Abstract

Background: Recent investigations have demonstrated the clear heterogeneity of sporadic colorectal cancer (CRC) with regard to CpG island methylation. Two unsupervised cluster analyses revealed that CRCs form three distinct DNA methylation subsets, which are referred to as the high-, intermediate- and low-methylation epigenotypes (HME, IME, and LME, respectively). A recent study by Yagi et al. found a fairly sensitive and specific identification of HME, IME and LME using two marker panels analysed by MALDI-TOF mass spectrometry (MassARRAY). However, the expensive equipment required for this method substantially increases the cost and complexity of the assay.

Findings: In this article, we demonstrate the assessment of HME, IME and LME in a group of 233 sporadic CRCs using seven markers proposed by Yagi et al. The DNA methylation of each marker was quantified using combined bisulphite restriction analysis (COBRA) together with an analysis of various genetic factors associated with CRC (the *BRAF* and *KRAS* mutations and microsatellite instability (MSI)). The baseline methylation of each marker was generated from pooled DNA isolated from 50 normal colon tissues. **Conclusions:** We demonstrate that the correlation of HME, IME and LME epigenotyped by COBRA using different molecular classifiers is similar to that achieved by MassARRAY. Therefore, epigenotyping CRCs using COBRA is a simple, specific and cost-effective method that has the potential to be widely used in CRC research.

Keywords: colorectal cancer, epigenotypes, methylation, methylator, COBRA

Findings

A loss of the expression of key genes may occur during carcinogenesis through genetic and epigenetic means. The best-defined epigenetic alteration of genes involves DNA hypermethylation of CpG islands in promoter regions associated with the transcriptional inactivation of the affected genes [1]. Studies on sporadic colorectal cancers (CRCs) have indicated the existence of a high frequency of aberrant promoter hypermethylation in a subset of CRCs known as the CpG island methylator phenotype (CIMP) and classified CRC cases into two groups: CIMP+ and CIMP- [2]. Tumors displaying CIMP were reported to have a distinct clinical, pathological and molecular profile, such as associations with proximal tumor location, poor tumor differentiation, *hMLH1* methylation, microsatellite instability (MSI), as well as high *BRAF*, low *p53* mutation rates and high LINE-1 methylation rates [3]. Two recent studies using unsupervised cluster analysis found that CRCs could be clearly classified into three epigenetic groups which were referred to by Yagi et al. [4] as the high-, intermediate- and low-methylation epigenotypes (HME, IME, and LME, respectively) [5]. HME CRCs were strongly associated with proximal localization, the *BRAF* mutation and MSI whereas IME were associated with the *KRAS* mutation. In 2010 Yagi et al. developed two panels consisting seven markers to distinguish between the HME and IME groups [4]. However, the technology that was adopted by Yagi et al. [4] to assess the methylation of markers (MALDI-TOF mass spectrometry) requires expensive equipment and, together with the complexity of the methodology, this can make it difficult to use in routine research [6]. Therefore, to quantify the methylation of each marker, we adopted combined bisulphite restriction analysis (COBRA), which is considered to be a simple and cost-effective method [7].

As the preferential amplification of unmethylated or methylated alleles may significantly influence the results of methylation quantification, to obtain methylation data in an unbiased fashion, for each measurement we applied the correction for bias described by Moskalev et al. [8]. To evaluate the methylation of 7 markers in CRC in relation to a normal colon tissue in an non-arbitrary manner, we measured the methylation of each marker in two pools of DNA extracted from proximally (n=25) and distally (n=25) resected normal colon tissues. We did not observe any differences in the level of methylation of each marker between these pools (data not shown). This led to the estimation of the average baseline methylation of each marker in a normal colon tissue (see Table 1) [9]. Applying this estimate, a tumor sample with a

methylation rate ≥5% above this baseline level was defined to be methylation positive.

Using the seven markers (*CACNA1G*, *SLC30A10*, *LOX*, *ELMO1*, *FBN2*, *THBD* and *HAND1*) and the classification described by Yagi et al. [4], we epigenotyped cohort sporadic CRCs consisting 233 samples that included 46% females and 54% males, 29% proximal and 71% distal tumors, 7% MSI tumors, 6% tumors with *hMLH1* hypermethylation, 9% *BRAF* V600E mutants and 24% *KRAS* (codon 12) mutants (Figure 1, Table 1). These proportions are consistent with previously observed large samples of sporadic colorectal cancers [2,3,10]. According to epigenotyping, the HME, IME and LME groups constituted 11%, 36% and 52% of the sample, respectively. We focused on various different classifiers that have previously been suggested to be associated with CIMP (Figure 2, Table 1) [11]. HME CRCs were significantly associated with proximal location, *BRAF* mutation, *hMLH1* methylation and MSI, when compared with IME CRCs, $p=9.1x10^{-6}$, $p=8.1x10^{-11}$ $3.5x10^{-10}$ and $1.3x10^{-7}$, respectively and when compared to LME CRCs, $p=2.1x10^{-6}$, $p=6.4x10^{-13}$ $5.9x10^{-10}$ and $8.7x10^{-11}$, respectively. IME CRCs were significantly associated with the *KRAS* mutation when compared with HME and LME CRCs (p=0.0082 and $p=7.2x10^{-5}$, respectively). These correlations are fully consistent with the results of studies where the CRCs were classified into three epigenetically distinct subsets [3,4,5].

In conclusion, the current study presents a simple and reliable method for the quantitative epigenotyping of CRCs, which could be an economical alternative to MethyLight or MassARRAY. However, the limitations of this method, when compared to aforementioned two, are its medium level of throughput and the low number of single CpG sites that can be analyzed.

Materials and methods

Surgically resected frozen tissues of 233 colorectal cancers and matched normal colon samples were obtained from the 2nd Department of General and Oncological Surgery, Wroclaw Medical University and the 1st Department of Surgical Oncology, Lower Silesian Oncology Center. The CRC patient group consisted entirely of Polish individuals (all Caucasians). Only patients with primary, sporadic colorectal cancer who had not received preoperative therapy were included in the studies. Informed consent was obtained from all the patients. The study was

accepted by the Wroclaw Medical University Ethics Committee. Genomic DNA was prepared using the Gentra Puregene Tissue Kit (Qiagen, Hilden, Germany)) according to the manufacturer's manual.

Detection of *BRAF* V600E in tumour tissues was carried out using the mutant allele-specific PCR amplification described by Sapio et al. [12]. Mutations at codon 12 of the *KRAS* gene were detected by PCR–RFLP as described by Miranda et al. [13]. Microsatellite instability was determined by pentaplex PCR, using the quasimonomorphic markers BAT-26, BAT-25, NR-21, NR-22, and NR-24, as described in detail by Buhard et al. [14].

The methylation of HME and IME markers was analyzed according to a combined bisulphite restriction assay (COBRA) consisting sodium bisulphite treatment followed by PCR, restriction digestion and quantitation [15]. Bisulphite treatment of 1 µg genomic DNA obtained from resected frozen tissues was carried out using the EpiTect kit (Qiagen). Primers were designed to include no CpG site. The primer sequences and amplification conditions are described in Table 2. Briefly, the PCR was carried out in a 12µl solution containing 50 ng of the bisulphite-treated DNA, 1× PCR buffer (Qiagen), 0.4mM dNTPs, 0.4µM forward and reverse primers and 0.75U HotStarTaq DNA Polymerase (Qiagen). PCR reactions were hot-started at 95°C for 15 min, subsequently denatured for 30 s at 95°C, with annealing for 30 s at the appropriate temperature for each primer (Table 1) and an extension for 30 s at 72°C. Thirty-five cycles were used to amplify the PCR products to the expected sizes in a MJ Mini Thermal Cycler (Bio-Rad). The PCR products were subsequently digested with 5U of the appropriate enzymes (see Table 2, MBI Fermentas) at 37°C overnight. These enzymes were selected using NEBcutter2 and Snake Charmer software [http://methdb.igh.cnrs.fr/cgrunau/methods/snake_charmer.html] on the basis of their specificity to a methylated sequence (i.e. only a methylated amplicon was digested) [16]. The products of digestion were then run on an 3% agarose gel stained with SyBr Green, and the band intensities were measured by MultiDoc-It Imaging System (UVP) using VisionWorksLS software (UVP). To assess the baseline methylation of each marker, we constructed two independent pools by combining equimolar amounts of DNA extracted from normal colon tissues: proximal (n=25) and distal (n=25). Each pool was quantified in triplicate. To correct biased methylation data, we used the method of cubic polynomial regression described by Moskalev et al. [8]. Briefly, fully methylated (CpG Methylated Jurkat Genomic DNA,

New England Biolabs) and unmethylated control DNA (generated by whole-genome amplification, REPLI-g Mini Kit, Qiagen) were bisulphite-treated and mixed in different ratios to obtain defined methylation levels (0%, 25%, 75% and 100%). The mixtures were then used to obtain a bias correction value for each marker, which was subsequently used to calculate accurate methylation levels for each marker in the sample. The statistical package R was used to carry out the necessary calucations. Epigenotypes were defined using a specific panel of markers and the criteria described by Yagi et al. [4]. Briefly, after the analysis of the methylation of a panel of three markers (*CACNA1G*, *SLC30A10* and LOX), HME tumors were defined as those with at least two methylated markers. The remaining tumors were screened using five markers (*SLC30A10, ELMO1*, *FBN2*, *THBD* and *HAND1)*, IME tumors were defined as those with at least three methylated markers. Tumors not classified as HME or IME were designated LME.

Methylation of the *hMLH1* promoter (the so called Deng C region) was analysed by methylation-specific PCR as described by Chan et al. [17].

The Mann-Whitney U test was used to compare mean age and Fisher's exact test was used to examine the association between epigenotypes and categorical variables. All p-values were two sided. Because multiple comparisons were carried out in this study, the Bonferroni correction was taken into account and therefore the statistical significance level was defined to be 0.7%.

List of abbreviations

CRC – colorectal cancer; CIMP - CpG island methylator phenotype; COBRA - combined bisulphite restriction analysis; MSI - microsatellite instability; HME – high methylation epigenotype; IME - intermediate methylation epigenotype; LME – low methylation epigenotype.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PK designed the study, optimized all the reactions, analyzed/interpreted the data and prepared the manuscript. ESz performed CRC epigenotyping. BM supervised the collection of the samples and analyzed clinical data. DR performed statistical

analyses and proofread the manuscript. PL analyzed the *BRAF* and *KRAS* mutations. MB, TS, and WK performed clinical sampling and gave pathological expertise. AJ and AL performed the analysis of microsatellite instability. MMS was responsible for the acquisition of funding, group supervision and drafting the manuscript. All the authors have read and approved the final version of the manuscript.

Acknowledgements

This study was supported by a grant from the State Committee for Scientific Research, Polish Ministry for Scientific Research and Information Technology No. 6041/B/P01/2010/38 2010–2013. PK was supported by research fellowship within "Development program of Wroclaw Medical University" funded from European Social Fund, Human Capital, National Cohesion Strategy No. UDA-POKL.04.01.01-00-010/08-01. DR was supported by Science Foundation Ireland under the BIO-SI project No. 07MI012.

References

1. Schuebel KE, Chen W, Cope L, Glöckner SC, Suzuki H, Yi JM, Chan TA, Van Neste L, Van Criekinge W, van den Bosch S, *et al*: **Comparing the DNA hypermethylome with gene mutations in human colorectal cancer.** *PLoS Genet* 2007 **3**:1709-23.
2. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D, *et al*: **CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer.** *Nat Genet* 2006 **38**:787-93.
3. Nosho K, Irahara N, Shima K, Kure S, Kirkner GJ, Schernhammer ES, Hazra A, Hunter DJ, Quackenbush J, Spiegelman D, *et al*: **Comprehensive biostatistical analysis of CpG island methylator phenotype in colorectal cancer using a large population-based sample.** PLoS One 2008 **3**:e3698.
4. Yagi K, Akagi K, Hayashi H, Nagae G, Tsuji S, Isagawa T, Midorikawa Y, Nishimura Y, Sakamoto H, Seto Y, *et al*: **Three DNA methylation epigenotypes in human colorectal cancer.** *Clin Cancer Res* 2010 **16**:21-33.

5. Shen L, Toyota M, Kondo Y, Lin E, Zhang L, Guo Y, Hernandez NS, Chen X, Ahmed S, Konishi K, *et al*: **Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer.** *PNAS* 2007 104:18654–9.
6. Kristensen LS, Hansen LL: **PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment.** *Clin Chem* 2009 **55**:1471-83.
7. Clark SJ, Statham A, Stirzaker C, Molloy PL, Frommer M: **DNA methylation: bisulphite modification and analysis.** *Nat Protoc* 2006 **1**:2353-64.
8. Moskalev EA, Zavgorodnij MG, Majorova SP, Vorobjev IA, Jandaghi P, Bure IV, Hoheisel JD: **Correction of PCR-bias in quantitative DNA methylation studies by means of cubic polynomial regression.** *Nucleic Acids Res* 2011 doi:10.1093/nar/gkr213 1–12.
9. Docherty SJ, Davis OS, Haworth CM, Plomin R, Mill J: **Bisulfite-based epityping on pooled genomic DNA provides an accurate estimate of average group DNA methylation**. *Epigenetics & Chromatin*. 2009 **2:**3.
10. Samowitz WS, Albertsen H, Herrick J, Levin TR, Sweeney C, Murtaugh MA, Wolff RK, Slattery ML: **Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. Gastroenterology.** 2005 **129:**837-45.
11. Ogino S, Goel A: **Molecular classification and correlates in colorectal cancer.** *J Mol Diagn* 2008 **10:**13-27.
12. Sapio MR, Posca D, Troncone G, Pettinato G, Palombini L, Rossi G, Fenzi G, Vitale M: **Detection of BRAF mutation in thyroid papillary carcinomas by mutant allele-specific PCR amplification (MASA)**. *Eur J Endocrinol* 2006 **154:**341-8.
13. Miranda E, Destro A, Malesci A, Balladore E, Bianchi P, Baryshnikova E, Franchi G, Morenghi E, Laghi L, Gennari L, Roncalli M: **Genetic and epigenetic changes in primary metastatic and nonmetastatic colorectal cancer.** *Br J Cancer* 2006 **95:**1101-7.
14. Buhard O, Cattaneo F, Wong YF, Yim SF, Friedman E, Flejou JF, Duval A, Hamelin R: **Multipopulation analysis of polymorphisms in five mononucleotide repeats used to determine the microsatellite instability status of human tumors.** *J Clin Oncol* 2006 **24:**241-51.

15. Xiong Z, Laird PW: **COBRA: a sensitive and quantitative DNA methylation assay.** *Nucleic Acids Res* 1997 **25:**2532-4.
16. Vincze T, Posfai J, Roberts RJ: **NEBcutter: A program to cleave DNA with restriction enzymes.** Nucleic Acids Res. 2003 **31:**3688-91.
17. Chan AO, Issa JP, Morris JS, Hamilton SR, Rashid A: **Concordant CpG island methylation in hyperplastic polyposis**. *Am J Pathol* 2002 **160**:529 – 36.

Figure Legends

Figure 1. Representative results of COBRA. PCR products were digested and separated on a 3% agarose gel. Lines: 1, 2 – HME tumors; 2, 3 – IME tumors; 4, 5 – LME tumors; 6,7 - distal and proximal normal colon DNA pools; 8 - blank control; 9 – GeneRuler100 bp DNA Ladder (Fermentas).

Figure 2. Distribution of various classifiers according to the epigenotype status of 233 CRCs.

Tables

Table 1. Clinical and molecular characteristics of the sample of 233 CRCs according to epigenotype status.

Clinical and molecular characteristics	Total *N* (%)	HME *N* (%)	IME *N* (%)	LME *N* (%)	p-value§ (HME vs IME)	p-value§ (HME vs LME)	p-value§ (IME vs LME)
	233 (100)	26 (11)	85 (36)	122 (52)			
Age mean ± SD	65.1 ±10.7	69.2 ±12.4	65.9 ±10.7	63.7 ±10.1	0.14	0.01	0.09
Female	107 (46)	16 (62)	36 (42)	55 (45)			
Male	126 (54)	10 (38)	49 (58)	67 (55)	0.12	0.14	0.78
Proximal	67 (29)	19 (73)	20 (24)	28 (23)			
Distal	166 (71)	7 (27)	65 (76)	94 (77)	**9.1×10^{-6}**	**2.1×10^{-6}**	1
BRAF mut.(+)	20 (9)	16 (62)	2 (2)	2 (2)			
BRAF mut.(-)	213 (91)	10 (38)	83 (98)	120 (98)	**8.1×10^{-11}**	**6.4×10^{-13}**	1
KRAS mut.(+)	55 (24)	3 (12)	34 (40)	18 (15)			
KRAS mut.(-)	178 (76)	23 (88)	51 (60)	104 (85)	0.01	0.11	**7.2×10^{-5}**
MSI (+)	17 (7)	13 (50)	3 (4)	1 (1)			
MSI (-)	216 (93)	13 (50)	82 (96)	121 (99)	**1.3×10^{-7}**	**8.7×10^{-11}**	0.31
hMLH meth.(+)	15 (6)	13 (50)	0 (100)	2 (2)			
hMLH meth.(-)	218 (94)	13 (50)	85 (0)	120 (98)	**3.5×10^{-10}**	**5.9×10^{-10}**	0.51

§ - Differences significant after the application of the Bonferroni correction for multiple testing $p\leq0.007$

Using the less conservative Benjamini-Hochberg procedure for multiple testing, all the p-values ≤0.01 given above are significant.

Table 2. Primer sequences, annealing temperature and restriction enznymes used in the COBRA analyses.

Gene	Primer sequence	PCR product coordinates (Build 35)	PCR product (bp)	Annealing temperature (°C)	$MgCl_2$ concentration (Mm)	Restriction enzyme	Digestion product size (bp)§	Baseline methylation (%)
CACNA1G	F: 5' GTTTTGGTTTAAGTAGAAGAAAATT 3' R: 5' ACCCACCAAATATACCCC 3'	45994019 45993711	309	56.5	1.5	*HinfI*	135; 91; 83§	38
SLC30A10	F: 5' GGGGGTAGTATTTGAATAGTT 3' R: 5' TCTAAATAAATCCCACCTCTAC 3'	216490380 216490590	211	56	1.5	*HinfI*	122; 89	5
LOX	F: 5' TTTTTTTATTGGATTTGTTGG 3' R: 5' ATTAAAAATTTCTCCTTCCCTC 3'	121441431 121441676	246	57	*3*	*RsaI*	129; 117	20
ELMO1	F: 5' TTTTAGGGAGGAAATAAAAGTG 3' R: 5' CAATTCCCCTTCAATAATAAC 3'	37261371 37261599	229	57	1.5	*TaqI*	131, 98	0
FBN2	F: 5' GTAAAATTTTAAAAAATGTGAATTTT 3' R: 5' AACCCTCCTCTTCTTCTAAAA 3'	127901540 127901701	161	55	1.5	*RsaI*	116; 45	0
THBD	F: 5' AAAGGAAGGAAGTGTTTGG 3' R: 5' ATTACAAAAATACTAACCTTCCCT 3'	22978370 22978560	191	55	1.5	*BstUI*	107; 84	0
HAND1	F: 5' TTTTTTAAGAATTTTTGGTTTTT 3' R: 5' AAAAATAACCCTACCCCC 3'	153838977 153839199	223	55	2.25	*HinfI*	172; 51	5