Quantitative profiling of CpG island methylation in human stool for colorectal cancer detection

Giles O Elliott, Ian T Johnson, Jane Scarll, Jack Dainty Elizabeth A Williams, D Garg, Amanda Coupe, David M Bradburn, John C Mathers and Nigel J Belshaw

G. O. Elliott I. T. Johnson J. Scarll J. Dainty N. J. Belshaw
Institute of Food Research, Norwich Research Park, Norwich, NR4 7UA, UK.

D Garg A. Coupe J. C. Mathers
Newcastle University, Human Nutrition Research Centre, Institute for Ageing and Health, Framlington Place, Newcastle-On-Tyne, NE2 4HH, UK.

Elizabeth A Williams
Department of Oncology, Faculty of Medicine, Dentistry & Health, University of Sheffield, Royal Hallamshire Hospital, Glossop Road, Sheffield, S10 2JF, UK.

D. M. Bradburn
Wansbeck Hospital, Woodhorn Lane, Ashington, Northumberland, UK.

Short title: CpG-island Methylation in stool DNA

Correspondence to: Dr G. O. Elliott, Institute of Food Research, Norwich Research Park, Norwich, NR4 7UA, UK. Email: giles.elliott@bbsrc.ac.uk
Abstract

**Purpose:** To investigate the use of quantitative CGI methylation data from stool DNA to classify colon cancer patients and to relate stool CGI methylation levels to those found in corresponding tissue samples.

**Methods:** We applied a quantitative methylation-specific PCR assay to determine CGI methylation levels of 6 genes, previously shown to be aberrantly methylated during colorectal carcinogenesis. Assays were performed on DNA from biopsies of ‘normal’ mucosa and stool samples from 57 patients classified as disease-free, adenoma or cancer by endoscopy, and in tumour tissue from cancer patients. Additionally CGI methylation was analysed in stool DNA from an asymptomatic population of individuals covering a broad age range (mean = 47 ±24 years).

**Results:** CGI methylation levels in stool DNA were significantly higher than in DNA from macroscopically normal mucosa, and a significant correlation between stool and mucosa was observed for *ESR1* only.

Multivariate statistical analyses using the methylation levels of each CGI in stool DNA as a continuous variable, revealed a highly significant ($p=0.003$) classification of cancer vs. non-cancer (adenoma+disease-free) patients (sensitivity = 65%, specificity = 81%).

**Conclusion:** CGI methylation profiling of stool DNA successfully identified patients with cancer despite the methylation status of CGIs in stool DNA not generally reflecting those in DNA from the colonic mucosa.

Key words: DNA methylation, Colorectal cancer, Stool, Biomarkers, Epigenetics
Introduction

Early detection of tumours is critical in determining successful treatment and improving survival rates in cancer patients. [1] The faecal occult blood test (FOBT) has been shown to reduce the risk of death from colorectal cancer, but is recognised to be of limited sensitivity, [2-6] and prone to false positives. The development of novel, non-invasive screening methods for colorectal cancer detection therefore remains an important area of research.

Mutations in cancer-related genes such as K-Ras [7,8] and APC [9] can be detected in stool, and could be used as an additional strategy for biomarker development. However, even when a panel of genes is used to cover the broad mutation spectrum low sensitivities were achieved (51.6% for detection of cancer [10] and 18-20% for neoplasia [10] [11].

Hypermethylation of CpG islands (CGIs) in a subset of genes is widely recognised as an early epigenetic event in the development of many cancers suggesting its utility as a biomarker for cancer detection. [12-17]

A number of studies have investigated the potential of CGI methylation profiling of DNA extracted from stool as a screening method for early detection of colorectal cancer. [18-30] Generally it appears that a panel of markers provides the most suitable way to achieve high detection rates (sensitivities of 70-94% for cancer and 68-72% for adenoma detection) along with high specificities (77-100%). [19,25,26,23] However, single gene methylation assays such as the vimentin gene detecting 46% of cancer patients with 90% specificity [20] and SFRP2 obtaining sensitivities of 87-94% for cancer and 46-61% for adenoma detection with specificities of up to 93%. [22,28,30] also show promise.

Several aspects of the potential use of CGI methylation in stool as a Colorectal cancer (CRC) biomarker remain to be addressed including the choice of gene(s), the quantitative comparison of methylation data from matching tissue and stool samples, and the determination of methylation levels in asymptomatic free-living individuals. Previous studies have selected genes which have a high frequency of methylation in tumour and/or adenoma samples compared with normal mucosa. This approach may select for late stage cancer unless the methylation also occurs in the normal mucosa as part of the field defect. Evidence indicates that stool DNA from cancer patients is of higher integrity [31] and more readily amplifiable [8] than that from healthy individuals. Furthermore, Klassens et al showed cancer patients stool contained more human DNA per gram of total extracted DNA than that from healthy individuals. [32] and speculated that this could be due to either,
decreased apoptosis of bowel cells and/or increased shedding of cancer cells/inflammatory cells into the colonic lumen, each of which may impact upon CGI methylation detected in stool samples.

We have used quantitative methylation-specific PCR (QMSP) to measure the methylation of 6 CGIs, in genes shown previously to be methylated in colorectal cancer,[33] in DNA from stool samples from 57 patients undergoing endoscopy, and diagnosed subsequently with cancer, adenomas or no neoplasia (disease-free). The selected genes include *ESR1, HPP1, APC* and *MLH1*, which have also been shown to undergo age-related methylation in normal mucosa suggesting a role in the generation of the field defect.[34-36,33,37,38]

Methylation of *p14* has been demonstrated to occur in 30-40% of colorectal cancers[39,40] and tends to be more methylated in multiple tumours[41] suggesting that it too may be involved in a field defect. Furthermore, *p14* methylation in UC patients may be associated with the early stages of neoplasia.[42] Lind et al showed *CDH1* methylation in approximately 40% of primary tumours[40] and increased methylation was also associated with inflammatory bowel disease.[43]

While CGI methylation data was able to significantly distinguish patients by disease, the levels of methylation in matching tissue (tumour and macroscopically normal mucosa) and stool samples showed that CGI methylation for five of the six genes did not correlate.
MATERIALS AND METHODS

Subject recruitment and collection of stool samples

Two groups of volunteers were recruited. The first were healthy volunteers recruited from the general population of North Eastern England. Individuals were recruited from among the staff and students of Newcastle University, and via posters in community centres, GP surgeries, and care homes for the elderly. The aim was to recruit 160 volunteers comprising 10 males and 10 females in each of the following age bands: <20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80+. The second group of volunteers were patients recruited from the gastroenterology outpatient and surgical lists of Wansbeck General Hospital, Ashington, Northumberland and included patients with colorectal cancer (Cancer), those with adenomatous polyps (Adenoma) and those in whom no large bowel pathology was detected (Disease-free). Out-patients were sent information detailing the study along with the notification of appointment from the hospital. Six to eight weeks after the endoscopy procedure the medical notes for each of the volunteers were reviewed, and the final diagnosis was recorded, based upon the findings of the pathology report and the conclusions of the responsible consultant. All volunteers were provided with a sealable pathology pot and an adapted bed-pan to facilitate stool collection, and asked to provide a faecal sample for laboratory analysis. Stool samples were collected at home from the healthy volunteers and from patients – the latter were obtained during the 2 weeks preceding endoscopy or surgery. Volunteers and patients were asked to contact the laboratory as soon as a stool sample became available, to enable transfer to the laboratory within 2 h. Stools were sub-sampled (250mg) into Eppendorf tubes and frozen at -80°C for subsequent DNA extraction and methylation analysis. Ethical approval for the project was received from the Northumberland Local Research Ethics Committee (Project reference NLREC2/2001) and samples were obtained after informed consent was received.

Stool DNA extraction, Bisulphite modification and quantification of CGI methylation

DNA was purified from stool sub-samples using the QIAamp DNA stool mini kit (QIAgen) according to the manufacturer’s protocol for enrichment of human DNA. Stool DNA (20µg) was bisulphite modified as described previously.[33] The efficiency of bisulphite modification was determined by sequencing cloned PCR fragments from the ESR1 CGI. A random PCR fragment per batch of bisulphite-modified samples was analysed. The average conversion efficiency was 98%.

CGI methylation was measured in all samples for 6 genes, HPP1/TMEFF2 (Hyperplastic polyposis protein 1/Transmembrane protein with EGF-like and two follistatin-like domains 2), ESR1 (Estrogen receptor
α), APC (Adenomatous Polyposis Coli), MLH1 (MutL homologue), p14ARF (cyclin-dependent kinase inhibitor) and CDH1 (E-cadherin) using a Quantitative Methylation-Specific PCR assay (QMSP) as described previously.[18]

**Statistical analysis**

Unpaired T-Tests were performed to determine the significance of differences between mean methylation levels for sample groups. Binomial or multinomial logistic regression models were used to identify differences in methylation patterns between the patient groups. Models were fitted by regressing the stool sample by patient type on a number of genes, as well as on age and sex. These variables were then “pruned” using both backwards elimination using ANOVA-type tests, and an automated stepwise procedure for optimising the Akaike Information Criterion (AIC). “Leave-one-out” cross-validation was used to estimate the classification error-rate. This was compared with the expected rate given by the proportional chance criterion using an exact binomial test (one-sided) to test the null-hypothesis that the given success rate of classification was no better than chance. Linear Discriminant Analysis (LDA) was also employed. The variables used in these models were both the full set of explanatory variables, and the sets of variables selected by the binomial regression models fitted previously to the data. As for the binomial regression models, “leave-one-out” cross-validation rates were compared with the expected rates from the proportional chance criterion. Statistical analyses were carried out using Minitab Release 14 (Minitab Inc, State College, PA) or the statistical programming environment “R”. [44]
Results

CGI methylation in stool DNA from endoscopy patients

CGI methylation levels were determined in stool samples from 57 endoscopy patients classified as either cancer (N=19 (11 males 8 females) Mean age=72.0±9.36), Adenoma (N=18 (15 males 3 females) Mean age=64.7±9.13), or disease free (N=20 (10 males 10 females) mean age=54.3±12.85). Mean stool methylation levels were lower in cancer patients than either adenoma or disease free patients for 4/6 of the CGIs studied (CDH1, HPP1, APC and ESR1) (Figure 1). This difference was statistically significant for ESR1 when comparing cancer patients with either adenoma patients (p=0.04) or with adenoma plus disease-free patients (p=0.02). Of the remaining CGIs analysed, mean stool methylation levels for both MLH1 and p14 were significantly higher in disease-free compared with adenoma patients (p=0.011 and p=0.015 respectively). Regression analysis did not reveal a significant correlation between subject age and stool methylation status for any of the genes across all patients.

Multinomial and LDA modelling techniques were used to determine if stool CGI methylation could distinguish between patient groups (cancer, adenoma and disease-free) (Table 1). LDA using the variables age and methylation of HPP1, ESR1 and p14 (selected by an automated stepwise procedure based on optimising the AIC), resulted in an overall correct classification rate of 61%, (sensitivity = 76%, specificity = 73% for cancer patients and sensitivity = 19%, specificity = 97% for adenoma patients). This was determined as significant (p=3.7x10^-5) using an exact one-sided test of 61% correctly classified versus the “proportional chance” criterion.

Stool CGI methylation also distinguished cancer from non-cancer patients (adenoma plus disease-free patients) (Table 1). Binomial modelling using the variables HPP1, ESR1, p14 and APC (selected by stepwise AIC optimisation) correctly classified 76% of samples (p=0.003) (sensitivity = 65%, specificity = 81%). Binomial modelling using backwards elimination for selection of variables was also performed and correctly classified 72% of subjects, using the variables HPP1, ESR1, and p14 (sensitivity = 41%, specificity = 86%). Whilst not as successful as the AIC optimised model, this classification was also statistically significant (p=0.015). LDA modelling using all variables (age and methylation of HPP1, ESR1, APC1, MLH1, p14 and CDH1) was performed to distinguish cancer from non-cancer and correctly classified 78% of samples (p=0.001), (sensitivity = 71%, specificity = 81%).

Finally, we attempted to distinguish adenoma vs. disease-free patients. LDA and binomial models were performed, which both gave the same classification rate using the variables age and methylation of p14 and
Overall 75.6% of samples were correctly classified with a sensitivity of 69% for adenoma detection and a specificity of 81%; the proportional chance criterion indicated that this classification was statistically significant \( (p=0.0018) \).

**Association between stool and tissue methylation**

Previously we have reported the methylation status of the CGIs studied here in DNA extracted from macroscopically normal biopsies of colorectal mucosa[33] for the same patients for whom gene methylation levels in stool samples are reported here. Comparison of the normal mucosa data with the stool data revealed a significant positive correlation for methylation of \( \text{ESR1} \) \( (p=0.003) \) only. Indeed, there was a significant \( (p=0.0002) \) inverse relationship for \( \text{APC} \) methylation between stool and normal mucosa. No significant correlations were observed for any of the other CGIs studied, in either the macroscopically normal mucosa or tumour samples. For all CGIs studied, mean methylation levels from the stool samples were significantly \( (p<0.0001) \) higher than those from the flat mucosa (Figure 2).

**CGI methylation in stool samples from healthy volunteers**

CGI methylation was also measured in stool samples collected from 169 healthy volunteers. Regression analysis indicated that there was a significant positive correlation of \( \text{CDH1} \) methylation with age \( (p=0.05) \) in the total population. This correlation was significant for males \( (p=0.04) \), but not females \( (p=0.45) \). There were no significant associations with age for any of the other genes.

The binomial model using AIC optimisation for classification of cancer vs. non-cancer developed using the patient group was applied to the healthy volunteers, resulting in 84% (142/169) of the “healthy volunteers” being classified as non-cancer. This translates into a potential false positive rate of 16% which is in agreement with the sensitivity obtained for the patient group. The LDA model using all variables classified 90% of the volunteers as “non-cancer” indicating a slightly better performance for this model in these subjects than for the patient group. Analysis of the mean methylation levels in the volunteer group after classification, i.e. comparing CGI methylation in those correctly classified as “non-cancer” with those incorrectly classified as “cancer”, using the binomial model indicated that there were significant differences in the four CGIs used for the binomial model with decreased methylation of \( \text{HPP1} \) \( (p=0.00018) \), \( \text{ESR1} \) \( (p=4.2\times10^{-15}) \)and \( \text{APC} \) \( (p=0.023) \) and increased methylation of \( \text{p14} \) \( (p=2.6\times10^{-5}) \) in those volunteers incorrectly classified as “cancer”.

---

MLH1.
Comparison of normal patients with age- and sex-matched healthy volunteers

A subset of the healthy volunteers was age- and sex-matched to the disease-free patients (2 per patient). Mann-Whitney t-tests were used to determine if stool CGI methylation levels in the healthy volunteers differed from those in the disease-free patients (Figure 3). In all cases mean CGI methylation levels were higher in stool from the healthy volunteers than from the matched disease-free patients and these were statistically significant for HPP1 \( (p<0.0001) \), ESR1 \( (p=0.028) \), MLH1 \( (p<0.0001) \) and p14 \( (p=0.0001) \).
Discussion

The role of CGI methylation in the development of cancer has been extensively documented and evidence is accumulating to support the hypothesis that epigenetic alterations may prior to disease initiation.[45] Such alterations are therefore under investigation as potential biomarkers for early stage detection. For CRC, a number of studies have explored stool DNA as a source of CGI methylation-related biomarkers and most have used non-quantitative techniques to determine the presence or absence of CGI methylation in stool from patients with or without cancer. We attempted to advance this area by quantifying CGI methylation of cancer-related genes in DNA from stool samples, and using these data as linear variables in multinomial and binomial statistical modelling. Furthermore, we made a quantitative comparison of CGI methylation in stool and corresponding tissue samples (macroscopically-normal mucosa and tumour tissue) from the same patients.

Contrary to expected, our results indicated that, in general, stool methylation profiles do not reflect the methylation status of either the corresponding macroscopically normal mucosa, nor the same genes in tumour tissue in cancer patients. The methylation patterns of genes in DNA recovered from stool are nonetheless related to the disease state of the individual, as evidenced by the statistically significant classifications of subjects according to disease status using multinomial and/or binomial modelling.

Previous studies of CGI methylation in stool predominantly used non-quantitative MSP methods.[46,20,24,25,29] Cancer patients are reported to have more human DNA of higher integrity in their faeces than non-cancer patients,[31,47,48] possibly due to the preferential shedding of intact tumour cells, resistant to apoptosis, into the gut lumen. The single-stage MSP assays applied to stool samples in previous studies, or the qPCR assay implemented by Hellebrekers et al.[49] may be biased in favour of detecting CGI methylation for cancer patients because their stools contain sufficiently high concentrations of intact DNA for amplification. Raising the possibility that negative results for disease-free subjects may simply reflect low levels of intact DNA insufficient for PCR amplification rather than the absence of CGI methylation.

An assumption made in using stool DNA as a source of biomarkers for CRC is that stool methylation reflects directly the mucosal methylation, either from the flat mucosa and/or from tumour tissue. Of the six genes analysed here, there was a significant positive correlation between matched stool and macroscopically normal tissue for methylation of $ESR1$ only, consistent with our previous study.[18] Also, there were no significant associations between stool and tumour methylation patterns. This suggests that the levels of CGI methylation in stool do not reflect those in the macroscopically normal (flat) mucosa and are not of solely of tumour origin, which is supported by Azuara and colleagues[50] who showed that patients with methylation
positive lesions do not always present methylation positive stool DNA. This lack of correlation may be explained in part if, only a small fraction of healthy epithelial cells are shed into the lumen, the majority having undergone apoptosis and recycling within the mucosal compartment.[51] Possibly human DNA obtained from stool is derived from a sub-set of apoptosis-resistant cells which differ in methylation patterns from those of the majority of cells in the normal epithelium, which in cancer patients may then also be mixed with cells shed from the tumour.

Here, as in our previous report,[18] we found that CGI methylation is significantly higher in stool samples than in the corresponding normal mucosa. Since this difference occurred in both cancer and non-cancer patients it cannot be attributed to tumour cells in the stool sample, but supports the hypothesis that cells containing methylated CGIs, are present in the fecal stream. Mechanistic support for this comes from observations that, in the intestine, cell detachment-induced apoptosis (anoikis) depends on Fas receptor signalling[52] and CGI methylation-induced silencing of Fas expression prevents Fas-ligand-induced apoptosis[53,54]. Alternatively, or in addition, highly methylated DNA domains (perhaps within heterochromatin) may resist digestion for longer than unmethylated DNA regions when exposed to bacterial nucleases in stool.

Mean stool CGI methylation levels in the free-living healthy population were higher for all genes studied compared with those in disease-free, but symptomatic, patients who had been referred for endoscopy. There is no obvious biological explanation for this observation but it may indicate that the origins and fate of human DNA in the stools from symptomatic patients are influenced by other (non-neoplastic) functional disorders; however this remains to be established. However, this difference may be important if screening protocols developed using symptomatic healthy individuals are to be applied to a healthy asymptomatic population.

Despite the lack of correlation between CGI methylation in stool and corresponding tissue, we obtained highly significant classifications of cancer versus non-cancer patients using the methylation status of a relatively small number of CGIs. Binomial models of cancer versus non-cancer patients (i.e. adenoma plus disease-free individuals) performed best, and when we used the combination of CGI methylation plus age and sex, we achieved a sensitivity of 71% and a specificity of 81%. When this model was applied to data from the free-living healthy volunteers, 90% were classified as non-cancer resulting in a higher specificity from an asymptomatic population using this model. However, age was included as a factor in this model, and the mean age of our cancer group was significantly (p<0.01) higher than for the adenoma and disease-free groups (72.0 v.
64.7 and 54.3 years respectively). This is a reflection of the fact that younger patients presenting for routine endoscopy are less likely to have cancer than are older patients. Although further statistical analysis revealed that age alone was particularly poor as a predictor of cancer (Table 1), we cannot exclude the possibility that its contribution to the model is factitious. Further testing of this model on an independent set of age- and sex-matched patients will be needed. Nevertheless, binomial regression with AIC optimisation generated a model that was independent of age and sex but was equally specific in separating cancer from non-cancer, with only a small (6%) decrease in sensitivity. We also obtained a highly significant classification of adenoma vs. disease-free patients (69% sensitivity, 81% specificity), using a model based solely on stool CGI methylation levels.

This study indicates there is a complex relationship between stool methylation and the epigenetic status of the corresponding colorectal mucosa, probably because the human DNA which can be harvested from stool represents only a small, and variable, sub-population of colorectal mucosal cells. Despite this complexity, the analysis of stool DNA may prove useful as a non-invasive research tool e.g. for investigating the effects of aging, diet and other important exposures on the epigenetic status of the alimentary tract in large population studies. In addition, this approach may ultimately provide clinically useful biomarkers of disease. However, further research will be necessary to determine both the origins and nature of the DNA fractions which survive in stool and to identify the most informative CpG islands for cancer diagnosis inorder to develop robust clinical screening methods.

Acknowledgments

We thank the UK Food Standards Agency (Project No. N12009) and the Biotechnology and Biological Sciences Research Council (42212A) for financial support.
References


Table 1. Multivariate statistical modelling for the classification of patient stool samples as cancer vs. non-cancer based on CGI methylation levels. *P*-values represent the significance of the overall classification compared with random chance based on the proportional chance criterion. Variables selected by the Akaike Information Criterion (AIC) method were methylation of *HPP1, ESR1, p14* and *APC*; variables selected by Backwards Elimination (BE) were methylation of *HPP1, ESR1* and *p14*. Success rates of using age alone in both models is also presented.

<table>
<thead>
<tr>
<th>Method</th>
<th>Overall classification</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Binomial</strong></td>
<td>AIC</td>
<td>76%</td>
<td>65%</td>
<td>81%</td>
</tr>
<tr>
<td></td>
<td>BE</td>
<td>72%</td>
<td>41%</td>
<td>86%</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>68%</td>
<td>12%</td>
<td>95%</td>
</tr>
<tr>
<td><strong>LDA</strong></td>
<td>All Variables</td>
<td>78%</td>
<td>71%</td>
<td>81%</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>69%</td>
<td>6%</td>
<td>97%</td>
</tr>
</tbody>
</table>
**Fig 1.** Mean CGI methylation levels for 6 genes in stool samples from Disease-free, Adenoma and Cancer patients undergoing endoscopy. Values are mean levels of methylation ± SEM for 19 cancer, 18 adenoma and 20 disease-free patients. * p≤0.05, ** p≤0.01.

**Fig 2.** CGI methylation for 6 genes in the macroscopically normal mucosa and stool from endoscopy patients. Values are mean levels of CGI methylation ± SEM for 57 patients (19 cancer, 18 adenoma and 20 disease-free patients). In all cases CGI methylation was significantly higher in stool than in the corresponding macroscopically normal mucosa (p < 0.0001).
Fig 3. Comparison of mean stool CGI methylation levels ± SEM for 6 genes in age and sex-matched normal patients and healthy volunteers. Statistical significance of the differences are indicated by * p=0.01, ** p=0.0001.