Collaborations between CpG sites in DNA methylation

You SONG, Honglei REN

School of Software, Beihang University, Beijing 100191, China

Jinzhi LEI

Zhou Pei-Yuan Center for Applied Mathematics,
MOE Key Laboratory of Bioinformatics,
Tsinghua University, Beijing 100084, China
jzlei@tsinghua.edu.cn

DNA methylation patterns have profound impacts on genome stability, gene expression, and development. The molecular base of DNA methylation patterns has long been focused at single CpG sites level. Here, we construct a kinetic model of DNA methylation with collaborations between CpG sites, from which a correlation function was established based on experimental data. The function consists of three parts that suggest three possible sources of the correlation: movement of enzymes along DNA, collaboration between DNA methylation and nucleosome modification, and global enzyme concentrations within a cell. Moreover, the collaboration strength between DNA methylation and nucleosome modification is universal for mouse early embryo cells. The obtained correlation function provide insightful understanding for the mechanisms of inheritance of DNA methylation patterns.

Keywords: DNA methylation; correlation; epigenetics

1. Introduction

DNA methylation is a process by which methyl groups are added to DNA segments. Methylation represents a key epigenetic modification that changes the activity of a DNA segment without changing the sequence. The DNA methylation patterns are dynamically regulated during development, and are important for stable silencing of gene expression, maintenance of genome stability, and establishment of genomic imprinting [1,18]. Methylation of the fifth position of cytosine (5-methylcytosine, 5mC) is found in most plant, animal, and fungal models, and is primarily restricted to palindromic CpG (CG/GC) dinucleotides [12,16]. Biologically, the molecular base of DNA methylation patterns has been focused at *single* CpG sites level, which is regulated by various types of enzymes, such as DNMT1/UHRF1 for the maintenance of methylation marks, TET families for the active demethylation, and DNMT3a/DNMT3b for the *de novo* methylation [18].

There are now substantial evidences that additional mechanisms should be at work to ensure the robust maintenance of methylation patterns. During mouse gametes and early embryos, most functional genomic elements undergo significant demethylation, except CpG islands (CGIs) and 5' untranslated regions (UTRs) whose methylation levels are already very low in gametes [17]. In zebrafish early embryos, the oocyte methylome is gradually discarded after 16-cell stage, and then progressively reprogrammed to a pattern similar to that of the sperm methylome [10]. DNA methylation patterns of human hematopoietic stem cells (HSCs) from four different sources show different profiles, and DNA methylation dynamics of myeloid-lymphoid lineage choice display an asymmetric pattern [6]. These collaborative changes in DNA methylation patterns suggest potential correlations between the kinetics of CpG sites methylation/demethylation.

A computation model has shown that dynamic collaboration between nearby CpG sites can provide strong error-tolerant inheritance of methylation states of a cluster of CpGs, and shown stable bimodal methylation patterns [8]. It was proposed that the collaboration can be achieved by recruitment of methylase or demethylase enzymes. However, the molecular based of the correlation remains unclear, and this proposed mechanism fail to explain the CpG site distance dependence correlation from experimental data (Figure 1b).

In this study, we examine experimental data for patterns of DNA methylation correlation between adjacent CpG sites, and establish a computational model of methylation/demethylation kinetics in which CpG site distance dependence correlations are involved. The correlation function is obtained from experimental data. Mechanisms of correlation are discussed in accordance with the obtained correlation function.

2. Results

2.1. *Pearson correlations between CpG site methylations*

To examine the patterns of DNA methylation, we study data from mouse early embryo (GSM1386021)[17], which includes 5mC reads of each CpG sites from different stages of mouse embryo from sperm/oocyte to E13.5. In mouse genome, the distances between adjacent CpG sites range from 2 to a few thousand bp. Different chromosomes show similar spectrum of CpG distance distributions and the frequency decreases with the CpG distance (Figure 1a). For adjacent pairs of CpG sites with given distance d (in bp), the Pearson correlation $C(d)$ was obtained as

$$C(d) = \frac{\sum_i (X_i - \langle X \rangle)(Y_i - \langle Y \rangle)}{\sqrt{\sum_i (X_i - \langle X \rangle)^2 \sum_i (Y_i - \langle Y \rangle)^2}}, \tag{1}$$

where (X_i, Y_i) is the DNA methylation level of a pair of CpG sites, $\langle X \rangle$ and $\langle Y \rangle$ are average of $\{X_i\}$ and $\{Y_i\}$, respectively.

We examine data from the parental strain chromosome 1 of 2-cell stage at mouse embryo (P-chr1-2-cell-mouse in short), and see clear correlation between methyla-
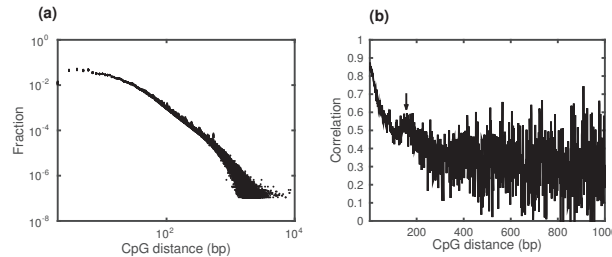
Fig. 1. Experimental data. (a) Distribution of the distances between adjacent CpG sites from chr1 to chr19 of mouse DNA. (b) Pearson correlations between DNA methylation levels of adjacent CpG sites of given distance (in bp). Data obtained from the parental strain chromosome 1 of 2-cell stage at mouse embryo (P-chr1-2-cell-mouse) (data source: GSM1386021).

tion levels of adjacent CpG sites. There is strong correlation with Pearson's $r = 0.84$ when the CpG distance is 2 bp, and the correlation decreases with the CpG distance (Figure 1b). The correlation displays a local maximum at the distance of about 160 bp (Figure 1, arrow). Moreover, the Pearson correlation is consistent when we examine data for different chromosomes in the same cell (Appendix A). While we calculate the correlation patterns from cells of different embryonic stages, the correlations show evolutionary dynamics during embryo development (Appendix B). These results suggest that the correlation pattern is universal for all chromosomes, however can be dependent on the cellular status.

2.2. *Kinetic model of DNA methylation*

To investigate the molecular mechanism of correlation between CpG sites, we construct a kinetic model aiming to uncover the molecular mechanism of CpG sites correlation in DNA methylation. Here, we introduce a key assumption that the correlation between methylation/demethylation of CpG sites is dependent on the CpG distance (Figure 2). Each CpG site can transit between fully un-metheylated (u), half-methyalted (h), and fully methylated (m) states (solid arrows in Figure 2). The transition rates are affected by the status of nearby CpGs (dashed arrows in Figure 2), and the correlation strength $\phi(d)$ is represented with a function of the distance (in bp) between the two CpGs. Our aim was to formulate the function $\phi(d)$ to reproduce the experimental correlation.

Our model was established based on the collaboration model in Haerter et al. (2014)[8]. The i'th CpG site can at one of the states u, h ,or m, and randomly methylate with a rate u_i^+ from the state u to h, or a rate h_i^+ from the state h to m. Moreover, the methyl mark can be removed from the CpG site with a rate m_i^- from m to h, or a rate h_i^- from h to u. The kinetic rates $(u_i^+, h_i^+, m_i^-, h_i^-)$ consist of two parts of contributions, the contribution from basal level rates (u^+, h^+, m^-, h^-), and the contribution due to collaboration between neighboring CpG sites. According the assumptions in Haerter et al. (2014)[8], when a CpG is at state m, it promotes the methylation of the nearby CpG, and hence the methylation rate from u to h

4 *SONG, REN, LEI*

Table 1. Default kinetic parameter values used in the simulations. Parameters are referred to Haerter et al.(2014).

u^+	h^+	m^-	h^-	h^{+h}	h^{+m}	u^{+m}	h^{-u}	m^{-u}
0.008	0.008	0.04	0.04	0.24	0.24	0.24	0.05	0.05

increases by $\phi(d)(u^{+m} - u^+)$, and the rate from h to m increases by $\phi(d)(h^{+m} - h^+)$. Similarly, a CpG of state h contributes an increase $\phi(d)(h^{+h} - h^+)$ to the methylation rate of the nearby CpG site h_i^+, and a CpG of state u promotes the demethylation processes, so that the demethylation rate m_i^- increases by $\phi(d)(m^{-u} - m^-)$, and the rate h_i^- increases by $\phi(d)(h^{-u} - h^-)$.

Mathematically, the kinetic rates of the i'th CpG in a DNA sequence are given by

$$
\begin{cases}
u_i^+ = u^+ + \sum_{j=1}^N p_{i,j} \cdot \chi_j(m) \cdot \phi(d_{i,j}) \cdot (u^{+m} - u^+), \\
h_i^+ = h^+ + \sum_{j=1}^N p_{i,j} \cdot \chi_j(m) \cdot \phi(d_{i,j}) \cdot (h^{+m} - h^+) \\
\qquad + \sum_{j=1}^N p_{i,j} \cdot \chi_j(h) \cdot \phi(d_{i,j}) \cdot (h^{+h} - h^+), \\
m_i^- = m^- + \sum_{j=1}^N p_{i,j} \cdot \chi_j(u) \cdot \phi(d_{i,j}) \cdot (m^{-u} - m^-), \\
h_i^- = h^- + \sum_{j=1}^N p_{i,j} \cdot \chi_j(u) \cdot \phi(d_{i,j}) \cdot (h^{-u} - h^-).
\end{cases}
\tag{2}
$$

Here $p_{i,j} = 1$ when the i'th CpG is affected by the j'th CpG, otherwise $p_{i,j} = 0$. The index $\chi_j(s) = 1$ $(s = u, h, m)$ if the j'th CpG has the status s, and $\chi_j(s) = 0$ if otherwise. We simply consider the nearby correlation so that $p_{i,j} = 1$ only when $j = i \pm 1$. Here we note that although only the nearby CpGs are involved in the correlation, however the distance between nearby CpG sites can be separated by hundreds base pairs. We note that the function $\phi(d)$ is assumed to be the same for all type correlations. This simplification was trying to construct a minimal model, however biologically can be different. In model simulations, we refer the kinetic models from Haerter et al.(2014) [8], which are listed at Table 1.

The kinetic rates in Eq. 2 define the dynamics of in each cell cycle. To include the effect of reallocation at DNA replication, we reset the status m to h at the beginning of each cell generation. For a given DNA sequence and a well defined function $\phi(d)$, stochastic simulation based on the above kinetic rates for ~ 20 generations lead to a stationary DNA methylation pattern, which give a model predicted correlation profile.
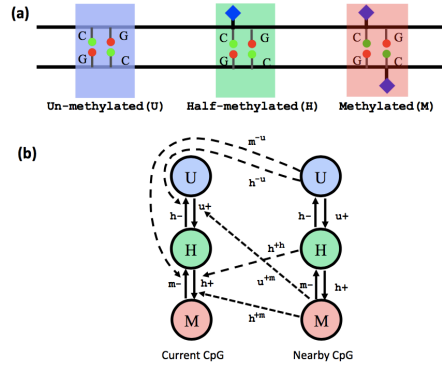
Fig. 2.   Schematic illustration of the model. (a) Three possible states of a CpG: un-methylated (U), half-methylated (H), or methylated (M). (b) Kinetics of the transitions between the methylation state of a CpG site. Solid arrows represent the transition between the states, dashed arrows show the correlations from the nearby CpG.

2.3. *Correlation function*

To obtain the function $\phi(d)$, we refer the experimental data from P-chr1-2-cell-mouse, denoted as $C(d)$ (Figure 1b). The procedure is given below.

(1) Set ϕ in model as a constant independent to the distance d and run the model simulation. Thus, the correlation defined in above is dependent on the value of ϕ, however is independent to the distance d (Figure 3(a)). Larger ϕ leads larger correlations (Figure 3(b)). Fitting the dependence of the simulated correlation with ϕ, we obtained the function $R_0(\phi)$ (Figure 3(b), red)

$$R_0(\phi) = \frac{\phi^{1.1}}{5.02 + \phi^{1.1}} + 0.65\sqrt{\phi}e^{-0.51\phi}. \tag{3}$$

In fitting the data, we first fit with Hill type function $\frac{\phi^s}{a+\phi^s}$ to obtain good fit with $\phi < 5$ to have $s = 1.1$, next fit the remaining part of data with the function of form $b\sqrt{\phi}e^{-c\phi}$ to obtain $c = 0.51$. Finally, we adjust the coefficients a and b to obtain a good fit for overall data. The function $R_0(\phi)$ is monotonous increasing.

(2) Consider the simulated correlation $R_0(\phi)$ and the correlation obtained from experimental data (Figure 1), which is denoted as $C(d)$. While assuming the two correlations equal to each other, *i.e.*, $R_0(\phi) = C(d)$, we obtain the approximation $\phi = R_0^{-1}(C(d))$ (Figure 3(c)).

(3) Substitute the function $\phi(d)$ obtained at the above step and run the model simulation to obtain the first approximation of correlation, which is denoted as $C_0(d)$ (Figure 3(d)). We note that the simulated correlation $C_0(d)$ has the same tendency as the experimental data $C(d)$, however $C_0(d)$ is in general larger than $C(d)$.

(4) Let the correlation $R_0(\phi)$ equals $C_0(d)$, which gives $R_0(\phi) = C_0(d)$. Solve this

6   *SONG, REN, LEI*

equations to have the correlation function $\phi(d) = R_0^{-1}(C_0(d))$ (Figure 3(e), blue dots). We note that the numerical data show power law decay in small d region, and a bell-shape bulb at around $d = 160$. Hence, we introduce a function that combines power law decay $\frac{1}{(\beta+\gamma d)^p}$, a bell-shape bulb function $\frac{\alpha}{30\sqrt{2\pi}}\exp\left[-\frac{(d-160)^2}{2\times 30^2}\right]$, and a constant tail, so that the the numerical data can be fitted with a correlation function of form (Figure 3(e), red line):

$$\phi(d) = \frac{1}{(\beta+\gamma d)^p} + \frac{\alpha}{30\sqrt{2\pi}}\exp\left[-\frac{(d-160)^2}{2\times 30^2}\right] + c. \qquad (4)$$

In numerical fitting, we can first find the power law p with data of small d region ($d < 130$), and then find the fitting coefficients through `FindFit` of `Mathematica`.

(5) Now, we substitute the correlation function (4) into model equations, and run the simulation to obtain the simulated correlation, which show good agreement with experimental results (Figure 3(f)).

Table 2.   Parameters for the correlation functions corresponding to different stages and strands from mouse embryo cells.

Para.	sperm	2-cell	4-cell	E6.5	E7.5	PGC E13.5
p	1.5	1	1	1	1	1
α	12	12	12	12	12	12
β	0.001	0.195	0.3	0.15	0.04	0.09
γ	0.0068	0.04	0.036	0.06	0.042	0.06
c	0.005	0.13	0.1	0.06	0.02	0.001
Para.	oocyte	2-cell	4-cell	E6.5	E7.5	PGC E13.5
p	1.5	1	1	1	1	1
α	12	12	12	12	12	12
β	0.07	0.0015	0.032	0.001	0.001	0.138
γ	0.0135	0.0092	0.0065	0.05	0.03	0.075
c	0.19	0.09	0.07	0.07	0.02	0.001

The above procedure gives the correlation function obtained from experimental data. Substituting the obtained $\phi(d)$ into Eq. 2 and perform stochastic simulation, we can nicely reproduce the correlation profile, in good agreement with experimental data (Fig. 3f).

We note that both kinetic parameters in Table 1 and experimental data are involved in the procedure of obtaining $\phi(d)$. The kinetic parameters usually affect the mean methylation level, and may alter the coefficients in the correlation function (detailed below). However, the formulation of the function $\phi(d)$ maintains for different sets of parameters. Here, as we intended to study the correlation, we always fixed the kinetic parameters and examined the dependence of $\phi(d)$ on experimental data. Nevertheless, the kinetic parameters are dependent on the enzyme activities, and can be alter from cell to cell.
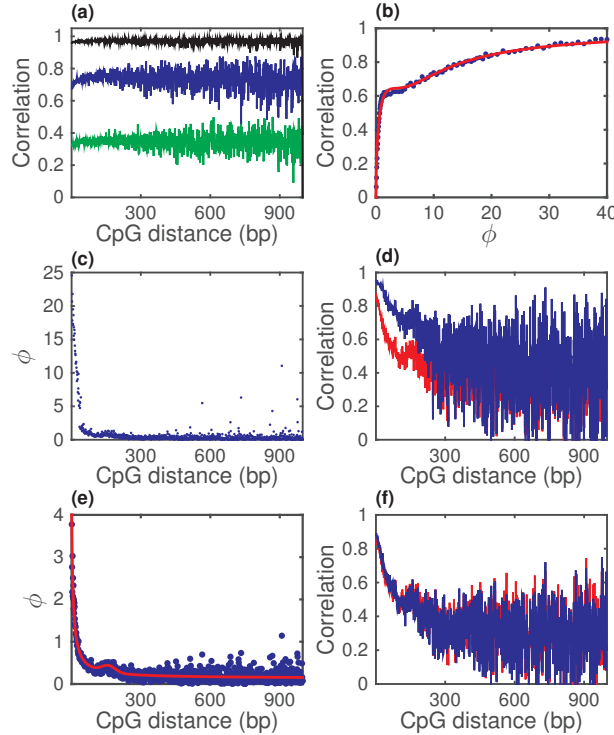
Fig. 3.   Procedure to obtain the correlation function $\phi(d)$. (a) Simulated Pearson correlations with fixed value ϕ: $\phi = 0.3$ (green), $\phi = 2$ (blue), and $\phi = 30$ (black). (b) Dependence of the Pearson correlation on the value ϕ. Red curve shows the fitting with equation (3). (c) The approximation correlation function calculated by $\phi = R_0^{-1}(C(d))$. (d) The Pearson correlation obtained from experimental data (red) and from model simulation with $\phi = R_0^{-1}(C(d))$ (blue). (e) The correlation function $\phi(d) = R_0^{-1}(C_0(d))$ (blue dots). Red curve shows the fitting with equation (4), with parameters $\alpha = 12, p = 1, \beta = 0.195, \gamma = 0.036, c = 0.13$. (f) Pearson correlations obtained from experimental data (red) and from model simulation with $\phi(d)$ given by equation (4) (blue).

We apply the procedure to experimental data from different stages of mouse early embryo. The function Eq. (4) is well fitting with different stages data, however the coefficients can be varied (Table 2)(Figure 5). Moreover, we apply the procedure to data from different tissues of mouse and human, and show small root-mean-square deviations (RMSDs) between experimental correlation and model simulation correlation (Appendix C, Table C1). These results imply that the correlation $\phi(d)$ of form Eq. (4) can provide insights to the dynamics of DNA methylation.

2.4. *Effects of kinetic parameters to the correlation function*

To examine the effects of changing kinetic parameters to the correlation function, we increase or decrease each kinetic parameter by 50% and examine the corresponding changes in the coefficients in the correlation function. Default values of the kinetic parameters are listed in Table 1, changes in the average methylation level and
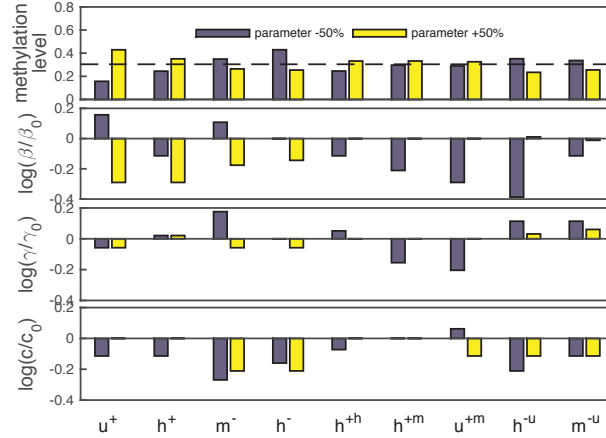
8   *SONG, REN, LEI*

Fig. 4. Effects of changes in the kinetics parameters. We either increased (blue bars) or decreased (yellow bars) each kinetic parameters by 50%, and examine the effects to the average methylation level and to the coefficients in the correlation function Eq. 4. Here the dashed line at the upper panel shows the methylation level with default parameters. Here β, γ, c represent the coefficients for modified parameters, and β_0, γ_0, c_0 represent the coefficients for default parameters, respectively. We always set $\alpha = 12$ in fitting the data. See the text for discussion. In the last three panels, the absent bars (zero) mean no difference in the parameter values.

parameters in the correlation function in response to changes in kinetic parameters, are given by Figure 4. Results show that changing the kinetic parameters can affect the average methylation level, as well as the coefficients β, γ and c in the correlation function given by Eq. 4 (here we fixed $p = 1$ and $\alpha = 12$).

Usually, decreasing the methyl adding rates u^+ and h^+ decrease the methylation level, and decreasing the methyl removing rates m^- and h^- increase the methylation level. The average methylation level is not sensitive to changes in the correlative parameters h^{+h}, h^{+m}, u^{+m}, h^{-u}, and m^{-u}. These observations are confirmed by Figure 4.

From Figure 4, the consequences of changing the kinetic parameters to the correlation function parameters β, γ, and c are complicated. According to the procedures of obtaining the correlation function, the kinetic parameters mainly affect the function $R_0(\phi)$ and the first order approximation $C_0(d)$. However, these two functions are obtained from model simulation, therefore the stochastic effects are non-negligible in obtaining these correlation function parameters.

2.5. *The parameter α in the correlation function*

For the parameter α, we have seen from Table 2 that $\alpha = 12$ can be applied to different stage embryonic cells. This suggests that the correlation may be insensitive with the parameter α. To test is hypothesis, we alter the value α to examine its effect to the simulated correlations. We increase or decrease α by 50% in Eq. 4 and keep other parameters unchanged and run the model simulation. We repeat

the simulations based on data of different embryonic cells, and found no obvious variations in obtained correlations (Figure 5). The RMSD of between simulated correlations and correlations obtained from experimental data are not sensitive with the parameter α Table 3. We further apply the produce to various mouse and human somatic tissue cells, the results give two catalogues cells, either $\alpha = 0.001$ or $\alpha = 4$ (Appendix C, Table C1). These results suggest that the α can be a universal parameter under certain conditions. Nevertheless, the mechanism waiting for further investigation through models with more kinetics details of the regulation of DNA methyaltion.
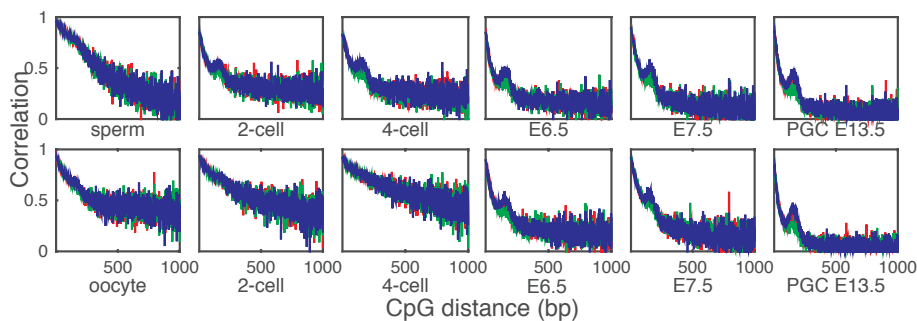

Fig. 5.   Correlations with $\alpha = 6.5$ (decreased by %50, green), $\alpha = 12$ (default, red) or $\alpha = 19.5$ (increased by %50, blue) based on different stages of mouse embryonic cells. Other coefficients or kinetic parameters were taken as default values as Table 1 and Table 2.

Table 3.   The Root Mean Squared errors of correlations from simulations ($\alpha = 6.5, 12$ or 19.5) relative to the correlations from experimental data of different stages or strains of mouse embryonic cells.

α	sperm	2-cell	4-cell	E6.5	E7.5	PGC E13.5
6.5	0.209	0.13	0.156	0.203	0.214	0.269
12	0.211	0.129	0.147	0.197	0.215	0.26
19.5	0.209	0.129	0.146	0.194	0.207	0.254

α	oocyte	2-cell	4-cell	E6.5	E7.5	PGC E13.5
6.5	0.179	0.212	0.263	0.182	0.185	0.282
12	0.18	0.217	0.266	0.172	0.188	0.274
19.5	0.183	0.218	0.263	0.17	0.184	0.271

2.6. *The parameter c in the correlation function*

The constant tail represents a global impact of the correlation. We examine the average methylation level versus the constant tail for different stages in mouse embryo cells and tissues. The results show a nonlinear dependence, except the stage of PGC at E13.5 when the DNA methylation is extreme low due to active demethylation

(Figure 6). Moreover, we test the data from human tissue cells, which reveal similar form nonlinear dependences (Figure 6). These results suggest that the constant tail in the correlation function is related to the global methylation level, however the molecular mechanism is waiting for further investigation.
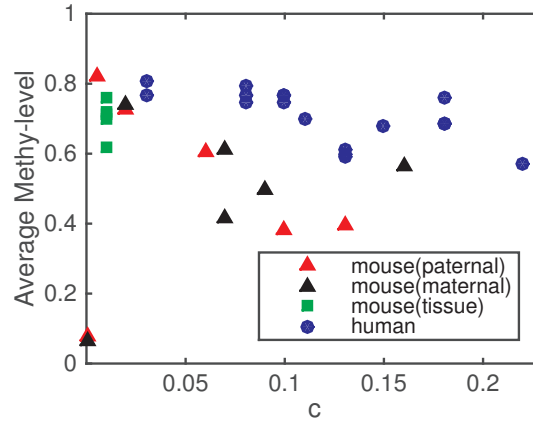

Fig. 6.   Experimental average methylation level versus constant tail c in the correlation function, data obtained form mouse early embryo cells (paternal, red triangles; maternal, blue triangles) and tissue (green squares), or human tissues (blue circles).

3. Discussion

We have established a kinetic model of DNA methylation, from which and combining with experimental data we obtained the correlation function for the collaboration effects of adjacent CpG sites methylation. The function $\phi(d)$ consists of three parts: the power law decay, a bell-shape bulb, and a constant tail. Here we discuss the biological sources of the three parts based on mathematical formulation.

It was proposed in [8] that collaboration between CpGs can be achieved by recruitment of methylase or demethylase enzymes, which akin to the situation of nucleosome modifications [19]. Here, the power law decay suggests an alternative mechanism, by which the enzymes perform random walk along DNA and self-enhance dissociation. Accordingly, the enzyme concentration along the DNA can be described with a diffusive equation of form

$$\frac{\partial A}{\partial t} = D\frac{\partial^2 A}{\partial x^2} - \gamma A^n, \ A(0) = A_0. \tag{5}$$

This equation gives the power-law decaying steady state solution of form $A = A_0(x/\epsilon + 1)^{-2/(n-1)}$ in the case of self-enhance dissociation ($n > 1$) [14]. Biologically, many experiments have revealed similar cohesin movement along DNA that play crucial roles in gene expression [4,11]. Nevertheless, evidences for the molecular based

June 2, 2017   6:7   WSPC/INSTRUCTION FILE   DNAMethy

of the correlation, either recruitment or movement along DNA, are waiting to further studies.

Interestingly, the bell-shape bulb is central at 160bp, right the distance of about one nucleosome. The strength coefficient $\alpha = 12$ is consistent for early mouse embryo cells from sperm/oocyte to PGC E13.5 (Table 2), and the correlation is insensitive with changes in the parameter α (Figure 5). Moreover, various mouse and human somatic tissue cells give two catalogues cells with either small or large α values (Appendix C, Table C1). These observations suggest that the term $\frac{\alpha}{30\sqrt{2\pi}} \exp[\frac{(d-160)^2}{2 \times 30^2}]$ may represent a collaborative effect between DNA methylation and nucleosome modification, the collaboration strength α can be different for different type cells, but is insensitive with the kinetic parameters. A type of nucleosome modification, histone H3 lysine 9 dimethylation (H3K9me2), often show nucleation sites overlap to CGIs [2], and is involved in DNA methylation [13,15,7]. Moreover, loss of DNA methylation enhances the removal of H3K3me3 under transcriptional stimulus [9]. These evidences support the correlation between nucleosome modification and DNA methylation, however the molecular mechanism is not yet clear.

Data analysis and model simulation have shown that the constant tail coefficient c is related to the global methylation level. Biologically, the global methylation level can be regulated by the enzyme activities that regulate DNA methylation. This observation remains further confirmation and theoretical consideration from the perspective of competition and collaboration of enzymes in methylation kinetics.

4. Conclusion

Inheritance of DNA methylation pattern is crucial during development. Here, driven from experimental data, we establish the formulation of the collaboration of DNA methylation. Applying the obtained function to a simple stochastic dynamic model can well reproduce the experimental observed correlation. The function reveal that there are three possible sources of the correlation: movement of enzymes along DNA, collaboration between DNA methylation and nucleosome modification, and global enzyme concentrations in the cell. Moreover, the collaboration strength between DNA methylation and nucleosome modification is universal in our study of mouse early embryo. These findings provide insightful understanding of the collaborations in DNA methylation, and for the mechanisms of inheritance of DNA methylation patterns. Nevertheless, despite the well support from experimental data, molecular details of the correlation in DNA methylation proposed here are waiting for further experimental studies.

Acknowledgments

Appendix A. Correlation patterns are similar for different chromosomes

12   *SONG, REN, LEI*

in a cell

We calculated the correlation patterns of DNA methylation from the mouse embryo MethylC-Seq data [17]. We choose the experimental data from the parental strain of mouse embryo at the 2-cell stage, and calculated the Pearson correlation of methylation levels between two adjacent CpG sites from chromosome 1 to 18. The relationship between the correlation and CpG distance are shown at Figure 1. From Figure 1, we can see that the correlation patterns are similar for different chromosomes.
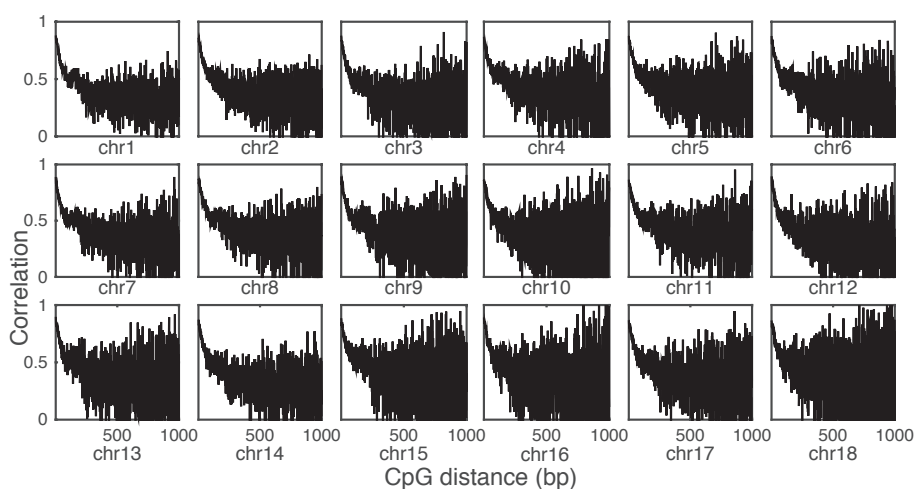

Fig. A1.   Pearson correlations, from chromosome 1 to 18, between DNA methylation levels of adjacent CpG sites as function of the distance (in bp) between CpG sites. Data obtained from the parental strain of mouse embryo at the 2-cell stage (P-2-cell-mouse, GEO: GSM1386021).

Appendix B. Correlation patterns are different at different embryo stages

We calculated the correlation patterns of DNA methylation from different mouse embryo stages (Figure 2). The correlation patterns of paternal and maternal chromosomes are distinct to each other before E6.5, and tend to be consistent in latter stages.

Appendix C. Correlation function for mouse and human tissues

Table C1 gives the coefficients and RMSD while we apply the procedure to experimental data of mouse and human tissues. Experimental data were obtained from the Whole Genome Bisulfite Sequencing (WGBS) data, GEO(ref. [5]), and ENCODE(ref. [3]). Fitting results are shown at Figure C1-C2.
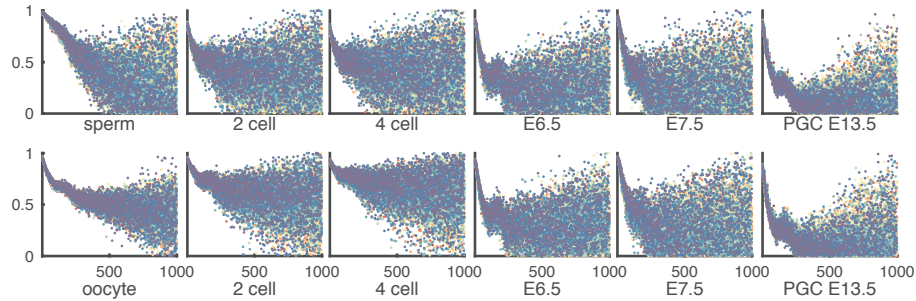
Fig. B1.   Pearson correlations between DNA methylation levels at adjacent CpG sites from different embryo stage of mouse (GEO: GSE56697). The correlation patterns of upper row are paternal chromosomes, and the bottom row are maternal chromosomes.
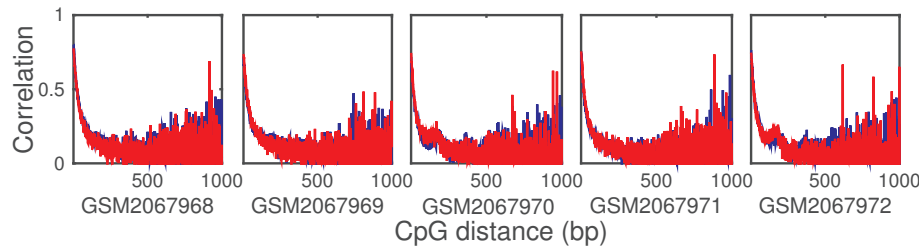

Fig. C1.   Correlations of mouse sample cells from Table C1. Blue for correlation obtained from experimental data, red for correlation obtained from models simulation.
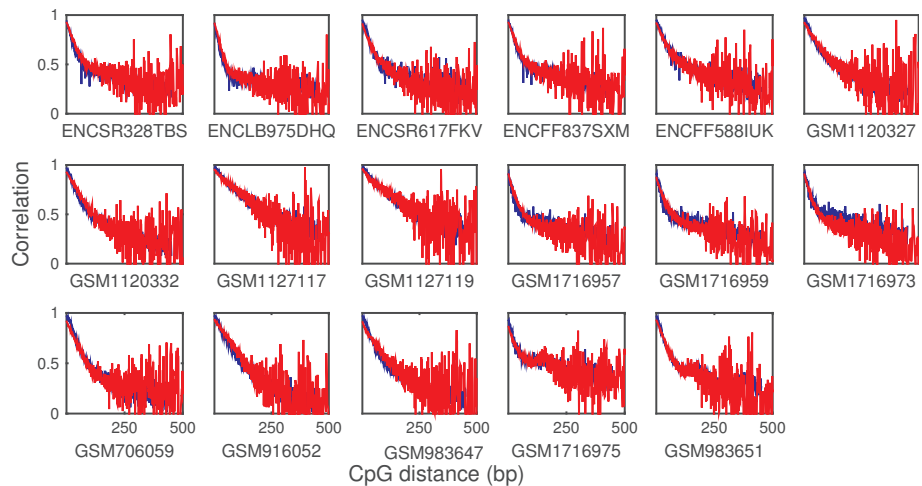

Fig. C2.   Correlations of human sample cells from Table C1. Blue for correlation obtained from experimental data, red for correlation obtained from models simulation.

1. Adrian Bird. DNA methylation patterns and epigenetic memory. *Genes Dev*, 16(1):6–21, January 2002.

14   *SONG, REN, LEI*

Table C1.   The data-source, average methylation levels and fitted coefficients of data.

Accession	Description	Methy-level	RMSD	c	α	β	γ	p
GSM2067968	mouse normal B1 cells	0.759	0.073	0.01	0.001	0.3	0.13	1
GSM2067969	mouse tumor Het CLL	0.716	0.071	0.01	0.001	0.3	0.13	1
GSM2067971	mouse normal CD8 cells	0.720	0.071	0.01	0.001	0.35	0.16	1
GSM2067970	mouse tumor mutant CLL	0.617	0.073	0.01	4	0.3	0.16	1
GSM2067972	mouse tumor PTCL	0.698	0.075	0.01	4	0.3	0.16	1
ENCSR328TBS	human skeletal muscle myoblast	0.686	0.112	0.18	0.001	0.001	0.035	2
ENCLB975DHQ	human mammary epithelial cell	0.677	0.104	0.15	0.001	0.08	0.045	2
ENCSR617FKV	human H1-hESC	0.767	0.124	0.1	0.001	0.001	0.035	1.5
ENCFF837SXM	human skeletal muscle myoblast	0.688	0.118	0.18	0.001	0.001	0.035	2
ENCFF588IUK	human smooth muscle cell	0.757	0.118	0.18	0.001	0.001	0.03	2
GSM1120327	human aorta male adult	0.748	0.129	0.1	0.001	0.01	0.0205	2
GSM1120332	human stomach male adult	0.748	0.114	0.08	0.001	0.01	0.0225	2
GSM1127117	human sperm male adult	0.768	0.123	0.08	0.001	0.01	0.0115	2
GSM1127119	human sperm male adult	0.768	0.123	0.1	0.001	0.01	0.0115	2
GSM1716957	human liver normal cell	0.588	0.115	0.13	0.001	0.001	0.045	1.5
GSM1716959	human liver tumor cell	0.599	0.107	0.13	0.001	0.1	0.04	1.5
GSM1716973	human lung normal cell	0.611	0.129	0.13	0.001	0.001	0.045	1.5
GSM706059	human neuronal stem cell	0.791	0.122	0.08	0.001	0.1	0.025	2
GSM916052	human common myeloid progenitor	0.809	0.12	0.03	0.001	0.01	0.0175	2
GSM983647	human lung female adult	0.764	0.137	0.03	0.001	0.01	0.0225	1.5
GSM1716975	human lung tumor cell	0.573	0.117	0.22	4	0.1	0.04	1.5
GSM983651	human pancreas male adult	0.696	0.127	0.11	4	0.01	0.03	2

2. Xiaoji Chen, Kyobi Skutt-Kakaria, Jerry Davison, Yang-Li Ou, Edward Choi, Punam Malik, Keith Loeb, Brent Wood, George Georges, Beverly Torok-Storb, and Patrick J Paddison. G9a/GLP-dependent histone H3K9me2 patterning during human hematopoietic stem cell lineage commitment. *Genes Dev*, 26(22):2499–2511, November 2012.

3. ENCODE Project Consortium et al. The encode (encyclopedia of dna elements) project. *Science*, 306(5696):636–640, 2004.

4. Iain F Davidson, Daniela Goetz, Maciej P Zaczek, Maxim I Molodtsov, Pim J Huis in't Veld, Florian Weissmann, Gabriele Litos, David A Cisneros, Maria Ocampo Hafalla, Rene Ladurner, Frank Uhlmann, Alipasha Vaziri, and Jan Michael Peters. Rapid movement and transcriptional re-localization of human cohesin on DNA. *EMBO J*, 35(24):2671–2685, December 2016.

5. Ron Edgar, Michael Domrachev, and Alex E Lash. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic acids research*, 30(1):207–210, 2002.

6. Matthias Farlik, Florian Halbritter, Fabian Müller, Fizzah A Choudry, Peter Ebert, Johanna Klughammer, Samantha Farrow, Antonella Santoro, Valerio Ciaurro, Anthony Mathur, Rakesh Uppal, Hendrik G Stunnenberg, Willem H Ouwehand, Elisa Laurenti, Thomas Lengauer, Mattia Frontini, and Christoph Bock. DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. *Cell Stem Cell*, November 2016.

7. Ehsan Habibi, Arie B Brinkman, Julia Arand, Leonie I Kroeze, Hindrik H D Kerstens, Filomena Matarese, Konstantin Lepikhov, Marta Gut, Isabelle Brun-Heath, Nina C Hubner, Rosaria Benedetti, Lucia Altucci, Joop H Jansen, Jörn Walter, Ivo G Gut, Hendrik Marks, and Hendrik G Stunnenberg. Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell stem cell*, 13(3):360–369, September 2013.

8. Jan O Haerter, Cecilia Lövkvist, Ian B Dodd, and Kim Sneppen. Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states. *Nucleic Acids Res*, 42(4):2235–2244, January 2014.

9. Nathaniel A Hathaway, Oliver Bell, Courtney Hodges, Erik L Miller, Dana S Neel, and Gerald R Crabtree. Dynamics and Memory of Heterochromatin in Living Cells. *Cell*, 149(7):1447–1460, June 2012.

10. Lan Jiang, Jing Zhang, Jing-Jing Wang, Lu Wang, Li Zhang, Guoqiang Li, Xiaodan

Yang, Xin Ma, Xin Sun, Jun Cai, Jun Zhang, Xingxu Huang, Miao Yu, Xuegeng Wang, Feng Liu, Chung-I Wu, Chuan He, Bo Zhang, Weimin Ci, and Jiang Liu. Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell*, 153(4):773–784, May 2013.

11. Mai Kanke, Eri Tahara, Pim J Huis in't Veld, and Tomoko Nishiyama. Cohesin acetylation and Wapl-Pds5 oppositely regulate translocation of cohesin along DNA. *EMBO J*, 35(24):2686–2698, December 2016.

12. Julie A Law and Steven E Jacobsen. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*, 11(3):204–220, February 2010.

13. Bernhard Lehnertz, Yoshihide Ueda, Alwin A H A Derijck, Ulrich Braunschweig, Laura Perez-Burgos, Stefan Kubicek, Taiping Chen, En Li, Thomas Jenuwein, and Antoine H F M Peters. Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Current Biology*, 13(14):1192–1200, July 2003.

14. Jinzhi Lei, Wing-Cheong Lo, and Qing Nie. Mathematical models of morphogen dynamics and growth control. *Annals of Mathematical Sciences and Applications*, 1(2):427–471, July 2016.

15. Toshinobu Nakamura, Yu-Jung Liu, Hiroyuki Nakashima, Hiroki Umehara, Kimiko Inoue, Shogo Matoba, Makoto Tachibana, Atsuo Ogura, Yoichi Shinkai, and Toru Nakano. PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature*, 486(7403):415–419, June 2012.

16. Stefanie Seisenberger, Julian R Peat, Timothy A Hore, Fátima Santos, Wendy Dean, and Wolf Reik. Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philos Trans R Soc Lond B Biol Sci*, 368(1609):20110330–20110330, January 2013.

17. Lu Wang, Jun Zhang, Jialei Duan, Xinxing Gao, Wei Zhu, Xingyu Lu, Lu Yang, Jing Zhang, Guoqiang Li, Weimin Ci, Wei Li, Qi Zhou, Neel Aluru, Fuchou Tang, Chuan He, Xingxu Huang, and Jiang Liu. Programming and inheritance of parental DNA methylomes in mammals. *Cell*, 157(4):979–991, May 2014.

18. Hao Wu and Yi Zhang. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell*, 156(1-2):45–68, January 2014.

19. Hang Zhang, Xiao-Jun Tian, Abhishek Mukhopadhyay, K S Kim, and Jianhua Xing. Statistical mechanics model for the dynamics of collective epigenetic histone modification. *Phys Rev Lett*, 112(6):068101, February 2014.