Mathematical Models of DNA Methylation Dynamics: Implications for Health and Ageing

Loukas Zagkos[a], Mark Mc Auley[b], Jason Roberts[a], Nikos I. Kavallaris[a,*]

[a] *Department of Mathematics*
School Of Science and Engineering
University of Chester
Thornton Science Park Pool Lane, Ince
Chester CH2 4NU, UK
[b] *Department of Chemical Engineering*
School Of Science and Engineering
University of Chester
Thornton Science Park Pool Lane, Ince
Chester CH2 4NU, UK

Abstract

DNA methylation is a key epigenetic process which has been intimately associated with gene regulation. In recent years growing evidence has associated DNA methylation status with a variety of diseases including cancer, Alzheimer's disease and cardiovascular disease. Moreover, changes to DNA methylation have also recently been implicated in the ageing process. The factors which underpin DNA methylation are complex, and remain to be fully elucidated. Over the years mathematical modelling has helped to shed light on the dynamics of this important molecular system. Although the existing models have contributed significantly to our overall understanding of DNA methylation, they fall short of fully capturing the dynamics of this process. In this paper we develop a linear and nonlinear model which captures more fully the dynamics of the key intracellular events which characterise DNA methylation. In particular the outcomes of our linear model result in gene promoter specific methylation levels which are more biologically plausible than those revealed by previous mathematical models. In addition, our nonlinear model predicts DNA methylation promoter bistability which is commonly observed experimentally. The findings from our models have implications for our current understanding of how changes to the dynamics which underpin DNA methylation affect ageing and health. We also propose how our ideas can be tested in the

*Corresponding author
Email addresses: z.loukas@chester.ac.uk (Loukas Zagkos), m.mcauley@chester.ac.uk (Mark Mc Auley), j.roberts@chester.ac.uk (Jason Roberts), n.kavallaris@chester.ac.uk (Nikos I. Kavallaris)

lab.

Keywords: Gene promoter bistability, Sensitivity analysis, *CpG* island, *CpG* dyads, Hypermethylation, Hypomethylation

2010 MSC: 92B05, 92D20, 93A30, 97M60

HIGHLIGHTS

- The linear model accounts for the overall epigenetic inheritance of DNA methylation patterns and dynamics.
- The nonlinear model can predict the hypomethylated and hypermethylated states of gene promoters.
- DNA methylation dynamics do not alter when the quantity of DNA methylation enzymes changes.

1. Introduction

DNA methylation is considered a key epigenetic mark for mammalian gene expression and regulation (Smith and Meissner, 2013). Intriguingly, a growing body of experimental evidence suggests age-related changes to the methylation status of the genome have a fundamental role to play in healthspan (Jones et al., 2015). For instance, age-related DNA methylation changes are closely correlated with cancer (Klutstein et al., 2016). Moreover, alterations to genomic methylation patterns with age have an emerging role to play in many age-related pathologies including, Alzheimers disease (De Jager et al., 2014), cardiovascular disease (Zhong et al., 2016) and osteoporosis/osteoarthritis (Delgado-Calle et al., 2013). It has also been postulated that genomic methylation status could be used to quantify intrinsic ageing by virtue of a methylation clock, where DNA methylation age is underscored by the cumulative effect of an epigenetic maintenance system (Horvath, 2013; Horvath et al., 2016). Taken together these findings suggest that perturbations to the molecular reactions which are responsible for preserving DNA methylation could significantly impact the trajectory of healthspan and possibly ageing. In mammals these reactions occur primarily at CpG dyads, where the methyl group is attached to the fifth carbon of the cytosine at the CpG site (Bird, 2002; Holliday and Pugh, 1975; Riggs, 1975; Sager and Kitchin, 1975). A CpG site being a $5' - 3'$ Cytosine - Guanine dinucleotide sequence within the DNA molecule with p indicating the phosphate group between the two nucleotides. A dyad consists of two CpG sites, one on each strand of the DNA molecule, while regions of DNA which contain a high frequency of CpG sites are referred to as

CpG islands ($CGIs$). Such islands consist of approximately $500 - 2000$ mainly unmethylated base pairs which cover in the region of 1% of the mammalian genome (Jones and Liang, 2009).

Despite the sparseness of $CGIs$ within the genome they are acutely important due to the association their methylation status has with gene promoter activity. Specifically, hypermethylation, of $CGIs$ is routinely correlated with the transcriptional silencing of gene promoters, a phenomenon which is often a feature of diseases such as cancer (Esteller et al., 2000; Kane et al., 1997). Moreover, advancing age has been associated with the hypermethylation of a wide variety of gene promoters belonging to genes which have been associated with ageing/longevity (Hunter et al., 2012; Scott et al., 2010). Thus, intrinsic ageing would appear to have an effect on the molecular dynamics of DNA methylation. This dynamic process is underpinned by several enzymes (Robertson, 2001). For instance, post replicatively, new CpG sites are attached to the complementary strand of the daughter cells, which are unmethylated. DNA methyltransferase ($DNMT1$) then uses $S-$Adenosyl methionine as a substrate to transfer methyl groups to the DNA molecule (Crider et al., 2012). As $DNMT1$ preferentially acts on hemimethylated DNA it is widely accepted as primarily a maintenance enzyme (Robertson et al., 1999). Consequently, additional enzymes are required for de novo DNA methylation. It is generally regarded that $DNMT3a$ and $DNMT3b$ are the enzymes responsible for this role. In addition to maintenance and de novo methylation, the DNA methylation cycle is counterbalanced by active and passive demethylation (Chen and Riggs, 2011). Passive demethylation usually takes place during replication and it is known that methylation levels can drop following several rounds of replication (Razin and Riggs, 1980). Conversely active methylation is thought to involve Ten-eleven Translocation (TET) dioxygenases, which oxidize the methyl groups of cytosine; a process which eventually culminates with the reintroduction of an unmethylated cytosine into the DNA molecule (Scourzic et al., 2015). Thus, DNA methylation status is the combined result of the complex interactions between maintenance/de novo methylation and passive/active demethylation. Moreover, it can be reasonably argued that the dysregulation of these processes have a role to play in the onset of aberrant gene promoter methylation and unravelling the factors which lead to this dysregulation is of fundamental importance to our understanding of healthspan and ageing.

Mathematical models have been a key to exploring the intricacies of DNA methylation status and its intersection with health (reviewed in Mc Auley et al. 2016). Early attempts to capture mathematically the variability associated with DNA methylation levels are grounded in the re-

duced probabilistic mathematical representation of methylation dynamics (known as the standard/ classical model) as proposed by Pfeifer et al. (1990) and further utilised by Riggs and Xiong (2004). More recently Jeltsch and Jurkowska (2014) extended this mathematical framework by introducing a term to represent the efficiency of maintenance methylation. In addition to this approach Haerter et al. (2013) also provided a probabilistic view of DNA methylation by using the Gillespie algorithm (Gillespie, 1976) to represent this system. Their model attempted to account for experimental data which has consistently shown non-random patterns of DNA methylation in CpG clusters, namely CpG clusters can be hypo-methylated or highly methylated (Illingworth and Bird, 2009). Based on the idea that DNA methylation levels strongly depend on the density of the CpG cluster, Loevkvist and colleagues attempted to mathematically capture the observed *bistability* in DNA methylation patterns of CpG clusters. Therefore, the explicit assumption was made that spatial dependent collaboration between CpG sites occurs. In particular, this idea was supported by the hypothesis that methylated CpG sites are able to recruit methylation enzymes that act on the CpG sites in the neighbouring area (Dodd et al., 2007). Moreover the assumption was made that unmethylated CpGs can recruit demethylation enzymes which predispose neighbouring CpGs to demethylation in that particular region (Loevkvist et al., 2016).

Methylation bistability of gene promoters is a key area of investigation with this work, as we: (i) describe the construction and examination of a deterministic model of DNA methylation, (ii) introduce a mathematical approach for representing a model of DNA methylation, (iii) outline the reasons why our model provides a more complete representation of the dynamics of DNA methylation and (iv) discuss the implications of the findings for our understanding of health and ageing.

2. Models and Methods

2.1. Existing Mathematical Models of DNA Methylation Dynamics.

Recently, McGovern and colleagues created a deterministic set of six homogeneous ordinary differential equations, introducing the incorporation of hydroxymethylation by the Ten-eleven Translocation (TET) enzymes (Scourzic et al., 2015). The authors considered six different states of a CpG dyad which were represented as follows: 1. unmethylated/unmethylated, 2. unmethylated/methylated, 3. methylated/methylated, 4. hydroxymethylated/methylated, 5. hydroxymethylated/hydroxymethylated, 6. hydroxymethylated/unmethylated. Mathematically each state was represented by $x_1(t)$, $x_2(t)$, $x_3(t)$, $x_4(t)$, $x_5(t)$ and $x_6(t)$, respectively. The rates of the transitions between the possible states of the CpG dyads were in turn represented by the rate constants k_j, $j = 1 \dots 6$. The authors also included the cell division rate d and the rate of cell loss l due to cell death in their model. The model is represented by the following set of differential equations.

$$
\begin{aligned}
\frac{dx_1(t)}{dt} &= (d - l - k_1)x_1(t) + dx_2(t) + dx_6(t) \\
\frac{dx_2(t)}{dt} &= k_1 x_1(t) - (l + k_2 + k_3)x_2(t) + 2dx_3(t) + dx_4(t) \\
\frac{dx_3(t)}{dt} &= k_2 x_2(t) - (d + l + k_4)x_3(t) \\
\frac{dx_4(t)}{dt} &= k_4 x_3(t) - (d + l + k_5)x_4(t) \\
\frac{dx_5(t)}{dt} &= k_5 x_4(t) - (d + l)x_5(t) \\
\frac{dx_6(t)}{dt} &= k_3 x_2(t) + dx_4(t) + 2dx_5(t) - lx_6(t).
\end{aligned}
$$

This model was applied to the global methylome and then to local epigenetic regions. Their model was able to predict the relative abundances of unmethylated, hemimethylated, fully methylated and hydroxymethylated CpG dyads in the DNA of cells.

Following the work of McGovern et al, Jeltsch and Jurkowska also focused on methylation dynamics and constructed a single ordinary differential equation, which was based on a classic model by Pfeifer et al (1990). This model included the interactions among $DNMT1, DNMT3a$ and $DNMT3b$. In addition they included hydroxymethylation by the TET protein family enzymes as the first step for active demethylation. The authors monitored the evolution of the fraction of

methylation at a random CpG site $\theta^i_m(t)$. They considered methylation (r^i_{met}) and demethylation (r^i_{demet}) rates, division rate (D) and efficiency of maintenance methylation as a fraction (f^i_{main}). However the authors did not dynamically simulate their model.

$$\frac{d\theta^i_m(t)}{dt} = r^i_{met}\left(1-\theta^i_m(t)\right) - \left(\frac{1}{2}D(1-f^i_{main}) + r^i_{demet}\right)\theta^i_m(t).$$

Although both of the preceding models are noteworthy additions to this field in our opinion, it is necessary to outline some of their limitations. For instance, McGovern and colleagues created a set of homogeneous, linear ordinary differential equations. The main drawback with systems of this form is that they produce only trivial equilibrium points. In other words, only the start of the axis can be a steady state solution of the system, that is, all coordinates equal to zero. This is biologically unreasonable because it would mean that the only stable state of the system happens when there are no CpG dyads of any state in the region of investigation. On the other hand, the Jeltsch and Jurkowska model calculates the fraction of CpG sites and not CpG dyads. This restricts the elucidation of the mechanisms underlying locus-specific DNA methylation. For example, $DNMT1$ shows affinity for hemimethylated CpG dyads, being responsible for maintenance DNA methylation (1). As a consequence this mechanism cannot be supported by this model.

2.2. A Linear Model of DNA Methylation

In order to address these specific limitations we created a new mathematical model. This model retains key elements of both previous models (McGovern et al., 2012; Jeltsch and Jurkowska, 2014). Significantly, however, we introduce three different types of molecules or chemical species of interest; unmethylated CpG dyads, hemimethylated CpG dyads and methylated CpG dyads. An unmethylated CpG dyad is a CpG dyad with none of the two CpG sites methylated. Analogously, a hemimethylated CpG dyad has only one methylated CpG site and the opposing unmethylated and a methylated CpG dyad has both opposing sites methylated, see Figure 1. The number of unmethylated, hemimethylated and methylated CpG dyads is denoted as $x_1(t), x_2(t)$ and $x_3(t)$, respectively. Transitions between the possible states of CpG dyads occur due to the methylation enzymes $DNMT1, DNMT3a$ and $DNMT3b$, demethylation enzymes TET family and DNA replication. k_1 is the methylation rate of unmethylated CpG dyads, k_2 the methylation rate of hemimethylated CpG dyads, k_3 the demethylation rate of hemimethylated CpG dyads, and k_4 is rate of DNA demethylation of methylated CpG dyads and D is the rate of cell division, see also

the diagram in Figure 2. For instance, the constant k_1 represents the rate that a methyl group is attached to one of the two opposing CpG sites of an unmethylated CpG dyad (x_1) to form a hemimethylated CpG dyad (x_2). All rate constants are defined analogously.

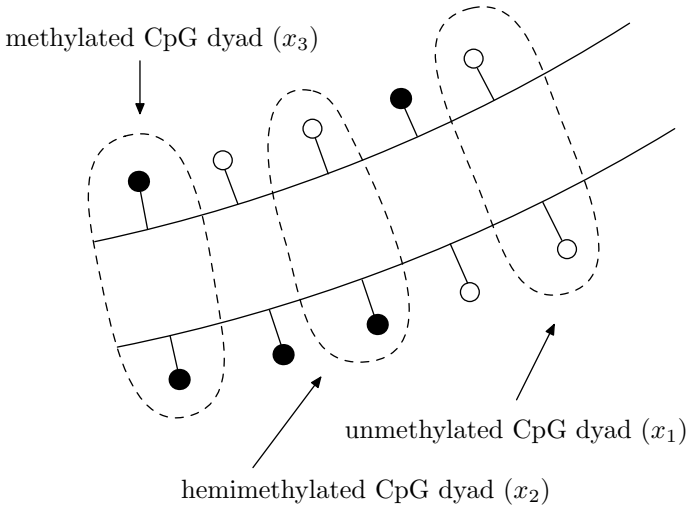

Figure 1: The three different states of a CpG dyad; unmethylated (x_1), hemimethylated (x_2) and methylated (x_3) CpG dyads. A white circle denotes an unmethylated CpG site whereas a black circle represents a methylated CpG site. An unmethylated (x_1) CpG dyad consists of two unmethylated opposite CpG sites, a hemimethylated dyad (x_2) has only one of the two sites methylated and a methylated dyad (x_3) has both opposing sites methylated.

As suggested experimentally interactions among the methylation enzymes $DNMT1, DNMT3a$ and $DNMT3b$ are considered important for maintenance methylation. Rates k_3 and k_4 are necessary to represent active demethylation. Mechanistically this is thought to occur by the TET enzymes hydroxymethylating the cytosine of the CpG sites and eventually reintroducing an unmethylated cytosine into them (Rasmussen and Helin, 2016). Thus, we consider the process of hydroxymethylation to be an important intermediate step leading to the active demethylation of a CpG site.

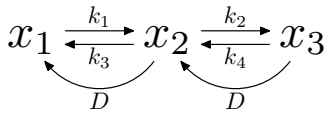

Figure 2: Methylation rates between x_1, x_2 and x_3.

During DNA replication, unmethylated DNA strands bond with the parental strands. Thus,

all methylated CpG dyads of the parental cell produce hemimethylated CpG dyads in the daughter cells (Waterland and Michels, 2007). The hemimethylated parental CpG dyads either become unmethylated, or remain hemimethylated in the daughter cells. Unmethylated dyads remain unmethylated, as outlined in Figure 3.

If the above processes are translated as a set of ordinary differential equations (ODEs), then the following system of ODEs is derived

$$\begin{aligned}
\frac{dx_1(t)}{dt} &= -k_1 x_1(t) + \left(k_3 + \frac{1}{2}D\right)x_2(t) \\
\frac{dx_2(t)}{dt} &= k_1 x_1(t) - \left(k_2 + k_3 + \frac{1}{2}D\right)x_2(t) + \left(k_4 + D\right)x_3(t) \\
\frac{dx_3(t)}{dt} &= k_2 x_2(t) - \left(k_4 + D\right)x_3(t).
\end{aligned} \tag{2.1}$$

The assumption was made that after replication half of the parental hemimethylated CpG dyads become unmethylated in the daughter cells and the other half remain hemimethylated. Thus, there is a $\frac{1}{2}$ coefficient in the terms $\frac{1}{2}Dx_2(t)$.

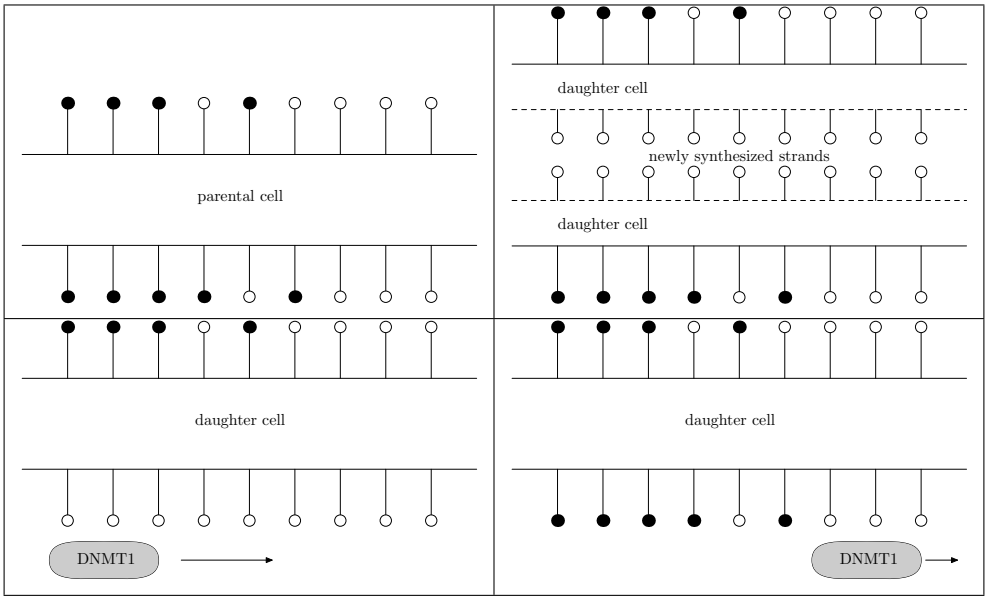

Figure 3: DNA replication and DNA maintenance methylation. Top left: A parental cell with specific methylation levels. Top right: DNA replication. New unmethylated CpG sites attach, synthesizing complementary strands and creating two daughter cells. Bottom left: $DNMT1$ is responsible for maintaining the methylation levels of the daughter cells. Bottom right: $DNMT1$ catalyzes the transfer of methyl groups to the unmethylated CpG sites.

Perturbations to the methylation levels of a variety of gene promoters have been associated with several diseases and ageing (Gopalakrishnan et al.,2008; Jung and Pfeifer, 2015). Therefore, our underlying aim with the proposed model (2.1) is to use it to elucidate the dynamics associated with gene promoter methylation. Moreover, as we are interested in the evolution of a population of CpG sites in a specific region of the genome with fixed length, the total number of CpG dyads has to be constant. Thus, we consider $x_1(t) + x_2(t) + x_3(t) = C$, with $C > 0$. Regardless of the level of methylation in the region of interest, the total number of CpG dyads is determined to be C. The importance of the above relation is that it reduces the size of the set of differential equations by one and leads to a non-homogeneous linear system. Significantly, its solution stabilizes each time to a non trivial equilibrium point, which is biologically consistent since the populations $x_i(t)$, $i = 1, 2, 3$ denote the number of CpG dyads and thus they should remain positive. Substituting the above

equation into the system (2.1), we deduce the equivalent non-homogeneous system

$$\begin{aligned}
\frac{dx_1(t)}{dt} &= -\left(k_1 + k_3 + \frac{1}{2}D\right)x_1(t) - \left(k_3 + \frac{1}{2}D\right)x_3(t) + C\left(k_3 + \frac{1}{2}D\right) \\
\frac{dx_3(t)}{dt} &= -k_2 x_1(t) - \left(k_2 + k_4 + D\right)x_3(t) + Ck_2 \\
x_2(t) &= C - x_1(t) - x_3(t),
\end{aligned} \tag{2.2}$$

which has non-trivial (non-zero) equilibrium solutions.

Gene promoters are differently methylated within the global genome. Indeed, methylation levels in gene promoters depend on the specific type of tissue under consideration (Lokk et al., 2014). It is likely however, that methylation mechanisms, on the contrary, are the same for every CpG island, following the same processes. Thus, it is possible to use our model to quantify the methylation level of a specific CpG island. In the first instance, we apply our model to determine methylation levels in a gene promoter. For that purpose we initially adopt the parameter values suggested by McGovern et al. Hence, $k_1 = 0.012D$, $k_2 = 99D$, $k_3 = 0.11D$, $k_4 = 0.08D$, is used for calculating methylation levels in gene promoters, and $k_1 = 0.205D$, $k_2 = 99D$, $k_3 = 0.04D$, $k_4 = 0.08D$, for estimating the global genome methylation levels. Note that in McGovern's model, the rates k_3 and k_4 denote the hydroxymethylation of hemimethylated and methylated dyads respectively. This meant that the same parameter values for active demethylation could be adopted because hydroxymethylation is the first step of the process that leads to active demethylation. The cell division rate is represented by D.

Although the linear model gives biologically satisfactory solutions, it cannot predict possible changes in the methylation levels due to the onset of disease or ageing. Experimental data suggest $DNMT1$, which is the enzyme primarily responsible for maintaining DNA methylation levels following cell replication, suffers from a decrease in activity during ageing (Li et al., 2010). Consequently it can be assumed that the passive demethylation rate increases. It can also be argued this increase would result in methylation levels in a promoter of a gene in a cancer cell being higher than in the same promoter in a healthy cell. Moreover, recent evidence indicates that methylation states in most average-sized promoters are bistable, namely, they would be either highly methylated or hypomethylated and hardly ever in an intermediate state (Haerter et al., 2014). For these reasons, it is reasonable to assume that the transition rates k_j are not constant but change with respect to time.

In the current subsection a mathematical model is constructed for predicting potential changes in methylation levels due to disease and ageing. A reasonable way to accomplish this is by making the cogent assumption that the methylation rates are functions of the population of CpG dyads, i.e. $k_j = k_j(x_i(t)), i, j = 1, \ldots, 4$. In particular this approach illuminates how de novo methylation, maintenance methylation and active demethylation can possibly change over time. The majority of gene promoters are either in a hypomethylated or in a highly methylated state. In the former state, the majority of the CpG dyads are unmethylated, whilst it has been experimentally found that when CpG dyads are in a highly methylated state, the majority of the CpG dyads are occupied with methyl groups (Haerter et al., 2014). In order to describe a possible transition between the two states it is necessary to appreciate the following biological arguments. If a scenario exists whereby there is an abundance of unmethylated CpG dyads (x_1) and the gene promoter is hypomethylated. It is then biologically plausible that with time unmethylated CpG dyads start to become methylated. This could happen due to fluctuating levels of $DNMT3a$ and $DNMT3b$ (denoted in the model by an increase in k_1 rate). As the number of unmethylated CpG dyads (x_1) drops, the methylation rate k_1 increases. While the number of unmethylated dyads decreases, then the number of hemimethylated dyads (x_2) increases. This can be interpreted as k_1 being a decreasing function of $x_1(t)$ or an increasing function of $x_2(t)$. It is not known if the transition between the two different states is either due to a rise in de novo methylation enzymes ($DNMT3a$ and $DNMT3b$) or because of a decrease in the active demethylation enzymes (TET protein family). To describe the latter, we can assume the similar argument, that as the number of unmethylated CpG dyads (x_1) decreases, the demethylation rate k_3 drops, due to the decrease in the TET enzymes and consequently the number of hemimethylated CpG dyads increases. This can be interpreted as denoting the k_3 rate as an increasing function of $x_1(t)$ or a decreasing function of $x_2(t)$. The exact same argument can be stated for the transition rates k_2 and k_4. A huge increase in methylated CpG dyads can be a result of either an increase in $DNMT1$ maintenance levels or a decrease in TET enzymes, namely either a k_2 increase or a k_4 drop. Therefore, k_2 can be a decreasing function of x_2 or an increasing function of x_3 and k_4 an increasing function of x_2 or a decreasing function of x_3.

Table 1: Transition rates k_{ij} are functions of populations of CpG dyads, x_1, x_2 and x_3. There are two different function selections that comply with the biological evidence. The arrow next to each formula denotes an increasing or decreasing function of x_i. All parameters k_{ij} are positive.

Selection I	**Selection II**
$\mathbf{k_j}$ **functions of** $\mathbf{x_1, x_3}$	$\mathbf{k_j}$ **functions of** $\mathbf{x_2}$
$k_1(x_1) = k_{11} - k_{12}x_1^2 \quad (\searrow)$	$k_1(x_2) = k_{11} + k_{12}x_2^2 \quad (\nearrow)$
$k_2(x_3) = k_{21} + k_{22}x_3^2 \quad (\nearrow)$	$k_2(x_2) = k_{21} - k_{22}x_2^2 \quad (\searrow)$
$k_3(x_1) = k_{31} + k_{32}x_1^2 \quad (\nearrow)$	$k_3(x_2) = k_{31} - k_{32}x_2^2 \quad (\searrow)$
$k_4(x_3) = k_{41} - k_{42}x_3^2 \quad (\searrow)$	$k_4(x_2) = k_{41} + k_{42}x_2^2 \quad (\nearrow)$

Following this rationale, two different mathematical formulae are introduced to account for the transition rate functions, (see Table 2).]A nonlinear model describing gene promoter bistability.

Although the linear model gives biologically satisfactory solutions, it cannot predict possible changes in the methylation levels due to the onset of disease or ageing. Experimental data suggest $DNMT1$, which is the enzyme primarily responsible for maintaining DNA methylation levels following cell replication, suffers from a decrease in activity during ageing (Li et al., 2010). Consequently it can be assumed that the passive demethylation rate increases. It can also be argued this increase would result in methylation levels in a promoter of a gene in a cancer cell being higher than in the same promoter in a healthy cell. Moreover, recent evidence indicates that methylation states in most average-sized promoters are bistable, namely, they would be either highly methylated or hypomethylated and hardly ever in an intermediate state (Haerter et al., 2014). For these reasons, it is reasonable to assume that the transition rates k_j are not constant but change with respect to time.

In the current subsection a mathematical model is constructed for predicting potential changes in methylation levels due to disease and ageing. A reasonable way to accomplish this is by making the cogent assumption that the methylation rates are functions of the population of CpG dyads, i.e. $k_j = k_j(x_i(t)), i, j = 1, \ldots, 4$. In particular this approach illuminates how de novo methylation, maintenance methylation and active demethylation can possibly change over time. The majority of gene promoters are either in a hypomethylated or in a highly methylated state. In the former

state, the majority of the CpG dyads are unmethylated, whilst it has been experimentally found that when CpG dyads are in a highly methylated state, the majority of the CpG dyads are occupied with methyl groups (Haerter et al., 2014). In order to describe a possible transition between the two states it is necessary to appreciate the following biological arguments. If a scenario exists whereby there is an abundance of unmethylated CpG dyads (x_1) and the gene promoter is hypomethylated. It is then biologically plausible that with time unmethylated CpG dyads start to become methylated. This could happen due to fluctuating levels of $DNMT3a$ and $DNMT3b$ (denoted in the model by an increase in k_1 rate). As the number of unmethylated CpG dyads (x_1) drops, the methylation rate k_1 increases. While the number of unmethylated dyads decreases, then the number of hemimethylated dyads (x_2) increases. This can be interpreted as k_1 being a decreasing function of $x_1(t)$ or an increasing function of $x_2(t)$. It is not known if the transition between the two different states is either due to a rise in de novo methylation enzymes ($DNMT3a$ and $DNMT3b$) or because of a decrease in the active demethylation enzymes (TET protein family). To describe the latter, we can assume the similar argument, that as the number of unmethylated CpG dyads (x_1) decreases, the demethylation rate k_3 drops, due to the decrease in the TET enzymes and consequently the number of hemimethylated CpG dyads increases. This can be interpreted as denoting the k_3 rate as an increasing function of $x_1(t)$ or a decreasing function of $x_2(t)$. The exact same argument can be stated for the transition rates k_2 and k_4. A huge increase in methylated CpG dyads can be a result of either an increase in $DNMT1$ maintenance levels or a decrease in TET enzymes, namely either a k_2 increase or a k_4 drop. Therefore, k_2 can be a decreasing function of x_2 or an increasing function of x_3 and k_4 an increasing function of x_2 or a decreasing function of x_3.

Table 2: Transition rates k_{ij} are functions of populations of CpG dyads, x_1, x_2 and x_3. There are two different function selections that comply with the biological evidence. The arrow next to each formula denotes an increasing or decreasing function of x_i. All parameters k_{ij} are positive.

Selection I	**Selection II**
$\mathbf{k_j}$ **functions of** $\mathbf{x_1, x_3}$	$\mathbf{k_j}$ **functions of** $\mathbf{x_2}$
$k_1(x_1) = k_{11} - k_{12}x_1^2 \quad (\searrow)$	$k_1(x_2) = k_{11} + k_{12}x_2^2 \quad (\nearrow)$
$k_2(x_3) = k_{21} + k_{22}x_3^2 \quad (\nearrow)$	$k_2(x_2) = k_{21} - k_{22}x_2^2 \quad (\searrow)$
$k_3(x_1) = k_{31} + k_{32}x_1^2 \quad (\nearrow)$	$k_3(x_2) = k_{31} - k_{32}x_2^2 \quad (\searrow)$
$k_4(x_3) = k_{41} - k_{42}x_3^2 \quad (\searrow)$	$k_4(x_2) = k_{41} + k_{42}x_2^2 \quad (\nearrow)$

Following this rationale, two different mathematical formulae are introduced to account for the transition rate functions, (see Table 2). The arrow next to each formula denotes an increasing or decreasing function of x_i. These reactions are akin to second order kinetics which are common in biochemical systems. Both selections for the transition rates were used to obtain the nonlinear system. The nonlinear system corresponding to the selection II transition rates was only able to predict one global equilibrium point for the CpG populations, thus leaving us with the same limitation as the linear system. Therefore, if Selection I is selected, we obtain the following nonlinear system

$$\begin{aligned} \frac{dx_1(t)}{dt} &= -A_1(x_1(t))x_1(t) - A_2(x_1(t))x_3(t) + A_3(x_1(t))C \\ \frac{dx_3(t)}{dt} &= -B_1(x_3(t))x_1(t) - B_2(x_3(t))x_3(t) + B_3(x_3(t))C \\ x_2(t) &= C - x_1(t) - x_3(t), \end{aligned} \tag{2.3}$$

where

$$
\begin{aligned}
A_1(x_1(t)) &= k_{11} - k_{12}x_1^2(t) + k_{31} + k_{32}x_1^2(t) + \frac{1}{2}D,\\
A_2(x_1(t)) &= k_{31} + k_{32}x_1^2(t) + \frac{1}{2}D,\\
A_3(x_1(t)) &= k_{31} + k_{32}x_1^2(t) + \frac{1}{2}D,\\
B_1(x_3(t)) &= k_{21} + k_{22}x_3^2(t),\\
B_2(x_3(t)) &= k_{21} + k_{22}x_3^2(t) + k_{41} - k_{42}x_3^2(t) + D,\\
B_3(x_3(t)) &= k_{21} + k_{22}x_3^2(t).
\end{aligned}
$$

The specific selection of the rate functions k_2 and k_3, is inspired by biological mechanisms and as it is shown below, leads to the bistability which has been observed experimentally in CpG clusters (Haerter et al., 2014). In particular, it has been suggested that due to the interaction between neighbouring CpG sites, methylated CpG sites affect the methylation status of nearby unmethylated sites; hence the rate of methylation k_2 increases as the population of methylated dyads x_3 grows. Conversely, an abundance of unmethylated CpG sites in a CpG cluster influences the demethylation of close methylated sites; thus, the demethylation rate k_3 increases while unmethylated CpG dyads x_1 increase. In addition, if we consider that intrinsic ageing has the potential to dysregulate $DNMT1$ this will result in a decrease in global methylation levels. In other words, a drop in the value of k_2 results in a concomitant drop in globally methylated CpG dyads x_3. Both functions k_2 and k_3, therefore, are described in terms of x_3 and x_1, respectively. The same argument can be stated as a reasonable reasoning for the selection of the rate functions k_1 and k_4. In both cases, a decrease in $DNMT3a$ and $DNMT3b$ results primarily in an abundance of unmethylated CpG dyads x_1, so k_1 would be more appropriately denoted as a decreasing function of x_1. The same arguments are valid for k_4 as well.

The analysis of the nonlinear model (2.3) was done by using Matlab 2017a. For the delivered simulations a specific interval of values was determined for each transition rate function k_j so that the model predicts the observed bistability of gene promoters. For that purpose and for a CpG cluster of a hundred CpG dyads size, i.e. $C = 100$, transition rate functions were selected in the following intervals, $1.9 \leq k_1 \leq 2.1$, $10 \leq k_2 \leq 110$, $1 \leq k_3 \leq 100$ and $2 \leq k_4 \leq 4$. In order to determine the values of the parameters, the expression of each rate function was substituted into the intervals. As a result the values listed in Table 3 were determined.

Table 3: Nonlinear model parameter values

k_{11}	k_{12}	k_{21}	k_{22}	k_{31}	k_{32}	k_{41}	k_{42}
2.1	2×10^{-5}	10	10^{-2}	1	10^{-2}	4	2×10^{-4}

3. Results and discussion

Mathematical analysis was conducted for both the linear and the nonlinear model. Equilibrium points were calculated for the linear system and stability analysis for both of the systems was delivered by constructing the phase plane portraits which provide graphically a representation of the solutions. In addition sensitivity analysis was performed to assess the impact of parameter perturbations. Matlab 2017a was used for producing the phase plane portraits and for performing the sensitivity analysis for parameters for the nonlinear model.

3.1. Linear model outcomes

We examined the equilibrium state of the system $\mathbf{x}(t) = (x_1(t), x_2(t), x_3(t))^T$; namely, the equilibrium point of the system was calculated when the condition $\frac{d}{dt}\mathbf{x}(t) = 0$ is solved. We then obtained the following expression for the equilibrium point with respect to the parameters

$$
\begin{aligned}
x_1^{ss} &= \frac{C(2k_3 + D)(D + k_4)}{D(2k_1 + 2k_3 + k_4) + 2(k_1k_2 + k_1k_4 + k_3k_4) + D^2}, \\
x_2^{ss} &= \frac{2k_1D + 2k_1k_4}{D(2k_1 + 2k_3 + k_4) + 2(k_1k_2 + k_1k_4 + k_3k_4) + D^2}, \\
x_3^{ss} &= \frac{2Ck_1k_2}{D(2k_1 + 2k_3 + k_4) + 2(k_1k_2 + k_1k_4 + k_3k_4) + D^2}.
\end{aligned}
$$

Mathematical analysis showed that all three components of the equilibrium point are always positive and less than $C = 100$, for any positive values of the parameters. In addition, stability analysis indicated that the equilibrium point is stable for all positive selection of the parameter values. The latter actually entails that in a region of investigation, like a CpG island, our model predicts that no matter what the initial methylation levels are, the number of unmethylated, hemimethylated and methylated CpG dyads will always converge to a steady-state value inside the interval $[0, 100]$.

In order to examine the model in a specific region of the genome, it would be necessary to have clear experimental data. In the absence of experimental data, the assumption that we were

working with a promoter of a gene was made. Data suggests that the values of the parameters should be considered as in the McGovern et al. paper (McGovern et al., 2012) for local epigenetic regions. Thus, $k_1 = 0.012D$, $k_2 = 99D$, $k_3 = 0.11D$, $k_4 = 0.08D, D = 1$ and $C = 100$. were selected. As shown in Figure 4, our model predicts that the equilibrium methylation state for this promoter would be: 38.2% unmethylated, 0.8% hemimethylated and 61% methylated CpG dyads. The fact the equilibrium point is stable ensures the heredity of post replicative maintenance methylation. Moreover, any perturbation in the availability of $DNMT1, DMNT3a, DNMT3b$ or the TET family enzymes during methylation or active demethylation is temporary and will eventually terminate. Therefore, the linear model accounts for the overall epigenetic inheritance of DNA methylation patterns and dynamics.

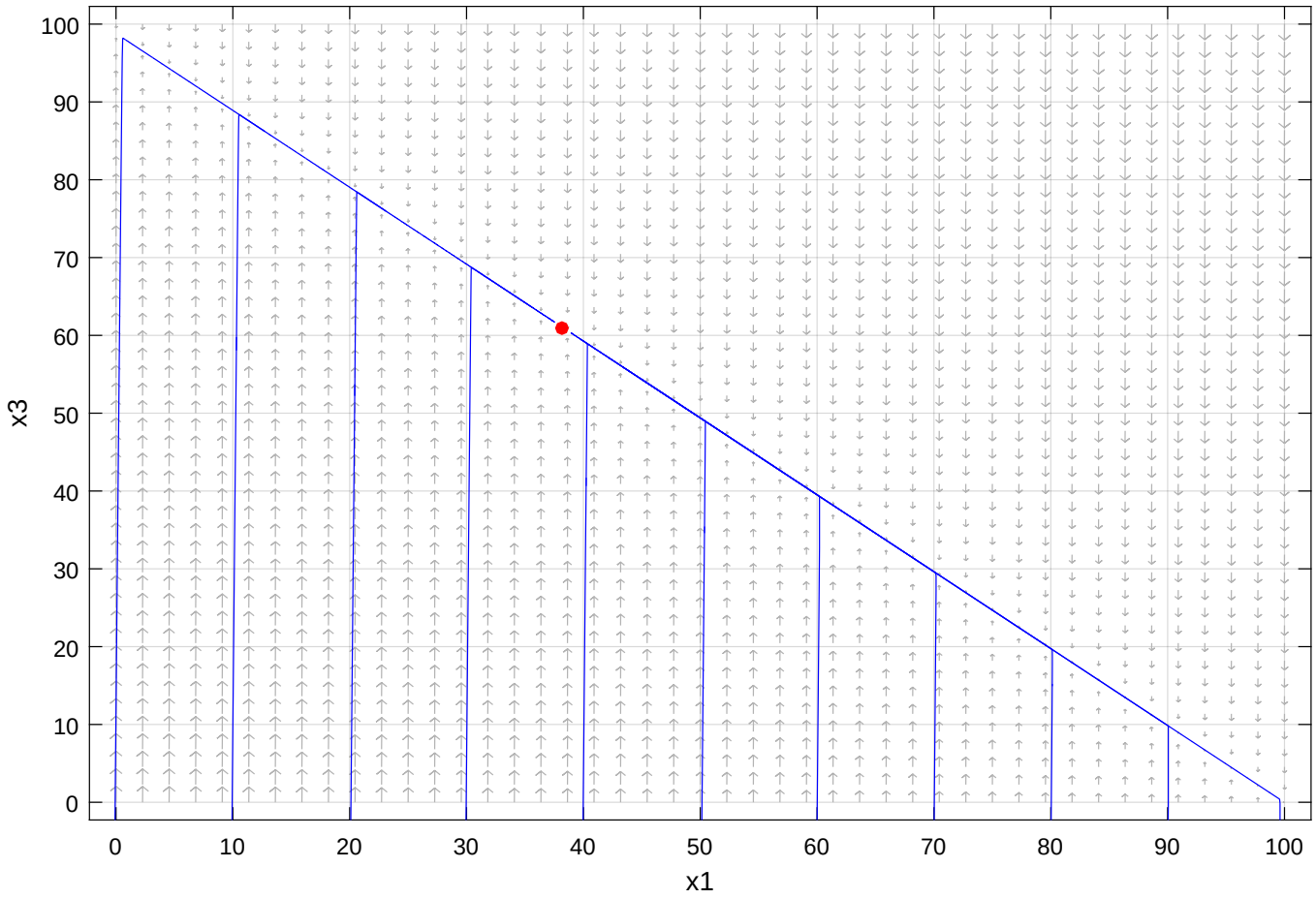

Figure 4: Phase plane portrait of solutions for the linear model (2.2). The above figure shows the percentage of unmethylated CpG dyads versus the unmethylated CpG dyads. For $k_1 = 0.012D$, $k_2 = 99D$, $k_3 = 0.11D$, $k_4 = 0.22D, D = 1$ and $C = 100$, the equilibrium point is $(x_1, x_3) = (38.2, 61)$.

3.2. Nonlinear model outcomes

The nonlinear model (2.3) was created to overcome some significant weaknesses of the linear model. Due to the presence of the nonlinear terms we had to use numerical methods to solve it. The 4th order Runge-Kutta method was implemented for defining the equilibrium points of the system, and for performing sensitivity analysis for the parameters perturbation.

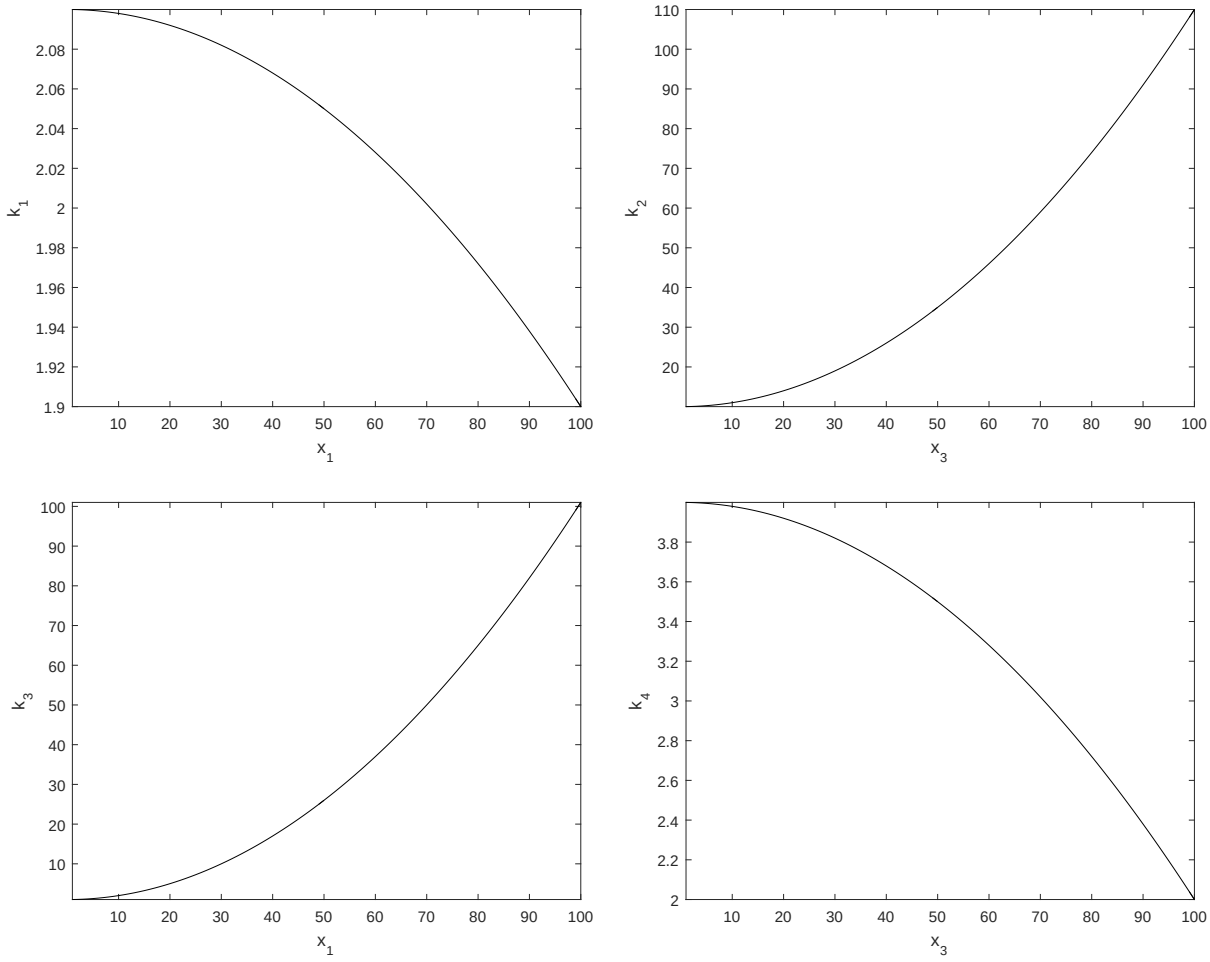

Figure 5: Transition rate functions k_j versus populations x_i for a *CpG* cluster of $C = 100$ *CpG* dyads in total. $k_1(x_1) = k_{11} - k_{12}x_1^2$, $k_2(x_3) = k_{21} + k_{22}x_3^2$, $k_3(x_1) = k_{31} + k_{32}x_1^2$ and $k_4(x_3) = k_{41} - k_{42}x_3^2$. with $k_{11} = 2.1$, $k_{12} = 2 \times 10^{-5}$, $k_{21} = 10$, $k_{22} = 10^{-2}$, $k_{31} = 1$, $k_{32} = 10^{-2}$, $k_{41} = 4$, $k_{42} = 2 \times 10^{-4}$.

3.2.1. Nonlinear Phase Plane Portrait

Notably the transition rates are not constants anymore but vary over time. We selected rates as functions of population of *CpG* dyads, namely $k_1(x_1) = k_{11} - k_{12}x_1^2$, $k_2(x_3) = k_{21} + k_{22}x_3^2$, $k_3(x_1) = k_{31} + k_{32}x_1^2$ and $k_4(x_3) = k_{41} - k_{42}x_3^2$, (see Figure 5).

For computational purposes and for the reasons explained in the previous section, we focus on the case of a gene promoter of $C = 100$ CpG dyads length, for which experimental evidence suggests that the following parameters $k_{11} = 2.1$, $k_{12} = 2 \times 10^{-5}$, $k_{21} = 10$, $k_{22} = 10^{-2}$, $k_{31} = 1$, $k_{32} = 10^{-2}$, $k_{41} = 4$, $k_{42} = 2 \times 10^{-4}$ should be considered. Then the model produces two stable equilibrium points $(x_1, x_2, x_3) = (2.2, 3.1, 94.7)$ and $(x_1, x_2, x_3) = (93.9, 2, 4.1)$, see Figure 6. This means that there are two possible scenarios regarding the eventual methylation level of the promoter. In the first case the promoter is hypermethylated, as the first equilibrium point dictates, with 94.7% of the total CpG dyads within the promoter being methylated, and only 2.2% remain unmethylated. The second scenario corresponds to the promoter being in a hypomethylated state, with only 4.1% of the total CpG dyads methylated and 93.9% unmethylated. Therefore it could be reasonably inferred that the first scenario represents the hypermethylation of the promoter, possibly due to ageing. The promoter we consider was assumed to be a generic promoter akin to a homebox gene. If the gene is responsible for suppressing genetic mutations in the cell, its silencing would potentially lead to cancer development. Depending on the gene function, silencing due to hypermethylation could potentially lead to different diseases. Analogously, the second case represents the normal hypomethylated state of the promoter, where the gene is active. There is an intermediate region which separates these two stable regions. This region denotes a methylation state threshold where if exceeded, methylation levels change properly in the promoter.

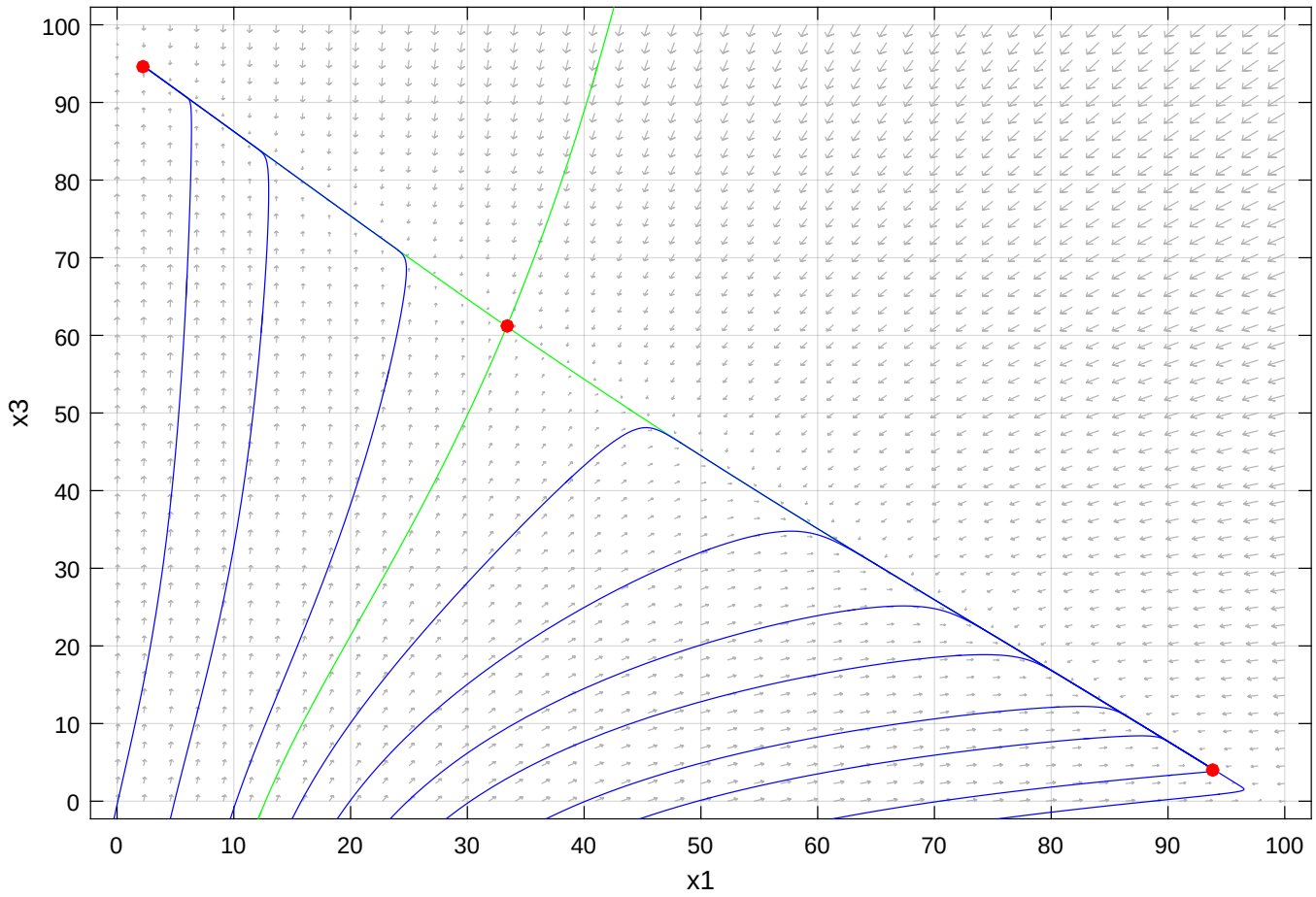

Figure 6: Phase plane portrait for the solutions of the nonlinear model (2.3). The above figure shows the percentage of unmethylated CpG dyads versus the unmethylated CpG dyads. Assuming that $k_{11} = 2.1$, $k_{12} = 2 \times 10^{-5}$, $k_{21} = 10$, $k_{22} = 10^{-2}$, $k_{31} = 1$, $k_{32} = 10^{-2}$, $k_{41} = 4$, $k_{42} = 2 \times 10^{-4}$, two stable equilibrium points occur; $(x_1, x_3) = (2.2, 94.7)$ and $(x_1, x_3) = (93.9, 4.1)$.

3.2.2. Sensitivity analysis

A sensitivity analysis for the parameters of the system was performed. This approach allowed us to identify how a variation in the value of the parameters of the model would influence the solution of the system. In other words, we used a sensitivity analysis to examine how a possible change in the availability of the methylation and demethylation enzymes would affect the methylation levels and the dynamics of the model under investigation. For that purpose we increased the parameters by $5\%, 10\%, 25\%, 50\%, 100\%$ and decreased them by $5\%, 10\%, 25\%, 50\%, 75\%$. Next we calculated

Table 4: Sensitivity analysis for the parameters of the nonlinear model (2.3). Each parameter was increased by 50%. The percentage change was calculated for each variable of the system. Parameters k_{11} and k_{32} had the most significant impact on both hemimethylated and methylated CpG dyads whereas k_{21} and k_{41} cause a noticeable change in methylated dyads.

50% change in parameter	% change in x_1	% change in x_2	% change in x_3
k_{11}	4.2	61.7	64.7
k_{12}	0.3	4.9	5
k_{21}	2.4	3.1	54.4
k_{22}	0.04	0.3	0.9
k_{31}	0.04	0.8	0.6
k_{32}	2.3	34.8	34.4
k_{41}	1.3	1.8	29.8
k_{42}	0.001	0.2	0.04

the percentage change in the solution for x_1, x_2 and x_3 and the results are presented in Table 4. After ranking the results, we calculated the Spearman's rank correlation coefficient, usually called Spearman's r.

The rankings between the various increases and decreases in parameters are compared for each variable, x_i. It is found that the order of the impact of parameters on x_2 and x_3 between 5% increases and 100% increases was exactly the same, see Table 5. This confirmed that none of the parameters of the system has a switching on effect, that is, when they reach at a particular threshold, their effect on the system is greatly increased, in relation to the other parameters. Having identified this, it is inferred that the DNA methylation mechanisms remained unaltered despite changes to the parameters. For example, the sensitivity analysis shows that a change in the quantity of the de novo methylation enzymes $DNMT3a$ and $DNMT3b$ causes a significant change in the number of hemimethylated and methylated CpG dyads, but not in the unmethylated dyads. The Spearman's rank analysis also indicates that hemimethylated and methylated dyads are more sensitive to a $DNMT3a$ and $DNMT3b$ change compared to the unmethylated ones.

Table 5: The Spearman's rank correlation coefficient r calculated indicates a positive correlation for $r = 1$ and no correlation for $r = 0$.

	x_1	x_2	x_3
Spearman's r	0.987	1	1

In order to explore the mechanics of DNA methylation further, we compare all x_i rankings to identify a correlation between them. Interestingly, comparing x_1 rankings with x_3 rankings yielded a positive correlation, i.e. $r = 1$. This means that the parameters of the nonlinear system (2.3) affect x_1 and x_3 with the same order of magnitude. In addition, both x_1 rankings compared to x_2 and x_2 rankings compared to x_3 yield a significant correlation, $r = 0.89$. Generally, changes in the quantity of the enzymes taking part in the DNA methylation processes cause changes in the number of the three different states of CpG dyads with the same order of magnitude.

4. Conclusion

Based on the biological assumptions made in pre-existing models in the literature, two mathematical models were constructed; the linear model (2.2) and the nonlinear one (2.3). Both models were built to address mathematical and biological limitations of the pre-existing models. Our key idea based on the fact that the total number of CpG dyads in a region of investigation is always constant. The linear model (2.2) demonstrated the methylation levels in any region of the genome, from CpG islands, to the whole genome. Each transition rate k_j was considered constant over time. The linear model, however, was unable to predict possible changes in methylation levels caused by perturbations in the enzymes taking part in the methylation processes. In order to address this limitation we constructed the nonlinear model (2.3). Transition rates, for the latter model, were considered as functions of populations x_i and thus varied over time. The outcomes of the nonlinear model are in agreement with experimental work, namely, the observed bistable methylation states found in gene promoters can be predicted.

As the existing models (McGovern et al., 2012) (Jeltsch & Jurkowska, 2014), both linear and nonlinear model were constructed considering continuous and deterministic dynamics. The advantage of the assumption that the transition rates vary over time is that the obtained nonlinear system possesses two stable equilibrium points, each one representing a stability state of methylation in the region of investigation. Past models were only able to attribute a single - sometimes trivial - equilibrium point, failing to portray the potentiality of change in methylation levels of CpG islands. Furthermore, McGovern and colleagues introduced the incorporation of hydroxymethylation by the Ten-eleven Translocation (TET) enzymes; we considered active demethylation rates instead, to account for the action of TET enzymes. In addition, another key difference between our nonlinear model and the model by Jeltsch & Jurkowska is the consideration of CpG dyads as the variables of our system, in contrast to the proposal of regarding CpG sites as the main variables in their work. In our opinion, investigating the methylation levels in CpG dyads provides a better understanding of the processes underlying DNA methylation; Moreover, we were able to dynamically simulate our model, mathematically analyze the stability of the equilibrium points and examine the sensitivity of the parameters of the system.

Additionally, one of the key predictions of our model it that DNA methylation dynamics do not alter when the quantity of DNA methylation enzymes changes. A worthwhile way to test the validity of this prediction would be to perform methylation assays using the three methylation enzymes, $DNMT1, DNMT3a$ and $DNMT3b$. In fact methylation assays for each of these enzymes have been identified (Poh et al., 2016). It would be relatively straightforward from an experimental standpoint to design an investigation which quantifies DNA methylation levels based on the different cellular concentration of each enzyme or to design an experiment which measures the level of DNA methylation in response to varying concentrations of different combinations of these enzymes. This experiment could be enriched further if conducted in tandem with a gene promoter whereby bistability has been observed. For instance, there is an array of genomic regions which display this phenomenon (Zhang et al., 2009)

Although, our nonlinear model (2.3) provides new insights into the dynamics of DNA methylation dynamics, it is important to recognize that it has a number of limitations that need to be addressed. Notably intrinsic and extrinsic noise are acknowledged factors intracellularly. Therefore, a deterministic model, as (2.3), can predict up to a level of certainty the levels of methylation by calculating the values of the variables of the system at any given time. Thus, uncertainty could be introduced in the model to describe the presence of intrinsic and extrinsic noise. This way it would be possible for models to account for the stochastic nature in cell dynamics. In addition, continuous and deterministic models are very convenient to construct, but they have limitations for biological processes and specifically for DNA methylation dynamics. In this work we considered cell division as a continuous process occurring over time although it only occurs in discrete time steps.

Moreover, it is possible to expand our model to include additional biological mechanisms. Firstly, a worthwhile addition to the model would be to incorporate the functional forms which describe the enzymatic reactions underpinning the rate laws in our model. For instance, the mathematics which characterize the enzymatic mechanism of action of $DNMT1$ have previously been proposed (Svedruzic, 2008). Secondly, we propose including the dynamics of folate cycle in our model, as ultimately the folate cycle is the source of the methyl groups which attach to the DNA molecule. Several ODE models of the folate cycle have been built previously and scope exists for connecting our model to these (Sora and Mc Auley, 2016; Duncan et al., 2013; Nijhout et al., 2004; Reed et al., 2006).

References

[Bird 2002] A. Bird, *DNA* methylation patterns and epigenetic memory, Genes Dev 16 (2002) 6–21, doi:10.1101/gad.947102

[Chen and Riggs 2011] Z. Chen and A.D. Riggs, *DNA* Methylation and Demethylation in Mammals, J Biol Chem 286 (2011) 18347-18353, doi:10.1074/jbc.R110.205286

[Crider et al. 2012] K.S. Crider, T.P. Yang, R.J. Berry and L. B. Bailey, Folate and *DNA* methylation: a review of molecular mechanisms and the evidence for folate's role, Adv Nutr 3 (2012) 21–38, doi:10.3945/an.111.000992

[De Jager et al. 2014] P. L. De Jager, G. Srivastava, K. Lunnon, J. Burgess, L. C. Schalkwyk, L. Yu, M. L. Eaton, B. T. Keenan, J. Ernst, C. McCabe, A. Tang, T. Raj, J. Replogle, W. Brodeur, S. Gabriel, H. S. Chai, C. Younkin, F. Zou, M. Szyf, C. B. Epstein, J. A. Schneider, B. E. Bernstein, A. Meissner, N. Ertekin-Taner, L. B. Chibnik, M. Kellis, J. Mill and D. A. Bennett, Alzheimer's disease: early alterations in brain *DNA* methylation at ANK1, BIN1, RHBDF2 and other loci, Nat Neurosci 17 (2014) 1156–63, doi:10.1038/nn.3786

[Delgado-Calle et al. 2013] J. Delgado-Calle, A. F. Fernandez, J. Sainz, M. T. Zarrabeitia, C. Sanudo, R. Garcia-Renedo, M. I. Perez-Nunez, C. Garcia-Ibarbia, M. F. Fraga and J. A. Riancho, Genome-wide profiling of bone reveals differentially methylated regions in osteoporosis and osteoarthritis, Arthritis Rheum 65 (2013) 197–205, doi:10.1002/art.37753

[Dodd et al. 2007] I. B. Dodd, M. A. Micheelsen, K. Sneppen and G. Thon, Theoretical analysis of epigenetic cell memory by nucleosome modification, Cell 129 (2007) 813–822, doi:10.1016/j.cell.2007.02.053

[Duncan et al. 2013] Duncan TM, Reed MC, Nijhout HF. A population model of folate-mediated one-carbon metabolism. Nutrients 2013; 5: 245774, doi:10.3390/nu5072457

[Esteller et al. 2000] M. Esteller, J. M. Silva, G. Dominguez, X. Matias-Guiu, E. Lerma, E. Bussaglia, J. Prat, I. C. Harkes, E. A. Repasky, E. Gabrielson, M. Schutte, S. B. Baylin and J. G. Herman, Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors, J Natl Cancer Inst 92 (2000) 564–569

[Gillespie 1976] D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, Journal of Computational Physics, 22 (1976) 403–434, doi:10.1016/0021-9991(76)90041-3

[Gopalakrishnan et al. 2009] S. Gopalakrishnan, E. O. Van Emburgh and K. D. Robertson, *DNA* methylation in development and human disease, Mutat Res 647 (2009) 30–38, doi:10.1016/j.mrfmmm.2008.08.006

[Haerter et al. 2014] J. O. Haerter, C. Loevkvist, I. B. Dodd and K. Sneppen, Collaboration between CpG sites is needed for stable somatic inheritance of *DNA* methylation states, Nucl Acids Res 42 (20014) 2235–2244, doi:10.1093/nar/gkt1235

[Holliday and Pugh 1975] R. Holliday and J. E. Pugh, *DNA* modification mechanisms and gene activity during development, Science 187 (1975) 226–232, doi:10.1126/science.187.4173.226

[Horvath 2013] S. Horvath, *DNA* methylation age of human tissues and cell types, Genome Biol 14 (2013) R115, doi:10.1186/gb-2013-14-10-r115

[Horvath et al. 2016] S. Horvath, M. Gurven, M. E. Levine, B. C. Trumble, H. Kaplan, H. Allayee, B. R. Ritz, B. Chen, A. T. Lu, T. M. Rickabaugh, B. D. Jamieson, D. Sun, S. Li, W. Chen, L. Quintana-Murci, M. Fagny, M. S. Kobor, P. S. Tsao, A. P. Reiner, K. L. Edlefsen, D. Absher, T. L. Assimes, An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease, Genome Biol, 17 (2016) 171, doi:10.1186/s13059-016-1030-0

[Hunter et al. 2012] A. Hunter, P. A. Spechler, A. Cwanger, Y. Song, Z. Zhang, G. S. Ying, A. K. Hunter, E. Dezoeten, J. L. Dunaief, *DNA* methylation is associated with altered gene expression in AMD, Invest Ophthalmol Vis Sci 53 (2012) 2089–2105, doi:10.1167/iovs.11-8449

[Illingworth and Bird 2009] R. S. Illingworth and A. P. Bird, CpG islands–"a rough guide", FEBS Lett 583 (2009) 1713–1720, doi:10.1016/j.febslet.2009.04.012

[Jeltsch and Jurkowska 2014] A. Jeltsch and R. Z. Jurkowska, New concepts in *DNA* methylation, Cell Press 39 (2014) 1–9, doi:10.1016/j.tibs.2014.05.002

[Jones et al. 2015] M. J. Jones, S. J. Goodman and M. S. Kobor, *DNA* methylation and healthy human aging, Aging Cell, 14 (2015) 924–932, doi:10.1111/acel.12349

[Jones and Liang 2009] P. A. Jones and G. Liang, Rethinking how *DNA* methylation patterns are maintained, Nat Rev Genet 10 (2009) 805–811, doi:10.1038/nrg2651

[Jung and Pfeifer 2015] M. Jung and P. P. Pfeifer, Aging and *DNA* methylation, BMC Biol 13 (2015) 1–8, doi:10.1186/s12915-015-0118-4

[Kane et al. 1997] M. F. Kane, M. Loda, G. M. Gaida, J. Lipman, R. Mishra, H. Goldman, J. M. Jessup and R. Kolodner, Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines, Cancer Res 57 (1997) 808–811

[Klutstein et al. 2016] M. Klutstein, D. Nejman, R. Greenfield and H. Cedar, *DNA* Methylation in Cancer and Aging, Cancer Res 76 (2016) 3446–3550, doi:10.1158/0008-5472.CAN-15-3278

[Landan et al. 2012] G. Landan, N. M. Cohen, Z. Mukamel, A. Bar, A. Molchadsky, R. Brosh, S. Horn-Saban, D. A. Zalcenstein, N. Goldfinger, A. Zundelevich, E. N. Gal-Yam, V. Rotter and A. Tanay, Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues, Nat Genet, 44 (2012) 1207–1214, doi:10.1038/ng.2442

[Li et al. 2010] Y. Li, Y. Liu, F. M. Strickland and B. Richardson, Age-dependent decreases in *DNA* methyltransferase levels and low transmethylation micronutrient levels synergize to promote overexpression of genes implicated in autoimmunity and acute coronary syndromes, Exp Gerontol. 45 (2010), 312–322, doi:10.1016/j.exger.2009.12.008

[Loevkvist et al. 2016] C. Loevkvist, I. B. Dodd, K. Sneppen and J. O. Haerter, *DNA* methylation in human epigenomes depends on local topology of CpG sites, Nucleic Acids Research 44 (2016) 1–9, doi:10.1093lnarlgkw124

[Lokk et al. 2014] K. Lokk, V. Modhukur, B. Rajashekar, K. Mrtens, R. Mgi, R. Kolde, M. Koltina, T. K. Nilsson, J. Vilo, A. Salumets and N. Tnisson, *DNA* methylome profiling of human tissues identifies global and tissue-specific methylation patterns, Genome Biol 15 (2014) 1–9, doi:10.1186/gb-2014-15-4-r54

[Mc Auley et al. 2016] M. T. Mc Auley, K. M. Mooney and J. E. Salcedo-Sora, Computational modelling folate metabolism and DNA methylation: implications for understanding health and ageing, Brief Bioinform. (2016) 1–15, doi:10.1093/bib/bbw116

[McGovern et al. 2012] A. P. McGovern, B. E. Powell and J. T. Chevassut, A dynamic multi-compartmental model of DNA methylation with demonstrable value in hematological malignancies, Journal of Theoretical Biology 310 (2012) 14–20, doi:10.1016/j.jtbi.2012.06.018

[Nijhout et al. 2004] Nijhout HF, Reed MC, Budu P, et al. A mathematical model of the folate cycle: new insights into folate homeostasis. J Biol Chem 2004;279:5500816, doi:10.1074/jbc.M410818200

[Pfeifer et al. 1990] G. P. Pfeifer, S. D. Steigerwald, R. S. Hansen, S. M. Gartler and A. D. Riggs, Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability, Proc Natl Acad Sci U.S.A. 87 (1990) 8252–8256

[Poh et al. 2016] W. J. Poh, C. P. P. Wee, Z. Gao, DNA Methyltransferase Activity Assays: Advances and Challenges, Theranostics 6(3), (2016) 369-391, doi:10.7150/thno.13438

[Rasmussen and Helin 2016] K. D. Rasmussen and K. Helin, Role of TET enzymes in DNA methylation, development, and cancer, Genes Dev. 30 (2016) 733–750, doi:10.1101/gad.276568.115

[Razin and Riggs 1980] A. Razin and A. D. Riggs, DNA methylation and gene function, Science 210 (1980) 604–610, doi:10.1126/science.6254144

[Reed et al. 2006] Reed MC, Nijhout HF, Neuhouser ML, et al. A mathematical model gives insights into nutritional and genetic aspects of folate-mediated one-carbon metabolism. J Nutr 2006; 136 : 2653-61, doi:10.1093/jn/136.10.2653

[Riggs 1975] A. D. Riggs, X inactivation, differentiation, and DNA methylation, Cytogenet Cell Genet 14 (1975) 9–25

[Riggs and Xiong 2004] A. D. Riggs and Z. Xiong, Methylation and epigenetic fidelity, Proc Natl Acad Sci U.S.A. 101 (2004) 4–5, doi:10.1073/pnas.0307781100

[Robertson et al. 1999] K. D. Robertson, E. Uzvolgyi, G. Liang, C. Talmadge, J. Sumegi, F. A. Gonzales and P. A. Jones, The human DNA methyltransferases ($DNMTs$)$1, 3a$ and $3b$: coordinate $mRNA$ expression in normal tissues and overexpression in tumors, Nucleic Acids Res. 27 (1999) 2291–2298

[Robertson 2001] K. D. Robertson, DNA methylation, methyltransferases, and cancer, Oncogene 20 (2001) 3139–3155, doi:10.1038/sj.onc.1204341

[Sager and Kitchin 1975] R. Sager and R. Kitchin, Selective silencing of eukaryotic DNA, Science 189 (1975) 426–433, doi:10.1126/science.189.4201.426

[Scott et al. 2010] J. L. Scott, C. Gabrielides, R. K. Davidson, T. E. Swingler, I. M. Clark, G. A. Wallis, R. P. Boot-Handford, T. B. Kirkwood, R. W. Taylor and D. A. Young, Superoxide dismutase downregulation in osteoarthritis progression and end-stage disease, Ann Rheum Dis 69 (2010) 1502–1510, doi:10.1136/ard.2009.119966

[Scourzic et al. 2015] L. Scourzic, E. Mouly and O. A. Bernard, TET proteins and the control of cytosine demethylation in cancer, Genome Med 7 (2015) 9, doi:10.1186/s13073-015-0134-6

[Svedruzic 2008] Z. M. Svedruzic, Mammalian cytosine DNA methyltransferase $DNMT1$: enzymatic mechanism, novel mechanism-based inhibitors, and $RNA-$ directed DNA methylation, Curr Med Chem 15 (2008) 1, doi:10.1186/s13073-015-0134-6

[Smith and Meissner 2013] Z. D. Smith and A. Meissner, DNA methylation: roles in mammalian development, Nat Rev Genet. 14 (2013) 204–220, doi:10.1038/nrg3354

[Sora and Mc Auley 2000] Enrique Salcedo-Sora J, Mc Auley MT, A mathematical model of microbial folate biosynthesis and utilisation: implications for antifolate development. Mol Biosyst 2016; 12 : 923-33, doi:10.1039/c5mb00801h

[Waterland and Michels 2007] R. A. Waterland and K. B. Michels, Epigenetic epidemiology of the developmental origins hypothesis, Ann Review of Nutr 27 (2007) 363–388

[Zhang et al. 2009] Y. Zhang, C. Rohde, S. Tierling, T. P. Jurkowski, C. Bock, D. Santacruz, S. Ragozin, R. Reinhardt, M. Groth, J. Walter, A. Jeltsch, DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution, PLoS Genet, (2009) 5(3), doi:doi:10.1371/journal.pgen.1000438

[Zhong et al. 2016] J. Zhong, G. Agha and A. A. Baccarelli, The Role of DNA Methylation in Cardiovascular Risk and Disease: Methodological Aspects, Study Design, and Data Analysis for Epidemiological Studies, Circ Res 118 (2016) 119–131, doi:10.1161/CIRCRESAHA.115.305206