



TADs enriched in histone H1.2 strongly overlap with the B compartment, inaccessible chromatin, and AT-rich Giemsa bands

Núria Serna-Pujol¹, Mónica Salinas-Pena¹, Francesca Mugianesi², Natalia Lopez-Anguaita^{1,*}, Francesc Torrent-Llagostera¹, Andrea Izquierdo-Bouldstridge¹, Marc A. Marti-Renom^{2,3,4,5}  and Albert Jordan¹ 

¹ Molecular Biology Institute of Barcelona (IBMB-CSIC), Spain

² CNAG-CRG, Centre for Genomic Regulation, The Barcelona Institute of Science and Technology, Spain

³ Centre for Genomic Regulation, The Barcelona Institute for Science and Technology, Spain

⁴ Pompeu Fabra University, Barcelona, Spain

⁵ ICREA, Barcelona, Spain

Keywords

genome compartments; Giemsa bands; histone H1 variants; linker histone; TADs

Correspondence

A. Jordan, IBMB-CSIC, Baldri Reixac 4, Barcelona 08028, Spain

Tel: +34 93 402 0487

E-mail: albert.jordan@ibmb.csic.es

*Present address

Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, Berlin, 14195, Germany

(Received 4 May 2020, revised 22 July 2020, accepted 1 September 2020)

doi:10.1111/febs.15549

Giemsa staining of metaphase chromosomes results in a characteristic banding useful for identification of chromosomes and its alterations. We have investigated *in silico* whether Giemsa bands (G bands) correlate with epigenetic and topological features of the interphase genome. Staining of G-positive bands decreases with GC content; nonetheless, G-negative bands are GC heterogeneous. High GC bands are enriched in active histone marks, RNA polymerase II, and SINEs and associate with gene richness, gene expression, and early replication. Low GC bands are enriched in repressive marks, lamina-associated domains, and LINEs. Histone H1 variants distribute heterogeneously among G bands: H1X is enriched at high GC bands and H1.2 is abundant at low GC, compacted bands. According to epigenetic features and H1 content, G bands can be organized in clusters useful to compartmentalize the genome. Indeed, we have obtained Hi-C chromosome interaction maps and compared topologically associating domains (TADs) and A/B compartments to G banding. TADs with high H1.2/H1X ratio strongly overlap with B compartment, late replicating, and inaccessible chromatin and low GC bands. We propose that GC content is a strong driver of chromatin compaction and 3D genome organization, that Giemsa staining recapitulates this organization denoted by high-throughput techniques, and that H1 variants distribute at distinct chromatin domains.

Databases

Hi-C data on T47D breast cancer cells have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number [GSE147627](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE147627).

Abbreviations

bp/hs, bands per haploid sequence; G band, Giemsa band; Gneg, negative (unstained) Giemsa bands; Gpos, Giemsa-positive (stained) bands; LAD, lamina-associated domain; mESCs, mouse embryonic stem cells; NAD, nucleolus-associated domain; PTM, post-translational modification; RNAPII, RNA-polymerase II; S/MAR, scaffold or matrix attachment region; TAD, topologically associating domain; TSS, transcription start site.

Introduction

Eukaryotic DNA is packaged into chromatin, whose repeating structural unit is the nucleosome. Each nucleosome consists of an octamer of core histones (H2A, H2B, H3, and H4) around which ~ 147 base pairs (bp) of DNA are wrapped. Histone H1 binds at both entry/exit sites to the linker DNA at the nucleosome, participating in the formation of higher-order chromatin structures [1]. Unlike core histones, H1 proteins are more evolutionary diverse. The human histone H1 family includes seven somatic subtypes (or variants) (H1.1 to H1.5, H1.0, and H1X), three testis-specific (H1t, H1T2, and H1LS1), and one oocyte-specific variant (H1oo) [2–4]. Among somatic variants, H1.1–H1.5 variants are expressed in a replication-dependent manner while H1.0 and H1X are replication-independent. Regarding their patterns of expression, H1.2 to H1.5 and H1X are ubiquitously expressed, H1.1 is restricted to certain tissues, and H1.0 accumulates in terminally differentiated cells.

This large repertoire of H1 variants leads to wonder whether somatic H1 variants are redundant or show specific properties in terms of functionality and genomic distribution. Classically, H1 has been seen as a structural component associated with chromatin compaction, but in recent years, several evidences support the idea of H1 playing a more dynamic role in chromatin regulation [4,5]. Previous studies have shown that histone H1 variants are involved in several nuclear processes including transcription, replication, genome stability, splicing, or heterochromatin maintenance, among others [6–10].

To fully characterize H1 variants specific functionality, it is important to address their genomic distribution, due to the growing evidence that chromatin organization is crucial to genome function. Reports point to a variant-specific genomic distribution among cell types. In mouse embryonic stem cells (ESCs), H1c and H1d (H1.2 and H1.3 orthologs) were found to be depleted from high GC, gene-rich regions, and abundant at major satellites [11]. By using DamID technology in human IMR90 cells, results showed that H1.2–H1.5 was depleted from CpG-dense and regulatory regions, whereas H1.1 had a distinct profile [12]. Besides, H1.5 was enriched in genic and intergenic regions in IMR90 cells but not in ESCs, suggesting that its genomic distribution depends on the differentiation state [13]. In human fibroblasts, mapping of H1.0 revealed its correlation with GC content and abundance at gene-rich chromosomes [14]. In breast cancer cells, H1.2 was the variant that showed the most specific pattern. H1.2 was found enriched in low

GC domains and lamina-associated domains (LADs) [15]. Moreover, combined depletion of H1.2 and H1.4 leads to the activation of heterochromatic repeats, supporting the role of H1.2 in heterochromatin organization [10]. Regarding replication-independent variants, H1.0 and H1X were more abundant at high GC, gene-rich chromosomes. H1.0 was also found enriched at nucleolus-associated domains (NADs) while H1X was more associated with coding regions and RNA polymerase II binding sites [16]. A general feature for all H1 variants, in all species, is its depletion from the transcription start site (TSS) of active genes, meaning that upon transcriptional activation H1 is removed from the TSS of genes.

Nevertheless, although uncovering specific features for H1 variants, data support that all H1 variants are distributed across the whole genome [15]. For this reason, methods to compartmentalize the genome could be useful to study and compare H1 variants genomic distribution. Due to the complex paradigm of chromatin organization, this compartmentalization has to be addressed by a multi-omics approach.

From a functional point of view, the genome has classically segregated into euchromatin and heterochromatin. Transcriptionally active euchromatin present an open state to facilitate accession of transcription machinery, replicates early within S-phase, and is abundant in SINE repetitive elements and active histone modifications. On the contrary, closed and transcriptionally silent heterochromatin is characterized by late replication timing, LINEs and inactive histone modifications [17]. Moreover, it is well established that chromosomes occupy a nonrandom regions in the nucleus (chromosome territories), where gene-poor regions are placed at the heterochromatic nuclear periphery and gene-rich ones to the euchromatic interior. Chromosome conformation capture techniques (such as Hi-C) have revealed the existence of topologically associating domains (TADs), self-organized chromatin domains in spatial proximity that interact more frequently within themselves than with the rest of the genome [18–20]. These structures are conserved across species and are relatively stable in different cell types [18,21]. Hi-C data also lead to the discovery of the so-called A and B genomic compartments, comprising active and inactive regions, respectively [22]. Independently, other chromatin domains participating in nucleus organization have been described, such as aforementioned LADs or NADs [23,24].

Other layers of chromosome architecture have also been studied for years. In 1970s, several staining methods of metaphase chromosomes arised, that is, Giemsa

staining [25]. Although the precise molecular basis of Giemsa has remained unknown for decades, it is widely accepted that staining correlates with AT-rich sequences and chromatin compaction [26,27]. Giemsa bands (G bands) have been useful in cytogenetics allowing detection of chromosomes rearrangements in diseased cells. However, they have not been much explored in relation to functional genomics. Staining of G-positive (Gpos) bands correlates with AT content; nonetheless, unstained or G-negative (Gneg) bands, expected to be GC-rich, are as heterogeneous in its GC or AT content as Gpos.

Here, we used G bands as epigenetic units to investigate the differential distribution of linker histones. We have *in silico* investigated how G bands correlate with epigenetic, accessibility and topological features of the interphase genome, taking advantage of previously published ChIP-seq, ATAC-seq, and newly generated Hi-C data in breast cancer cells. Our results show a heterogeneous and opposite distribution of histones H1.2 and H1X within G bands, being H1.2 associated with low GC bands and H1X with high GC bands. We have found a strong correlation between B compartment, TADs presenting a high H1.2/H1X ratio, low GC bands, and compact chromatin. To our knowledge, this is the first report including an extensive characterization of G bands based on a wide repertoire of genome-wide data, including H1 variants abundance or Hi-C experiments, among others. Moreover, the balance between two H1 variants has never been considered as an epigenetic feature nor related to genome topology before. Overall, this work represents a comprehensive attempt to further investigate how chromatin is organized within the nucleus, integrating histone H1 variants as putative chromatin organizers.

Results

Characterization of Giemsa bands with epigenetic features and GC content dependency

Giemsa staining of metaphase chromosomes results in an alternating dark and light banding pattern that became useful for identifying individual chromosomes and their abnormalities in diseased cells (Fig. 1A). After the sequencing of the human genome and with the help of a dynamic programming algorithm employing data from thousands of fluorescence *in situ* hybridization experiments, the boundaries of each of the bands were estimated [28]. The estimated starting and ending position of each of the Giemsa-positive bands, classified into four groups according to its increasing staining intensity (Gpos25-Gpos100), and

intergenic bands (Gneg), was obtained from the UCSC human genome database. The number of Gpos bands ranged from 81 to 121. Gpos bands occupied from 7.6% to 17.6% of the genome and Gneg bands a 46% (Fig. 1B).

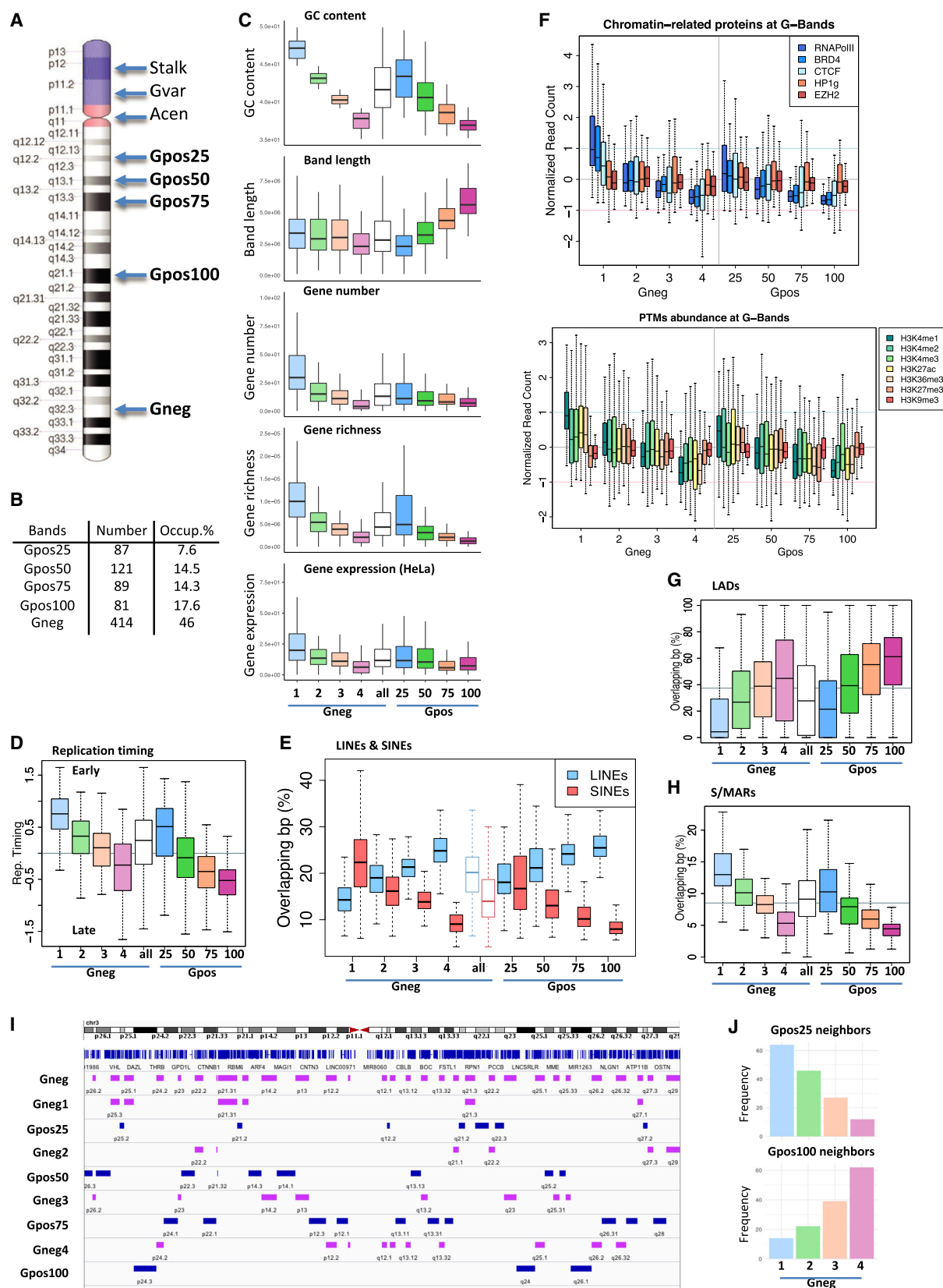
The molecular basis of cytogenetic bands is not well understood. Banding was thought to correspond to GC-poor (dark bands) and GC-rich (light bands) regions. However, Gpos100 bands were consistently AT-rich, but Gpos25 and particularly Gneg bands were highly heterogeneous in its GC content (Fig. 1C). Gneg bands presented a mean GC content intermediate between Gpos25 and Gpos50, indicating that banding could not be explained only by the base composition. Therefore, we wanted to investigate whether banding could be explained by epigenetic features such as core histone marks or linker histone variants.

Darker bands (Gpos100) were longer on average and have been associated with chromatin condensation. Accordingly, they contained the lowest gene content, gene richness, and average gene expression of all the bands (Fig. 1C), as well as longer introns (data not shown). Gneg bands presented intermediate features. As a consequence, we decided to split the Gneg bands in four equivalent groups according to their GC content (Gneg1–4). Gene richness and gene expression correlated positively with GC content (Fig. 1C).

Replication occurs first at active/open chromatin and later at compact chromatin. Data on replication timing for HeLa cells are available, and we used it to calculate the average replication timing at each G band. As expected, within G-positive bands, replication timing was lower (late) at Gpos100 (Fig. 1D). Within G-negative bands, replication timing correlated with the GC content; high GC bands replicated the earliest.

It was previously reported that darker bands are enriched in LINEs and G-negative bands are enriched in SINEs [28]. We have calculated, per chromosome, the percentage of bases in each of the eight band types that is contained within LINEs and SINEs (Fig. 1E). The abundance of SINEs correlated with the GC content, whereas abundance of LINEs correlated with AT content, more than with the darkness of G bands.

Next, we explored the abundance of core histone post-translational modifications (PTMs) and transcription or chromatin-related proteins (from breast cancer T47D cells publically available data) at Gneg and Gpos bands (Fig. 1F). On the one hand, the abundance of PTMs related to gene activation and factors such as RNA polymerase II (RNAPII), BRD4, or CTCF decreased accordingly to the GC content, that is, being high within Gneg1 and Gpos25 and low at



Gneg4 and Gpos100 bands. On the other hand, repressive marks such as H3K27me3 or H3K9me3, as well as EZH2 methyl transferase and heterochromatin protein HP1 gamma, did not follow this pattern following GC content. Instead, they were more abundant than active marks at Gpos75 and Gpos100 bands, but also at Gneg4. The overlap between LADs and G bands also increased at low GC bands, particularly at Gpos100 (Fig. 1G), coinciding with H3K9me3 enrichment over active marks.

Scaffold (metaphase) or matrix (interphase) attachment regions (S/MARs) are involved in control of gene expression, replication, DNA repair, and chromatin to chromosome transition. By linking DNA to the nuclear scaffold, they generate structural and functional loops that span ≈ 20 –100 kb. S/MARs are relatively short sequences (100–1000 bp long) containing one or several of these features: AT richness ($\approx 70\%$), OriC, kinked or curved DNA, TG richness, and topoisomerase-II sites [29]. Because of their AT richness, it was initially proposed that S/MARs were present densely within dark G bands [30]. Mapping of human S/MARs using ChIP-seq data of 14 S/MAR binding proteins was recently achieved [31]. These sites were confirmed to contain the previously described features including AT richness. Nonetheless, we found that they were enriched within high GC bands, both Gneg and Gpos bands (Fig. 1H), as expected for elements involved in the control of gene expression and replication. Accordingly, S/MAR density was found to correlate with gene density [31]. Moreover, S/MARs also correlated with retrovirus integration sites [31]. Accordingly, we found that hotspots for retroviral integration were enriched within high GC bands (data not shown).

As a consequence of this analysis, Gneg interbands were seen epigenetically heterogeneous, being its GC content an important predictive factor of its

characteristics, but not of its lack of Giemsa staining. We then hypothesized that Gneg bands surrounding Gpos bands with a particular GC content could have similar GC values, forming patches of bands with similar features, as shown in Fig. 1I. In fact, we determined that the most abundant neighbors of Gpos25 bands were Gneg1 bands, and Gpos100 bands were preferably surrounded by Gneg4 bands (Fig. 1J). In conclusion, Giemsa-stained bands were surrounded by unstained bands of similar GC content, gene content, and other features, except that were shorter.

Correlation of Giemsa staining with AT content is enhanced along chromosome condensation

Along chromosome condensation, the initially observed banding of 850 bands per haploid sequence (bphs) (prometaphase) gets condensed down to 400 bphs (metaphase) (Fig. 2A) [32]. We predicted that neighbor Gpos and Gneg bands (at 850 bphs) with similar GC content would become either dark (stained) or white (unstained) bands at 400 bphs, depending on its GC content, as shown in Fig. 2B. As an example, bands p25.1, p25.3 (Gneg) and p25.2 (Gpos25) of chromosome 3 became band p25 (white, high GC) at 400 bphs, while bands p14.1, p14.3 (Gpos50), and p14.2 (Gneg) became band p14 (dark, low GC) (Fig. 2B). As a consequence, the difference between GC content at stained versus not-stained bands was increased at 400 bphs compared with 850 bphs, that is upon chromosome condensation (Fig. 2C).

Analyzing what proportion of each of the G bands at 850 bphs became dark or white at 400 bphs, we obtained that a big proportion of Gneg4 became dark, and a big amount of Gpos25 and Gpos50 became white (Fig. 2D). A circular permutation of G bands at 850 bphs confirmed that this observation depends on

Fig. 1. Characterization of Giemsa bands. (A) Ideogram of a human metaphase chromosome showing banding after Giemsa staining. G-positive bands are classified into four types (Gpos25 to Gpos100) according to increasing staining intensity. Unstained bands or interbands are called G negative (Gneg). Ideograms are from NCBI's Genome Decoration Page. (B) Table indicating the number of bands of each type existing in human chromosomes and the percentage of base pair occupancy in the genome. (C) Box plots showing the GC content, base pair length, gene number content, gene richness (gene number/base pair length), and average gene expression (from HeLa cells) of each G band for each band type. Gneg bands were divided into four equal groups according to GC content. (D) Box plot showing replication timing at each G band (normalized by band length), for each band type. HeLa-S3 public Repli-seq data were used. (E) Box plot showing the proportion of overlapping base pairs between LINEs or SINEs and each G band, for each band type. (F) Box plots showing abundance of chromatin-related proteins or histone PTMs at each G band, for each band type. Enrichment was calculated by computing the average normalized read count of the peaks mapped at each G band. Publically available data from T47D cells were used, except for EZH2 and H3K27me3 that correspond to HeLa cells. (G) Box plot showing overlapping base pairs between LADs and each G band, for each band type. (H) Box plot showing overlapping base pairs between S/MARs and each G band, for each band type. (I) Browser snapshot of human chromosome 3 showing the position of Gpos and Gneg bands. (J) Bar plots showing the frequency of Gneg band groups that are neighbors of Gpos25 or Gpos100 bands.

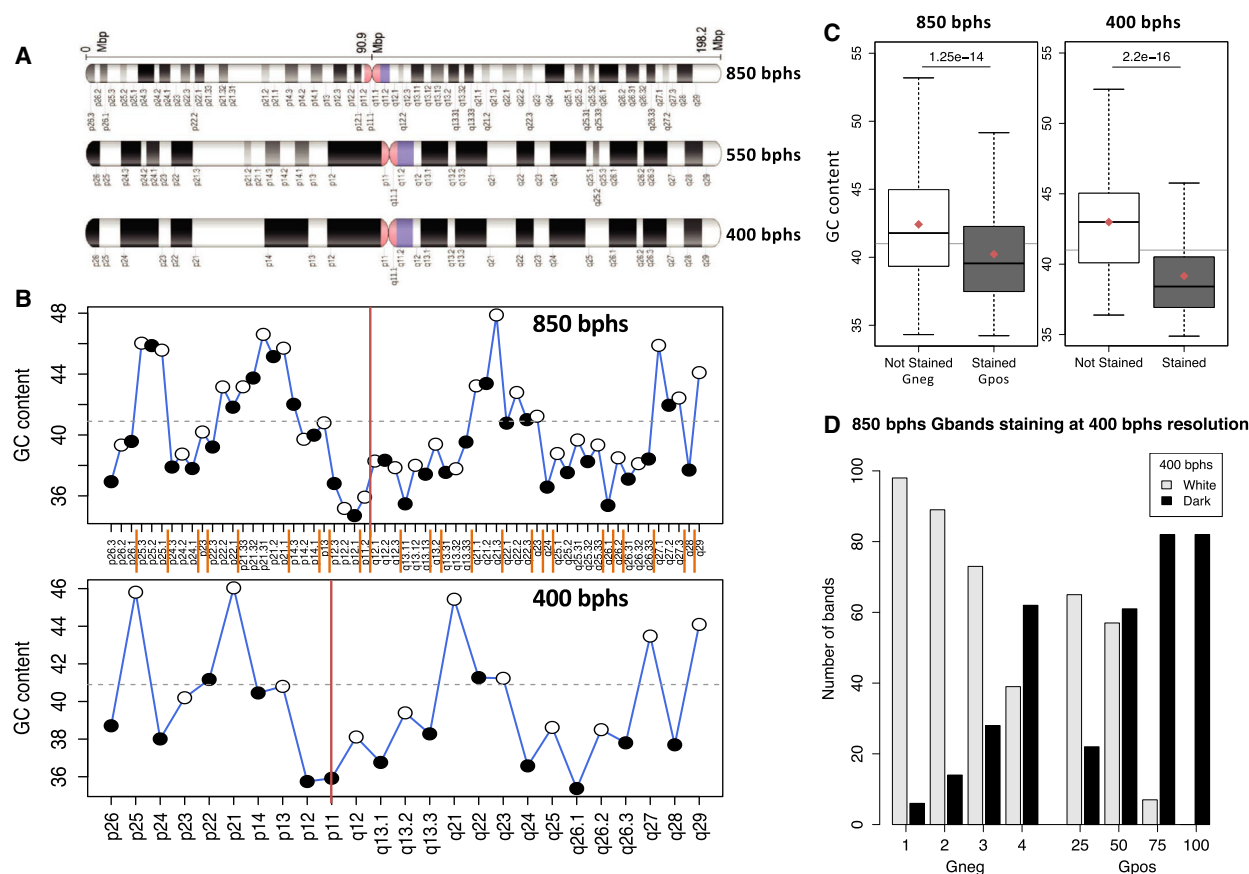
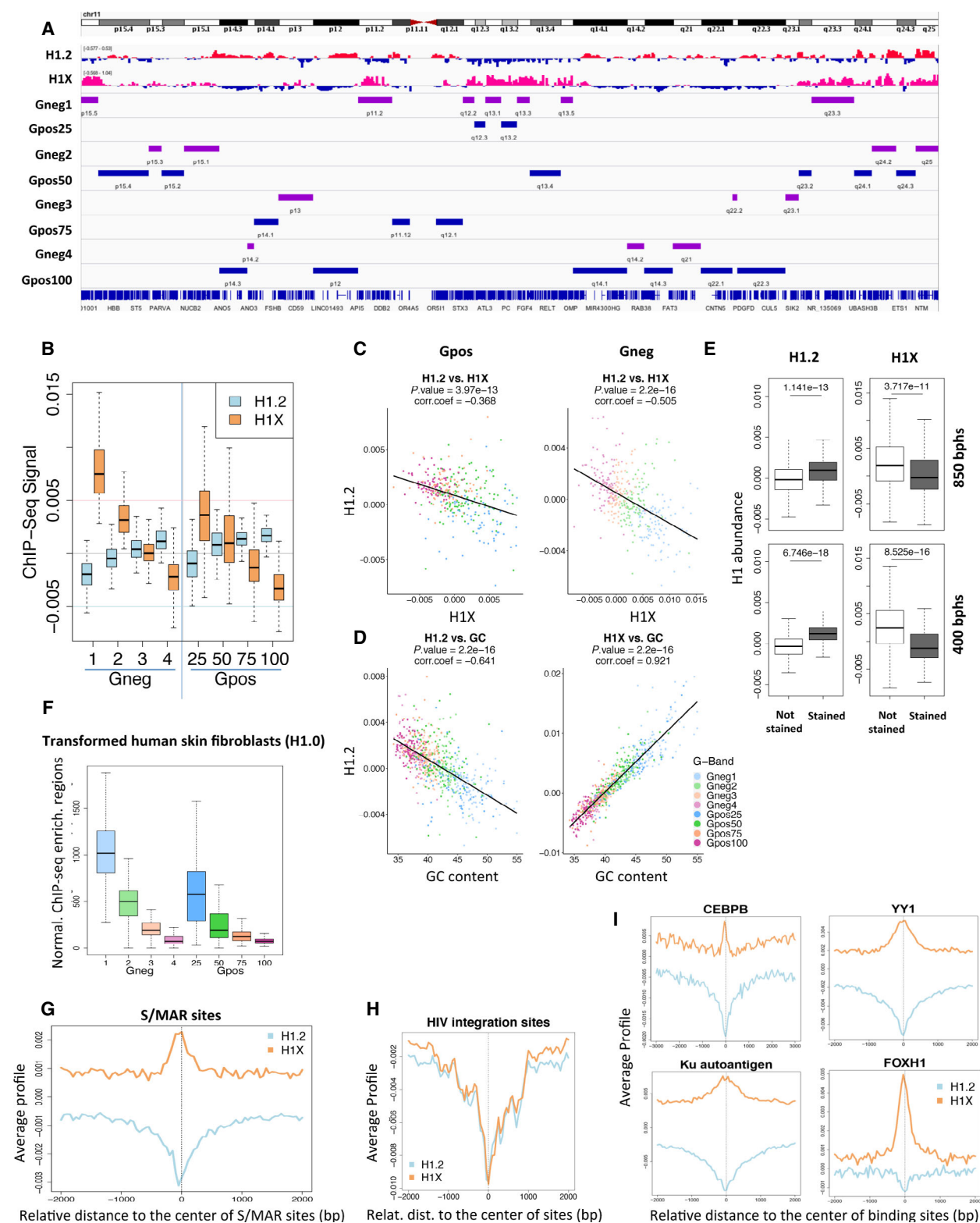


Fig. 2. Correlation of Giemsa staining with AT content is enhanced along chromosome condensation. (A) Ideograms of human chromosome 3 at 850, 550, and 400 bands per haploid sequence (bps) resolution. Along metaphase condensation, the number of G bands (resolution) decreases, and bands are classified just as stained (dark) or unstained (light). Ideograms are from NCBI's Genome Decoration Page. (B) Representation of GC content of Gpos/stained (dark circle) and Gneg/unstained (light circle) bands along chromosome 3, at 850 and 400 bps resolution. Clusters of bands at 850 bps that are merged to a single band at 400 bps are separated by orange lines. (C) Box plots showing the GC content of Gneg and Gpos bands at 850 and 400 bps. The Wilcoxon test was used to evaluate the significance of the differences in GC content. (D) Bar plot showing the frequency of bands of each type at 850 bps that end up stained (dark) or unstained (white) at 400 bps.

the actual position of the bands (data not shown). Bands that changed their staining status along condensation and ended stained or not as expected according

to its GC content (Gneg4 and Gpos25, respectively) are the shortest bands on average (Fig. 1C). Then, these bands could be seen as short interbands

Fig. 3. Histone H1 variants distribute heterogeneously among G bands. (A) Browser snapshot of human chromosome 11 showing H1.2 and H1X input-subtracted ChIP-seq signal from T47D cells and the position of Gpos and Gneg bands. (B) Box plots showing H1.2 and H1X input-subtracted ChIP-seq abundance within G bands, for each band type. (C) Scatter plots of H1.2 and H1X input-subtracted ChIP-seq abundance at each Gpos (left) or Gneg (right) band. Pearson's correlation coefficient is shown as well as *P*-value. (D) Scatter plots of H1.2 or H1X input-subtracted ChIP-seq abundance against GC content at each Gneg and Gpos band. Pearson's correlation coefficient is shown as well as *P*-value. (E) Box plots showing H1.2 and H1X input-subtracted ChIP-seq abundance within unstained (light) or stained (dark) G bands at 850 or 400 bps. The Wilcoxon test was used to evaluate the significance of the differences in H1.2 and H1X enrichment. (F) Box plots showing the normalized number of H1.0 ChIP-seq enrichment regions from *in vitro* transformed human skin fibroblasts (GSE66169) within G bands. (G–I) Abundance of H1 variants at retroviral integration sites and S/MAR protein binding sites. Average, input-subtracted ChIP-seq signal of H1.2 and H1X around the center of S/MARs sites (mapped in [31]) (G), HIV-1 integration sites (H), or around the center of the S/MAR binding protein sites indicated (I).



inadequately stained initially (850 bphs), that mimic surrounding, larger bands later (400 bphs), forming larger patches stained or not according to their GC content. These results reinforced the notion that Giemsa staining depends on AT richness, but this is better seen in highly condensed chromosomes. Still, correlation is not perfect because, even at 400 bphs, some stained bands have higher GC content than some unstained bands (Fig. 2B). Nonetheless, locally, stained bands always have lower GC than neighbor unstained bands. This suggests that chromosomes are partitioned into a small number of large domains of high or low GC content and, within them, relative differences in GC dictate band staining.

Another possibility to explain the lack of correlation between staining and GC content at 850 bphs could be that staining was more sensitive to the existence of long AT tracks than to the average AT content. Nonetheless, we obtained that AT content and abundance of AT tracks correlated (correlation coefficient = 0.994, P -value < 0.001), and the number of AT tracks of different lengths was not more different between Gpos and Gneg bands than its average AT content, discarding this hypothesis (data not shown).

In summary, AT content is partially responsible for the intensity of Giemsa banding of metaphase chromosomes and correlates with epigenetic features of chromatin already present at interphase chromosomes.

Histone H1 variants in breast cancer cells distribute heterogeneously among G bands

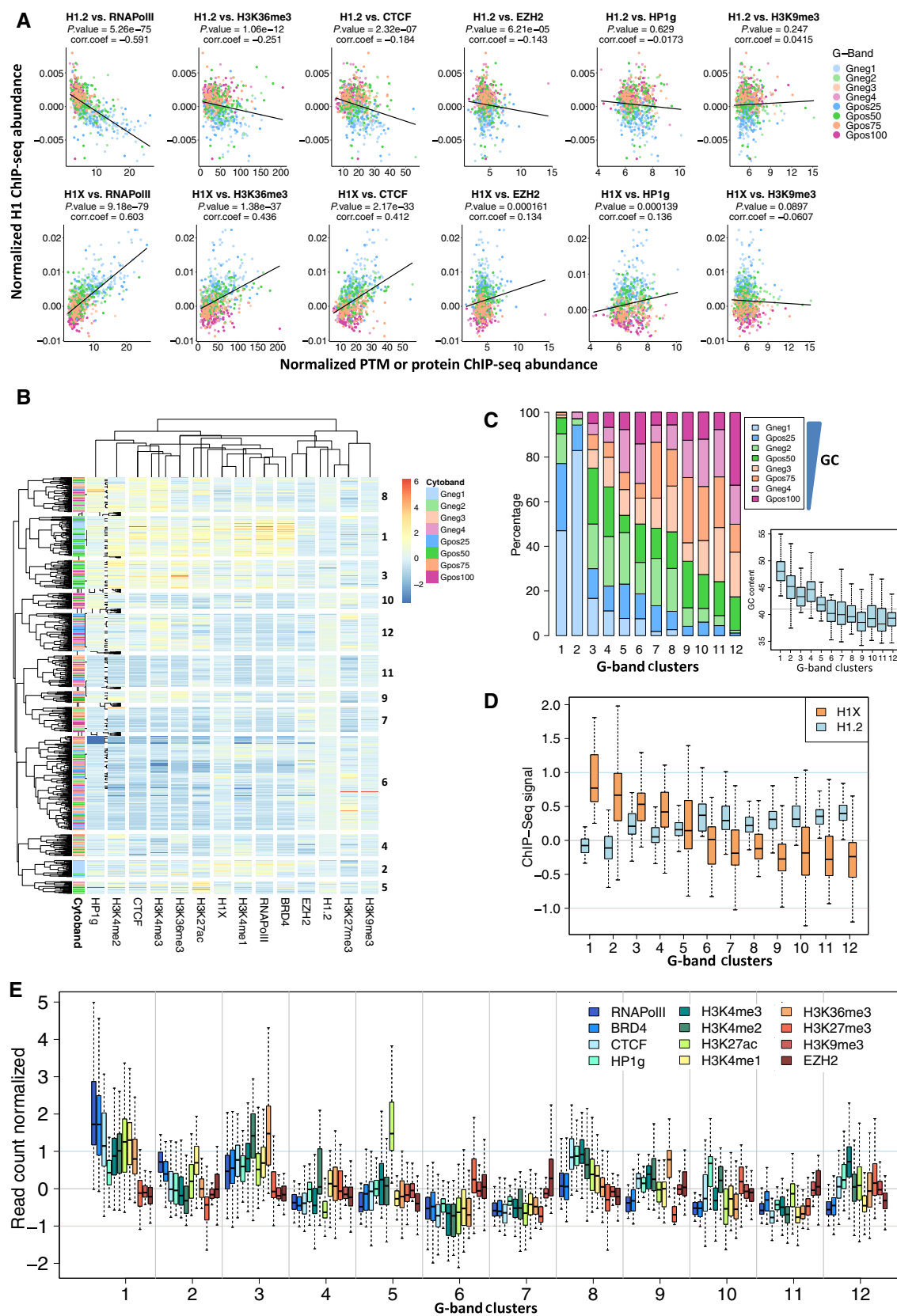
We have previously reported that histone H1 variants distribute heterogeneously along the human genome in T47D breast cancer cells, being H1.2 the variant that is more abundant within closed and intergenic regions, and H1X the most abundant within RNA polymerase II-enriched regions [15,16]. Then, we interrogated whether the abundance of these two H1 variants differed among G bands. In a genome browser, it was apparent that H1X was enriched at Gneg1 and Gpos25, while H1.2 was more abundant at Gneg4 and Gpos100, suggesting a relation with GC content

(Fig. 3A). Indeed, H1X was enriched at G bands with high GC content, while H1.2 was rich at low GC bands, both G-positive and G-negative (Fig. 3B). As a consequence, H1.2 and H1X abundance at both types of G bands correlated inversely (Fig. 3C). The positive correlation between H1X and the GC content of bands was stronger than the inverse correlation between H1.2 and GC content (correlation coefficient 0.92 versus −0.64) (Fig. 3D). Differences in H1 variants abundance at stained versus not-stained G bands were enhanced at 400 bphs compared with 850 bphs (Fig. 3E), as it occurred with GC content (Fig. 2C). H1.2 was significantly enriched at stained bands, and H1X was more abundant at nonstained bands.

Parallel to profiling the distribution of endogenous H1.2 and H1X with variant-specific antibodies, we had profiled H1.0 and H1.4 C-terminally tagged with the hemagglutinin (HA) peptide, stably expressed in T47D cells, with anti-HA antibodies. H1.4-HA and H1.0-HA distribution was similar to H1X and different to H1.2 [15,16]. We calculated the abundance of these two HA-tagged variants into G bands. We obtained that both were enriched toward high GC bands, being H1.0-HA the one that was more similar to H1X, in agreement with our previous reports (data not shown). Moreover, using published data on H1.0 profiling in human skin fibroblast [14], we determined that H1.0 was enriched at high GC Gpos and Gneg bands as well (Fig. 3F). Therefore, we decided to focus on endogenous H1.2 and H1X for further studies, as representatives of the different H1 profiles observed.

Because S/MARs and H1X were enriched at high GC bands in a very similar way (Figs 1H and 3B), we compared the abundance of H1.2 and H1X around the center of mapped S/MAR sites. H1X was enriched at S/MAR sites while H1.2 was clearly depleted (Fig. 3G). As mentioned above, S/MARs correlated with retrovirus integration sites [31], which were enriched within high GC bands. Instead, both H1.2 and H1X were found depleted from putative HIV-1 and HTLV-1 integration sites, suggesting that retroviruses integrate at H1-depleted loci (Fig. 3H and data not shown). This showed that H1X was not enriched at all features that are enriched

Fig. 4. Clustering of G bands according to H1 variants and epigenetic features. (A) Scatter plots of H1.2 or H1X input-subtracted ChIP-seq abundance against abundance of the indicated histone marks or chromatin-associated proteins at each Gpos and Gneg band. Pearson's correlation coefficient is shown as well as P -value. (B) Heat map and dendrogram of the abundance of H1 variants, histone marks, and chromatin-associated proteins at Gpos and Gneg bands. Twelve clusters of G bands are shown, ordered from high to low proportion of high GC content bands (Gneg1 + Gpos25 + Gneg2 + Gpos50). (C) Bar plot showing the proportion of each G band type within the 12 clusters of bands generated in (B). GC content at clusters is also shown. (D) Box plots showing H1.2 and H1X input-subtracted ChIP-seq abundance within G bands, for each G bands cluster. (E) Box plots showing abundance of the indicated histone marks or chromatin-associated proteins within G bands, for each G bands cluster.



within high GC bands, but functional selectivity exists. Next, we calculated the abundance of H1s around the binding sites of the proteins that were used to define S/MAR [31]. H1X was more abundant than H1.2 at all S/MAR protein binding sites, but different profiles were observed (Fig. 3I and data not shown). For CEBPB, YY1, Ku antigen, and FOXH1, H1X was locally enriched around the center of the binding site, but not for the others (CTCF, NMP4, BRIGHT, BRCA1, SAF-A/hnRNP-U, SATB1, SMAR1). H1.2 was depleted from all tested sites. This was previously observed for RNA polymerase II [15,16]. All together, these data confirmed that H1X is present at places where transcription and replication initiate.

Clustering of G bands according to H1 variants and epigenetic features

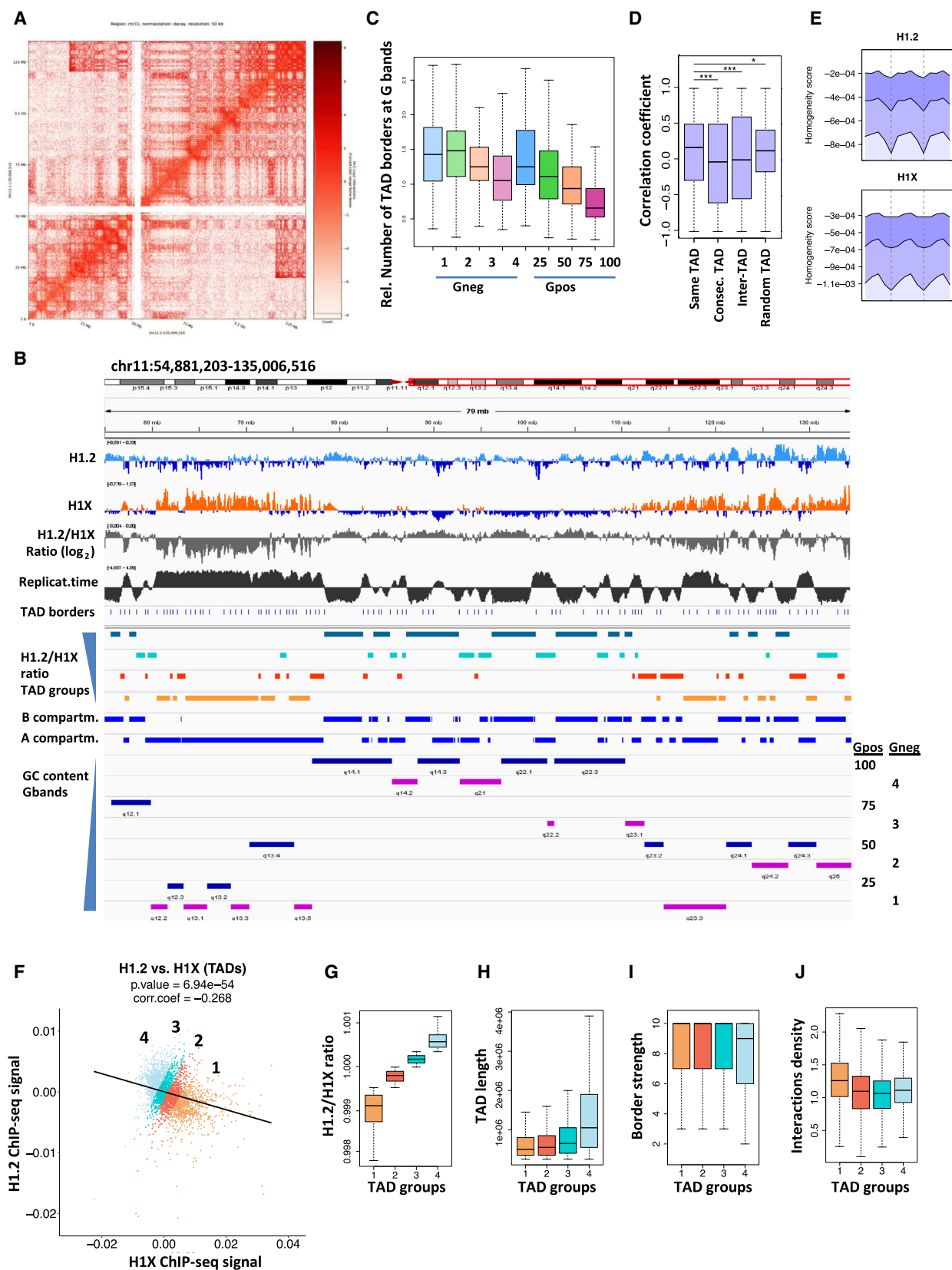
To study the colocalization of the different H1 variants with epigenetic factors within the G bands, the abundance of H1.2 and H1X at Gpos and Gneg bands in T47D cells was compared with the abundance of PTMs and chromatin-associated factors (Fig. 4A and Fig. S1A,B). H1.2 correlated negatively with active histone marks, such as H3K27ac, H3K4me1, or H3K36me3, and RNAPII, BRD4, or CTCF. Significant negative correlation was also observed between H1.2 and EZH2, related to transcriptional repression. The repressive marks H3K9me3 or H3K27me3 showed no correlation with H1.2 abundance, nor HP1 gamma. Instead, H1X correlated positively with all histone marks and chromatin-associated factors tested, except H3K9me3 and H3K27me3. All these results were

similar when the abundance of H1s and PTMs at Gpos or Gneg bands was used separately (Fig. S1C–F) and confirms that H1.2 localizes at inactive G bands whereas H1X is more abundant at high GC content bands enriched in active chromatin.

Next, the calculated abundance of H1 variants, core histone marks, and chromatin-associated factors at Gpos and Gneg bands was used to cluster the G bands and, consequently, compartmentalize the human genome according to epigenetically relevant features (Fig. 4B). Active marks, RNAPII, CTCF, and H1X clustered together, as did H3K9me3, H3K27me3, and EZH2 with H1.2. Next, G bands were clustered into 12 groups with 6 clusters enriched in active epigenetic features and 6 in repressive marks. Each cluster contained a different proportion of G band types; clusters were named from 1 to 12 according to decreasing proportion of high GC content bands (Fig. 4C). As expected, GC content decreased along the defined clusters (see insert in Fig. 4C).

Next, the abundance of H1 variants and epigenetic features at G bands contained in each of the 12 clusters was calculated (Fig. 4D,E). H1 variants increased or decreased progressively according to the GC content of the bands included in each cluster, particularly H1X, as H1.2 was similarly abundant at clusters 6 to 12. Clusters 1 to 4 were enriched in H1X while clusters 6 to 12 were enriched in H1.2. RNAPII or active histone marks were enriched toward the high GC content clusters, in particular clusters 1 to 3, but also cluster 8. Repressive marks or EZH2 was enriched in clusters 6, 7, and 9 to 12. Interestingly, cluster 2 contained predominantly Gneg1 bands and was enriched in H1X,

Fig. 5. Clustering of TADs according to its content in histone H1 variants. (A) Hi-C interaction map of chromosome 11 in T47D cells, at the resolution of 50 kb. The map is normalized, corrected by decay, and in Log2 scale. (B) Representative IGV snapshot of human chromosome 11 (partial). Tracks refer as follows (from top to bottom): H1.2 and H1X input-subtracted ChIP-seq signal from T47D cells; the calculated H1.2/H1X ratio (log2) over 100-kb bins; replication timing of the genome from T47D cells (smoothed signal of early/late S-phase read counts in 5 kb windows); TAD borders obtained by Hi-C in T47D cells; the extension of TADs classified into four groups according to H1.2/H1X ratio as described in (F–G); the extent of A/B compartments obtained by Hi-C; and the position of Gpos and Gneg bands. (C) Box plot showing the number of TAD borders within each G band, corrected by band length, for each G band type. (D, E) TADs as homogeneous units of H1 variants abundance. (D) Distributions of pairwise correlation coefficients of H1 profiles between 100-kb genome bins located within the same TAD, within consecutive or randomly picked TADs (inter-TADs), or within a similar randomly defined domain (*** $P < 0.001$; * $P < 0.05$; Wilcoxon test). (E) Homogeneity score of linker histones enrichment between consecutive subsegments over three successive TADs. For this analysis, TADs were divided into five subsegments of equal size. The opposite of the absolute difference of the H1 variants ChIP-Seq signal was calculated for two consecutive subsegments on three consecutive TADs. Higher scores indicate higher similarity between the consecutive subsegments. The 25th, 50th, and 75th percentiles (black lines from top to bottom, respectively) of the 14 consecutive values were computed genome-wide. Dashed lines correspond to the TADs borders. (F) Scatter plot of H1.2 and H1X ChIP-seq abundance at each individual TAD. Pearson's correlation coefficient is shown as well as P -value. TADs corresponding to the four groups defined in (G) according to H1.2/H1X ratio are differentially colored. (G) Box plot showing the ChIP-seq H1.2/H1X ratio within TADs in the four groups generated with equal count of TADs in each. (H–J) Box plots showing the base pairs length (H), border strength (I), and interactions density (J) of TADs belonging to the four groups defined according to H1.2/H1X ratio.



RNAPII, H3K4me1, also in BRD4 and H3K27ac, but not other active marks. Cluster 1 was clearly the most active one with absence of repressive features, while cluster 6 showed the highest abundance of repressive features and absence of active ones. Cluster 8 was enriched in particular features that formed a cluster in the dendrogram: CTCF, HP1 gamma, and H3K4me2/3. Cluster 5 was highly enriched in H3K27ac, whereas it was quite neutral on the rest of active/inactive features, including H1.2 and H1X.

In summary, clustering of G bands according to epigenetic features and H1 variants content compartmentalized the human genome and identified different types of chromatin units. Interestingly, when clusters were ordered according to the abundance of high GC bands or GC content, H1 variants decreased or increased progressively, something not clearly seen for the other epigenetic features or proteins, indicating that histone H1 best correlates with the GC content of the genome.

Overlap between TADs defined by the abundance of H1 variants, G banding, A/B compartments, replication timing, and ATAC-seq accessibility regions

Chromosome conformation capture techniques, such as Hi-C [22], allow to detect local and distal contacts within the genome and to establish the position of borders flanking the so-called TADs. We performed Hi-C experiments in T47D cells, and we calculated the position of TADs within the genome, obtaining a total of 3247 TADs. Figure 5A shows the normalized Hi-C interaction map of chromosome 11 at the resolution of 50 kb as an example.

The comparison of the positions of TADs and G bands denoted that often the limits of G bands were in close proximity to TAD borders (Fig. 5B); therefore, we further investigated the coincidences between

these two features and in relation to H1 variants abundance. First, we calculated the frequency of TAD borders that fell into each of the G bands normalized by their length. Gneg bands and, in general, high GC bands, showed a higher relative frequency of TAD borders than Gpos100 (Fig. 5C). Moreover, Gpos100 is longer on average than other G bands (Fig. 1C). As a consequence, Gneg and high GC Gpos bands are shorter and contain several short TADs, whereas Gpos100 (and Gpos75) contains one or a few long TADs (Fig. 5B).

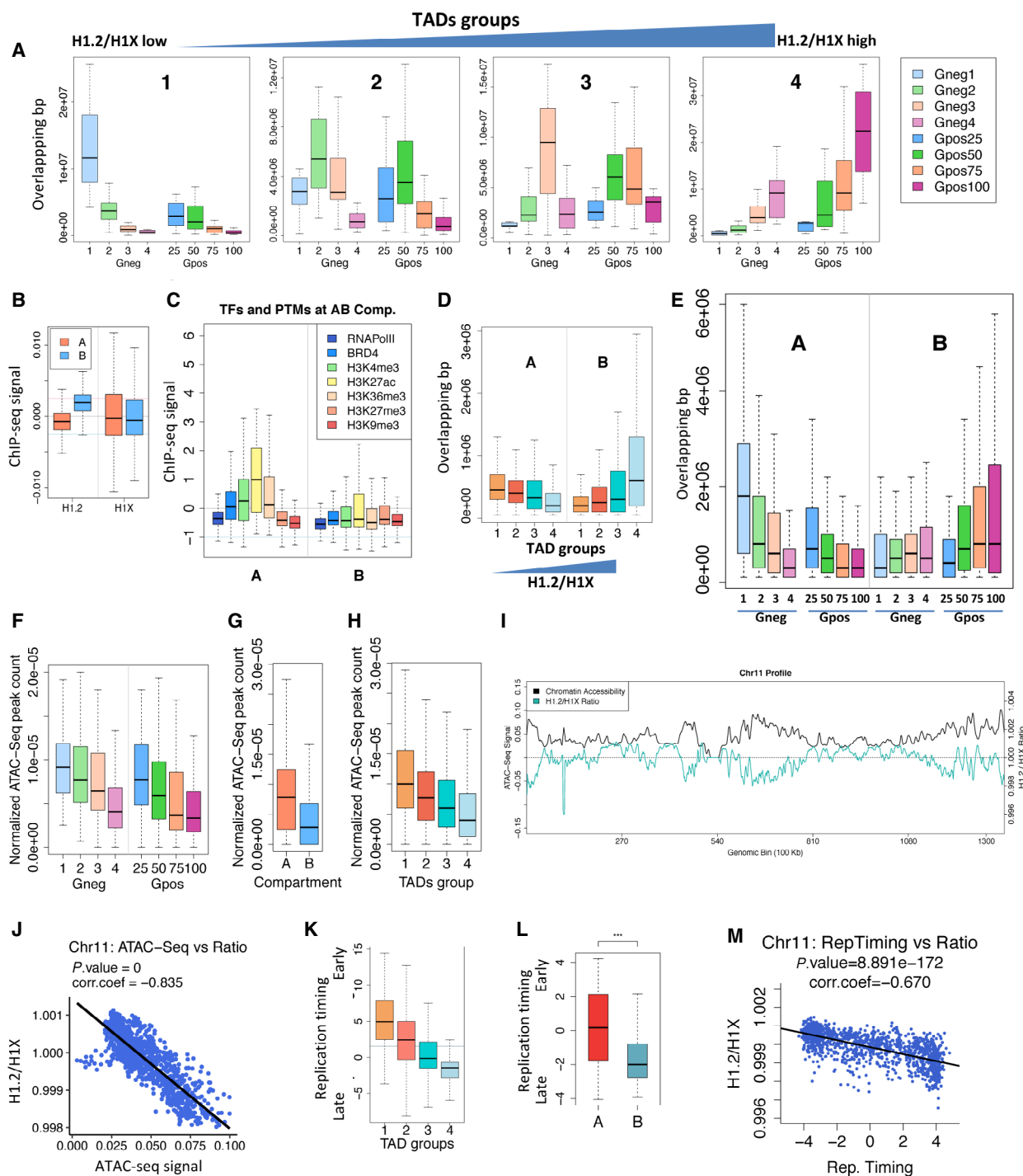
We observed that shifts on the distribution of H1 variants often coincided also with TAD borders (Fig. 5B). Before using TADs as units to compare the distribution of H1.2 and H1X variants, we asked whether this distribution (calculated within 100-kb bins) was more homogeneous within the same TAD than between consecutive TADs, randomly picked TADs or randomly defined domains. Correlation coefficient between the two H1 variants was significantly higher within the same TAD than any other comparison, suggesting that H1 variants were more homogeneous within than between TADs and that transitions between variants occurred preferentially at the borders (Fig. 5D). Besides, we performed 5000 randomizations of TAD borders to further confirm whether the relationship between H1s occupancy depends on these genomic units. Our results showed that the average correlation coefficient between the histones was significantly higher within the real TAD borders compared with the distribution of the average correlation coefficients calculated for the random domains (data not shown). This hypothesis was additionally tested by dividing TADs into subsegments and computing a homogeneity score of linker histones enrichment, which was higher between intra-TAD subsegments (Fig. 5E). Given that TADs and G bands tend to overlap (Fig. 5B), we also performed this analysis for G bands. We found that linker histones distribution

Fig. 6. Overlap between TAD groups defined by H1.2/H1X ratio, G bands, A/B compartments, ATAC-seq accessibility regions, and replication timing. (A) Box plot showing overlapping base pairs between TADs classified according to H1.2/H1X ratio (from low, Group 1; to high, Group 4) and the G bands. (B) Box plot showing the occupancy of H1.2 and H1X variants (input-subtracted ChIP-seq signal) within A/B compartments. (C) Boxplots showing the average normalized read count of the peaks mapped at A or B compartments of each histone PTM or chromatin-associated protein indicated. (D) Box plot showing overlapping base pairs between TADs classified according to H1.2/H1X ratio (Groups 1 to 4) and the A/B compartments ($N_A = 1098$, $N_B = 1098$). (E) Box plot showing overlapping base pairs between G bands and the A/B compartments. (F–H) Box plots showing the relative number of ATAC-seq peaks within G bands (F), A/B compartments (G), or TADs classified according to H1.2/H1X ratio (Groups 1 to 4), normalized by TAD length (H). (I) Profiles of ATAC-seq accessibility and H1.2/H1X abundance ratio along chromosome 11, calculated within 100 kb bins. (J, M) Scatter plots between ChIP-seq H1.2/H1X abundance ratio and ATAC-seq accessibility (J) or replication timing (M) within 100-kb bins along chromosome 11. Pearson's correlation coefficient is shown as well as *P*-value. (K, L) Box plot showing the T47D replication timing (ENCODE) (normalized by TAD length) within TADs classified according to H1.2/H1X ratio (Groups 1 to 4) (K), or within A/B compartments (L).

was more homogeneous within the same G band than between consecutive, alternate, or within similar random genomic regions (data not shown).

Next, we calculated the abundance of H1.2 and H1X within each TAD (Fig. 5F). As expected, an

inverse correlation was observed. The ratio between H1.2 and H1X abundance was calculated for each TAD and used to generate four equal groups of TADs, from low to high H1.2/H1X ratio (Fig. 5G). TADs with a high H1.2/H1X ratio, presumably more



compacted, were much larger in average (Fig. 5H) and were enriched in Gpos bands, especially Gpos100 (Fig. 6A). Instead, TADs with the lowest H1.2/H1X ratio were enriched in high GC bands, mainly Gneg1. In fact, TADs with similar H1.2/H1X ratios are seen as clusters that resemble the G bands (Fig. 5B). Long stretches of TADs with a high H1.2/H1X ratio greatly overlap with Gpos100 bands and so on. This allows us to propose that G bands extension and staining correlate with the relative abundance of two histone H1 variants with opposite genomic distribution and are related to the topology of the genome, which has been proposed to be highly conserved between cell types, as occurs for G banding.

Hi-C data allow to compute the relative strength of each TAD border and the relative intra-TAD interactions density in which each TAD is involved. TAD border strength was slightly higher in TADs with the lowest H1.2/H1X ratio (Fig. 5I). Those TADs also presented a major abundance of TADs with a high interaction density (Fig. 5J). Border strength and interactions density within TADs correlated positively (correlation coefficient = 0.274, P -value < 0.001). In conclusion, TADs with low H1.2/H1X ratio, the GC-rich ones, are better defined according to their border strength and present a higher relative number of interactions given their size, as expected from open chromatin genome regions. TADs with high H1.2/H1X ratio, within AT-rich G bands, are not defined as well and present less interactions, probably because they are immersed in closed chromatin regions, as shown below.

Hi-C experiments also allow to establish a division of the genome into two compartments, A (active) and B (repressive). We hypothesized that A/B compartments could also reflect differences in the abundance of H1.2 and H1X and maybe greatly overlap with the stretches of TADs defined by the H1.2/H1X ratio or even with the G bands staining (Fig. 5B). From our Hi-C data, we established the A/B compartments and calculated the abundance of the H1 variants in each A or B compartment fragment. B compartment was greatly enriched in H1.2, whereas H1X was only slightly increased in A compartment (Fig. 6B). Instead, A compartment was enriched in active histone H3 marks and transcription factors (Fig. 6C). B compartment was highly enriched in the group of TADs containing a high H1.2/H1X ratio. Instead, A compartment was enriched in TADs with low H1.2/H1X ratio (Fig. 6D). As expected, B compartment greatly overlapped with the Gpos bands (Gpos75 and Gpos100), whereas A compartment overlapped with high GC bands (Gneg1, Gneg2, and Gpos25)

(Fig. 6E). Moreover, G bands that present a higher base pair overlap with the B compartment showed a higher AT content (correlation coefficient = 0.56, P -value < 0.001).

Next, we used accessibility data of T47D cells previously obtained by ATAC-seq [10] to calculate its overlap with G bands, A/B compartments, and TADs classified according to H1.2/H1X ratio. High GC content G bands showed a major density of accessibility peaks (Fig. 6F). Interestingly, Gpos50, 75, and 100 were particularly deprived of accessibility peaks. As predicted, A compartment was also enriched in high accessibility regions compared with B compartment (Fig. 6G). Moreover, accessibility peaks were enriched within TADs presenting a low H1.2/H1X ratio, denoting that H1.2-rich TADs are more compact (Fig. 6H). This was further confirmed by profiling along chromosomes the ATAC-seq accessibility and ChIP-seq H1 variants abundance within 100-kb bins; it was evident that H1X correlates strongly with accessibility, while H1.2 or the H1.2/H1X ratio correlated negatively with accessibility (Fig. 6I,J and data not shown). This reinforces the relationship of H1.2 and H1X with repressed and active genomic regions, respectively.

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

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

Finally, we studied whether replication timing correlated with all features described here. We showed above that high GC cytobands replicated earlier than low GC bands (Fig. 1D). Replication timing represented in a browser formed clusters that clearly overlapped the TAD clusters defined by H1.2/H1X ratio, genome compartments, and G bands (Fig. 5B). Late replicating regions overlapped with TADs enriched in H1.2 and the B compartment (Fig. 6K,L). A strong inverse correlation existed between replication timing and the H1.2/H1X ratio within 100-kb bins along chromosomes (Fig. 6M).

In summary, topological domains enriched in H1.2 compared with H1X or other variants correspond to poorly accessible, late replicating regions that overlap with the B compartment of the 3D genome and with the low GC Giemsa bands of the metaphase chromosomes (Table 1).

Correlation between epigenetic scores and chromatin accessibility within G bands clustered according to H1 variants and epigenetic features

Taking advantage of the topological and accessibility data available, we further analyzed the (12) G bands clusters generated to compartmentalize the genome using histone PTMs, H1 variants, and chromatin proteins (Fig. 4B). First, we calculated the base pair overlap between A/B compartments and the G bands included in each of the 12 clusters. Bands within clusters 1 to 5 with high GC content (abundant Gneg 1, Gneg2, and Gpos25) and high abundance of H1X and active marks were located mainly in A compartment. Bands within clusters 8 to 12 with low GC content (abundant Gpos100, Gpos75, and Gneg4) and high H1.2 abundance showed major overlap with the B compartment (Fig. 7A).

Further, we represented in a 3D plot the twelve G bands clusters according to their H1.2/H1X ratio, a calculated compartment B/A ratio, and a repressive or heterochromatic 'epigenetic score' obtained from the ratio between the average abundance of repressive versus active histone marks or chromatin factors (Fig. 7B). As expected, clusters with high H1.2/H1X ratio also showed high B/A compartments ratio and repressive epigenetic score, that is, clusters 9 to 12. Nonetheless, this representation denoted particularities of several clusters that have been described above. Cluster 6 presented the highest repressive epigenetic score, and cluster 8 an epigenetic score lower than expected according to its H1.2/H1X and B/A ratios. Among the clusters with low H1.2/H1X and B/A ratios, that is, 1 to 4, there is some heterogeneity on

the epigenetic score, being clusters 1 and 3 those showing the highest proportion of active marks (Figs 4E and 7B,D).

When computing the ATAC-seq accessibility within the 12 G bands clusters described above, it was observed that initial clusters enriched in H1X and located within the A compartment were more accessible than clusters of bands enriched in H1.2 (Fig. 7C). Still, the best correlation of accessibility occurred with the calculated epigenetic score; clusters 1, 3, and 8 presented the lowest repressive epigenetic score (or the highest active/euchromatic epigenetic score) and the highest accessibility (Fig. 7C,D). Pairwise correlations between the different parameters studied in the 12 clusters of G bands confirmed that the active/euchromatic epigenetic score correlated the best with ATAC-seq accessibility (Spearman's correlation = 0.76), but not as well with GC content, A compartment or H1X (Fig. 7E).

In conclusion, by first dividing the heterogeneous Gneg bands in four groups according to GC content and, later, all Giemsa bands into 12 clusters according to the abundance of H1 variants and other epigenetic features, we functionally compartmentalized the genome in a way that allowed to search for correlations with accessibility and topological data. Each cluster contained G bands of different types that presented common features. The GC content within each cluster was not more homogeneous than inside each of the five original Giemsa bands categories (Gneg, Gpos25–100), indicating that GC content was not the main parameter dictating clustering once epigenetic features were used. All together, we propose that the clustering made here including H1 variants may be useful to identify and characterize different functional chromatin units inside the human genome.

The overlap between H1.2-rich TADs, the B compartment, and AT-rich G bands is extensive to mouse ESCs

Finally, we asked whether the correlations described here were extensive to other cell types or species. Different H1 variants correlate with high or low GC content in different studies [4,11–16]. Genomic localization data on H1.2 are not available elsewhere, except for DamID studies of H1 variants in IMR90 human fibroblasts and ChIP-seq of tagged variants in knock-in mouse ESCs [11,12]. In both cases, H1.2 was abundant at low GC DNA as in T47D cells. We obtained available data on mouse ESCs H1 ChIP-seq and Hi-C [11,33], together with the coordinates of mouse Giemsa bands from UCSC server, to test the

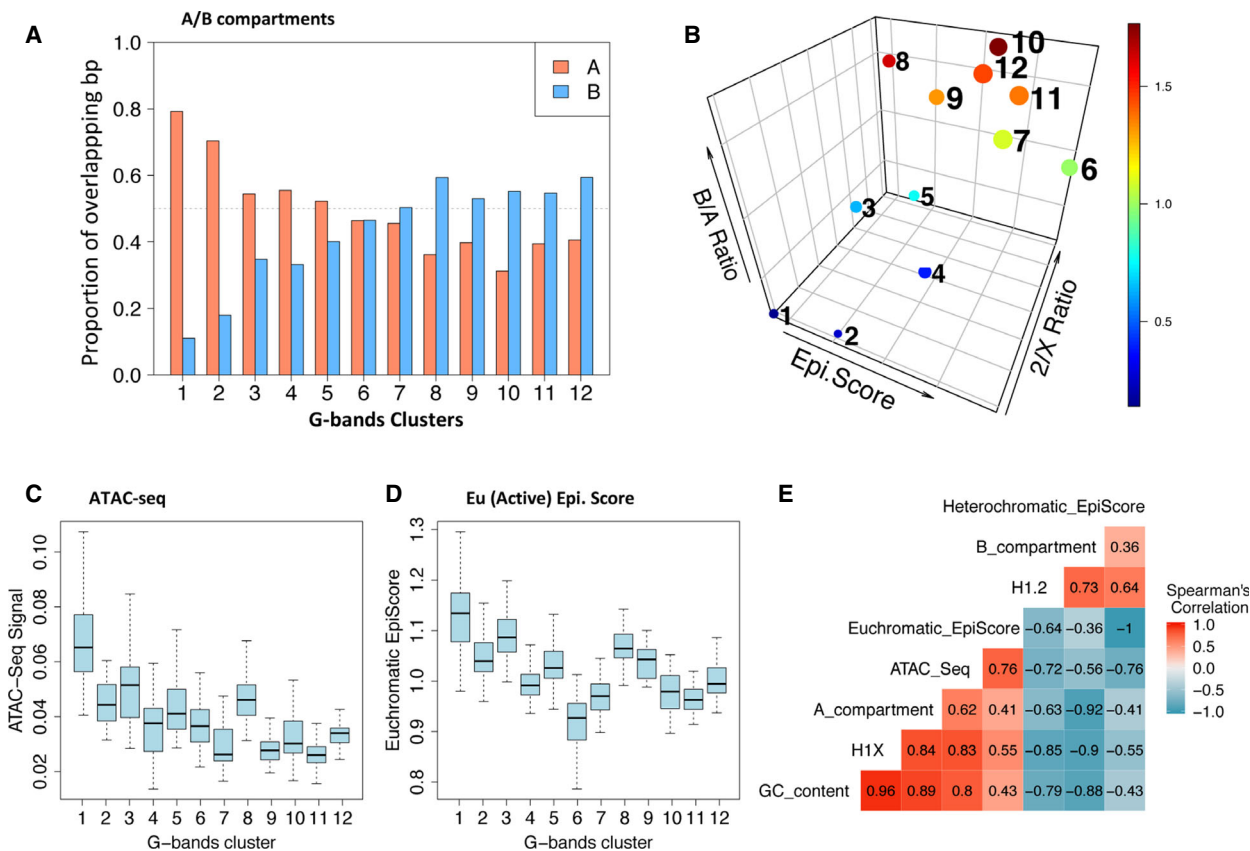
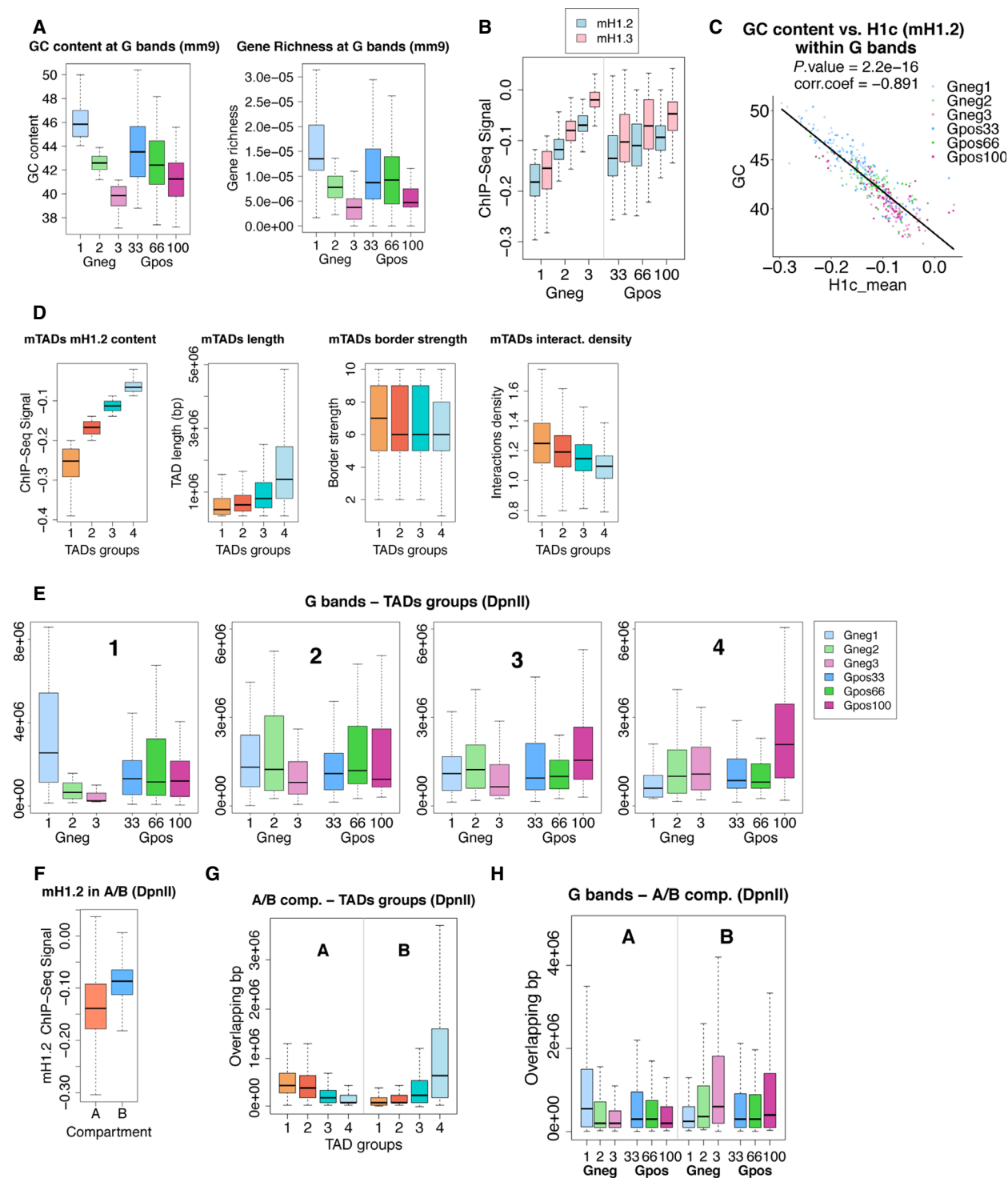


Fig. 7. Correlations between epigenetic scores, H1 variants abundance, and chromatin accessibility within G bands clusters. (A) Proportion of overlapping base pairs between A/B compartments and the G bands clustered according to histone marks, H1 variants, and chromatin proteins (Clusters 1 to 12; Fig. 4). (B) 3D plot of G band clusters according to its H1.2/H1X ratio, a calculated B/A compartments ratio, and a repressive 'epigenetic score' obtained from the ratio between the average abundance of repressive versus active histone marks or chromatin factors. Color scale refers to the B/A compartments ratio, and size of dots refers to the H1.2/H1X ratio. (C) ATAC-seq accessibility within the twelve G bands clusters. (D) Active/euchromatic epigenetic score within the twelve G bands clusters, calculated as the inverse of the repressive epigenetic score defined for B, for better comparison with ATAC-seq accessibility. (E) Correlation matrix of the different parameters studied in the 12 clusters of G bands. The graph shows the pairwise correlation coefficient between the average within the clusters of the following variables: GC content, ATAC-seq signal, euchromatic and heterochromatic epigenetic scores, H1.2 and H1X abundances, and A and B compartments overlapping.

Fig. 8. Overlap between TAD groups defined by H1.2 content, G bands, and A/B compartments from mouse ESCs. (A) Box plots showing the GC content and gene richness (gene number/base pair length) of each mouse G band type. Mouse G-positive bands are classified into three types (Gpos33 to Gpos100) according to increasing staining intensity. Gneg bands were divided into three equal groups according to GC content. (B) Box plots showing mouse Myc-H1.2 (H1c) and Flag-H1.3 (H1d) input-subtracted ChIP-seq abundance from mESCs (GSE46134) within G bands, for each band type. (C) Scatter plot of mouse H1.2 ChIP-seq abundance and GC content at each individual G band. Pearson's correlation coefficient is shown as well as *P*-value. (D) Box plots showing mouse H1.2 ChIP-seq abundance, TAD length, border strength, and interactions density of TADs (*N* = 2460) from mouse ESCs (GSE75426) divided into four groups according to their H1.2 content. (E) Box plot showing overlapping base pairs between TADs classified according to mouse H1.2 content (from low, Group 1; to high, Group 4) and the mouse G bands. (F) Box plot showing the occupancy of mouse H1.2 (input-subtracted ChIP-seq signal) within A/B compartments obtained by Hi-C in mESCs cells (GSE75426). (G) Box plot showing overlapping base pairs between TADs classified according to mouse H1.2 content (Groups 1 to 4) and the A/B compartments (*N_A* = 1367, *N_B* = 1418). (H) Box plot showing overlapping base pairs between G bands and the A/B compartments.

described correlations. Mouse G-positive bands are classified into four groups: 85 Gpos33, 44 Gpos66, 1 Gpos75, and 83 Gpos100 bands. We divided G-negative 190 bands into three equal groups according to

their GC content. Gpos and Gneg bands with the lowest GC content (Gneg3 and Gpos100) presented the lowest gene richness (Fig. 8A) and the highest abundance of mouse H1c (H1.2) and H1d (H1.3) (Fig. 8B).



H1.2 correlated negatively with GC content within G bands (Fig. 8C). Because H1X was not profiled in mESCs, we were unable to calculate the H1.2/H1X ratio. From the wild-type mESCs Hi-C data, we calculated the location of TADs and A/B compartments using the same protocol used in T47D cells. Abundance of H1s within individual TADs was calculated and four groups of TADs were generated according to the H1.2 content (Fig. 8D). TADs enriched in H1.2 were longer and presented low border strength and interactions density (Fig. 8D). TADs with the highest H1.2 content (group 4) were enriched at low GC bands, particularly Gpos100, whereas TADs with the lowest H1.2 content (group 1) were enriched at Gneg1 bands (Fig. 8E). Abundance of H1s within compartments was also calculated. H1.2 was enriched at the B compartment (Fig. 8F). Moreover, TADs with a high H1.2 content were enriched at the B compartment (Fig. 8G), and this compartment was enriched on low GC Gpos bands (Fig. 8H). Altogether, these results confirmed that the overlap between TADs enriched in histone H1.2 (among others), the B compartment, and gene-poor, AT-rich Giemsa bands is also observed in mouse ESCs and we anticipate that it might be, at least, widely extended. A remaining issue is which mammalian H1 variants accumulate at high and low GC compartments, in different cell types, to establish whether the variant preference is universal or depends on cell type or differentiation stage, or on H1 variants content. From the data available up to date, H1.2 is preferentially located at low GC, compacted or inactive regions. Whether H1X or other variants prefer high GC, active regions, extensively, needs further studies.

Discussion

It is well established that the eukaryotic genome is topologically compartmentalized inside the nucleus at several levels including chromosome territories, active and inactive compartments, TADs, and loops [34]. Initial evidences of the nonhomogeneous nature of the interphase genome came from different physico-chemical techniques that identified two major forms of chromatin, euchromatin and heterochromatin, with distinct compaction properties and location inside the nucleus, back to the 1960s. In the 1970s, several staining methods of metaphase chromosomes identified characteristic and well-conserved bands that later were associated with different features or sequences of DNA, including GC content. Here, we have combined available data on mapping of Giemsa bands and ChIP-seq data on epigenetic features with our histone H1 variants ChIP-

seq, Hi-C, and ATAC-seq data in breast cancer cells to fully characterize the overlap between genome compartments defined by these classical and state-of-the-art high-throughput methodologies. By comparing the location of G bands to chromatin accessibility maps (ATAC-seq) and the location of the A and B compartments and TADs (Hi-C) classified according to the relative abundance of different H1 variants, we have found strong correlations that support the biological relevance of these techniques to establish different compaction/activity states of the genome compartments. Besides, we demonstrate that genomic properties of compartments established in the interphase genome are in agreement with those shown by the characteristic banding of metaphase chromosomes, and vice versa. This supports the reversibility of chromosome architecture through the cell cycle, which may be sustained by the retention of architectural proteins (CTCF, cohesins) allowing the recovery of the original interphase chromatin loop structure at the end of mitosis [35].

In our previous studies, we mapped somatic H1 variants in breast cancer cells to study their specific genomic distribution. To date, specific ChIP-grade antibodies were only available for human H1.2 and H1X variants, so, for the remaining variants, HA-tagged H1 variants were overexpressed in the cells [15]. Regarding endogenous H1.2 and H1X, data uncovered some specific features for both variants. More recently, we realized that patches of enrichment of H1.2 and H1X greatly overlap with the classical chromosomal bands resulting from Giemsa staining (G bands). In this work, we have characterized G bands at several epigenetic levels to use them as genomic units to compartmentalize the genome and evaluate histone H1 variants genomic distribution (Table 1). High GC bands are enriched in active histone marks, RNA polymerase II and SINEs, and associate with gene richness, gene expression, and early replication. Low GC bands are enriched in repressive histone marks, LADs, LINEs, and late replication domains. Our results support a heterogeneous distribution of histones H1.2 and H1X within G bands that is reinforced at highly condensed chromosomes. Thus, H1.2 was found enriched in low GC bands whereas H1X was more abundant at high GC bands. From our data on HA-tagged H1 variants or elsewhere data available, we have shown that H1.0 and H1.4 are also enriched at high GC bands. Consequently, evaluating the abundance of H1 variants within G bands allows to easily compare the genomic preferences of different variants within a cell type, or to compare a variant between cell types.

We rapidly realized that both G-positive and G-negative bands were heterogeneous and not highly differentiated among them in all features investigated initially, including GC content, gene richness, replication timing, epigenetic marks, and histone H1 variants content. Gpos bands were already categorized according to staining intensity (Gpos25–Gpos100), and this was inversely correlated to GC content, gene richness, replication timing, SINEs, S/MARs, active core histone marks, transcription factors, and histone H1X. When Gneg bands were classified into four groups according to GC content, we realized they also showed the same correlations, indicating that all features obey to the GC content of regional domains of the genome (Fig. 1). These observations opened a question mark, as Gneg and Gpos bands with similar GC content and epigenetic features stained differently, at least at 850 bphs resolution, while historically it was suggested that Giemsa was staining AT-rich regions [26,27]. To solve this paradox, others suggested that the banding pattern may be related to the differences in GC content between neighboring regions [30]. We observed that Gneg and Gpos bands that were located close to each other presented similar GC contents and, upon chromosome compaction (400 bphs resolution), became stained or remained unstained more consistently with their GC content, that is, neighbor Gpos100 and Gneg4 became stained, and neighbor Gpos25 and Gneg1 did not. In other words, most of low GC Gneg bands (Gneg4) become stained at 400 bphs, while most of high GC Gpos bands (Gpos25) remain unstained. Thus, the correlation of staining with AT content is reinforced at 400 bphs compared with 850 bphs, upon chromosome compaction (Fig. 2C,D). Still, Giemsa banding cannot be explained only by the difference in base composition, especially within the Gneg bands. Instead, GC content correlates with almost every epigenetic and topological feature studied here, specially H1 variants abundance (discussed below).

One difference between Gpos and Gneg bands having a similar GC content was the average band length (Fig. 1C). Gpos100 and Gpos75 bands were longer than any Gneg band. Besides, they contained a reduced number of TAD borders within them, and those TADs with a high proportion of H1.2 were also longer than others (Fig. 5C,H). As a consequence, there was a relatively good overlap between Gpos100 bands and TADs with high H1.2 abundance. In addition, TADs within the B compartment were longer on average than TADs within the A compartment (data not shown). From all these observations, we can conclude that the domains of repressed or compacted

chromatin tend to form longer patches than active or open chromatin. Therefore, heterochromatin is less compartmentalized than euchromatin, and probably, compartmentalization (TAD borders) is needed for the proper regulation of active chromatin and gene expression occurring inside.

In general, differences between Gpos and Gneg bands with a similar GC content increased notably when topological features from the Hi-C data were analyzed. For instance, Gpos100 bands, but not Gneg4, highly overlapped with the B compartment and with TADs enriched in H1.2 (Fig. 6A,E). On the contrary, Gneg1 bands, but not Gpos25, overlapped with the A compartment and with TADs enriched in H1X (low H1.2/H1X ratio). As a consequence, Giemsa staining seems to better correspond to topological and compaction properties of genome domains.

Still, within Gneg or Gpos bands, topological features correlated to some extent with their GC content. For instance, within Gneg bands that were classified entirely based on GC content herein, their overlap with the A compartment, or with the different TAD groups based on H1.2/H1X ratio, depended greatly on GC content. Whether GC content is a prior determinant of the epigenetic and topological features of genomes, or the base pair composition of the genome has evolved as a consequence of the existence of compartments with high or low activity/accessibility, is an interesting issue that would need further debate. Assuming that low GC content is favorable for compaction, if a region is under functional constraint to maintain a compact chromatin structure, an increase in GC content would be selectively disadvantageous or an increase in AT content would be advantageous. Alternatively, GC to AT derive through evolution may occur spontaneously more often at inactive/compact regions.

We have also described that S/MARs, which in general are AT-rich sequences, are densely present in both Gpos and Gneg high GC bands. DNA molecules that are rich in AT stretches are flexible and prone to strand separation, properties needed for S/MAR functions, but these elements do not need to be immersed in AT-rich bands or domains. Apparently, S/MARs are short AT-rich stretches within GC-rich environments such as the high GC cytobands, where gene expression occurs and replication starts. S/MARs and H1X follow a similar distribution within G bands, so it would be interesting to further investigate which is the involvement of histone H1X in the function of S/MARs and, in general, in controlling gene expression and replication. We already reported that H1X is enriched at RNAPII binding sites [16]. Now, we have

found that H1X is enriched around S/MAR proteins binding sites, while H1.2 is deprived.

Another interesting observation we made was that histones H1.2 and, especially, H1X correlate with the GC content of G bands, both Gpos and Gneg, more consistently than any other epigenetic feature we investigated (i.e., core histone marks, transcription factors, etc.). This is still clearer when we generated 12 clusters of G bands according to epigenetic features including H1 variants. Upon classifying them according to the decreasing proportion of high GC bands and, consequently, decreasing GC content, H1X also decreased proportionally and H.2 increased, but the other features did not follow a clear pattern across the 12 clusters although there was a tendency. Active marks and transcription-related proteins accumulated over repressive ones at the initial clusters, and the opposite occurred toward the final clusters. This behavior may be due to the fact that histones H1 distribute uniformly along chromatin as every nucleosome may contain one linker histone and, consequently, each variant may paint a particular G band or chromosome domain uniformly according to its characteristics and GC content. Transcription factors and most of core histone marks occupy better defined positions at promoters, enhancers, coding regions, etc., and some variability may exist within a G band despite having some general behavior dictated by GC content and location within chromosome territories, among others. Obviously not all genes within a G band may be in the same state, especially because their transcriptional activity depends on the expression program of each cell type at every moment of the development or in response to diverse stimuli. Instead, the nature of G bands and even chromosome territories seems to be widely conserved across cell types.

Clustering of G bands according to epigenetic features and H1 content was a useful method to compartmentalize the genome, similar to previous initiatives based on epigenetic profiling of the genome divided in size-defined bins, resulting in defined clusters that were named the ‘colors’ of chromatin [36,37]. Here, genome segments (G bands) are much longer but the compartmentalization method proposed, although based on Giemsa staining, indirectly underlies multiple functional properties, including GC content. Further, this is the first time that H1 variants with different distribution have been used as an epigenetic feature. This method gave rise to several clusters with particular combinations of epigenetic features that might be functionally relevant and would need further investigation.

Moreover, we represented in a 3-axis diagram the characteristics of these 12 clusters based on a

repressive epigenetic score, its H1.2/H1X ratio, and a calculated compartment B/A ratio that was useful to identify clusters where the three parameters correlate, and clusters where some of the parameter deviates from the expected result, allowing to further identify and characterize particular regions of the genome. Therefore, the methods described here allow combining epigenetic data with topological information to better investigate the diversity that may be found within genome compartments. Notably, GC content, H1 variants content, and overlap with A/B compartments showed a strong correlation among clusters, whereas the epigenetic score (calculated from the abundance of histone marks and chromatin factors) presented the best correlation with ATAC-seq accessibility. This suggests that the first parameters may be related to the division of the genome in the classical euchromatin and heterochromatin compartments, and the second group of parameters may be occurring due to local changes in chromatin related to genome functions including gene expression.

Our previous studies showed that combined H1 depletion in breast cancer cells causes induction of repetitive elements, such as satellites [10]. In this last study, one of the variants depleted was H1.2 that here, we have found to be enriched in B compartment and compact TADs, characteristics presumably associated with heterochromatin. Moreover, Hi-C data in H1 triple knockout ES mouse cells revealed that reduced levels of histone H1 result in altered epigenetic and topological organization at the most active chromosomal domains [33]. Altogether, these data suggest that histone H1 levels are crucial for maintenance of the global genome topological organization, both at active and at inactive compartments. Indeed, our data show that H1.2 and H1X inversely correlate with genome topology parameters, so it is reasonable to hypothesize that altering H1 variants homeostasis could have different consequences on genome topology, in a H1-variant-dependent manner. This work supports the notion of H1 variants functional specificity, not only at the linear level but also in correspondence with the 3D genome.

We have found that H1.2/H1X ratio is closely related to G bands and genome topology. Both G banding and genome topology are expected to be highly conserved among different cell types, but this is not happening with H1 variants distribution. Several studies point to a cell type-specific distribution of H1 variants [4,11–16], so further research will be needed to elucidate if H1.2/H1X ratio correlation with G bands and topology found in breast cancer cells is maintained across cell types. If not conserved, other

H1 variants could be responsible for the mentioned correlation, in a cell type-specific manner. We believe that an extensive study of the abundance and genome distribution of all H1 variants in different cell types would be of great interest to understand H1 function and specificity in genome organization. In mouse ESCs, H1.2 and H1.3 present a similar distribution, enriched at low GC regions [11]. We have shown that TADs enriched in H1.2 are longer, present low interactions density, and correlate with the B compartment and AT-rich cytobands, indicating that the model exposed here is extensive to other cell types and species.

In conclusion, our study shows that linker histones are involved in compartmentalization of the genome. We have detected differences between H1 variants distribution within G bands, TADs, and A/B compartments that correlate with the epigenetic landscape as well as with genome sequence properties, such as GC content or the abundance of repetitive elements. Therefore, we hypothesize that H1 variants are organized according to a nonrandom clustering of the genome required to physically delineate regions with distinct functionalities.

Materials and methods

Cells culturing conditions

Breast cancer T47D-MTVL (carrying one stably integrated copy of luciferase reporter gene driven by the MMTV promoter) derivative cells were grown at 37 °C with 5% CO₂ in RPMI 1640 medium, supplemented with 10% FBS, 2 mM L-glutamine, 100 U·mL⁻¹ penicillin, and 100 µg·mL⁻¹ streptomycin, as described previously [38]. These cell lines are a model to study gene expression regulation by steroid hormones and the interplay of chromatin components and states including histone H1.

G bands characterization

Genome-wide GC content and G bands coordinates at 850 bands per haploid sequence (bphs) resolution were obtained from the UCSC human genome database. G bands average GC content was calculated with BEDTools Map to subsequently split Gneg bands into four subgroups according to their decreasing GC content. We used in-house scripts to calculate the G bands percentage of genomic occupancy as well as their average gene content, band length, gene richness, and gene expression.

LINEs, SINEs, and LADs coordinates were retrieved from the UCSC server. HeLa-S3 and T47D replication timing data, S/MARs coordinates and HIV-1 and HTLV-1 integration sites were obtained from the ENCODE,

MARome [31], and RID [39] databases, respectively. The overlapping coordinates between G bands and these regions were calculated with BEDTools Intersect and subsequently analyzed with in-house R scripts.

Since G bands coordinates at 400 bphs resolution are not available, we computed their expected starting and ending positions merging the bands at 850 bphs that give rise to each 400-bphs band according to the available ideograms (as an example, bands p24.1, p24.2, and p24.3 give rise to band p24). The properties of 400-bphs G bands, such as average GC content and H1 variants enrichment, were calculated with BEDTools as described previously for the 850 bphs bands. Next, in order to calculate the proportion of consecutive A or T nucleotides per G band, the DNA sequences of the human chromosomes were obtained from the NCBI database. We designed an R script which iterates along chromosome sequences and subtracts the fragment corresponding to each G band. It finally calculates the proportion of 1 to 5 or more consecutive A/T nucleotides at G bands as well as their total average AT content.

H1 variants ChIP-Seq analysis

Histone H1 ChIP-Seq data from T47D included in the Gene Expression Omnibus (GEO) dataset GSE49334 has been reprocessed for this study. Single-end reads were quality-checked via FASTQC v0.11.9 (S. Andrews, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and aligned to the human GRCh37/hg19 reference genome using BOWTIE2 v2.3.5.1 [40] with default options. Next, SAMTOOLS v1.9 [41] utilities were used to sort the alignments and filter out the low-quality ones with the flag 3844. Input and H1 variant genome coverage was calculated with BEDTOOLS v2.28.0 [42]. Genome coverage was normalized by reads per million and regions with zero coverage were also reported in the ChIP-Seq annotation (*genomecov -ibam -bga -scale* options). MACS2 (Model-based Analysis of ChIP-Seq) v2.1.2 [43] was used to subtract input coverage from H1 variants and to generate signal tracks (*bdgcmp -m subtract* option). We used BEDTools Map to determine the enrichment of histone H1 variants within the eight groups of G bands. ChIP signals around the center of S/MARs and HIV-1/HTLV-1 integration sites were calculated by using 'Sitepro' script of CEAS package [44] with normalized input-subtracted-average tags in 50-bp bins in a set window.

PTMs and chromatin-associated proteins analysis

We conducted our epigenetic analysis for T47D cells by downloading and reprocessing PTMs and chromatin-associated proteins raw data from the GEO database. GEO accession numbers are GSE109229 (RNAPolIII, BRD4), GSE41617 (H3K4me1, H3K4me3), GSE120162 (CTCF, H3K9ac, H3K27ac), GSE63109 (H3K4me2, H3K9me2,

H3K36me3), GSE64467 (HP1 γ), and GSE29611 (EZH2, H3K27me3). ChIP-Seq reads were processed as described [45] with minor modifications. Briefly, reads were aligned to the reference human genome (GRCh37/hg19) using BOWTIE2 v2.3.5.1 with default parameters. Mapped reads were sorted and filtered to discard the low-quality ones with SAMTOOLS. HOMER (Hypergeometric Optimization of Motif EnRichment) v4.11 [46] was used to call peaks using an input from T47D cells as a control. The ‘-style histone’ option was specified for PTMs and the ‘-style factor’ option for transcription factors and some specific histone marks which are known to develop narrow peaks (e.g., H3K4me3 or H3K9ac). The enrichment of PTMs and chromatin-associated proteins within G bands was calculated by mapping the normalized read count onto G bands with BEDTools Map.

Clustering of G bands

We designed an R script to calculate the Pearson’s correlation between H1 variants and the analyzed epigenetic factors within G bands, to establish the 12 clusters of bands and to finally characterize them. Specifically, we computed the clusters’ Gpos and Gneg bands proportion and we used the previously generated files to study the distribution of the GC content, the H1 variants, and the epigenetic factors. The packages pheatmap, ggplot2, and plot3D were used to visualize the results.

In situ Hi-C analysis

Hi-C libraries were generated from untreated derivative T47D cells as previously described [47,48]. In brief, adherent cells were cross-linked with 1% formaldehyde in PBS for 10 min at room temperature and glycine 0.125 M was added for 5 min at room temperature and for 15 min at 4 °C to stop the crosslink reaction. Before permeabilization, cells were treated for 5 min with trypsin. Nuclei digestion was performed with 400 units of MboI restriction enzyme. The ends of restriction fragments were labeled using biotinylated nucleotides and ligated with T4 DNA ligase. After reversal of crosslinks, DNA was purified and sheared (Diagenode BioruptorPico, Seraing, Belgium) to obtain 300–500 bp fragments and ligation junctions were pull down with streptavidin beads. Hi-C libraries were finally amplified, controlled for quality, and sequenced on an Illumina HiSeq 2500 sequencer (Illumina, Inc., San Diego, CA, USA).

Hi-C data preprocessing, normalization, and generation of interaction matrices

The analysis of Hi-C data, from FASTQ files mapping to genome segmentation into A/B compartments and TADs, was performed using TADBIT software [49]. TADbit pipeline starts by performing a quality control on the raw data in FASTQ format. Next, sequencing reads were mapped to the

reference genome (GRCh37/hg19) applying a fragment-based iterative strategy and using the GEM MAPPER [50]. Mapped reads were filtered to remove those resulting from unspecified ligations, errors, or experimental artifacts. Specifically, nine different filters were applied using the default parameters in TADbit: self-circles, dangling ends, errors, extra dangling ends, over-represented, too short, too long, duplicated, and random breaks [49]. Hi-C data were normalized with OneD correction [51] at the resolutions of 1 Mb, 500 kb, 100 kb, and 10 kb, to remove Hi-C biases and artifacts. Filtered read-pairs were binned at the resolutions of 1 Mb, 500 kb, 100 kb, and 10 kb, applying biases from the normalization step and decay correction to generate interaction matrices.

Hi-C data on T47D breast cancer cells have been deposited in NCBI’s Gene Expression Omnibus and are accessible through GEO Series accession number GSE147627.

Genome segmentation into topologically associating domains

We identified TADs at the resolution of 50 kb using TADbit with default parameters. TADbit segments the genome into constitutive TADs after analyzing contact distribution along the genome. TADbit employs a BIC-penalized breakpoint detection algorithm based on probabilistic interaction frequency model that returns the optimal segmentation of the chromosome [52]. This algorithm leads to a ~ 99% average genome coverage. In the output, TADbit also describes TADs border strength and TADs density. TADs border strength is the algorithm likelihood corresponding to each border (the higher the strength, the higher the algorithm confidence). TADs density represents the number of interactions within each TAD compared with the others (the higher the density, the higher the number of interactions within the TAD).

Genome segmentation into A/B compartments

We segmented the genome into A/B compartments at 100 kb resolution on *OneD*-normalized and decay-corrected matrices, using HOMER software [46]. Briefly, HOMER calculates correlation between the contact profiles of each bin against each other and performs principal component analysis (PCA) on chromosome-wide matrices. Normally, A compartment is assigned to genomic bins with positive first principal component (PC1), and B compartment is assigned to genomic bins with negative PC1.

Computing the overlap between G bands, TADs, A/B compartments, and ATAC-Seq regions

BEDTools Map was used to calculate the average H1.2 and H1X enrichment within TADs and A/B compartments while the overlapping coordinates between G bands, TADs, and A/B compartments were computed with BEDTools

Intersect. We also computed the H1.2 and H1X abundance within 100-kb bins to confirm that those located within the same TAD are more homogeneous in their H1 variants content than bins located within consecutive or alternate TADs or within similar random domains. Then, we used R to define four groups of TADs according to their increasing H1.2/H1X ratio and calculate their average length, border strength, and interactions density. We also developed a function to calculate the overlapping base pairs between two sets of intersected coordinates and therefore calculated the total overlapping nucleotides between G bands, the four groups of TADs, and the A/B compartments. This function was also used for calculating the overlapping base pairs between A/B compartments and the G bands included in each of the 12 clusters. The overlapping coordinates between the ATAC-Seq peaks and the four groups of TADs were calculated with BEDTools Intersect to subsequently compute the average number of peaks per TAD.

ATAC-Seq analysis

We reprocessed our ATAC-Seq data identified by the accession number GSE100762 as described [53] with slight modifications. Paired-end sequencing reads were quality-checked via FASTQC v0.11.9, trimmed, and subsequently aligned to the human GRCh37/hg19 reference genome using BOWTIE2 v2.3.5.1. SAMTOOLS v1.9 was used to sort and filter out the low-quality alignments with the flag 1796, remove reads mapped in the mitochondrial chromosome, and discard those reads with a MAPQ score below 30. The peak calling was performed with MACS2 v2.1.2 by specifying the *-BAMPE* mode. Filtered BAM files were also used to compute the ATAC-Seq genome coverage, which was normalized by reads per million (*bedtools genomecov -ibam -bga -scale* options). BEDTools Map was used to compute the average ATAC-Seq signal within 100-kb genomic bins as well as within G bands.

Analysis of data on mouse ESCs

mESCs GC content, G bands coordinates, and transcript annotation were obtained from the UCSC database while data on genome 3D organization and H1 variants distribution were downloaded from the GEO server. FASTQ files from Hi-C experiments performed in mESCs (GSE75426) were processed as described before for human T47D cells to compute TADs and A/B compartments coordinates. Processed input-subtracted ChIP-Seq files (GSE46134) were used to calculate the average abundance of histones H1c and H1d within G bands, TADs, and A/B compartments by using BEDTools utilities.

Acknowledgements

This work was supported by the Spanish Ministry of Science and Innovation [BFU2017-82805-C2-1-P to

AJ, BFU2017-85926-P to MAM-R (AEI/FEDER, UE)]. This research was partially funded by the European Union's Seventh Framework Programme ERC grant agreement 609989 to MAM-R, European Union's Horizon 2020 research and innovation program grant agreement 676556 to MAM-R. We also acknowledge the Generalitat de Catalunya Suport Grups de Recerca AGAUR 2017-SGR-597 to AJ and 2017-SGR-468 to MAM-R. CRG acknowledges support from 'Centro de Excelencia Severo Ochoa 2013–2017', SEV-2012-0208, and the CERCA Programme/Generalitat de Catalunya. We acknowledge Generalitat de Catalunya for an AGAUR-FI predoctoral fellowship [to MS-P and to FM].

Conflict of interest

The authors declare no conflict of interest.

Author contributions

NS-P and MS-P designed research, performed the experiments, analyzed data, and wrote the paper; NL-A, FT-L, FM, and MAM-R analyzed data; AI-B performed the experiments; AJ designed research, analyzed data, and wrote the paper; and NS-P, MS-P, MAM-R, AI-B, and AJ contributed to discussion.

References

- 1 Bednar J, Horowitz RA, Grigoryev SA, Carruthers LM, Hansen JC, Koster AJ & Woodcock CL (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.
- 2 Happel N & Doenecke D (2009) Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene* **431**, 1–12.
- 3 Izzo A, Kamieniarz K & Schneider R (2008) The histone H1 family: specific members, specific functions? *Biol Chem* **389**, 333–343.
- 4 Millán-Ariño L, Izquierdo-Bouldstridge A & Jordan A (2016) Specificities and genomic distribution of somatic mammalian histone H1 subtypes. *Biochim Biophys Acta* **1859**, 510–519.
- 5 Fyodorov DV, Zhou B-R, Skoultchi AI & Bai Y (2018) Emerging roles of linker histones in regulating chromatin structure and function. *Nat Rev Mol Cell Biol* **19**, 192–206.
- 6 Laybourn PJ & Kadonaga JT (1991) Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. *Science* **254**, 238–245.

- 7 Almeida R, Fernández-Justel JM, Santa-María C, Cadoret JC, Cano-Aroca L, Lombraña R, Herranz G, Agresti A & Gómez M (2018) Chromatin conformation regulates the coordination between DNA replication and transcription. *Nat Commun* **9**, 1590.
- 8 Bayona-Feliu A, Casas-Lamesa A, Reina O, Bernués J & Azorín F (2017) Linker histone H1 prevents R-loop accumulation and genome instability in heterochromatin. *Nat Commun* **8**, 283.
- 9 Glaich O, Leader Y, Lev Maor G & Ast G (2019) Histone H1.5 binds over splice sites in chromatin and regulates alternative splicing. *Nucleic Acids Res* **47**, 6145–6159.
- 10 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 *et al.* (2017) Histone H1 depletion triggers an interferon response in cancer cells via activation of heterochromatic repeats. *Nucleic Acids Res* **45**, 11622–11642.
- 11 Cao K, Lailier N, Zhang Y, Kumar A, Uppal K, Liu Z, Lee EK, 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.
- 12 Izzo A, Kamieniarz-Gdula K, Ramírez F, Noureen N, Kind J, Manke T, van Steensel B & 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.
- 13 Li JY, Patterson M, Mikkola HKA, Lowry WE & Kurdistani SK (2012) Dynamic distribution of linker histone H1.5 in cellular differentiation. *PLoS Genet* **8**, e1002879.
- 14 Torres CM, Biran A, Burney MJ, Patel H, Hensler-Brownhill T, Cohen AHS, 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. <https://doi.org/10.1126/science.aaf1644>.
- 15 Millán-Ariño L, Islam ABMMK, Izquierdo-Bouldstridge A, Mayor R, Terme JM, Luque N, Sancho M, López-Bigas N & 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.
- 16 Mayor R, Izquierdo-Bouldstridge A, Millán-Ariño L, Bustillos A, Sampaio C, Luque N & 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.
- 17 Solovei I, Thanisch K & Feodorova Y (2016) How to rule the nucleus: divide et impera. *Curr Opin Cell Biol* **40**, 47–59.
- 18 Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS & Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380.
- 19 Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, Van Berkum NL, Meisig J, Sedat J *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385.
- 20 Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A & Cavalli G (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472.
- 21 Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680.
- 22 Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293.
- 23 Nemeth A, Conesa A, Santoyo-Lopez J, Medina I, Montaner D, Peterfia B, Solovei I, Cremer T, Dopazo J & Langst G (2010) Initial genomics of the human nucleolus. *PLoS Genet* **6**, e1000889.
- 24 Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, 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.
- 25 Caspersson T, Lomakka G & Zech L (1971) The 24 fluorescence patterns of the human metaphase chromosomes – distinguishing characters and variability. *Hereditas* **67**, 89–102.
- 26 Comings DE (1978) Mechanisms of chromosome banding and implications for chromosome structure. *Annu Rev Genet* **12**, 25–46.
- 27 Holmquist G, Gray M, Porter T & Jordan J (1982) Characterization of Giemsa dark- and light-band DNA. *Cell* **31**, 121–129.
- 28 Furey TS & Haussler D (2003) Integration of the cytogenetic map with the draft human genome sequence. *Hum Mol Genet* **12**, 1037–1044.
- 29 Singh GB, Kramer JA & Krawetz SA (1997) Mathematical model to predict regions of chromatin attachment to the nuclear matrix. *Nucleic Acids Res* **25**, 1419–1425.
- 30 Niimura Y & Gojobori T (2002) In silico chromosome staining: Reconstruction of Giemsa bands from the whole human genome sequence. *Proc Natl Acad Sci USA* **99**, 797–802.

- 31 Narwade N, Patel S, Alam A, Chattopadhyay S, Mittal S & Kulkarni A (2019) Mapping of scaffold/matrix attachment regions in human genome: a data mining exercise. *Nucleic Acids Res* **47**, 7247–7261.
- 32 Costantini M, Clay O, Federico C, Saccone S, Auletta F & Bernardi G (2007) Human chromosomal bands: nested structure, high-definition map and molecular basis. *Chromosoma* **116**, 29–40.
- 33 Geeven G, Zhu Y, Kim BJ, Bartholdy BA, Yang SM, Macfarlan TS, Gifford WD, Pfaff SL, Verstegen MJAM, Pinto H *et al.* (2015) Local compartment changes and regulatory landscape alterations in histone H1-depleted cells. *Genome Biol* **16**: 289. <https://doi.org/10.1186/s13059-015-0857-0>.
- 34 Cavalli G & Misteli T (2013) Functional implications of genome topology. *Nat Struct Mol Biol* **20**, 290–299.
- 35 Bernardi G (2015) Chromosome architecture and genome organization. *PLoS One* **10**, e0143739.
- 36 Filion GJ, van Bommel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro IJ, Kerkhoven RM, Bussemaker HJ *et al.* (2010) Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**, 212–224.
- 37 Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330.
- 38 Sancho M, Diani E, Beato M & Jordan A (2008) Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. *PLoS Genet* **4**, e1000227.
- 39 Shao W, Shan J, Kearney MF, Wu X, Maldarelli F, Mellors JW, Luke B, Coffin JM & Hughes SH (2016) Retrovirus Integration Database (RID): a public database for retroviral insertion sites into host genomes. *Retrovirology* **13**, 47.
- 40 Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359.
- 41 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G & Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
- 42 Quinlan AR & Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842.
- 43 Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137.
- 44 Shin H, Liu T, Manrai AK & Liu SX (2009) CEAS: cis-regulatory element annotation system. *Bioinformatics* **25**, 2605–2606.
- 45 Zhang G, Zhao Y, Liu Y, Kao LP, Wang X, Skerry B & Li Z (2016) Foxa1 defines cancer cell specificity. *Sci Adv* **2**, e1501473.
- 46 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H & Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576–589.
- 47 Vara C, Paytavi-Gallart A, Cuartero Y, Le Dily F, Garcia F, Salva-Castro J, Gomez HL, Julia E, Moutinho C, Aiese Cigliano R *et al.* (2019) Three-dimensional genomic structure and cohesin occupancy correlate with transcriptional activity during spermatogenesis. *Cell Rep* **28**, 352–367.e9.
- 48 Pascual-Reguant L, Blanco E, Galan S, Le Dily F, Cuartero Y, Serra-Bardenys G, Di Carlo V, Iturbide A, Cebrià-Costa JP, Nonell L *et al.* (2018) Lamin B1 mapping reveals the existence of dynamic and functional euchromatin lamin B1 domains. *Nat Commun* **9**, 3420.
- 49 Serra F, Baù D, Goodstadt M, Castillo D, Filion G & Marti-Renom MA (2017) Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput Biol* **13**, e1005665.
- 50 Marco-Sola S, Sammeth M, Guigó R & Ribeca P (2012) The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat Methods* **9**, 1185–1188.
- 51 Vidal E, le Dily F, Quilez J, Stadhouers R, Cuartero Y, Graf T, Marti-Renom MA, Beato M & Filion GJ (2018) OneD: increasing reproducibility of Hi-C samples with abnormal karyotypes. *Nucleic Acids Res* **46**, e49.
- 52 Le Dily FL, Baù D, Pohl A, Vicent GP, Serra F, Soronellas D, Castellano G, Wright RHG, Ballare C, Filion G *et al.* (2014) Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev* **28**, 2151–2162.
- 53 Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B *et al.* (2017) An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959–962.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Correlations between H1 variants and epigenetic features within G bands.