Journal of Theoretical Biology 336 (2013) 87–95

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Bayesian inference supports a location and neighbour-dependent model of DNA methylation propagation at the MGMT gene promoter in lung tumours

Nicolas Bonello [a], James Sampson [b], John Burn [b], Ian J. Wilson [b], Gail McGrown [c], Geoff P. Margison [c], Mary Thorncroft [c], Philip Crossbie [d], Andrew C. Povey [d], Mauro Santibanez-Koref [b], Kevin Walters [a,*]

[a] School of Mathematics and Statistics, University of Sheffield, Sheffield S3 7RH, UK
[b] Institute of Genetic Medicine, International Centre for Life, University of Newcastle, Newcastle NE1 3BZ, UK
[c] Paterson Institute for Cancer Research, University of Manchester, Manchester M20 4BX, UK
[d] Centre for Occupational and Environmental Health, University of Manchester, Manchester M13 9PL, UK

HIGHLIGHTS

- We model epigenetic changes in blood cells that may lead to lung tumour formation.
- We propose four models of methylation propagation at the MGMT gene promoter.
- We use approximate Bayes factors to choose between competing models.
- We use rejection sampling to obtain posterior distributions for model parameters.
- Our data suggests a location and neighbour dependent model of pattern propagation.

ARTICLE INFO

Article history:
Received 14 December 2012
Received in revised form
12 July 2013
Accepted 19 July 2013
Available online 30 July 2013

Keywords:
Bayesian analysis
Model selection
DNA methylation
MGMT
Lung cancer

ABSTRACT

We exploit model-based Bayesian inference methodologies to analyse lung tumour-derived methylation data from a CpG island in the O6-methylguanine-DNA methyltransferase (MGMT) promoter. Interest is in modelling the changes in methylation patterns in a CpG island in the first exon of the promoter during lung tumour development. We propose four competils of methylation state propagation based on two mechanisms. The first is the location-dependence mechanism in which the probability of a gain or loss of methylation at a CpG within the promoter depends upon its location in the CpG sequence. The second mechanism is that of neighbour-dependence in which gain or loss of methylation at a CpG depends upon the methylation status of the immediately preceding CpG. Our data comprises the methylation status at 12 CpGs near the 5′ end of the CpG island in two lung tumour samples for both alleles of a nearby polymorphism. We use approximate Bayesian computation, a computationally intensive rejection-sampling algorithm to infer model parameters and compare models without the need to evaluate the likelihood function. We compare the four proposed models using two criteria: the approximate Bayes factors and the distribution of the Euclidean distance between the summary statistics of the observed and simulated datasets. Our model-based analysis demonstrates compelling evidence for both location and neighbour dependence in the process of aberrant DNA methylation of this MGMT promoter CpG island in lung tumours. We find equivocal evidence to support the hypothesis that the methylation patterns of the two alleles evolve independently.

1. Introduction

One of the most common adducts in human DNA is 5 methyl cytosine (5meC). In mammals this modification preferentially affects cytosine residues immediately preceding a guanine residue, in particular in 5′ CG 3′ (CpG) sequence motifs (Bird, 1980). Mammalian genomes show a lack of 5′ CG 3′ sequence motifs when compared to the frequency expected from the base composition. Particular regions of the genome, so-called “CpG islands”, tend to show no or less of a CpG frequency depression and the cytosine residues in these islands tend to be unmethylated (Bird et al., 1985). This contrasts with CpG cytosines in the rest of the genome, which tend to be methylated (Bird et al., 1985). CpG islands tend to occur in the proximity of the transcription start

* Corresponding author. Tel.: +44 114 222 3720.
E-mail address: k.walters@sheffield.ac.uk (K. Walters).

0022-5193/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2013.07.019

sites of genes (Gardiner-Garden and Frommer, 1987). The ability of 5meC to modulate interactions between proteins, such as transcription factors, and DNA (Lee and Lee, 2012; Tate and Bird, 1993) provides a base for the involvement of such residues in the regulation of transcription. Methylation of cytosine residues in CpG islands is often associated with decreased transcription (Jaenisch and Bird, 2003). In malignant tissues, inactivation of certain genes is frequently observed (Peltomaki, 2012). Often this affects genes whose products directly or indirectly impair the ability of cells to proliferate. For example, decreased activity of O6-methylguanine-DNA-methyltransferase, (MGMT) can be found in a variety of tumours including malignancies of the skin, lung and of the brain (Egyhazi et al., 1997; Esteller and Herman, 2004; Myong, 2010; Ostrowski et al., 1991).

MGMT itself is involved in the repair of alkylation damage at the O6-position of guanine residues in DNA by damage reversal. Lack of MGMT results in increased sensitivity towards alkylating agents, a group of compounds that includes drugs used in cancer chemotherapy such as temozolomide (Agarwala and Kirkwood, 2000; Lashkari et al., 2011). In the past few years this has spurred efforts to adapt chemotherapeutical regime therapies according to tumour MGMT activity (Hau et al., 2007; Medeiros et al., 2012).

The MGMT promoter includes a CpG island that covers the first exon (Harris et al., 1991). An increased frequency of 5meC residues in the promoter of the MGMT gene can lead to reduced transcription (Harris et al., 1994), however, the correlation between methylation and lack of gene expression varies with cytosine residue in the promoter. There is low correlation between methylation and transcription at certain sites whilst recent results in glioblastoma suggest that sites closer to the first exon show the strongest correlation (Malley et al., 2011). The relatively high error rate in the transmission of methylation between mother and daughter cells allows us to use 5meC to track ancestry in somatic tissues. Over the past decade methylation patterns of sampled cells have been used to assess stem cell dynamics in colon crypts and infer the number of cell divisions (Graham et al., 2011; Walters, 2009, 2012). Malignant development is associated with changes in methylation, this is reflected, on the one hand, in decreased methylation of inter-genic regions and, on the other hand, increased methylation of certain promoters (Bergman and Cedar, 2013). In normal tissues MGMT is ubiquitously expressed. In many tumours including those of the lung, expression levels decrease (Chen et al., 1992). Repeated passaging of immortalised cell can be accompanied by transcriptional silencing. The relationship between promoter methylation was originally unclear (Margison et al., 2003 and references there in) but it is now well established that reduced activity correlates with methylation of specific sites within the promoter (Christians et al., 2012). However these sites are 5′ of the sites analysed in this study. Therefore one of the motivations for this study was to investigate the emergence of new methylation patterns in a situation where the cellular environment changes.

The boundary between a CpG island and the rest of the genomic DNA represents a particularly interesting region since on one side CpGs tend to be methylated whilst on the other they tend to be considerably less methylated. During tumour development we might expect the methylation state to change and modelling these changes will allow us to infer the nature of the factors involved. The 5′ boundary of the CpG island also contains a polymorphism with a high minor allele frequency (rs1625649, MAF 0.37 Hapmap CEU population). In heterozygous individuals this polymorphism allows us to follow each of the two alleles separately and assess the contribution of factors that affects both or each of them. Among the formers we would include trans-acting factors reflecting the cellular environment whilst the latter would include cis-acting factors, for example the influence of the methylation state of adjacent sites. Allele-specific CpG methylation was determined at 12 CpGs at the CpG island boundary of the MGMT gene in lung tumour cells and peripheral blood mononuclear cells from each of the two lung cancer patients for both alleles (A and C). The minimum and maximum number of DNA sequences in these 8 cell groups were 88 and 224 respectively and the mean was 123 sequences. The data was provided by the Patterson Institute for Cancer Research in Manchester. In this paper we propose several probabilistic models for the somatic propagation of the DNA methylation pattern of this 12-CpG sequence at the boundary of the MGMT promoter. We use approximate Bayes factors to choose between these competing

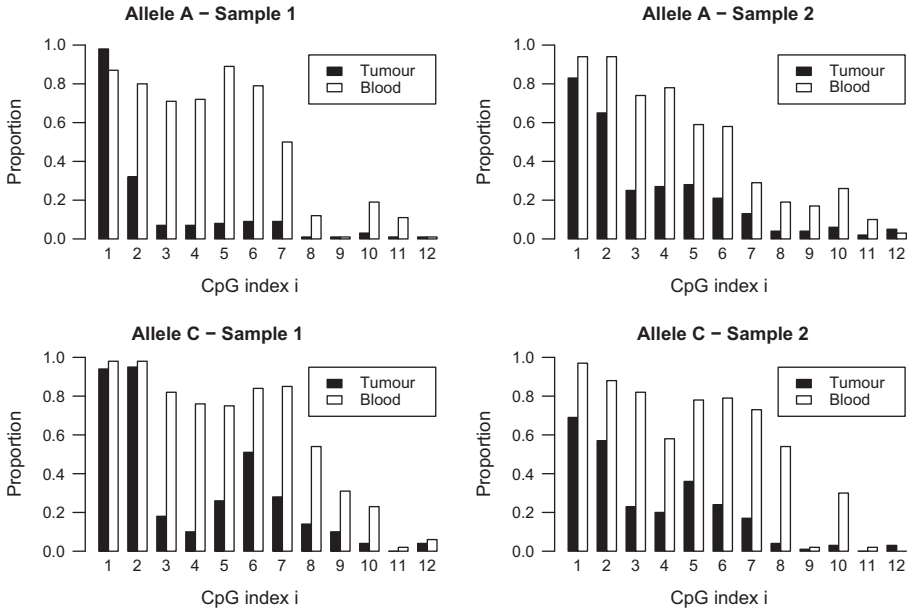

Fig. 1. The proportion of methylated cytosines on the *i*th CpG for the two methylation datasets derived from tumour tissue and blood, with results for allele A displayed on top and sample 1 on the left.