# Diffusion controls local versus dispersed inheritance of histones during replication and shapes epigenomic architecture

Archit Singh[1] and Shaon Chakrabarti[1,*]

[1] Simons Centre for the Study of Living Machines,
National Centre for Biological Sciences,
Tata Institute of Fundamental Research,
Bangalore 560065, India.
[*]Author for correspondence: shaon@ncbs.res.in

**Abstract:**
The dynamics of inheritance of histones and their associated modifications across cell divisions can have major consequences on maintenance of the cellular epigenomic state. Recent experiments contradict the long-held notion that histone inheritance during replication is always local, suggesting that active and repressed regions of the genome exhibit fundamentally different histone dynamics independent of transcription-coupled turnover. Here we develop a stochastic model of histone dynamics at the replication fork and demonstrate that differential diffusivity of histones in active versus repressed chromatin is sufficient to quantitatively explain these recent experiments. Further, we use the model to predict patterns in histone mark similarity between pairs of genomic loci that should be developed as a result of diffusion, but not from PRC2 mediated mark spreading or transcriptional processes. Interestingly, using a combination of CHIP-seq, replication timing and Hi-C datasets we demonstrate that all the computationally predicted patterns are consistently observed for both active and repressive histone marks in two different cell lines. Our results suggest that in contrast to current models that posit histone transfer exclusively between parental and daughter DNA strands, dislodged histones in euchromatin and facultative heterochromatin diffuse within larger "Diffusion-Accessible-Domains" (DADs), leading to redistribution of epigenetic marks both within and across chromosomes. Preservation of the epigenomic state across cell divisions therefore might be achieved not by passing on strict positional information of histone marks, but by maintaining the marks in somewhat larger DADs of the genome.

**Introduction:**
Differences in cellular identities are thought to be maintained by varied gene expression programs, which are in turn controlled by a complex interplay of modifications on DNA and histones[1,2]. These modifications are key determinants of chromatin architecture in the nucleus, which repress or facilitate gene expression leading to cell-state specification. Depending on the context, these cellular states might need to be plastic as in stem cells during development, maintained faithfully over time such as in adult cells or even exhibit periodic switching, for example in a variety of cancer cell types[1]. How the interplay between forces of regulatory plasticity and stable inheritability is achieved remains an intriguing unsolved question in development and disease.

In dividing cells, DNA replication provides a potential platform for these varied forces to play out. Cellular replication over multiple generations presents a significant challenge to the maintenance of chromatin architecture and DNA/histone modification patterns[2]. The nucleosomal structures ahead of the replication fork need to be dismantled and eventually distributed between the newly created DNA strands. Whether and how the parental distributions of histones along with their accompanying post-translational modifications (PTMs) get re-established in the daughter cells, remains incompletely understood even after decades of research[2–7]. In particular, the question of how histones get dislodged, reassembled and repositioned has a rich history[2,3]. A number of studies have suggested that the histone transfer process is local (within 200-400 bp of the replication fork, along the same chromosome)[5,8–10], is usually symmetric[11–13], and that "epigenetic memory" in mammalian cells is likely maintained by the stable tetramer formed by H3 (H3.1 or H3.2 replicative variants) and H4 components of the histone complex[14–20]. The non-replicative H3.3 variant is deposited at genomic regions of high turnover such as actively transcribed sites[20], hence is unlikely to play a role in epigenetic inheritance. More recently, the question of how the histone PTMs get transferred and maintained during replication has received a lot of attention[21–27], with reports suggesting that positional information of the marks is accurately transmitted irrespective of the active or repressed nature of the chromatin domain[26].

A recent study[28] however challenged the general consensus on accurate and local transfer of histone-PTM positional information, suggesting that inter-generational inheritance is local only in repressed genomic regions (heterochromatin), while wide-spread dispersion of histones in active regions (euchromatin) might prevent faithful epigenetic information transfer. Following biotin-tagged H3.1 and H3.2 histones over time in specific active and repressed loci of mouse embryonic stem cells, this work demonstrated significantly faster kinetics of histone dilution in active as compared to repressed regions. Importantly, this difference did not arise from transcription-coupled histone turnover[28]. While a number of potential mechanisms have been suggested such as differences in histone chaperones at the replication fork[2,29], differential replication rates[28] or concentrations of newly synthesized histones[30], it remains unclear which one of these mechanisms, if any, can fully explain the distinct histone dynamics in active versus repressed regions. Important to note is that the new model of "dispersed" histones still allows for histone re-deposition only on the two daughter DNA strands on the same chromosome[2,28], further from the replication fork compared to current estimates of 200-400 bp.

Here we provide a new hypothesis, and then evidence from stochastic models coupled with bioinformatics analyses of available datasets, to demonstrate that the physical phenomenon of diffusion is sufficient to explain these recent experimental results. Furthermore, we find evidence for diffusion-driven histone mark dispersion *across* different chromosomes in euchromatin and facultative heterochromatin, therefore not limiting histone re-deposition to just the daughter DNA strands. The basic hypothesis is based upon observations that many proteins often diffuse in the nucleus before binding to DNA[31], and that the diffusivity of such proteins, including core histone components, vary depending on local chromatin architecture[32,33]. Additionally, chromatin itself exhibits diffusive motion[34] with increased mobility in active domains[35–37]. We suggest that the increased protein and DNA mobility in active chromatin over repressed chromatin is sufficient to explain the dispersed vs local histone inheritance recently observed[28].

Our hypothesis is distinct from all previous models of histone inheritance at the replication fork (both experimental[2,3,6] and theoretical[10,38–40]) since we suggest that parental histone marks do not necessarily land on the leading and lagging daughter strands only, but via diffusion land also on *any* DNA segment that is spatially proximal and is being replicated at the same time. To make this hypothesis testable, we develop a lattice model of cell division with diffusive spatial dynamics of histone-PTM complexes, and demonstrate that allowing for higher diffusive dynamics in active regions of DNA can quantitatively explain the histone dilution data reported by Escobar et al[28]. Furthermore, we use our computational model to predict three distinct patterns of histone-PTM similarity between pairs of DNA segments within chromosome compartments, that are expected to develop as a consequence of differential diffusivity. Remarkably, we show that the computationally predicted patterns can indeed be observed consistently in two different cell lines (GM06990 and K562) after combining HiC, replication timing and CHIP-seq information. Furthermore, the patterns can be seen both in *intra* as well as *inter*-chromosomal loci pairs, and we argue why such patterns are unlikely to be generated by either PRC2 mediated histone mark spreading or transcriptional processes. Our model therefore suggests that in contrast to the current descriptions of histone transfer between parental and daughter DNA strands, dislodged histones along with their modifications diffuse within "Diffusion-Accessible-Domains" (DADs), leading to redistribution of epigenetic marks within these domains. This phenomenon happens faster and in a larger domain for active chromatin regions where diffusivity of proteins and mobility of chromatin is higher, and importantly, might redistribute histone marks across different chromosomes. Interestingly, we find that these signatures of diffusion are absent in constitutive heterochromatin marked by H3K9me3, consistent with the idea that these histone marks are very well conserved across replication[41–45]. This "DAD" (Diffusion Accessible Domain) model adds a new dimension to our understanding of nucleosome dynamics[46] and suggests that positional information of histone marks may not be maintained as accurately as previously thought in euchromatin and facultative heterochromatin[26,47].

**Methods:**

*A stochastic model for quantifying the Diffusion-Accessible-Domain hypothesis (Model 1)*

We begin with a two-dimensional "parent" matrix $\overline{P}$, representing the parent cell undergoing replication. The rows of $\overline{P}$, indexed by $i$ ($i \in 1, 2, \ldots, R$), represent $R$ different regions or loci within the same or different chromosomes. The columns of $\overline{P}$, indexed by $j$ ($j \in 1, 2, \ldots, N$), represent $N$ nucleosomal sites along each genomic loci. $\overline{P}(i, j)$ is initialized with the value "1" denoting the presence of a histone at each site $(i, j)$. Only along one special row denoted by $i = R^*$ are the "1"s followed by a "*", representing tagged parental histones in accordance with the experimental design of Escobar et al. Two daughter matrices $\overline{D_1}$ and $\overline{D_2}$ are initialized with "0"s as entries; these represent the newly formed naked DNA strands for each loci. Next, physical distances between nucleosomal sites in one column of $\overline{P}$ and sites in the same column of $\overline{D_1}$ or $\overline{D_2}$ are defined based on a simple rule: $x_{i,i'} = |i - i'|$, for any $j$, where $i$ represents the row number of $\overline{P}$, $i'$ the row number of $\overline{D_1}$ (or $\overline{D_2}$) and $j$ denotes the column number of $\overline{P}$.

The processes of replication fork progression and histone transfer are modeled by choosing columns serially from $j = 1$ to $j = N$, and moving the entries in each column of $\overline{P}$ ("1" or "1*")

to sites in the same column of either $\overline{D_1}$ or $\overline{D_2}$ at the same time. This movement of entries in a column from $\overline{P}$ to a randomly chosen row in the same column of either $\overline{D_1}$ or $\overline{D_2}$ is the crucial component of our model – we create a probabilistic rule for this move to mimic diffusion among the positions of this column (see **SI** for details):

$$p_{i,i'} = \frac{1}{\sqrt{4\pi DT}} e^{-x_{i,i'}^2/4DT}, \qquad (1)$$

where $p_{i,i'}$ is the probability of a histone in row $i$ of $\overline{P}$ moving to row $i'$ of $\overline{D_1}$ or $\overline{D_2}$ (within the same column $j$), $D$ is the diffusion constant and $T$ represents some characteristic time associated with histone diffusion. Once all entries of all columns of $\overline{P}$ have been transferred to $\overline{D_1}$ and $\overline{D_2}$, sites in the two daughters that remain empty are filled up with new "1"s, representing newly synthesized histones. This series of steps represents the completion of one cell cycle event. We then randomly (with probability 0.5) choose one of the two daughter matrices $\overline{D_1}$ or $\overline{D_2}$ and assign it to be the parent $\overline{P}$, and the whole procedure is repeated many times to mimic multiple cell division events.

*Modeling the impact of diffusion on the emergence of histone modification patterns (Model 2)*

To include histone modifications in the simple model above, we simply re-interpret the "1"s in the parent matrix $\overline{P}$ as histones with a particular modification (say H3K27ac), and "0"s as empty sites or newly synthesized histones without any parental marks. We add another label "2" to denote histones with marks other than the one of interest. We also modify the description of $\overline{P}$ by assigning two labels to the rows: the first $R/2$ rows are labelled "$E$", denoting loci from early replicating regions, while the last $R/2$ rows are labelled "$L$" to denote late replicating loci. In this model, Early and Late loci serve as proxies for active and repressed chromatin regions respectively. As before, we assign distances to every pair of rows $(i, j)$, but with the following modifications: row-pairs $(i, j)_{EE}$ where both the rows belong to $E$, have a larger average distance than $(i, j)_{LL}$ pairs where both rows belong to $L$. This ensures that the model is consistent with the fact that late replicating loci tend to lie in heterochromatin which is more compact than euchromatin[23,37], and hence are expected to have smaller inter-loci distances on average. The distance assignment is explained in detail in the **SI**. The probabilities of histone mark re-distribution during cell division are assigned using Eq. (1) as before. The only difference now is that two distinct diffusion constants are used for $E - E$ and $L - L$ transitions: $D_{EE}$ and $D_{LL}$, where $D_{EE} > D_{LL}$. This inequality in the diffusion constants is consistent with experimental data suggesting faster diffusion of proteins and DNA in euchromatin as opposed to heterochromatin[32,34,35]. Finally, we initialize $\overline{P}$ with equal number of "1"s and "2"s (on average) between $E$ and $L$ rows. For any given histone mark this equal distribution is not realistic, but we use this initial condition simply to avoid bias and demonstrate that diffusion is sufficient for the emergence of specific patterns in the histone modifications. Changing the initial distribution of histone marks between $E$ and $L$ rows to more accurately reflect CHIP-seq datasets makes no difference to the central results of this work.

This modified model is then simulated for many cell cycles as described for Model 1, using Eq. (1) to probabilistically move both the histone marks ("1" and "2") to either one of the daughter matrices. Transitions within $E - E$ or $L - L$ rows differ simply in the use of different values of the diffusion constant. Unlike Model 1, in Model 2 we also implement a mark-copying model at the end of each cell cycle, to prevent the histone mark levels from diminishing to zero. A simple mark copying model is implemented, where all empty sites in the daughters are located, and one of the two nearest neighbor's marks is randomly assigned to the empty site (details in SI). This process is carried out till no more empty site remains in the daughters. Finally, a daughter is randomly selected between $\overline{D_1}$ and $\overline{D_2}$ and assigned to $\overline{P}$. After the required number of cell cycles, signal values are calculated for each row of the final matrix as: $signal_i =$ number of 1's in row $i$, and the similarity of histone modification is calculated as given below.

*Measuring similarity of a histone modification pattern between two loci in simulations*

The (dis)similarity of a particular histone mark between two loci or rows $(i, j)$ is given by:

$$Q(i,j) = Log_{10}\left|\left(signal_i/signal_j\right)\right| \qquad (2)$$

where $signal_i =$ number of 1's in row $i$, and the vertical bars represent the absolute value. In the context of measuring histone modification similarity, $Q(i,j)$ has been used earlier and was called the histone modification distance[48]. When two loci have identical signal values, $Q(i,j) = 0$ and as the signals become increasingly different, $Q(i,j)$ becomes larger. Therefore, the more dissimilar two genomic loci are in terms of their histone modification signals, the larger $Q(i,j)$ will be.

*Bioinformatics analysis to quantify histone modification similarity patterns*

To investigate whether the histone modification patterns predicted from the stochastic model can be observed in real data, we identified two publicly available cell lines, GM06990 and K562, both of which had HiC, replication timing as well as CHIP-seq datasets available. Replication timing was used as a measure of active versus repressed chromatin, since this correlation is strong and very well established. For 1Mb resolution HiC datasets, the corresponding 1Mb genomic fragments were defined and replication sequencing data was used to compute the total early and late scores for each 1Mb segment. A cutoff was then defined for the difference in early and late scores – genomic segments with score higher than +cutoff were classified as early, while segments with score less than -cutoff were classified as late. Unless mentioned otherwise, the cutoff value was chosen to be 1000 for all figures. Choosing the cutoff to be 500 did not change any results, as shown in the SI. These early and late segments were then intersected with CHIP-seq datasets for various histone modification marks, and net histone modification signal calculated for each mark over the 1Mb segment (described below). Finally, all pairs of intra and inter-chromosomal genomic segments $(i, j)$ were iterated over, and $Q(i,j)$ as well as HiC distance values computed for each pair.

To compute the Median$Q(i,j)_{chance}$ values expected by chance for any particular histone mark, random pairs of 1 Mb segments were chosen from the early group 10000 times and $Q(i,j)$ values

calculated each time. The median $Q(i,j)$ from these 10000 random pairs is Median$Q(i,j)_{chance}$, the expected median $Q(i,j)$ given the average histone modification signal values for early segments, when HiC proximity is not accounted for. A similar procedure was followed for late segments. Deviations from these expected values when pairs of segments are grouped by HiC distance would indicate effects of diffusion as predicted by our stochastic model.

*Measuring similarity of histone modification signals between two loci from CHIP-seq datasets*
To compute $Q(i,j)$ from CHIP-seq datasets, we use Eq. (2) as in the computer simulations, but define the signal value from each genomic locus differently. The signal value of a particular histone modification $m$ from a locus $i$ is defined as:

$$signal_{i,m} = \sum_{p} H_p \times W_p \ ,$$

where the index $p$ runs over all the CHIP-seq peaks in the genomic locus $i$, $H_p$ and $W_p$ are the height and width of the $p^{th}$ peak respectively and $m$ could be any of the histone modifications H3K27me3, H3K27ac, etc. We also compute a total signal score for each genomic loci, which is the sum of signals from all histone modifications:

$$signal_i = \sum_{m} signal_{i,m}$$

**Results:**
*The Diffusion-Accessible-Domain (DAD) hypothesis.*
As shown in **Fig. 1A**, the classic model of histone inheritance implicitly posits that 100% of the histones are transferred from the parent DNA to the leading and lagging daughter DNA strands, most likely in a symmetric fashion[2,11–13]. However, there seems to be no direct investigation on the possibility of parental histones along with their PTMs binding to nearby (non-daughter) chromatin after being dislodged during replication. This is somewhat surprising given proteins in other contexts are known to combine 3D diffusion with sliding on DNA tracks to find stable binding sites[31,49] or to be transferred between different DNA segments[50]. Core histones have also been demonstrated to exhibit diffusion[33]. Therefore although a number of active chromatin remodeling motors have been implicated in histone disassembly, sliding and exchange processes[51,52], and models for $(H3\text{-}H4)_2$ tetramer transfer coupled to the replication fork via the chaperones NPM1, CAF1, ASF1, MCM2 and POLE3/4 have been proposed[2,29], some degree of passive diffusion of histones cannot be ruled out. Indeed, it is well understood that ATP-dependent activity of motors and chaperones can result in increased diffusivity of other proteins within the same medium[53].

These experimental results on protein diffusion allowed us to formulate the first part of the DAD hypothesis – dislodged parental histones may exhibit some level of undirected diffusion while being transported to the destination DNA site. An immediate consequence of this hypothesis is that a small fraction of parental histones will diffuse to non-daughter DNA sites on neighboring

chromatin regions. If DNA is being replicated in the destination sites at the same time, this diffusion process could end with some histones binding to non-daughter DNA strands. Two conditions must therefore be satisfied for histone redistribution over neighboring chromosomal loci – (1) the origin and destination sites on DNA must be in close physical proximity, and (2) both the sites should belong to chromatin regions undergoing replication at the same time. **Fig. 1B** provides a schematic of the DAD hypothesis.

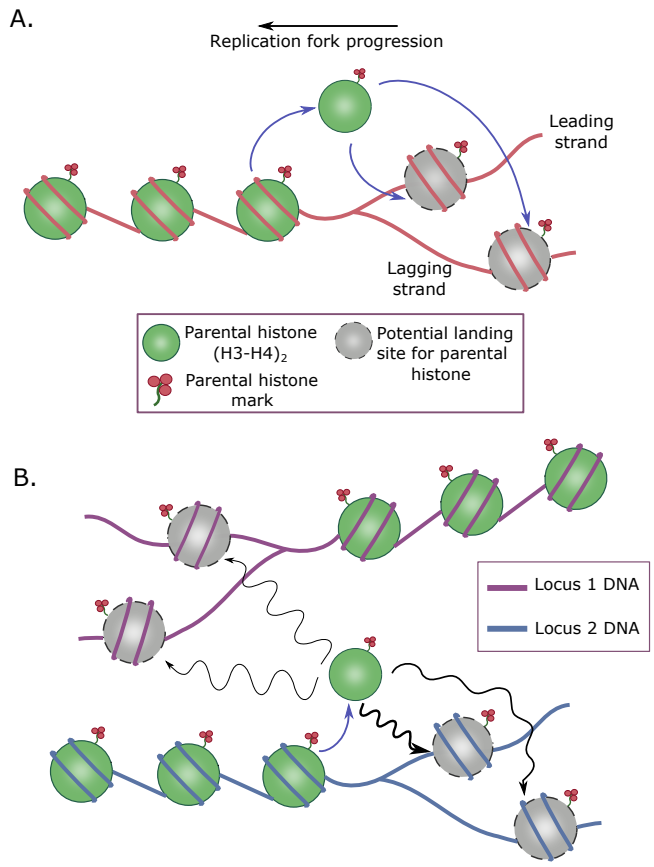


**Figure 1**: The Diffusion-Accessible-Domain (DAD) hypothesis. (A) Classical view of histone transfer at the replication fork, where the parental histones on being dislodged from the DNA land exclusively on the leading and lagging daughter strands. (B) Our new DAD hypothesis which states that due to diffusion, some parental histones (for example on locus 2 DNA) may end up landing on non-daughter strands (locus 1 DNA), as long as the loci are in close physical proximity as well as replicated at the same time. Thicker arrows indicate higher probability transitions due to physical proximity. The wiggly nature of the arrows are meant to indicate diffusive histone transfer. Note that the undirected nature of diffusion in this model results in symmetric distribution of parental histones to the daughter DNA strands.

The second part of the hypothesis is that diffusion of proteins as well as DNA in active chromatin regions (euchromatin) is faster than that in repressed regions (heterochromatin). The intuitive reason is that euchromatin is less tightly packed as compared to heterochromatin, thereby providing a relatively unhindered space for diffusion. Measurements of diffusivity in active versus repressed chromatin as well as in spatially heterogeneous chromatin directly confirm this hypothesis for test and core histone proteins[32,33]. Enhanced diffusivity of DNA in euchromatin[34,35] has also been experimentally observed. This second part of the hypothesis is crucial for explaining

the local versus dispersed histone transfer process observed in the recent work by Escobar et al[28], as we quantitatively demonstrate in the next section.

*A stochastic model of the DAD hypothesis reproduces experimental data on differential histone dilution kinetics at the replication fork.*

We then converted the DAD hypothesis as formulated above into a quantitative model (Model 1), to investigate whether the observed experimental data on differential histone dilution kinetics in active versus repressed chromatin[28] can be recapitulated. A detailed description of the model along with its assumptions and limitations is provided in the **Methods** and **SI**; here we provide a brief description. We set up a lattice model of the replication fork where the rows of a 2D matrix $\overline{P}$ (the parent "cell") represent different genomic loci and columns represent positions of nucleosomes along each loci (**Fig. 2A**). Replication fork progression is simulated by serially removing histones from each column and redistributing them stochastically in two new matrices $\overline{D_1}$ and $\overline{D_2}$ representing daughter cells, following the laws of diffusion (**Methods**, Eq. (1)). One special locus $R^*$ starts with marked histones which are tracked over cell cycles, allowing quantification of the kinetics of histone dilution and mimicking the experiments by Escobar et al[28] (**Fig. 2B**). The only free parameter in this model which is varied to fit the experimental results is the diffusion constant $D$ in Eq. (1).

The results of simulating Model 1 are shown in **Fig. 2C**, where decay in the marked histone level at the locus $R^*$ is followed over time. Best fit results to experimental data on active genes are shown in **Fig. 2C** top row**,** while those for repressed genes are shown in **Fig. 2C** bottom row**.** The good fits of the model to the data suggest that differential diffusion is sufficient to explain the large differences in the histone dilution kinetics at the replication fork. The best fit diffusion constants are also shown in the respective plots, and as expected, are larger for active genes as compared to repressed genes (**Fig. 2D**). Interestingly, the diffusion constants for all the active genes were quite similar to each other while those for the repressed genes were smaller but similar to each other, suggesting little variation in diffusivity given the genomic context. While these results suggest that histone dilution kinetics could be governed by differential diffusivity between active and repressed regions, this model by itself does not provide any means of checking whether histone redeposition occurs on other loci, including on different chromosomes. To explore the possibility of this phenomenon, we modify the model to include a description of histone marks and predict histone mark similarity patterns between pairs of loci in the genome, as described in the next section.

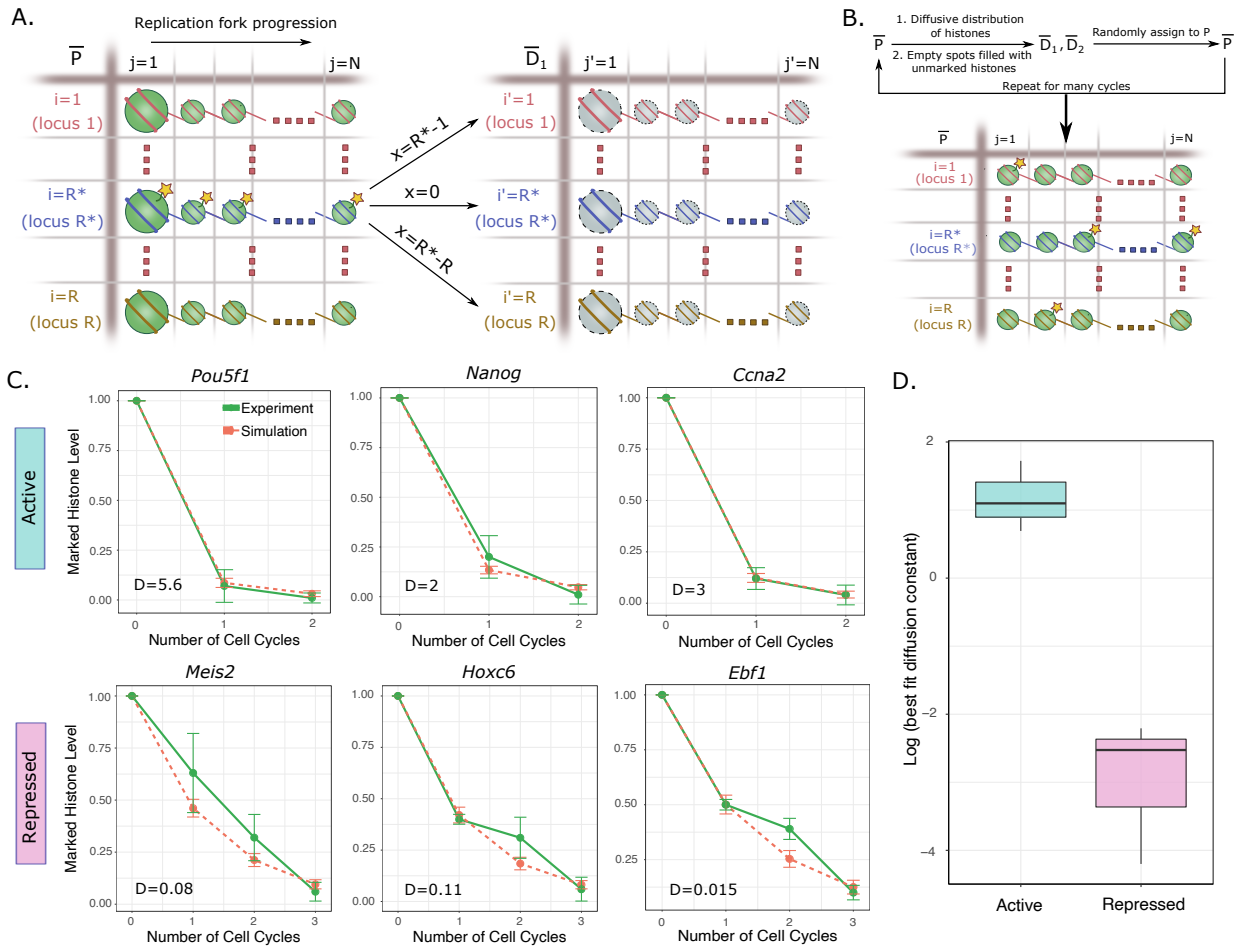


**Figure 2**: Differential histone diffusivity between chromatin compartments explains histone dilution kinetics during replication. **(A)** A model of histone dilution at the replication fork incorporating the DAD hypothesis (Model 1). Each row of the parent matrix $\overline{\mathrm{P}}$ represents a different DNA locus (either on the same or on different chromosomes) and columns represent histone sites. Histones are shown as green spheres. The yellow stars represent biotin tags on the histones at one particular locus of interest $\mathrm{R}^*$, mimicking the initial conditions of the experimental protocol in Escobar et al[28]. The arrows show examples of possible movement of histones between loci from the parent $\overline{\mathrm{P}}$ to either daughter $\overline{\mathrm{D}_1}$ or $\overline{\mathrm{D}_2}$. The probabilities of these transitions are based upon laws of diffusion, which accounts for the distance between parent and daughter loci $\mathrm{x}_{\mathrm{i},\mathrm{i}'} = |\mathrm{i} - \mathrm{i}'|$ (details in Methods and SI). **(B)** Flowchart of one run of the simulation for multiple cell cycles. Histone are distributed from parent to the two daughter matrices based on the laws of diffusion, and the process repeated many times. Empty spots arising out of histone dilution are filled with untagged histones, thereby leading to a decrease in number of tagged histones at the locus $\mathrm{R}^*$. **(C)** Comparison of model (pink dashed lines) with experimental results[28] (green solid lines) for both active and repressed genes. Results from three active genes (Pou5f1, Nanog, Ccna2) and three repressed genes (Meis2, Hoxc6, Ebf1) are shown here. The model results represent best fit curves, with the diffusion constant $\mathrm{D}$ as the only free parameter. **(D)** As expected, fitted diffusion constants from panel (C) are higher for active genes than repressed genes, demonstrating the reasonability of our model.

## *Diffusion of histones is predicted to generate both distance-dependent and independent patterns of histone mark similarity in early and late-replicating genomic loci*

We argued that if histone diffusion within DADs and the concomitant redeposition is indeed prevalent during replication, there must be signatures of this process left behind on the PTMs associated with the histones. In particular, pairs of genomic loci might show histone modification patterns that are indicative of diffusive processes. To develop a quantitative expectation for the nature of patterns that might be generated by diffusion of histones, we extended the model from

the last section to include a description of histone modifications. This allowed us to quantify average levels of a particular modification on every loci in the model, and compare the similarity of these levels between loci within active/repressed regions (**Methods**). Note that in this and subsequent parts of the paper, we use Early (E) and Late (L) replication timing regions as proxies for active and repressed genomic regions respectively, since this connection is well established[54].

Briefly, in this extended DAD model (Model 2; **Fig. 3A**) we start with equal average signal of a histone modification in E and L loci and allow for diffusive exchange between early-early (EE) and late-late (LL) loci. This exchange or redistribution once again follows the law of diffusion (Eq. (1)), and the diffusion constants follow the inequality $D_{EE} > D_{LL}$. At the end of each cell cycle, mark copying[24,38,55] is implemented to assign marks to newly synthesized histones – this prevents dilution of histone mark levels across cell divisions. Many cell cycles are simulated and finally the similarity in histone modification signals $Q(i,j)$ from EE and LL pairs of loci are calculated. When signals from two segments are identical, $Q(i,j) = 0$, and it increases monotonically as the signals become less similar (**Methods and SI**).

Predictions of Model 2 are shown in **Fig 3B.** Histone modification similarity $Q(i,j)$ as a function of distance at four time points (after 0, 2, 8 and 50 cell cycles) is shown. When no cell divisions have occurred, the pattern is set simply by the initial conditions, which we specifically chose to avoid any biases among the EE and LL segments. With increasing number of cell cycles, the diffusive process redistributes histone marks, leading to decreasing $Q(i,j)$ (indicating increasing similarity of loci pairs driven by diffusion) and emergence of a final stable pattern (**Fig. 3B**). Note that the mark-copying process is crucial for stability of the pattern over many cell cycles, since the pattern would have become flat over many cell cycles in the presence of diffusion and no mark-copying. Three distinct and important aspects of the pattern can be identified in **Figs. 3B** and **3C**:

(1) a distance independent pattern: at any given value of the distance $x$, the median $Q(i,j)$ satisfies the inequality $Q(i,j)_{EE} < Q(i,j)_{LL}$, implying more similarity between EE segments compared to LL segments. In **Fig. 3B**, this pattern develops even after just 2 cell divisions.

(2) a distance-dependent pattern: pairs of loci that are closer in distance are more similar than pairs that are further away, resulting in a non-zero slope of the Median$Q(i,j)$ vs $x$ line. This is true irrespective of whether the loci come from early-early or late-late pairs. This slope is evident in the EE segments already after 2 cell divisions, but takes more time to develop in the LL segments (**Fig. 3B**).

(3) The slope of the distance dependence is steeper for the lower diffusion constant. Hence the LL pairs exhibit a more negative slope compared to the EE pairs once the steady pattern emerges after sufficient number of cell cycles (**Fig. 3C**).

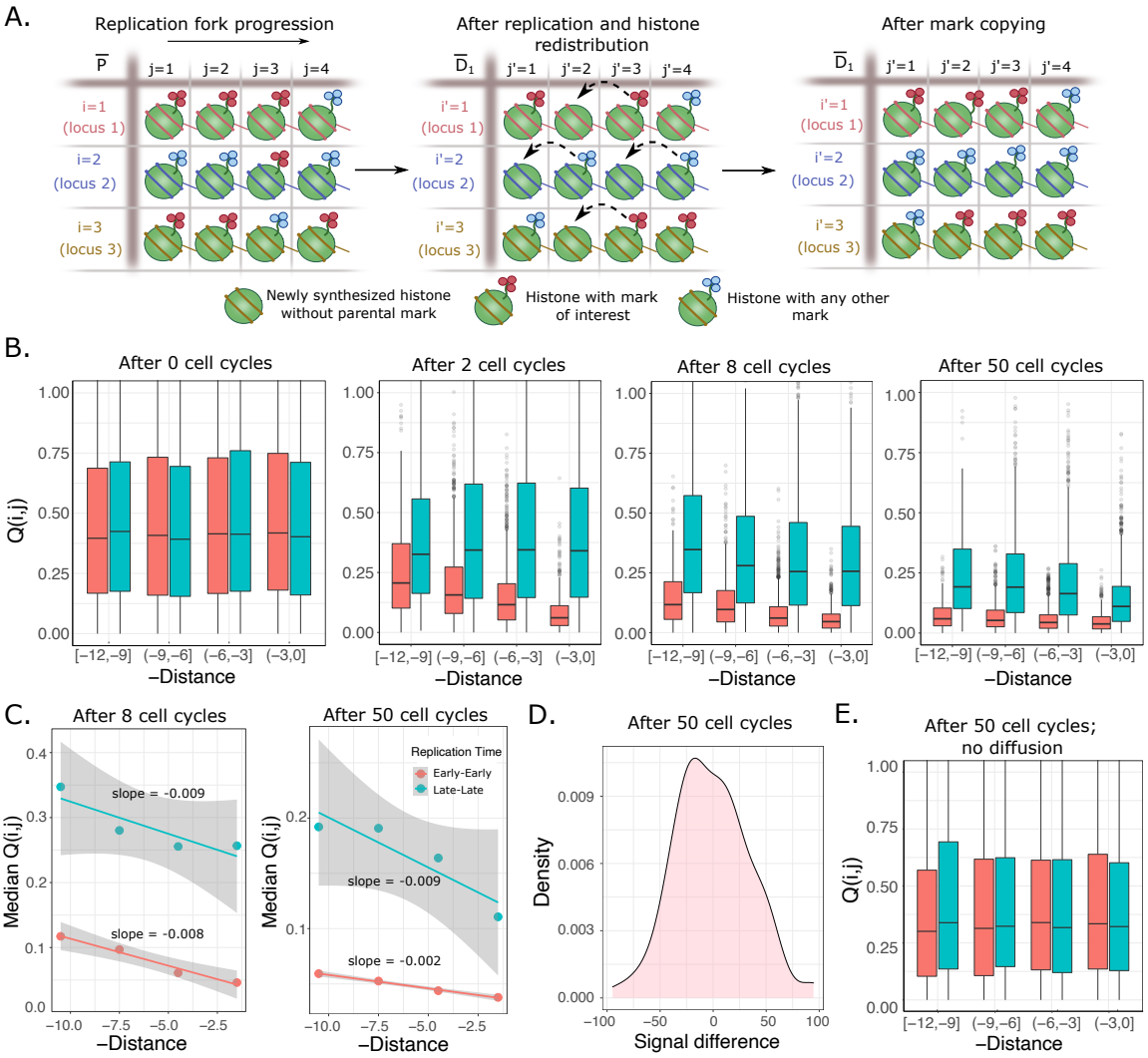


**Figure 3**: DAD model predictions of diffusion-induced histone mark patterns in the epigenome. **(A)** A modified version of the model shown in Fig. 2A, that incorporates a description of histone marks and mark copying (Model 2). Since any locus will have various types of histone marks, we incorporate a mark of interest (red clover symbol) and other marks (blue clover). Once the parent $\overline{P}$ divides, empty spots are left in the daughters $\overline{D_1}$ and $\overline{D_2}$, which get filled with newly synthesized histones carrying neither blue nor red marks. These sites are then filled by a simple mark copying rule where the nearest neighbor sites determine the mark to be assigned (dashed lines; details in Methods and SI). Rules of diffusion and transition of histones from parent to daughter cells remain identical to Model 1. **(B)** Results from simulations of Model 2, for various number of cell cycles. Histone modification similarity between pairs of loci $Q(i,j)$ as a function of negative distance ($-x_{i,i'}$) is plotted. Negative distance was used to qualitatively match the spatial proximity measure in HiC datasets, where physical proximity increases towards the right of the x-axis. The red box- plots correspond to Early-Early (EE) segments while blue corresponds to Late-Late (LL) segments. For all simulations performed, the matrices $\overline{P}$, $\overline{D_1}$ and $\overline{D_2}$ have 20 rows and 200 columns, where the first 10 rows correspond to Early segments (faster diffusion, $D = 1.2$) and the last 10 rows to Late segments (slower diffusion, $D = 0.07$). Each row was assigned a number of histone marks of interest drawn from an exponential distribution with parameter 130. Statistics for each plot were generated from 100 repetitions. Beginning with no difference in $Q(i,j)$ between EE and LL segments at 0 cell cycles, three distinct patterns emerge and stabilize over increasing cell cycles: $Q(i,j)_{LL}$ is larger than $Q(i,j)_{EE}$ for any value of the x-axis, there is a distinct distance dependence in the $Q(i,j)$ values for both EE and LL loci pairs, and finally the slope of the distance dependence is higher for LL pairs. **(C)** Quantification of the slopes of the distance dependent $Q(i,j)$ patterns using linear regression, demonstrating a steeper slope for the LL loci pairs (blue line) once the patterns have stabilized after 50 cell cycles. **(D)** Distribution of the difference between average Early (averaged over the 10 Early rows) and average Late signal after 50 divisions, demonstrating no biases in the simulations. Since these simulations are started with equal numbers of histone marks of interest (on average) between Early and Late rows, no differences emerge after any number of divisions. Yet, the $Q(i,j)$ values develop the distinct patterns shown in panel (B), highlighting the role of diffusion in generating non-trivial patterns. **(E)** Simulations of Model 2 where diffusion is not allowed and parental histones from a particular locus land on the identical locus in either one of the daughter DNA strands. Evidently, none of the three patterns from panel (B) develop when diffusion is prevented from occurring.

Note that since there is no directionality to the diffusion process, no positive or negative bias in signal difference between EE and LL pairs emerges after many cell cycles (**Fig. 3D**). Yet, differences in signal similarity get stably established simply due to the differences in diffusion constants – faster diffusion within EE loci allows for histones to reach further distances in a given amount of time $T$, thereby increasing overall similarity of these segments. If diffusion is disallowed in the model and a histone evicted at a parental site is allowed to land *only* on the same site in either one of the two daughter strands, none of the three patterns are observed as shown in **Fig 3E.** This simple model serves to clearly demonstrate the patterns that emerge purely out of diffusion of histones at the replication fork. None of the above results change qualitatively if the simulations are started with unequal average signal of the histone modification between E and L regions, which is the more realistic scenario (**Fig. S1**). In the Discussion section below, we provide detailed arguments for why these patterns are unlikely to be generated by other processes such as PRC2-mediated histone mark spreading or histone turnover during transcription.

*DAD model predictions are consistently observed in datasets of the GM06990 and K562 cell lines, providing evidence for diffusion-mediated redistribution of histone marks*

Having established quantitative expectations for diffusion-induced histone modification patterns in the genome using stochastic simulations, we next investigated whether such patterns can be found in real datasets. We used data from lymphoblastoid (GM06990) and erythroleukemia (K562) cell lines for which Hi-C, replication timing and CHIP-seq datasets for various histone modifications were available in public repositories (see **Methods** and **SI** for details; **Fig. 4A** provides a summary of the bioinformatics analysis). We first computed the net early and late replication timing scores for each 1 Mb segment (for Hi-C data of 1 Mb resolution), and noticed that the distribution of the difference in score across all 1 Mb segments was bimodal, with peaks at positive and negative values (**Fig. 4B**). This allowed us to classify coarse-grained 1 Mb loci as either early or late, based on whether the score difference was positive or negative. To be confident that a 1 Mb segment comprised mostly early or mostly late signals, we used positive and negative cutoffs respectively for the classification (**Fig. 4B**), and checked that the precise value of the cutoff does not make any qualitative difference to the main results (**Fig. S2**). We then quantified the similarity of histone marks $Q(i,j)$ between pairs of these coarse-grained loci $(i,j)$, grouped by the Hi-C distance (spatial proximity) between them. Note that this spatial proximity metric is not a physical distance, but a correlation measure[56]. Values of spatial proximity closer to 1 indicate more frequent contacts between loci, implying high proximity in physical space. While this difference between $x$ in our computational models and spatial proximity from Hi-C maps prevents a direct quantitative comparison, we expect to be able to make qualitative comparisons on the patterns of histone modifications.

**Fig. 4C** shows one example each of an active mark (H3K27ac), a repressive mark (H3K27me3) and a mark that has been suggested to be associated with poised chromatin regions (H3K4me2[57]). The strong qualitative similarity of the $Q(i,j)$ patterns in **Fig. 4C** to computational predictions of the DAD model (in **Fig. 3B**) is immediately apparent across both cell types GM06990 and K562. We found signs of all three predicted patterns – $Q(i,j)_{EE} < Q(i,j)_{LL}$ for any value of spatial proximity (**Fig. 4C**), the non-zero slope in the $Q(i,j)$ vs spatial proximity plot (**Fig. 4C,D**), as well

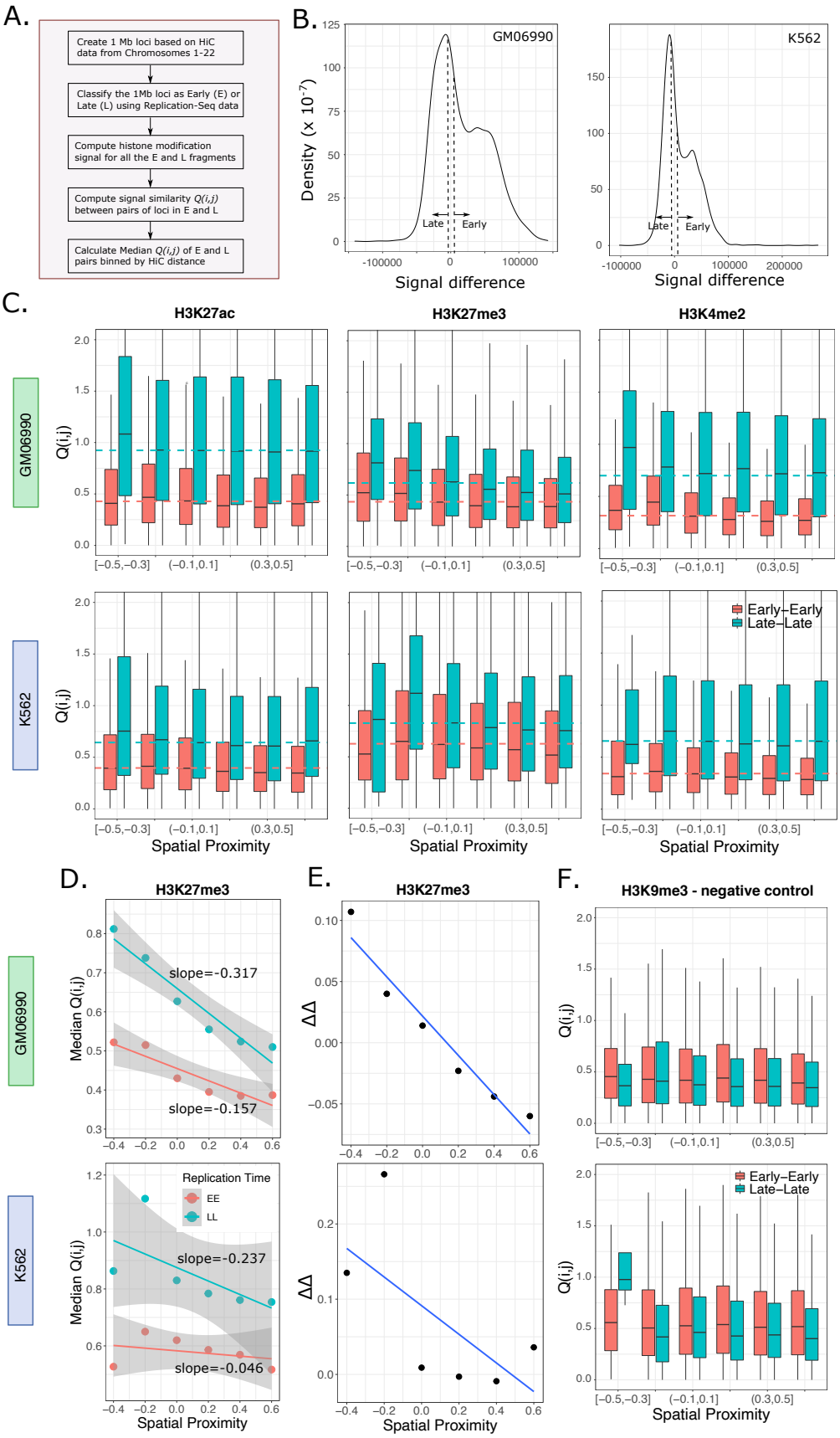
A.
Create 1 Mb loci based on HiC data from Chromosomes 1-22
Classify the 1Mb loci as Early (E) or Late (L) using Replication-Seq data
Compute histone modification signal for all the E and L fragments
Compute signal similarity Q(i,j) between pairs of loci in E and L
Calculate Median Q(i,j) of E and L pairs binned by HiC distance
B.
GM06990
K562
Density (x 10^-7)
Late
Early
Signal difference
C.
H3K27ac
H3K27me3
H3K4me2
GM06990
K562
Q(i,j)
Early–Early
Late–Late
[−0.5,−0.3]
(−0.1,0.1]
(0.3,0.5]
Spatial Proximity
D.
H3K27me3
Median Q(i,j)
slope=-0.317
slope=-0.157
Replication Time
EE
LL
slope=-0.237
slope=-0.046
E.
H3K27me3
ΔΔ
F.
H3K9me3 - negative control

**Figure 4**: *Histone modifications in the GM06990 and K562 cell lines consistently exhibit the predicted patterns from the DAD model.* **(A)** An overview of the bioinformatics pipeline for investigating the presence of the three predicted histone modification patterns in public datasets. **(B)** Classification of 1 Mb genomic segments as early or late, depending on the replication timing score difference (early score – late score) for each segment. A cutoff value of 1000 was used in all plots in this figure. If the score difference was larger than +1000, the segment was classified as early. If the score difference was less than -1000, the segment was classified as late. Results with another cutoff value of 500 are shown in **Fig. S**. **(C)** $Q(i,j)$ vs spatial proximity plots for H3K27ac, H3K27me3 and H3K4me2 modifications from the GM06990 (top row) and K562 (bottom row) cell lines. The dashed lines represent the $\mathrm{Median}Q(i,j)_{\mathrm{chance}}$ values for early (red) and late (blue) loci. The three predicted patterns from the DAD model are clearly visible here. **(D)** Linear regression of Median$Q(i,j)$ against spatial proximity for EE and LL loci pairs, for the H3K27me3 mark in both cell types. The x-axis points are the middle values of the bins from panel (C). The third pattern – a steeper slope for LL segments is evident in these plots. **(E)** The difference $\Delta\Delta = \Delta\mathrm{Median}Q(i,j) - \Delta\mathrm{Median}Q(i,j)_{\mathrm{chance}}$ plotted as a function of spatial proximity, where $\Delta\mathrm{Median}Q(i,j) = \mathrm{Median}Q(i,j)_{LL} - \mathrm{Median}Q(i,j)_{EE}$. As per the DAD model predictions, $\Delta\mathrm{Median}Q(i,j)$ is expected to be larger at lower values of spatial proximity (left side of the plot) and smaller at higher values of spatial proximity (right side of the plot). $\Delta\mathrm{Median}Q(i,j)_{\mathrm{chance}}$ was subtracted from $\Delta\mathrm{Median}Q(i,j)$ to account for the observed baseline differences in average values of histone modification signal between early and late segments for any mark. Positive $\Delta\Delta$ values at low spatial proximity demonstrate consistency with the DAD model predictions. **(F)** A negative control histone mark H3K9me3, that shows none of the three predicted patterns in $Q(i,j)$. This is likely a result of H3K9me3 marking constitutive heterochromatin regions, where diffusion is expected to be minimal or even absent.

as the higher slope for LL segments (**Fig. 4D** shows H3K27me3 slopes as an example). We noticed that in these datasets the average histone modification levels were different between the early and late segments, for each modification (**Fig. S3**). To make sure that the three patterns of interest were not spuriously generated due to this baseline difference in average levels, we calculated for both early and late segments the quantity $\mathrm{Median}Q(i,j)_{\mathrm{chance}}$, the median similarity expected given their average modification levels. To compute this quantity, we randomly sampled pairs of loci from within the early (late) segments, with no Hi-C information, and computed the median of the $Q(i,j)$ values of these random pairs. These are shown as dashed lines in **Fig. 4C**. Clearly, when grouped by Hi-C distance, Median$Q(i,j)$ at low spatial proximity (left side of the plots) are systematically higher than $\mathrm{Median}Q(i,j)_{\mathrm{chance}}$ while at high spatial proximity (right side of the plots), Median$Q(i,j)$ are lower than $\mathrm{Median}Q(i,j)_{\mathrm{chance}}$. This is true both for early and late loci pairs. Furthermore, the difference in Median$Q(i,j)$ values between late and early segments, $\Delta\mathrm{Median}Q(i,j)$), are consistently larger than $\Delta\mathrm{Median}Q(i,j)_{\mathrm{chance}}$ at lower spatial proximity (**Fig. 4E**), becoming smaller (or even negative) towards higher spatial proximity values as expected from the DAD model. This suggests that baseline differences between average histone mark signal in early vs late segments is not sufficient to explain the trends we observe. Note that strictly speaking the DAD model predicts positive $\Delta\mathrm{Median}Q(i,j)$ at any spatial proximity; hence the negative values observed in **Fig. 4E** are at odds with our model results. This is likely because $\Delta\mathrm{Median}Q(i,j)$ is expected to become close to zero at close proximity (see **Fig. 3B**, 50 cell cycles), and such small positive differences can be hard to detect in noisy datasets.

We also investigated the existence of the same patterns when loci pairs were separately analyzed based on *intra* versus *inter*-chromosomal contacts. Expectedly, the intra-chromosomal contacts consistently showed all three patterns in both cell lines (**Fig. S4**). More interestingly, some of the inter-chromosomal contacts also showed clear signs of the diffusion-induced patterns, particularly in the GM06990 datasets (**Fig. S5**). Consistent with the expectation that diffusion across larger distances is easier in active chromatin, the inter-chromosomal diffusion patterns were clearer in the Early-Early pairs (**Fig. S5** top row, red boxplots). These results suggest that

while diffusion is more prevalent between intra-chromosomal regions due to shorter distances, redistribution of histone marks can also occur between *different* chromosomes. Note however, that since inter-chromosomal contacts occur much less frequently as compared to intra-chromosomal contacts, the patterns for the former are noisier as is evident in **Fig. S5**.

Finally, we wondered whether we could find a "negative control" histone mark where the predicted patterns cannot be observed. We argued that a good candidate might be the constitutive heterochromatin mark H3K9me3, since constitutive heterochromatin regions are expected to be most densely packed with minimal diffusion, and the marks are known to be very well conserved across cellular generations[41–45]. As per our expectation, neither in the GM06990 nor in the K562 cell types could we see evidence of any of the three patterns in H3K9me3 (**Fig. 4F**). In fact the late-late loci were more similar than the early-early loci and there was no distance-dependence in $Q(i,j)$. These patterns did not show up even if the intra-chromosomal contacts were separately analyzed (**Fig. S6**). In summary, the consistent presence of our computationally predicted patterns in most datasets point to a widespread influence of diffusion in shaping the landscape of histone modification patterns during cell division.

**Discussion**

How various active and passive biological processes combine to give rise to the observed architecture of the genome has garnered immense interest in the last decade. Advances in experimental techniques combined with theoretical and computational modeling have provided quantitative explanations for the emergence of features of the genome such as chromatin loops[58–62], topologically associated domains (TADs)[56,63,64], lamin-associated domains (LADs)[65,66] and phase separated active and repressed domains[67–69]. However, much less is understood about what patterns in the epigenome exist and how they are maintained over time and across cell division events. Motivated by simple physical principles, here we hypothesized the existence of a new kind of domain which we termed "Diffusion Accessible Domain" or DAD, that may shape the architecture of the epigenome. Our DAD hypothesis posits that during replication, dislodged histones along with their post-translational modifications undergo some level of undirected diffusion (potentially as complexes with various histone chaperones). These diffusing histones may get captured by any genomic loci in the vicinity which are also in the process of being replicated. While the majority of parental histones from any locus will get captured *in cis* by the spatially proximal leading or lagging strands of the same locus, occasionally some histones will get deposited *in trans*, leading to a redistribution of histone marks. Since the diffusion is undirected, deposition of the histones is still symmetric[6,12,13] between the leading and lagging strands in this model. Of note is that this hypothesis relies purely on passive, ATP independent processes, and does not yet account for asymmetries that have been observed recently in Drosophila germ cells[70]. Intriguingly, results from a recent minimal experimental system tracking single nucleosomes suggested that passive, chaperone-independent processes can indeed lead to nucleosome transfer at the replication fork[71].

We combined a lattice model for the replicating fork with stochastic rules for histone transfer to convert the DAD hypothesis into a computational model that can be used to explore the ramifications of this diffusion-based hypothesis. We first demonstrated that simply by allowing

the diffusion constants of histones to be different between active and repressed regions, we can quantitatively explain recent results[28] on the disparate kinetics of histone dilution at various genomic loci. We next predicted three characteristic patterns of histone mark similarity between pairs of loci, which are expected to be developed as a result of diffusion. While one of these patterns is distance independent, two of the emergent patterns are distance-dependent. We carefully demonstrated that even in the absence of any differences between early and late loci that may bias the simulations (except the differences in diffusion constants), these three patterns emerged over the course of cell divisions. Perhaps most importantly, we find the existence of all three patterns in histone modification datasets from two different cell lines, strongly suggesting that diffusion plays an important role during replication. Rather remarkably, our analysis provides evidence not just for intra-chromosomal, but also *inter*-chromosomal diffusion of histone marks – a result that is fundamentally different from the current models of histone mark inheritance, including those that allow for dispersion of histones in active genomic regions[2,28]. Some earlier studies have reported more similarity in histone mark signals at closer physical distances[48], but the replication timing resolved patterns have not been reported to the best of our knowledge. Crucially, the reason for the existence of such patterns is poorly understood, and here we provide a plausible model based on simple physical principles.

Interestingly, H3K9me3, which is a hallmark of constitutive heterochromatin, was the only modification that did not exhibit any of the expected patterns in either the GM06990 or the K562 datasets. This is fully consistent with many past studies, which have demonstrated strict conservation of this constitutive heterochromatin mark via suppression of histone turnover by a plethora of proteins such as ASF1, SNF2 family proteins and FACT[44,45,72], This would also be consistent with the idea that constitutive heterochromatin is the most condensed of all chromatin states[73], thereby minimizing diffusion. Additionally, it is interesting to note that H3K27me3, which is a hallmark of facultative heterochromatin[2], does show all the signs of diffusion-mediated patterning. This also seems consistent with the observation that facultative heterochromatin regions are less condensed than constitutive heterochromatin and are poised to convert into euchromatin[28], therefore allowing for more diffusion of histones.

Could other models of histone mark transfer potentially explain all or some of the similarity patterns we predict via simulations and observe in datasets? The most obvious alternate model would be that of PRC2 mediated long-range histone mark spreading, which would indeed explain why chromosomal regions in close proximity tend to have similar levels of H3K27me3[74,75] (pattern 2 in **Figs. 3B,C**). However, if the *only* factor determining distance-based histone modification similarity (for H3K27me3) was PRC2-dependent spreading, then one would not be able to explain why for the same distance, we see more similarity in Early replicating loci pairs as compared to Late replicating loci pairs (pattern 1, **Fig. 3B** and **Fig. 4C**). Furthermore, the PRC2 model alone cannot explain why the slope of the distance dependence is higher in Late as compared to Early replicating regions (pattern 3; Fig. 3B and Fig. 4C). And since we see the same pattern in marks other than H3K27me3 where PRC2-like long-range spreading mechanisms have not yet been reported (**Fig. 4C**), this argues for the existence of additional, more general mechanisms that could be at play. Additionally, while there is strong evidence for H3K27me3 spreading from "nucleation" sites to both proximal (cis) and distal (trans) "spreading" sites, the

precise mechanism of this spreading is currently not well understood. In a recent work by Oksuz and colleagues, mouse embryonic stem cells with H3K27me3 depleted genomes were induced to express PRC2 and consequently develop H3K27me3 de novo[74]. It was demonstrated that while PRC2 occupied and traversed the distance between nucleosomes at nucleation sites where the initial H3K27me3 deposition occurred, PRC2 occupancy in distal sites (though close in physical distance due to DNA looping) remained negligible over time even though the H3K27me3 marks spread to these sites[74]. Furthermore, the idea that PRC2 might bring together distal chromatin domains and form loops[76] may not be true everywhere in the genome, since mutating the EED domain of PRC2 did not lead to any observable differences in the 4C interactome of 11 nucleation sites[74]. These results argue that the precise mechanism(s) leading to trans spreading of H3K27me3 remain unclear, and it is plausible that mechanisms distinct from PRC2-mediated spreading exist. Indeed, since the cells used in the Oksuz et al study (and other recent H3K27me3 spreading studies[75]) were stem cells from humans and mice, the division times would be of the order of 12 hours, thereby making diffusion mediated spreading a distinct possibility as well. In summary, it is likely that there exist additional and more general mechanisms for long-range rearrangements of histone marks besides the canonical one proposed for H3K27me3 via the PRC2 complex. Diffusion mediated histone mark spreading could potentially be one such mechanism affecting not just H3K27me3, but other marks as well, as our results suggest.

A second natural question that arises is whether similar patterns could arise due to histone turnover during transcription. This however is unlikely because of two reasons: first, many past studies[45,77–79] and also recent cryo-EM structures[80] have suggested that histones that are displaced during transcription, are mostly replaced in their original sites after passage of the RNA Pol-II polymerase. This is true for most of the genome, except small regions around the edges of genes and in very highly transcribed genes[81]. While histone turnover is largely minimized, correct phasing of histones is achieved during transcription[82]. Second, even if we were to assume that a small fraction of histones dislodged during transcription do not get replaced in their original positions, this is not sufficient to explain the distance dependent patterns in the K562 and GM06990 datasets. Transcriptional events across genes in a single cell occur in unsynchronized, stochastic bursts[83,84]. As a result, dislodged histones that are not replaced after RNAP-II passage are unlikely to diffuse and reassemble on physically proximal genomic regions. Simultaneous transcription at the proximal DNA would be unlikely, leaving no naked DNA for the histones to land on. This is in contrast to replication, where entire genomic regions are known to replicate in phase, broadly divided into Early and Late replicating regions[54]. This synchronized availability of naked genomic regions is a critical requirement of the DAD model, as mentioned earlier. Indeed, in a recent study where replication timing patterns of the genome were randomized by knocking out the protein RIF1, replication dependent alterations were observed in the histone modification patterns[85]. This experiment suggests that the replication timing program is crucial for maintaining the boundaries of euchromatin and heterochromatin, and the DAD model requiring synchronized replication among genomic loci is consistent with these results. The DAD model is also consistent with the idea of sharp heterochromatin boundaries, since replication timing differences would minimize diffusion between early and late replication timing regions which mirror euchromatin and heterochromatin respectively.

While our model predictions are recapitulated in a remarkable number of datasets, there are a number of limitations of the current model arising from simplifying assumptions (discussed in detail in the **SI**). Primary among them is the fact that the diffusion constants used in the model cannot be directly compared to their experimental counterparts. This is partly because of the simplicity of our model – we use a 1D diffusion equation and the lattice model has distances only in 1D instead of 3D. Our models also make the assumption that a dislodged set of histones must find their new spots before the next set of parental histones get dislodged. This may be reasonable, since a number of studies in the past have demonstrated that nucleosome positioning is rapidly re-established behind the replicating fork[5]. Another major limitation of our results is that since Hi-C provides correlates of physical proximity and not real distances, it is challenging to map our distance $x$ in simulations to the spatial proximity measure in Hi-C datasets. This limitation also currently prevents us from estimating how large the DADs could be and how the DAD sizes differ between active and repressed chromatin regions. Methods to infer physical distances from Hi-C contact maps are being currently developed, and might help resolve these intriguing questions in future studies[86,87]. Indeed, since the slope of the median $Q(i,j)$ vs $x$ plot is directly related to the diffusion constant, estimating the diffusion constants in various genomic regions from these plots presents an exciting possibility for the future.

**Conclusion**

While much is known about the physical forces that shape the genome, far less is understood about the emergence of the epigenome. Our results in this work suggest that diffusive events redistribute histone modifications during cell division both in active and repressed genomic regions, in contrast to the "text-book" picture of parental histones being transferred only to the leading or lagging daughter strands. This redistribution is more prevalent between intra-chromosomal genomic contacts, but also occurs between inter-chromosomal genomic loci. The consequence of such a model for our understanding of epigenetic inheritance could be immense – maintenance of cellular states might be achieved not by passing on strict positional information of histone marks, but by maintaining histone marks in somewhat larger Diffusion Accessible Domains (DADs). Additionally, we surmise that these replication-coupled diffusive events might be a potential mechanism to generate plasticity in epigenomic states, for example in cancer or stem cells. Future experiments will clarify whether this model and its implications hold – whether diffusion is indeed a physical phenomenon that shapes the architecture and plasticity of the epigenome.

**Author Contributions:** S.C. conceptualized and designed the research, A.S. and S.C. conducted the research and analyzed the results, S.C. wrote the manuscript with inputs from A.S.

**Acknowledgements:** S.C. would like to thank Anjana Badrinarayanan, Mukund Thattai, Sabarinathan Radhakrishnan, Dave Thirumalai, P.V. Shivaprasad, Madan Rao, Satyanarayan Rao, Mahipal Ganji and Sahana Holla for detailed discussions and feedback that helped improve the manuscript. S.C. would also like to thank Franziska Michor and her lab for thoughts and suggestions in the early stages of the project.

**References:**


1. Cavalli, G. & Heard, E. Advances in epigenetics link genetics to the environment and disease. *Nature* **571**, 489–499 (2019).
2. Escobar, T. M., Loyola, A. & Reinberg, D. Parental nucleosome segregation and the inheritance of cellular identity. *Nat. Rev. Genet.* **22**, 379–392 (2021).
3. Annunziato, A. T. Split decision: what happens to nucleosomes during DNA replication? *J. Biol. Chem.* **280**, 12065–12068 (2005).
4. Groth, A. *et al.* Regulation of replication fork progression through histone supply and demand. *Science* **318**, 1928–1931 (2007).
5. Annunziato, A. T. Assembling chromatin: the long and winding road. *Biochim. Biophys. Acta* **1819**, 196–210 (2013).
6. Stewart-Morgan, K. R., Petryk, N. & Groth, A. Chromatin replication and epigenetic cell memory. *Nat. Cell Biol.* **22**, 361–371 (2020).
7. Parmar, J. J. & Padinhateeri, R. Nucleosome positioning and chromatin organization. *Curr. Opin. Struct. Biol.* **64**, 111–118 (2020).
8. Herman, T. M., DePamphilis, M. L. & Wassarman, P. M. Structure of chromatin at deoxyribonucleic acid replication forks: location of the first nucleosomes on newly synthesized simian virus 40 deoxyribonucleic acid. *Biochemistry* **20**, 621–630 (1981).
9. Ransom, M., Dennehey, B. K. & Tyler, J. K. Chaperoning histones during DNA replication and repair. *Cell* **140**, 183–195 (2010).
10. Radman-Livaja, M. *et al.* Patterns and Mechanisms of Ancestral Histone Protein Inheritance in Budding Yeast. *PLOS Biol.* **9**, e1001075 (2011).
11. Probst, A. V., Dunleavy, E. & Almouzni, G. Epigenetic inheritance during the cell cycle. *Nat. Rev. Mol. Cell Biol.* **10**, 192–206 (2009).
12. Petryk, N. *et al.* MCM2 promotes symmetric inheritance of modified histones during DNA replication. *Science* **361**, 1389–1392 (2018).
13. Yu, C. *et al.* A mechanism for preventing asymmetric histone segregation onto replicating DNA strands. *Science* **361**, 1386–1389 (2018).
14. Prior, C. P., Cantor, C. R., Johnson, E. M. & Allfrey, V. G. Incorporation of exogenous pyrene-labeled histone into Physarum chromatin: a system for studying changes in nucleosomes assembled in vivo. *Cell* **20**, 597–608 (1980).
15. Jackson, V. Deposition of newly synthesized histones: hybrid nucleosomes are not tandemly arranged on daughter DNA strands. *Biochemistry* **27**, 2109–2120 (1988).
16. Kimura, H. & Cook, P. R. Kinetics of Core Histones in Living Human Cells. *J. Cell Biol.* **153**, 1341–1354 (2001).
17. Xu, M. *et al.* Partitioning of Histone H3-H4 Tetramers During DNA Replication–Dependent Chromatin Assembly. *Science* (2010) doi:10.1126/science.1178994.
18. Campos, E. I., Stafford, J. M. & Reinberg, D. Epigenetic inheritance: histone bookmarks across generations. *Trends Cell Biol.* **24**, 664–674 (2014).
19. Francis, N. J. & Sihou, D. Inheritance of Histone (H3/H4): A Binary Choice? *Trends Biochem. Sci.* **46**, 5–14 (2021).

20. Mendiratta, S., Gatto, A. & Almouzni, G. Histone supply: Multitiered regulation ensures chromatin dynamics throughout the cell cycle. *J. Cell Biol.* **218**, 39–54 (2018).
21. Reinberg, D. & Vales, L. D. Chromatin domains rich in inheritance. *Science* **361**, 33–34 (2018).
22. Zhang, T., Cooper, S. & Brockdorff, N. The interplay of histone modifications - writers that read. *EMBO Rep.* **16**, 1467–1481 (2015).
23. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229–244 (2018).
24. Alabert, C. *et al.* Domain Model Explains Propagation Dynamics and Stability of Histone H3K27 and H3K36 Methylation Landscapes. *Cell Rep.* **30**, 1223-1234.e8 (2020).
25. Alabert, C. *et al.* Two distinct modes for propagation of histone PTMs across the cell cycle. *Genes Dev.* **29**, 585–590 (2015).
26. Reverón-Gómez, N. *et al.* Accurate Recycling of Parental Histones Reproduces the Histone Modification Landscape during DNA Replication. *Mol. Cell* **72**, 239-249.e5 (2018).
27. Nickels, J. F. *et al.* Establishment of heterochromatin in domain-size-dependent bursts. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2022887118 (2021).
28. Escobar, T. M. *et al.* Active and Repressed Chromatin Domains Exhibit Distinct Nucleosome Segregation during DNA Replication. *Cell* **179**, 953-963.e11 (2019).
29. Escobar, T. M. *et al.* Inheritance of repressed chromatin domains during S phase requires the histone chaperone NPM1. *Sci. Adv.* **8**, eabm3945 (2022).
30. Gruszka, D. T., Xie, S., Kimura, H. & Yardimci, H. Single-molecule imaging reveals control of parental histone recycling by free histones during DNA replication. *Sci. Adv.* **6**, eabc0330 (2020).
31. Liu, Z. & Tjian, R. Visualizing transcription factor dynamics in living cells. *J. Cell Biol.* **217**, 1181–1191 (2018).
32. Di Bona, M. *et al.* Measuring Mobility in Chromatin by Intensity-Sorted FCS. *Biophys. J.* **116**, 987–999 (2019).
33. Bhattacharya, D., Mazumder, A., Miriam, S. A. & Shivashankar, G. V. EGFP-Tagged Core and Linker Histones Diffuse via Distinct Mechanisms within Living Cells. *Biophys. J.* **91**, 2326–2336 (2006).
34. Bohn, M. & Heermann, D. W. Diffusion-Driven Looping Provides a Consistent Framework for Chromatin Organization. *PLOS ONE* **5**, e12218 (2010).
35. Whitefield, D. B., Spagnol, S. T., Armiger, T. J., Lan, L. & Dahl, K. N. Quantifying site-specific chromatin mechanics and DNA damage response. *Sci. Rep.* **8**, 18084 (2018).
36. Shi, G., Liu, L., Hyeon, C. & Thirumalai, D. Interphase human chromosome exhibits out of equilibrium glassy dynamics. *Nat. Commun.* **9**, 3161 (2018).
37. Shinkai, S., Nozaki, T., Maeshima, K. & Togashi, Y. Dynamic Nucleosome Movement Provides Structural Information of Topological Chromatin Domains in Living Human Cells. *PLoS Comput. Biol.* **12**, e1005136 (2016).
38. Dodd, I. B., Micheelsen, M. A., Sneppen, K. & Thon, G. Theoretical Analysis of Epigenetic Cell Memory by Nucleosome Modification. *Cell* **129**, 813–822 (2007).
39. Zerihun, M. B., Vaillant, C. & Jost, D. Effect of replication on epigenetic memory and consequences on gene transcription. *Phys. Biol.* **12**, 026007 (2015).
40. Jadhav, U. *et al.* Replicational Dilution of H3K27me3 in Mammalian Cells and the Role of Poised Promoters. *Mol. Cell* **78**, 141-151.e5 (2020).

41. Hall, I. M. *et al.* Establishment and Maintenance of a Heterochromatin Domain. *Science* **297**, 2232–2237 (2002).
42. Wang, X. & Moazed, D. DNA sequence-dependent epigenetic inheritance of gene silencing and histone H3K9 methylation. *Science* **356**, 88–91 (2017).
43. Hathaway, N. A. *et al.* Dynamics and Memory of Heterochromatin in Living Cells. *Cell* **149**, 1447–1460 (2012).
44. Taneja, N. *et al.* SNF2 Family Protein Fft3 Suppresses Nucleosome Turnover to Promote Epigenetic Inheritance and Proper Replication. *Mol. Cell* **66**, 50-62.e6 (2017).
45. Holla, S. *et al.* Positioning Heterochromatin at the Nuclear Periphery Suppresses Histone Turnover to Promote Epigenetic Inheritance. *Cell* **180**, 150-164.e15 (2020).
46. Ahmad, K., Henikoff, S. & Ramachandran, S. Managing the Steady State Chromatin Landscape by Nucleosome Dynamics. *Annu. Rev. Biochem.* **91**, null (2022).
47. Ramakrishnan, N., Pillai, S. R. B. & Padinhateeri, R. High fidelity epigenetic inheritance: Information theoretic model predicts threshold filling of histone modifications post replication. *PLOS Comput. Biol.* **18**, e1009861 (2022).
48. Khrameeva, E. E., Mironov, A. A., Fedonin, G. G., Khaitovich, P. & Gelfand, M. S. Spatial Proximity and Similarity of the Epigenetic State of Genome Domains. *PLOS ONE* **7**, e33947 (2012).
49. Mirny, L. *et al.* How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. Math. Theor.* **42**, 434013 (2009).
50. Broek, B. van den, Lomholt, M. A., Kalisch, S.-M. J., Metzler, R. & Wuite, G. J. L. How DNA coiling enhances target localization by proteins. *Proc. Natl. Acad. Sci.* **105**, 15738–15742 (2008).
51. Gurard-Levin, Z. A., Quivy, J.-P. & Almouzni, G. Histone Chaperones: Assisting Histone Traffic and Nucleosome Dynamics. *Annu. Rev. Biochem.* **83**, 487–517 (2014).
52. Zhou, C. Y., Johnson, S. L., Gamarra, N. I. & Narlikar, G. J. Mechanisms of ATP-Dependent Chromatin Remodeling Motors. *Annu. Rev. Biophys.* **45**, 153–181 (2016).
53. Marchetti, M. C. *et al.* Hydrodynamics of soft active matter. *Rev. Mod. Phys.* **85**, 1143–1189 (2013).
54. Rhind, N. & Gilbert, D. M. DNA Replication Timing. *Cold Spring Harb. Perspect. Biol.* **5**, a010132 (2013).
55. Micheelsen, M. A., Mitarai, N., Sneppen, K. & Dodd, I. B. Theory for the stability and regulation of epigenetic landscapes. *Phys. Biol.* **7**, 026010 (2010).
56. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
57. Orford, K. *et al.* Differential H3K4 Methylation Identifies Developmentally Poised Hematopoietic Genes. *Dev. Cell* **14**, 798–809 (2008).
58. Takaki, R., Dey, A., Shi, G. & Thirumalai, D. Theory and simulations of condensin mediated loop extrusion in DNA. *Nat. Commun.* **12**, 5865 (2021).
59. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci.* **112**, E6456–E6465 (2015).
60. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci.* **115**, E6697–E6706 (2018).

61. Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* (2019) doi:10.1126/science.aaz3418.
62. Ganji, M. *et al.* Real-time imaging of DNA loop extrusion by condensin. *Science* (2018) doi:10.1126/science.aar7831.
63. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).
64. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
65. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
66. Kind, J. *et al.* Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153**, 178–192 (2013).
67. Bickmore, W. A. & van Steensel, B. Genome architecture: domain organization of interphase chromosomes. *Cell* **152**, 1270–1284 (2013).
68. Ganai, N., Sengupta, S. & Menon, G. I. Chromosome positioning from activity-based segregation. *Nucleic Acids Res.* **42**, 4145–4159 (2014).
69. Agrawal, A., Ganai, N., Sengupta, S. & Menon, G. I. Nonequilibrium Biophysical Processes Influence the Large-Scale Architecture of the Cell Nucleus. *Biophys. J.* **118**, 2229–2244 (2020).
70. Wooten, M. *et al.* Asymmetric histone inheritance via strand-specific incorporation and biased replication fork movement. *Nat. Struct. Mol. Biol.* **26**, 732–743 (2019).
71. Brennan, L. D., Forties, R. A., Patel, S. S. & Wang, M. D. DNA looping mediates nucleosome transfer. *Nat. Commun.* **7**, 13337 (2016).
72. Clément, C. *et al.* High-resolution visualization of H3 variants during replication reveals their controlled recycling. *Nat. Commun.* **9**, 3181 (2018).
73. Saksouk, N., Simboeck, E. & Déjardin, J. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin* **8**, 3 (2015).
74. Oksuz, O. *et al.* Capturing the Onset of PRC2-Mediated Repressive Domain Formation. *Mol. Cell* **70**, 1149-1162.e5 (2018).
75. Kraft, K. *et al.* Polycomb-mediated genome architecture enables long-range spreading of H3K27 methylation. *Proc. Natl. Acad. Sci.* **119**, e2201883119 (2022).
76. Lavigne, M., Francis, N. J., King, I. F. G. & Kingston, R. E. Propagation of Silencing: Recruitment and Repression of Naive Chromatin In trans by Polycomb Repressed Chromatin. *Mol. Cell* **13**, 415–425 (2004).
77. Jamai, A., Puglisi, A. & Strubin, M. Histone Chaperone Spt16 Promotes Redeposition of the Original H3-H4 Histones Evicted by Elongating RNA Polymerase. *Mol. Cell* **35**, 377–383 (2009).
78. Jeronimo, C., Poitras, C. & Robert, F. Histone Recycling by FACT and Spt6 during Transcription Prevents the Scrambling of Histone Modifications. *Cell Rep.* **28**, 1206-1218.e8 (2019).
79. Formosa, T. & Winston, F. The role of FACT in managing chromatin: disruption, assembly, or repair? *Nucleic Acids Res.* **48**, 11929–11941 (2020).
80. Ehara, H., Kujirai, T., Shirouzu, M., Kurumizaka, H. & Sekine, S.-I. Structural basis of nucleosome disassembly and reassembly by RNAPII elongation complex with FACT. *Science* **377**, eabp9466 (2022).

81. Svensson, J. P. *et al.* A nucleosome turnover map reveals that the stability of histone H4 Lys20 methylation depends on histone recycling in transcribed chromatin. *Genome Res.* **25**, 872–883 (2015).
82. Vasseur, P. *et al.* Dynamics of Nucleosome Positioning Maturation following Genomic Replication. *Cell Rep.* **16**, 2651–2665 (2016).
83. Suter, D. M. *et al.* Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. *Science* **332**, 472–474 (2011).
84. Dar, R. D. *et al.* Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci.* **109**, 17454–17459 (2012).
85. Klein, K. N. *et al.* Replication timing maintains the global epigenetic state in human cells. *Science* **372**, 371–378 (2021).
86. Shi, G. & Thirumalai, D. From Hi-C Contact Map to Three-Dimensional Organization of Interphase Human Chromosomes. *Phys. Rev. X* **11**, 011051 (2021).
87. Kumari, K., Duenweg, B., Padinhateeri, R. & Prakash, J. R. Computing 3D Chromatin Configurations from Contact Probability Maps by Inverse Brownian Dynamics. *Biophys. J.* **118**, 2193–2208 (2020).