

## RESEARCH ARTICLE

# High fidelity epigenetic inheritance: Information theoretic model predicts threshold filling of histone modifications post replication

Nithya Ramakrishnan<sup>1</sup>, Sibi Raj B. Pillai<sup>2\*</sup>, Ranjith Padinhateeri<sup>1\*</sup>

**1** Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Mumbai, India,

**2** Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India

\* [bsraj@ee.iitb.ac.in](mailto:bsraj@ee.iitb.ac.in) (SRBP); [ranjithp@iitb.ac.in](mailto:ranjithp@iitb.ac.in) (RP)



## OPEN ACCESS

**Citation:** Ramakrishnan N, Pillai SRB, Padinhateeri R (2022) High fidelity epigenetic inheritance: Information theoretic model predicts threshold filling of histone modifications post replication. *PLoS Comput Biol* 18(2): e1009861. <https://doi.org/10.1371/journal.pcbi.1009861>

**Editor:** Vladimir B Teif, University of Essex, UNITED KINGDOM

**Received:** June 28, 2021

**Accepted:** January 24, 2022

**Published:** February 17, 2022

**Copyright:** © 2022 Ramakrishnan 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.

**Data Availability Statement:** All relevant data are within the manuscript and its [Supporting information](#) files. The software code in R and associated documentation are available online (<https://github.com/nitka/threshold-filling>).

**Funding:** This study was supported by the Science and Engineering Research Board, Department of Science and Technology India, (RP), Grant number: EMR/2016/005965 and by the Department of Biotechnology, Ministry of Science and Technology India, (RP), Grant number: BT/

## Abstract

During cell division, maintaining the epigenetic information encoded in histone modification patterns is crucial for survival and identity of cells. The faithful inheritance of the histone marks from the parental to the daughter strands is a puzzle, given that each strand gets only half of the parental nucleosomes. Mapping DNA replication and reconstruction of modifications to equivalent problems in communication of information, we ask how well enzymes can recover the parental modifications, if they were ideal computing machines. Studying a parameter regime where realistic enzymes can function, our analysis predicts that enzymes may implement a critical threshold filling algorithm which fills unmodified regions of length at most  $k$ . This algorithm, motivated from communication theory, is derived from the maximum a posteriori probability (MAP) decoding which identifies the most probable modification sequence based on available observations. Simulations using our method produce modification patterns similar to what has been observed in recent experiments. We also show that our results can be naturally extended to explain inheritance of spatially distinct antagonistic modifications.

## Author summary

Chromatin is essentially the DNA that is folded and packaged with the help of proteins. While the nucleotide sequence in the DNA codes genetic information, the packaging of the DNA into chromatin encodes extra layer of information. This epigenetic code regulates reading of the genetic code and provides identity to cells—whether the cell is a skin cell or a brain cell, for example. In this work we examine a long standing puzzle, that is, how the epigenetic code in the form of histone modification patterns may get inherited, when a cell divides. Using theoretical arguments, we present an algorithm that enzymes could be executing so that the epigenetic code can be inherited with minimal error, soon after DNA replication.

HRD/NBA/39/12/2018-19. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

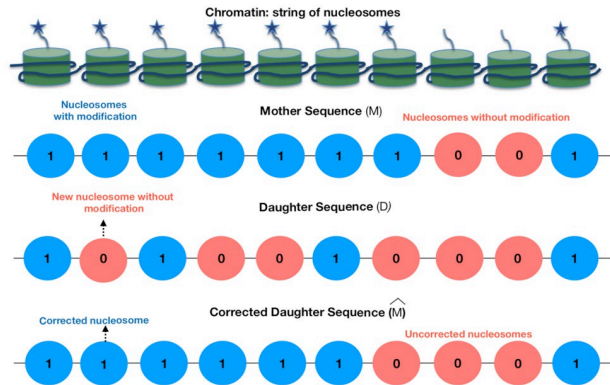
## Introduction

Why do our bone cells behave very differently from our muscle cells or cells of other types even though they all have the same genetic code? To explain the emergence of such diverse cell types, one would need to note that, beyond the genetic code, there are multiple layers of information encoded by the wrapping and folding of the DNA into chromatin with the help of many proteins [1, 2]. Most of the DNA is wrapped around octamers of histone proteins, making the chromatin, essentially, like a string of beads made of nucleosomes (DNA+histones) [3–6]. Each nucleosome carries chemical modifications, like acetylations and methylations [7], forming a pattern of histone marks along the chromatin polymer contour [2, 8–11] (see Fig 1A top panel). This pattern encodes a crucial layer of information regulating accessibility, transcription, replication and other cellular processes [11]. Even though the entire histone modification code is not deciphered yet, we understand it in parts. For example, H3K27me3/H3K9me3 represses reading of the local DNA region where the modification is present, H3K9ac/H3K4me3 encodes for local gene activation and so on [7, 12, 13]. The activation and repression of genes collectively decide the gene expression pattern and hence determine the function and fate of a cell [10, 14–16].

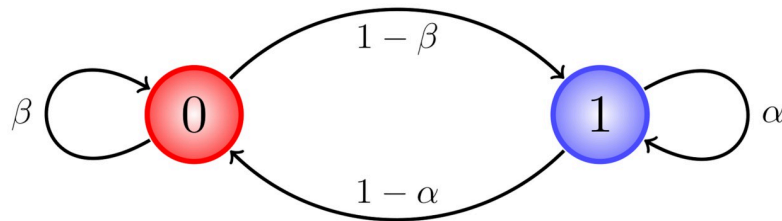
While preparing to divide, cells copy their genetic code via the DNA replication process. For DNA to be copied, the chromatin has to be unfolded and histone proteins need to be disassembled [17, 18]. This would disrupt the pattern of histone modifications. Recent studies have shown that the (H3 – H4)<sub>2</sub> tetramer from the parent remains intact [19] and randomly gets deposited onto either of the newly synthesized DNA strands [20–22]. That is, daughter strands will have only some ( $\approx 50\%$ ) of its nucleosomes from the parent; the rest need to be assembled from the pool of new histone proteins made afresh [23, 24]. Since the new nucleosomes will not carry the histone marks present on the parental chromatin, half the parental marks are missing in each daughter chromatin [25]. Since the histone modification patterns can decide the state (repressed/active) of all genes and the identity of the cell itself, it is crucial that the newly made chromatin re-establishes the pattern immediately after replication [26]. Recent experiments show that many of the histone modification patterns—patterns in the repressed gene regions, in particular—are “inherited” from the parental chromatin [27–31]. We know from literature that the heterochromatin region is inherited faithfully during both mitosis and meiosis [27, 32, 33]. It is known that there are specific enzymes to read and write histone marks [31, 33–35]. While molecular details of some of these enzymes are known [36–38], precisely what strategies they use to re-establish the histone modification pattern after replication are not fully understood.

Previous theoretical studies have investigated average properties of modifications and whether long-range interactions are necessary or short-range interactions would suffice [39, 40]. Dodd et al [40] present a stochastic model and investigate the bistability similar to what is observed in silent mating-type region in *S. pombe*. This model has been further extended to discuss the likelihood of bivalency and bistability in poised chromatin [41]. Additionally, there have also been various models exploring the maintenance of chromatin state in specific types of cells [42–44]. Another set of models investigated establishment of stable modification patterns near a nucleation site [45, 46]. The antagonism between certain modification pairs, leading to domains in the chromatin, have also been quantitatively explored [47]. Several other studies have discussed the possible nature of the spread of the histone modifications and maintenance of the chromatin states [48–52]. However, all aspects of the experimentally observed [28] high-fidelity re-establishment of the modification patterns post replication along the genome (spatial pattern) is not explained by the existing models.

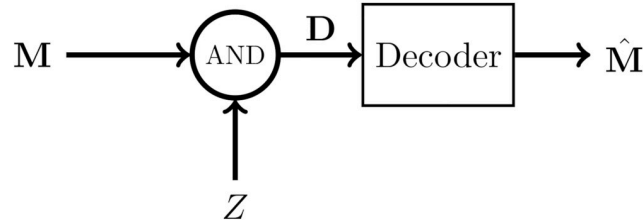
(A)



(B)



(C)



**Fig 1.** Schematic description of the problem (A) Row 1 from top: chromatin as a string of nucleosomes with and without the histone modification (star) of our interest. This can be mapped to a string of binary numbers indicating the presence (1) or absence (0) of the modification (row 2), giving us  $M$ . Row 3: one typical realization of a daughter chromatin  $D$  produced from  $M$  above, via a process mimicking DNA replication, where only a fraction of the modifications (1s) will end up in the daughter chromatin, stochastically; the rest do not have the modification of interest (0) [25]. Post replication, certain enzymes will insert modifications correcting  $D$  to a mother-like sequence  $\hat{M}$  (row 4). Since these are stochastic processes, we expect some errors. (B) The mother sequence ( $M$ ) is modeled as a first order Markov chain having sequence of 0s and 1s.  $\alpha$  and  $\beta$  are probabilities of finding a 1 followed by a 1, and a 0 followed by a 0, respectively.  $1 - \alpha$  and  $1 - \beta$  are probabilities of finding a 1 followed by a 0 (note arrowheads), and a 0 followed by a 1, respectively. (C) The daughter sequence  $D$  is obtained by a mother sequence  $M$  getting logically ANDed with an independent and identically distributed (IID) binary sequence  $Z$  (noise). A mother-like sequence  $\hat{M}$  is reconstructed by passing  $D$  through a decoder. The plausible ways by which enzymes could act as decoders is the subject of this study.

<https://doi.org/10.1371/journal.pcbi.1009861.g001>

In the past, information theory has been applied to biological systems to study several aspects of gene regulation and signalling [53–55]. The problem of loss of information in the modification pattern during replication and its retrieval within the daughter chromatin is very similar to data loss and error correction in telecommunication. In such systems, a transmitted signal gets exposed to noise and consequently becomes, error-prone at the receiving end. The decoder at the receiver detects and corrects these errors using techniques from information and communication theory [56, 57]. This viewpoint immediately poses the following questions: Can we use known decoding algorithms from communication theory to analyze chromatin modification loss and retrieval? How well can the best known algorithms correct the missing modifications and re-establish the modification patterns? What is the best possible correction strategy enzymes could use if they were ideal computing machines? Are these algorithms compatible with the biological processes that realistic cellular enzymes can conceivably do? In this paper, we address these questions using ideas from information theory. We consider one of the daughter chromatins to be a noise-corrupted signal created at the replication fork, while the enzymes and other molecular agents help to correct this error using mathematical techniques. In this model, the inheritance of the mother's pattern is approached using Bayesian decoding techniques. We show that our model can reconstruct histone modification patterns present in publicly available experimental data [28], suggesting that the model is relevant.

## Model and methods

Consider a region on a mother chromatin having  $N$  nucleosomes. We are interested in studying the inheritance of one histone modification at a time. Since many of the repressive marks are known to be inherited accurately [28] after replication, we will consider one such repressive mark (e.g., H3K27me3) and its pattern along a chromatin. This pattern can be represented by a vector  $\mathbf{M} = (m_1, m_2, \dots, m_N)$ , where  $m_i$  can take values 1 or 0 indicating the presence or absence of the modification on the  $i^{\text{th}}$  nucleosome (see Fig 1A).

For brevity, we use the following notations:

- $m_i^j$  represents the sequence of modification values  $(m_i, m_{i+1}, \dots, m_j)$  between location  $i$  and  $j$  with  $j > i$ ; Thus,  $\mathbf{M} = m_1^N$  represents the entire mother chromatin modification sequence.
- a region with  $k$  consecutive modifications present (ones) or absent (zeros) will be denoted as  $1_k$  or  $0_k$ , respectively. Extending this, the sequence  $(1, 0, \dots, 0, 1)$  representing an island of  $k$  consecutive zeros between two ones will be denoted as  $(1, 0_k, 1)$ .

Models with nearest neighbor interactions (Markov model of order 1) are commonly used to represent the genetic/epigenetic sequences and other biophysical phenomena like protein-binding [58]. For instance, Zhang et al [39] have used a nearest neighbor interaction model to study epigenetic memory. The modeling of experimental data by Hodges and Crabtree [45] also assumes that each nucleosome interacts with only its nearest neighbour and that the modification is passed on to the nearest neighbour with a certain probability. Poepsel et al [59] show that histone methyl transferase protein PRC2 simultaneously engages with two nucleosomes—one unmodified nucleosome and one H3K27me3-containing nucleosome—suggesting a modification spreading mechanism among neighbouring nucleosomes. Many of the existing models have the underlying assumption of nearest neighbor interactions (Markov model of order 1). In our work, we make a similar assumption: Since modification on a nucleosome is very likely related to its immediate neighbors, we model the pattern  $\mathbf{M}$  along the mother chromatin as a binary-valued random walk, having neighbourhood interactions corresponding to a first order homogeneous Markov chain. More specifically, given the

modifications  $m_{i-1}$  and  $m_{i+1}$ , the modification  $m_i$  is assumed to be independent of all other modification values. Equivalently, the conditional probability law is

$$\mathbb{P}(m_i | m_1, \dots, m_{i-1}) = \mathbb{P}(m_i | m_{i-1}), \quad i \geq 2, \tag{1}$$

where  $m_1$  is the modification on the first nucleosome of the region of our interest. This is also similar to the 1-dimensional Ising model where the interaction energy of the  $i^{\text{th}}$  site is coupled only to the immediate neighboring sites [60, 61]. The state-space evolution of the Markov chain  $\mathbf{M}$  is as follows: given  $m_i = 1$ , let  $\alpha$  and  $1 - \alpha$  be the probabilities for obtaining  $m_{i+1} = 1$  and  $m_{i+1} = 0$  respectively. Similarly, if  $m_i = 0$ , let  $\beta$  and  $1 - \beta$  be the probabilities for having  $m_{i+1} = 0$  and  $m_{i+1} = 1$  respectively. The sequence  $\mathbf{M}$  can be seen as a random walk on the state space shown in Fig 1B. The parameters  $\alpha$  and  $\beta$  are functions of the mean contiguous length of the modified and unmodified regions respectively and can be computed from the experimental data. For example, when  $\alpha$  and  $\beta$  values are close to 1, the pattern would often contain long runs of either 1s (presence of modification) or 0s (absence of modification). See S1 and S2 Text.

From the mother chromatin  $\mathbf{M}$ , the generation of a daughter chromatin having histone modification sequence  $\mathbf{D} = d_1^N$  is modeled as follows. During replication, with probability  $\frac{1}{2}$ , each nucleosome on a daughter chromatin is either directly inherited from its parental counterpart (i.e.  $d_i = m_i$ ) or newly deposited (i.e.  $d_i = 0$ ) from a pool of fresh histones assembled de novo [25]. We consider both the histones in the  $(H3 - H4)_2$  tetramer to be symmetrically modified in our model based, on recent evidence that a large proportion of the histones are symmetrically modified [19, 62]. This process is equivalent to doing a logical AND operation of the mother sequence  $\mathbf{M}$  with an independent binary vector  $\mathbf{Z}$  (noise), which is generated by independent tosses of a fair coin. Thus,  $d_i = m_i \cdot z_i$ ,  $1 \leq i \leq N$  (see Fig 1C), where  $\mathbf{Z} = (z_1, z_2, z_3, \dots, z_N)$  has Independent and Identically Distributed (IID) entries. This biological process leads to a memoryless model with the conditional probability

$$\mathbb{P}(d_1^N | m_1^N) = \prod_{i=1}^N \mathbb{P}(d_i | m_i). \tag{2}$$

Owing to the independent and random nature of the placement of the parental nucleosomes along the newly formed daughter chromatin, the above relation is justified. In biology, the question is, given a daughter sequence  $d_1^N$  post replication, how can a cell reconstruct a mother-like sequence  $\hat{\mathbf{M}} = \hat{m}_1^N$ ? In other words, is it possible to build a decoder that would reconstruct a  $\hat{\mathbf{M}}$  from  $\mathbf{D}$ , as depicted in Fig 1C? Ideally, a cell would want to choose a binary sequence  $\hat{\mathbf{M}}$  having the minimum deviation from  $\mathbf{M}$ . A simple way to quantify this deviation is to compute the fraction of errors in the reconstructed sequence as:

$$\Delta(\mathbf{M}, \hat{\mathbf{M}}) = \frac{1}{N} \sum_{i=1}^N (m_i - \hat{m}_i)^2. \tag{3}$$

Since we are comparing bitwise, this deviation metric is effectively the bit error rate (BER) when  $N$  becomes large [63]. In communication, BER is the proportion of bits that are in error among the total transmitted/received bits. In the present context, BER is essentially the fraction of nucleosomes in a daughter sequence that differ in their histone mark from the parental sequence. Thus the chosen  $\hat{\mathbf{M}}$  should minimize the BER with respect to the actual sequence  $\mathbf{M}$ , while obeying the transition law in Eq (1). This is similar to data communication through an erroneous channel. It is well known that Bayesian estimation schemes minimize the average detection error probability at the receiver. In particular, a decoder choosing the input sequence

having the Maximum A posteriori Probability (MAP) is optimal in minimizing the message error probability in communication [57, 63] (see S1 Text). We call this the Sequence MAP (SMAP) decoder, which identifies the most probable sequence  $\hat{\mathbf{M}} = (\hat{m}_1, \dots, \hat{m}_N)$  based on the observations  $d_1^N$  as

$$(\hat{m}_1, \dots, \hat{m}_N) = \operatorname{argmax}_{m_1, \dots, m_N} \mathbb{P}(m_1^N | d_1^N). \tag{4}$$

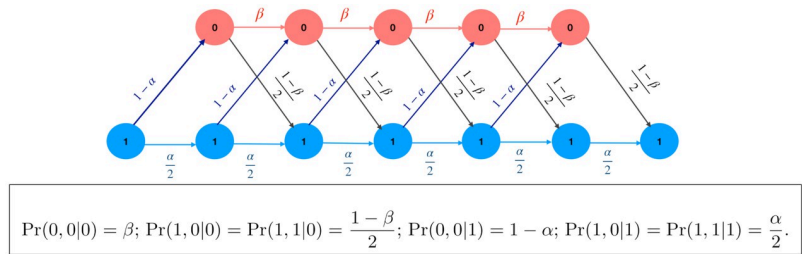
The above equation essentially says that, given a daughter sequence, the sequence that maximizes the conditional probability should be declared as the reconstructed mother-like sequence. SMAP decoding is known to have very good BER performance and good analytical tractability in many contexts [63]. Though SMAP decoding is not directly minimizing BER, it nevertheless is known to have good BER performance as well.

**Implementation of SMAP algorithm:** To implement the SMAP algorithm, one has to maximize the conditional probability  $\mathbb{P}(\mathbf{M}|\mathbf{D})$  over possible sequences of  $\mathbf{M}$ , based on the given daughter sequence observations  $\mathbf{D} = d_1^N$ . Applying Bayes' rule [64], along with Eqs (1) and (2) (see S1 Text) one gets

$$\mathbb{P}(\mathbf{M}|\mathbf{D}) = \frac{1}{\mathbb{P}(\mathbf{D})} \prod_{i=1}^N \mathbb{P}(m_i | m_{i-1}) \mathbb{P}(d_i | m_i), \tag{5}$$

where we took  $m_0 = \emptyset$  (empty set) for notational convenience. Since  $\mathbb{P}(\mathbf{D})$  does not play a role in the maximization over  $\mathbf{M}$ , it can be ignored for our purposes. The equation says that the reconstruction process of the mother-like sequence needs to account for two facts: the modification status of the immediate neighbours ( $\mathbb{P}(m_i | m_{i-1})$ ) and the information embedded in the replication process ( $\mathbb{P}(d_i | m_i)$ ). Knowing Eq (5), ideal computing machines can now implement the SMAP algorithm using the idea of trellis decoding [56], which is closely related to the well known Viterbi Algorithm in coding theory [63]. While the memory and computational power requirement for trellis decoding is high in general, we find that decoding procedure for our model can be broken down to smaller sub-sequences, each corresponding to a different run of zeros in  $\mathbf{D}$ . In particular, SMAP decoding can be applied separately on sub-sequences of the form  $(1, 0_k, 1)$ , which has  $k$  consecutive zeros in between two ones. Mathematically, we can show that SMAP decoding will choose a sequence having  $\hat{m}_i^j$  according to  $\hat{m}_i, \dots, \hat{m}_j = \operatorname{argmax} \mathbb{P}(m_i, \dots, m_j | d_i = 1, d_{i+1}^{j-1}, d_j = 1)$ ; here  $i$  and  $j$  be two positions (with  $j > i$ ) where the daughter sequence has ones (see S1 and S2 Text. Our analysis suggests that to decide on a bit at position  $l$  where the daughter has inherited a zero, we need to only consider the smallest daughter segment containing the position  $l$ , and flanked by ones at both ends. Only daughter segments with at least one intermediate zero are to be considered; otherwise there is nothing to decode. Without loss of generality, we will take  $\mathbf{D} = (1, 0_k, 1)$  for the rest of the exposition, corresponding to a run of  $k$  zeros, and perform trellis decoding on this sequence.

**Trellis Decoding:** Trellis decoding [56] is a technique that can identify the most probable mother-like histone modification sequence, according to the theory described above. In this study, a sequence of states of a Markov chain (here, histone modification sequence) is called a path or a trajectory. Given the states at the start and end of a possible path, a trellis diagram [56] can be used to find the joint probability of each path with the given observation  $\mathbf{D}$ . Since  $\mathbf{D}$  has the form  $(1, 0_k, 1)$ , the start and end states of the trellis are ones. Fig 2 demonstrates the trellis diagram for a run of 5 zeros ( $k = 5$ ). Starting from the initial state 1 (left bottom in Fig 2), the trellis diagram assigns a conditional probability (branch metric) to each subsequent transition (arrow), based on the transition probability law and observed daughter state. For



**Fig 2. The trellis diagram used for illustrating how the MAP algorithm chooses the path of maximum a posteriori probability given a daughter sequence (1, 0<sub>5</sub>, 1).** From each Markov state (0 or 1), there are two possible arrows to transition to the next state. The values  $\mathbb{P}(m_i, d_i|m_{i-1})$  given in the box represent probabilities associated with each arrow. Starting from state 1 at the left bottom, for each possible path moving along the arrows, one can compute the path probabilities (path metrics) using Eq (5). We finally choose the path with the largest path metric. This is based on the Viterbi algorithm in communication theory.

<https://doi.org/10.1371/journal.pcbi.1009861.g002>

transitions from state at  $i - 1$  to  $i$ , the branch metric is  $\mathbb{P}(m_i, d_i|m_{i-1})$ , which can be evaluated as  $\mathbb{P}(m_i|m_{i-1})\mathbb{P}(d_i|m_i)$ , where  $d_i = 0$  for  $2 \leq i \leq N - 1$ . While the sequence  $\mathbf{D}$  is easily seen to be generated by a hidden Markov Model (HMM), the branch probability metrics are explicitly given inside the box of Fig 2 (also see S1 Text). Notice that we have to find  $\mathbb{P}(m_i, d_i = 0|m_{i-1})$  for  $(m_{i-1}, m_i) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ , and  $\mathbb{P}(m_i = 1, d_i = 1|m_{i-1})$  for  $m_{i-1} \in \{0, 1\}$  for the last stage (see Fig 2). The computation of the branch and path metrics and the choice of the path with the largest path metric is performed using the Viterbi Algorithm in communication theory [56].

Notice that each possible mother sequence can be identified as a path in the trellis, with the labels identifying the branch metrics. Using Eq (5), the product of corresponding branch metrics will yield the path metric of each possible sequence, and then the path maximizing the SMAP metric can be chosen.

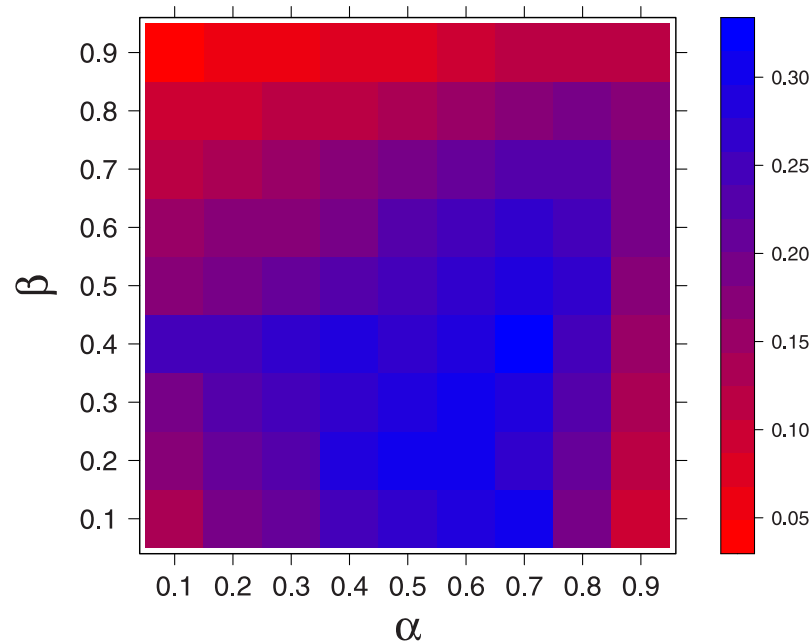
### Results

In this section, we will be using the ideas developed in the Model and Methods to answer how one can reconstruct a mother-like modification sequence  $\hat{\mathbf{M}}$ , given the daughter chromatin sequence  $\mathbf{D}$ . We will discuss how well algorithms like the Sequence MAP (SMAP) decoding will compute  $\hat{\mathbf{M}}$ , and whether realistic enzymes can implement this in practice.

### Ideal enzymes implementing SMAP decoding

From recent experiments, it has been established that a plethora of enzymes and histone chaperones (including FACT, MCM2) are involved in the inheritance of the histone modifications post replication [20, 21, 31]. Yet, how the enzyme machinery uses the parental information and tries to reproduce the exact pattern of the histone modification, post-replication, is not clear. In this work, we discuss how well can the communication theory-inspired algorithms reconstruct a mother-like sequence. To test this, one can imagine some *ideal enzymes*—computing machines—constructed to implement the SMAP algorithm in Eq (4); they will achieve this by maximizing the conditional probability  $\mathbb{P}(\mathbf{M}|\mathbf{D})$  as mentioned above.

To do this, on a computer, we generated several mother sequences for different values of  $\alpha$  and  $\beta$  parameters in the Markov model. The details of how the mother sequences were generated are provided in S2 Text. We then generated several daughter sequences, for each of the mother sequences, by flipping 1s to 0s randomly with probability 0.5; that is,  $d_i = m_i \cdot z_i$ , with  $z_i$  representing an element in an IID binary sequence generated by an unbiased coin. Each of the



**Fig 3.** The average deviation between the original mother and the mother-like corrected daughter sequences ( $\bar{\Delta} =$  ensemble averaged  $\Delta(\mathbf{M}, \hat{\mathbf{M}})$ ) is plotted for different  $\alpha$  and  $\beta$  values as a heatmap (see color bar on the side). The error is averaged over the error of 300 mother sequences and 200 daughter sequences corresponding to each mother sequence—that is, 60000  $\Delta(\mathbf{M}, \hat{\mathbf{M}})$  values.  $\alpha$  represents the conditional probability of a nucleosome staying in 1 (modified state):  $P(m_i = 1 | m_{i-1} = 1)$ .  $\beta$  represents the conditional probability of a nucleosome staying in 0 (unmodified state):  $P(m_i = 0 | m_{i-1} = 0)$ .

<https://doi.org/10.1371/journal.pcbi.1009861.g003>

daughter sequences was corrected with the trellis-based SMAP decoding to generate the corresponding estimate  $\hat{\mathbf{M}}$ ; the error  $\Delta(\mathbf{M}, \hat{\mathbf{M}})$  was computed (Eq (3)). The average of  $\Delta(\mathbf{M}, \hat{\mathbf{M}}) = \bar{\Delta}$  (averaged over many realizations) for fixed pairs of  $\alpha$  and  $\beta$ , is presented as a heatmap in Fig 3. The mean deviation ( $\bar{\Delta}$ ) between mother and corrected mother-like daughter is low for very high values of  $\alpha$  and  $\beta$ —a region dominated by long islands of ones (modified nucleosomes) separated by long islands of zeros (unmodified nucleosomes). There are other regions too where  $\bar{\Delta}$  is relatively small—regions that appear more red in Fig 3—like regions with high  $\beta$  and low  $\alpha$  and vice-versa. (also see S2 Text, S1 Fig). Overall, this result shows how well an ideal computing machine that employs state of the art information theory inspired algorithms can recover the original mother sequence. The remaining question is, can a real enzyme do as good as this computing algorithm?

### Threshold- $k$ model: Enzymes filling unmodified islands of size at most $k$ maintain chromatin fidelity

Biological enzymes being equipped to do complex SMAP computations like trellis decoding by themselves is arguably hypothetical. Nevertheless, we show that in certain biologically relevant parameter regimes, the decoding rule can be simple enough for enzymes to potentially execute. Among the known histone modification patterns, it is common to have regions densely filled by a certain modification (e.g., H3K27me3), and other regions where the modification is totally absent. Since  $\alpha$  defines the probability of finding a modified nucleosome followed by another modified nucleosome, and  $\beta$  defines the probability of finding an unmodified nucleosome followed by another unmodified nucleosome, regions densely filled with a



given modification would correspond to higher value of  $\alpha$ ; similarly, regions where the modification is totally absent would correspond to higher values of  $\beta$  in our Markov model (see Fig 1B; also see S2 Text and S2 Fig). Below we show that in this regime, the SMAP algorithm simplifies to tasks that the enzymes may easily carry out.

Consider an island of  $k$  unmodified nucleosomes in the daughter chromatin, giving the pattern  $(1, 0_k, 1)$ . From the trellis diagram (Fig 2), it can be shown that, if

$(\frac{\alpha}{2})^{k+1} > \frac{1}{2}(1 - \alpha)\beta^{k-1}(1 - \beta)$ , the probability of having the all-one path  $(1, 1_k, 1)$  at the mother is greater than that of  $(1, 0_k, 1)$ . In other words, when the function  $g_k(\alpha, \beta) = (\frac{\alpha}{2})^{k+1} - \frac{1}{2}(1 - \alpha)\beta^{k-1}(1 - \beta)$  is positive, an all-ones path is preferred over a run of  $k$  zeros by SMAP decoding. We can characterize the values of  $\alpha$  and  $\beta$  for which the above condition holds true. The expression  $g_k(\alpha, \beta) = 0$  is easy to solve if we take  $k$  to be a real value, this yields the root  $k^*$  as:

$$k^* = \frac{\log\left(\frac{(1 - \alpha)(1 - \beta)}{(\alpha^2/2)}\right)}{\log(\alpha/2\beta)} + 1. \tag{6}$$

Notice that the solution for  $k^*$  is unique when  $0 < \alpha < 1$  and  $0 < \beta < 1$ . Since  $g_1(\alpha, \beta) = \frac{1}{4}\alpha^2 - \frac{1}{2}(1 - \alpha)(1 - \beta)$ , the condition  $g_1(\alpha, \beta) > 0$  implies that the sequence  $(1, 1, 1)$  is preferred over  $(1, 0, 1)$ . In addition, this condition will also imply that the numerator of Eq (6) is positive. The uniqueness of  $k^*$  will now have the following implications:

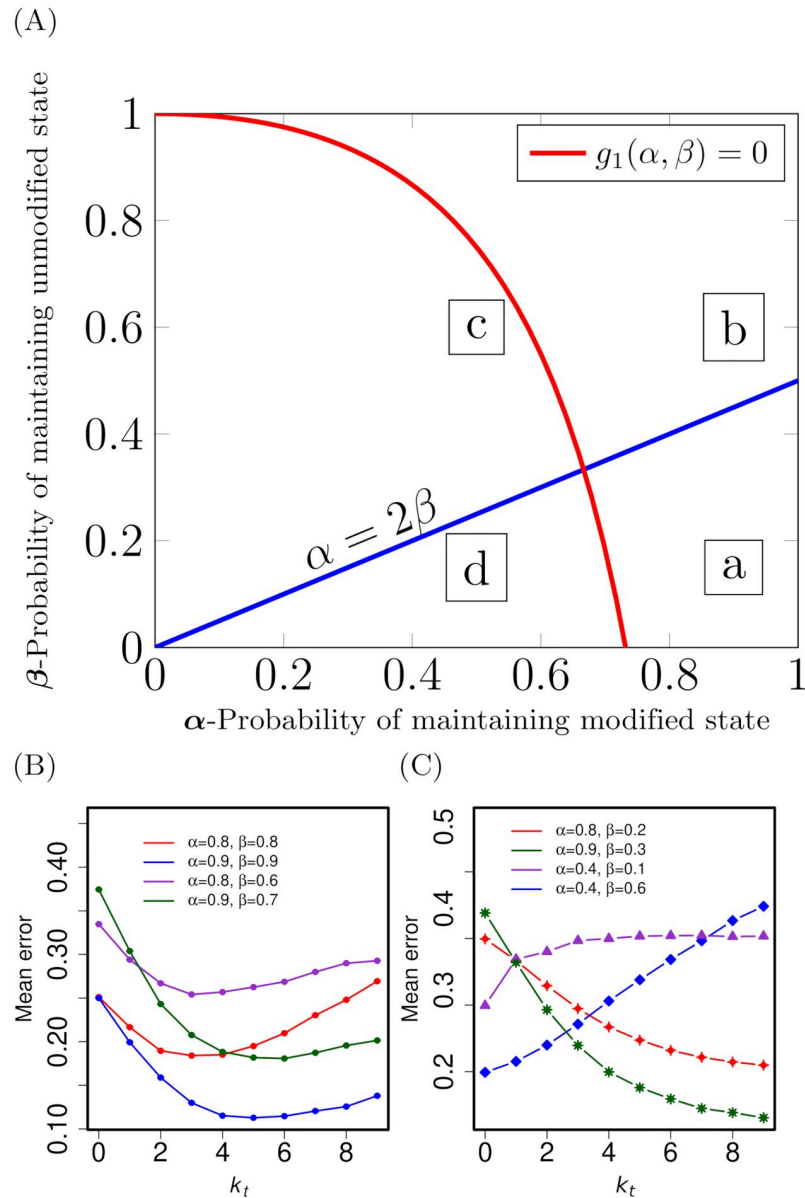
On a unit square region of parameters  $(\alpha, \beta) \in (0, 1) \times (0, 1)$ ,

- When  $\alpha < 2\beta$  and  $g_1(\alpha, \beta) > 0$ , one gets  $k^* \geq 1$ ; we find that  $g_k(\alpha, \beta) > 0$  for all positive integers  $k \leq k^*$  in Eq (6). This suggests that the SMAP algorithm will replace  $(1, 0_k, 1)$  with  $(1, 1_k, 1)$ , if and only if  $k \leq k^*$ .
- When  $\alpha > 2\beta$  and  $g_1(\alpha, \beta) > 0$ , we find that  $g_k(\alpha, \beta) > 0$  for any positive integer  $k$ ; hence the SMAP algorithm will replace  $(1, 0_k, 1)$  with  $(1, 1_k, 1)$ , for any value of  $k$ .

Notice that when every path of at most  $k^*$  zeros between two ones has less path metric than the corresponding all ones path, clearly any possible path other than all ones cannot have the maximum SMAP metric, while decoding sequences of length less than  $k^*$ .

The above analysis based on trellis decoding suggests two simple ways for enzymes to work. Enzymes of Type-I would simply modify all unmodified nucleosomes (0s) between two modified nucleosomes (1s). Such enzymes may be preferred when the modification pattern can be modeled by parameters  $\alpha$  and  $\beta$  that corresponds to region  $a$  in Fig 4A; notice that this has large  $\alpha$  and small  $\beta$ . An enzyme of Type-II would fill an unmodified nucleosome (0), if and only if it falls in the region of unmodified region of length atmost  $\leq k^*$ . That is, replace  $(1, 0_k, 1)$  by  $(1, 1_k, 1)$ , if the island size is  $k \leq k^*$ . Thus long islands of 0s are left unfilled. We call this a **threshold- $k$  filling model**, which becomes active in region  $b$  of Fig 4A. Notice that when both  $\alpha$  and  $\beta$  are close to 1, the modification is expected to have long domains (islands) with its presence, followed by islands with no modification. Biologically, this is a realistic regime for many modifications where enzymes can do threshold- $k$  filling.

We tested the threshold- $k$  filling model on a computer using the following procedure. For various values of  $\alpha$  and  $\beta$ , we generated several mother sequences using the Markov model. The daughter sequences were generated by simulating the replication process, by randomly flipping the non-zero values using independent realizations of an unbiased coin. (See Model and Methods and S2 Text). Each daughter sequence was corrected using the threshold- $k$  filling algorithm—that is, we filled all islands of 0s, having size  $k \leq k_t$ , by 1s; here  $k_t$  is taken as a variable. Corresponding mean error ( $\bar{\Delta}$ )—bit by bit difference between mother and daughter



**Fig 4.** Behavior of our algorithms across the parameter spaces (A) Four (a, b, c, d) regions in  $(\alpha, \beta)$  parameter space. The curves shown are  $g_1(\alpha, \beta) = 0$ , and  $\alpha - 2\beta = 0$ . In region a ( $g_1 > 0, \alpha > 2\beta$ ), the SMAP will replace every 0 with 1. In the region b ( $g_1 > 0, \alpha < 2\beta$ ), enzymes can implement threshold- $k$  filling (see text). This parameter regime is realistic, biologically. Region c has small values of  $\alpha$  and relatively higher values of  $\beta$ . As per the SMAP algorithm, this region will not be filled with 1s. In region d, where  $\alpha$  is very high compared to  $\beta$ , there is intermediate filling of 1s in the daughter sequence. (B) The mean error ( $\bar{\Delta}$ ) when we fill all islands of 0s having size at most  $k_t$ . All these curves have  $(\alpha, \beta)$  values in region b, and  $\bar{\Delta}$  is non-monotonic having a finite optimum  $k_t = k^*$ . (C)  $\bar{\Delta}$  for parameter values in region a (red and green curves) are monotonically decreasing suggesting that the optimal  $k_t$  is unbounded; hence the least error would be when all 0s are replaced with 1s. In regimes c, d (blue and violet curves) the mean error is minimal when nothing is filled suggesting that threshold- $k$  filling is not suitable here. The standard errors here are smaller than the size of the points.

<https://doi.org/10.1371/journal.pcbi.1009861.g004>

sequences—averaged over many realizations, was computed for a given  $(\alpha, \beta)$ . In Fig 4B, all the curves correspond to  $(\alpha, \beta)$  in regime b of Fig 4A ( $g_1 > 0$  and  $\alpha < 2\beta$ ). In this interesting regime, the mean error has a non-monotonic behaviour with a minimum at a particular value of  $k_t = k^*$ , where  $k^*$  is predicted by Eq (6). Note that  $k^*$  values are around 3 to 6: enzyme of

Type-II can fill small unmodified islands having 3 to 6 zeros, and leave much longer unmodified islands unfilled.

In Fig 4C, the two monotonically decreasing curves belong to the parameter regime *a*, that is, ( $g_1 > 0, \alpha > 2\beta$ ). The mean error is decreasing as we increase  $k_t$ , suggesting that there is no finite  $k^*$ . If the modification patterns were in this regime, the corresponding enzymes should attempt to fill every unmodified region, however small or big that may be. The other two curves in Fig 4C correspond to regimes *c* and *d* in Fig 4A. For both the curves, the mean error is increasing, suggesting that the threshold- $k$  filling algorithm is not suitable in these parameter regimes.

### Threshold- $k$ filling model can obtain inheritance patterns similar to what is observed experimentally

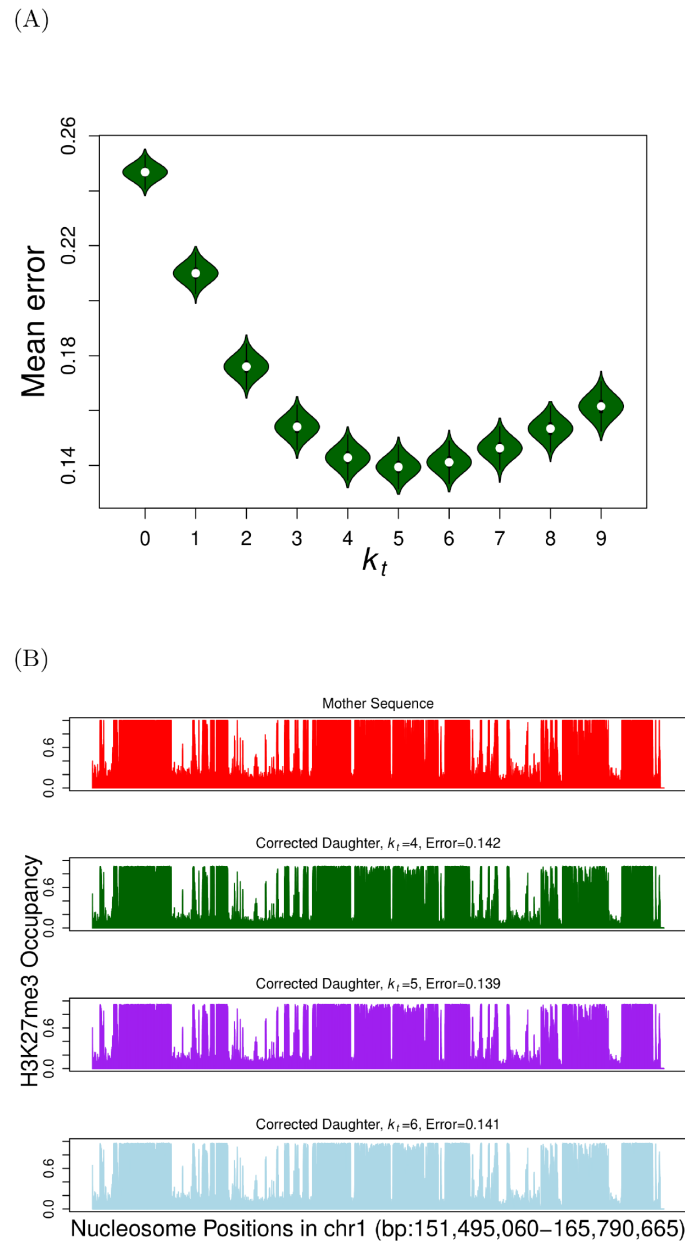
To examine the biological relevance of the findings leading to a threshold- $k$  filling model, we took publicly available experimental histone modification data, and compared with our simulation results. The inheritance of modification H3K27me3 has been systematically studied recently [28] by measuring modification occupancy before and after DNA replication. Since the experimental data is population-averaged, we used a simple randomized discretization algorithm to generate many binary sequences of the available data (see S3 Text and S3 Fig).

This discretized binary sequence is the mother sequence  $m_i$ . From the mother sequence, a daughter sequence ( $\mathbf{D} = \{d_i\}$ ) is created by converting each nucleosome to an unmodified state with a probability 0.5. This is equivalent to doing the mathematical operation  $d_i = m_i \cdot z_i$  as mentioned earlier. The daughter sequence  $\mathbf{D}$  was corrected to a mother-like modification sequence  $\hat{\mathbf{M}}$  using the threshold- $k$  algorithm for different  $k = k_t$ . This was repeated several times (100  $\mathbf{M}$  sequences, and 100  $\mathbf{D}$  for each  $\mathbf{M}$ ), and the mean error ( $\bar{\Delta}$ ) is plotted as a function of  $k_t$  in Fig 5A. The results show that when  $k_t = 5$  the mean error in the corrected H3K27me3 pattern is minimized. Note that this is very similar to the curves in Fig 4B for large values of  $\alpha$  and  $\beta$ . We independently verified that the parameters corresponding to the original mother sequence is  $\alpha \approx 0.81$  and  $\beta \approx 0.815$  (see S3 Text) implying that a biologically relevant modification falls in parameter regime *b*. In S4 Fig, we show the discretized mother sequences, daughter and corrected daughter sequences for a different modification (H3K4me3) whose original data was obtained from the same study [28].

In Fig 5B we plot the population-averaged modification pattern for different values  $k_t$ . Note that islands of high and low modification occupancy regions are present in both the mother as well as the corrected daughter sequences. This indicates that our threshold  $k$ -filling algorithm can reproduce biologically relevant data. Thus, our information theory-inspired algorithm predicts that there might be enzymes that simply fill short segments (4 or 5 nucleosomes) of unmodified regions, but leave the longer unmodified regions ( $> 5$  nucleosomes) unfilled. This helps in maintaining the fidelity during epigenetic inheritance. The two interesting biological questions in this context, namely (i) what are the plausible biological processes/mechanisms that could facilitate such a threshold filling process and (ii) how enzymes know what is the optimal threshold for filling, are discussed further in the Discussion section.

### Spatially distinct antagonistic modifications

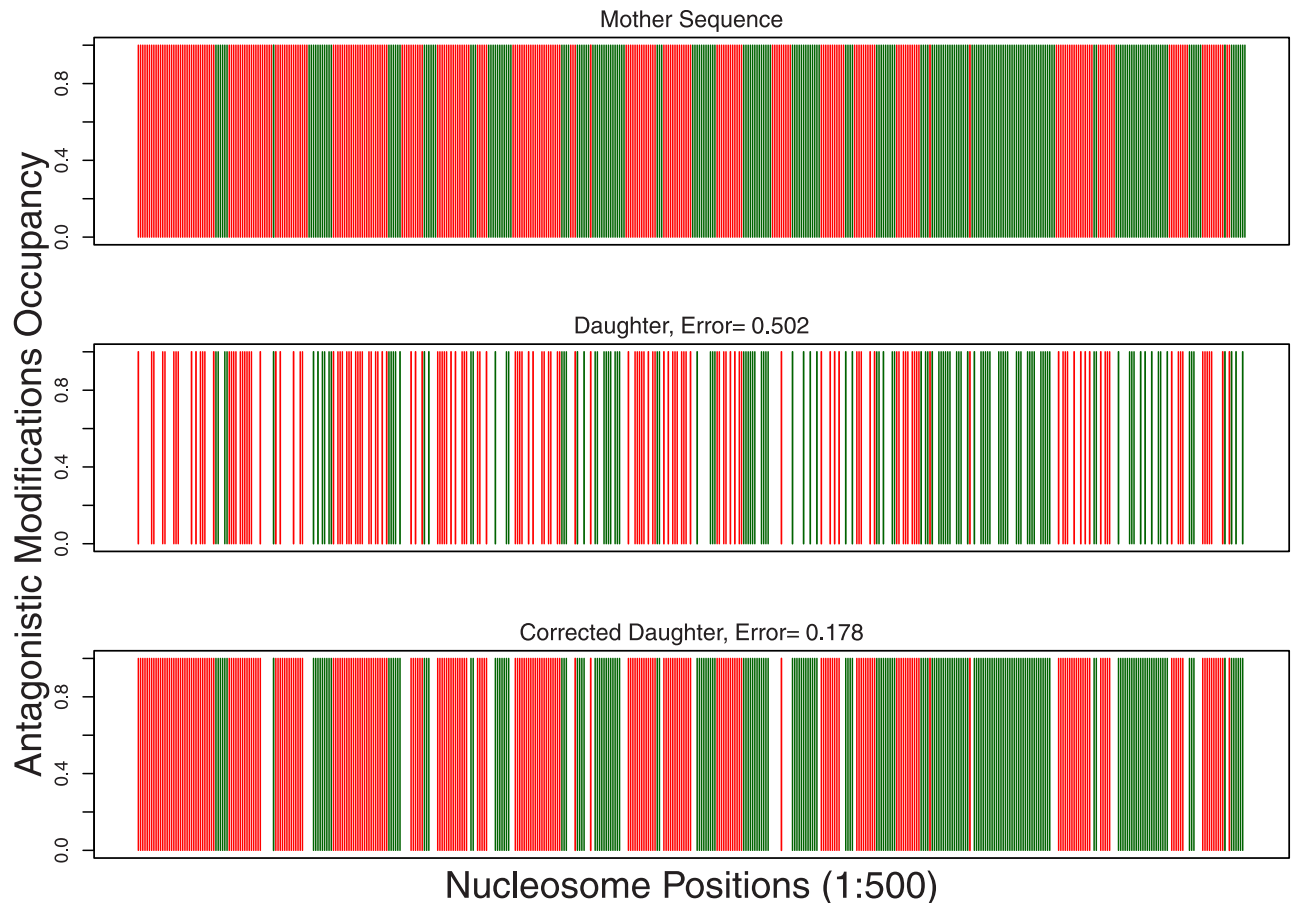
The threshold- $k$  filling model can be naturally extended to study two (or multiple) modifications that are antagonistic, spatially distinct (the same nucleosome will not have both the modifications simultaneously), and to be acted upon by very different enzymes. Specifically, we consider a case of a long stretch of repressive modification followed by a long stretch of activating modification in distinct neighboring regions of the chromatin. In the context of the



**Fig 5.** Error Correction in H3K27me3 data (A) Mean deviation ( $\bar{\Delta}$ ) between corrected daughters and corresponding mothers, where the experimental population-averaged parental data for H3K27me3 is from [28] (see database GEO: GSE110354). Error correction was performed using the threshold- $k$  filling algorithm for different  $k_t$  values. (B) The population averaged histone modification occupancy for H3K27me3 is plotted for mother sequence (top), and corrected daughter sequences corresponding to different values of  $k_t$ .

<https://doi.org/10.1371/journal.pcbi.1009861.g005>

bivalent/bistable state of the chromatin discussed in [41], it has to be noted that the antagonistic model we consider is largely not bivalent. As per our model, for an enzyme-1 responsible for modification-1, the nucleosomes having the second modification are not “visible” and would appear as a long stretch of 0s. For enzyme-2, similarly the modification-1 nucleosomes appear as a long stretch of 0s.



**Fig 6. Simulating antagonistic modifications: Typical realizations of the modification patterns from the study of two spatially distinct/antagonistic modifications, spread over 500 nucleosomes.** The red and green regions represent the modifications 1 and 2, respectively, and the white regions represent absence of both modifications (value 0). Corrections were performed using the threshold- $k$  algorithm with an optimum  $k_t = 6$ .

<https://doi.org/10.1371/journal.pcbi.1009861.g006>

We generated a sequence having two spatially distinct modifications. Since very high values of  $\alpha$  and  $\beta$  are known to produce long regions of 1s alternating with long regions of 0s, we used this approach to generate mother sequences with long stretches of two distinct modifications—we assign them values 1 and 2 indicating the presence of modification-1 or modification-2. These are represented by red and green colors in Fig 6. Taking this as the mother sequence, daughter sequences were then obtained by simulating the replication procedure of randomly flipping the non-zero values to zero using independent realizations of an unbiased coin. Note that both 1 and 2 here are flipped to 0 indicating the fact that freshly assembled nucleosomes will have neither of these modifications. One realization of the resulting daughter sequence is shown in Fig 6 (middle panel). Then the following version of the threshold- $k$  algorithm was applied to each enzyme separately. Whenever there is an island of size  $k \leq k_t$  between two nucleosomes having modification-1, the region was filled with modification-1, similarly for modification-2. Anything else was left unfilled. This gave us results as shown in Fig 6. In Fig 6, it may be observed that while a typical daughter sequence has an error  $\sim 0.5$ , the corrected daughter had an error less than 0.2. These are occupancies from a typical realization (not averaged over the population) and hence the values are either 0 or 1.

It has to be noted that this is only a preliminary extension of our single modification inheritance model to a pair of antagonistic modifications. In reality, the process of filling, in case of

multiple modifications may involve intermediate steps by several enzymes. This could be explored potentially as part of a future work.

## Discussion

In this work, we proposed that the problem of the daughter chromatin retrieving histone modification patterns, to achieve a mother-like chromatin state, can be mapped to a communication theory problem of receiving noisy signal and correcting it to retrieve the original signal. Using ideas from information theory, we argued that if enzymes were ideal computing machines, they could mimic a MAP decoding algorithm to get back a mother-like sequence. We showed how well this algorithm would reconstruct the mother—the error can be as low as 5% in certain parameter regimes. However, the question was whether realistic enzymes can practically do SMAP decoding. We showed that in a biologically relevant parameter regime, MAP decoding algorithm is equivalent to a threshold- $k$  filling algorithm. That is, the enzymes could simply insert modifications in  $k$ -sized or shorter unmodified stretches (0s). The fact that a detailed theory simplifies to a process that is potentially executable by enzymes makes this result attractive. Below, we discuss some aspects of our model, predictions and future directions, in the context of the existing literature.

**Our results in the context of existing models:** Currently, there are different categories of models that study various aspects of inheritance. First category belongs to computational models that study regions that are bistable. Work from Sneppen, Dodd, Ringrose, Howard and others have been discussing aspects of bistability [40, 41, 44]. This approach, in particular, is useful to explain many regions on stem cell chromatin, bivalent regions, specific regions like mating type locus in yeast and so on. In this model, the region of interest switches between two states—e.g., active and inactive—as a function of time; this would also imply that, at any given time, there are two populations of cells in a large ensemble. In the work of Sneppen and co-workers, “state” is quantified by finding the mean modification (over all modified/unmodified nucleosomes) in a given region and is independent of the parental chromatin state as long as there is a sufficient number of modified nucleosomes. They do not study the precise spatial pattern in detail. In contrast, there are regions that are stably maintained in a certain modified state and have specific pattern of histone modifications. This leads to a second category of models: e.g., models by Hodges and Crabtree, and others [45, 46] that study the spreading and maintaining of histone modifications as stable modified patterns over a finite length region in the steady state. This is achieved by introducing a nucleation site (which could be non-epigenetic like a DNA sequence-dependent location) that acts like a source for modifications/modifying enzymes. While these models explain maintenance of modifications, they require a nucleation site and have a decaying pattern from the nucleation site.

The above two categories of models start with certain physics-based kinetic moves for enzymes and ask what the steady state is. We take a different approach—an information/communication theory based approach—and ask the following reverse question: What must enzymes do to achieve a high-fidelity inheritance? We find that threshold- $k$  filling is the answer that information/communication theory provides. While we do not have kinetics in our model, we show (in S5(A) Fig) that if the parental population of cells have two stable states (bistable), the inherited population will also have two stable states. This can be considered as the ensemble interpretation of bistable states [40, 41, 44]. We also show (S5(B) Fig) that if the parental population has a stable modified state, it will also be stably inherited like in the case discussed in [45]. See S3 Text.

**Physical/Biological Relevance of the Model:** We modeled the mother chromatin as a first order Markov process. This is a reasonable model as there are just two parameters,  $\alpha$  and  $\beta$ ,

which are explicitly related to the experimentally measurable properties of the modification patterns such as the mean contiguous length of modified and unmodified regions (see [S2 Text](#)). By varying the threshold  $k_t$ , as shown in [Figs 4B](#) and [5A](#), we can determine the optimal configurations and obtain insights about how enzymes might work without a-priori knowledge of the statistical parameters. Note that even in the large  $\alpha$  regime (region *a* in [Fig 4A](#)), if enzymes settle for threshold filling with a finite  $k_t$  (e.g., 5 or 6), it becomes a pragmatic modification correction solution as the resulting error is relatively low (see red and green curves in [Fig 4B](#)).

It has to be noted that our model picks configurations (sequences) that maximize certain conditional probabilities, analogous to the minimum energy (or “equilibrium”) solution in statistical physics [[60](#), [61](#)]. The model does not involve precise kinetic moves. There could be multiple different kinetic events leading to the same static configuration. Hence, the predictions of our model may be interpreted in terms of various known kinetic events in the field. For example, the determination of critical threshold  $k_t$  as well as the decision to fill the gaps would involve local read-write mechanisms as well as feedback loops. The presence of a modification (1) inducing more modifications nearby, is an example of positive feedback. The decision to leave long gaps unfilled also may involve certain feedback mechanisms.

While enzymes fill the unmodified nucleosomes based on modified inherited nucleosomes, there is a natural question: how far away can the modification spread, from a modified nucleosome. Our model gives an answer to this question: it can spread upto  $k$  sites from a modified nucleosome. Even starting with nearest neighbour Markov model, there is a correlation length emerging that is  $k$  sites long. Our threshold- $k$  model provides a natural length-scale (of length  $k_t$ ) for inter-nucleosomal interactions. Moreover, our model has the advantage that it reproduces a spatial pattern similar to that of the parent.

**How plausible is threshold- $k$  filling in chromatin?:** What might be the biological processes that keep a group of nucleosomes in a modified state or unmodified state is an interesting question. A region can be kept unfilled as a combined effect of modification and de-modification (de-methylation, de-acetylation etc.) events. When the rate of de-modification in a region is much higher than the modification rate, the region can remain unfilled. Physically, this could happen in several ways. Enzymes could cooperatively act on a small region. For example, interesting recent experimental studies [[65](#)] suggest that enzymes that act on histone modifications can phase separate to form a droplet around a group of nucleosomes, and may collectively modify/demodify all the nucleosomes together. It has also been shown that chromatin itself can form nano-domains having size of a few nucleosomes paving way for collectively maintaining a certain modification state in these groups of nucleosomes [[66–68](#)]. The potential enzymes/complexes that could do these activities could be JmjC domain proteins, UTX, NuRD, Fbxl10 and JARID1A [[41](#), [69](#), [70](#)]. In case of H3K27me3 in particular, experimental evidence supports that EZH2 and H2AK119Ub also contribute to the modification of the neighboring nucleosomes [[26](#), [71](#)]. Approaching this question from a different angle, currently, it is well-known that there exist certain correlations and anti-correlations between modifications [[14](#)]. There are biological mechanisms which are known to maintain these correlated states between modifications [[38](#)]. Hence, the mechanism that may cause the absence of a modification could also cause the presence of an opposite modification as we have discussed in the antagonistic modifications section. For example, recent experiments from [[47](#), [72](#)] suggest that there is mutual antagonism between H3K27 and H3K36 methylation. This antagonism could also be one of the reasons for a long set of nucleosomes in which one modification is absent.

How enzymes know what is the optimal threshold for filling, is another interesting question. In recent experiments, it was shown that a critical density of H3K9me3 was required for

stable transgenerational inheritance of heterochromatin [31]. It has to be noted that our proposed model in which a critical threshold of parental modifications is required for filling the gaps of unmodified nucleosomes, is strengthened with this result. There are different families of enzymes, and each may have evolved differently. Some enzymes may be naturally adept at filling short regions, i.e. optimal threshold value ( $k^*$ ) could be hard-wired into such enzymes, via evolution. Another possibility is that other phenomena like local looping, phase separation etc. decide the threshold  $k^*$ , by bringing unfilled nucleosomes together [73, 74]. These need to be understood in more detail in the future. As we showed, the information theory ideas suggest that the threshold- $k$  filling algorithm emerges as a natural primary solution. However, further reconstruction may involve secondary mechanisms like boundary determination (e.g. CTCF [75]), synergistic modification between correlated modifications, and DNA methylation maintenance over the course of the cell-cycle [25, 38].

**Inheritance over many generations:** Another interesting aspect is the fidelity of inheritance over many generations. Error propagation across generations will depend on how well reconstruction happens in the very first generation. In the Figs 3 and 4, we have shown that the error (if we compute it as bit error rate—BER) is a non-zero number. Any non-zero error will propagate over generations. None of the published models so far, to the best of our knowledge, compare the mother and daughter modifications bit by bit and compute the error, they rather compute the mean modifications (modifications averaged over a long stretch) or the population-averaged pattern. In this work, we consider both the mean modification/mean pattern and bit by bit error (see S3 Text).

As pointed out above, we argue that threshold- $k$  filling algorithm emerges as a natural primary solution (immediately post-replication) based on information theoretic principles. However, in addition to this, the reconstruction may involve secondary mechanisms like boundary determination via binding of proteins like CTCF, formation of nucleosome free regions [20, 25, 76–79] and so on. In reality, all these mechanisms, together, will contribute to reconstruction and propagation of epigenetic information with minimum error.

Even in the absence of secondary mechanisms, we show that our model approximately preserves the pattern of modifications across a few generations (S6 Fig). We have simulated the case of repeated replication over 3 generations for a biologically realistic regime of  $\alpha$  and  $\beta = 0.9$  each. When the deviation (BER) is computed as against the original set of mother sequences with the above statistical parameters, we observe that deviation is reasonably low for up to 5 generations for  $k_t = 6$  (see S3 Text and S6 Fig). We also computed the block error—error averaged over blocks of neighboring nucleosomes—and plotted the results in S6(A) Fig. From the figure, one can observe that the error decreases with increasing block size. The BER and the mean block error remain very small across generations (see S3 Text). We also show in S6(B) Fig that the mean modification pattern is maintained across generations, using our algorithm on the H3K27me3 data. The details are provided in S3 Text.

Finally, it would also be of great interest to study how the polymer nature of the chromatin would work in tune with the results from information theory. It has to be noted that the epigenetic code might involve an interplay between the one-dimensional histone codes and polymer dynamics of the chromatin [80, 81]. The fact that we have two 1s at the boundaries suggests some potential role of looping or micro phase separation in far-away regions coming together. Electrostatic interactions among nucleosomes could also play an important role here [82–87]. For example, the charge distribution of a bunch of de novo nucleosomes can lead to a loose clustering of  $k$  nucleosomes, thereby facilitating the threshold- $k$  filling. Our own earlier studies [68] hint to us that small patches of unmodified nucleosomes (newly inserted nucleosomes) may lead to small clusters, influencing the kinetics of the modification process itself. Additionally it might be interesting to analyse how the inheritance is affected in cases where there may



be unequal distribution of the parental nucleosomes between the leading and lagging strands [21]. These are questions that await future studies.

## Supporting information

**S1 Fig. Analysis of Mean Error after correction.** In (A) and (C), we plot the mean error after correction ( $\bar{\Delta}$ ) using the SMAP algorithm, as a function of  $\alpha$  for different values of  $\beta$ . In (B) and (D), we plot the mean error after correction as a function of  $\beta$  for different values of  $\alpha$ . The error bars representing SEM are smaller than the size of the points in the plots.  $\alpha$  represents the conditional probability of a nucleosome staying in 1 (modified state) :  $P(m_i = 1 | m_{i-1} = 1)$ .  $\beta$  represents the conditional probability of a nucleosome staying in 0 (unmodified state) :  $P(m_i = 0 | m_{i-1} = 0)$ . The error is averaged over a simulation of 300 mother sequences each of which has 200 daughter sequences (a total of 60000 cases). (E) represents the variation of error after correction for different lengths of the sequence, for  $\alpha$  and  $\beta$  values of 0.9 each. (F) and (G) represent normalized mean error after correction obtained varying against  $k_t$ —these plots are similar to the Fig 4B and 4C in the manuscript but are normalized with respect to the optimum error.

(EPS)

**S2 Fig. Simulated and Binarized sequences.** Binarized sequences (top panel) for H3K27me3 and H3K4me3 from HeLa cells (discretized using the algorithm mentioned in S3 Text) and simulated sequences (bottom panel) corresponding to the same  $\alpha$  and  $\beta$ . The mean length of the modified and unmodified regions (runs) are distributed in a statistically similar manner in both the cases, suggesting that Markov model is a reasonable starting point.

(EPS)

**S3 Fig. Discretization Algorithm.** The algorithm used to discretize the population-averaged parental modification data ( $X$ ) from [28] into a binary realization—equivalent of a single cell representation indicating the presence (1) or absence (0) of the modification. In our case,  $X$  is H3K27me3 or H3K4me3 data obtained from [28].

(EPS)

**S4 Fig. H3K4me3 Experimental Verification.** Mother, Daughter, and the Corrected sequences for  $k_t = 4, 5, 6$  for the H3K4me3 values in the HeLa cells from a region of the Chr19 of the human genome. The data was obtained from [28] and is averaged over 200 binary realizations of the mother, each having 100 daughters.

(EPS)

**S5 Fig. Simulation in the context of existing models.** The histograms of the mother and corrected daughter sequences in the case where the mother sequences exhibit a bistable population of cells (i.e), roughly half of the cells have high modification levels whereas the remaining half of the cells have low modification levels. The replicated daughter cells after correction by our algorithm also show a similar bistable distribution. In (B), the plots of the mean mother and corrected daughter sequences for the case of a mother having a pattern similar to the one observed in [45] are shown. After correction by our algorithm, the daughter sequences obtain a pattern very similar to what is observed in the mother.

(EPS)

**S6 Fig. Error across generations.** (A)—The plot of the mean block error between the original mother sequence and the corrected daughter sequences across subsequent generations for  $\alpha$  and  $\beta$  values of 0.9 each. With the optimum value of  $k_t$  for this range of  $\alpha$  and  $\beta$  ( $k_t = 6$ ), the mean error increases minimally over subsequent replications. With increasing block size, the

mean error decreases monotonically. When the block size is 1, the error is the Bit Error Rate (BER). In (B), the plots of the mean mother and corrected daughter sequences for different generations are shown. The data is obtained from [28] and is the population-averaged binarized sequences of H3K27me3 ChIP-seq data in the Chr1 region.  
(EPS)

**S1 Text. SMAP Algorithm, Markov Process, Trellis Diagram.**  
(PDF)

**S2 Text. Computation of Statistical Properties.**  
(PDF)

**S3 Text. Discretization Algorithm, Experimental Data Simulation, Error across Generations.**  
(PDF)

## Author Contributions

**Conceptualization:** Sibi Raj B. Pillai, Ranjith Padinhateeri.

**Data curation:** Nithya Ramakrishnan.

**Funding acquisition:** Ranjith Padinhateeri.

**Investigation:** Nithya Ramakrishnan, Sibi Raj B. Pillai, Ranjith Padinhateeri.

**Methodology:** Nithya Ramakrishnan, Sibi Raj B. Pillai, Ranjith Padinhateeri.

**Software:** Nithya Ramakrishnan.

**Supervision:** Sibi Raj B. Pillai, Ranjith Padinhateeri.

**Validation:** Nithya Ramakrishnan.

**Visualization:** Nithya Ramakrishnan.

**Writing – original draft:** Nithya Ramakrishnan, Sibi Raj B. Pillai, Ranjith Padinhateeri.

**Writing – review & editing:** Nithya Ramakrishnan, Sibi Raj B. Pillai, Ranjith Padinhateeri.

## References

1. Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. *Cell*. 2007; 128(4):635–638. <https://doi.org/10.1016/j.cell.2007.02.006> PMID: 17320500
2. Allis CD, Jenuwein T, Reinberg D. Epigenetics. vol. 61. CSHL Press; 2007.
3. Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. *Science*. 1974; 184(4139):868–871. <https://doi.org/10.1126/science.184.4139.868> PMID: 4825889
4. Kornberg RD, Lorch Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*. 1999; 98(3):285–294. [https://doi.org/10.1016/S0092-8674\(00\)81958-3](https://doi.org/10.1016/S0092-8674(00)81958-3) PMID: 10458604
5. Van Holde KE. Chromatin. Springer Science & Business Media; 2012.
6. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 1997; 389(6648):251–260. <https://doi.org/10.1038/38444> PMID: 9305837
7. Kouzarides T. Chromatin modifications and their function. *Cell*. 2007; 128(4):693–705. <https://doi.org/10.1016/j.cell.2007.02.005> PMID: 17320507
8. Jenuwein T, Allis CD. Translating the histone code. *Science*. 2001; 293(5532):1074–1080. <https://doi.org/10.1126/science.1063127> PMID: 11498575

9. Richards EJ, Elgin SC. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell*. 2002; 108(4):489–500. [https://doi.org/10.1016/S0092-8674\(02\)00644-X](https://doi.org/10.1016/S0092-8674(02)00644-X) PMID: 11909520
10. Rando OJ. Global patterns of histone modifications. *Current Opinion in Genetics & Development*. 2007; 17(2):94–99. <https://doi.org/10.1016/j.gde.2007.02.006> PMID: 17317148
11. Noma Ki, Allis CD, Grewal SI. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science*. 2001; 293(5532):1150–1155. <https://doi.org/10.1126/science.1064150> PMID: 11498594
12. Alberts B. *Molecular Biology of The Cell*. 6th ed. New York: Garland Science, Taylor and Francis Group; 2014.
13. Taverna SD, Ueberheide BM, Liu Y, Tackett AJ, Diaz RL, Shabanowitz J, et al. Long-distance combinatorial linkage between methylation and acetylation on histone H3 N termini. *Proceedings of the National Academy of Sciences*. 2007; 104(7):2086–2091. <https://doi.org/10.1073/pnas.0610993104> PMID: 17284592
14. Weiner A, Hsieh THS, Appleboim A, Chen HV, Rahat A, Amit I, et al. High-resolution chromatin dynamics during a yeast stress response. *Molecular Cell*. 2015; 58(2):371–386. <https://doi.org/10.1016/j.molcel.2015.02.002> PMID: 25801168
15. Zentner GE, Henikoff S. Regulation of nucleosome dynamics by histone modifications. *Nature Structural & Molecular Biology*. 2013; 20(3):259. <https://doi.org/10.1038/nsmb.2470> PMID: 23463310
16. Seligson DB, Horvath S, Shi T, Yu H, Tze S, Grunstein M, et al. Global histone modification patterns predict risk of prostate cancer recurrence. *Nature*. 2005; 435(7046):1262–1266. <https://doi.org/10.1038/nature03672> PMID: 15988529
17. Margueron R, Reinberg D. Chromatin structure and the inheritance of epigenetic information. *Nature Reviews Genetics*. 2010; 11(4):285–296. <https://doi.org/10.1038/nrg2752> PMID: 20300089
18. Madamba EV, Berthet EB, Francis NJ. Inheritance of histones H3 and H4 during DNA replication in vitro. *Cell reports*. 2017; 21(5):1361–1374. <https://doi.org/10.1016/j.celrep.2017.10.033> PMID: 29091772
19. Xu M, Long C, Chen X, Huang C, Chen S, Zhu B. Partitioning of histone H3-H4 tetramers during DNA replication-dependent chromatin assembly. *Science*. 2010; 328(5974):94–98. <https://doi.org/10.1126/science.1178994> PMID: 20360108
20. Groth A, Rocha W, Verreault A, Almouzni G. Chromatin challenges during DNA replication and repair. *Cell*. 2007; 128(4):721–733. <https://doi.org/10.1016/j.cell.2007.01.030> PMID: 17320509
21. Petryk N, Dalby M, Wenger A, Stromme CB, Strandsby A, Andersson R, et al. MCM2 promotes symmetric inheritance of modified histones during DNA replication. *Science*. 2018; 361(6409):1389–1392. <https://doi.org/10.1126/science.aau0294> PMID: 30115746
22. Yu C, Gan H, Serra-Cardona A, Zhang L, Gan S, Sharma S, et al. A mechanism for preventing asymmetric histone segregation onto replicating DNA strands. *Science*. 2018; 361(6409):1386–1389. <https://doi.org/10.1126/science.aat8849> PMID: 30115745
23. Alabert C, Barth TK, Reverón-Gómez N, Sidoli S, Schmidt A, Jensen ON, et al. Two distinct modes for propagation of histone PTMs across the cell cycle. *Genes & development*. 2015; 29(6):585–590. <https://doi.org/10.1101/gad.256354.114> PMID: 25792596
24. Bameta T, Das D, Padinhateeri R. Coupling of replisome movement with nucleosome dynamics can contribute to the parent–daughter information transfer. *Nucleic Acids Research*. 2018; 46(10):4991–5000. <https://doi.org/10.1093/nar/gky207> PMID: 29850895
25. Probst AV, Dunleavy E, Almouzni G. Epigenetic inheritance during the cell cycle. *Nature Reviews Molecular Cell Biology*. 2009; 10(3):192–206. <https://doi.org/10.1038/nrm2640> PMID: 19234478
26. Reinberg D, Vales LD. Chromatin domains rich in inheritance. *Science*. 2018; 361(6397):33–34. <https://doi.org/10.1126/science.aat7871> PMID: 29976815
27. Hall IM, Shankaranarayana GD, Noma Ki, Ayoub N, Cohen A, Grewal SI. Establishment and maintenance of a heterochromatin domain. *Science*. 2002; 297(5590):2232–2237. <https://doi.org/10.1126/science.1076466> PMID: 12215653
28. Reverón-Gómez N, González-Aguilera C, Stewart-Morgan KR, Petryk N, Flury V, Graziano S, et al. Accurate recycling of parental histones reproduces the histone modification landscape during DNA replication. *Molecular Cell*. 2018; 72(2):239–249. <https://doi.org/10.1016/j.molcel.2018.08.010> PMID: 30146316
29. Escobar TM, Oksuz O, Saldaña-Meyer R, Descostes N, Bonasio R, Reinberg D. Active and repressed chromatin domains exhibit distinct nucleosome segregation during DNA replication. *Cell*. 2019; 179(4):953–963. <https://doi.org/10.1016/j.cell.2019.10.009> PMID: 31675501

30. Ragunathan K, Jih G, Moazed D. Epigenetic inheritance uncoupled from sequence-specific recruitment. *Science*. 2015; 348(6230):1258699. <https://doi.org/10.1126/science.1258699> PMID: 25831549
31. DiPiazza ARC, Taneja N, Dhakshnamoorthy J, Wheeler D, Holla S, Grewal SI. Spreading and epigenetic inheritance of heterochromatin require a critical density of histone H3 lysine 9 tri-methylation. *Proceedings of the National Academy of Sciences*. 2021; 118(22).
32. Grewal SI, Klar AJ. Chromosomal inheritance of epigenetic states in fission yeast during mitosis and meiosis. *Cell*. 1996; 86(1):95–101. [https://doi.org/10.1016/S0092-8674\(00\)80080-X](https://doi.org/10.1016/S0092-8674(00)80080-X) PMID: 8689692
33. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. *Nature Reviews Genetics*. 2016; 17(8):487–500. <https://doi.org/10.1038/nrg.2016.59> PMID: 27346641
34. Chi P, Allis CD, Wang GG. Covalent histone modifications—miswritten, misinterpreted and mis-erased in human cancers. *Nature Reviews Cancer*. 2010; 10(7):457–469. <https://doi.org/10.1038/nrc2876> PMID: 20574448
35. Hyun K, Jeon J, Park K, Kim J. Writing, erasing and reading histone lysine methylations. *Experimental & Molecular Medicine*. 2017; 49(4):e324–e324. <https://doi.org/10.1038/emm.2017.11> PMID: 28450737
36. Sabari BR, Zhang D, Allis CD, Zhao Y. Metabolic regulation of gene expression through histone acylations. *Nature reviews Molecular Cell Biology*. 2017; 18(2):90–101. <https://doi.org/10.1038/nrm.2016.140> PMID: 27924077
37. Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury WJ Iii, et al. Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*. 2009; 461(7265):762–767. <https://doi.org/10.1038/nature08398> PMID: 19767730
38. Guo Y, Zhao S, Wang GG. Polycomb Gene Silencing Mechanisms: PRC2 Chromatin Targeting, H3K27me3 'Readout', and Phase Separation-Based Compaction. *Trends in Genetics*. 2021;. <https://doi.org/10.1016/j.tig.2020.12.006> PMID: 33494958
39. Zhang H, Tian XJ, Mukhopadhyay A, Kim K, Xing J. Statistical mechanics model for the dynamics of collective epigenetic histone modification. *Physical Review Letters*. 2014; 112(6):068101. <https://doi.org/10.1103/PhysRevLett.112.068101> PMID: 24580708
40. Dodd IB, Micheelsen MA, Sneppen K, Thon G. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell*. 2007; 129(4):813–822. <https://doi.org/10.1016/j.cell.2007.02.053> PMID: 17512413
41. Sneppen K, Ringrose L. Theoretical analysis of Polycomb-Trithorax systems predicts that poised chromatin is bistable and not bivalent. *Nature communications*. 2019; 10(1):1–18. <https://doi.org/10.1038/s41467-019-10130-2>
42. Michieletto D, Chiang M, Coli D, Papantonis A, Orlandini E, Cook PR, et al. Shaping epigenetic memory via genomic bookmarking. *Nucleic acids research*. 2018; 46(1):83–93. <https://doi.org/10.1093/nar/gkx1200> PMID: 29190361
43. Angel A, Song J, Dean C, Howard M. A Polycomb-based switch underlying quantitative epigenetic memory. *Nature*. 2011; 476(7358):105–108. <https://doi.org/10.1038/nature10241> PMID: 21785438
44. Berry S, Dean C, Howard M. Slow chromatin dynamics allow polycomb target genes to filter fluctuations in transcription factor activity. *Cell systems*. 2017; 4(4):445–457. <https://doi.org/10.1016/j.cels.2017.02.013> PMID: 28342717
45. Hodges C, Crabtree GR. Dynamics of inherently bounded histone modification domains. *Proceedings of the National Academy of Sciences*. 2012; 109(33):13296–13301. <https://doi.org/10.1073/pnas.1211172109> PMID: 22847427
46. Sandholtz SH, Kannan D, Beltran BG, Spakowitz AJ. Chromosome Structural Mechanics Dictates the Local Spreading of Epigenetic Marks. *Biophysical Journal*. 2020; 119(8):1630–1639. <https://doi.org/10.1016/j.bpj.2020.08.039> PMID: 33010237
47. Alabert C, Loos C, Voelker-Albert M, Graziano S, Forné I, Reveron-Gomez N, et al. Domain Model Explains Propagation Dynamics and Stability of Histone H3K27 and H3K36 Methylation Landscapes. *Cell Reports*. 2020; 30(4):1223–1234. <https://doi.org/10.1016/j.celrep.2019.12.060> PMID: 31995760
48. Zhang T, Cooper S, Brockdorff N. The interplay of histone modifications—writers that read. *EMBO reports*. 2015; 16(11):1467–1481. <https://doi.org/10.15252/embr.201540945> PMID: 26474904
49. Zheng Y, Sweet SM, Popovic R, Martinez-Garcia E, Tipton JD, Thomas PM, et al. Total kinetic analysis reveals how combinatorial methylation patterns are established on lysines 27 and 36 of histone H3. *Proceedings of the National Academy of Sciences*. 2012; 109(34):13549–13554. <https://doi.org/10.1073/pnas.1205707109> PMID: 22869745
50. Fischle W, Wang Y, Allis CD. Binary switches and modification cassettes in histone biology and beyond. *Nature*. 2003; 425(6957):475–479. <https://doi.org/10.1038/nature02017> PMID: 14523437
51. Ramachandran S, Henikoff S. Replicating nucleosomes. *Science Advances*. 2015; 1(7):e1500587. <https://doi.org/10.1126/sciadv.1500587> PMID: 26269799

52. Nam KM, Gyori BM, Amethyst SV, Bates DJ, Gunawardena J. Robustness and parameter geography in post-translational modification systems. *PLOS Computational Biology*. 2020; 16(5):e1007573. <https://doi.org/10.1371/journal.pcbi.1007573> PMID: 32365103
53. Tkačik G, Bialek W. Information processing in living systems. *Annual Review of Condensed Matter Physics*. 2016; 7:89–117. <https://doi.org/10.1146/annurev-conmatphys-031214-014803>
54. Mora T, Bialek W. Are biological systems poised at criticality? *Journal of Statistical Physics*. 2011; 144(2):268–302.
55. Tkačik G, Walczak AM. Information transmission in genetic regulatory networks: a review. *Journal of Physics: Condensed Matter*. 2011; 23(15):153102. PMID: 21460423
56. Madhoo U. *Fundamentals of digital communication*. Cambridge University Press; 2008.
57. Cover TM, Thomas JA. *Elements of information theory*. John Wiley & Sons; 2012.
58. Teif VB, Rippe K. Statistical–mechanical lattice models for protein–DNA binding in chromatin. *Journal of Physics: Condensed Matter*. 2010; 22(41):414105. PMID: 21386588
59. Poepsel S, Kasinath V, Nogales E. Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. *Nature structural & molecular biology*. 2018; 25(2):154–162. <https://doi.org/10.1038/s41594-018-0023-y> PMID: 29379173
60. Chaikin PM, Lubensky TC, Witten TA. *Principles of condensed matter physics*. vol. 10. Cambridge university press Cambridge; 1995.
61. Bialek W. *Biophysics: searching for principles*. Princeton University Press; 2012.
62. Sneppen K, Dodd IB. A simple histone code opens many paths to epigenetics. *PLoS Comput Biol*. 2012; 8(8):e1002643. <https://doi.org/10.1371/journal.pcbi.1002643> PMID: 22916004
63. Lin S, Costello D. *Error Control Coding*. Pearson Education; 2005.
64. Stewart WJ. *Probability, Markov chains, queues, and simulation: the mathematical basis of performance modeling*. Princeton university press; 2009.
65. Gallego LD, Schneider M, Mittal C, Romanauska A, Carrillo RMG, Schubert T, et al. Phase separation directs ubiquitination of gene-body nucleosomes. *Nature*. 2020; 579(7800):592–597. <https://doi.org/10.1038/s41586-020-2097-z> PMID: 32214243
66. Thorn GJ, Clarkson CT, Rademacher A, Mamayusupova H, Schotta G, Rippe K, et al. DNA sequence-dependent formation of heterochromatin nanodomains. *bioRxiv*. 2020;. <https://doi.org/10.1101/2020.10.07.307546> PMID: 33052340
67. Szabo Q, Donjon A, Jerković I, Papadopoulos GL, Cheutin T, Bonev B, et al. Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nature Genetics*. 2020; 52(11):1151–1157. <https://doi.org/10.1038/s41588-020-00716-8> PMID: 33077913
68. Bajpai G, Padinhateeri R. Irregular chromatin: packing density, fiber width, and occurrence of heterogeneous clusters. *Biophysical Journal*. 2020; 118(1):207–218. <https://doi.org/10.1016/j.bpj.2019.11.004> PMID: 31810656
69. Swygert SG, Peterson CL. Chromatin dynamics: interplay between remodeling enzymes and histone modifications. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*. 2014; 1839(8):728–736. <https://doi.org/10.1016/j.bbagr.2014.02.013> PMID: 24583555
70. Shi Y. Histone lysine demethylases: emerging roles in development, physiology and disease. *Nature reviews genetics*. 2007; 8(11):829–833. <https://doi.org/10.1038/nrg2218> PMID: 17909537
71. Blackledge NP, Farcas AM, Kondo T, King HW, McGouran JF, Hanssen LL, et al. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell*. 2014; 157(6):1445–1459. <https://doi.org/10.1016/j.cell.2014.05.004> PMID: 24856970
72. Streubel G, Watson A, Jammula SG, Scelfo A, Fitzpatrick DJ, Oliviero G, et al. The H3K36me2 methyltransferase Nsd1 demarcates PRC2-mediated H3K27me2 and H3K27me3 domains in embryonic stem cells. *Molecular cell*. 2018; 70(2):371–379. <https://doi.org/10.1016/j.molcel.2018.02.027> PMID: 29606589
73. Mir M, Bickmore W, Furlong EE, Narlikar G. Chromatin topology, condensates and gene regulation: shifting paradigms or just a phase? *Development*. 2019; 146(19):dev182766. <https://doi.org/10.1242/dev.182766> PMID: 31554625
74. Rowley MJ, Corces VG. Organizational principles of 3D genome architecture. *Nature Reviews Genetics*. 2018; 19(12):789–800. <https://doi.org/10.1038/s41576-018-0060-8> PMID: 30367165
75. Narendra V, Bulajić M, Dekker J, Mazzoni EO, Reinberg D. CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes & development*. 2016; 30(24):2657–2662. <https://doi.org/10.1101/gad.288324.116> PMID: 28087711
76. Ranjan A, Mizuguchi G, FitzGerald PC, Wei D, Wang F, Huang Y, et al. Nucleosome-free region dominates histone acetylation in targeting SWR1 to promoters for H2A. Z replacement. *Cell*. 2013; 154(6):1232–1245. <https://doi.org/10.1016/j.cell.2013.08.005> PMID: 24034247

77. Bonasio R, Tu S, Reinberg D. Molecular signals of epigenetic states. *Science*. 2010; 330(6004):612–616. <https://doi.org/10.1126/science.1191078> PMID: 21030644
78. Allshire RC, Madhani HD. Ten principles of heterochromatin formation and function. *Nature Reviews Molecular Cell Biology*. 2018; 19(4):229. <https://doi.org/10.1038/nrm.2017.119> PMID: 29235574
79. Torres-Garcia S, Yaseen I, Shukla M, Audergon PN, White SA, Pidoux AL, et al. Epigenetic gene silencing by heterochromatin primes fungal resistance. *Nature*. 2020; 585(7825):453–458. <https://doi.org/10.1038/s41586-020-2706-x> PMID: 32908306
80. Michieletto D, Orlandini E, Marenduzzo D. Polymer model with epigenetic recoloring reveals a pathway for the de novo establishment and 3D organization of chromatin domains. *Physical Review X*. 2016; 6(4):041047. <https://doi.org/10.1103/PhysRevX.6.041047>
81. Jost D, Vaillant C. Epigenomics in 3D: importance of long-range spreading and specific interactions in epigenomic maintenance. *Nucleic acids research*. 2018; 46(5):2252–2264. <https://doi.org/10.1093/nar/gky009> PMID: 29365171
82. Woodcock CL, Skoultchi AI, Fan Y. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Research*. 2006; 14(1):17–25. <https://doi.org/10.1007/s10577-005-1024-3> PMID: 16506093
83. Fan Y, Nikitina T, Zhao J, Fleury TJ, Bhattacharyya R, Bouhassira EE, et al. Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell*. 2005; 123(7):1199–1212. <https://doi.org/10.1016/j.cell.2005.10.028> PMID: 16377562
84. Routh A, Sandin S, Rhodes D. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proceedings of the National Academy of Sciences*. 2008; 105(26):8872–8877. <https://doi.org/10.1073/pnas.0802336105> PMID: 18583476
85. Cherstvy AG, Teif VB. Electrostatic effect of H1-histone protein binding on nucleosome repeat length. *Physical biology*. 2014; 11(4):044001. <https://doi.org/10.1088/1478-3975/11/4/044001> PMID: 25078656
86. Harshman SW, Young NL, Parthun MR, Freitas MA. H1 histones: current perspectives and challenges. *Nucleic acids research*. 2013; 41(21):9593–9609. <https://doi.org/10.1093/nar/gkt700> PMID: 23945933
87. Luzhetskaya OP, Sedykh SE, Nevinsky GA. How Human H1 Histone Recognizes DNA. *Molecules*. 2020; 25(19):4556. <https://doi.org/10.3390/molecules25194556> PMID: 33028027