Collaborations between CpG sites in DNA methylation

You Song and Honglei Ren

*School of Software, Beihang University,
Beijing 100191, P. R. China*

Jinzhi Lei*

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

Received 2 May 2017

Revised 28 May 2017

Accepted 1 June 2017

Published 6 June 2017

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 provides insightful understanding for the mechanisms of inheritance of DNA methylation patterns.

Keywords: DNA methylation; correlation; epigenetics.

PACS numbers: 87.16.af, 87.10.Mn, 05.20.Dd

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

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

*Corresponding author.

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 displays 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 correlation remains unclear, and this proposed mechanism is not enough to explain the CpG site distance dependence correlation from experimental data [Fig. 1(b)].

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 two to a few thousand bp. Different

Fig. 1. Experimental data. (a) Distribution of the distances between adjacent CpG sites from chr1 to chr19 of mouse DNA, and (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).
Source: GSM1386021.

chromosomes show similar spectrum of CpG distance distributions and the frequency decreases with the CpG distance [Fig. 1(a)]. For adjacent pairs of CpG sites with given distance d (in bp), the Pearson correlation $C(d)$ was calculated by

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

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 methylation 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 [Fig. 1(b)]. The correlation displays a local maximum at the distance of about 160 bp (Fig. 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

Fig. 2. (Color online) Schematic illustration of the model. (a) Three possible states of a CpG: unmethylated (U), half-methylated (H) or methylated (M) and (b) Kinetics of the transitions between the methylation state of a CpG site. Solid arrows represent the transition between the states and the dashed arrows show the correlations from the nearby CpG.

CpG distance (Fig. 2). Each CpG site can transit between fully unmethylated (u), halfmethylated (h), and fully methylated (m) states (solid arrows in Fig. 2). The transition rates are affected by the status of nearby CpGs (dashed arrows in Fig. 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 to 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 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 methyla-

tion 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

$$\left\{ \begin{array}{l} u_i^+ = u^+ + \sum_{j=1}^N p_{i,j} \cdot \chi_j(m) \cdot \phi(d_{i,j}) \cdot (u^{+m} - u^+), \\ h_i^+ = h^+ + \sum_{j=1}^N p_{i,j} \cdot \chi_j(m) \cdot \phi(d_{i,j}) \cdot (h^{+m} - h^+) \\ \quad + \sum_{j=1}^N p_{i,j} \cdot \chi_j(h) \cdot \phi(d_{i,j}) \cdot (h^{+h} - h^+), \\ m_i^- = m^- + \sum_{j=1}^N p_{i,j} \cdot \chi_j(u) \cdot \phi(d_{i,j}) \cdot (m^{-u} - m^-), \\ h_i^- = h^- + \sum_{j=1}^N p_{i,j} \cdot \chi_j(u) \cdot \phi(d_{i,j}) \cdot (h^{-u} - h^-). \end{array} \right. \quad (2)$$

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 hundred 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 parameters from Haerter *et al.* (2014),⁸ which are listed in Table 1.

The kinetic rates in Eq. (2) define the dynamics 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 gives a model predicted correlation profile.

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

2.3. Correlation function

To obtain the function $\phi(d)$, we refer the experimental data from *P*-chr1-2-cell-mouse, denoted as $C(d)$ [Fig. 1(b)]. 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 [Fig. 3(a)]. Larger ϕ leads to larger correlations [Fig. 3(b)]. Fitting the dependence of the simulated correlation with ϕ , we obtained the function $R_0(\phi)$ [Fig. 3(b), red]

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

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 monotonously increasing.

Fig. 3. (Color online) 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 Eq. (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 Eq. (4), with parameters $\alpha = 12, p = 1, \beta = 0.195, \gamma = 0.036, c = 0.13$ and (f) Pearson correlations obtained from experimental data (red) and from model simulation with $\phi(d)$ given by Eq. (4) (blue).

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

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

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 shows good agreement with experimental results [Fig. 3(f)].

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

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 altered from cell to cell.

We apply the procedure to experimental data from different stages of mouse early embryo. The function Eq. (4) is well fitting with different stages of data, however the coefficients can be varied (Table 2, Fig. 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

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

correlation (Appendix C, Table C.1). 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 and changes in the average methylation level and parameters in the correlation function in response to changes in kinetic parameters are given in Fig. 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^+ decreases the methylation level, and decreasing the methyl removing rates m^- and h^- increases 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 in Fig. 4.

From Fig. 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 nonnegligible 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 stages of embryonic cells. This suggests that the correlation may be insensitive with the parameter α . To test this hypothesis, we alter the value α to

Fig. 4. (Color online) Effects of changes in the kinetic 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.

Fig. 5. (Color online) Correlations with $\alpha = 6$ (decreased by 50%, green), $\alpha = 12$ (default, red) or $\alpha = 18$ (increased by 50%, blue) based on different stages of mouse embryonic cells. Other coefficients and kinetic parameters were taken as default values given in Tables 1 and 2.

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 the obtained correlations (Fig. 5). The RMSD between simulated correlations and correlations obtained from experimental data is not sensitive with the parameter α (Table 3). We further apply the procedure to various mouse and human somatic tissue cells where the results give two catalogue cells, either

Table 3. The Root Mean Squared errors of correlations from simulations ($\alpha = 6, 12$ or 18) 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	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
18	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	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
18	0.183	0.218	0.263	0.17	0.184	0.271

$\alpha = 0.001$ or $\alpha = 4$ (Appendix C, Table C.1). 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 kinetic details of the regulation of DNA methylation.

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 embryonic cells and tissues. The results show a nonlinear dependence, except the stage of PGC at E13.5 when the DNA methylation is extremely low due to active demethylation (Fig. 6). Moreover, we test the data from human tissue cells, which reveal similar form of nonlinear dependences (Fig. 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. (Color online) Experimental average methylation level versus constant tail c in the correlation function, data obtained from mouse early embryonic 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-shaped bulb and a constant tail. Here, we discuss the biological sources of the three parts based on mathematical formulation.

It was proposed in Ref. 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

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

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 correlation, either recruitment or movement along DNA, are waiting for further studies.

Interestingly, the bell-shaped bulb is central at 160 bp, right at 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 (Fig. 5). Moreover, various mouse and human somatic tissue cells give two catalogue cells with either small or large α values (Appendix C, Table C.1). 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 types of cells, but is insensitive with the kinetic parameters. A type of nucleosome modification, histone H3 lysine 9 dimethylation (H3K9me2), often shows nucleation sites overlap to CGIs,² and is involved in DNA methylation.^{7,13,15} Moreover, loss of DNA methylation enhances the removal of H3K9me3 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 the 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 reveals 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-supported experimental data, molecular details of the correlation in DNA methylation proposed here are waiting for further experimental studies.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (91430101 and 11272169).

Appendix A. Correlation Patterns are Similar for Different Chromosomes in a Cell

We calculated the correlation patterns of DNA methylation from the mouse embryonic MethylC-Seq data.¹⁷ We choose the experimental data from the parental

Fig. 7. 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).

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 is shown in Fig. 7. From Fig. 7, we can see that the correlation patterns are similar for different chromosomes.

Appendix B. Correlation Patterns are Different at Different Embryo Stages

We calculated the correlation patterns of DNA methylation from different mouse embryonic stages (Fig. B.1). The correlation patterns of paternal and maternal chromosomes are distinct to each other before E6.5 and tend to be consistent in latter stages.

Fig. B.1. (Color online) Pearson correlations between DNA methylation levels at adjacent CpG sites from different embryonic stages of mouse (GEO: GSE56697). The correlation patterns of upper row are paternal chromosomes and the bottom row are maternal chromosomes.

Appendix C. Correlation Function for Mouse and Human Tissues

Table C.1 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 in Figs. C.1 and C.2.

Table C.1. 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

Table C.1. (Continued)

Accession	Description	Methy-level	RMSD	c	α	β	γ	p
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

Fig. C.1. (Color online) Correlations of mouse sample cells from Table C.1. Blue for correlation obtained from experimental data and red for correlation obtained from model simulation.

Fig. C.2. (Color online) Correlations of human sample cells from Table C.1. Blue for correlation obtained from experimental data and red for correlation obtained from model simulation.

References

1. A. Bird, *Genes Dev.* **16**, 6 (2002).
2. X. Chen *et al.*, *Genes Dev.* **26**, 2499 (2012).
3. ENCODE Project Consortium *et al.*, *Science* **306**, 636 (2004).
4. I. F. Davidson *et al.*, *EMBO J.* **35**, 2671 (2016).
5. R. Edgar, M. Domrachev and A. E. Lash, *Nucleic Acids Res.* **30**, 207 (2002).
6. M. Farlik *et al.*, *Cell Stem Cell* **19**, 808 (2016).
7. E. Habibi *et al.*, *Cell Stem Cell* **13**, 360 (2013).
8. J. O. Haerter *et al.*, *Nucleic Acids Res.* **42**, 2235 (2014).
9. N. A. Hathaway *et al.*, *Cell* **149**, 1447 (2012).
10. L. Jiang *et al.*, *Cell* **153**, 773 (2013).
11. M. Kanke *et al.*, *EMBO J.* **35**, 2686 (2016).
12. J. A. Law and S. E. Jacobsen, *Nat. Rev. Genet.* **11**, 204 (2010).
13. B. Lehnertz *et al.*, *Curr. Biol.* **13**, 1192 (2003).
14. J. Lei, W.-C. Lo and Q. Nie, *Ann. Math. Sci. Appl.* **1**, 427 (2016).
15. T. Nakamura *et al.*, *Nature* **486**, 415 (2012).

16. S. Seisenberger *et al.*, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20110330 (2013).
17. L. Wang *et al.*, *Cell* **157**, 979 (2014).
18. H. Wu and Y. Zhang, *Cell* **156**, 45 (2014).
19. H. Zhang *et al.*, *Phys. Rev. Lett.* **112**, 068101 (2014).