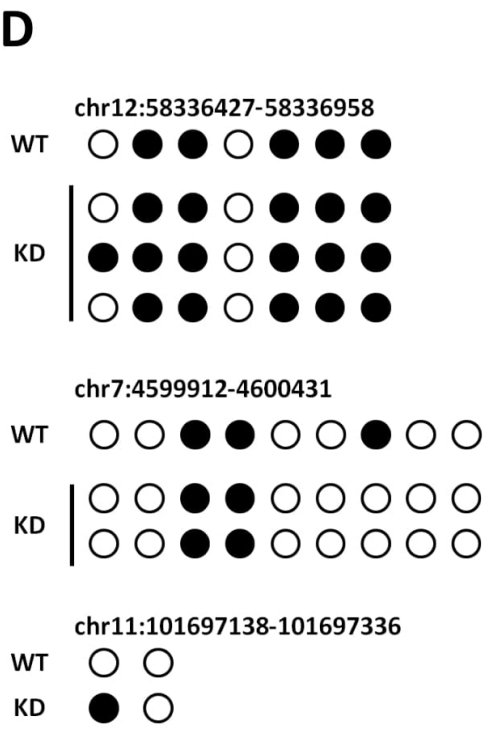
Izquierdo et al. Supplementary Figure 9





C

COL24A1	RFX3-AS1
IL34	SPATA24
KCNQ4	TCF7L2
MORN1	TMEFF2
PTPRN2	TWIST1



Supplementary Table 1. List of primers used for RT-qPCR.

Gene	Primers	Sequence (from 5' to 3')
H1.0	forward reverse	CCTGCGGCCAAGCCCAAGCG AACTTGATCTGCGAGTCAGC
H1.1	forward reverse	CTCCTCTAAGGAGCGTGGTG GAGGACGCCTTCTTGTTGAG
H1.2	forward reverse	GGCTGGGGGTACGCCT TTAGGTTTGGTTCCGCCC
H1.3	forward reverse	CTGCTCCACTTGCTCCTACC GCAAGCGCTTTCTTAAGC
H1.4	forward reverse	GTCGGGTTCCTTCAAACTCA CTTCTTCGCCTTCTTTGGG
H1.5	forward reverse	CATTAAGCTGGGCCTCAAGA TCACTGCCTTTTTCGCCCC
H1X	forward reverse	TTCCTTCAAGCTCAACCG TGCCTTCTTCGCTTTGTG
GAPDH	forward reverse	GAGTCAACGGATTTTGGTCGT TTGATTTTGGAGGGATCTCG
MAVs	forward reverse	GGAGCAGCAGAAATGAGGAG AAGGCCCCTATTCTCAGAGC
MDA5	forward reverse	TCTGGGGCATGGAGAATAA AAGGCCTGAGCTGGAGTTCT
STING	forward reverse	CAGGCACTGAACATCCTCCT GTTATCAGGCACCCCACAGT
IFNAR1	forward reverse	GCTCAGATTGGTCCTCCAGA CCATCCAAAGCCCACATAAC
IFNβ	forward reverse	AGCACTGGCTGGAATGAGAC TCCTTGGCCTTCAGGTAATG
IFNα	forward reverse	CAAAGACTCTCACCCCTGCT CACAGTGTAAAGGTGCACATGA
IL-29	forward reverse	AATTGGGACCTGAGGCTTCT AGCTGGGAGAGGATGTGGT
DDX60	forward reverse	AAGGTGTTCCTTGATGATCTCC TGACAATGGGAGTTGATATTCC
SET	forward reverse	TGGTTCCCGATATGGATGAT ATCCTCATCCCCTTCTTCGT
IFI27	forward reverse	TGCTCTCACCTCATCAGCAGT CACAACTCCTCCAATCACAACT
IFIT1	forward reverse	GCCTCCTTGGGTTCGTCTATAA TCAAAGTCAGCAGCCAGTCTCA
IFIT2	forward reverse	ACGGTATGCTTGGAACGATTG AACCCAGAGTGTGGCTGATG
IFIT3	forward reverse	CGGAACAGCAGAGACACAGA ATGGCATTTCAGCTGTGGA
IFI6	forward reverse	CTGTGCCCATCTATCAGCAG GGGCTCCGTCACTAGACCTT
STAT1	forward reverse	GTTTTCATGACCTCCTGTCAC GTGGACTCCTCCATGTTCATC
IRF7	forward reverse	GAGTGGCCATAGGTGGCTC ACCACTTCGTGAACAGACAGA
IRF9	forward reverse	CCGTGATAATCGTGTCCTGAAA CCTGGGTTCACACCATTTGG

TLR3	forward reverse	TCACTTGCTCATTCTCCCTT GACCTCTCCATTCCTGGC
IFIH1	forward reverse	ACCAAATACAGGAGCCATGC GCGATTTCCTTCTTTTGCAG
OASL	forward reverse	GGGACAGAGATGGCACTGAT AAATGCTCCTGCCTCAGAAA
RSAD2	forward reverse	GTGAGCAATGGAAGCCTGAT TCCCTACACCACCTCCTCAG
MX1	forward reverse	ACCTACAGCTGGCTCCTGAA CGGCTAACGGATAAGCAGAG
OAS2	forward reverse	GAGTGGCCATAGGTGGCTC ACCACTTCGTGAACAGACAGA
ISG15	forward reverse	TGTCGGTGTCAGAGCTGAAG GCCCTTGTTATTCCTCACCA
IRF1	forward reverse	TTTGTATCGGCCTGTGTGAATG AAGCATGGCTGGGACATCA
SATa	forward reverse	AAGGTCAATGGCAGAAAAGAA CAACGAAGGCCACAAGATGTC
D4Z4	forward reverse	CCGCGTCCGTCCGTGAAA TCCGTCGCCGTCCTCGTC
MER4D	forward reverse	CCCTAAAGAGGCAGGACACC TCAAGCAATCGTCAACCAGA
MER21C	forward reverse	GGAGCTTCCTGATTGGCAGA ATGTAGGGTGGCAAGCACTG
MLT1C49	forward reverse	TATTGCCGTACTGTGGGCTG TGGAACAGAGCCCTTCCTTG
HERVK	forward reverse	GAGAGCCTCCCACAGTTGAG TTTGCCAGAATCTCCCAATC
LTR57	forward reverse	TCCCTGAGAACCCAAACATC GTGCCACCTCCCATCTCTAA
env-Fc1	forward reverse	CTCCATTAGTAGCAGTTCCTCTCC GAGAATAGTGGGACCTGTCCTTT
MLT1J2	forward reverse	CCTGGGTCCCTGAGTCACTA TGCCAGCTGCTGTAACAAAC
MLT1H2	forward reverse	AGCAAATGTGATGCAAGCAG ATGTCCAAGATGGCCTCTCA
SST1	forward reverse	AACCACTGTGACGGGAGAAA CTGGGACAGGACGAGACAC
MER52A	forward reverse	CCAGTTCCGTGGACAAGAGT TCTGGCTGAGTCTGGGACTT

Izquierdo et al. Supplementary Table 2

Supplementary Table 2. Summary of gene expression changes upon H1 variant KD (total, up- and down-regulated genes) and overlap between variants obtained in microarray (FC≥1.4, q-value≤0.05) or RNAseq (FC≥1.4, p adjusted≤0.05) analysis. Only genes present in both data sets are being considered. Genes that appeared deregulated in control (RDsh) experiments have been subtracted.

	microarrays				RNAseq		
	H1.0	**H1.2**	**H1.3**	**H1.5**	**H1.4**	**mH1**	
	253	**76**	**116**	**76**	**7**	**78**	Total
H1.0	82	10	12	25	2	8	Up
	171	65	99	51	5	64	Down
		395	**96**	**46**	**16**	**95**	
H1.2		104	17	7	8	9	
		291	70	35	4	69	
			441	**60**	**6**	**123**	
H1.3			139	2	0	3	
			302	51	5	90	
				193	**5**	**56**	
H1.5				93	3	10	
				100	2	44	
					146	**64**	
H1.4					84	36	
					62	20	
						1595	
mH1						732	
						863	

DISCUSSION

DISCUSSION

Using variant-specific antibodies to H1 and hemagglutinin (HA)-tagged recombinant H1 variants in T47D breast cancer cells, the distribution of six somatic H1 variants was previously investigated by ChIP-chip and ChIP-Seq methods (*Appendix I*). Specifically, ChIP-Seq data from two replication-dependent (H1.2 and H1.4) and replication-independent H1 variants (H1.0 and H1X) together with core histone H3 was produced. Shortly, H1.2 was found less abundant than other H1 variants at transcription start sites (TSS) of inactive genes and promoters enriched in H1.2 tend to be more repressed. Additionally, H1.2 was found enriched at GC-poor, gene-poor and intergenic chromosomal domains in addition to lamin-associated domains (LADs), compared to the other three studied H1 variants.

We further investigated their distribution in regions not included in the reference genome such as repetitive sequences and ribosomal DNA (rDNA) or chromatin domains such as nucleolus-associated domains (NADs) in addition to deeply analyse genic regions, introns and exons. Strikingly, we found that distribution of replication-independent H1 variants (H1.0 and H1X) is distinct.

H1.0 is enriched at nucleolar chromatin

We found that H1.0 is enriched in nucleolus-related features compared to the other studied H1 variants (H1.2, H1.4 and H1X). The nucleolus is the largest structure within the nucleus where ribosome biogenesis is performed. Nucleolus is formed around specific genetic loci called nucleolus organizer regions (NORs), which are found at short arms of acrocentric chromosomes (in human, chromosomes 13, 14, 15, 21 and 22) organized in tandem repeats. NORs contain rDNA, the 45S single transcription unit, lately processed to 18S, 5.8S and 28S rRNA, flanked by non-transcribed spacers, where regulatory elements are located. Some copies of rDNA are highly transcribed and frequently found inside nucleoli. Instead, repressed ones are at the periphery, where a shell of highly compacted heterochromatic DNA is found surrounding nucleoli. In addition to contain NORs, nucleolar heterochromatin also contains other specific sequences which have been recently identified and named nucleolus-associated domains (NADs) [35, 36]. NADs are gene-poor regions, AT-rich and contain a high proportion of silent genes, features that are also found at LADs. Indeed, a high overlap of NADs and LADs has been described and at single-cell level, DNA can be stochastically found in one domain after mitosis [35-37, 33]. Although sharing common characteristics, we found that their content in H1 variants is different, being H1.2

enriched at LADs and H1.0 at NADs, suggesting a possible H1 variant-specific role in maintaining, stablishing or organizing those heterochromatic domains (*Chapter I*, Figures 1 and 3).

In addition to NADs, we found an H1.0 enrichment at rDNA, specifically at non-transcribed spacers, in addition to the 5S rDNA, which is found encoded in tandem arrays, the largest one on chromosome 1. Other specific heterochromatic repetitive elements such as SINE-VNTR-SVA (SVA) retrotransposons, telomeric and acromeric satellites are also enriched with H1.0 (*Chapter I*, Figure 2). Thus, H1.0 seems to be related to the nucleolus structure and/or function. Consistently, using optical and electron microscopy, it was observed that H1.0 is present at highly compacted chromatin and perinucleolar regions in certain differentiated tissues [255]. In agreement with the presence of H1.0 at nucleolar chromatin, Kalashnikova et al. showed by LC-MS/MS that H1.0 is interacting with an extensive network of proteins found in the nucleolus such as core splicing factors, proteins related to rRNA biogenesis, ribosomal proteins and proteins involved in cellular transport [204].

Recently, an interesting study showed that within a tumour, cancer stem cells (CSCs) have a reduced content of H1.0, which correlates with patient survival [179]. Its low abundance at CSCs, its nucleolar distribution and its enrichment in specific repetitive sequences may show that H1.0 is critical for the proper control of heterochromatic regions near nucleoli that when uncontrolled may lead to the activation of an oncogenic program. Nevertheless, what is the specific function of H1.0 in the nucleolus is still not known and other H1 variants or PTMs have been related to nucleoli as well. Proteomic analysis of the T-cell nucleolus showed that other linker histone H1 variants are also present, specifically H1.1, H1.2, H1.5 and H1X [256]. Moreover, interphase phosphorylated H1.4 is enriched at active 45S rDNA gene promoter and is rapidly induced by steroid hormone treatment [257]. H1.2 and H1.4 phosphorylation is associated to RNA polymerase I activity and rRNA biogenesis. Besides, H1 testis-specific variant H1T has been shown to be expressed in mESCs, cancer and some somatic normal cells and it was found to accumulate at nucleoli by immunofluorescence and predominantly targeted rDNA repeats by ChIP-Seq [269].

It has been previously reported that H1X is accumulated at nucleoli at the cell cycle G1-phase. It was stated that the differential localization, cell cycle-dependent, of H1X may provide a mechanism for controlling H1X activity by changing its nuclear subcompartment localization rather than a controlled turnover of protein levels [258, 259]. Although our cells are not cell cycle synchronized, by immunofluorescence and cellular fractionation we also see an enrichment of H1X at nucleoli and nucleolar fraction, respectively. This H1X enrichment is smaller than H1.0 by cellular

fractionation and at rDNA their distribution is distinct, being H1X mainly found at the transcribed 45S rDNA and H1.0 at the non-transcribed spacer.

How replication-independent H1 variants are recruited to nucleoli and which is their role in this compartment it is still not known. In addition, H1.0 and H1X are not exclusively found at nucleolus and they are also frequently localized with other somatic replication-dependent H1 variants genome-wide. Further, as the abundance of linker histone H1 variants greatly varies between cell types and through development and H1.0 is restricted to differentiated cells it is tempting to speculate that replication-independent variants may have distinct functions depending on cellular context. Thus, study their nucleolar distribution and function in differentiated cells would be of great interest.

H1X is associated with actively transcribed chromatin

Linker histone H1 has classically been regarded as a mere structural component, stabilising and/or maintaining chromatin, related to gene-repression and intergenic regions where a more compacted DNA is found. However, as explained, new evidences are showing that histone H1 variants have a more dynamic and gene-specific role (See *Introduction, 2.2. Linker histone H1 function*).

Intriguingly, we found a specific enrichment of H1X variant at actively transcribed chromatin in T47D breast cancer cells. Specifically, we observed that H1X is enriched at RNAPII peaks, accumulates at coding regions, mainly at exons of expressed genes (*Chapter I*, Figures 4 and 5). Nevertheless, this enrichment is not seen at promoter regions where the typical “H1 valley” at TSS is observed for all H1 variants. In this regard, H1X colocalization with RNAPII might be related to the elongation process, suggested by it enrichment towards the 3’ end. Or, secondly, to splicing as it occurs cotranscriptionally and we do see that H1X is enriched at exons, especially at included alternatively spliced exons (ASEs) and retained introns.

H1X was firstly discovered as an interacting partner of the WD40 repeat of TFIID, a critical transcription factor for the establishment of the pre-initiation complex of RNAPII [259]. Unluckily, their association was not explored functionally and might be related to our findings. Indeed, other linker histone H1 variants have been associated to the transcription elongation process [208]. In HeLa cell line, H1.2 was found to stably interact with Cul4A E3 ubiquitin ligase, PAF1 elongation complexes and serine 2 phosphorylated RNAPII, potentiating transcription elongation via induction of H4K31ubiquitylation, H3K4me3 and H3K79me3. Interestingly, this H1.2 association with elongation cannot be specifically attributed to H1.2 as WDR5, another substrate adaptor of the Cul4 E3 ligase, was co-purified with six linker histone H1 variants and

H1X was not included in the analyses. Thus, we cannot exclude a similar H1X function in T47D cell line similar to the aforementioned H1.2 association in HeLa cell line, regulating transcriptional elongation.

In addition to elongation, H1X may be related to the splicing process. H1X pattern along the coding regions resembles H3K36me3 distribution, which has been involved in defining exons and in regulating alternatively splicing events. Exons are enriched in nucleosomes in general and they contain certain core histone modifications, including H3K79, H4K20 and especially H3K36me3 [260]. Apart from defining exons, H3K36me3 have been shown to directly control alternative splicing events, as an overexpression of its methyltransferase (SET2) leads to different splicing outcomes [46]. Similar to H3K36me3, we observed an H1X enrichment in exons compared to introns and even higher at included ASE and retained introns (*Chapter I*, Figure 5). Further experiments are needed to determine how H1X influences splicing events, directly interacting with splicing machinery or H3K36me3. In addition, H1X might modulate RNAPII kinetics along the coding regions, which has been proved to influence splicing outcomes (See *Introduction, 1.5. Chromatin and splicing*).

In IMR90 human lung fibroblasts, the genomic distribution of replication-dependent somatic H1 variants (H1.1 to H1.5) showed no differences at exons compared to introns, although a more variable binding at exons is observed [217]. Unfortunately, replication-independent variants (H1.0 and H1X) were not included in this study and they might have a distinct distribution, especially H1X at coding regions similar to our results in T47D breast cancer cells.

Regarding CpG islands, we have shown that H1 variants distribution is heterogeneous, being replication-independent variants (H1.0 and H1X) clearly overrepresented in contrast to other studied H1 variants, which are underrepresented (*Chapter I*, Figures 6A and 6B). It could be hypothesised that as replication-independent variants are replacing other H1 variants in a cell-specific manner, this enrichment at CpG might be showing their specific regulatory roles. Further, CpG islands overlapping H1-depleted islands are hypomethylated, in agreement with a general H1 function in repressive chromatin. However, when assessing those CpG islands overlapping H1-enriched islands, H1X peaks coincide with hypomethylated CpG islands, again relating H1X to active transcription at these regulatory elements. In agreement, as we have previously reported H1.2 is the variant that better correlates with gene repression (*Appendix I*), H1.2-enriched islands overlapping CpG coincide with a more hypermethylated status and thus, repressive chromatin states.

It is known that in cancer, a global genome-wide DNA hypomethylation is observed at intergenic regions and particular promoter regions are locally hypermethylated, some of them leading to the repression of tumour suppressor genes. Ruike et al. analysed those hyper- and hypomethylated regions in T47D cancer cells compared to normal mammary epithelial cells [261] and we have analysed the distribution of H1 variants at those defined regions (*Chapter I*, Figure 6C). H1.2 was found enriched at hypomethylated regions in agreement with an intergenic H1.2 enrichment (*Appendix I*) and, as said, a genome-wide hypomethylation in cancer. Instead, replication-independent variants (H1.0 and H1X) were found enriched at hypermethylated regions and depleted from hypomethylated ones in contrast to the other studied H1 variants (H1.2 and H1.4). However, our results in breast cancer cells need to be further explored to determine if co-localization of histone H1 variants with DNA methylation changes during cancer are due to a shared chromatin environment or to a direct cross-talk between two epigenetic mechanisms both highly altered in cancer [179, 180]. Indeed, in mice ESCs, Yang et al. demonstrated that DNA methyltransferases DNMT1 and DNMT3B specifically interact with some H1 variants, which promote methylation of imprinted genes [210].

A summary of all genomic features with a particular linker histone H1 variant distribution in T47D breast cancer cell line is shown in Table 7. Furthermore, a review of the specificities and genomics distribution of somatic mammalian histone H1 variants is included in *Appendix II*.

Table 7. Genomic features found specifically enriched in a particular linker histone H1 variant in T47D breast cancer cell line.

FEATURE		**H1.2**	**H1.4, H1.0, H1X**
Presence at promoter/TSS	Active Genes	no	no
	Inactive Genes	no	yes
Coincidence with Nuc+1		no	yes
Depletion at TTS		no	yes
Presence at enhancers and insulators		no	no
Coincidence with repressive core histone PTMs		yes	no
Negative correlation with gene expression		yes	medium
Coincidence with high GC content		no	yes
Coincidence with CpG		no	yes
Coincidence with LADs		yes	no
Enriched regions		intergenic	genes and promoters
Abundance at gene-rich chromosomes		no	yes
Abundance at gene-poor chromosomes		yes	no
FEATURE		**H1X**	**H1.2, H1.4, H1.0**
Coincidence with RNAPII binding sites		yes	no
Methylation state of H1-enriched CpG islands		medium	highest
Enrichment at active coding regions (towards 3' end)		yes	no
Enriched at included exons and retained introns		yes	no
Enriched at short (exon-rich) genes		yes	no
FEATURE		**H1.0**	**H1.2, H1.4, H1X**
Enrichment at NADs		yes	no
Abundance at NAD-rich chromosomes		yes	no
Abundance at rDNA and specific repetitive elements		yes	no
Enrichment at nucleoli		yes	no

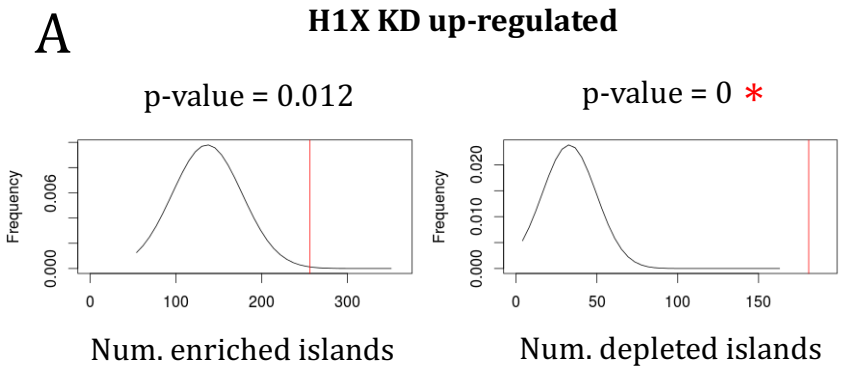
Histone H1 depletion alters gene expression through an unknown mechanism

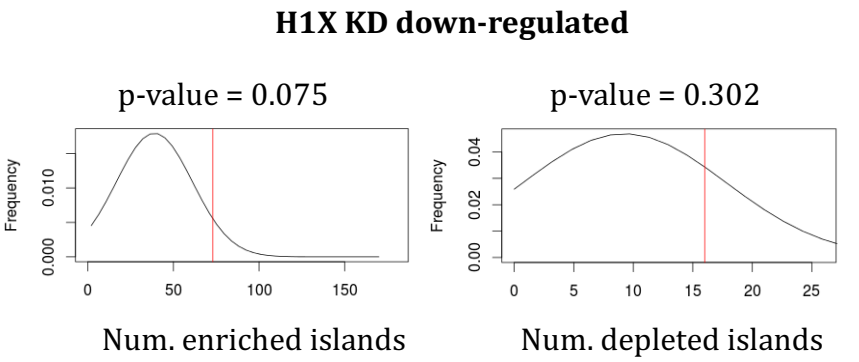
Linker histone H1 has classically been regarded as a structural component of chromatin but we have previously reported that a small subset (1-2% transcriptome) of genes is differentially expressed (DE) upon single H1 variant KD [92]. The subset of DE genes was variant-specific with no special differences in the proportion of up- and down-regulated genes (*Chapter II*, Supplementary Table 2). In addition, when several H1 variants were depleted simultaneously, a bigger number of genes are deregulated (10% transcriptome). In multiple H1 variants depleted cells (H1.2 and H1.4), the total H1 content is reduced roughly 70%, despite an H1.0 up-regulation (*Chapter II*, Figure 1). Although higher numbers of DE genes (1595) are observed in multiple H1 depleted cells compared to single H1 KD, we discard a global chromatin dysregulation where massive transcription is expected. Thus, our results in single and multiple H1 variant depleted cells suggest a variant-specific role in transcription.

Considering the specific distribution of H1X at actively transcribed chromatin (See *Discussion, H1X is associated with actively transcribed chromatin)* and that H1X was missing in our previous analysis [92], we set to analyse transcriptional changes upon H1X single KD. The proportion of up- and down-regulated genes (149 and 45 genes, respectively) upon H1X did not reflect its abundance at active chromatin. Next, we analysed H1 variants abundance at promoters and gene bodies of DE genes upon H1X KD. DE genes did not show a specific H1X enrichment at promoter (*Chapter I*, Figure 8C). Indeed, other studied H1 variants were present at similar levels with a rather global H1 depletion.

Moreover, regarding H1X enrichment at coding regions, we analysed gene bodies of DE genes upon H1X KD. The observed H1X content at exons and introns did not show a specific H1X enrichment (*Chapter I*, Figure 8D). Further, at gene bodies of up-regulated genes in single H1X KD cells, a general low H1 content was observed (Figure 24). And, we were not able to see a specific H1X enrichment or depletion at up- nor down-regulated genes, compared to other H1 variants (Figure 24B).

A





B

Up-regulated genes	Enriched islands	Depleted islands	
H1.0-HA	0,042	0	
H1.2-HA	0,051	0	
H1.2endo_r2	0,003 *	0	
H1.2endo_r3	0,002 *	0	*
H1.4-HA	0,004 *	0	
H1Xendo	0,012	0	
H3	0,011	0	

Down-regulated genes	Enriched islands	Depleted islands
H1.0-HA	0,223	0,034
H1.2-HA	0,834	0,034
H1.2endo_r2	0,098	0,323
H1.2endo_r3	0,41	0,024
H1.4-HA	0,038	0,158
H1Xendo	0,075	0,302
H3	0,04	0,573

Figure 24. Abundance of H1X enriched or depleted islands within DE genes (up- and down-regulated) in single H1X KD cells. (**A**) Frequency of the abundance of H1X enriched or depleted islands within the transcriptome (1000 permutations, samples containing the same number of genes). The *p*-value indicated was computed as the fraction of times the difference between the mean and the random sample is equal or more extreme than the observed difference between the mean and the group of differentially expressed genes. (**B**) p-values computed as in (A) for the abundance of all H1 enriched or depleted islands in H1X KD up- and down-regulated genes.
p-values < 0.01 are marked with red asterisks.

Furthermore, stated as data not shown in the publication and showed here in Figure 25, genes up-regulated with higher fold-changes, upon H1X depletion, did not gain the canonical activation core histone PTM (H3K4me3). Besides, up-regulated genes did not change H1.2 abundance at promoter regions, suggesting that although present, other histone H1 do not form the "H1 valley" seen in active genes, upon H1X depletion (*Chapter I*, Figure 8).

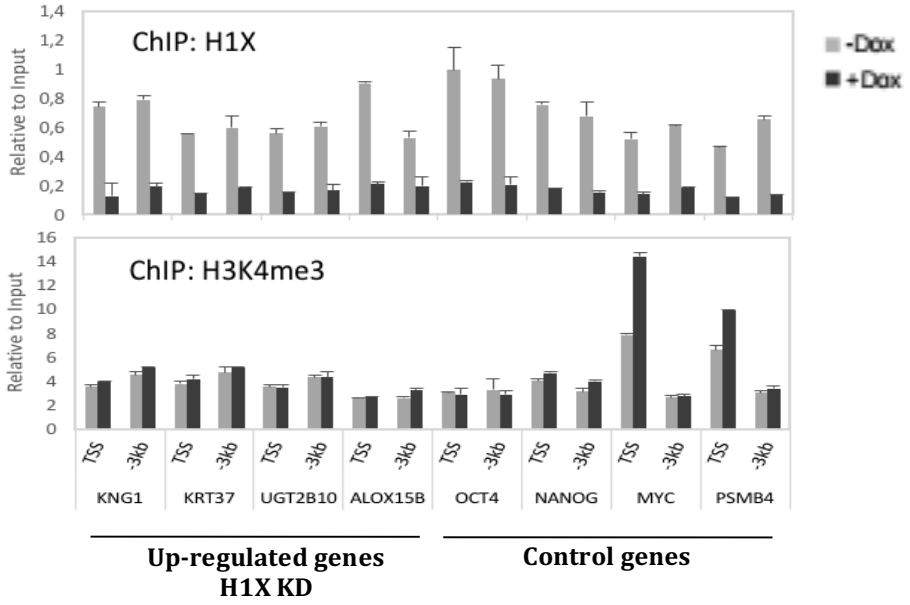


Figure 25. H1X and H3K4me3 abundance at up-regulated gene promoters upon H1X KD. ChIP-qPCR on up-regulated genes and control genes that did not change their expression (silent: OCT4 and NANOG, and active: MYC and PSMB4) upon H1X KD.

How those genes get up-regulated upon single H1X KD independently of core histone PTMs remains unsolved. One possibility could be that H1X is influencing other aspects of gene expression such as elongation or splicing. There is no doubt that further experiments are needed to explore this unknown mechanism of gene activation when depleting H1X protein levels that in our hands seems to be independent of core histone PTMs.

As mentioned, in multiple H1 depleted cells, the number of up- and down-regulated genes (732 and 863, respectively) does not reflect a massive increase in transcription (*Chapter II*, Figure 2B). Interestingly, upon reducing the H1 content a 70% in multiH1 depleted cells, core histone levels and binding remained clearly unchanged. ChIP-qPCR experiments in genic and intergenic regions showed that core histones H3

and H4 are clearly present. Further, multiH1 depleted chromatin extracts (normalized by DNA content) showed no changes in core histone binding levels (*Chapter II*, Figure 6C). Furthermore, global H3 acetylation marking active transcription and repressive PTMs (H3K9me3, H3K27me2) remain globally unchanged. Moreover, repetitive elements, with an increased transcription upon multiH1 KD, neither showed changes in core histone PTMs (*Chapter II*, Figure 6A).

However, assay for transposase-accessible chromatin coupled to high throughput sequencing (ATAC-Seq) in multiH1 KD cells showed a genome-wide chromatin opening. But, we were not able to see specific changes in nucleosome accessibility at up- and down-regulated genes compared to unaltered genes with similar basal expression (*Chapter II*, Figures 7A and 7B).

Hence, our results in breast cancer cells show a similar mechanism of gene induction independent of core histone PTMs, in single H1X and multiH1 depleted cells (Figure 25 and *Chapter II*, Figure 6A). Additionally, up-regulated genes upon single H1X and multiH1 (H1.2 and H1.4) depletion do not show a specific H1X and H1.2 enrichment, respectively (Figures 24 and *Chapter II*, Supplementary Figure 7E). Intriguingly, despite a global increase in chromatin accessibility in multiH1 depleted cells, we were not able to see specific nucleosome changes at DE genes (*Chapter II*, Figure 7B).

Similar results were obtained in *in vitro* transformed fibroblasts, depleted from H1.0 [179]. Assessed by FAIRE-Seq an altered nucleosome occupancy is induced by H1.0 loss and mapping of H3K27ac and H3K27me3 by ChIP-Seq indicated that it was uncoupled from changes in core histone PTMs. However, other studies have reported a direct cross-talk between core histone PTMs and linker histone H1 in different cells. Specifically, in other breast cancer cells (MCF7) it was observed that H1.2 gets recruited to distinct chromatin regions in a manner dependent on EZH2-mediated H3K27me3, and inhibits transcription of multiple growth suppressive genes [207]. However, in 293T cell line, the same group showed H1.2 acting as a gene activator by directly interacting with elongation complexes and serine 2 phosphorylated RNAPII, inducing H3K4me3 among others core histone PTMs [208]. Certainly, the cross-talk between core histone PTMs and linker histone H1 variants needs to be deeply explored to analyse direct or indirect consequences of linker histone H1 presence at promoter regions in different cells.

Disruption of specific transcription factors binding, that might be variant-specific or depending on H1 PTMs might be sufficient to alter gene expression in a core histone PTMs-independent manner.

Another explanation of this activation mechanism seen upon histone H1 depletion could be changes in chromatin structure that creates an environment which facilitates machinery recruitment, independently of specific H1 interactions. Intriguingly, BLACK chromatin, which covers a 48% of *Drosophila melanogaster* genome, assessed by Filion et al. shows a lack of classical heterochromatic marks, no active core histone PTMs (H3K4me3 and H3K79me2) and an extremely low transcription [25]. Linker histone H1 was instead clearly present together with lamin among other proteins. And, we observed an H1.2 variant enrichment in lamin-associated domains (LADs) in breast cancer cells. So, we cannot discard that a disorganized LADs structure, due to histone H1 depletion, is responsible of this unknown mechanism of gene activation in multiH1 KD cells.

Topologically associated-domains (TADs) have been shown to modulate transcriptional environment by long- and short-range DNA interactions (See *Introduction, 1.4. Chromatin spatial organization*). As explained, TADs boundaries were found mainly unchanged in triple KO mice ESC and only a small number of new long-range interaction were observed [102]. Large epigenetic changes were observed within gene-dense TADs but those may reflect indirect effects of a long-term 50% reduction of the H1 content. Specifically, an increase in marks of active promoters and potential enhancers (H3K4me3 and H3K4me1) is observed together with a gain in DNA accessibility, assessed by DNase treatment. Intriguingly, no changes in repressive heterochromatic marks (H3K9me3 and H3K27me3) is seen. As our H1 variants (single and multiple) KD experiments are inducible and depletion is assessed in short time periods (6 days), we cannot discard that the induced genic transcription is due to a disruption of TADs structure and/or distribution that cellular mechanisms cannot rapidly solve.

Among other consequences, TADs disruption leads to new enhancer-promoter interactions that may explain the small subset of DE genes in single and multiple H1 KD. It is also tempting to speculate that specific H1 variants might be repressing specific enhancer regions independently of it 3D chromatin organization. Among enhancers, other regulatory elements are shown to modulate specifically gene transcription through its expression: non-coding RNA (ncRNAs). Indeed, H1X was found enriched at two types of ncRNAs, microRNA (miRNA) and small nucleolar RNA (snoRNA) (*Chapter I*, Figure 2B), which are frequently found in introns [274, 275]. The specific H1X distribution along active genic regions (See *Discussion, H1X is associated*

with actively transcribed chromatin) and its relative enrichment in intronic-regulatory elements, such as miRNA and snoRNA, needs to be further explored.

To wrap up, our results in breast cancer cell line point to specific functions of linker histone H1 variants in activating gene expression. Transcription is induced independently of their abundance at differentially expressed genes and core histone post-translational modifications. We were able to show that multiple H1 depletion leads to a gain in nucleosome accessibility genome-wide that does not fully correspond to transcriptional changes at genic regions.

Although just a mere "open" chromatin might facilitate machinery recruitment independently of histone H1 variants, appealing possibilities appear for this gene activation mechanism upon H1 depletion. As follows, disruption of H1 variants interacting with particular transcription factors might explain our results. In nuclear domains such as lamin- and nucleolus-associated domains (LADs and NADs) we could observe a H1.2 and H1.0 enrichment, respectively. If nuclear domains such as NADs and LADs are disorganized upon histone H1 depletion needs to be further explored together with other unexplored nuclear domains like topological-associated domains (TADs). Beyond promoter regions where histone H1 content, especially H1.2, clearly correlates with gene repression, other transcriptional-related processes might be modulated by linker histone H1 variants. Our results of H1X distribution at coding regions points to elongation and splicing. Besides, another appealing possibility is regulation of specific-regulatory elements like enhancers and non-coding RNAs by specific histone H1 variants.

Multiple H1 variant depletion triggers activation of an interferon response

Interestingly, single H1 variant KD unable cell growth to some extent but, upon multiple H1 variants depletion, more drastic effects on cell proliferation are observed (*Chapter II*, Figure 1F). The majority of DE genes (36%) in multiH1 KD cells and those with a higher induction (fold-changes from 180 to 2) are related to the interferon (IFN) response (*Chapter II*, Figure 2C), named interferon-stimulated genes (ISGs), which may trigger those anti-proliferative effects. ISGs induction in multiH1 KD cells was confirmed by RT-qPCR and time course experiments. Furthermore, removal of doxycycline after 3 days, which leads to a reactivation of H1.2 and H1.4 expression allowed reversion of ISGs expression, showing a great correlation between H1 depletion and ISGs induction (*Chapter II*, Figure 3).

In order to understand how multiH1 depletion induced expression of ISGs, we analysed their H1 variant content at promoters and coding regions. Genes up-regulated showed an increased H1 content at promoter and coding regions compared to down-regulated genes. Nevertheless, up-regulated genes are expressed at lower levels, in basal conditions, than down-regulated ones. Further, we have previously reported that H1 content at promoters and coding regions greatly correlates with expression levels of those genes (*Appendix I*). Thus, the differences in the H1 content observed between up- and down-regulated genes correspond to their basal gene expression (*Chapter II*, Supplementary Figure 7A and 7B). In addition, we could not observe a variant-specific enrichment at ISGs at gene body, nor at promoter regions (*Chapter II*, Figures 6 and Supplementary 7C and 7D). So, we discard a specific H1 role in ISGs repression.

Indeed, interferon is secreted extracellularly in multiH1 depleted cells. As well, chemical or shRNA inhibition of IFN-signalling intermediates inhibits ISGs transcription (*Chapter II*, Figure 4). Further, the fold-changes observed upon multiH1 KD in ISGs and those observed in interferon treatments correlate to some extent (*Chapter II*, Figure 2D). One report described a H1-specific role in ISGs transcription [98] but our results in breast cancer cells point to a normal interferon response due to multiH1 variants depletion.

ISGs transcription in multiH1 depleted cells is achieved without changes in core histone PTMs neither active such as H3 and H4 total acetylation and H3K4me3, nor repressive (H3K9me3, H4K20me3 and H3K27me3) (*Chapter II*, Figure 6D). Although no changes in core histone PTMs are also observed in up-regulated genes in single H1X KD, we believe that ISGs are genes that respond rapidly to stimulus, independently of changes in core histone PTMs. When analysing the core histone PTMs at ISGs we could observe significant amounts of H3K4me3 at their promoter under basal conditions in publicly available data from UCSC browser in several cell lines. Although core histone PTMs are present at promoter regions, upon multiH1 KD, ISGs get transcribed without changes in core histone PTMs. Certainly, if ISGs change their core histone PTMs upon IFN treatment needs to be further addressed. However, developmentally regulated genes, which also respond rapidly to stimulus do not contain core histone PTMs at their promoter regions [13]. During development, those genes get transcribed and no changes in core histone PTMs are observed, similar to our results. Those results and ours break the notion that a direct relationship exists between changes of core histone PTMs and gene transcription in specific genes. Rather than epigenetic mechanisms, transcription factors recruitment and RNAPII activation might play predominant roles in these contexts.

Next, we seek to analyse variant-specific roles in ISGs induction. Multiple H1 depletion is achieved by one inducible shRNA targeting several H1 variants mRNAs, specifically H1.2-H1.5. Nevertheless, reduced protein levels are only seen for H1.2 and H1.4. The fact that there is little overlap between genes deregulated upon H1.2 or H1.4 single KD and multiH1 KD (*Chapter II*, Supplementary Table 2) lead us to combine two specific shRNA for H1.2 and H1.4 KD to discard off-target effects. On the other hand, double H1.2 and H1.4 KD induced ISGs transcription. Induction of ISGs transcription was not observed in single H1 variant KD. Next, H1.5 inducible shRNA was introduced in single H1.2 or H1.4 KD cells. Simultaneous depletion of H1.2 and H1.5 induced ISGs transcription although to a much lesser extent than in H1.2 and H1.4 KD cells. No ISGs transcription was observed in double H1.5 and H1.4 KD cells, suggesting that H1.2 is crucial but not sufficient to induce ISGs transcription in T47D breast cancer cells. Despite, rescue experiments showed that single H1.2-4 variant overexpression was able to reduce ISGs induction in multiH1 KD cells (*Chapter II*, Figure 3).

As explained, linker histone H1 variants may have specific functions depending on the cellular state or process and their specific variant genomic distribution is also distinct. By analysing transcriptomic data in other breast cancer cells (MCF7) and normal human embryonic kidney cells (293T) depleted from H1.2 [207, 208], we were able to detect an unreported induction of the IFN response and a better overlap of DE genes with our multiH1 KD than with single H1.2 KD in T47D breast cancer cells. Thus, H1.2 seems to have a predominant role in prompting an IFN response. In MCF7 and 293T single H1.2 KD cells an IFN response could be observed. However, in T47D cell line, H1.2 needs to be combined with other H1 variants, especially H1.4, to induce an IFN response.

In pancreatic ductal adenocarcinoma (PDAC) an anti-viral state, seen by an ISGs deregulation, segregates two molecular phenotypes [266]. mRNA expression from eight xenografted primary PDAC, three primary PDAC bulk tissues, three pancreatitis and three normal tissues was compared and the specific subset of ISGs dividing adenocarcinomas in two molecular phenotypes was indicated, being some ISGs up- but also down-regulated. We seek to analyse linker histone H1 variant expression in these datasets, however whole microarray raw data was not available. Thus, expression of those ISGs, together with linker histone H1 variants, was assessed in another subset of publically available pancreatic adenocarcinomas (GSE16515), which contains 36 tumours and 16 normal pancreatic tissue samples [267]. The classification of groups depending on ISGs (defined by Monssurrò et al. [266]) expression succeeded in this new data set of pancreatic adenocarcinomas (*Chapter II*, Supplementary Figure 8). We were able to show that the group containing cancer samples with a higher ISGs

induction has a significant increase in replication-independent variants (H1.0 and H1X) and many replication-dependent variants are down-regulated (*Chapter II*, Figure 8). These results are similar to ours in T47D cell line where an IFN response is triggered upon multiH1 replication-dependent KD (H1.2 and H1.4).

In cancer cells, a heterogeneous dysregulation in H1 variants levels is observed; in general, cancer cells show a higher expression of H1 compared to healthy tissues. Colorectal cancer cells show a particular pattern, being H1.2 clearly downregulated in all tumour samples (*Introduction*, Figure 20) [180]. In this regard, we seek to analyse ISGs induction in colorectal cancers using the same method and distinct results were obtained. Despite the loss of H1.2, ISGs transcription is not observed in accordance with single H1.2 or H1.4 KD T47D breast cancer cells. In fact, different to pancreatic cancer, colorectal cancer samples clustering failed to show a subgroup with high ISGs dysregulation. Certainly, if the IFN response seen in pancreatic cancer tissues is due to an aberrant expression of multiple linker histone H1 variants or due to an indirect effect of the activation of the oncogenic program needs to be tightly explored.

To further understand how interferon response is triggered, we analysed possible pathways induced upon depletion of multiH1 variants that prompted ISGs transcription. Pathway sensing nucleic acid toxicity due to increased dsRNA, mediated by RIG-I and MDA5, is significantly induced (See *Introduction, 3. Interferon response and chromatin*). In contrast, the STING-mediated pathway which is prompted upon the presence of dsDNA does not, suggesting that dsRNA molecules produced upon multiH1 KD might be triggering the IFN response. Concomitantly, KD experiments of dsRNA sensors (RIG-I and MDA5) and dsDNA sensors (STING) combined with multiH1 depletion showed that STING unable IFN response to a less extent than RIG-I and MDA5 (*Chapter II*, Figure 4C). Interestingly, we were able to show the presence of cytoplasmic dsRNA and transcription of heterochromatic regions mainly repetitive elements, which might be sensed by cytoplasmic receptors and sensors, triggering an innate immune response.

Multiple H1 variant depletion prompts transcription of repetitive elements

Although new evidences are pointing to a more dynamic and gene-specific function for linker histone H1, its most well-known function is in heterochromatin establishment and/or maintenance (See *Introduction, 2.2. Linker histone H1 function*). Together with DNA methylation, heterochromatin is a key player in repetitive elements silencing that when not properly controlled can lead to an aberrant expression of repetitive elements, triggering an IFN response and high rates of

mutations and chromosomal rearrangement, which can eventually provoke cell death. We were able to show by RT-qPCR that heterochromatic repetitive regions including transposable elements like ERVs and satellites like D4Z4 or alpha satellites are highly transcribed upon multiH1 depletion. Also, the presence of cytoplasmic dsRNA is increased and intergenic transcription is observed in RNA-Seq data (*Chapter II*, Figure 5).

Repetitive elements changed to a more open chromatin state upon multiH1 KD, assessed by ATAC-Seq (*Chapter II*, Figure 7). As explained for DE genes in H1X depleted cells, how these repetitive elements get activated upon multi H1 KD is one remaining open question as we were able to show a gain in DNA accessibility but no changes in core histone PTMs globally nor at ISGs promoters nor at repetitive elements (*Chapter II*, Figure 6). A main component of heterochromatin, HP1α (heterochromatin protein 1) was globally reduced in multiH1 KD chromatin in agreement with a gain in nucleosome accessibility (*Chapter II*, Figure 6C). As methylated H1.4K26 has been shown to be needed in its recruitment [149], H1.4 depletion could have an impact on chromatin-bound HP1α, potentially impacting heterochromatin integrity and repeat repression. Although it is worth mentioning, that we could not observe an IFN response in single H1.4 KD cells

One possibility could be that DNA methylation levels at repetitive elements are changed and a direct cause of its transcriptional activation. Actually, as explained, Yang et al. demonstrated that DNA methyltransferases DNMT1 and DNMT3B specifically interact with some H1 variants in mice [210] and several reports showed that a reduction in DNA methylation, by treatment with DNMT inhibitors such as 5-aza-2'-deoxycytidine (aza-dC), leads to endogenous retroviruses (ERVs) transcription and dsRNA formation, triggering an IFN response [241, 242]. Upon multiH1 KD in breast cancer cells, the DNA methylation levels remained globally unchanged assessed by a genome-wide CpG array. Only six out of 850.000 studied CpG changed their methylation level and their associated gene did not change its expression. Three HERV-K elements were also tested by bisulphite PCR and no changes were found. Treatment with aza-dC for three days did not induced an IFN response and transcription of repetitive elements was heterogeneous although when combined with multiH1 KD, enhanced (*Chapter II*, Supplementary Figure 9). Thus, our results suggest that in T47D breast cancer cell line, linker histone H1 is responsible for repetitive elements silencing, independently of DNA methylation.

Recently published results showed that depletion of the single somatic H1 variant in *Drosophila melanogaster* (dH1) leads to increased levels of γH2Av indicating DNA damage, specifically double-strand breaks (DSB), preferentially located at heterochromatic elements [265]. They showed that dH1 depletions leads to an abnormal accumulation in heterochromatin of DNA:RNA hybrids (R-loops) detectable in G1-phase. In humans, H1 has been shown to couple initiation and amplification of ubiquitin signalling after DNA damage [155] and H1.2 to specifically induce apoptosis upon DSB in a p53-dependent manner [105]. We cannot discard that the IFN response seen in human multiH1 KD is due to DSB and DNA:RNA hybrids accumulation, which can also trigger an innate immune response (See *Introduction, 3. Interferon response and chromatin*) and are also modulated by linker histone H1.

Besides, one of the interacting partners of linker histone H1.2 is DNA-PK [206]. Phosphorylation at T146 by DNA-PK has been shown to modulate apoptosis by disrupting a complex containing H1.2 and p53 [129]. In addition, DNA-PK has been shown to directly interact with STING and induce an IFN response upon DSB (*Introduction,* Figure 22) [245]. We could not observe an induction of dsDNA sensing pathways neither an increase in DNA-PK in our RNA-Seq data and in the single H1.2 KD an IFN response is not prompted. Thus, we believe that DNA-PK does not play a major role in the IFN response seen upon depletion of multiH1 variants.

Another unsolved question is if the IFN response is due to transcription of several repetitive elements or only because of one/some of them with great induction. Upon multiH1 KD, from the repetitive elements tested, the one with a higher fold-change is satellite alpha (*Chapter I*I, Figure 5). As explained, satellite alpha is mainly located at human centromeres and Maida et al. recently showed that their transcription, together with other transposons, influenced heterochromatin formation at centromeres [59]. In mice triple KO (TKO) mice of linker histone H1 variants (H1c, H1d and H1e; H1.2, H1.4 and H1.3 in human, respectively) produced an increased expression of major satellite repeats, mainly located at mice centromeres, independent of multiple epigenetic mark and DNA methylation, similar to our obtained results in human [216]. Satellite repeats cluster around centromeres, attract pericentromeric heterochromatin and aggregate into nuclear chromocenters in mice [264]. In TKO mice, they also showed by FISH techniques that chromocenters are significantly clustered together resulting in a reduced number. These results in TKO mice and the high expression of alpha satellites seen in our study, using multiH1 KD T47D breast cancer cell, needs to be further explored as we cannot discard the possibility that the observed IFN response is sole due to alpha satellite transcription and/or a disorganized centromere establishment, organization or maintenance.

Regarding the predominant role of one repetitive element in triggering an IFN response in multiH1 KD cells, one report analysing IRF1 and STAT1 binding genome-wide recently showed that those transcription factors, which are activated by IFN, bind to transposable elements. CRISPR-Cas9 deletion of a subset of these ERV elements impaired the expression of the adjacent ISG and revealed a regulatory role of those elements in the IFN response. It was stated that those elements constitute a dynamic reservoir of IFN-inducible enhancers. Other reports (reviewed in [62]) showed other regulatory activities of transposable elements. Certainly, a deep screen of repetitive elements is needed to discard indirect effects and maybe find new specific regulatory functions of linker histone H1 in these repetitive elements that, as recently shown, may act as regulatory enhancers.

Enrichment of particular H1 variants at repetitive DNA

Our genome-wide analysis of linker histone H1 variant distribution have shown that some H1 variants are enriched in specific repetitive elements compared to other linker histone H1 variants, pointing to a possible variant-specific repression mechanism. Among telomeric and ACRO1 satellites, a high abundance of H1.0 was found in SINE-VNTR-Alu (SVA), a recently evolved, hominid-specific non-LTR retrotransposon (*Chapter I*, Figure 2D). Together with other non-autonomous LTR-retrotransposons such as *Alu*, SVA retrotransposition relay in LINE1 machinery acting in *trans*. The several thousand copies of SVA in the human genome are relatively small compared with 500.000 copies of LINE1 and 1.100.000 copies of Alu. However, all of them have been found active in humans, causing complex human diseases when not properly controlled [271, 272].

The specific enrichment of H1.0 at SVA repetitive elements and their active transcription in human suggests a specific repression mechanism by this variant. The role of H1.0 in repressing repetitive elements is intriguingly. We have shown H1.0 enrichment at specific repetitive elements (SVA) and in perinucleolar heterochromatin, which is abundant in repetitive elements. As mentioned, H1.0 is only found in differentiated cells where repression of repetitive elements is tightly controlled. Pluripotent ESCs, depleted from H1.0, have a high transcription of repetitive elements without deleterious effects [69]. And, transcription of repetitive elements is naturally modulated in development achieving a compete repression before implantation [251]. During reprogramming to pluripotency, an uncontrolled transcription of endogenous retroelements is observed and it is already known that H1.0 depletion impairs differentiation [71, 178]. Indeed, H1.0 promoter contains a bivalent domain of pluripotent ESCs and histone H1 variants are differentially expressed and incorporated

through reprogramming to induced pluripotent stem cells (iPSCs) and differentiation of hESCs, especially H1.0 [178].

It is thus, tempting to speculate that H1.0 is specifically regulating repetitive elements transcription during differentiation as happens in pluripotent ESCs, depleted of H1.0, where the H3.3 variant is required for ERVs silencing [273]. It is worth mentioning H3.3 inhibits histone H1 binding keeping diverse genomic site in an open chromatin state in *Drosophila melanogaster* [96]. In this regard, interesting results were obtained when overexpressing H1(0) in mice ESC. Like H1c and H1d (H1.2 and H1.4 in humans, respectively), overexpressed H1(0) was found enriched at major satellite however, other repetitive elements like minor satellites and LINEs showed a specific H1(0) enrichment [216]. In agreement, LINE-1 and other retrotransposons (like intracisternal A particle (IAP) family) are developmentally regulated [251]. Further experiments are of great interest as low expression of H1.0 has been described as a intratumour marker of CSCs and a main feature of cancer cells is its genomic instability, in part due to an aberrant expression of repetitive elements [179, 253, 272, 276].

Multiple H1 depletion leads to increased levels of H1.0 variant that depend on core histone acetylation

The only H1 variant that is greatly up-regulated at mRNA and protein levels upon changing the levels of H1 variants is H1.0, and this does not happen for the other replication-independent variant H1X (*Chapter II*, Figures 1C, 2A and 6C). Further, other replication-independent variants repressed in these cells, like testis-specific H1T, neither change its expression (data not shown, RT-qPCR and RNA-Seq data). As we have previously reported [92], single H1 variant depletion does not induce great changes in the other H1 variant protein levels but we cannot discard redistribution. It is worth mentioning that H1.0 is not involved in the interferon response seen in multiH1 depleted cells. Depletion of H1.0 did not alter ISGs induction in multiH1 depleted cells.

The great increase of H1.0 variant in multiH1 depleted cells lead us to analyse by ChIP-qPCR core histone PTMs in the promoter region of the gene encoding for H1.0 (*H1F0*) (*Chapter II*, Figure 6A). It could be observed that under normal conditions, H1.0 locus has typical features of an active gene compared to repressed genes (NANOG) or intergenic sequences: low H1.2 content, increase of H3K4me3, H3 and H4 acetylation and reduced levels of repressive marks, mainly H3K9me3 and H3K27me3 at TSS. Upon multiH1 depletion a specific gain in H3 and H4 acetylation is observed at TSS. So, H1.0 promoter is regulated by histone acetylation, in agreement with its induction by HDAC

inhibitors like Trichostatin A (TSA) (Figure 26). Apart from core histone acetylation, other tested active and repressive marks that rely in core histone methylation like H3K4me3 and H3K9me3, H3K27me and H4K20me3 did not change.

Genes of replication-dependent variants are clustered together in contrast to replication-independent (H1.0 and H1X). In agreement with their localization, our results show a different promoter regulation in replication-independent variants upon TSA treatment.

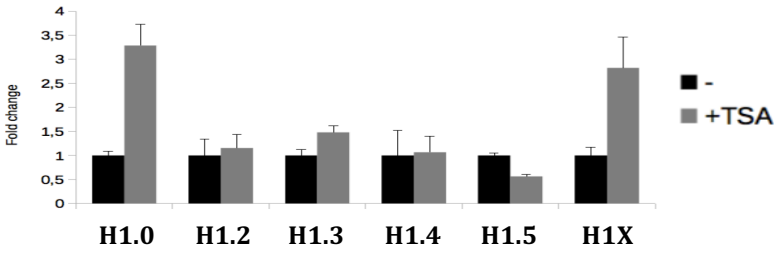


Figure 26. mRNA expression levels of somatic linker histone H1 variants in T47D breast cancer cell line treated or not with HDAC inhibitor Trichostatin A (TSA). RT-qPCR values are normalized to GAPDH expression relative to untreated cells indicated as fold change.

H1.0 and, probably, H1X are synthesised independently of DNA replication, and accumulate in differentiated cells when cells stop proliferating may be replacing replication-dependent variants. In our hands, upon depletion of H1.2, H1.4 or H1.2+H1.4 variants we observed an induction of H1.0, which might be sensing H1 levels (*Chapter II*, Figures 1A and 2A). Besides, an aberrant proportion of histone H1 variants has been observed in different cancer types and in cancer stem cells (CSCs) within a tumour might have a reduced content of H1.0 [179, 180]. Changes in H1 variant levels, which are mainly sensed by H1.0 variant in our hands need to be certainly always assessed in order to attribute H1 variant-specific roles.

So, histone H1.0 role in sensing H1 levels, and its specific enrichment at nucleolar domains and SVA retrotransposons, in cellular processes such as differentiation, reprogramming and cancer are still unsolved and interesting questions in the field (See *Discussion, H1.0 is enriched at nucleolar chromatin* and *Enrichment of particular H1 variants at repetitive DNA*).

In summary, replication-independent variants (H1.0 and H1X) in T47D breast cancer cells have a distinct genomic distribution compared to the other studied replication-dependents variants (H1.2 and H1.4) (Chapter I). H1.0 was found enriched at nucleoli-associated features such as NADs, NORs encoding for the 45S rDNA, specifically at non-transcribed spacers and in 5S rDNA. Specific repetitive sequences such as SINE-VNTR-Alu (SVA) retrotransposons and telomeric and ACRO1 satellites showed also a specific enrichment of H1.0.

In contrast, H1X has been associated to actively transcribed chromatin indicated by a colocalization with RNAPII-enriched regions and an enrichment towards the 3' end of active genes. In addition, all coding regions that are included in the final mRNA (constitutive exons, included ASE and retained introns) are enriched in H1X. Further, specific non-coding RNA (miRNA and snoRNA), mainly found at introns showed an H1X enrichment. Our results point to a potential role of H1X in elongation, splicing or non-coding RNA regulation, which might be prompting gene transcription without changes in core histone PTMs.

Depletion of multiple H1 variants triggers an interferon response due to an aberrant transcription of repetitive elements seen by RT-qPCR, increase in cytoplasmic dsRNA and transcription of intergenic regions (Chapter II). Although H1.2 and H1.4 are critical in the observed phenotype, rescue experiments showed redundant functions for H1 variants in breast cancer cells. The molecular mechanism that leads to transcription of repetitive elements upon multiH1 KD, as happens for DE genes upon single or multiH1 variants KD, is still unsolved. We were able to show an increase in nucleosome accessibility genome-wide that did not fully correlate with transcriptional changes, and core histone PTMs remained unchanged.

Specific molecular mechanisms, involved in transcriptional modulation, that might be regulated by a particular H1 variant (or H1 variant combinations) are appealing possibilities. Among them, establishment, maintenance or organization of nuclear domains (LADs, NADs or TADs), chromosome structures (centromeres) or localised heterochromatin regions (transposons). Beyond promoters where histone H1 content clearly correlate with repression, other transcription-related processes might be regulated by specific H1 variants. Processes influenced by RNAPII (elongation or splicing) and other regulatory elements (non-coding RNAs or enhancers) need to be certainly explored in a histone H1 variant(s) depletion context.

Upon single and multiple H1 variants depletion, H1.0 is induced in a regulated manner that may depend in histone acetylation, assessed by ChIP-qPCR at promoter regions and by treatments with histone deacetylase inhibitor (TSA). Further experiments are needed to elucidate relocation of histone replication-independent H1 variants, mainly H1.0 upon changing H1 stoichiometry and during differentiation, reprogramming and cancer.

CONCLUSIONS

CONCLUSIONS

Genomic distribution of linker histone H1 variants (ChIP-Seq) shows H1.2 is the variant with a more particular distribution, being enriched at lamin-associated domains (LADs) and at GC-poor, gene-poor and intergenic regions. Moreover, H1.2 is the variant that better correlates with gene repression at promoter and coding regions (*Appendix I*).

Further focus shows particularities for replication-independent variants H1.0 and H1X:

1. H1.0 is associated with nucleolar domains
    - **a.** H1.0 is enriched at nucleolus-associated domains
    - **b.** H1.0 is enriched at nucleolus organizer regions, which contain the 45S ribosomal DNA, specifically at non-transcribed spacers
    - **c.** H1.0 is enriched at 5S ribosomal DNA

2. H1.0 is enriched at SINE-VNTR-Alu retrotransposon, ACRO1 and telomeric satellites compared to H1X, H1.2 and H1.4

3. H1X is associated with actively transcribed chromatin
    - **a.** H1X is the H1 variant that better colocalises with RNA polymerase II-enriched regions
    - **b.** H1X is enriched towards the 3' end of active genes
    - **c.** H1X is enriched at exons compared to introns
    - **d.** Included alternatively spliced exons and retained introns have a higher H1X content than constitutive introns and exons
    - **e.** H1X enriched-islands overlap with CpG islands is higher than for other variants
    - **f.** CpG islands enriched in H1X are hypomethylated

4. H1X is enriched at micro RNAs and small nucleolar RNAs compared to H1.0, H1.2 and H1.4

Histone H1 depletion triggers an interferon response in cancer cells via activation of heterochromatic repeats:

5. Multiple H1 variants depletion triggers an interferon response
    a. RNA-Seq data showed that 36% of up-regulated genes are interferon-stimulated genes
    b. Linker histone H1 variants are not particularly enriched at coding region and promoters of interferon-stimulated genes
    c. Histone H1 content is inversely correlated with basal gene expression
    d. Gene Ontology and motif analyses showed enrichment of genes that respond to IFN type I signalling

6. dsRNA sensing pathways are induced in multiple H1 variants depleted cells

7. Multiple H1 variants depletion induces transcription of heterochromatic repetitive elements
    a. RT-qPCR in repetitive elements show an increased expression
    b. Immunofluorescence experiments show an increase of cytoplasmic dsRNA
    c. Increased intergenic transcription is observed by RNA-Seq

8. Upon multiple H1 variants depletion, ATAC-Seq experiments showed a global increase in nucleosome accessibility

9. Compared to uninduced genes with same basal gene expression, differentially expressed genes, upon multiple H1 variants depletion, do not show particular changes in chromatin accessibility (ATAC-Seq)

10. Multiple H1 variants depletion (as well as single H1X depletion) induce gene expression without changes in core histone post-translational modifications at promoters of up-regulated genes, nor at induced repetitive elements

11. Upon multiple H1 variants depletion, H1.0 is induced at mRNA and protein levels, concomitant to increased acetylation of H3 and H4 at the H1.0 gene promoter

REFERENCES

REFERENCES

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

2. Fyodorov DV, Zhou BR, Skoultchi AI and Bai Y (2017) Emerging roles of linker histones in regulation chromatin structure and function. *Nat Rev Mol Cell Biol*, doi: 10.1038/nrm.2017.94

3. Thoma F, Koller T, Klug A. (1979) Involvement of linker histone H1 in the organization of the nucleosome and the salt-dependent superstructures of chromatin. *J Cell Biol*, 83: 403-27

4. Li B, Carey M and Workman JL. (2007) The role of chromatin during transcription. *Cell*, 128: 707-19

5. Luger K, Mäder AW, Richmond RK, Sargent DF and Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8A resolution. *Nature*, 389: 251-60

6. Olins, DE and Olins AL. (2003) Chromatin history: over view from the bridge. Nature Rev *Mol Cell Biol*, 4: 811

7. Smolle M and Workman JL (2013) Transcription-associated histone modifications and cryptic transcription. *Biochim Biophys Acta*, 1829: 84-97

8. Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek A, Chen T, Li E Jenuwein T and Peters AH (2003) Suv39-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellites repeats at pericentric heterochromatin. *Curr Biol*, 13: 1192-2000

9. Rougeulle C, Chaumeil J, Sarma K, Allis CD, Reinberg D, Avner P and Heard E (2004) Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome. *Mol Cell Biol*, 24: 5475-84

10. Barth TK and Imhof A (2010) Fast signal and slow marks: the dynamics of histone modifications. *Trends Biochem Sci*, 35: 618-26

11. Calo E and Wysocka J (2013) Modification of enhancer chromatin: what, how and why? *Mol Cell*, 49: 825-37

12. Strahl BD and Allis CD (2000) The language of covalent histone modifications. *Nature*, 403: 41-5

13. Perez-LLuch S, Blanco E, Tilgner H, Curado J, Ruiz-Romero M, Corominas M and Guigó R (2015) Absence of canonical marks of active chromatin in developmentally regulated genes. *Nat Genet*, 47: 1158-67

14. Pusarla RH and Bhargava P (2005) Histones in functional diversification. Core histone variants. *FEBS J*, 272: 5149-68

15. Loyola A and Almouzni G (2007) Marking histone H3 variants: how, when and why? *Trends Biochem Sci*, 31: 425-33

16. Szenker E, Ray-Galler D and Almouzni G (2011) The double face of the histone variant H3.3. *Cell Res*, 21: 421-34

17. Bönisch C and Hake SB (2012) Histone H2A variants in nucleosomes and chromatin: more or less stable. *Nucleic Acids Res*, 40: 10719-41

18. Marques M, Laflamme L, Gervais AL and Gaudreau L (2010) Reconciling the positive and negative roles of histone H2A.Z in gene transcription. *Epigenetics*, 5: 267-72

19. Han W, Li X and Fu X (2011) The macro domain protein family: structure, functions, and their potential therapeutic implications. *Mutat Res*, 727: 86-103

20. Pasque V, Gillich A, Garrett N and Gurdon JB (2011) Histone variant macroH2A confers resistance to nuclear reprogramming. *EMBO J*, 30: 2373-87

21. Tolstorukov MY, Goldman JA, Gilbert C, Orgyzko V, Kingston RE and Park PJ (2012) Histone variant H2A.Bbd is associated with active transcription and mRNA processing in human cells. *Mol Cell*, 47: 596-607

22. Ishibashi T, Li S, Eirin-López JM, Zhao M, Missaen K, Abbott DW, Meistrich M, Hendzel MJ and Ausió J (2010) H2A.Bbs: an X-chromosome-encoded histone involved in mammalian spermiogenesis. *Nucleic Acids Res*, 38: 1780-9

23. Mescher AL (2009) Junqueira's Basic Histology: Text Atlas, 12th edition

24. Puschendorf M, Terranova R, Boutsma E, Mao X, Isono K, Brykczynska U, Kolb C, Otte AP, Koseki H, Orkin SH, van Lohuizen M and Peters AH (2008) PRC1 and Suv39h specify parental asymmetry at constitutuve heterochromatin in early mouse embryos. *Nat Genet*, 40: 411-20

25. Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro I, Kerkhoven RM, Bussemaker HJ and van Steensel B (2010) Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell*, 143: 212-24

26. Ernst J and Kellis M (2010) Discovery and characterization of chromatin states for systemic annotation of the human genome. *Nat Biotechnol*, 28: 817-25

27. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M and Bernstein BE (2011) Mapping and analysis of chromatin state dynamics in nine cell types. *Nature*, 473: 43-9

28. Enrst J and Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods*, 9: 215-6

29. Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA and Noble WS (2012) Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods*, 9: 473-6

30. Pombo A and Dillon N (2015) Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Biol*, 16: 245-57

31. Guelen L, Pagie L, Brasset e, Meuleman W, Faza MB, Talhout W, Eussen BH, de Kein A, Wessels L, de Laat W and van Steensel B (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, 453: 948-51

32. Meuleman W, Peric-Hupkes D, Kind J, Beaudry JB, Pagie L, Kellis M, Reinders M, Wessels L and van Steensel B (2013) Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res*, 23: 270-80

33. van Steensel B and Belmont AS (2017) Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell*, 169: 780-91

34. Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, Noushmeher H, Lange CP, van Dijk CM, Tollenaar AR, van Den Ber D and Laird PW (2011) Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet*, 44: 40-6

35. Németh A, Conesa A, Santoyo-Lopez J, Medina I, Montaner D, Peterfia B, Solovei I, Cremer T, Dopazo J and Langst G (2010) Initial genomics of the human nucleolus. *PLoS Genet*, 6: e1000889

36. van Koningsbruggen S, Gierlinski M, Schofield P, Martin D, Barton GJ, Ariyurek Y, den Dunnen JT and Lamond AI (2010) High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Mol Biol Cell*, 21:3735-48

37. Kind J, Pagie L, Ortabozkoyun H, Boyle S, de Vries SS, Janssen H, Amendola M, Nolen LD, Bickmore WA and van Steensel B (2013) Single-cell dynamics of genome-nuclear lamina interactions. *Cell*, 153: 178-92

38. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS and Ren B (2012) Topological domains in mammalian genomes identifies by analysis of chromatin interactions. *Nature*, 485: 376-80

39. Cremer T and Cremer M (2010) Chromosome territories. *Cold Spring Harb Perspect Biol*, 2: a003889

40. Cubeñas-Potts C and Corces VG (2015) Topologically associating domains: an invariant framework or a dynamic scaffold? *Nucleus*, 6: 430-4

41. Schwartz S, Meshorer E and Ast G (2009) Chromatin organization marks exon-intro structure. *Nat Struct Mol Biol*, 16: 990-5

42. Andersson R, Enroth S, Rada-Iglesias A, Wadelius C and Komorowski J (2009) Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genomes Res*, 19: 1732-41

43. Hnilicová J and Stanek D (2011) Where splicing joins chromatin. *Nucleus*, 2(3): 182-188

44. Nogues G, Kadener S, Cramer P, Bentley D and Kornblihtt AR (2002) Transcriptional activator differ in their abilities to control alternative splicing. *J Biol Chem*, 277: 43110-4

45. Luco RF, Allo M, Schor IE, Kornblihtt AR and Misteli T (2011) Epigenetics in alternative pre-mRNA splicing. *Cell*, 144: 16-26

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

47. Jelinic P, Pellegrino J and David G (2011) A novel mammalian complex containing Sin3B mitigates histone acetylation and RNA polymerase II progression within transcribed loci. *Mol Cell Biol*, 31: 54-62

48. Saint-André V, Batsché R, Rachez C and Muchardt C (2011) Histone H3 lysine 9 trimethylation and HP1γ favour inclusion of alternative exons. *Nat Struct Mol Biol*, 18: 337-44

49. Smallwood A, Hon GC, Jin F, Henry RE, Espinosa JM and Ren B (2012) CBX3 regulates efficient RNA processing genome-wide. *Genome Res*, 22: 1426-36

50. de Almeida SF, Grosso AR, Koch F, Fenouil R, Carvalho S, Andrade J, Levezinho H, Gut M, Eick D, Gut I, Andrau JC, Ferrier P and Carmo-Fonseca M (2011) Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys 36. *Nat Struct Mol Biol*, 18: 977-83

51. Kim S, Kim H Fong N, Erickson B and Bentley DL (2011) Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc Natl Acad Sci U S A*, 108:13564-9

52. Saksouk N, Simboeck E and Déjardin J (2015) Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin*, 8: 3

53. Adega F, Guedes-Pinto H and Chaves R (2009) Satellite DNA in the karyotype evolution of domestic animals—clinical considerations. *Cytogenet Genome Res*, 126: 12-20

54. Gasser SM (2016) Selfish DNA and epigenetic repression revisited. *Genetics*, 204: 837-839

55. Aldrup-MacDonald and Sullivan BA (2014) The past, present and future of human centromere genomics. *Genes*, 5: 33-50

56. McKinley KL and Cheeseman IM (2016) The molecular basis for centromere identity and function. *Nat Rev Mol Cell Biol*, 17: 16-29

57. Henikoff S, Ahmad K and Malik HS (2001) The centromere paradox, stable inheritance with rapidly evolving DNA. *Science*, 293: 1098–102

58. Biscotti MA, Canapa A, Froconi M, Olmo E and Barucca M (2015) Transcription of tandemly repetitive DNA: functional roles. *Chromosome Res*, 23: 463-77

59. Maida Y, Yasukawa M, Okamoto N, Ohka S, Kinoshita K, Totoki Y, Ito TK, Minamino T, Nakamura H, Yamaguchi S, Shibata T and Masutomi K (2014) Involvement of telomerase reverse transcriptase in heterochromatin maintenance. *Mol Cell Biol*, 34: 1576-93

60. Quénet D and Dalal Y (2014) A long noncoding RNA is required for targeting centromeric protein A to the human centromere. *eLife*, 3: e03254

61. Ayarpadikannan and Kim HS (2014) The impact of transposable elements in the genome evolution and genetic instability and their implications in various diseases. *Genomics Inform*, 12: 98-104

62. Chuong EB, Elde NC and Feschotte C (2017) Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet*, 18: 71-86

63. Walter M, Tessandier A, Pérez-Palacios R and Bourc'his D (2016) An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife*, 5: e11418

64. Tuan D and Pi W (2014) In human β-globin gene locus, ERV-9 LTR retrotransposon interacts with and activates β- but not γ-globin gene. *Blood*, 124: 2686

65. Emera D, Casola C, Lynch VJ, Wildman DE, Agnew D and Wagner GP (2012) Convergent evolution of endometrial prolaction expression in primates, mice, and elephants through the independent recruitment of transposable elements. *Mol Biol Evol*, 29: 239-47

66. Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D and Pfaff SL (2012) Emryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, 487: 57-63

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

68. Santoni FA, Guerra J and Luban J (2012) HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology*, 9: 111

69. Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoeppner DJ, Dash C, Bazett-Jones DP, Le Grice S, McKay RDG, Buetow KH, Gingeras TR and Mistelli T (2008), Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell*, 2: 437-47

70. Ricci MA, Manzo C, García-Parajo MF, Lakadamyali M and Cosma MP (2015) Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell*, 160: 1145-58

71. Friedli M, Turelli P, Kapopoulou A, Rauwel B, Castro-Díaz N, Rowe HM, Ecco G, Unzu C, Planet E, Lombardo A, Mangeat B, Wildhaber BE, Naldini L and Trono D (2014) Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genomes Res*, 24: 1251-9

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

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

74. Shimamura A, Sapp M, Rodriguez-Campos A, Worcel A (1989) Histone H1 represses transcription from minichromosomes assembled in vitro. *Mol Cell Biol*, 9: 5573-84

75. O'Neill TE, Meersseman G, Pennings S, Bradbury EM (1995) Deposition of H1 onto reconstituted nucleosome arrays inhibits both initiation and elongation of transcripts by T7 RNA polymerase. *Nucleic Acids Res*, 23: 1075-82

76. Pennings S, Meerssemann G, Bradbury EM (1994) Linker histones prevent the mobility of positioned nucleosomes. *Proc Natl Acad Sci U S A*, 91: 10275-9

77. Hill DA (2001) Influence of linker histone H1 on chromatin remodelling. *Biochem Cell Biol*, 79: 317–24

78. Horn PJ, Carruthers LM, Logie C et al. (2002) Phosphorylation of linker histones regulates ATP-dependent chromatin remodelling enzymes. *Nat Struct Biol*, 9: 263-7

79. Ramachandran A, Omar M, Cheslock P, Schnitzler GR (2003) Linker histone H1 modulates nucleosome remodeling by human SWI/SNF. *J Biol Chem*, 278: 48590-601

80. Weintraub H (1985) Histone-H1-dependent chromatin superstructures and the suppression of gene activity. *Cell*, 38: 17-27

81. Bresnick EH, Bustin M, Marsaud V, Richard-Foy H, Hager GL (1992) The transcriptionally-active MMTV promoter is depleted of histone H1. *Nucleic Acids Res*, 20: 5278-417

82. Smith CL and Hager GL (1997) Transcriptional regulation of mammalian genes in vivo. *J Biol Chem,* 272: 27493-6

83. Ricci MA, Manzo C, García-Parajo MF, Lakadamyali M and Cosma MP (2015) Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell*, 160: 1145-58

84. Nishiyama N, Skoultchi AI and Nakayama KI (2012) Histone H1 recruitment by CDH8 is essential for supression of the Wnt-β-catenin signaling pathway. *Mol Cell Biol*, 32: 501-12

85. Nishiyama M, Oshikawa K, Tsukada Y, Nakagawa T, Iemura S, Natsume T, Fan Y, Kikuchi A, Skoultchi AI and Nakayama KI (2009) CDH8 suppresses p53-mediated apoptosis through histone H1 recruitment during early embryogenesis. *Nat Cell Biol*, 11: 172-82

86. Maclean JA, Bettegowda A, Kim BJ, Lou CH, Yang SM, Bhardwaj A, Shanker S, Hu Z, Fan Y, Eckardt S, McLaughlin KJ, Skoultchi AI and Wilkinson MF (2011) The *rhox homeobox* gene cluster is imprinted and selectively targeted for regulation by histone H1 and DNA methylation. *Mol Cell Biol*, 31: 1275-87

87. Misteli T, Gunjan A, Hock R, Bustin M and Brown DT (2000) Dynamic binding of histone H1 to chromatin in living cells. *Nature*, 408: 877-81

88. Lever MA, Th'ng JP, Sun X and Hendzel MJ (2000) Rapid exchange of histone H1.1 on chromatin in living human cells. *Nature*, 408: 873-6

89. Brown DT (2003) Histone H1 and the dynamic regulation of chromatin function. *Biochem Cell Biol*, 81: 221-7

90. Shen X and Gorovsky MA (1996) Linker histone H1 regulates specific gene expression but not global transcription in vivo. *Cell*, 86: 475-83

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

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

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

94. Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts JA, Mahony S, Pugh BF, Lee D, Kaestner KH and Zaret KS (2016) The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol Cell*, 62:79-91

95. Ghosh RP, Horowitz-Scherer RA, Nikitina T, Shlyakhtenko LS and Woodcock CL (2010) MeCP2 binds cooperatively to its substrate and compete with histone H1 for chromatin binding. *Mol Cell Biol*, 30: 4656-70

96. Braunschweig U, Hogan GJ, Pagie L and van Steensel B (2009) Histone H1 binding is inhibited by histone variant H3.3. *EMBO J*, 28: 3635-45

97. Chen ZH, Zhu M, Yang J, Liang H, He J, He S, Wang P, Kang X, McNutt MA, Yin Y and Shen WH (2015) PTEN interacts with histone H1 and controls chromatin condensation. *Cell Rep*, 8: 2003-14

98. Kadota S and Nagata K (2014) Silencing of IFN-stimulated gene transcription is regulated by histone H1 and its chaperone TAF-I. *Nucleic Acids Res*, 42: 7642-53

99. Lu X, Wontakal SN, Kavi H, Kim BJ, Guzzardo PM, Emelyanov AV, Xu N, Hannon GJ, Zavadil J, Fyodorov DV and Skoultchi AI (2013) *Drosophila* H1 regulates the genetic acativity of heterochromatin by recruitment of Su(var)3-9. *Science*, 340: 78-81

100. Andreyeva EN, Bernardo TJ, Kolesnikova TD, Lu X, Yarinich LA, Bartholdy BA, Guo X, Posukh OV, Healton S, Willcockson MA, Pindyurin AV, Zhimulev IF, Skoultchi AI and Fyodorov DV (2017) Regulatory functions and chromatin loading dynamics of linker histone H1 during endoreplication in *Drosophila*. *Genes Dev*, 31: 603-16

101. Xu N, Emelyanov AV, Fyodorov DV and Skoultchi AI (2014) *Drosophila* linker histone H1 coordinates STAT-dependent organization of heterochromatin and suppresses tumorigenesis caused by hyperactive JAK-STAT signalling. *Epigenetics Chromatin*, 7: 16

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

103. Zlatanova JS, Srebreva LN, Banchev TB, Tasheva BT and Tsaney RG (1990) Cytoplasmatic pool of histone H1 in mammalian cells. *J Cell Sci*, 96: 461-8

104. Ye X, Feng C, Gao T, Mu G, Zhu W and Yang Y (2017) Linker histone in diseases. *Int J Biol Sci*, 13: 1008-18

105. Konishi A, Shimizu S, Hirota J, Takao T, Fan Y, Matsuoka Y, Zhang L, Yoneda Y, Fujii Y, Skoultchi AI and Tsujimoto Y (2003) Involvement of histone H1.2 in apoptosis induced by DNA double-strand breaks. *Cell*, 114:673-88

106. Hiemstra PS, Eisenhauer PB, Harwig SS, van den Barselaar MT, van Furth R and Lehrer RI (1993) Antimicrobal proteins of murine macrophages. *Infect Immun*, 61: 3038-46

107. Parseghian MH and Luhrs KA (2006) Beyond the walls of the nucleus: the role of histones in cellular signaling and innate immunity. *Biochem Cell Biol*, 84: 589-604

108. Brix K, Summa W, Lottspeich F and Herzog V (1998) Extracellularly occurring histone H1 mediates the binding of thyroglobulin to the cell surface of mouse macrophages. *J Clin Invest*, 102: 283-93

109. Rose FR, Bailey K, Keyte JW, Chan WC, Greenwood D and Mahida YR (1998) Potential role of epithelial cell-derived histone H1 proteins in innate antimicrobial defense in the human gastrointestinal tract. *Infect Immun*, 66: 3255-63

110. Zhu G, Chen H, Choi BK, Del Piero F and Schifferli DM (2005) Histone H1 proteins act as receptors for the 987P fimbriae of enterotoxigenic Escherichia coli. *J Biol Chem*, 280: 23057-65

111. Tamura M, Natori K, Kobayashi M, Miyamura T and Takeda N (2003) Inhibition of attachment of virions of Norwalk virus to mammalian cells by soluble histone molecules. *Arch Virol*, 148: 1659-70

112. Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, Weinrauch Y and Zychlinsky A (2004) Neutrophil extracellular traps kill bacteria. *Science*, 303: 1532-5

113. Lüders T, Birkemo GA, Nissen-Meyer J, Andersen O and Nes IF (2005) Proline conformation-dependent antimicrobial activity of a proline-rich histone h1 N-terminal peptide fragment isolated from the skin mucus of Atlantic salmon. *Antimicrob Agents Chemother*, 49:2399-406

114. Minota S, Morino N, Sakurai H, Yamada A and Yazaki Y (1993) Interrelationship between autoepitope, DNA-binding domain, and CRP-binding domain on a histone H1 molecule. *Clin Immunol Immunopathol*, 66: 269-71

115. Fan Y, Sirotkin A, Russell RG, Ayala J and Skoultchi AI (2001) Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the $H1_0$ replacement subtype. *Mol Cell Biol*, 21: 7933-43

116. Parseghian MH, Newcomb RL and Hamkalo BA (2001) Distribution of somatic H1 subtypes is non-random on active vs. inactive chromatin II: distribution in human adult fibroblasts. *J Cell Biochem*, 83: 643-59

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

118. Izzo A, Kamieniarz K and Schenider R (2008) The histone H1 family: Specific members, specific functions? *Biol Chem*, 389: 333-43

119. Bednar J, Garcia-Saez I, Boopathi R, Cutter AR, Papai G, Reymer A, Syed SH, Lone IN, Tonchev O, Crucifix C, Meoni H, Papin C, Skoufias DA, Kurumizaka H, Lavery R, Hamiche A, Hayes JJ, Schultz P, Angelov D, Petosa C and Dimitrov S (2017) Structure and dynamics of 197 bp nucleosome in complex with linker histone H1. *Mol Cell*, 66: 384-97

120. Wallace IM, O'Sullivan O, Higgins DG and Notredame C (2006) M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res*, 34: 1692-9

121. Ramakrishnan V (1997) Histone H1 and chromatin higher-order structure. *Crit Rev Eukaryot Gene Expr*, 7: 215-30

122. Ponte I, Romero D, Yero D, Suau P and Roque A (2017) Complex evolutionary history of the mammalian histone H1.1-H1.5 gene family. *Mol Biol Evol*, 34: 545-58

123. Ponte I, Vila R and Suau P (2003) Sequence complexity of histone H1 subtypes. *Mol Biol Evol*, 20: 371-80

124. McBryant Sj and Hansen JC (2012) Dynamic fuzziness during linker histone action. *Adv Exp Med Biol*, 725: 15-26

125. Caterino TL, Fang H and Hayes JJ (2011) Nucleosome linker DNA contacts and induces specific folding of the intrinsically disordered H1 carboxyl-terminal domain. *Mol Cell Biol*, 31: 2341-8

126. Caterino TL and Hayes JJ (2011) Structure of the H1 C-terminal domain and function in chromatin condensation. *Biochem Cell Biol*, 89: 35-44

127. Misteli T, Gunjan A, Hock R, Bustin M and Brown DT (2000) Dynamic binding of histone H1 to chromatin in living cells. *Nature*, 408: 877-81

128. Harshman SW, Young NL, Parthun MR and Freitas MA (2013) H1 histones: current prespectives and challenges. *Nucleic Acids Res*, 41: 9593-9609

129. Balhorn R, Chalkley R and Granner D (1972) Lysine-rich histone phosphorylation. A positive correlation with cell replication. *Biochemistry*, 11: 1094-8

130. Garcia BA, Busby SA, Barber CM, Shabanowitz J, Allis CD and Hunt DF (2004) Characterization of phosphorylation site on histone H1 isoforms by tandem mass spectrometry. *J Proteome Res*, 3: 1219-27

131. Deterding LJ, Bunger MK, Banks GC, Tomer KB and Archer TK (2008) Global changes in and characterization of specific site of phosphorylation in mouse and human H1 isoforms upon CDK inhibitor treatment using mass spectrometry. *J Proteomes Res*, 7: 2368-79

132. Maresca TJ, Freedman BS and Heald R (2005) Histone H1 is essential for mitotic chromosome architecture and segregation in *Xenopus laevis* egg extracts. *J Cell Biol*, 169: 859-69

133. Gurley LR, Valdez JG and Buchanan JS (1995) Characterization of the mitotic specific phosphorylation site of histone H1. Absence of a consensus sequence for the p34cdc2/cyclin B kinase. *J Biol Chem*, 270: 27653-60

134. Th'ng JP, Guo XW, Swank RA, Crissman HA and Bradbury EM (1994) Inhibition of histone phosphorylation by staurosporine leads to chromosome decondensation. *J Biol Chem*, 269:9568-73

135. Alexandrow MG and Hamlin JL (2005) Chromatin decondensation in S-phase involves recruitment of Cdk2 by Cdc45 and histone H1 phosphorylation. *J Cell Biol*, 168: 875-86

136. Telu KH, Abbaoui B, Thomas-Ahner JM, Zynger DL, Clinton SK, Freitas MA and Mortazavi A (2013) Alterations of histone H1 phosphorylation during bladder carcinogenesis. *J Proteome Res*, 12: 3317–26

137. Vincent GP, Nacht AS, Font-Mateu J, Castellano G, Gaveglia L, Ballaré C and Beato M (2011) Four enzymes cooperate to displace histone H1 during the first minute of hormonal gene activation. *Genes Dev*, 25: 845-62

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

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

140. Zheng Y, John S, Pesavento JJ, Schultz-Norton JR, Schiltz RL, Baek S, Nardulli AM, Hager GL, Kelleher NL and Mizzen CA (2010) Histone H1 phosphorylation is associated with transcription by RNA polymerases I and II. *J Cell Biol*, 189: 407-15

141. Kim K, Jeong KW, Kim H, Choi J, Lu W, stallcup MR and An W (2012) Functional interplay between p53 acetylation and H1.2 phosphorylation in p53-regulated transcription. *Oncogene*, 31: 4290-301

142. Roque A, Ponte I, Arrondo JL and Suau P (2008) Phosphorylation of the carboxy-terminal domain of histone H1: effects on secondary structure and DNA condensation. *Nucleic Acids Res*, 36: 4719-26

143. Lopez R, Sarg B, Lindner H, Bartolomé S, Ponte I, Suau P and Roque A (2015) Linker histone partial phosphorylation: effects on secondary structure and chromatin condensation. *Nucleic Acids Res*, 43: 4463-76

144. Bonet-Costa C, Vilaseca M, Diema C, Vujatovic O, Vaquero A, Omeñaca N, Castejón L, Bernués J, Giralt E and Azorín F (2012) Combined bottom-up and top-down mass spectrometry analyses of the pattern of psot-translational modification of *Drosophila melanogaster* linker histone H1. *J Proteomics*, 75: 4124-38

145. Trojer P, Zhang J, Yonezawa M, Schmidt A, Zheng H, Jenuwein T and Reinberg D (2009) Dynamic histone H1 isotype 4 methylation and dimethylation by the histone lysine methyltransferase G9a/KMT1C and the Jumonji domain-containing JMJD2/KDM4 proteins. *J Biol Chem*, 284: 8395-405

146. Weiss T, Hergeth S, Zeissler U, Izzo A, Tropberger P, Zee BM, Dundr M, Garcia BA, Daujat S and Schneider R (2010) Histone H1 variant-specific lysine methylation by G9a/KMT1C and Glp1/KMT1D. *Epigenetics Chromatin*, 3: 7

147. Kuzmichev A, Jenuwein T, Tempst P and Reinberg D (2004) Different EZH2-containing complexes target methylation of histone H1 or nucleosomal H3. *Mol Cell*, 14: 183-93

148. Hergeth SP and Schenider R (2015) The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Rep*, 16: 1439-53

149. Daujat S, Zeissler U, Waldmann T and Happel N (2005) HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation block HP1 binding. *J Biol Chem*, 280: 38090-95

150. Vaquero A, Scher M, Lee D, Erdjument-Bromage H, Tempst P and Reinberg D (2004) Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. *Mol Cell*, 1681: 93-105

151. Wisniewski JR, Zougman A, Krüger A and Mann M (2007) Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. *Mol Cell Proteomics*, 6: 72-8

152. Kamieniarz K, Izzo A, Dundr M, Tropberger P, Ozretic L, Kirfel J, Scheer E, Tropel P, Wisniewski JR, Tora L, Viville S, Buettner R and Schneider R (2012) A dual role of linker histone H1.4 Lys 34 acetylation in transcriptional activation. *Genes Dev*, 26: 797-802

153. Terme JM, Millán-Ariño Ll, Mayor R, Luque N, Izquierdo-Bouldstridge A, Bustillos A, Sampaio C, Canes J, Font I, Sima N, Sancho M, Torrente L, Forcales S, Roque A, Suau P and Jordan A (2014) Dynamics and dispensability of variant-specific histone H1 Lys-26/Ser-27 and Thr-165 post-translational modifications. *FEBS Lett*, 588: 2353-62

154. Christophorou MA, Castelo-Branco G, Halley-Stott RP, Oliveira CS, Loos R, Radzisheuskaya A, Mowen KA, Bertone P, Silva JC, Zernicka-Goetz M, Nielsen ML, Gurdon JB and Kouzarides T (2014) Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature*, 507: 104-8

155. Thorslund T, Ripplinger A, Hoffmann S, Wild T, Uckelmann M, Villumsen B, Narita T, Sixma TK, Choudhary C, Bekker-Jensen S and Mailand N (2015) Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage. *Nature*, 527: 389-93

156. Krishnakumar R, Gamble MJ, Frizzell KM, Berrocal JG, Kininis M and Kraus WL (2008) Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. *Science*, 319: 819-21

157. Krishnakumar R and Kraus WL (2010) PARP-1 regulates chromatin structure and transcription through a KDM5B-dependent pathway. *Mol Cell*, 39: 736-49

158. Wright RH, Castellano G, Bonet J, Le Dily F, Font-Mateu J, Ballare C, Nacht AS, Soronellas D, Oliva B and Beato M (2012) CDK2-dependent activation of PARP-1 is required for hormonal gene regulation in breast cancer cells. *Genes Dev*, 26: 1972-83

159. Shan L, Li X, Liu L, Ding X, Wang Q, Zheng Y, Duan Y, Xuan C, Wang Y, Yang F, Shang Y and Shi L (2014) GATA3 cooperates with PARP1 to regulate CCND1 transcription through modulating histone H1 incorporation. *Oncogene*, 33: 3205-16

160. Kassner I, Barandun M, Fey M, Rosenthal F and Hottiger MO (2013) Crosstalk between SET7/9 dependent methylation and ARTD1-mediated ADP-ribosylation of histone H1.4. *Epigenetics Chromatin*, 6: 1

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

162. Ponte I, Vial-Taboada JM and Suau P (1998) Evolution of the vertebrate H1 histone class: evidence for the functional differentiation of the subtypes. *Mol Biol Evol*, 15: 702-8

163. Eriín-López JM, González-Tizón AM, Martínez A and Méndez J (2004) Birth-and-death evolution with strong purifying selection in the histone H1 multigene family and the origin of orphon H1 genes. *Mol Biol Evol*, 21: 1992-2003

164. Franke K, Drabent B and Doenecke D (1998) Testicular expression of the mouse histone H1.1 gene. *Histochem Cell Biol*, 109: 383-90

165. Rasheed BK, Whisenant EC, Ghai RD, Papaioannou VE and Bhatnagar YM (1989) Biochemical and immunocytochemical analysis of a histone H1 variant from the mouse testis. *J Cell Sci*, 94: 61-71

166. Franke K, Drabent B and Doenecke D (1998) Testicular expression of the mouse histone H1.1 gene. *Histochem Cell Biol*, 109: 383-90

167. Lin Q, Sirotkin A and Skoultchi AI (2000) Normal spermatogenesis in mice lacking the testis-specific linker H1t. *Mol Cell Biol*, 20: 2122-8

168. Martianov I, Brancorsini S, Catena R, Gansmuller A, Kotaja N, Parvinen M, Sassone-Corsi P and Davidson I (2005) Polar nuclear localization of H1T2, a histone H1 variant, required for spermatid elongation and DNA condensation during spermiogenesis. *Proc Natl Acad Sci U S A*, 102: 2808-13

169. Yan W, Ma L, Burns KH and Matzuk MM (2003) HILS1 is a spermatid-specific linker histone H1-like protein implicated in chromatin remodeling during mammalian spermiogenesis. *Proc Natl Acad Sci U S A*, 100: 10546-51

170. Tanaka H, Iguchi N, Isotani A, Kitamura K, Toyama Y, Matsuoka Y, Onishi M, Masai K, Maekawa M, Toshimori K, Okabe M and Nishimune Y (2005) HANP1/H1T2, a novel histone H1-like protein involved in nuclear formation and sperm fertility. *Mol Cell Biol*, 25: 7107-19

171. Iguchi N, Tanaka H, Yomogida K and Nishimune Y (2003) Isolation and characterization of a novel cDNA encoding a DNA-binding protein (Hils1) specifically expressed in testicular haploid germ cells. *Int J Androl*, 26: 354-65

172. Clarke HJ, McLay DW and Mohamed OA (1998) Linker histone transitions during mammalian oogenesis and embryogenesis. *Dev Genet*. 22: 17-30

173. Clarke HJ, Bustin M and Oblin C (1997) Chromatin modifications during oogenesis in the mouse: removal of somatic subtypes of histone H1 from oocyte chromatin occurs post-natally through a post- transcriptional mechanism. *J Cell Sci*, 110: 477-87

174. Fu G, Ghadam P, Sirotkin A, Khochbin S, Skoultchi AI and Clarke HJ (2003) Mouse oocytes and early embryos express multiple histone H1 subtypes. *Biol Reprod*, 68: 1569-76

175. Pan C and Fan Y (2015) Role of H1 linker histones in mammalian development and stem cell differentiation. *Biochim Biophys Acta*, 1859: 496-509

176. Helliger W, Linder H, Grubl-Knosp O and Puschendorf B (1992) Alternations in proportions of histone H variants during differentiation of murine erythroleukaemic cells. *Biochem J*, 288: 747-51

177. Zalatanova J and Doenecke D (1994) Histone H1 zero: a major player in cell differentiation? *FASEB J*, 8: 1260-8

178. Terme JM, Sesé B, Millán-Ariño Ll, Mayor R, Izpisúa Belmonte JC, Barrero MJ and Jordan A (2011) Histone H1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency. *J Biol Chem*, 286: 35347-57

179. Morales C, Biran A, Burney MJ, Patel H, Henser-Brownhill T, Cohen AHS, Li Y, Ben-Hamo R, Nye E, Spencer-Dene B, Chakravarty P, Efroni S, Matthews N, Misteli T, Meshorer E and Scaffidi P (2017) The linker histone H1.0 generates epigenetic and functional intratumor heterogeneity. *Science*, 353: aaf1644

180. Scaffidi P (2016) Histone H1 alternations in cancer. *Biochim Biophys Acta*, 1859: 533-9

181. Hechtman JF, Beasley MB, Kinoshita Y, Ko HM, Hao K and Burstein DE (2013) Promyelocytic leukaemia zinc-finger and histone H1.5 differentially stain low- and high-grade

pulmonary neuroendocrine tumors: a pilot immunohistochemical study. *Hum Pathol*, 44: 1400-5

182. Khachaturov V, Xiao GQ, Kinoshita Y, Unger PD and Burstein DE (2014) Histone H1.5, a novel prostatic cancer marker: an immunohistochemical study. *Hum Pathol*, 45: 2115-9

183. Medrzycki M, Zhang Y, McDonald JF and Fan Y (2012) Profiling of linker histone variants in ovarian cancer. *Front Biosci (Landmark Ed)*, 17: 396-406

184. Lennox RW and Cohen LH (1983) The histone H1 complements of dividing and nondividing cells of the mouse. *J Biol Chem*, 258: 262-8

185. Meergans T, Albig W and Doenecke D (1997) Varied expression patterns of human H1 histone genes in different cell lines. DNA Cell Biol, 16: 1041-9

186. Piña B and Suau P (1987) Changes in the proportions of histone H1 subtypes in brain cortical neurons. *FEBS Lett*, 210: 161-4

187. Parseghian MH and Hamkalo BA (2001) A compendium of the histone H1 family of somatic subtypes: an elusive cast of characters and their characteristics. *Biochem Cell Biol*, 79: 289-304

188. Domínguez V, Piña B and Suau P (1992) Histone H1 subtype synthesis in neurons and neuroblasts. *Development*, 115: 181-5

189. Winter E, Palatnik CM, Williams DL, Coles LS, Wells JR and Gordon JS (1985) Changes in the H-1 histone complement during myogenesis. II. Regulation by differential coupling of H-1 variant mRNA accumulation to DNA replication. *J Cell Biol*, 101: 175-81

190. Stoldt S, Wenzel D, Schulze E, Doenecke D and Happel N (2007) G1 phase-dependent nucleolar accumulation of human histone H1x. *Biol Cell*, 99: 541-52

191. Warneboldt J, Haller F, Horstmann O, Danner BC, Füzesi L, Doenecke D and Happel N (2008) Histone H1x is highly expressed in human neuroendocrine cells and tumours. *BMC Cancer*, 8: 388

192. Sepsa A, Levidou G, Gargalionis A, Adamopoulos C, Spyropoulou A, Dalagiorgou G, Thymara I, Boviatsis E, Themistocleous MS, Petraki K, Vrettakos G, Samaras V, Zisakis A, Patsouris E, Piperi C and Korkolopoulou P (2015) Emerging role of linker histone variant H1x as a biomarker with prognostic value in astrocytic gliomas. A multivariate analysis including trimethylation of H3K9 and H4K20. *PLoS One*, 10: e0115101

193. Shahhoseini M, Favaedi R, Baharvand H, Sharma V and Stunnenberg HG (2010) Evidence for a dynamic role of the linker histone H1x during retinoic acid-induced differentiation of NT2 cells. *FEBS Lett*, 584: 4661-4

194. Vyas P and Brown DT (2012) N- and C- terminal domains determine differential nucleosomal binding geometry and affinity of linker histone isotypes H1(0) and H1c. *J Biol Chem*, 287: 11778-87

195. Clausell J, Happel N, Hale TK, Doenecke D and Beato M (2009) Histone H1 subtypes differentially modulate chromatin condensation without preventing ATP-dependent remodelling by SWI/SNF or NURF. *PLoS One*, 4: e0007243

196. Orrego M, Ponte I, Roque A, Buschati N, Mora X and Suau P (2007) Differential affinity of mammalian histone H1 somatic subtypes for DNA and chromatin. *BMC Biol*, 5: 22

197. Hergeth SP, Dundr M, Tropberger P, Zee BM, Garcia BA, Daujat S and Schneider R (2011) Isoform-specific phosphorylation of human linker histone H1.4 in mitosis by the kinase Aurora B. *J Cell Sci*, 124: 1623-8

198. Kamieniarz K, Izzo A, Dundr M, Tropberger P, Ozretic L, Kirfel J, Scheer E, Tropel P, Wisniewski JR, Tora L, Viville S, Buettner R and Schneider R (2012) A dual role of linker histone H1.4 Lys 34 acetylation in transcriptional activation. *Genes Dev*, 26: 797-802

199. Contreras A, Hale TK, Stenoien DL, Rosen JM, Mancini MA and Herrera RE (2003) The dynamic mobility of histone H1 is regulated by cyclin/CDK phosphorylation. *Mol Cell Biol*, 23: 8626-36

200. Raghuram N, Strickfaden H, McDonald D, Williams K, Fang H, Mizzen C, Hayes JJ, Th'ng J and Hendzel MJ (2013) Pin1 promotes histone H1 dephosphorylation and stabilizes its binding to chromatin. *J Cell Biol*, 203: 57-71

201. Munro S, Hookway ES, Floderer M, Carr SM, Konietzny R, Kessler BM, Oppermann U and La Thangue NB (2017) Linker histone H1.2 directs genome-wide chromatin association of retinoblastoma tumor suppressor protein and facilitates its function. *Cell Rep*, 19: 2193-201

202. Lee H, Habas R and Abate-Shen C (2004) MSX1 cooperates with histone H1b for inhibition of transcription and myogenesis. *Science*, 304: 1675-8

203. Mackey-Cushman SL, Gao J, Holmes DA, NunoyaJ, Wang R, Unutmaz D and Su L (2011) FoxP3 interacts with linker histone H1.5 to modulate gene expression and program Treg cell activity. *Genes Immun*, 12: 559-67

204. Kalashnikova AA, Winkler DD, McBryant SJ, Henderson RK, Herman JA, DeLuca JG, Luger K, Prenni JE and Hansen JC (2013) Linker histone H1.0 interacts with an extensive network of proteins found in the nucleolus. *Nucleic Acids Res*, 41: 4026-35

205. Szerlong HJ, Herman JA, Krause CM, DeLuca JG, Skoultchi A, winger QA, Prenni JE and Hansen JC (2015) Proteomic characterization of the nucleolar linker histone H1 interaction network. *J Mol Biol*, 427: 2056-71

206. Kim K, Choi J, Heo K, Kim H, Levens D, Kohno K, Johnson EM, Brock HW and An W (2008) Isolation and characterization of a novel H1.2 complex that acts as a repressor of p53-mediated transcription. *J Biol Chem*, 283: 9113-26

207. Kim JM, Kim K, Puni V, Liang G, Ulmer TS, Lu W and An W (2015) Linker histone H1.2 establishes chromatin compaction and gene silencing through recognition of H3K27me3. *Sci Rep*, 5: 16714

208. Kim K, Lee B, Kim J, Choi J, Kim JM, Xiong Y, Roeder R and An W (2013) Linker histone H1.2 cooperates with Cul4A and PAF1 to drive H4K31 ubiquitylation-mediated transactivation. *Cell Rep*, 5: 1690-1703

209. Bhan S, Way W, Warren SL and Sittman DB (2008) Global gene expression analysis reveals specific and redundant role for H1 variants, H1c and H1(0). *Gene*, 414: 10-18

210. Yang SM, Kim BJ, Toro LN and Skoultchi AI (2013) H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation. *Proc Natl Acad Sci U S A*, 110: 1708-13

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

212. Parseghian MH, Clark RF, Hauser LJ, Dvorkin N, Harris DA and Hamkalo BA (1993) Fractionation of human H1 subtypes and characterization of subtype-specific antibody exhibiting non-uniform nuclear staining. *Chromosome Res*, 1: 127-39

213. Parseghian MH, Harris DA, Rishwan DR and Hamkalo BA (1994) Characterization of a set of antibodies specific for three human histone H1 subtypes. *Chromosoma*, 103: 198-208

214. Th'ng JP, Sung R, Ye M and Hendzel MJ (2005) H1 family histones in the nucleus. Control of the binding and localization by the C-terminal domain. *J Biol Chem*, 280: 27809-14

215. Li JY, Patterson M, Mikkola HKA, Lowry WE and Kurdistani SK (2012) Dynamic distribution of linker histone H1.5 in cellular differentiation. *PLoS Genet*, 8: e1002879

216. Cao K, Lailler N, Zhang Y, Kumar A, Uppal K, Liu Z, Lee EK, Wu H, Medrzycki M, Pan C, Ho PY, Cooper GP, Dong JX, Bock C, Bouhassira EE and Fan Y (2013) High-resolution mapping of H1 linker histone variants in embryonic stem cells. *PLoS Genet*, 9: e1003417

217. Izzo A, Kamieniarz-Gdula K, Ramírez F, Noureen N, Kind J, Manke T van Steensel B and Schenider R (2013) The genomic landscape of the somatic linker histone subtypes H1.1 to H1.5 in human cells. *Cell Rep*, 3: 2142-54

218. Millán-Ariño Ll, Islam AB, Izquierdo-Bouldstridge A, Mayor R, Terme JM. Luque N, Sancho M, López-Bigas N and Jordan A (2014) Mapping of six somatic linker histone H1 variants in human breast cancer cells uncovers specific features of H1.2. *Nucleic Acids Res*, 42: 4474-93

219. Trollope AF, Sapojnikova N, Throne AW, Crane-Robinson C and Myers FA (2010) Linker histone subtypes are not generalized gene repressors. *Biochim Biophys Acta*, 1799: 642-52

220. Schlee M and Hartmann G (2016) Discriminating self from non-self in nucleic acid sensing. *Nat Rev Immunol*, 16: 566-80

221. Unterholzner L (2013) The interferon response to intracellular DNA: why so many receptors? *Immunobiology*, 218: 1312-21

222. Reikine S, Nguyen JB and Modis Y (2014) Pattern recognition and signalling mechanisms of RIG-I and MDA5. *Front Immunol*, 5: 342

223. Kato H, Takeuchi O, Sato S, Yoneyama M, Yamamoto M, Matsui K, Uematsu S, Jung A, Kawai T, Ishii KJ, Yamaguchi O, Otsu K, Tsujimura T, Koh CS, Reis e Sousa C, Matsuura Y, Fujita T and Akira S (2006) Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature*, 441: 101-5

224. Pichlmair A, Schulz O, Tan CP, Näslund TI, Liljeström P, Weber F, Reis e Sousa C (2006) RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science*, 314:997-1001

225. Kawai T, Takahashi K, Sato S, Coban C, Kumar H, Kato H, Ishii KJ, Takeuchi O and Akira S (2005) IPS-1, an adaptor triggering RIG-I and MDA5-mediated type I interferon induction. *Nat Immunol*, 6: 981-8

226. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, 409: 860-921

227. Kassiotis G (2014) Endogenous retroviruses and the development of cancer. *J Immunol*, 192: 1343-9

228. Hurst TP and Magiorkinis G (2015) Activation of the innate immune response by endogenous retroviruses. *J Gen Virol*, 96:1207-18

229. Colmegna I and Garry RF (2006) Role of endogenous retroviruses in autoimmune diseases. *Infect Dis Clin Nort Am*, 20: 913-29

230. Yu P (2016) The potential role of retroviruses in autoimmunity. *Immunol Rev*, 269: 85-99

231. Groh S and Schotta G (2017) Silencing of endogenous retroviruses by heterochromatin. *Cell Mol Life Sci*, 74:2055-65

232. Nishibuchi G and Déjardin J (2017) The molecular basis of the organization of repetitive DNA-containing constitutive heterochromatin in mammals. *Chromosome Res*, 25: 77-87

233. Lee SA, Kwak MS, Kim S and Shin JS (2014) The role of high mobility group box 1 in innate immunity. *Yonsei Med J*, 55: 1165-76

234. Yanai H, Ban T and Taniguchi T (2012) High-mobility group box family of proteins ligand and sensor for innate immunity. *Trends Immunol*, 33: 633-40

235. Yanai H, Ban T, Wang Z, Choi MK, Kawamura T, Megishi H, Nakasato M, Lu Y, Hangai S, Koshiba R, Savitsky D, Ronfani L, Akira S, Bianchi ME, Honda K, Tamura T, Kodama T and Taniguchi T (2009) HMGB proteins function as universal sentinels for nucleic-acid-mediated innate immune responses. *Nature*, 462: 99-103

236. Mankan AK, Schmidt T, Chauhan D, Goldeck M, Höning K, Gaidt M, Kubarenko AV, Andreeva L, Hopfner KP and Homung V (2014) Cytosolic RNA:DNA hybrids activate the cGAS-STING axis. *EMBO J*, 33: 2937-46

237. Sollier J and Cimprich KA (2015) Breaking bad: R-loops and genome intergrity. *Trends Cell Biol*, 25: 514-22

238. Aguilera A and García-Muse T (2012) R loops: from transcription by-products to threats to genome stability. *Mol Cell*, 46: 115-24

239. Lim YW, Sanz LA, Xu X, Hartono SR and Chédin F (2015) Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutières syndrome. *eLife*, 16: 4

240. Leonova K, Brodsky L, Lipchick B, Pal M, Novototskava L, Chenchik AA, Sen GC, Komarova EA and Gudkov AV (2013) p53 cooperates with DNA methylation and a suicidal interferon

response to maintain epigenetic silencing of repeats and noncoding RNAs. *Proc Natl Acad Sci U S A*, 110: E89-98

241. Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, Hein A, Rote NS, Cope LM, Snyder A, Makarov V, Budhu S, Slamon DJ, Wolchok JD, Pardoll DM, Beckmann MW, Zahnow CA, Merghoub T, Chan TA, Baylin SB and Strick R (2015) Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell*, 162: 974-86

242. Roulois D, Yau HL, Singhania R, Wang Y, Danesh A, Shen SY, Han H, Liang G, Jones PA, Pugh TJ, O'Brien C and de Carvalho DD (2015) DNA – demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous Ttranscripts. *Cell*, 162:961-73

243. Lenova Y, Safina A, Nesher E, Sandlesh P, Pratt R, Burkhart C, Lipchick B, Frangou C, Koman I, Wang J, Kirsanov K, Yakubovskaya M and Gudkov AV (2017) Transcription of repeats activates interferon (TRAIN) in response to chromatin destabilization induced with anti-cancer small molecule. Preprint, https://doi.org/10.1101/142471

244. Nakad R and Schumacher B (2016) DNA damage response and immune defense: links and mechanisms. *Front Genet*, 7: 147

245. Kuo LJ and Yang LX (2008) GammaH2AX – a novel biomarker for DNA double-strand breaks. *In Vivo*, 22: 305-9

246. Tallis M, Morra R, Brakauskaite E and Ahel I (2014) Poly(ADP-ribosyl)ation in regulation of chromatin structure and the DNA damage response. *Chromosoma*, 123: 79-90

247. Crisciona SW, Zhang Y, Thompson W, Sedivy JM and Neretti (2014) Transcriptional landscape of repetitive elements in normal and cancer human cells. BMC Genomics, 15: 583

248. Pal S and Tyler JK (2016) Epigenetics and aging. *Sci Adv*, 2: e1600584

249. Fuertes MB, Woo SR, Burnett B, Fu YX and and Gajewski TF (2014) Type I IFN response and innate immune sensing of cancer. *Trends Immunol*, 34: 67-73

250. He L, Chen Y, Wu Y, Xu Y, Zhang Z and Liu Z (2017) Nucleic acid sensing pattern recognition receptors in the development of colorectal cancer and colitis. *Cell Mol Life Sci*, 74: 2395-411

251. Fadloun A, Le Gras S, Jost B, Ziegler-Birling C, Takahasi H, Gorab E, Carninci P and Torres-Padilla ME (2013) Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat Struct Mol Biol*, 20: 332-8

252. Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holdbrook AE, Solter D and Knowles BB (2004) Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell*, 7: 597-606

253. Gomez NC, Hepperla A, Raluca D, Simon JM, Fang F and Davis IJ (2016) Widespread chromatin accessibility at repetitive elements links stem cells with human cancer. *Cell Rep*, 17: 1607-20

254. Chuong E, Elde NC and Feschotte C (2016) Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science*, 351: 1083-7

255. Gorka C, Fakan S and Lawrence JJ (1993) Light and electron microscope immunocytochemical analyses of histone H1(0) distribution in the nucleus of Friend erythroleukemia cells. *Exp Cell Res*, 205: 152-8

256. Jarboui MA, Wynne K, Elia G, Hall WW and Gautier VW (2011) Proteomic profiling of the human T-cell nucleolus. *Mol Immunol*, 49: 441-52

257. Zheng Y, John S, Pesavento JJ, Schultz-Norton JR, Schiltz RL, Baek S, Nardulli AM, Hager GL, Kelleher NL and Mizzen CA (2010) Histone H1 phosphorylation is associated with transcription by RNA polymerases I and II. *J Cell Biol*, 189: 407-15

258. Stoldt S, Wenzel D, Schulze E, Doenecke D and Happel N (2007) G1 phase-dependent nucleolar accumulation of human histone H1x. *Biol Cell*, 99: 541-52

259. Takata H, Matsunaga S, Morimoto A, Ono-Maniwa R, Uchiyama S and Fukui K (2007) H1.X with different properties from other linker histones is required for mitotic progression. *FEBS Lett*, 581: 3783-8

260. Schwartz S, Meshorer E and Ast G (2009) Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol*, 16: 990-5

261. Ruike Y, Imanaka Y, Sato F, Shimizu K and Tsujimoto G (2010) Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. *BMC Genomics*, 11: 137

262. Höld M and Basler K (2012) Transcription in the absence of histone H3.2 and H3K4 methylation. *Curr Biol*, 22: 2253-7

263. Zhang H, Gao L, Anandhakumar J and Gross DS (2014) Uncoupling transcription from covalent histone modification. *PLoS Genet*, 10: e1004202

264. Wijchers PJ, Geeven G, Eyres M, Bergsma AJ, Janssen M, Verstegen M, Zhu Y, Schell Y, Vermeulen C, de Wit E and de Laat W (2015) Characterization and dynamics of pericentromere-associated domains in mice. *Genomes Res*, 25: 958-969

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

266. Monsurrò V, Beghelli S, Wang R, Barbi S, Coin S, Di Pasquale G, Bersani S, Castellucci M, Sorio C, Eleuteri S, Worschech A, Chiorini JA, Pederzoli P, Alter H, Marincola FM and Scarpa A (2010) Anti-viral state segregates two molecular phenotypes of pancreatic adenocarcinoma: potential relevance for adenoviral gene therapy. *J Transl Med*, 8: 10

267. Pei H, Li L, Fridley BL, Jenkins GD, Kalari KR, Lingle W, Petersen G, Lou Z and Wang L (2009) FKBP51 affects cancer cell response to chemotherapy negatively regulating Akt. *Cancer Cell*, 16: 259-66

268. Tani R, Hayakawa K, Tanaka S and Shiota K (2016) Linker histone variant H1T targets rDNA repeats. *Epigenetics*, 11: 288-302

269. Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, Martin L, Ware CB, Blish CA, Chang HY, Pera RA and Wysocka J (2015) Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature*, 522: 221-5

270. Németh A and Längst G (2011) Genome organization in and around the nucleolus. *Trends Genet*, 27: 149-56

271. Ayarpadikannan and Kim HS (2014) The impact of transposable elements in the genome evolution and genetic instability and their implications in various diseases. *Genomics Inform*, 12: 98-104

272. Ostertag EM, Goodier JL, Zhang Y and Kazazian Jr HH (2003) SVA elements are nonautonomus retrotransposons that cause disease in humans. *Am J Hum Genet*, 73: 1444-51

273. Elsässer SJ, Noh KM, Diaz N, Allis CD and Banasynski LA (2015) Histone H3.3 is required for endogenous retroviral elements silencing in embryonic stem cells. *Nature*, 522: 240-4

274. Scott MS and Ono M (2011) From snoRNA to miRNA: Dual function regulatory non-coding RNAs. *Biochimie*, 93: 1987-92

275. Mattick JS (2003) Challenging the dogma: the hidden layer of non-protein-codin RNAs in complex organisms. *Bioessays*, 25: 930-9

276. Criscione SW, Zhang Y, Thompson W, Sedivy JM and Neretti N (2014) Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics*, 15:583

APPENDIX I

4474–4493 *Nucleic Acids Research, 2014, Vol. 42, No. 7* *Published online 29 January 2014*
doi:10.1093/nar/gku079

Mapping of six somatic linker histone H1 variants in human breast cancer cells uncovers specific features of H1.2

Lluís Millán-Ariño[1], Abul B. M. M. K. Islam[2,3], Andrea Izquierdo-Bouldstridge[1], Regina Mayor[1], Jean-Michel Terme[1], Neus Luque[1], Mónica Sancho[4], Núria López-Bigas[2,5] and Albert Jordan[1,*]

[1]Department of Molecular Genomics, Institut de Biologia Molecular de Barcelona (IBMB-CSIC), Barcelona, E-08028 Spain, [2]Research Programme on Biomedical Informatics, Universitat Pompeu Fabra, Barcelona, E-08003 Spain, [3]Department of Genetic Engineering, Biotechnology, University of Dhaka, Dhaka-1000, Bangladesh, [4]Centro de Investigación Príncipe Felipe, Valencia, E-46012 Spain and [5]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, E-08010 Spain

Received October 31, 2013; Revised December 20, 2013; Accepted December 30, 2013

ABSTRACT

Seven linker histone H1 variants are present in human somatic cells with distinct prevalence across cell types. Despite being key structural components of chromatin, it is not known whether the different variants have specific roles in the regulation of nuclear processes or are differentially distributed throughout the genome. Using variant-specific antibodies to H1 and hemagglutinin (HA)-tagged recombinant H1 variants expressed in breast cancer cells, we have investigated the distribution of six H1 variants in promoters and genome-wide. H1 is depleted at promoters depending on its transcriptional status and differs between variants. Notably, H1.2 is less abundant than other variants at the transcription start sites of inactive genes, and promoters enriched in H1.2 are different from those enriched in other variants and tend to be repressed. Additionally, H1.2 is enriched at chromosomal domains characterized by low guanine–cytosine (GC) content and is associated with lamina-associated domains. Meanwhile, other variants are associated with higher GC content, CpG islands and gene-rich domains. For instance, H1.0 and H1X are enriched at gene-rich chromosomes, whereas H1.2 is depleted. In short, histone H1 is not uniformly distributed along the genome and there are differences between variants, H1.2 being the one showing the most specific pattern and strongest correlation with low gene expression.

INTRODUCTION

Eukaryotic DNA is packaged into chromatin through its association with histone proteins. The fundamental repeat unit of chromatin is the nucleosome, which consists of 146 bp of DNA wrapped around an octamer of core histone proteins H2A, H2B, H3 and H4. Linker histone H1 sits at the base of the nucleosome near the entry and exit sites and is involved in the folding and stabilization of the 30-nm chromatin fiber, allowing a higher degree of DNA compaction (1–4). Histone H1 is a family of lysine-rich proteins that consists of three domains: a short basic N-terminal tail, a highly conserved central globular domain and a long positively charged C-terminal tail. Like in core histones, these tails are posttranslationally modified, mainly by phosphorylation, but also by acetylation, methylation, ubiquitination and formylation (5–10). Due to its role in the formation of higher-order chromatin structures, H1 has classically been seen as a structural component related to chromatin compaction and inaccessibility to transcription factors, RNA polymerase and chromatin remodeling enzymes (11,12). However, in recent years, the view that H1 plays a more dynamic and gene-specific role in regulating gene expression is gaining strength. Knock-out or knock-down studies in several organisms have revealed that only a few genes change in expression on complete depletion of H1, some being up- and some downregulated (13–22).

Unlike core histones, the H1 histone family is more evolutionary diverse and many organisms have multiple H1 variants or subtypes, making the study of these proteins more complex. In humans, the histone H1 family includes 11 different H1 variants with 7 somatic subtypes (H1.1 to H1.5, H1.0 and H1X), three testis-specific variants (H1t,

*To whom correspondence should be addressed. Tel: +34 93 402 0487; Fax: +34 93 403 4979; Email: albert.jordan@ibmb.csic.es

H1T2 and HILS1) and one oocyte-specific variant (H1oo). Among the somatic histone H1 variants, H1.1 to H1.5 are expressed in a replication-dependent manner, whereas H1.0 and H1X are replication-independent. H1.2 to H1.5 and H1X are ubiquitously expressed, H1.1 is restricted to certain tissues, and H1.0 accumulates in terminally differentiated cells (23).

It is still far from clear why there are so many H1 variants and great efforts have been made recently to elucidate whether they play specific roles or have redundant functions. Single or double H1 variant knock-out studies in mice did not identify any specific phenotype and this was attributed to the compensatory upregulation of other subtypes, favoring the view that there is redundancy between H1 variants (18). Despite these observations, there is growing evidence supporting the view that histone H1 variants do have specific functions. H1 subtypes present cell type and tissue-specific expression patterns and their expression is regulated over the course of differentiation and development (24–31). Different H1 subtypes have also been differentially related with cancer processes (32–35). Chromatin binding affinity and residence time vary between H1 subtypes owing to differences mainly in the C-t tail, but also in the N-t tail (36–44). Furthermore, H1 subtypes are differently posttranslationally modified and these modifications modulate their interaction with different partners. This could explain some reported specific functions for certain H1 variants (45–57). Finally, global gene expression analyses in various cell types reveal that histone H1 variants control the expression of different subsets of genes, pointing to a specific role of H1 variants in gene regulation (58,59).

To fully understand the function of histone H1 and its variants, several groups have explored the genomic distribution of H1 *in vivo*. Initial biochemical and microscopy-based approaches suggested a nonuniform distribution of H1 in the cell nucleus and found differences between variants (44,60,61). However, due to the lack of specific ChIP-grade antibodies for most of the H1 variants, it has been challenging to identify the precise mapping of H1 variants in the genome. Genome-wide studies with histone H1 started with ChIP-chip experiments in MCF7 cells using an antibody for total H1 (62) and continued using DamID technique for the unique *Drosophila* histone H1 (63). Recently, some groups succeeded in obtaining the first genome maps for H1 variants. The genome-wide distribution of human H1.5 in IMR90 fibroblasts reveals that there are zones of enrichment in genic and intergenic regions of differentiated human cells, but not in embryonic stem cells, associated with gene repression and chromatin compaction (64). Furthermore, analysis of tagged H1c and H1d variants in knock-in mouse embryonic stem cells (ESCs) by ChIP-seq shows depletion of these variants from guanine–cytosine (GC)- and gene-rich regions and active promoters, and positive and negative correlations with H3K9me3 and H3K4me3, respectively, as well as an overrepresentation in major satellites (65). Finally, using DamID technology, the genomic mapping of human H1.1 to H1.5 variants was also achieved in IMR90 cells (66). While H1.2 to H1.5 showed, in general, similar distributions and were depleted from CpG-dense and regulatory regions, H1.1 showed a district profile, pointing to a specific role of this variant in chromatin function.

In this study, we investigated the distribution of the different H1 somatic variants in breast cancer cells by chromatin immunoprecipitation (ChIP) combined with quantitative polymerase chain reaction (qPCR), tiling promoter arrays and high-resolution sequencing. We combined the use of specific antibodies for some variants and hemagglutinin (HA)-tagged recombinant H1 variants expressed in cell lines to study the genome-wide distribution of H1.0, H1.2 to H1.5 and also H1X, a more recently identified and distantly related H1 variant. H1.1 was omitted from our analysis, it being the only somatic H1 variant not present in many cell types, including the cells used here. We also compared H1 distribution with the nucleosome distribution in our T47D human breast cancer cell lines, by H3 immunoprecipitation. Our data support the view that all H1 variants occur across the genome, but also uncover specific features for H1.2, both at promoters and genome-wide. Interestingly, H1.2 enrichment correlates the most closely with gene repression, structural domains of chromatin such as lamina-associated domains (LADs) and regions of low GC content. Overall, the distribution of H1.2 along chromosomes differs from that of other variants including H1.0 and H1X, the two variants most structurally distant within the somatic H1 family. This work represents a comprehensive attempt to investigate for the first time the occurrence and relevance of the different histone H1 variants in the genome of human cancer cells, and provides valuable data to clarify our understanding of the functionalities and heterogeneity of H1.

MATERIALS AND METHODS

Cell lines and culturing conditions

Breast cancer T47D-MTVL cells (carrying one stably integrated copy of luciferase reporter gene driven by the MMTV promoter), or derivative cells stably expressing HA-tagged H1 variants (H1-HA), were grown at 37°C with 5% CO_2 in RPMI 1640 medium, supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin and 100 μg/ml streptomycin, as described previously (59). HeLa cell line was grown at 37°C with 5% CO_2 in Dulbecco's modified Eagle's medium GlutaMax medium containing 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. MCF7 cell line was grown at 37°C with 5% CO_2 in Minimum Essential Medium (MEM) medium containing 10% FBS, 1% penicillin/streptomycin, 1% nonessential amino acids, 1% sodium pyruvate and 1% glutamine.

For Phorbol myristate acetate (PMA) experiments, serum-containing Roswell Park Memorial Institute (RPMI) 1640 media was replaced by serum-free media. After 24 h under serum-free conditions, cells were treated with PMA (100 nM) for the indicated time at 37°C.

Stable expression of HA-Tagged H1 variants

Generation of T47D-MTVL stably expressing HA-Tagged H1 variants was achieved as described previously (59). Briefly, human histone H1 variants were PCR-amplified from genomic DNA and cloned into pCDNA4-HA vector provided by D. Reinberg's group (NYU Medical School). The complete H1-HA cassette was cloned into the lentiviral expression vector pEV833.GFP provided by E. Verdin (Gladstone Institute) upstream an internal ribosome entry site (IRES)-GFP cassette. Viruses were then produced and cells were infected with pEV833-derived lentivirus. HA-tagged H1 variants-expressing cell lines were selected by sorting in a FACSvantageSE or FACS caliber machine (Becton Dickinson) for green fluorescent protein (GFP)-positive fluorescence.

ChIP assays

ChIP assays were performed as described previously (67). Briefly, cells were fixed using 1% formaldehyde, harvested and sonicated using a Diagenode Bioruptor to generate chromatin fragments between 200 and 500 bp. To perform the ChIP, 30 µg of chromatin was immunoprecipitated overnight using the indicated antibody. Rabbit IgG (Santa Cruz Biothechnology) was used as a control for nonspecific interaction of DNA. Input was prepared with 10% of the chromatin material used for an immunoprecipitation. Immunocomplexes were recovered using 20 µl of Protein-A magnetic beads from Millipore. Beads with bound antibody/protein/DNA complexes were washed, decross-linked at 65°C overnight and immunoprecipitated DNA was recovered using the IPure Kit from Diagenode.

The following antibodies were used in this study: anti-H1.2 (Abcam 4086), anti-H1X (Abcam 31972), anit-H3 (Abcam 1791) and anti-HA tag (Abcam 9110).

ChIP-qPCR

Real-time PCR was performed on ChIP and input DNA using EXPRESS SYBR GreenER qPCR SuperMix Universal (Invitrogen) and specific oligonucleotides in a Roche 480 Lightcycler. ChIP values were corrected by the correspondent input chromatin sample. All oligonucleotide sequences used for the amplifications are available on request.

ChIP-chip assays with Nimblegen promoter array

At least 10 ng of ChIP and input DNA was amplified using GenomePlex Complete Whole Genome Amplification Kit (Sigma) and eluted with GenElute PCR Clean-Up Kit (Sigma). For ChIP-on-chip experiments we used Nimblgen HG18 Refseq Promoter 3x720K array. One microgram of ChIP and input DNA was directly labeled by Klenow random priming with Cy5 and Cy3 nonamers with Nimblegen Dual-color DNA Labeling Kit following manufacturer's user's guide Chip-chip arrays v6.2, and the labeled DNA was precipitated with 1 volume isopropanol. Hybridization mix including 15 µg of labeled DNA was prepared using Nimblegen Hybridization Kit. Arrays were hybridized in Nimblegen Hybridization System 4 Station for 16–18 h at 42°C, and then washed in 1× Wash solution I, II and III. Hybridization buffers and washes were completed using manufacturer's protocols. Arrays were scanned on a Nimblegen MS 200 Scanner per manufacturer's protocol.

ChIP-on-chip raw data was normalized and differential intensity of each probe compared with input control was calculated using the Nimblegen software DEVA. Average fold change (ChIP versus input) each 50 bp bin for a range of −3.2 kb upstream and 800 bp downstream window from RefSeq transcription start sites (TSS) were calculated using in-house Perl script. LOESS smoothed line plot around the TSS were plotted using in-house script written in R statistical programming language. For ChIP-signal heat map, similarly fold change average for each individual RefSeq transcript was calculated and then data were visualized with Java Treeview (68). Functional annotation of target genes based on Gene Ontology was performed using DAVID Software (Database for Annotation, Visualization and Integrated Discovery).

ChIP-seq

Library preparation for sequencing: ChIP and genomic library preparation was performed using standard Illumina protocols. Libraries were prepared with the ChIP-seq Sample Preparation Kit (Illumina) according to the manufacturer's instructions. Briefly, 10 ng of ChIP and input DNA were repaired to overhang a 3′-dA and then adapters were ligated to the end of DNA fragments. DNA fragments with proper size (usually 100–300 bp, including adaptor sequence) were selected after PCR amplification, obtaining qualified library for sequencing.

Sequencing, mapping and peak detection: Sequencing was performed with Illumina HiSeq 2000 system. Raw sequence reads containing >10% of 'N', or bases with $Q \leq 20$ account for >50% of the total were removed and adaptor sequences were trimmed. Identified clean reads were uniquely aligned allowing at best two mismatches to the UCSC (The Genome Sequencing Consortium) reference genome (human hg18) using the program SOAP (version 2.21) (69). Sequences matching exactly more than one place with equal quality were discarded to avoid bias. Read length and read counts of each library are listed in Supplementary Table S1. Peak caller program for histone, SICER (version 1.1) (70), was used with following parameters: redundancy threshold = 1, window size = 200, fragment size = 150, effective genome fraction = 0.75, gap size = 200, false discovery rate (FDR) = 0.01 and Fold Change at least 2. Input subtracted normalized (total mapped library size) WIG files were produced from duplicate removed aligned reads using the program javaGenomicsToolKit.

Binding sites to gene feature annotation: Enriched peaks were annotated to nearest gene (RefSeq genes) using Bioconductor package ChIPpeakAnno (71). Distribution of enriched and depleted regions (peaks) to various genomic features, and continuous ChIP signal profile distribution of reads along the meta-gene were performed using software CEAS (72) and in-house Python and Perl scripts.

Regulatory regions, histone modification peaks, CpG and LADs abundance: Input-subtracted normalized average H1 variants read density in each enriched locations of regulatory regions, histone modification peaks, CpG and LADs were calculated, and representation in box-plot were made using in-house scripts. As a control, a random sample of genomic windows with equal width was used to perform the significance test (Kolmogorov–Smirnov test).

Publicly available genome-wide location data analysis: Public ChIP-seq data, which includes H3K4me1, H3K4me2, H3K4me3, H3K27me3, H3K27ac, H3K9me3, H3K9ac, H3K36me3, P300, CTCF, FAIRE and DNase enriched genomic locations, are taken from ENCODE project. CpG island genomic location information (hg18) and the coordinates of LADs (73) were taken from UCSC database. Publicly available whole-genome data if not available on hg18 version, they were first remapped to the human genome version hg18 using the UCSC coordinate conversion tool (http://genome.ucsc.edu/cgi-bin/hgLiftOver).

Overlap analysis: Overlap of genomic position range data was done using BedTools (74). Overlap means two genomic range data overlap by at least one base.

Average ChIP signal profile: For sequencing data, ChIP signal around center of each given genomic location were calculated by using normalized input subtracted-average tags number in each 50 bp bins in a set window. Relative distance of each tag from above-mentioned position and average signal was determined by using 'Sitepro' script of CEAS package (72) and plotting was done in R programming language.

Occupancy of H1 variants at individual chromosomes: Occupancy of H1 variants at all human chromosomes is an average of the input-subtracted ChIP-seq signal in 50 bp windows. Heat map and dendrogram were done with in-house R scripts. Correlation between the occupancy of H1 variants (input-subtracted ChIP-seq signal average of 50 bp genomic windows) and gene expression and gene richness coefficient was done with in-house R scripts. Gene expression for each chromosome was computed as the average of the expression of all the available expressed genes. The gene-richness coefficient (GRC) for each chromosome was calculated as the ratio between the percentage of total genes present in each chromosome and the percentage of base pairs of each chromosome to the total human genome.

Agilent expression arrays

Total RNA was extracted using High Pure RNA isolation Kit (Roche) according to the manufacturer's instructions. cDNA was obtained from 100 ng of total RNA using SuperScript VILO cDNA synthesis Kit (Invitrogen). High RNA integrity was assessed by Bioanalyzer nano 6000 assay. For each sample, 100 ng of total were reverse transcribed into cDNA with a T7 promoter and the cDNA was *in vitro* transcribed into cRNA in the presence of Cy3-CTP using the Low input quick Amp kit (Agilent). Labeled samples were purified using RNeasy mini spin columns (Qiagen). Then, 600 ng of cRNA were preblocked and fragmented in Agilent fragmentation buffer and mixed with Agilent GEx Hybridization mix. Hybridization mix was laid onto each sector of subarray gasket slide and sandwiched against an $8 \times 65K$ format oligonucleotide microarray (Human v1 Sureprint G3 Human GE 8x60k Microarray, Agilent design ID 028004) inside a hybridization chamber, which was hybridized overnight at 65°C. Subsequently array chambers were disassembled submerged in Agilent Gene Expression Buffer 1 and washed 1 min in another dish with the same solution with a magnetic stirrer at 200 rpm at room temperature, followed by 1 min in Agilent Gene Expression Buffer 2 with a magnetic stirrer at 200 rpm at 37°C and immediate withdrawal from the solution and air drying. Fluorescent signal was captured into TIF images with an Agilent scanner using recommended settings with Scan Control software (Agilent). Signal intensities were extracted into a tabulated text file using Feature Extraction software (Agilent) using the appropriate array configuration and annotation files. The normalized log2 intensities were obtained using quantile method with normalized expression background correction the Bioconductor Limma package in R.

Human H1 variants nomenclature

The correspondence of the human H1 variants nomenclature with its gene names is as follows: H1.0, HIF0; H1.1, HIST1H1A; H1.2, HIST1H1C; H1.3, HIST1H1D; H1.4, HIST1H1E; H1.5, HIST1H1B; H1X, HIFX.

RESULTS

All H1 variants are nonspecifically present at gene promoters and are depleted from TSS in active genes or on induced gene activation

To determine whether the genomic distribution of human histone H1 differs between variants, we used ChIP combined with semiquantitative PCR (ChIP-qPCR), promoter array hybridization (ChIP-on-chip) and massive sequencing (ChIP-seq). Because there is a limited number of H1-variant–specific ChIP-grade antibodies (only H1.2 and H1X in our hands), we developed T47D-derived cell lines stably expressing HA-tagged versions of each of the five somatic H1 variants expressed in most cell types (H1.0, H1.2, H1.3, H1.4 and H1.5) (see 'Materials and Methods' section) (59). These cell lines proliferated similarly to parental cells (data not shown). HA-tagged H1 variants (H1-HA) were expressed at levels lower than or similar to their corresponding endogenous histone, comparably across the different H1 variant-expressing cell lines, and they were incorporated into chromatin (Supplementary Figure S1). In ChIP-qPCR experiments, an anti-HA antibody was used to specifically pull down H1-associated chromatin fragments in cells expressing H1-HAs (Supplementary Figure S2). H1-associated chromatin included gene promoters, coding regions and repetitive DNA, irrespective of which H1-HA variant was immunoprecipitated (Supplementary Figure S3). A few differences were observed between

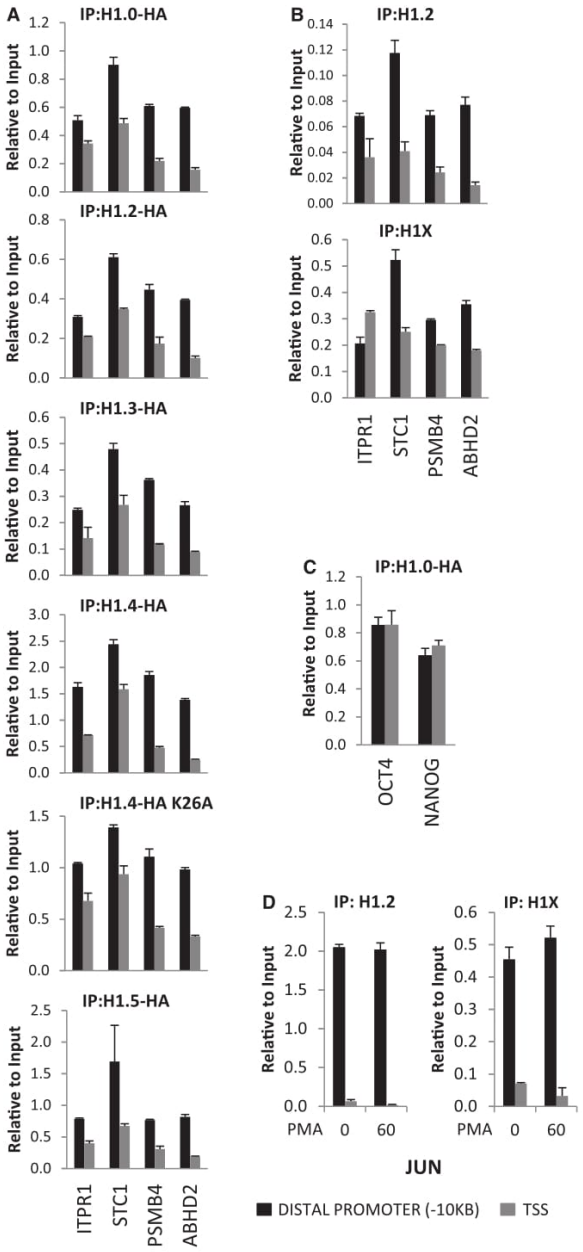


Figure 1. All H1 variants are present at gene promoters and depleted from TSS. (**A** and **C**) ChIP experiments were performed in T47D-derived cells stably expressing HA-tagged H1 variants, wild-type or a K26A mutant of H1.4 with anti-HA antibody and the abundance of IPed material was quantified by qPCR with oligonucleotides for the indicated promoters (−10 kb distal promoter or TSS), and corrected by input DNA amplification with the same primer pair. (**B**) ChIP experiments were performed in parental T47D cells with H1 variant-specific antibodies against H1.2 and H1X and the IPed material was quantified as in (A). (**D**) An H1 valley was performed at TSS of the JUN gene and increased on mitogenic stimulation. T47D cells were treated with PMA 100 nM for 60 min or left untreated and ChIP was performed with H1.2 and H1X antibodies. The abundance of IPed material was quantified by qPCR with oligonucleotides for the JUN promoter (−10 kb distal promoter or TSS), and corrected by input DNA amplification. Representative experiments performed in triplicate are shown.

variants, e.g. there were relatively less H1.3 but more H1.4 and H1.5 at alphoid repeats.

The specificity of H1 variant distribution was investigated in more detail at gene promoters previously shown to contain H1 in distal regions located 10 kb upstream of their TSS and depletion of H1 at the TSS (H1 valley) (62). All the H1 variants were detected at all distal promoter regions tested, in similar proportions, and a similar degree of H1 depletion was observed at the TSS of all genes for all the H1 variants, including an H1.4 mutant (K26A) at a residue targeted by acetyl and methyl transferases and reported to be involved in recruiting chromatin proteins (Figure 1A) (5,6,46,75). Moreover, local depletion of H1 at TSS was also observed by immunoprecipitating endogenous histones with specific H1.2 and H1X antibodies (Figure 1B). The ChIP specificity of these antibodies was confirmed in H1.2 and H1X inducible knock-down cells (Supplementary Figure S4). Interestingly, the TSS-associated H1 valley was not observed at genes inactive in these cells, i.e. OCT4 and NANOG (Figure 1C), while the H1 valley was evident at genes being expressed, as indicated by mRNA accumulation measured by RT-qPCR. Moreover, the H1 valley correlated with H3K4me3 enrichment at the TSS compared with a 10-kb upstream region, an open chromatin state at TSS measured by formaldehyde-assisted isolation of regulatory elements (FAIRE)-qPCR and nucleosome depletion (H3 ChIP) (Supplementary Figure S5). Furthermore, under stimulating conditions H1 depletion at the TSS was increased at inducible promoters, such as steroid hormone responsive promoters (MMTV) or genes induced by mitogenic agents (JUN and FOS) (Supplementary Figures S6 and S7). Noteworthily, in these early response genes, there was already an H1 valley in noninducing conditions and this became deeper on stimulation (Figure 1D and Supplementary Figure S7).

Extended depletion of H1 at promoters is dependent on the transcriptional status of the gene and shows differences between variants

To explore the genome-wide distribution of the different H1 variants across gene promoters, we hybridized ChIP material obtained with variant-specific antibodies or corresponding to HA-tagged H1 variant-associated chromatin with a promoter tiling array containing probes for 30 893 transcripts (−3200 to +800 bp to the TSS) arising from 22 542 human promoters. The average log2 ratio of probe intensity for all transcripts was plotted against the relative distance to the TSS for each variant and an H1 valley close to the TSS was apparent in all cases. Interestingly, in the two H1.2 samples (endogenous H1.2 and H1.2-HA), the valley was more pronounced and slightly shifted toward the TSS, compared with that for the other H1 variants (endogenous H1X and H1.0/3/4/5-HA) (Figure 2A).

Subsequently, this ChIP-chip data was combined with gene expression data for ca. 20 000 of the transcripts, obtained with the parental cell line in a human expression array (Agilent) (Supplementary Figure S8), and heat maps representing binding intensity were constructed for each

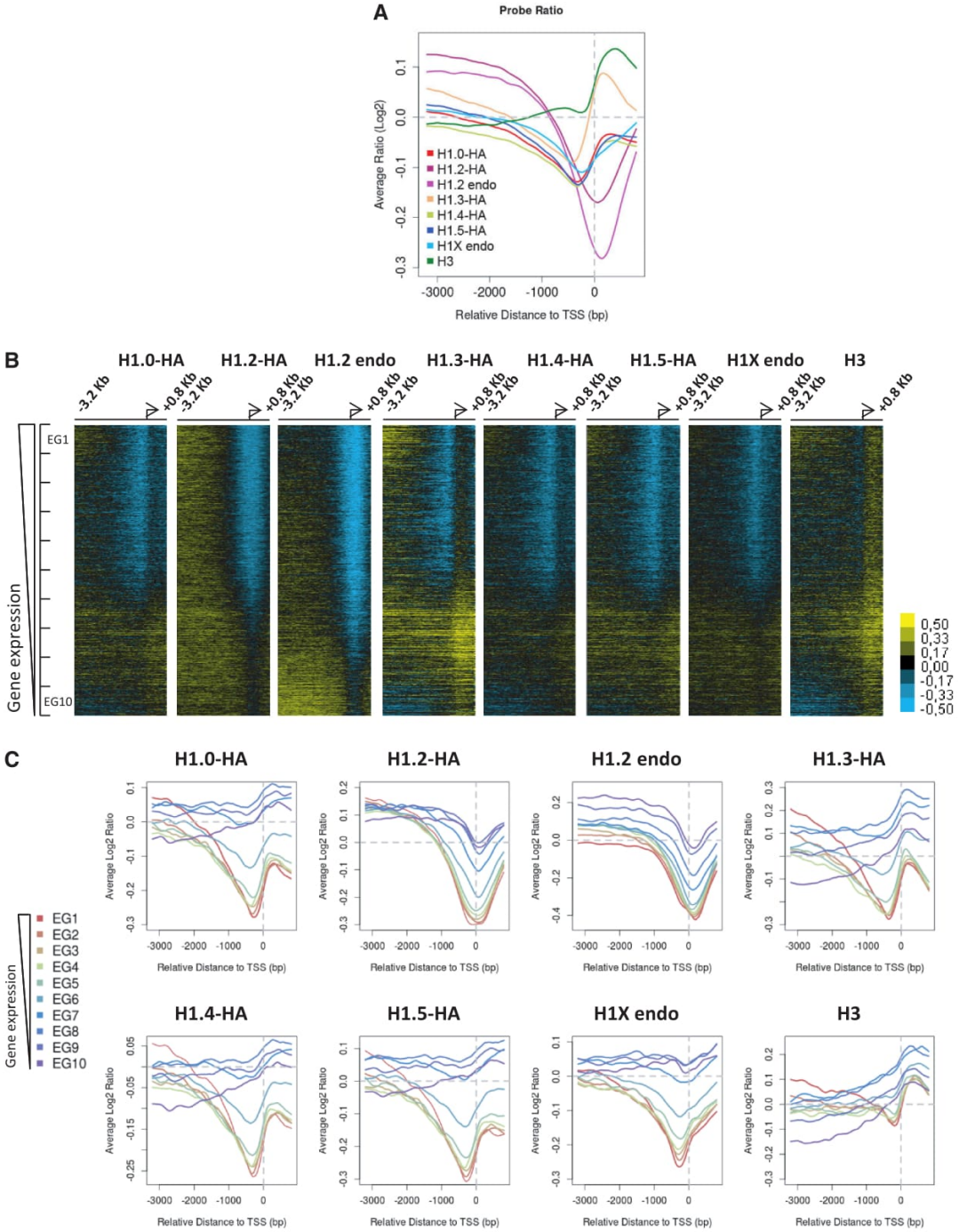


Figure 2. The extension of H1 depletion at promoters is transcription status-dependent and variant-specific. (**A**) Average log2 enrichment ratio of ChIP-chip probe intensity for all transcripts was represented regarding the relative distance to TSS for each variant. (**B**) Heat maps of ChIP-chip probe intensity around TSS (−3200 to +800 bp) for 20 338 transcripts from which the expression rate was determined. Genes are ordered from highest to lowest gene expression. (**C**) Average log2 ratio of ChIP-chip probe intensity represented regarding the relative distance to TSS for all transcripts classified according to expression in 10 groups containing a same number of transcripts, from highest (EG1) to lowest (EG10) expression. Representative ChIP-chip experiments are shown.

4480 Nucleic Acids Research, 2014, Vol. 42, No. 7

variant, ranking promoters from highest to lowest gene expression (Figure 2B). An H1 valley was clearly seen for at least the top 50–60% most highly expressed transcripts in all variants. Notably, the valley extended toward the least expressed genes in H1.2 samples. Then all the transcripts considered were divided into 10 groups from high to low expression, and average log2 ratio of ChIP-chip probe intensity was plotted against the relative distance to the TSS for each expression group and each variant (Figure 2C). These graphs confirmed that H1 depletion at promoters is dependent on the transcriptional status of the gene. The H1 valley around TSS was deeper and wider for H1.2 than for the other variants, irrespective of whether endogenous or HA-tagged histone was measured. In general, H1 depletion extended to some degree at least 1 kb upstream of the TSS of active genes, further than the predicted extent of the reported nucleosome-free region (NFR) that lies upstream of the TSS. To confirm this result, ChIP-chip for the core histone H3 was also performed and showed that H3 was depleted at active genes and more locally than H1 (Figure 2B and C). H3 and all H1s except H1.2 presented a marked enrichment peak immediately downstream of the TSS, which may correspond to a positioned nucleosome as previously reported (76,77). ChIP-qPCR on selected promoters confirmed some of these observations, namely, in some repressed promoters there was high H1.0 but low H1.2 content around the TSS (Supplementary Figure S9).

In addition to protein-coding genes, the promoter array contained 1145 noncoding transcripts, including structural RNAs and transcribed pseudogenes, that overall presented a low expression rate compared with the complete transcriptome. An H1 valley at the TSS was only apparent on the ChIP-chip heat maps for endogenous and HA-tagged H1.2, in agreement with our observation that an H1.2 valley occurs even at weakly expressed promoters (Supplementary Figure S10).

H1.2 abundance at distal promoters is a mark of transcriptional inactivity and negatively correlated with the presence of other H1 variants

Noteworthily, H1.2 abundance at distal promoter regions (−3200 to −2000 bp from TSS) was inversely proportional to gene expression, being more abundant at repressed promoters (Figure 2C). This was also observed to some extent for the other H1 variants and H3 with the exception of the ca. 10% most and least strongly expressed genes that showed the opposite trend. In agreement with this, when gene promoters were ranked from weakest to strongest H1 enrichment at the distal promoter region, a negative correlation with gene expression was seen especially for H1.2 (Figure 3A). Genes with the highest distal promoter H1.2 content (top 10%) mainly fell among those with the lowest expression, whereas genes with the lowest H1.2 content (bottom 10%) fell among those with the highest expression (Figure 3A, right panel). This was partially true also for H1X but less evident for the H1-HAs. Gene ontology analysis of H1 variant-enriched (top 10%) or -deprived (bottom 10%) promoters revealed that different biological processes were regulated by the different variants in T47D cells. For example, genes with the lowest content of H1X at promoters included active genes involved in chromatin organization, and those with the lowest H1.2 content in these regions included genes involved with cell–cell signaling or regionalization. On the other hand, genes with the highest H1X and H1-2 content at promoters included those involved in pattern formation and repressed genes involved in sensory perception, respectively (Supplementary Table S2).

Moreover, H1.2 abundance at distal promoter regions was inversely correlated with H3, H1X and H1-HA abundance, while H1.2-HA showed an intermediate pattern (Figure 3B and Supplementary Figure S11). This indicates that there is a preferential binding of H1.2 in some promoters (mostly repressed genes) compared with the other variants, and vice versa, many promoters are devoid of H1.2 but contain other H1 variants.

Venn diagrams were drawn for the top 10% genes with high or low H1.2 and high or low H1X at the distal promoter to identify genes presenting high2/lowX and vice versa (Supplementary Figure S12). The largest overlaps were between low2/highX promoters (553 genes), mainly corresponding to expressed genes (Figure 3C). Representative genes of the two groups were randomly selected (TMEM204 and TUBGCP5 for low2/highX, and COL4A3 and CUGBP2 for high2/lowX-containing promoters) and used to confirm by ChIP-qPCR that some promoters preferentially bind with particular variants (Figure 3D). Similarly, Venn diagram comparisons of the top 10% genes with high or low H1.2 versus high or low H1.0-HA showed that the largest overlaps were low2/high0 with 716, and high2/low0 with 276 genes (Supplementary Figure S13). Taken together, our data indicated that promoters having few H1.2 variants are loaded with large amounts of other variants, not only with exogenously expressed H1.0-HA but also endogenous H1X. Expression analysis of such groups of genes found that genes with few H1 variants at distal promoters are highly expressed, and vice versa, but also that H1.2 content is the strongest predictor of gene expression (Figures 3C and Supplementary Figures S12C and S13C).

The universality of the relative H1.2/H1X abundance at representative genes was tested in two additional cell lines by ChIP-qPCR (Figure 3E and Supplementary Figure S14). HeLa cells showed results similar to T47D, i.e. H1.2/H1X ratios were higher in COL4A3 and CUGBP2 genes than TMEM204 and TUBGCP5, although ratios in all genes were higher than in T47D reflecting a higher relative abundance of H1.2 in HeLa cells (Supplementary Figure S14). On the other hand, H1.2/H1X ratios in MCF7 were similar in all four genes, due to higher H1X signals in COL4A3 and CUGBP2 genes. This result indicated that relative abundances between variants at promoters were not fully conserved between cell types, although the patterns in T47D and HeLa were similar. In relation to this, ChIP-chip of H1.2 and H1X in HeLa confirmed that these two variants do not coexist at exactly the same distal promoters (Supplementary Figure S15).

Next, we plotted heat maps of H1.2 abundance at the promoters of genes ranked according to their position along several human chromosomes (Figure 4A).

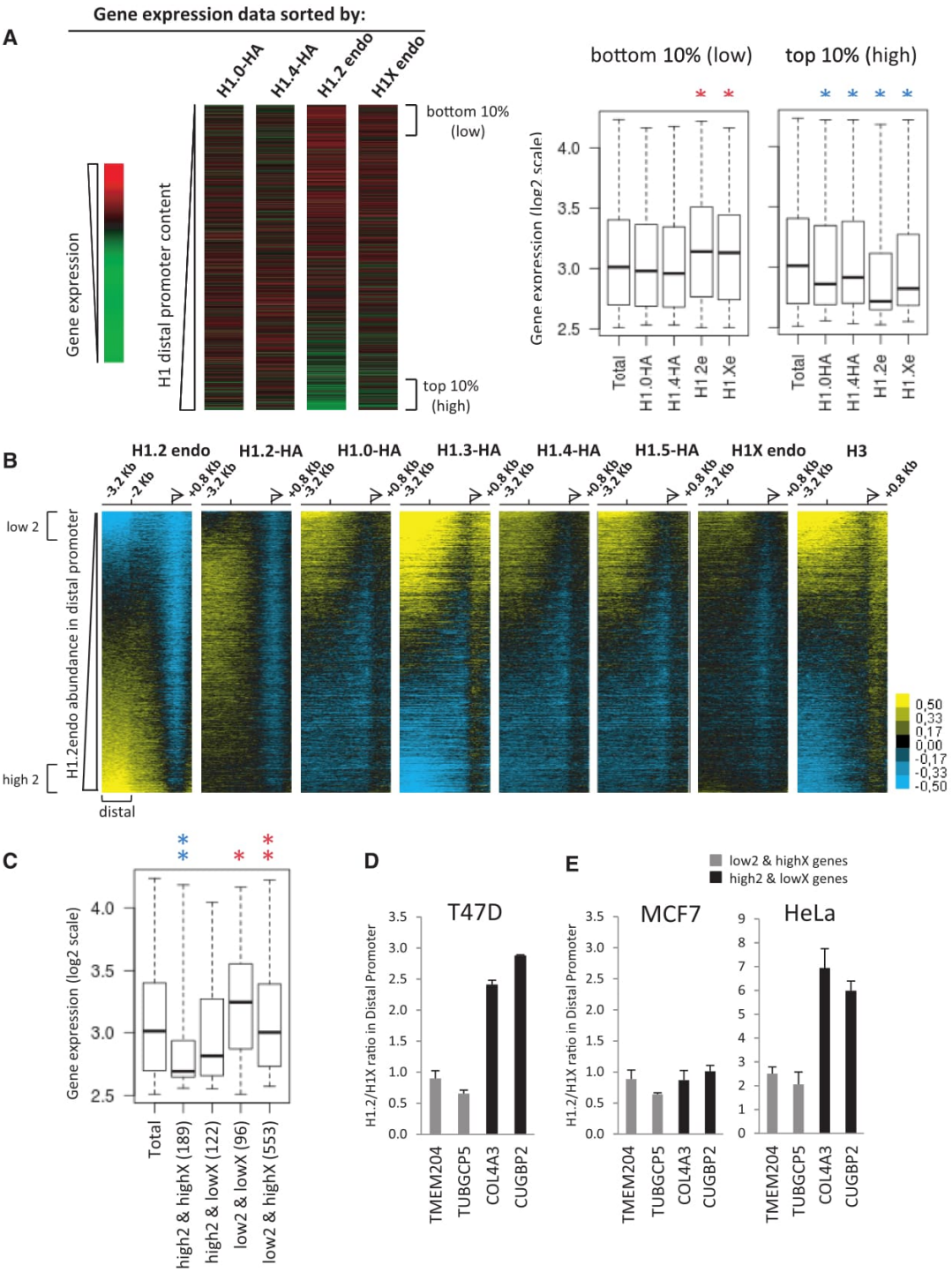


Figure 3. H1.2 abundance at distal promoter regions negatively correlates with gene expression and abundance of other variants. (**A**) Heat maps of gene expression data for 20 338 transcripts ordered from lowest to highest H1 content at distal promoter regions (−3200 to −2000 bp relative to TSS), for each of the H1 variants indicated. (Right panel) Expression levels of genes presenting the highest or lowest H1 variant content at distal promoter is shown as a box plot. Significance was tested using the Kolmogorov–Smirnov test. Enrichment and depletion is marked with red and blue asterisks, respectively. $*P < 0.001$. (**B**) Heat maps of H1 ChIP-chip probe intensity around TSS (−3200 to +800 bp) for 20 338 transcripts from which the expression rate was determined. Genes are ordered from lowest to highest H1.2 content at distal promoter regions. Genes with the top or lowest distal H1 content are indicated. These genes (2050 genes for each group, 10% of the total) were used to determine the number of coinciding genes as

(continued)

Interestingly, several domains of high H1.2 abundance were detected along these chromosomes, correlating with clusters of differential gene expression. Notably, chromosome 19, the most gene-rich chromosome, showed overall high gene expression and low H1.2 content at promoters, as did chromosome 17. On the other hand, the least gene-rich chromosome, chromosome 13, presented low gene expression and high H1.2 content (Figure 4A and Supplementary Figure S16). The observed clustered distribution was well conserved between cell lines, but differed between H1 variants. H1X and H1.0-HA abundances were not clustered with the same pattern as gene expression. Rather, these variants were abundant at promoters located on gene-rich chromosomes 17 and 19, and depleted on the gene-poor chromosome 13 (Supplementary Figure S16). In summary, H1.2 content at promoters is the best H1 reporter of gene expression.

H1 variants are differentially depleted from regulatory regions and enriched at CpG sites

To further explore whether the genomic distribution of H1 variants is heterogeneous, we combined ChIP of endogenous H1.2, H1X, H3 and HA-tagged H1.0, H1.2 and H1.4 with high-resolution sequencing (ChIP-seq) of up to 50 million reads per sample (Supplementary Table S1). To confirm the results obtained by ChIP-chip, we focused first on the input-subtracted normalized average ChIP signal obtained around coding regions of genes grouped according to basal expression as before (Figure 5A). Again, the H1 valley at the TSS depended on expression rates and differences were seen between H1 variants, mainly the abundance of H1.2 at the TSS of nonexpressed genes being lower than that of the other subtypes, which showed high levels toward nucleosome +1. Transcription termination sites (TTS) also showed differences between variants, being depleted of H1 subtypes except for H1.2. Interestingly, the H1 content of gene bodies increased toward the end and also depended on gene expression rates. While H3 levels were uniform, those of H1 variants such as H1.2 were lower at the 5′ moiety of highly active genes (Figure 5A).

In addition to the local displacement of H1 from active promoters, H1 variants were markedly depleted from other regulatory regions along the genome, namely, CC CTC-binding factor (CTCF) binding sites corresponding to insulators, and p300 binding sites associated with enhancers, but little affected at DNase hypersensitivity sites and FAIRE-identified regions representing open chromatin (Figure 5B and Supplementary Figure S17). When we calculated the input-subtracted coverage of H1 variants across the peaks of selected core histone modifications, depletion of H1.0 and H1.2, and to some extent of H1.4 but not H1X, was associated with positive histone marks linked to strong enhancers such as H3K4me1, H3K4me2 and H3K27ac (Supplementary Figure S17). H1 abundance at H3K4me3 and H3K9ac sites, enriched at TSS of active promoters, differed between variants, reflecting H1.2 depletion at the TSS of most genes but local enrichment of the other variants immediately after the TSS. No strong enrichment of H1 was found at negative histone marks such as H3K9me3 or H3K27me3. It is also worth noting that H1.2 abundance was lower at active marks than at those related with repression and chromatin compaction, in agreement with the observed correlation between H1.2 content and gene repression.

Next, we investigated whether the location of H1 variants coincided with CpG regions across the genome. As seen in Figure 5C, H1.0, H1X and H1.4 were clearly overrepresented in CpG regions compared with H1.2. Because CpG are mostly localized at gene promoters, this finding may reflect the overall higher abundance of those variants compared with H1.2 around TSS, considering the weakly expressed genes. Alternatively, it is not possible to rule out a certain relationship between H1.0 (and other variants apart from H1.2) and CpG or DNA methylation.

Differential prevalence of H1 variants along the genome

To further correlate ChIP-chip data of H1 abundance at promoters with ChIP-seq signals, regions of clustered genes with high H1.2 content such as the ones marked with asterisks in Figure 4A (chromosomes 1 and 12) were explored for input-subtracted H1 variant content using the UCSC genome browser (Figure 4B). The whole domain was enriched in H1.2 ChIP-seq signal compared with neighboring regions, indicating that H1.2 enrichment was not limited to the promoters of repressed genes therein. Interestingly, this domain was characterized by low GC content and the presence of LADs reported to anchor chromatin segments to the nuclear periphery (73). LADs are typified by low gene-expression levels, representing a repressive chromatin environment. Notably, the distribution of the other variants analyzed by ChIP-seq was not as clearly delimited to this domain as H1.2. Further examination of H1 variant signals across several regions containing LADs using the UCSC genome browser showed that H1.2 was the variant most strongly correlated with LAD positions and had fairly well delimited borders of enrichment (Supplementary Figure S18). When the input-subtracted coverage of H1 variants across LADs was calculated, H1.2 was the only variant showing enrichment (Figure 4C).

We then examined individual chromosomes for the presence of the input-subtracted signal of the different H1 variants. Abundance of H1 was heterogeneous along

Figure 3. Continued
shown in Supplementary Figures S12 and S13. (**C**) Expression levels of coinciding genes in the comparisons between genes presenting the highest or lowest H1.2 or H1X (h2/l2/hX/lX, respectively) content at distal promoter is shown as a box plot. The number of common genes for each comparison is indicated. Significance was tested using the Kolmogorov–Smirnov test. $^{**}P < 0.001$ and $^{*}P < 0.005$. (**D**) ChIP-qPCR confirmed that some genes are enriched in H1.2 or H1X at distal promoter. TMEM204 and TUBGCP5 genes were randomly chosen among the group of genes presenting low H1.2 and high H1X (553 genes), and COL4A3 and CUGBP2 genes among the genes presenting high H1.2 and low H1X (189 genes) (see Supplementary Figure S12). After ChIP-qPCR of H1.2 and H1X abundance at distal promoter regions of these genes in T47D cells, the relative ratio H1.2/H1X was calculated. (**E**) The differential ratio between H1.2 and H1X abundance at selected genes observed in T47D cells is conserved in HeLa cells but not in MCF7. Representative ChIP-qPCR experiments performed in triplicate are shown.

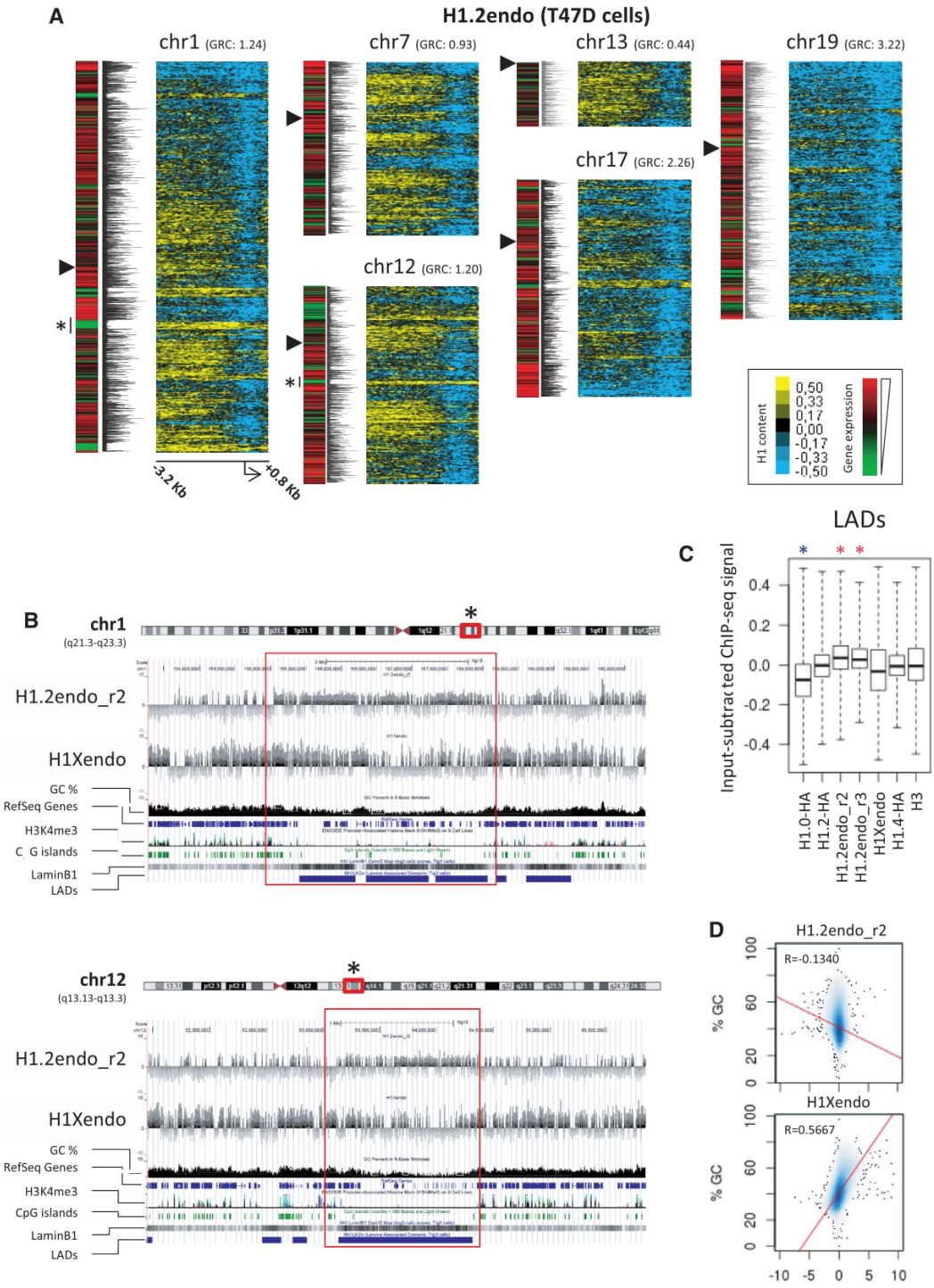


Figure 4. H1 variant content at gene promoters along human chromosomes and relation of H1 variants with LADs and GC content. (**A**) Heat maps of H1.2 ChIP-chip probe intensity around TSS (−3200 to +800 bp) for genes ordered according to their position along several human chromosomes. Gene expression levels for each gene in T47D cells is represented in the left in two different ways (as a heat map and graphical representation of log 2 ratios). A GRC for each chromosome, calculated as the ratio between the percentage of genes present in each chromosome and the percentage of base pairs of each chromosome to the total human genome, is indicated. The centromere location is marked with a triangle. Regions of interest are marked with an asterisk and viewed in the UCSC genome browser in (**B**). (B) Distribution of H1 variants along selected regions of chromosome 1 and 12. Input-subtracted H1.2 and H1X ChIP-seq signal viewed in the UCSC genome browser together with GC content, RefSeq genes, H3K4me3

(continued)

chromosomes, showing extensive patches of enrichment or depletion of H1 compared with the input (Supplementary Figure S19). Interestingly, the H1.2 pattern was the most different from the other variants, endogenous H1X and HA-tagged variants showing patterns that were similar to each other and to that of H3, while the pattern for HA-tagged H1.2 was more similar to endogenous H1.2 than to other HA-tagged H1s. It is worth noting that long genome patches of low GC content were found to be devoid of all H1 variants except H1.2, which was enriched. We next performed genome-wide correlation analysis between the input-subtracted H1 variant signal and GC content. Low CG content was associated with high occupancy of H1.2 but low occupancy of the other variants, including H1X, and vice versa (Figure 4D and Supplementary Figure S20).

Further comparison of the overall abundance of H1 variants at each individual chromosome revealed unique patterns creating corresponding clusters of chromosomes and H1 variants (Figure 6). Interestingly, chromosomes were clustered in a manner that was related to their gene-richness and the overall expression of genes they contained. Gene-rich chromosomes showed H1.0 and H1X enrichment, and H1.2 depletion, whereas the opposite was found at gene-poor chromosomes, in agreement with the promoter ChIP-chip data described above. Correlation analysis confirmed these conclusions (Supplementary Figure S21). Notably, H1 variants were clustered differently depending on whether they were replication-independent (H1.0 and H1X) or synthesized over the course of DNA replication only (H1.2, H1.4 and the core H3 histone). Further, H1.0 and H1X had a more heterogeneous distribution between chromosomes (data not shown).

Genomic annotation of enriched or depleted regions of individual H1 variants shows that H1.2 is associated with intergenic regions and repressed genes

Next, we searched specific regions of the genome either enriched or depleted for each H1 variant signal over input DNA with a fold change ≥ 2 using SICER software (Supplementary Table S3). Most H1-enriched regions were inside genes (arbitrarily defined as from −5 kb upstream to +3 kb downstream of the TTS), whereas H1.2 peaks were more abundant at intergenic regions (Supplementary Table S3 and Supplementary Figure S22). On the other hand, all H1-depleted regions were more abundant inside genes, especially for H1.2. Within genes, H1.2-enriched regions were disfavored at promoters (−5 kb to +1 kb flanking TSS) compared with other H1 peaks, whereas H1.2-depleted regions were strongly favored, in agreement with ChIP-chip data presented in Figure 2. In agreement with our aforementioned data, the GC content in H1.2-enriched regions was lower than in the other variants (Supplementary Figure S20). Next, we analyzed the overlap between H1-enriched and depleted regions with CpG islands. CpG islands were enriched at H1.2-depleted regions and at regions enriched for the other variants, confirming the inverse correlation between CpG islands and H1.2 described above (Supplementary Figures S22 and S23). As expected, regions overlapping with CpG sites were preferentially located at promoters. For example, 42% of H1.0- or H1X-enriched regions located at promoters overlapped with a CpG island, while this was the case for only 4–8% of regions enriched in these variants located at intergenic regions.

To identify H1 variant target genes we looked for genes that had at least one H1-enriched region from −5 kb to +3 kb from the TTS. H1.2 was the variant that was found to have the smallest number of target genes (Supplementary Table S3). Overlap analysis disclosed the number of genes containing peaks of a single variant or several variants (Supplementary Figure S24), and expression analysis revealed that genes with only H1.2 peaks were less expressed than target genes containing peaks of any other H1 variant (Figure 7A), in agreement with data above showing lower expression of genes containing elevated levels of H1.2 at distal promoter or coding regions. In those genes, the peak tended to be outside the promoter for H1.2, but at the promoter for the other single variant target genes (Supplementary Table S3). On the other hand, genes presenting H1.2-depleted regions were highly expressed, while genes with depleted regions of H1.0, H1.4 or H1X were expressed at lower levels than the total transcriptome average (Figure 7A).

We further investigated whether the identified H1-enriched regions fell within genes, proximal regulatory regions or distal intergenic regions using CEAS software (70). Again, H1.2 was more differently distributed than the other variants analyzed. H1.0-HA, H1X and H1.4-HA peaks were overrepresented in promoters, UTRs, exons and downstream regulatory regions, and underrepresented in distal intergenic regions compared with the complete genome, whereas H1.2-enriched regions were overrepresented in intergenic regions and underrepresented in exons and promoters (Figure 7B and Supplementary Figure S25). Except for those for H1.2, H1 peaks were as abundant in introns as in distal intergenic regions. On the other hand, depleted regions were similarly distributed across compartments in the different H1 variants, except H1.2-depleted regions, which were more abundant at promoters and less so at intergenic regions.

In summary, our data shows that histone H1 is not uniformly distributed along the genome and there are differences between variants, H1.2 being the one showing the most specific pattern and

Figure 4. Continued
(ENCODE average of 9 cell lines), CpG and LADs (data from Tig3 lung fibroblasts). (**C**) Box plots showing the occupancy of H1 variants (input-subtracted ChIP-seq signal) within LADs. Significance was tested using the Kolmogorov–Smirnov test taking as a control a random sample of windows with equal width to the LADs. Enrichment and depletion is marked with red and blue asterisks, respectively. $*P < 0.001$. (**D**) Genome-wide correlation scatterplots of H1.2 and H1X variants versus GC content. X axes: average input-subtracted H1 signal (normalized to 1000 bp window). Y axes: GC%. R: Pearson's correlation coefficient.

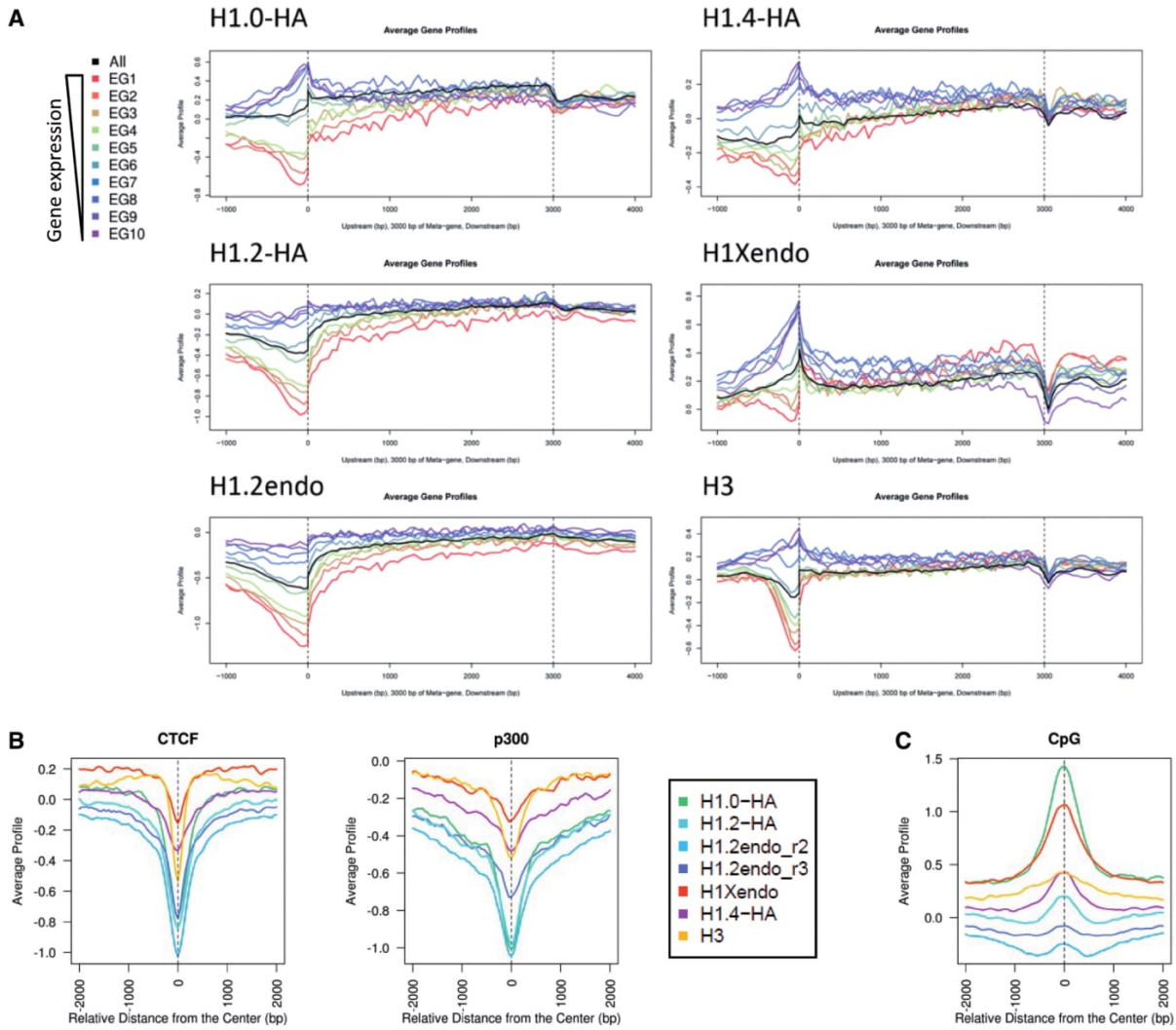


Figure 5. H1 is depleted from regulatory regions but present at CpG sites in a variant-specific manner. (**A**) Average, input-subtracted ChIP-seq signal of H1 variants around gene bodies flanked by TSS and transcription termination site (TTS), grouped according to basal expression (10% of total genes in each group). EG1 represents top expressed genes and EG10 genes with the lowest expression. Average for all genes is shown in black. Genic regions are represented as a 3-kb-long meta-gene surrounded by 1 kb region upstream TSS and 1 kb downstream TTS. (**B**) Average, input-subtracted ChIP-seq signal of H1 variant around the center of genomic CTCF and p300 binding sites (data from T47D cells). (**C**) Average, input-subtracted ChIP-seq signal of H1 variant around the center of CpG islands (as defined in UCSC database).

strongest correlation with low gene expression in breast cancer cells.

DISCUSSION

Mapping of H1 variants by ChIP with variant-specific antibodies and protein tagging uncovers differences between H1.2 and the other variants in breast cancer cells

Herein, we have investigated the distribution of all somatic histone H1 variants present in breast cancer cells, i.e. H1.0, H1X and H1.2 to H1.5 by combining ChIP with genomic technologies such as tiling promoter array hybridization and high-resolution sequencing. After testing several H1 variant-specific antibodies that we and others have produced, only H1.2 and H1X commercial antibodies were found to be useful in the ChIP-qPCR experiments, and variant specific, as shown by performing ChIP experiments in H1.2 and H1X knockdown (KD) cells. Consequently, we generated stable cell lines expressing HA-tagged versions of the H1 variants at protein levels close to or below endogenous levels, despite mRNA levels of exogenous H1 forms being higher (data not shown). This suggests that H1 is tightly posttranscriptionally regulated to control the overall levels of H1 and the proportion between variants, which vary considerably across cell types and cell lines. HA-tagging allowed us to perform ChIP of all variants with the same antibody,

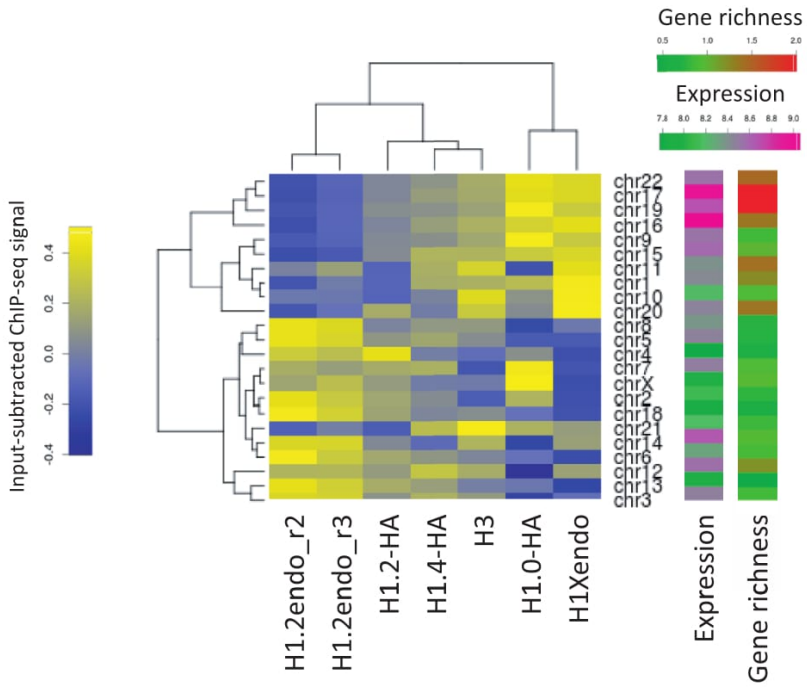


Figure 6. Human chromosomes show enrichment of different H1 variants in relation to average gene expression and gene richness. Heat map and dendrogram of the occupancy of H1 variants (input-subtracted ChIP-seq signal average of 50 bp genomic windows) at individual chromosomes. Gene expression and gene richness coefficient (GRC) of all chromosomes are shown as heat maps. GRC ≥ 2 are shown in the same color.

ruling out the variability being due to diverse antibody specificity or affinity. We found that all the H1 variants studied are widely distributed along the genome and within promoters with few differences between HA-tagged H1.0, H1.3, H1.4 and H1.5. In contrast, endogenous H1.2 presents striking differences. We rule out the possibility that the differential distribution is due to antibody usage or protein overexpression, as endogenous H1X presented an occurrence similar to HA-tagged variants and exogenous H1.2-HA resembled its endogenous counterpart more closely than the other H1-HAs.

On this basis, we report that, in the cell line investigated, H1.2 presents a variant-specific distribution and may have differential functions. In fact, we reported elsewhere that H1.2 KD produces unique effects, namely, cell cycle arrest at G1 and decreased nucleosome spacing, not seen in other H1 KDs, and these were observed not only in T47D cells but also in MCF7 cells (59). Nonetheless, this feature was not general, as it was not seen in other cell types tested, including HeLa cells in which H1.2 is highly abundant, indicating that H1 variants may have cell type-dependent specific effects. Instead, our data cannot rule out that the other variants studied may have redundant functions and distribution in breast cancer cells. A recent report on the genomic distribution of Dam-H1.1 to H1.5 in lung fibroblasts IMR90 cells found that H1.1 is the only subtype showing divergent features (66). H1.1 is not expressed in breast cancer cells or in many other cell types. Instead, H1.2 and H1.4 are the only variants that have been found in all cell lines tested to date (29,78). Additionally, mRNA levels of these two variants are maintained in nondividing cells and along differentiation, compared with H1.3 and H1.5 levels that are reduced (31,79). Although too small a sample, these results suggest that different H1 subtypes may play different roles in different cell types, over the course of development and in cancer cells, inviting further investigation of H1 variants occurrence.

We have noticed that H1.2-HA was not distributed in exactly the same way as endogenous H1.2 and showed intermediate features somewhat similar to the other H1-HAs. We believe that this recombinant protein has the H1.2 structural features that direct it to the natural H1.2-occupied sites, but owing to its overexpression it may also locate at distinct sites normally occupied by other H1 variants. We have observed, by ChIP, that on knock down of endogenous H1.2, H1.2-HA occupancy increased (data not shown), suggesting a relocation to H1.2 sites. Overall, we believe that caution should be taken when interpreting data generated with exogenous histone variants fused either to the Dam domain or to peptide tags.

H1 depletion from promoters and coding regions is more pronounced than H3 depletion and shows differences between H1 variants

Our analysis has also shown that all H1s are removed from active promoters, with maximum depletion close to TSS but extending several nucleosomes upstream, beyond the reported NFR, and within the coding regions. These

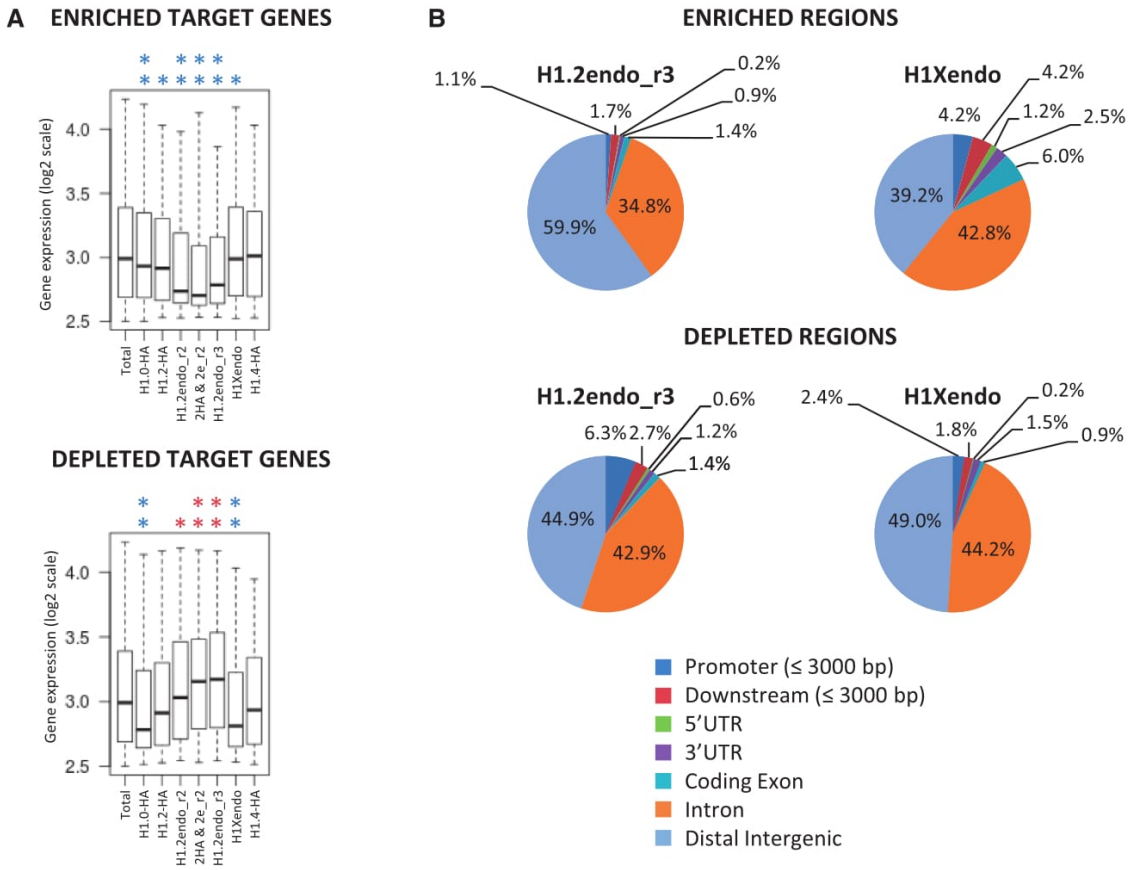


Figure 7. Genomic annotation of regions found to be enriched or depleted of individual H1 variants and expression of target genes. (**A**) The expression profiles of target genes containing enriched or depleted regions for a unique variant are shown as box plots. The profile of genes containing both H1.2-HA and H1.2endo (replica 2) enriched or depleted regions are also shown (2HA & 2e_r2). Significance was tested using the Kolmogorov–Smirnov test. Enrichment and depletion is marked with red and blue asterisks, respectively. $^{**}P < 0.001$ and $^{*}P < 0.005$. (**B**) Genomic annotation of regions enriched or depleted of endogenous H1.2 or H1X. Pie diagram of distribution of H1 variants enriched regions at genes, proximal regulatory regions and distal intergenic regions. Promoter and downstream regions are defined as 3000 bp upstream TSS or downstream TTS, respectively. The proportions of the H1.2 and H1X enriched or depleted regions in several genomic features were significantly different from the whole genome proportions of those features ($P < 2.2\text{e-}16$). Significance was tested using in-house R scripts.

regions containing nucleosomes but not H1 may coincide with H2A.Z and H3.3-containing nucleosomes, as both H2A.Z and H3.3 have been reported to locate at active promoters surrounding the NFR, where they positively regulate transcription (80–82). Additionally, other authors have observed weaker histone H1 binding in H2A.Z-containing nucleosomes (83) and a negative genome-wide correlation between H1 and H3.3 (63). These observations support the view that H1 removal is part of the chromatin remodeling events that occur on promoter activation to facilitate binding of transcription factors and the RNA polymerase machinery (49,84–86). Furthermore, the shape of the H1.2 (and H1.2-HA) valley at the TSS in ChIP-chip and ChIP-seq data (Figures 2 and 5) was slightly different from that of other H1 variants. Unlike the signals for other variants, the H1.2 signal did not show local enrichment immediately after the TSS. This local enrichment may coincide with a well-positioned nucleosome (+1), flanked by phased nucleosomes. This indicates that such a nucleosome may contain any H1 variant except H1.2. Additionally, H1.2 was not abundant around the TSS of repressed genes, suggesting that TSS of genes are epigenetically marked, including the absence of H1.2. Overall, we have shown a strong rejection of H1.2 from the TSS of most genes.

Interestingly, we have found that immediate-early responsive promoters, under nonstimulating conditions, are prepared to respond to stimuli by keeping the TSS free of H1, indicating that mechanisms other than transcription initiation might dictate H1 clearance. In this case, there is also histone H3 depletion at the TSS compared with at the distal promoter in the absence of stimuli, indicating that the NFR might be maintained to allow rapid response after stimulation. Supporting our hypothesis, it has been recently proposed that transcription factors interact with DNA in a dynamic way, and

some transcription factor–DNA interactions are established before the stimuli, especially at immediate-early genes (87).

Comparison of H1 occupancy with H3 has shown that all H1s except H1.2 follow the distribution of the core histone, whether this represents nucleosome enrichment, stability or defined positioning through the cell population. Nonetheless, H1 depletion at promoters and regulatory sites (CTCF or p300 binding sites) is more extensive than H3, denoting that nucleosomes might be ejected from delimited sites such as the NFR at the TSS, but H1 might be depleted from larger regions encompassing several nucleosomes. This is in agreement with previous reports showing that dips of low H1 occupancy at TSS and regulatory sites are not due to a lack of nucleosomes as they show enrichment of the core histone variant H3.3 (63). Moreover, at coding regions, the differential content of H1 in active versus repressed genes is more pronounced than those of H3, especially toward the 5′ of genes. Consequently, gene-rich domains might adopt an overall decondensed chromatin structure. Nonetheless, at active genes H1 is less abundant in promoters than coding regions, indicating that H1 presence might be more restrictive for transcription initiation than for elongation.

Initial ChIP-qPCR experiments indicated that all H1 variants were present at all tested promoters. Nonetheless, hybridization of ChIP material with a promoter array revealed that promoters might present differential H1 variant abundance (Figure 3). The most striking difference is between H1.2 and the other H1s, including H1X. Subsets of genes with the highest abundance of one variant and the lowest of another have been identified, i.e. those with a high or low H1.2/H1X ratio. Overall, expression of genes presenting these features is different, relating H1 variant content with gene expression. Notably, the relative abundance of H1.2 and H1X in the selected promoters was conserved in the distant HeLa cell line, but not in MCF7 cells. Thus, we propose that the relative promoter abundance of H1 variants may be related to, among other factors, their relative H1 variant content in a given cell type.

Two types of H1-containing chromatin are present in breast cancer cells with different association with gene density and expression

The negative correlation observed between gene activity and H1.2 content found at promoters extended upstream toward the whole genomic region. Patches of H1.2 enrichment seem to be associated with gene repression, gene-poor regions (including entire chromosomes, such as chromosome 13), low GC content or LADs, features related to chromatin compaction (Figure 4). Moreover, H1.2-enriched regions were frequently found at intergenic regions. Similar results were found in previous studies, linking histone H1 to repressive and compacted regions of the genome and suggesting a role for H1 in 3D organization of the genome. Some of these features were described by Cao et al. for mouse $H1c^{Myc}$ and $H1d^{FLAG}$ in ESCs, the closest orthologs of human H1.2 and H1.3, by Li et al. for human H1.5 in differentiated IMR90 fibroblasts, and by Izzo et al. for human Dam-H1.2 to H1.5 in IMR90 cells also (64–66). However, in the last of these, H1.1 presented a DamID binding profile distinct from the other subtypes that, in some extent, resembles the distribution of H1 other than H1.2 in our analysis in breast cancer cells, that is, they were more closely associated with higher GC content, genes, its promoters and CpG islands, and were not enriched in LADs. Interestingly, in the study of Cao *et al.* when single peaks for H1c and H1d in mouse ESCs were compared, H1d (H1.3) was more closely related to GC-rich sequences and LINES, and H1c (H1.2) to AT-rich sequences, Giemsa-positive regions and satellite DNA. It is conceivable that there are at least two groups of H1 variants with different distributions in each cell type, such that taken together histone H1 variants cover the whole genome, being present in most of the nucleosomes.

Whether a single variant may present distinct features in different cell types rather than having intrinsic properties is an intriguing question. Factors involved may be the relative and absolute abundance of each variant and whether a genome needs more plasticity or is progressively silenced, i.e. pluripotency versus terminal differentiation. In this sense, Li et al. described the existence of zones of H1.5 enrichment in differentiated fibroblasts but not in ESCs (64), and it has been reported that architectural proteins, such as HP1 and H1, are hyperdynamic and bind loosely to chromatin in ESCs (88,89). Additionally, we have previously reported progressive changes in the expression and abundance of H1 variants over the course of differentiation of human embryonic stem cells and of reprogramming of differentiated cells to Induced pluripotent stem cells (iPS), i.e. the opposite direction (31). Thus, considering the importance of H1 in chromatin structure and compaction, differential expression and/or distribution of H1 variants could mediate the transition between different chromatin states, and explain the more 'open' chromatin state of undifferentiated cells, which contributes to the maintenance of pluripotency by creating a poised chromatin state that leads to rapid activation of lineage-specific genes when differentiation is induced. In fact, it has been proposed that different 'anti-silencing' mechanisms, including incorporation of specific histone variants such as H3.3, are involved in the maintenance of open chromatin in ES cells (90).

Cancer is another cellular state in which global chromatin rearrangement is observed. In fact, alterations in nuclear morphology are one of the characteristics of cancer cells. Tumor-originated cells accumulate genetic and/or epigenetic differences compared with nontumor cells, and chromatin is reorganized leading to altered gene expression programs and higher plasticity. The hallmark of cancer is dedifferentiation and gene dysregulation. DNA methylation and histone modifications are two epigenetic mechanisms that are altered in cancer cells. Moreover, large organized chromatin K (lysine) modifications are reduced in cancer (91), and genes encoding proteins of the nuclear membrane present altered expression in many cancer types (92), indicating that LADs might be partially disorganized in cancer in accordance with the large-scale chromatin

decondensation. Thus, it is conceivable that the distribution of histone H1 variants could be different in such reorganized nuclei, to that observed in nonmalignant cells. In turn, this could be the reason why in our study most of the H1 variants in genome regions were found to be associated with more active and open chromatin. Moreover, given the association of H1 with LADs reported here and by Izzo et al., we hypothesize that H1 could be a key player in establishing LADs in normal cells, and could also participate in the rearrangement of such domains in cancer cells due to a different prevalence of H1 variants within these domains. Alternatively, LAD reorganization in cancer cells could cause H1 variant redistribution in these genomic domains.

Tumor cells are characterized by a different methylome from that of normal cells [reviewed in (93)]. There is both global CpG hypomethylation, causing genomic instability, and hypermethylation of particular promoters including tumor-suppressor genes. In our analysis, we found that CpG islands contain H1.0, H1X and to a lesser extent H1.4, but not H1.2. This might reflect the relative abundance of these variants at promoters and suggests that promoter occupancy by H1 variants other than H1.2 is more permissive for transcription regulation in breast cancer cells. Alternatively, as H1.2 prevalence in intergenic CpG islands is also lower than that of other variants, we cannot rule out a direct role of the different H1 variants in CpG island regulation in breast cancer cells.

Similarly, within a long region of genomic sequence, genes are often characterized by having a higher GC content than the background GC content of the entire genome. We found that H1 variants except H1.2 are associated with higher GC content regions, consistent with the preferential location of H1-enriched regions within genes. H1.2 presents an inverse correlation with GC content at a genome-wide level and H1.2-enriched regions associate with lower GC content than other variants. In our analysis, H3 also associates preferentially with higher GC-content regions, in agreement with reports describing greater nucleosome-space occupancy coinciding with active transcription and higher GC contents (94).

Altogether, it seems that H1 variants are differentially associated with CpG islands and GC content in breast cancer cells. Our data are not completely consistent with previous reports showing low amounts of H1 in CpG islands (65,95). However, mouse H1d was more closely associated with GC-rich regions than H1c in the study of Cao *et al.* (65). Additionally, another study showed H1 variant-dependent interaction with DNMTs (96). In that study, it was found that, unlike other H1 variants, H1c (H1.2) does not interact with DNMT1 and DNMT3B. Based on the differential association of H1 variants with CpG islands and GC-rich regions in T47D breast cancer cells, we hypothesize that a redistribution of most of histone H1 variants in cancer may help to establish a differential chromatin state, but also an altered methylation pattern. In fact, H1 variants are differentially related to several types of cancer (33,97). Additionally, comparison of human mammary epithelial cells with breast cancer cell lines including T47D (98) showed global massive hypomethylation at CpG-poor regions, and hypermethylation at CpG-rich gene-related regions, proximal to the TSS, where local enrichment of all H1 variants except H1.2 is observed in our data. Moreover, hypomethylated regions in breast cancer cells coincide with repressive chromatin, gene silencing, repressive histone posttranslational modifications (PTMs), intergenic regions and LADs (99), which in turn coincides with an enrichment of H1.2 found in our analysis. Further investigation of the DNA methylation profile of T47D breast cancer cells could confirm a differential role of H1 variants in establishing or maintaining DNA methylation in breast cancer.

Chromatin containing H1 variants other than H1.2 might support a level of compaction that facilitates a rapid conversion into either an active or a repressed state and, consequently, these variants are allowed at TSS of genes before activation. In fact, a particular posttranslational modification in H1.4 (K34Ac) has been found to locate around the TSS of active genes (49). Instead, we have described that H1.2 occupancy at distal promoters is the best predictor of gene repression. Moreover, genes presenting H1.2-enriched regions are clearly less strongly expressed than average. This study points toward the inclusion of H1.2 as a repression mark and to it being associated with closed chromatin. In this regard, H1.2 has been found to be included in a p53-containing repressive complex in HeLa cells (50), and murine H1.2 has been found to be developmentally upregulated in the retina, promoting facultative heterochromatin formation in mature rod photoreceptors (100).

Several studies have compared the chromatin binding affinity and residence time on chromatin of the different H1 subtypes in different organisms or cell lines, as well as its nuclear localization, obtaining diverse, if not controversial, results on the functional heterogeneity of H1 variants. In general, H1.2 is among the variants presenting intermediate or low affinity for chromatin and, consequently, elevated mobility. Instead, H1.4 has been mostly associated with high affinity, low mobility and colocalization with heterochromatin (40,101–103). We do not fully understand how these properties may relate or contradict our observation of H1.2 being enriched in repressed and gene-poor chromatin in breast cancer cells. Certainly, different experimental approaches performed in the same cell model would facilitate to reconcile the different observations.

There is nowadays increasing evidence of a 3D organization of the genome within the cell nucleus. Interphase chromatin is organized in large chromosome territories defined as 'topological domains', which can interact despite being several megabases apart (104,105). These domains are stable across different cell types and highly conserved across species. It has already been reported that embedded genes in these domains are in a transcriptionally similar state and associated with transcriptionally related histone marks and chromatin features. Hence, it is not unreasonable to speculate that H1 could be involved in the formation or maintenance of such domains due to its role in chromatin structure. High-throughput profiling

of chromatin marks and components has recently made it possible to define chromatin states (106,107). In *Drosophila* cells, five principal chromatin types have been described, H1 being present in all of them in different proportions (107). Although this may reflect the general features of H1 occurrence, in cells presenting several H1 subtypes a differential distribution of subtypes between different chromatin types may occur, as is suggested in our study. We have found that H1.2 is the variant most closely associated with LADs, low GC content and gene-poor regions and chromosomes that are normally located at the periphery of the nucleus, features related to chromatin compaction, while chromatin associated with the other variants presents features of a more plastic chromatin. Interestingly, gene-rich chromosomes, presumably with a more dynamic chromatin and histone H1 exchange, and located toward the center of the nucleus, are enriched in H1 variants synthesized all through the cell cycle, namely H1.0 and H1X. It would be interesting to further analyze the colocalization of the different human H1 variants with chromatin marks and components that better define the diverse chromatin states, although these types of comparisons are limited by the availability of high-throughput data on the same or related cell types.

ACCESSION NUMBERS

The data sets are available in the Gene Expression Omnibus (GEO) database under the accession number GSE49345.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Drs M. Martínez-Balbas and F. Azorín for comments on the manuscript, and Dr Lee Krauss for sharing oligonucleotide sequences. They thank Institut de Recerca de Barcelona Functional Genomics core facility, BGI Tech and Dr Lauro Sumoy for the genomic experiments.

FUNDING

Ministerio de Ciencia e Innovación of Spain (MICINN) and FEDER [BFU2011-23057 to A.J.]; Ministerio de Economía y Competitividad [SAF2012-36199 to N.L.-B.]; Generalitat de Catalunya [2009-SGR-1222 to A.J.]; JAE-Doc contract from CSIC-MICINN (to J.-M.T.); TA contract from CSIC-MICINN (to R.M.); FPU predoctoral fellowship from MICINN (to L.M.-A.). Funding for open access charge: Ministerio de Ciencia e Innovación of Spain (MICINN) and FEDER [BFU2011-23057 to A.J.].

Conflict of interest statement. None declared.

REFERENCES

1. Thoma,F., Koller,T. and Klug,A. (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell Biol.*, **83**, 403–427.
2. Bednar,J., Horowitz,R.A., Grigoryev,S.A., Carruthers,L.M., Hansen,J.C., Koster,A.J. and Woodcock,C.L. (1998) Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc. Natl Acad. Sci. USA*, **95**, 14173–14178.
3. Brown,D.T. (2003) Histone H1 and the dynamic regulation of chromatin function. *Biochem. Cell Biol.*, **81**, 221–227.
4. Bustin,M., Catez,F. and Lim,J.H. (2005) The dynamics of histone H1 function in chromatin. *Mol. Cell*, **17**, 617–620.
5. Garcia,B.A., Busby,S.A., Barber,C.M., Shabanowitz,J., Allis,C.D. and Hunt,D.F. (2004) Characterization of phosphorylation sites on histone H1 isoforms by tandem mass spectrometry. *J. Proteome Res.*, **3**, 1219–1227.
6. Wisniewski,J.R., Zougman,A., Kruger,S. and Mann,M. (2007) Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. *Mol. Cell. Proteomics*, **6**, 72–87.
7. Wood,C., Snijders,A., Williamson,J., Reynolds,C., Baldwin,J. and Dickman,M. (2009) Post-translational modifications of the linker histone variants and their association with cell mechanisms. *FEBS J.*, **276**, 3685–3697.
8. Bonet-Costa,C., Vilaseca,M., Diema,C., Vujatovic,O., Vaquero,A., Omenaca,N., Castejon,L., Bernues,J., Giralt,E. and Azorin,F. (2012) Combined bottom-up and top-down mass spectrometry analyses of the pattern of post-translational modifications of *Drosophila melanogaster* linker histone H1. *J. Proteomics*, **75**, 4124–4138.
9. Lesner,A., Kartvelishvili,A., Lesniak,J., Nikolov,D., Kartvelishvili,M., Trillo-Pazos,G., Zablotna,E. and Simm,M. (2004) Monoubiquitinated histone H1B is required for antiviral protection in CD4(+)T cells resistant to HIV-1. *Biochemistry*, **43**, 16203–16211.
10. Wisniewski,J.R., Zougman,A. and Mann,M. (2008) Nepsilon-formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. *Nucleic Acids Res.*, **36**, 570–577.
11. Hill,D.A. (2001) Influence of linker histone H1 on chromatin remodeling. *Biochem. Cell Biol.*, **79**, 317–324.
12. Schlissel,M.S. and Brown,D.D. (1984) The transcriptional regulation of Xenopus 5s RNA genes in chromatin: the roles of active stable transcription complexes and histone H1. *Cell*, **37**, 903–913.
13. Hashimoto,H., Takami,Y., Sonoda,E., Iwasaki,T., Iwano,H., Tachibana,M., Takeda,S., Nakayama,T., Kimura,H. and Shinkai,Y. (2010) Histone H1 null vertebrate cells exhibit altered nucleosome architecture. *Nucleic Acids Res.*, **38**, 3533–3545.
14. Vujatovic,O., Zaragoza,K., Vaquero,A., Reina,O., Bernues,J. and Azorin,F. (2012) *Drosophila melanogaster* linker histone dH1 is required for transposon silencing and to preserve genome integrity. *Nucleic Acids Res.*, **40**, 5402–5414.
15. Sera,T. and Wolffe,A.P. (1998) Role of histone H1 as an architectural determinant of chromatin structure and as a specific repressor of transcription on *Xenopus oocyte* 5S rRNA genes. *Mol. Cell. Biol.*, **18**, 3668–3680.
16. Alami,R., Fan,Y., Pack,S., Sonbuchner,T.M., Besse,A., Lin,Q., Greally,J.M., Skoultchi,A.I. and Bouhassira,E.E. (2003) Mammalian linker-histone subtypes differentially affect gene expression *in vivo*. *Proc. Natl Acad. Sci. USA.*, **100**, 5920–5925.
17. Fan,Y., Nikitina,T., Zhao,J., Fleury,T.J., Bhattacharyya,R., Bouhassira,E.E., Stein,A., Woodcock,C.L. and Skoultchi,A.I. (2005) Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell*, **123**, 1199–1212.
18. Fan,Y., Sirotkin,A., Russell,R.G., Ayala,J. and Skoultchi,A.I. (2001) Individual somatic H1 subtypes are dispensable for mouse

development even in mice lacking the H1(0) replacement subtype. *Mol. Cell. Biol.*, **21**, 7933–7943.
19. Hellauer,K., Sirard,E. and Turcotte,B. (2001) Decreased expression of specific genes in yeast cells lacking histone H1. *J. Biol. Chem.*, **276**, 13587–13592.
20. Lin,Q., Inselman,A., Han,X., Xu,H., Zhang,W., Handel,M.A. and Skoultchi,A.I. (2004) Reductions in linker histone levels are tolerated in developing spermatocytes but cause changes in specific gene expression. *J. Biol. Chem.*, **279**, 23525–23535.
21. Shen,X. and Gorovsky,M.A. (1996) Linker histone H1 regulates specific gene expression but not global transcription *in vivo*. *Cell*, **86**, 475–483.
22. Takami,Y., Nishi,R. and Nakayama,T. (2000) Histone H1 variants play individual roles in transcription regulation in the DT40 chicken B cell line. *Biochem. Biophys. Res. Commun.*, **268**, 501–508.
23. Happel,N. and Doenecke,D. (2009) Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene*, **431**, 1–12.
24. Zhang,Y., Cooke,M., Panjwani,S., Cao,K., Krauth,B., Ho,P.Y., Medrzycki,M., Berhe,D.T., Pan,C., McDevitt,T.C. *et al.* (2012) Histone h1 depletion impairs embryonic stem cell differentiation. *PLoS Genet.*, **8**, e1002691.
25. Helliger,W., Lindner,H., Grubl-Knosp,O. and Puschendorf,B. (1992) Alteration in proportions of histone H1 variants during the differentiation of murine erythroleukaemic cells. *Biochem. J.*, **288(Pt.3)**, 747–751.
26. Lennox,R.W. and Cohen,L.H. (1983) The histone H1 complements of dividing and nondividing cells of the mouse. *J. Biol. Chem.*, **258**, 262–268.
27. Pina,B., Martinez,P. and Suau,P. (1987) Changes in H1 complement in differentiating rat-brain cortical neurons. *Eur. J. Biochem.*, **164**, 71–76.
28. Pina,B. and Suau,P. (1987) Changes in the proportions of histone H1 subtypes in brain cortical neurons. *FEBS Lett.*, **210**, 161–164.
29. Meergans,T., Albig,W. and Doenecke,D. (1997) Varied expression patterns of human H1 histone genes in different cell lines. *DNA Cell Biol.*, **16**, 1041–1049.
30. Parseghian,M.H. and Hamkalo,B.A. (2001) A compendium of the histone H1 family of somatic subtypes: an elusive cast of characters and their characteristics. *Biochem. Cell Biol.*, **79**, 289–304.
31. Terme,J.M., Sese,B., Millan-Arino,L., Mayor,R., Izpisua Belmonte,J.C., Barrero,M.J. and Jordan,A. (2011) Histone H1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency. *J. Biol. Chem.*, **286**, 35347–35357.
32. Sato,S., Takahashi,S., Asamoto,M., Nakanishi,M., Wakita,T., Ogura,Y., Yatabe,Y. and Shirai,T. (2012) Histone H1 expression in human prostate cancer tissues and cell lines. *Pathol. Int.*, **62**, 84–92.
33. Medrzycki,M., Zhang,Y., McDonald,J.F. and Fan,Y. (2012) Profiling of linker histone variants in ovarian cancer. *Front. Biosci. (Landmark Ed).*, **17**, 396–406.
34. Sjoblom,T., Jones,S., Wood,L.D., Parsons,D.W., Lin,J., Barber,T.D., Mandelker,D., Leary,R.J., Ptak,J., Silliman,N. *et al.* (2006) The consensus coding sequences of human breast and colorectal cancers. *Science*, **314**, 268–274.
35. Telu,K.H., Abbaoui,B., Thomas-Ahner,J.M., Zynger,D.L., Clinton,S.K., Freitas,M.A. and Mortazavi,A. (2013) Alterations of histone H1 phosphorylation during bladder carcinogenesis. *J. Proteome Res.*, **12**, 3317–3326.
36. Clausell,J., Happel,N., Hale,T.K., Doenecke,D. and Beato,M. (2009) Histone H1 subtypes differentially modulate chromatin condensation without preventing ATP-dependent remodeling by SWI/SNF or NURF. *PLoS One*, **4**, e0007243.
37. Caterino,T.L. and Hayes,J.J. (2011) Structure of the H1 C-terminal domain and function in chromatin condensation. *Biochem. Cell Biol.*, **89**, 35–44.
38. Oberg,C. and Belikov,S. (2012) The N-terminal domain determines the affinity and specificity of H1 binding to chromatin. *Biochem. Biophys. Res. Commun.*, **420**, 321–324.
39. Vila,R., Ponte,I., Collado,M., Arrondo,J.L., Jimenez,M.A., Rico,M. and Suau,P. (2001) DNA-induced alpha-helical structure in the NH2-terminal domain of histone H1. *J. Biol. Chem.*, **276**, 46429–46435.
40. Vyas,P. and Brown,D.T. (2012) N- and C-terminal domains determine differential nucleosomal binding geometry and affinity of linker histone isotypes H1(0) and H1c. *J. Biol. Chem.*, **287**, 11778–11787.
41. Lever,M.A., Th'ng,J.P., Sun,X. and Hendzel,M.J. (2000) Rapid exchange of histone H1.1 on chromatin in living human cells. *Nature*, **408**, 873–876.
42. Misteli,T., Gunjan,A., Hock,R., Bustin,M. and Brown,D.T. (2000) Dynamic binding of histone H1 to chromatin in living cells. *Nature*, **408**, 877–881.
43. Orrego,M., Ponte,I., Roque,A., Buschati,N., Mora,X. and Suau,P. (2007) Differential affinity of mammalian histone H1 somatic subtypes for DNA and chromatin. *BMC Biol.*, **5**, 22.
44. Th'ng,J.P., Sung,R., Ye,M. and Hendzel,M.J. (2005) H1 family histones in the nucleus. Control of binding and localization by the C-terminal domain. *J. Biol. Chem.*, **280**, 27809–27814.
45. Mackey-Cushman,S.L., Gao,J., Holmes,D.A., Nunoya,J.I., Wang,R., Unutmaz,D. and Su,L. (2011) FoxP3 interacts with linker histone H1.5 to modulate gene expression and program Treg cell activity. *Genes Immun.*, **12**, 559–567.
46. Daujat,S., Zeissler,U., Waldmann,T., Happel,N. and Schneider,R. (2005) HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. *J. Biol. Chem.*, **280**, 38090–38095.
47. Hale,T.K., Contreras,A., Morrison,A.J. and Herrera,R.E. (2006) Phosphorylation of the linker histone H1 by CDK regulates its binding to HP1alpha. *Mol. Cell*, **22**, 693–699.
48. Hergeth,S.P., Dundr,M., Tropberger,P., Zee,B.M., Garcia,B.A., Daujat,S. and Schneider,R. (2011) Isoform-specific phosphorylation of human linker histone H1.4 in mitosis by the kinase Aurora B. *J. Cell Sci.*, **124**, 1623–1628.
49. Kamieniarz,K., Izzo,A., Dundr,M., Tropberger,P., Ozretic,L., Kirfel,J., Scheer,E., Tropel,P., Wisniewski,J.R., Tora,L. *et al.* (2012) A dual role of linker histone H1.4 Lys 34 acetylation in transcriptional activation. *Genes Dev.*, **26**, 797–802.
50. Kim,K., Choi,J., Heo,K., Kim,H., Levens,D., Kohno,K., Johnson,E.M., Brock,H.W. and An,W. (2008) Isolation and characterization of a novel H1.2 complex that acts as a repressor of p53-mediated transcription. *J. Biol. Chem.*, **283**, 9113–9126.
51. Kim,K., Jeong,K.W., Kim,H., Choi,J., Lu,W., Stallcup,M.R. and An,W. (2012) Functional interplay between p53 acetylation and H1.2 phosphorylation in p53-regulated transcription. *Oncogene*, **31**, 4290–4301.
52. Konishi,A., Shimizu,S., Hirota,J., Takao,T., Fan,Y., Matsuoka,Y., Zhang,L., Yoneda,Y., Fujii,Y., Skoultchi,A.I. *et al.* (2003) Involvement of histone H1.2 in apoptosis induced by DNA double-strand breaks. *Cell*, **114**, 673–688.
53. Kuzmichev,A., Jenuwein,T., Tempst,P. and Reinberg,D. (2004) Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol. Cell*, **14**, 183–193.
54. Lee,H., Habas,R. and Abate-Shen,C. (2004) MSX1 cooperates with histone H1b for inhibition of transcription and myogenesis. *Science*, **304**, 1675–1678.
55. Nishiyama,M., Oshikawa,K., Tsukada,Y., Nakagawa,T., Iemura,S., Natsume,T., Fan,Y., Kikuchi,A., Skoultchi,A.I. and Nakayama,K.I. (2009) CHD8 suppresses p53-mediated apoptosis through histone H1 recruitment during early embryogenesis. *Nat. Cell Biol.*, **11**, 172–182.
56. Vaquero,A., Scher,M., Lee,D., Erdjument-Bromage,H., Tempst,P. and Reinberg,D. (2004) Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. *Mol. Cell*, **16**, 93–105.
57. Weiss,T., Hergeth,S., Zeissler,U., Izzo,A., Tropberger,P., Zee,B.M., Dundr,M., Garcia,B.A., Daujat,S. and Schneider,R. (2010) Histone H1 variant-specific lysine methylation by G9a/KMT1C and Glp1/KMT1D. *Epigenetics Chromatin*, **3**, 7.
58. Bhan,S., May,W., Warren,S.L. and Sittman,D.B. (2008) Global gene expression analysis reveals specific and redundant roles for H1 variants, H1c and H1(0), in gene expression regulation. *Gene*, **414**, 10–18.

59. Sancho,M., Diani,E., Beato,M. and Jordan,A. (2008) Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. *PLoS Genet.*, **4**, e1000227.
60. Orthaus,S., Klement,K., Happel,N., Hoischen,C. and Diekmann,S. (2009) Linker histone H1 is present in centromeric chromatin of living human cells next to inner kinetochore proteins. *Nucleic Acids Res.*, **37**, 3391–3406.
61. Parseghian,M.H., Newcomb,R.L., Winokur,S.T. and Hamkalo,B.A. (2000) The distribution of somatic H1 subtypes is non-random on active vs. inactive chromatin: distribution in human fetal fibroblasts. *Chromosome Res.*, **8**, 405–424.
62. Krishnakumar,R., Gamble,M.J., Frizzell,K.M., Berrocal,J.G., Kininis,M. and Kraus,W.L. (2008) Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. *Science*, **319**, 819–821.
63. Braunschweig,U., Hogan,G.J., Pagie,L. and van Steensel,B. (2009) Histone H1 binding is inhibited by histone variant H3.3. *EMBO J.*, **28**, 3635–3645.
64. Li,J.Y., Patterson,M., Mikkola,H.K., Lowry,W.E. and Kurdistani,S.K. (2012) Dynamic distribution of linker histone H1.5 in cellular differentiation. *PLoS Genet.*, **8**, e1002879.
65. Cao,K., Lailler,N., Zhang,Y., Kumar,A., Uppal,K., Liu,Z., Lee,E.K., Wu,H., Medrzycki,M., Pan,C. *et al.* (2013) High-resolution mapping of h1 linker histone variants in embryonic stem cells. *PLoS Genet.*, **9**, e1003417.
66. Izzo,A., Kamieniarz-Gdula,K., Ramirez,F., Noureen,N., Kind,J., Manke,T., van Steensel,B. and Schneider,R. (2013) The genomic landscape of the somatic linker histone subtypes H1.1 to H1.5 in human cells. *Cell Rep.*, **3**, 1–13.
67. Strutt,H. and Paro,R. (1999) Mapping DNA target sites of chromatin proteins in vivo by formaldehyde crosslinking. *Methods Mol. Biol.*, **119**, 455–467.
68. Saldanha,A.J. (2004) Java Treeview–extensible visualization of microarray data. *Bioinformatics*, **20**, 3246–3248.
69. Li,R., Yu,C., Li,Y., Lam,T.W., Yiu,S.M., Kristiansen,K. and Wang,J. (2009) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, **25**, 1966–1967.
70. Zang,C., Schones,D.E., Zeng,C., Cui,K., Zhao,K. and Peng,W. (2009) A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics*, **25**, 1952–1958.
71. Zhu,L.J., Gazin,C., Lawson,N.D., Pages,H., Lin,S.M., Lapointe,D.S. and Green,M.R. (2010) ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics*, **11**, 237.
72. Shin,H., Liu,T., Manrai,A.K. and Liu,X.S. (2009) CEAS: cis-regulatory element annotation system. *Bioinformatics*, **25**, 2605–2606.
73. Guelen,L., Pagie,L., Brasset,E., Meuleman,W., Faza,M.B., Talhout,W., Eussen,B.H., de Klein,A., Wessels,L., de Laat,W. *et al.* (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, **453**, 948–951.
74. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
75. Trojer,P., Zhang,J., Yonezawa,M., Schmidt,A., Zheng,H., Jenuwein,T. and Reinberg,D. (2009) Dynamic histone H1 isotype 4 methylation and demethylation by histone lysine methyltransferase G9a/KMT1C and the jumonji domain-containing JMJD2/KDM4 proteins. *J. Biol. Chem.*, **284**, 8395–8405.
76. Schones,D.E., Cui,K., Cuddapah,S., Roh,T.Y., Barski,A., Wang,Z., Wei,G. and Zhao,K. (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell*, **132**, 887–898.
77. Kelly,T.K., Liu,Y., Lay,F.D., Liang,G., Berman,B.P. and Jones,P.A. (2012) Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.*, **22**, 2497–2506.
78. Kratzmeier,M., Albig,W., Meergans,T. and Doenecke,D. (1999) Changes in the protein pattern of H1 histones associated with apoptotic DNA fragmentation. *Biochem. J.*, **337(Pt. 2)**, 319–327.
79. Wang,Z.F., Sirotkin,A.M., Buchold,G.M., Skoultchi,A.I. and Marzluff,W.F. (1997) The mouse histone H1 genes: gene organization and differential regulation. *J. Mol. Biol.*, **271**, 124–138.
80. Barski,A., Cuddapah,S., Cui,K., Roh,T.Y., Schones,D.E., Wang,Z., Wei,G., Chepelev,I. and Zhao,K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
81. Jin,C., Zang,C., Wei,G., Cui,K., Peng,W., Zhao,K. and Felsenfeld,G. (2009) H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat. Genet.*, **41**, 941–945.
82. Luk,E., Ranjan,A., Fitzgerald,P.C., Mizuguchi,G., Huang,Y., Wei,D. and Wu,C. (2010) Stepwise histone replacement by SWR1 requires dual activation with histone H2A.Z and canonical nucleosome. *Cell*, **143**, 725–736.
83. Thakar,A., Gupta,P., Ishibashi,T., Finn,R., Silva-Moreno,B., Uchiyama,S., Fukui,K., Tomschik,M., Ausio,J. and Zlatanova,J. (2009) H2A.Z and H3.3 histone variants affect nucleosome structure: biochemical and biophysical studies. *Biochemistry*, **48**, 10852–10857.
84. Vicent,G.P., Nacht,A.S., Font-Mateu,J., Castellano,G., Gaveglia,L., Ballare,C. and Beato,M. (2011) Four enzymes cooperate to displace histone H1 during the first minute of hormonal gene activation. *Genes Dev.*, **25**, 845–862.
85. Wright,R.H., Castellano,G., Bonet,J., Le Dily,F., Font-Mateu,J., Ballare,C., Nacht,A.S., Soronellas,D., Oliva,B. and Beato,M. (2012) CDK2-dependent activation of PARP-1 is required for hormonal gene regulation in breast cancer cells. *Genes Dev.*, **26**, 1972–1983.
86. Krishnakumar,R. and Kraus,W.L. (2010) PARP-1 regulates chromatin structure and transcription through a KDM5B-dependent pathway. *Mol. Cell*, **39**, 736–749.
87. Garber,M., Yosef,N., Goren,A., Raychowdhury,R., Thielke,A., Guttman,M., Robinson,J., Minie,B., Chevrier,N., Itzhaki,Z. *et al.* (2012) A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol. Cell*, **47**, 810–822.
88. Melcer,S., Hezroni,H., Rand,E., Nissim-Rafinia,M., Skoultchi,A., Stewart,C.L., Bustin,M. and Meshorer,E. (2012) Histone modifications and lamin A regulate chromatin protein dynamics in early embryonic stem cell differentiation. *Nat. Commun.*, **3**, 910.
89. Meshorer,E., Yellajoshula,D., George,E., Scambler,P.J., Brown,D.T. and Misteli,T. (2006) Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell*, **10**, 105–116.
90. Goldberg,A.D., Banaszynski,L.A., Noh,K.M., Lewis,P.W., Elsaesser,S.J., Stadler,S., Dewell,S., Law,M., Guo,X., Li,X. *et al.* (2010) Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell*, **140**, 678–691.
91. McDonald,O.G., Wu,H., Timp,W., Doi,A. and Feinberg,A.P. (2011) Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition. *Nat. Struct. Mol. Biol.*, **18**, 867–874.
92. Chow,K.H., Factor,R.E. and Ullman,K.S. (2012) The nuclear envelope environment and its cancer connections. *Nat. Rev. Cancer*, **12**, 196–209.
93. Kulis,M. and Esteller,M. (2010) DNA methylation and cancer. *Adv. Genet.*, **70**, 27–56.
94. Cui,P., Zhang,L., Lin,Q., Ding,F., Xin,C., Fang,X., Hu,S. and Yu,J. (2010) A novel mechanism of epigenetic regulation: nucleosome-space occupancy. *Biochem. Biophys. Res. Commun.*, **391**, 884–889.
95. Izzo,A., Kamieniarz-Gdula,K., Ramirez,F., Noureen,N., Kind,J., Manke,T., van Steensel,B. and Schneider,R. (2013) The genomic landscape of the somatic linker histone subtypes H1.1 to H1.5 in human cells. *Cell Rep.*, **3**, 2412–2154.
96. Yang,S.M., Kim,B.J., Norwood Toro,L. and Skoultchi,A.I. (2013) H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation. *Proc. Natl Acad. Sci. USA*, **110**, 1708–1713.
97. Hechtman,J.F., Beasley,M.B., Kinoshita,Y., Ko,H.M., Hao,K. and Burstein,D.E. (2013) Promyelocytic leukemia zinc finger and histone H1.5 differentially stain low- and high-grade pulmonary neuroendocrine tumors: a pilot immunohistochemical study. *Hum. Pathol.*, **44**, 1400–1405.

98. Ruike,Y., Imanaka,Y., Sato,F., Shimizu,K. and Tsujimoto,G. (2010) Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. *BMC Genomics*, **11**, 137.
99. Hon,G.C., Hawkins,R.D., Caballero,O.L., Lo,C., Lister,R., Pelizzola,M., Valsesia,A., Ye,Z., Kuan,S., Edsall,L.E. *et al.* (2012) Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.*, **22**, 246–258.
100. Popova,E.Y., Grigoryev,S.A., Fan,Y., Skoultchi,A.I., Zhang,S.S. and Barnstable,C.J. (2013) Developmentally regulated linker histone H1c promotes heterochromatin condensation and mediates structural integrity of rod photoreceptors in mouse retina. *J. Biol. Chem.*, **288**, 17895–17907.
101. Th'ng,J.P., Sung,R., Ye,M. and Hendzel,M.J. (2005) H1 family histones in the nucleus. Control of binding and localization by the C-terminal domain. *J. Biol. Chem.*, **280**, 27809–27814.
102. Orrego,M., Ponte,I., Roque,A., Buschati,N., Mora,X. and Suau,P. (2007) Differential affinity of mammalian histone H1 somatic subtypes for DNA and chromatin. *BMC Biol.*, **5**, 22.
103. Parseghian,M.H., Newcomb,R.L. and Hamkalo,B.A. (2001) Distribution of somatic H1 subtypes is non-random on active vs. inactive chromatin II: distribution in human adult fibroblasts. *J. Cell. Biochem.*, **83**, 643–659.
104. Dixon,J.R., Selvaraj,S., Yue,F., Kim,A., Li,Y., Shen,Y., Hu,M., Liu,J.S. and Ren,B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
105. Bickmore,W.A. and van Steensel,B. (2013) Genome architecture: domain organization of interphase chromosomes. *Cell*, **152**, 1270–1284.
106. Ernst,J., Kheradpour,P., Mikkelsen,T.S., Shoresh,N., Ward,L.D., Epstein,C.B., Zhang,X., Wang,L., Issner,R., Coyne,M. *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, **473**, 43–49.
107. Filion,G.J., van Bemmel,J.G., Braunschweig,U., Talhout,W., Kind,J., Ward,L.D., Brugman,W., de Castro,I.J., Kerkhoven,R.M., Bussemaker,H.J. *et al.* (2010) Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell*, **143**, 212–224.

Supplementary Data

Figure S1. Expression of HA-tagged somatic histone H1 variants in breast cancer cells. Chromatin was prepared from T47D-derived cells stably expressing HA-tagged H1 variants, wild-type or a K26A mutant of H1.4, and loaded into a 10% SDS-PAGE. Western blot hybridization was performed with anti-HA antibody (A), or with H1 variant-specific antibodies (B).

Figure S2. Specificity of the anti-HA antibody. ChIP was performed in cells expressing H1-HA or mock infected with the empty retroviral expression vector with anti-HA antibody or unrelated immunoglobulins (IgG), and the abundance of IPed material was quantified by qPCR with ß-actin promoter oligonucleotides. A representative experiment performed in triplicate is shown.

Figure S3. HA-tagged H1 variants are associated with gene promoters, coding regions and repetitive DNA. ChIP was performed in cells expressing H1-HA with anti-HA antibody and the abundance of IPed material was quantified by qPCR with oligonucleotides for the indicated promoters, coding or heterochromatic regions, and corrected by input DNA amplification with the same primer pair. A representative experiment performed in triplicate is shown.

Figure S4. Specificity of H1.2 and H1X antibodies in ChIP experiments determined in H1 variant-specific knock-down cells. T47D-derived cells stably harboring an inducible system for shRNA expression against H1.2 or H1X expression, were treated with doxycycline for 6 days or left untreated. Then, ChIP was performed with H1 variant-specific antibodies against H1.2 and H1X (or control IgG), and the IPed material was quantified by qPCR with oligonucleotides for the indicated promoters (-10 kb distal promoter or TSS), and corrected by input DNA amplification. A representative experiment performed in triplicate is shown. Western blot hybridizations showing the rate of H1 depletion upon doxycycline treatment are shown.

Figure S5. An H1 valley at TSS is found at genes being expressed that show FAIRE-measured open chromatin and increased H3K4me3 at TSS. Several genes representing different levels of expression according to microarrays data were chosen to analyze

mRNA abundance by RT-qPCR, chromatin accessibility at TSS by FAIRE-qPCR, and distribution of H1, H3 and H3K4me3 by ChIP at distal (upstream) promoter compared to TSS. ND, not determined. RT-PCR values for each gene were normalized to GAPDH expression and genomic DNA amplification with the same set of primers. ChIP and FAIRE PCRs were normalized to input DNA. A representative experiment performed in triplicate is shown.

Figure S6. H1 depletion at TSS of hormone-responsive promoters upon stimulation with ligand. T47D-derived cells harboring an MMTV-luciferase construct and expressing different H1-HA were treated with a progestin (R5020 10nM) for the indicated time and ChIP was performed with anti-HA antibody. The abundance of IPed material was quantified by qPCR with specific oligonucleotides for the MMTV promoter (nucleosome B). A representative experiment performed in triplicate is shown.

Figure S7. H1 depletion at TSS of inducible promoters. The H1 valley at TSS of JUN and FOS genes is preformed and increases upon mitogenic stimulation. T47D cells were treated with PMA 100nM for 60 min or left untreated and ChIP was performed with H1, H1.2, H1X and H3 antibodies. The abundance of IPed material was quantified by qPCR with oligonucleotides for the indicated promoters (-10 kb distal promoter or TSS), and corrected by input DNA amplification. JUN and FOS were responsive to PMA as shown in the RT-qPCR experiment (right panels). PSMB4 and OCT4 are non-responsive control genes, active and repressed, respectively, in T47D cells. The table below shows the H1 valley ratio (calculated as distal/TSS) at 0 and 60 min, or the relative H1 valley ratio at 60 min compared to 0 min. A representative experiment performed in triplicate is shown.

Figure S8. Global gene expression profile in T47D cells. (A) Overlap between unique transcript IDs at the Nimblegen promoter array and Agilent expression microarray. (B) Gene expression profile in T47D cells of 62,976 transcripts, from highest to lowest expression, obtained by hybridization with an Agilent microarray.

Figure S9. Some repressed genes show an H1.2 valley at the TSS. ChIPed material from H1.0-HA cells or with the specific H1.2 antibody was quantified by qPCR with oligonucleotides for the indicated promoter regions and corrected by input DNA

amplification. Selected genes belong to the indicated expression profile percentiles. A representative experiment performed in triplicate is shown.

Figure S10. Non protein-coding transcripts show an H1.2 valley at TSS. (A) Heat maps of ChIP-chip probe intensity around TSS (-3200 to +800 bp) for 1,145 non protein-coding transcripts (NRs) from which the expression rate was determined (B). NRs are ordered from highest to lowest gene expression. (B) Expression levels of NRs are shown as a box plot compared to total transcriptome included in the Agilent expression microarray. Significance was tested using the Kolmogorov-Smirnov test. p-value < 2.2e-16.

Figure S11. Correlation between the abundance of different H1 variants at distal promoter regions.
(A) Heat map and dendrogram of the Pearson's correlation coefficient between all H1 variants ChIP-chip samples.
(B) Scatter plots of the abundance of different H1 variants at distal promoter regions. X and Y axis represent mean probe intensity at distal promoter regions (-3200 to -2000 bp relative to TSS), for the indicated H1 variants. R: Pearson's correlation coefficient.

Figure S12. Coincidence between genes presenting the highest or lowest H1.2 or H1X content at distal promoter. (A) Heat maps of H1 ChIP-chip probe intensity around TSS. Genes are ordered from lowest to highest H1.2 (left) or H1X (right) content at distal promoter regions. Genes with the top or lowest distal H1 content are indicated. These genes (2050 genes for each group, 10% of the total) were used to determine the number of coinciding genes as shown in Venn diagrams (B). A streaking coincidence exist between genes presenting few H1.2 but high H1X.
(C) Expression levels of coinciding genes in the four comparisons depicted in (B). Genes were classified in five expression groups (EG) similar to Figure 2, and the percentage of coinciding genes belonging to each of these groups was determined. (Right panel) Expression levels of coinciding genes is also shown as a box plot. Significance was tested using the Kolmogorov-Smirnov test. Enrichment and depletion is marked with red and blue asterisks, respectively. ** p-value<0,001 and * p-value<0,005.

Figure S13. Coincidence between genes presenting highest or lowest H1.2 or H1.0-HA content at distal promoter. See Figure S12 legend.

Figure S14. The ratio between H1.2 and H1X abundance at selected genes is conserved among T47D and HeLa cells, but not in MCF7, in relation to the cell abundance of each variant.
(A) ChIP-qPCR of H1.2 and H1X abundance at TMEM204, TUBGCP5, COL4A3 and CUGBP2 distal promoter regions in T47D, HeLa and MCF7 cells. A representative experiment performed in triplicate is shown.
(B) Abundance of H1 variants in T47D, HeLa and MCF7 chromatin determined by immunoblot with specific antibodies.
(C) Expression of H1 variants in T47D, HeLa and MCF7 cells determined by RT-qPCR. cDNA levels for the indicated H1 variants were corrected by GAPDH expression and amplification of genomic DNA with the same PCR primers. A representative experiment performed in triplicate is shown.

Figure S15. Comparison of H1.2 and H1X promoter content among T47D and HeLa cell lines. Heat maps of H1.2 and H1X ChIP-chip probe intensity around TSS (-3200 to +800 bp) for 20,338 transcripts. Genes in all heat maps shown are ordered from lowest to highest H1.2 content at distal promoter regions in T47D cells.

Figure S16. H1 variant abundance around TSS for genes ordered according to their position along selected human chromosomes.
(A) Cluster dendrogram of all human chromosomes depending on their gene richness coefficient and gene expression. The gene-richness coefficient for each chromosome was calculated as described in Figure 4. Average gene expression for individual chromosomes was obtained from microarray data obtained in T47D cells. GRC≥2 are shown in the same color.
(B) Heat maps of H1.2, H1.0-HA and H1X ChIP-chip probe intensity around TSS (-3.2 to +0.8 kbp) for genes ordered according to their position along selected gene-poor or gene-rich chromosomes in T47D cells.
(C) Heat maps of H1.2 and H1X ChIP-chip probe intensity around TSS (-3.2 to +0.8 kbp) for genes ordered according to their position along chromosome 1 in T47D and HeLa cells. Heat map of H1.0-HA in T47D is also included.

Figure S17. Occupancy of H1 variants at regions enriched for different genome features and histone marks.
(A) Box plots showing the occupancy of H1 variants (average, input-subtracted ChIP-Seq signal) at regions enriched for the indicated genome features: DNase hypersensitivity sites (data from T47D cells), FAIRE regions (HeLa data), CTCF and p300 binding sites (T47D data).
(B) Box plots showing the occupancy of H1 variants (average, input-subtracted ChIP-Seq signal) at regions enriched for the indicated histone marks (data from HeLa cells).
(C) Box plots showing the occupancy of H1 variants (average, input-subtracted ChIP-Seq signal) at CpG islands. Significance was tested using the Kolmogorov-Smirnov test taking as a control a random sample of windows with equal width to the indicated histone mark. Enrichment and depletion is marked with red and blue asterisks, respectively. * p-value<0,001.

Figure S18. Distribution of H1 variants along selected chromosome regions containing LADs. Input-subtracted H1 variants and H3 ChIP-seq data viewed in the UCSC genome browser together with GC content, RefSeq genes, H3K4me3 (ENCODE average of 9 cell lines), CpG and LADs.

Figure S19. Distribution of H1 variants along the entire length of chromosomes 3, 4, 10, 13, 17 and 21. Input-subtracted H1 variants and H3 ChIP-seq signal viewed in the UCSC genome browser together with GC content, RefSeq genes, H3K4me3 (ENCODE average of 9 cell lines), CpG and LADs.

Figure S20. GC content of H1-enriched chromatin. (A) Genome-wide correlation scatter plots of H1 variants versus GC content. X axes: average input-subtracted H1 variants and H3 ChIP-seq signal (normalized to 1000bp window). Y axes: GC%. R: Pearson's correlation coefficient. (B) Box plot of the percentage of GC in regions found enriched and depleted for each H1 variant.

Figure S21. Correlation scatter plots between the occupancy of H1 variants at all chromosomes and the gene richness coefficient or average gene expression. Correlation

between gene richness coefficient and gene expression of all chromosomes is also shown. R: Pearson's correlation coefficient.

Figure S22. Annotation of H1 enriched and depleted regions to promoters, genes and intergenic regions and overlap with CpG islands. (A) H1 variants enriched and depleted regions over input were mapped to promoters (defined as -5 kb to +1 kb from TSS), genes (-5kb from TSS to +3kb from TTS) or intergenic regions (rest of the genome). The percentages of identified regions for the different variants falling into each of these three categories are represented. Notice that 100% is the sum of genic and intergenic regions. The theoretical occupancy of these compartments in the UCSC genome is shown as percentages, as well as occupancy of total ChIP-seq input reads (right panel). (B) Co-localization of H1 enriched or depleted regions with CpG islands expressed as percentage of total H1 regions overlapping CpG sites. Areas of enrichment or depletion of H1 variants compared to input derived from ChIP-seq data with a fold-change ≥ 2 were considered.

Figure S23. Co-localization of H1 enriched regions with CpG islands. Distribution of H1 variants along selected regions of the genome containing CgG sites and H1 enriched regions identified using SICER software. Input-subtracted H1 variants and H3 ChIP-seq signal viewed in the UCSC genome browser together with GC content, RefSeq genes, H3K4me3 (ENCODE average of 9 cell lines) and CpG islands.

Figure S24. Venn diagrams showing the overlap between genes containing enriched or depleted regions of the different H1 variants. H1 enriched or depleted regions mapping within genes (-5kb from TSS to +3kb from TTS) were used to identify target genes. Comparisons between different H1 variants are shown in two clusters for clarity coinciding with different ChIP-seq experiments. The expression profiles of target genes containing enriched or depleted regions for a unique variant are shown as box plots in Figure 7.

Figure S25. Genomic annotation of enriched or depleted regions of individual H1 variants. Pie diagram of the distribution of H1 variants enriched regions at genes, proximal regulatory regions, and distal intergenic regions. Promoter and downstream regions are defined as 3,000bp upstream TSS or downstream TTS, respectively. As a

control, the annotation of all genome base pairs is shown. Two replicas of endogenous H1.2 ChIP-seq experiments are shown.

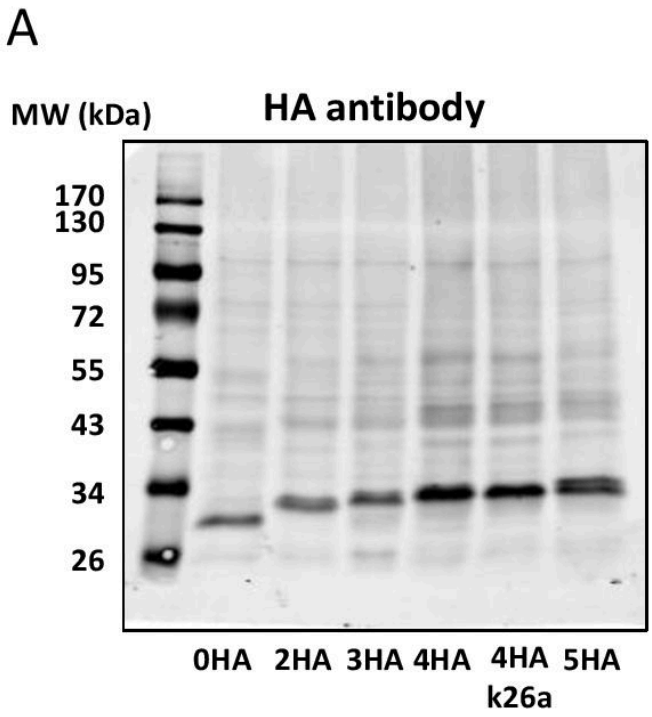
Table S1. Summary of samples analyzed by ChIP-seq in three independent experiments (r1, r2, r3; replica 1, 2 and 3). Read length, number of total reads obtained, and number of mappable reads to the human genome version hg18, as well as mapped rate, are shown.

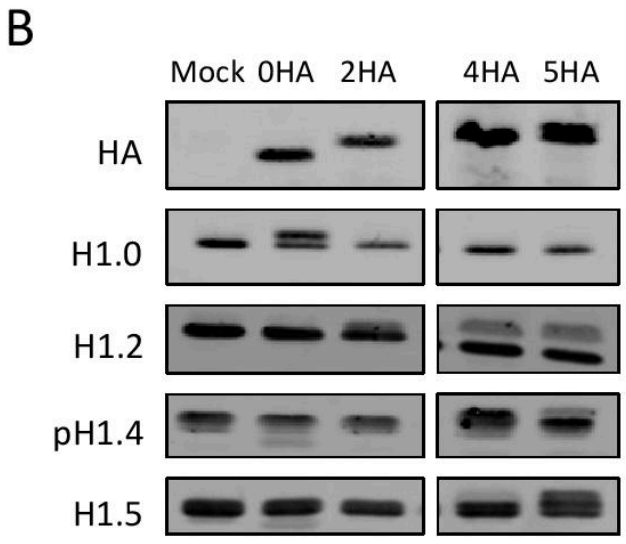
Table S2. Gene ontology of genes presenting the highest (top 10%) or lowest (bottom 10%) H1.2 or H1X content at distal promoter (-3200 to -2000 bp relative to TSS) according to ChIP-chip data shown in Figure 3. P-value (adjusted for multiple testing by Benjamini method) and false discovery rate are shown.

Table S3. Summary of enriched or depleted regions of individual H1 variants and its target genes. Areas of enrichment or depletion of H1 variants compared to input derived from ChIP-seq data with a fold-change equal or greater than 2 were considered. Genes were defined as comprised between -5kb from TSS to +3kb from TTS, and promoters from -5kb upstream TSS to +1kb downstream TSS.

Supplementary Methods. Cell treatments. Antibodies. H1 extraction, gel electrophoresis and immunoblotting. FAIRE assays. RNA extraction and RT-PCR.

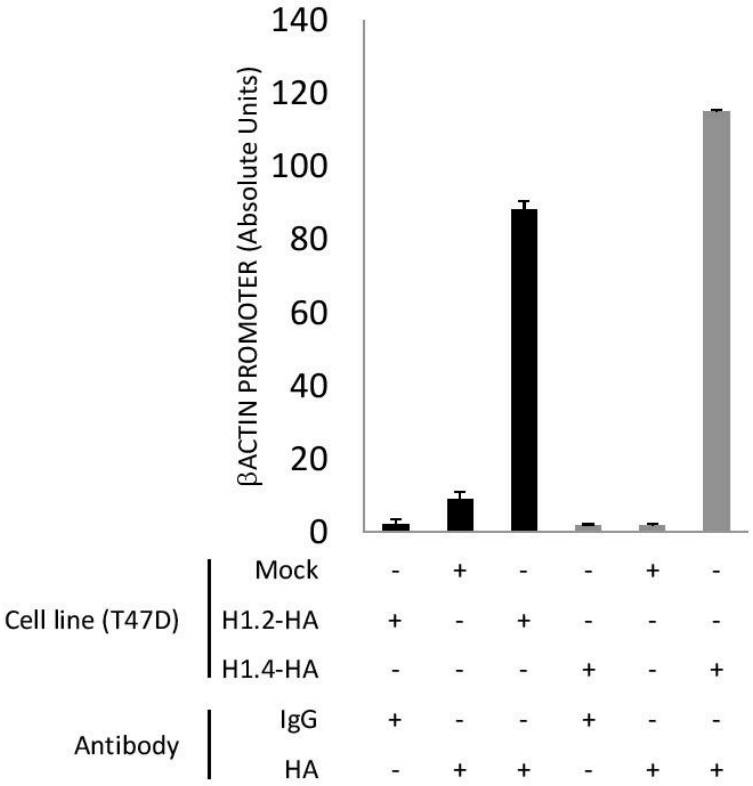
Figure S1





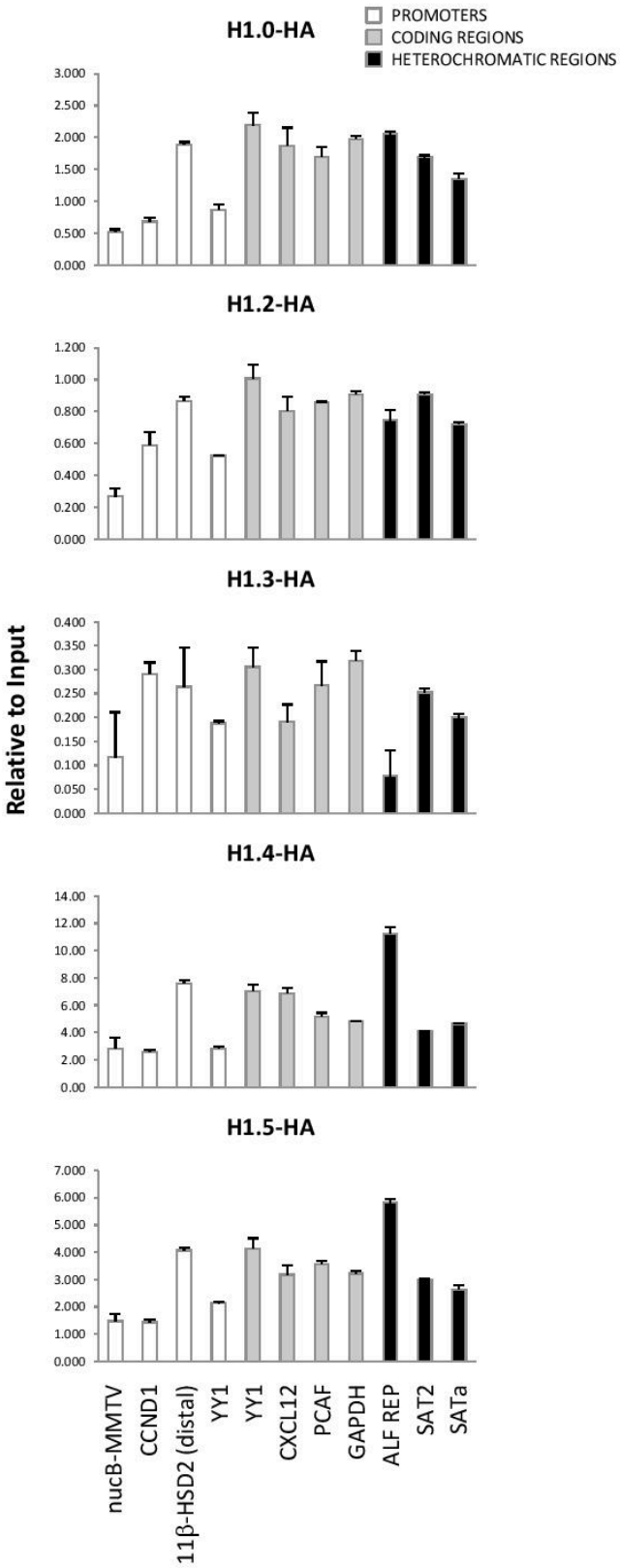
Millán-Ariño, Ll., et al.

Figure S2



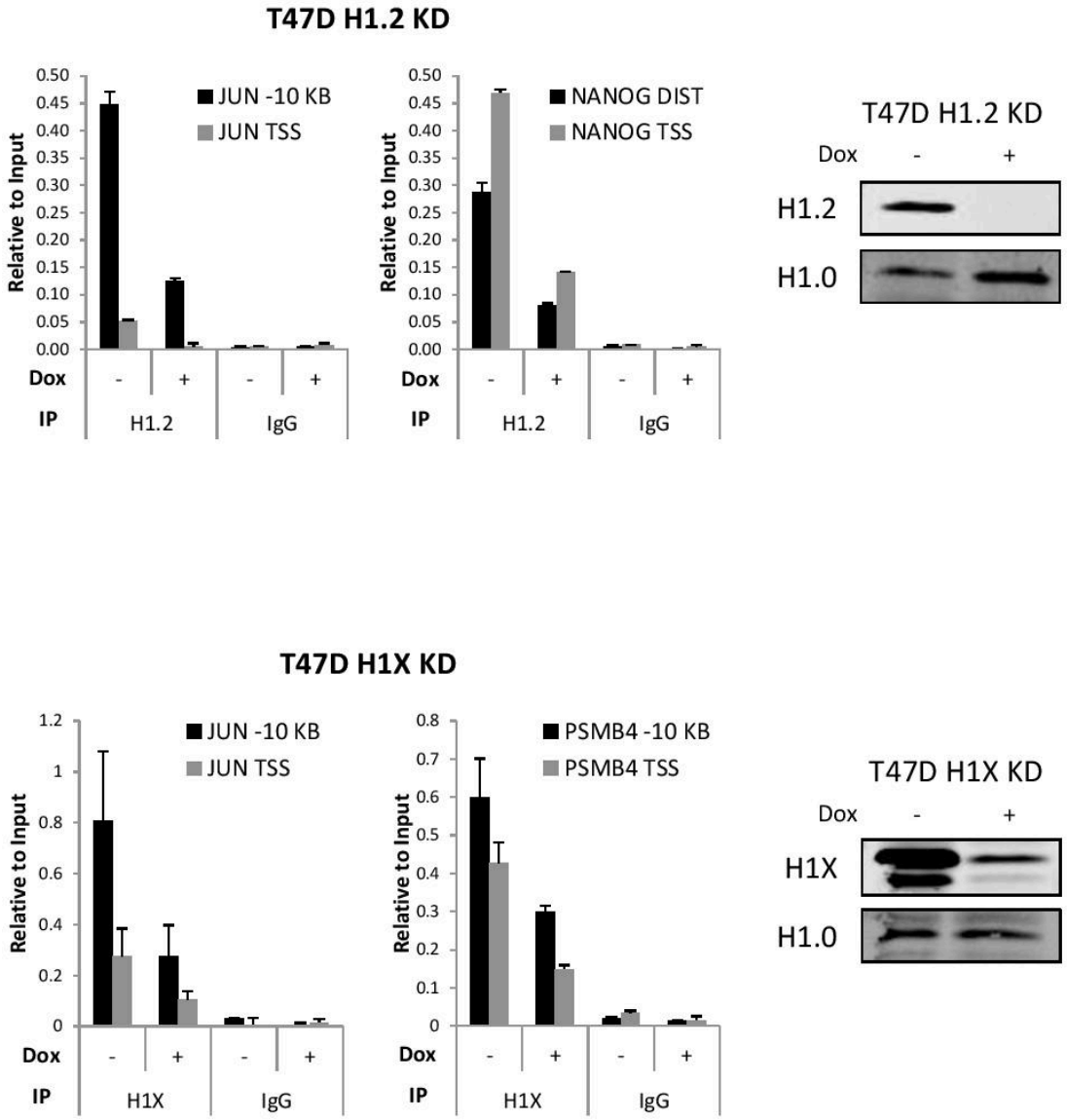
Millán-Ariño, Ll., et al.

Figure S3



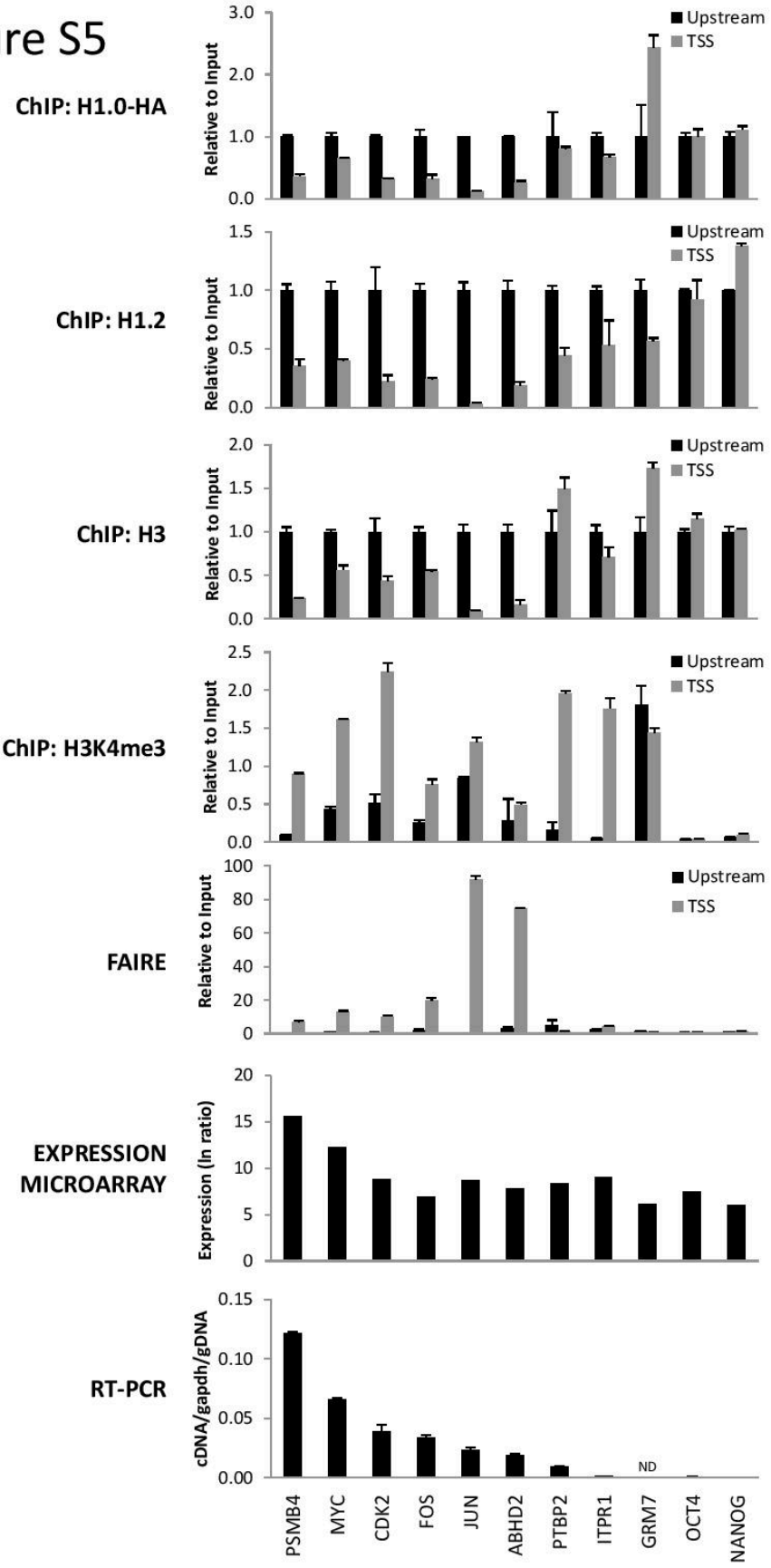
Millán-Ariño, Ll., et al.

Figure S4



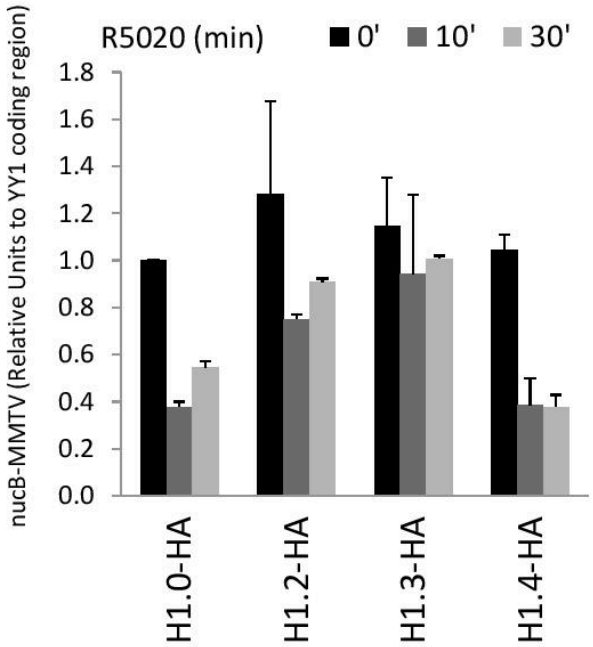
Millán-Ariño, Ll., et al.

Figure S5



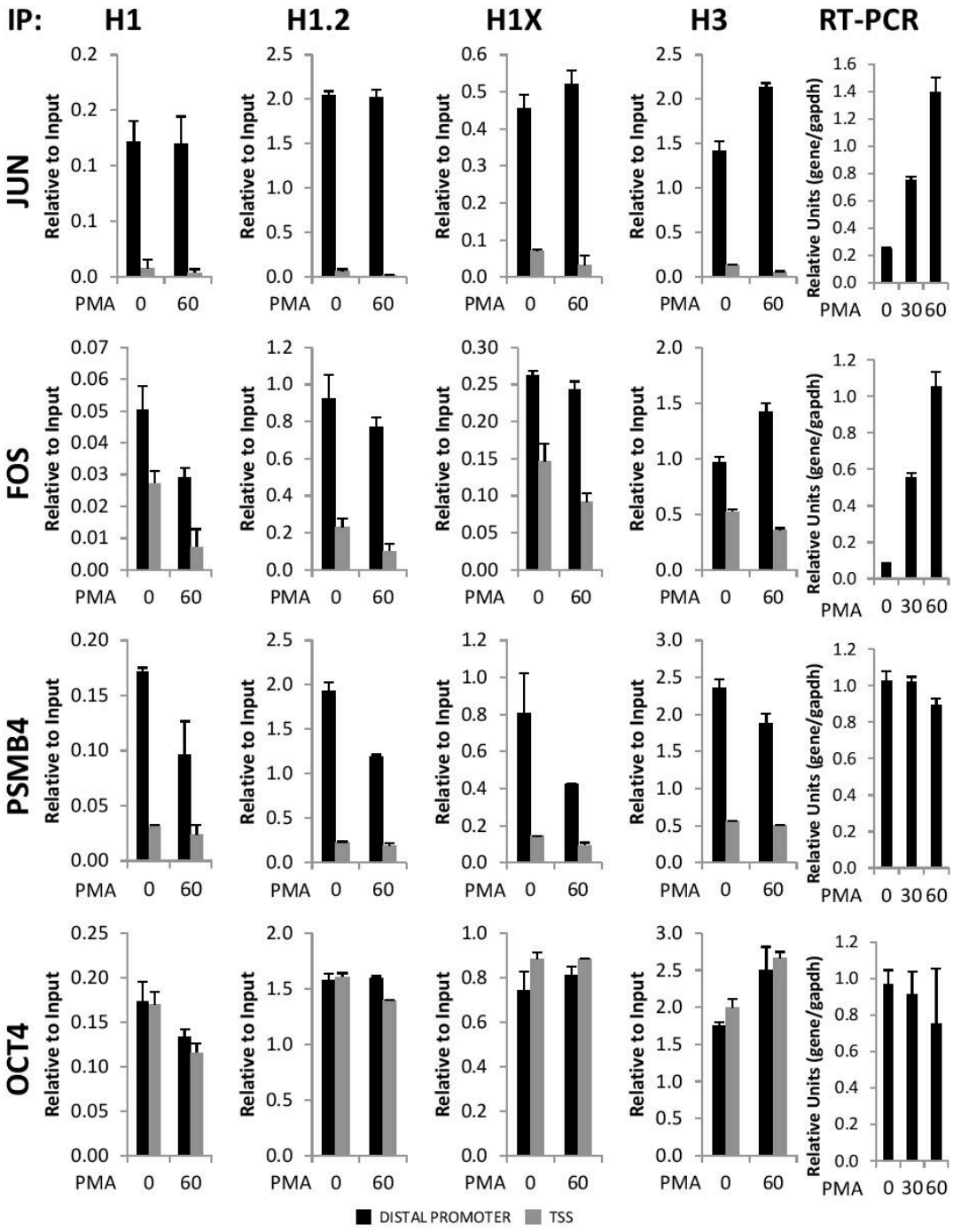
Millán-Ariño, Ll., et al.

Figure S6



Millán-Ariño, Ll., et al.

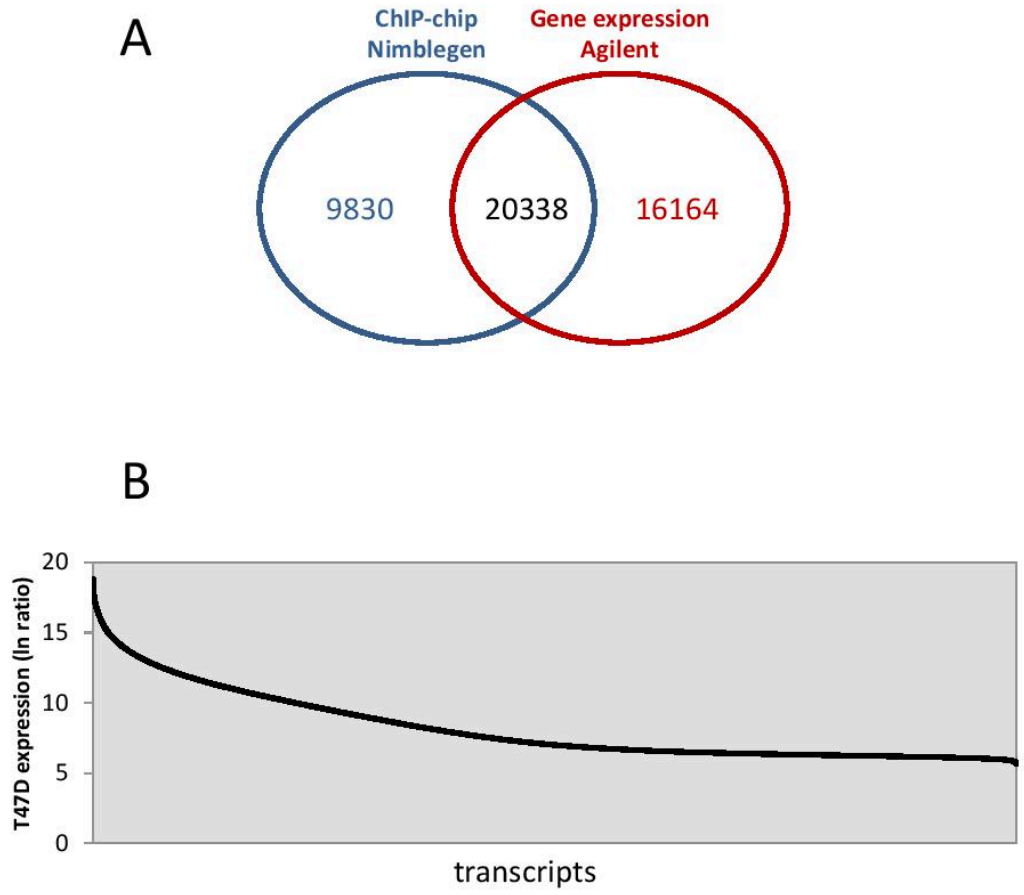
Figure S7



	IP	H1		H1.2		H1X		H3	
	PMA (min)	0	60	0	60	0	60	0	60
JUN	valley Fc	15,18	29,64	30,68	115,42	6,41	16,01	10,56	39,45
	relative Fc	1,00	1,95	1,00	3,76	1,00	2,50	1,00	3,73
FOS	valley Fc	1,84	4,02	4,01	7,49	1,79	2,65	1,83	3,94
	relative Fc	1,00	2,18	1,00	1,87	1,00	1,48	1,00	2,15
PSMB4	valley Fc	5,45	4,06	8,70	6,30	5,59	4,46	4,20	3,76
	relative Fc	1,00	0,75	1,00	0,72	1,00	0,80	1,00	0,89
OCT4	valley Fc	1,02	1,15	0,98	1,14	0,84	0,92	0,87	0,94
	relative Fc	1,00	1,12	1,00	1,16	1,00	1,09	1,00	1,07

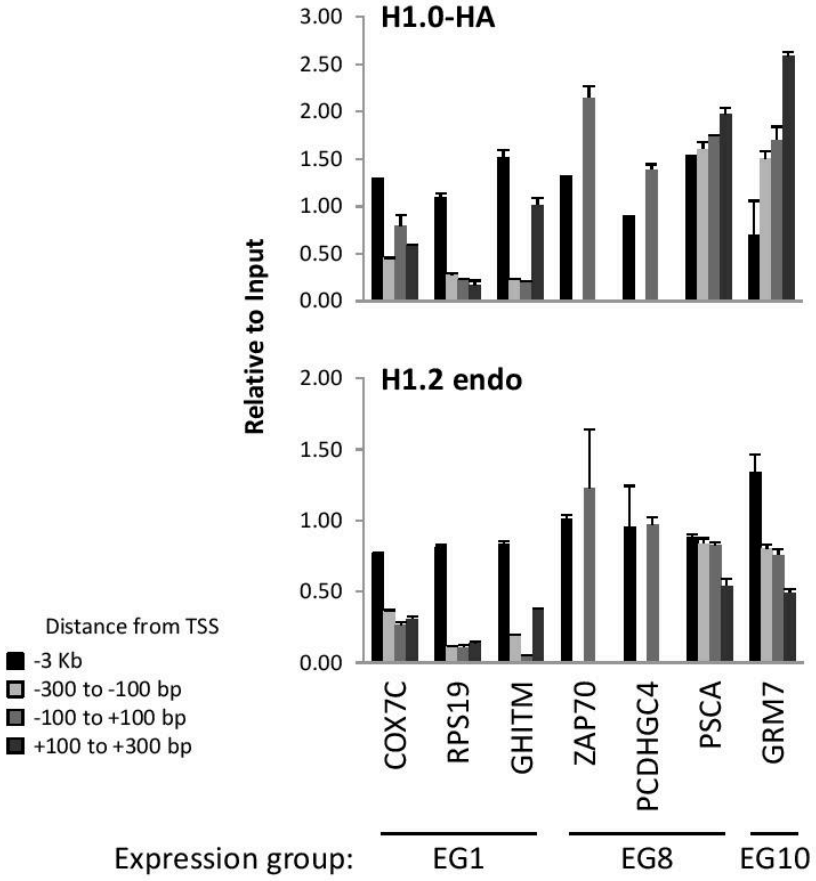
Millán-Ariño, Ll., et al.

Figure S8



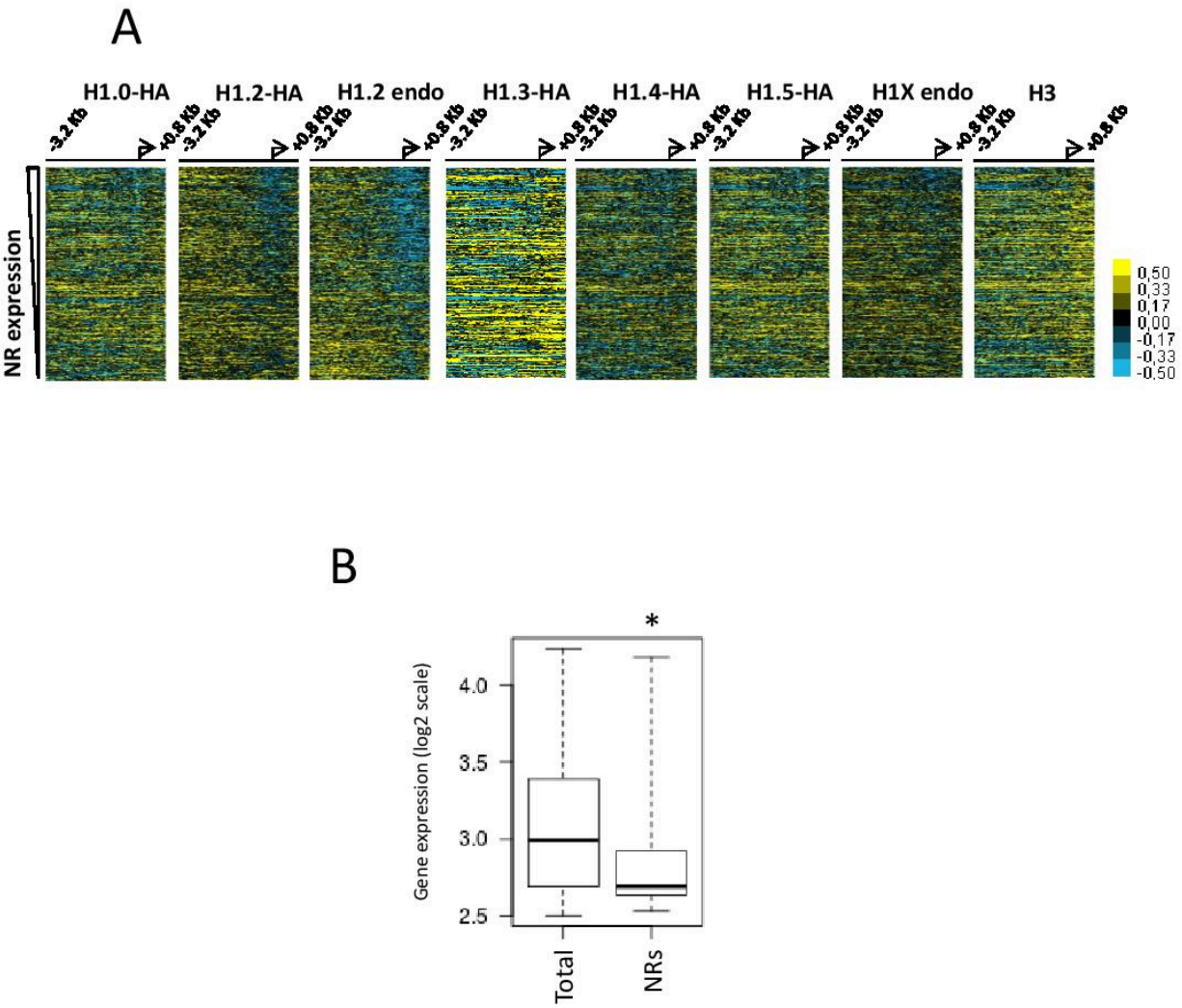
Millán-Ariño, Ll., et al.

Figure S9



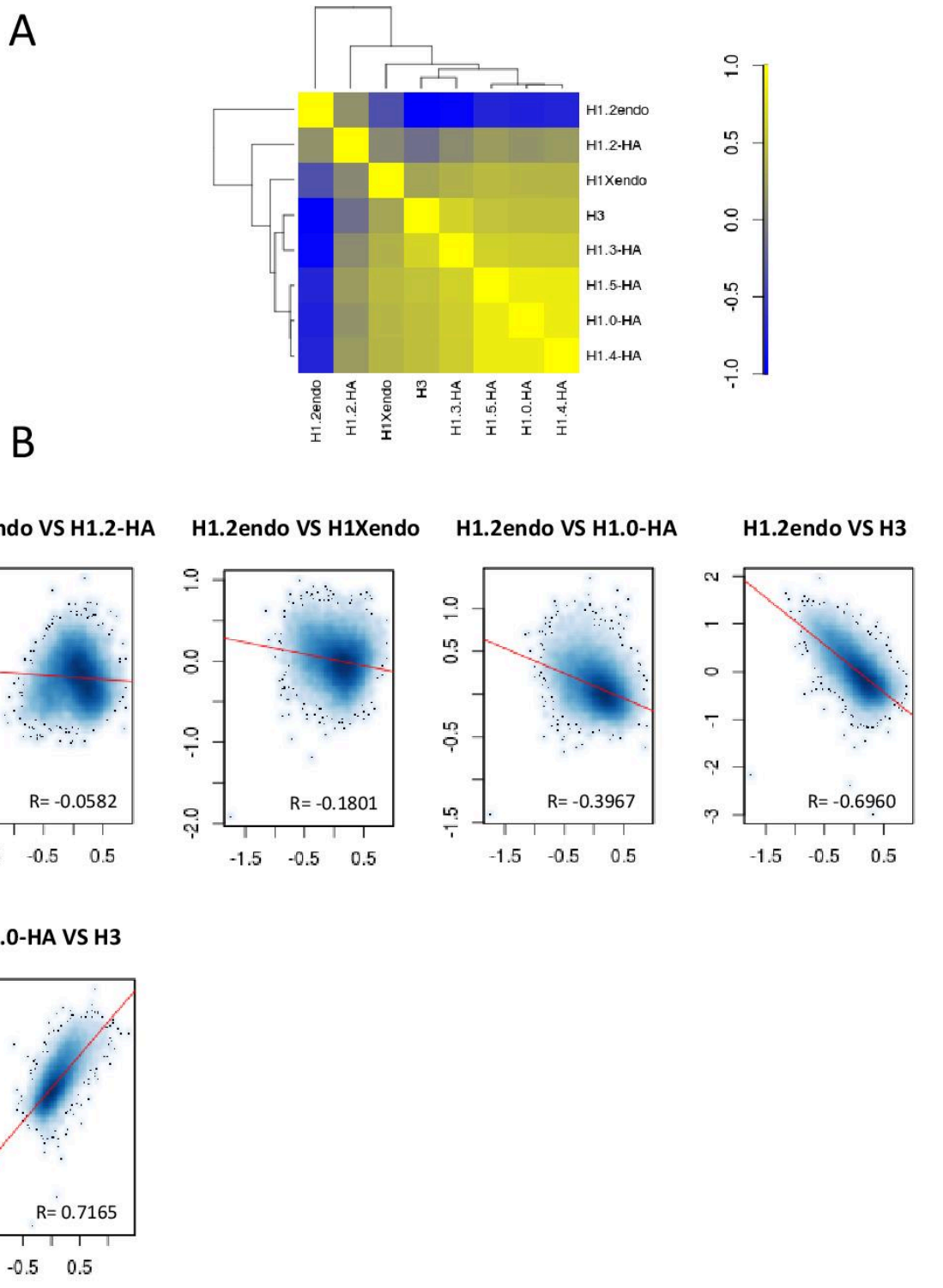
Millán-Ariño, Ll., et al.

Figure S10



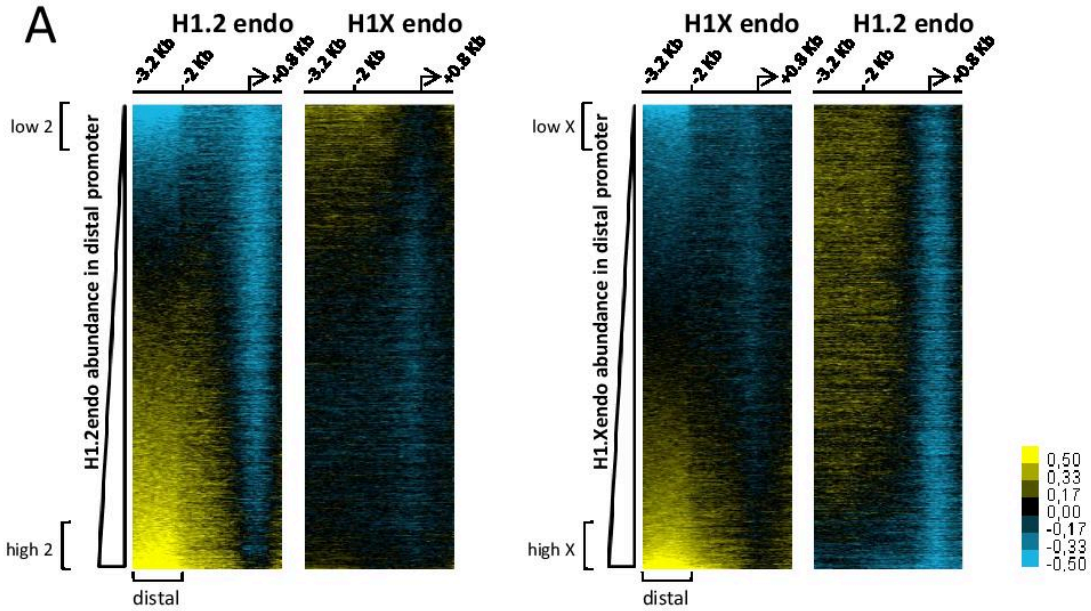
Millán-Ariño, Ll., et al.

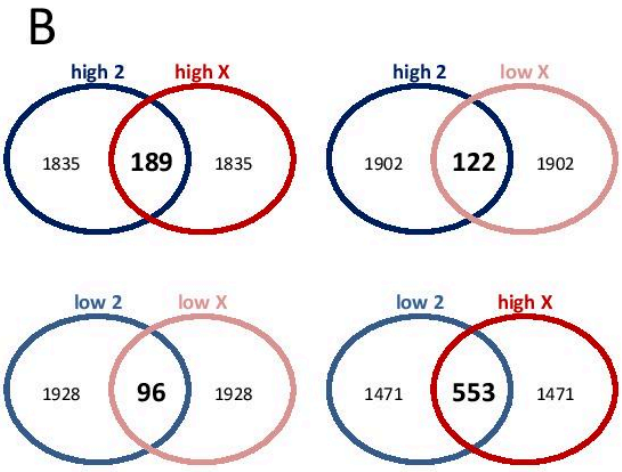
Figure S11

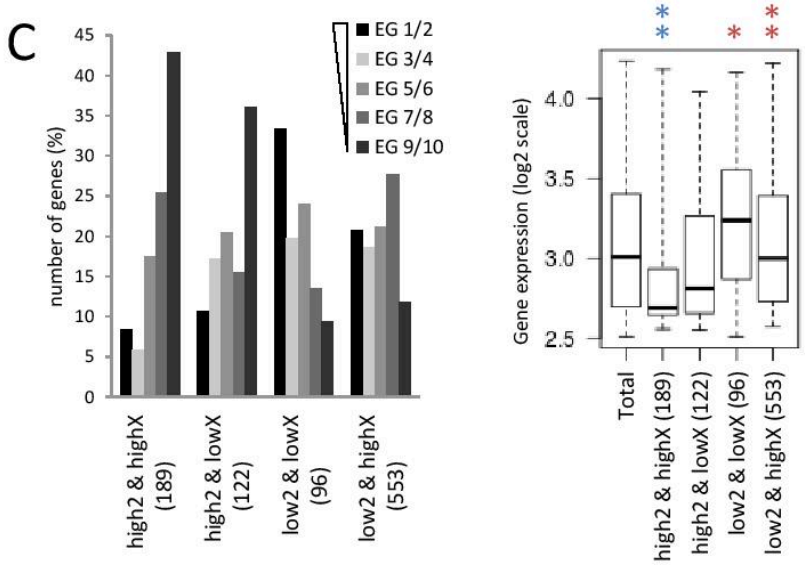


Millán-Ariño, Ll., et al.

Figure S12

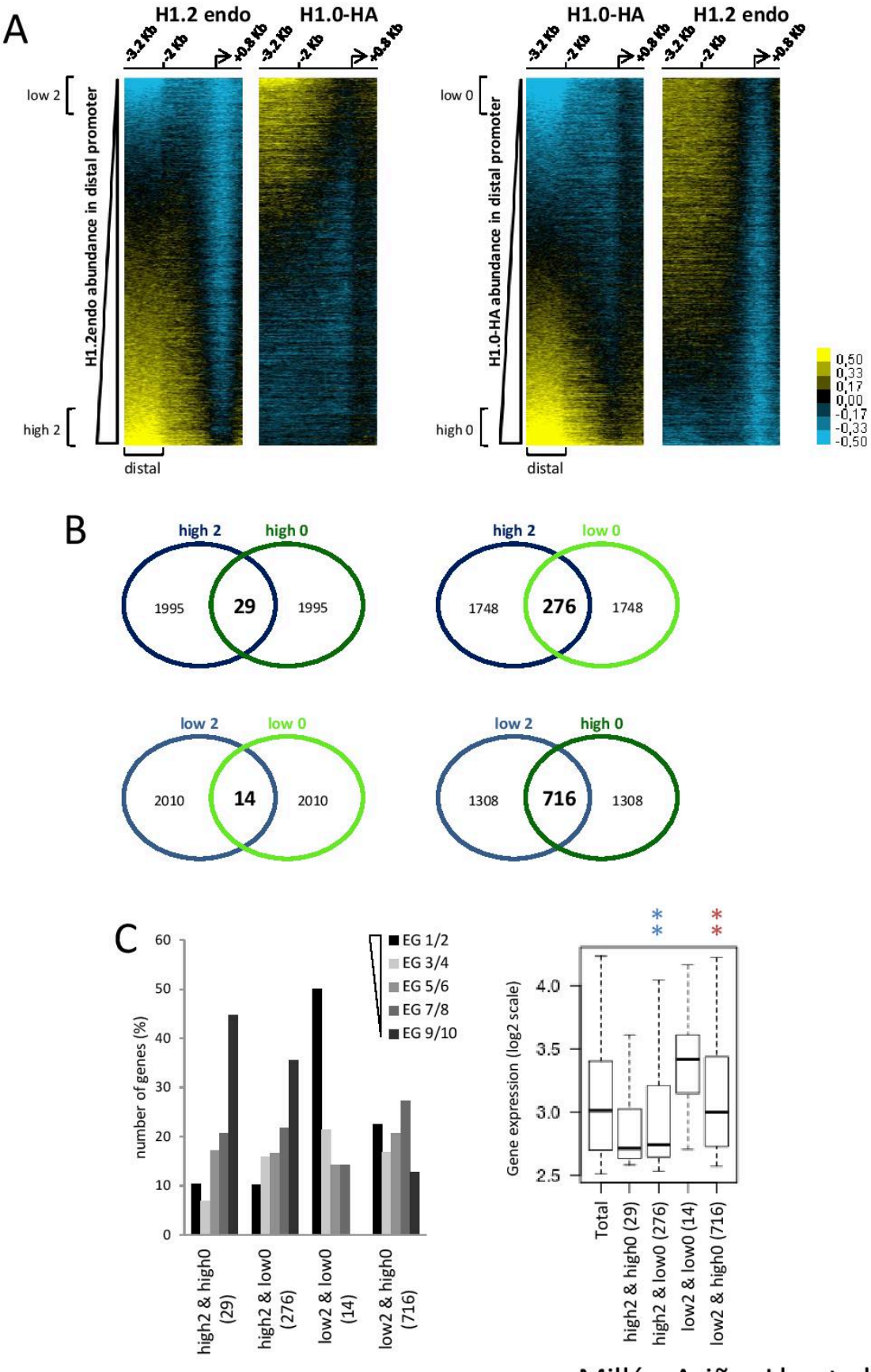






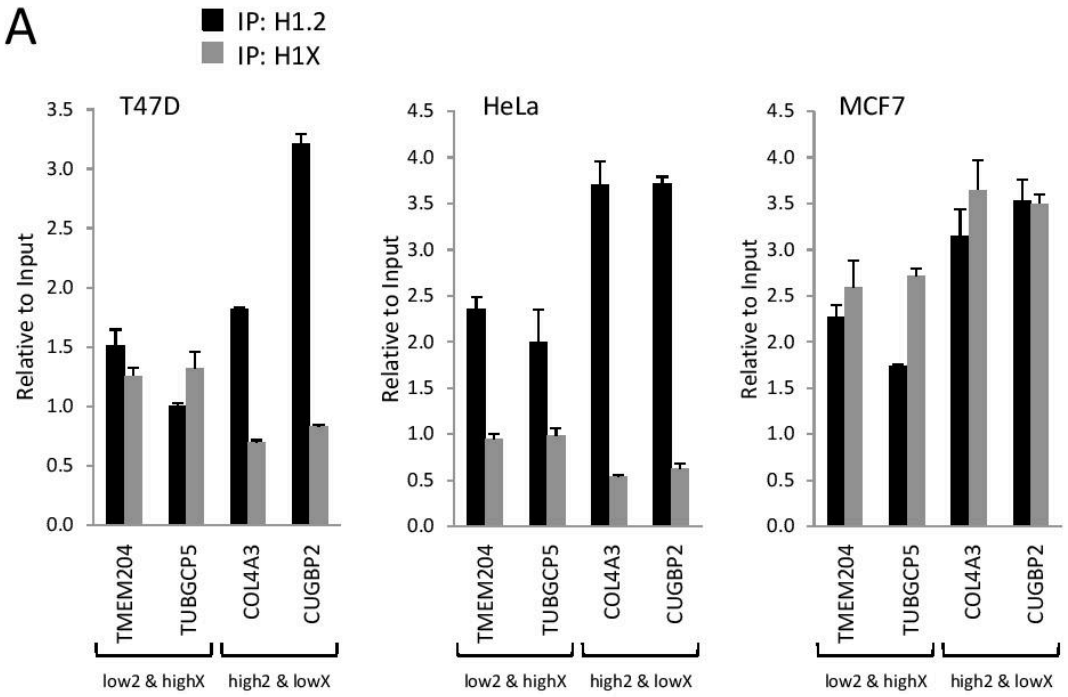
Millán-Ariño, Ll., et al.

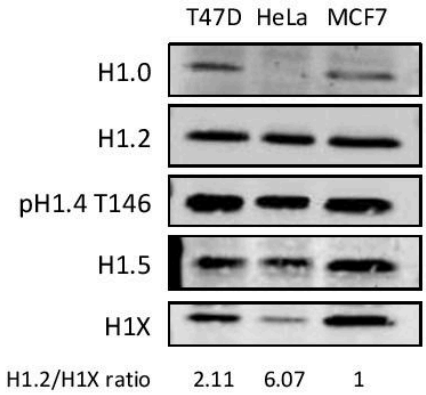
Figure S13

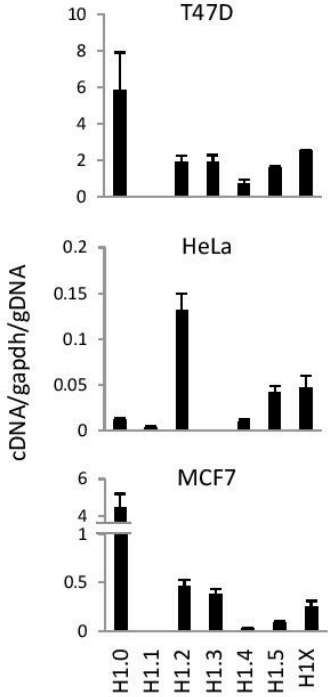


Millán-Ariño, Ll., et al.

Figure S14

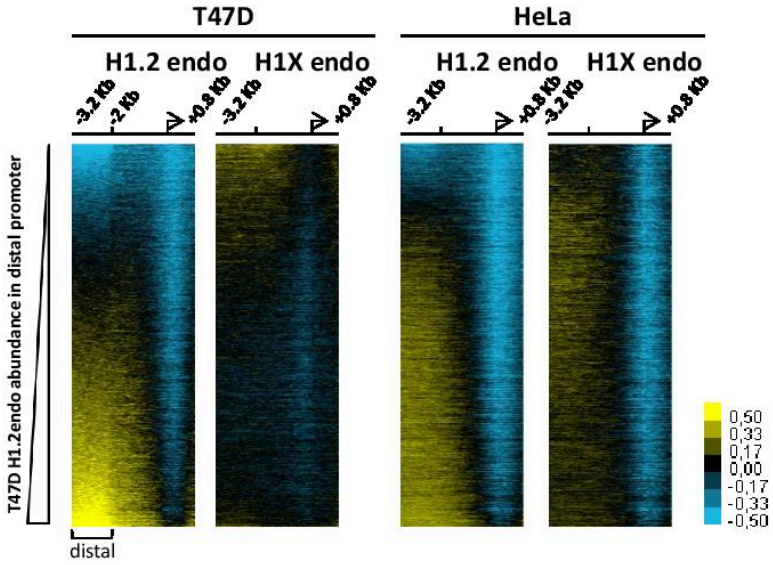






Millán-Ariño, Ll., et al.

Figure S15



Millán-Ariño, Ll., et al.

Figure S16

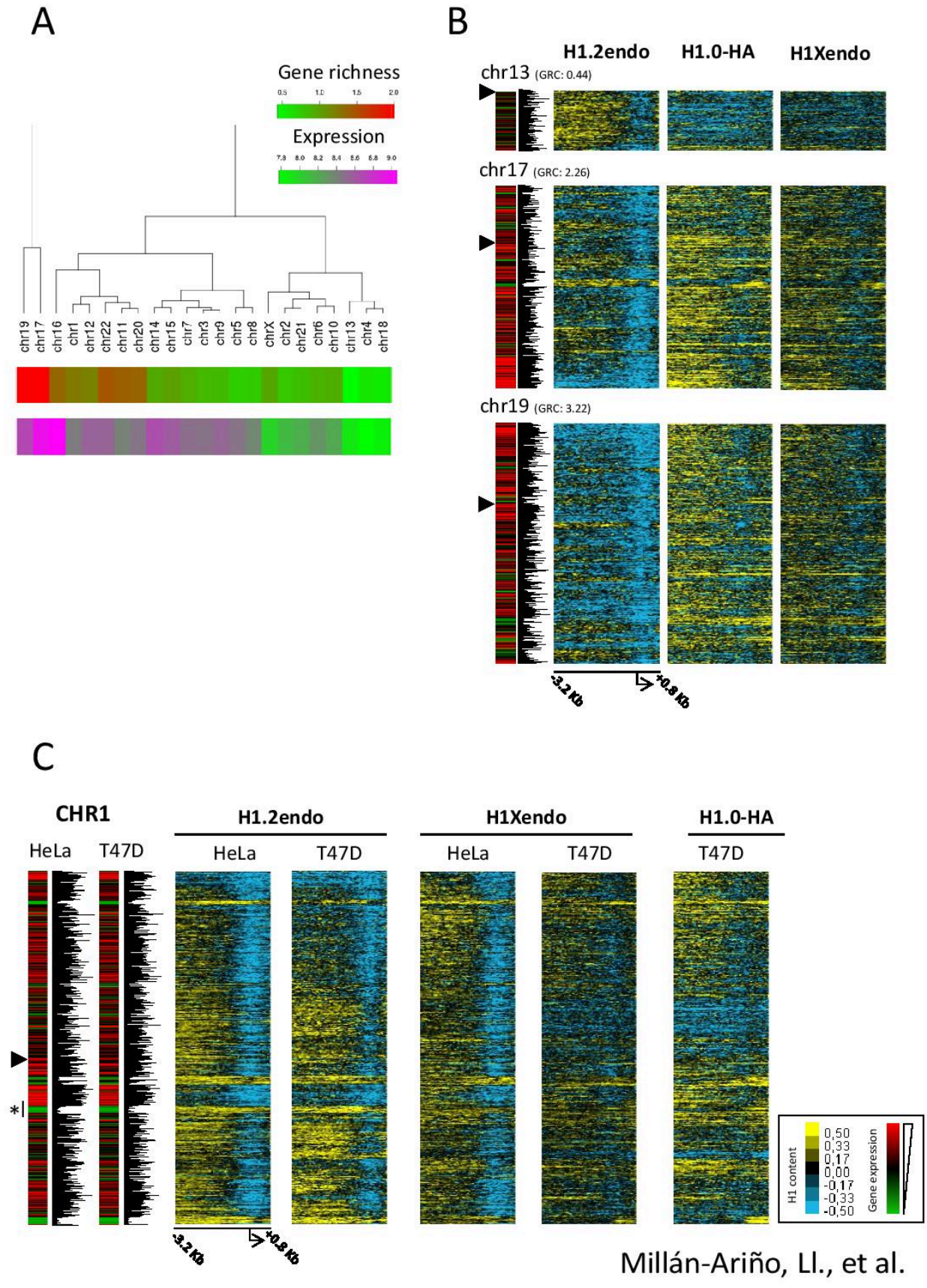
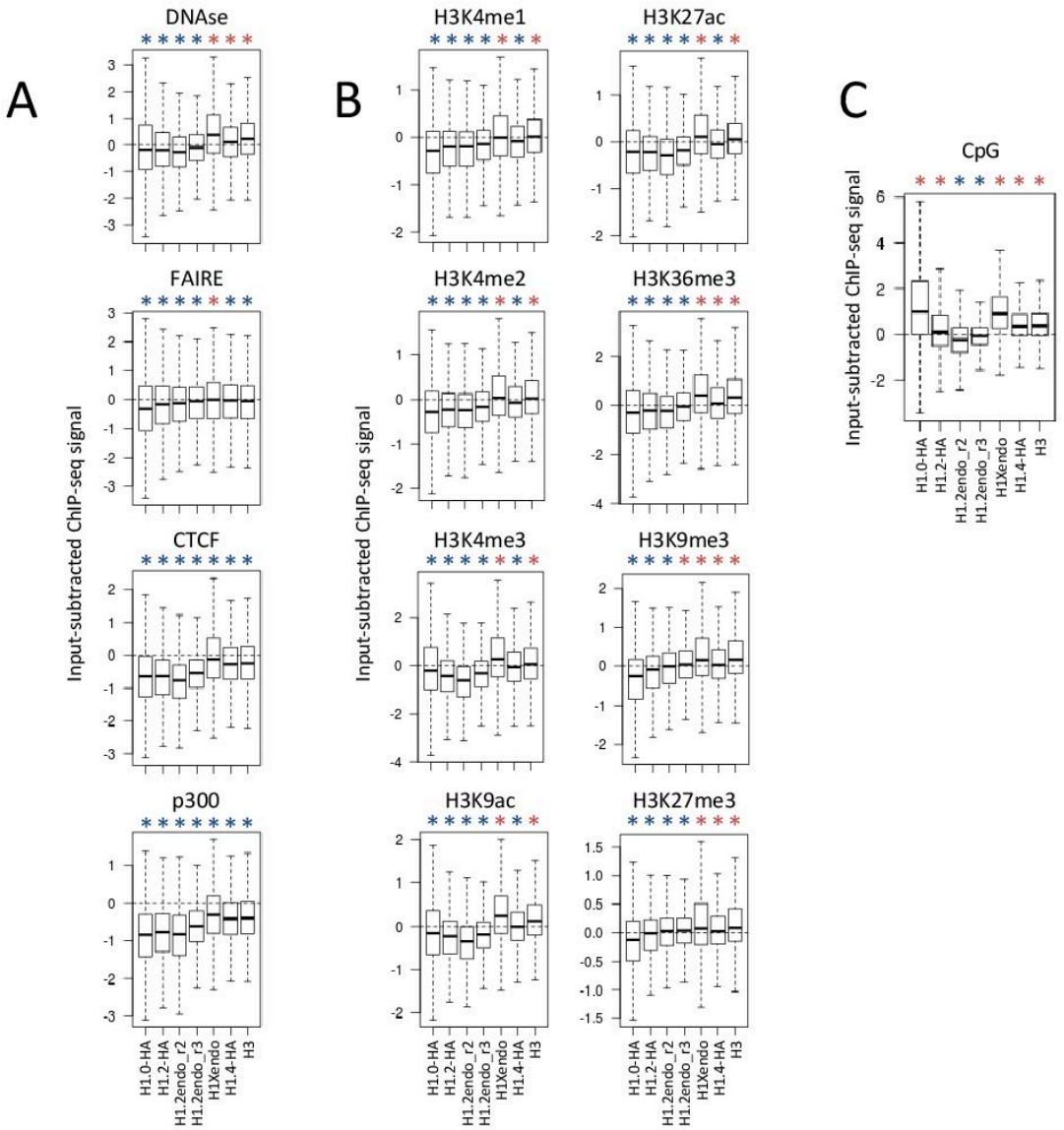
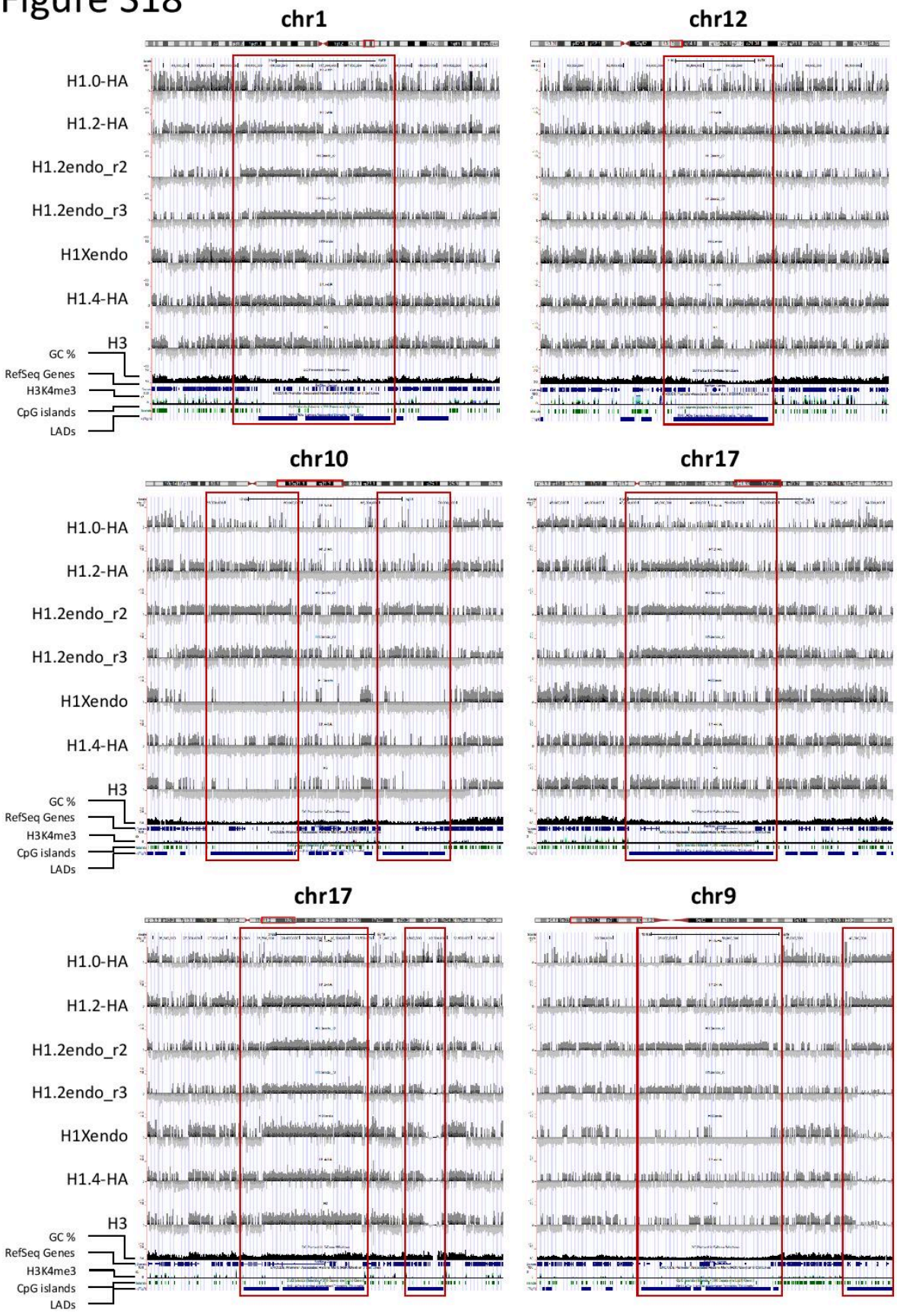

A
Gene richness
Expression
chr19 chr17 chr16 chr1 chr12 chr22 chr11 chr20 chr14 chr15 chr7 chr3 chr9 chr5 chr8 chrX chr2 chr21 chr6 chr10 chr13 chr4 chr18
B
H1.2endo
H1.0-HA
H1Xendo
chr13 (GRC: 0.44)
chr17 (GRC: 2.26)
chr19 (GRC: 3.22)
-3.2 Kb
+0.8 Kb
C
CHR1
HeLa
T47D
H1.2endo
HeLa
T47D
H1Xendo
HeLa
T47D
H1.0-HA
T47D
*
-3.2 Kb
+0.8 Kb
H1 content
0,50
0,33
0,17
0,00
-0,17
-0,33
-0,50
Gene expression
Millán-Ariño, Ll., et al.

Figure S17



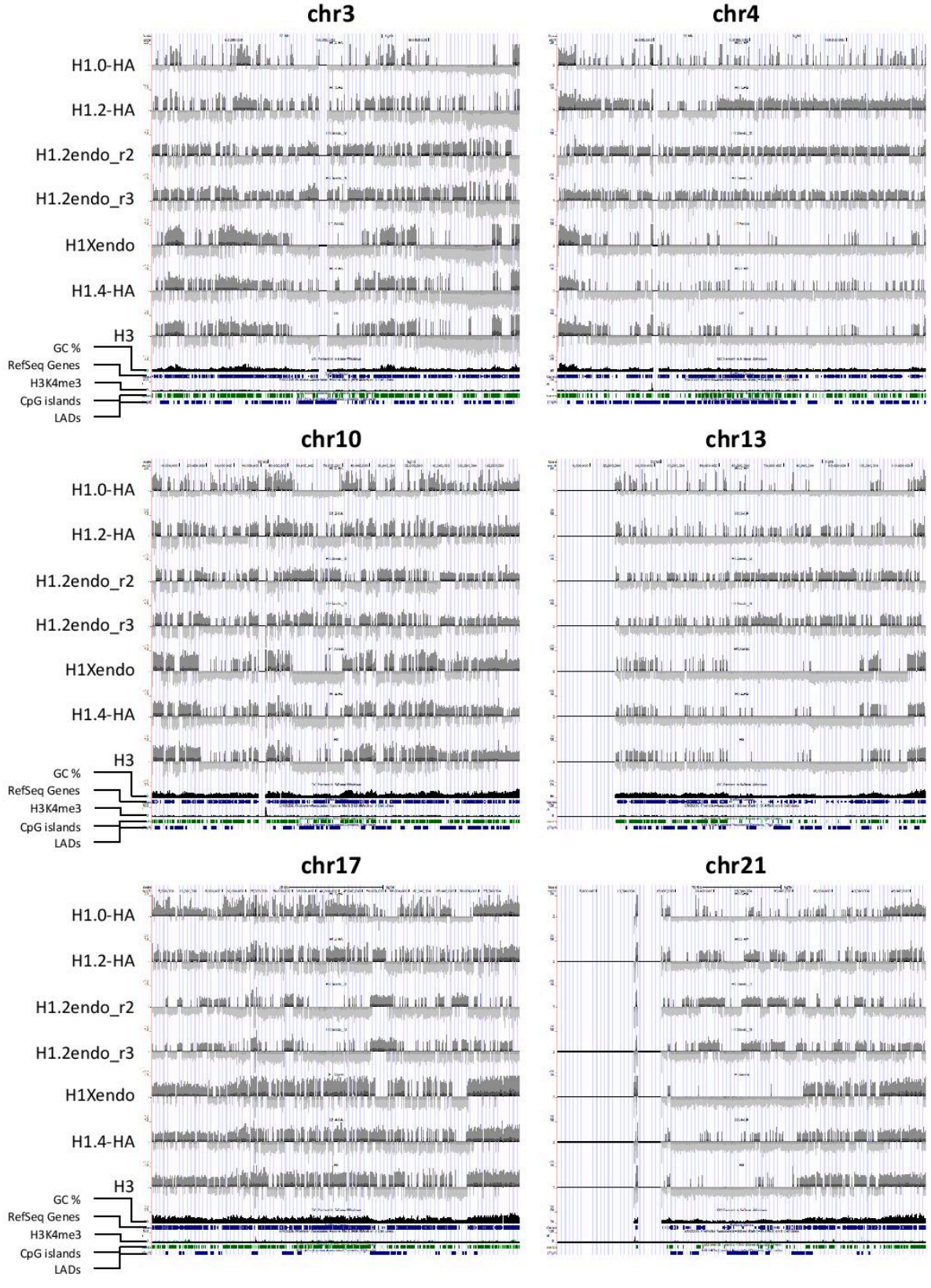
Millán-Ariño, Ll., et al.

Figure S18



Millán-Ariño, Ll., et al.

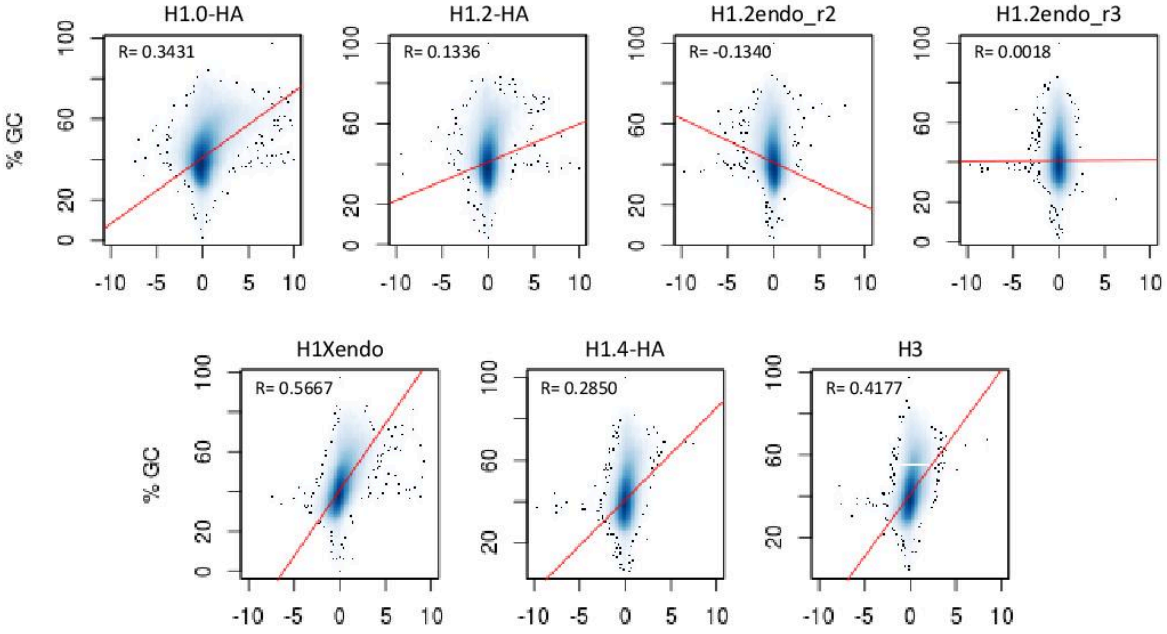
Figure S19



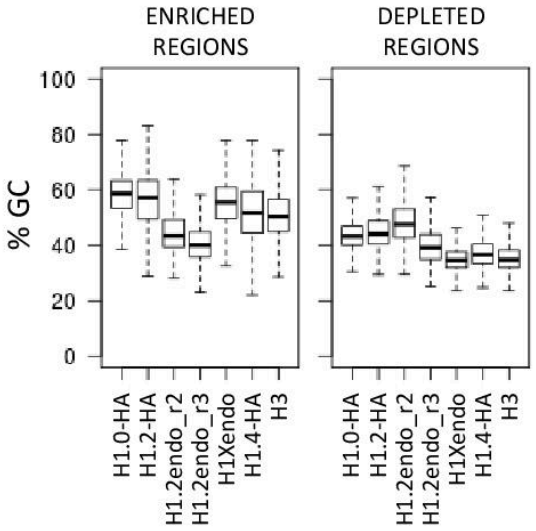
Millán-Ariño, Ll., et al.

Figure S20

A

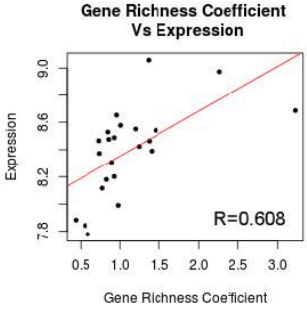


B

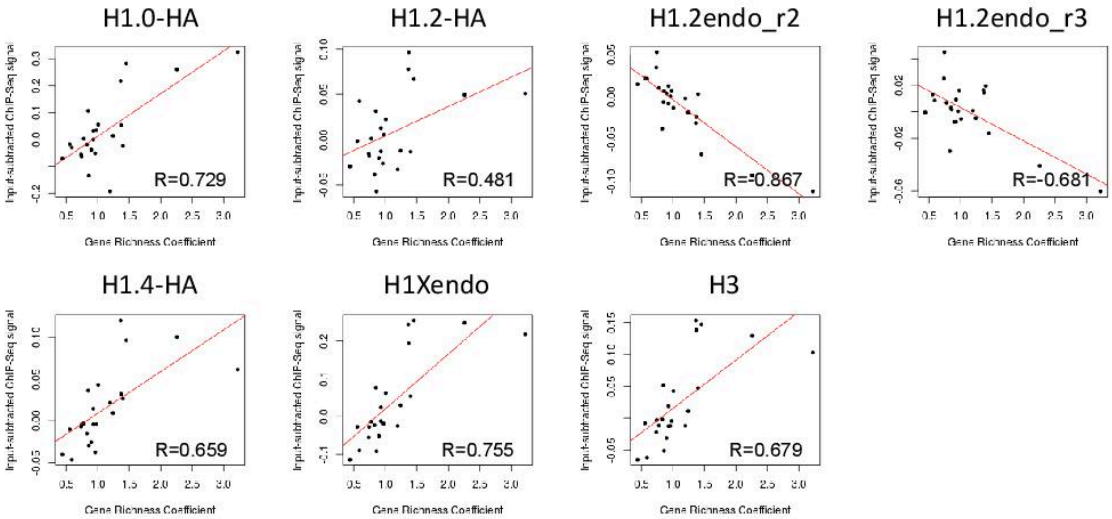


Millán-Ariño, Ll., et al.

Figure S21



Gene richness vs H1 abundance



Gene expression vs H1 abundance

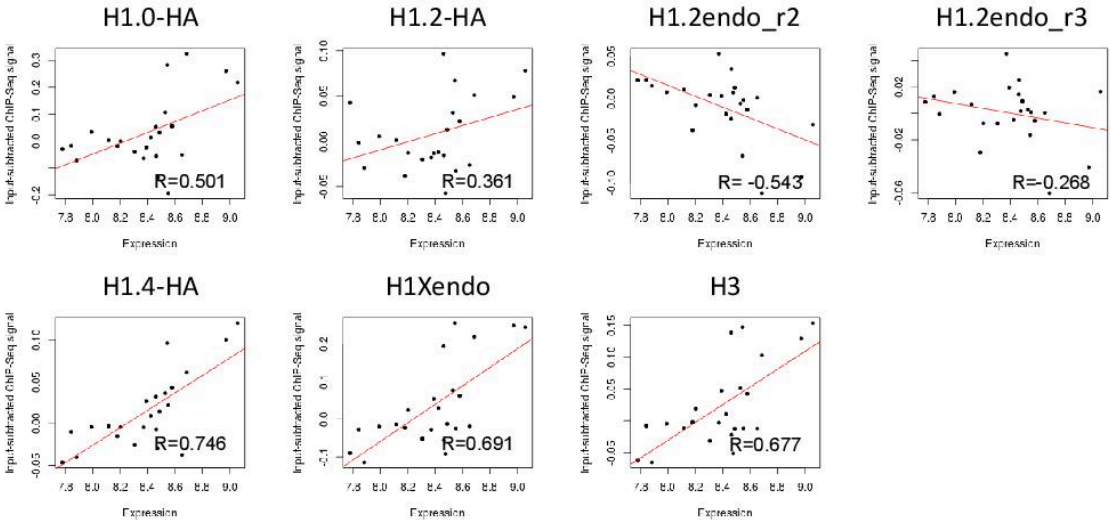
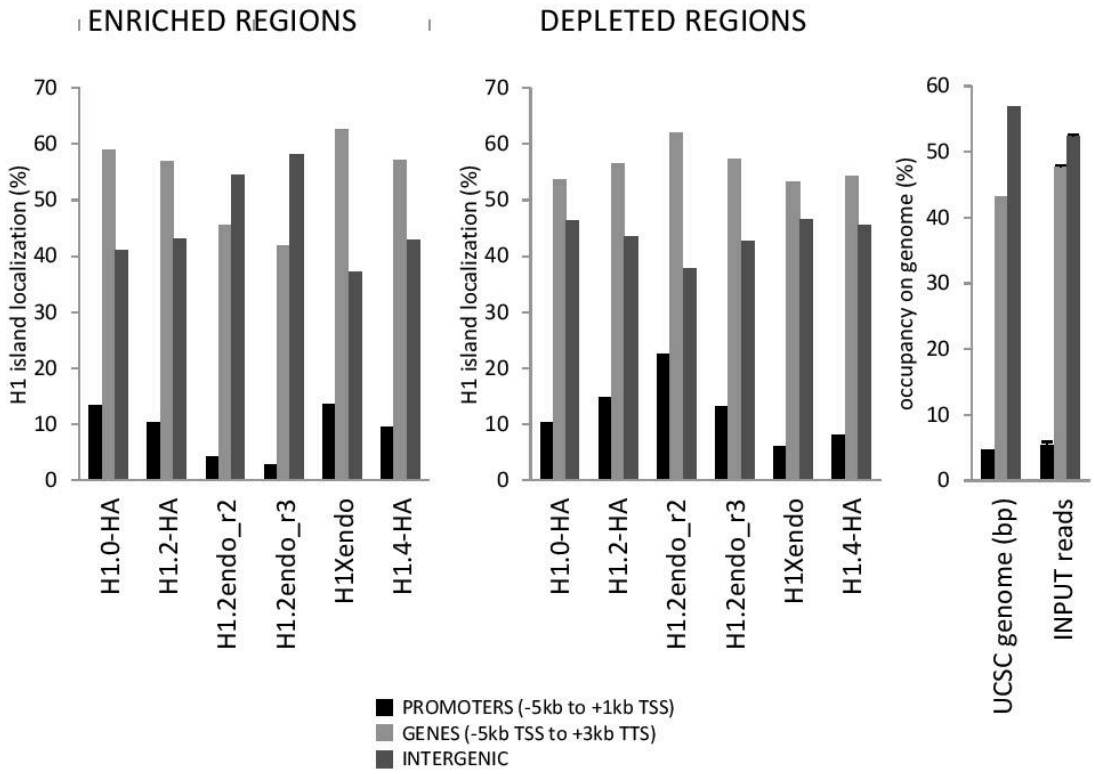
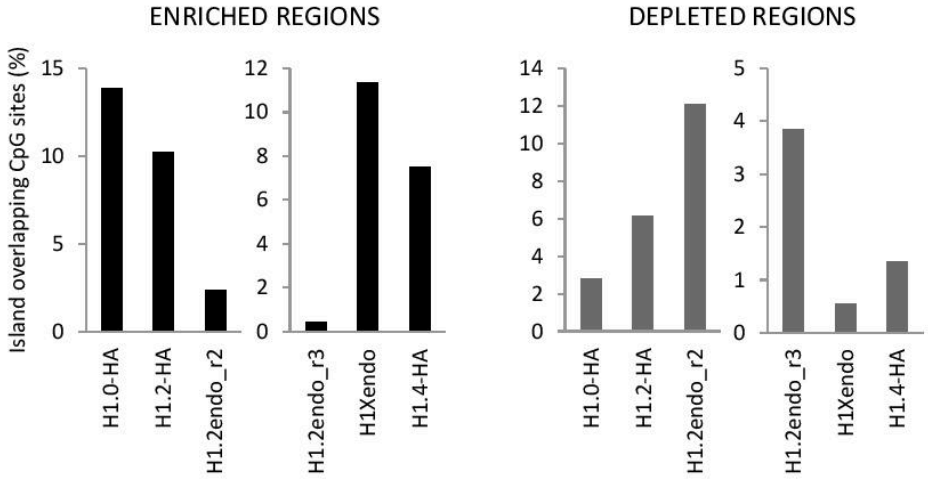


Figure S22

A

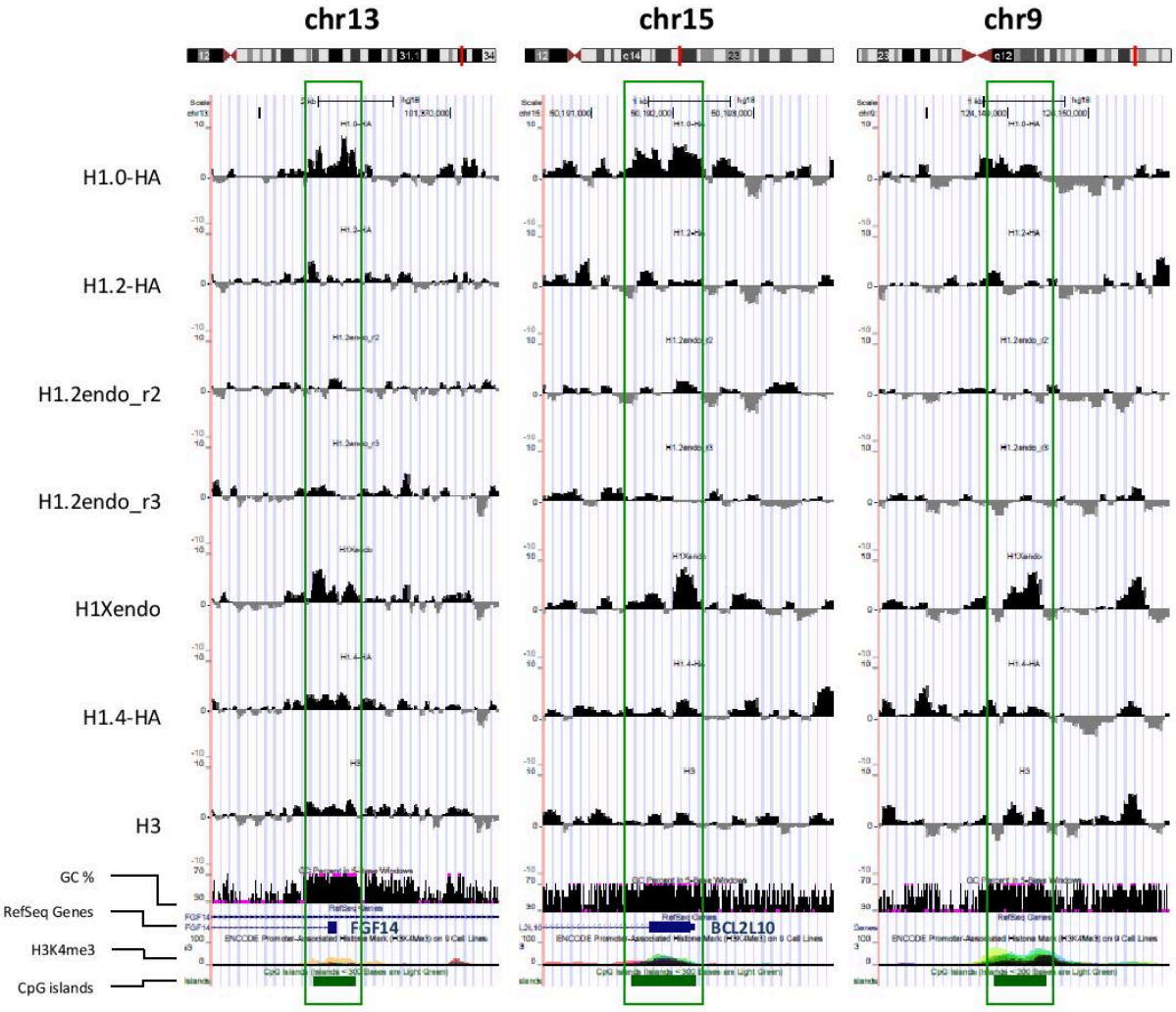


B



Millán-Ariño, Ll., et al.

Figure S23



Millán-Ariño, Ll., et al.

Figure S24

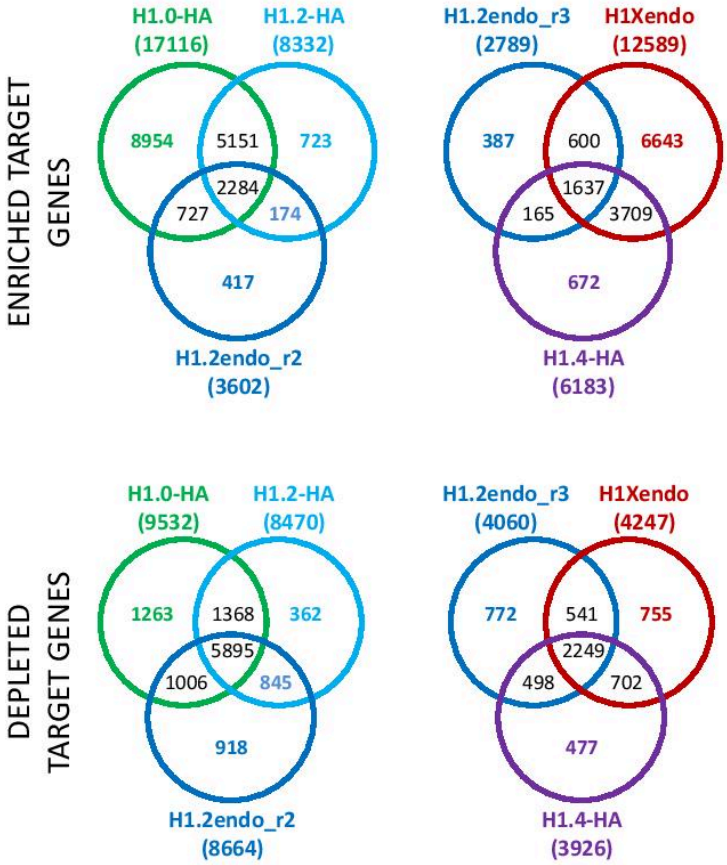
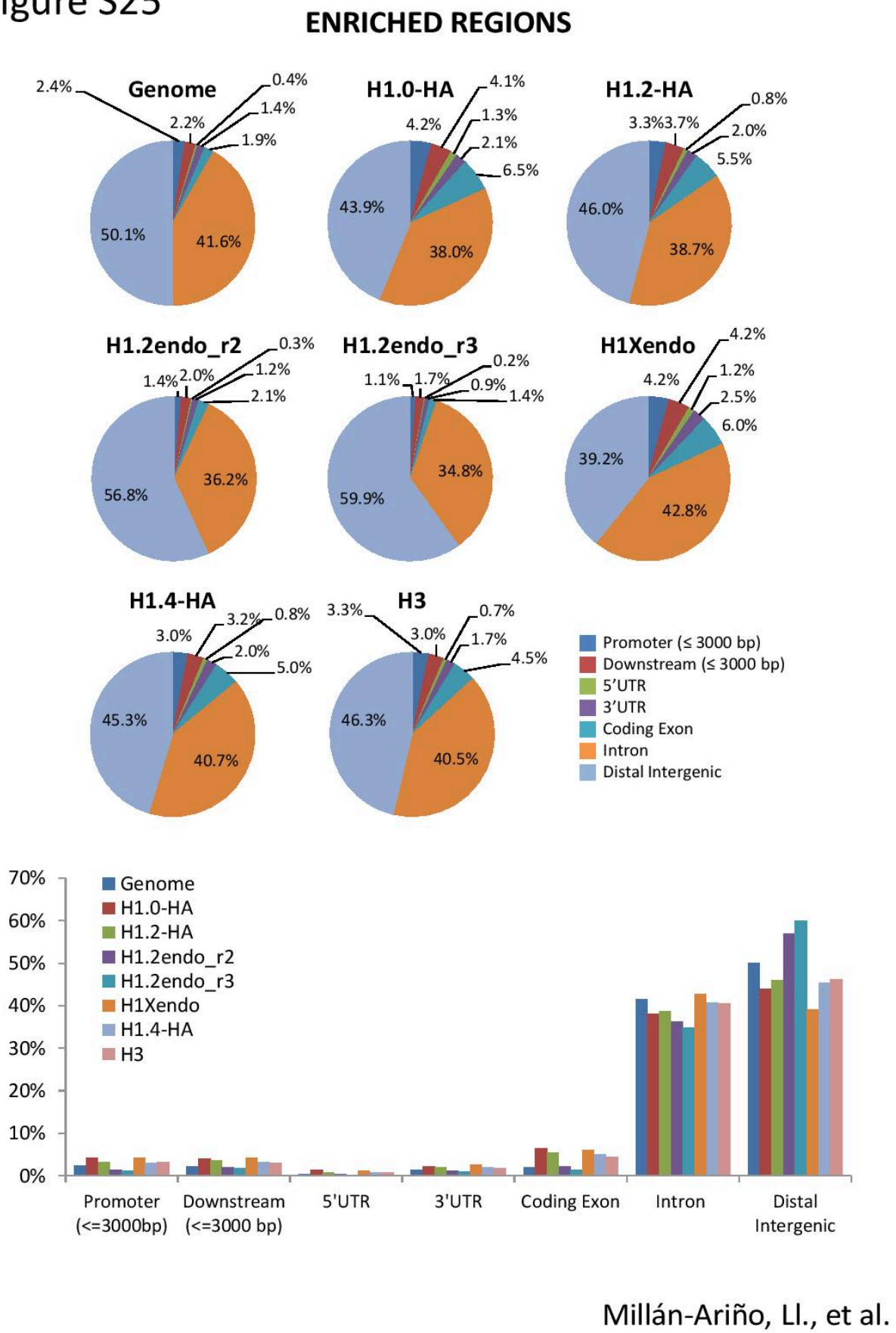
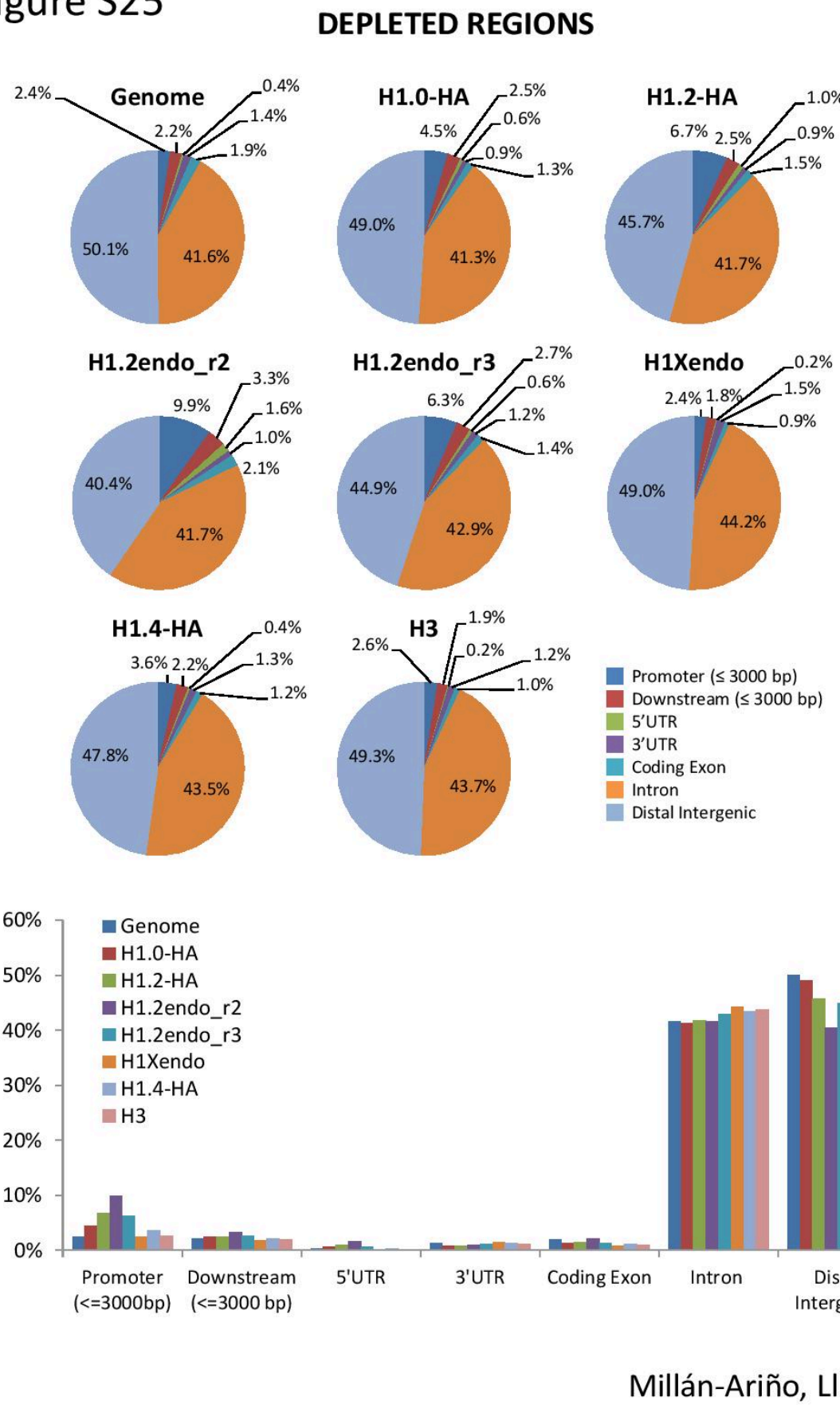

ENRICHED TARGET GENES
H1.0-HA (17116)
H1.2-HA (8332)
8954
5151
723
2284
727
174
417
H1.2endo_r2 (3602)
H1.2endo_r3 (2789)
H1Xendo (12589)
387
600
6643
1637
165
3709
672
H1.4-HA (6183)
DEPLETED TARGET GENES
H1.0-HA (9532)
H1.2-HA (8470)
1263
1368
362
5895
1006
845
918
H1.2endo_r2 (8664)
H1.2endo_r3 (4060)
H1Xendo (4247)
772
541
755
2249
498
702
477
H1.4-HA (3926)

Figure S25



Millán-Ariño, Ll., et al.

Figure S25



Millán-Ariño, Ll., et al.

Table S1

List of read lenght, counts, and total mappable reads (to hg18) of the libraries

Library	Read lenght (bp)	Total reads	Mapped reads	Mapped rate (%)
H1.0-HA_r1	49	50,088,474	47,468,471	94.77
H1.2-HA_r1	49	49,572,689	46,963,903	94.74
H1.0-HA_r2	49	49,298,091	47,071,182	95.48
H1.2-HA_r2	49	50,163,521	48,076,368	95.84
H1.2endo_r2	49	50,380,892	48,487,915	96.24
INPUT_r2	49	51,302,633	48,832,661	95.19
H1.2endo_r3	49	51,112,949	49,120,181	96.10
H1Xendo_r3	49	51,304,089	49,152,136	95.81
H1.4-HA_r3	49	51,537,524	49,354,356	95.76
H3_r3	49	50,125,868	48,383,270	96.52
INPUT_r3	49	52,386,794	49,803,697	95.07

Millán-Ariño, Ll., et al.

Table S2

Gene ontology of H1.2endo top10% enriched promoters.

Biological Process	P-Value	Benjamini
SENSORY PERCEPTION OF SMELL	5.9E-45	2.1E-41
SENSORY PERCEPTION OF CHEMICAL STIMULUS	3.7E-43	6.4E-40
SENSORY PERCEPTION	9.0E-35	1.0E-31
NEUROLOGICAL SYSTEM PROCESS	4.9E-34	4.3E-31
COGNITION	1.2E-33	8.5E-31
G-PROTEIN COUPLED RECEPTOR PROTEIN SIGNALLING PATHWAY	3.9E-28	2.3E-25
CELL SURFFACE RECEPTOR LINKED SIGNAL TRANSDUCTION	1.9E-23	9.3E-21
CELL ADHESION	2.1E-5	9.1E-3
BIOLOGICAL ADHESION	2.2E-5	8.6E-3
Cellular Component	**P-Value**	**Benjamini**
PLASMA MEMBRANE	7.6E-11	3.9E-8
INTEGRAL TO PLASMA MEMBRANE	3.7E-10	9.5E-8
INTRINSIC TO PLASMA MEMBRANE	1.9E-9	3.3E-7
EXTRACELLULAR REGION	1.0E-5	1.4E-3
EXTRACELLULAR REGION PART	6.3E-5	6.5E-3

Gene ontology of H1.Xendo top10% enriched promoters.

Biological Process	P-Value	Benjamini
NEURON DIFFERENTIATION	5.6E-9	2.0E-5
G-PROTEIN SIGNALLING, COUPLET TO CYCLIC NUCLEOTIDE SECOND MESSENGER	6.8E-7	1.2E-3
CELL-CELL SIGNALING	1.1E-6	1.3E-3
SECOND-MESSENGER-MEDIATED SIGNALLING	1.4E-6	1.3E-3
ANTERIOR/POSTERIOR PATTERN FORMATION	1.9E-6	1.4E-3
NEURON DEVELOPMENT	2.9E-6	1.8E-3
CYCLIC-NUCLEOTIDE-MEDIATED SIGNALLING	3.2E-6	1.7E-3
EMBRYONIC ORGAN DEVELOPMENT	4.1E-6	1.9E-3
PATTERN SPECIFICATION PROCESS	4.5E-6	1.8E-3
Cellular Component	**P-Value**	**Benjamini**
INTRINSIC TO PLASMA MEMBRANE	1.0E-8	5.4E-6
PLASMA MEMBRANE PART	1.5E-8	3.9E-6
PLASMA MEMBRANE	2.8E-8	4.9E-6
INTEGRAL TO PLASMA MEMBRANE	4.5E-8	6.0E-6
EXTRACELLULAR REGION	2.5E-7	2.6E-5

Gene ontology of H1.2endo bottom10% enriched promoters.

Biological Process	P-Value	Benjamini
REGULATION OF SYSTEM PROCESS	1.1E-5	4.2E-2
CELL-CELL SIGNALLING	1.2E-5	2.4E-2
SENSORY ORGAN DEVELOPMENT	6.9E-5	8.6E-2
NEGATIVE REGULATION OF CATABOLIC PROCESS	1.1E-4	1.1E-1
REGIONALIZATION	2.7E-4	1.9E-1
CIRCULATORY SYSTEM PROCESS	8.5E-4	4.2E-1
BLOOD CIRCULATION	8.5E-4	4.2E-1
Cellular Component	**P-Value**	**Benjamini**
PLASMA MEMBRANE PART	1.1E-3	4.6E-1
TROPONIN COMPLEX	4.3E-3	7.1E-1
DNA-DIRECTED RNA POLYMERASE II, CORE COMPLEX	7.7E-3	7.7E-1
PLASMA MEMBRANE	1.5E-2	8.8E-1
ACTIN CYTOSKELETON	1.6E-2	8.4E-1

Gene ontology of H1Xendo bottom10% enriched promoters.

Biological Process	P-Value	Benjamini
CHROMOSOME ORGANIZATION	1.0E-11	3.4E-8
CHROMATIN ORGANIZATION	8.7E-11	1.4E-7
NUCLEOSOME ORGANIZATION	3.2E-6	3.5E-3
CHROMTAIN MODIFICATION	6.9E-6	5.6E-3
NUCLEOSOME ASSEMBLY	2.2E-5	1.4E-2
DNA PACKING	2.4E-5	1.3E-2
PROTEIN-DNA COMPLEX ASSEMBLY	2.5E-5	1.2E-2
Cellular Component	**P-Value**	**Benjamini**
INTRACELLULAR ORGANELLE LUMEN	7.2E-7	4.2E-4
ORGANELLE LUMEN	1.2E-6	3.4E-4
MEMBRANE-ENCLOSED LUMEN	2.5E-6	4.8E-4
NUCLEAR LUMEN	8.8E-6	1.3E-3
CHROMOSOME	1.3E-5	1.5E-3

Millán-Ariño, Ll., et al.

Table S3

ENRICHED ISLANDS	H1.0-HA	H1.2-HA	H1.2endo_r2	H1.2endo_r3	H1Xendo	H1.4-HA
Total islands	49320	16059	7500	6911	38782	12478
% intergenic/total	40.97	43.10	54.44	58.10	37.29	42.84
% genic/total	59.03	56.90	45.56	41.90	62.71	57.16
% promoter/gene	22.68	18.01	9.13	6.63	21.75	16.49
Target genes	17116	8332	3602	2789	12589	6183
Target promoters	7591	2176	490	257	6284	1715
% CpG overlap	13.88	10.22	2.37	0.42	11.35	7.53

DEPLETED ISLANDS	H1.0-HA	H1.2-HA	H1.2endo_r2	H1.2endo_r3	H1Xendo	H1.4-HA
Total islands	25459	15224	12305	5714	10205	6864
% intergenic/total	46.41	43.44	37.89	42.65	46.65	45.63
% genic/total	53.59	56.56	62.11	57.35	53.35	54.37
% promoter/gene	19.28	26.13	36.35	23.02	11.34	15.06
Target genes	9532	8470	8664	4060	4247	3926
Target promoters	4105	3783	4610	1412	1014	1046
% CpG overlap	2.78	6.15	12.11	3.85	0.55	1.35

Millán-Ariño, Ll., et al.

Supplementary Methods

Treatments

For H1 knock-down cell lines, doxycycline (Sigma) was added at 2.5 μg/ml when indicated. Along a 6-day treatment with Dox, cells were passaged at day 3.

For hormone treatment experiments with R5020 (PerkinElmer Life Sciences), cells were plated in phenol red-free medium supplemented with 10% dextran-coated charcoal-treated FBS and, 24 hours later, the medium was replaced by fresh serum-free medium. After 24 hours under serum-free conditions, cells were treated with R5020 (10 nM) for different times at 37°C.

Antibodies

Polyclonal antibodies specifically recognizing human H1 variants, are: anti-H1.0 (Abcam 11079), anti-H1 phospho-T146 (Abcam 3596), anti-H1.5 (Abcam 18208). Other antibodies used are: anti-H3K4me3 (Abcam 8580), and anti-total H1 (Millipore, clone AE-4, 05-457).

H1 Extraction, Gel Electrophoresis and Immunoblotting

Histone H1 was purified by 5% perchloric acid lysis for 1 hour at 4°C. Soluble acid proteins were precipitated with 30% trichloroacetic acid over-night at 4°C, washed twice with 0.5 ml of acetone and reconstituted in water. Protein concentration was determined by Micro BCA protein assay (Pierce).

Chromatin or purified H1 histones were subjected to 12% SDS-PAGE, transferred to a PVDF membrane, blocked with Odissey blocking buffer (LI-COR Biosciences) for 1 hour, incubated with primary antibodies over-night at 4°C and with secondary antibodies conjugated to fluorescence (IRDye 680 goat anti-rabbit IgG and IRDye 800CW goat anti-mouse IgG, LI-COR) for 1 hour at room temperature. Bands were visualized with the Odissey Infrared Imaging System.

Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) Assays

Cells were fixed using 1% formaldehyde, harvested and sonicated using a Diagenode Bioruptor to generate chromatin fragments between 200 and 500 bp. To prepare input DNA an aliquot of chromatin was taken treated with RNAase A, de-crosslinked overnight at 65°C, purified by phenol/chloroform extraction, and run on a gel to ensure

average fragment sizes of 200-500 bp. FAIRE DNA was prepared processing twice chromatin by phenol/chloroform extraction to purify DNA not bound by nucleosome in the water phase. The samples were later treated with RNase A, de-crosslinked by overnight incubation at 65ºC, and purified by GenElute PCR Clean-Up Kit (Sigma). Real-time PCR was performed on FAIRE and input DNA using EXPRESS SYBR GreenER qPCR SuperMix Universal from Invitrogen and specific oligonucleotides in a Roche 480 Lightcycler. All oligonucleotide sequences used for the amplifications are available on request.

RNA extraction and RT-PCR

Total RNA was extracted using High Pure RNA isolation kit (Roche) according to the manufacturer's instructions. cDNA was obtained from 100 ng of total RNA using SuperScript VILO cDNA synthesis (Invitrogen). Gene products were analyzed by qPCR using EXPRESS SYBR GreenER qPCR SuperMix Universal (Invitrogen) and specific oligonucleotides in a Roche 480 Lightcycler. Each value was corrected by human GAPDH and expressed as relative units. Gene-specific oligonucleotide sequences are available on request.

APPENDIX II

ARTICLE IN PRESS
BBAGRM-00951; No. of pages: 10; 4C:
Biochimica et Biophysica Acta xxx (2015) xxx–xxx

Contents lists available at ScienceDirect
Biochimica et Biophysica Acta
journal homepage: www.elsevier.com/locate/bbagrm

Specificities and genomic distribution of somatic mammalian histone H1 subtypes

Lluís Millán-Ariño [1], Andrea Izquierdo-Bouldstridge, Albert Jordan *

Institut de Biologia Molecular de Barcelona (IBMB-CSIC), Barcelona E-08028, Spain

ARTICLE INFO

Article history:
Received 14 July 2015
Received in revised form 13 October 2015
Accepted 14 October 2015
Available online xxxx

Keywords:
Histone H1
Linker
Variants
Genomics

ABSTRACT

Histone H1 is a structural component of chromatin that may have a role in the regulation of chromatin dynamics. Unlike core histones, the linker histone H1 family is evolutionarily diverse and many organisms have multiple H1 variants or subtypes, distinguishable between germ-line and somatic cells. In mammals, the H1 family includes seven somatic H1 variants with a prevalence that varies between cell types and over the course of differentiation, H1.1 to H1.5 being expressed in a replication-dependent manner, whereas H1.0 and H1X are replication-independent. Until recently, it has not been known whether the different variants had specific roles in the regulation of nuclear processes or were differentially distributed across the genome. To address this, an increasing effort has been made to investigate divergent features among H1 variants, regarding their structure, expression patterns, chromatin dynamics, post-translational modifications and genome-wide distribution. Although H1 subtypes seem to have redundant functions, several reports point to the idea that they are also differently involved in specific cellular processes. Initial studies investigating the genomic distribution of H1 variants have started to suggest that despite a wide overlap, different variants may be enriched or preferentially located at different chromatin types, but this may depend on the cell type, the relative abundance of the variants, the differentiation state of the cell, or whether cells are derived from a neoplastic process. Understanding the heterogeneity of the histone H1 family is crucial to elucidate their role in chromatin organization, gene expression regulation and other cellular processes. This article is part of a Special Issue entitled: Histone H1, edited by Dr. Albert Jordan.

1. Introduction

For many years, most work on chromatin structure and function has been focused on the nucleosome core particle, composed of the core histones H2A, H2B, H3 and H4, and less attention has been paid to the linker histone H1, regarding its structure, function, post-translational modifications (PTMs), etc. Since the linker histone H1 was discovered in calf thymus as a lysine-rich family of histone proteins that differed in sequence composition [1], the study of these subtypes has remained challenging because of their high heterogeneity, which makes it difficult to obtain specific antibodies for all members of this family.

Compared with core histones, which are highly conserved in evolution, the linker histone H1 family is more divergent [2] and in several organisms many subtypes or variants are known to exist due to gene duplication events during evolution, from one variant in simple eukaryotes to eleven variants in humans or mice. The diversity of histones and the identification of an increasing number of variants have led to confusion in naming. Over the years, many attempts have been made to unify nomenclature (Table 1). Recently, in 2012, a unified phylogeny-based nomenclature was proposed for histone variants [3].

In humans and mice, the H1 family comprises eleven variants or subtypes, products of different paralog genes (Table 1). They can be classified according to various criteria: expression, cell cycle dependence, and gene location in the genome (see [4–6] for recent reviews). Specifically, seven human H1 variants are somatic subtypes (H1.1 to H1.5, H1.0 and H1X), while others are restricted to germ cells, with three testis-specific variants (H1t, H1T2 and HILS1) and one oocyte-specific variant (H1oo). For an extended review of germline-specific H1 variants in different species, see [7]. Among the somatic histone H1 variants, H1.1 to H1.5 are expressed in a replication-dependent manner through the cell cycle, whereas H1.0 and H1X are replication-independent. H1.2 to H1.5 and H1X are ubiquitously expressed, H1.1 is restricted to certain tissues and cell types (liver, kidney, lung, lymphocytes from thymus and spleen, neurons and germ cells), and H1.0 accumulates in terminally differentiated cells. Regarding gene location, H1.1 to H1.5-encoding genes are clustered in a region of chromosome 6, together with the core histone genes, whereas H1X and H1.0 are located on chromosome 3 and 22, respectively.

* Corresponding author.
E-mail address: albert.jordan@ibmb.csic.es (A. Jordan).
[1] Present address: Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden.

http://dx.doi.org/10.1016/j.bbagrm.2015.10.013
1874-9399/

Please cite this article as: L. Millán-Ariño, et al., Specificities and genomic distribution of somatic mammalian histone H1 subtypes, Biochim. Biophys. Acta (2015), http://dx.doi.org/10.1016/j.bbagrm.2015.10.013

Table 1
Nomenclature for the mammalian histone H1 variants.

Unified name [3]	Human gene symbol	Previous human [105]	Mouse gene symbol	Previous mouse [23,106]	Gene location (human)	Expression	Cell cycle dependence
H1.1	HIST1H1A	H1.1	Hist1h1a	H1a	6p21.3	Somatic	Replication-dependent
H1.2	HIST1H1C	H1.2	Hist1h1c	H1c	6p21.3	Somatic	
H1.3	HIST1H1D	H1.3	Hist1h1d	H1d	6p21.3	Somatic	
H1.4	HIST1H1E	H1.4	Hist1h1e	H1e	6p21.3	Somatic	
H1.5	HIST1H1B	H1.5	Hist1h1b	H1b	6p22.1	Somatic	
(TS) H1.6	HIST1H1T	H1t	Hist1h1t	H1t	6p21.3	Testis	
H1.0	H1F0	H1.0, $H1^0$	H1f0	H1(0)	22q13.1	Somatic	Replication-independent
(TS) H1.7	H1FNT	H1T2	H1fnt	H1t2	12q13.1	Testis	
(OO) H1.8	H1FOO	H1oo	H1foo	H1oo	3q22.1	Oocyte	
(TS) H1.9	HILS1		Hils1	TISP64	17q21.33	Testis	
H1.10	H1FX	H1x	H1fx	H1X	3q21.3	Somatic	

Genes located in large clusters on chromosome 6 (6p21–p22) are encoded by individual intronless genes, with short 5′ and 3′ ends. Transcripts lack polyA tails but contain a 3′ stem-loop sequence that allows for rapid translation during DNA replication. On the other hand, isolated genes such as H1.0 and H1X are also intronless, but their mRNA is polyadenylated. It is interesting to note that, although clustered genes share the same chromosome location and gene structure, they are not expressed equally, H1.1 and H1t ((TS) H1.6) showing tissue specificity and the expression of other subtypes fluctuating differently through the cell cycle [8]. Thus, it seems to be that transcription of different H1 variants is tightly regulated in order to achieve proper expression of each variant in different tissues or cells, but also during the cell cycle and differentiation. However, the exact molecular mechanisms by which this happens have yet to be identified. Little is known about histone H1 transcriptional regulation, but it is reported that specific sequences in their promoters modulate binding of transcription factors and chromatin proteins [9–12].

2. Histone H1 variants: more than a redundant family of structural chromatin proteins

Due to its role in the formation of higher-order chromatin structures, H1 has been classically seen as a structural component related to chromatin compaction and inaccessibility to transcription factors, RNA polymerase, and chromatin remodeling enzymes [13,14]. Many studies support this view, as the presence of H1 in promoter regions impairs transcription of the associated gene [15–17]. However, in recent years, the view that H1 plays a more dynamic and gene-specific role in regulating gene expression is gaining strength in the field. In fact, overexpression and knock-out or knock-down studies in several organisms have revealed that only a few genes change in expression upon alteration of H1 content, some being up- and some down-regulated; this points to a complex positive or negative gene-specific function of H1, rather than a general repressive function. The study of H1 function has been addressed by manipulating the amount of H1 in different organisms (Table 2).

Another controversial issue regarding the linker histone H1 concerns the functional specificity of its variants. It is still not well understood why there are so many H1 variants and many efforts have recently been made to ascertain whether they play specific roles or there is redundancy in their functions. In particular, single or double H1 variant knock-out mice were generated and, interestingly, they presented no apparent phenotype [18]. This was attributable to the compensatory up-regulation of other subtypes that maintained a normal H1-to-nucleosome stoichiometry, favoring the view that H1 variants are redundant. However, triple somatic H1 knock-out (H1c, H1d and H1e, i.e. mouse H1.2, H1.3 and H1.4) was embryonic lethal during gestation causing severe developmental defects, because the up-regulation of the remaining subtypes did not fully compensate for the severe lack of H1. It can be concluded that correct total H1 levels are critical for proper mammalian development [19]. Furthermore, mouse embryonic stem cells derived from triple KO mice presenting an ~50% reduction in total H1 showed a shorter nucleosome repeat length (NRL) and reduced chromatin compaction. Moreover, these cells had an altered level of gene expression in a small number of genes, mainly regulated by DNA methylation (imprinted or X chromosome genes), pointing to a role of H1 in maintaining or establishing DNA methylation patterns [20]. This was attributed to the decrease in total H1, not to the lack of any particular variant.

Despite previous results in KO mice, increasing evidence supports the hypothesis of specific functions for histone H1 variants. Reports of their expression preferences or different affinities for chromatin, as well as their specific role in various cellular processes, support the idea that the heterogeneity of H1 histone family in several organisms is linked to the specialization of these variants in certain cellular processes. Evidence supporting specific functions for H1 variants is described below.

3. Evidence for H1 subtype specialization

3.1. H1 variant sequence conservation

H1 variants are paralog genes, as they originate from gene duplication events. On the other hand, the corresponding variants within two species are orthologs, because they share a common ancestor before the event of speciation. H1 ortholog genes are much more conserved than paralog genes; this means that the primary sequence of a given H1 variant is more conserved across species than within variants from the same species. This evolutionary effort to conserve the sequence of a given H1 subtype indicates that H1 variants specialized in their function after they had been generated in a common ancestor. Moreover, estimates of the rates of nucleotide substitution also support functional differentiation between H1 variants [21,22]. This observation is the single most compelling evidence to date for H1 variant functional specialization.

3.2. Differential expression patterns of H1 variants

There is 40 years' worth of work devoted to identifying expression patterns of H1 variants in different tissues and under different physiological conditions [4–8], concluding that H1 subtypes present cell type and tissue-specific expression patterns. It is well established that some histone H1 variants are ubiquitously expressed, while others are tissue restricted. Moreover, the relative ratio of a given variant can vary between cell types, and only H1.2 and H1.4 are expressed in all investigated cells [23–26].

It should also be noted that differential expression of H1 variants is regulated over the course of differentiation and development. In fact, H1.0 was considered a replacement histone variant because of its accumulation in terminally differentiated cells that stopped dividing [27,28]. A recent study, based on in vitro experiments on human embryonic stem cell (ESC) differentiation and on keratinocyte reprogramming to induced pluripotent cells (iPS), confirms that

Please cite this article as: L. Millán-Ariño, et al., Specificities and genomic distribution of somatic mammalian histone H1 subtypes, Biochim. Biophys. Acta (2015), http://dx.doi.org/10.1016/j.bbagrm.2015.10.013

Table 2
Overview of phenotypes observed after manipulation of H1 expression in different organisms. # KO = knock out; RE = reduced expression; OE = overexpression; LoF = loss of function. Adapted from [6].

Organism	Variant	#	Phenotype	References
S. cerevisiae	Hho1	KO	Effect on expression of specific genes (up or down)	[107]
		KO	Inhibition of DNA repair, effect on life span	[108]
		KO	Compromised sporulation efficiency	[109]
		LoF	Suppression of DNA repair by homologous recombination	[108]
T. thermophila		KO	Effect on expression of specific genes (up or down)	[110]
A. immersus		KO	Effect on life span (shorter)	[111]
C. elegans	H1.1	RE	Defect in development and transgene silencing	[112]
	H1.1	KO/LoF	Globally higher levels of H3K4 and lower levels of H3K9 methylation. Specific positive and negative effects on gene expression, embryonic lethality, reduced fertility, defects in germline development and differentiation Impaired innate immune system	[113–115]
	H1.X	RE	Uncoordinated and egg-laying defective worms	[116]
N. tabacum	H1A, H1B	RE	Defects in flower development and male gametogenesis	[117]
A. thaliana		RE	Aberrant development, DNA hypomethylation	[118]
D. melanogaster	dH1	RE	Compromised fly viability, lethality at larval stage Effect on expression of specific genes in a regional manner (transposons), genome instability and proliferation defects	[59,60]
	dBigH1	LoF	Impaired zygotic genome activation, increase in DNA damage and mitotic defects	[119]
	Mst77F	LoF	Sterility and morphological defects in spermatids	[120]
G. gallus (*DT 40 cells*)	01H1	KO	Changed protein patterns	[67]
	02H1		Enhanced expression of remaining H1s	
	03H1			
	10H1			
	H1L, H1R	KO	Enhanced expression of remaining H1s	[64]
	H1R	KO	Accumulation of IR-induced chromosomal aberrations, reduction in gene targeting efficiencies, impaired sister chromatid exchange	
	H1-null	KO	Shorter NRL, expanded nuclear volumes, higher chromosome aberration rates, alteration of transcription (mostly down-regulation)	[40]
X. laevis	H1A	RE/OE	Activation/repression of oocytes 5S gene	[121–123]
			Defects in the control of mesoderm differentiation	[124]
			Defects in mitotic chromosome architecture and segregation	
M. musculus	H1.0, H1c	OE	Longer NRL, differential alterations in the cell cycle and transcription levels	[125,126]
	H1a, H1b, H1c, H1d,	Single	Mice develop normally	[18,127]
	H1e, H1.0	KO	Positive and negative effects on gene expression	
	H1.0	KO	Mice develop normally	[128,129]
			Defect in immune system	
	H1t	KO	No phenotype in spermatogenesis	[62,63,65,66]
			Higher expression of H1.1, H1.2 and H1.4	
			Effect on expression of specific genes	
	H1T2	KO	Reduced male fertility	[130,131]
	H1c + H1d + H1e	Triple	50% lower H1-to-nucleosome ratio, embryonic lethal	[19,20]
		KO	Shorter NRL, genes up- and down-regulated, in particular genes regulated by methylation	
			Hyperresistance to DNA damage	[132]
			Impaired ESC differentiation	[61,133]
	H1c + H1e + H1.0	Triple KO	Impairment of chromatin condensation and rod cell structural integrity	[134]
H. sapiens	H1.0, H1.2, H1.3, H1.4, H1.5, H1X	RE	H1.2 and H1.4 caused defects in proliferation. H1.2 knock down caused G1 arrest, shorter NRL. Variant specific changes in global gene expression (breast cancer cells)	[44,74]
			H1.0 knock down impaired ESC differentiation	[29]
	H1.5	RE	Lower cell growth, SIRT1 and H3K9me2 loss, higher chromatin accessibility, deregulation of gene expression (fibroblasts)	[45]

H1 variants are differentially expressed over the course of these transitions [29]. Pluripotent cells have lower levels of H1.0 and H1X, and higher levels of H1.1, H1.3 and H1.5, than differentiated cells, where H1.0 represents ~80% of the H1 transcripts. It is also worth noting that H1.0 knock-down in ESCs impaired their differentiation.

Finally, expression of H1 variants is altered during cancer progression (see review by Scaffidi in this issue for a detailed discussion of this topic). For example, in ovarian cancer, H1.3 expression is elevated, while expression of H1.0, H1.1, H1.4 and H1X is lower in malignant adenocarcinomas than in benign adenomas [30]. Additionally, H1.5 expression is positively correlated with the grade of pulmonary neuroendocrine tumors [31]. As a consequence, histone H1 variants could be used as biomarkers, and perhaps even as a therapeutic target, if proven to be drivers in cancer.

3.3. Chromatin binding affinity

Originally, H1 was considered to be constantly associated with chromatin because of its structural role in chromatin formation. However, this view changed after the development of in vivo approaches to studying H1 binding to chromatin. Fluorescence recovery after photobleaching (FRAP) with recombinant H1s fused to green-fluorescent protein showed that they are much more mobile than core histones, but less so than their competitors in chromatin binding: high mobility group (HMG) proteins [32,33].

Further, FRAP experiments and biochemical studies in vitro showed that chromatin binding affinity and residence time on chromatin differed between H1 subtypes due to differences not only in the C-terminal tail, but also in the N-terminal tail [34]. FRAP experiments in human SK-N-SH neuroblastoma cells by Th'ng et al. [35], showed that H1.4 and H1.5 are the variants with the highest affinity to chromatin, followed by H1.3 and H1.0, and with H1.1 and H1.2 presenting the highest mobility. Interestingly, H1 mobility is also dependent on specific PTMs, for example, mutations in the residues associated with a particular PTM being linked to altered FRAP recovery kinetics [38,39]. In an in vitro competitive assay, in which the binding of the H1 subtypes to long chromatin fragments and to scaffold-associated regions (SARs) was determined, H1.1 was the subtype with the lowest affinity, while

Please cite this article as: L. Millán-Ariño, et al., Specificities and genomic distribution of somatic mammalian histone H1 subtypes, Biochim. Biophys. Acta (2015), http://dx.doi.org/10.1016/j.bbagrm.2015.10.013

H1.2 and H1.5 showed intermediate and H1.3, H1.4 and H1.0 high affinity to both SAR and non-SAR DNA [36]. Moreover, a study using chromatin assembled in *Drosophila* embryo extracts showed that human H1 variants differ in their ability to reconstitute nucleosome arrays in vitro —(H1.5, H1.4) > (H1.3, H1.2, H1.0) > H1.1 > H1X— [37]. Unfortunately, despite many more studies in this field, there is still no consensus on how the individual variants quantitatively interact with chromatin, partly because of the different parameters measured by the different techniques used.

3.4. Contribution to the NRL

Linker histone H1 is important in determining the nucleosomal repeat length (NRL), i.e. the distance between consecutive nucleosomes. The paradigm of one histone H1 molecule per nucleosome is no longer valid. It has been established that different cell types or cellular states differ in histone H1-to-nucleosome ratio, and that the total level of H1 affects the length of the linker DNA. Thus, high H1-to-nucleosome ratios are associated with longer NRLs. On the other hand, low amounts of H1 per nucleosome are associated with short spacing between nucleosomes. This is characteristic of active chromatin domains or rapidly growing cells, such as ESCs (ratio ~ 0.5), unlike mature cells in which chromatin is more compact (ratio ~ 0.8) [19,20,40–42]. The effect of different H1 variants on nucleosome spacing has been challenged by cytosolic microinjection of chicken and human H1 subtype mRNAs into *Xenopus* oocytes. It was demonstrated that H1 subtypes differ in their effect on nucleosome spacing in vivo, suggesting that they have different roles in the organization of the chromatin fiber [43]. Moreover, in breast cancer cells, depletion of H1.2, but not of other variants, caused a decrease in NRL, pointing to specificities of the variants in chromatin organization [44]. Hence, more evidence is required to unequivocally establish whether different H1 variants contribute differently to nucleosome spacing.

3.5. Variant-specific histone H1 PTMs and protein–protein interactions

Like core histones and most other proteins, H1 subtypes are post-translationally modified, mainly by phosphorylation, but also by acetylation, methylation, ubiquitination, formylation, citrullination and PARylation [46–49]. Some modifications are specific to a given H1 variant, as the involved residue is not conserved among variants, and may mediate interactions with specific protein partners. Although some research groups have started to perform proteomic analysis on H1 [50,51], the complete interactome for the different H1 variants has yet to be established. However, several studies have revealed that H1 variants interact with different partners, and these different associations could account for some variant-specific histone H1 functions. For instance, the transcription factor Msx1 interacts with mouse H1.5 to inhibit *MyoD* transcription and muscle differentiation [52]. H1.5 also interacts with FoxP3 via the leucine zipper domain to alter its FoxP3 binding to target genes, modulating gene expression and programming Treg function [53]. Furthermore, it has been proposed that H1.2 is part of a complex that acts as a repressor of p53-mediated transcription through modulation of chromatin remodeling. The association of H1.2 with p53 is disrupted upon DNA damage by H1.2 phosphorylation in T146 by DNA-PK, followed by p300-mediated p53 acetylation, resulting in increased transcription [54,55]. Several reports have also noted the importance of H1.4 modifications in K26 and S27 for HP1 binding and heterochromatin formation [56–58]. Further, pioneering proteomic analysis of H1 reveals that H1.0 interacts with an extensive network of proteins related to the nucleoli [50,51]. All these studies on H1 variant modifications and interactions with different partners, discussed in much more detail in the reviews in this current special issue by Izzo et al. and Kalashnikova et al., respectively, indicate that H1 variants may present different functions or locations based on these differences.

3.6. Phenotypic effects of H1 variant expression manipulation

As mentioned before, manipulating the abundance of linker histone H1 has been the main strategy for studying the function of H1 and its variants. A summary of the phenotypic effects observed after individual H1 variant deregulation in different species is presented in Table 2. The observations in many species that an adequate amount of linker histone H1 is essential for embryonic viability [19,59,60] and that H1 depletion impairs differentiation of ESCs [29,61] point to this histone playing a crucial role in many aspects of cellular function. However, the effects of depletion of an individual H1 variant are not so drastic, probably due to compensatory mechanisms involving other subtypes [44,62–71]. Still, several groups have reported a direct relation between a lack of specific somatic variants and cell cycle progression in various different species. In human IMR90 fibroblasts, H1.5 is required for normal cell growth [45]. A similar observation has been reported for H1.2 and H1.4 in T47D breast cancer cells, but not for the other variants. H1.2 knock-down impairs the proliferation of these cells by altering the expression of some cell cycle-related genes and arresting the cell cycle in the G1 phase [44]. In the same work, the use of an inducible system for the expression of H1 variant-specific interfering short-hairpin RNAs did not allow the cells to compensate for the lack of a variant by inducing other variants. Only the replacement variant H1.0 was up-regulated when other variants (H1.2 or H1.4) were knocked-down [44]. The fact that this occurs shortly after knock-down is complete suggests that there is transcriptional regulation via crosstalk between the different H1 genes and this would be an interesting area for further research.

In other studies, H1 variant overexpression has been observed to impair normal cell growth. Specifically, H1.3 overexpression inhibits ovarian cancer cell proliferation and colony formation capacity [72].

3.7. Effects of altering H1 variant levels on global gene expression

Given its structural role in mediating higher-order chromatin structure formation, histone H1 could be seen as a component mediating suppression of DNA-mediated processes such as gene transcription. Nonetheless, histone H1 depletion in various different biological systems has shown that, instead of general gene de-repression, a limited number of genes are up- or down-regulated. This suggests that H1 is not merely a general transcriptional repressor, but has a role in gene expression regulation in both senses and may also have a positive role in the transcription of particular genes. Moreover, global gene expression analyses in various cell types have revealed that histone H1 variants control the expression of different subsets of genes, indicating a specific role of H1 variants in gene regulation. For example, this was found when studying the effect on global gene expression of overproduction of H1.2 (H1c) and H1.0 in 3 T3 cells [73]. Many of the genes were uniquely targeted either by H1.2 or H1.0, indicating H1 variant specificity. Interestingly, H1.0 repressed more genes than H1.2, supporting the idea that H1.0 is a stronger repressor of transcription. In human breast cancer cells, individual inducible knock-down of each of the H1 variants caused deregulation of a very limited fraction of the genome [44,74]. A small proportion of these genes were affected by the depletion of more than one variant, suggesting a redundant or indirect role of the variants in those genes. However, most of genes were affected only by one variant, pointing to variant-specific regulation. Moreover, the proportion of genes down- versus up-regulated also differed between variants, from 1:1 for H1.5 to 2.7:1 for H1.2, suggesting that H1.2 may have a positive role in the expression of certain genes. As mentioned above, many genes deregulated upon H1.2 knock down were related to cell cycle progression.

Regarding this question, it would be interesting to determine how H1 variants regulate the expression of particular genes. It could be that H1 variants are locally enriched at certain promoters or in regulatory regions in order to control the expression of the associated gene. So

Please cite this article as: L. Millán-Ariño, et al., Specificities and genomic distribution of somatic mammalian histone H1 subtypes, Biochim. Biophys. Acta (2015), http://dx.doi.org/10.1016/j.bbagrm.2015.10.013

far, no evidence has been found of enrichment of a specific variant at the promoter of genes deregulated upon depletion of that variant [74]. Nonetheless, one study described a functional interaction of H1.2 with Cul4A E3 ubiquitin ligase, PAF1 elongation complexes and the serine 2 phosphorylated form of RNA polymerase II, indicating that H1.2 may be involved in the elongation process [75]. Alternatively, instead of selective enrichment of particular variants at certain promoters or regulatory regions, changes in the state of certain H1 variant PTMs induced by modifying enzymes located specifically at particular locations could be significant. Finally, we cannot rule out that deregulation of certain genes upon H1 depletion may be caused by indirect effects as a result of cellular changes such as cell cycle arrest, or changes in nucleosome spacing or chromatin compaction.

4. Genomic distribution of somatic histone H1 variants

To fully understand the biology of histone H1 and whether its variants locate distinctly, which might reflect specific functions, several groups have sought to explore the genomic distribution of H1 in vivo. However, due to the current lack of specific chromatin immunoprecipitation (ChIP)-grade antibodies for most of the H1 variants, the precise mapping of H1 variants in the genome has been challenging.

4.1. Nuclear localization

Initial biochemical and microscope approaches pointed to a non-uniform distribution of H1 in the cell nucleus and indicated differences between variants. Specifically, a first analysis by indirect immunofluorescence approaches with specific polyclonal antibodies showed H1.5 localizing at the periphery of the nucleus, where chromatin is more compact [76]. In a subsequent study, the same research group showed that H1.2 was distributed in parallel with DNA concentration, and H1.3 and H1.4 presented a punctuated pattern [77]. Sometime later, ChIP studies and PCR analysis of selected loci in human fetal fibroblasts confirmed previous observations, suggesting that active chromatin is depleted of total histone H1. Moreover, some authors reported differences in the distribution of the somatic H1s on active versus inactive genes and heterochromatic regions [78,79]. In particular, active and poised chromatin was characterized by H1.3 and H1.4 depletion, while all variants were present within inactive genes and heterochromatin. Unfortunately, antibodies used in those early studies are no longer available and subsequently different strategies have been adopted. A study using fluorescence microscopy showed that H1 variants fused to GFP at their N-terminal region were differently associated with euchromatin and heterochromatin in mouse fibroblast cells: while H1.0, H1.1, H1.2 and H1.3 were more closely associated with euchromatin, H1.4 and H1.5 were preferentially located within heterochromatin [35].

4.2. Genome-wide analysis of H1 distribution

Despite all these previous observations supporting a differential distribution of H1 variants in the genome, further analysis was needed because of the low resolution of the data. In particular, extensive genome-wide localization of H1 variants was needed to provide high resolution data and the possibility of interrogating the whole genome instead of particular loci. During recent years, the explosion in high-throughput sequencing technologies in the study of chromatin has provided valuable information about the distribution of core histones in the genome and their post-translational modifications, as well as of transcription factors and other DNA-binding proteins. However, the study of H1 distribution in the genome has been more difficult because of the heterogeneity of this histone family. Moreover, histone H1 is expected to be found all over the genome in parallel to nucleosome occupancy and is highly dynamic, in contrast to the discrete and defined position of transcription factors and core histone PTMs, making this mission even more challenging. Despite this, the first total H1 and H1 variant maps have started to emerge.

4.2.1. Total histone H1 distribution

Genome-wide studies with histone H1 started with ChIP-chip experiments in breast cancer MCF7 cells using an antibody for total H1 and custom promoter-containing arrays [80]. That work showed a clear depletion of H1 (an "H1 valley") near the transcription start site (TSS) of active genes, which was not seen in repressed promoters. Furthermore, as microarrays also contained ENCODE regions, other significant troughs and peaks of H1 could be observed within intergenic regions. Moreover, it was proposed that H1 PARylation by PARP-1 mediates H1 displacement from promoters, leading to chromatin remodeling and transcription activation. In fact, PARP-1 and H1 are mutually exclusive in active promoters and depletion of PARP-1 increases the binding of H1 at many target regions [80,81]. In support of this observation, activation of PARP-1 by CDK2 contributes to the displacement of histone H1 from progesterone responsive promoters in breast cancer cells [82]. PARP-1 also acts as a coactivator of GATA3 in breast cancer cells to regulate CCND1 transcription by ejecting H1 from the promoter [83].

H1 genome-wide mapping was also achieved for the unique somatic H1 in *Drosophila* [84]. By performing DNA adenine methyltransferase identification (DamID) coupled with microarray hybridization in Kc167 cells, it was shown that H1 was bound throughout the genome without significant differences between euchromatin and heterochromatin. However, like in the previous report on MCF7 cells, H1 was excluded from both active promoters and other intergenic regulatory regions. Interestingly, these authors also showed that H3.3 binding in promoters is inversely correlated with H1 presence, suggesting that H3.3 may contribute to H1 exclusion from promoters to maintain chromatin in an open state when transcription must take place. H1 binding to DNA is also negatively correlated with the occupancy of the chromatin architectural HMG proteins in Drosophila S2 cells [85], which have similar DNA and chromatin binding properties to H1 [86–88]. HMGs relax higher-order chromatin structures to facilitate the binding of nuclear regulatory factors to their binding sites, and the mutually exclusive binding with H1 to these sites controls specific transcriptional programs.

4.2.2. Specific distribution of histone H1 variants

Despite previous approaches to studying the distribution of H1 in the genome, and given the specific functions that H1 subtypes are now known to play in various processes, it is important to elucidate the precise mapping of individual variants within the genome in order to understand their contribution to chromatin organization and gene regulation. To achieve this goal, some groups recently succeeded in obtaining the first maps of H1 variant distribution within the human or mouse genomes (Fig. 1). Specifically, the genome-wide distribution of human H1.5 in IMR90 fibroblasts analyzed by ChIP-seq with a variant-specific antibody revealed that this variant presents blocks of enrichment in genic and intergenic regions of differentiated human cells, but not in ESCs, suggesting that the H1.5 pattern depends on the cellular differentiation state [45]. Moreover, H1.5 target genes are enriched in gene families which are clustered together in the genome and are in a transcriptionally repressed state. Moreover, gene repression is associated with H1.5 binding, and this variant is necessary for SIRT1 binding, H3K9me2 enrichment, and chromatin compaction in these cells. Comparing the distributions of H1.5 and H1.3, these features seemed to be specific to H1.5, although H1.3 antibody validation was not reported.

The first attempt to systematically compare the genome-wide distribution of different H1 variants focused on the mouse H1.2 (H1c) and H1.3 (H1d) variants [89]. ChIP-seq analysis of N-terminal tagged (with Myc or FLAG) H1.2 and H1.3 variants in knock-in mouse ESCs showed depletion of these variants from GC- and gene-rich regions, and from active promoters, presenting the characteristic "H1 valley" around the TSS. Both variants presented a positive and negative correlation with

Please cite this article as: L. Millán-Ariño, et al., Specificities and genomic distribution of somatic mammalian histone H1 subtypes, Biochim. Biophys. Acta (2015), http://dx.doi.org/10.1016/j.bbagrm.2015.10.013

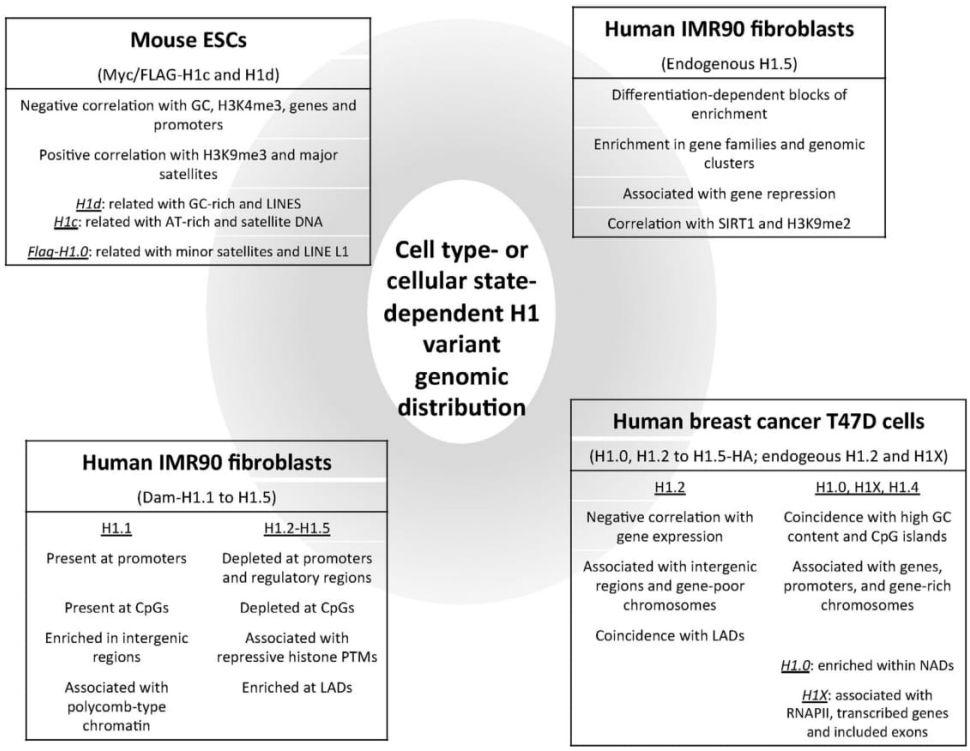


Fig. 1. Summary of recent reports on the genome-wide distribution of mammalian histone H1 variants, focusing on the differential distribution of variants within each cellular model. Information was extracted from [45,74,89–91].

H3K9me3 and H3K4me3, respectively, and they were overrepresented at major satellites. The H1 enrichment at major satellites seemed to contribute to the increased NRL observed in pericentromeric regions compared to that in bulk chromatin. When unique peaks for H1.2 and H1.3 were compared, H1.3 was more related to GC-rich sequences and LINES, whereas H1.2 was more closely associated with AT-rich sequences, Giemsa positive regions and satellite DNA. Finally, the genomic distribution of overexpressed FLAG-tagged H1.0 mostly resembled that of H1.2 and H1.3, although, besides overrepresentation at major satellites, it was also enriched at minor satellites and LINE L1 elements. This suggests notably differential binding preferences for this variant. However, as H1.0 is present at low levels in undifferentiated wild-type ESCs, to confirm these differences it would be necessary to extend these observations to differentiated cells.

Another group exploited DamID technology to map human H1.1 to H1.5 variants in human lung IMR90 fibroblasts [90]. These researchers showed that H1.2 to H1.5 were similarly distributed and were depleted from CpG-dense regions and active regulatory regions. All these variants showed dips at promoters, enhancers, and CTCF binding sites. Furthermore, H1 abundance was negatively correlated with "active" histone marks and positively correlated with "repressive" ones. Interestingly, the observation that most H1s are overrepresented at lamin-associated domains (LADs) and that different combinations of H1 variants associated with functionally distinct topological domains pointed to a possible role of H1 in the three-dimensional organization of the genome, with different H1 variants contributing to the establishment of particular chromatin states. Finally, it is worth noting that H1.1 showed a distinct binding profile from all other variants (H1.2 to H1.5), suggesting a special role of this subtype in chromatin function in those cells. Specifically, it was more abundant at promoters and CpGs than other variants, was not depleted from regulatory regions, and was enriched in intergenic regions. Moreover, it was not associated with LADs, but it was found enriched in polycomb-type chromatin domains.

The distribution of six somatic H1 subtypes has also been studied in the human breast cancer cell line T47D comparing available variant-specific antibodies (H1.2 and H1X) with hemagglutinin (HA)-tagged recombinant H1 variants (H1.2-H1.5 and H1.0). By combining ChIP-chip and ChIP-seq experiments, it was shown that the genomic distribution of H1.2 differed from that of the other variants [91]. H1.2 was the variant that best correlated with low gene expression and low GC content and was also more closely associated with LADs. Comparatively, other H1 variants tended to be associated with higher GC content, CpG islands, and gene-rich regions and chromosomes. Additionally, although an H1 valley was present at active promoters for all variants, H1.2 was also depleted around TSS of repressed genes. H1 valleys were shown to affect promoters more extensively than over a single nucleosome free region upstream of TSS, indicating that H1 removal from promoters is necessary to accommodate the RNA polymerase machinery and promote transcription. Finally, comparison of the relative abundance of endogenous H1.2 and H1X at specific loci in different cell lines (beyond T47D) showed that the abundance of H1 variants at those loci is not conserved between cell lines, pointing to a differential genomic distribution of H1 variants between cell types, probably related to their relative H1 variant protein abundance.

A subsequent study examined in more detail the genomic distribution of H1.0 and H1X, the two variants most structurally distant within the somatic H1 family [74]. H1X was found to be enriched in active chromatin and hence there was found to be a large overlap with RNA polymerase II-enriched regions, coding regions and hypomethylated CpG islands, the highest enrichment being toward the 3′ end of highly expressed genes. Further, H1X was shown to be enriched within constitutively and included alternatively spliced exons and retained introns in the breast cancer cell line T47D. Taking into account these observations,

Please cite this article as: L. Millán-Ariño, et al., Specificities and genomic distribution of somatic mammalian histone H1 subtypes, Biochim. Biophys. Acta (2015), http://dx.doi.org/10.1016/j.bbagrm.2015.10.013

it was suggested that H1X could be an additional player in the functional interconnections between chromatin structure, transcriptional elongation and alternative splicing regulation, in the breast cancer cell line analyzed. A link between H1.2 and elongation has already been mentioned above [75]. Another study using liquid chromatography–tandem mass spectrometry in four different cell lines found that core-splicing factors were H1.0-binding proteins [50], again pointing to a possible role of H1 in splicing, as has already been shown for several chromatin-binding proteins and core histone PTMs [92–94]. A way to reconcile all these reports would be to assume that different variants play distinct roles in different cell types.

Moreover, for the other replication-independent H1 variant, H1.0, an accumulation was shown in nucleolus-associated DNA, including nucleolus associated domains (NADs), ribosomal DNA, and acromeric and telomeric satellites [74]. Cellular fractionation confirmed enrichment of H1.0 in the nucleolus, in agreement with Kalashnikova et al. who showed that nearly all H1.0-binding proteins are found in this nuclear subdomain [50]. These authors suggested that H1.0 might be involved in the stabilization of perinucleolar late-replicating heterochromatin. It is worth noting that other H1 variants have also been identified in proteomic profiling of the human nucleolus [95], and some PTMs of H1.2 and H1.4 are found in nucleoli-associated processes such as RNA Pol I activity and rRNA biogenesis [96]. In addition, an H1X accumulation in the nucleoli was found in the G1 phase of the cell cycle although it was suggested that H1X is mainly located at inactive ribosomal genes [97,98]. Hence, it will be interesting to analyze whether there is a specific nucleolar location and/or function of H1 variants that might differ between cell lines.

Most of the aforementioned studies used tagged recombinant forms of H1 variants, either with small epitope tags or the Dam domain, to elucidate their genomic distribution. Thus, after the reporting of few differences in the occupancy of endogenous versus HA-tagged H1.2 [91], it is still important to be careful when interpreting the results. H1 variant overexpression or structural protein changes due to the tagging could to some extent affect the genomic localization of these exogenous proteins. Nonetheless, until better antibodies or other techniques become available, this remains the best approach. Moreover, Cao et al. have shown that knocked-in tagged H1 proteins are functionally equivalent to the endogenous H1 in vivo [89].

Overall, by combining the data of all these studies and comparing what is known about the genomic distribution of H1 subtypes in different cell types and conditions, it seems that the occupancy of a given subtype, although maintaining a general pattern, differs slightly from cell line to cell line and between conditions (stem cells vs. differentiated cells vs. cancer cells). A clear example of this is the fact that H1.5 distribution differs in hESCs and differentiated fibroblasts, and compact blocks of H1.5 enrichment arise only when cells differentiate [45]. This would suggest that H1 could be involved in the establishment of peripheral heterochromatin during the process of differentiation. In fact, in two other studies, H1 has been found to be associated with LADs, also suggesting that H1 plays a role in spatial chromatin organization [90,91]. This variability in the H1 genomic distribution during differentiation could be related to the relative abundance of H1 variants at different stages of differentiation [29].

Additionally, by comparing studies in normal fibroblasts [90] and breast cancer cells [91], it seems that different groups of H1 variants present different distributions in each case, the variant with the most specific pattern being H1.1 in IMR90 cells, but H1.2 in breast cancer cells. In breast cancer cells, H1.2 presented a genomic distribution more associated with a repressive chromatin environment, similar to the distribution of H1s other than H1.1 in fibroblasts. On the other hand, the H1.1 distribution in fibroblasts resembled, to some extent, the distribution of H1 variants other than H1.2 in breast cancer cells.

Similarly, a ChIP study of the presence of the six chicken histone H1 subtypes and the variant histone H5 across selected loci showed that active genes carry high levels of some but not all linker histone subtypes, indicating that they are not intrinsically inhibitory to gene expression, and, again, that there are differences between cell types [99].

To sum up, the first approaches to studying H1 variant genomic distribution indicate that single variants may present distinct features in different cell types or conditions, rather than having intrinsic properties that would confer them with a universal genomic binding pattern. To prove that, further studies specifically comparing the H1 variant distribution in various cell types and conditions should be performed. This observation could also be partially explained by the existence of specific histone H1 PTMs that would provide differential histone binding ability to different parts of the genome. Thus, other genome-wide studies focused on the distribution of histone H1 PTMs would help to clarify this diversity in the genomic distribution of linker histone H1. To date, only one report has emerged focusing on a specific acetylation on H1.4 [39]. Kamieniarz et al. showed that H1.4-K34 acetylation by GCN5 is associated with promoters of active genes and regulatory regions. This modification seems to positively regulate transcription, both by increasing H1 mobility and by recruiting transcription factors (TAF1). Interestingly, they also showed that H1.4-K34 acetylation is dynamic during spermatogenesis and marks human seminomas.

5. Concluding remarks

Although histone H1 variants show partial redundancy, the most recent research highlights that they also have specific functions in certain cellular processes and they present differences regarding regulation of gene expression and/or chromatin compaction, which can be explained by differential structural properties, interaction with specific partners due, in part, to differential post-translational modifications, or a heterogeneous genomic distribution. Based on the many studies described herein, one can conclude that histone H1 variants are differentially involved in transitions between different cellular states, such as development and differentiation or during cancer progression. Thus, further progress in studying the properties of individual H1 variants might help us to understand how cellular identity is determined, but also how some diseases such as cancer arise, with potential implications for prognosis and treatment.

Regarding their genomic distribution, it is conceivable that rearrangement of H1 variant distribution in the genome as a result of imbalanced relative expression and/or specific PTMs is related to chromatin reorganization and, particularly, the establishment of peripheral heterochromatic regions (LADs and LOCKs) observed during the differentiation process [100]. It has recently been proposed that this chromatin rearrangement, together with changes in the DNA methylation status in these regions, mediates the transition between developmental processes by conferring differential genomic plasticity to the cells [101]. Similarly, differential H1 variant distribution in the genome could be associated with cancer progression in a similar manner by reorganizing the chromatin architecture and conferring cancer cells with a greater hypervariability that would favor tumor cell heterogeneity. This epigenetic reprogramming has already been related to different cancer types [102] and also to the epithelial–mesenchymal transition [103,104]. Thus, H1 variants appear to be potential targets for cancer treatment or, at least, a way to understand how genome plasticity may be altered during the course of cancer and other diseases.

Hence, it will be important in the future to extend the precise mapping of H1 variant distribution to other cell types and conditions, because the relative abundance of these variants in a given cell line or the cellular state (i.e. undifferentiated vs. differentiated vs. tumorigenic) could influence on their distribution in the genome. This will definitively help us to understand the function of this histone and its variants, and improve our understanding of chromatin organization and regulation.

Transparency document

The Transparency document associated with this article can be found, in the online version.

Please cite this article as: L. Millán-Ariño, et al., Specificities and genomic distribution of somatic mammalian histone H1 subtypes, Biochim. Biophys. Acta (2015), http://dx.doi.org/10.1016/j.bbagrm.2015.10.013

Acknowledgements

We thank members of the lab for their feedback on the manuscript. This work was supported by the Spanish Ministry of Science and Innovation (MICINN) and the European Regional Development Fund (grant BFU2014-52237).

References

[1] J.M. Kinkade Jr., R.D. Cole, A structural comparison of different lysine-rich histones of calf thymus, J. Biol. Chem. 241 (1966) 5798–5805.
[2] H.E. Kasinsky, J.D. Lewis, J.B. Dacks, J. Ausio, Origin of H1 linker histones, FASEB J. 15 (2001) 34–42.
[3] P.B. Talbert, K. Ahmad, G. Almouzni, J. Ausio, F. Berger, P.L. Bhalla, W.M. Bonner, W.Z. Cande, B.P. Chadwick, S.W. Chan, G.A. Cross, L. Cui, S.I. Dimitrov, D. Doenecke, J.M. Eirin-Lopez, M.A. Gorovsky, S.B. Hake, B.A. Hamkalo, S. Holec, S.E. Jacobsen, K. Kamieniarz, S. Khochbin, A.G. Ladurner, D. Landsman, J.A. Latham, B. Loppin, H.S. Malik, W.F. Marzluff, J.R. Pehrson, J. Postberg, R. Schneider, M.B. Singh, M.M. Smith, E. Thompson, M.E. Torres-Padilla, D.J. Tremethick, B.M. Turner, J.H. Waterborg, H. Wollmann, R. Yelagandula, B. Zhu, S. Henikoff, A unified phylogeny-based nomenclature for histone variants, Epigenetics Chromatin 5 (2012) 7.
[4] S.W. Harshman, N.L. Young, M.R. Parthun, M.A. Freitas, H1 Histones: Current Perspectives and Challenges, Nucleic Acids Res. (2013).
[5] N. Happel, D. Doenecke, Histone H1 and its isoforms: contribution to chromatin structure and function, Gene 431 (2009) 1–12.
[6] A. Izzo, K. Kamieniarz, R. Schneider, The histone H1 family: specific members, specific functions? Biol. Chem. 389 (2008) 333–343.
[7] S. Perez-Montero, A. Carbonell, F. Azorin, Germline-specific h1 variants: the "sexy" linker histones, Chromosoma (2015).
[8] N. Happel, J. Warneboldt, K. Hanecke, F. Haller, D. Doenecke, H1 subtype expression during cell proliferation and growth arrest, Cell Cycle 8 (2009) 2226–2232.
[9] L.S. Coles, J.R. Wells, An H1 histone gene-specific 5′ element and evolution of H1 and H5 genes, Nucleic Acids Res. 13 (1985) 585–594.
[10] D. Doenecke, W. Albig, H. Bouterfa, B. Drabent, Organization and expression of H1 histone and H1 replacement histone genes, J. Cell. Biochem. 54 (1994) 423–431.
[11] A. Eilers, H. Bouterfa, S. Triebe, D. Doenecke, Role of a distal promoter element in the S-phase control of the human H1.2 histone gene transcription, Eur. J. Biochem./FEBS 223 (1994) 567–574.
[12] D. Gokhman, I. Livyatan, B.S. Sailaja, S. Melcer, E. Meshorer, Multilayered chromatin analysis reveals E2f, Smad and Zfx as transcriptional regulators of histones, Nat. Struct. Mol. Biol. 20 (2013) 119–126.
[13] D.A. Hill, Influence of linker histone H1 on chromatin remodeling, Biochem. Cell Biol. 79 (2001) 317–324.
[14] M.S. Schlissel, D.D. Brown, The transcriptional regulation of Xenopus 5 s RNA genes in chromatin: the roles of active stable transcription complexes and histone H1, Cell 37 (1984) 903–913.
[15] M. Nishiyama, A.I. Skoultchi, K.I. Nakayama, Histone H1 recruitment by CHD8 is essential for suppression of the Wnt-beta-catenin signaling pathway, Mol. Cell. Biol. 32 (2012) 501–512.
[16] M. Nishiyama, K. Oshikawa, Y. Tsukada, T. Nakagawa, S. Iemura, T. Natsume, Y. Fan, A. Kikuchi, A.I. Skoultchi, K.I. Nakayama, CHD8 suppresses p53-mediated apoptosis through histone H1 recruitment during early embryogenesis, Nat. Cell Biol. 11 (2009) 172–182.
[17] J.A. Maclean, A. Bettegowda, B.J. Kim, C.H. Lou, S.M. Yang, A. Bhardwaj, S. Shanker, Z. Hu, Y. Fan, S. Eckardt, K.J. McLaughlin, A.I. Skoultchi, M.F. Wilkinson, The rhox homeobox gene cluster is imprinted and selectively targeted for regulation by histone h1 and DNA methylation, Mol. Cell. Biol. 31 (2011) 1275–1287.
[18] Y. Fan, A. Sirotkin, R.G. Russell, J. Ayala, A.I. Skoultchi, Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the H1(0) replacement subtype, Mol. Cell. Biol. 21 (2001) 7933–7943.
[19] Y. Fan, T. Nikitina, E.M. Morin-Kensicki, J. Zhao, T.R. Magnuson, C.L. Woodcock, A.I. Skoultchi, H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo, Mol. Cell. Biol. 23 (2003) 4559–4572.
[20] Y. Fan, T. Nikitina, J. Zhao, T.J. Fleury, R. Bhattacharyya, E.E. Bouhassira, A. Stein, C.L. Woodcock, A.I. Skoultchi, Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation, Cell 123 (2005) 1199–1212.
[21] J.M. Eirin-Lopez, A.M. Gonzalez-Tizon, A. Martinez, J. Mendez, Birth-and-death evolution with strong purifying selection in the histone H1 multigene family and the origin of orphon H1 genes, Mol. Biol. Evol. 21 (2004) 1992–2003.
[22] I. Ponte, J.M. Vidal-Taboada, P. Suau, Evolution of the vertebrate H1 histone class: evidence for the functional differentiation of the subtypes, Mol. Biol. Evol. 15 (1998) 702–708.
[23] R.W. Lennox, L.H. Cohen, The histone H1 complements of dividing and nondividing cells of the mouse, J. Biol. Chem. 258 (1983) 262–268.
[24] T. Meergans, W. Albig, D. Doenecke, Varied expression patterns of human H1 histone genes in different cell lines, DNA Cell Biol. 16 (1997) 1041–1049.
[25] B. Pina, P. Suau, Changes in the proportions of histone H1 subtypes in brain cortical neurons, FEBS Lett. 210 (1987) 161–164.
[26] M.H. Parseghian, B.A. Hamkalo, A compendium of the histone H1 family of somatic subtypes: an elusive cast of characters and their characteristics, Biochem. Cell Biol. 79 (2001) 289–304.
[27] W. Helliger, H. Lindner, O. Grubl-Knosp, B. Puschendorf, Alteration in proportions of histone H1 variants during the differentiation of murine erythroleukaemic cells, Biochem. J. 288 (Pt 3) (1992) 747–751.
[28] J. Zlatanova, D. Doenecke, Histone H1 zero: a major player in cell differentiation? FASEB J. 8 (1994) 1260–1268.
[29] J.M. Terme, B. Sese, L. Millan-Arino, R. Mayor, J.C. Izpisua Belmonte, M.J. Barrero, A. Jordan, Histone H1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency, J. Biol. Chem. 286 (2011) 35347–35357.
[30] M. Medrzycki, Y. Zhang, J.F. McDonald, Y. Fan, Profiling of linker histone variants in ovarian cancer, Front. Biosci. 17 (2012) 396–406.
[31] J.F. Hechtman, M.B. Beasley, Y. Kinoshita, H.M. Ko, K. Hao, D.E. Burstein, Promyelocytic leukemia zinc finger and histone H1.5 differentially stain low- and high-grade pulmonary neuroendocrine tumors: a pilot immunohistochemical study, Hum. Pathol. 44 (2013) 1400–1405.
[32] M.A. Lever, J.P. Th'ng, X. Sun, M.J. Hendzel, Rapid exchange of histone H1.1 on chromatin in living human cells, Nature 408 (2000) 873–876.
[33] T. Misteli, A. Gunjan, R. Hock, M. Bustin, D.T. Brown, Dynamic binding of histone H1 to chromatin in living cells, Nature 408 (2000) 877–881.
[34] P. Vyas, D.T. Brown, N- and C-terminal domains determine differential nucleosomal binding geometry and affinity of linker histone isotypes H1(0) and H1c, J. Biol. Chem. 287 (2012) 11778–11787.
[35] J.P. Th'ng, R. Sung, M. Ye, M.J. Hendzel, H1 family histones in the nucleus. Control of binding and localization by the C-terminal domain, J. Biol. Chem. 280 (2005) 27809–27814.
[36] M. Orrego, I. Ponte, A. Roque, N. Buschati, X. Mora, P. Suau, Differential affinity of mammalian histone H1 somatic subtypes for DNA and chromatin, BMC Biol. 5 (2007) 22.
[37] J. Clausell, N. Happel, T.K. Hale, D. Doenecke, M. Beato, Histone H1 subtypes differentially modulate chromatin condensation without preventing ATP-dependent remodeling by SWI/SNF or NURF, PLoS One 4 (2009), e0007243.
[38] S.P. Hergeth, M. Dundr, P. Tropberger, B.M. Zee, B.A. Garcia, S. Daujat, R. Schneider, Isoform-specific phosphorylation of human linker histone H1.4 in mitosis by the kinase Aurora B, J. Cell Sci. 124 (2011) 1623–1628.
[39] K. Kamieniarz, A. Izzo, M. Dundr, P. Tropberger, L. Ozretic, J. Kirfel, E. Scheer, P. Tropel, J.R. Wisniewski, L. Tora, S. Viville, R. Buettner, R. Schneider, A dual role of linker histone H1.4 Lys 34 acetylation in transcriptional activation, Genes Dev. 26 (2012) 797–802.
[40] H. Hashimoto, Y. Takami, E. Sonoda, T. Iwasaki, H. Iwano, M. Tachibana, S. Takeda, T. Nakayama, H. Kimura, Y. Shinkai, Histone H1 null vertebrate cells exhibit altered nucleosome architecture, Nucleic Acids Res. 38 (2010) 3533–3545.
[41] C.L. Woodcock, A.I. Skoultchi, Y. Fan, Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length, Chromosom. Res. 14 (2006) 17–25.
[42] A. Gunjan, B.T. Alexander, D.B. Sittman, D.T. Brown, Effects of H1 histone variant overexpression on chromatin structure, J. Biol. Chem. 274 (1999) 37950–37956.
[43] C. Oberg, A. Izzo, R. Schneider, O. Wrange, S. Belikov, Linker histone subtypes differ in their effect on nucleosomal spacing in vivo, J. Mol. Biol. 419 (2012) 183–197.
[44] M. Sancho, E. Diani, M. Beato, A. Jordan, Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth, PLoS Genet. 4 (2008), e1000227.
[45] J.Y. Li, M. Patterson, H.K. Mikkola, W.E. Lowry, S.K. Kurdistani, Dynamic distribution of linker histone H1.5 in cellular differentiation, PLoS Genet. 8 (2012), e1002879.
[46] B.A. Garcia, S.A. Busby, C.M. Barber, J. Shabanowitz, C.D. Allis, D.F. Hunt, Characterization of phosphorylation sites on histone H1 isoforms by tandem mass spectrometry, J. Proteome Res. 3 (2004) 1219–1227.
[47] J.R. Wisniewski, A. Zougman, S. Kruger, M. Mann, Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue, Mol. Cell. Proteomics 6 (2007) 72–87.
[48] C. Wood, A. Snijders, J. Williamson, C. Reynolds, J. Baldwin, M. Dickman, Post-translational modifications of the linker histone variants and their association with cell mechanisms, FEBS J. 276 (2009) 3685–3697.
[49] M.A. Christophorou, G. Castelo-Branco, R.P. Halley-Stott, C.S. Oliveira, R. Loos, A. Radzisheuskaya, K.A. Mowen, P. Bertone, J.C. Silva, M. Zernicka-Goetz, M.L. Nielsen, J.B. Gurdon, T. Kouzarides, Citrullination regulates pluripotency and histone H1 binding to chromatin, Nature 507 (2014) 104–108.
[50] A.A. Kalashnikova, D.D. Winkler, S.J. McBryant, R.K. Henderson, J.A. Herman, J.G. DeLuca, K. Luger, J.E. Prenni, J.C. Hansen, Linker histone H1.0 interacts with an extensive network of proteins found in the nucleolus, Nucleic Acids Res. 41 (2013) 4026–4035.
[51] H.J. Szerlong, J.A. Herman, C.M. Krause, J.G. DeLuca, A. Skoultchi, Q.A. Winger, J.E. Prenni, J.C. Hansen, Proteomic characterization of the nucleolar linker histone h1 interaction network, J. Mol. Biol. 427 (2015) 2056–2071.
[52] H. Lee, R. Habas, C. Abate-Shen, MSX1 cooperates with histone H1b for inhibition of transcription and myogenesis, Science 304 (2004) 1675–1678.
[53] S.L. Mackey-Cushman, J. Gao, D.A. Holmes, J.I. Nunoya, R. Wang, D. Unutmaz, L. Su, FoxP3 interacts with linker histone H1.5 to modulate gene expression and program Treg cell activity, Genes Immun. 12 (2011) 559–567.
[54] K. Kim, J. Choi, K. Heo, H. Kim, D. Levens, K. Kohno, E.M. Johnson, H.W. Brock, W. An, Isolation and characterization of a novel H1.2 complex that acts as a repressor of p53-mediated transcription, J. Biol. Chem. 283 (2008) 9113–9126.
[55] K. Kim, K.W. Jeong, H. Kim, J. Choi, W. Lu, M.R. Stallcup, W. An, Functional interplay between p53 acetylation and H1.2 phosphorylation in p53-regulated transcription, Oncogene 31 (2012) 4290–4301.

Please cite this article as: L. Millán-Ariño, et al., Specificities and genomic distribution of somatic mammalian histone H1 subtypes, Biochim. Biophys. Acta (2015), http://dx.doi.org/10.1016/j.bbagrm.2015.10.013

[56] S. Daujat, U. Zeissler, T. Waldmann, N. Happel, R. Schneider, HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding, J. Biol. Chem. 280 (2005) 38090–38095.
[57] A. Vaquero, M. Scher, D. Lee, H. Erdjument-Bromage, P. Tempst, D. Reinberg, Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin, Mol. Cell 16 (2004) 93–105.
[58] T. Weiss, S. Hergeth, U. Zeissler, A. Izzo, P. Tropberger, B.M. Zee, M. Dundr, B.A. Garcia, S. Daujat, R. Schneider, Histone H1 variant-specific lysine methylation by G9a/KMT1C and Glp1/KMT1D, Epigenetics Chromatin 3 (2010) 7.
[59] X. Lu, S.N. Wontakal, A.V. Emelyanov, P. Morcillo, A.Y. Konev, D.V. Fyodorov, A.I. Skoultchi, Linker histone H1 is essential for Drosophila development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure, Genes Dev. 23 (2009) 452–465.
[60] O. Vujatovic, K. Zaragoza, A. Vaquero, O. Reina, J. Bernues, F. Azorin, Drosophila melanogaster linker histone dH1 is required for transposon silencing and to preserve genome integrity, Nucleic Acids Res. 40 (2012) 5402–5414.
[61] Y. Zhang, M. Cooke, S. Panjwani, K. Cao, B. Krauth, P.Y. Ho, M. Medrzycki, D.T. Berhe, C. Pan, T.C. McDevitt, Y. Fan, Histone h1 depletion impairs embryonic stem cell differentiation, PLoS Genet. 8 (2012), e1002691.
[62] B. Drabent, P. Saftig, C. Bode, D. Doenecke, Spermatogenesis proceeds normally in mice without linker histone H1t, Histochem. Cell Biol. 113 (2000) 433–442.
[63] D.A. Fantz, W.R. Hatfield, G. Horvath, M.K. Kistler, W.S. Kistler, Mice with a targeted disruption of the H1t gene are fertile and undergo normal changes in structural chromosomal proteins during spermiogenesis, Biol. Reprod. 64 (2001) 425–431.
[64] H. Hashimoto, E. Sonoda, Y. Takami, H. Kimura, T. Nakayama, M. Tachibana, S. Takeda, Y. Shinkai, Histone H1 variant, H1R is involved in DNA damage response, DNA Repair (Amst) 6 (2007) 1584–1595.
[65] Q. Lin, A. Inselman, X. Han, H. Xu, W. Zhang, M.A. Handel, A.I. Skoultchi, Reductions in linker histone levels are tolerated in developing spermatocytes but cause changes in specific gene expression, J. Biol. Chem. 279 (2004) 23525–23535.
[66] Q. Lin, A. Sirotkin, A.I. Skoultchi, Normal spermatogenesis in mice lacking the testis-specific linker histone H1t, Mol. Cell. Biol. 20 (2000) 2122–2128.
[67] Y. Takami, R. Nishi, T. Nakayama, Histone H1 variants play individual roles in transcription regulation in the DT40 chicken B cell line, Biochem. Biophys. Res. Commun. 268 (2000) 501–508.
[68] Y. Fan, A. Sirotkin, R.G. Russell, J. Ayala, A.I. Skoultchi, Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the H1(0) replacement subtype, Mol. Cell. Biol. 21 (2001) 7933–7943.
[69] R. Alami, Y. Fan, S. Pack, T.M. Sonbuchner, A. Besse, Q. Lin, J.M. Greally, A.I. Skoultchi, E.E. Bouhassira, Mammalian linker-histone subtypes differentially affect gene expression in vivo, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 5920–5925.
[70] D.I. Gabrilovich, P. Cheng, Y. Fan, B. Yu, E. Nikitina, A. Sirotkin, M. Shurin, T. Oyama, Y. Adachi, S. Nadaf, D.P. Carbone, A.I. Skoultchi, H1(0) histone and differentiation of dendritic cells. A molecular target for tumor-derived factors, J. Leukoc. Biol. 72 (2002) 285–296.
[71] A.M. Sirotkin, W. Edelmann, G. Cheng, A. Klein-Szanto, R. Kucherlapati, A.I. Skoultchi, Mice develop normally without the H1(0) linker histone, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 6434–6438.
[72] M. Medrzycki, Y. Zhang, W. Zhang, K. Cao, C. Pan, N. Lailler, J.F. McDonald, E.E. Bouhassira, Y. Fan, Histone h1.3 suppresses h19 noncoding RNA expression and cell growth of ovarian cancer cells, Cancer Res. 74 (2014) 6463–6473.
[73] S. Bhan, W. May, S.L. Warren, D.B. Sittman, Global gene expression analysis reveals specific and redundant roles for H1 variants, H1c and H1(0), in gene expression regulation, Gene 414 (2008) 10–18.
[74] R. Mayor, A. Izquierdo-Bouldstridge, L. Millan-Arino, A. Bustillos, C. Sampaio, N. Luque, A. Jordan, Genome distribution of replication-independent histone H1 variants shows H1.0 associated with nucleolar domains and H1X associated with RNA polymerase II-enriched regions, J. Biol. Chem. 290 (2015) 7474–7491.
[75] K. Kim, B. Lee, J. Kim, J. Choi, J.M. Kim, Y. Xiong, R.G. Roeder, W. An, Linker histone H1.2 cooperates with Cul4A and PAF1 to drive H4K31 ubiquitylation-mediated transactivation, Cell Rep. 5 (2013) 1690–1703.
[76] M.H. Parseghian, R.F. Clark, L.J. Hauser, N. Dvorkin, D.A. Harris, B.A. Hamkalo, Fractionation of human H1 subtypes and characterization of a subtype-specific antibody exhibiting non-uniform nuclear staining, Chromosom. Res. 1 (1993) 127–139.
[77] M.H. Parseghian, D.A. Harris, D.R. Rishwain, B.A. Hamkalo, Characterization of a set of antibodies specific for three human histone H1 subtypes, Chromosoma 103 (1994) 198–208.
[78] M.H. Parseghian, R.L. Newcomb, B.A. Hamkalo, Distribution of somatic H1 subtypes is non-random on active vs. inactive chromatin II: distribution in human adult fibroblasts, J. Cell. Biochem. 83 (2001) 643–659.
[79] M.H. Parseghian, R.L. Newcomb, S.T. Winokur, B.A. Hamkalo, The distribution of somatic H1 subtypes is non-random on active vs. inactive chromatin: distribution in human fetal fibroblasts, Chromosom. Res. 8 (2000) 405–424.
[80] R. Krishnakumar, M.J. Gamble, K.M. Frizzell, J.G. Berrocal, M. Kininis, W.L. Kraus, Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes, Science 319 (2008) 819–821.
[81] R. Krishnakumar, W.L. Kraus, PARP-1 regulates chromatin structure and transcription through a KDM5B-dependent pathway, Mol. Cell 39 (2010) 736–749.
[82] R.H. Wright, G. Castellano, J. Bonet, F. Le Dily, J. Font-Mateu, C. Ballare, A.S. Nacht, D. Soronellas, B. Oliva, M. Beato, CDK2-dependent activation of PARP-1 is required for hormonal gene regulation in breast cancer cells, Genes Dev. 26 (2012) 1972–1983.
[83] L. Shan, X. Li, L. Liu, X. Ding, Q. Wang, Y. Zheng, Y. Duan, C. Xuan, Y. Wang, F. Yang, Y. Shang, L. Shi, GATA3 cooperates with PARP1 to regulate CCND1 transcription through modulating histone H1 incorporation, Oncogene (2013).
[84] U. Braunschweig, G.J. Hogan, L. Pagie, B. van Steensel, Histone H1 binding is inhibited by histone variant H3.3, EMBO J. 28 (2009) 3635–3645.
[85] N. Nalabothula, G. McVicker, J. Maiorano, R. Martin, J.K. Pritchard, Y.N. Fondufe-Mittendorf, The chromatin architectural proteins HMGD1 and H1 bind reciprocally and have opposite effects on chromatin structure and gene regulation, BMC Genomics 15 (2014) 92.
[86] S.S. Ner, T. Blank, M.L. Perez-Paralle, T.A. Grigliatti, P.B. Becker, A.A. Travers, HMG-D and histone H1 interplay during chromatin assembly and early embryogenesis, J. Biol. Chem. 276 (2001) 37569–37576.
[87] S.S. Ner, A.A. Travers, HMG-D, the Drosophila melanogaster homologue of HMG 1 protein, is associated with early embryonic chromatin in the absence of histone H1, EMBO J. 13 (1994) 1817–1822.
[88] J. Zlatanova, K. van Holde, Linker histones versus HMG1/2: a struggle for dominance? BioEssays 20 (1998) 584–588.
[89] K. Cao, N. Lailler, Y. Zhang, A. Kumar, K. Uppal, Z. Liu, E.K. Lee, H. Wu, M. Medrzycki, C. Pan, P.Y. Ho, G.P. Cooper Jr., X. Dong, C. Bock, E.E. Bouhassira, Y. Fan, High-resolution mapping of h1 linker histone variants in embryonic stem cells, PLoS Genet. 9 (2013), e1003417.
[90] A. Izzo, K. Kamieniarz-Gdula, F. Ramirez, N. Noureen, J. Kind, T. Manke, B. van Steensel, R. Schneider, The genomic landscape of the somatic linker histone subtypes H1.1 to H1.5 in human cells, Cell Rep. (2013).
[91] L. Millan-Arino, A.B. Islam, A. Izquierdo-Bouldstridge, R. Mayor, J.M. Terme, N. Luque, M. Sancho, N. Lopez-Bigas, A. Jordan, Mapping of six somatic linker histone H1 variants in human breast cancer cells uncovers specific features of H1.2, Nucleic Acids Res. 42 (2014) 4474–4493.
[92] J. Hnilicova, D. Stanek, Where splicing joins chromatin, Nucleus 2 (2011) 182–188.
[93] P. Kolasinska-Zwierz, T. Down, I. Latorre, T. Liu, X.S. Liu, J. Ahringer, Differential chromatin marking of introns and expressed exons by H3K36me3, Nat. Genet. 41 (2009) 376–381.
[94] A. Yearim, S. Gelfman, R. Shayevitch, S. Melcer, O. Glaich, J.P. Mallm, M. Nissim-Rafinia, A.H. Cohen, K. Rippe, E. Meshorer, G. Ast, HP1 is involved in regulating the global impact of DNA methylation on alternative splicing, Cell Rep. 10 (2015) 1122–1134.
[95] M.A. Jarboui, K. Wynne, G. Elia, W.W. Hall, V.W. Gautier, Proteomic profiling of the human T-cell nucleolus, Mol. Immunol. 49 (2011) 441–452.
[96] Y. Zheng, S. John, J.J. Pesavento, J.R. Schultz-Norton, R.L. Schiltz, S. Baek, A.M. Nardulli, G.L. Hager, N.L. Kelleher, C.A. Mizzen, Histone H1 phosphorylation is associated with transcription by RNA polymerases I and II, J. Cell Biol. 189 (2010) 407–415.
[97] S. Stoldt, D. Wenzel, E. Schulze, D. Doenecke, N. Happel, G1 phase-dependent nucleolar accumulation of human histone H1x, Biol. Cell./under the auspices of the European Cell Biology Organization 99 (2007) 541–552.
[98] H. Takata, S. Matsunaga, A. Morimoto, R. Ono-Maniwa, S. Uchiyama, K. Fukui, H1.X with different properties from other linker histones is required for mitotic progression, FEBS Lett. 581 (2007) 3783–3788.
[99] A.F. Trollope, N. Sapojnikova, A.W. Thorne, C. Crane-Robinson, F.A. Myers, Linker histone subtypes are not generalized gene repressors, Biochim. Biophys. Acta 1799 (2010) 642–652.
[100] B. Wen, H. Wu, Y. Shinkai, R.A. Irizarry, A.P. Feinberg, Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells, Nat. Genet. 41 (2009) 246–250.
[101] K.L. Reddy, A.P. Feinberg, Higher order chromatin organization in cancer, Semin. Cancer Biol. 23 (2013) 109–115.
[102] K.D. Hansen, W. Timp, H.C. Bravo, S. Sabunciyan, B. Langmead, O.G. McDonald, B. Wen, H. Wu, Y. Liu, D. Diep, E. Briem, K. Zhang, R.A. Irizarry, A.P. Feinberg, Increased methylation variation in epigenetic domains across cancer types, Nat. Genet. 43 (2011) 768–775.
[103] O.G. McDonald, H. Wu, W. Timp, A. Doi, A.P. Feinberg, Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition, Nat. Struct. Mol. Biol. 18 (2011) 867–874.
[104] W. Timp, A.P. Feinberg, Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host, Nat. Rev. Cancer 13 (2013) 497–510.
[105] W. Albig, T. Meergans, D. Doenecke, Characterization of the H1.5 gene completes the set of human H1 subtype genes, Gene 184 (1997) 141–148.
[106] S.M. Seyedin, W.S. Kistler, H1 histone subfractions of mammalian testes. 1. Organ specificity in the rat, Biochemistry 18 (1979) 1371–1375.
[107] K. Hellauer, E. Sirard, B. Turcotte, Decreased expression of specific genes in yeast cells lacking histone H1, J. Biol. Chem. 276 (2001) 13587–13592.
[108] J.A. Downs, E. Kosmidou, A. Morgan, S.P. Jackson, Suppression of homologous recombination by the Saccharomyces cerevisiae linker histone, Mol. Cell 11 (2003) 1685–1692.
[109] J.M. Bryant, J. Govin, L. Zhang, G. Donahue, B.F. Pugh, S.L. Berger, The linker histone plays a dual role during gametogenesis in Saccharomyces cerevisiae, Mol. Cell. Biol. 32 (2012) 2771–2783.
[110] X. Shen, M.A. Gorovsky, Linker histone H1 regulates specific gene expression but not global transcription in vivo, Cell 86 (1996) 475–483.
[111] J.L. Barra, L. Rhounim, J.L. Rossignol, G. Faugeron, Histone H1 is dispensable for methylation-associated gene silencing in *Ascobolus immersus* and essential for long life span, Mol. Cell. Biol. 20 (2000) 61–69.
[112] M.A. Jedrusik, E. Schulze, A single histone H1 isoform (H1.1) is essential for chromatin silencing and germline development in *Caenorhabditis elegans*, Development 128 (2001) 1069–1080.
[113] M.A. Jedrusik, E. Schulze, Linker histone HIS-24 (H1.1) cytoplasmic retention promotes germ line development and influences histone H3 methylation in *Caenorhabditis elegans*, Mol. Cell. Biol. 27 (2007) 2229–2239.

Please cite this article as: L. Millán-Ariño, et al., Specificities and genomic distribution of somatic mammalian histone H1 subtypes, Biochim. Biophys. Acta (2015), http://dx.doi.org/10.1016/j.bbagrm.2015.10.013

[114] M. Studencka, A. Konzer, G. Moneron, D. Wenzel, L. Opitz, G. Salinas-Riester, C. Bedet, M. Kruger, S.W. Hell, J.R. Wisniewski, H. Schmidt, F. Palladino, E. Schulze, M. Jedrusik-Bode, Novel roles of *Caenorhabditis elegans* heterochromatin protein HP1 and linker histone in the regulation of innate immune gene expression, Mol. Cell. Biol. 32 (2012) 251–265.
[115] M. Studencka, R. Wesolowski, L. Opitz, G. Salinas-Riester, J.R. Wisniewski, M. Jedrusik-Bode, Transcriptional repression of Hox genes by *C. elegans* HP1/HPL and H1/HIS-24, PLoS Genet. 8 (2012), e1002940.
[116] M.A. Jedrusik, S. Vogt, P. Claus, E. Schulze, A novel linker histone-like protein is associated with cytoplasmic filaments in *Caenorhabditis elegans*, J. Cell Sci. 115 (2002) 2881–2891.
[117] M. Prymakowska-Bosak, M.R. Przewloka, J. Slusarczyk, M. Kuras, J. Lichota, B. Kilianczyk, A. Jerzmanowski, Linker histones play a role in male meiosis and the development of pollen grains in tobacco, Plant Cell 11 (1999) 2317–2329.
[118] A.T. Wierzbicki, A. Jerzmanowski, Suppression of histone H1 genes in arabidopsis results in heritable developmental defects and stochastic changes in DNA methylation, Genetics 169 (2005) 997–1008.
[119] S. Perez-Montero, A. Carbonell, T. Moran, A. Vaquero, F. Azorin, The embryonic linker histone H1 variant of drosophila, dBigH1, regulates zygotic genome activation, Dev. Cell 26 (2013) 578–590.
[120] S. Jayaramaiah Raja, R. Renkawitz-Pohl, Replacement by *Drosophila melanogaster* protamines and Mst77F of histones during chromatin condensation in late spermatids and role of sesame in the removal of these proteins from the male pronucleus, Mol. Cell. Biol. 25 (2005) 6165–6177.
[121] P. Bouvet, S. Dimitrov, A.P. Wolffe, Specific regulation of xenopus chromosomal 5S rRNA gene transcription in vivo by histone H1, Genes Dev. 8 (1994) 1147–1159.
[122] H. Kandolf, The H1A histone variant is an in vivo repressor of oocyte-type 5S gene transcription in *Xenopus laevis* embryos, Proc. Natl. Acad. Sci. U. S. A. 91 (1994) 7257–7261.
[123] O.C. Steinbach, A.P. Wolffe, R.A. Rupp, Somatic linker histones cause loss of mesodermal competence in xenopus, Nature 389 (1997) 395–399.
[124] T.J. Maresca, B.S. Freedman, R. Heald, Histone H1 is essential for mitotic chromosome architecture and segregation in *Xenopus laevis* egg extracts, J. Cell Biol. 169 (2005) 859–869.
[125] D.T. Brown, B.T. Alexander, D.B. Sittman, Differential effect of H1 variant overexpression on cell cycle progression and gene expression, Nucleic Acids Res. 24 (1996) 486–493.
[126] A. Gunjan, B.T. Alexander, D.B. Sittman, D.T. Brown, Effects of H1 histone variant overexpression on chromatin structure, J. Biol. Chem. 274 (1999) 37950–37956.
[127] R. Alami, Y. Fan, S. Pack, T.M. Sonbuchner, A. Besse, Q. Lin, J.M. Greally, A.I. Skoultchi, E.E. Bouhassira, Mammalian linker-histone subtypes differentially affect gene expression in vivo, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 5920–5925.
[128] D.I. Gabrilovich, P. Cheng, Y. Fan, B. Yu, E. Nikitina, A. Sirotkin, M. Shurin, T. Oyama, Y. Adachi, S. Nadaf, D.P. Carbone, A.I. Skoultchi, H1(0) histone and differentiation of dendritic cells. A molecular target for tumor-derived factors, J. Leukoc. Biol. 72 (2002) 285–296.
[129] A.M. Sirotkin, W. Edelmann, G. Cheng, A. Klein-Szanto, R. Kucherlapati, A.I. Skoultchi, Mice develop normally without the H1(0) linker histone, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 6434–6438.
[130] I. Martianov, S. Brancorsini, R. Catena, A. Gansmuller, N. Kotaja, M. Parvinen, P. Sassone-Corsi, I. Davidson, Polar nuclear localization of H1T2, a histone H1 variant, required for spermatid elongation and DNA condensation during spermiogenesis, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 2808–2813.
[131] H. Tanaka, N. Iguchi, A. Isotani, K. Kitamura, Y. Toyama, Y. Matsuoka, M. Onishi, K. Masai, M. Maekawa, K. Toshimori, M. Okabe, Y. Nishimune, HANP1/H1T2, a novel histone H1-like protein involved in nuclear formation and sperm fertility, Mol. Cell. Biol. 25 (2005) 7107–7119.
[132] M. Murga, I. Jaco, Y. Fan, R. Soria, B. Martinez-Pastor, M. Cuadrado, S.M. Yang, M.A. Blasco, A.I. Skoultchi, O. Fernandez-Capetillo, Global chromatin compaction limits the strength of the DNA damage response, J. Cell Biol. 178 (2007) 1101–1108.
[133] G.D. Nguyen, S. Gokhan, A.E. Molero, S.M. Yang, B.J. Kim, A.I. Skoultchi, M.F. Mehler, The role of H1 linker histone subtypes in preserving the fidelity of elaboration of mesendodermal and neuroectodermal lineages during embryonic development, PLoS One 9 (2014), e96858.
[134] E.Y. Popova, S.A. Grigoryev, Y. Fan, A.I. Skoultchi, S.S. Zhang, C.J. Barnstable, Developmentally regulated linker histone H1c promotes heterochromatin condensation and mediates structural integrity of rod photoreceptors in mouse retina, J. Biol. Chem. 288 (2013) 17895–17907.

Please cite this article as: L. Millán-Ariño, et al., Specificities and genomic distribution of somatic mammalian histone H1 subtypes, Biochim. Biophys. Acta (2015), http://dx.doi.org/10.1016/j.bbagrm.2015.10.013

APPENDIX III

APPENDIX III – Bioinformatics in chromatin research

By definition, bioinformatics is an interdisciplinary field that develops and combines computational methods to extract information from biological/medical data. Statistics, computer science, mathematics and a deep knowledge of the biological question are needed for the analysis and interpretation of complex biological processes. Bioinformatics is mainly used in molecular biology for analysing genomes, identifying proteomes, modelling of three-dimensional biomolecules and in biological systems.

Genome-wide high throughput experiments use classical cell biology techniques such as chromatin immunoprecipitation, chromatin digestion with enzymes (micrococcal nuclease, hyperactive Tn5 transposase…), RT-qPCR… coupled to massive parallel DNA sequencing. This combination produces enormous quantities of data, which allows answering biological questions that are otherwise unattainable using conventional methods.

The main methods used in chromatin research are those analysing chromatin-binding proteins (ChIP-Seq), DNA methylation (Bisulphite sequencing, methyl-Seq and methylCpG arrays), chromatin accessibility (DNase-Seq, FAIRE-Seq and ATAC-Seq) and nucleosome positioning (MNase-Seq and ATAC-Seq), chromosome conformation captures (ChIA-PET, Hi-C, 5C and 4C) and gene regulation through chromatin partners (RNA-Seq, CLIP-Seq, GRO-Seq and RIP-Seq) (Table 1). This section is only focused in the bioinformatics methods used in both publications (*Chapter I* and *II*), bolded in Table 1.

Table 1: High throughput experiments used in chromatin research.

Chromatin feature	High throughput experiment
Binding locations	**Chromatin immunoprecipitation (ChIP-Seq)**
DNA methylation	Whole genome bisulphite sequencing (WGBS) Reduced representation bisulphite sequencing (RRBS) methylCpG array and methyl-Seq
Chromatin accessibility	DNase I digestion coupled to sequencing (DNase-Seq) Formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-Seq) **Assay for transposase accessible chromatin sequencing (ATAC-Seq)**
Nucleosome positioning	Micrococcal nuclease digestion coupled to sequencing (MNase-Seq) **Assay for transposase accessible chromatin sequencing (ATAC-Seq)**
Chromosome conformation capture	Chromatin interaction analysis by paired-end tag (ChIA-PET) Hi-C (all vs all) 5C (many vs many) 4C (one vs all)
Transcriptional changes	**RNA sequencing (RNA-Seq)** Cross-linking immunoprecipitation sequencing (CLIP-Seq) RNA immunoprecipitation sequencing (RIP-Seq) Global run-on sequencing (GRO-Seq)

1. Sequencing strategies

With a next generation sequencer billions of short sequences, called reads, are obtained. Two strategies of sequencing a DNA fragment (200bp-800bp) are the most commonly used: single end where only one end of the fragment is sequenced and paired-end were both ends are sequenced (Figure 1). The main advantage of paired-end sequencing is that the fragment size can be inferred once the reads are mapped in the genome. Paired-end sequencing can detect easier genomic rearrangements and repetitive sequences, as well as isoforms, gene fusions and novel transcripts. Specific experiments such as ATAC-Seq can only be performed with paired-end sequencing.

Furthermore, another strategy exists, allowing to use longer DNA fragments (2kb- 5kb): mate pair sequencing. With mate pair strategy, fragment ends are biotinylated, circularized and purified. Those pseudo-fragments, containing both ends of long fragments, are next sequenced (Figure 1).

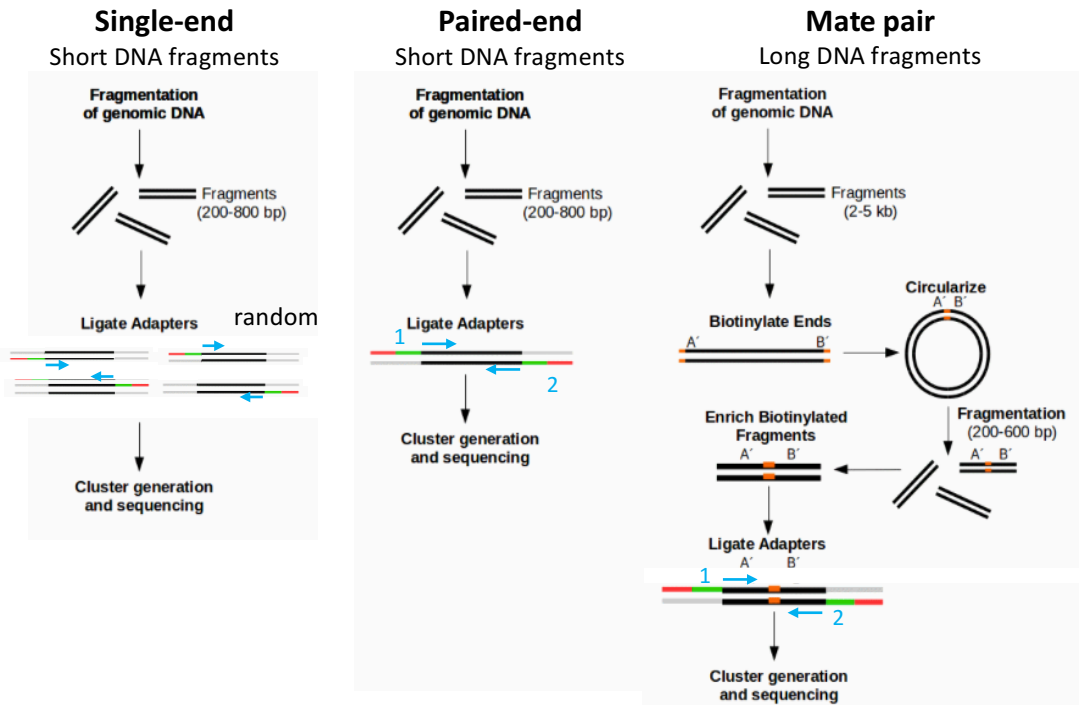


Figure 1: Sequencing strategies with short DNA fragments (single-end and paired-end sequencing) and long DNA fragments (mate pair sequencing).

2. Read quality control

Before any downstream analysis, the quality of the identified reads needs to be explored. Quality controls allow removing, trimming of poor quality reads to obtain a good quality of the final raw data. This is done by means of several softwares such as FastQC [1].

Some basic steps are as follows:

1. Exogenous sequences

Check if adapters, primers or other sequence contaminants are still remaining in reads.

2. Per base quality

Readout files from sequencing platforms, mainly in fastQ format, include a quality measure of the identification of each nucleotide.

3. Per sequence quality

The quality score distribution over all reads allows the identification of possible subsets of sequences with low quality measure. They should represent a small portion of the overall reads.

4. Per base sequence content

Allows exploring the relative number of bases at each read position. Their content should reflect the reference genome ratio or acceptable little imbalance.

5. GC content per base and per sequence

6. Per base N content

Inspection of Ns, which represent nucleotides that were not identified. It is usual to find a small portion of Ns at the end of the read.

7. Over-represented sequences

Although high coverage at target sequence is expected, the number of reads having the same sequence that will down-stream map at exactly the same position should be small. High number of duplications might reflect PCR over-amplification biases. However, care should be taken if analysing repetitive sequences were over-represented sequences should be expected.

3. ChIP-Seq – Chromatin immunoprecipitation sequencing

Chromatin immunoprecipitation with massively parallel DNA sequencing (ChIP-Seq) allows the analysis of genome-wide distribution of chromatin-binding proteins and histone PTMs in any organism with a sequenced genome.

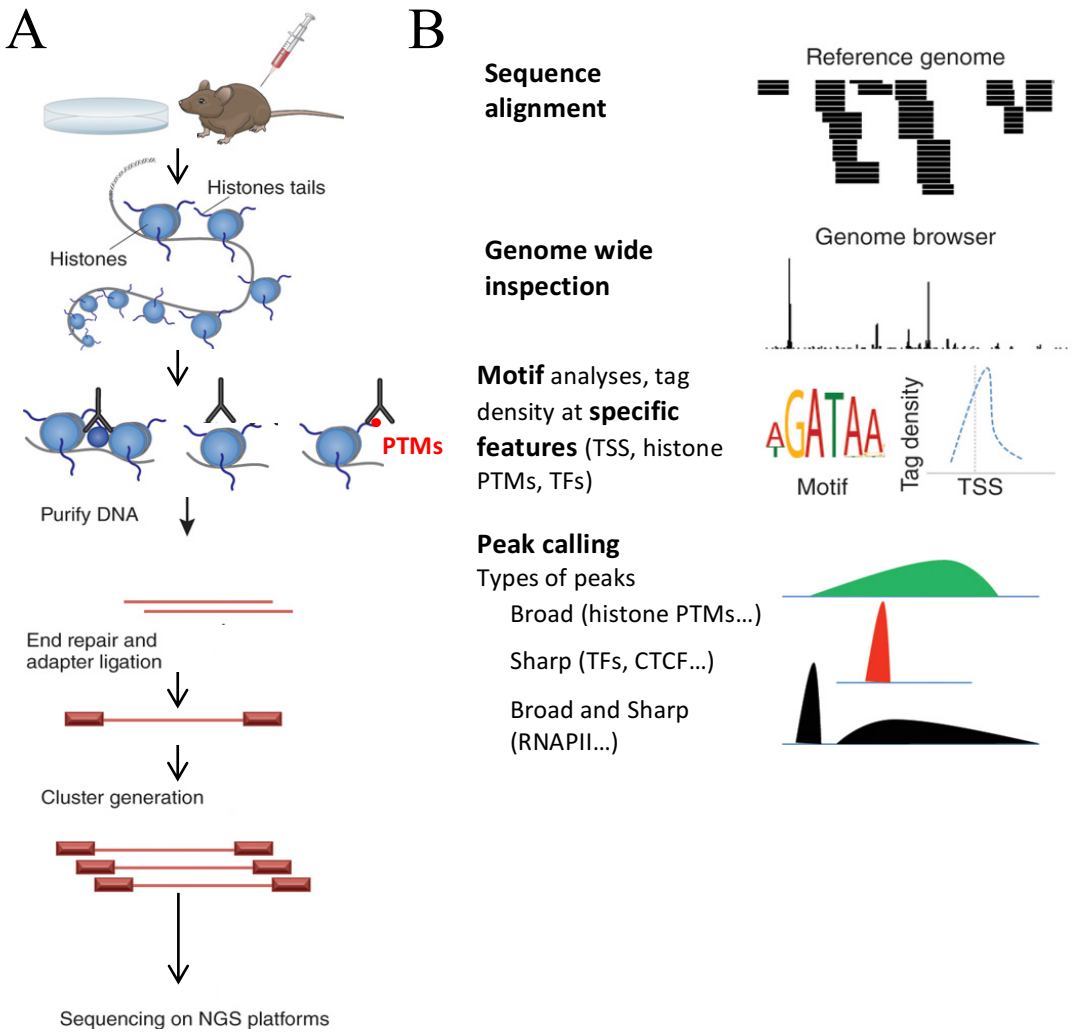


Figure 2: Chromatin immunoprecipitation coupled to DNA sequencing (ChIP-Seq). (**A**) Experimental design. (**B**) Computational analysis. PTMs: post-translational modifications; TFs: transcription factors; TSS: transcription start site; RNAPII: RNA polymerase II. Figure adapted from [2].

Shortly, chromatin is formaldehyde cross-linked, fragmented with sonication and immunoprecipitated with antibodies against a target protein or post-translational modification. Next, DNA is purified, amplified and then hybridized to adapter sequences. Sequencing and bioinformatic analysis is finally performed (Figure 2). A key point in the experimental design is fragmentation of chromatin to obtain DNA of about 150-300bp,

approximately a mono- and a dinucleosome. This produces high resolution of binding sites and a manageable size for next generation sequencing platforms. Another key point is the control sample, which allows to get rid, in downstream analyses, of artefacts done in the experimental procedure. Three control samples exist: input DNA, mock IP DNA (DNA obtained without antibody) and DNA from non-specific IP (using IgG antibodies for example). It is worth saying that the input DNA is the most commonly used.

The typical workflow of a bioinformatic analysis of ChIP-Seq data is composed of two main steps: align reads to a reference genome (mapping) and identification of binding or enriched sites (peak calling) (Figure 2). Downstream analyses are specifically performed depending on the biological question.

Mapping to the reference genome is performed with those reads that passed the overall quality control. Trimming a subset of reads might be sufficient to achieve a good quality data.

3.1. Read mapping

Several challenges need to be addressed when mapping reads to the reference genome. Firstly, reads have to be mapped accurately and quickly while consuming less memory. Secondly, read sequence can slightly vary compared to reference genome. And thirdly, reads can map to several positions (multi-reads) due to repetitive sequences. There are several ways to deal with multi-reads. Usually when not studying repetitive elements, multi-reads are discarded or randomly allocated.

There are several algorithms available specifically to align short reads in public repositories such as Bowtie, BWA or MAQ aligners [3-5]. Almost all short-read aligners rely on the same principle of a first bypass “heuristic” match, which rapidly finds a short list of possible positions. Then, a complex “local alignment” algorithm is performed at those candidate locations. To address the first step quickly, aligners perform a computational strategy called “indexing”. An index of a large DNA sequence (reference genome) is used to find shorter sequences within it in small times. Indeed, softwares such as Bowtie and MAQ use, as an input of the reference genome, only an index.

The choice of the alignment parameters is crucial and has profound influence on the number of mapped reads and thus, the coverage of the reference genome (number of positions with at least one read). The easy option to begin with is to consider parameters used in large consortiums such as ENCODE for similar proteins. However, assess the outcome with different parameters is always the best option.

Normally, in aligner parameters a unique position is allowed for every read and those with multiple positions are discarded. One of the aligner reports shows the

proportion of unique and multiple mapped reads together with unaligned reads, which need to be carefully inspected. When dealing with proteins unknown or proteins that may bind to repetitive elements, aligner parameters need to be certainly taken in to account (See *3.5. Enrichment analysis in repetitive sequences*).

3.2. Peak calling

After successful mapping, the next step is "peak calling", which reports the regions that are enriched in a ChIP sample relative to the control (input, normally) with statistical significance.

Before identifying peaks, pre-process of ChIP and input mapping files is needed if not done by the aligner or the peak calling software (Figure 3A). If not done in the quality control by identic sequence, reads mapping exactly at the same genomic location need to be discarded as they may reflect PCR over-amplification artifacts.

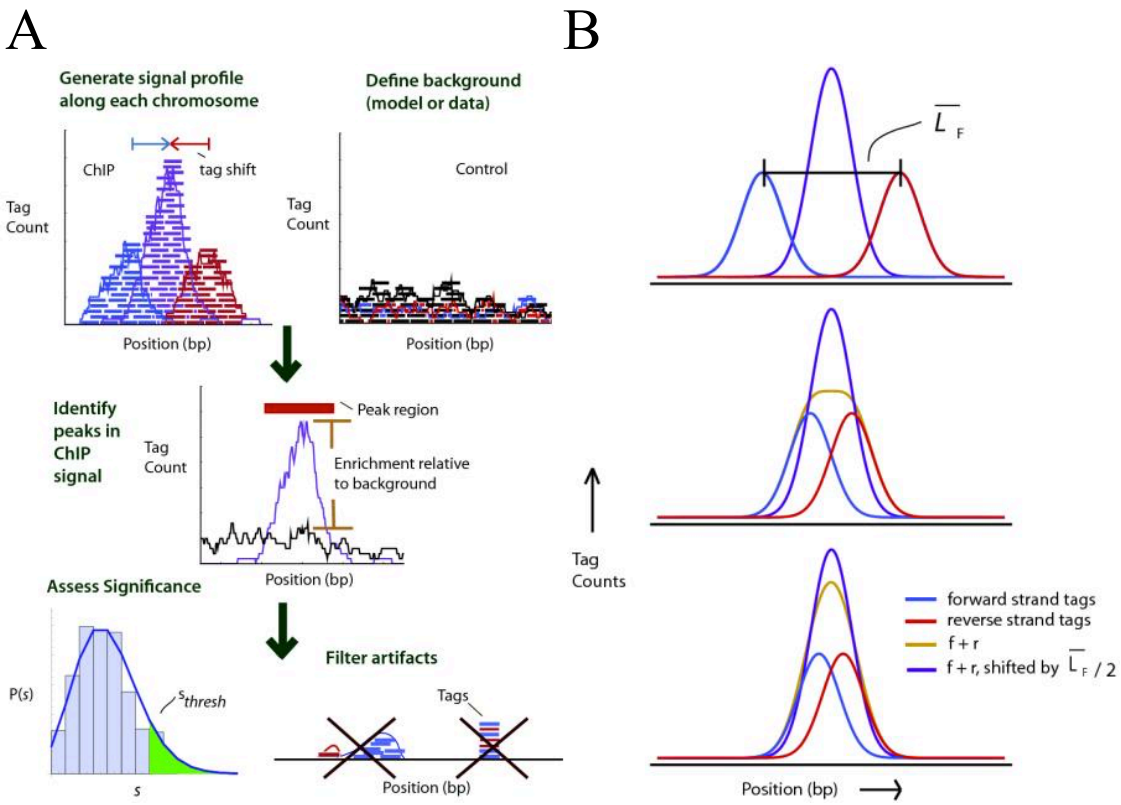


Figure 3: (**A**) ChIP-Seq peak calling steps. (**B**) The influence of fragment size in ChIP-Seq peak calling. L_F: fragment length. Figure adapted from [6].

Most of the ChIP-Seq experiments are done using single-end reads that is, sequenced from one of the two strands in the 5' to 3' direction. This feature is reflected in the mapping output. Especially for transcription factors, a bimodal shape distribution

is produced. A shift procedure is needed in the ChIP sample to precisely identify binding sites (Figure 3B). Fragment size obtained after chromatin sonication is a great factor at this point. If sonication was not performed correctly, large fragments are obtained, which are later sequenced from each end. The expected bimodal distribution is lost and two peaks (representing a same fragment) might be reported. An unbalanced bimodal distribution may also reflect PCR over-amplification artifacts and need to be also inspected (Figure 3A).

However, with proteins such as CTCF a bimodal shape is expected and shifting should not be performed. In addition, the explained bimodal distribution is only expected for transcription factors and specific proteins, whose binding region is precisely located in a genomic region, showing sharp peaks (Figure 2). Motif analysis in this kind of proteins are interesting as they normally bind to specific DNA sequences.

Other proteins, especially chromatin-related such as histones or post-translational modifications show broader regions of enrichment (broad peaks) and shifting should not be done. Some peak calling softwares such as MACS2 can identify both narrow (sharp) and broad peaks [7]. However, SICER was specifically designed to assess broad regions of enrichment and it is especially suitable for histones and histone PTMs ChIP-Seq data [8].

Regarding input sample, some algorithms define a background model but some studies use directly the mapped data to avoid over-manipulation and indirect effect in down-stream analyses.

When assessing broad and narrow peaks, the number of reads mapped at a particular genomic region is compared between ChIP and input samples. The number of sequenced reads might not be the same and a normalization step is needed when assessing significance. Finally, aligners report a list of enriched regions (narrow or broad peaks) with p-value, FDR, q-value, fold changes… measures. A filtering step to report significantly enriched regions will lead to the final binding locations, used in down-stream analysis.

3.3. Depth of sequencing

Sequencing depth is an important parameter in ChIP-Seq experiments as actual binding site might not be captured due to a limited number of sequenced reads. The number of reads sufficient to achieve high resolution data greatly depends on the genome size, the protein and the biological question. When the addition of more reads does not change the number of identified reads a saturation point is reached and that is the desired coverage (Figure 4) [9].

ChIP-Seq data producing narrow peaks (transcription factors) need lower sequencing depths than those producing broad peaks (histones and PTMs, mainly).

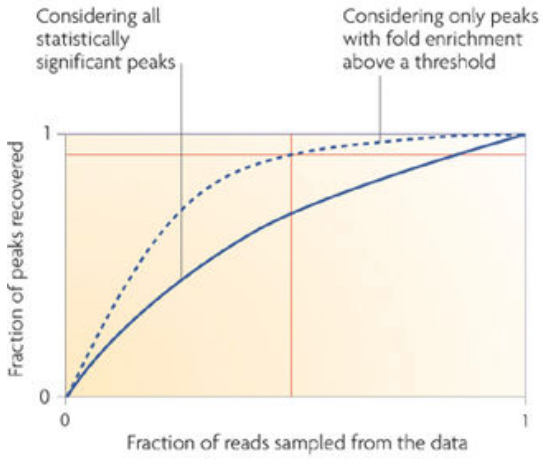


Figure 4: Determination of depth sequencing. Figure adapted from [9].

3.4. Down-stream analyses

Further analyses are directly influenced by the biological question. All of them will use as inputs: ChIP-Seq peaks and tag density along whole genome. Tag density file is the input subtracted ChIP-Seq data normalized by depth sequencing. Most common analyses are shortly explained, as follows:

Annotation to genomic features

One basic analysis is to locate enriched regions to known genomic features such as TSS, UTRs, gene-bodies, promoters, exons, introns, intergenic regions, bidirectional promoters, 3' ends of genes among many others. Specific softwares such as Bioconductor package ChIPpeakAnno, CEAS and BEDTools among others are useful to perform those analyses [10-12].

Motif analysis

For proteins that directly bind to DNA, motif analyses are particularly suitable. Those proteins, mainly transcription factors (TFs) will produce narrow peaks in specific DNA sequences (named motifs). Databases of motifs are used to assess enrichment of known-motifs in peaks. New regulatory pathways can be described for TFs by this method. In addition, several algorithms perform new motif discovery and co-occurrence of various motifs and TFs binding sites can be also analysed.

Sequence conservation

Sequence underlying a ChIP-Seq peak that is highly conserved may suggest a functional role. Peaks can be ranked by this parameter to investigate possible conserved functions. Nucleotide level conservation can be obtained from PhastCons or PhyloP [13].

Correlation to expression

ChIP-Seq peaks can be also classified by expression data. This approach is helpful to investigate the role of transcription factors in different tissues for example. Histone and PTMs are also normally investigated correlating them with expression data. For instance, H3K4me3 peaks are more frequently found around TSS of active genes in contrast to H3K27me3 peaks, which are found at TSS of developmentally repressed genes.

To perform ChIP-Seq correlation to expression data two approaches are normally implemented (Figure 5). First, the overlap of ChIP-Seq peaks with genic regions, regarding their expression. Second, using the tag density (Input-subtracted ChIP-Seq signal) along a specific genic feature, also regarding its expression. For example, assess abundance of H3K4me3 around TSS of all genes divided by expression ranges. With the tag density approach, it can be inspected the distribution of our protein around specific features while with the peak approach only the number of overlapping peaks can be inspected.

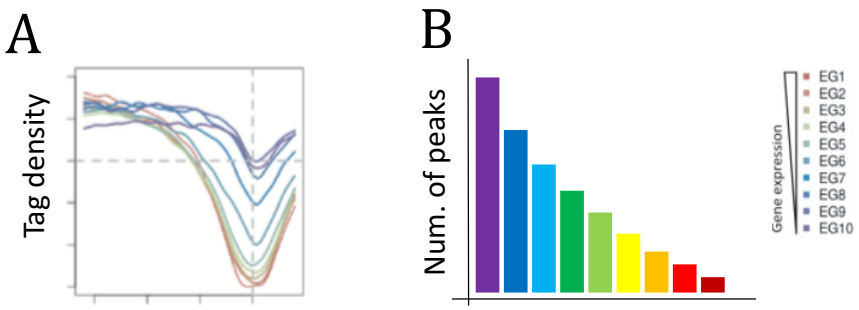


Figure 5: ChIP-Seq data correlation to expression. Genes are divided in 10 groups according to their expression (**A**) using tag density approach (**B**) using number of overlapping peaks.

Functional and pathway analyses

Genes overlapping ChIP-Seq peaks can be investigated to discover possible functional roles. Overrepresentation of genes belonging to a same functional pathway is further investigated to describe new regulatory roles (See *4.5. Functional and pathway enrichment analysis*).

Overlap analysis

Binding locations (peaks) of transcription factors or histones can be compared to those of other features. Enrichment peak location in non-genic regions such as enhancers, promoters, centromeric and telomeric regions, repetitive elements, binding site of other proteins (ChIP-Seq peaks), CpG islands… Transcription factors can be correlated with combinations of core histone PTMs that define chromatin regions such as poised enhancers or bivalent domains found at pluripotent ESCs, among others.

Two types of questions can be answered using peaks or tag density, respectively. Firstly, are target genomic regions (ChIP-Seq peaks) overlapping more than expected a specific genomic region (promoters, CpGs, repetitive elements…)? (Figure 6A). To assess if the overlap between two genomic regions is significant, a permutation test is the best option, using softwares such as regioneR [14]. Basically, a user-defined number of random samples (permutations) containing the same number of peaks and of the same length as the ChIP-Seq peaks of the target protein is computed. The number of overlaps of those permutations with the specific genomic region (promoters, CpGs, repetitive elements…) is measured and their distribution plotted. The permutation distribution (expected) is then used to compute the p-value of the overlap of the target protein binding regions (observed) (Figure 6B).

As said, a second question can be asked: are the levels of my target protein (tag density: input-subtracted ChIP-Seq signal, DNA methylation…) significantly different in these specific regions? (Figure 6C). Similar to the overlap analysis, a permutation test is performed but instead of counting the number of overlaps (evaluate function: *numOverlaps*), the mean of the tag density (evaluate function: *meanInRegions*) is compared between an expected distribution and the observed value.

Furthermore, permutation can be performed using as universe the reference genome (randomization: *randomizeRegions*) or a subset of specific regions (genes, promoter…). For example, when analysing DNA methylation in active and inactive genes, the permutations need to be done only taking gene regions and not intergenic. Thus, a resampling method is done using all genic regions (randomization: *resampleRegions*).

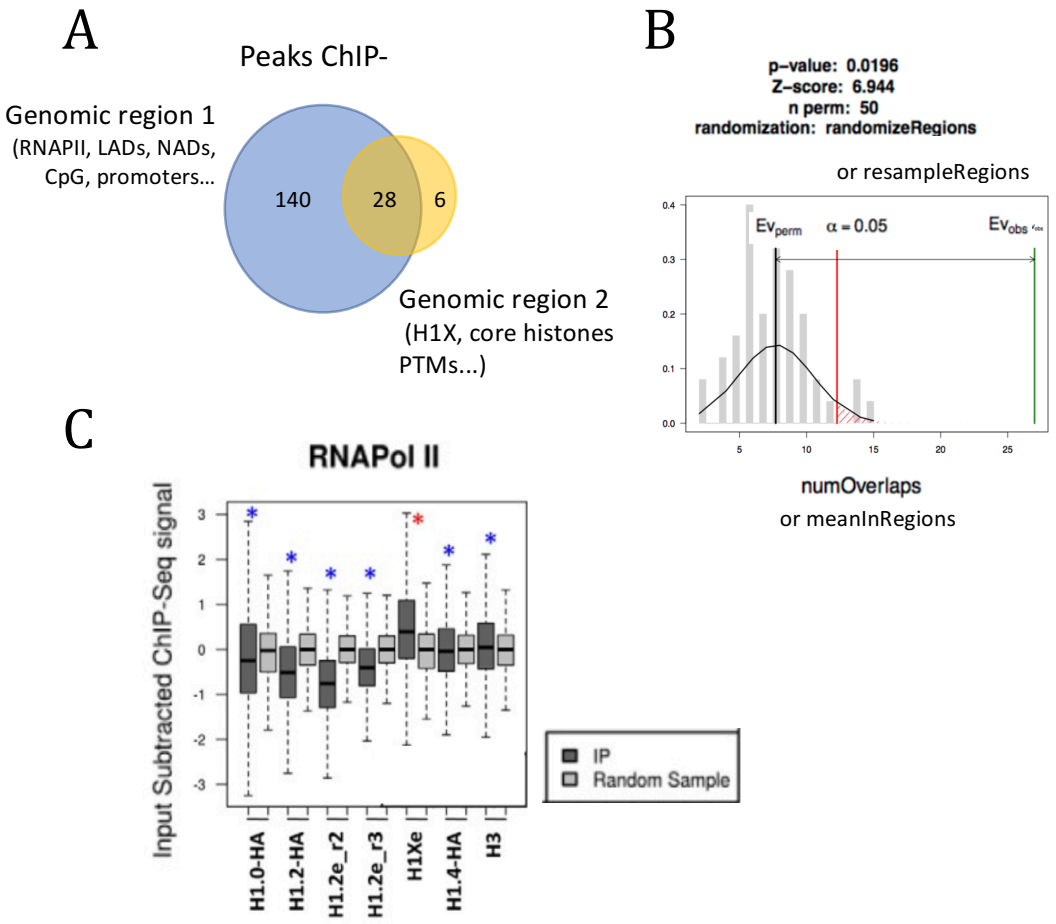


Figure 6: Overlap analysis. (**A**) Venn diagram showing the peak overlap of two genomic regions. (**B**) Permutation test (evaluate function: *numOverlaps*) performed using 50 random samples (*n perm: 50*), across the reference genome (randomization: *randomizeRegions*), regioneR software. (**C**) Tag density (Input subtracted ChIP-Seq signal) using RNA pol II peaks or a random sample (same number of RNA pol II and of the same length). Figure adapted from [14] and *Chapter I*.

3.5. Enrichment analysis in repetitive elements

Individual repeat sequences have accumulated specific mutations and are flanked by unique regions and annotated in the reference genome. Indeed, the percentage of short reads that map to unique locations on the human genome is typically reported to be 70-80%. By contrast, the repeat content in the genome is 50%. This discrepancy shows that repetitive elements mapped in the reference genome are or contain unique sequences. So, some of them can be analysed as other non-repetitive regions with unique read position mapping. However, this is a simple analysis as reads mapping to multiple positions (multi-reads) are discarded. Therefore, unique mapping analysis cannot fully explore repetitive elements as well as multi-gene families.

Multiple positioning mapping to the reference genome

Allowing reads to map in multiple positions is a used strategy for solving this issue and analyse repetitive sequences that are annotated in the reference genome. Two options are possible when performing multiple mapping: report the best-match or all matches (Figure 7A).

A better estimate of repetitive element coverage is obtained reporting the best match. However, allowing multi-reads to map to all possible positions avoids making decisions in the alignment parameters as the alignment score has lower confidence levels when sequences are more similar (Figure 7B) [15].

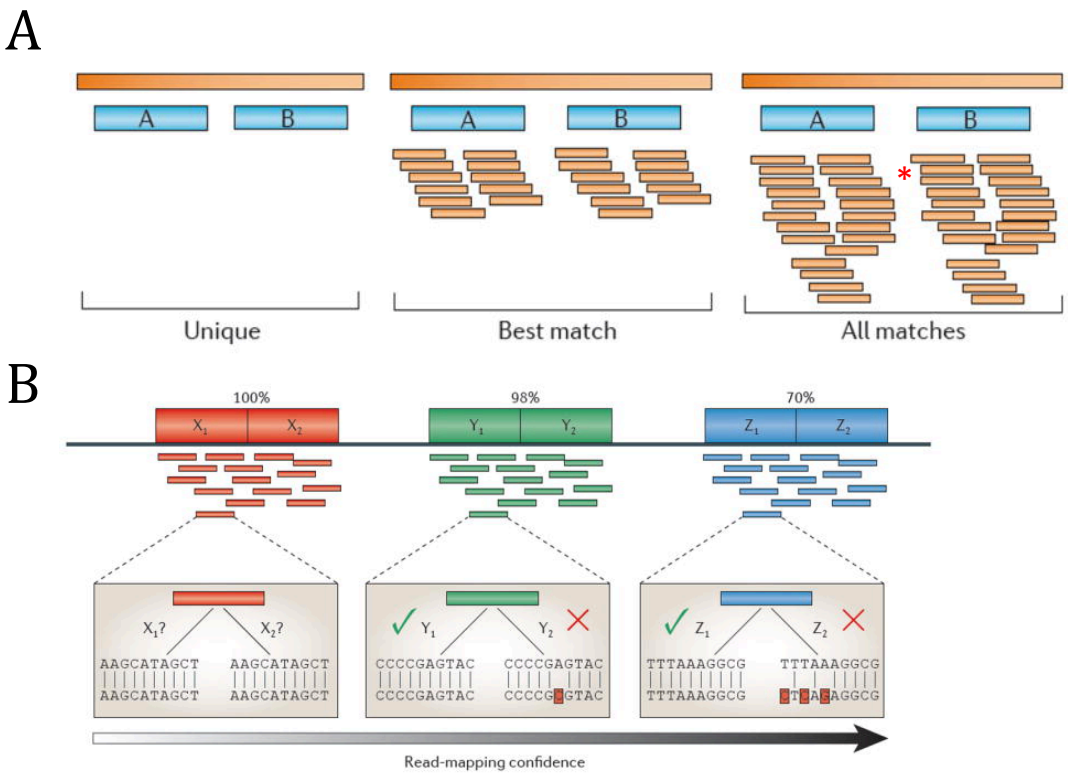


Figure 7: Multiple positioning mapping strategy. (**A**) Three strategies for mapping multi-reads in two identical regions (A and B). Unique positioning mapping; multi-reads are discarded. Multiple positioning mapping reporting: best match (high alignment score) or all matches. Red asterisk marks two reads mapping to the exact same location. (**B**) Ambiguities in read mapping. The more similar are two repetitive elements the lower is the read mapping confidence. Figure adapted from [15].

With multiple mapping strategy reporting all matches, a read can be taken into account if all of the regions to which it aligns are copies of the same repeat. Down-stream analysis will not indicate the exact genomic binding but for sure an enrichment can be reported for that repeat type. As the number of copies for a given repeat is

typically unknown and can vary between cells, the enrichment calculation will rely on input sequencing to normalize the read counts [16].

Moreover, when performing multiple positioning mapping, the probability that two reads have the same sequence and map to the exact same location increases (Figure 7A, red asterisk). A common standard quality control is to remove reads with exact sequence as they are considered PCR over-amplification biases (See *2. Read quality control*). Further, in the alignment, reads mapped to the same location are also commonly discarded (Figure 3A). These two common strategies in quality control and mapping are not recommended when analysing enrichment in repetitive elements.

Analysis of repetitive elements not annotated in the reference genome

However not all repetitive elements are included in the reference genome and a specific pipeline is needed. Repbase is a database of prototypic (consensus) sequences representing repetitive DNA from different eukaryotic species [17]. Repbase describes many families of repeats unreported anywhere else. By aligning reads directly to Repbase it can be estimated enrichments at particular repeat families or classes. However, several steps are needed before as reads mapping to unique positions in the reference genome have to be discarded.

RepeatMasker is a software that screens DNA sequences for interspersed repeats and low complexity DNA sequences [18]. The repeat sequence database used to identify repeats by RepeatMasker software is Repbase. The output of the program is a detailed annotation of the repeats (Figure 8) that are present in the query sequence as well as a modified version of the query sequence in which all the annotated repeats have been masked (replaced by Ns).

```
==================================================
file name: A-355G7.fasta
sequences:          1
total length:  139958 bp
GC level:       41.03 %
bases masked    91491 bp ( 65.37 %)
==================================================
                number of      length   percentage
                elements*    occupied  of sequence
--------------------------------------------------
SINEs:              46        12182 bp     8.70 %
      ALUs          41        11603 bp     8.29 %
      MIRs           5          579 bp     0.41 %

LINEs:              42        52641 bp    37.61 %
      LINE1         38        52296 bp    37.37 %
      LINE2          4          345 bp     0.25 %

LTR elements:       20        13441 bp     9.60 %
      MaLRs         10         5618 bp     4.01 %
      Retrov.        4         5131 bp     3.67 %
      MER4_group     3         1439 bp     1.03 %

DNA elements:        8         1741 bp     1.24 %
      MER1_type      7         1114 bp     0.80 %
      MER2_type      1          627 bp     0.45 %
      Mariners       0            0 bp     0.00 %

Unclassified:        5         9215 bp     6.58 %

Total interspersed repeats:  89220 bp    63.75 %


Small RNA:           0            0 bp     0.00 %

Satellites:          0            0 bp     0.00 %
Simple repeats:     20         1647 bp     1.18 %
Low complexity:      9          437 bp     0.31 %
==================================================

* most repeats fragmented by insertions or deletions
  have been counted as one element

The sequence(s) were assumed to be of primate origin.
RepeatMasker version  11/06/98               default
ProcessRepeats version  06/16/98
```

Figure 8: RepeatMasker report. Figure adapted from [18].

The four main classes (SINEs, LINEs, LTRs and DNA elements) are stated to be well defined and form a good basis for a summary or visual presentation of the repeats in a locus. Among subclasses, some uncertainty in the classification remains and not all the subclasses are listed and the total number of classes is often higher than the sum of subclasses. Consequently, RepeatMasker report cannot be directly used in down-stream enrichment analysis and an alignment to Repbase is needed.

So, RepeatMasker software is used to mask the reference genome and replaced by Ns all repetitive elements found in Repbase. Interestingly, currently over 56% of human genomic sequence is identified and masked by RepeatMasker software.

Once the genome is masked, the reads are aligned to Repbase database. As mentioned, raw data (reads) needs to be quality checked but removal of overrepresented sequences (putative PCR over-amplification biases) is not performed when analysing repetitive elements. In addition, the alignment is done permitting multiple positions and, as Repbase uses consensus sequences, more lax parameters dealing with mismatches are preferred. Obviously, input and ChIP samples are treated equally to finally obtain normalized read counts per repetitive sequence. Significant enrichment in each repetitive element is calculated using softwares dealing with RNA-Seq data (See *4.2. Normalization and estimation of transcript abundance*).

In summary, the pipeline to analyse enrichment at repetitive elements using Repbase database, including those not annotated in the reference genome is as follows:

1. Mask the reference genome to “cover” (with Ns) repetitive elements found in Repbase database with RepeatMasker software.

2. Perform standard read alignment with unique positioning and report reads unaligned and with multiple positioning.

3. Index Repbase database, necessary for aligners such as bowtie [3].

4. Align unaligned and with multiple positioning reads to Repbase database (allowing multiple positioning and more mismatches).

5. Report read counts per consensus repeat sequence.

6. Calculate significant enrichment using input and ChIP samples and normalized read counts, similar to an RNA-Seq experiment (See *4.2. Normalization and estimation of transcript abundance*).

See Table 2 for a summary of the three strategies described for enrichment analysis in repetitive elements.

Table 2: Summary of three strategies to analyse binding enrichment (ChIP-Seq) in repetitive elements.

	Mapping positioning	Masked genome	Paired-end reads	All repetitive elements	Known location	Used reads
Reference genome	Unique	-	+	-	+	unique reads
	Multiple	-	+	-	-	multi-reads
Repbase database	Multiple	+	-/+	+	-	unaligned and multi-reads

ribosomal DNA (rDNA) is a repetitive element not included in the reference genome and it might be of special interest for some biological questions. To perform enrichment analysis at rDNA, a costume-made reference genome is constructed. It is important to add the rDNA to a known reference genome to avoid false positives due to the small length of the rDNA sequence (roughly, 40 kb). Unique mapping can be used to investigate the 45S transcription unit. However, the non-transcribed spacer (NTS) contain a high number of repetitive elements and multiple mapping might be more suitable to assess distribution at NTSs. See in Figure 9 the tag density distribution at rDNA assessed with unique positioning mapping. A valley in the repeats region can be observed and care needs to be taken with this region in down-stream analysis.

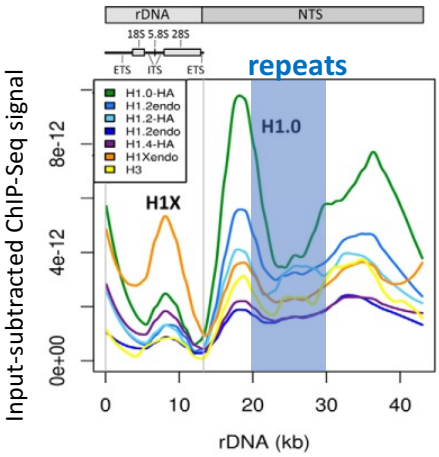


Figure 9: Linker histone H1 variants tag density distribution at ribosomal DNA (rDNA) and non-transcribed spacer (NTS) using unique positioning mapping. Figure from *Chapter I*.

4. RNA-Seq – RNA sequencing

RNA-Seq is another revolutionary high throughput experiment (Table 1) to analyse transcriptome, revealing the presence and quantity of RNA in a biological sample at a given moment time.

Briefly, RNAs are first converted into cDNA fragments through either RNA fragmentation or DNA fragmentation (Figure 10). Subsequently, adapter sequences are added to each cDNA fragments and reads are obtained using high throughput sequencing technology [19]. Reads are then mapped to the transcriptome or genome and the RNA abundance is reported.

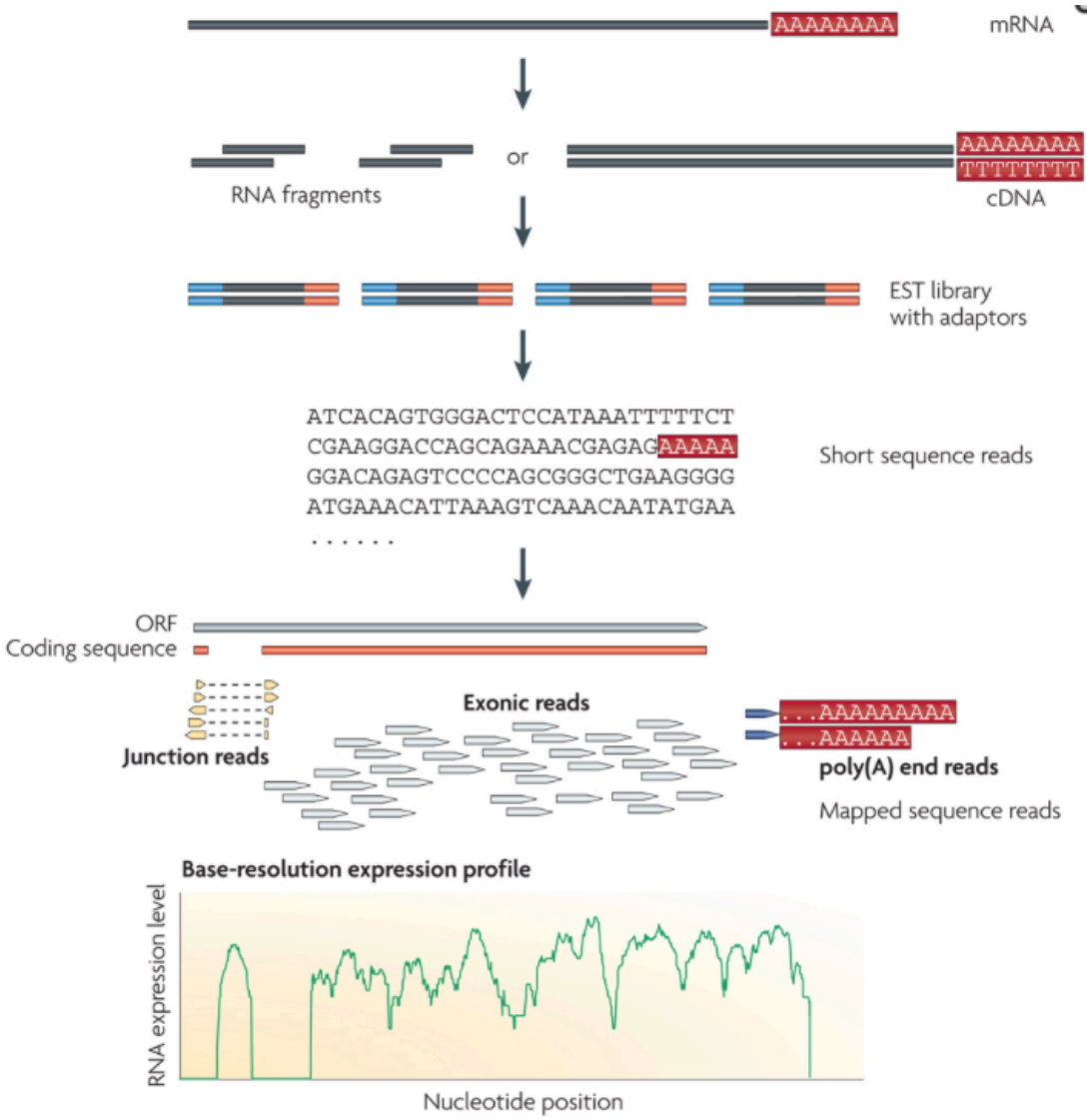


Figure 10: Schematic representation of a RNA-Seq experiment. Figure adapted from [19].

When designing an RNA-Seq experiment it is very important the biological question as several sequencing methods can be used (single or paired-end). Further, pre-processing of the total RNA preparation can increase or remove specific RNA species in the sample.

In a total RNA preparation, ribosomal RNA (rRNA) constitutes the majority (>98%). To avoid wasting sequencing reads, it is recommended to remove rRNA before preparing RNA libraries for deep sequencing. Specific kits like Ribo-Zero are normally used. In addition to rRNA, total RNA preparation contains mitochondrial ribosomal RNA (mtrRNA). The abundance of mtrRNA will greatly vary, depending on the expression of mitochondrial 12S and 16S rRNA genes as well as the number of mitochondria, which varies widely across cell types and differentiation. Specific analysis will need to remove mtrRNA, using the Ribo-Zero Gold kit that will remove mtrRNA in addition to rRNA [20] (Figure 11).

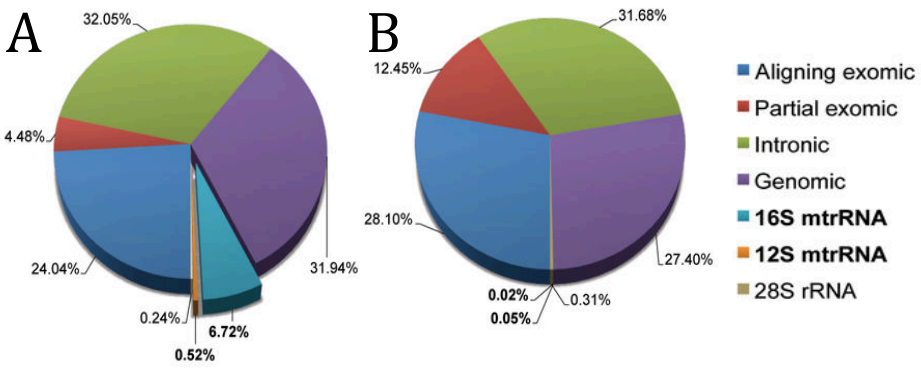


Figure 11: Profiles of RNA-Seq libraries prepared with (**A**) Ribo-Zero and (**B**) Ribo-Zero Gold kits. Figure adapted from [20].

In addition, as mRNAs contain a ploy(A) tail, a poly(T) adapter sequence coupled to magnetic beads is used to enrich the RNA fraction. Care needs to be taken when sequencing only poly(A) mRNA as other species of RNA like non-coding RNAs do not contain poly(A) tails. The technology will change from “Total RNA-Seq” to “mRNA-Seq”, the later using poly(A) RNAs extraction.

Further, specific transcriptome analyses need to know the direction of transcription and in which strand is occurring. Sequencing technologies allow the report for every read, its strand and direction.

See Table 3 for a summary of the transcriptome analysis and its corresponding sequencing technology and pre-processing of total RNA preparation.

Table 3: Transcriptome analysis and recommended sequencing technology and pre-process of total RNA preparation. Depth of sequencing when analysing transcription of repetitive elements will depend on the number of copies and its expression that will greatly vary across samples.

RNA-Seq Analysis	poly(A) mRNA extraction	rRNA extraction	Direction	Strand	Paired-end reads	Sequencing depth
poly(A) **mRNA-Seq** (alignment to transcriptome)	+	+				
Total RNA-Seq (alignment to genome and transcriptome)	-	+				
ribosomal RNA studies	-	-	+/-	+/-		Lower
Differentially expressed gene analysis	+	+	+/-	+/-	-	Lower
Splicing and isoform detection	+	+	+/-	+/-	+	Higher
Transcription of repetitive elements (alignment to Repbase)	-	+	-	-	+/-	?
Gene fusion detection, novel transcript detection…	-	+	+	+	+	Higher

4.1. Read mapping

Mapping of RNA to the genome is different from mapping DNA. Final RNA transcripts are normally spliced and only contain exons. Thus, RNA-Seq reads may contain parts of two exons that in the genome are separated by one intron (Figure 12).

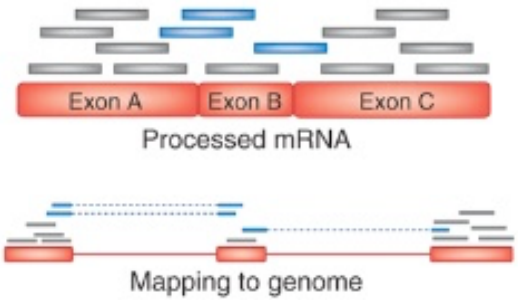


Figure 12: RNA-Seq reads represent processed mRNA, intronless. Mapping needs to take into account that introns are present in the reference genome. Figure adapted from [54].

Alignment to the genome of RNA-Seq reads is done with specific spliced-read mappers such as TopHat or STAR softwares [21, 22]. Furthermore, when mapping RNA-Seq reads to the reference genome, the corresponding annotation of transcripts is also used. Transcript annotation file will contain all the transcriptomic information such as gene, exons, introns… that will be used in down-stream analyses. Different annotations exist such as RefSeq or GENCODE, the annotations for the ENCODE project [23, 43].

It is worth mentioning that in poly(A) mRNA-Seq, the alignment can be performed to the transcriptome directly.

4.2. Normalization and estimation of transcript abundance

The transcript abundance could be estimated by simply counting the number of reads or fragments (paired-end reads). However, counts are biased due to the length of the mRNA as well as its expression. Therefore, normalization of counts is necessary and it is normally performed by softwares such as RSEM [26]. RSEM output consists of two files: one for isoform-level estimates (counts) and the other for gene-level estimates. Isoform-level estimates will be used by differential expression methods (See *4.3. Differentially expressed genes analysis*).

The gene-level estimates reported by RSEM are the estimated fractions of transcripts made up by a given isoform or gene. This measure can be used directly or can be multiplied by 10^6 to obtain transcripts per million (TPM). Other measures, apart from TPM, are normally used: reads per kilobase million (RPKM) for single-end reads and the equivalent for paired-end reads, fragments per kilobase million (FPKM). TPM and RPKM measure calculations are shown in Figure 13. Note that FPKM is calculated equally to RPKM but counting fragments instead of reads.

$$\boldsymbol{RPM}_{gene\ i}\ (reads\ per\ million) = \frac{read\ counts_{gene\ i}}{Total\ mapped\ reads} \cdot 10^6$$

$$\boldsymbol{RPKM}_{\boldsymbol{gene\ i}}\ (reads\ per\ kb\ million) = \frac{RPM_{gene\ i}}{length_{gene\ i}\ (kb)}$$

$$\boldsymbol{RPK}_{gene\ i}\ (reads\ per\ kb) = \frac{read\ counts_{gene\ i}}{length_{gene\ i}\ (kb)}$$

$$\boldsymbol{TPM}_{\boldsymbol{gene\ i}}\ (transcripts\ per\ million) = \frac{RPK_{gene\ i}}{Total\ number\ of\ RPK\ values} \cdot 10^6$$

Figure 13: RNA-Seq normalised measures of read counts for a given $gene_i$. Reads per kb million (**RPKM**) for single-end reads. Transcripts per million (**TPM**). For paired-end reads, fragments are counted instead of reads (**FPKM**).

As an example, you have sequenced one library of 5M reads. Among them, 4M mapped to the reference genome and 5000 reads matched to a given *gene* with a length of 2kb. There were 10K genes with mapped reads.

$$RPM_{gene} = \frac{5000}{4 \cdot 10^6} \cdot 10^6 = 1250 \qquad \boldsymbol{RPKM_{gene}} = \frac{1250}{2} = 625$$

$$RPK_{gene} = \frac{5000}{2} = 2500 \qquad \boldsymbol{TPM_{gene}} = \frac{2500}{10.000} \cdot 10^6 = 0{,}25 \cdot 10^6$$

The only difference when calculating TPM is that you normalise for gene length first and then normalise for sequencing depth second. This has profound effects on the finally measure as TPM represents the proportion of a given transcript in the total pool of transcripts (values range from 0 to 1, per million). Therefore, TPM values are more comparable across samples and species and are preferred over RPKM and FPKM measures that are dependent of the mean expressed transcript length [26].

4.3. Differentially expressed genes analysis

The aim of differentially expression analysis is to find genes that significantly changed their RNA abundance between two experimental conditions. As mentioned, normalized isoform counts are used as input and statistical testing is performed. Several softwares offer differential gene analysis but most of them are based in two statistical distributions: Poisson (DEGseq) and negative binomial (DEseq, edgeR) [27-29]. However, count data is discrete and skewed and softwares based in a negative binomial distribution such as DEseq are normally preferred.

Finally, a list significant differentially expressed (DE) genes between two conditions is reported together with normalized counts in the two conditions, p-values (multiple-testing corrected), fold-changes and/or fold discovery rate (FDR).

Once the significantly DE genes are reported, several down-stream analyses can be performed to retrieve biological information, as follows.

4.4. Discovery and search of regulatory motifs

The subset of DE genes may reflect regulatory functions of proteins that recognise specific DNA sequences (motifs) such as transcription factors or splicing machinery. So, motif enrichment in DE genes at specific regions such as promoters or splice junctions is commonly performed. Two types of motif analysis can be performed: scan of known motifs or *de novo* motif discovery. Several motif databases are publically

available and used by softwares like MEME or Homer to retrieve significantly enriched motif in DE genes [30, 31].

4.5 Functional and pathway enrichment analysis

Two main methods are normally used for functional enrichment analyses in RNA-Seq experiments. First one uses the subset of significantly expressed genes (Gene Ontology, KEGG pathways enrichment). The second method does not rely in the subset of significantly DE genes but on whole genic expression changes (gene set enrichment analysis, GSEA).

Briefly, **Gene Ontology** (GO) project has developed three structured ontologies that describe gene products in terms of their associated biological process, cellular component and molecular function in a species-independent manner [32]. For each gene, several GO terms are associated and publically available at GO term database. Other databases such as KEGG pathway contain information on molecular interactions, reaction and relation networks for metabolism, genetic and environmental information processing, cellular processes, human diseases, drug development… [33]. So, enrichment of functional pathways is normally performed using several databases such as GO terms or KEGG pathways, among others.

Taking significantly DE genes and their associated annotation, statistical tests can be performed to retrieve significantly enriched biological process, pathways, cellular components and molecular functions (Figure 14).

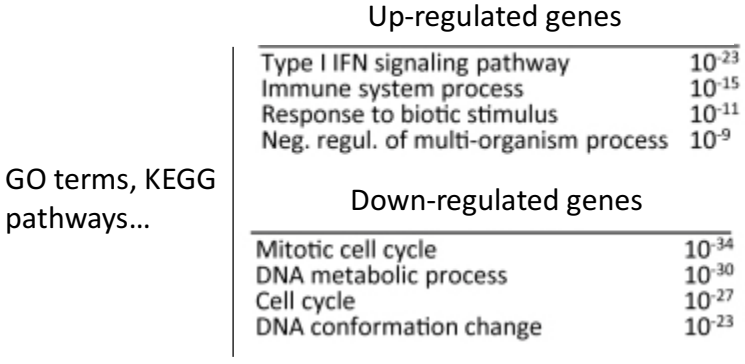


Figure 14: Typical report of a GO term enrichment analysis in significantly up- and down-regulated genes. The indicated enriched GO terms are reduced by semantic similarity using REVIGO software [34]. Figure adapted from *Chapter II*.

GO term database is highly redundant and grouping of GO terms is highly recommended. Several algorithms are implemented and publically available. As an

example, REVIGO software uses semantic similarity among GO terms to group them and user can define the stringency parameters (Figure 15) [33].

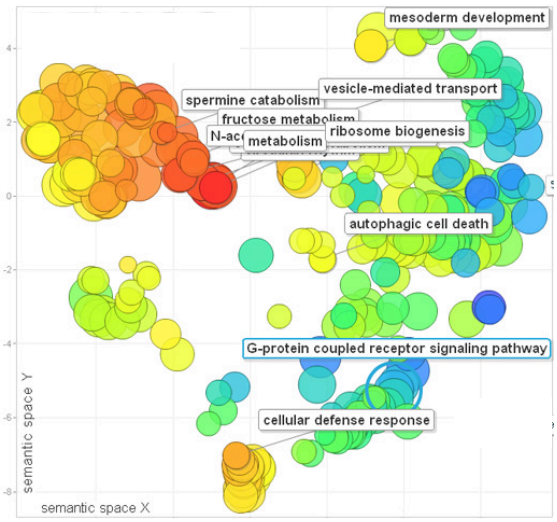


Figure 15: Output report of REVIGO software, clustering GO terms by semantic similarity. Figure adapted from [34].

Gene set enrichment analysis (GSEA) is a method that retrieves similar information although it does not rely on pre-selected DE genes [35]. Genes are ranked according to their fold-changes and compared to other transcriptome experiments (gene set of DE genes) (Figure 16). Pre-selection of genes is avoided and pathways as a whole can be significantly enriched without taking into account individual gene changes.

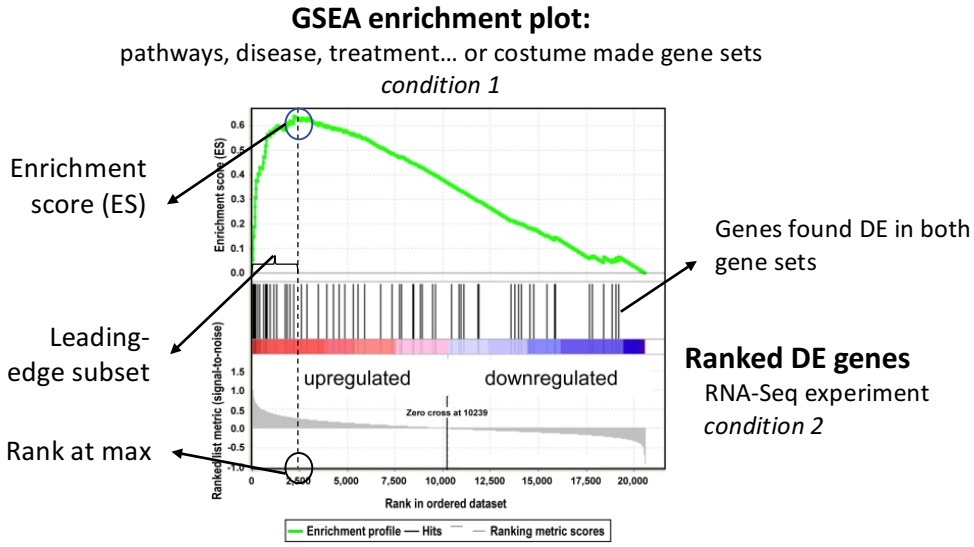


Figure 16: Enrichment plot from a gene set enrichment analysis (GSEA) comparing gene sets (DE genes) of a known experimental *condition 1* and an RNA-Seq experiment under *condition 2*. Figure adapted from [35].

The enrichment score (ES) is used to retrieve significantly enriched gene sets in an RNA-Seq experiments (Figure 16). Further, normalised ES (NES) is used for comparing several GSEA experiments.

The leading-edge subset of a gene set is the subset of genes that contribute most for the ES. For a positive ES, the leading-edge subset are those genes that appear in the ranked list prior to the peak score (Figure 16). It is hypothesized that this specific subset is the main responsible of the reported enrichment. Thus, GSEA allows to inspect the leading-edge subset specifically. The expression of leading-edge subset genes in other gene sets can be explored (Figure 17A). In addition, a set-to-set comparison is also reported, regarding the changes in expression in the leading-edge subset (Figure 17B).

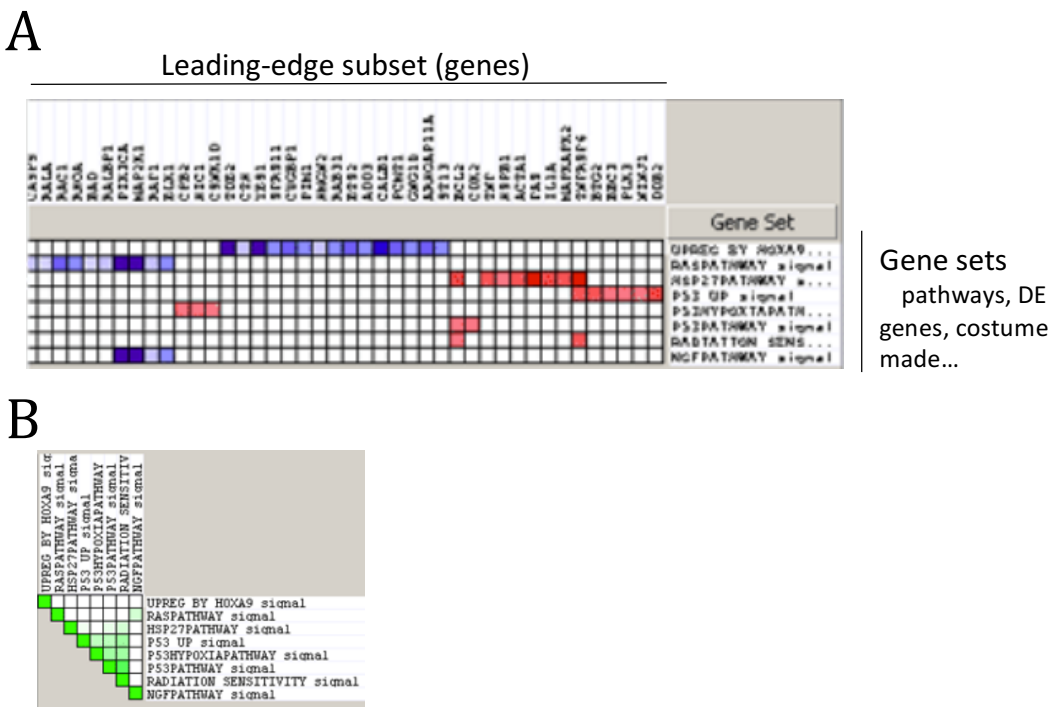


Figure 17: Leading-edge analysis in gene set enrichment analysis (GSEA). (**A**) Heatmap showing the expression of the leading-edge subset of an RNA-Seq experiment in several genes sets (pathways, DE genes or costume made...). (**B**) Set-to-set comparison. Specifically, the intensity of the cell for sets A and B corresponds to the X/Y ratio where X is the number of leading-edge genes from set A and Y is the union of the leading-edge genes in sets A and B. Figure adapted from [35].

Interestingly, GSEA allows to introduce costume made gene sets using a pre-ranked gene list and/or their expression values. As an example, two experimental conditions such as knock-down or knock-out of a specific protein (shRNA and CRISPR/Cas9 methodologies, among others) performed with two different methods

(RNA-Seq and microarray) can be compared using GSEA and costume made pre-ranked gene sets.

4.6. Splicing analysis and isoform detection

pre-mRNAs are post-transcriptionally processed to mature mRNA via a process known as splicing. After splicing, introns are removed and exons are joined together. In many cases, splicing creates unique proteins by varying the exon composition of the same pre-mRNA, known as alternative splicing. Those unique proteins, coming from the same gene, are known as isoforms. Isoform abundance is very tissue- and process-dependent and in some RNA-Seq experiments of great interest.

By analysing reads from an RNA-Seq experiments two kinds of reads are obtained: reads overlapping a single exon or reads containing a splice junction and thus, representing two exons. For paired-end reads, the same occurs with fragments in addition to reads. New splice junctions can be reported by analysing those reads. In addition, those reads allow the analysis of alternative splicing and relative isoform abundance.

Isoform detection is performed using specific softwares like MISO [36]. MISO software uses an annotation database of all possible alternative splicing events (isoforms), which are the following:

1. Skipped exons (ES)
2. Alternative 3'/5' splice sites (A3SS, A5SS)
3. Mutually exclusive exons (MXE)
4. Tandem 3' UTRs (TandemUTR)
5. Retained introns (RI)
6. Alternative first exons (AFE)
7. Alternative last exons (ALE)

Reads aligning to the alternative exon or to its junctions with adjacent constitutive exons (inclusion reads) provide support for the inclusion isoform, whereas reads aligning to the junction between adjacent constitutive exons (exclusion reads) support the exclusion isoform. For each possible alternative splicing event, MISO software will calculate the percentage spliced in (PSI or ψ) that denotes the fraction of mRNA that represent the inclusion isoform (ψ = inclusion reads / exclusion + inclusion reads) (Figure 18).

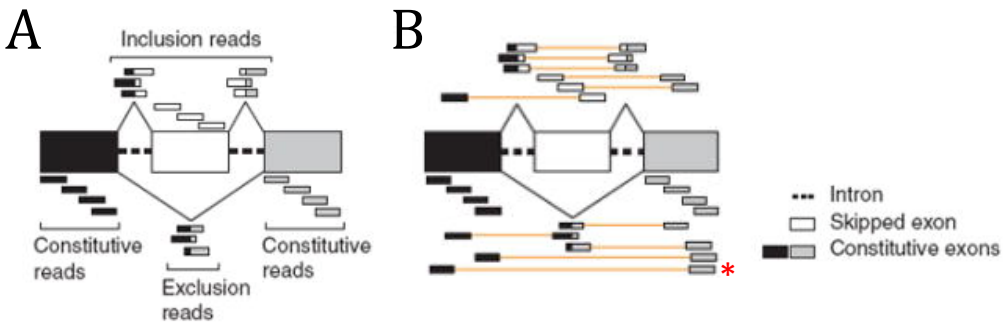


Figure 18: Types of reads (**A**) and fragments (**B**) in an RNA-Seq experiment, regarding splicing (constitutive exons, skipped exons and introns). Red asterisk marks splicing information obtained only with fragments (paired-end sequencing). Figure adapted from [36].

In RNA-Seq experiments analysing splicing, paired-end reads are highly recommended. Fragments can provide more information than reads as they can span two constitutive exons separated by a skipped exon (Figure 18B, marked with a red asterisk). This information cannot be obtained by analysing reads and they are all considered constitutive reads instead of exclusion reads. In addition, as some isoforms might be expressed at a very low level, a high depth of sequencing is also recommended (Table 3).

ENCODE consortium considers that quantification of individual transcript isoforms, being much more complex, can differ substantially depending on the processing pipeline employed and are of unknown accuracy. Therefore, alignments and gene quantification in an RNA-Seq can be used confidently, while transcript quantifications should be used with care.

4.7. Transcription of repetitive elements

Transcription of repetitive elements in an RNA-Seq experiment can be explored to assess genomic stability due to a high transcription of repetitive elements. In addition, RNAs from repetitive elements have been shown to be developmentally regulated and have regulatory functions, acting as non-coding RNA (ncRNA) or enhancer RNA (eRNA) (See *Introduction, 1.6. Heterochromatin and repetitive elements* and *3. Interferon response and chromatin*). RNA-Seq experiments are crucial in those analyses as not all repetitive elements are included in microarrays.

When designing an RNA-Seq experiment with the aim to analyse transcription of repetitive elements, total RNA preparations are not ploy(A) selected as repetitive

elements do not contain poly(A) tails. However, rRNA should be removed to increase the sample with target RNA species (Table 3).

Furthermore, an important challenge when analysing transcription of repetitive elements is depth of sequencing. Transcription of a repetitive element is highly dependent on the number of copies and its expression. Thus, depth of sequencing is very dependent on the sample (expression level) and the specific repetitive element (number of genomic copies).

Some repetitive elements contain unique sequences and are included in the reference genome. So, they can be analysed as gene transcripts using read counts (See *4.2. Normalization and estimation of transcript abundance*). In contrast, other repetitive elements are not included in the reference genome and a specific pipeline is performed (similar to a ChIP-Seq experiment, See *3.5. Enrichment analysis in repetitive elements*).

In short, the pipeline to analyse transcription of repetitive elements not included in the reference genome in a RNA-Seq experiment is the following:

1. Mask the reference genome to "cover" (with Ns) repetitive elements found in Repbase database with RepeatMasker software [17, 18].
2. Perform read alignment using spliced-read mapper (like TopHat or STAR [21, 22]) and report unaligned and multiple mapped reads.
3. Index Repbase database, necessary for some aligners such as Bowtie [3].
4. Align unaligned and with multi-reads to Repbase database (allowing multiple positioning and more mismatches). As RNAs from repetitive elements are not spliced, no spliced-read aligners are used (like Bowtie or BWA [3, 4]).
5. Report read counts per consensus repeat sequence.
6. Normalise counts as done for gene transcripts (See *4.2. Normalization and estimation of transcript abundance*).
7. Perform statistical testing between two conditions in each repetitive element using contingency tables and a Fisher exact test.

The consensus sequence of a repetitive element might be small and fragments might expand several repeat copies. When dealing with paired-end reads, fragment information cannot be obtained and paired-end reads need to be treated as single-end reads. Although fragment information is lost, counts are expected to increase roughly two times using paired-end reads.

As happens with ChIP-Seq experiments analysing repetitive elements, using this pipeline does not report the location of the repetitive element. It can only be reported

binding enrichment or transcription (ChIP-Seq and RNA-Seq, respectively) at specific repetitive elements (families or classes) (Table 2).

5. ATAC-Seq – Assay for transposase accessible chromatin sequencing

Assay for transposase accessible chromatin high-throughput sequencing (ATAC-Seq) is a method for mapping chromatin accessibility genome-wide (Table 1) [37, 38]. This method uses a hyperactive Tn5 transposase that cuts DNA and inserts sequencing adapters, mainly into accessible regions in chromatin (Figure 19A). Compared to other experiments assessing chromatin accessibility, like DNase-Seq or FAIRE-Seq, the number of cells used is smaller and the sample preparation time is shorter (Figure 19B).

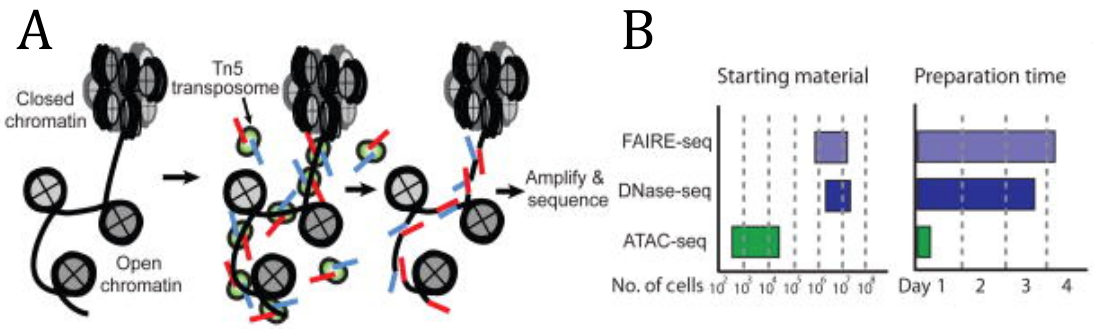


Figure 19: Assay for transposase accessible chromatin high-throughput sequencing (ATAC-Seq). (**A**) ATAC-Seq reaction schematic. Transposase is represented in green and adapters in red and blue. (**B**) Approximate reported input material and sample preparation times in genome-wide methods assessing chromatin accessibility. Figure adapted from [36].

In addition to DNA accessibility, ATAC-Seq reads can be used to infer nucleosome positioning. Paired-end reads are used as the fragment length is needed to determine nucleosome positioning. Fragment size distribution has a clear periodicity of approximately 200 base pairs, the size of a single nucleosome (Figure 20A). The more abundant fragments are those smaller than ~200bp, representing nucleosome-free regions (Figure 20A). Interestingly, the fragment length can be used as a measure of chromatin accessibility and its distribution clear correlates with specific chromatin features (Figure 20B).

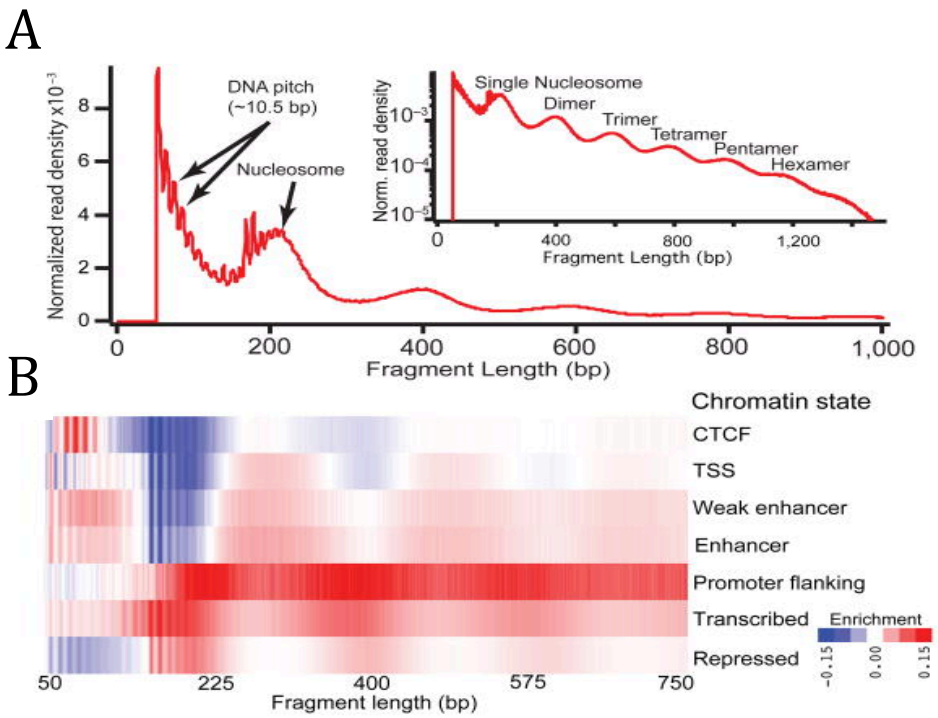


Figure 20: Fragment size distribution in an ATAC-Seq experiment. (**A**) Fragment size distribution shows a periodicity of 200bp. (**B**) Smaller fragments (<~200bp) represent nucleosome-free regions, differentially found in distinct chromatin features. Figure adapted from [37].

So, ATAC-Seq reads are divided in two groups using a simple heuristic model that positively weights nucleosome associated fragments and negatively weights nucleosome free fragments. Two datasets (tracks) are obtained: nucleosome-free and nucleosome signal. Chromatin accessibility is explored using nucleosome free track; instead, nucleosome positioning uses the nucleosome signal track. Indeed, their relative fraction shows nucleosome-free and nucleosome regions are clearly enriched at TSS and distal sites, respectively (Figure 21D).

It is worth mentioning that in ATAC-Seq experiments, the depth of sequencing needs to be carefully taken into account, depending on the biological question. A clear correlation exists between the size of the fragment and the number of reads, as larger is the fragment, lesser it is represented (low read density) (Figure 20A). Consequently, to assess nucleosome positioning in an ATAC-Seq experiment, high depth of sequencing is needed to obtain a high number of larger reads and consequently, a good coverage in the nucleosome signal track. In contrast, chromatin accessibility in ATAC-Seq experiments can be explored with lower depths of sequencing, as free-nucleosome fragments are highly abundant.

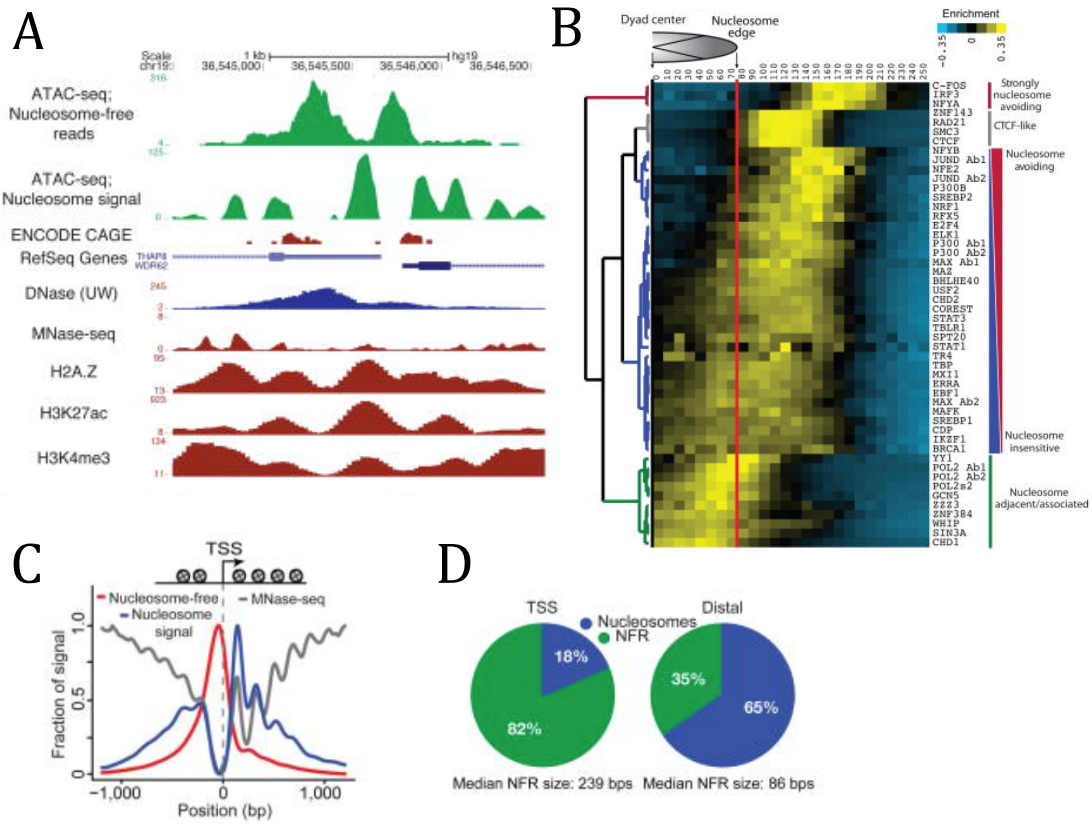


Figure 21: ATAC-Seq provide information on nucleosome positioning and chromatin accessibility. (**A**) Genome browser snapshot of nucleosome free and nucleosome signal at a gene containing two TSS. (**B**) Hierarchical clustering of DNA binding factor position with respect to the nearest nucleosome dyad. (**C**) Nucleosome-free and nucleosome signal tracks at TSS, representing chromatin accessibility and nucleosome positioning, respectively. (**D**) Relative fraction of nucleosomes vs nucleosome-free regions (NFR) in TSS and distal sites. Figure adapted from [37].

As an example, at a locus that contains a bidirectional promoter with two TSS separated by ~700bp, ATAC-Seq data shows in fact two distinct nucleosome free regions, separated by a single well-positioned mononucleosome (Figure 21A). Compared to MNase-Seq, ATAC-Seq data is more amenable detecting nucleosomes within putative regulatory regions as the majority of ATAC-Seq reads are concentrated within accessible regions of chromatin (Figure 21C). By averaging ATAC- and MNase-Seq signal at all active TSSs, it can be observed that ATAC-Seq nucleosome track is clearly enriched near TSS and its signal decreases at +2, +3 and +4 nucleosomes in contrast to MNase-Seq (Figure 21C). Thus, ATAC-Seq data can provide high-resolution of nucleosome-free regions and nucleosome positioning, mainly in regulatory elements genome-wide. An unsupervised hierarchical clustering of ChIP-Seq of a variety of DNA

binding factors and ATAC-Seq data can reveal major classes of binding with respect to the proximal nucleosome (Figure 21B).

Because ATAC-Seq reads are enriched at open chromatin sites, to assess nucleosome positioning genome-wide MNase-Seq is more suitable. However, ATAC-Seq can provide high-resolution information of chromatin accessibility and nucleosome positioning at open chromatin sites, putative regulatory regions genome-wide. Other methods like DNase- or FAIRE-Seq can only report chromatin accessibility and MNase-Seq is used for nucleosome positioning genome-wide (Figure 22).

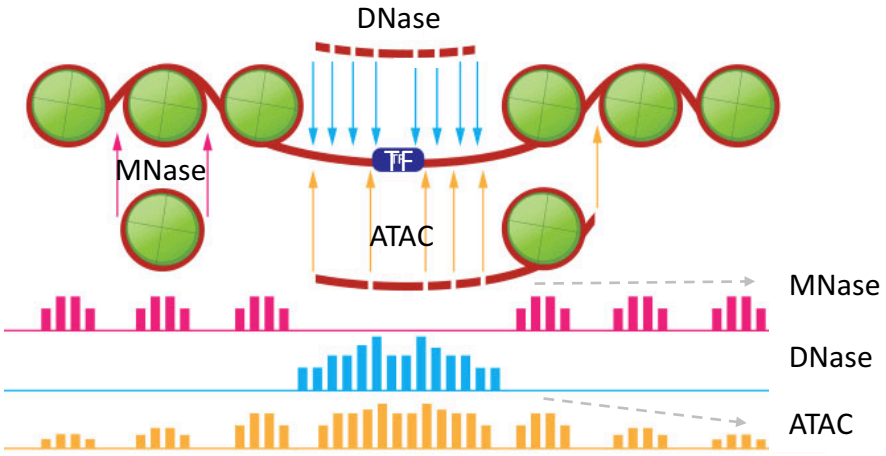


Figure 22: DNase-Seq and ATAC-Seq are used to sequence and map exposed regions of DNA, whereas MNase-Seq maps regions that are protected by nucleosomes. TF: transcription factor. Note that ATAC-Seq signal decreases when moving away from accessibility sites compared to MNase-Seq signal. Figure adapted from [39].

When comparing several genome-wide methods to assess chromatin accessibility, DNase- and ATAC-Seq do not provide data that perfectly complement those of MNase-Seq. The reason is that those methods provide snapshots of a dynamic process that is averaged across many thousands of cells. The fact that chromatin accessibility and nucleosome positioning at open chromatin sites can be explored, at high-resolution, within the same ATAC-Seq experiment, using a small number of cells (500 to 50.000 cells) is of great advantage until single-cell methodologies are developed.

Briefly, the pipeline to analyse an ATAC-Seq experiment is as follows. Paired-end reads are mapped to the reference like a ChIP-Seq experiment (See *3.1. Read mapping*). ATAC-Seq reads are then divided in two subsets (nucleosome and nucleosome-free signal) and analysed independently.

On the one hand, ATAC-Seq reads coming from larger fragments (> ~200bp) are analysed using algorithms to report nucleosome positioning like in a MNase-Seq experiment. As mentioned, ATAC-Seq will only report nucleosome positioning at accessible chromatin sites and only with high depths of sequencing.

On the other hand, reads representing nucleosome-free regions are analysed like a ChIP-Seq experiment. Peak calling is performed to assess accessibility sites using the same ChIP-Seq peak calling methods (See *3.2. Peak calling*). Narrow and broad peaks will represent accessibility in small regions (small nucleosome changes like those of transcription factors) or broad regions (exploring chromatin organization), respectively. For instance, pluripotent ESCs, characterized by a more "open" chromatin, will be enriched in ATAC-Seq broad peaks rather than in narrow peaks. Peak calling softwares like MACS2 allow to compute and compare both (narrow and broad) peaks at the same time [7].

Interestingly, when assessing peak calling within a sample of an ATAC-Seq experiment, the vast majority of peaks are mapped to constitutive active chromatin (Figure 23A). This kind of experiments assessing chromatin accessibility sites can be used to compare different cell types (ESCs vs differentiated cells) by overlapping peaks analysis.

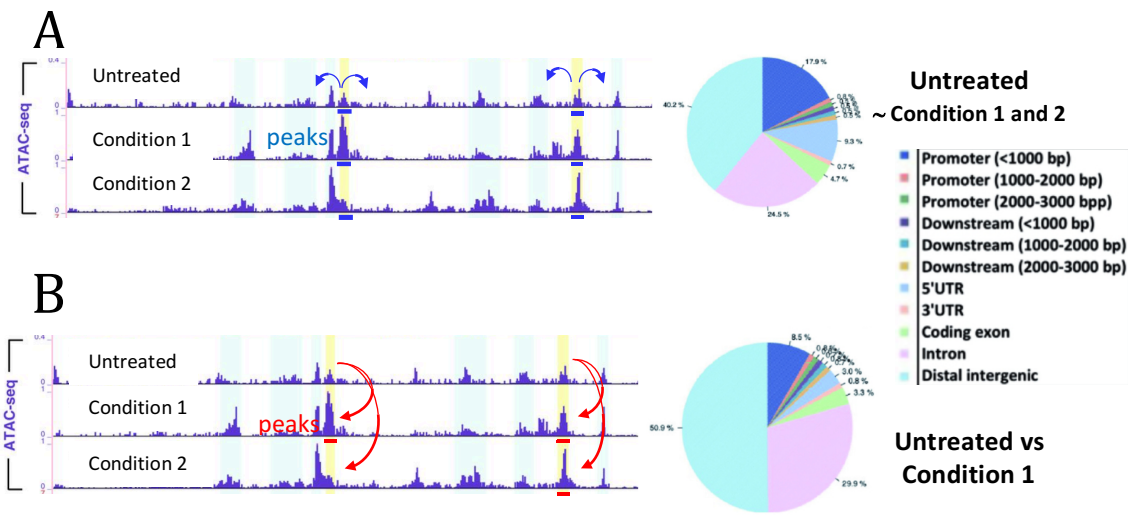


Figure 23: ATAC-Seq peak calling: ATAC-Seq signal browser snapshot and pie chart of the genomic annotation of ATAC-Seq peaks (**A**) Peak calling within the same sample to assess chromatin accessibility sites. Note the vast majority of peaks are on open genic regions (**B**) Peak calling between two conditions (differential peak calling). Figures adapted from ([40] and *Chapter II*).

Within the same cellular type, drastic changes (from fully closed to fully opened chromatin) are not observed under different conditions. So, differential peak calling

analysis needs to be performed to report differences between ATAC-Seq signals in the same cell type (Figure 23B). Thus, in the peak calling software one track is introduced as the background (untreated, input in a ChIP-Seq experiment), which is compared to a second ATAC-Seq track under a specific condition. It is worth mentioning that as ATAC-Seq experiments can only be performed with paired-end sequencing, read shifting in the peak calling is avoided and fragments are used instead of reads (Figure 3).

Down-stream analyses such as correlation of chromatin accessibility with expression, accessibility (peak overlap) at repetitive elements, annotation to genomic features, motif analysis… are performed as in a ChIP-Seq experiment using the ATAC-Seq signal and peaks.

6. References

1. Andrews S et al. (2010) FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

2. Kidder BL, Hu G and Zhao K (2011) ChIP-Seq: Technical considerations for obtaining high quality data. *Nat Immunol*, 12: 918-22

3. Langmead B, Trapnell C, Pop M and Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10: R25

4. Li H and Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25: 1754-60

5. Ruan J and Durbin R (2008) Mapping of short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*, 18(11): 1851-8

6. Pepke S, Wold B and Mortazavi A (2009) Computation for ChIP-seq and RNA-seq studies. *Nat Methods*, 6: S22-S32

7. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DA, Bernstein BE, Nusbaum C, Meyers RM, Brown M, Li W and Liu XS (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, 9: R137

8. Xu S, Grullon S, Ge K and Peng W (2014) Spatial clustering for identification of ChIP-enriched regions (SICER) to map regions of histone methylation patterns in embryonic stem cells. *Methods Mol Biol*, 1150: 97-111

9. Park PJ (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*, 10: 669-80

10. Zhu L, Gazin C, Lawson N, Pages H, Lin S, Lapointe D and Green M (2010) ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics*, 11: 237

11. Shin H, Liu T, Manrai AK and Liu XS (2009) CEAS: cis-regulatory element annotation system. *Bioinformatics*, 25: 2605-6

12. Quinlan AR and Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26: 841-2

13. Siepel A, bejerano G, Pedersen JS, Hinrichs A, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W and Haussled D (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genomes Res*, 15: 1034-50

14. Gel B, Diez-Villanueva A, Serra E, Buschbeck M, Peinado MA and Malinverni R (2016) regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics*, 32: 289-91

15. Treangen TJ and Salzberg SL (2011) Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet*, 13: 36-46

16. Day DS, Luquette LJ, Park PJ and Kharchenko PV (2010) Estimating enrichment of repetitive elements from high-throughput sequence data. *Genome Biol*, 11: R69

17. Jurka J (2000) Repbase Update: a database and an electronic journal of repetitive elements. *Trends Genet*, 9: 418-20

18. Smith AFA, Hubley R and Green P, RepeatMasker at http://repeatmasker.org

19. Wang Z, Gerstein M and Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10: 57-63

20. Benes, V Blake J and Doyle K (2011) Ribo-Zero Gold Kit: improved RNA-seq results after removal of cytoplasmic and mitochondrial ribosomal RNA. *Nature Methods*, 8: http://dx.doi.org/10.1038/nmeth.f.352

21. Trapnell C, Pachter L and Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25: 1105-11

22. Dobin A, Davis AC, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M and Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29: 15-21

23. Pruitt K, Brown G, Tatusova T and Maglott D (2002) The Reference Sequence (RefSeq) Database.

24. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken Bl, barrel D, Zadissa S, Searle S, Barnes I, Bignell A et al. (2012) GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res*, 22: 1760-74

25. Trapnell C and Salzberg SL (2009) How to map billions of short reads onto genomes. *Nat Biotechnol*, 27: 455-7

26. Li B and Dewey CN (2011) RSEM: accurate transcript qualification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12: 323

27. Wang L, Feng Z, Wang X, Wang X and Zhang X (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics*, 26: 136-8

28. Love MI, Huber W and Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15: 550

29. Robinson MD, McCarthy DJ and Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26: 139-40

30. Bailey TL, Bodén M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW and Noble WS (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*, 37: W202-8

31. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H and Glass CK (2010) Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell*, 38: 576-89

32. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardon JE, Ringwald M, Rubin GM and Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consotrium. *Nat Genet*, 25: 25-9

33. Kanehisa M, Furumichi, M, Tanabe M, Sato Y and Morishima K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45: D353-D361

34. Supek F, Boanjak M, Skunca N and Smuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, 6: e21800

35. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulocih A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102: 15545-50

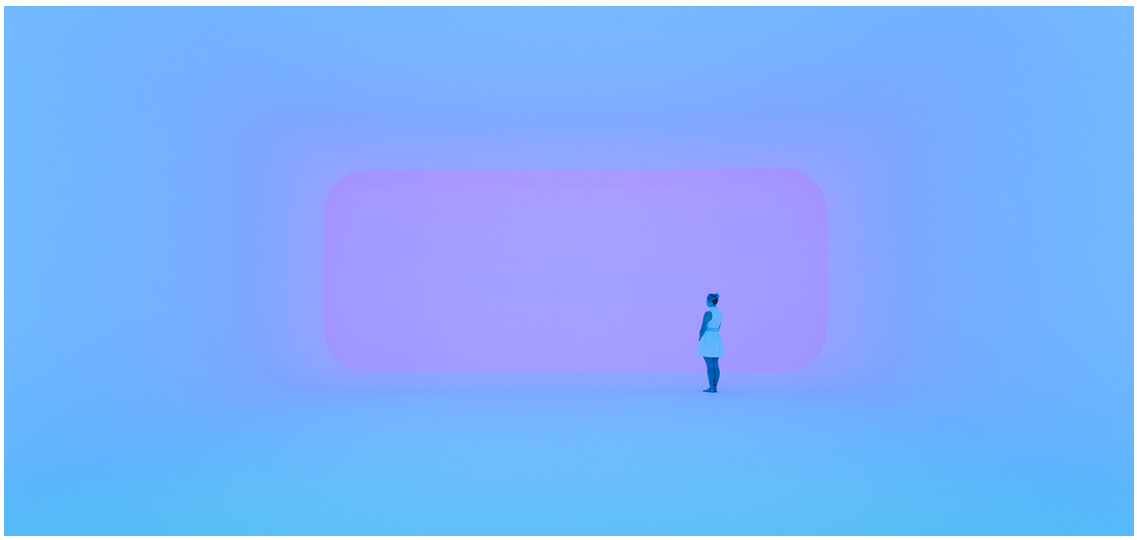
36. Katz Y, Wang ET, Airoldi EM and Burge CB (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods*, 7: 1009-15

37. Buenrostro JD, Giresi PG, Zaba LC, Chang HY and Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome positioning. *Nat Methods*, 10: 1213-8

38. Buenrostro JD, Wu B, Chang HY and Greenleaf WJ (2015) ATAC-Seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol*, 109: 21.29. 1-9

39. Kelly Rae Chi (2016) Reveling in the revealed. Based on epigenetic chromatin, 7:33

40. Minoux M, Holwerda S, Vitobello A, Kitazawa T, Kohler H, Stadler MB and Rijli FM (2017) Gene bivalency at Polycomb domains regulates cranial crest positional identity. Science, 355: eaal2913.



James Turrell