

RESEARCH ARTICLE

Embryonic Stem Cell Specific “Master” Replication Origins at the Heart of the Loss of Pluripotency

Hanna Julienne^{1,2}, Benjamin Audit^{1,2*}, Alain Arneodo^{1,2*}

1 Université de Lyon, Lyon, France, **2** Laboratoire de Physique, CNRS UMR 5672, Ecole Normale Supérieure de Lyon, Lyon, France

* benjamin.audit@ens-lyon.fr (BA); alain.ameodo@ens-lyon.fr (AA)

Abstract

Epigenetic regulation of the replication program during mammalian cell differentiation remains poorly understood. We performed an integrative analysis of eleven genome-wide epigenetic profiles at 100 kb resolution of Mean Replication Timing (MRT) data in six human cell lines. Compared to the organization in four chromatin states shared by the five somatic cell lines, embryonic stem cell (ESC) line H1 displays (i) a gene-poor but highly dynamic chromatin state (EC4) associated to histone variant H2AZ rather than a HP1-associated heterochromatin state (C4) and (ii) a mid-S accessible chromatin state with bivalent gene marks instead of a polycomb-repressed heterochromatin state. Plastic MRT regions ($\approx 20\%$ of the genome) are predominantly localized at the borders of U-shaped timing domains. Whereas somatic-specific U-domain borders are gene-dense GC-rich regions, 31.6% of H1-specific U-domain borders are early EC4 regions enriched in pluripotency transcription factors NANOG and OCT4 despite being GC poor and gene deserts. Silencing of these ESC-specific “master” replication initiation zones during differentiation corresponds to a loss of H2AZ and an enrichment in H3K9me3 mark characteristic of late replicating C4 heterochromatin. These results shed a new light on the epigenetically regulated global chromatin reorganization that underlies the loss of pluripotency and lineage commitment.



OPEN ACCESS

Citation: Julienne H, Audit B, Arneodo A (2015) Embryonic Stem Cell Specific “Master” Replication Origins at the Heart of the Loss of Pluripotency. *PLoS Comput Biol* 11(2): e1003969. doi:10.1371/journal.pcbi.1003969

Editor: Christos A. Ouzounis, The Centre for Research and Technology Hellas, GREECE

Received: February 10, 2014

Accepted: October 6, 2014

Published: February 6, 2015

Copyright: © 2015 Julienne et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Centre National de la Recherche Scientifique (CNRS) and by the Agence Nationale de la Recherche under project REFOPOL (ANR 10 BLAN 1615). The funders had no role in study design, data collection, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Author Summary

During development, embryonic stem cell (ESC) enter a program of cell differentiation eventually leading to all the necessary differentiated cell types. Understanding the mechanisms responsible for the underlying modifications of the gene expression program is of fundamental importance, as it will likely have strong impact on the development of regenerative medicine. We show that besides some epigenetic regulation, ubiquitous master replication origins at replication timing U-domain borders shared by 6 human cell types are transcriptionally active open chromatin regions specified by a local enrichment in nucleosome free regions encoded in the DNA sequence suggesting that they have been selected

during evolution. In contrast, ESC specific master replication origins bear a unique epigenetic signature (enrichment in CTCF, H2AZ, NANOG, OCT4, . . .) likely contributing to maintain ESC chromatin in a highly dynamic and accessible state that is refractory to polycomb and HP1 heterochromatin spreading. These ESC specific master origins thus appear as key genomic regions where epigenetic control of chromatin organization is at play to maintain pluripotency of stem cell lineages and to guide lineage commitment to somatic cell types.

Introduction

One of the most remarkable phenomenon in biology is the generation of a whole organism containing a large and phenotypically diverse collection of cells and tissues from a single totipotent cell. This tremendous level of diversity in cellular functions originates from a unique genomic DNA sequence. Since the original sequencing of the human genome a decade ago [1], it has become clear that the functional role of the primary DNA sequence is not only to code for proteins which represent less than 5% of the mammalian genomes, but also to contribute to controlling the spatial structure of DNA in chromatin and in turn to regulate nuclear functions including transcription and replication [2, 3]. But as development goes on, the use of the DNA sequence has to be altered to enable lineage commitment. Epigenetic mechanisms including DNA methylation [4], histone modifications [5–13] and chromatin structure and dynamics [14–25] have been proposed to play a key role in regulation of embryonic development, the maintenance of pluripotency and self-renewal of ESCs, lineage specification and the maintenance of cellular identity during differentiation [26–30]. For years, transcriptional and chromatin changes during mammalian development have been attracting increasing interest. Among noteworthy advances, let us mention the identification of pluripotency markers including NANOG/SOX2/OCT4 [13, 31] and of trithorax proteins and polycomb complexes [32–36] as major actors in developmental gene regulation, the identification of the neural restrictive silencer factor NRSF that represses transcription of several neuronal genes in neural development [37]. Also, as differentiation progresses, chromatin structure switches from a highly dynamic, accessible and permissive euchromatin in ESCs to a less open chromatin riddled with accumulating highly condensed transcriptionally inactive heterochromatin regions [26–28, 38, 39].

In contrast to this overwhelming activity concerning the interplay between chromatin structure and transcription regulation during development, only little attention has been paid to replication and its potential role in lineage commitment and fidelity. There exist however objective reasons to believe that replication may provide some molecular handle on the study of epigenetic programming and reprogramming during development. Specific properties of the ESC cycle such as a high proliferation rate and a shortened G1 phase, that are necessary for self-renewal and the maintenance of pluripotency [29, 40], could explain differences observed between chromatin landscapes and gene expression profiles in pluripotent ESCs and in somatic cells [26–28, 38, 39]. In metazoan genomes, thousand of replication origins are prepared in G1-phase which is more than actively needed in S-phase [41–43]. Epigenetic mechanisms likely take part in the spatial and temporal control of origin usage and efficiency in relation with gene expression [43–47]. In particular, replicon size [48], which is dictated by the spacing between active origins, was shown to correlate with the length of chromatin loops [49, 50] and to be smaller in ESCs than in differentiated cells [51]. DNA replication is also an occasion to act upon the underlying primary chromatin structure at the moment of new histone incorporation

or by the spatial reorganization of pre-existent histone marks [45]. The shorter G1-phase and cell cycle duration in ESCs may thus explain the highly dynamic plastic chromatin in pluripotent cells as a lack of time for transcriptionally inactive heterochromatin regions observed in somatic cells to establish [26, 28, 38, 39]. For many years, elucidating the determinants that specify replication program in mammals has been hampered by the limited number of well established origins [43, 46, 52, 53]. The recent availability of genome-wide mean replication timing (MRT) data in various organisms [54–55], including mouse [56–58] and human [59–62], has given a new impetus to establish links between chromatin structure, transcription and replication [3, 11, 47, 53]. In pioneering studies, in mouse [57, 58] and human [59, 60], replication domains along chromosomes were delineated in constant timing regions (CTRs) of coordinated origin firings and timing transition regions (TTRs) as origin-less regions [63, 64]. In good agreement with previous studies in *Drosophila* [65, 66], these CTRs regions in different mammalian cell types revealed a correlation with epigenetic modifications [67]. Early CTRs tend to be enriched in open chromatin marks, whereas late CTRs are mostly associated with repressive HP1-associated marks [57, 68]. Actually, each cell type presents specific replication timing patterns with mouse ESCs showing a clear MRT pluripotency fingerprint [51]. Differentiation induces important changes in MRT profiles in chromosomal units of size ~400–800 kb [57, 58, 68]. Early to late (EtoL) MRT changes were associated with loss of pluripotency and largely preceded, in development, late to early (L to E) changes associated with germ-layer specific transcriptional activation [68]. Importantly, these dynamic changes in MRT come along with some sub-nuclear repositioning [57, 58, 68–70]. EtoL (resp. LtoE) transitions occur simultaneously with a movement from (resp. towards) interior of the nuclei towards (resp. from) a more peripheral location or near nucleoli [14, 16, 71, 72]. Recent experimental studies of long-range chromatin interactions using chromosome conformation capture techniques [68, 73–75] have confirmed that 3D chromatin tertiary structure plays an important role in regulating replication timing. But as questioned in a detailed analysis of replication fork polarity [76], the above dichotomic picture with early and late replicating loci occurring in separated compartments of open and closed chromatin respectively [68, 73, 74], is somehow a too simple approximation of the information contained in MRT data. The recent analysis of genome-wide MRT data in seven human cell types including ES, somatic and HeLa cells [77, 78], revealed that, in each cell type, about half of the genome can be paved by the so-called replication U-domains where the MRT is U-shaped and its derivative N-shaped like the nucleotide compositional asymmetry in the germline skew N-domains [79–83]. These peculiar N-shaped patterns were shown to be the signature of the existence of large-scale gradients of replication fork polarity [77, 84, 85] originating from transcriptionally active early initiation euchromatin zones (~200–300 kb) separated by megabase size genome distances [77, 86–89]. These “master” replication origins at U/N-domain borders were further shown to be long-range interconnected hubs of chromatin interactions delineating topological domains of self-interacting chromatin [75, 77, 90]. Here, our aim is to show that, these “master” replication origins at U/N-domain borders [77, 86–91] are a possible clue to the understanding of the plasticity of the spatio-temporal replication program, gene expression and chromatin organization across different cell lines during development, lineage commitment and fidelity.

In a previous study [92], with the aim at quantifying the influence of epigenetic modifications on the spatio-temporal replication program, we used principal component analysis [93] and clustering method [94] to analyze thirteen epigenetic mark maps in the K562 human cell line at the 100-kb resolution of MRT data. This study revealed that the huge combinatorial epigenetic complexity could in fact be reduced to a rather small number of prevalent chromatin states that interestingly shared strong similarities with the ones previously found in *Arabidopsis thaliana* [95], *Caenorhabditis elegans* [96] and *Drosophila* [66, 97]. These four main chromatin

states were further shown to replicate at distinct periods of the S-phase, namely from early to late replicating, a gene rich transcriptionally active euchromatin state (C1) enriched in insulator binding protein CTCF, a polycomb repressed facultative heterochromatin state (C2), a silent heterochromatin state (C3) not enriched in any available marks and a gene-poor HP1-associated heterochromatin state (C4). When mapping these chromatin states inside the corresponding megabase-sized U/N-domains, we found that as the signature of an increasing firing frequency during S-phase [98], the accelerating replication wave [76] actually proceeds along a directional path through the four chromatin states, from the open euchromatin state C1 at the “master” replication origins bordering U/N-domains, successively followed by the three silent chromatin states C2, C3 and C4 at U/N-domain centers [92]. The complete analysis of the other half of the human genome that is complementary to U/N-domains [92, 99] turned out to be more consistent with the above mentioned dichotomic picture proposed in pioneering studies of the mouse [56–58] and human [60, 68, 74] genomes. About 25% of the human genome are covered by megabase-sized GC-rich (C1+C2) chromatin blocks that on average replicate early by multiple rather synchronous randomly positioned origins with almost equal proportions of forks coming from both directions which explains that the skew has not accumulated in these gene-rich, high GC isochore-like regions devoid of skew N-domains [79–81]. The last 25% of the human genome corresponds to megabase-sized gene deserts, low GC isochore-like regions of extended (C3+C4) heterochromatin states or long C4 domains that on average replicate late by again multiple almost coordinated origins [92, 99]. Here we extend this study to different cell types including the ESC H1hesc, three hematopoietic cell lines (K562, Gm1278, Monocyte CD14+), a mammary epithelial cell line (Hmec) and an adult fibroblast cell line (Nhdfad). By investigating the global reorganization of replication U/N-domains in these different cell types in relation to coordinated changes in chromatin state and gene expression, we shed a new light on the chromatin-mediated epigenetic regulation of transcription and replication during human differentiation. Because they are likely to be the cornerstone to better understanding of pluripotency maintenance, developmental specification and lineage fidelity, we will pay special attention to the “master” replication initiation zones that border U/N-domains and specially to those that are specific to ESCs.

Materials and Methods

Histone marks, H2AZ, CTCF, CHD1, NANOG and OCT4 ChIP-seq data

ChIP-seq data were retrieved for the following cell lines: an ESC line (H1hesc), an immature myeloid cell line (K562), a monocytes-CD14+ (monocd14ro1746), a lymphoblastoid cell line (Gm12878), a mammary epithelial cell line (Hmec), an adult dermal fibroblast cell line (Nhdfad).

For all ChIP-seq data, we downloaded data in the Encode standard format “broadpeaks” (<http://genome.ucsc.edu/FAQ/FAQformat.html>). Broadpeaks format is a table of significantly enriched genomic intervals. The signal value associated with each enriched intervals is the fold enrichment compared to a uniform distribution of reads [100]. For all cell types, we downloaded the broadpeak tables for the following antibodies: CTCF, H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K9me3, H2AZ, H3K79me2, H4K20me1. For the H1hesc cell line, we downloaded these additional broadpeak genomic intervals: H3K9ac, CHD1, EZH2, NANOG and OCT4.

Most of the data correspond to the release 3 (August 2012) of the Broad histone track, downloaded from: <http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone/> The NANOG and OCT4 data corresponds to the release 3 (September 2012) of the HAIB TFBS track. Tables were downloaded from: <http://hgdownload.cse.ucsc>.

[edu/goldenPath/hg19/encodeDCC/wgEncodeHaibTfbs/](http://www.encodeproject.org/goldenPath/hg19/encodeDCC/wgEncodeHaibTfbs/) All these tables in hg19 coordinates were converted to hg18 coordinates using LiftOver.

Note that cell line H1hesc (part of ENCODE Tier 1) is the only embryonic stem cell line for which we could gather a large chromatin mark dataset as described above. Hence, we could not include in the present work the analysis of another ESC line to assess whether the unique properties observed in H1hesc are also valid for other ESC lines.

Epigenetic profile computation at 100 kb resolution

For each ChIP-seq data and each cell line, we computed a profile at the 100 kb resolution for the 28465 non-overlapping windows corresponding to the sequenced part of the genome. For an antibody, the score in a 100kb window was computed as the sum of the coverage of each significantly enriched interval multiplied by its score; it is thus a read density.

Construction of a shared epigenetic space for differentiated cell lines

For the five differentiated cell lines (K562, Monocd14ro1746, Gm12878, Hmec and Nhdaf), we constructed a shared epigenetic space. All epigenetic profiles at 100kb of the same mark were concatenated together to obtain one vector of $5 \times 28405 = 170970$ windows per mark.

Treatment of H1hesc data set

We took into account the specificity of H1hesc cell line epigenetics by applying the clustering pipeline described in [92] apart from other cell lines but considering the same eleven epigenetic marks as for differentiated cell lines. The number of clusters was set to four because it led to the most qualitatively different chromatin states.

Rank transformation and Spearman correlation matrix

All statistical computations were performed using the R software (<http://www.r-project.org/>).

In order to compute the Spearman correlation matrix, the epigenetic profiles at 100 kb resolution were transformed with the R function *rank* with option *ties.method = max*. Then we computed the Pearson correlation matrix on the transformed data set. To reorder the matrix (Fig. 1 and S1 Fig.), we computed the Spearman correlation distance *dSCor* as:

$$dSCor(X, Y) = 1 - SCor(X, Y), \quad (1)$$

where *SCor* is the Spearman correlation. Then, a dendrogram was computed using the R function *hclust* with option *method = average* and with *dSCor* as dissimilarity (Fig. 1). Note that for H1hes cell line, we performed an additional correlation analysis taking into account H3K9ac, CHD1, EZH2, NANOG and OCT4 ChIP-seq data (Fig. 1, top panel).

Principal component analysis

Principal component analysis was performed on the rank transformed dataset using the function *dudi.pca* from the R package *ade4* (see <http://pbil.univ-lyon1.fr/ADE-4> and [101]) with the option *scale = TRUE* (i.e. each variable was centered and normalized before the PCA computation). The first four components were retained which accounts for 86% of the dataset variance (S2A and A' Fig.). Clustering was performed in this 4D space.

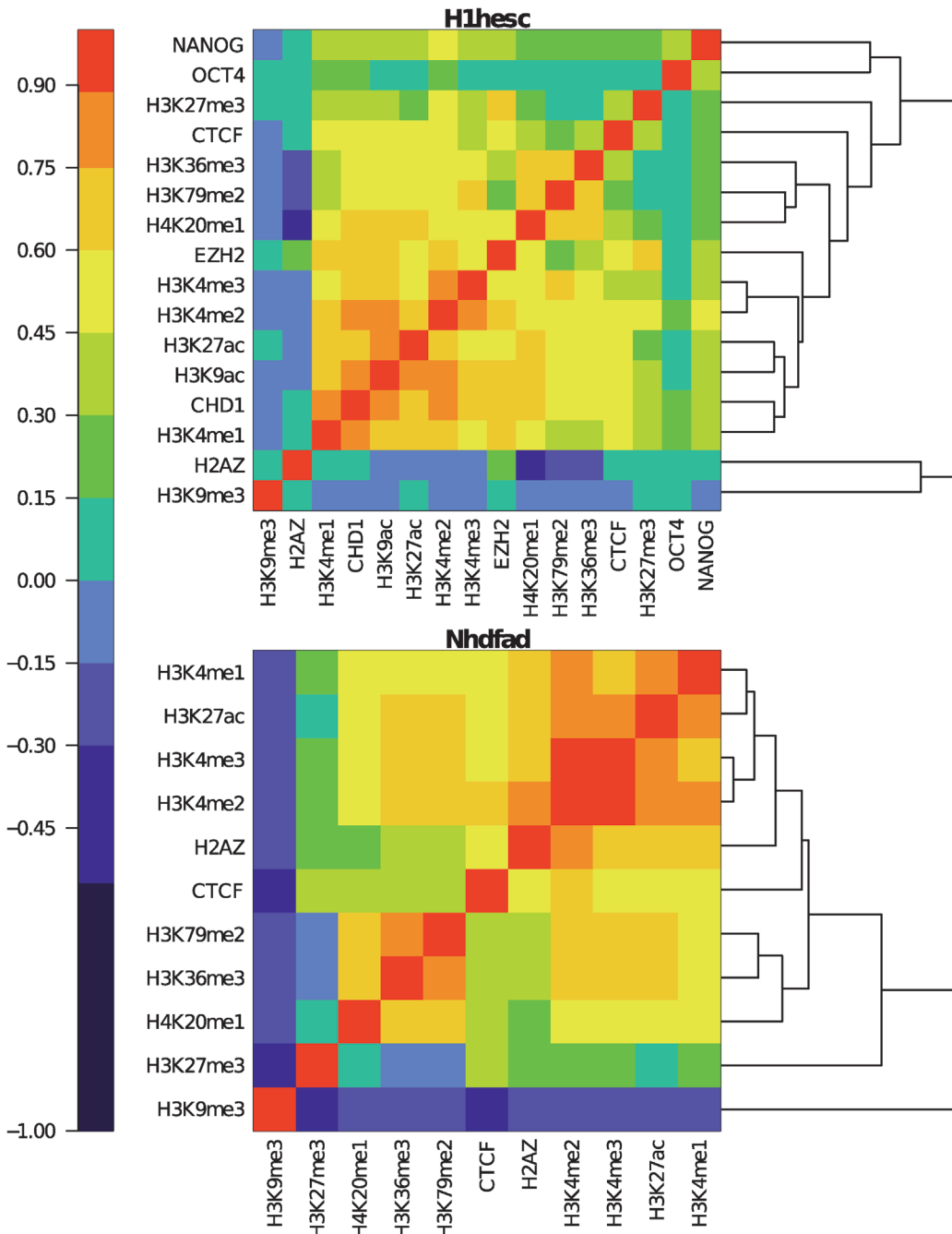


Figure 1. Spearman correlation matrix between epigenetic marks in H1hes (top) and Nhdad (bottom). For each cell line, we computed the Spearman correlation over all 100 kb non-overlapping windows with a valid score. Spearman correlation value is color coded using the color map shown on the left. Lines for the epigenetic marks were reorganized by a hierarchical ordering using Spearman correlation distances [Equation (1)] as illustrated by the dendrograms on the right of the corresponding matrices. This ordering implies that highly correlated epigenetic marks are close to each other.

doi:10.1371/journal.pcbi.1003969.g001

Clustering strategy

We used Clara algorithm [94] which is an optimization of k-means for large data set. We used the *clara* function implemented in the R package *cluster*. The options were set to: *stand* = *FALSE*, *sampsize* = 500, *samples* = 20, *metric* = *euclidean*.

For the shared differentiated cells, the number of clusters was set to the number of prevalent chromatin states detected in [92]. Previously to the merging of dataset into one shared epigenetic space, we checked that, when applied to each cell individually, the analysis pipeline led to qualitatively the same epigenetic states (data not shown).

Poorly clustered data points were removed from the set of chromatin states. The silhouette value [102] is a way to quantify how well a point is clustered.

Definition 1 Given a particular clustering, C_1, C_2, \dots, C_k , of the data in k clusters, let i be a data point and $d(i, C_j)$ the average distance of the data point i to the members of the cluster C_j . Let i be a member of cluster C_l and

$$a_i = d(i, C_l), \quad b_i = \min_{j \neq l} (d(i, C_j)). \quad (2)$$

The silhouette value of the data point i is defined as:

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}. \quad (3)$$

A silhouette value below 0 means that the data point is actually closer in average to the points from another cluster than to the ones it has been assigned to. Points with a negative silhouette value are questionable assignments. We decided to remove those points from the set of identified chromatin states. Hence chromatin states are groups (clusters) with homogeneous epigenetic features. 91% (resp. 94%) of all 100 kb non-overlapping windows of the human genome were assigned to one of the four chromatin states C1, C2, C3 or C4 (resp. EC1, EC2, EC3 and EC4) in the differentiated (resp. H1hesc) cell lines.

Mean replication timing data and replication U-domain coordinates

Mean replication timing (MRT) determined in 100 kb non-overlapping windows in hg18 coordinates for an ESC line (BG02), a lymphoblastoid cell line (GM06990), a skin fibroblast cell line (BJ), an immature myeloid cell line K562 and a HeLa cell line were obtained from the authors [77]. The coordinates of the 1534 (BG02), 882 (GM06990), 1150 (BJ), 876 (K562) and 1498 (HeLa) replication U-domains were also obtained from these authors. Replication data in BG02, GM06990 and BJ were used as surrogates of replication data in H1hesc, GM12878 and Nhdad, respectively. We had previously observed a very strong conservation of MRT between lymphoblastoid cell lines including GM06990 [75], as well as between fibroblast cell lines BJ and IMR90 (unpublished). Note that, due to a lack of data, we do not have evidence of such a conservation between ESC lines.

DNase Hypersensitive site data

DNaseI hypersensitive sites (DHSs) data were downloaded in the Encode standard format “narrowpeaks” (<http://genome.ucsc.edu/FAQ/FAQformat.html>). DHS narrowpeaks are genomic intervals identified as hypersensitive zones to DNaseI within a false discovery rate of 0.5%. We downloaded the tables from: <http://hgdownload.cse.ucsc.edu/goldenPath/hg18/encodeDCC/wgEncodeUwDnaseSeq/>

Annotation and Expression data

As human gene coordinates, we used the UCSC Known Genes table. When several genes presenting the same orientation overlapped, they were merged into one gene whose coordinates correspond to the union of all the overlapping gene coordinates, resulting in 23818 distinct genes.

Expression data were retrieved from the Genome Browser of the University of California Santa Cruz (UCSC). We downloaded expression values from the release 2 of Caltech RNA-seq track (ENCODE project at UCSC): <http://hgdownload.cse.ucsc.edu/goldenPath/hg18/encodeDCC/wgEncodeCaltechRnaSeq/>

To determine the genome coordinates of each gene (labeled by its RefSeq identifier), we used RefSeq Genes track. For genes associated to more than one splicing variant, we merged exons coordinates by taking their union. Hence the transcription start site (TSS) was placed at the beginning of the first exon. We obtained a table of 23329 genes.

Expression for one gene is given in reads per kilobase of exon model per million mapped reads (RPKM) [103]. RPKM is defined as:

$$R = \frac{10^9 C}{NL}, \quad (4)$$

where C is the number of mappable reads that fall into gene exons (union of exons for genes with alternative splicing), N is the total number of mappable reads in the experiment, and L is the total length of the exons in base pairs. We associated 17872 genes with a valid RPKM value in K562 and Gm12878 and 17463 in H1hesc.

CpG o/e computation and GC content

CpG observed/expected ratio (CpG o/e) was computed as $\frac{n_{CpG}}{L-l} \times \frac{l^2}{n_C n_G}$, where n_C , n_G and n_{CpG} are the numbers of C, G and dinucleotides CG, respectively, counted along the sequence, L is the number of nonmasked nucleotides and l is the number of masked nucleotide gaps plus one, *i.e.* $L-l$ is the number of dinucleotide sites. The CpG o/e was computed over the sequence after masking annotated CGIs.

The GC content was computed on the native sequence.

Nucleosome free regions (NFR)

The coordinates of the NFRs predicted by the physical model defined in [104–107] were obtained from the authors [108]. This theoretical model amounts to compute the energy required for nucleosome formation based on sequence-dependent bending properties [3].

Chromatin state blocks

We detected contiguous windows of the same chromatin state (C1 to C4 and EC1 to EC4). We then kept the coordinates of the blocks of contiguous windows. To form chromatin state blocks of states (1+2), we simply detected contiguous windows of state 1 or 2. The same procedure was applied to define chromatin blocks of states (3+4). For chromatin blocks (1+2) and (3+4), we authorized the inclusion of isolated windows which did not belong to any chromatin state so to not disrupt very long blocks.

Replication N-domains

The coordinates of the 678 human replication N-domains for assembly NCBI35/hg17 were obtained from the authors [81] and mapped using LiftOver to hg18 coordinates; we kept only the 663 N-domains that had the same size after conversion [77].

Index of conservation for U-domain borders

To identify MRT U-domain borders which are common to several cell lines, we constructed a counting signal and we attributed a conservation index as follows:

1. We created a merged data set of the coordinates of all U-domain borders detected in [77] and of skew N-domain borders detected in [81]. U-domains were detected in the following cell line (BG02, K562, GM06990, H0287, TL010, BJ, HeLa). GM06990, H0287, TL010 are three lymphoblastoid cell lines. To avoid lymphoblastoid cell specific U-domains getting an artificially high conservation index, we took only GM06690 into account. To avoid MRT to be a confounding factor, we excluded late U-domain borders with $MRT > 0.5$.
2. Then, we slid a 200 kb window along the genome with 10 kb incremental steps. At each position, we retrieved the number of cell lines that have a domain border in the window. By doing so, we constructed the counting signal called the conservation index. For instance, if a U-domain border of K562 has a conservation index of 3, it means that this border together with 2 domain borders from other cell lines are contained within a common 200 kb window.

Results

Combinatorial analysis of chromatin marks

We investigated relationships between the genome-wide distributions of nine histone modifications, one histone variant and one binding protein at 100 kb resolution in five sommatic cell types including an immature myeloid cell line (K562), a monocyte cell line (Monocd14ro1746), a lymphoblastoid cell line (Gm12878), a mammary epithelial cell line (Hmec), an adult dermal fibroblast cell line (Nhdfad) and an ESC line (H1hesc). As a first step, we computed the Spearman correlation coefficient of each mark with each other (Materials and Methods). We next represented the resulting matrix as a heatmap after having reorganized rows and columns with a hierarchical clustering based on the Spearman distance [Equation (1)] (Fig. 1 and S1 Fig.). This analysis was very enlightening since, on the one hand it revealed that the correlation matrices obtained for the five sommatic cell lines strongly resemble to the one obtained in K562 in our previous study [92] (Fig. 1, bottom panel and S1 Fig.), and on the other hand it clearly discriminated the pluripotent H1hesc cell line for having a different correlation structure between epigenetic marks (Fig. 1).

In the epigenetic mark matrices obtained for the differentiated cell lines Nhdfad (Fig. 1, bottom panel), Hmec, Monocd14ro1746, K562 and Gm12878 (S1 Fig.), all histone modifications that are known to be involved in transcription positive regulation, namely H3K4me1, H3K4me2, H3K4me3, H3K27ac, H3K36me3, H3k79me2 and H4K20me1, form a block that also includes the histone variant H2AZ and the transcription factor CTCF, meaning that all these marks are all correlated with each other and are likely to occupy similar regions in the genome [6, 12]. In fact, two lines are clearly apart in all correlation matrices as illustrated on the hierarchical clustering dendrogram (Fig. 1, bottom panel). They correspond to the repressive chromatin marks H3K27me3 and H3K9me3 that are respectively associated with the so-called facultative and constitutive heterochromatins [92]. These two marks are recognized by the chromodomains of polycomb (Pc) proteins and heterochromatin protein 1 (HP1) respectively, components of distinct gene silencing mechanisms which may explain that they are anti-correlated with each other. While H3K9me3 behaves quite independently if not anticorrelated with most of the active chromatin marks (except for Gm12878 where some positive correlations were observed), H3K27me3 correlates to some of them in a cell line dependent fashion but more systematically to CTCF and H4K20me1 (Fig. 1, bottom panel and S1 Fig.). This consistency of epigenetic mark correlations in the five differentiated cell lines prompted us to build a

“shared” epigenetic space ([Materials and Methods](#)). This consisted in pooling data points of all differentiated cell lines together and then in applying PCA and clustering algorithm to reduce the dimensionality of the data. We used the first four principal components along which the data display some meaningful pattern emerging from the noisy background and which together account for 86% of the total data set variance ([S2A’ Fig.](#)). As previously experienced with K562 [[92](#)], we fixed the number of clusters to four in Clara algorithm [[94](#)] ([Materials and Methods](#)). When labeling each of the four main chromatin states with a color, we obtained four domains in the (PC1, PC2, PC3, PC4) space that have common boundaries as illustrated on the (PC1, PC2) projection plane ([S2B’ and C’ Fig.](#)). To improve the quality of our clustering procedure, we filtered out poorly clustered data points that were closer to another cluster than the one they belong to and had a negative silhouette [[102](#)] ([Materials and Methods](#)). Note that the classification obtained for K562 in our “shared” epigenetic space is 74% identical to the one previously reported for K562 alone [[92](#)]. A similar quantitative identity holds for the four other somatic cell types (data not shown).

The correlation matrix obtained for the same 11 epigenetic mark profiles of the pluripotent H1hesc cell line ([Fig. 1](#), top panel) displays important differences from the ones previously obtained for differentiated cell lines. Among others, let us mention the repressive polycomb-associated mark H3K27me3 which now strongly correlates with most of the active marks including H3K4me3 as the probable signature of bivalent ESC chromatin [[6](#), [9](#), [26](#), [28](#), [32](#), [33](#)]. Also the histone variant H2AZ that now correlates as much with both the repressive marks H3K27me3 and H3K9me3 as with some of the active marks, which is likely an indication of specific highly dynamic and accessible chromatin of pluripotent cells [[6](#), [9](#), [11](#), [26](#), [28](#), [39](#)]. When reproducing our PCA and clustering analysis on the H1hesc epigenetic data, we again found that four PCs were enough to account for 86% of the total variance ([S2A Fig.](#)), and that one could still reduce the ESC epigenetic complexity to four chromatin states ([S2B and C Fig.](#)) but, as described in the next sub-section, these chromatin states are distinct from the ones delineated in somatic cells confirming that ESCs and differentiated cells have different epigenomes [[5](#), [6](#), [9](#), [11](#), [26](#), [39](#)]. Note that we have confirmed the conclusions of our correlation matrix analysis ([Fig. 1](#), top panel) when including in our study the distributions of the ATP-dependent helicase CHD1, the EZH2 subunit of polycomb repressive complex 2 (PRC2) and the two pluripotency transcription factors NANOG and OCT4.

Epigenetic content of prevalent chromatin states in ESCs versus differentiated cells

The four chromatin states so identified in the five differentiated cell lines are quite similar to the ones previously found in K562 [[92](#)] (see also [[28](#)]). C1 is a transcriptionally active chromatin state enriched in the histone modifications H3K27ac, H3K4me1, H3K4me3, H3K36me3 ([Fig. 2](#)) and H3K4me2, H3K27me2, H4K20me1 ([S3 Fig.](#)), as well as in the histone variant H2AZ whose binding level was shown to correlate with gene activity in human [[6](#)] ([Fig. 2](#)). C2 is notably associated with the histone modification H3K27me3 ([Fig. 2](#)) and hence corresponds to a polycomb repressed chromatin state [[6](#), [109](#)]. C3 can be compared to the “null” or “black” silent heterochromatin regions devoid of chromatin marks previously found in *Drosophila* [[66](#), [97](#)] and *Arabidopsis* [[95](#)]. C4 corresponds to the HP1-associated heterochromatin state with all C4 100 kb-loci containing H3K9me3 and almost only that repressive mark ([Fig. 2](#)) [[6](#), [109](#)]. Note that CTCF which is known to establish chromatin boundaries to prevent the spreading of heterochromatin into transcriptionally active regions [[110](#), [111](#)] was found in C1 and to a slightly less extend in C2 loci ([Fig. 2](#)).

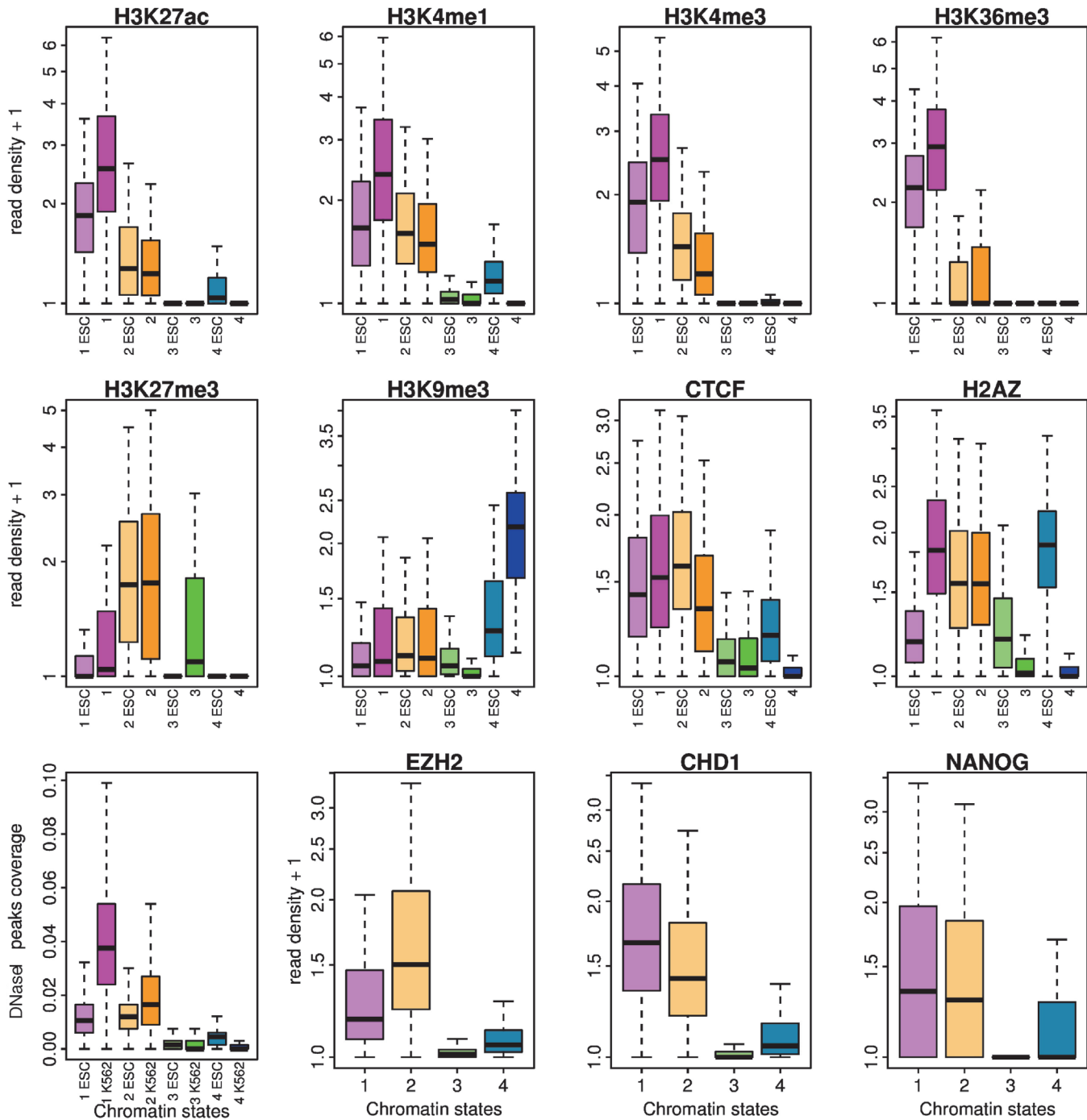


Figure 2. (First two rows) Repartition of epigenetic marks in the four prevalent chromatin states of H1hesec cell line: EC1 (light pink), EC2 (light orange), EC3 (light green), EC4 (light blue) and of differentiated cell lines: C1 (pink), C2 (orange), C3 (green), C4 (blue). Boxplots of the decimal logarithm of histone mark ChIP-seq read density in 100 kb non-overlapping windows per chromatin state. (Third row) Boxplots of the coverage of DNaseI hypersensitive peaks in 100 kb non-overlapping windows per chromatin state in H1hesec and K562 cell lines and the decimal logarithm of EZH2, CHD1 and NANOG ChIP-seq read density in 100 kb non-overlapping windows per chromatin state in H1hesec cell lines. Same color coding as above.

doi:10.1371/journal.pcbi.1003969.g002

Chromatin states in pluripotent H1hesc cell line (EC1, EC2, EC3, EC4) are different even though they display some similarities with the above described differentiated chromatin states (C1, C2, C3, C4). As for differentiated C1 state but to a slightly less extent, more than 75% of 100kb-loci in EC1 state contain all the active histone modification marks considered (Fig. 2, S3 Fig.). 91% of 100 kb loci in EC2 like in C2 are marked by H3K27me3 which is deposited by polycomb complex PRC2 and then enhances PRC1 targeting [35, 36, 112] (Fig. 2). Consistently, EZH2, which is a subunit of PRC2 containing a SET domain that acts upon H3K27 as a methyltransferase, was abundantly found in EC2 confirming the polycomb activity of this state (Fig. 2). CTCF is also present in both EC1 and EC2 as previously seen in C1 and C2 but in slightly reverse importance, EC2 being more enriched than C2 and vice versa for EC1 and C1 (Fig. 2). C1, C2 and EC1, EC2 being the most genic chromatin states in differentiated and ESCs, this result is coherent with the correlation observed between CTCF positioning and gene density [113]. H4K20me1 which was recently shown to strongly correlate with gene activation [6], was consistently found in EC1 and C1 but more surprisingly also in EC2 and C2 which are silent chromatin states (S3 Fig.). Interestingly, PR-Set 7 that is involved in the deposition of H4K20me1, was recently shown to play an important role in the control of replication origin firing in mammalian cells [114].

However the epigenetic chromatin states in pluripotent ESCs and differentiated cells bear more differences than similarities. Systematically the differentiated C1 state is more enriched in active histone marks than the pluripotent EC1 state, and this for all histone modifications but H4K20me1 (Fig. 2 and S3 Fig.). Relatively to EC1, EC2 contains more H3K4me3 than C2 relatively to C1 (Fig. 2), which, with the enrichment of EC2 in H3K27me3, is an indication of bivalent heterochromatin. But the most striking difference concerns the pluripotent state EC4 whose epigenetic content is qualitatively and quantitatively different from the one of C4. Noticeably, H2AZ is highly present in more than 99% of EC4 100kb loci which contrasts with its scarcity in C4 (Fig. 2). As compared to C4 which is enriched in the HP1-associated heterochromatin mark H3K9me3, EC4 contains significantly less H3K9me3 concomitantly with an enrichment in H2AZ (Fig. 2). As recently observed in human [28], the enrichment of the ESCs in the histone variant H2AZ associated with nucleosome exchange and remodeling [6, 24, 115, 116] is likely to contribute to the highly dynamic properties of pluripotent chromatin and its refractory character to both HP1- and polycomb heterochromatin extension [6, 11, 26, 28, 38]. This interpretation is strengthened by the observation that in contrast to C4, EC4 is enriched in CTCF (Fig. 2), which besides its insulator properties [110, 111], is also known to mediate long-range intra- and inter-chromosomal interactions [110, 113, 117–120]. Thus, the accessible and more relaxed EC4 chromatin might be more central in the nucleus than the HP1-associated heterochromatin C4 state that likely corresponds to the emergence of compact chromatin at the nuclear periphery [14–26, 39].

To get a better comprehension of ESC chromatin states, we looked at two additional epigenetic marks known for their implication in pluripotency. Globally all chromatin remodelers are over expressed in ESC [121] but only some knockdown are known to impair pluripotency. The ATP-dependent helicase CHD1 is one of these [122]. In mouse [122, 123], CHD1 helps to maintain a globally more loose chromatin in ESCs. Interestingly, CHD1 is present in EC1 and EC2 (Fig. 2) which makes sense since both these chromatin states contain most of the human genes (discussed below) whose expression can possibly be altered by CHD1 in pluripotent cells [122]. But CHD1 is also present in 84% of EC4 100 kb loci indicating that this remodeler likely contributes to prevent HP1-associated constitutive C4 heterochromatin spreading and compaction [122]. The pluripotent OCT4/SOX2/NANOG network enables self-renewal properties of ESCs, and ectopic expression of these factors together with additional factors or mechanisms was shown to reprogram somatic cells into pluripotent cells (iPS cells) [124, 125]. NANOG was found to be the same extent in EC1 and EC2 (Fig. 2) which is consistent with the fact that

NANOG regulates roughly the same number of expressed genes and silent genes [31, 126]. NANOG is surprisingly present in the gene-poor EC4 state suggesting that it may play a role in promoting the relative openness of this pluripotent chromatin state.

Chromatin state coverages and chromatin state changes between cell lines

When comparing the genome coverages, *i.e.* the percentages of the 28465 100 kb non-overlapping windows corresponding to the sequenced part of the human genome that belong to the previously identified prevalent chromatin states, we found that whatever the considered cell line, less than 16% of these windows were not properly classified in any chromatin state (Table 1). In H1hesc cells, EC1 and EC2 coverages are about the same (~20%) and are quite similar to the C1 and C2 coverages (~15–23%) generally observed in the five differentiated cells. If the EC3 (27%) and EC4 (24%) coverages are comparable in the ESCs, the C3 and C4 coverages in the differentiated cells are much more variable from 12% to 36%, with a total (C3+C4) coverage ~45% (Table 1).

To study changes in chromatin states between different cell lines, among all possible pairs of cell lines (S4 Fig.), we focus on two representative transitions from ES to somatic cell lines and from somatic to somatic cell lines. The changes obtained from H1hesc chromatin states to Nhdad chromatin states (Fig. 3A) reveals that the transcriptionally active state is highly conserved: 80% of EC1 100 kb-loci in H1hesc are C1 loci in Nhdad as compared to 13% that experience a repression by polycomb to C2 and only 4% and 3% that transit towards the heterochromatin states C3 and C4 respectively. The bivalent state EC2 directs towards either the active euchromatin state C1 (29%) or the polycomb repressed state C2 (51%) which is coherent with initial bivalency adding flexibility in transcription regulation during development [5–12, 33, 127]. The unmarked state EC3 mainly leads to the heterochromatin states C3 (30%) and C4 (51%) and almost never to the active state C1 (5%). EC4 does not transit much to the active state C1 (7%) but distributes almost equally into C2 (34%), C3 (21%) and C4 (39%). Even though they are quite different in terms of epigenetic marks (Fig. 2, S3 Fig.), these three states are silent (Table 2) [92, 99]. Therefore EC4 state in pluripotent cells appears prepared to silencing during differentiation. Now when looking at chromatin state changes from differentiated cell lines K562 to Nhdad (Fig. 3B), we observed that a majority of 100 kb loci in C1 (73%), C2 (55%), C4 (66%) and to a lesser extent C3 (40%) are conserved. Indeed, a noticeable difference is that the constitutive heterochromatin state C4 rarely transits to the active euchromatin state C1 (4%) and to the polycomb repressed state C2 (12%), which confirms that the pluripotent state EC4, if prepared

Table 1. Proportions of 28465 sequenced 100-kb windows that belong to the EC1, EC2, EC3 and EC4 chromatin states in H1hesc and to the C1, C2, C3 and C4 chromatin states in differentiated cell lines.

Chromatin states	EC1	EC2	EC3	EC4	ED
H1hesc	0.21	0.19	0.27	0.24	0.09
Chromatin states	C1	C2	C3	C4	D
K562	0.21	0.14	0.28	0.26	0.11
Monocd14ro1746	0.23	0.18	0.13	0.30	0.16
Gm12878	0.21	0.15	0.36	0.12	0.16
Hmec	0.22	0.21	0.25	0.16	0.16
Nhdad	0.21	0.22	0.19	0.25	0.13

ED and D correspond to 100-kb windows that were not classified in any chromatin states (Materials and Methods).

doi:10.1371/journal.pcbi.1003969.t001

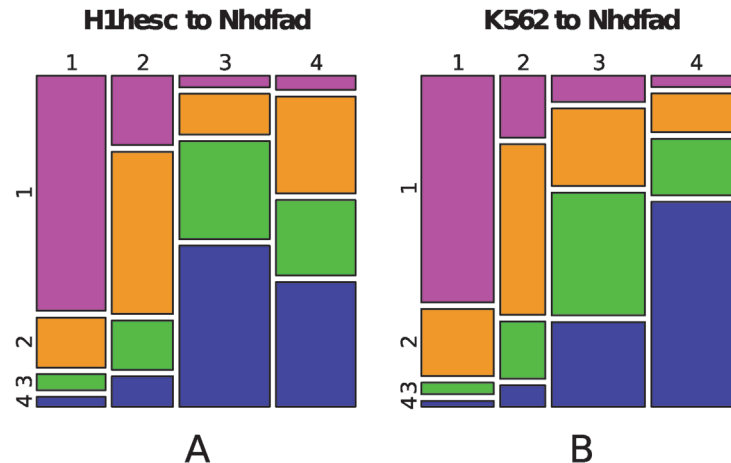


Figure 3. (A) Mosaic plot representing the probabilities of transition from H1hesec chromatin states to Nhdfad chromatin states. The width of columns corresponds to the proportion of chromatin states in H1hesec. The segmentation for the i^{th} column follows the proportion of windows in state EC_i in H1hesec that become C_j in Nhdfad. In other words, if we take the first pink rectangle of the first column, its width is proportional to the probability for a 100 kb window to be in chromatin state EC_1 in H1hesec and its height is proportional to the probability for a 100 kb window to be in C_1 in Nhdfad given that it is in EC_1 in H1hesec. The area of this rectangle (product of the previously mentioned probability) is proportional to the probability for a window to be in state EC_1 in ESC and C_1 in Nhdfad. (B) Same as (A) for the chromatin state changes from the cell line K562 towards Nhdfad.

doi:10.1371/journal.pcbi.1003969.g003

to silencing, is not as C_4 , a compactly repressed heterochromatin state. Note that overall, chromatin states are highly dynamic since only 48% (resp. 57%) of 100 kb loci are conserved from H1hesec (resp. K562) to Nhdfad. Merging the genic chromatin states EC_1+EC_2 (resp. C_1+C_2) significantly increases the conservation rate to 83% (resp. 69%). The merging of EC_3+EC_4 (resp. C_3+C_4) also displays high conservation rate 74% (resp. 89%).

Replication timing of chromatin states

Consistent with our preliminary analysis of the K562 cell line [92], we confirmed that there exists a strong correlation between the four prevalent chromatin states and the MRT, and this for both the pluripotent (H1hesec) and the differentiated (K562, Gm12878, Nhdfad) cell lines (Fig. 4). The transcriptionally active euchromatin states C_1 and EC_1 replicate early in S-phase in agreement with previous studies of open chromatin marks in human and mouse [57, 60, 62, 67, 68, 128]. The pluripotent bivalent EC_2 state and the differentiated polycomb repressed C_2

Table 2. Gene content in the four prevalent chromatin states of H1hesec and K562 cell lines.

Chromatin states	EC1	EC2	EC3	EC4	C1	C2	C3	C4
promoter count	7653	5047	1138	1505	9588	3089	2334	827
promoter density per Mb	13.127	9.278	1.497	2.16	15.861	7.812	2.953	1.133
mean gene expression (RPKM)	30.537	7.76	0.798	1.299	18.529	2.307	0.696	0.244
median gene expression (RPKM)	8.527	0.988	0.008	0.025	5.58	0.29	0.005	0
mean gene length (kb)	50.999	72.705	56.335	67.538	43.557	58.535	86.301	173.962

For each chromatin state the following information is given: (i) the total number of promoters per chromatin state, (ii) the density of promoters per Mb, (iii) the mean level of expression per chromatin state in RPKM (Materials and Methods), (iv) the median level of expression per chromatin state in RPKM, (v) the mean gene length per chromatin state in kb.

doi:10.1371/journal.pcbi.1003969.t002

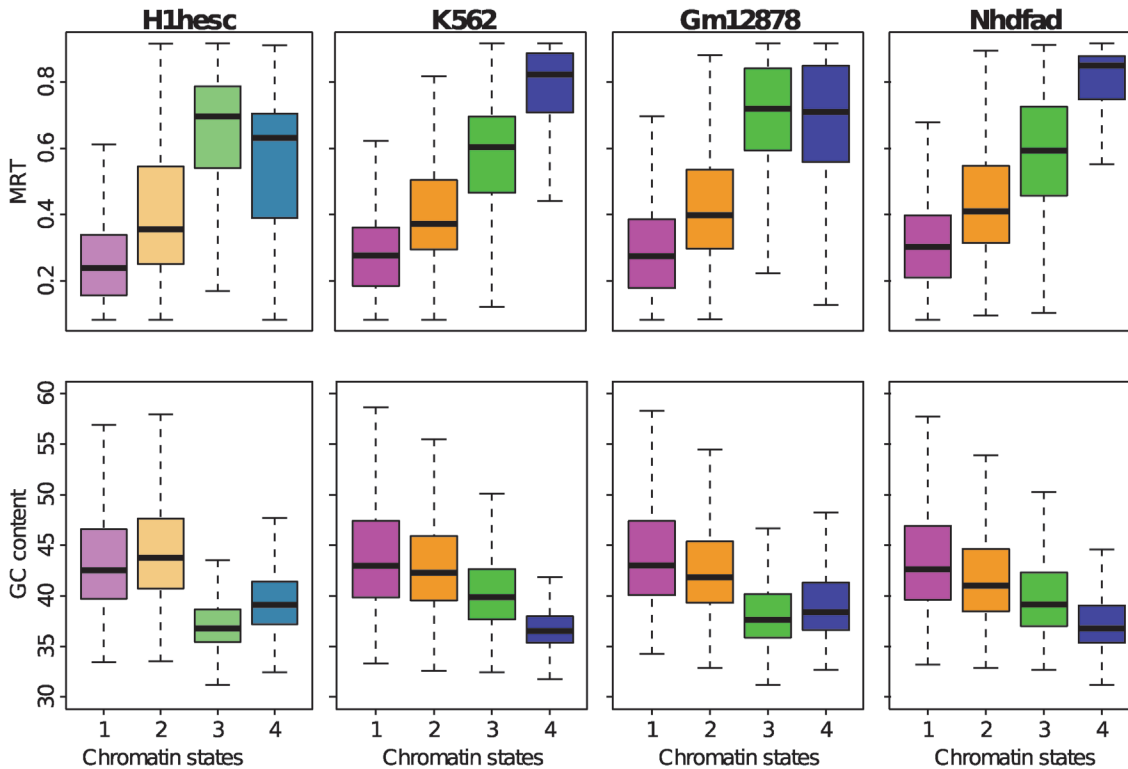


Figure 4. MRT and GC distributions in the four chromatin states for H1hesc and three differentiated cell lines (K562, Gm12878, Nhd fad). (First row) Boxplot of MRT computed in 100 kb non-overlapping windows per chromatin state. (Second row) Boxplots of GC content computed in 100 kb non-overlapping windows per chromatin state. Same color coding as in Fig. 2.

doi:10.1371/journal.pcbi.1003969.g004

heterochromatin state both replicate slightly later in mid-S phase which contrasts with previous report of the existence of high correlation between late replication and the repressive chromatin mark H3K27me3 [68, 129]. The silenced unmarked EC3 and C3 states as well as the pluripotent chromatin state EC4 prepared to heterochromatinization and the HP1-associated heterochromatin state C4 all replicate much latter up to the end of S-phase. Interestingly, whereas (EC1, C1) and (EC2, C2) have clear different MRT, they have almost the same high mean GC content as expected for gene-rich states [1]. In contrast, a clear correlation between MRT and mean GC content was observed for the late replicating chromatin states. When C3 replicates before C4 (K562, Nhd fad), C3 has a higher GC content than C4 and vice-versa when C3 replicates after C4 (H1hesc, Gm12878) (Fig. 4). There is however a major difference between MRT of pluripotent and differentiated cell lines. EC4 exhibits a much wider MRT distribution than C4 with a non-negligible proportion of early replicating (MRT < 0.5) 100-kb loci, namely 35.7% (H1hesc) as compared to 5.5% (K562), 19.2% (Gm12878) and 4.2% (Nhd fad). This can be seen as an additional indication that EC4 is sufficiently accessible and open to enable origin firing and early replication. This is confirmed by non negligible coverage by DNaseI hypersensitive sites (DHS) in H1hesc EC4 (DHS coverage median at 0.45%) as in EC1 (1.1%) and EC2 (1.2%), which contrasts with the abundance of DHS in differentiated C1 (3.8%) and C2 (1.7%) states and their virtual absence in the heterochromatin states C3 (0%) and C4 (0%) (Fig. 2). The highly dynamic and accessible character of pluripotent chromatin states likely facilitates the access of the replication machinery to DNA and thus prevents having to replicate long (EC3+EC4) threads at the end of S-phase.

Gene content of chromatin states

To address the issue of gene content of pluripotent and differentiated prevalent chromatin states, we focused on H1hesc and K562 cell lines. We took advantage of our previous detailed integrative analysis of epigenetic marks, MRT and gene expression data in K562 [92, 99] that showed that the euchromatin state C1 is highly genic and contains almost all expressed genes and a non negligible proportion of inactive genes that almost equals the total number of genes found in C2 as mostly repressed by polycomb complexes. As compared to these high-GC (Fig. 4), gene rich C1 and C2 states, the low-GC C3 and C4 states were found to be gene deserts with scarce long genes. In the pluripotent H1hesc cell line, the gene rich chromatin states are still EC1 and EC2. But there are some noticeable differences with respect to K562. There are less promoters per Mb in EC1 (13.1 promoters/Mb) than in C1 (15.9 promoters/Mb), and in compensation more in EC2 (9.3 promoters/Mb) than in C2 (7.8 promoters/Mb) (Table 2). Moreover the relative distributions of RPKM values [Equation (4)] (S5A and B Fig.) revealed that relative to C1, EC1 contains more expressed genes with RPKM > 1 as well as EC2 relative to C2. Indeed, both mean and median RPKM values are higher in EC1 and EC2 than in C1 and C2 respectively (Table 2). This is consistent with the extensive presence of bivalent genes in EC2 that was previously shown to be more accessible and less compact than the polycomb repressed C2 state in differentiated cell lines [28, 127]. This is also in agreement with previous report on the higher global transcription activity in ESCs with only sporadic tissue-specific gene expression as compared to differentiated cells [130]. Note that, in that respect, EC4 is slightly permissive to expression whereas C4 is the most repressive heterochromatin state (only 25% of genes with a non-null RPKM) with by far the lowest gene density and largest gene mean length (Table 2).

The coupling between MRT and gene expression has been extensively studied in *Drosophila* [55, 65] and mammals [56, 57, 59, 60]. We found that in both H1hesc and K562, a vast majority of expressed genes are in the early replicating EC1 and C1 chromatin states which confirms the link between MRT and expressed gene density previously reported in mouse [56–58] and human [59, 60, 62, 92]. Even more, in [99] we showed that the activation of one gene in K562 was almost always sufficient for its 100 kb environment to be in an early C1 chromatin state. But the presence of an important number of inactive genes in early C1 regions and to a less extent in early EC1 regions (Table 2, S5A and B Fig.), suggests that there is no causal link between an early replicating region and a high expression level yet many recently identified early replication origins are strongly associated with CGI and active CpG-rich gene promoters [43, 47, 131–139]. If almost all genes in the late replicating heterochromatic C3 and C4 states are silent with few exceptions, there is a slightly larger number of expressed genes in the pluripotent EC4 state (25% of the few genes in EC4 100kb windows have a non null RPKM). Recent studies in mammals have further shown that the dynamic of MRT through differentiation is only loosely coupled with gene expression dynamic [57, 58, 60, 65].

Spatial organization of chromatin states along human chromosomes

Once mapped on the genome (Fig. 5), the organization of the four prevalent chromatin states looks different in the pluripotent H1hesc cell line as compared to the one in the five differentiated cell lines (Table 3). In H1hesc, the four chromatin states EC1, EC2, EC3 and EC4 do not differ so much in their genome coverage (Table 1). Moreover, the blocks of adjacent 100-kb-loci in the same chromatin state have similar length distributions (Table 3, S6A Fig.). In Nhdad, in agreement with previous analysis in K562 [92], the HP1-associated heterochromatin state C4 has a block length distribution that displays a fat tail not observed in the C1, C2 and C3 block length distributions (S7A Fig.) nor in the corresponding H1hesc block length

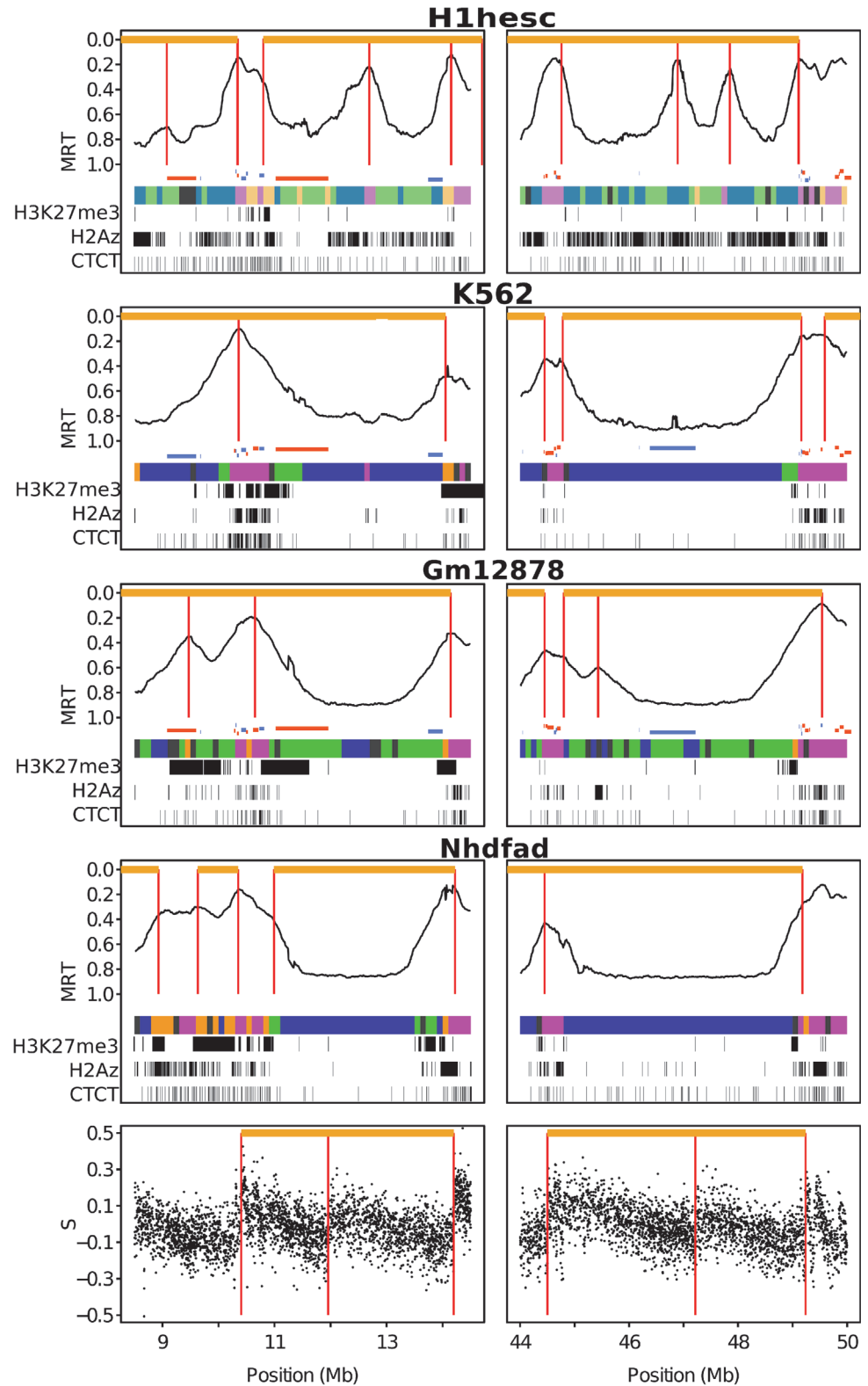


Figure 5. Genome-wide spatial distribution of chromatin states in ESCs and differentiated cells. MRT profile along two Mb long fragments of human chromosome 5 (left) and 14 (right). U-domains are marked by an horizontal orange line and their borders by vertical red lines. Below the MRT profile, gene positions are indicated by a horizontal segment (blue: not expressed, orange: expressed) as well as the chromatin state of

each 100 kb window is represented using the same color coding as in Fig. 2. At the bottom of the plot, intervals significantly enriched in H3K27me3, H2AZ and CTCF are represented in black. At the bottom of the figure, the last panel represents the compositional skew $S = S_{GC} + S_{TA} = \frac{G-C}{G+C} + \frac{T-A}{T+A}$ [79–82] in 1 kb windows of repeat masked sequences; germline replication skew N-domains are marked by an horizontal orange line and their borders by a vertical red line.

doi:10.1371/journal.pcbi.1003969.g005

distributions (S6A Fig.). This fat tail explains that the mean C4 block lengths ($\bar{L} = 894$ kb) is significantly larger than the mean block length of C1 ($\bar{L} = 312$ kb), C2 ($\bar{L} = 262$ kb) and C3 ($\bar{L} = 281$ kb) (Table 3). This peculiar length property of C4 blocks is shared by all differentiated cell lines except Gm12878 where C3 blocks are larger ($\bar{L} = 576$ kb) as compared to C4 blocks ($\bar{L} = 276$ kb). Interestingly, as originally observed in K562 [92], for all differentiated cell lines as well as for the ESC line H1hesc, the association of C1+C2 (resp. EC1+EC2) on one side and C3+C4 (resp. EC3+EC4) on the other side, results in large scale blocks of surprisingly similar length distributions (Table 3, S6B and S7B Figs.). The block length distributions obtained for differentiated cells have a fat tail up to blocks larger than 10 Mb (S7B Fig. and also Fig. 9 in [92]). These very long

C1+C2 blocks actually replicate very early (S7C Fig.). Within these long blocks, C2 loci replicate later than C1 loci suggesting that C2 loci are replicated passively from fork coming from neighboring C1 active loci. On the contrary, long C3+C4 blocks replicate very late (S7D Fig.) as expected for gene desert low-GC heterochromatin regions. These results are quite consistent with the statistical model proposed in [60] where MRT is predicted from the distance to the nearest active promoter. In H1hesc, the long EC1+EC2 (resp. EC3+EC4) blocks also correspond to early (reps. late) replicating regions (S6C and D Figs.). Interestingly, their maximal length (~5Mb) is significantly shorter than in differentiated cells (~12 Mb), which might be related to a shorter cell cycle in ESC.

Mean replication timing dynamics during differentiation

Consistently with previous studies in mouse [57, 58], most 100 kb loci ($\geq 80\%$) do not present significant MRT change ($|\Delta MRT| < 0.2$) when comparing two cell types (Fig. 6A). Whereas there are as many positive (EtoL) as negative (LtoE) MRT changes between two somatic cell types, there is an excess of EtoL transitions from H1hesc to somatic cell types (Fig. 6A). The MRT conservation level of C1+C2 (EC1+EC2) chromatin blocks (Fig. 6B) and of C3+C4 (EC3+EC4) chromatin blocks (Fig. 6C) is clearly larger for the largest blocks. For the former, the conservation level increases with the size of the chromatin block indicating that mostly small C1+C2 (EC1+EC2) fragments a few hundred kb long switch MRT from one cell type to another. In contrast, MRT conservation level for small C3+C4 (EC3+EC4) presents a minimum around ~600 kb. This means that not only the largest but also the very small blocks have a

Table 3. Mean length of chromatin state blocks per chromatin state in kb (Materials and Methods) in ES H1hesc and differentiated cells (see Table 1).

Chromatin state	EC1	EC2	EC3	EC4	EC1+EC2	EC3+EC4
H1hesc	296.14	241.89	398.56	300.81	569.34	740.2
Chromatin state	C1	C2	C3	C4	C1+C2	C3+C4
K562	327.43	191.14	437.69	881.56	567.19	869.6
Monocd14ro1746	357.48	215.22	210.33	582.15	642.93	628.52
Gm12878	336.87	198.34	576.08	276.3	551.83	702.88
Hmec	322.34	231.96	357.57	610.08	608.46	563.51
Nhdfad	312.11	261.89	281.14	894.38	675.45	637.58

doi:10.1371/journal.pcbi.1003969.t003

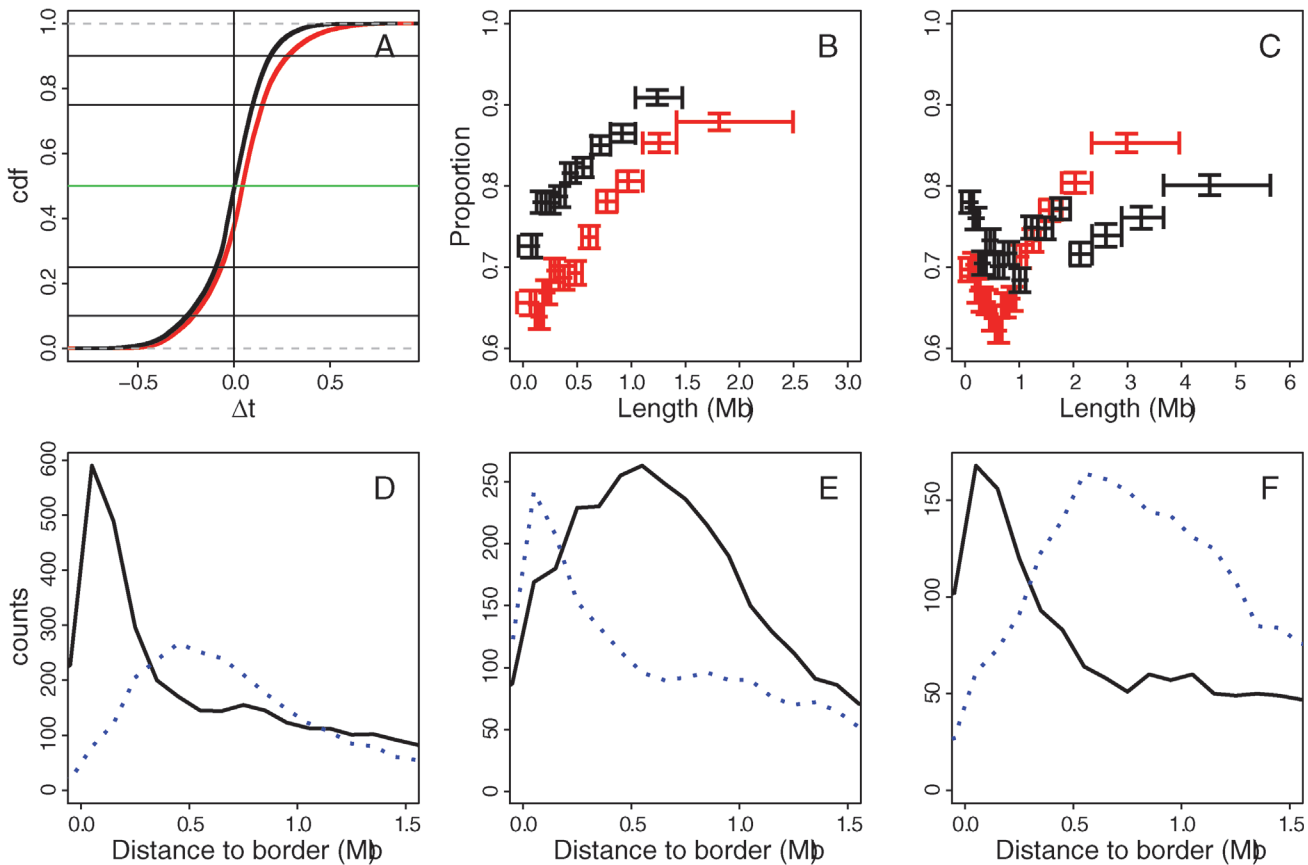


Figure 6. (A) Cumulative distribution function (cdf) of MRT difference Δt between two cell lines: $\Delta t = \text{MRT}(\text{Nhd fad}) - \text{MRT}(\text{H1hesc})$ (red) and $\Delta t = \text{MRT}(\text{Nhd fad}) - \text{MRT}(\text{K562})$ (black). Horizontal lines mark from bottom to top the first decile, first quartile, median, last quartile and last decile, respectively. (B) Proportion of 100 kb windows in C1+C2 blocks that have a conserved MRT ($|\Delta t| < 0.2$) with respect to block length: H1hesc blocks with MRT compared to Nhd fad (red) and K562 blocks with MRT compared to Nhd fad timing (black). (C) Same as (B) for C3+C4 blocks. (D) Histogram of distances to H1hesc U-domain borders for loci that have a $\Delta t = \text{MRT}(\text{Nhd fad}) - \text{MRT}(\text{H1hesc}) > 0.2$ (line) and < -0.2 (dots). (E) Histogram of distances to Nhd fad U-domain borders for loci that have a $\Delta t = \text{MRT}(\text{Nhd fad}) - \text{MRT}(\text{H1hesc}) > 0.2$ (line) and < -0.2 (dots). (F) Histogram of distances to K562 U-domain borders for loci that have a $\Delta t = \text{MRT}(\text{Nhd fad}) - \text{MRT}(\text{K562}) > 0.2$ (line) and < -0.2 (dots).

doi:10.1371/journal.pcbi.1003969.g006

robust MRT between cell types. These small C3+C4 (EC3+EC4) blocks have a similar conservation level as their surrounding C1+C2 (EC1+EC2) blocks that likely contribute to stabilize their MRT via passive replication [140, 141].

Altogether, the largest chromatin state blocks correspond to mega-base sized MRT domains that are well conserved between pluripotent and differentiated cell lines (Fig. 6B and C). These constant MRT regions cover about half of the human genome in agreement with the dichotomic view proposed in early studies of the mouse [56–58] and human [60, 68, 74] genomes, where early and late replicating regions occur in separated compartments of open and close chromatin, respectively. About 25% of the human genome are covered by megabase sized gene-rich, high-GC EC1+EC2 (resp. C1+C2) chromatin blocks in H1hesc (resp. differentiated) cells, that on average replicate early (S6C and S7C Figs.). Since the replication fork polarity is reflected in the MRT derivative [77, 84], each locus along these MRT plateaus is replicated by an equal proportion of forks coming from both directions originating from multiple early firing origins. Similarly, about 25% of the genome are covered by megabase sized gene-poor, low-GC

EC3+EC4 (resp. C3+C4) chromatin blocks in H1hesc (resp. differentiated) cells, that on average replicate late (S6D and S7D Figs.) by again multiple almost coordinated origins. Note that despite the difference in chromatin properties of the silent states EC4 (dynamically accessible) and C4 (compact heterochromatin), these late MRT plateaus present a similar MRT conservation level as the early EC1+EC2 (resp. C1+C2) plateaus.

In all the cell lines examined in this work, the other half of the human genome complementary to the mega-base sized early and late MRT domains was shown to be paved by U-shaped MRT domains (Fig. 5) [77]. However their number (N) and mean length (\bar{L}) drastically differ in H1hesc (N = 1534, \bar{L} = 1.09 Mb) and in the differentiated cell lines K562 (N = 876, \bar{L} = 1.42 Mb), Gm12878 (N = 882, \bar{L} = 1.52 Mb) and Nhdfad (N = 1150, \bar{L} = 1.19 Mb). MRT U-domains are more numerous and shorter in the ESC line than in the differentiated cell lines. The corresponding excess of early replicating U-domain borders in H1hesc possibly underlies the excess of EtoL transitions observed from H1hesc to somatic cell types (Fig. 6A). Interestingly, if these MRT U-domains are a robust feature of the spatio-temporal replication program in human, they indeed correspond to the most dynamical regions of the genome during differentiation for MRT changes [77]. The 100 kb loci that present an EtoL MRT switch from H1hesc to Nhdfad are preferentially located (resp. depleted) within ± 300 kb of replication U-domain borders in H1hesc (resp. Nhdfad) (Fig. 6D and E). MRT plasticity is thus concomitant with the disappearance of MRT U-domain borders consistently with the replication consolidation scenario previously reported in mouse [57, 58]. An active early replication initiation zone in ESCs that no longer fires early results in the merging of two neighboring MRT U-domains in H1hesc into a larger MRT U-domain in Nhdfad. In contrast, the situation is the opposite for the 100 kb loci that present an LtoE MRT switch from H1hesc to Nhdfad that actually corresponds to the breaking of one H1hesc U-domain into two Nhdfad U-domains via the appearance of an early initiation region (Fig. 6D and E). Note that a similar U-domain border dynamic is also observed between somatic cell types (Fig. 6F).

Chromatin state organization inside replication U-domains

Replication U-domains were detected as regions bordered by two early replicating regions having a U-shaped MRT profile [77, 78], so that these domains capture the spatial coherence of MRT distribution along the genome. In this respect, we reasoned that mapping the organization of the four prevalent chromatin states within replication U-domains can provide complementary (but not independent) information on the genomic organization of chromatin states and on the modifications of this organization during cell differentiation, compared to the results described above about the spatial distribution and the replication timing of chromatin states.

When concentrating our study on the replication U-domains identified in H1hesc (Fig. 7A and B) and Nhdfad as a representative of differentiated cell lines (Fig. 7A' and B'), we revealed some remarkable organization of the four prevalent chromatin states with some notable differences that distinguish the global dynamical and accessible character of pluripotent chromatin from the expanding HP1-associated heterochromatin in differentiated cells. Consistent with the organization found in K562 [92], the highly expressed gene-rich open euchromatin state C1 was found to be confined in a closed (≈ 150 kb) neighborhood of the master replication origins that border each individual U-domains (Fig. 7A') and this independently of the domain size [92] (data not shown). Significantly enriched in DHS and CTCF (Figs. 2 and 8B), C1 can thus be seen as specifying the early initiation zones that border U-domains and that were further shown to delimit topological domains on genome-wide (4C, Hi-C) chromatin state conformation data [75, 77, 90]. The polycomb repressed state C2 was mainly found occupying the

mid-S phase 200–300 kb region away from U-domain borders (see also Fig. 9 in [92]). Remarkably, U-domain borders are significantly depleted in unmarked (C3) and constitutive (C4) heterochromatin states (Fig. 7A). C3 homogeneously occupies large U-domain centers. C4 is abundantly found in the center of large U-domains (≥ 1 Mb). These results for Nhdfad and K562 [92] suggest that a replication “wave” starting from the early initiation zones at U-domain borders and propagating inside these domains via the progressive activation of secondary origins [76, 89], actually progress in a gradient of chromatin structures from openness (C1) to compactness (C3, C4), via the polycomb repressed state C2 [92].

In the smaller H1hesc U-domains, the concentration of EC1 around the bordering master replication initiation zones and the distribution of EC2 nearby in mid-S phase proximal regions (Fig. 7A) resembles to the organization of high-GC, gene-rich chromatin states (C1, C2) in differentiated cells (Fig. 7A). However the distributions of EC3 and EC4 (Fig. 7A) are different from those of C3 and C4 in Nhdfad (Fig. 7A') and K562 [92]. EC3 is still depleted at U-domain borders and mainly covers the center of the largest U-domains. Importantly, unlike C4, EC4 is now found at many U-domain borders as well as inside these domains. As addressed in the “Discussion”, this homogeneous distribution of the gene-poor silent EC4 state inside

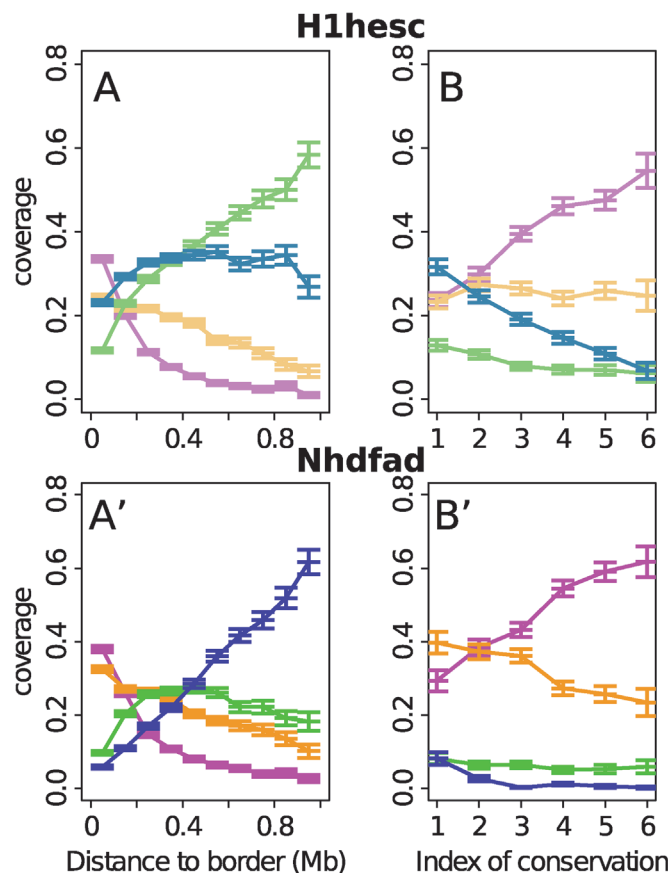


Figure 7. Distribution of chromatin states inside replication timing U-domains of H1hesc and Nhdfad. (A) Mean coverage of chromatin states with respect to the distance to the closest U-domain border in the H1hesc cell line. (B) Mean coverage of ESC chromatin state in the 100kb window containing a U-domain border with respect to the conservation index n of the U-domain border (Materials and Methods). (A', B') same as (A, B) for the Nhdfad cell line. Same color coding as in Fig. 2.

doi:10.1371/journal.pcbi.1003969.g007

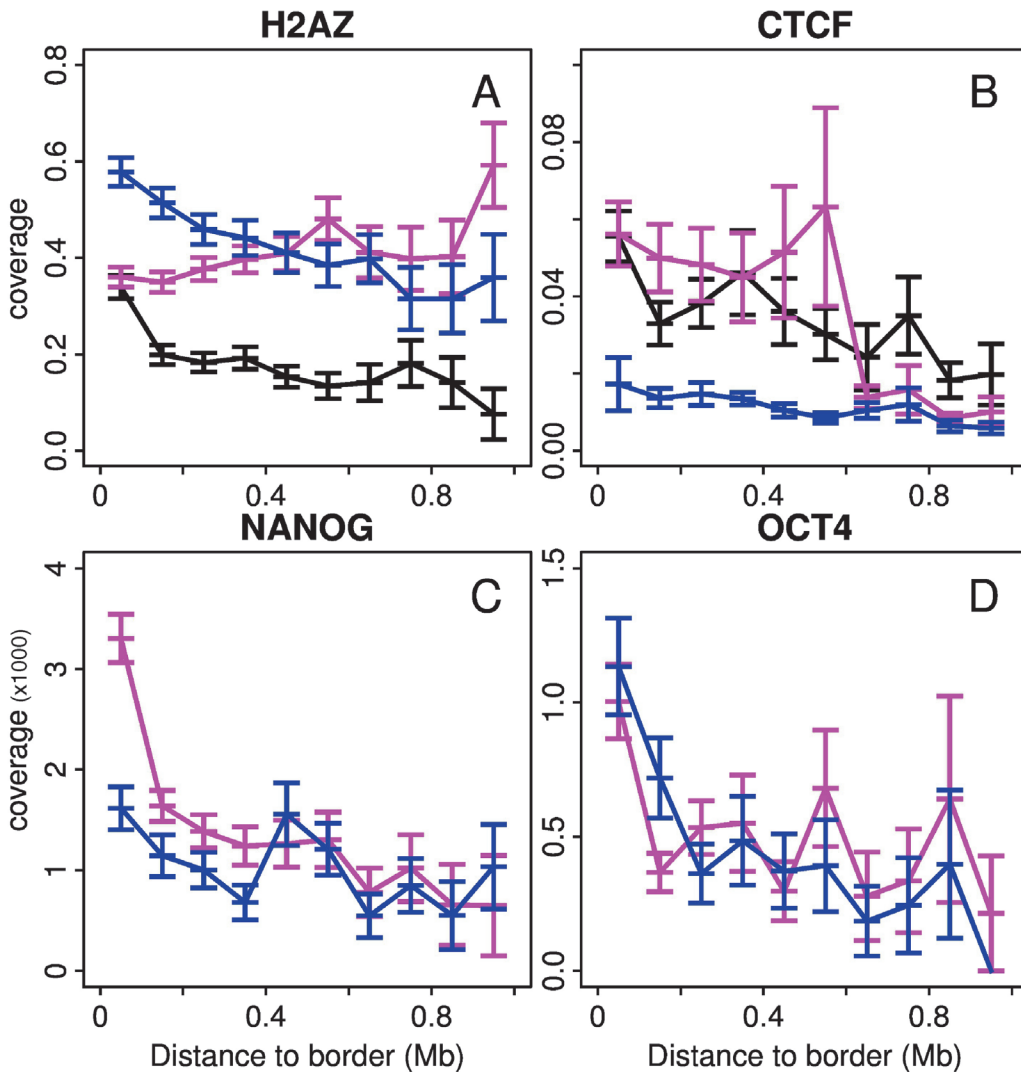


Figure 8. Epigenetic marks enrichment in specific MRT U-domains of H1hes and Nhdad. (A) Mean coverage of H2AZ enriched intervals with respect to the distance to the closest U-domain border specific to the cell line. The different colors correspond to specific U-domains of Nhdad (black), specific U-domains of H1hes whose border is in EC1 or EC2 (red) and specific U-domains of H1hes whose border is in EC4 (blue). (B), (C) and (D) are as (A) for respectively CTCF, NANOG and OCT4. Note that in (C) and (D) coverage are per thousand.

doi:10.1371/journal.pcbi.1003969.g008

replication U-domains actually reflects the almost uniform covering (inside U-domains as well as outside) of the human genome by the histone variant H2AZ in pluripotent cells (Fig. 8A) [28].

Discussion

Specific genome-wide histone signature of pluripotent plastic chromatin

Our integrative analysis of epigenetic marks confirmed the existence of fundamental differences between the pluripotent and differentiated chromatin states (Fig. 2). These differences account for the changes observed in epigenetic landscapes in ESC and lineage committed cells (Fig. 5) [6, 9, 11, 26, 28, 39]. In general, histone modifications show two distinct types of spatial distributions: small localized peaks and large spreading domains. The histone variant H2AZ associated with nucleosome exchange and remodeling [6, 11, 23, 24, 28, 115, 116], was typically

found confined to promoters and distal elements in differentiated cells [6, 28] which explains its abundance in the gene-rich chromatin states C1 and C2 (Fig. 2). Its binding level was further shown to correlate with gene expression in human [6] which is consistent with its highest enrichment in the transcriptionally active state C1. Remarkably, the global H2AZ distribution diverges markedly between pluripotent and differentiation cells. In H1hesc, 92% of the overall 100-kb loci contain the H2AZ mark as compared to smaller coverages in K562 (61%), Gm12878 (65%), Hmec (76%) and Nhdfad (79%) (Note the important covering found for Monocd14ro (94%)) (S1 Table). Thus in ESCs, H2AZ marks promoters and distal elements but it is also distributed throughout intergenic regions which explains its presence in the gene-rich chromatin states EC1 and EC2 and in addition, its specific abundance in the gene-poor chromatin state EC4 (Fig. 2). This broad H2AZ distribution suggests that chromatin exchange and remodeling are prevalent throughout human chromosomes in ESCs [28]. This highly dynamic and potentially accessible properties of pluripotent chromatin are further strengthened by the presence of the ATP-dependent remodeler CHD1 not only in EC1 and EC2 but also in EC4 as an inhibitory factor to HP1-heterochromatin (Fig. 2) [122]. In addition, we showed that the active marks H3K4me1 has a broad dispersion in H1hesc (85% coverage) as compared to differentiated cells types K562 (55%), Monocd14ro1746 (61%), Gm12878 (62%), Hmec (77%) and Nhdfad (72%) (S1 Table), likely resulting from a more restrictive confinement at promoters and enhancers in differentiated cells [28]. This is consistent with the abundance of H3K4me1 in the gene-rich EC1 and EC2 chromatin states and also with its presence in the gene-poor EC4 state contrasting its absence in the heterochromatin state C4 (Fig. 2). There is another histone modification, namely H3K27me3, that distributes quite differently in pluripotent and differentiated cells. In ESCs, this surrogate of polycomb activity is narrowly distributed (37% coverage) relative to the much broader distributions of H3K27me3 in K562 (54%), Monocd14ro1746 (65%), Gm12878 (55%), Hmec (54%) and Nhdfad (60%). In pluripotent cells, H3K27me3 is known to be mainly confined to “bivalent” promoters that also carry H3K4me3 [32–36, 112]. As indicated by the co-presence of H3K27me3 and H2AZ in the bivalent chromatin state EC2 (Fig. 2) and the observed local surrounding of H3K27me3 marks by H2AZ variants in the H1hesc epigenetic landscape (Fig. 5), the highly dynamic pluripotent chromatin is likely refractory to polycomb facultative heterochromatin formation and spreading [9, 28]. The smaller mean size of EC4 blocks ($\bar{L} = 301$ kb) in H1hesc as compared to C4 blocks in K562 ($\bar{L} = 882$ kb), Hmec ($\bar{L} = 610$ kb) and Nhdfad ($\bar{L} = 894$ kb) (Table 3), suggests that the gene-poor H2AZ marked accessible EC4 chromatin is incompatible with the stable interactions involved in the H3K9me3 enriched HP1 heterochromatin compaction and spreading (Fig. 2). All the other histone marks known to be involved in transcription positive regulation, including H3K27ac and H3K36me3, have a similar distribution with a similar coverage of the gene-rich genome regions in H1hesc (EC1+EC2) and differentiated cells (C1+C2) (S1 Table).

Distinct epigenetic mechanisms of heterochromatin expansion during differentiation

There are mainly two epigenetic mechanisms of heterochromatin expansion during differentiation that correspond to the transitions towards the polycomb repressed state C2 and towards the HP1-associated heterochromatin state C4 (Fig. 3A). For the former mechanism, there are indeed two possible scenarios according to whether the pluripotent chromatin state that switches to C2 is EC2 or EC4. As previously described, EC2 is a bivalent chromatin state that is enriched in gene promoters that carry both the active mark H3K4me3 and the polycomb associated mark H3K27me3. This second mark has repressive effect on gene expression and

contributes to maintain repression of bivalent genes including developmental genes [32–36, 112]. Some of these bivalent genes get activated during differentiation and switch from EC2 to the open euchromatin state C1 (Fig. 3A). The other ones experience some repression to the facultative chromatin state C2 via the expansion of H3K27me3 to often cover the entire gene and frequently neighboring gene loci [9]. But there is another category of genes that face this facultative heterochromatinization which are the genes that are in the H2AZ rich accessible chromatin state EC4. These EC4 regions are actually lying nearby EC2 regions and get involved in the repressive expansion of H3K27me3 which dictates their switch to the polycomb repressed state C2. Note that this H3K27me3 spreading over several kb or tens of kb is locally at the expense of H2AZ which confirms that, in pluripotent cells, this histone variant is refractory to the compaction associated with polycomb repression [28]. Interestingly, the polycomb repressed scenario from EC4 to C2 mainly corresponds to MRT changes from late to early replicating loci (data not shown).

The second mechanism corresponds to transitions from the silent unmarked (EC3) and H2AZ rich accessible (EC4) states to the HP1-associated heterochromatin state C4 (Fig. 3A). This mechanism corresponds to a dramatic redistribution of the histone modification H3K9me3 which, although present in the pluripotent EC4 state, expands into large (from several 100 kb to a few Mb) late replicating highly compacted heterochromatin (Table 3, Figs. 4 and 5). H3K9me3 is important for the formation of the constitutive heterochromatin via the anchoring of the α and β isoforms of the HP1 protein [142, 143]. There is also evidence of some crosstalk between H3K9 methyltransferase (HKMT) and DNA methyltransferase (DNMT) [144, 145] that might explain the correlation observed between H3K9me3 and DNA methylation and the contribution of the later to the long-term maintenance of these large domains of late replicating C4 heterochromatin devoid of H2AZ and of any other histone modification but H3K9me3 [4, 9]. Knockout studies of H3K9 methyltransferases and H3K27 methyltransferases have led to differentiation or development defects [146–151], confirming that the epigenetic mechanisms underlying heterochromatin expansion play a critical role in cell fate determination.

Master replication origins at U/N-domain borders are determinants of cell-fate commitment

We found that MRT changes induced by differentiation resulted in an important change in the number and size of replication U/N-domains [77]. Small neighboring U/N-domains merged to become one large coordinately replicated domains (2 and 3 domains merged to 1 in Fig. 5 left and right column respectively). This replication domain consolidation [57, 58] is thus the consequence of an active early replication initiation zone in ESCs that no longer fires early in somatic cells. To characterize this consolidation phenomenon from pluripotent to differentiated cell lines as well as between differentiated cell lines, we defined an index of conservation n (Materials and Methods) that quantifies the number of U-domain borders in a given cell line that were also shared by $n-1$ other cell lines. To the sets of U-domains of the cell types considered so far, we added those previously identified in HeLa cells [61, 77] and the germline replication skew N-domains [77–83] (Fig. 5). For each cell type, about half U-domains are shared by at least another cell line, namely H1hesc (38.4%), K562 (61%), Gm12878 (59.2%), Nhdfad (51.6%) and Ndom (50.2%). Note that the smallest matching percentage was obtained for H1hesc as a direct consequence of the largest number of U-domains in this ESC line. When looking at U-domain borders individually (peaks in replication timing [61]), we got the following percentages of matching with at least another U/N-domain borders in another cell line: H1hesc (78.8%), K562 (88.1%), Gm12878 (88.9%), Nhdfad (85.6%) and Ndom (87.9%). As

originally revealed in skew N-domains [51, 88] and further confirmed in MRT U-domains [77], there exists a remarkable gene organization inside these replication domains that turns out to be robust in each cell type. Expressed genes are confined in the euchromatin C1 (resp. EC1) environment of the bordering master replication origins whereas non expressed genes are distributed rather uniformly inside these domains (S8 Fig.) independently of the gradient of chromatin states (Fig. 7A') (resp. Fig. 7A). When comparing the gene content nearby replication U/N-borders for increasing index of conservation (S8C Fig.), we found that the density as well as the distribution of non-expressed genes were quite insensitive to the degree of ubiquity of the nearby master replication origin. In other words, non-expressed genes seem to have no knowledge of the replication wave initiating at U/N-domain borders. We got the opposite for expressed genes with a significant enhancement of gene density when increasing the conservation index n (S8C Fig.). Ubiquitous master replication origins are surrounded by a C1 euchromatin environment which is hypomethylated (Fig. 9D), GC-high (Fig. 9E, Table 4), significantly enriched in DHS and CTCF (Table 4) and more importantly in nucleosome free regions (NFRs) (Fig. 9B) coded in the DNA sequence via high energy barriers that impair nucleosome formation (Material and Methods) [87, 104–108]. Thus these ubiquitous master replication origins are specified by an open chromatin structure which is to some extent encoded in the DNA sequence [3, 87]. This also provides some understanding of the local clustering of highly expressed genes with strong CpG rich promoters including house-keeping genes (S8B Fig.). As exemplified with the Nhdad cell line, master replication origins that are specific to a differentiated cell line are still GC high (Fig. 9E) but no longer enriched in NFRs (Fig. 9A) suggesting that these early firing regions are epigenetically regulated and no longer favored by the DNA sequences. Indeed, Nhdad specific master replication origins are hypomethylated (Fig. 9C), and significantly enriched in H2AZ (Fig. 8A) and CTCF (Fig. 8B) epigenetic marks. They are mainly surrounded by tissue specific genes with weak CpG poor promoters. Our results are consistent with previous reports that most genes do not change expression during domain consolidation in the MRT profile [57, 58, 60, 65]. When examining the joint distribution of gene expression in H1hesc and K562 (S5C Fig.), for 100kb loci that experience a EtoL transition and reversely for those that change from LtoE, we confirmed that most (~55%) genes lying in dynamic MRT regions do not change expression, suggesting that, at the 100 kb scale, phenotypic differences between cell types are better reflected by epigenetic properties including the MRT than by transcriptional differences.

ESC specific master replication origins as the cornerstone of pluripotency maintenance

Master replication origins that are specific to the pluripotent H1hesc cell line actually correspond to lineage-independent switches in MRT that are stably maintained after the late epiblast stage. These pluripotent master replication origins (N = 483) are almost equally distributed in the chromatin states EC1 (N = 113), EC2 (N = 131) and EC4 (N = 149) and only few are in the unmarked state EC3 (N = 51) and in the discarded set D (N = 41) (Table 5). Those that are gene rich in EC1 and EC2 environments display very similar properties than master replication origins specific to differentiated cell lines. They are hypomethylated (Fig. 9C), enriched in CTCF (Fig. 8B) and DHS (Table 5), their GC content is high (Fig. 9E) but they are not enriched in constitutive NFRs (Fig. 9A) as an indication of epigenetic regulation. Interestingly, these H1hesc specific (EC1, EC2) master replication origins are enriched in the key pluripotency transcription factors NANOG (Fig. 8C) and OCT4 (Fig. 8D) (Table 5). Note that they are also highly covered by the H2AZ mark but they are nonetheless depleted compared to the very high level coverage of the genome (Fig. 8A). Somatic specific master replication origins have the

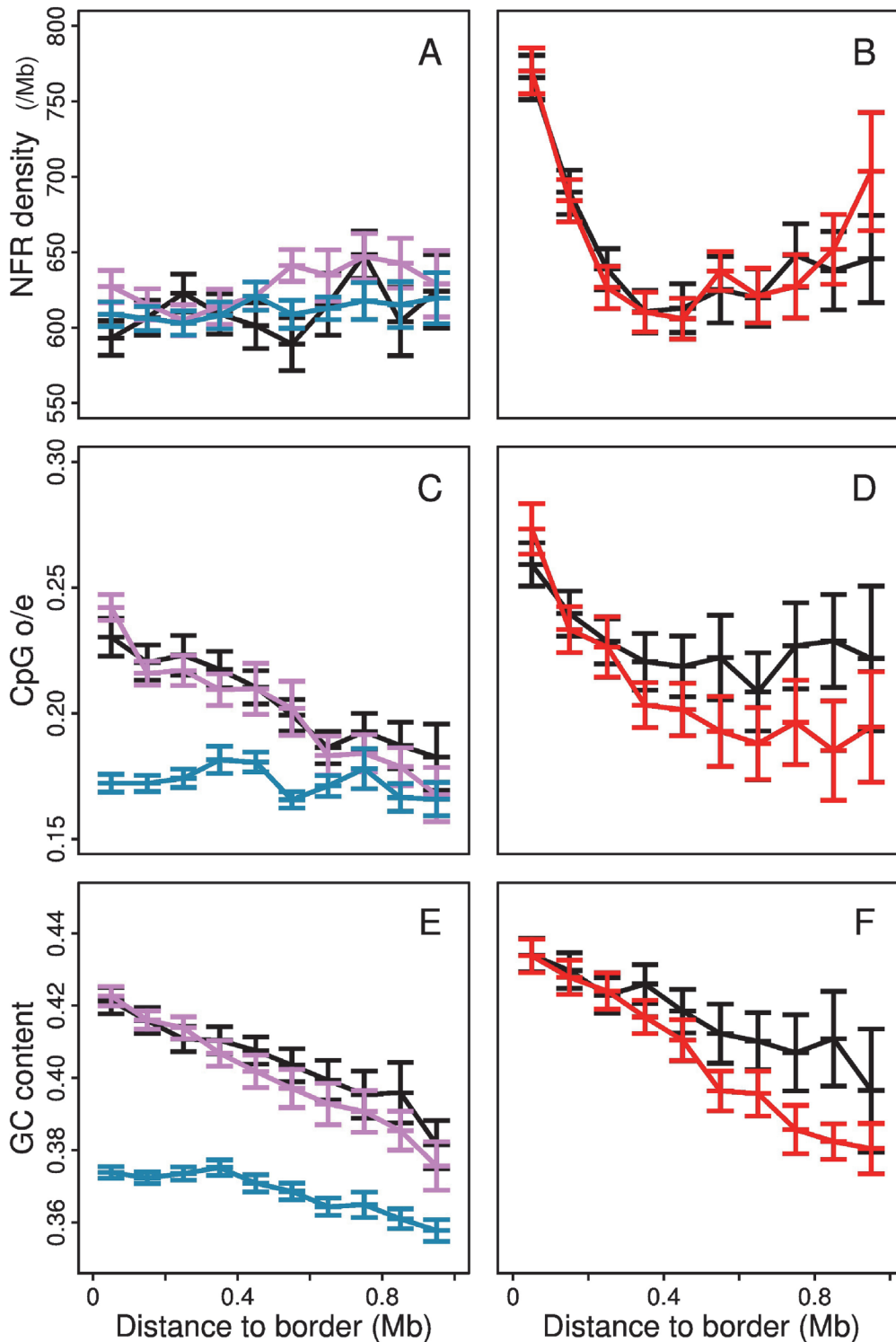


Figure 9. Sequence characteristics of MRT U-domains of H1hesec and Nhdfad. (A) Density of nucleosome free regions (NFRs) with respect to the distance to the closest U-domains border specific ($n = 1$) to the cell line. The different colors correspond to specific U-domains of Nhdfad (black), specific U-domains of H1hesec whose border is in EC1 or EC2 (red) and specific U-domains of H1hesec whose border is in EC4 (blue). (B) Density of (NFRs) with respect to the distance to the closest conserved ($n = 6$) U-domains border. (C) same as (A) for the CpG o/e. (D) same as (B) for the CpG o/e. (E) same as (A) for the GC content. (F) same as (B) for the GC content.

doi:10.1371/journal.pcbi.1003969.g009

Table 4. Sequence, epigenetic and gene characteristics of conserved (n = 6) replication U-domain borders of H1hesec that switch from state ECi to Cj.

	Nb	GC	CpGo/e	H2AZ	CTCF	NANOG	OCT4	DHS	Promoters/border	Expressed promoters/border
1 to 1	38	0.445	0.233	0.331	0.663	1.321	0.335	0.016	3.947	3.237
1 to 2	3	0.458	0.235	0.346	0.472	0.733	0.145	0.018	3.333	3
2 to 1	4	0.444	0.198	1.214	1.066	1.254	0.133	0.025	3	1.75
2 to 2	6	0.427	0.206	0.606	0.872	0.556	0.458	0.021	2.667	2
2 to 3	1	0.436	0.179	0.468	0.248	0.277	0	0.014	2	2
3 to 2	1	0.418	0.16	0.765	0.419	0	0	0.008	3	3
4 to 2	1	0.391	0.143	0.526	0.261	0	0	0	2	2
4 to 3	1	0.445	0.192	1.227	0.049	0.234	0	0.004	1	1

For each border type, the table provides number (Nb), G+C content (GC), coverage by significantly enriched intervals relative to the genome average coverage for H2AZ (genome average 33%), CTCF (4.5%), NANOG (0.2%) and OCT4 (0.07%), coverage by DNase I hypersensitive sites (DHS), as well as the numbers of promoters and expressed promoters per border.

doi:10.1371/journal.pcbi.1003969.t004

same coverage in H2AZ than specific ESC ones, but in contrast they are enriched compared to the genome background (Fig. 8A).

More surprising is the non negligible proportion (30.9%) of specific H1hesec origins that belong to a EC4 environment and that mainly consolidate into a C4 heterochromatin domain (Table 5). These EC4 master replication origins indeed correspond to the early replicating EC4 regions that experience a EtoL transition mostly towards the HP1-associated heterochromatin state C4 (Table 5). They have totally different epigenetic and sequence properties. They are methylated (Fig. 9C), no longer enriched in CTCF (Fig. 8B) and DHS (Table 5), their GC content is low (Fig. 9E) and they are still not enriched in constitutive NFRs (Fig. 9A). Actually they are mainly epigenetically regulated by a local enrichment of H2AZ (Fig. 8A) that turns out to play an unexpected specific role in regulating the spatio-temporal replication program in pluripotent cells. Notably, these pluripotent specific EC4 master replication initiation zones are gene deserts: only 20/82 (~24%) contain a gene promoter as compared to 82/99 (resp. 75/100)

Table 5. Sequence, epigenetic and gene characteristics of specific (n = 1) replication U-domain borders of H1hesec that switch from state ECi to Cj.

	Nb	GC	CpGo/e	H2AZ	CTCF	NANOG	OCT4	DHS	Promoters/border	Expressed promoters/borde
1 to 1	58	0.442	0.226	0.361	0.602	1.129	0.159	0.014	3.379	2.655
1 to 2	15	0.429	0.21	0.44	0.655	2.689	0.759	0.013	1.667	1.4
1 to 3	12	0.42	0.201	0.241	0.488	0.998	0.199	0.013	1.917	1.25
1 to 4	14	0.399	0.194	0.469	0.212	0.783	0.13	0.009	0.643	0.643
2 to 1	9	0.418	0.224	0.672	0.886	1.474	0.076	0.019	1.889	1
2 to 2	56	0.441	0.205	0.804	0.903	1.507	0.355	0.017	2.357	1.25
2 to 3	16	0.43	0.181	0.582	0.482	2.418	0.691	0.014	0.875	0.5
2 to 4	19	0.408	0.214	0.977	0.503	2.968	0.76	0.012	0.684	0.158
3 to 2	5	0.374	0.15	0.276	0.336	0.074	0	0.003	1.4	0.6
3 to 3	11	0.371	0.143	0.228	0.069	0.257	0	0.003	0.273	0.273
3 to 4	14	0.352	0.165	0.352	0.068	0.217	0.165	0.004	0.214	0.143
4 to 1	1	0.399	0.231	0.189	0.288	0	0.389	0.008	0	0
4 to 2	13	0.391	0.155	0.824	0.314	0.244	0.148	0.008	1	0.308
4 to 3	22	0.388	0.164	1.007	0.226	0.577	0.257	0.006	0.864	0.409
4 to 4	46	0.376	0.172	0.984	0.14	1.232	0.408	0.006	0.174	0.087
4 to 4 intergenic	39	0.374	0.17	1.013	0.145	1.402	0.481	0.006	0	0

The information provided for each border category is the same as in Table 4.

doi:10.1371/journal.pcbi.1003969.t005

for those in EC1 (resp. EC2). Nevertheless, they are enriched in NANOG and OCT4 (Fig. 8C and D, Table 5), even in the intergenic ones, which suggests that these transcription factors are also involved in the regulation of replication in pluripotent cells. Note that the unusual principle of chromatin folding during development reported in [70] likely results from the C4 domain consolidation of these early ESC specific EC4 master replication origins (see for example one of them at position 47 Mb on the right panels of Fig. 5). As discussed in previous works [57, 58, 68, 70], the EtoL transitions associated with the consolidation of pluripotent specific EC1 (see for example one of them at position 12.5 Mb on the left panel of Fig. 5), EC2 and EC4 to HP1-associated C4 heterochromatin likely coincide with the emergence of compact chromatin near the nuclear periphery and with a dramatic large-scale 3D genome reorganization that may constitute an epigenetic barrier to cellular reprogramming. In that respect, the master-replication origins bordering ESC specific replication U/N-domains are likely to be major determinants in the maintenance of pluripotency.

Conclusion/Perspectives

In summary, the integrative analysis of genome-wide epigenetic marks, expression and MRT data at 100 kb resolution in an ESC and several differentiated human cell lines, shows that the combinatorial complexity of these epigenetic data can be significantly reduced consistently with previous studies in *Drosophila* [66, 97], *Arabidopsis* [95] and *human* [28, 92, 99]. The epigenetic landscapes of pluripotent and differentiated cells are different even though, in both cases, four but distinct prevalent chromatin states are enough to characterize the diversity in chromatin environment along human chromosomes. Among these four states, only one is transcriptionally active and three are silent. The first one is a gene rich euchromatin state that is shared by pluripotent (EC1) and differentiated (C1) cells as well as the “unmarked” states EC3 and C3 that correspond to a silent state not enriched in any available epigenetic marks. The two other states are different as the signature of the global accessible character of the pluripotent chromatin [26]: H2AZ and H3K4me1 marks are broadly distributed [28] in the bivalent state EC2 containing bivalent genes and in the gene-poor accessible EC4 state as compared to the polycomb repressed state C2 and the HP1-associated heterochromatin state C4 that respectively result from the spreading of H3K27me3 and H3K9me3 in differentiated cells [9, 28]. When looking at the way these chromatin states are distributed along human chromosomes with a special focus on the regions where the MRT changes significantly during differentiation, we show that the master replication origins that border megabase-sized MRT U/N-domains [77, 78] are major determinants in cell-fate commitment and lineage fidelity. The minority (5.3%) that are conserved in all cell lines have a peculiar high GC hypomethylated (EC1, C1) euchromatin environment highly enriched in open marks including H2AZ, CTCF, DNase HS and also in NFRs encoded in the DNA sequence suggesting that these ubiquitous master replication origins have been selected during evolution. In these particularly highly decondensed regions are also found numerous CpG rich promoters of highly expressed genes including house-keeping genes. Most of the master replication origins that are cell type specific or shared by a few cell types, still correspond to GC-rich euchromatin mainly regulated epigenetically and no longer favored by a local abundance of NFRs encoded in the DNA sequence. They are mainly surrounded by highly expressed tissue-specific genes. A majority of master replication origins specific to ESCs have rather similar epigenetic properties with a high density of neighboring genes that are likely regulated by the pluripotency factors NANOG/OCT4. But what our study has revealed is the existence of a class of ESC specific master replication origins that fire early in a GC-low, gene desert EC4 environment and that experiences a change to a compact HP1-associated C4 heterochromatin environment during differentiation. These master

origins have a specific epigenetic regulation that sheds a new light on the unexpected role of both H2AZ and the transcription factors NANOG/SOX2/OCT4 in the maintenance of the replication spatio-temporal program in pluripotent cells. An important proportion (67.4%) of the ESC specific master replication origins indeed correspond to EtoL transitions likely associated with some repositioning towards the nuclear periphery and some large-scale 3D chromatin rearrangements that may hinder cell reprogramming [57, 58, 68, 70]. As reported in previous studies of 4C [75] and Hi-C [77, 90] data in differentiated cell lines, these master replication zones at MRT U/N-domain borders act on the one hand as insulators that delimit topological domains of self-interacting chromatin [152], and on the other hand as long-distance interconnected hubs in the intra- and inter- chromosome interaction network [119]. This highlights the interest of performing a similar comparative analysis of Hi-C data in ESCs. Besides confirming the key role played by ESC specific master replication origins in the 3D chromatin regulation and control of pluripotency, it will likely bring new elements of discussion concerning the hypothesized influence of longer G1-phase enabling targeting of loci to the nuclear periphery and providing more time for nuclei to reorganize their genome before replication initiates in differentiated cells [57, 58]. As fundamental structural and functional units underlying the plasticity of replication domain organization in relation to gene expression and chromatin states, the replication timing U/N-domains together with the bordering master replication origins provide a framework for further studies in different cell types and different organisms, in both health and disease.

Supporting Information

S1 Table. Genome coverage by epigenetic marks in ESCs and differentiated cells. Percentage of 100-kb windows that contain a given epigenetic mark.

(TIFF)

S1 Fig. Spearman correlation matrix between epigenetic marks in Hmec, Mono-cd14ro1746, K562 and Gm12878 cell lines. Same color map as in Fig. 1.

(EPS)

S2 Fig. PCA analysis and clustering procedure for ESC line (A-C) and the five differentiated cell lines (A'-C'). (A, A') Percentage of variance accounted by the eleven principal components ordered according to their corresponding variance (eigenvalues). (B, B') Scatterplot of the data points when projected on the (PC1, PC2) plane; color dots indicate the four chromatin states as found by our clustering procedure. (C, C') Density of data points on the (PC1, PC2) plane using the same color coding as in (B, B'). In (B, C) the colors have the following meaning: EC1 (light pink) transcriptionally active chromatin, EC2 (light orange) bivalent chromatin, EC3 (light green) silent unmarked chromatin, EC4 (light blue) dynamically accessible chromatin poised to HP1-heterochromatin expansion. In (B', C') the colors correspond to: C1 (pink) transcriptionally active chromatin, C2 (orange) chromatin repressed by polycomb, C3 (green) silent unmarked chromatin, C4 (blue) HP1-associated heterochromatin. In (B, B', C, C') the points in dark grey are not classified in any chromatin state (Materials and Methods).

(EPS)

S3 Fig. Repartition of the histone modifications H3K4me2, H3K79me2 and H4K20me1 in the four prevalent chromatin states of H1hesc cell line (EC1, EC2, EC3, EC4, same color coding as in S2B and C Fig.) and differentiated cell lines (C1, C2, C3, C4, same color coding as in S2B' and C' Fig.). Boxplots of the decimal logarithm of epigenetic mark ChIP-seq read

density in 100 kb non-overlapping windows per chromatin state.
(EPS)

S4 Fig. Mosaic plots representing the probabilities of transition between the chromatin states of two different cell lines (from line 1 to line 2). The width of columns corresponds to the proportion of chromatin states in line 1. The segmentation for the i^{th} column follows the proportion of windows in state (E)C_i in line 1 that become C_j in line 2. In other words, if we take the first pink rectangle of the first column, its width is proportional to the probability for a 100 kb window to be in chromatin state (E)C₁ in line 1 and its height is proportional to the probability for a 100 kb window to be in C₁ in line 2 given that it is in (E)C₁ in line 1. The area of this rectangle (product of the previously mentioned probability) is proportional to the probability for a window to be in state (E)C₁ in line 1 and C₁ in line 2.
(EPS)

S5 Fig. Gene expression in the H1hesc and K562 chromatin states. (A) Density of promoters in the 4 chromatin states of the H1hesc cell line as a function of gene expression (genes were grouped into bins of width 0.05 in log₁₀(RPKM) unit). Same color coding as in [S2B and C Fig.](#) (B) Density of promoters in the 4 chromatin states of the K562 cell line as a function of gene expression. Same color coding as in [S2B' and C' Fig.](#) (C) 2D representation of the joint density of gene expression in H1hesc (X-axis) and K562 (Y-axis) when focusing on EtoL (blue) and LtoE (magenta) MRT transitions. For comparison is shown as a control (black), the joint density obtained for comparable size sets of randomly chosen genes.
(EPS)

S6 Fig. Spatial organization of chromatin states in H1hesc. (A) Histogram of chromatin state (EC₁, EC₂, EC₃, EC₄) block length in a logarithmic representation (Materials and Methods). (B) same as (A) for chromatin blocks formed by states EC₁ and EC₂ (EC₁+EC₂, light red) or by states EC₃ and EC₄ (EC₃+EC₄, light blue). (C) MRT in chromatin state blocks EC₁+EC₂ with respect to their length. Each 100 kb window in a chromatin state block is represented by the color of its state defined in [S2B and C Fig.](#) The mean profile was obtained by (i) ordering data points according to their block length, (ii) grouping them in classes of equal number of data points and (iii) computing the average length and MRT over each class. Vertical bars represent the standard deviation. Horizontal bars represent the range of length over each class. (D) Same as (C) for chromatin state blocks EC₃+EC₄.
(EPS)

S7 Fig. Spatial organization of chromatin states in Nhd₁fad. (A) Histogram of chromatin state (C₁, C₂, C₃, C₄) block length in a logarithmic representation (Materials and Methods). (B) same as (A) for chromatin states C₁ and C₂ (C₁+C₂, light red) or by states C₃ and C₄ (C₃+C₄, light blue). (C) MRT in chromatin state blocks C₁+C₂ with respect to their length. Each 100 kb window in a chromatin state block is represented by the color of its state defined in [S2B' and C' Fig.](#) The mean profile was obtained by (i) ordering data points according to their block length, (ii) grouping them in classes of equal number of data points and (iii) computing the average length and MRT over each class. Vertical bars represent the standard deviation. Horizontal bars represent the range of length over each class. (D) Same as (C) for chromatin state blocks C₃+C₄.
(EPS)

S8 Fig. Distribution of expressed (orange) and not expressed (blue) gene promoters inside replication timing U-domains of H1hesc (solid line) and Nhd₁fad (dashed line). (A) Mean

density of gene promoters with respect to the distance to the closest U-domain border specific to the cell line ($n = 1$). (B) Mean density of gene promoters with respect to the distance to the closest U-domain common to all cell lines ($n = 6$). (C) Mean density of gene promoters in the 100 kb windows containing a U-domain border versus its conservation index n (Materials and Methods).

(EPS)

S9 Fig. Spearman correlation matrix between epigenetic marks in HeLa. Same color coding as in Fig. 1.

(EPS)

Acknowledgments

We are very grateful to Y. d’Aubenton-Carafa, A. Baker, JC. Cadoret, E. Cascales, CL. Cheng, L. Duret, A. Goldar, O. Hyrien, M. Kahli, N. Petryk, F. Picard, MN. Prioleau, C. Thermes and C. Vaillant for helpful discussions.

Author Contributions

Performed the experiments: HJ. Analyzed the data: HJ BA AA. Contributed reagents/materials/analysis tools: HJ BA. Wrote the paper: HJ BA AA.

References

1. Lander ES et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, 409: 860–921. doi: [10.1038/35057062](https://doi.org/10.1038/35057062) PMID: [11237011](https://pubmed.ncbi.nlm.nih.gov/11237011/)
2. Lander ES (2011) Initial impact of the sequencing of the human genome. *Nature*, 470: 187–197. doi: [10.1038/nature09792](https://doi.org/10.1038/nature09792) PMID: [21307931](https://pubmed.ncbi.nlm.nih.gov/21307931/)
3. Arneodo A, Vaillant C, Audit B, Argoul F, d’Aubenton-Carafa Y et al. (2011) Multi-scale coding of genomic information: From DNA sequence to genome structure and function. *Phys. Rep.*, 498: 45–188. doi: [10.1016/j.physrep.2010.10.001](https://doi.org/10.1016/j.physrep.2010.10.001)
4. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J et al. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, 454: 766–770. doi: [10.1038/nature07107](https://doi.org/10.1038/nature07107) PMID: [18600261](https://pubmed.ncbi.nlm.nih.gov/18600261/)
5. Bernstein BE, Meissner A, and Lander ES (2007) The mammalian epigenome. *Cell*, 128: 669–681. doi: [10.1016/j.cell.2007.01.033](https://doi.org/10.1016/j.cell.2007.01.033) PMID: [17320505](https://pubmed.ncbi.nlm.nih.gov/17320505/)
6. Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, et al. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, 129: 823–837. doi: [10.1016/j.cell.2007.05.009](https://doi.org/10.1016/j.cell.2007.05.009) PMID: [17512414](https://pubmed.ncbi.nlm.nih.gov/17512414/)
7. Schones DE and Zhao K. (2008) Genome-wide approaches to studying chromatin modifications. *Nat. Rev. Genet.*, 9, 179–191. doi: [10.1038/nrg2270](https://doi.org/10.1038/nrg2270) PMID: [18250624](https://pubmed.ncbi.nlm.nih.gov/18250624/)
8. Wang Z, Schones DE, and Zhao K (2009) Characterization of human epigenomes. *Curr. Opin. Genet. Dev.*, 19: 127–134. doi: [10.1016/j.gde.2009.02.001](https://doi.org/10.1016/j.gde.2009.02.001) PMID: [19299119](https://pubmed.ncbi.nlm.nih.gov/19299119/)
9. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, et al. (2010) Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell*, 6, 479–491. doi: [10.1016/j.stem.2010.03.018](https://doi.org/10.1016/j.stem.2010.03.018) PMID: [20452322](https://pubmed.ncbi.nlm.nih.gov/20452322/)
10. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, et al. (2010) The NIH roadmap epigenomics mapping consortium. *Nat. Biotechnol.*, 28: 1045–1048. doi: [10.1038/nbt1010-1045](https://doi.org/10.1038/nbt1010-1045) PMID: [20944595](https://pubmed.ncbi.nlm.nih.gov/20944595/)
11. Zhou VW, Goren A, and Bernstein BE (2011) Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.*, 12: 7–18. doi: [10.1038/nrg2905](https://doi.org/10.1038/nrg2905) PMID: [21116306](https://pubmed.ncbi.nlm.nih.gov/21116306/)
12. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, et al. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473: 43–49. doi: [10.1038/nature09906](https://doi.org/10.1038/nature09906) PMID: [21441907](https://pubmed.ncbi.nlm.nih.gov/21441907/)

13. The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489: 57–74. doi: [10.1038/nature11247](https://doi.org/10.1038/nature11247) PMID: [22955616](https://pubmed.ncbi.nlm.nih.gov/22955616/)
14. Cook PR (1999) The organization of replication and transcription. *Science*, 284: 1790–1795. doi: [10.1126/science.284.5421.1790](https://doi.org/10.1126/science.284.5421.1790) PMID: [10364545](https://pubmed.ncbi.nlm.nih.gov/10364545/)
15. Cremer T and Cremer C (2001) Chromosome Territories, Nuclear Architecture and Gene Regulation in Mammalian Cell. *Nat. Rev. Genet.*, 2, 292–301. doi: [10.1038/35066075](https://doi.org/10.1038/35066075) PMID: [11283701](https://pubmed.ncbi.nlm.nih.gov/11283701/)
16. Berezney R. (2002) Regulating the mammalian genome: the role of nuclear architecture. *Adv. Enzyme Regul.*, 42: 39–52. doi: [10.1016/S0065-2571\(01\)00041-3](https://doi.org/10.1016/S0065-2571(01)00041-3) PMID: [12123705](https://pubmed.ncbi.nlm.nih.gov/12123705/)
17. Chakalova L, Debrand E, Mitchell JA, Osborne CS, and Fraser P. (2005) Replication and transcription: shaping the landscape of the genome. *Nat. Rev. Genet.*, 6, 669–677. doi: [10.1038/nrg1673](https://doi.org/10.1038/nrg1673) PMID: [16094312](https://pubmed.ncbi.nlm.nih.gov/16094312/)
18. Gilbert N, Gilchrist S, and Bickmore WA (2005) Chromatin organization in the mammalian nucleus. *Int. Rev. Cytol.*, 242: 283–336. doi: [10.1016/S0074-7696\(04\)42007-5](https://doi.org/10.1016/S0074-7696(04)42007-5) PMID: [15598472](https://pubmed.ncbi.nlm.nih.gov/15598472/)
19. Misteli T. (2007) Beyond the sequence: cellular organization of genome function. *Cell*, 128: 787–800. doi: [10.1016/j.cell.2007.01.028](https://doi.org/10.1016/j.cell.2007.01.028) PMID: [17320514](https://pubmed.ncbi.nlm.nih.gov/17320514/)
20. Sexton T, Schober H, Fraser P, and Gasser SM (2007) Gene regulation through nuclear organization. *Nat. Struct. Mol. Biol.*, 14: 1049–1055. doi: [10.1038/nsmb1324](https://doi.org/10.1038/nsmb1324) PMID: [17984967](https://pubmed.ncbi.nlm.nih.gov/17984967/)
21. Branco MR and Pombo A. (2007) Chromosome organization: new facts, new models. *Trends Cell Biol.*, 17: 127–134. doi: [10.1016/j.tcb.2006.12.006](https://doi.org/10.1016/j.tcb.2006.12.006) PMID: [17197184](https://pubmed.ncbi.nlm.nih.gov/17197184/)
22. Fraser P. and Bickmore W. (2007) Nuclear organization of the genome and the potential for gene regulation. *Nature*, 447: 413–417. doi: [10.1038/nature05916](https://doi.org/10.1038/nature05916) PMID: [17522674](https://pubmed.ncbi.nlm.nih.gov/17522674/)
23. Kouzarides T. (2007) Chromatin modifications and their function. *Cell*, 128: 693–705. doi: [10.1016/j.cell.2007.02.005](https://doi.org/10.1016/j.cell.2007.02.005) PMID: [17320507](https://pubmed.ncbi.nlm.nih.gov/17320507/)
24. Rando OJ and Chang HY (2009) Genome-wide views of chromatin structure. *Annu. Rev. Biochem.*, 78: 245–271. doi: [10.1146/annurev.biochem.78.071107.134639](https://doi.org/10.1146/annurev.biochem.78.071107.134639) PMID: [19317649](https://pubmed.ncbi.nlm.nih.gov/19317649/)
25. Bickmore WA and vanSteensel B (2013) Genome architecture: domain organization of interphase chromosomes. *Cell*, 152: 1270–1284. doi: [10.1016/j.cell.2013.02.001](https://doi.org/10.1016/j.cell.2013.02.001) PMID: [23498936](https://pubmed.ncbi.nlm.nih.gov/23498936/)
26. Meshorer E and Misteli T (2006) Chromatin in pluripotent embryonic stem cells and differentiation. *Nat. Rev. Mol. Cell Biol.*, 7, 540–546. doi: [10.1038/nrm1938](https://doi.org/10.1038/nrm1938) PMID: [16723974](https://pubmed.ncbi.nlm.nih.gov/16723974/)
27. Teif VB, Vainshtein Y, Caudron-Herger M, Mallm J-P, Marth C, et al. (2012) Genome-wide nucleosome positioning during embryonic stem cell development. *Nat. Struct. Mol. Biol.*, 19: 1185–1192. doi: [10.1038/nsmb.2419](https://doi.org/10.1038/nsmb.2419) PMID: [23085715](https://pubmed.ncbi.nlm.nih.gov/23085715/)
28. Zhu J, Adli M, Zou JY, Verstappen G, Coyne M, et al. (2013) Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell*, 152: 642–654. doi: [10.1016/j.cell.2012.12.033](https://doi.org/10.1016/j.cell.2012.12.033) PMID: [23333102](https://pubmed.ncbi.nlm.nih.gov/23333102/)
29. Cantone I and Fisher AG (2013) Epigenetic programming and reprogramming during development. *Nat. Struct. Mol. Biol.*, 20, 282–289. doi: [10.1038/nsmb.2489](https://doi.org/10.1038/nsmb.2489) PMID: [23463313](https://pubmed.ncbi.nlm.nih.gov/23463313/)
30. Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, et al. (2013) Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell*, 153: 1149–1163. doi: [10.1016/j.cell.2013.04.037](https://doi.org/10.1016/j.cell.2013.04.037) PMID: [23664763](https://pubmed.ncbi.nlm.nih.gov/23664763/)
31. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine, et al. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122: 947–956. doi: [10.1016/j.cell.2005.08.020](https://doi.org/10.1016/j.cell.2005.08.020) PMID: [16153702](https://pubmed.ncbi.nlm.nih.gov/16153702/)
32. Azuara V, Perry P, Sauer S, Spivakov M, Jørgensen HF, et al. (2006) Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.*, 8, 532–538. doi: [10.1038/ncb1403](https://doi.org/10.1038/ncb1403) PMID: [16570078](https://pubmed.ncbi.nlm.nih.gov/16570078/)
33. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, et al. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125: 315–326. doi: [10.1016/j.cell.2006.02.041](https://doi.org/10.1016/j.cell.2006.02.041) PMID: [16630819](https://pubmed.ncbi.nlm.nih.gov/16630819/)
34. Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, and Cavalli G. (2007) Genome regulation by polycomb and trithorax proteins. *Cell*, 128: 735–745. doi: [10.1016/j.cell.2007.02.009](https://doi.org/10.1016/j.cell.2007.02.009) PMID: [17320510](https://pubmed.ncbi.nlm.nih.gov/17320510/)
35. Margueron R, Justin N, Ohno K, Sharpe ML, Son, et al. (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*, 461: 762–767. doi: [10.1038/nature08398](https://doi.org/10.1038/nature08398) PMID: [19767730](https://pubmed.ncbi.nlm.nih.gov/19767730/)
36. Pirotta V and Li H-B (2012) A view of nuclear Polycomb bodies. *Curr. Opin. Genet. Dev.*, 22: 101–109. doi: [10.1016/j.gde.2011.11.004](https://doi.org/10.1016/j.gde.2011.11.004) PMID: [22178420](https://pubmed.ncbi.nlm.nih.gov/22178420/)

37. Schoenherr CJ and Anderson DJ (1995) The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science*, 267: 1360–1363. doi: [10.1126/science.7871435](https://doi.org/10.1126/science.7871435) PMID: [7871435](https://pubmed.ncbi.nlm.nih.gov/7871435/)
38. Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT, et al. (2006) Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell*, 10: 105–116. doi: [10.1016/j.devcel.2005.10.017](https://doi.org/10.1016/j.devcel.2005.10.017) PMID: [16399082](https://pubmed.ncbi.nlm.nih.gov/16399082/)
39. Prikrylová T, Pacherník J, Kozubek S, and Bártošová E. (2013) Epigenetics and chromatin plasticity in embryonic stem cells. *World J. Stem Cells*, 5, 73–85. doi: [10.4252/wjsc.v5.i3.73](https://doi.org/10.4252/wjsc.v5.i3.73) PMID: [23951389](https://pubmed.ncbi.nlm.nih.gov/23951389/)
40. Ruiz S, Panopoulos AD, Herreras A, Bissig K-D, Lutz M, et al. (2011) A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. *Curr. Biol.*, 21: 45–52. doi: [10.1016/j.cub.2010.11.049](https://doi.org/10.1016/j.cub.2010.11.049) PMID: [21167714](https://pubmed.ncbi.nlm.nih.gov/21167714/)
41. Gilbert DM (2001) Making sense of eukaryotic DNA replication origins. *Science*, 294: 96–100. doi: [10.1126/science.1061724](https://doi.org/10.1126/science.1061724) PMID: [11588251](https://pubmed.ncbi.nlm.nih.gov/11588251/)
42. Bell SP and Dutta A. (2002) DNA replication in eukaryotic cells. *Annu. Rev. Biochem.*, 71: 333–374. doi: [10.1146/annurev.biochem.71.110601.135425](https://doi.org/10.1146/annurev.biochem.71.110601.135425) PMID: [12045100](https://pubmed.ncbi.nlm.nih.gov/12045100/)
43. Méchali M. (2010) Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat. Rev. Mol. Cell Biol.*, 11: 728–738. doi: [10.1038/nrm2976](https://doi.org/10.1038/nrm2976) PMID: [20861881](https://pubmed.ncbi.nlm.nih.gov/20861881/)
44. Bogan JA, Natale DA, and Depamphilis ML (2000) Initiation of eukaryotic DNA replication: conservative or liberal?. *J. Cell. Physiol.*, 184: 139–150. doi: [10.1002/1097-4652\(200008\)184:2<139::AID-JCP1>3.0.CO;2-8](https://doi.org/10.1002/1097-4652(200008)184:2<139::AID-JCP1>3.0.CO;2-8) PMID: [10867638](https://pubmed.ncbi.nlm.nih.gov/10867638/)
45. McNairn AJ and Gilbert DM (2003) Epigenomic replication: linking epigenetics to DNA replication. *Bioessays*, 25: 647–656. doi: [10.1002/bies.10305](https://doi.org/10.1002/bies.10305) PMID: [12815720](https://pubmed.ncbi.nlm.nih.gov/12815720/)
46. Aladjem MI (2007) Replication in context: dynamic regulation of DNA replication patterns in metazoans. *Nat. Rev. Genet.*, 8, 588–600. doi: [10.1038/nrg2143](https://doi.org/10.1038/nrg2143) PMID: [17621316](https://pubmed.ncbi.nlm.nih.gov/17621316/)
47. Maric C and Prioleau M-N (2010) Interplay between DNA replication and gene expression: a harmonious coexistence. *Curr. Opin. Cell Biol.*, 22: 277–283. doi: [10.1016/j.ceb.2010.03.007](https://doi.org/10.1016/j.ceb.2010.03.007) PMID: [20363609](https://pubmed.ncbi.nlm.nih.gov/20363609/)
48. Berezney R, Dubey DD, and Huberman JA (2000) Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. *Chromosoma*, 108: 471–484. doi: [10.1007/s004120050399](https://doi.org/10.1007/s004120050399) PMID: [10794569](https://pubmed.ncbi.nlm.nih.gov/10794569/)
49. Buongiorno-Nardelli M, Micheli G, Carri MT, and Marilley M. (1982) A relationship between replicon size and supercoiled loop domains in the eukaryotic genome. *Nature*, 298: 100–102. doi: [10.1038/298100a0](https://doi.org/10.1038/298100a0) PMID: [7088157](https://pubmed.ncbi.nlm.nih.gov/7088157/)
50. Courbet S, Gay S, Arnoult N, Wronka G, Anglana M, et al. (2008) Replication fork movement sets chromatin loop size and origin choice in mammalian cells. *Nature*, 455: 557–560. doi: [10.1038/nature07233](https://doi.org/10.1038/nature07233) PMID: [18716622](https://pubmed.ncbi.nlm.nih.gov/18716622/)
51. Ryba T, Hiratani I, Sasaki T, Battaglia D, Kulik M, et al. (2011) Replication timing: a fingerprint for cell identity and pluripotency. *PLoS Comput. Biol.*, 7, e1002225. doi: [10.1371/journal.pcbi.1002225](https://doi.org/10.1371/journal.pcbi.1002225) PMID: [22028635](https://pubmed.ncbi.nlm.nih.gov/22028635/)
52. Hamlin JL, Mesner LD, and Dijkwel PA (2010) A winding road to origin discovery. *Chromosome Res.*, 18: 45–61. doi: [10.1007/s10577-009-9089-z](https://doi.org/10.1007/s10577-009-9089-z) PMID: [19859818](https://pubmed.ncbi.nlm.nih.gov/19859818/)
53. Gilbert DM (2010) Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat. Rev. Genet.*, 11: 673–684. doi: [10.1038/nrg2830](https://doi.org/10.1038/nrg2830) PMID: [20811343](https://pubmed.ncbi.nlm.nih.gov/20811343/)
54. Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, et al. (2001) Replication dynamics of the yeast genome. *Science*, 294: 115–121. doi: [10.1126/science.294.5540.115](https://doi.org/10.1126/science.294.5540.115) PMID: [11588253](https://pubmed.ncbi.nlm.nih.gov/11588253/)
55. MacAlpine DM, Rodriguez HK, and Bell SP (2004) Coordination of replication and transcription along a *Drosophila* chromosome. *Genes Dev.*, 18, 3094–3105. doi: [10.1101/gad.1246404](https://doi.org/10.1101/gad.1246404) PMID: [15601823](https://pubmed.ncbi.nlm.nih.gov/15601823/)
56. Farkash-Amar S, Lipson D, Polten A, Goren A, Helmstetter C, et al. (2008) Global organization of replication time zones of the mouse genome. *Genome Res.*, 18, 1562–1570. doi: [10.1101/gr.079566.108](https://doi.org/10.1101/gr.079566.108) PMID: [18669478](https://pubmed.ncbi.nlm.nih.gov/18669478/)
57. Hiratani I, Ryba T, Itoh M, Yokochi T, Schwaiger M, et al. (2008) Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.*, 6, e245. doi: [10.1371/journal.pbio.0060245](https://doi.org/10.1371/journal.pbio.0060245) PMID: [18842067](https://pubmed.ncbi.nlm.nih.gov/18842067/)
58. Hiratani I, Ryba T, Itoh M, Rathjen J, Kulik M, et al. (2010) Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. *Genome Res.*, 20: 155–169. doi: [10.1101/gr.099796.109](https://doi.org/10.1101/gr.099796.109) PMID: [19952138](https://pubmed.ncbi.nlm.nih.gov/19952138/)

59. Woodfine K, Fiegler H, Beare DM, Collins JE, McCann OT, et al. (2004) Replication timing of the human genome. *Hum. Mol. Genet.*, 13, 191–202. doi: [10.1093/hmg/ddh016](https://doi.org/10.1093/hmg/ddh016) PMID: [14645202](https://pubmed.ncbi.nlm.nih.gov/14645202/)
60. Desprat R, Thierry-Mieg D, Lailier N, Lajugie J, Schildkraut C, et al. (2009) Predictable dynamic program of timing of DNA replication in human cells. *Genome Res.*, 19: 2288–2299. doi: [10.1101/gr.094060.109](https://doi.org/10.1101/gr.094060.109) PMID: [19767418](https://pubmed.ncbi.nlm.nih.gov/19767418/)
61. Chen C-L, Rappailles A, Duquenne L, Huvet M, Guilbaud G, et al. (2010) Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome Res.*, 20: 447–457. doi: [10.1101/gr.098947.109](https://doi.org/10.1101/gr.098947.109) PMID: [20103589](https://pubmed.ncbi.nlm.nih.gov/20103589/)
62. Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, et al. (2010) Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl. Acad. Sci. USA*, 107: 139–144. doi: [10.1073/pnas.0912402107](https://doi.org/10.1073/pnas.0912402107) PMID: [19966280](https://pubmed.ncbi.nlm.nih.gov/19966280/)
63. Karnani N, Taylor C, Malhotra A, and Dutta A. (2007) Pan-S replication patterns and chromosomal domains defined by genome-tiling arrays of ENCODE genomic areas. *Genome Res.*, 17: 865–876. doi: [10.1101/gr.5427007](https://doi.org/10.1101/gr.5427007) PMID: [17568004](https://pubmed.ncbi.nlm.nih.gov/17568004/)
64. Ryba T, Battaglia D, Pope BD, Hiratani I, and Gilbert DM (2011) Genome-scale analysis of replication timing: from bench to bioinformatics. *Nat. Protoc.*, 6, 870–895. doi: [10.1038/nprot.2011.328](https://doi.org/10.1038/nprot.2011.328) PMID: [21637205](https://pubmed.ncbi.nlm.nih.gov/21637205/)
65. Schwaiger M, Stadler MB, Bell O, Kohler H, Oakeley EJ, et al. (2009) Chromatin state marks cell-type- and gender-specific replication of the *Drosophila* genome. *Genes Dev.*, 23, 589–601. doi: [10.1101/gad.511809](https://doi.org/10.1101/gad.511809) PMID: [19270159](https://pubmed.ncbi.nlm.nih.gov/19270159/)
66. Fillion GJ, van Bommel JG, Braunschweig U, Talhout W, Kind J, et al. (2010) Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell*, 143, 212–224. doi: [10.1016/j.cell.2010.09.009](https://doi.org/10.1016/j.cell.2010.09.009) PMID: [20888037](https://pubmed.ncbi.nlm.nih.gov/20888037/)
67. Farkash-Amar S and Simon I. (2010) Genome-wide analysis of the replication program in mammals. *Chromosome Res.*, 18, 115–125. doi: [10.1007/s10577-009-9091-5](https://doi.org/10.1007/s10577-009-9091-5) PMID: [20205353](https://pubmed.ncbi.nlm.nih.gov/20205353/)
68. Ryba T, Hiratani I, Lu J, Itoh M, Kulik, et al. (2010) Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.*, 20: 761–770. doi: [10.1101/gr.099655.109](https://doi.org/10.1101/gr.099655.109) PMID: [20430782](https://pubmed.ncbi.nlm.nih.gov/20430782/)
69. Zhou J, Ermakova OV, Riblet R, Birshtein BK, and Schildkraut CL (2002) Replication and subnuclear location dynamics of the immunoglobulin heavy-chain locus in B-lineage cells. *Mol. Cell Biol.*, 22: 4876–4889. doi: [10.1128/MCB.22.13.4876-4889.2002](https://doi.org/10.1128/MCB.22.13.4876-4889.2002) PMID: [12052893](https://pubmed.ncbi.nlm.nih.gov/12052893/)
70. Takebayashi S, Dileep V, Ryba T, Dennis JH, and Gilbert DM (2012) Chromatin-interaction compartment switch at developmentally regulated chromosomal domains reveals an unusual principle of chromatin folding. *Proc. Natl. Acad. Sci. U.S.A.*, 109: 12574–12579. doi: [10.1073/pnas.1207185109](https://doi.org/10.1073/pnas.1207185109) PMID: [22807480](https://pubmed.ncbi.nlm.nih.gov/22807480/)
71. Zink D, Bornfleth H, Visser A, Cremer C, and Cremer T. (1999) Organization of early and late replicating DNA in human chromosome territories. *Exp. Cell Res.*, 247: 176–188. doi: [10.1006/excr.1998.4311](https://doi.org/10.1006/excr.1998.4311) PMID: [10047460](https://pubmed.ncbi.nlm.nih.gov/10047460/)
72. Grasser F, Neusser M, Fiegler H, Thormeyer T, Cremer M, et al. (2008) Replication-timing-correlated spatial chromatin arrangements in cancer and in primate interphase nuclei. *J. Cell Sci.*, 121: 1876–1886. doi: [10.1242/jcs.026989](https://doi.org/10.1242/jcs.026989) PMID: [18477608](https://pubmed.ncbi.nlm.nih.gov/18477608/)
73. Lieberman-Aiden E, vanBerkum NL, Williams L, Imakaev M, Ragoczy T, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326: 289–293. doi: [10.1126/science.1181369](https://doi.org/10.1126/science.1181369) PMID: [19815776](https://pubmed.ncbi.nlm.nih.gov/19815776/)
74. Yaffe E, Farkash-Amar S, Polten A, Yakhini Z, Tanay A, et al. (2010) Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. *PLoS Genet.*, 6, e1001011. doi: [10.1371/journal.pgen.1001011](https://doi.org/10.1371/journal.pgen.1001011) PMID: [20617169](https://pubmed.ncbi.nlm.nih.gov/20617169/)
75. Moindrot B, Audit B, Klous P, Baker A, Thermes C, et al. (2012) 3D chromatin conformation correlates with replication timing and is conserved in resting cells. *Nucleic Acids Res.*, 40: 9470–9481. doi: [10.1093/nar/gks736](https://doi.org/10.1093/nar/gks736) PMID: [22879376](https://pubmed.ncbi.nlm.nih.gov/22879376/)
76. Guilbaud G, Rappailles A, Baker A, Chen C-L, Arneodo A, et al. (2011) Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. *PLoS Comput. Biol.*, 7, e1002322. doi: [10.1371/journal.pcbi.1002322](https://doi.org/10.1371/journal.pcbi.1002322) PMID: [22219720](https://pubmed.ncbi.nlm.nih.gov/22219720/)
77. Baker A, Audit B, Chen C-L, Moindrot B, Leleu A, et al. (2012) Replication fork polarity gradients revealed by megabase-sized U-shaped replication timing domains in human cell lines. *PLoS Comput. Biol.*, 8, e1002443. doi: [10.1371/journal.pcbi.1002443](https://doi.org/10.1371/journal.pcbi.1002443) PMID: [22496629](https://pubmed.ncbi.nlm.nih.gov/22496629/)
78. Audit B, Baker A, Chen C-L, Rappailles A, Guilbaud G, et al. (2013) Multiscale analysis of genome-wide replication timing profiles using a wavelet-based signal-processing algorithm. *Nat. Protoc.*, 8, 98–110. doi: [10.1038/nprot.2012.145](https://doi.org/10.1038/nprot.2012.145) PMID: [23237832](https://pubmed.ncbi.nlm.nih.gov/23237832/)

79. Brodie of Brodie E.-B, Nicolay S, Touchon M, Audit B, d'Aubenton-Carafa, et al. (2005) From DNA sequence analysis to modeling replication in the human genome. *Phys. Rev. Lett.*, 94, 248103. doi: [10.1103/PhysRevLett.94.248103](https://doi.org/10.1103/PhysRevLett.94.248103) PMID: [16090582](https://pubmed.ncbi.nlm.nih.gov/16090582/)
80. Touchon M, Nicolay S, Audit B, Brodie of Brodie E.-B, d'Aubenton-Carafa Y, et al. (2005) Replication-associated strand asymmetries in mammalian genomes: toward detection of replication origins. *Proc. Natl. Acad. Sci. USA*, 102: 9836–9841. doi: [10.1073/pnas.0500577102](https://doi.org/10.1073/pnas.0500577102) PMID: [15985556](https://pubmed.ncbi.nlm.nih.gov/15985556/)
81. Huvet M, Nicolay S, Touchon M, Audit B, d'Aubenton-Carafa Y, et al. (2007) Human gene organization driven by the coordination of replication and transcription. *Genome Res.*, 17, 1278–1285. doi: [10.1101/gr.6533407](https://doi.org/10.1101/gr.6533407) PMID: [17675363](https://pubmed.ncbi.nlm.nih.gov/17675363/)
82. Baker A, Nicolay S, Zaghoul L, d'Aubenton-Carafa Y, Thermes C, et al. (2010) Wavelet-based method to disentangle transcription- and replication-associated strand asymmetries in mammalian genomes. *Appl. Comput. Harmon. Anal.*, 28, 150–170. doi: [10.1016/j.acha.2009.10.003](https://doi.org/10.1016/j.acha.2009.10.003)
83. Chen C-L, Duquenne L, Audit B, Guilbaud G, Rappailles A, et al. (2011) Replication-associated mutational asymmetry in the human genome. *Mol. Biol. Evol.*, 28, 2327–2337. doi: [10.1093/molbev/msr056](https://doi.org/10.1093/molbev/msr056) PMID: [21368316](https://pubmed.ncbi.nlm.nih.gov/21368316/)
84. Baker A, Julienne H, Chen CL, Audit B, d'Aubenton-Carafa Y, et al. (2012) Linking the DNA strand asymmetry to the spatio-temporal replication program. I. About the role of the replication fork polarity in genome evolution. *Eur. Phys. J. E*, 35, 92. doi: [10.1140/epje/i2012-12092-y](https://doi.org/10.1140/epje/i2012-12092-y) PMID: [23001787](https://pubmed.ncbi.nlm.nih.gov/23001787/)
85. Baker A, Chen CL, Julienne H, Audit B, d'Aubenton-Carafa Y, et al. (2012) Linking the DNA strand asymmetry to the spatio-temporal replication program: II. Accounting for neighbor-dependent substitution rates. *Eur. Phys. J. E*, 35, 123. doi: [10.1140/epje/i2012-12123-9](https://doi.org/10.1140/epje/i2012-12123-9) PMID: [23179013](https://pubmed.ncbi.nlm.nih.gov/23179013/)
86. Audit B, Nicolay S, Huvet M, Touchon M, d'Aubenton-Carafa Y, et al. (2007) DNA replication timing data corroborate in silico human replication origin predictions. *Phys. Rev. Lett.*, 99, 248102. doi: [10.1103/PhysRevLett.99.248102](https://doi.org/10.1103/PhysRevLett.99.248102) PMID: [18233493](https://pubmed.ncbi.nlm.nih.gov/18233493/)
87. Audit B, Zaghoul L, Vaillant C, Chevereau G, d'Aubenton-Carafa Y, et al. (2009) Open chromatin encoded in DNA sequence is the signature of “master” replication origins in human cells. *Nucleic Acids Res.*, 37, 6064–6075. doi: [10.1093/nar/gkp631](https://doi.org/10.1093/nar/gkp631) PMID: [19671527](https://pubmed.ncbi.nlm.nih.gov/19671527/)
88. Zaghoul L, Baker A, Audit B, and Arneodo A (2012) Gene organization inside replication domains in mammalian genomes. *C. R. Mécanique*, 340: 745–757. doi: [10.1016/j.crme.2012.10.023](https://doi.org/10.1016/j.crme.2012.10.023)
89. Hyrien O, Rappailles A, Guilbaud G, Baker A, Chen C-L, et al. (2013) From simple bacterial and archaeal replicons to replication N/U-domains. *J. Mol. Biol.*, 425: 4673–4689. doi: [10.1016/j.jmb.2013.09.021](https://doi.org/10.1016/j.jmb.2013.09.021) PMID: [24095859](https://pubmed.ncbi.nlm.nih.gov/24095859/)
90. Boulos R, Arneodo A, Jensen P, and Audit B. (2013) Revealing long-range interconnected hubs in human chromatin interaction data using graph theory. *Phys. Rev. Lett.*, 111, 118102. doi: [10.1103/PhysRevLett.111.118102](https://doi.org/10.1103/PhysRevLett.111.118102) PMID: [24074120](https://pubmed.ncbi.nlm.nih.gov/24074120/)
91. Lemaitre C, Zaghoul L, Sagot M-F, Gautier C, Arneodo A, et al. (2009) Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. *BMC Genomics*, 10: 335. doi: [10.1186/1471-2164-10-335](https://doi.org/10.1186/1471-2164-10-335) PMID: [19630943](https://pubmed.ncbi.nlm.nih.gov/19630943/)
92. Julienne H, Zoufir A, Audit B, and Arneodo A. (2013) Human genome replication proceeds through four chromatin states. *PLoS Comput. Biol.*, 9, e1003233. doi: [10.1371/journal.pcbi.1003233](https://doi.org/10.1371/journal.pcbi.1003233) PMID: [24130466](https://pubmed.ncbi.nlm.nih.gov/24130466/)
93. Izenman AJ (2008) *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*, Springer, New York.
94. Kaufman L and Rousseeuw PJ (1984) *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, New York.
95. Roudier F, Ahmed I, Bérard C, Sarazin A, Mary-Huard T, et al. (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J.*, 30: 1928–1938. doi: [10.1038/emboj.2011.103](https://doi.org/10.1038/emboj.2011.103) PMID: [21487388](https://pubmed.ncbi.nlm.nih.gov/21487388/)
96. Liu T, Rechtsteiner A, Egelhofer TA, Vielle A, Latorre I, et al. (2011) Broad chromosomal domains of histone modification patterns in *C. elegans*. *Genome Res.*, 21: 227–236. doi: [10.1101/gr.115519.110](https://doi.org/10.1101/gr.115519.110) PMID: [21177964](https://pubmed.ncbi.nlm.nih.gov/21177964/)
97. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, et al. (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*, 148: 458–472. doi: [10.1016/j.cell.2012.01.010](https://doi.org/10.1016/j.cell.2012.01.010) PMID: [22265598](https://pubmed.ncbi.nlm.nih.gov/22265598/)
98. Goldar A, Marsolier-Kergoat M-C, and Hyrien O. (2009) Universal temporal profile of replication origin activation in eukaryotes. *PLoS One*, 4, e5899. doi: [10.1371/journal.pone.0005899](https://doi.org/10.1371/journal.pone.0005899) PMID: [19521533](https://pubmed.ncbi.nlm.nih.gov/19521533/)
99. Julienne H, Zoufir A, Audit B, and Arneodo A. (2013) Epigenetic regulation of the human genome: coherence between promoter activity and large-scale chromatin environment. *Front. Life Sci.*, 7, 44–62. doi: [10.1080/21553769.2013.832706](https://doi.org/10.1080/21553769.2013.832706)

100. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, et al. (2012) ChIP-seq guidelines and practices of the encode and modencode consortia. *Genome Res.*, 22: 1813–1831. doi: [10.1101/gr.136184.111](https://doi.org/10.1101/gr.136184.111) PMID: [22955991](https://pubmed.ncbi.nlm.nih.gov/22955991/)
101. Chessel D, Dufour A, and Thioulouse J. (2004) The ade4 package -I- One-table methods. *R News*, 4, 5–10.
102. Rousseeuw P. (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, 20: 53–65. doi: [10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
103. Mortazavi A, Williams B, McCue K, Schaeffer L, and Wold B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods*, 5, 621–628. doi: [10.1038/nmeth.1226](https://doi.org/10.1038/nmeth.1226) PMID: [18516045](https://pubmed.ncbi.nlm.nih.gov/18516045/)
104. Vaillant C, Audit B, and Arneodo A. (2007) Experiments confirm the influence of genome long-range correlations on nucleosome positioning. *Phys. Rev. Lett.*, 99, 218103. doi: [10.1103/PhysRevLett.99.218103](https://doi.org/10.1103/PhysRevLett.99.218103) PMID: [18233262](https://pubmed.ncbi.nlm.nih.gov/18233262/)
105. Chevereau G, Palmeira L, Thermes C, Arneodo A, and Vaillant C. (2009) Thermodynamics of intragenic nucleosome ordering. *Phys. Rev. Lett.*, 103: 188103. doi: [10.1103/PhysRevLett.103.188103](https://doi.org/10.1103/PhysRevLett.103.188103) PMID: [19905836](https://pubmed.ncbi.nlm.nih.gov/19905836/)
106. Milani P, Chevereau G, Vaillant C, Audit B, Haftek-Terreau Z, et al. (2009) Nucleosome positioning by genomic excluding-energy barriers. *Proc. Natl. Acad. Sci. USA*, 106: 22257–22262. doi: [10.1073/pnas.0909511106](https://doi.org/10.1073/pnas.0909511106) PMID: [20018700](https://pubmed.ncbi.nlm.nih.gov/20018700/)
107. Vaillant C, Palmeira L, Chevereau G, Audit B, d'Aubenton-Carafa Y, et al. (2010) A novel strategy of transcription regulation by intragenic nucleosome ordering. *Genome Res.*, 20: 59–67. doi: [10.1101/gr.096644.109](https://doi.org/10.1101/gr.096644.109) PMID: [19858362](https://pubmed.ncbi.nlm.nih.gov/19858362/)
108. Chevereau G, Arneodo A, and Vaillant C. (2011) Influence of the genomic sequence on the primary structure of chromatin. *Front. Life Sci.*, 5, 29–68. doi: [10.1080/21553769.2012.708882](https://doi.org/10.1080/21553769.2012.708882)
109. Chandra T, Kirschner K, Thuret J-Y, Pope BD, Ryba T, et al. (2012) Independence of repressive histone marks and chromatin compaction during senescent heterochromatic layer formation. *Mol. Cell*, 47, 203–214. doi: [10.1016/j.molcel.2012.06.010](https://doi.org/10.1016/j.molcel.2012.06.010) PMID: [22795131](https://pubmed.ncbi.nlm.nih.gov/22795131/)
110. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, et al. (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, 128: 1231–1245. doi: [10.1016/j.cell.2006.12.048](https://doi.org/10.1016/j.cell.2006.12.048) PMID: [17382889](https://pubmed.ncbi.nlm.nih.gov/17382889/)
111. Phillips JE and Corces VG (2009) CTCF: master weaver of the genome. *Cell*, 137: 1194–1211. doi: [10.1016/j.cell.2009.06.001](https://doi.org/10.1016/j.cell.2009.06.001) PMID: [19563753](https://pubmed.ncbi.nlm.nih.gov/19563753/)
112. Simon JA and Kingston RE (2009) Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.*, 10: 697–708. doi: [10.1038/nrm2763](https://doi.org/10.1038/nrm2763) PMID: [19738629](https://pubmed.ncbi.nlm.nih.gov/19738629/)
113. Merckenschlager M and Odom DT (2013) CTCF and cohesin: linking gene regulatory elements with their targets. *Cell*, 152: 1285–1297. doi: [10.1016/j.cell.2013.02.029](https://doi.org/10.1016/j.cell.2013.02.029) PMID: [23498937](https://pubmed.ncbi.nlm.nih.gov/23498937/)
114. Tardat M, Brustel J, Kirsh O, Lefevbre C, Callanan M, et al. (2010) The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. *Nat. Cell Biol.*, 12: 1086–1093. doi: [10.1038/ncb2113](https://doi.org/10.1038/ncb2113) PMID: [20953199](https://pubmed.ncbi.nlm.nih.gov/20953199/)
115. Mizuguchi G, Shen X, Landry J, Wu W-H, Sen S, et al. (2004) ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science*, 303: 343–348. doi: [10.1126/science.1090701](https://doi.org/10.1126/science.1090701) PMID: [14645854](https://pubmed.ncbi.nlm.nih.gov/14645854/)
116. Talbert PB and Henikoff S. (2010) Histone variants-ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.*, 11: 264–275. doi: [10.1038/nrm2861](https://doi.org/10.1038/nrm2861) PMID: [20197778](https://pubmed.ncbi.nlm.nih.gov/20197778/)
117. Hou C, Dale R, and Dean A. (2010) Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. USA*, 107: 3651–3656. doi: [10.1073/pnas.0912087107](https://doi.org/10.1073/pnas.0912087107) PMID: [20133600](https://pubmed.ncbi.nlm.nih.gov/20133600/)
118. Ohlsson R, Lobanenkov V, and Klenova E. (2010) Does CTCF mediate between nuclear organization and gene expression?. *Bioessays*, 32: 37–50. doi: [10.1002/bies.200900118](https://doi.org/10.1002/bies.200900118) PMID: [20020479](https://pubmed.ncbi.nlm.nih.gov/20020479/)
119. Botta M, Haider S, Leung IX Y, Lio P, et al. (2010) Intra- and inter-chromosomal interactions correlate with CTCF binding genome wide. *Mol. Syst. Biol.*, 6, 426. doi: [10.1038/msb.2010.79](https://doi.org/10.1038/msb.2010.79) PMID: [21045820](https://pubmed.ncbi.nlm.nih.gov/21045820/)
120. Handoko L, Xu H, Li G, Ngan CY, Chew E, et al. (2011) CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.*, 43: 630–638. doi: [10.1038/ng.857](https://doi.org/10.1038/ng.857) PMID: [21685913](https://pubmed.ncbi.nlm.nih.gov/21685913/)
121. Gaspar-Maia A, Alajem A, Meshorer E, and Ramalho-Santos M. (2011) Open chromatin in pluripotency and reprogramming. *Nat. Rev. Mol. Cell Biol.*, 12: 36–47. doi: [10.1038/nrm3036](https://doi.org/10.1038/nrm3036) PMID: [21179060](https://pubmed.ncbi.nlm.nih.gov/21179060/)

122. Gaspar-Maia A, Alajem A, Polesso F, Sridharan R, Mason MJ, et al. (2009) CHD1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature*, 460: 863–868. doi: [10.1038/nature08212](https://doi.org/10.1038/nature08212) PMID: [19587682](https://pubmed.ncbi.nlm.nih.gov/19587682/)
123. Ho L and Crabtree, GR (2010) Chromatin remodelling during development. *Nature*, 463: 474–484. doi: [10.1038/nature08911](https://doi.org/10.1038/nature08911) PMID: [20110991](https://pubmed.ncbi.nlm.nih.gov/20110991/)
124. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, et al. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 131: 861–872. doi: [10.1016/j.cell.2007.11.019](https://doi.org/10.1016/j.cell.2007.11.019) PMID: [18035408](https://pubmed.ncbi.nlm.nih.gov/18035408/)
125. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, et al. (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science*, 318: 1917–1920. doi: [10.1126/science.1151526](https://doi.org/10.1126/science.1151526) PMID: [18029452](https://pubmed.ncbi.nlm.nih.gov/18029452/)
126. Jaenisch R and Young R. (2008) Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell*, 132: 567–582. doi: [10.1016/j.cell.2008.01.015](https://doi.org/10.1016/j.cell.2008.01.015) PMID: [18295576](https://pubmed.ncbi.nlm.nih.gov/18295576/)
127. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448: 553–560. doi: [10.1038/nature06008](https://doi.org/10.1038/nature06008) PMID: [17603471](https://pubmed.ncbi.nlm.nih.gov/17603471/)
128. Hiratani I, Takebayashi S, Lu J, and Gilbert DM (2009) Replication timing and transcriptional control: beyond cause and effect part II. *Curr. Opin. Genet. Dev.*, 19, 142–149. doi: [10.1016/j.gde.2009.02.002](https://doi.org/10.1016/j.gde.2009.02.002) PMID: [19345088](https://pubmed.ncbi.nlm.nih.gov/19345088/)
129. Thurman RE, Day N, Noble WS, and Stamatoyannopoulos JA (2007) Identification of higher-order functional domains in the human ENCODE regions. *Genome Res.*, 17: 917–927. doi: [10.1101/gr.6081407](https://doi.org/10.1101/gr.6081407) PMID: [17568007](https://pubmed.ncbi.nlm.nih.gov/17568007/)
130. Efroni S, Dutttagupta R, Cheng J, Dehghani H, Hoepfner DJ, et al. (2008) Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell*, 2, 437–447. doi: [10.1016/j.stem.2008.03.021](https://doi.org/10.1016/j.stem.2008.03.021) PMID: [18462694](https://pubmed.ncbi.nlm.nih.gov/18462694/)
131. Cadoret J-C, Meisch F, Hassan-Zadeh V, Luyten I, Guillet C, et al. (2008) Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc. Natl. Acad. Sci. USA*, 105: 15837–15842. doi: [10.1073/pnas.0805208105](https://doi.org/10.1073/pnas.0805208105) PMID: [18838675](https://pubmed.ncbi.nlm.nih.gov/18838675/)
132. Sequeira-Mendes J, Diaz-Uriarte R, Apedaile A, Huntley D, Brockdorff N, et al. (2009) Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet.*, 5, e1000446. doi: [10.1371/journal.pgen.1000446](https://doi.org/10.1371/journal.pgen.1000446) PMID: [19360092](https://pubmed.ncbi.nlm.nih.gov/19360092/)
133. Karnani N, Taylor CM, Malhotra A, and Dutta A. (2010) Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Mol. Biol. Cell*, 21: 393–404. doi: [10.1091/mbc.E09-08-0707](https://doi.org/10.1091/mbc.E09-08-0707) PMID: [19955211](https://pubmed.ncbi.nlm.nih.gov/19955211/)
134. Cayrou C, Coulombe P, Vigneron A, Stanojic S, Ganier O, et al. (2011) Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res.*, 21: 1438–1449. doi: [10.1101/gr.121830.111](https://doi.org/10.1101/gr.121830.111) PMID: [21750104](https://pubmed.ncbi.nlm.nih.gov/21750104/)
135. Valenzuela MS, Chen Y, Davis S, Yang F, Walker RL, et al. (2011) Preferential localization of human origins of DNA replication at the 5'-ends of expressed genes and at evolutionarily conserved DNA sequences. *PLoS One*, 6, e17308. doi: [10.1371/journal.pone.0017308](https://doi.org/10.1371/journal.pone.0017308) PMID: [21602917](https://pubmed.ncbi.nlm.nih.gov/21602917/)
136. Besnard E, Babled A, Lapasset L, Milhavel O, Parrinello H, et al. (2012) Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.*, 19, 837–844. doi: [10.1038/nsmb.2339](https://doi.org/10.1038/nsmb.2339) PMID: [22751019](https://pubmed.ncbi.nlm.nih.gov/22751019/)
137. Cayrou C, Coulombe P, Puy A, Rialle S, Kaplan N, et al. (2012) New insights into replication origin characteristics in metazoans. *Cell Cycle*, 11: 658–667. doi: [10.4161/cc.11.4.19097](https://doi.org/10.4161/cc.11.4.19097) PMID: [22373526](https://pubmed.ncbi.nlm.nih.gov/22373526/)
138. Méchali M, Yoshida K, Coulombe P, and Pasero P. (2013) Genetic and epigenetic determinants of DNA replication origins, position and activation. *Curr. Opin. Genet. Dev.*, 23, 124–131. doi: [10.1016/j.gde.2013.02.010](https://doi.org/10.1016/j.gde.2013.02.010) PMID: [23541525](https://pubmed.ncbi.nlm.nih.gov/23541525/)
139. Picard F, Cadoret J-C, Audit B, Arneodo A, Alberti A, et al. (2014) The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. *PLoS Genet.*, 10: e1004282. doi: [10.1371/journal.pgen.1004282](https://doi.org/10.1371/journal.pgen.1004282) PMID: [24785686](https://pubmed.ncbi.nlm.nih.gov/24785686/)
140. deMoura APS, Retkute R, Hawkins M, and Nieduszynski CA (2010) Mathematical modelling of whole chromosome replication. *Nucleic Acids Res.*, 38, 5623–5633. doi: [10.1093/nar/gkq343](https://doi.org/10.1093/nar/gkq343)
141. Yang SC-H, Rhind N, and Bechhoefer J. (2010) Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Mol. Syst. Biol.*, 6, 404. doi: [10.1038/msb.2010.61](https://doi.org/10.1038/msb.2010.61) PMID: [20739926](https://pubmed.ncbi.nlm.nih.gov/20739926/)
142. Jones DO, Cowell IG, and Singh PB (2000) Mammalian chromodomain proteins: their role in genome organisation and expression. *Bioessays*, 22: 124–137. doi: [10.1002/\(SICI\)1521-1878\(200002\)22:2<124::AID-BIES4>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1521-1878(200002)22:2<124::AID-BIES4>3.0.CO;2-E) PMID: [10655032](https://pubmed.ncbi.nlm.nih.gov/10655032/)

143. Nielsen AL, Oulad-Abdelghani M, Ortiz JA, Remboutsika E, Chambon P, et al. (2001) Heterochromatin formation in mammalian cells: interaction between histones and HP1 proteins. *Mol. Cell*, 7, 729–739. doi: [10.1016/S1097-2765\(01\)00218-0](https://doi.org/10.1016/S1097-2765(01)00218-0) PMID: [11336697](https://pubmed.ncbi.nlm.nih.gov/11336697/)
144. Tamaru H. and Selker EU (2001) A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature*, 414: 277–283. doi: [10.1038/35104508](https://doi.org/10.1038/35104508) PMID: [11713521](https://pubmed.ncbi.nlm.nih.gov/11713521/)
145. Fuks F, Hurd PJ, Deplus R, and Kouzarides T. (2003) The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res.*, 31, 2305–2312. doi: [10.1093/nar/gkg332](https://doi.org/10.1093/nar/gkg332) PMID: [12711675](https://pubmed.ncbi.nlm.nih.gov/12711675/)
146. Faust C, Lawson KA, Schork NJ, Thiel B, and Magnuson T. (1998) The Polycomb-group gene *eed* is required for normal morphogenetic movements during gastrulation in the mouse embryo. *Development*, 125: 4495–4506. PMID: [9778508](https://pubmed.ncbi.nlm.nih.gov/9778508/)
147. O'Carroll D, Erhardt S, Pagani M, Barton SC, Surani MA, et al. (2001) The polycomb-group gene *ezh2* is required for early mouse development. *Mol. Cell Biol.*, 21: 4330–4336. doi: [10.1128/MCB.21.13.4330-4336.2001](https://doi.org/10.1128/MCB.21.13.4330-4336.2001) PMID: [11390661](https://pubmed.ncbi.nlm.nih.gov/11390661/)
148. Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, et al. (2001) Loss of the *suv39h* histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*, 107: 323–337. doi: [10.1016/S0092-8674\(01\)00542-6](https://doi.org/10.1016/S0092-8674(01)00542-6) PMID: [11701123](https://pubmed.ncbi.nlm.nih.gov/11701123/)
149. Tachibana M, Sugimoto K, Nozaki M, Ueda J, Ohta T, et al. (2002) G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.*, 16, 1779–1791. doi: [10.1101/gad.989402](https://doi.org/10.1101/gad.989402) PMID: [12130538](https://pubmed.ncbi.nlm.nih.gov/12130538/)
150. Dodge JE, Kang Y-K, Beppu H, Lei H, and Li E. (2004) Histone H3-K9 methyltransferase ESET is essential for early development. *Mol. Cell Biol.*, 24, 2478–2486. doi: [10.1128/MCB.24.6.2478-2486.2004](https://doi.org/10.1128/MCB.24.6.2478-2486.2004) PMID: [14993285](https://pubmed.ncbi.nlm.nih.gov/14993285/)
151. Pasini D, Bracken AP, Hansen JB, Capillo M, and Helin K. (2007) The polycomb group protein *suz12* is required for embryonic stem cell differentiation. *Mol. Cell Biol.*, 27, 3769–3779. doi: [10.1128/MCB.01432-06](https://doi.org/10.1128/MCB.01432-06) PMID: [17339329](https://pubmed.ncbi.nlm.nih.gov/17339329/)
152. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, et al. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485: 376–380. doi: [10.1038/nature11082](https://doi.org/10.1038/nature11082) PMID: [22495300](https://pubmed.ncbi.nlm.nih.gov/22495300/)