Published online 29 February 2016 *Nucleic Acids Research, 2016, Vol. 44, No. 11* **5123–5132**
doi: 10.1093/nar/gkw124

DNA methylation in human epigenomes depends on local topology of CpG sites

Cecilia Lövkvist[1], Ian B. Dodd[2], Kim Sneppen[1,*] and Jan O. Haerter[1,*]

[1]Center for Models of Life, Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, DK-2100, Copenhagen, Denmark and [2]Department of Molecular and Cellular Biology, University of Adelaide, SA 5005, Australia

Received August 19, 2015; Revised February 17, 2016; Accepted February 20, 2016

ABSTRACT

In vertebrates, methylation of cytosine at CpG sequences is implicated in stable and heritable patterns of gene expression. The classical model for inheritance, in which individual CpG sites are independent, provides no explanation for the observed non-random patterns of methylation. We first investigate the exact topology of CpG clustering in the human genome associated to CpG islands. Then, by pooling genomic CpG clusters on the basis of short distances between CpGs within and long distances outside clusters, we show a strong dependence of methylation on the number and density of CpG organization. CpG clusters with fewer, or less densely spaced, CpGs are predominantly hypermethylated, while larger clusters are predominantly hypo-methylated. Intermediate clusters, however, are either hyper- or hypo-methylated but are rarely found in intermediate methylation states. We develop a model for spatially-dependent collaboration between CpGs, where methylated CpGs recruit methylation enzymes that can act on CpGs over an extended local region, while unmethylated CpGs recruit demethylation enzymes that act more strongly on nearby CpGs. This model can reproduce the effects of CpG clustering on methylation and produces stable and heritable alternative methylation states of CpG clusters, thus providing a coherent model for methylation inheritance and methylation patterning.

INTRODUCTION

Cytosine methylation in vertebrates occurs predominantly at CG dinucleotide sequences (1), termed CpG sites. The intense experimental interest in this modification is due to its potential to provide epigenetic regulation of gene expression (2,3). To qualify as an epigenetic mark, the CpG methylation state needs to be stable and heritable through cell division. The symmetry of the CpG sequence has served as the basis for a simple model where the methylation state of a single CpG can be inherited without dependence on the state of neighboring DNA (4,5). During replication, DNA polymerase inserts non-methylated cytosines, copying an unmethylated CpG site to unmethylated sites on the two daughter strands, and copying a fully methylated CpG site into two hemimethylated sites. The fully methylated state is then re-established by efficient recognition of these hemimethylated sites by DNA methyltransferases (DNMTs).

However, a number of observations indicate that this 'classical' model is now untenable (6–8). First, the model requires a high fidelity of methylation of hemimethylated sites as well as non-methylation of unmethylated sites, features that are not matched by the activity of DNMTs *in vitro* (8) or *in vivo* (9) and are compromised by active removal of methyl groups by demethylation pathways (8,10–15). Indeed, the frequencies of hemimethylated CpG sites observed *in vivo* by hairpin bisulfite polymerase chain reaction (16) indicate high error rates for individual CpG sites. Second, CpG sites display group behavior that is not predicted from a model where CpG sites are independent. Measurement of methylation patterns among clusters of CpG sites *in vivo* reveal bimodality of methylation—different clusters tend to be either hyper- methylated or hypo-methylated, infrequently existing in intermediate methylation states (17–20). Bimodal methylation is often displayed by the same CpG cluster, with the cluster being in distinct methylation states in different cells, or even in different alleles in the same cell (17).

An alternative class of model is to assume that CpGs are not independent, rather that the methylation of a given CpG site is affected by the methylation of the surrounding CpG sites. We have proposed a model where methylated and hemimethylated CpG sites recruit DNMTs, and unmethylated CpGs recruit demethylases, with the recruited enzymes acting on CpG sites in the vicinity (7). Simulations show that this positive feedback could allow CpG sites to collaborate to dynamically maintain either an overall hyper- or hypo-methylated state of a cluster. This bimodal methy-

*To whom correspondence should be addressed. Tel: +45 353 25352; Fax: +45 353 25425; Email: sneppen@nbi.dk
Correspondence may also be addressed to Jan O. Haerter. Tel: +45 353 33519; Email: haerter@nbi.dk

© The Author(s) 2016. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

lation arises naturally as a result of the inherent bistability of the system. Importantly, the hyper- or hypo-methylated state of a CpG cluster could each be robustly inherited over many cell generations, even in the presence of high error rates.

The availability of genome-wide methylation mapping, for example whole genome bisulfite sequencing (21,22), allows examination of how CpG collaboration could operate on a genomic scale. The 28 million CpG sites in the human genome are predominantly methylated and occur at low frequencies (on average 1/100 bp) across the genome (16,23). Strong interest has however been drawn by the methylation patterns of comparably dense regions of CpG sites. These regions, termed CpG-islands (CGIs) (24), have traditionally been considered to be largely unmethylated (24–26), but more recent evidence is supportive of a picture where CpG islands can also be in predominantly methylated states (1,17,27). Some of the interest in CGIs stems from the association of their methylation patterns with promoter activity (28–30). Common to a range of definitions and descriptions of CGIs, the density of CpG content (24,31) is the crucial parameter used to identify CGIs. Overall, the level of methylation has been considered to be anti-correlated with CpG density (17,32).

Effects of CpG topology on methylation are a natural corollary of collaborative models, since they propose that the methylation status of a CpG site is dependent on the methylation status of nearby CpGs. To understand how the topology of CpG sites affects their methylation, we systematically analyzed the clustering of CpG sites in the human genome, finding that a large fraction of the CpGs can be defined as existing in isolated 'clusters' of 1–60 sites with inter-CpG distances <25 bp and separated by at least 65 bp from surrounding CpG sites. Examining the methylation status of these and other clusters in four human methylomes, we find the expected bimodal methylation pattern, where clusters were either hypo- or hyper-methylated. We also saw a strong trend where the probability of hypo-methylation increases with increasing number and density of CpGs in the cluster. We show that these geometric effects on methylation can be reproduced by a modified collaborative model, in which the efficiencies of the recruitment-based methylation and demethylation reactions decay differently with increasing separation between CpGs. Our work suggests that ubiquitous collaborative interactions between CpGs could provide much of the patterning of genomic methylation and would allow clusters of moderate size to exist stably in heritable alternative methylation states to support epigenetic gene regulation.

MATERIALS AND METHODS

Distances and positions of CpGs were analyzed for the human genome (hg18, downloaded from http://genome.ucsc.edu/) (33). $d = 2$ bp for adjacent CpGs. IMR90 methylome data were from http://neomorph.salk.edu/human_methylome/data.html (IMR90 C basecalls) (21), and brain tissue methylome data (22) were from http://www.ncbi.nlm.nih.gov/geo GEO accessions: GSM1163695 fetal frontal cortex, GSM1164630 and GSM1164632 middle frontal gyrus from 12 and 25 year old males. Data for CpGs with coverage of at least 10 was used for methylation averages, except for Figure 3B.

We simulate a CpG cluster including its surroundings using a collaborative distance-dependent model (Figure 3A). In the limit of an infinite number of CpG sites and assuming that each CpG site interacts equally with any other CpG site (mean-field assumption) the equations describing the fraction of CpG sites in u (unmethylated), h (hemimethylated) and m (methylated), are:

$$h = 1 - m - u \tag{1}$$

$$\frac{du}{dt} = \mu \cdot h - \beta \cdot u + \kappa_2 \cdot u \cdot h - \sigma_1 \cdot m \cdot u \tag{2}$$

$$\frac{dm}{dt} = \beta \cdot h - \mu \cdot m - \kappa_1 \cdot u \cdot m + \sigma_3 \cdot h^2 + \sigma_2 \cdot h \cdot m. \tag{3}$$

Using the parameters $\{\beta = 0.005, \mu = 0.01, \sigma_1 = 0.2, \sigma_2 = 0.8, \sigma_3 = 0.8, \kappa_1 = 0.8, \kappa_2 = 0.8\}$ (Figure 3A) the stable steady states for Equations (1)– (3) are $\{u = 0.0007, h = 0.0129, m = 0.9864\}$ and $\{u = 0.99373, h = 0.00619, m = 0.00008\}$. We simulate a CpG cluster of N_C CpG sites with CpG-CpG distances of d and a distance of D (varying values in Figure 4) to the surrounding $N_{\text{out}} = 200$ CpG sites (100 CpG sites on each side of the cluster). The CpG–CpG distances between any two neighboring CpGs in the surroundings are $D^* = 100$ bp. The system is initialized with a random methylation pattern, i.e. each site has equal probability to be in either of the three states u, h or m. We use a standard Gillespie algorithm to update the state of the CpG sites according to the nine different reactions (Figure 3A). First, a reaction is chosen according the standard Gillespie step and a target CpG site is chosen and random. If the reaction is collaborative, a recruiting CpG site is also chosen. The probability of choosing a specific recruiting CpG site is dependent on its distance from the target site. For the collaborative demethylation reactions the probability is calculated from an exponential probability distribution, $b \cdot exp(-d/d_0))$ where d is the distance between the two CpG sites (Figure 3C), $d_0 = 174$ bp and $b = 5.525$. For the methylation reactions, the probability for the recruiting sites is calculated from a power law probability distribution, $(a/(d + \alpha))$, $\alpha = 196$ bp and $a = 650$ bp. A cell generation in the simulations consists of on average 0.5 reaction attempts of the reaction μ per CpG site. In the end of each generation all CpG sites are replicated. All sites in m are then converted to h, all in h to u or h with equal probability 0.5 and all sites in u remain in u. The status of the system is recorded before each replication event. The parameters above are used as rates in our simulations (Figure 4). As in our previous model (7), bistability of the cluster requires the collaborative methylation reactions (σ_1,σ_2,σ_3) to be strong relative to the non-collaborative 'noise' reaction (β,μ). The collaborative demethylation reactions (κ_1,κ_2), while not necessary for cluster bistability in the absence of outside CpGs (7), are needed for the cluster to maintain the hypo-methylated state in the face of methylation pressure from the surrounding hyper-methylated DNA. Slight reductions in the strength of the collaborative methylation reactions, or increases in the collaborative demethylation reactions, reduce the N^* of

the cluster, that is, smaller clusters were able to exist in the high u-state. Stronger reduction, respectively increases of these two reaction types (>10%) causes the CpGs inside and the surroundings to become stably hypo-methylated. Conversely, increasing the strength of the methylation reactions, or decreasing the strength of the demethylation reactions, increased the N^*, with strong increases (>10%) causing a loss of bistability. For a cluster of size $N_C = 28$ and $d = 10$ bp and $D = 65$ bp an increase/decrease of each parameter by 10%, while at the same time keeping the others fixed, gives the following relative change in the methylation average of the cluster (for $N_C = 28$ the methylation average is 0.47):

change	β	μ	σ_1	σ_2	σ_3	κ_1	κ_2
+10%	9.7%	−39%	85%	79%	32%	−58%	−68%
−10%	−9.7%	36%	−73%	−68%	−12%	72%	94%

Alternations in the distance parameters (a, b, d_0 and α) affect how the inside and outside CpG densities and cluster sizes control the inside and outside methylation status. Increasing α to $\alpha \approx 1000$ bp makes the demethylation reactions stronger and smaller islands become unmethylated, i.e. no N^* would be found. Decreasing α to $\alpha \approx 100$ bp makes larger islands more methylated. Increasing d_0 to $d_0 \approx 300$ bp leads to unmethylated small islands and thereby no N^* is found. Decreasing d_0 to $d_0 \approx 120$ bp leads to methylated islands where the demethylation reactions are weaker than the methylation reactions. With low a ($a \approx 100$ bp) the demethylation reactions are stronger and the islands are predominantly unmethylated independent of cluster size. The opposite is observed for higher a ($a \approx 750$ bp). Methylated islands dominate when b is decreased to $b \approx 4$ and unmethylated islands dominate when b is increased to $b \approx 11$. Generally, the transition from methylated small islands to unmethylated large islands is lost when the parameters are perturbed and consequently no N^* is found.

RESULTS

Clustering of CpG sites in the human genome

Systematic analyses of the distribution of CpG sites within vertebrate genomes have shown a highly non-random pattern, with the frequencies of short and long distances between CpG sites enhanced at the cost of intermediate distances (34,35). This is shown in Figure 1A and B, which compares the frequencies of observed CpG–CpG distances in the human genome (33) with that expected from a random arrangement of the same number of CpG sites. The distribution of the null model (Figure 1A) approaches an exponential distribution and there is a small peak at distances close to 10 bp, a distance that is observed in dense regions of CpG sites (36). However, such analyses only partially capture the clustering of CpGs because they do not address higher order clustering due to correlations between neighboring CpG–CpG distances.

We thus counted the occurrences of each possible combination of successive CpG distances (i.e. CpG-d_1-CpG-d_2-CpG, where d_1 and d_2 denote distances between the CpG sites) in the human genome and compared these to the case where all observed CpG–CpG distances are maintained but are randomly arranged (Figure 1C). This randomization leaves the observed frequencies of distances intact while removing correlations between neighboring distances. Plotting the ratio between the observed d_1–d_2 counts and those in the randomized genome (Figure 1D) shows that short-short and long-long distance combinations are strongly enhanced, while short-long and long-short combinations are under-represented. The enhanced regions in Figure 1D set natural scales for CpG clusters; considering the lines of unit ratio, clustering of the distances occurs for distances less than ∼25 bp and for distances greater than ∼65 bp.

Accordingly, genomic CpGs can be captured by a definition of a CpG cluster that requires every pair of neighboring CpG sites in the cluster to be separated by a distance shorter than a threshold $d_{\max} = 25$ bp, and the terminal CpG sites of the cluster to be separated by a distance larger than a threshold $D_{\min} = 65$ bp from both flanking CpG sites (Figure 1E). (Note that a single CpG that is $> D_{\min}$ from both neighboring CpGs is scored as a 'cluster' of 1). This definition includes ∼30% of the CpG sites in the human genome as existing in cluster sizes N_C ranging from 1 to 60 CpG sites, with the majority of the CpG clusters in the N_C range 1–11 (Figure 1F).

Plotting the average distance as a function of CpG position in and around the pooled $d_{\max} = 25$ bp, $D_{\min} = 65$ bp clusters of size $N_C = 15$ (Figure 1G) shows that a 'boundary' of 65 bp around the clusters causes them to be surrounded by typical CpG densities, since the average inter-CpG distances $\langle d \rangle$ around the cluster immediately return to close to the genomic average. Thus, on average, these clusters are not strongly associated with other clusters. In contrast, the larger set of clusters defined by use of a smaller boundary $D_{\min} = 45$ bp tend to be surrounded by regions of higher CpG density, indicating that this definition includes many clusters that are nearby other clusters (Figure 1H).

Effect of CpG clustering on methylation

We examined the methylation of the $d_{\max} = 25$ bp, $D_{\min} = 65$ CpG clusters in four human methylomes obtained by whole genome bisulphite sequencing of a fetal lung fibroblast cell line (IMR90), and fetal, juvenile and adult brain cell samples (21,22). Thus each CpG cluster was represented four times. We used the average methylation values, ranging from 0 to 1, for individual CpGs that had been covered at least 10 times within each methylome dataset (∼22 million CpGs of the 28 million in hg18).

The mean methylation of each cluster, calculated as the average of the methylation fractions of each CpG in the cluster, was strongly dependent on the number of CpGs in the cluster, N_C. The distributions of mean cluster methylation in Figure 2A display a strong bimodal pattern, with clusters either hyper- or hypo-methylated but rarely in intermediate methylation states. However, clusters containing few CpGs are almost invariably highly methylated, while clusters with increasing numbers of CpGs become increasingly likely to be hypo-methylated (Figure 2A). Thus, 'lone' CpGs, which occupy the largest fraction of the genome (Figure 1F), are predominantly hyper-methylated, while very large clusters are predominantly hypo-methylated. Importantly, there is no clear

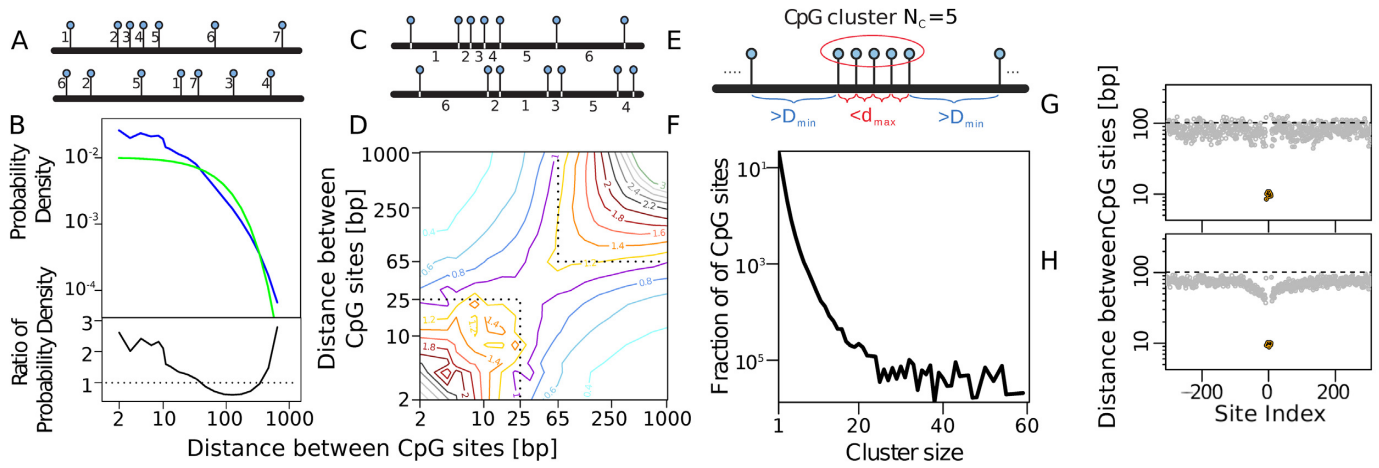

Figure 1. Distances between CpG sites in the human genome. (**A**) Schematic of randomization of CpG positions used to produce an equal number of CpG sites but remove all spatial correlations between the CpG positions. The position of each of the 28 million CpGs in the genome was randomly assigned a new position (avoiding overlapping of CpG sites) within a 'blank genome' of 28 billion positions. (**B**) The observed CpG–CpG distance frequencies for the data (blue), and after CpG randomization (green). The standard errors of the mean for 12 separate genome randomizations lie within the thickness of the green line. The lower panel shows the ratio between the real and randomized distance frequencies. (**C**) Schematic of randomization of distances between CpG sites, keeping each individual distance unchanged but removing the correlation between distances, i.e. the distances are preserved. Effectively, an array of the 28 million genomic CpG–CpG distances was shuffled to produce a random sequence of these distances. (**D**) Frequencies of distances (d_1) and subsequent distances (d_2) are divided by the corresponding frequencies after distance randomization, showing enhancement of short-short and long-long distance combinations. (**E**) Schematic of CpG cluster criteria. (**F**) Distribution of cluster sizes N_C in the genome for $d_{max} = 25$ bp and $D_{min} = 65$ bp. (**G**) The genome contains 21 000 clusters of size $N_C = 15$ with $d_{max} = 25$ bp and $D_{min} = 65$ bp. In the plot, the point at site index = 1 is the distance between the central CpG (site index = 0) and the first CpG to the right (site index = 1) averaged across all clusters. The point at site index = 2 is the average of the distances between the first and second CpGs on the right, and so on. Average successive CpG–CpG distances going leftward from the central CpG are given by negative site indices. Black points show average distances between CpG sites within the cluster, with the average distances outside the cluster shown in gray. (**H**) As (G) but for $D_{min} = 45$ bp. Note the correlation between CpG distances surrounding the island. Note the logarithmic vertical axes in (F, G and H) and the double logarithmic axes in (B, top) and (D).

demarcation between high methylation-favoring and low methylation-favoring regimes, suggesting that current criteria for defining CpG islands are somewhat arbitrary.

To check that these effects are not particular to our choice of $d_{max} = 25$ bp, $D_{min} = 65$ bp, we tested clusters with various d_{max} and D_{min} combinations. We kept $D_{min} > d_{max}$ so that all clusters are set within lower density regions. However, we note that low D_{min} values mean that it becomes more likely that the cluster is nearby other clusters (Figure 1H). The effect of N_C was measured for each D_{min}/d_{max} combination by determining N^*, the N_C at which the average methylation of the clusters crosses 0.5 (e.g. for the $d_{max} = 25$ bp, $D_{min} = 65$ bp clusters, $N^* = 29$, Figure 2A). In all cases, the methylation versus N_C trend was the same, with methylation favored when $N_C < N^*$ and unmethylation favored when $N_C > N^*$, as shown for $d_{max} = 25$ bp, $D_{min} = 45$ bp (Figure 2B). Plotting the N^* values against average d, $\langle d \rangle$, for each D_{min}/d_{max} combination shows a CpG density effect; decreasing average distances between CpGs give lower N^* values i.e. clusters of fewer CpGs are able to exist in an unmethylated state if they are more dense (Figure 2C). Thus, the points in Figure 2C define a transition between a lower CpG number/lower CpG density regime where hyper-methylation is favored (lower right), and a higher CpG number/higher CpG density regime where hypo-methylation is favored (upper left). We note that the actual change in methylation preference across this transition region is gradual. Interestingly, N^* only weakly increases with D_{min}.

Dynamical model for spatial collaboration

The observed strong bimodality of cluster methylation is a natural feature of the bistability that can result when collaboration involves positive feedback, that is, when methylated CpGs foster methylation of nearby CpGs and unmethylated CpGs foster demethylation of nearby CpGs. Some effect of CpG number and density on cluster methylation is also expected because of the interactions between nearby CpGs. However, we wanted to test whether the asymmetry of the effect of CpG number and density, with hypo-methylation favored in larger, denser clusters, could also be explained by a collaborative model.

Our previous model (7) invoked a number of methylation and demethylation reactions that interconvert fully methylated (m), hemimethylated (h) and unmethylated (u) CpG sites (Figure 3A). Interconversions can be non-collaborative, that is occur independently of other CpGs (black and gray arrows, Figure 3A) or collaborative, where the particular reaction at a target CpG involves a nearby mediator CpG in a particular methylation state (curved arrows, Figure 3A). For example, the methylation of a hemimethylated CpG could depend on the presence of a nearby fully methylated CpG (dark red arrow, Figure 3A). The most robust heritable bistability was obtained with the positive feedback collaborative reactions shown in Figure 3A, where m and h sites act to foster methylation of nearby u and h sites (maintaining the hyper-methylated state), and u sites act to foster demethylation of nearby h and m sites (maintaining the hypo-methylated state) (7). However, this

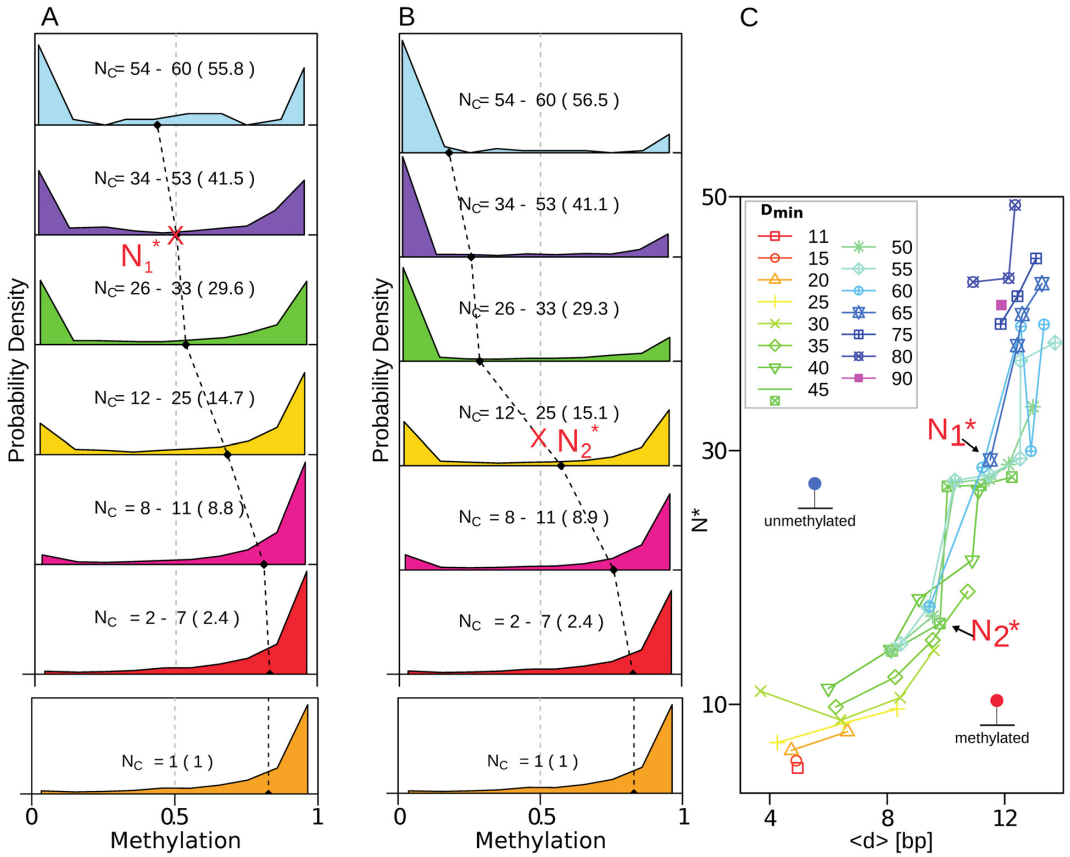

Figure 2. Empirical CpG distance and methylation distributions. (**A**) Distributions of average methylation of clusters sized $1 \leq N_C \leq 60$ with $d_{max} = 25$ bp and $D_{min} = 65$ bp. Panels show different N_C ranges (with the mean N_C in parentheses). The black dashed line shows the average methylation of each distribution. (**B**) As (A) but $D_{min} = 45$ bp. (**C**) For each pool of clusters defined with specific values of D_{min} and d_{max}, there is a critical cluster size, N^*, at which the methylation distribution is maximally bimodal (e.g. N_1^* and N_2^*, mark the maximal bimodality obtained in (A and B)). For each cluster pool, N^* is plotted against the average inter-CpG distance $\langle d \rangle$ in that pool. Each line corresponds to a particular value of D_{min} (as indicated in the inset), with each point on the line derived from a cluster with a distinct d_{max} value.

basic model does not predict that CpG number or density should affect which state is favored.

A simple and plausible way to allow an effect of CpG site topology in this model is to introduce a distance scaling of the collaboration reactions, where the probability of interaction between a target CpG and an enzyme recruited to a mediator CpG is dependent on the DNA distance between the two CpG sites. The relationship between contact probability and distance on chromatin *in vivo* is poorly understood. Hi-C experiments show that at long distances (>1000 bp), relative contact probability between two sites in human DNA *in vivo* generally falls with increasing distance, roughly as $1/d$ (37). However, at shorter distances, contact can be sub-optimal because of the stiffness of DNA and the nature of its packaging. A study of FLP recombination in mouse cells found recombination frequency increased as d was increased from 74 to 200 bp, followed by a steady decrease in recombination as d was increased to 15 kb (38). This effect of short distances on reaction probability is likely to be different for different enzymes, as it depends on the flexibility of the protein and the steric requirements for the reaction. Thus, different collaborative reactions may have quite different sensitivities to the distance between the mediator and target CpGs.

The bias toward hyper-methylation for less dense CpG clusters (Figure 2A) suggests that collaborative methylation reactions generally act more efficiently than collaborative demethylation reactions over longer CpG–CpG distances. Conversely, the bias toward hypo-methylation for more dense CpG clusters suggests that collaborative demethylation reactions are favored at shorter CpG–CpG distances. Different ranges for these reactions are supported by analysis of CpG clusters surrounded by at least 2.4 kb of low CpG density on both sides (Figure 3B). Hyper-methylated clusters are associated with a large zone of increased methylation, while hypo-methylated clusters seem to have effects over only a small region. To implement these different distance sensitivities in the model, we chose two mathematically convenient probability density functions (Figure 3C). For the collaborative methylation reactions, the probability that a DNMT recruited by a mediator CpG converts a target CpG that is d bp away from the mediator, scaled as $1/(d + \alpha)$. Here, α is an offset that produces a less steep decrease of probability over distances of $d < \alpha$ but approaches a $1/d$ power law as $d \gg \alpha$. For the collaborative demethylation reactions, we used a simple exponential function, $\exp(-d/d_0)$, which provides a steeper decay of probability with distance that favors short-range collaboration

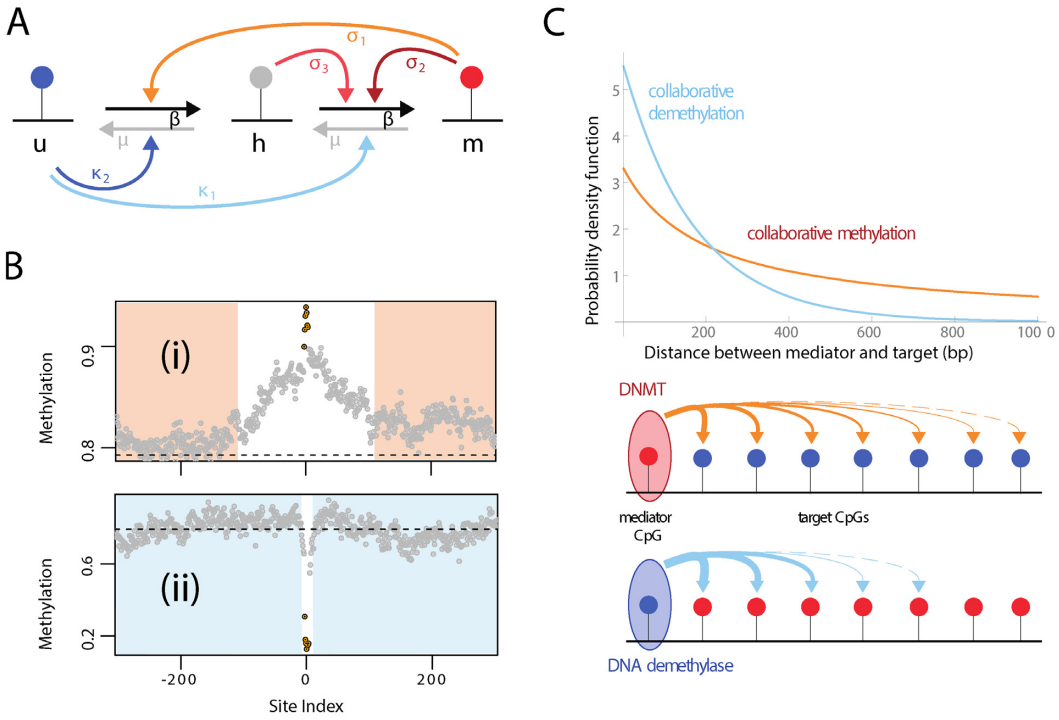

Figure 3. Model design. (**A**) Collaborative model (7). Straight arrows (gray and black) are non-collaborative reactions, curved arrows are collaborative reactions (start at mediator CpG, end at the reaction stimulated). See text. (**B**) Average CpG methylation for genomic regions containing CpG-clusters consisting of seven CpG sites with the average inter-CpG distance $\langle d \rangle < 12.5$ (black points) with a low density of surrounding CpG sites (gray points). Specifically, clusters were selected where 30 CpGs on each side of the cluster are spaced on average at least 80 bp apart. Clusters are sorted into those that are hyper-methylated (upper panel) or are hypo-methylated (lower panel). As in Figure 1G, site index is the ordinate position of the CpG relative to the central CpG of the cluster. (**C**) Introducing distance-dependent collaboration—short-range demethylation and long-range methylation. Plots show the reaction probability density function as a function of the distance between mediator and target in the new model. Methylation reactions have a power-law distance dependence $\sim a/(d + \alpha)$ with d the distance and $\alpha = 196$ bp; an offset. Demethylation reactions have an exponential distance dependence $\sim b \cdot \exp(d/d_0)$, with $d_0 = 174$ bp the range of the interaction. The parameters $a = 650$ bp and $b = 5.525$ are scaling factors.

(Figure 3). Here, d_0 is used to scale this distance sensitivity. We stress that it is unlikely that these functions accurately describe the d versus probability relationships for each of the collaborative reactions; they are used here simply to test the idea that the cluster size/density bias could be explained by a difference in distance sensitivity in the competing methylation/demethylation reactions.

We tested the behavior of the new model by simulating 20 kb DNA regions containing a single CpG cluster (a portion of such a region is shown in Figure 4A). In different simulations we varied the number of CpG sites in the cluster N_C and the distance d bp between the sites. The DNA surrounding the cluster contained CpG sites spaced 100 bp apart (the genomic average), except that the first CpG on each side of the cluster was a distance D bp from the cluster. As with our previous modeling, simulations involved iterating the five collaborative and four non-collaborative methylation and demethylation reactions (Figure 3A), randomly chosen according to defined reaction probabilities. For each reaction attempt, a target CpG and for the collaborative reactions also a mediator CpG, are randomly chosen. If the methylation status of these CpGs (u, h or m) is correct for the chosen reaction, then the target CpG is converted, otherwise the target is unchanged. However, in the new model, the collaborative reactions were also subjected to a distance test where the probability for the reaction to occur is determined from the distance between the concerning CpG sites. The probability for a methylation reaction is determined from a probability density function of a power law ($a/(d + \alpha)$), while a demethylation is determined from an exponential ($b \cdot exp(-d/d_0)$), where d is the distance between mediator and target CpGs, and a and b are scaling factors. Each generation comprised on average 100 reaction attempts per CpG, after which a replication event was simulated by making the replacements $m \rightarrow u$, $u \rightarrow u$ (unchanged) and $h \rightarrow u$ or $h \rightarrow h$ with equal probability. Simulations were carried out for 1000 generations. Parameters were adjusted to test if the systems could replicate the response of real clusters to CpG number and density (Figure 2).

Spatial collaboration can recapitulate genomic patterns

Figure 4A shows a system with $N_C = 23$, $d = 25$ bp and $D = 65$ bp where the cluster is bistable, able to exist stably and heritably in either a hyper- or hypo-methylated state, while the surrounding low density CpG region remains predominantly hyper-methylated. This overall pattern was attained whatever the initial state of the system, the simulation was begun with all CpGs in random states. Thus, in the model a single cluster can display the bimodality characteristic of real CpG clusters. Note that in the hyper-methylated state, a zone of mixed and rapidly varying methylation occurs in

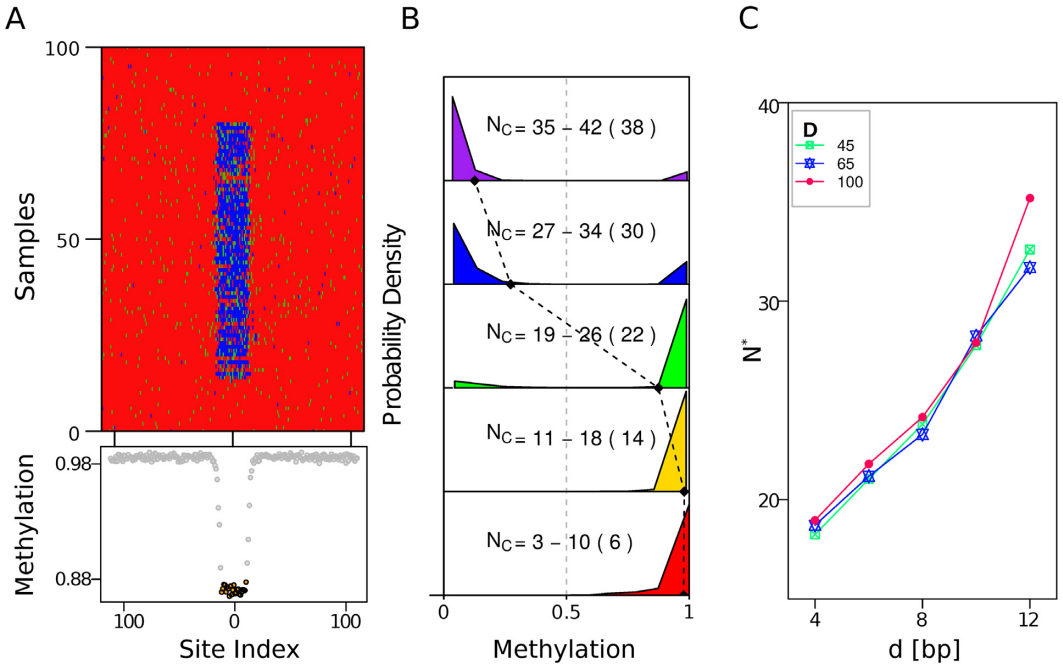

Figure 4. Model results. (**A**) Space-time plot for a bistable system of cluster size $N_C = 23$ (dense region in center of plot), $d = 10$ bp, $D = 65$ bp. m, h and u sites shown in red, green and blue. CpG sites within a cluster are spaced at a distance d, separated by D from the $N_{out} = 200$ CpG sites spaced $D^* = 100$ bp apart (the genomic average), with periodic boundary conditions (a ring of $N_C + N_{out}$ CpG sites). Nine different reactions (Figure 3A) with rates $\{\beta = 0.005, \mu = 0.01, \sigma_1 = 0.2, \sigma_2 = 0.8, \sigma_3 = 0.8, \kappa_1 = 0.8, \kappa_2 = 0.8\}$ were used in a standard Gillespie algorithm (see text). Collaborative reactions were subject to a distance test (see Figure 3 and text). The state of all sites were recorded just before replication, with a subset of 100 out of 3000 simulated generations shown. (**B**) Methylation distributions for simulations for varying CpG-cluster sizes using $D = 65$ bp and $d = 10$ bp. Compare with Figure 2A. (**C**) Modeled dependence of N^* on d. Compare with Figure 2C.

the regions adjacent to the cluster, reminiscent of CpG island shores (39).

Varying the number of CpGs in the cluster, while keeping all other parameters fixed, produced the trend seen for the methylome data, where smaller clusters were predominantly hyper-methylated and the probability of the hypo-methylated state increased with increasing N_C (Figure 4B). In the model this comes about because the long-range property of the collaborative methylation reactions allows the sparsely distributed CpGs outside the cluster to collaborate with each other to sustain their own hyper-methylation, but also to act within the cluster. A few clustered CpGs cannot overcome this pervasive methylating 'force'. However, increasing the number of CpGs in a cluster allows the short-range collaborative demethylation reactions to build up an interaction field that is able to resist methylation. In large clusters the demethylation reactions can dominate to the extent that only the hypo-methylated state is possible.

We also systematically tested the effect of cluster density $1/d$ and the separation between the cluster and the surroundings D, on N^*, the N_C at which the cluster was equally likely to be in the high or low methylation states (Figure 4C). We saw the same trend as seen in the methylomes, where N^* was smaller for more dense clusters (low d). That is, increasing cluster density allowed clusters with fewer CpGs to access the unmethylated state. As in the methylomes, the effect of D was small. These effects are understandable, as increased density of the cluster favors CpG interactions via short-range collaborative demethylation, while having little effect on collaborative methylation. The long-range activity of collaborative methylation reactions means that the effect of the outside CpGs on the cluster is not sensitive to changes in D that are relatively small compared to α.

For clusters that display bistable behavior, i.e. where $N_C \sim N^*$, the stability of each of the states is an important factor when comparing the clusters in the model with clusters in methylomes. If the state of a particular cluster were to flip back and forth rapidly, then in a sample of DNA from many cells, that cluster would be hyper-methylated in some DNAs and hypo-methylated in others, giving intermediate average methylation levels. In order to produce the bimodal pattern seen in the methylomes (Figure 2), each specific cluster must be in just one of the two possible states within most of the cells sampled, implying high stabilities. The stabilities of the hyper- and hypo-methylated states in the model vary depending on cluster size and density, but the stability of the unfavored state ranges from 0 to 500 generations while the stability of the favored state ranges from 100 to >1000 generations. The average number of consecutive generations spent in each state for $N_C = 28$ is ∼100 generations in the methylated state and similarly 100 generations in the unmethylated state (out of 3000 simulated generations). However, both states are stable for at least 300 generations. Thus for many clusters, the stability of methylation states seen in the modeling may be insufficient by itself to explain the bimodality seen in the methylome data. We imagine two possible explanations. First, our model may for some reason underestimate the stabilities of each methylation state.

For example, we know that reducing the rate of the non-collaborative reactions in the model can increase stability (7). Decreasing the non-collaborative reactions by 5% increases the average consecutive generations spent in the unmethylated state to 150 generations and 280 generations for the methylated state. Second, many or most of the clusters may not be bistable in the cells studied. CpG number and density cannot be the only determinants of methylation state, and each individual cluster is likely to be subject to sequence-specific factors that affect the rates of the methylation or demethylation reactions and bias the cluster toward one of the states. In some clusters this bias could favor the hypo-methylated state, in others the hyper-methylated state, so that many clusters which might be bistable in other cell types remain stably in one state.

DISCUSSION

We proposed the collaborative model of CpG methylation as a mechanism to provide the robust maintenance and inheritance of alternative methylation states required for a true epigenetic mark (7). We have shown here that a simple extension of this model, in which the methylation and demethylation reactions are differentially sensitive to the distance between interacting CpGs, is able to reproduce the general relationship between CpG clustering and CpG methylation in the human genome.

Mechanisms of distance-dependent CpG collaboration

Although there is some evidence for collaborative methylation and demethylation reactions, little is known about their distance-dependence. However, we expect that the required distance-dependent collaboration would not be difficult to achieve mechanistically. For example, the UHRF1 protein binds a hemi-methylated CpG site via its SRA domain and recruits DNMT1 (40). This recruited DNMT1 is thought to methylate other hemi-methylated CpGs (41), providing one of the required collaborative $h \rightarrow m$ reactions ((7); Figure 3). It is possible that this DNA-tethered UHRF1-DNMT1 complex is not flexible enough to allow equal access of the DNMT1 catalytic domain to all CpGs in nearby chromatin, possibly giving a bias against short-range interactions.

Ten-eleven translocation methylcytosine dioxygenase 1 (TET) proteins are the prime candidates for CpG demethylases, catalyzing oxidation of 5mC to 5-hydroxymethyl-C and initiating a complex pathway for removal of the methylated cytosine (15). Consistent with the collaborative model, TET1 preferentially associates with CGIs, which are largely unmethylated (12) but this recruitment is poorly understood. Recruitment of TET2 by IDAX, which contains a CXXC domain that recognizes DNA containing unmethylated CpG and is enriched at sites with high CpG content (42), could in theory provide collaborative demethylation. A DNA-tethered IDAX–TET2 complex may be sufficiently flexible to oxidize CpGs close by on the DNA, providing the short-range collaboration required by the model. In theory, methylation or demethylation collaboration may be achieved by more complex recruitment reactions, potentially involving other chromatin marks such as histone modifications or other DNA modifications (8), each with their own characteristic distance dependencies.

An alternative mechanism to the short-range collaborative demethylation reactions in our model is suggested by the study of Thomson *et al.* (43). They proposed that recruitment of the CXXC protein Cfr1 to unmethylated CpG clusters could inhibit DNMT action on the cluster and maintain the unmethylated state. We have tested this type of mechanism by simulations and have shown that it is indeed able to substitute for the collaborative demethylation reactions in our model (44). Bistability is possible if recruitment of the inhibitor protein by unmethylated CpGs is cooperative, and if the inhibition of methylation extends to the neighbors of the unmethylated CpGs to which the protein is bound. Thus, the principle of short-range collaboration between unmethylated CpGs is shared by both mechanisms. However, when more long-ranged reactions are required, there may be limitations to such a short-ranged cooperative protection mechanism.

The relationship between CpG topology and methylation could be tested experimentally by inserting large DNA fragments containing synthetic CpG clusters set within a low CpG density sequence, into gene-free genomic regions in a suitable cell line, followed by assessment of their methylation states. Use of clusters of different sizes, densities and initial methylation status, would allow systematic determination of the general geometric rules for DNA methylation.

Implications of the model

The model provides a different way of thinking about CpG islands, one that is more strongly tied to the bistability that underpins epigenetic memory. Our analysis suggests that clusters ranging from ∼10 CpG sites within a region of ∼80 bp to ∼40 CpG sites within ∼500 bp are intrinsically bistable. Thus, even a small cluster may be capable of carrying epigenetic memory, being able to be in either a hyper- or hypo-methylated state by transient signals and retaining that state once the signals disappear. Our results argue for a stronger focus on small CpG clusters.

In contrast individual, isolated CpGs and small, sparse clusters are predicted to be unable to maintain hypomethylation in the absence of a sequence-specific external factor. Similarly, very large, dense clusters, or clusters of clusters, may not be able to stably maintain hypermethylation. This intrinsic property of large clusters may explain the failure of maintenance of targeted CpG methylation within a large CGI (a cluster of clusters with 198 CpG within 2220 bp) at the human VEGF-A promoter (45). Even if methylation of all of this cluster could be achieved by targeting, the intrinsic bias toward demethylation may be too strong for methylation to persist after targeting. Our modeling suggests that targeting methylation at small, isolated CpG clusters is more likely to induce stable changes.

The collaborative model also has important implications for the origin of clustering in vertebrate genomes. CpG clustering is proposed to be a by-product of a high mutation rate for 5mC residues causing CpG sites that are more often methylated in the germ line to be lost faster than those that are more often unmethylated (46). In the collaborative model, the feedback between CpG density and methylation state should tend to make this mutation rate-driven evolution of clustering more rapid, since loss of a CpG site will

enhance methylation and thus loss of nearby sites, while gain of a CpG site will help nearby CpG sites be unmethylated and thus survive. In addition, the functionality of CpG clustering in collaboration means that there would likely be significant selective pressure for gain or loss of CpG sites in order to optimize methylation states (43).

The generation of CpG methylation patterns and epigenetic memory

The classical model does not by itself predict any effects of CpG clustering on methylation state. Variants of the classical model invoke locus-specific individual CpG methylation and demethylation reaction rates (8,47), which can in theory explain, but not predict, clustering effects. In contrast, our collaborative model is generic, invoking relatively few global parameters that apply equally to all CpG sites and allows some prediction of methylation status from CpG number and density alone. However, additional sequence-specific factors are clearly needed to generate the full temporal and spatial patterning seen in methylomes.

In the non-collaborative models, there is a single equilibrium methylation level for any CpG under any given set of conditions. Sequence-specific factors can change the position of this equilibrium but do not automatically generate bimodal methylation patterns. In contrast, the positive feedback in the collaborative model provides an intrinsic force that pushes a cluster away from intermediate methylation levels toward either hyper- or hypo-methylation. Sequence-specific factors act to change the probability of occupation of these alternative states.

The lack of bistability in the non-collaborative models means that if a methylation state of a cluster is set by sequence-specific signals, it will inexorably revert to its default methylation level once the signals disappear. In contrast, the collaborative model predicts that some CpG clusters, once set into the hyper- or hypo-methylated state, can remain in that state stably and heritably in the absence of the signal, providing epigenetic memory.

ACKNOWLEDGEMENT

C.L., J.O.H., I.B.D. and K.S. acknowledge financial support from the Danish National Research Foundation through the Center for Models of Life.

FUNDING

Australian NHMRC [GNT1025549]. Funding for open access charge: Danish National Research Foundation.
Conflict of interest statement. None declared.

REFERENCES

1. Jaenisch,R. and Bird,A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.*, **33**, 245–254.
2. Felsenfeld,G. (2014) A brief history of epigenetics. *Cold Spring Harb. Perspect. Biol.*, **6**, doi:10.1101/cshperspect.a018200.
3. Bird,A. (2007) Perceptions of epigenetics. *Nature*, **447**, 396–398.
4. Holliday,R. and Pugh,J.E. (1975) DNA modification mechanisms and gene activity during development. *Science*, **187**, 226–232.
5. Riggs,A.D. (1975) X inactivation, differentiation, and DNA methylation. *Cytogenet. Genome Res.*, **14**, 9–25.
6. Jones,P.A. and Liang,G. (2009) Rethinking how DNA methylation patterns are maintained.. *Nat. Rev. Genet.*, **10**, 805–811.
7. Haerter,J.O., Lövkvist,C., Dodd,I.B. and Sneppen,K. (2014) Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states. *Nucleic Acids Res.*, **42**, 2235–2244.
8. Jeltsch,A. and Jurkowska,R.Z. (2014) New concepts in DNA methylation. *Trends Biochem. Sci.*, **39**, 310–318.
9. Lorincz,M.C., Schübeler,D., Hutchinson,S.R., Dickerson,D.R. and Groudine,M. (2002) DNA methylation density influences the stability of an epigenetic imprint and Dnmt3a/b-independent de novo methylation. *Mol. Cell. Biol.*, **22**, 7572–7580.
10. Cedar,H. and Bergman,Y. (2012) Programming of DNA methylation patterns. *Annu. Rev. Biochem.*, **81**, 97–117.
11. Williams,K., Christensen,J. and Helin,K. (2011) DNA methylation: TET proteins—guardians of CpG islands? *EMBO Rep.*, **13**, 28–35.
12. Williams,K., Christensen,J., Pedersen,M.T., Johansen,J.V., Cloos,P.A., Rappsilber,J. and Helin,K. (2011) TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*, **473**, 343–348.
13. Xu,Y., Wu,F., Tan,L., Kong,L., Xiong,L., Deng,J., Barbera,A.J., Zheng,L., Zhang,H., Huang,S. *et al.* (2011) Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell*, **42**, 451–464.
14. He,Y.-F., Li,B.-Z., Li,Z., Liu,P., Wang,Y., Tang,Q., Ding,J., Jia,Y., Chen,Z., Li,L. *et al.* (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, **333**, 1303–1307.
15. Wu,H. and Zhang,Y. (2014) Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell*, **156**, 45–68.
16. Laird,C.D., Pleasant,N.D., Clark,A.D., Sneeden,J.L., Hassan,K.M.A., Manley,N.C., Vary,J.C., Morgan,T., Hansen,R.S. and Stöger,R. (2004) Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 204–209.
17. Zhang,Y., Rohde,C., Tierling,S., Jurkowski,T.P., Bock,C., Santacruz,D., Ragozin,S., Reinhardt,R., Groth,M., Walter,J. *et al.* (2009) DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. *PLoS Genet.*, **5**, e1000438.
18. Eckhardt,F., Lewin,J., Cortese,R., Rakyan,V.K., Attwood,J., Burger,M., Burton,J., Cox,T.V., Davies,R., Down,T.A. *et al.* (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.*, **38**, 1378–1385.
19. Weber,M., Hellmann,I., Stadler,M.B., Ramos,L., Pääbo,S., Rebhan,M. and Schübeler,D. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.*, **39**, 457–466.
20. Meissner,A., Mikkelsen,T.S., Gu,H., Wernig,M., Hanna,J., Sivachenko,A., Zhang,X., Bernstein,B.E., Nusbaum,C., Jaffe,D.B. *et al.* (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, **454**, 766–770.
21. Lister,R., Pelizzola,M., Dowen,R.H., Hawkins,R.D., Hon,G., Tonti-Filippini,J., Nery,J.R., Lee,L., Ye,Z., Ngo,Q.-M. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315–322.
22. Lister,R., Mukamel,E.A., Nery,J.R., Urich,M., Puddifoot,C.A., Johnson,N.D., Lucero,J., Huang,Y., Dwork,A.J., Schultz,M.D. *et al.* (2013) Global epigenomic reconfiguration during mammalian brain development. *Science*, **341**, 1237905.
23. Bird,A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
24. Gardiner-Garden,M. and Frommer,M. (1987) CpG islands in vertebrate genomes. *J. Mol. Biol.*, **196**, 261–282.
25. Cooper,D.N., Taggart,M.H. and Bird,A.P. (1983) Unmethlated domains in vertebrate DNA. *Nucleic Acids Res.*, **11**, 647–658.
26. Bird,A., Taggart,M., Frommer,M., Miller,O.J. and Macleod,D. (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, **40**, 91–99.
27. Rollins,R.A., Haghighi,F., Edwards,J.R., Das,R., Zhang,M.Q., Ju,J. and Bestor,T.H. (2006) Large-scale structure of genomic methylation patterns. *Genome Res.*, **16**, 157–163.

28. Saxonov,S., Berg,P. and Brutlag,D.L. (2006) 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.
29. Larsen,F., Gundersen,G., Lopez,R. and Prydz,H. (1992) CpG islands as gene markers in the human genome. *Genomics*, **13**, 1095–1107.
30. Varley,K.E., Gertz,J., Bowling,K.M., Parker,S.L., Reddy,T.E., Pauli-Behn,F., Cross,M.K., Williams,B.A., Stamatoyannopoulos,J.A., Crawford,G.E. *et al.* (2013) Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.*, **23**, 555–567.
31. Takai,D. and Jones,P.A. (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 3740–3745.
32. Edwards,J.R., O'Donnell,A.H., Rollins,R.A., Peckham,H.E., Lee,C., Milekic,M.H., Chanrion,B., Fu,Y., Su,T., Hibshoosh,H. *et al.* (2010) Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. *Genome Res.*, **20**, 972–980.
33. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
34. Hackenberg,M., Previti,C., Luque-Escamilla,P.L., Carpena,P., Martínez-Aroza,J. and Oliver,J.L. (2006) CpGcluster: a distance-based algorithm for CpG-island detection. *BMC Bioinformatics*, **7**, 446.
35. Glass,J.L., Thompson,R.F., Khulan,B., Figueroa,M.E., Olivier,E.N., Oakley,E.J., Van Zant,G., Bouhassira,E.E., Melnick,A., Golden,A. *et al.* (2007) CG dinucleotide clustering is a species-specific property of the genome. *Nucleic Acids Res.*, **35**, 6798–6807.
36. Antequera,F. (2007) CpG Islands and DNA Methylation. *eLS Encyclopedia of Life Sciences*, doi:10.1002/9780470015902.a0005027.
37. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
38. Ringrose,L., Chabanis,S., Angrand,P.O., Woodroofe,C. and Stewart,A.F. (1999) Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. *EMBO J.*, **18**, 6630–6641.
39. Irizarry,R.A., Ladd-Acosta,C., Wen,B., Wu,Z., Montano,C., Onyango,P., Cui,H., Gabo,K., Rongione,M., Webster,M. *et al.* (2009) The human colon cancer methylome shows similar hypo-and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.*, **41**, 178–186.
40. Bostick,M., Kim,J.K., Estève,P.O., Clark,A., Pradhan,S. and Jacobsen,S.E. (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science*, **317**, 1760–1764.
41. Bashtrykov,P., Jankevicius,G., Smarandache,A., Jurkowska,R.Z., Ragozin,S. and Jeltsch,A. (2012) Specificity of Dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain. *Chem. Biol.*, **19**, 572–578.
42. Ko,M., An,J., Bandukwala,H.S., Chavez,L., Äijö,T., Pastor,W.A., Segal,M.F., Li,H., Koh,K.P., Lähdesmäki,H. *et al.* (2013) Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature*, **497**, 122–126.
43. Thomson,J.P., Skene,P.J., Selfridge,J., Clouaire,T., Guy,J., Webb,S., Kerr,A.R., Deaton,A., Andrews,R., James,K.D. *et al.* (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*, **464**, 1082–1086.
44. Sormani,G., Haerter,J.O., Lövkvist,C. and Sneppen,K. (2016) Stabilization of epigenetic states of CpG islands by local cooperation. *Mol. Biosyst.*, doi:10.1039/C6MB00044D.
45. Kungulovski,G., Nunna,S., Thomas,M., Zanger,U.M., Reinhardt,R. and Jeltsch,A. (2015) Targeted epigenome editing of an endogenous locus with chromatin modifiers is not stably maintained. *Epigenet. Chromatin*, **8**, 1–11.
46. Cooper,D.N. and Krawczak,M. (1989) Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Hum. Genet.*, **83**, 181–188.
47. Genereux,D.P., Miner,B.E., Bergstrom,C.T. and Laird,C.D. (2005) A population-epigenetic model to infer site-specific methylation rates from double-stranded DNA methylation patterns. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 5802–5807.