Acta Neuropathol (2011) 121:651–661
DOI 10.1007/s00401-011-0803-5

ORIGINAL PAPER

A distinct region of the *MGMT* CpG island critical for transcriptional regulation is preferentially methylated in glioblastoma cells and xenografts

Deborah S. Malley · Rifat A. Hamoudi · Sylvia Kocialkowski · Danita M. Pearson · Vincent Peter Collins · Koichi Ichimura

Received: 3 December 2010 / Revised: 19 January 2011 / Accepted: 20 January 2011 / Published online: 3 February 2011

Abstract O^6-Methylguanine-DNA methyltransferase (MGMT) is a DNA repair protein that removes alkyl DNA adducts such as those induced by alkylating agents. Loss of *MGMT* expression through transcriptional silencing by hypermethylation of its CpG island (CGI) is found in diverse human cancers including glioblastomas. Glioblastomas that have *MGMT* methylation respond to temozolomide, an alkylating agent, resulting in improved survival. Consequently, assessment of *MGMT* methylation has become a therapy response and prognostic indicator. However, it is not clear whether the region of the *MGMT* CGI commonly analysed is the critical region involved in transcriptional control. We measured methylation levels at each CpG site for the entire *MGMT* CGI using bisulfite modification and pyrosequencing, and compared them with *MGMT* mRNA expression in glioblastoma cell lines, xenografts and normal brain tissues (41 samples). Two critical regions were identified (DMR1 and DMR2). DMR2 encompasses the commonly analysed region and was always methylated when DMR1 was methylated. A luciferase reporter assay showed that substitutions of several specific CpG sites within DMR2 significantly attenuated the promoter activity of the *MGMT* CGI. Our results indicate that several CpG sites within DMR2 play a critical role in the transcriptional control of *MGMT*, making DMR2 the optimal target for methylation testing. However, given the highly variable patterns of *MGMT* methylation associated with transcriptional silencing observed in this region among the tumours in this study, methylation levels need to be measured at a number of individual CpGs within DMR2 to confidently predict transcriptional silencing and thus sensitivity to alkylating agents.

Keywords MGMT · Temozolomide · Methylation · Pyrosequencing · Luciferase assay · Nucleosome positioning

Abbreviations

CGI	CpG island
DMR	Differentially methylated region
FDR	Benjamini and Hochberg false discover rate
MGMT	O^6-Methylguanine-DNA methyltransferase
MSP	Methylation-specific PCR
PCR	Polymerase chain reaction
qPCR	Real-time quantitative PCR
TF	Transcription factor
TMZ	Temozolomide
TSS	Transcription start site

Introduction

O^6-Methylguanine-DNA methyltransferase (MGMT) is an ubiquitously expressed protein that repairs O^6-alkylguanine (e.g., O^6-methylguanine), which can be induced by various environmental carcinogens and alkylating agents (reviewed in [11, 19]). O^6-alkylguanine is a highly mutagenic DNA adduct as its persistence leads to mis-pairing with thymine

Electronic supplementary material The online version of this article (doi:10.1007/s00401-011-0803-5) contains supplementary material, which is available to authorized users.

D. S. Malley · R. A. Hamoudi · S. Kocialkowski · D. M. Pearson · V. P. Collins · K. Ichimura (✉)
Division of Molecular Histopathology, Department of Pathology, Level 3, Lab Block, Addenbrooke's Hospital, University of Cambridge, Box 231, Cambridge CB2 0QQ, UK
e-mail: ki212@cam.ac.uk

by DNA polymerase during DNA replication. One MGMT molecule can repair only one alkyl adduct as MGMT removes the alkyl group from the DNA lesion to its cysteine residue, thereby permanently inactivating itself while restoring the DNA. Thus, the capacity to repair O^6-alkylguanine depends on the amount and production rate of MGMT. Loss of *MGMT* expression is a common finding in diverse human cancers (reviewed in [8]), generally as a result of transcriptional silencing induced by DNA hypermethylation of the CpG island (CGI) of *MGMT* [1, 15, 24, 26]. Silencing of *MGMT* is expected to lead to increased sensitivity to alkylating agents.

Glioblastomas are one of the commonest and most malignant primary brain tumours in adults. Current standard treatment with surgery followed by combined radiotherapy and temozolomide (TMZ), an oral alkylating agent, result in a median survival of 14.6 months [30]. Several large-scale studies have found that patients with glioblastomas with *MGMT* methylation responded better to TMZ and had significantly longer progression-free survival and overall survival than those without *MGMT* methylation [16, 30, 34]. Thus, *MGMT* methylation is considered an important therapy response and prognostic indicator for glioblastomas. Consequently, the methylation status of *MGMT* is often routinely examined using various methods to stratify patients in clinical studies of glioblastomas. Most frequently methylation-specific PCR (MSP) of bisulfite-modified genomic DNA is used [17].

Currently, it is not known which or how many CpGs in the *MGMT* CGI have a major impact on expression and should be analysed in a clinical setting [27]. The majority of methods examine a very limited number of CpG sites within *MGMT* CGI. It is unknown whether the commonly used MSP primers for *MGMT* [7] are targeting a crucial region [9]. MSP primers are designed to specifically amplify only fully methylated sequences, and the effect of heterogeneous methylation on the results is unpredictable [6, 23]. Direct sequencing of bisulfite-modified PCR-amplified sequences is also used and can semi-quantitatively measure methylation levels at individual CpG sites [23]. However, this is a labour-intensive method, which makes it less suitable for clinical testing. Pyrosequencing is a sequence-by-synthesis method which can quantitatively assess methylation levels of individual CpGs in bisulfite-modified, PCR-amplified sequences with a high efficiency. As such, it overcomes many of the limitations in other methods and is emerging as a robust and accurate technique for methylation detection [32].

To investigate the impact of individual CpGs in the *MGMT* CGI on the transcriptional control, with the aim to define the region suitable for clinical *MGMT* methylation testing, we studied the methylation status of the entire CpG island of *MGMT* using pyrosequencing and compared it with *MGMT* mRNA expression in a series of glioblastoma cell lines and xenografts. We identified two separate regions in which the methylation status was significantly correlated with expression. Further investigation using a luciferase reporter assay suggested that several CpG sites within one of the two regions, the differentially methylated region 2 (DMR2), is critical for transcriptional control.

Materials and methods

Cell lines and tissue materials, DNA/RNA extraction and cDNA synthesis

Glioblastomas from 22 patients were subcutaneously xenografted and serially transplanted in nude mice (nu/nu, Balb/C; Harlan Sprague Dawley Inc., Indianapolis, USA) as has been described [12]. A xenografted tumour from the second to fourth serial transplantation was chosen in each case and used in the study. In addition, the following 13 established glioma cell lines were included: A172, H4, T98G, TP267, TP336, TP365, TP483, U118MG, U138MG, U178MG, U251MG, U373MG and U87MG. The U251MG and U373MG glioma cell lines used in this study were obtained directly from the original depositor, and their distinct genotypes have been verified [10]. Six normal brain tissue samples from cerebral subcortex were also included. One of them was purchased from Cambridge BioSciences (Cambridge, UK). DNA/RNA extractions from the xenografted tumour tissues and cell lines were performed as described previously [18]. DNA and RNA from normal brain tissues were extracted using a Qiagen AllPrep^{TM} DNA/RNA/Protein Mini kit (Qiagen, Crawley, UK). cDNA was generated as described [22]. The study was approved by the Ethical Committee of the Sahlgrenska University Hospital (S339:01), the Karolinska Hospital (no. 91:16), and the Cambridge Research Ethics Committee (Cambridge, UK; NRES Cambridgeshire 2 REC reference 03/115).

Pyrosequencing

Bisulfite modification of genomic DNA (500 ng) was performed using an EZ DNA methylation kit (Zymo Research, Orange, CA, USA) as per manufacturer's recommendation. As a control for methylated or unmethylated DNA, chemically methylated or de-methylated human genomic DNAs (EpiTect PCR Control DNA Set, Qiagen, Crawley, UK) were used. Templates for pyrosequencing were prepared by amplifying bisulfite-modified DNA with primers that were biotinylated for template strands. Preparation of single-stranded DNA template, annealing to the pyrosequencing primer and pyrosequencing run were

performed using PyroGold Q96 SQA reagents on the PyroMark ID pyrosequencer (Qiagen, Crawley, UK) as per manufacturer's recommendation. The pyrosequencing data were analysed using a Pyro Q-CpG software (Qiagen, Crawley, UK). All primer sequences for PCR and pyrosequencing, the PCR conditions and pyrosequencing assays are listed in Supplementary Table 1.

Real-time quantitative PCR

mRNA expression levels of *MGMT* were determined with a forward primer located in exon 3 and a reverse primer in exon 5 (Supplementary Table 1) by real-time quantitative PCR (qPCR) using LightCycler 480 SYBR Green I Master (Roche Diagnostics, Burgess Hill, UK) and the SYBR Green I (483–533 nm) detection format on a LightCycler 480 (Roche Diagnostics, Burgess Hill, UK) as described [35]. An additional step of 82°C incubation for 5 s was included after the elongation step at 72°C to denature primer dimers before measuring the signal intensity. The expression level of 18S rRNA was used as internal references for normalisation [35]. Relative quantification analyses were performed using a LightCycler 480 software version 1.2 (Roche Diagnostics, Burgess Hill, UK).

Luciferase assay

Luciferase reporter constructs were generated based on a pGL3-Control Vector (pGL3C, Promega, Madison, WI, USA) with a minor modification in which the *Xba*I site of the original pGL3C vector was removed (pGL3C-Xb(−), see below) to allow generation of deletion mutants by restriction enzyme digestion at the *Xba*I site within the insert. The CpG island (CGI) of *MGMT* spans 762 bp (The UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly, chr10:131,264,949-131,265,710, http://genome.ucsc.edu/cgi-bin/hgTracks?org=human) encompassing *MGMT* exon 1. The entire *MGMT* CGI was PCR amplified using primers that contain *Bgl*II or *Nco*I sites (PC6108/09; Supplementary Table 1) and subcloned into pGL3C at the corresponding sites, replacing the SV40 promoter of the original pGL3C. A primer pair that amplifies the same region with reciprocally tagged restriction sites was used to clone the *MGMT* CGI in the reverse orientation (PC6132/33; Supplementary Table 1).

Two sets of mutant *MGMT* CGI luciferase reporter constructs were generated from the wild-type construct (Supplementary Table 3). In the first set (C>T mutation constructs), selected CpG site(s) were replaced with TpG in the constructs using a complementary pair of oligonucleotides that contain the specific C>T mutation(s) and a QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies, La Jolla, CA, USA) as per manufacturer's recommendation. In the second set (deletion constructs), various lengths of *MGMT* CGI were removed by restriction enzyme digestion combining either *Sac*II/*Xba*I, *Xba*I/*Pst*I or *Sac*II/*Pst*I, blunting the overhangs with T4 DNA polymerase (New England BioLabs, Hitchin, UK) and re-ligating the final constructs with Quick-Stick Ligase (Bioline, London, UK).

For the luciferase assay, the U373MG cells were seeded at a density of 10^4 cells/well in triplicate or quadruplicate in a 96-well plate with F-12K medium (Invitrogen, Paisley, UK) supplemented with 10% foetal bovine serum, 24 h prior to transfection. Fifty nanograms of the pGL3C constructs were mixed with 0.35 µl of Lipofectamine LTX and 0.1 µl of PLUS reagent (Invitrogen, Paisley, UK) in 20 µl of F-12K without serum and transfected to the U373MG cells. After 24 h, a luciferase assay was performed using a Dual-Luciferase® Reporter Assay System (Promega, Madison, WI, USA) as per manufacturer's recommendation. Briefly, the medium was removed, cells washed with 1× PBS and lysed with Passive Lysis Buffer. Cell lysates were transferred to a new white 96-well plate, the Luciferase Assay Buffer II added, and luminometry was performed immediately using a fluorescence plate reader (Wallac 1420 Victor3 Plate Reader, Perkin Elmer, Cambridge, UK) to quantify the expression of the firefly luciferase.

The cell lysates were diluted 20 times with water and used as templates for qPCR to quantify the amount of transfected constructs in each well. qPCR was performed with a primer pair that specifically amplified a small region of the firefly luciferase gene so that only the firefly luciferase constructs were detected (Supplementary Table 1) using the SYBR Green I detection format on LightCycler 480 (Roche Diagnostics, Burgess Hill, UK) as above. The relative copy numbers of the constructs were calculated using the LightCycler 480 software version 1.2. The firefly luciferase luminescence was then normalised by the relative copy numbers obtained above, replicas averaged, and the expression of the firefly luciferase relative to that of the wild-type *MGMT* CGI construct was calculated.

Bioinformatic and statistical analyses

Unsupervised hierarchical clustering using Euclidean distance measure and average linkage method was carried out on raw methylation percentage data for 96 CpG sites within the *MGMT* CGI in the R statistical analysis environment version 2.8.0 (http://www.r-project.org).

Nucleosome positioning was predicted for all the sequences including 5 kb upstream and downstream of *MGMT* CGI using the online software available at Segal lab of Computational Biology, Department of Computer Science and Applied Mathematics, Weizmann Institute of

Science, Israel (http://genie.weizmann.ac.il/software/nucleo_prediction.html) [20]. Transcription factor binding sites were predicted using TFSEARCH (http://molsun1.cbrc.aist.go.jp/research/db/TFSEARCH.html) and PATCH public 1.0 (http://www.gene-regulation.com/cgi-bin/pub/programs/patch/bin/patch.cgi), which utilise the TRANSFAC database (http://www.gene-regulation.com/pub/databases.html).

Statistical analyses were performed using IBM SPSS Statistics 18 (SPSS UK Ltd., Surrey, UK). Mann–Whitney *U* test was used to compare levels of methylation at each site between two groups of samples divided according to their *MGMT* expression levels. Multiple statistical tests were corrected using the Benjamini and Hochberg false discover rate in the R statistical analysis environment version 2.8.0 (http://www.r-project.org).

Results

Pyrosequencing of the *MGMT* CpG island

The first goal of this study was to comprehensively profile the methylation status of the entire *MGMT* CpG island (CGI) in glioblastoma cells by pyrosequencing on bisulfite-modified DNA and correlate the findings to expression levels of the transcripts. As glioblastoma tissue may contain considerable numbers of non-neoplastic cells such as vascular endothelial cells and microglia, known to express *MGMT* that could mask diminished expression in tumour cells [27, 28], we used 13 established glioblastoma cell lines and 22 human glioblastoma xenografts grown subcutaneously in nude mice as neither of them contains human non-neoplastic cells. Six normal brain tissues were also included in the study.

The *MGMT* CGI, which spans 762 bp (chr10: 131264949–131265710) and contains 98 CpG dinucleotides, was defined according to the UCSC Human Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly (http://genome.ucsc.edu). Consecutive numbers were assigned to the CpGs from 5′ to 3′ on the coding strand of *MGMT*, starting from CpG1 (−452 bp from the transcription start site (TSS) [13]) through CpG98 (+308 bp). The relative positions of all CpGs studied are listed in Supplementary Table 2. Five pairs of PCR primers were designed to collectively amplify the entire CpG Island to generate templates for pyrosequencing (Fig. 1a; Supplementary Table 1). A total of 11 pyrosequencing primers and the corresponding pyrosequencing assays were designed to cover all CpG sites in the *MGMT* CGI except CpG21 and CpG49 (see below), for which no dependable pyrosequencing assay could be designed. One SNP, rs2782888 (T/G), is located 73 bp upstream of the TSS, where the reference genomic sequence was T. However, all samples used in this study had only the G allele (data not shown). As the nucleotide preceding this position is a cytosine, the G allele creates an additional CpG site. This site was not included in the analysis.

As a result, the methylation status of 96 CpG sites in the *MGMT* CGI were assessed in all 41 samples by pyrosequencing (Fig. 2; Supplementary Table 2). There was no statistically significant difference in the mean levels of methylation at either all 96 CpG sites or at any regions between glioma cell lines and xenografts ($p = 0.101$, Mann–Whitney *U* test). The mean methylation levels for all CpGs in each sample varied from 5.8 to 72.9%. Although the methylation levels across the 96 CpGs were heterogeneous, several distinct regions could be identified where a cluster of CpGs was often concurrently methylated (Fig. 2). The region containing CpG1–24 was most frequently methylated (Figs. 1b, 2). This region was always methylated when any other regions were methylated and could also be the only methylated region as exemplified by GB179X3 or TP336. Two separate regions, containing CpG25–50 and CpG73–90 respectively, were highly methylated in a subset of samples. When CpG25–50 was methylated, CpG73–90 was always methylated as well, while several samples showed high levels of methylation of CpG73–90, but not of CpG25–50. The intervening region (CpG51–72) was generally unmethylated, with only a few exceptions (e.g., GB185X3). The most distal region (CpG91–98) was rarely methylated. Interestingly, some of these distinct regions of preferential methylation were not clearly depicted in the profile of the mean methylation levels at each site across all the samples (Fig. 1b).

One-way clustering analysis of methylation levels at the 96 CpG sites divided all 41 samples into six groups (Fig. 2). One group of samples (group 1; Fig. 2) was distinct from all the others showing high levels of methylation at CpG1–24, 25–50 and 73–90. The other tumours were further subdivided into five groups, with group 2 showing high levels of methylation within CpG1–24 and 73–90, but not in 25–50, group 3 moderate levels of methylation in all three regions, group 4 and 6 high level of methylation only in CpG1–24, and group 5 no or low methylation in any region. Five of the six normal brain samples were clustered in group 5, reflecting their general lack of methylation. These results show that (1) the patterns of methylation across the *MGMT* CGI are non-random and some CpGs are preferentially methylated, (2) glioblastoma cells may be sub-grouped according to the methylation patterns in these regions and (3) the two regions, CpG25–50 and CpG73–90, are often concurrently methylated although some samples showed high-level methylation only in CpG73–90, but not in CpG25–50. Two-way clustering grouped most CpGs

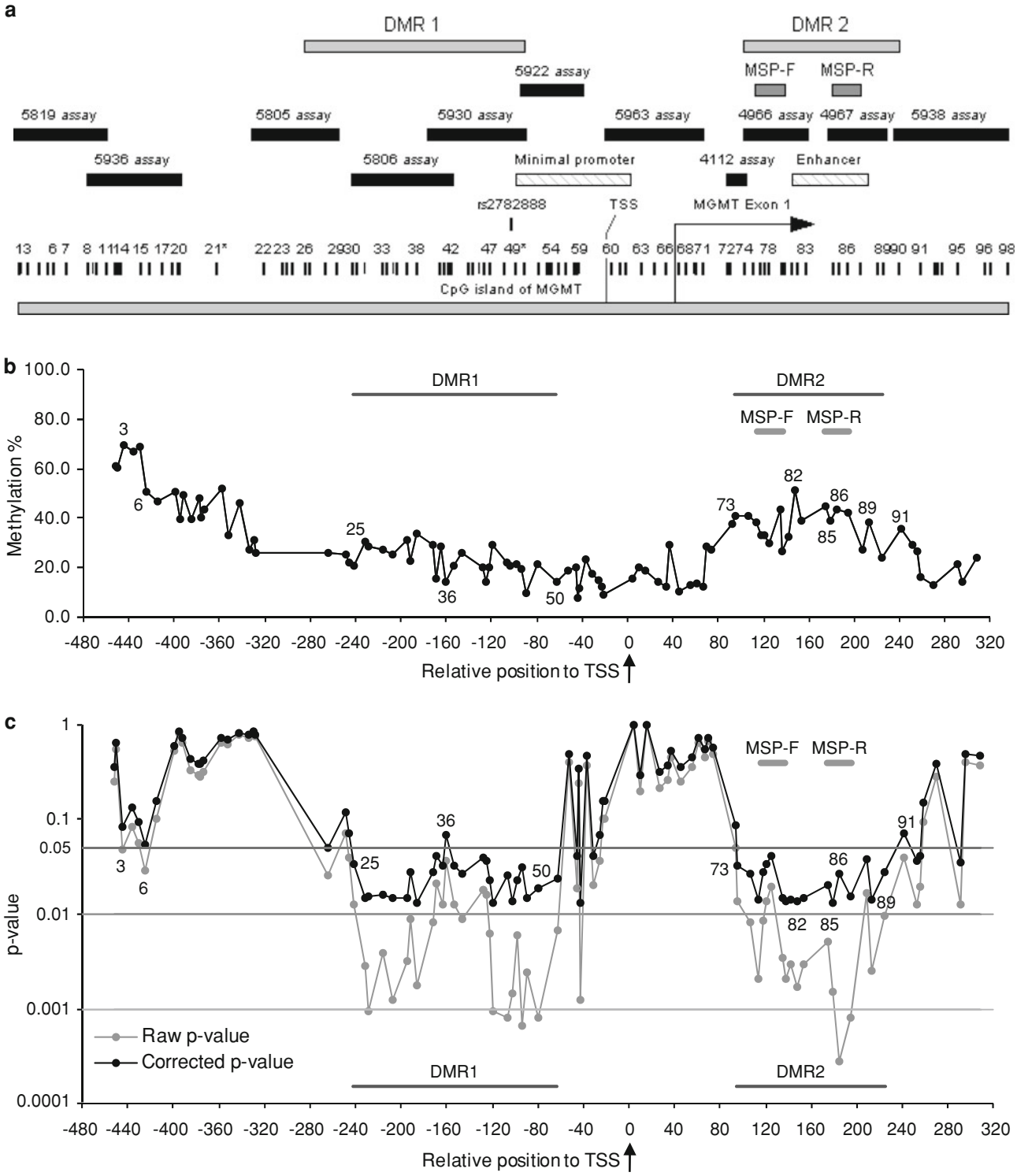

Fig. 1 **a** A diagram of *MGMT* CGI (−452 to +309 relative to the transcription start site (*TSS*) defined by Harris et al. [13]). The relative positions of each CpG site, the SNP site (rs2782888), the minimal promoter and enhancer region [14], pyrosequencing assays, the MSP primers (MSP-F and MSP-R) and the differentially methylated regions (*DMR*) defined in this study are indicated. **b** Mean methylation levels at each CpG site plotted against their distance from TSS (0, *arrow*). The CpG numbers (see "Results") are indicated at selected sites. The positions of DMRs and the MSP primers are indicated. **c** Superimposed plots of the raw *p* values obtained from Mann–Whitney *U* test (*grey*) and the multiple test-corrected *p* values using the Benjamini and Hochberg false discover rate (*black*) at each CpG site against their distance from TSS (0, *arrow*) to indicate statistically significant correlation between the methylation and expression levels. *Grid lines* represent $p = 0.05$, 0.01 and 0.001. The CpG numbers are indicated at selected sites. The positions of DMRs and the MSP primers are shown by *grey bars*

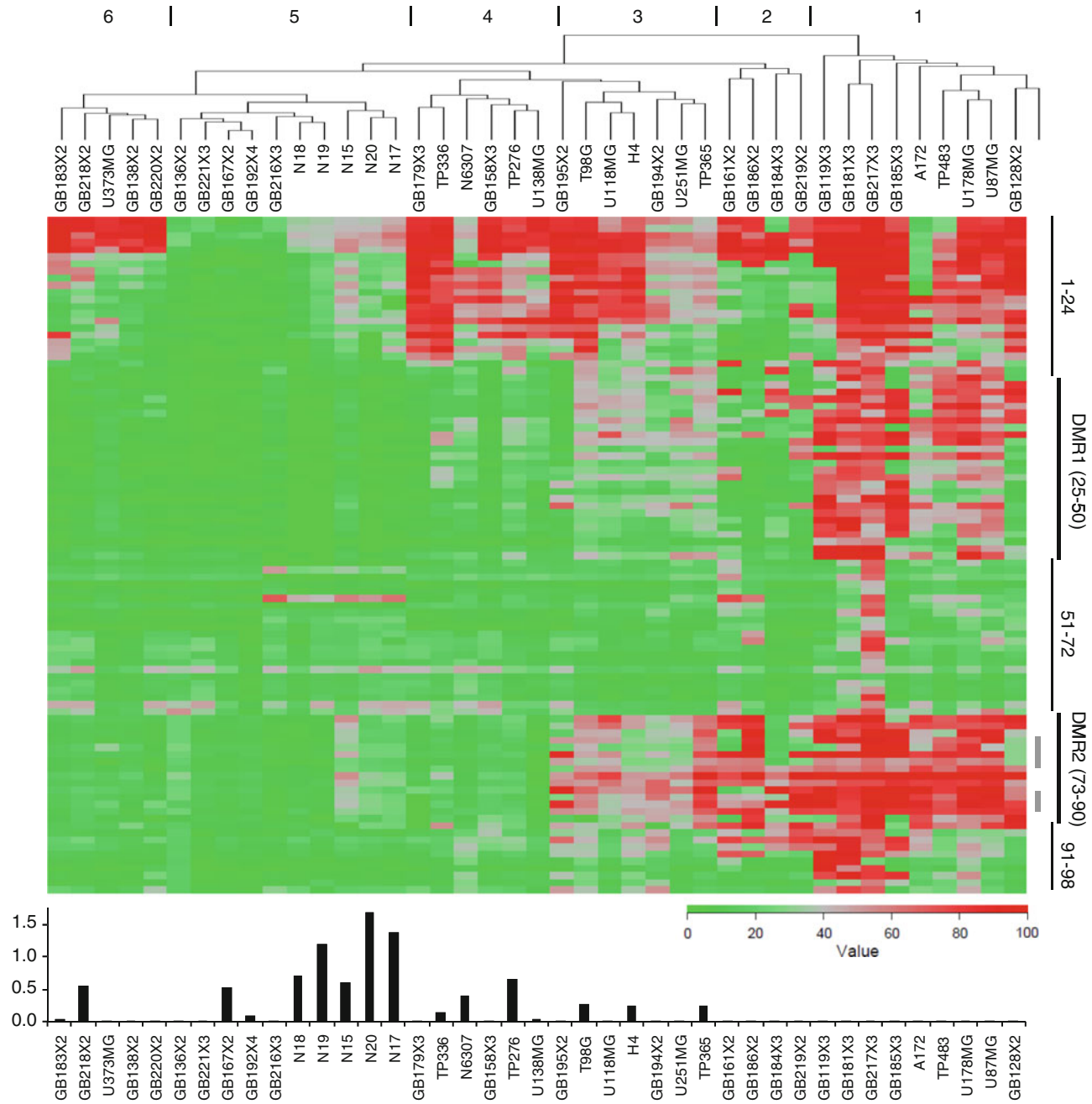

Fig. 2 One-way clustering and heat map of *MGMT* CGI methylation as determined by pyrosequencing (*red* high methylation, *green* low methylation). Each *row* represents one CpG ordered from CpG1 (*top*) to CpG98 (*bottom*), and each *column* represents individual samples. Five regions within CGI including DMR1 and DMR2 are indicated as *vertical lines* on the right of the heat map (see "Results"). The *grey vertical bars* indicate the position of the MSP primers. The mRNA expression levels of *MGMT* measured by real-time quantitative PCR are shown as relative to the mean of six normal brain expressions in linear scale below the corresponding columns of the heat map

from CpG73–90 within the same group, as well as CpG25–50, while these two groups of CpGs were clustered next to each other (Supplementary Fig. 1). CpG1–20 were grouped separately from these two groups of CpGs. The results support the notion that the *MGMT* CGI may be divided into several blocks.

Correlation of methylation to mRNA expression

To determine whether mRNA expression of *MGMT* correlated to the methylation status of each CpG, *MGMT* mRNA expression was measured by real-time quantitative reverse-transcriptase PCR (qPCR) in all 41 samples. The

expression levels of *MGMT* varied considerably among the samples (Fig. 2, bottom panel). All six normal brain samples showed strong expression of *MGMT*, while the majority of glioma cell lines/xenografts had very low levels or absence of expression. All 41 samples were then divided into two groups on the basis of their expression levels using an arbitrary threshold, i.e., *MGMT* expression higher or lower than 1% of the mean expression levels of the six normal brains. These two groups were then compared using the Mann–Whitney *U* test at each CpG site. Multiple testing was corrected using the Benjamini and Hochberg FDR. This analysis identified two separate regions of CpG sites where methylation was significantly correlated to *MGMT* mRNA expression ($p < 0.05$ in FDR; Fig. 1c). The first region (differentially methylated region 1, DMR1) spans CpG25–50 and the second region (differentially methylated region 2, DMR2) CpG73–90. The profile was similar to this when either 30 or 0.1% of the mean normal brain expression was used as an alternative threshold, *p* values being even smaller with 0.1% threshold, indicating that the correlation between methylation and expression is not dependent on a particular threshold (Supplementary Fig. 2). These two regions coincide with two of the frequently methylated regions described above. The methylation status of CpG1–24, the most frequently methylated region (Fig. 1b), was not significantly correlated to mRNA expression. These results suggested that the status of the CpGs within DMR1 and DMR2 may be critical for the promoter activity of *MGMT*.

Luciferase assay

To further investigate whether alterations at any individual CpGs affect the promoter activity, a luciferase reporter assay was performed. The entire *MGMT* CGI was amplified and subcloned into a firefly luciferase reporter vector. Two sets of mutant *MGMT* CGI reporter constructs were generated (Fig. 3). In the first set, selected CpG(s) were replaced by TpG(s) using site-directed mutagenesis. We investigated a part of DMR2 in detail, i.e., CpG82–89 because (1) DMR2 was always methylated when DMR1 was methylated (see above), suggesting that DMR2 may be playing the main role in transcriptional regulation, (2) CpG82–89 were generally methylated at higher levels than CpG73–81 (Fig. 1b; Supplementary Table 2), (3) CpG82–89 include the CpGs covered by the MSP primers. In total, 14 C>T mutant constructs were generated (Fig. 3; Supplementary Table 3). These include eight constructs with each one of CpG82–89 replaced by TpG; four constructs in which consecutive CpGs that are interrogated by the MSP primers (MSP-F, CpG 76–80 and MSP-R, CpG 84–87) [7] were mutated in various combinations (CpG76–80C>T, CpG84–87C>T, CpG76–80:84–87C>T or CpG76–87C>T) and two constructs with C>T mutations either at CpG3–6 or at CpG43–48 (DMR1). The second set of mutant constructs harboured deletions within *MGMT* CGI involving either DMR1 or DMR2, or both. Constructs with either the SV40 promoter, *MGMT* CGI in reverse orientation or no promoter were used as positive and negative controls. All constructs were designed to carry a G allele at the position corresponding to SNP sr2782888 because all cell lines and xenografts used in this study had only the G allele (see above). To investigate whether this polymorphism has an effect on promoter activity, a construct with a T allele at sr2782888 was also generated.

The wild-type *MGMT* CGI reporter construct (MGMT wt; Fig. 3) showed high levels of firefly luciferase activity comparable to that of the SV40 promoter construct (see Supplementary Table 3 for the normalised raw data). The constructs with the reversed *MGMT* CGI or without a promoter showed no luciferase activity, confirming the previously reported unidirectional promoter activity of *MGMT* CGI [13, 24]. Among the eight single CpG mutation constructs, the CpG83C>T and CpG86C>T ($p < 0.001$, *t* test) as well as CpG87C>T ($p = 0.029$) mutants showed moderate decrease of luciferase activity, while the CpG89C>T mutant showed dramatically decreased luciferase activity as compared to MGMT wt ($p < 0.0001$; Fig. 3). Among the six multiple CpG mutation constructs, luciferase activities of CpG84–87C>T, CpG76–80:84–87C>T (C>T at 76–80 and 84–87) and CpG76–87C>T were significantly lowered ($p < 0.0001$). Luciferase activity of the CpG76–80C>T mutant also showed moderate decrease ($p = 0.004$); however, the CpG3–6C>T and CpG43–48C>T constructs did not show any significant changes. Deletions including either CpG21–62 (encompassing DMR1) or CpG63–90 (encompassing DMR2) significantly decreased luciferase activity of the constructs. Deletion of CpG21–90 (DMR1 and 2) completely abolished any activity (Fig. 3). The T-allele construct at rs2782888 showed comparable activity to the G-allele wild-type counterpart.

Nucleosome positioning and TF-binding prediction

Finally, to see if any specific biological features relevant to expression control are present in the DNA sequence of the differentially methylated regions, nucleosome positioning and transcription factor (TF)-binding sites were predicted and correlated to the methylation pattern in the *MGMT* CGI (Supplementary Fig. 3). Two of the frequently methylated regions, CpG1–20 and DMR1, coincided with high nucleosome occupancies, while the intervening CpG-depleted region between these two regions as well as the unmethylated region around TSS showed low occupancy (Supplementary Fig. 3). The latter region contains the

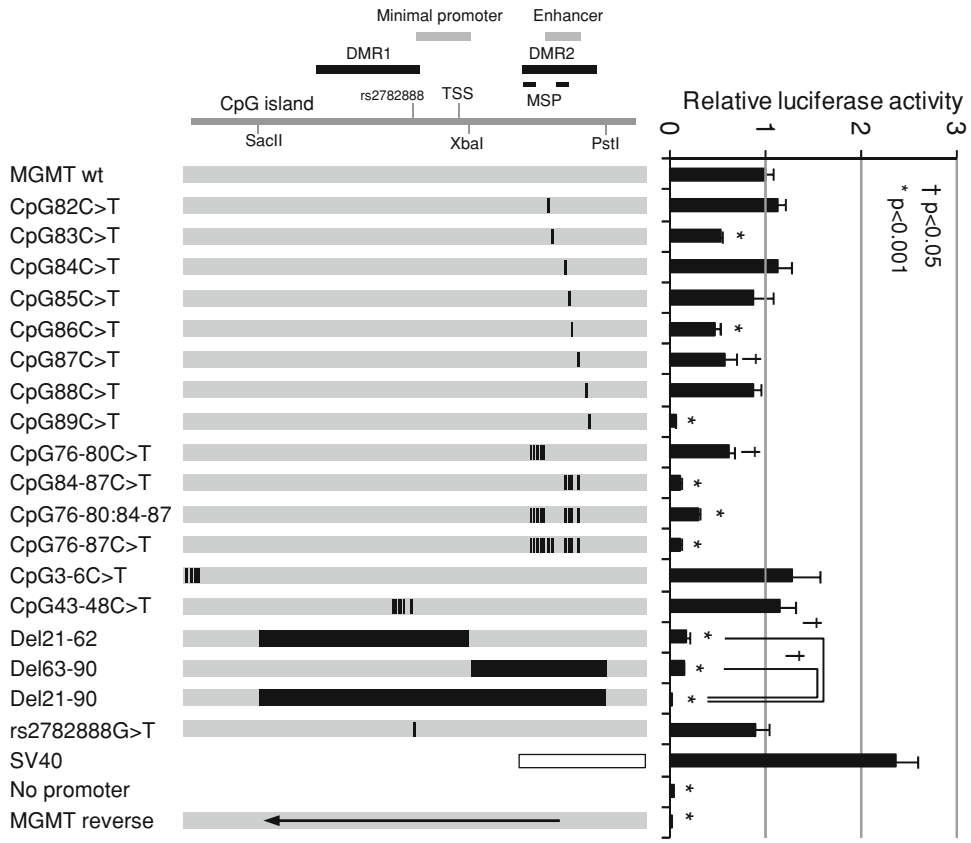

Fig. 3 A schematic diagram of the firefly luciferase reporter constructs and the luciferase assay results. The *left panel* shows schematic diagram of the firefly luciferase reporter constructs. The wild-type *MGMT* CGI is indicated by a *grey horizontal bar*. The C>T mutations introduced by site-directed mutagenesis are indicated by *black vertical lines*. The regions removed by a combination of restriction enzyme digestions (see Supplementary Table 3) are indicated by *black horizontal bars*. The SV40 promoter is represented by an *open box*. The *MGMT* CGI in reverse orientation is indicated by a *grey box* containing an *arrow*. The relative positions of DMR1, DMR2, MSP primers (MSP), minimum promoter, enhancer, TSS [14], rs2782888 SNP and three restriction sites used to generate deletion constructs are shown in a schematic drawing (*top panel*). The *right panel* shows luciferase assay results. The expression levels of the firefly luciferase for each construct relative to that of the wild-type *MGMT* construct (MGMT wt) are shown. The means of three independent experiments, each containing triplicate to quadruplicate data point for each construct, are shown. The *error bars* indicate standard error. $^{\dagger}p < 0.05$; $^{*}p < 0.001$ (*t* test). The numbers followed by C>T indicate the constructs where CpGs were replaced with TpG at that precise position(s) (see also Supplementary Table 3). "Del" followed by numerical values indicates the constructs where the region including the corresponding CpG sites was deleted. SV40 is a positive control that contains the SV40 promoter instead of the *MGMT* CGI. No promoter is the construct without any promoter. *MGMT reverse* the construct that contains *MGMT* CGI in reverse orientation

minimal promoter (−69 to +18) defined by Harris et al. [13] and a number of SP1-binding sites previously predicted by Costello et al. [4]. The DMR2 largely coincided with a gap between two predicted nucleosomes. This gap region encompasses CpG83–89 and contains the minimal enhancer (+143 to +201) [14] and predicted binding sites for several TFs including STAT3 (see "Materials and methods").

Discussion

In this study, we investigated the *MGMT* CGI in a series of glioma cell lines and xenografts to identify a set of CpGs critical for transcriptional control of *MGMT* with the aim of defining a biologically relevant and clinically practicable region within the CGI for the *MGMT* methylation assay for human gliomas. Through comprehensive pyrosequencing and mRNA expression analysis, we identified two distinct regions within *MGMT* CGI, DMR1 and DMR2, where the methylation status is strongly correlated with *MGMT* expression. These two regions did not show the highest mean methylation levels within the *MGMT* CGI (Fig. 1b), suggesting that not all methylated CpGs are directly involved in transcriptional regulation. Deletions of either region in the luciferase reporter constructs resulted in significant reduction of the promoter activity, and the deletion of both regions resulted in further attenuation, indicating that both play a role in transcriptional control. These results

are in agreement with the data presented by Nakagawachi et al. [24]. On the other hand, we found that DMR2 is always methylated when DMR1 is methylated, while in some samples, only DMR2 was highly methylated (e.g., GB168X2; Fig. 2; Supplementary Table 2). This indicates that, while both DMR1 and DMR2 may be important for the activity of the promoter, DMR2 alone is practically sufficient for the assessment of *MGMT* CGI methylation in clinical testing.

We identified several specific CpG sites (CpG 83, 86, 87 and 89) within DMR2 which seem to play a critical role in transcriptional control of *MGMT*. The reporter assay showed that mutations of either the individual CpGs or multiple consecutive CpGs significantly attenuated the promoter activity of the *MGMT* CGI. In particular, mutation of CpG89 alone almost completely abolished the promoter activity (Fig. 3). Mutations of CpGs within DMR1 (CpG43–48) or outside of the differentially methylated region (CpG3–6) did not adversely affect the transcriptional activity.

A few studies have investigated the methylation status at individual CpGs within *MGMT* CGI and compared them with gene expression. The regions defined in these studies in comparison to ours are listed in Supplementary Table 4. Watts et al. [33] studied the methylation status of 108 CpGs across *MGMT* CGI using bisulfite sequencing in a pair of isogenic multiple myeloma cell lines with or without *MGMT* expression. They identified three distinct regions within CGI where high levels of methylation were found only in the *MGMT* non-expressing cell line. These regions roughly overlap with CpG1–25, DMR1 and DMR2 in our study. Nakagawachi et al. [24] studied 97 CpGs in 19 cell lines of various origins that do not express *MGMT* for methylation of *MGMT* CGI using bisulfite sequencing and defined two distinct highly methylated regions, an upstream highly methylated region (UHMR) and downstream highly methylated region (DHMR). They showed by in vitro methylation reporter assay that complete methylation of either region leads to transcriptional silencing. Although their UHMR and DHMR, respectively, include the DMR1 and DMR2 as defined in this study, the regions are large, and individual CpGs were not examined. Everhard et al. [9] studied methylation at 68 CpG sites within *MGMT* CGI by pyrosequencing of 54 glioblastomas and 24 normal brains. For the 52 CpGs where methylation was observed only in tumours, methylation levels were compared with expression. They found that methylation of CpGs at −228, −186, +95, +113, +135, +137 (CpG27, 32, 73, 75, 79 and 80 in our study) were most significantly correlated with expression, and that methylation status at CpGs −186 to −172 (CpG32–33) and +93 to +153 (CpG72–83) are most concordant with expression. All of them fall into either DMR1 or DMR2 in this study, showing concordance between the two studies. However, some other CpGs where methylation was found to be correlated with expression found in this study were not identified. Several potential reasons for the discordance could be considered.

The study by Everhard et al. utilised the concordance between methylation and expression to define the critical CpGs. An assessment of mRNA expression in primary tumour tissues as has been used in their study may be hampered by the presence of non-neoplastic cells in the tissue which may express *MGMT*. We circumvented this problem by using glioma cell lines and xenografts which do not contain any human normal cells. We have generally observed higher *MGMT* expression in patient glioblastoma tissue with varying methylation levels than in glioma cell lines or xenografts with equivalent levels (the primers used for the expression analysis only amplify human sequences, data not shown). We consider this to be attributable to *MGMT* expression in the non-neoplastic cells present in patient tumour tissue. In addition, transcriptional silencing may be caused by mechanisms other than DNA methylation such as histone H3K9 dimethylation [36], which could also lead to discordance between findings.

In this study, the critical CpGs were defined using a reporter assay, an approach that may provide more direct functional evidence as to the contribution of each CpG or CpG region in transcriptional control. It must be noted, however, that the effect of a C>T mutation in the promoter activity of the *MGMT* CGI may not directly imitate that of cytosine methylation. Reporter assays directly using in vitro-methylated DNA fragments re-ligated into the construct have successfully been used in some other studies [1, 15, 24]. This method, however, relies on the presence of suitable restriction sites within the target region and is unlikely to decipher the effect of methylation for each individual CpG site. It still remains a technical challenge to construct an expression vector that retains cytosine methylation at one or more specific positions. Whether the results generated from C>T mutations as a surrogate to emulate cytosine methylation can be extrapolated to those produced by cytosine methylation remains to be shown.

Nevertheless, our findings provide evidence that at least several specific CpGs within DMR2 may play an important role in the promoter activity of the *MGMT* CGI, either individually or in cooperation with other sites. Therefore, DMR2 appears to be the optimal target for clinical *MGMT* methylation testing. Some of the commonly used methods including the MSP examine several CpGs in this region. However, it is possible that any of these critical CpGs within DMR2 can be specifically methylated in all possible configurations (e.g., GB138X2 shows a moderate level of methylation at CpG89 alone; Supplementary Table 2). Given the fact that the methylation levels are heterogeneous within DMR2 and that methylation at even a single CpG

may have an impact on transcription, it is important to thoroughly examine methylation levels at all individual CpGs within DMR2.

The most widely used methylation detection method is MSP [27, 31]. The commonly used MSP primers for *MGMT* bind to +115 to +137 (forward) and +174 to +195 (reverse). These two regions fall into DMR2. Our luciferase assay showed that replacement of four CpGs covered by MSP-R primers (CpG84–87) resulted in significantly reduced transcriptional activity (Fig. 3). Thus, the CpGs examined by the commonly used MSP primers are among the CpGs identified as critical for transcriptional regulation. However, these MSP primers are designed to amplify only the sequence methylated at all CpGs covered by the primers, and therefore, it would theoretically detect only sequences fully methylated at all 9 CpGs interrogated by the primer pair. The effect of partial methylation among those CpGs on amplification efficiency of MSP is unpredictable, regardless of the PCR strategy, e.g., quantitative or qualitative, direct or nested, and it is likely that methylation of certain patterns will go undetected. We have compared pyrosequencing and MSP in a large series of patients' glioma samples and indeed found a number of tumours where heterogeneous methylation observed by pyrosequencing was missed by MSP (Ichimura et al., in preparation).

Pyrosequencing overcomes these problems [6, 21, 23, 32]. It provides quantitative information at each individual CpG and may produce more accurate data than bisulfite sequencing, for which the accuracy depends on the labour-intensive sequencing of many individual clones. The current limitation of pyrosequencing is the lack of consensus how the data should be interpreted. Some studies use the mean methylation levels to represent the overall methylation status; however, there is a risk that the impact of methylation at specific CpGs may be masked by averaging. A clinically practicable and biologically relevant threshold to convert quantitative methylation value into a qualitative score also has to be established as a binary classification (i.e., methylated or unmethylated) is ultimately required for clinical testing. We have addressed this issue in a separate *MGMT* methylation study using a large number of astrocytic and oligodendroglial tumours, which will be published elsewhere (Ichimura et al., in preparation).

The association between the patterns of methylation and nucleosome positioning is intriguing. It has been shown that nucleosomal DNA is generally methylated at a higher level than flanking DNA [3]. This may reflect the fact that methyltransferases preferentially target nucleosome-bound DNA where the molecule is topographically exposed and more accessible to modifying enzymes [3]. This observation could explain the coincidence of the predicted nucleosome positions with the two highly methylated regions, CpG1–24 and DMR1. Nucleosomes are on the other hand generally depleted near TF-binding and/or transcription start sites [20], as has also been observed in this study. Nucleosome positioning is one of the major mechanisms to control gene expression because nucleosome-bound DNA is much less accessible than linker DNA for regulatory proteins such as transcription factors [29]. It is possible that the small region within DMR2 that harbours CpG83–89 may be critical in promoting transcription and normally left relatively free from nucleosome binding and methylation. Aberrant methylation of CpGs in this region may prevent binding of regulatory proteins to this region by either direct blockage or by remodelling chromatin through binding of methyl-CpG binding proteins such as MeCP2, which may recruit histone deacetylase and methyltransferase and lead to chromatin condensation [24, 33]. In addition, a nuclease accessibility analysis showed that an open chromatin structure was lost in *MGMT* methylated cells and the TSS became inaccessible even if the region around the TSS itself was not methylated [25, 33]. This suggests that methylation outside of the TSS (such as DMR2) may result in blockage of the TSS through chromatin remodelling, leading to transcriptional silencing. Some of the TFs predicted to bind to this region, such as STAT3, have been linked to gliomagenesis [5]. Whether these TFs actually bind to this region needs to be experimentally confirmed before their potential roles in transcriptional regulation of *MGMT* are considered.

The ultimate evaluation for the efficacy of the *MGMT* methylation assay is to test its ability to predict therapeutic response and/or patients' survival in a large number of clinical cases. The pyrosequencing assay developed in this study is being applied to several independent studies of a large series of malignant astrocytic tumours including the translational part of the BR12 Clinical Trial [2] (Collins et al., in preparation) and the French Multicentre Study (Quillien et al., in preparation). We believe that our findings will help establish an accurate and robust clinical test for *MGMT* methylation in gliomas.

Acknowledgments This work was supported by funds from Samantha Dickson Brain Tumour Trust and Cambridge Fund for the Prevention of Disease (CAMPOD).

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Bhakat KK, Mitra S (2003) CpG methylation-dependent repression of the human O6-methylguanine-DNA methyltransferase gene linked to chromatin structure alteration. Carcinogenesis 24:1337–1345

2. Brada M, Stenning S, Gabe R et al (2010) Temozolomide versus procarbazine, lomustine, and vincristine in recurrent high-grade glioma. J Clin Oncol 28:4601–4608
3. Chodavarapu RK, Feng S, Bernatavichute YV et al (2010) Relationship between nucleosome positioning and DNA methylation. Nature 466:388–392
4. Costello JF, Futscher BW, Kroes RA, Pieper RO (1994) Methylation-related chromatin structure is associated with exclusion of transcription factors from and suppressed expression of the O-6-methylguanine DNA methyltransferase gene in human glioma cell lines. Mol Cell Biol 14:6515–6521
5. de la Iglesia N, Puram SV, Bonni A (2009) STAT3 regulation of glioblastoma pathogenesis. Curr Mol Med 9:580–590
6. Dunn J, Baborie A, Alam F et al (2009) Extent of MGMT promoter methylation correlates with outcome in glioblastomas given temozolomide and radiotherapy. Br J Cancer 101:124–131
7. Esteller M, Hamilton SR, Burger PC, Baylin SB, Herman JG (1999) Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia. Cancer Res 59:793–797
8. Esteller M, Herman JG (2004) Generating mutations but providing chemosensitivity: the role of O6-methylguanine DNA methyltransferase in human cancer. Oncogene 23:1–8
9. Everhard S, Tost J, El Abdalaoui H et al (2009) Identification of regions correlating MGMT promoter methylation and gene expression in glioblastomas. Neuro Oncol 11:348–356
10. Fuxe J, Akusjarvi G, Goike HM, Roos G, Collins VP, Pettersson RF (2000) Adenovirus-mediated overexpression of p15INK4B inhibits human glioma cell growth, induces replicative senescence, and inhibits telomerase activity similarly to p16INK4A. Cell Growth Differ 11:373–384
11. Gerson SL (2004) MGMT: its role in cancer aetiology and cancer therapeutics. Nat Rev Cancer 4:296–307
12. Goike HM, Asplund AC, Pettersson EH, Liu L, Ichimura K, Collins VP (2000) Cryopreservation of viable human glioblastoma xenografts. Neuropathol Appl Neurobiol 26:172–176
13. Harris LC, Potter PM, Tano K, Shiota S, Mitra S, Brent TP (1991) Characterization of the promoter region of the human O6-methylguanine-DNA methyltransferase gene. Nucleic Acids Res 19:6163–6167
14. Harris LC, Remack JS, Brent TP (1994) Identification of a 59 bp enhancer located at the first exon/intron boundary of the human O6-methylguanine DNA methyltransferase gene. Nucleic Acids Res 22:4614–4619
15. Harris LC, Remack JS, Brent TP (1994) In vitro methylation of the human O6-methylguanine-DNA methyltransferase promoter reduces transcription. Biochim Biophys Acta 1217:141–146
16. Hegi ME, Liu L, Herman JG et al (2008) Correlation of O6-methylguanine methyltransferase (MGMT) promoter methylation with clinical outcomes in glioblastoma and clinical strategies to modulate MGMT activity. J Clin Oncol 26:4189–4199
17. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci USA 93:9821–9826
18. Ichimura K, Bolin MB, Goike HM, Schmidt EE, Moshref A, Collins VP (2000) Deregulation of the p14ARF/MDM2/p53 pathway is a prerequisite for human astrocytic gliomas with G1-S transition control gene abnormalities. Cancer Res 60:417–424
19. Kaina B, Christmann M, Naumann S, Roos WP (2007) MGMT: key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. DNA Repair (Amst) 6:1079–1099
20. Kaplan N, Moore IK, Fondufe-Mittendorf Y et al (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 458:362–366
21. Karayan-Tapon L, Quillien V, Guilhot J et al (2010) Prognostic value of O6-methylguanine-DNA methyltransferase status in glioblastoma patients, assessed by five different methods. J Neurooncol 97:311–322
22. Liu L, Backlund LM, Nilsson BR et al (2005) Clinical significance of EGFR amplification and the aberrant EGFRvIII transcript in conventionally treated astrocytic gliomas. J Mol Med 83:917–926
23. Mikeska T, Bock C, El-Maarri O et al (2007) Optimization of quantitative MGMT promoter methylation analysis using pyrosequencing and combined bisulfite restriction analysis. J Mol Diagn 9:368–381
24. Nakagawachi T, Soejima H, Urano T et al (2003) Silencing effect of CpG island hypermethylation and histone modifications on O6-methylguanine-DNA methyltransferase (MGMT) gene expression in human cancer. Oncogene 22:8835–8844
25. Patel SA, Graunke DM, Pieper RO (1997) Aberrant silencing of the CpG island-containing human O6-methylguanine DNA methyltransferase gene is associated with the loss of nucleosome-like positioning. Mol Cell Biol 17:5813–5822
26. Qian XC, Brent TP (1997) Methylation hot spots in the 5′ flanking region denote silencing of the O6-methylguanine-DNA methyltransferase gene. Cancer Res 57:3672–3677
27. Riemenschneider MJ, Jeuken JW, Wesseling P, Reifenberger G (2010) Molecular diagnostics of gliomas: state of the art. Acta Neuropathol 120:567–584
28. Sasai K, Nodagashira M, Nishihara H et al (2008) Careful exclusion of non-neoplastic brain components is required for an appropriate evaluation of O6-methylguanine-DNA methyltransferase status in glioma: relationship between immunohistochemistry and methylation analysis. Am J Surg Pathol 32:1220–1227
29. Segal E, Widom J (2009) What controls nucleosome positions? Trends Genet 25:335–343
30. Stupp R, Hegi ME, Mason WP et al (2009) Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. Lancet Oncol 10:459–466
31. Tabatabai G, Stupp R, van den Bent MJ et al (2010) Molecular diagnostics of gliomas: the clinical perspective. Acta Neuropathol 120:585–592
32. Tost J, El abdalaoui H, Gut IG (2006) Serial pyrosequencing for quantitative DNA methylation analysis. Biotechniques 40:721–722, 724, 726
33. Watts GS, Pieper RO, Costello JF, Peng YM, Dalton WS, Futscher BW (1997) Methylation of discrete regions of the O6-methylguanine DNA methyltransferase (MGMT) CpG island is associated with heterochromatinization of the MGMT transcription start site and silencing of the gene. Mol Cell Biol 17:5612–5619
34. Weller M, Felsberg J, Hartmann C et al (2009) Molecular predictors of progression-free and overall survival in patients with newly diagnosed glioblastoma: a prospective translational study of the German Glioma Network. J Clin Oncol 27:5743–5750
35. Zeng N, Liu L, McCabe MG, Jones DT, Ichimura K, Collins VP (2009) Real-time quantitative polymerase chain reaction (qPCR) analysis with fluorescence resonance energy transfer (FRET) probes reveals differential expression of the four ERBB4 juxtamembrane region variants between medulloblastoma and pilocytic astrocytoma. Neuropathol Appl Neurobiol 35:353–366
36. Zhao W, Soejima H, Higashimoto K et al (2005) The essential role of histone H3 Lys9 di-methylation and MeCP2 binding in MGMT silencing with poor DNA methylation of the promoter CpG island. J Biochem 137:431–440