A Physics-based Model Reveals Mechanisms of Epigenetic Memory

Erik J. Navarro[†1], William Poole[†1], Ariane Lismer[2], Jieqiong Qu[2], Ioannis Kafetzopoulos[2], Aled Parry[2], Jannat Ijaz[2], Alice Santambrogio[2], Ryan Spangler[1], Francisco M. Martín-Zamora[3], Ken Raj[2], Wolf Reik[2], Hana El-Samad[1], Carlos F. Lopez[3], Simone Bianco [*1]

[1]Altos Labs - Institute of Computation [2]Altos Labs - Institute of Science

[3]Altos Labs - Institute of Technology [*]Corresponding author. Email: sbianco@altoslabs.com

[†]These authors contributed equally to this work.

In multi-cellular eukaryotic organisms, cell type and specific functional identity are defined by the epigenetic patterning of chemical modifications to DNA and chromatin that modulate the expression and silencing of specific genes. When a cell divides, histones containing important epigenetic marks are distributed between the two daughter strands leading to a temporary dilution of epigenetic information and cell identity. In this work we introduce a physics-based model of epigenetic memory that explains how cells restore and maintain H3K9me3 and H3K27me3 histone methylation patterning after cell division. We demonstrate that emergence and maintenance of the epigenetic program is driven by an evolved mechanism that makes use of the biophysics of polymers, phase condensates and enzymatic activity. We validate our model via genome-wide epigenetic time-course simulation and comparison to experimental epigenetic data from multiple donors, multiple cell types, and for multiple epigenetic marks. Finally, we use our model as a conceptual framework to understand cellular reprogramming by hypothesizing that these processes first contend with and later utilize

somatic epigenetic maintenance programs.

1 Introduction

Chromatin, the highly dynamic molecular complex composed of DNA, RNA, and proteins, is involved in many aspects of cellular behavior, ranging from adaptation and response to perturbations to organismal physiology (*1*). Epigenetics modulate the expression of genes through physically and chemically modifying chromatin in a cell-type specific manner patterned over the course of development (*2*). The progressive loss of epigenetic stability has been linked to aging and disease, and recent evidence has shown that the epigenome is reset during partial cellular reprogramming and rejuvenation (*3*).

Among the epigenetic modifications that occur during development, histone modifications are of particular importance. Histones are the proteins that bind to DNA forming molecular octamers called nucleosomes, which are one of the main structural elements of chromatin. Histones are frequently chemically modified via a complex code of acetylation, methylation, ubiquitination, and phosphorylation marks which help determine cell type by modulating gene expression (*4*). Two chromatin-silencing histone methylation marks are known to be of particular importance in determining cell state: H3K9me3 and H3K27me3, which are catalyzed (in humans) by methyl-transferases SUV39 and PRC2, respectively (*5*). Their loss is directly linked to aging and results in chromatin instability and increased cellular stress (*6*). Additionally, partial reprogramming has been associated with restoration of epigenetic marks (*7*).

Unlike the inheritance of genetic information encoded as a DNA sequence, histone marks like H3K9me3 and H3K27me3 are not directly copied during DNA replication. In humans it has been established that modified parental histones bearing H3K9me3 and H3K27me3 modifications are distributed symmetrically (meaning randomly with equal probability) between the two daughter cells during cell division (*8*). This means that after division, each daughter cell receives a random combination of epigenetically modified histones and new synthesized naive histones not yet bearing epigenetic modifications (*8, 9*). The daughter cells are then tasked with restoring the full epigenetic pattern from this diluted signal in order to preserve cellular identity and function. On the other hand, full and partial reprogramming aim to rewrite and restore the epigenome, respectively. It is unlikely

that partial or full reprogramming via Yamanaka factors fundamentally changes the mechanisms of epigenetic inheritance. Therefore, the formation and maintenance of epigenetic memory during cellular division is not merely a characteristic of cellular health, but a fundamental process that is both utilized and challenged by cellular reprogramming.

The mechanisms of epigenetic inheritance, stability, and memory across generations are still major outstanding questions in cellular biology. Early studies have aimed to model the spread of epigenetic marks from a single nucleation site (*10, 11*). It has been postulated that epigenetic patterns may be maintained through epigenetic boundary elements that constrain and maintain epigenetic domains (*12, 13*), or by invoking 3D polymer interactions coupled with additional effects such as co-localization of methyltransferases to methylated histones (*14*), domain-specific kinetic parameters (*15*), cross-talk between histone marks and DNA methylation (*16*), and attractive forces between methylated histones coupled with chromatin compression during mitosis (*17*). However, these mechanisms are primarily supported by highly idealized, small-scale, and/or artificial computational validation studies. There is little evidence that these mechanisms are acting at the genome wide level due to the lack of genome-wide validation using epigenetic modalities like ChIP-seq or CUT&RUN. One notable exception achieves fewer than 10 generations of pattern maintenance (*14*) on such a validation.

Our work builds on these efforts to model the inheritance and restoration of H3K9me3 and H3K27me3 methylation marks through the cell cycle as illustrated in Figure 1 A. We propose a much simplified model, which we call the Looping, Enzyme And Phase-Condensate (LEAP) model, involving polymer looping mediated catalysis and co-localization of methyltransferases to densely methylated genomic regions. We prove this model to be sufficient in generating stable epigenetic patterns that persist across many generations. Importantly, this model does not rely on epigenetic boundary elements, domain-specific parameters, or large density differences between chromatin domains, which have recently been called into question (*18, 19*). Additionally, we validate our model on genome-scale H3K9me3 and H3K27me3 data from multiple cell types.

2 The Looping, Enzyme And Phase-condensate (LEAP) Model

Conceptual Model: The Looping, enzyme, and phase-condensate (LEAP) model of epigenetic restoration combines three experimental observations about histone methyltransferases. First, these enzymes are known to bind at higher frequency to their own reaction product. This implies positive feedback or a strong preference for the enzymes to interact with the methylated state rather than the unmethylated or other state (*20, 21*). Second, these enzymes' catalytic activity is mediated by cooperatively binding to a methylated histone and an unmethylated histone simultaneously (*20, 21*). Essentially, by binding to its own product, the histone methyltransferase can methylate adjacent histones and ensure that the mark is maintained, reinforced or potentially spread across the local chromatin structure. It is well known that polymers in solution, including chromatin, produce loops which bring together linearly separated regions of DNA. The observed probability of two monomers coming into contact scales like an inverse power law with the linear distance and an exponent that depends on the solvent conditions and may vary between cell and organism types (*22*).

Third, it has been observed that H3K9me3 and H3K27me3 methyltransferases co-localize in high density condensates that form around regions of dense H3K9 and H3K27 methylation (*23, 24*). The formation of these condensates potentially increases the local reaction rate and ensures epigenetic specificity by enforcing a spatial organization. After cell division, even though only half the original number of methylation histones are present, there is still sufficient density for enzyme condensates to localize around epigenetic domains. These regions are then able to reliably restore themselves due to polymer-polymer looping which ensures that methylation is contained in a relatively local region around these dense condensates. This model is illustrated in Figure 1 B.

Idealized Model: We mathematically model an idealized epigenetic code consisting of a single epigenetic mark (methylation) which resides along the monomers of a polymer chain representing chromatin. This polymer has N monomers (nucleosomes containing a modifiable histone), each denoted by $x_i \in \{0, 1\}$, which indicates whether the histone i is unmodified ($x_i = 0$) or methylated ($x_i = 1$). The ordering of these monomers in a linear chain $X = (x_1, ..., x_i, ..., x_N)$ constitutes a specific epigenetic pattern. The pattern degrades due dilution at cell division, modeled as binomial partitioning occurring regularly with a frequency $1/t_{div}$. The pattern restores itself via loop-mediated methyltransferase activity. Loops are allowed to form between sites i and j with

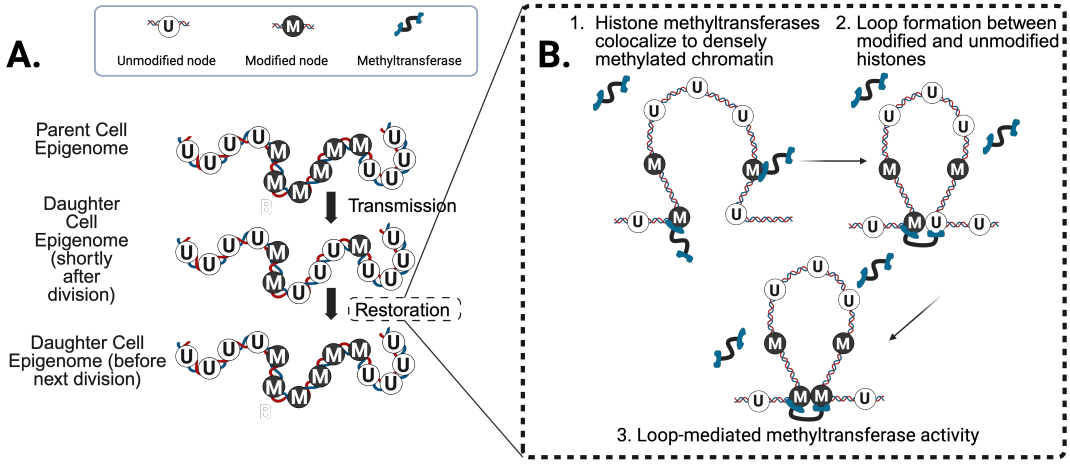

Figure 1: A. Cartoon illustration of epigenetic dilution due to cell-division and subsequent restoration. B. A cartoon of the proposed mechanism of epigenetic restoration; (1) histone methyltransferase co-localization to methylated chromatin, (2) loop formation, and (3) loop mediated catalysis.

probability P_L through the polymer physics derived inverse-power law (*25*). Methyltransferases (E) may be bound to any given methylated monomer with probability P_E modeled as a sigmoid function of the local marked-monomer density along the chain, representing the non-linear effects of methyltransferase phase condensation and co-localization to methylated histones. The resultant rate at which a monomer i may gain a mark, $Q(x_i = 0 \rightarrow x_i = 1)$, is given by:

$$Q(x_i = 0 \rightarrow x_i = 1) = \sum_j k_m P_L(i,j) P_E(j) \quad \text{with} \quad P_L(i,j) = k_c \frac{1}{d_{ij}^{\alpha}},$$

$$P_E(j) = \frac{\sigma(j)}{Z} E_b^*, \qquad \sigma(j) = \frac{\rho(j)^n}{K_A^n + \rho(j)^n} + b(j,t), \quad \text{and} \quad Z = \sum_j \sigma(j).$$

Here d_{ij} is the distance between monomers i and j along the polymer chain and $b(j,t)$ is a time-dependent bias term assumed to be 0 unless otherwise noted. The local monomer density ρ_j is computed with a Gaussian kernel density estimate with bandwidth λ. Additionally, t_{div} is the division time, k_m is the methylation rate of the methyltransferase, k_c is the contact rate related to polymer mobility, α is the scaling exponent of the contact law, K_A and n parameterize the non-linearity σ, and $E_b^* = \frac{1}{2}\left((N_M + E + K_D) - \sqrt{(N_M + E + K_D)^2 - 4EN}\right)$ is the total number of bound enzymes which depends on the total number of enzymes E, the dissociation constant K_D, and the number of marked monomers $N_M = \sum_i x_i$. This model is formalized and simulated as a

continuous time Markov chain (CTMC) as elaborated in the SI A.

Coarse Grained Model: simulating the above model at the genomic scale is prohibitively expensive from a computational standpoint due to the quadratic number of terms required to evaluate $Q(x_i = 0 \rightarrow x_i = 1)$ for all i. The human genome has around 30×10^6 nucleosomes and larger chromosomes may have over 10^6 nucleosomes which would result in an equation with $\sim 10^{12}$ terms evaluated at every simulation timestep. In order to carry out genome-scale simulations, we developed a coarse grained model which allows us to simulate sets of marks pooled together into a single monomer. Briefly, instead of simulating dynamics for $X \in \{0, 1\}^N$ where each monomer represents a single marked or unmarked position of chromatin, we simulate $Y \in \{0, 1, ...\mu\}^\nu$ where μ are the maximum number of marks at any given position and $\nu = N/\mu$ are the number of coarse grained monomers. For example, $\mu = 2$ would be appropriate for a nucleosome-level simulation where each nucleosome can have 0, 1, or 2 marks. If $\mu > 2$, this indicates that multiple nucleosomes are coarse-grained into a single simulated monomer. For a detailed description of this model and simulation details see SI C.

3 Results

3.1 The mechanisms captured by the idealized LEAP model are sufficient to generate epigenetic memory

We show that epigenetic patterns can be stably maintained over hundreds of cell cycles using an illustrative pattern simulated with the idealized LEAP model. These dynamics are visualized over a single generation, multiple generations, and as transition rates in Figure 2 A, B, and D. We quantify epigenetic drift by computing the auto-correlation between the initial pattern and later patterns in Figure 2 C. Notice that this correlation is much higher when the data is smoothed because the stochastic nature of the model results in high variance at the single mark scale despite epigenetic domains being largely stable.

The simulations demonstrate that polymer looping and methyltransferase methyl-mark co-condensate formation are both necessary to stably maintain epigenetic patterns. Removing either the methylation density dependence non-linearity σ or the inverse power law contact dependence

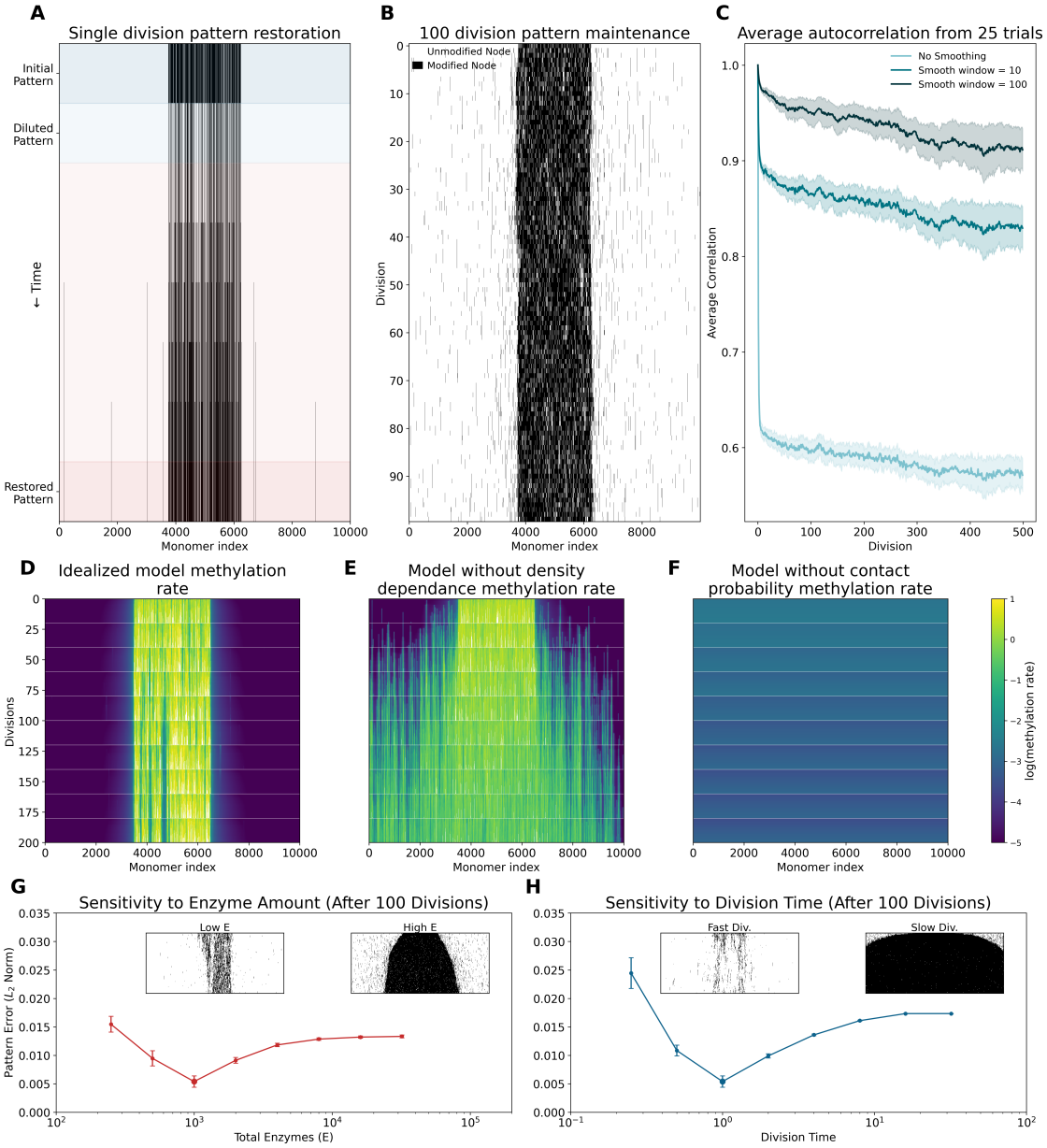

Figure 2: A. Simulation results for a single dilution-restoration cycle. B. Simulation results for 100 dilution-restoration cycles. C. Auto-correlation computed between the initial condition and subsequent time-points; correlations are computed using data smoothed with a uniform window. D. The methylation rate simulated with the idealized model. E. The methylation rate simulated using a model without a density dependence term. F. the methylation rate simulated using a model without the contact probability term. G. Idealized model sensitivity to the enzyme amount. H. Idealized model sensitivity to the division time.

P_L results in the inability for patterns to be maintained as shown in Figure 2 E and F. The contact probability term helps patterns localize along the chain but cannot constrain them from rapidly spreading. The co-condensate formation provides a non-linearity that stabilizes the mark locally and dramatically diminishes spreading but is alone insufficient to localize patterns. Furthermore, these two features are together sufficient for spontaneous pattern formation, as investigated in SI B.

3.2 The coarse Grained LEAP model shows remarkable accuracy in predicting epigenetic marks for hundreds of generations in biological experiments, across different ages and cell types

We apply the coarse LEAP model with $\mu = 10^4$, corresponding to 50 nucleosomes per coarse grained monomer, to simulate real H3K9me3 and H3K27me3 data collected from passaged HUVECs and human derived fibroblasts (HDFs), see SI G and G.4. Example snapshots and simulated dynamics of the model can be seen in the first two columns of Figure 3. This model is applied to HDF H3K27me3 (first row), HDF H3K9me3 (second row), and HUVEC H3K27me3 (third row) data. Additional simulation output and implementation details can be found in SI D. Notably, the model is able to recapitulate H3K9me3 and H3K27me3 epigenetic patterns in 8 different HDFs for over 100 generation with high accuracy, as measured via the correlation between the initial pattern and the simulated data after 100 generations Figure 3 B and F. We further compute the auto-correlation times for each mark in each cell type (Figure 3 C, G, K) using both a smoothed (uniform smoothing with a window of 100 coarse monomers which is equivalent to 1Mb) and unsmoothed state to compute the correlation. The increased auto-correlation time of the smoothed data indicates that that larger epigenetic patterns on the scale of Mb (which correspond to topologically associated domains (TADs) (*26*)) can be accurately maintained over hundreds of generations. TADs are thought to constrain and control interactions between enhancers and their target genes, and their disruption is linked with loss of cell type specific epigenetic memory (*27*). Finally, due to having time-course epigenetic data from HUVEC cells as a function of cumulative population doubling, we compare the model dynamics to the epigenetic dynamics. We find that our model predicts more epigenetic drift in HUVECs than we see in the actual data illustrated in Figure 3 L by the observation that young simulations (which have not drifted far from data derived from young

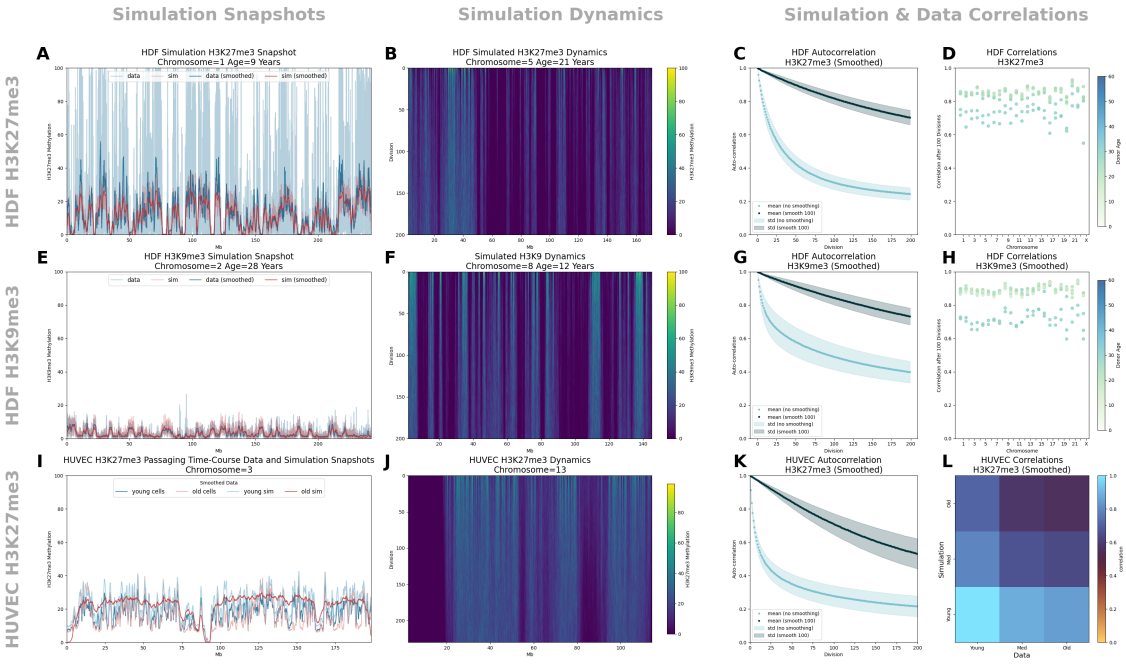

Figure 3: A. Example H3K27me3 data and simulation output after 100 divisions for a single HDF donor and chromosome. B. Example H3K27me3 simulation dynamics for a single HDF donor and chromosome. C. Auto-correlation of the HDF H3K27me3 simulations averaged across all donors and chromosomes shown for unsmoothed data and smoothed data (uniform smoothing with a window of 100). D. Correlation between HDF H3K9me3 smoothed simulation output after 100 generations and smoothed data split by donor and chromosome and colored by age. E. Example H3K9me3 data and simulation output after 100 divisions for a single HDF donor and chromosome. F. Example H3K9me3 simulation dynamics for a single HDF donor and chromosome. G. Auto-correlation of the HDF H3K9me3 simulations averaged across all donors and chromosomes shown for unsmoothed data and smoothed data (uniform smoothing with a window of 100). H. Correlation between HDF H3K27me3 smoothed simulation output after 100 generations and smoothed data split by donor and chromosome and colored by age. I. Young (CPD = 34, simulated divisions = 10) , medium (CPD = 150, simulated divisions = 150) , and old (CPD = 263, simulated divisions = 225) H3K27me3 data and simulation output. CPD is the cumulative population doubling score. J. Example H3K27me3 simulation dynamics for a single HUVEC chromosome beginning at a young state. K. HUVEC H3K27me3 auto-correlation of the simulation dynamics with the initial state for 25 replicates. L. Correlation between young, medium, and old simulations and data (as described in the caption for panel I).

cells) correlate more strongly to data derived from old passaged cells than old simulations. We conceptualize this discrepancy by noting that our model provides a baseline for epigenetic drift but does not take into account any other biological processes which may be actively maintaining and modulating histone methylation patterns as cells age. Indeed, HUVECs are common in-vitro model of aging in part because they have a well-characterized senescence phenotype that emerges over cell passages (*28*); this cell state transition may be responsible for discrepancies with the LEAP model.

3.3 Insights into the therapeutic effects of reprogramming

To represent the remodeling of the epigenome during development and reprogramming processes (*29*), we use the idealized model with a bias term $b(j,t) \geq 0$. Both development and cellular reprogramming are known to involve epigenetic changes in H3K9me3 and H3K27me3 marks (*30, 31*). The bias term can be interpreted as a change in the density of methyltransferases at specific sites. This effect mimics important observations of transcription factor mediated reprogramming. It has been observed that reprogramming factors (particularly Oct4) cause reorganization of phase condensates via H3K27 acetylation (*32*), which is mutually exclusive to H3K27me3 (*33*). Moreover, pioneering transcription factors like OSKM are known to recruit chromatin modifying enzymes to the site of binding, effectively biasing the density of enzymes across the genome. In general, our model predicts that any epigenetic modification due to stress, disease, or intervention, may persist in the absence of post-modification correction. Therefore, we use simulations of the idealized model to conceptualize how epigenetic stress (red shaded regions in Figure 4) can create changes in epigenetic patterns that are propagated after the stress is removed, which has been observed experimentally (*30*). Similarly, therapeutic strategies such as rejuvenation via partial reprogramming (*7*) can be seen as epigenetic modifications that restore patterns to a healthier state, which has also been observed via tools such as epigenetic clocks (*34*). Figure 4A-C show how different dosages of epigenetic therapies can result in different outcomes ranging from partial reprogramming to full reprogramming that changes cell identity. Figure 4D shows the bias terms $b(j,t)$ applied in these simulations.

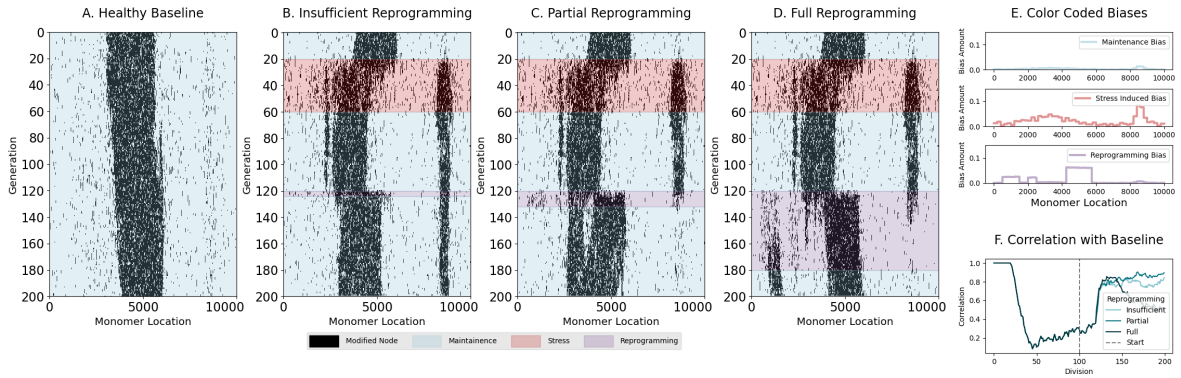

Figure 4: A. Idealized model simulation of a healthy baseline without stress or reprogramming. B. Idealized model simulation with stress and a short reprogramming intervention. C. Idealized model simulation with stress and medium length reprogramming intervention. D. Idealized model simulation with stress and a long reprogramming intervention. E. Bias terms used for the maintenance, stress, and reprogramming periods of the simulation. F. Correlation of the different reprogramming interventions with the healthy baseline as a function of cell divisions using smoothed output (window = 100).

4 Discussion

The LEAP model illustrates how the fundamental biophysics of polymer looping, enzyme kinetics, and phase condensate formation can underlie a robust and re-writable epigenetic patterning mechanism. At a conceptual level, the LEAP model shows how epigenetics can be viewed as an analog memory system. This perspective follows from a long tradition of understanding chromatin regulation as a form of neural computation (*35, 36*). A recent trend has been towards theoretically modeling chromatin in terms of energy landscapes; (*37*) suggested that reprogramming can be thought of as annealing on a chromatin energy landscape, (*38*) connect liquid-liquid phase condensates to programmable spatial Hopfield networks, and (*39, 40*) interpret enzyme-mediated polymer looping as an energy-based generative model. The LEAP model stands out from these works because it moves beyond theoretical frameworks to model specific biological molecules and epigenetic marks. In the process, we find that cells can evolve a flexible computational system capable of maintaining a wide range of patterns without specifying an external (or learned) energy

function or landscape.

Additionally, the LEAP model makes several qualitative predictions which agree with previous literature. First, the LEAP model predicts that the total amount of methylation is approximately conserved (see SI F) which is consistent with experimentally observed histone mark redistribution (*41–44*). This can be seen in the sensitivities of epigenetic patterns to the amount of enzyme E and the division time which conceptually to control the effective size and density of epigenetic regions as depicted in Figure 2 G and H. The sensitivity analysis suggests that faster dividing cells must either have higher levels of enzymes to compensate, or lower levels of the methylation mark. Quantitative measurements of PRC2 proteins in different cell types show an approximately ten fold higher number of enzymes in mouse embryonic stem cells, which are reported to divide every $\sim 4-12$ hours (*45,46*) compared to both glioma and human embryonic kidney cells (HEK), which divide roughly every 48 hours (*47*). Comparing glioma cells with HEK cells, which have roughly equal doubling times, HEK cells have about two times more enzymes than glioma cells, and a also higher percentage of methylated histones (*47*). Our model is also consistent with the findings of Jhadev at al. (*48*), who experimentally decreased the levels of methyltransferase, ledading to replicative dilution of the mark; in addition they observed that dilution per cell cycle is dependant on replication speed, with faster dividing cells diluting faster.

We demonstrate the generalizability of our model via the first ever genome-wide histone methylation mechanistic simulation using data from two cell types, two methylation marks, and 8 different donors. The correlations between the simulations and data indicate that the LEAP model can reliably explain epigenetic maintenance at the TAD scale. However, the LEAP model performs considerably worse on small length scales and tends to over-estimate epigenetic drift when compared to cell-passage experiments. Jointly these results indicate that there are additional biological processes biasing the formation of and maintenance of epigenetic patterns.

Finally, we capitalize on this framework model to build a simple model to mimic epigenetic rejuvenation that captures the capability of metozoan cells to be reprogrammed. Our simulations demonstrate a mechanism by which epigenetic patterns can persist after the reprogramming signal has been removed. In this model, a bias signal promotes methylation in the regions overlapping with the original pattern. This is consistent with the idea that pioneering transcription factors, such as the Yamanaka factors, when dosed and timed appropriately can reactivate the developmental

programs that established the somatic patterns (*49–51*). The model can also be instructive about chemical reprogramming based on the observation that small molecules may disrupt chromatin phase condensates (*52*). Rejuvenation in our model arises naturally through a dynamic reorganization of the epigenome caused by accumulation of epigenetic modifying enzymes in regions becoming increasingly methylated through the action of the bias signal. This causes regions to which epigenetic marks have drifted to slowly lose their marks through competitive process similar to transcription factor theft (*53*), without a need to recognize, keep track of, or know *a priori* in any way the genomic location of epigenetic drift. A consequence of this fact is the observation that reprogramming can occur in many different cell types. This hints at the possibility of partial reprogramming exploiting the precise regulation afforded by the co-evolution of transcription factors, epigenetic regulators, and the physical substrate they bind to during development.

References and Notes

1. P. Sen, P. P. Shah, R. Nativio, S. L. Berger, Epigenetic Mechanisms of Longevity and Aging. Cell **166** (4), 822–839 (2016), doi:10.1016/j.cell.2016.07.050.

2. K. K. Duclos, J. L. Hendrikse, H. A. Jamniczky, Investigating the evolution and development of biological complexity under the framework of epigenetics. Evolution & Development **21** (5), 276–293 (2019), doi:10.1111/ede.12301.

3. A. D. Yücel, V. N. Gladyshev, The long and winding road of reprogramming-induced rejuvenation. Nature Communications **15** (1), 1941 (2024), doi:10.1038/s41467-024-46020-5.

4. B. Carter, K. Zhao, The epigenetic basis of cellular heterogeneity. Nature Reviews Genetics **22** (4), 235–250 (2021).

5. S. I. Grewal, The molecular basis of heterochromatin assembly and epigenetic inheritance. Molecular cell **83** (11), 1767–1785 (2023).

6. N. Yang, et al., A hyper-quiescent chromatin state formed during aging is reversed by regeneration. Molecular cell **83** (10), 1659–1676 (2023).

7. D. Gill, et al., Multi-omic rejuvenation of human cells by maturation phase transient reprogramming. eLife **11**, e71624 (2022), doi:10.7554/elife.71624.

8. C. Alabert, et al., Two distinct modes for propagation of histone PTMs across the cell cycle. Genes & Development **29** (6), 585–590 (2015), doi:10.1101/gad.256354.114.

9. N. Reverón-Gómez, et al., Accurate recycling of parental histones reproduces the histone modification landscape during DNA replication. Molecular cell **72** (2), 239–249 (2018).

10. N. A. Hathaway, et al., Dynamics and memory of heterochromatin in living cells. Cell **149** (7), 1447–1460 (2012).

11. F. Erdel, E. C. Greene, Generalized nucleation and looping model for epigenetic memory of histone modifications. Proceedings of the National Academy of Sciences **113** (29), E4180–E4189 (2016).

12. I. B. Dodd, M. A. Micheelsen, K. Sneppen, G. Thon, Theoretical Analysis of Epigenetic Cell Memory by Nucleosome Modification. Cell **129** (4), 813–822 (2007), doi:10.1016/j.cell.2007.02.053.

13. J. F. Nickels, K. Sneppen, Confinement Mechanisms for Epigenetic Modifications of Nucleosomes. PRX Life **1** (1), 013013 (2023), doi:10.1103/prxlife.1.013013.

14. S. H. Sandholtz, D. Kannan, B. G. Beltran, A. J. Spakowitz, Chromosome Structural Mechanics Dictates the Local Spreading of Epigenetic Marks. Biophysical Journal **119** (8), 1630–1639 (2020), doi:10.1016/j.bpj.2020.08.039.

15. A. Z. Abdulla, C. Vaillant, D. Jost, Painters in chromatin: a unified quantitative framework to systematically characterize epigenome regulation and memory. Nucleic Acids Research **50** (16), 9083–9104 (2022), doi:10.1093/nar/gkac702.

16. S. Palacios, et al., Analog epigenetic memory revealed by targeted chromatin editing. Cell Genomics **5** (11) (2025).

17. J. A. Owen, D. Osmanović, L. Mirny, Design principles of 3D epigenetic memory systems. Science **382** (6672), eadg3053 (2023), doi:10.1126/science.adg3053.

18. T. Nozaki, et al., Condensed but liquid-like domain organization of active chromatin regions in living human cells. Science Advances **9** (14), eadf1488 (2023).

19. H. K. Prajapati, Z. Xu, P. R. Eriksson, D. J. Clark, Nucleosome dynamics render heterochromatin accessible in living human cells. Nature Communications **16** (1), 4577 (2025).

20. C. Saab, J. Stephan, E. Akoury, Structural insights into the binding mechanism of Clr4 methyltransferase to H3K9 methylated nucleosome. Scientific Reports **14** (1), 5438 (2024), doi:10.1038/s41598-024-56248-2.

21. R. Leicher, et al., Single-Molecule Investigation of PRC2 Non-Adjacent Nucleosome Bridging. Biophysical Journal **118** (3), 380a (2020).

22. E. Lieberman-Aiden, et al., Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science **326** (5950), 289–293 (2009), doi:10.1126/science.1181369.

23. M. M. Müller, B. Fierz, L. Bittova, G. Liszczak, T. W. Muir, A two-state activation mechanism controls the histone methyltransferase Suv39h1. Nature chemical biology **12** (3), 188–193 (2016).

24. J. M. Eeftens, M. Kapoor, D. Michieletto, C. P. Brangwynne, Polycomb condensates can promote epigenetic marks but are not required for sustained chromatin compaction. Nature communications **12** (1), 5888 (2021).

25. H. S. Chan, K. A. Dill, Intrachain loops in polymers: Effects of excluded volume. The Journal of Chemical Physics **90** (1), 492–509 (1989), doi:10.1063/1.456500.

26. J. R. Dixon, et al., Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature **485** (7398), 376–380 (2012).

27. F. Paldi, G. Cavalli, 3D genome folding in epigenetic regulation and cellular memory. Trends in Cell Biology (2025).

28. R. Hastings, M. Qureshi, R. Verma, P. Lacy, B. Williams, Telomere attrition and accumulation of senescent cells in cultured human endothelial cells. Cell proliferation **37** (4), 317–324 (2004).

29. K. Takahashi, S. Yamanaka, Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell **126** (4), 663–676 (2006), doi:10.1016/j.cell.2006.07.024.

30. S. Buckberry, et al., Transient naive reprogramming corrects hiPS cells functionally and epigenetically. Nature **620** (7975), 863–872 (2023), doi:10.1038/s41586-023-06424-7.

31. K. Huang, et al., Dynamically reorganized chromatin is the key for the reprogramming of somatic cells to pluripotent cells. Scientific reports **5** (1), 17691 (2015).

32. J. Wang, et al., Phase separation of OCT4 controls TAD reorganization to promote cell fate transitions. Cell Stem Cell **28** (10), 1868–1883 (2021).

33. P. Dasgupta, P. Prasad, S. K. Bag, S. Chaudhuri, Dynamicity of histone H3K27ac and H3K27me3 modifications regulate the cold-responsive gene expression in Oryza sativa L. ssp. indica. Genomics **114** (4), 110433 (2022).

34. L. P. de Lima Camillo, M. H. Asif, S. Horvath, E. Larschan, R. Singh, Histone mark age of human tissues and cell types. Science Advances **11** (1), eadk9373 (2025).

35. E. Mjolsness, D. H. Sharp, J. Reinitz, A connectionist model of development. Journal of theoretical Biology **152** (4), 429–453 (1991).

36. N. E. Buchler, U. Gerland, T. Hwa, On schemes of combinatorial transcription logic. Proceedings of the National Academy of Sciences **100** (9), 5136–5141 (2003).

37. S. Memczak, J. C. Izpisua Belmonte, T. Graepel, Escaping ageing through Cell Annealing—a phenomenological model. Cell Research pp. 1–4 (2025).

38. C. Chalk, S. Buse, K. Shrinivas, A. Murugan, E. Winfree, Learning and Inference in a Lattice Model of Multicomponent Condensates (2024).

39. W. Poole, T. Ouldridge, M. Gopalkrishnan, E. Winfree, Detailed balanced chemical reaction networks as generalized Boltzmann machines. arXiv preprint arXiv:2205.06313 (2022).

40. W. Poole, T. E. Ouldridge, M. Gopalkrishnan, Autonomous learning of generative models with chemical reaction network ensembles. Journal of the Royal Society Interface **22** (222), 20240373 (2025).

41. Y. Liu, et al., DNA hypomethylation promotes UHRF1-and SUV39H1/H2-dependent crosstalk between H3K18ub and H3K9me3 to reinforce heterochromatin states. Molecular Cell **85** (2), 394–412.e12 (2025), doi:10.1016/j.molcel.2024.11.009.

42. E. Y. Basenko, et al., Genome-wide redistribution of H3K27me3 is linked to genotoxic stress and defective growth. Proceedings of the National Academy of Sciences **112** (46), E6339–E6348 (2015), doi:10.1073/pnas.1511377112.

43. J. P. Reddington, et al., Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. Genome Biology **14** (3), R25 (2013), doi:10.1186/gb-2013-14-3-r25.

44. V. Ramesh, F. Liu, M. S. Minto, U. Chan, A. E. West, Bidirectional regulation of postmitotic H3K27me3 distributions underlie cerebellar granule neuron maturation dynamics. eLife **12**, e86273 (2023), doi:10.7554/elife.86273.

45. J. Wang, P. Alexander, S. McKnight, Metabolic specialization of mouse embryonic stem cells, in Cold Spring Harbor symposia on quantitative biology (Cold Spring Harbor Laboratory Press), vol. 76 (2011), pp. 183–193.

46. S. Bertels, et al., Geometrically defined environments direct cell division rate and subcellular YAP localization in single mouse embryonic stem cells. Scientific Reports **11** (1), 9269 (2021).

47. J. M. Stafford, et al., Multiple modes of PRC2 inhibition elicit global chromatin alterations in H3K27M pediatric glioma. Science Advances **4** (10), eaau5935 (2018), doi:10.1126/sciadv.aau5935.

48. U. Jadhav, et al., Replicational Dilution of H3K27me3 in Mammalian Cells and the Role of Poised Promoters. Molecular Cell **78** (1), 141–151.e5 (2020), doi:10.1016/j.molcel.2020.01.017.

49. X. Liu, et al., Yamanaka factors critically regulate the developmental signaling network in mouse embryonic stem cells. Cell Research **18** (12), 1177–1189 (2008), doi:10.1038/cr.2008.309.

50. J. Huang, et al., More synergetic cooperation of Yamanaka factors in induced pluripotent stem cells than in embryonic stem cells. Cell research **19** (10), 1127–1138 (2009).

51. K. Takahashi, S. Yamanaka, A developmental framework for induced pluripotency. Development **142** (19), 3274–3285 (2015).

52. T. Wang, et al., Chemical-induced phase transition and global conformational reorganization of chromatin. Nature Communications **14** (1), 5556 (2023).

53. R. Patrick, et al., The activity of early-life gene regulatory elements is hijacked in aging through pervasive AP-1-linked chromatin opening. Cell metabolism **36** (8), 1858–1881 (2024).

54. T. Zhang, S. Cooper, N. Brockdorff, The interplay of histone modifications – writers that read. EMBO reports **16** (11), 1467–1481 (2015), doi:10.15252/embr.201540945.

55. W. Marshall, et al., Interphase chromosomes undergo constrained diffusional motion in living cells. Current Biology **7** (12), 930–939 (1997), doi:10.1016/s0960-9822(06)00412-x.

56. J. R. Dixon, et al., Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature **485** (7398), 376–380 (2012), doi:10.1038/nature11082.

57. K. Zhang, K. Mosch, W. Fischle, S. I. S. Grewal, Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. Nature Structural & Molecular Biology **15** (4), 381–388 (2008), doi:10.1038/nsmb.1406.

58. M. Tyagi, N. Imam, K. Verma, A. K. Patel, Chromatin remodelers: We are the drivers!! Nucleus **7** (4), 388–404 (2016), doi:10.1080/19491034.2016.1211217.

59. D. Li, X. Shu, P. Zhu, D. Pei, Chromatin accessibility dynamics during cell fate reprogramming. The EMBO Reports **22** (2), EMBR202051644 (2021), doi:10.15252/embr.202051644.

60. D. F. Anderson, T. G. Kurtz, Continuous time Markov chain models for chemical reaction networks, in Design and analysis of biomolecular circuits: engineering approaches to systems and synthetic biology (Springer), pp. 3–42 (2011).

61. Y. Cao, D. T. Gillespie, L. R. Petzold, Efficient step size selection for the tau-leaping simulation method. The Journal of chemical physics **124** (4) (2006).

62. Y. Zheng, J. D. Tipton, P. M. Thomas, N. L. Kelleher, S. M. M. Sweet, Site-specific human histone H3 methylation stability: fast K4me3 turnover. PROTEOMICS **14** (19), 2190–2199 (2014), doi:10.1002/pmic.201400060.

63. B. Al-Sady, H. D. Madhani, G. J. Narlikar, Division of Labor between the Chromodomains of HP1 and Suv39 Methylase Enables Coordination of Heterochromatin Spread. Molecular Cell **51** (1), 80–91 (2013), doi:10.1016/j.molcel.2013.06.013.

64. N. Kepper, D. Foethke, R. Stehr, G. Wedemann, K. Rippe, Nucleosome Geometry and Internucleosomal Interactions Control the Chromatin Fiber Conformation. Biophysical Journal **95** (8), 3692–3705 (2008), doi:10.1529/biophysj.107.121079.

65. C. Saab, J. Stephan, E. Akoury, Structural insights into the binding mechanism of Clr4 methyltransferase to H3K9 methylated nucleosome. Scientific Reports **14** (1), 5438 (2024).

66. M. Doi, S. F. Edwards, S. F. Edwards, The theory of polymer dynamics, vol. 73 (oxford university press) (1988).

67. N. Goldenfeld, Lectures on phase transitions and the renormalization group (CRC Press) (2018).

68. A. Nordin, G. Zambanini, P. Pagella, C. Cantù, The CUT&RUN suspect list of problematic regions of the genome. Genome Biology **24** (1), 185 (2023).

69. D. T. Gillespie, Approximate accelerated stochastic simulation of chemically reacting systems. The Journal of chemical physics **115** (4), 1716–1733 (2001).

70. P. A. Ewels, et al., The nf-core framework for community-curated bioinformatics pipelines. Nature biotechnology **38** (3), 276–278 (2020).

71. P. Di Tommaso, et al., Nextflow enables reproducible computational workflows. Nature biotechnology **35** (4), 316–319 (2017).

72. F. Simon Andrews, A. Segonds-Pichon, L. Biggins, C. Krueger, S. Wingett, FastQC: a quality control tool for high throughput sequence data (2010), `http://www.bioinformatics.babraham.ac.uk/projects/fastqc/`.

73. F. Krueger, Trim Galore!: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. Babraham Institute (2015), `https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/`.

74. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal **17** (1), 10–12 (2011).

75. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. Nature methods **9** (4), 357–359 (2012).

76. H. Li, et al., The sequence alignment/map format and SAMtools. bioinformatics **25** (16), 2078–2079 (2009).

77. P. Toolkit, Broad Institute; 2019. Broad Institute, GitHub Repository. Available online: https://broadinstitute. github. io/picard/(accessed on 28 April 2022) .

78. T. Daley, A. D. Smith, Predicting the molecular complexity of sequencing libraries. Nature methods **10** (4), 325–327 (2013).

79. F. Ramírez, et al., deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic acids research **44** (Web Server issue), W160 (2016).

80. W. J. Kent, et al., The human genome browser at UCSC. Genome research **12** (6), 996–1006 (2002).

Supplemental Information: A Physics-based Model Reveals Mechanisms of Epigenetic Memory and Reprogramming

Erik J. Navarro[†1], William Poole[†1], Ariane Lismer[2], Jieqiong Qu[2], Ryan Spangler[1], Francisco M. Martín-Zamora[3], Ioannis Kafetzopoulos[2], Aled Parry[2], Jannat Ijaz[2], Alice Santambrogio[2], Ken Raj[2], Wolf Reik[2], Hana El-Samad[1], Carlos F. Lopez[3], Simone Bianco [*1]1Altos Labs - Institute of Computation[2]Altos Labs - Institute of Science[3]Altos Labs - Institute of Technology[*]Corresponding author. Email: sbianco@altoslabs.com[†]These authors contributed equally to this work.

A Idealized LEAP Model

Let us consider a simplified epigenetic code consisting of a single epigenetic mark which resides along the monomers of a polymer chain representing chromatin. This polymer has N monomers (nucleosomes containing a modifiable histone), each denoted by $x_i \in \{0, 1\}$, which indicates whether the histone i is unmodified ($x_i = 0$) or modified ($x_i = 1$). The ordering of these monomers in a linear chain $X = (x_1, ..., x_i, ..., x_N)$ constitutes a specific epigenetic pattern.

The process of epigenetic modification occurs through the action of enzymes capable of both reading and writing biochemical modifications on histones (*54*). A common property of reader/writer enzymes is that they "spread" modifications from modified to unmodified histones that are co-localized (*54*). For this to happen an unmodified nucleosome has to come into contact with a modified nucleosome, and a reader/writer enzyme has to be present to catalyze the reaction. We postulate that epigenetic pattern restoration is rooted in two physical properties: 1) an inverse power law decay with separation distance in chromatin contact probabilities; And, 2) the ability for epigenetic reader/writer enzymes to co-localize with their own reaction product. We now turn our attention to these two requirements.

Contact Probability

We first seek the probability that an unmodified nucleosome comes into contact with a modified nucleosome. Chromatin is at thermal equilibrium with the rest of the cell which gives rise to Brownian motion (*55*). This diffusion driven motion of the polymer chain allows the polymer to explore its conformational space, effectively sampling the probability distribution that one

chromatin locus comes into contact with another chromatin locus. The entropic cost of bringing together two monomers on a polymer is a logarithmic function of the number of monomers in the loop. By combining the Gibbs free energy change associated with this change in entropy with the Boltzmann distribution, a scaling relationship between the probability of loop formation and the length of a loop is obtained. It is well established that polymer contact probabilities decay with separation distance as an inverse power law (*25*):

$$P_L(i, j) = k_c \frac{1}{d_{ij}^{\alpha}}. \tag{1}$$

Here, $P_L(i, j)$ is the contact probability of two polymer loci, i and j, $d_{ij} = |i - j|$ is the linear separation distance between the loci, k_c is the contact probability rate related to the diffusivity of the chain and and α is a scaling exponent. Experiments measuring chromatin contact probabilities genome-wide have confirmed this behavior for a wide range of length scales (*22, 56*).

Enzyme Localization

Next, we seek the probability that an enzyme is present at locus j to catalyze the modification of locus i. Experimental observations (*20, 57*) have shown that enzyme binding is preferential for already modified histones. First, we compute the concentration of bound enzymes as a function of the total available enzymes $E = E_f + E_b$ and total monomers $N = N_M + N_U$ where E_f, E_b, N_M and N_U denote the free enzymes, bound enzymes, number of modified monomers and number of unmodified monomers, respectively. In the limit of large N_M and large E, we model the binding of an enzyme to any modified monomer as an ordinary differential equation:

$$\frac{\mathrm{d}E_b}{\mathrm{d}t} = k_b E_f N_M - k_u E_b. \tag{2}$$

Here, k_b is the binding rate of a free enzyme to a modified monomer and k_u is the unbinding rate. We next solve for the steady state solution assuming that binding and unbinding reactions occur quickly:

$$\frac{\mathrm{d}E_b}{\mathrm{d}t} = 0 \quad \Longrightarrow \quad E_b^* = \frac{1}{2}\left((N_M + E + K_D) - \sqrt{(N_M + E + K_D)^2 - 4EN}\right). \tag{3}$$

Here, E_b^* is the steady state concentration of bound enzymes and $K_D = \frac{k_u}{k_b}$. Additionally, we note that the binding curve between the catalyzing enzyme and modified histones has a sigmoidal

shape suggestive of cooperative binding (*57*). We conjecture that this effect is due to preferential accumulation of enzymes at densely modified regions in the 3D polymer chain conformation, which we model phenomenological with a Hill function of the local modified monomer density:

$$\sigma(j) = \frac{\rho(j)^n}{K_A^n + \rho(j)^n} + b(j,t). \tag{4}$$

Here $\rho(j)$ is the local density of modified monomer around j, n is a Hill coefficient representing the non-linear nature of enzyme localization, and K_A is the histone methylation density that produces half max of the enzyme count. The term $b(j,t) > 0$ is a position and time dependent bias term which can be used to model sequence dependence or introduce external perturbations; however, unless otherwise noticed assume $b(j,t) = 0$. We chose to make this term additive (as opposed to multiplicative) because there is substantial evidence that epigenetic remodeling factors can act on both methylated and unmethylated regions (*58*). For example, Yamanaka factor signalling networks are known to target euchromatin and hetereochromatin (*59*), while H3K9 and H3K27 methylation occurs only in hetereochromatin. Therefore, we allow a bias which is independent of the methylation density. Combining equations (3) and (4) we arrive at an expression for the probability of an enzyme at monomer j:

$$P_E(j) = \frac{\sigma(j)}{Z} E_b^* \qquad Z = \sum_j \sigma(j). \tag{5}$$

Note that normalizing by Z is required to ensure the correct total number of bound enzymes: $\sum_j P_E(j) = E_b^*$.

Continuous Time Markov Chain

Next, we model the process of epigenetic mark restoration. In this process, marks are stochastically added one at a time to unmarked monomers x_i^U after they come into contact with an enzyme bound to marked monomer x_j^M. These microscopic spreading rates are given by:

$$T_{ij} = k_m P_L(i,j) P_E(j). \tag{6}$$

Here, k_m is the enzymatic modification rate. Finally, we combine all the microscopic spreading rates into a continuous time Markov chain (CTMC) (*60*) representing a stochastic process of epigenetic

mark restoration between cell divisions. The transition rates from state X to X' are given by:

$$Q(X \to X') = \begin{cases} \sum_j T_{ij} & X' - X = 1_i \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Here, T_{ij} are the spreading events where an enzyme at j catalyzes the addition of a mark at i, and $X' - X = 1_i$ denotes that the difference between X' and X is the addition of a single modification at monomer i. This transition rate matrix governs the dynamics of a probability distribution of the epigenetic pattern $\mathbb{P}(X)$:

$$\frac{\mathrm{d}\mathbb{P}(X)}{\mathrm{d}t} = \sum_{X'} \mathbb{P}(X')Q(X' \to X) - \mathbb{P}(X)Q(X \to X'). \tag{8}$$

Computational Implementation

We begin each simulation as a sequence X that represents a particular epigenetic pattern. This sequence is used as the initial state in the CTMC represented by equation (7). This CTMC is simulated for a fixed amount of time, akin to cell division time, t_{div} via the τ-leaping method (*61*). The density $\rho(j)$ in equation (4) is approximated as Gaussian kernel density estimate along the polymer chain estimate at monomer j :

$$\rho(j) \approx D_\lambda(j) \tag{9}$$

where D_λ is a one-dimensional Gaussian kernel density estimate with bandwidth λ. This estimate is updated every iteration of the τ-leaping algorithm.

When t_{div} is reached, the cell divides resulting the stochastic partitioning of all epigenetic marks randomly between the two daughter cells. Because only a single cell is tracked in our simulation, this amounts to randomly erasing each epigenetic mark with a probability of 50%. The resulting post division state, X', is then used as the initial state that can be simulated as a CTMC until the next division.

Model Parameterization

We used the following parameters for our simulations which are estimated from literature. The number of monomers, N, is set to 10000 monomers as a representative sequence length. The

number of histone methyltransferases, E vary widely by cell type and organism, but are always one or more orders of magnitude less than the number of histones. Therefore, we choose $E = 1000$ enzymes (*47*). The methyltransferase-histone dissociation constant was estimated to be $K_D = 300$ (which is in units of molecular counts) based on (*20*). Measurements for the enzymatic rate constant vary by orders of magnitude, from $1e{-}3$ as measured in (*62*) to 30 based as measured in (*63*) and we used the intermediate value of $k_{meth} = 1.0$ (in units of per day). The division time of a Eukaryotic cell ranges from about a day for rapidly dividing cells to years for slowly dividing cells, but experiments are typically conducted on rapidly dividing cells to shorten the length of experiments. We therefore choose a division time of 1 day. The density bandwidth λ measures the length scale over which nearby nucleosomes compact and should be roughly the persistence length of the nucleosome chain, which is approximately 10 nucleosomes (*64*). Finally, we estimated the Hill function parameters $K_A = 0.00025$ (which is in units of monomer density along the sequence) and $n = 3.3$ from the binding curve for the H3K9me3 methyltransferase SUV39H1 measured in (*65*).

B Spontaneous Pattern Formation

In this section, we investigate some properties the idealized LEAP model and discuss how they give rise to spontaneous pattern formation.

We begin by briefly commenting on some of the mathematical properties of this model. First, this model is trivially ergodic - meaning any state can in theory be reached by any other state - due to the ability to add individual marks at any location and then partition the marks randomly into two arbitrary new states (provided no new marks are created). Second, the dynamics of this system are monomer location invariant (neglecting edge effects of the chain) provided that the bias term is 0. Jointly, these properties suggest that the steady state of a very large N idealized model should be almost uniform.

These observations are not at odds with the fact that the LEAP model maintains patterns. Indeed, steady state dynamics of a CTMC are defined as an infinite time average which does not imply that the short term dynamics of any given trajectory should appear uniform. We can observe aspects of this behavior via numerical experiments. In Figure 5 we show a sampling of idealized

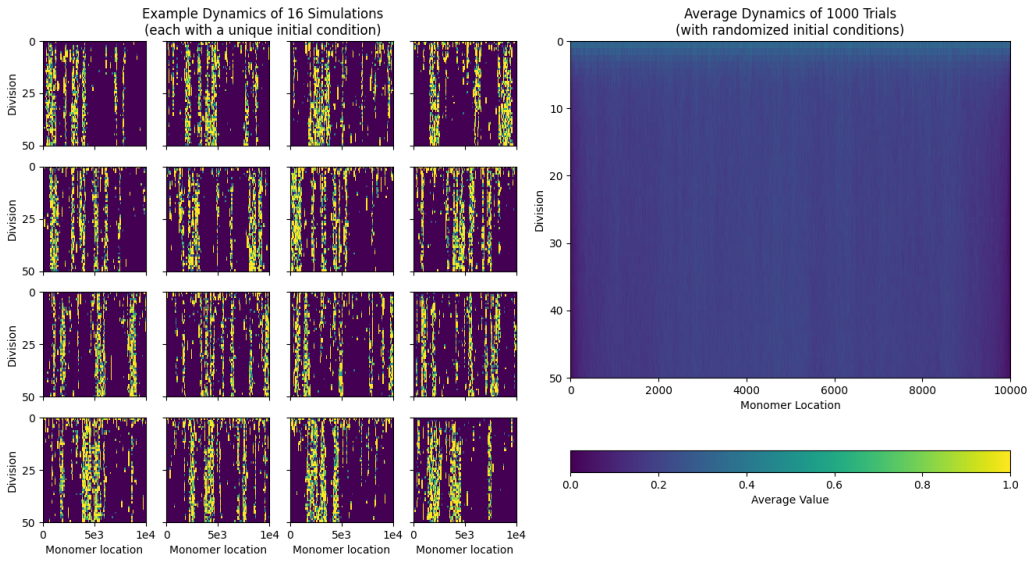

Figure 5: Left. Example simulations with randomized initial conditions spontaneously form banded patterns. Right. Average of 1000 randomized simulations indicates that the steady state distribution is uniform.

model simulations from random initial conditions. Fluctuations in initial condition density coupled with the stochastic nature of the model reliably give rise to persistent banded structures, a form of spontaneous pattern generation. Indeed, a fundamental hypothesis implicit in our model is that nature has evolved to use the spontaneous pattern formation and restoration properties intrinsic in looping, enzymes, and phase condensates as a flexible and modulatable epigenetic memory system. While beyond the scope of this paper, we note that the ability to form epigenetic patterns spontaneously has potential implications for epigenetic evolution.

C Coarse Grained LEAP Model

In this section, we described the coarse grained model used for genome-scale epigenetic simulations. In addition to the coarse-graining approach, this model also involves multiple chromosomes and several assumptions had to be made. First, we (very approximately) assume that the chromosomes do not interact in 3D space, meaning there is no contact probability between different chromosomes. Although this is obviously not accurate, it is a necessary approximation for our simple model to be

applicable. However, we do couple the chromosomes via a shared pool of enzymes. Specifically, we partition the total enzymes E between the different chromosomes in a way proportional to the 3D volume occupied by each chromosome. In the proceeding section, we will derive the coarse grained model from the un-coarse grained *detailed* model. Note that due to having multiple chromosomes, the notation used in this section is slightly different than the rest of the paper.

The detailed state X_c of a chromosome c with N_c monomers each with 0 or 1 marks is denoted by:

$$X_c = (x_{c,1}, ..., x_{c,i}, ..., x_{c,N_c}) \in \{0,1\}^{N_c} \qquad x_{c,i} \in \{0,1\}. \tag{10}$$

We coarse grain the detailed polymer X_c into a shorter polymer Y_c of length $\nu_c = N_c/\mu$ where each monomer may have up to μ marks. In other words, the coarse graining procedure combines μ binary monomers in X_c into a single coarse-grained monomer in Y which may have up to μ marks. Denote the coarse grained polymer:

$$Y_c = (y_{c,1}, ..., y_{c,i}, ..., y_{c,\nu_c}) \in \{0,1,...\mu\}^{\nu_c} \qquad y_{c,i} \in \{0,1,...,\mu\}. \tag{11}$$

The value of each individual coarse grained monomer $y_{c,i}$ is a sum of monomers from the detailed state $x_{c,i}$:

$$y_{c,i}^{\eta_i} = \sum_{l=\mu*i}^{\mu*(i+1)} x_{c,l}. \tag{12}$$

Next, coarse grained chromosomes are coupled via a shared pool of enzymes partitioned proportional to their occupied volume V_c:

$$V_c \sim (R_c^g)^3 \sim (\nu_c^{0.6})^3 \tag{13}$$

Where R_c^g is the radius of gyration of chromosome c and the final scaling law related the radius of gyration to the number of monomers in a self-avoiding polymer chain (*66*). Using the result from 3, the total number of bound enzymes to chromosome c, E_c, is given by:

$$E_c = \frac{1}{2}\left((N_{M,c} + E_c^{tot} + K_D) - \sqrt{(N_M + E_c^{tot} + K_D)^2 - 4E_c^{tot}N}\right) \qquad E_c^{tot} = \frac{V_c}{\sum_{c'} V_c'}E \tag{14}$$

Here, $N_{M,c}$ are the number of marked monomers in a chromosome c and E_c^{tot} are the total number of enzymes available to chromosome c.

We now have the necessary pieces to to write exact transition rates between coarse grained states Y and Y':

$$\Phi(Y \to Y') = \begin{cases} \sum_{l=\mu*i}^{\mu*(i+1)} \sum_{j=1}^{N_c} T_{lj} & Y' - Y = 1_{c,i} \quad \text{and} \quad y'_{c,i} < \mu \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Here, transitions add a single mark to a coarse-grained monomer at a time which has at most μ marks. The term $1_{c,i}$ indicates a that Y and Y' differ by the addition of a single mark at chromosome c monomer i. The double sum first sums over the monomers that are combined and then over the transition rates of those individual monomers as defined in (6). Next, in order to efficiently simulate at the coarse-grained scale, the double sum must be approximated in terms of the coarse-grained polymer:

$$\sum_{l=\mu*i}^{\mu*(i+1)} \sum_{j=1}^{N_c} T_{lj} = \sum_{l=\mu*i}^{\mu*(i+1)} \sum_{j=1}^{N_c} k_m P_L(l,j) P_E(c,j) = \sum_{l=\mu*i}^{\mu*(i+1)} \sum_{j=1}^{N_c} k_c k_m d_{jl}^{-\alpha} P_E(c,j) \tag{16}$$

$$\approx \sum_{j=1}^{\nu_c} \left(k_c k_m \mu \tilde{P}_e(c,j) \sum_{l=\mu*i}^{\mu*(i+1)} d_{jl}^{-\alpha} \right) = \sum_{j=1}^{\nu_c} \tilde{k} \tilde{P}_e(c,j) \tilde{d}(j,i,\mu,\alpha) = \sum_{j=1}^{\nu_c} \tilde{T}_{ij}. \tag{17}$$

Where

$$\tilde{T}_{ij} = \tilde{k} \tilde{P}_e(j) \tilde{d}(j,i,\mu,\alpha), \qquad \tilde{k} = k_c k_m \mu, \quad \text{and} \quad \tilde{d}(j,i,\mu,\alpha) = \sum_{l=\mu*i}^{\mu*(i+1)} d_{jl}^{-\alpha}. \tag{18}$$

Here, we have rewritten the sum over all monomers from the detailed state into a sum over coarse grained monomers. Two approximations were made to get this result. First, the probability of enzyme localization $P_E(c,j)$ of the detailed model is approximated by the probability of enzyme localization of the coarse-grained monomer, $\tilde{P}_e(c,j)$. Inspired by the renormalization group theory of the Ising model (*67*), this function is assumed to be constant across the set of monomers in the sum and have the same functional form as P_E:

$$\tilde{P}_e(c,j) = \frac{\sigma(j)}{Z} E_c \qquad Z = \sum_j \sigma(j) \qquad \sigma(j) = \frac{\rho(j)^{\tilde{n}}}{\tilde{K}_A^{\tilde{n}} + \rho(j)^{\tilde{n}}} + \tilde{b}(j,t). \tag{19}$$

Here, $\rho(j)$ is the marked monomer density at location j. Additionally, the sigmoid parameters $\tilde{K}_A$, and $\tilde{n}$ must be derived from data and are not necessarily the same as the detailed model's parameters. Finally, we have let $\tilde{b}(j,t) = 0$ for simplicity. Second, the inverse power law is replaced by a surrogate function which adds up the contributions of the power law over each coarse-grained

region of the polymer. Importantly, due to being density-independent, $\tilde{d}$ can be pre-computed as a look-up table for computational efficiency. This allows us to write an approximate coarse-grained transition rate function $\tilde{\Phi}(Y \to Y')$:

$$\tilde{\Phi}(Y \to Y') = \begin{cases} \sum_{j=1}^{\nu_c} \tilde{T}_{ij} & Y' - Y = 1_{c,i} \quad \text{and} \quad y'_{c,i} < \mu \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

The CTMC for the coarse grained model will then follow the following master equation:

$$\frac{\mathrm{d}\mathbb{P}(Y)}{\mathrm{d}t} = \sum_{Y'} \mathbb{P}(Y')\tilde{\Phi}(Y' \to Y) - \mathbb{P}(Y)\tilde{\Phi}(Y \to Y'). \tag{21}$$

D Full Genome Epigenetic Simulations

D.1 Full Genome Simulation Results

HDF full genome simulations were carried out by coarse-graining the HDF data (H3K9me3 and H3K27me3 CUT&RUN from 8 donors) and then simulating the entire genome of each donor for 200 generations after initializing the model to the coarse-grained data from that donor. Ten replicates were carried out per donor, resulting in a total of 160 full genome simulations. Large scale patterns (on the order of megabases) were faithfully maintained over the entire duration of the simulation. The average output from the 10 replicates of each full genome simulation and the coarse grained data is shown in Figures 6 and 7.

HUVEC full genome simulations were carried out analogously to the HDF simulations by initializing the the LEAP model to the young passaged HUVEC H3K27me3 data and simulating 25 replicates for 230 generations. The simulated epigenetic dynamics were then compared coarse grained data from the medium and old passaged HUVECs. Although we see greater epigenetic drift in the model than in the data, the model is still able to accurately predict maintenance of large-scale epigenetic domains.

D.2 Coarse Graining Data

In order to compare the coarse grained model to real epigenetic data, it is necessary to coarse-grain the bulk epigenetic data in an analogous way to the model. This procedure is also used to generate

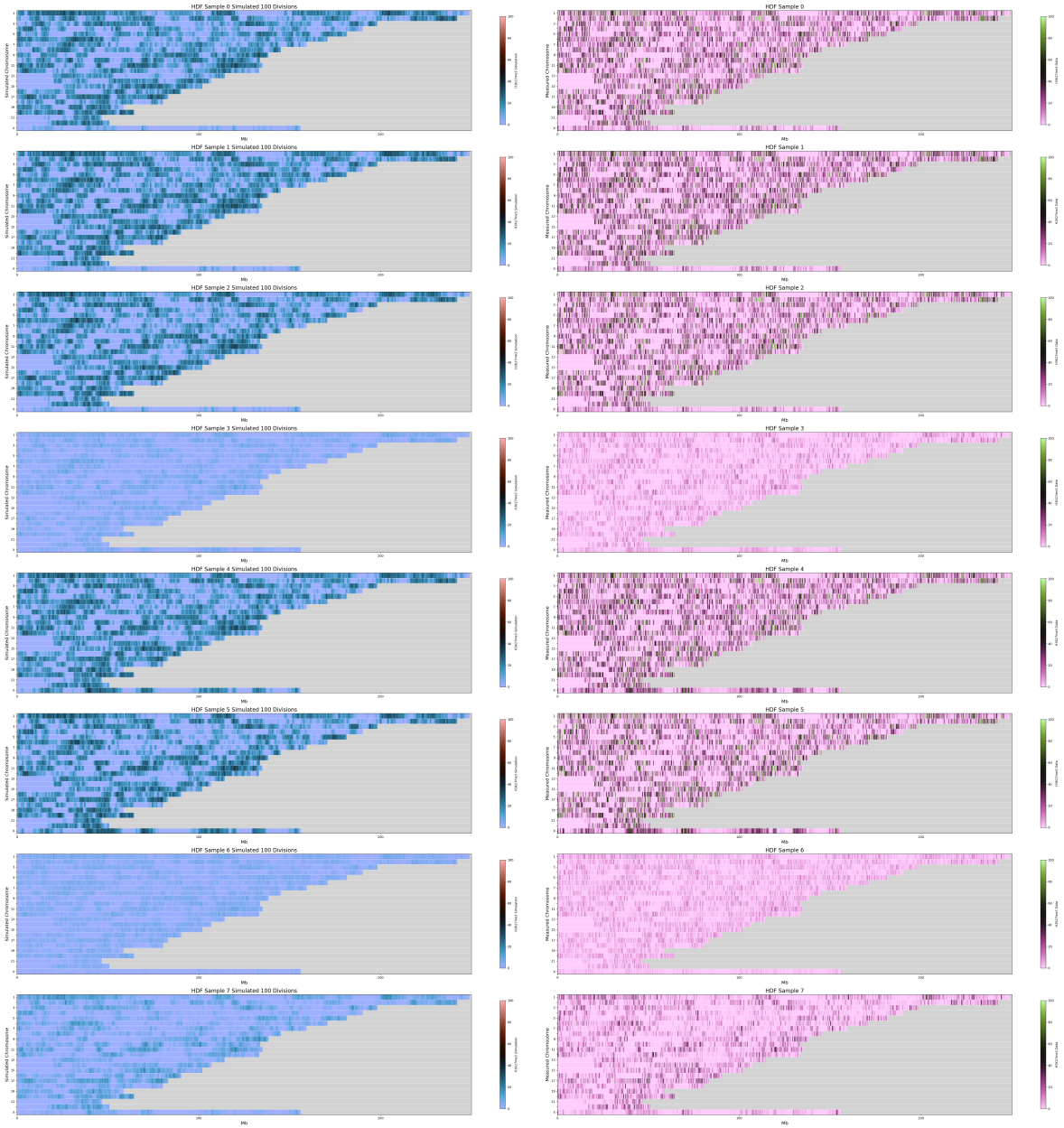

Figure 6: HDF H3K27me3 coarse simulation (blue heatmaps) results after 100 divisions and coarse-grained data (pink heatmaps) by donor and chromosome. Not smoothed.

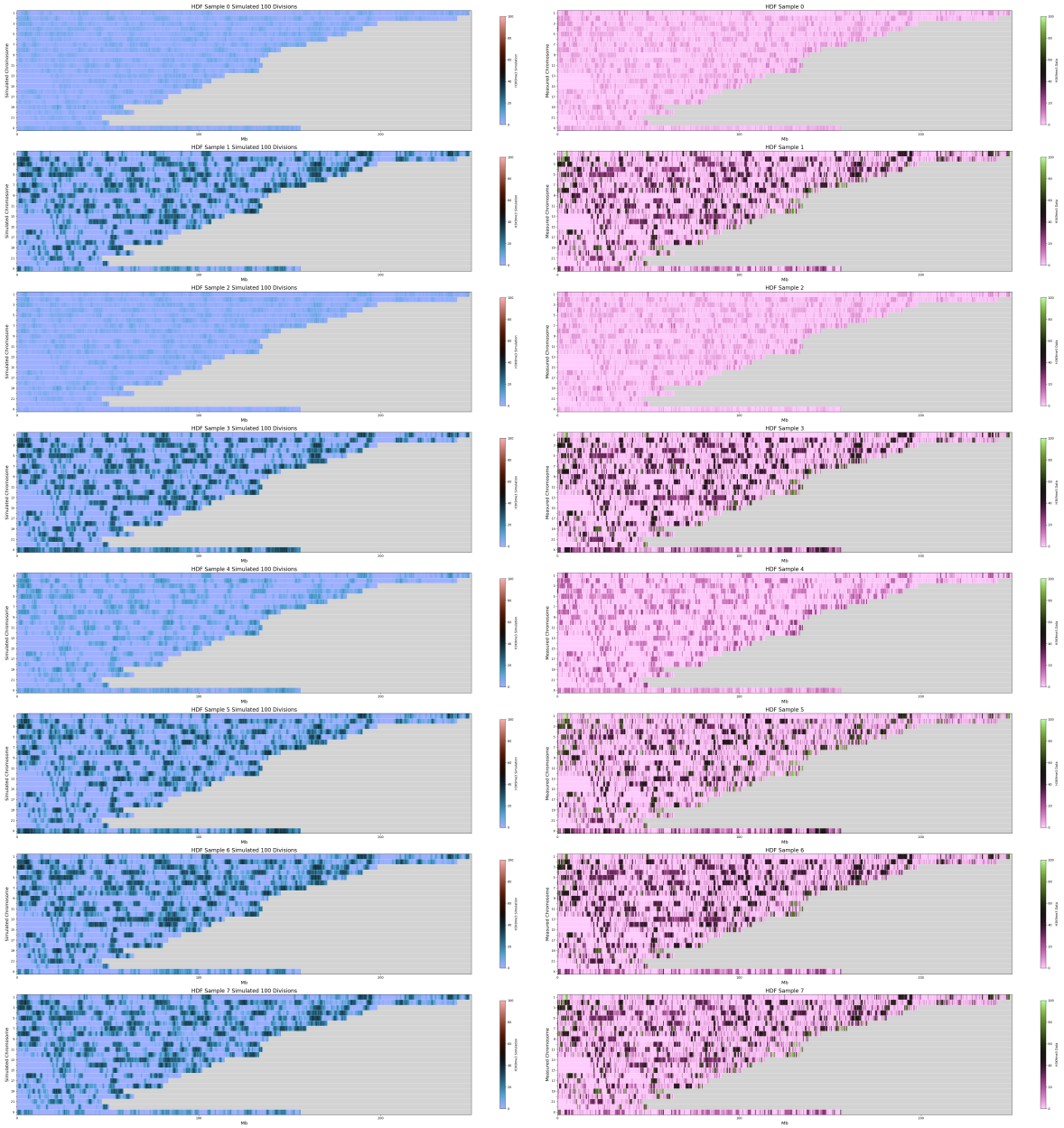

Figure 7: HDF H3K9me3 coarse simulation results (blue heatmaps) after 100 divisions and coarse-grained data (pink heatmaps) by donor and chromosome. Not smoothed.

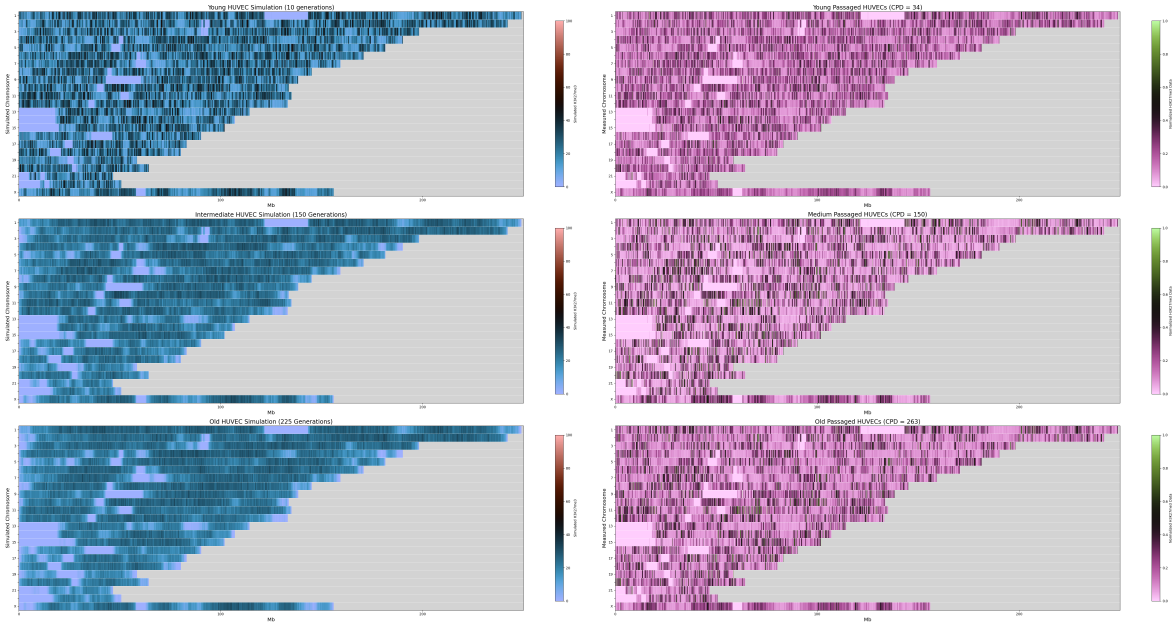

Figure 8: HUVEC H3K27me3 coarse grained data and simulation results at different passages and divisions. Not smoothed.

initial conditions for the coarse-grained model from data. Briefly, raw reads from the Cut&Run experiment are aligned to a reference genome and converted to counts per nucleotide as described in SI G.4. This data is effectively a noisy and unit-less quantity proportional to the number of marked histones at any given genomic location across a large population of cells. First, we bin and compress the raw underlying data Z to compute a one-normalized local density $\overline{\rho}$:

$$\overline{\rho}(i) = \min\left[\frac{1}{b(z_{max} - z_{min})}\sum_{j=b}^{b(i+1)} Z(j) - z_{min}, 1\right] \tag{22}$$

Here, b is the bin-size, chosen to match the resolution of the coarse grained model, and z_{max} and z_{min} are empirically determined parameters. We found that $z_{max} = 4$ and $z_{min} = 0$ gave reasonable results for our data set. In particular, $z_{max} = 4$ was chosen to compress the extreme values found in blacklist regions (*68*) as we cannot discard these regions of each chromosome when running a mechanistic simulations.

D.3 Generating Initial Conditions from Data

We generate initial conditions $Y^0 = (y_1^0, ...y_\nu^0)$ by sampling a binomial distribution $y_i^0 \sim B(n, p_i)$. Here, $n = \mu$ are the maximum number of marks per coarse-grained monomer and $p_i = \overline{\rho}(i)$ is the probability any particular site in the coarse-grained monomer being occupied.

D.4 Coarse CTMC Simulation Details

We simulate the coarse grained CTMC using the τ-leaping algorithm (*69*) with E_c and $\tilde{T}$ recomputed at each τ-step. Analogously to the idealized model, marks are partitioned at regular cell divisions based upon the binomial distribution.

All simulations were run using custom python code on 48-CPU AWS nodes running Linux with 384 GIG of RAM. Simulations were run sequentially on any given node but were accelerated due to the automatic parallelization of Numpy array operations.

We were able to run hundreds of full genome simulations using just a 4 nodes in a little over a week. For each of 8 donors, 10 replicates of H3K9me3 and H3K27me3 epigenetic models were simulated for 200 cell divisions starting at initial conditions derived from CUT&RUN data. Additionally, 20 replicates of HUVEC H3K27me3 simulations were conducted for 230 divisions with initial conditions derived from the young (low passage) HUVEC samples.

D.5 Coarse CTMC Simulation Parameterization

The parameters used for the coarse-grained simulations are as follows. The bin size was set to $10,000$ base-pairs, which amounts to coarse graining the model by a factor of 50 nucleosomes (200bp per nucleosome). Min_val and max_val were set to 0.0 and 4.0 respectively, which we found reduced the incidence of problematic peaks in the signal, often corresponding to blacklisted CUT and RUN regions (*68*). K_A sigmoid was set to $3.5e-7$ (in units of methylation density per coarse-grained monomer), n in the sigmoid was set to 3.3 as in the idealized model. λ, the bandwidth used in the density estimate was set to 10, which in the coarse grained model amounts to 100kb in genomic space.The methyltransferase-histone dissociation constant K_D was set to 200 molecules. The division time t_d was set to 1.667 days which is the average time between division for our HUVEC cells. Our timestep was set to .667 days, which amounts to 10 timesteps per

division and was chosen to both maintain accuracy and minimize computational requirements. α, the inverse power law decay exponent was set to 1.5 as in the idealized model and falls within experimentally measured values of α (*22,56*). The total number of enzymes across all chromosomes set to $E = N * (bin - size)/((base - pairs - per - mark) * 100))$ which amounts to choosing E to be 100 times less than the number of histone substrates. Lastly we use a methylation rate k_meth of 0.001657 per day for the HDF cell simulations, and a value of .001653 per day for the HUVEC cell simulations.

Importantly, we note that by measuring histone marks along the genome using CUT&RUN, we don't necessarily know the true amount of the mark at each genomic location, but rather the number of reads that were mapped to each genomic location. It is thus likely that those parameters in our model that interact to give rise to a total methylation are somewhat disconnected from the true biological values. We hypothesize that this is what allowed us to use almost identical parameter sets in both cell types, both types of marks, and for all donors.

E Correlations and Autocorrelation Analyses

E.1 Computing Correlations

Correlations between a simulation state and data were computed using Scipy's to compute of Pearson's correlation between a simulated epigenetic state and the one-normalized density $\overline{\rho}(i)$.

Correlations between smoothed states were computed using a Pearson correlation on smoothed epigenetic states and smoothed one-normalized densities. Smoothing was carried out with a 1-dimensional uniform smoothing using window sizes of 100 unless otherwise noted.

Note that we excluded chromosome Y from these analyses because in the majority of simulations it lost all methylation and therefore we cannot compute correlations.

E.2 Computing Auto-correlations

Auto-correlation functions of epigenetic simulations were computed numerically by computing the Pearson correlation between the initial state and subsequent states at each time point t averaged together across multiple replicates (in the case of HUVECs) and replicates and donors (in the case

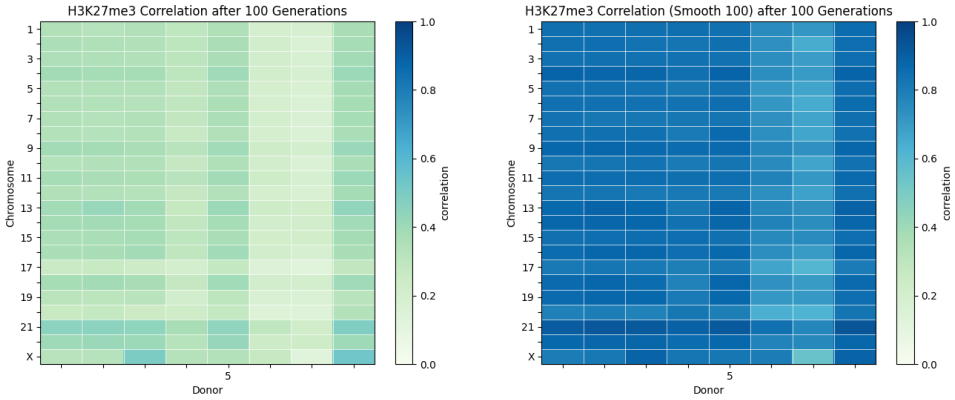

Figure 9: Correlations between coarse-grained data and simulation results (after 100 divisions) by donor and chromosome for smoothed and un-smoothed H3K27me3.

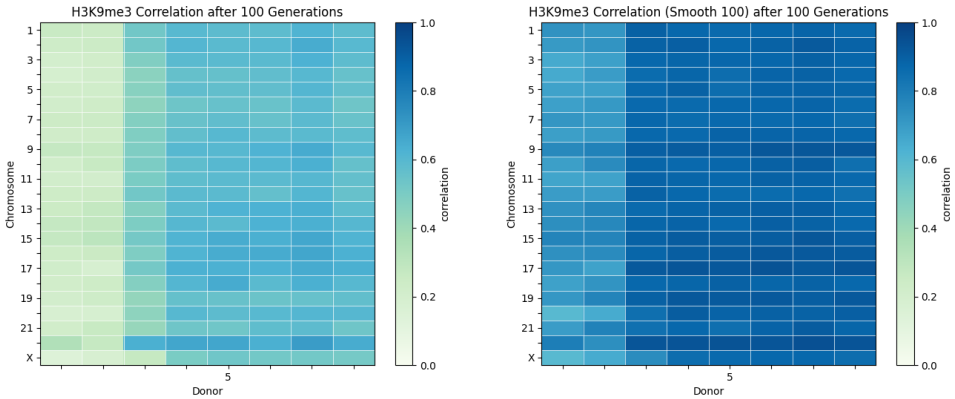

Figure 10: Correlations between coarse-grained data and simulation results (after 100 divisions) by donor and chromosome for smoothed and un-smoothed H3K9me3.

of HDFs).

Auto-correlation functions of smoothed epigenetic simulations were computed numerically by computing the Pearson correlation between the smoothed initial state and subsequent smoothed states at each time point t averaged together across multiple replicates (in the case of HUVECs) and replicates and donors (in the case of HDFs). Smoothing was carried out with a 1-dimensional uniform smoothing using window sizes of 100 unless otherwise noted.

F Conservation of Epigenetic Marks Analysis

First we comment that the CTMC described in equation (8) does not explicitly include mark degradation; this was included as regular binomial partitioning in our simulations to better mimic cell division. However, our implementation of binomial partitioning would be complex to consider in full CTMC formalism. Thus, we first simplify the model and allow marks to degrade at a rate $\delta = \frac{\log 2}{T_{div}}$ where T_{div} is the time between cell divisions. This value of δ ensures that in T_{div} simulation, each mark has a fifty percent chance of being degraded, consistent with our previous model. Second, instead of considering the full distribution of epigenetic marks, we consider a simplified bulk model. In this model, the total number of marks are partitioned into marked and unmarked continuous variables which are governed from the following ODE:

$$\frac{\mathrm{d}N_M}{\mathrm{d}t} = k_{eff} N_M N_U - \delta N_M \quad \text{subject to the constraint} \quad N = N_M + N_U \tag{23}$$

where N is the polymer length, N_M is the amount of marked monomers, N_U is the amount of unmarked monomers, and k_{eff} is an effective methylation rate. Solving for steady state results in:

$$\frac{\mathrm{d}N_M}{\mathrm{d}t} = 0 \quad \Longrightarrow \quad N_M = N \frac{k_{eff}}{k_{eff} + \delta}. \tag{24}$$

The effective methylation rate will be state dependent. Indeed, for a specific state X, it can be written exactly by summing over all possible methylation events:

$$k_{eff}(X) = \sum_{X'} Q(X \to X') = \sum_{i,j} k_m P_L(i,j) P_E(j). \tag{25}$$

Both P_E and P_L both depend on many details of the state X so we must approximate them. The contact probability $P_c(i,j)$ depends on the specific distance between the monomers i and j. We

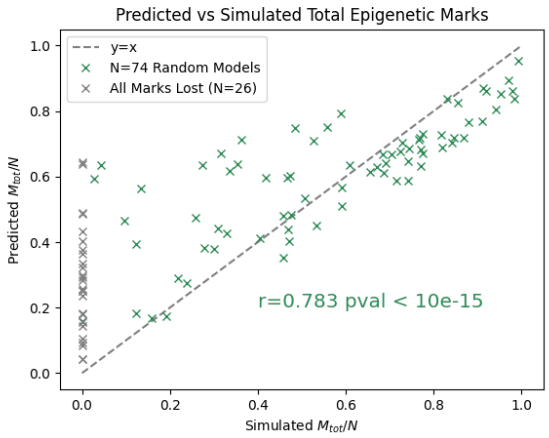

Figure 11: Steady state total epigenetic marks for randomly parameterized models compared to analytic prediction.

approximate this by assuming i is in the middle of the polymer and averaging over possible values of j:

$$\zeta = \frac{2}{N}\sum_{i=1}^{N/2}\frac{1}{i^{\alpha}} \qquad P_L \approx k_c\zeta. \tag{26}$$

The bound enzyme probability depends on the total number of marked monomers. However, the total number of bound enzymes can never exceed E. Therefore, we bound $\sum_j P_E(j) < E$. Combining these approximation we arrive at:

$$k_{eff} \approx k_m k_c \zeta E \qquad \Longrightarrow \qquad N_M \approx N\frac{k_m k_c \zeta E}{k_m k_c \zeta E + \delta} \tag{27}$$

G Experimental Methods

G.1 Cell preparation and culture

Human Dermal Fibroblast lines were purchased from the following vendors: 9F (Corriell, Catalogue No. GM00038), 12M (Corriel, Catalogue No. AG16409), 21F (Corriel, Catalogue No. AG09309), 22Ma (Corriel, Catalogue No. GM05294), 22Mb (Corriel, Catalogue No. GM23815), 28Fa (Cell Applications, Catalogue No. 106-05a, Lot No. 2146), 28Fb (Cell Applications, Catalogue No. 106-05a, Lot No. 2049), 29M (Corriel, Catalogue No. AG04054). Cells were maintained in a 37°C

hypoxic incubator (5% CO2, 5% O2) using DMEM/F-12 GlutaMAX media (Gibco, Catalogue No. 31331093) supplemented with 10% heat inactivated Fetal Bovine Serum (Thermo Fisher, Catalogue No. 17593595), 1X MEM-NEAA (Gibco, Catalogue No. 11140035), 50 µM 2-Mercaptoethanol (Gibco, Catalogue No. 31350010), and 16 ng/ mL FGF-2 (Gibco, Catalogue No. PHG0360).

For nuclei isolation, cells were seeded and harvested two days after. Briefly, media was removed, cells were washed once with 1X DPBS (Gibco, Catalogue No. 14190169), and trypsinised using TrypLE Express (Gibco, Catalogue No. 12605010). After cells were detached, media was added to quench the TrypLE Express solution. Cells were transferred to a 15 mL Eppendorf tube and washed once with 1X DPBS. Nuclei were extracted using freshly prepared Nuclei Extraction Buffer (20 mM HEPES pH 7.9, 10 mM KCl, 0.1% Triton 10X, 20% Glycerol, 1 mM MnCl2, 0.5 mM Spermidine, 1X Protease Inhibitor Cocktail) and CUTANA ChIC/ Cut&Run manufacturer recommendations (Epicypher; Catalogue No. 14-1048). 600,000 cells per Cut&Run experiment were slow-frozen in an Eppendorf DNA LoBind tube (Eppendorf, Catalogue No. 0030108051) and stored in a -80°C freezer.

Human umbilical vein endothelial cells (HUVECs; Cell Applications) were cultured in Endothelial Cell Growth Medium (Cell Applications, 211–500) at 37 °C in a humidified incubator with 5% CO. Culture vessels were pre-coated with 0.001% fibronectin (Sigma-Aldrich, F0895) in HBSS (Gibco, 14175-053) for at least 1 h prior to seeding. The medium was replaced every other day. Cells were subcultured at 70–90% confluency using Trypsin-EDTA (Gibco, 15400054), neutralized with soybean trypsin inhibitor (Invitrogen, 17075-029), and reseeded at 5 × 10 cells per T75 flask. Continued passaging induces epigenetic aging, consistent with prior reports linking DNA methylation changes to cellular aging (Add reference: The relationship between epigenetic age and the hallmarks of aging in human cells). For CutRun assays, HUVECs were harvested two days post-seeding, pelleted by centrifugation at 1200 rpm for 5 min, at room temperature, and resuspended in cold freezing medium consisting of 10% DMSO (Sigma-Aldrich, D2650) and 90% FBS (Gibco, 10500064). Samples were cryopreserved using a controlled-rate freezing container at 80 °C for 24 h before transfer to liquid nitrogen storage. 250,000 cryo-preserved HUVEC cells were used for each CutRun reaction.

G.2 Cut&Run

Cut&Run was performed using the CUTANA ChIC/ Cut&Run Kit (Epicypher; Catalogue No. 14-1048) and manufacturer recommendations. Briefly, Wash Buffer, Cell Permeabilization Buffer, and Antibody Buffer, were made fresh on the day of the experiment and kept at 4°C. Concavalin A (ConA) Beads were washed twice with cold Bead Activation Buffer. After the last wash, activated ConA beads were resuspended in cold Bead Activation Buffer, distributed to 8-strip tubes, and placed on ice.

Cryopreserved HUVECs were thawed and washed twice with HBSS, followed by centrifugation at 600 × g for 3 min at room temperature. The cell pellet was then washed with room-temperature Wash Buffer and resuspended in Wash Buffer, before adding to ConA beads.

Human Dermal Fibroblast nuclei in Nuclei Extraction Buffer were thawed on ice and spun down. Supernatant was removed, nuclei were resuspended in freshly prepared Nuclei Extraction Buffer, and added to 8-strip tubes containing ConA bead mixture. Samples were incubated for 10 minutes at room temperature. Tubes were placed on magnetic rack to allow slurry to clear, and supernatant was discarded. Nuclei bound to ConA beads were resuspended in cold Antibody Buffer and 0.5 µg of either H3K9me3 (Active Motif, Catalogue No. 39161) or H3K27me3 (EpiCypher, Catalogue No. 13-0055) antibody was added. Samples were gently vortexed and spun down before being placed on a nutator at 4°C overnight.

The next day, samples were washed twice with cold Cell Permeabilization Buffer on a magnetic rack. Tubes were taken off the magnetic rack and samples were resuspended in cold Cell Permeabilization Buffer. Subsequently, pAG-MNase was added to the mix and tubes were gently vortexed. Samples were then incubated at room temperature for 10 minutes and washed another two times with cold Cell Permeabilization Buffer. Samples were resuspended in cold Cell Permeabilization Buffer and calcium chloride was added to a final concentration of 2 mM. Tubes were incubated on a nutator for 2 hours at 4 ° C. After, Stop Buffer was added and samples were incubated at 37°C for 10 minutes. Tubes were placed on a magnetic rack, and supernatants of Cut&Run enriched DNA were transferred to new 8-strip tubes. SPRIselect beads were added at a concentration of 1.4X and incubated at room temperature for 5 minutes. Tubes were placed on magnetic rack and supernatant was removed. Two 85% EtOH washes were performed and SPRIselect beads were air-dried for 2

minutes at room temperature. 0.1X TE Buffer was used to elute the DNA and Cut&Run purified DNA transferred to new 8-strip tubes. Samples were quantified using the High Sensitivity Qubit dsDNA Quantification Kit (Thermo Fisher, Catalogue No. Q32851) before being stored in a -20°C freezer.

G.3 Library Preparation

Cut&Run library preparation was performed using the CUTANA Cut&Run Library Prep Kit (Epicypher; Catalogue No. 14-1001 and 14-1002) and manufacturer recommendations. In brief, samples and kit reagents were thawed on ice. 10 ng of Cut&Run enriched DNA was transferred to new 8-strip tubes and volume was adjusted to 25 µL with 0.1X TE Buffer. Samples were End Repaired, and Adaptors for Illumina were ligated followed by U-Excision digestion. SPRIselect beads were added at a 1X concentration and incubated at room temperature for 5 minutes. Tubes were placed on magnetic rack and supernatant was removed. Two 85% EtOH washes were performed and SPRIselect beads were air-dried for 2 minutes at room temperature. 0.1X TE Buffer was used to elute the DNA. Indexing PCR was performed by adding unique i7 and i5 Primer combinations to samples followed by Hot Start 2X PCR Master Mix. 14 PCR cycles were used for Human Dermal Fibroblast and 11 PCR cycles were used for HUVEC samples to amplify the End Repaired and Adaptor Ligated DNA. A final PCR cleanup was performed with SPRIselect beads added at a 1X concentration, followed by two 85% Ethanol washes and resuspension in 0.1X TE Buffer. Samples were quantified using the High Sensitivity Qubit dsDNA Quantification Kit (Thermo Fisher, Catalogue No. Q32851) and amplified DNA was visualised using a High Sensitivity D5000 ScreenTape Assay for TapeStation Systems (Agilent, Catalogue No. 5067-5592).

G.4 CUT&RUN Sequencing and Data Processing

CUT&RUN libraries for HUVECs were sequenced on an Illumina NovaSeq 6000 system on S1 flow cells in 2 × 50 bp mode. Libraries for HDFs were sequenced on an Illumina NextSeq 2000 system on P2 flow cells in 2 × 150 bp mode (H3K9me3) or on P3 flow cells in 2 × 50 bp mode (H3K27me3). CUT&RUN data was then processed using the nf-core/cutandrun pipeline (v3.2.2) (*70*) implemented in Nextflow (v24.04.4) (*71*). The pipeline was executed on paired-

end sequencing data from H3K27me3 and H3K9me3 experiments performed on human dermal fibroblasts (HDFs) from donors of different ages, as well as H3K27me3 experiments on in vitro passaged HUVECs. Raw sequencing reads were initially assessed for quality using FastQC (v0.12.1) (*72*). Adapter trimming and quality control were performed using Trim Galore (v0.6.6) (*73*) with Cutadapt (v2.10) (*74*) for adapter removal. The trimmed reads were aligned to the human reference genome assembly GRCh38.p13 (GCA_000001405.28) using Bowtie2 (v2.4.4) (*75*) with default parameters. Post-alignment processing was performed using SAMtools (v1.17) (*76*) for sorting, indexing, and generating alignment statistics. Duplicate reads were marked and removed using Picard MarkDuplicates (v3.1.0) (*77*). Quality control metrics of the aligned data were assessed using multiple approaches: Preseq (v3.1.1) (*78*) was used for library complexity estimation, and deepTools (v3.5.1) (*79*) was employed for generating quality metrics including correlation plots, PCA analysis, and fingerprint plots. Coverage tracks were generated using deepTools bamCoverage and converted to bigWig format using UCSC tools (v377/445) (*80*) for visualization.