PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Modeling DNA methyltransferase function to predict epigenetic correlation patterns in healthy and cancer cells

Ariana Y. Tse[a] and Andrew J. Spakowitz[b,1]

Affiliations are included on p. 8.

Edited by Jeff Wereszczynski, Illinois Institute of Technology, Chicago, IL; received August 1, 2024; accepted November 16, 2024, by Editorial Board Member Karolin Luger

DNA methylation is a crucial epigenetic modification that orchestrates chromatin remodelers that suppress transcription, and aberrations in DNA methylation result in a variety of conditions such as cancers and developmental disorders. While it is understood that methylation occurs at CpG-rich DNA regions, it is less understood how distinct methylation profiles are established within various cell types. In this work, we develop a molecular-transport model that depicts the genomic exploration of DNA methyltransferase within a multiscale DNA environment, incorporating biologically relevant factors like methylation rate and CpG density to predict how patterns are established. Our model predicts DNA methylation-state correlation distributions arising from the transport and kinetic properties that are crucial for the establishment of unique methylation profiles. We model the methylation correlation distributions of nine cancerous human cell types to determine how these properties affect the epigenetic profile. Our theory is capable of recapitulating experimental methylation patterns, suggesting the importance of DNA methyltransferase transport in epigenetic regulation. Through this work, we propose a mechanistic description for the establishment of methylation profiles, capturing the key behavioral characteristics of methyltransferase that lead to aberrant methylation.

DNA methylation | facilitated diffusion | epigenetic modifications | polymer physics modeling | cell differentiation

Significance

DNA methylation plays a fundamental role in regulating gene expression and maintaining genomic stability. Errors in methylation patterning result in impaired cellular function and abnormal gene expression, which leads to cancer, neurodevelopmental disorders, and autoimmune conditions. Determining the factors that influence the establishment of a distinct methylation profile is important for unraveling the intricacies of cellular differentiation and the mechanisms responsible for epigenetic-related diseases. In this work, we focus on how the kinetic and transport behavior of DNA methyltransferase leads to methylation patterns associated with cancer. The findings of this work provide fundamental guidance for future researchers to develop therapies that target the mechanistic causes for diseases associated with aberrant DNA methylation patterns.

Within the nucleus of a eukaryotic cell, genomic DNA is organized within a structure called chromatin whose multiscale behavior includes local wrapping into nucleosomes (1, 2), intermediate-scale looped motifs arising from cohesin function (3–5), and large-scale compartmentalization (6–8). Chromatin organization plays a pivotal role in determining cellular identity and function by altering the accessibility of DNA to regulatory proteins and transcriptional machinery. The large-scale organization of chromatin is determined by epigenetic modifications—chemical marks on the DNA and histone proteins that encode for regions to adopt more open or closed chromatin states. Variation in these chemical modifications results in altered expression of genes, leading to different cell types and behaviors despite having the exact same DNA sequence.

A key chemical modification is methylation, which has been found to serve as a recruiter for various chromatin remodelers that condense the DNA and alter transcription (9–13). DNA methylation describes a covalent addition of a methyl mark to cytosine nucleotides at key locations within the genome. In mammals, methylation most prevalently occurs at cytosine-phosphate guanine dinucleotides, known as CpG sites, that are unevenly distributed across the genomes (14–18). Most of the genome is sparse in CpG sites outside of select high-density CpG regions known as CpG islands.

Upon cell replication, DNA methyltransferases (DNMTs) serve to recapitulate and maintain the methylation pattern across the DNA sequence. There are three classes of DNMTs (dubbed DNMT1, DNMT2, and DNMT3). While the role of DNMT2 has not been extensively studied, there is more understanding of the functions of DNMT1 and DNMT3. It is currently held that DNMT3 is responsible for de novo methylation, which is most prevalent during embryonic development (19–21). In contrast, DNMT1 typically serves as DNA-methylation maintenance proteins, ensuring that the methylation pattern of newly synthesized DNA matches that of the parent strand (22–26). The proper patterning of methyl marks is required to ensure the cells function correctly, and aberrations in DNMT1 function can result in conditions such as cancers (27–30), neurodevelopmental disorders (31), and other health conditions (32–34). While these ideas have been generally accepted, the mechanisms that result in these unique methylation patterns are less understood.

Author contributions: A.Y.T. and A.J.S. designed research; A.Y.T. and A.J.S. performed research; A.Y.T. contributed new reagents/analytic tools; A.Y.T. analyzed data; and A.Y.T. and A.J.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. J.W. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: ajspakow@stanford.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2415530121/-/DCSupplemental.

Published January 10, 2025.

There exist several critical questions regarding how DNA methylation patterns are conserved after cell replication and the factors that contribute to unique methylation patterns necessary for cell differentiation. A crucial focus of research on DNMTs centers around their kinetic behavior, encompassing their processive tendencies (35, 36), and how this behavior may impact the protein's methylation patterning. Some previous works have focused on elucidating relative contributions of DNMT1 and DNMT3 on the establishment of a methylation pattern (37), determining postreplication DNA methylation rates (38), quantifying the correlation of methylation rates and states (39, 40), and determining important interactions for methylation patterning like PCNA-DNMT1 (41). The relationship between methylation patterns and chromatin structure has been explored in previous works to understand the structural effects of epigenetic modifications. Namely, methylation profiles have been measured across genomic sequences to characterize unique cell types. These studies reveal how methylation patterns may result in cell differentiation (42–49), reflecting the close relationship between DNA methylation, chromatin organization, and gene expression (40). Understanding the complexities of the dynamic behavior of DNMTs is crucial for understanding methylation patterning, and, in turn, provides insight into epigenetic regulation, cell differentiation, chromatin remodeling, and transcription.

These questions will be explored using a facilitated diffusion model. Our previous work on facilitated diffusion (50–52) is extended to capture necessary biological factors relevant to DNA methylation. Our facilitated-diffusion model was originally developed to determine the search efficiency of the target-site search process of DNA-binding proteins, revealing how protein binding rate can influence the timescale for target localization. While facilitated diffusion models have been used extensively for studying the temporal statistics of a DNA search process, the spatial metrics have been less explored. Specifically, the model by Díaz de la Rosa et al. (50) can characterize the genomic positioning of a DNA-binding protein during its search process by predicting the average amount of time a protein is localized to a specific genomic position on the DNA, effectively spatially mapping the protein. This application for this model allows for prediction of DNA chemical modifications and epigenetic patterns, like DNA methylation.

In this work, we consider methylation patterning arising from the transport behavior and kinetics of DNA methyltransferase. That is, we determine how the properties related to a DNA methyltransferase's exploration of the genome relate to the resulting methylation patterns. This differs from the approach of other mathematical models which focus on the kinetics, rather than the transport (38, 41, 53–55). We develop a model for a DNA methyltransferase undergoing a facilitated diffusion process to characterize the exploration behavior of the protein. Facilitated diffusion models (56–58) have been typically used for predicting the target search times of DNA-binding proteins like transcription factors (59–64) and to determine factors that aid in a DNA-binding protein's ability to find its target (50–52, 65). With this framework, we aim to develop a model to describe the localization of DNA methyltransferase within a complex DNA environment and to make predictions for the resultant methylation patterns given the protein's transport behavior.

To most accurately describe the DNA methyltransferase's exploration of the genome, we model the DNA environment in a way that captures the complex behavior of chromosomal DNA in vivo. In this model, we adopt a polymer-physics-based statistical description of the DNA, which has been found to exhibit multiple scaling-law behaviors at various length scales (66–70). We utilize experimentally measured looping probabilities of in vivo DNA to define a polymer model that captures the multiscale behavior of DNA (71). This model allows us to predict the transport of DNA methyltransferase undergoing facilitated diffusion in this complex DNA environment and to explore the impact of transport behaviors on the establishment of methyl marks. With this model, we determine the impact of the structure and behavior of chromosomal DNA on the transport of DNA methyltransferase. We then show how we can leverage this descriptive model to predict distinct methylation state correlation distributions of a variety of cancerous cell types and determine the kinetic and transport behavior of DNA methyltransferases that lead to characteristics of the methylome associated with cancer.

Model

In our study, we develop a model that captures the transport behavior of DNA methyltransferase undergoing a facilitated diffusion process throughout a multiscale DNA environment. The DNA methyltransferase is modeled as a particle that alternates between a DNA-bound state and an unbound state, while the DNA is modeled as a polymer chain that exhibits the biologically relevant behaviors seen in chromosomal DNA in vivo, specifically the multiscale behavior described by Chan et al. (71). The bound state of the particle captures the processivity exhibited by DNA methyltransferase, which allows the protein to methylate multiple sites while bound (39, 72–74). The details of modeling each of these components are described in detail below.

The model defines a diffusing particle which engages in an alternating process of binding and unbinding to a polymer chain, illustrated in Fig. 1. The model captures methylation correlation, governed by the joint probability that a second site at position s is methylated given that the first site $s_0 = 0$ is methylated. For the initial site to be methylated, the particle is bound to the polymer

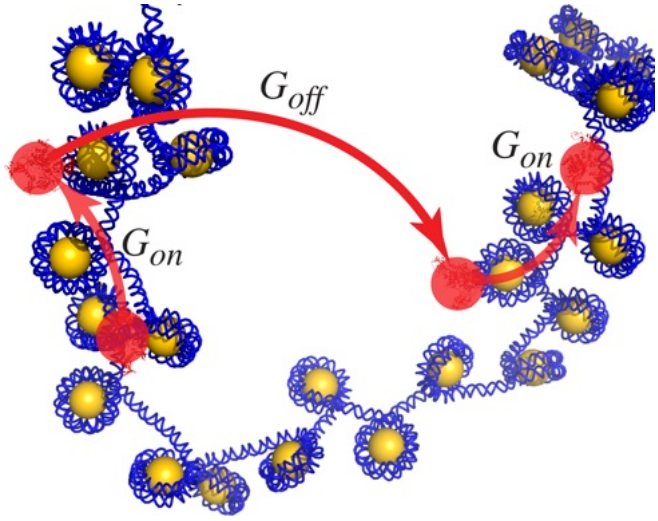

Fig. 1. Schematic of a facilitated diffusion process in which the particle engages in 1D diffusion along the DNA, described by the function G_{on}, and unbinds and diffuses three-dimensionally throughout the nucleoplasm, described by the function G_{off}. This schematic only shows two steps of a process which can contain an infinite number of switches between the bound and unbound states. The transport of the protein while bound depends on its bound diffusivity D_{on} and its unbinding rate k_u. Similarly, its unbound transport is dependent on its diffusivity throughout the nucleoplasm D_{off} and a binding rate k_b.

chain at position $s_0 = 0$ at time $t = 0$ as an initial reference point. This initial reference point indicates the genomic position of the DNA methyltransferases' first binding occurrence before transporting along the chain via a facilitated diffusion process. From this initial reference point, the particle engages in the one-dimensional random walk along the chain with a diffusivity D_{on}.

The particle has a rate for unbinding from the chain k_u and can unbind at any location of the polymer chain. When the particle unbinds, it freely diffuses in three-dimensions throughout the environment with a diffusivity D_{off}. While diffusing the particle can rebind to any segment of the polymer chain that is within a distance a with a binding rate of k_b. The transport of the particle while bound is governed by the probability function (or Green's function) G_{on} and the transport of the particle while unbound is governed by G_{off}. The overall transport process can alternate between the bound state and the unbound state any number of times.

Polymer Model Description. We first introduce the polymer-physics model used to describe the DNA. To capture a biologically relevant DNA environment, we develop a polymer-physics model that captures the multiscale behavior observed in experiments. We use the description discussed by Chan et al. (71), which identifies three different behavioral regimes at varying length scales of chromosomal DNA (71), illustrated in Fig. 2. These three behavioral regimes were determined by analyzing the DNA's looping probability at different genomic lengths. It was noted that at short lengths (less than 30 kb), the DNA exhibited the power-law scaling for looping probabilities akin to a Gaussian chain. At an intermediate length scale (approximately 30 to 400 kb), chromosomal DNA behaves as a fractal globule attributed to the presence of cohesin, which actively form loops in the genome. At large length scales (greater than 400 kb), the DNA returns to behaving like a Gaussian chain. From these looping probabilities, we determine the scaling behavior of the mean square end-to-end distance of the DNA $\langle R^2 \rangle$ with varying length, shown in Fig. 2. This is determined through the relationship $P_l \propto \langle R^2 \rangle^{-3/2}$.

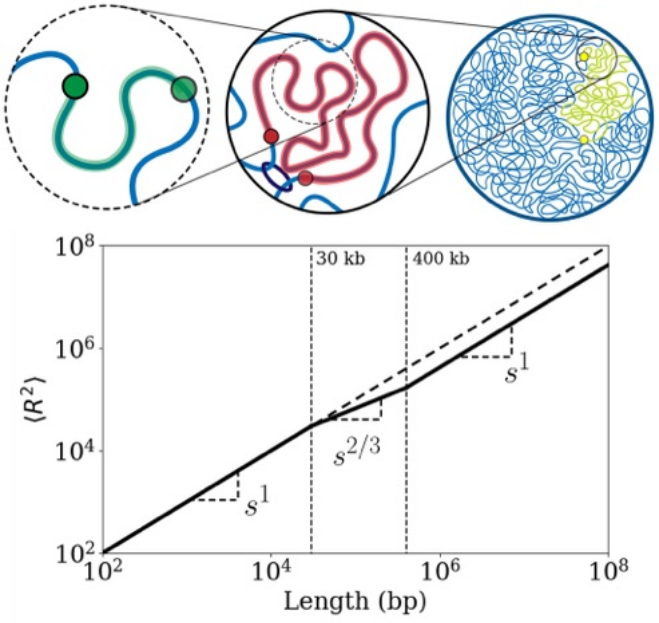

Fig. 2. Schematic depicting the multiscale nature of chromosomal DNA. DNA behaves as a random walk at short length scales but behaves like a fractal globule in intermediate length scales due to the function of cohesin. Chromosomal DNA returns to behaving like a random walk at large length scales. The plot shows average end-to-end distance squared $\langle R^2 \rangle$ versus s, showing the three scaling behaviors: $\beta_1 = 1$, $\beta_2 = 2/3$, $\beta_3 = 1$. For comparison, the mean squared end-to-end distance is shown as a dashed line for a pure Gaussian chain.

With the average end-to-end distance, we describe the DNA's average conformation through the Fourier-transformed Green's function

$$\hat{G}_D(\vec{k}; s) = \exp\left(-\frac{1}{6}\vec{k}^2\langle R^2\rangle\right), \quad [1]$$

which describes the segmental density of the chain in space. Here, $\vec{k}$ is the Fourier-transformed variable for a spatial position $\vec{R}$. The end-to-end distance of polymer $\langle R^2 \rangle$ is dependent on the polymer length s, such that

$$\langle R^2 \rangle = \begin{cases} R_1^2 \left(\frac{s}{s_1}\right)^{\beta_1} & \text{for } s < s_1 \\ R_1^2 \left(\frac{s}{s_1}\right)^{\beta_2} & \text{for } s_1 \le s < s_2 \\ R_2^2 \left(\frac{s}{s_2}\right)^{\beta_3} & \text{for } s \ge s_3 \end{cases}, \quad [2]$$

Here, s_1 and s_2 are the genomic lengths associated with transitions in scaling behavior and are found to be 30 kb and 400 kb, respectively. The scaling exponents β_1 and β_3 are 1 the Gaussian chain behavior, and the intermediate scaling exponent β_2 is 2/3 for the fractal globule regime. To ensure continuity of $\langle R^2 \rangle$ between behavioral regimes, we assert

$$R_2^2 = R_1^2 \left(\frac{s_2}{s_1}\right)^{\beta_2}. \quad [3]$$

Transport Model Description. We now define the protein's transport along the chain. Evaluating the overall transport of the protein requires defining the probability of binding or unbinding at a given time t as dictated by the binding and unbinding rates. From the unbinding rate, we define the probability that the particle will remain bound to the polymer at time t as

$$P_{\text{on}}(t) = \exp\left(-k_u t\right). \quad [4]$$

As the protein spends more time bound, it becomes more likely to unbind. We define the rate of unbinding to be its negative derivative over time

$$S_{\text{on}}(t) = -\frac{dP_{\text{on}}(t)}{dt}. \quad [5]$$

The description for the transport of the protein can be broken into two parts: transport while bound and transport while unbound to the chromosomal DNA. The transport of the particle bound to the chain can be defined by a Green's function G_{on}, which describes the probability of a particle beginning at position s_0, and diffusing to position s in time t. The Green's function G_{on} is governed by 1D-diffusion and is given by

$$G_{\text{on}}(s|s_0; t) = \frac{1}{\sqrt{4\pi D_{\text{on}} t}} \exp\left[-\frac{(s-s_0)^2}{4D_{\text{on}} t}\right]. \quad [6]$$

Notably, the transport in the bound state is not dependent on the polymer configuration and only depends on the bound diffusivity of the particle. Once the particle unbinds and is diffusing throughout the nucleoplasm, its probability of rebinding to the polymer is dependent on its binding rate but additionally

the chromosomal DNA conformation. That is, the protein is more likely to rebind to the polymer if it remains in close spatial proximity to the surrounding DNA. Therefore, the unbound transport is dependent on both the DNA conformation (governed by G_D) and the protein's diffusion (governed by G_p), resulting in the off-state transport equation

$$G_{\text{off}}(s|s_0;t) = \frac{1}{M(t)} \int d\vec{R}_1 \int d\vec{R}_2 G_D(\vec{R}_1; s - s_0) \times G_p(\vec{R}_2; t) H(a - |\vec{R}_1 - \vec{R}_2|). \quad [7]$$

The off-state Green's function $G_{\text{off}}(s|s_0;t)$ gives the joint probability of a polymer segment and the protein being found within a binding distance a at time t at genomic position s given that the particle leaves the polymer at s_0 at time $t = 0$. The Heaviside function restricts the off-state trajectories of the particle to be within the binding distance of the polymer segment s at time t. In other words, this equation describes the transport of a particle diffusing throughout the nucleoplasm until it comes to some spatial position $\vec{R}_1$ that is within binding distance a of a segment of the DNA, characterized by its genomic position s and spatial position $\vec{R}_2$. In this equation, $M(t)$ serves as a normalization constant and gives the total length of polymer within the binding distance a of the particle at time t. G_p describes the spatial probability density for the protein which diffuses three-dimensionally throughout the nucleoplasm and is written as

$$\hat{G}_p(\vec{k};t) = \exp\left(-D_{\text{off}}\vec{k}^2 t\right). \quad [8]$$

From this defined framework, we describe the probability of the particle rebinding based upon the total length of polymer the particle comes in contact at a given time t and the binding rate k_b, leading to the differential equation

$$\frac{dP_{\text{off}}}{dt} = -k_b M(t) P_{\text{off}}. \quad [9]$$

This equation specifies that the particle becomes less likely to stay unbound if it encounters more of the polymer during its diffusion. The solution to this differential equation is

$$P_{\text{off}}(t) = \exp\left[-k_b \int_0^t dt' M(t')\right]. \quad [10]$$

We define complete trajectories of the protein using the component transport terms described above. Individual transport processes include transport of the particle along the polymer from position s_0 to s_1 in time t_1, governed by the on-state Green's function $G_{\text{on}}(s_1|s_0,t_1)$, the rate of unbinding $S_{\text{on}}(t_1)$, and the probability of remaining bound $P_{\text{on}}(t_1)$. Similarly, the transport of the particle while unbound from position s_1 to s_2 in time t_2 is governed by the off-state Green's function $G_{\text{off}}(s_2|s_1,t_2)$, the rate of binding $S_{\text{off}}(t_2)$, and the probability of remaining in the off state $P_{\text{off}}(t_2)$. From this framework, we determine the statistical weight of trajectories containing any number of transitions between the bound and unbound states. This involves the enumeration of an infinite number of genomic trajectories and state trajectories accounting for all possible paths along the polymer. Defining all trajectories requires integration over all possible transition times and all possible polymer positions where the transition may occur. This results in a convolutional structure that is greatly simplified with the use of a Fourier–Laplace transform. We define the Fourier–Laplace on-state Green's function

$$\hat{\tilde{G}}^{(S)}_{\text{on}}(\omega,\nu) = \mathcal{F}_{s\to\omega}\left\{\mathcal{L}_{t\to\nu}\left[G_{\text{on}}(s|s_0;t)S_{\text{on}}(t)\right]\right\}, \quad [11]$$

where we perform a Fourier transform over the polymer distance variable s to the Fourier variable ω and a Laplace transform over the time variable t to the Laplace variable ν. Similarly, we define the terms $\hat{\tilde{G}}^{(P)}_{\text{on}}$, $\hat{\tilde{G}}^{(S)}_{\text{off}}$, and $\hat{\tilde{G}}^{(P)}_{\text{off}}$. Utilizing this convolutional structure, we forgo the multiple explicit integrations over s and t, allowing us to define any possible trajectory as a product of Fourier–Laplace transformed variables. Therefore, we define the possible transport of the particle through the infinite summation of the possible trajectories. Considering the trajectories in which the methylation ends in an on-state results in the expression

$$\hat{\tilde{G}}(\omega;\nu) = \frac{\hat{\tilde{G}}^{(P)}_{\text{on}}}{1 - \hat{\tilde{G}}^{(S)}_{\text{on}}\hat{\tilde{G}}^{(S)}_{\text{off}}}. \quad [12]$$

We focus our analyses on the long-time and long-length behavior, which is equivalent to the small ω and ν behavior in Fourier–Laplace space. In this limit, the on-state functions are found to be

$$\hat{\tilde{G}}^{(S)}_{\text{on}}(\omega;\nu) = 1 - \tau_{\text{on}}\nu - l^2_{\text{on}}\omega^2 + \mathcal{O}(\omega^4,\nu^2), \quad [13]$$

and

$$\hat{\tilde{G}}^{(P)}_{\text{on}}(\omega;\nu) = l^2_{\text{on}} + \mathcal{O}(\omega^4,\nu^2). \quad [14]$$

Here, $l_{\text{on}} = \sqrt{1/\kappa_u}$, where l_{on} describes the average translocation distance of the particle during a single binding event. The on-time τ_{on} gives the average time spent in the on-state and is defined as $\tau_{on} = 1/\kappa_u$. Here, the dimensionless unbinding rate $\kappa_u = t_{\text{diff}}k_u$ is the unbinding rate k_u nondimensionalized by the on-state diffusion time $t_{\text{diff}} = b^2/D_{\text{on}}$.

The off-state Green's function $\hat{\tilde{G}}^{(S)}_{\text{off}}$ is calculated through an expansion for small ω and ν. This analysis results in the equation

$$\hat{\tilde{G}}^{(S)}_{\text{off}} = 1 - \Delta_D l^{1/2}_{\text{off}} - \nu\tau_{\text{off}}, \quad [15]$$

where Δ_D defines the fluctuations over ω, which emerge from the polymer model description $\hat{G}_D$. The complete derivation and expression for Δ_D is presented in *SI Appendix*. The term l_{off} is the average translocation distance of the particle during a single off state and defined as $l_{\text{off}} = (\kappa_b\tau_{\text{off}})^2$, where $\kappa_b = t_{\text{diff}}bk_b$ is the binding rate k_b nondimensionalized by $t_{\text{diff}} = b^2/D_{\text{on}}$ and the Kuhn length b. The off time τ_{off} is the average time the particle spends in the off state and is defined as

$$\tau_{\text{off}} = \int_0^\infty d\tau \exp\left[-\frac{\kappa_b}{\gamma}\int_0^{\tau\gamma} d\tau' M(\tau')\right], \quad [16]$$

where $\gamma = D_{\text{off}}/D_{\text{on}}$. Replacing these terms in Eq. **12** results in the final equation

$$\hat{\tilde{G}}(\omega;\nu) = \frac{\tau_{\text{on}}}{(\tau_{\text{on}} + \tau_{\text{off}})\nu + l^2_{\text{on}}\omega^2 + l^{1/2}_{\text{off}}\Delta_D}. \quad [17]$$

The inverse Fourier–Laplace transform of this equation from ω to s and ν to t elicits a probability density describing the probability of the protein being found at genomic position s at time t.

Results and Discussion

We determine the average time the DNA methyltransferase spends on each segment of DNA using Eq. **17**. We take the inverse Fourier–Laplace transform of Eq. **17** and integrate the resultant equation with respect to t to find the average time T that the particle resides at position s over the time period t. The Laplace inversion of $\hat{\tilde{T}}$ gives the expression

$$\hat{T}(\omega;t)=\frac{\tau_{\text{on}}}{l_{\text{on}}^2\omega^2+l_{\text{off}}^{1/2}\Delta_D}\left[1-\exp\left(-t\frac{l_{\text{on}}^2\omega^2+l_{\text{off}}^{1/2}\Delta_D}{\tau_{\text{on}}+\tau_{\text{off}}}\right)\right], \quad [18]$$

and a subsequent Fourier inversion from ω to s is performed numerically. Given a protein that starts at position $s=0$ and time t to engage in the facilitated diffusion process described in the previous section, this equation describes the total time the protein spends on a genomic segment s distance away from its starting position. The average time T is shown in Fig. 3 for varying time t.

The average time is characterized by a plateau followed by a decay that exhibits multiple scaling behaviors associated with the DNA behavioral regimes described in Eq. **2**. Beyond the plateau, the short length-scale regime exhibits a power-law scaling of $-3/2$. The intermediate length-scale regime shows -1 scaling, followed by a transition back to $-3/2$ scaling in the large length-scale regime. These scaling behaviors relate to the polymer scaling behavior as $-3\beta/2$. These scaling behaviors coincide with the looping probabilities for the polymer in these three regimes, and the exploration time at short times is restricted to polymer segments that are looped with the $s_0=0$ segment of polymer.

From Fig. 3, we see that when time increases, the plateau in the average time extends to further distances as the protein explores further regions of the polymer. At sufficiently long time, the average-time plot converges to a distribution describing the protein's total time spent at each polymer position at infinite time. This steady-state exploration time depends on the DNA methyltransferase transport properties and is predictive of a steady-state methylation pattern established after cell replication and maintained during interphase. The steady-state exploration time distributions can be resolved through integration of $G(s;t)$ over infinite time or simply by asserting $\nu=0$ in Eq. **17**, resulting in

$$\hat{T}(\omega;t\to\infty)=\hat{\tilde{G}}(\omega;\nu=0)=\frac{l_{\text{on}}^2}{l_{\text{on}}^2\omega^2+l_{\text{off}}^{1/2}\Delta_D}, \quad [19]$$

where we note that $\tau_{\text{on}}=l_{\text{on}}^2$ in our dimensionless units. This equation involves the two transport parameters for the protein: l_{on} and l_{off}. We explore how these two transport properties impact the protein's exploration distribution. In doing so, we characterize how the methyltransferase's kinetic properties influence its methylation patterning.

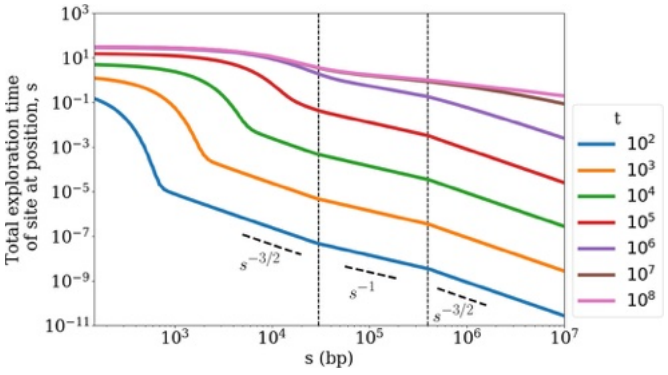

Fig. 3. Average time of genomic exploration of DNA methyltransferase at time t. This plot shows the total amount of time that a DNA methyltransferase spends at each genomic position s when engaging in the described facilitated diffusion process over the time t. The plot shows that as the amount of time increases, the methyltransferase explores further genomic regions. At sufficiently long time, the distribution saturates, indicative of a steady-state exploration distribution.

Impact of DNA Methyltransferase Transport Properties on Genomic Exploration. We assess how the binding distance l_{on} influences the exploration distribution. In Fig. 4*A*, we show the steady-state exploration distributions for a protein with varying binding distances for the biologically relevant multiscale model for the chromosome (solid curves) and a Gaussian-chain polymer (dashed curves) that exhibits a single mean-square end-to-end distance scaling $\beta=1$. The equations for the exploration distributions in the pure Gaussian environment are consistent with the work of Díaz de la Rosa et al. (50).

In both the pure random walk and the multiscale polymer model, the exploration distributions exhibit plateaus which extend to further distances when l_{on} increases, illustrating that a protein which travels further distances while bound will spend more time exploring proximal regions of the genome before transporting to distal regions through off-state jumps. Additionally, in both cases, the exploration distributions decline into $-1/2$ scaling at long distances. This $-1/2$ scaling behavior relates to the polymer end-to-end distance scaling behavior β as $3\beta/2-2$. However, it is notable that the transition into this $-1/2$ scaling differs for the two polymer models. Specifically, the exploration distributions for the protein in the multiscale environment have variable transition behaviors into the long-distance scaling. For example, exploration distributions of proteins with a low l_{on} value have a plateau which transitions into scaling exponents of -1 and $-1/3$ before settling into the $-1/2$ scaling, showing scaling behavior that transitions nonmonotonically. This contrasts with the exploration distributions of the proteins in the Gaussian environment, which has scaling behavior that monotonically decreases to this $-1/2$ scaling. In the multiscale environment, we note that when l_{on} is a large enough value, the exploration distributions in the multiscale environment do not distinguish themselves from the exploration distributions of a corresponding protein in a pure Gaussian chain environment. In these cases, the plateaus in the exploration distributions extend through the intermediate behavioral regime and do not exhibit the nonmonotonic scaling behavior. In these instances, the bound protein slides along the DNA for distances that are larger than the length scale of the intermediate fractal globule regime. Because the protein stays bound through those distances, its genomic exploration is not affected by the chromatin structures associated with those scaling behaviors.

Similar observations can be made in the exploration distributions of proteins with varying hopping distances l_{off}. Fig. 4*B* shows that proteins with shorter hopping distance exhibit exploration distributions with longer plateaus. That is, l_{off} contributes to an equal but opposite effect of l_{on}. Particularly, a protein with a hopping distance of $l_{\text{off}}=0$ does not unbind, leading to $l_{\text{on}}\to\infty$, illustrating their inverse relationship. This effect reveals that the shape of the exploration distribution is

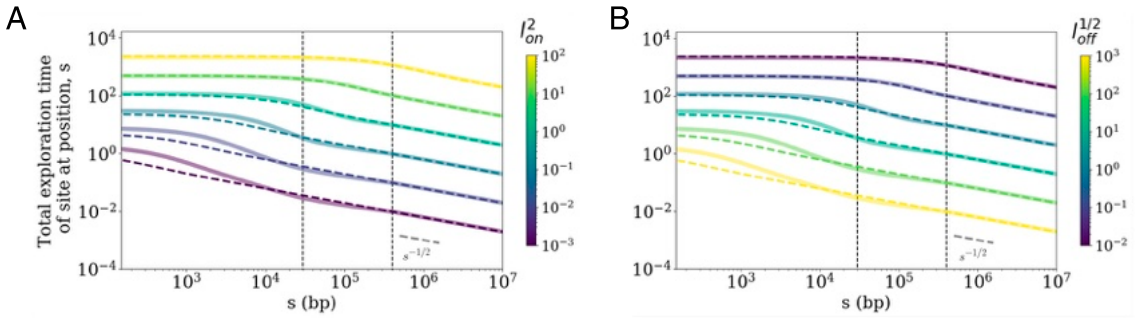

Fig. 4. Steady-state exploration distribution of methyltransferases with varying bound translocation distance l_{on} with $l_{off} = 10$ (*A*) and hopping distance l_{off} with $l_{on}^2 = 0.1$, (*B*). The solid line is the exploration distributions of the methyltransferase within the multiscale DNA environment, while the dotted lines are that of corresponding DNA methyltransferase in a DNA environment behaving as a pure Gaussian chain. We see that at large l_{on} values and small l_{off} values, the exploration distributions do not differ for the different DNA environments.

dependent on the ratio of l_{on} and l_{off}. Akin to the l_{on} analysis, we see that proteins with small l_{off} values have exploration distributions that are ambivalent to the multiscale environment and behave indistinguishably from proteins in the Gaussian chain environment. These proteins are more likely to hop short genomic distances that are not long enough to span across the chromatin structural features responsible for the intermediate behavioral regime.

Interestingly, we note that the proteins in the multiscale environment have elevated exploration times at distances close to the starting position in comparison to the proteins in the Gaussian chain environment. This is because the fractal globule intermediate scaling results in DNA structures that have a larger average density than that of a pure Gaussian chain. The probability of protein binding to the polymer S_{off} captures the increased likelihood for protein binding with increasing DNA length $M(t)$ with which the protein encounters. Therefore, we hypothesize that the increased time at short distances may be a result of increased binding due to a more DNA-dense environment. Additionally, this denser DNA environment also describes structures in which far genomic regions are spatially closer. This closer spatial proximity and increased binding is predicted to cause the protein to be more likely to re-explore genomic segments in its immediate spatial region, rather than hopping sufficiently large spatial distances that pulls it away from its starting position. This increases the likelihood that the protein stays spatially close to its starting position, encouraging re-exploration of nearby genomic regions and larger exploration times at short distances. Through this observation, we hypothesize that spatial regions that are more dense in DNA may localize the protein. Fluorescent microscopy experiments have shown DNA-binding proteins involved in RNA transcription become localized in denser DNA regions within the nucleus (75), though this has not yet been explored in DNA methyltransferase. This theoretical observation provides some physical intuition on how the interactions between a DNA environment and a DNA-binding protein may lead to localization behaviors.

Predicting Fraction of Methylation from DNA Methyltransferase Transport and Kinetics. The mathematical description of our model describes the transport of the DNA methyltransferase, which allows us to calculate the total time the protein spends on each DNA segment. However, it does not specify the establishment of methyl marks based on methyltransferase activity. This can be incorporated into the model by defining a methylation rate k_m. We describe a protein which deposits a methyl mark at some rate k_m while bound. Whether a site becomes methylated is dependent on the amount of time the protein spends on a site. From this, we define the fraction of methylation f_m, which is given by

$$f_m(s) = 1 - \exp\left[-\kappa_m T(s; t \to \infty)\right], \qquad [20]$$

where we define the dimensionless methylation rate $\kappa_m = t_{diff} k_m$. This equation describes the fraction of methylation at a specific genomic site. We present the methylation fraction with varying κ_m in Fig. 5. The shape of the distribution is consistent with that of the exploration time distribution, shown in a dotted line. We see that increasing methylation rate corresponds with an overall increase in methylation across the genome, showing that a DNA methyltransferase that catalyzes methylation at faster rates will methylate more of the genome. The methylation fraction saturates at 1, describing complete methylation or certainty of methylation. Large methylation rates cause the methylation fraction distribution to have an extended plateau at the saturated methylation fraction of 1, showing the significant effect that methylation rate has on establishment of a methylation pattern.

Assuming that the genomic position at $s = 0$ is methylated (i.e., a methylation fraction of 1), these distributions depict the spatial correlation of methylation states across the genome. Methylation correlation has been experimentally measured in a range of cells and has been distinguished as distinct between cell types (39, 40). As is, this methylation-state correlation

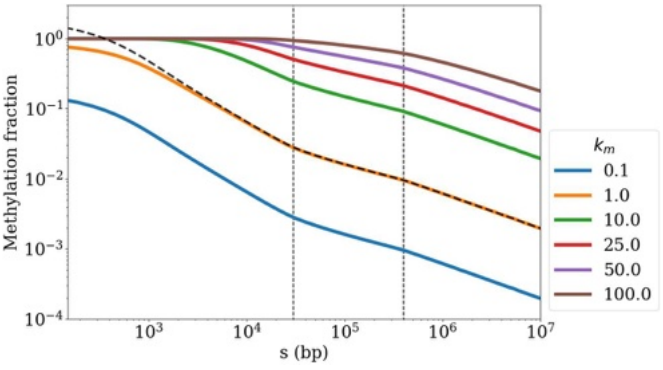

Fig. 5. Steady-state methylation fraction distribution determined using total exploration time distributions and variable methylation rates κ_m. The total exploration time of the protein is shown in the black-dotted line to illustrate its relationship to the methylation fraction. Methylation fraction saturates to 1 with high methylation rates, showing the impact of methylation rate on establishing methylation pattern.

is predicted exclusively from DNA conformational structure and the protein's transport and kinetic properties. However, to develop a more biologically accurate model for methylation, we expand our model to include genetic information. Specifically, we consider the presence of CpG sites and CpG islands, the predominant genomic regions for DNA methylation. While it is understood that DNA methylation does not exclusively occur at CpG sites, CpG sites make up a significant portion of methylation targets, accounting for approximately 98% of methylated regions (76). For this reason, we expand our model to consider this significant factor in methylation patterning. In doing so, our model is representative of the biological system when predicting methylation correlation distributions. These predictions are then compared with methylation state correlations that have been experimentally measured through whole genome bisulfite sequencing.

Theory Predicts Transport Properties Which Correspond to Unique Cell Type Methylation Correlation Profiles. To leverage this theory to predict methylation correlation, we expand the model to include a description for CpG site correlation. We assert that methylation is dependent on methyltransferase transport to a location and the density of CpG at that location, leading to the probably of methylation to be defined

$$p_m(s) = p_c(s) f_m(s), \qquad [21]$$

where p_c is the probability of a CpG site and f_m is the methylation fraction, presented in the previous section. The presence of CpG islands, which have been determined to span 300 to 3,000 bp in length through sequencing experiments, will result in higher correlations at those length scales. We assume that CpG islands lead to a correlation such that

$$\langle p_c(s) p_c(0) \rangle - \langle p_c \rangle^2 = p_0(1 - p_0) \exp(-\lambda |s|). \qquad [22]$$

Here, we describe the CpG correlation established within a CpG island. The average CpG density p_0 gives the average CpG content within a CpG island and ensures that the condition in which a genomic region is either void of CpG ($p_0 = 0$) or entirely composed of CpG ($p_0 = 1$). We specify the average CpG density to be $p_0 = 0.5$, consistent with the threshold used by Gardiner-Garden for defining CpG islands (77). We define $1/\lambda$, the average CpG island length, to be 1,000 bp as supported by genomic sequencing (78, 79). Assuming that the CpG probability p_c and the methylation fraction f_m are not correlated, we define the correlation of methylation to be

$$\begin{aligned} C_m(s,0) &= \langle p_c(s) p_c(0) \rangle \langle f_m(s) f_m(0) \rangle - p_0^2 f_0^2 \\ &= \frac{p_0(1-p_0) f_0(1-f_0)}{p_0 f_0 (1 - p_0 f_0)} \bar{C}_c \bar{C}_f + \frac{f_0^2 p_0 (1-p_0)}{p_0 f_0 (1 - p_0 f_0)} \bar{C}_c \\ &\quad + \frac{p_0^2 f_0 (1 - f_0)}{p_0 f_0 (1 - p_0 f_0)} \bar{C}_f. \end{aligned} \qquad [23]$$

In this equation, $\bar{C}_c$ and $\bar{C}_f$ are the normalized correlation functions for CpG density and methylation fraction, and the methylation fraction f_0 describes the average steady-state methylation fraction of the genome. The value for the average methylation fraction is dependent on the transportation and methylation kinetics of DNA methyltransferase but also encompasses the effect of other possible proteins like TET proteins, which demethylate or inhibit methylation. In other words, while the model's description of DNA methyltransferase activity may predict a specific methylome, other proteins not discretely modeled in this description can influence the methylome through modulating methylation or demethylating genomic regions, resulting in a variable average methylation fraction that cannot be predicted exclusively from theoretical DNA methyltransferase transport and kinetics. This average methylation fraction value f_0 asserts that if the DNA methyltransferase achieved complete methylation ($f_0 = 1$), methylation correlation would only be dependent on CpG correlation C_c.

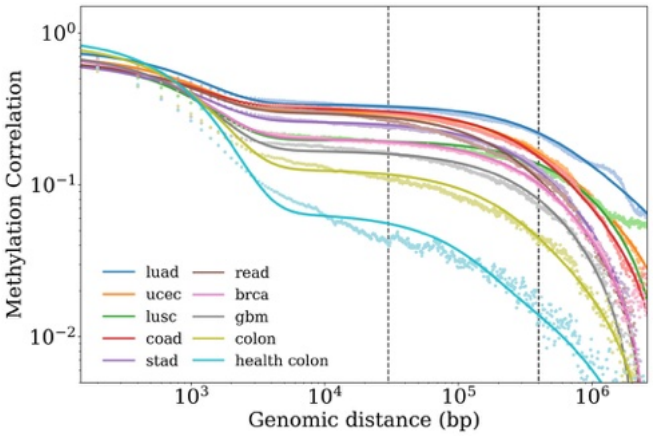

Fig. 6. Theory predicted methylation correlations fitted to methylation correlation data for chromosome 1 of 9 cancerous cell types and a healthy colon cell type generated by Zhang et al. (40). Each theoretical correlation distribution is fitted to the collected data using three parameters: transport ratio $l_{\text{off}}^{1/2}/l_{\text{on}}^2$, methylation rate κ_m, and fraction of methylation f_m. Theory shows close fits that recapitulate the qualitative and quantitative features of the experimental data.

In this model, we specify the average CpG density $p_0 = 0.5$ and the average CpG island length $1/\lambda = 1{,}000$ bp. We introduce the transport ratio $l_{\text{off}}^{1/2}/l_{\text{on}}^2$, the methylation rate k_m, and the average methylation fraction f_0, which we fit to existing methylation correlation data for Chromosome 1 of 9 cancerous human cell types and a healthy colon tissue cell collected from whole genome bisulfite sequencing experiments, shown in Fig. 6. These methylation correlation plots were generated by Zhang et al. (40). We conducted a parameter sweep through 1,000 transport ratios spanning 10^{-2} and 10, 100 methylation rates spanning 10^{-5} and 10^{-2}, and 100 methylation fractions spanning 0 and 1. The best fit was determined through minimizing mean squared logarithmic error. Through this fitting, we determine the transport and kinetic parameters of DNA methyltransferase that result in the distinct methylation correlation distribution for the unique cell types. These values are shown in Table 1.

The theoretical fits are both qualitatively and quantitatively representative of the methylation correlation data, describing a distinct transport behavior, methylation behavior, and average methylation fraction for each of the cell types. Specifically, the predicted methylation correlations resemble the experimentally measured methylation correlations by the characteristic high methylation correlation at distances less than 3,000 bp, which can be accounted to CpG islands, and a plateau region which varies in height and length for the various cell types. Our ability to recapitulate this methylation correlation data from a DNMT transport model demonstrates the strong relationship between methylation profiles and DNA methyltransferase behavior, showing how different methylation correlation patterns may be due to variable binding behavior of DNA methyltransferase.

Table 1. Parameter values for best fit theoretical methylation correlations for nine cancerous cell types and one healthy colon cell type

	$l_{off}^{1/2}/l_{on}^{2}$	κ_m	f_0	MSLE
luad	0.134	8.59E-04	0.21	2.44E-03
ucec	0.197	9.20E-04	0.18	1.23E-03
lusc	0.047	2.84E-04	0.43	1.33E-01
coad	0.210	8.83E-04	0.02	8.30E-02
stad	0.180	7.28E-04	0.22	1.62E-03
read	0.357	1.22E-03	0.13	1.33E-01
brca	0.149	6.38E-04	0.48	1.23E-03
gbm	0.152	6.21E-04	0.57	5.96E-03
colon	0.404	1.37E-03	0.76	1.07E-02
Healthy colon	1.403	3.42E-03	0.90	4.77E-02

Each cell type has a unique transport ratio, methylation rate, and methylation fraction.

From our model, we determine how DNMT behavior might vary across cell types. In particular, this model predicts that DNA methyltransferase in healthy cells has a transport ratio that is approximately a magnitude larger than that in cancerous cells, suggesting that DNMT remains unbound for larger distances in the healthy cells in comparison to those in the cancerous cells. Additionally, our model predicts that DNA methyltransferases in the healthy cells have a higher methylation rate than that in the cancerous cells. This suggests that DNMT in healthy cells diffuse throughout the nucleoplasm but methylate genomic regions when bound more readily. This contrasts with the DNMT behavior predicted for cancerous cells, which can be characterized by prevalent binding but reduced methylation reactions.

Differences in binding behavior may be accounted to a variety of biological influences. Variation in DNMT behavior among cell types is likely to arise from possible mutations in the genes responsible for encoding DNMTs, which may result in variable DNMT activity, or due to variable interactions with different proteins. There has been a variety of proteins which have been found to interact with DNMT1, regulate DNMT behavior, and effect recruitment and localization of DNMTs. These include MBD3 (80), PCNA (81), UHRF1 (82, 83), G9a (84, 85), and MeCp2 (86). For example, UHRF1 has been found to affect cell differentiation and the proliferation of cancerous cells through targeted recruitment of DNMT1 (87, 88). Additionally, it had been observed that DNMT1 is targeted to retrotransposons—genetic sequences that can move in the host genome (89). Therefore, from the predictions of our model, we suggest that genetic mutations, variability in the levels of different proteins, and behavior of transposable elements result in varying DNMT behaviors that may lead to the development of cancerous cells. We suggest that these factors may result in DNMTs in cancerous cells having impaired targeting and recruitment.

In addition, our model predicts the average methylation fraction of the genome for these cell types. We observe that the healthy cell type has the highest degree of methylation and is significantly more methylated than many of the cancerous cells. This agrees with the observation that carcinogenesis results in an overall global loss of methylation (90–92). More precisely, cancer cells have been found to be largely hypomethylated with hypermethylated sections in promoter regions (93–95). Because our model predicts more prevalent DNMT binding in cancerous cells, we hypothesize that this behavior causes certain regions to be wrongly methylated. Overall, our model concludes that DNMT behavior that presents less specificity in methylation results in cancerous cells.

Additionally, *SI Appendix* presents two alternative models that utilize different assumptions for the biological system. These alternatives were considered to assess certain biological influences on methylation but to also explore how the model may be modified to consider other hypotheses or research questions. In one of these alternative models, we model DNA as a pure Gaussian chain to investigate the effect of DNA organization on methylation. Additionally, we present a model that assumes DNMT behavior (transport ratio and methylation rate) is consistent among all cell types and only the methylation fraction varies across cell types. The resulting model relies on an assumption that protein interactions or DNMT mutations have limited variability across cell types or have negligible influence on the behavior of DNMTs. Both alternative models also resulted in fits that captured the qualitative features of the experimental methylation correlation distributions but with overall higher mean-squared logarithmic errors.

The Gaussian chain methylation model predicts larger average methylation fractions and transport ratios for the healthy colon cell, in agreement with the multiscale behavioral model. This is consistent with our hypothesis that DNMT in cancerous cells binds more prevalently, resulting in less specificity in methylation. From both additional models, we recapitulate the qualitative features of DNA methylation data spanning large genomic distances from a transport model that describes the binding, unbinding, and diffusion of DNA methyltransferases. In doing so, we predict DNMT transport behavior that coincides with biological observations. We attest that while the correlation between aberrant methylation and cancer has been observed is many previous works, this model provides a physical description for the establishment of methylation patterns and a mathematical framework for relating methylation patterns to DNMT behavior.

Summary

In this work, we present a model for the transport of a DNA-binding protein throughout a multiscale DNA environment. Through the expansion of this theory to include crucial biological aspects related to methylation (e.g., methylation rate and CpG density), our model recapitulates the methylation correlation distributions of various cell types. The quantitative agreement between the theory and the data suggests that the transport and kinetic behavior of DNA methyltransferase may be crucial aspects for methylation patterning. Our model predicts significantly different transport behavior for methyltransferase in cancerous cells from that in a healthy cell. This suggests that aberrant methylation may be a result of heightened activity in DNA binding and decreased methylation specificity. Our work demonstrates how transport properties, like bound translocation distance and hopping distance, and kinetic properties like methylation rate can have distinctive effects to a methylation profile. These findings shed light on the processes of methylome establishment, the mechanisms for cell differentiation, and potential causes for aberrant methylation patterns that result in cancer.

Data, Materials, and Software Availability. The python functions used as the basis for our theoretical analyses are available on the Spakowitz group website (96). All other data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. Financial support for A.Y.T. and A.J.S. is provided by the NSF, Physics of Living Systems Program (PHY-2102726).

Author affiliations: [a]Department of Materials Science, Stanford University, Stanford, CA 94305; and [b]Department of Chemical Engineering, Stanford University, Stanford, CA 94305

1. A. L. Olins, D. E. Olins, Spheroid chromatin units (v bodies). *Science* **183**, 330–332 (1974).
2. R. D. Kornberg, Chromatin structure: A repeating unit of histones and DNA. *Science* **184**, 868–871 (1974).
3. S. S. Rao *et al*., A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
4. S. Schoenfelder, P. Fraser, Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
5. G. Fudenberg *et al*., Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
6. J. R. Dixon *et al*., Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
7. B. Bonev, G. Cavalli, Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016).
8. J. Dekker, L. Mirny, The 3D genome as moderator of chromosomal communication. *Cell* **164**, 1110–1121 (2016).
9. M. Curradi, A. Izzo, G. Badaracco, N. Landsberger, Molecular mechanisms of gene silencing mediated by DNA methylation. *Mol. Cell. Biol.* **22**, 3157–3173 (2002).
10. A. Jeltsch, Beyond watson and crick: Dna methylation and molecular enzymology of dna methyltransferases. *Chembiochem* **3**, 274–293 (2002).
11. H. Liu *et al*., DNA methylation dynamics: Identification and functional annotation. *Brief. Funct. Genomics* **15**, 470–484 (2016).
12. A. R. Elhamamsy, DNA methylation dynamics in plants and mammals: Overview of regulation and dysregulation. *Cell Biochem. Funct.* **34**, 289–298 (2016).
13. J. A. Hackett, M. A. Surani, DNA methylation dynamics during the mammalian life cycle. *Phil. Trans. R Soc. B: Biol. Sci.* **368**, 20110328 (2013).
14. A. Bird, DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
15. R. J. Klose, A. P. Bird, Genomic DNA methylation: The mark and its mediators. *Trends Biochem. Sci.* **31**, 89–97 (2006).
16. J. O. Haerter, C. Lövkvist, I. B. Dodd, K. Sneppen, Collaboration between CPG sites is needed for stable somatic inheritance of DNA methylation states. *Nucleic Acids Res.* **42**, 2235–2244 (2014).
17. Y. Song, H. Ren, J. Lei, Collaborations between CPG sites in DNA methylation. *Int. J. Mod. Phys. B* **31**, 1750243 (2017).
18. S. Rao *et al*., Systematic prediction of DNA shape changes due to CPG methylation explains epigenetic effects on protein-DNA binding. *Epigenet. Chromatin* **11**, 1–11 (2018).
19. M. Okano, D. W. Bell, D. A. Haber, E. Li, DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257 (1999).
20. K. E. Bachman, M. R. Rountree, S. B. Baylin, Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. *J. Biol. Chem.* **276**, 32282–32287 (2001).
21. Z. D. Smith, A. Meissner, Roles in mammalian development. DNA methylation. *Nat. Rev. Genet.* **14**, 204–220 (2013).
22. M. F. Robert *et al*., DNMT1 is required to maintain CPG methylation and aberrant gene silencing in human cancer cells. *Nat. Genet.* **33**, 61–65 (2003).
23. S. Feng, S. E. Jacobsen, W. Reik, Epigenetic reprogramming in plant and animal development. *Science* **330**, 622–627 (2010).
24. R. Goyal, R. Reinhardt, A. Jeltsch, Accuracy of DNA methylation pattern preservation by the DNMT1 methyltransferase. *Nucleic Acids Res.* **34**, 1182–1188 (2006).
25. N. Wanner *et al*., DNA methyltransferase 1 controls nephron progenitor cell renewal and differentiation. *J. Am. Soc. Nephrol.* **30**, 63–78 (2019).
26. S. Adam *et al*., DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. *Nat. Commun.* **11**, 3723 (2020).
27. S. B. Baylin, P. A. Jones, A decade of exploring the cancer epigenome-biological and translational implications. *Nat. Rev. Genet.* **11**, 726–734 (2011).
28. P. A. Jones, G. Liang, Rethinking how DNA methylation patterns are maintained. *Nat. Rev. Genet.* **10**, 805–811 (2009).
29. S. Saghafinia, M. Mina, N. Riggi, D. Hanahan, G. Ciriello, Pan-cancer landscape of aberrant DNA methylation across human tumors. *Cell Rep.* **25**, 1066–1080 (2018).
30. R. Lakshminarasimhan, G. Liang, The role of DNA methylation in cancer. *Adv. Exp. Med. Biol.* **945**, 151–172 (2016).
31. E. Aref-Eshghi *et al*., Genomic DNA methylation signatures enable concurrent diagnosis and clinical genetic variant classification in neurodevelopmental syndromes. *Am. J. Hum. Genet.* **102**, 156–174 (2018).
32. A. A. Johnson *et al*., The role of DNA methylation in aging, rejuvenation, and age-related disease. *Rejuvenation Res.* **15**, 483–494 (2012).
33. G. Landan *et al*., Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat. Genet.* **44**, 1207–1214 (2012).
34. E. Li, T. H. Bestor, R. Jaenisch, Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
35. M. Željko, N. O. Svedružić, Reich,, Mechanism of allosteric regulation of DNMT1's processivity. *Biochemistry* **44**, 14977–14988 (2005).
36. A. Hermann, R. Goyal, A. Jeltsch, The DNMT1 DNA-(cytosine-c5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J. Biol. Chem.* **279**, 48350–48359 (2004).
37. J. Arand *et al*., In vivo control of CPG and non-CPG DNA methylation by DNA methyltransferases. *PLoS Genet.* **8**, 1–11 (2012).
38. L. Busto-Moner *et al*., Stochastic modeling reveals kinetic heterogeneity in post-replication DNA methylation. *PLoS Comput. Biol.* **16**, 1–23 (2020).
39. H. Ren, R. B. Taylor, T. L. Downing, E. L. Read, Locally correlated kinetics of post-replication DNA methylation reveals processivity and region specificity in DNA methylation maintenance. *J. R. Soc. Interface* **19**, 1–16 (2022).
40. L. Zhang *et al*., DNA methylation landscape reflects the spatial organization of chromatin in different cells. *Biophys. J.* **113**, 1395–1404 (2017).
41. K. Utsey, J. P. Keener, A mathematical model for inheritance of DNA methylation patterns in somatic cells. *Bull. Math. Biol.* **82**, 1–23 (2020).
42. P. A. Jones, S. M. Taylor, Cellular differentiation, cytidine analogs and DNA methylation. *Cell* **20**, 85–93 (1980).
43. K. Huang, G. Fan, DNA methylation in cell differentiation and reprogramming: An emerging systematic view. *Regen. Med.* **5**, 531–544 (2010).
44. Y. Tadokoro, H. Ema, M. Okano, E. Li, H. Nakauchi, De novo DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells. *J. Exp. Med.* **204**, 715–722 (2007).
45. C. Bock *et al*., DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Mol. Cell* **47**, 633–647 (2012).
46. M. Farlik *et al*., DNA methylation dynamics of human hematopoietic stem cell differentiation. *Cell Stem Cell* **19**, 808–822 (2016).
47. L. T. Kaaij *et al*., DNA methylation dynamics during intestinal stem cell differentiation reveals enhancers driving gene expression in the villus. *Genome Biol.* **14**, 1–15 (2013).
48. S. Banerjee, M. Bacanamwo, DNA methyltransferase inhibition induces mouse embryonic stem cell differentiation into endothelial cells. *Exp. Cell Res.* **316**, 172–180 (2010).
49. S. Luo *et al*., Folic acid acts through DNA methyltransferases to induce the differentiation of neural stem cells into neurons. *Cell Biochem. Biophys.* **66**, 559–566 (2013).
50. M. A. De La Díaz Rosa, E. F. Koslover, P. J. Mulligan, A. J. Spakowitz, Dynamic strategies for target-site search by DNA-binding proteins. *Biophys. J.* **98**, 2943–2953 (2010).
51. E. F. Koslover, M. A. D. D. L. Rosa, A. J. Spakowitz, Theoretical and computational modeling of target-site search kinetics in vitro and in vivo. *Biophys. J.* **101**, 856–865 (2011).
52. E. F. Koslover, M. D. de la Rosa, A. J. Spakowitz, Crowding and hopping in a protein's diffusive transport on DNA. *J. Phys. A: Math. Theor.* **50**, 074005 (2017).
53. L. Zagkos, M. M. Auley, J. Roberts, N. I. Kavallaris, Mathematical models of DNA methylation dynamics: Implications for health and ageing. *J. Theor. Biol.* **462**, 184–193 (2019).
54. A. P. McGovern, B. E. Powell, T. J. Chevassut, A dynamic multi-compartmental model of DNA methylation with demonstrable predictive value in hematological malignancies. *J. Theor. Biol.* **310**, 14–20 (2012).
55. L. B. Sontag, M. C. Lorincz, E. G. Luebeck, Dynamics, stability and inheritance of somatic DNA methylation imprints. *J. Theor. Biol.* **242**, 890–899 (2006).
56. O. G. Berg, R. B. Winter, P. H. von Hippel, Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948 (1981).
57. O. G. Berg, Diffusion-controlled protein-DNA association: Influence of segemental diffusion of the DNA. *Biopolymers* **23**, 1869–1889 (1984).
58. O. G. Berg, P. H. von Hippel, Diffusion-controlled macromolecular interactions. *Ann. Rev. Biophy. Biophys. Chem.* **14**, 131–158 (1985).
59. Y. Meroz, I. Eliazar, J. Klafter, Facilitated diffusion in a crowded environment: From kinetics to stochastics. *J. Phys. A: Math. Theor.* **42**, 434012 (2009).
60. K. V. Klenin, H. Merlitz, J. Langowski, C. X. Wu, Facilitated diffusion of DNA-binding proteins. *Phys. Rev. Lett.* **96**, 018104 (2006).
61. M. Bauer, R. Metzler, Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophys. J.* **102**, 2321–2330 (2012).
62. M. Coppey, O. Bénichou, R. Voituriez, M. Moreau, Kinetics of target localization of a protein on DNA: A stochastic approach. *Biophys. J.* **87**, 1640–1649 (2004).
63. J. Gorman, E. C. Greene, Visualizing one-dimensional diffusion of proteins along DNA. *Nature Struct. Mol. Biol.* **15**, 768–774 (2008).
64. C. Loverdo *et al*., Quantifying hopping and jumping in facilitated diffusion of DNA-binding proteins. *Phys. Rev. Lett.* **102**, 188101 (2009).
65. L. Mirny *et al*., How a protein searches for its site on DNA: The mechanism of facilitated diffusion. *J. Phys. A: Math. Theor.* **42**, 434013 (2009).
66. F. Valle, M. Favre, P. D. L. Rios, A. Rosa, G. Dietler, Scaling exponents and probability distributions of DNA end-to-end distance. *Phys. Rev. Lett.* **95**, 158105 (2005).
67. A. Amitai, D. Holcman, Polymer physics of nuclear organization and function. *Phys. Rep.* **678**, 1–83 (2017).
68. A. Bancaud, C. Lavelle, S. Huet, J. Ellenberg, A fractal model for nuclear organization: Current evidence and biological implications. *Nucleic Acids Res.* **40**, 8783–8792 (2012).
69. M. D. Pierro, B. Zhang, E. L. Aiden, P. G. Wolynes, J. N. Onuchic, Transferable model for chromosome architecture. *Proc. Natl. Acad. Sci.* **113**, 12168–12173 (2016).
70. L. A. Mirny, The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.* **19**, 37–51 (2011).
71. B. Chan, M. Rubinstein, Activity-driven chromatin organization during interphase: Compaction, segregation, and entanglement suppression. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2401494121 (2024).
72. G. Vilkaitis, I. Suetake, S. Klimašauskas, S. Tajima, Processive methylation of hemimethylated CPG sites by mouse DNMT1 DNA methyltransferase. *J. Biol. Chem.* **280**, 64–72 (2005).
73. A. J. Pollak, N. O. Reich, DNA adenine methyltransferase facilitated diffusion is enhanced by protein-DNA "roadblock" complexes that induce dna looping. *Biochemistry* **54**, 2181–2192 (2015).
74. M. A. Surby, N. O. Reich, Facilitated diffusion of the *Eco*RI DNA methyltransferase is described by a novel mechanism. *Biochemistry* **35**, 2209–2217 (1996).
75. T. E. Kuhlman, E. C. Cox, Gene location and DNA density determine transcription factor distributions in *Escherichia coli*. *Mol. Syst. Biol.* **8**, 610 (2012).
76. B. Jin, Y. Li, K. D. Robertson, DNA methylation: Superior or subordinate in the epigenetic hierarchy? *Genes Cancer* **2**, 607–617 (2011).
77. M. Gardiner-Garden, M. Frommer, CPG islands in vertebrate genomes. *J. Mol. Biol.* **196**, 261–282 (1987).
78. A. M. Deaton, A. Bird, CPG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).
79. S. Saxonov, P. Berg, D. L. Brutlag, A genome-wide analysis of CPG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1412–1417 (2006).
80. Y. Cui, J. Irudayaraj, Dissecting the behavior and function of MBD3 in DNA methylation homeostasis by single-molecule spectroscopy and microscopy. *Nucleic Acids Res.* **43**, 3046–3055 (2015).
81. A. Nishiyama *et al*., Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation. *Nat. Commun.* **11**, 1222 (2020).
82. X. Liu *et al*., UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat. Commun.* **4**, 1563 (2013).

83. C. Bronner, M. Alhosin, A. Hamiche, M. Mousli, Coordinated dialogue between UHRF1 and DNMT1 to ensure faithful inheritance of methylated DNA patterns. *Genes* **10**, 65 (2019).
84. P. O. Estève *et al*., Direct interaction between DNMT1 and G9A coordinates DNA and histone methylation during replication. *Genes Dev.* **20**, 3089–3103 (2006).
85. K. Y. Kim *et al*., UHRF1 regulates active transcriptional marks at bivalent domains in pluripotent stem cells through Setd1a. *Nat. Commun.* **9**, 2583 (2018).
86. H. Kimura, K. Shiota, Methyl-CPG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, DNMT1. *J. Biol. Chem.* **278**, 4806–4812 (2003).
87. B. C. Wang *et al*., UHRF1 suppression promotes cell differentiation and reduces inflammatory reaction in anaplastic thyroid cancer. *Oncotarget* **9**, 31945 (2018).
88. H. Sakai *et al*., UHRF1 governs the proliferation and differentiation of muscle satellite cells. *iScience* **25**, 103928 (2022).
89. C. Haggerty *et al*., DNMT1 has de novo activity targeted to transposable elements. *Nat. Struct. Mol. Biol.* **28**, 594–603 (2021).
90. W. Timp *et al*., Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. *Genome Med.* **6**, 1–11 (2014).
91. B. P. Berman *et al*., Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear laminag-associated domains. *Nat. Genet.* **44**, 40–46 (2012).
92. N. Besselink *et al*., The genome-wide mutational consequences of DNA hypomethylation. *Sci. Rep.* **13**, 6874 (2023).
93. D. J. Smiraglia *et al*., Excessive CPG island hypermethylation in cancer cell lines versus primary human malignancies. *Hum. Mol. Genet.* **10**, 1413–1419 (2001).
94. J. Roman-Gomez *et al*., Promoter hypermethylation of cancer-related genes: A strong independent prognostic factor in acute lymphoblastic leukemia. *Blood* **104**, 2492–2498 (2004).
95. D. Sproul *et al*., Tissue of origin determines cancer-associated CPG island promoter hypermethylation patterns. *Genome Biol.* **13**, 1–16 (2012).
96. A. Tse, Theoretical model of DNA methyltransferase function to predict epigenetic correlation patterns. Spakowitz Research Group. https://web.stanford.edu/~ajspakow/data_and_code/data_and_code.html. Deposited 15 December 2024.