

Supplementary Information:

Antagonistic histone post-translational modifications improve the fidelity of epigenetic inheritance - a Bayesian perspective

B. N. BALAKRISHNA PRABHU, ADITYA NAMAN SONI, SIBI RAJ B. PILLAI, and NITHYA RAMAKRISHNAN

1 Model and Methods

In our study, we analyze two binary sequences: $M^A = (m_1^A, m_2^A, \dots, m_n^A)$ and $M^B = (m_1^B, m_2^B, \dots, m_n^B)$. Sequence M^A represents a specific histone Post Translational Modification (PTM), while sequence M^B denotes an antagonistic PTM to M^A (Fig. 1A). Our goal is to estimate the transition probabilities α and β within sequence M^A , and the conditional probability μ between sequences M^A and M^B . Referring to Fig. 1B of the main manuscript, α is the probability that a site remains modified (represented as 1) given that it was modified on the preceding nucleosome. Mathematically, this can be expressed as:

$$\alpha = Pr(m_i^A = 1 | m_{i-1}^A = 1). \quad (S1)$$

Similarly, β represents the conditional probability that a site remains unmodified (represented as 0) *i.e.*

$$\beta = Pr(m_i^A = 0 | m_{i-1}^A = 0). \quad (S2)$$

In Fig. 1C we take the parameter μ as

$$\mu = Pr(m_i^B = 1 | m_i^A = 0), \quad (S3)$$

identically for all $1 \leq i \leq n$. Thus, μ quantifies the statistical relationship between sequences M^A and M^B , specifically the likelihood of observing an antagonistic modification (in M^B) given the absence of modification (in M^A).

For the case of **non-exclusive** antagonistic modifications A and C , we define γ as follows:

$$\gamma = Pr(m_i^B = 0 | m_i^A = 0) = \gamma = Pr(m_i^B = 1 | m_i^A = 1). \quad (S4)$$

1.1 Computation of α , β , μ and γ

Given a pair of binary sequences $M^A = \{m_1^A, m_2^A, \dots, m_n^A\}$, and $M^B = \{m_1^B, m_2^B, \dots, m_n^B\}$, we define the transition counts necessary for estimating α and β :

$$\begin{aligned} n_{11} &= \text{Number of transitions from } m_i^A = 1 \text{ to } m_{i+1}^A = 1, \\ n_{10} &= \text{Number of transitions from } m_i^A = 1 \text{ to } m_{i+1}^A = 0, \\ n_{01} &= \text{Number of transitions from } m_i^A = 0 \text{ to } m_{i+1}^A = 1, \\ n_{00} &= \text{Number of transitions from } m_i^A = 0 \text{ to } m_{i+1}^A = 0. \end{aligned}$$

The probabilities α and β for sequence M^A are then calculated as follows:

$$\alpha = \frac{n_{11}}{n_{11} + n_{10}}, \quad (S5)$$

$$\beta = \frac{n_{00}}{n_{00} + n_{01}}. \quad (S6)$$

Estimating μ and γ from M^A and M^B

The probability μ is given by:

$$\mu = \frac{\sum_{i=1}^n \mathbb{1}(m_i^A=0 \text{ AND } m_i^B=1)}{\sum_{i=1}^n \mathbb{1}(m_i^A=0)}, \quad (S7)$$

where n is the length of the sequences, and $\mathbb{1}(\cdot)$ is the indicator function, yielding 1 if the condition inside is true, and 0 otherwise. For the case of positively correlated pair of PTM's the probability γ is given by:

$$\gamma = \frac{\sum_{i=1}^n \mathbb{1}(m_i^A=m_i^B)}{\sum_{i=1}^n \mathbb{1}(m_i^A)}. \quad (S8)$$

1.2 Derivation of steady-state probabilities

We represent the joint sequence $M^{AB} := \{(m_i^A, m_i^B), 1 \leq i \leq n\}$ as a Markov model and derive the steady-state probabilities. From the state diagram (Fig. 1C), we consider the transition probability matrix representing the 3 states S , T and U as follows:

	S	T	U
S	α	$(1-\alpha)\mu$	$(1-\alpha)(1-\mu)$
T	$(1-\beta)$	$\beta\mu$	$\beta(1-\mu)$
U	$(1-\beta)$	$\beta\mu$	$\beta(1-\mu)$

where $S = (m_i^A = 1, m_i^B = 0)$, $T = (m_i^A = 0, m_i^B = 1)$ and $U = (m_i^A = 0, m_i^B = 0)$. The steady-state probability vector, denoted by $[\Pi_S, \Pi_T, \Pi_U]$, obeys

$$[\Pi_S \quad \Pi_T \quad \Pi_U] \cdot \begin{bmatrix} \alpha & (1-\alpha)\mu & (1-\alpha)(1-\mu) \\ (1-\beta) & \beta\mu & \beta(1-\mu) \\ (1-\beta) & \beta\mu & \beta(1-\mu) \end{bmatrix} = [\Pi_S \quad \Pi_T \quad \Pi_U]. \quad (S9)$$

Using $\Pi_S + \Pi_T + \Pi_U = 1$ and solving the balance equations, we get the steady-state probabilities as follows:

$$\Pi_S = \frac{1-\beta}{2-\alpha-\beta}, \quad (S10)$$

$$\Pi_T = \frac{(1-\beta)(1-\alpha-\beta)\mu}{2-\alpha-\beta} + \beta\mu, \quad (S11)$$

$$\Pi_U = \frac{(1-\beta)(1-\alpha-\beta)(1-\mu)}{2-\alpha-\beta} + \beta(1-\mu). \quad (S12)$$

Here, Π_S , Π_T and Π_U represent the stationary probabilities corresponding to the presence of PTMs A and B in a nucleosome.

1.3 Non-exclusive antagonistic modifications

In the mathematical model for the system, we considered two antagonistic pairs which are mutually exclusive, i.e., the possibility of both the antagonistic marks (A and B) being simultaneously present in a nucleosome was neglected. Let us now describe how this model and the results can be generalized to the non-exclusive antagonistic model with $\Pr(m_i^B = 1 | m_i^A = 1) = \nu > 0$. Assume that the sequence (M^A, M^B) evolves according to Equation (4) in the main manuscript. After replication, each daughter strand directly acquires the marks from half the parental nucleosomes. As before, our objective is to recover the primary PTM mark A at one of the daughter chromatids. The possible observations at each nucleosome is $(d_i^A, d_i^B) \in \{(0, 0), (1, 0), (0, 1), (1, 1)\}$. By calling the state $(1, 1)$ as $1'$, we can take $d^{AB} \in \{0, 1, 2, 1'\}$ as the output observation for each daughter nucleosome, see Section 2.2 of the main manuscript.

2 Simulation Results

Fig. S1 shows the plots showing the effect of varying parameters α , β , and μ on the simulated mother sequences. Using the SMAP decoding algorithm, we obtain the difference in BER between the original mother sequences and the reconstructed mother sequences for $\alpha, \beta \in (0, 1)$ and $\mu = 0$ (non-antagonistic case). In Fig. S1 E and F, data segments from the H9 cell-line data (GSM667622 and GSM605310) that fall into region B of Fig 4C of the main manuscript are plotted, to illustrate the prevalence of the parametric regimes. The heatmap in Fig. S2 (A) was obtained by averaging over 50000 simulated sequences for each (α, β) pair and reproduces the result from our earlier work[1]. The violin plot in fig. S2 (B) compares the no-antagonism case with our current model. Fig. S2 (C,D,E) represent the ΔBER between the no-antagonism and antagonism cases for $\mu = (0.2, 0.5, 0.8)$ and represent the color-divergent heatmaps in Fig 3(D-F) of the main manuscript.

Fig. S3 shows the heatmaps of the average BER for simulations comparing μ and γ for different values of (α, β) .

Fig. S4 shows the variation for threshold- $\hat{k}$ values for different (α, β, μ) combinations for Case 1(Sec. 2.2 of the main manuscript).

Fig. S5 shows the variations of the boundaries with μ for Cases 2, 3, and 4(Sec. 2.2 of the main manuscript).

Additionally, we define two specialized metrics to capture specific types of errors:

Bit Error Rate for 1-to-0 Transitions (BER_1): This metric measures the fraction of nucleosomes where a 1 in the original sequence M^A is incorrectly estimated as 0 in $\hat{M}^A$. It is given by:

$$BER_1 = \frac{\sum_{i=1}^n (m_i^A = 1 \wedge \hat{m}_i^A = 0)}{\sum_{i=1}^n (m_i^A = 1)}, \quad (S13)$$

Bit Error Rate for 0-to-1 Transitions (BER_0): This metric measures the fraction of nucleosomes where a 0 in the original sequence M^A is incorrectly estimated as 1 in $\hat{M}^A$. It is given by:

$$BER_0 = \frac{\sum_{i=1}^n (m_i^A = 0 \wedge \hat{m}_i^A = 1)}{\sum_{i=1}^n (m_i^A = 0)}. \quad (S14)$$

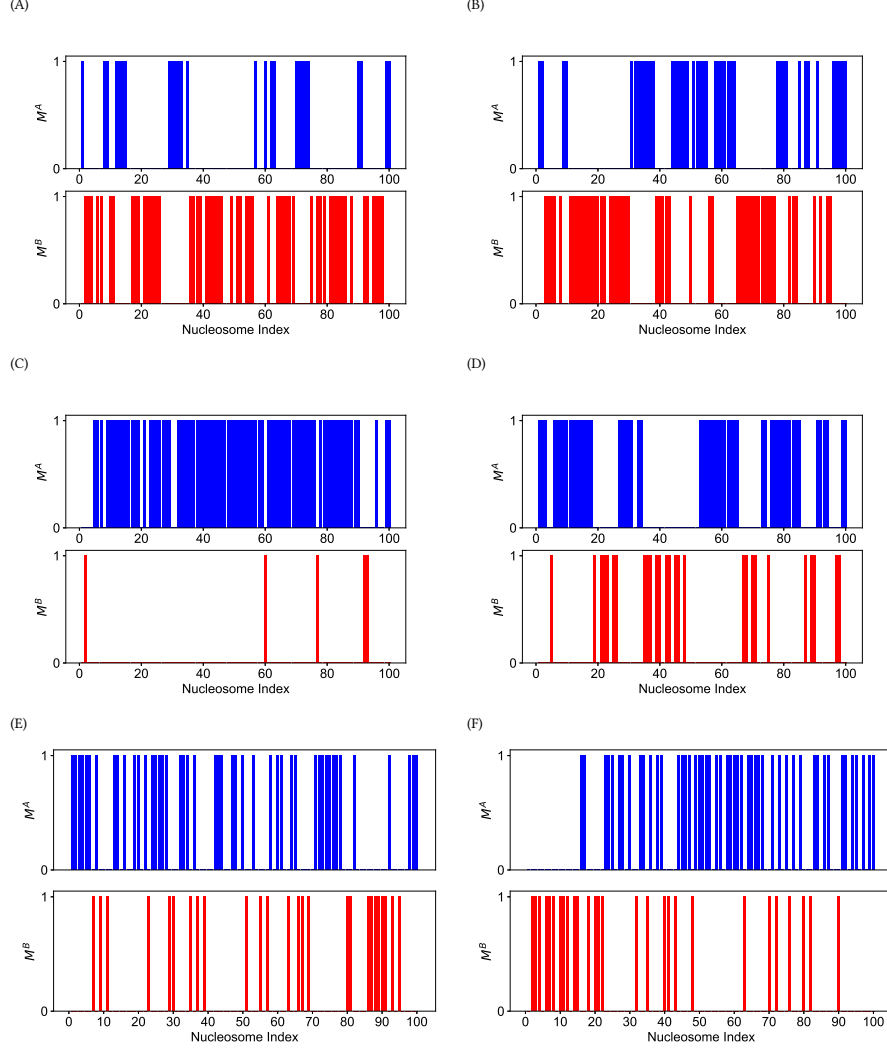


Fig. S1. Simulation plots and H9 cell-line plots: Plots showing the effect of varying parameters α , β , and μ on the simulated sequences. (A) $\alpha = 0.6$, $\beta = 0.8$, $\mu = 0.8$, (B) $\alpha = 0.8$, $\beta = 0.8$, $\mu = 0.8$, (C) $\alpha = 0.9$, $\beta = 0.4$, $\mu = 0.2$, and (D) $\alpha = 0.9$, $\beta = 0.9$, $\mu = 0.5$, (E and F) H9 cell-line data segments (E) chr 1 (105042162 onwards) $\alpha = 0.58$, $\beta = 0.64$, $\mu = 0.49$, (F) chr 1 (110042171 onwards) $\alpha = 0.55$, $\beta = 0.54$, $\mu = 0.56$.

3 Discretization Algorithm

The PTM data from the experiments obtained from [2] included processed bed (browser extensible data) files from the ChIP-Seq experiments. They are transformed into binary sequences that indicate the presence or absence of the PTMs in a nucleosome. The algorithm is detailed in the flowchart of Fig. S6. Binarised versions of PTMs H3K27me3 and H3K36me3 represent sequences M^A and M^B respectively in our work. Additional modifications—H3K27me1, H3K27me2, and H3K36me2 are also considered in detecting the presence of H3.

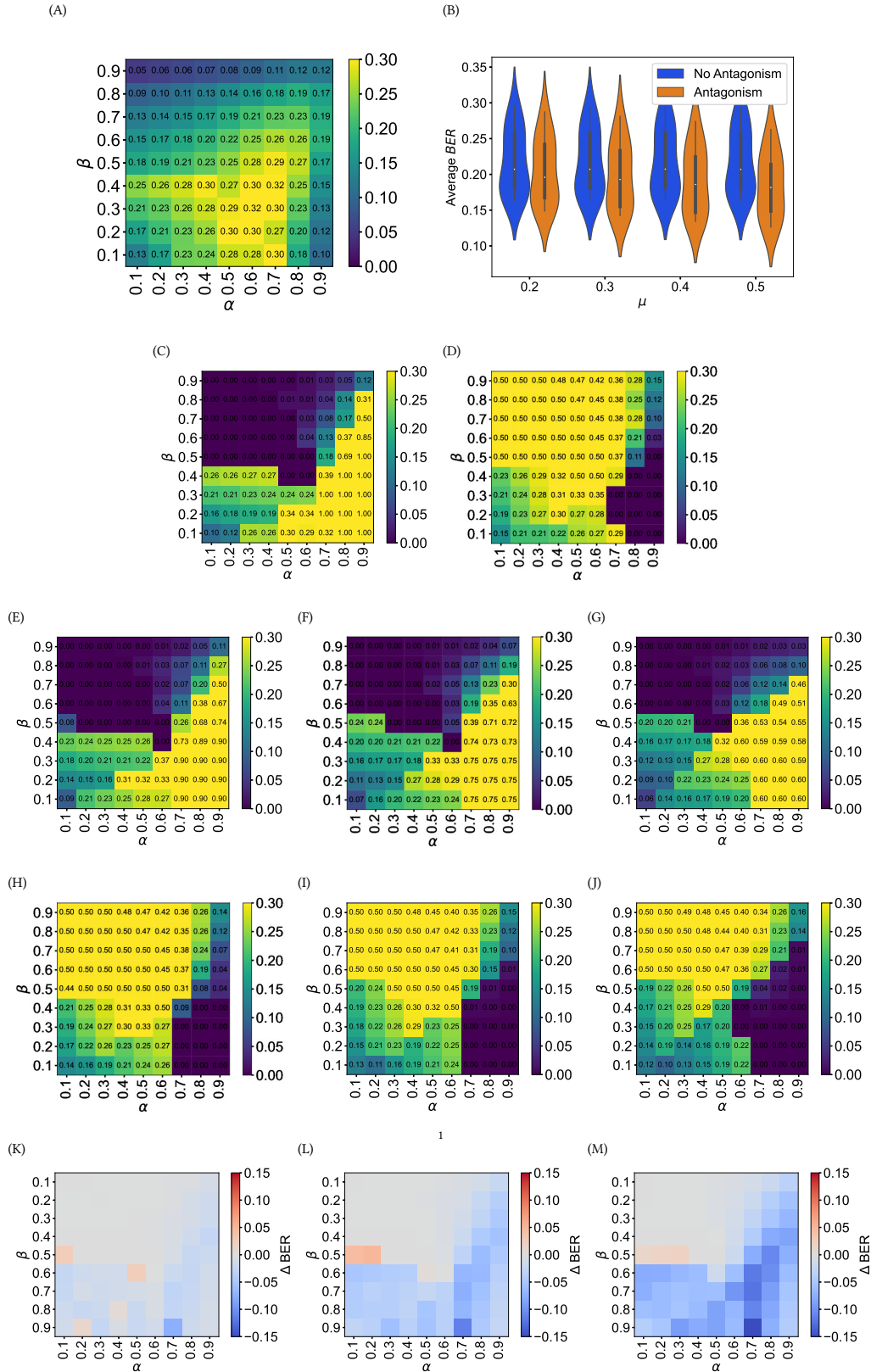


Fig. S2. BER in simulations: (A) Heatmaps for the BER between the original mother sequence (M_i^A) and the reconstructed mother sequence ($\hat{M}_i^A$) obtained using SMAP decoding for sequences without antagonism ($\mu = 0$). (B) Average BER of reconstructed non-antagonistic sequences and the antagonistic sequences obtained through simulations averaged over 1000 sequences of length 1000 each, with parameters $\alpha \in [0.7, 0.9]$, $\beta \in [0.5, 0.8]$ and $\mu \in [0.2, 0.6]$. Figures (C,E,F,G) are the BER_0 and (D, H, I, J) are the BER_i plots corresponding to Equations Eq. S14 and Eq. S13 plots for ($\mu = 0.0, 0.2, 0.5, 0.8$) respectively. Figures (K, L, M): Color divergent heatmaps corresponding to the hinton plots of the figures 3(D-F) in the main manuscript.

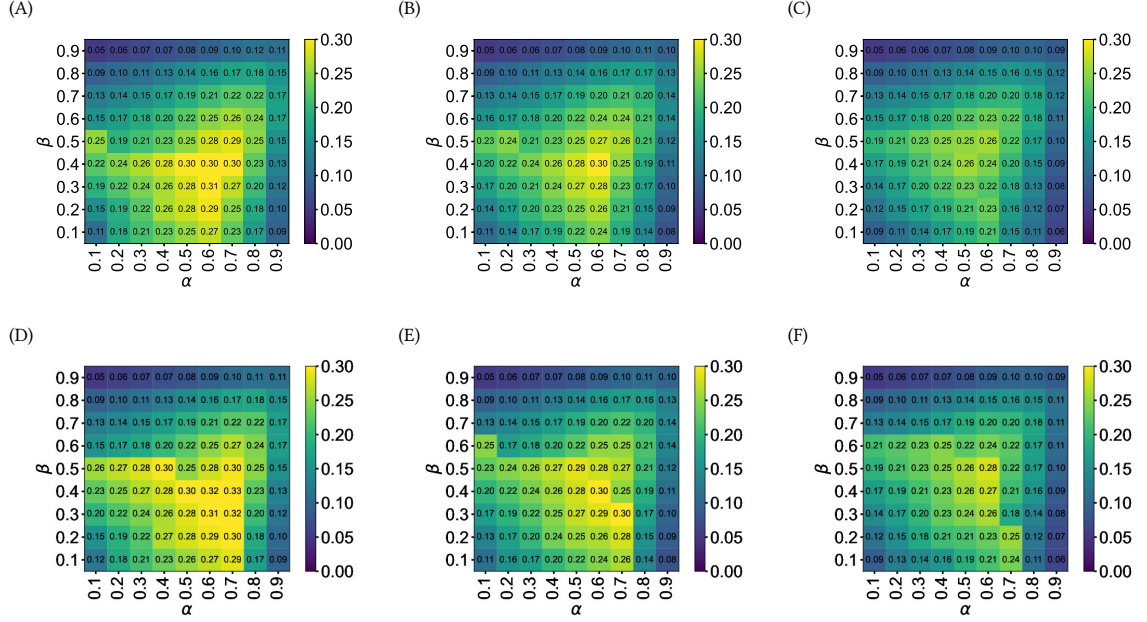


Fig. S3. Generalization to non-antagonistic pair of modifications: The top row (A, B and C) shows the heatmaps for average *BER* values (over 10000 sequences) for the different (α, β, μ) combinations, for the reconstruction performed with the SMAP algorithm. The bottom row (D, E, and F) shows the corresponding plots for α, β, γ values (non-antagonistic cases). In (A, D - $\mu = 0.2, \gamma = 0.8$, in B, E - $\mu, \gamma = 0.5$, C, F - $\mu = 0.8, \gamma = 0.2$).

The PTM data [2] initially exhibits a 500-base pair overlap between consecutive entries, leading to the convolution of the measured enrichment values across the base-pair segments. This is corrected by a deconvolution step, which involves taking the difference over the two overlapping base-pair segments. We start deconvolution from a segment where the enrichment is negligible.

In the base pair segments with negligible enrichment values, to distinguish between the cases where (i) the nucleosome is itself absent or (ii) the PTM is absent in the nucleosome, we use the following approach: If the enrichment values in any of the 5 modifications exceed 0.5 or the sum of enrichments for the 5 PTMs exceeds 1.25, then we consider it the latter case.

Following deconvolution and nucleosome detection, to address the abrupt changes in the enrichment values, we apply a median filter with a window size of 5 nucleosomes for smoothening. The smoothed data is then binarized with a threshold of 0.5.

4 Derivation of Phase Transition Equations

Refer to the trellis diagram in the main manuscript (Fig. 2A). In Case-1, given a gap of k zeros $(1, 0_k, 1)$, with $\rho = 0.5$, if an all-one path has a higher path metric than an all-zero path, we get,

$$\left(1 - \mu + \frac{\mu}{2}\right) \cdot (1 - \alpha) \cdot \left(\left(1 - \mu + \frac{\mu}{2}\right) \cdot \beta\right)^{k-1} \cdot \frac{1 - \beta}{2} \leq \left(\frac{\alpha}{2}\right)^{k+1}, \quad (\text{S15})$$

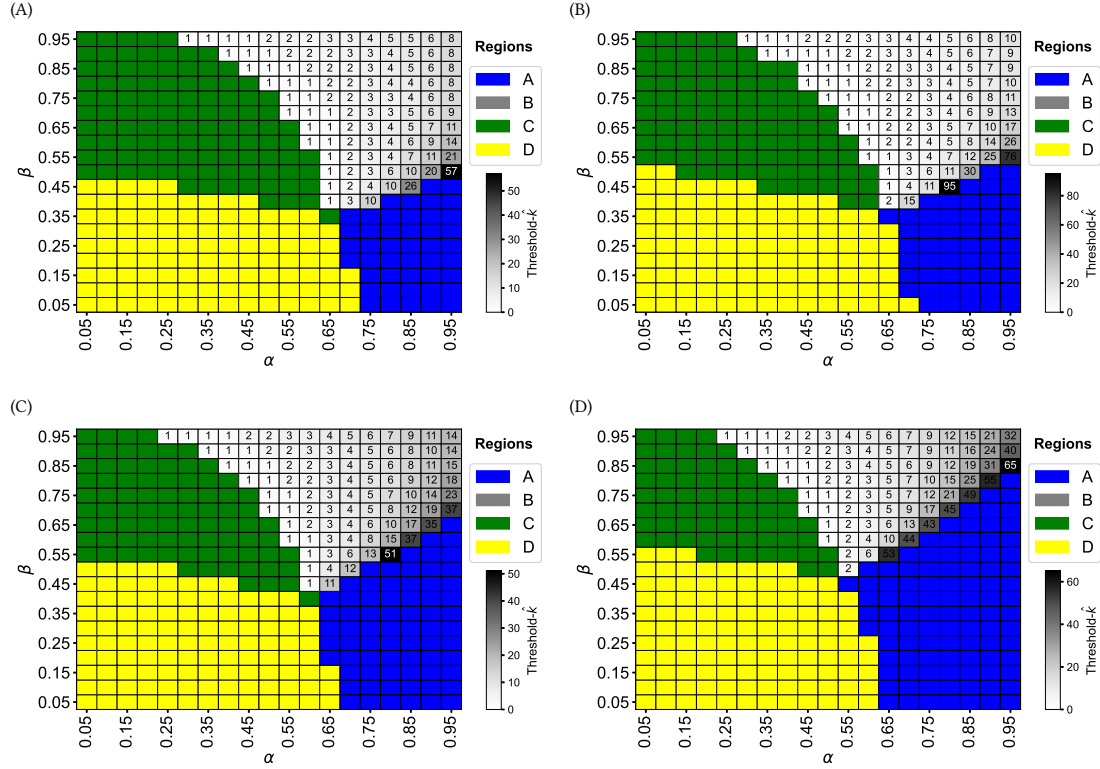


Fig. S4. Threshold- $\hat{k}$ values vary with μ : Reconstruction patterns for (A) $\mu = 0.0$, (B) $\mu = 0.2$, (C) $\mu = 0.5$, (D) $\mu = 0.8$.

Simplifying,

$$\frac{(1-\alpha)(1-\beta)}{2} \leq \frac{\left(\frac{\alpha}{2}\right)^{k+1}}{\left(1-\frac{\mu}{2}\right)^k \cdot \beta^{k-1}}. \quad (\text{S16})$$

The boundary of the region where the gaps are left unfilled can be evaluated by substituting $k = 1$ in Eq. (S16), to obtain the condition

$$\frac{(1-\alpha)(1-\beta)}{2} = \frac{\left(\frac{\alpha}{2}\right)^2}{\left(1-\frac{\mu}{2}\right)}. \quad (\text{S17})$$

Simplifying, we get

$$\frac{\alpha^2}{2} - (1-\alpha)(1-\beta) \left(1 - \frac{\mu}{2}\right) = 0. \quad (\text{S18})$$

4.1 Derivation of threshold- $\hat{k}$

Taking logarithm on both sides of Eq. (S16) and simplifying, we get

$$\log\left(\frac{(1-\alpha)(1-\beta)}{2}\right) = k \log\left(\frac{\frac{\alpha}{2}}{\left(1-\frac{\mu}{2}\right)\beta}\right) + \log\left(\frac{\alpha\beta}{2}\right). \quad (\text{S19})$$

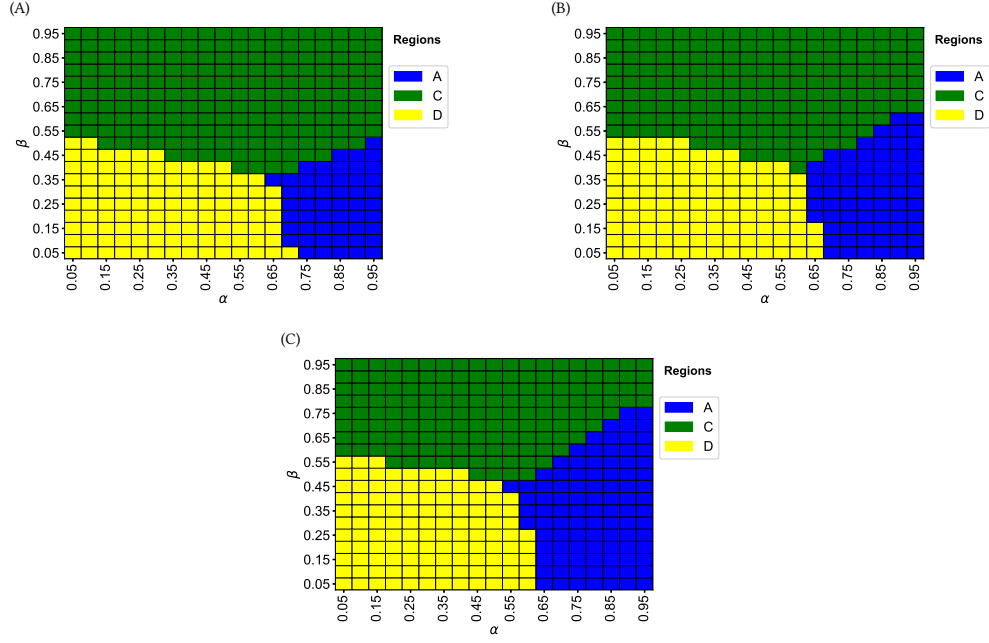


Fig. S5. Region boundaries of reconstruction vary with μ : Reconstruction patterns for Cases 2, 3, 4 (A) $\mu = 0.2$, (B) $\mu = 0.5$, (C) $\mu = 0.8$.

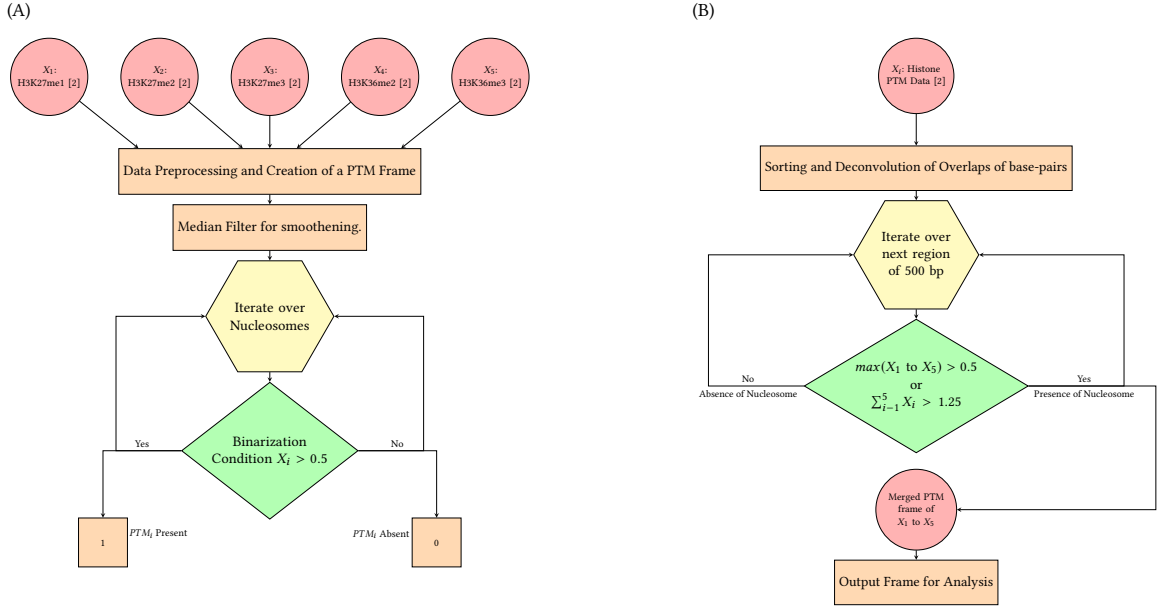
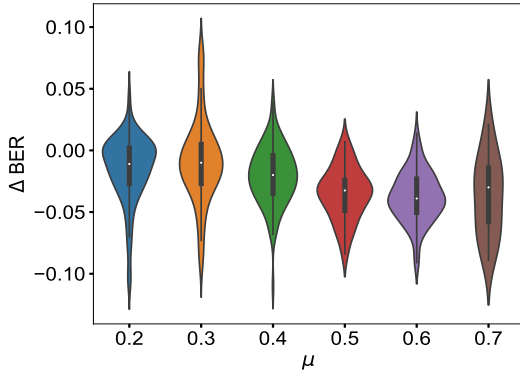


Fig. S6. (A) Discretization algorithm for the binarization of the ChIP-Seq data obtained from the [2]. (B) The pipeline for preprocessing the data includes deconvolution and nucleosome detection.

(A)



(B)

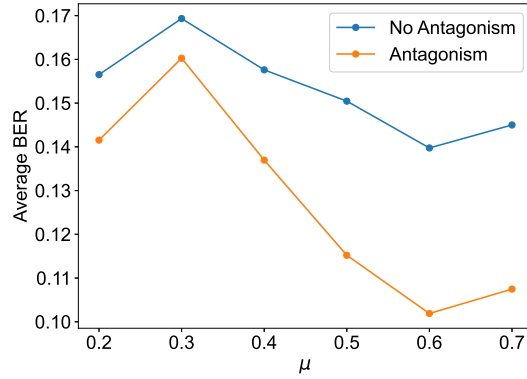


Fig. S7. Experimental Data: (Chr-18) (A) Violin plot of the difference in *BER* between the antagonistic and non-antagonistic cases and (B) line plot of the means of *BER* distributions for the non-antagonistic and antagonistic cases for the sequences binarized from the ChIP-Seq data in Chr 18 of the mESCs obtained from [2]. The selected chromatin regions were found to have the statistical parameters: $\alpha \in [0.7, 1.0]$, $\beta \in [0.5, 0.8]$, $\mu \in [0.2, 0.7]$.

Therefore,

$$k = \frac{\log \left(\frac{(1-\alpha)(1-\beta)}{\alpha\beta} \right)}{\log \left(\frac{\frac{\alpha}{2}}{(1-\frac{\mu}{2})\beta} \right)}. \quad (\text{S20})$$

The straight line shown in Fig. 4C corresponds to the denominator in Eq. (S20) being 0. This yields

$$\alpha = (2 - \mu)\beta. \quad (\text{S21})$$

4.2 Delta BER Plots

Following the discretization procedure, we plotted the ΔBER (Fig. S7A) and the means of *BER* (Fig. S7B) for the antagonistic and non-antagonistic cases for the data in a specific region - Chr18 of the mESCs dataset from [2].

5 Simulations of different cases of daughter chromatin

For the simulations described in Sec. 3 of the main manuscript, we analyse the results for no-antagonism ($\mu = 0.0$) and different ranges of antagonism ($\mu = 0.2, 0.5, 0.8$). The results are shown in Fig. S4 (Case 1) and Fig. S5 (Cases 2, 3, 4). Since Cases 2, 3, and 4 are having similar results, we show a representative case (Case 3) in Fig. S5.

From Fig. S4 one can infer that region A expands with μ . additionally, it can be observed that the threshold- $\hat{k}$ increases with μ for high values of (α, β) . As in Fig. 4 of the main manuscript, region B (threshold- $\hat{k}$) can be observed only in Case 1.

6 Variation of *BER* across multiple generations

To analyze the changes in *BER* across multiple generations, we performed the following simulation: 50 mother sequences of length 10000, α and β values both set to 0.7 were generated for each μ in the set $\{0.2, 0.5, 0.8\}$.

100 daughter sequences were produced from each mother sequence as described in Sec. 2.1 of the main manuscript. Correction based on SMAP decoding (which amounts to threshold- $\hat{k}$ filling, since it is region B) led to the average *BER* (with respect to the first generation mother sequence) plots as shown in Fig. S8. While the average *BER* (without any further secondary corrections over the cell cycle) increases with generations as expected, this fidelity deterioration is less pronounced in the presence of antagonistic PTMs.

7 *BER* across different parametric ranges

Fig. S9 shows different regions of the chromatin (from mESCs obtained from GSE135029) with PTM *A* and PTM *B* represented as binary sequences (the lines representing 1). The daughter sequences before and after correction are also shown. The parameter values, *BER* and the regions they fall under (from Fig. 4C of the main manuscript) are also provided. The red box in Fig. S9D shows an instance where the presence of the antagonistic mark in the vicinity of the primary marks aids in the correct prediction of the primary PTM.

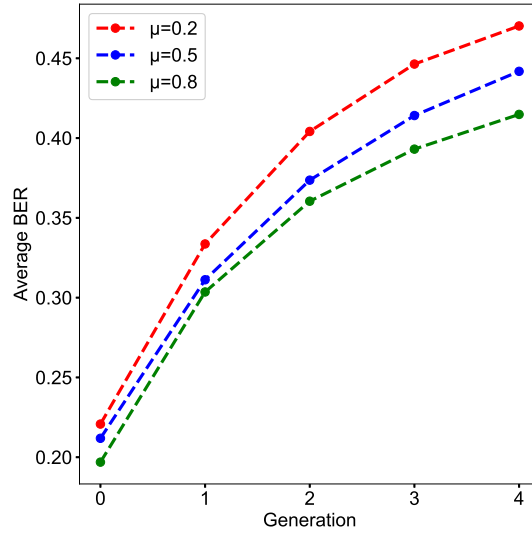
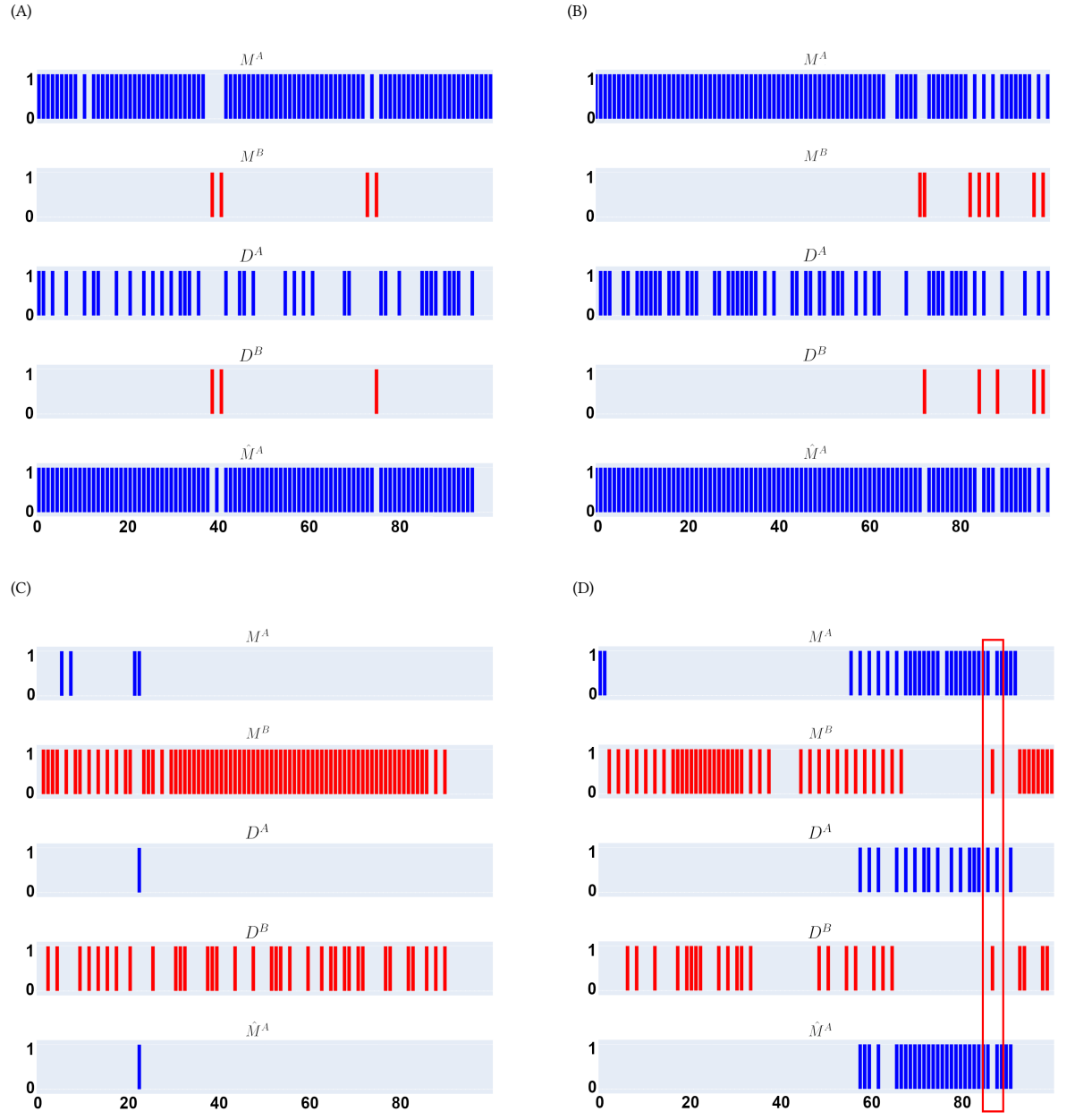


Fig. S8. Simulation plots of average *BER* for 5 generations (Generation 0 represents the set of original mother sequences). α and β values are set to 0.7 each. The *BER* values were averaged over 5000 daughter sequences each of length 10000.



References

- [1] Nithya Ramakrishnan, Sibi Raj B Pillai, and Ranjith Padinhateeri. High fidelity epigenetic inheritance: Information theoretic model predicts threshold filling of histone modifications post replication. *PLOS Computational Biology*, 18(2):e1009861, 2022.
- [2] Constance Alabert, Carolin Loos, Moritz Voelker-Albert, Simona Graziano, Ignasi Forné, Nazaret Reveron-Gomez, Lea Schuh, Jan Hasenauer, Carsten Marr, Axel Imhof, et al. Domain model explains propagation dynamics and stability of histone h3k27 and h3k36 methylation landscapes. *Cell reports*, 30(4):1223–1234, 2020.