DNA methylation biomarkers for non-invasive diagnostic of 
Colorectal Cancer

F. Javier Carmona1,*, Daniel Azuara2,*, Antonio Berenguer-Llergo3,4,*, Agustin F. Fernández1,5, Sebastiano Biondo6,7, Javier de Oca6,7, Francisco Rodriguez-Moranta8, Ramón Salazar2,4, Alberto Villanueva2, Mario F. Fraga5,9, Jordi Guardiola8, Gabriel Capellá2,10, Manel Esteller1,7,11, Victor Moreno3,4,7

* These three authors contributed equally to the study

1 Cancer Epigenetics and Biology Program (PEBC), Catalan Institute of Oncology (ICO-IDIBELL), L'Hospitalet del Llobregat, Barcelona, Spain
2 Translational Research Laboratory, Catalan Institute of Oncology (ICO-IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain
3 Biomarkers and Susceptibility Unit, Catalan Institute of Oncology (ICO-IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain
4 CIBER de Epidemiologia y Salud Pública (CIBERESP), Instituto de Salud Carlos III, Spain.
5 Cancer Epigenetics Laboratory, Instituto Universitario de Oncología del Principado de Asturias (IUOPA), HUCA, Universidad de Oviedo, Oviedo, Spain
6 Department of General and Digestive Surgery, Colorectal Unit, University Hospital of Bellvitge (HUB-IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain
7 School of Medicine, University of Barcelona, Barcelona, Spain.
8 Department of Gastroenterology, University Hospital of Bellvitge (HUB-IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain
9 Department of Immunology and Oncology, Centro Nacional de Biotecnología/CNB-CSIC, Cantoblanco, Madrid, Spain
10 Hereditary Cancer Program, Catalan Institute of Oncology (ICO-IDIBELL), L’Hospitalet de Llobregat, Barcelona, Spain
11 Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

Running title: DNA methylation in colorectal cancer diagnosis.

Keywords: colorectal cancer, DNA methylation, pyrosequencing, biomarkers, screening

Correspondence to: Victor Moreno. Biomarkers and Susceptibility Unit, Catalan Institute of Oncology (ICO-IDIBELL). Av. Gran Vía 199-203, Hospitalet del Llobregat 08908, Barcelona, Spain. Tel: 0034-932607434 Fax: 0034-932607188. E-mail: v.moreno@iconcologia.net

Disclaimers: The authors disclose no conflicts.

Word count: 4.251
Number of Tables 3
Number of Figures 2
ABSTRACT

DNA methylation biomarkers for non-invasive diagnosis of colorectal cancer (CRC) and precursor lesions have been extensively studied. Different panels have been reported attempting to improve current protocols in clinical practice, although no definite biomarkers have been established.

In the present study we have examined patient biopsies starting from a comprehensive analysis of DNA methylation differences between paired normal and tumor samples in known cancer-related genes aiming to select the best performing candidates informative for CRC diagnosis in stool samples.

Five selected markers were considered for subsequent analyses in independent biological cohorts and in silico datasets. Among the five selected genes, three of them (AGTR1, WNT2, and SLIT2) were validated in stool DNA of affected patients with a detection sensitivity of 78% (95% CI: 56% - 89%). As a reference, DNA methylation of VIM and SEPT9 was evaluated in a subset of stool samples yielding sensitivities of 55% and 20% respectively. Moreover, our panel may complement histological and endoscopic diagnosis of Inflammatory Bowel Disease (IBD)-associated neoplasia, since it was also efficient detecting aberrant DNA methylation in non-neoplastic tissue samples from affected patients.

This novel panel of specific methylation markers can be useful for early diagnosis of CRC using stool DNA, and may help in then follow-up of high risk IBD patients.
INTRODUCTION

Colorectal cancer (CRC) is a leading cause of cancer-related death in developed countries (1). Several screening options are available for the early detection, being fecal occult blood test (FOBT) a cost-effective procedure to reduce cancer-related mortality when used programmatically (2). However, this test still needs substantial improvement to detect early-stage lesions. Population screening based on flexible sigmoidoscopy or colonoscopy are more sensitive alternatives (3-5), yet the invasive nature of these procedures and the high cost associated have prompted the development of alternative non-invasive screening methods, such as stool- and serum-based screening tests.

Sporadic CRC arises as a consequence of the accumulation of genetic and epigenetic alterations that transform colonic epithelial cells into colon adenocarcinoma cells (6). Stool DNA testing has emerged as a biologically rational and user-friendly strategy for the non-invasive detection of both CRC and critical precursor lesions (7-9). Stool DNA testing offers a biologically rational approach based on the fact that neoplastic cells are continuously shed into the colonic lumen and mixed with stool. This method is non-invasive, requires no diet or medication restriction, and has potential for high accuracy of neoplasm detection.

Hypermethylation of an increasing number of genes has been associated with human colorectal tumorigenesis (10, 11) and the detection of CpG island methylation in human DNA isolated from stool has been proposed as a promising approach for non-invasive screening and early diagnosis of colorectal cancer.
neoplasia (7, 12-14). Aberrantly methylated genes are attractive tumor markers because of their relatively high frequency in neoplasms (15, 16). Ideally, PCR-based methods for detecting methylated markers in bodily fluids would be able to detect minute amounts of methylated gene copies derived from early altered cells against an essentially unmethylated background from normal tissue.

The selection of the correct genes to analyze is crucial to improve the sensitivity and specificity of methylation DNA tests. Studies carried so far attempting to improve the diagnostic accuracy for tumor detection have mainly been based on candidate-gene approaches. Current microarray technology like the GoldenGate Methylation BeadArray (Illumina, Inc.) provides the opportunity for high-throughput unbiased methylation analysis of a large number of CpG sites (17).

The aim of the present study was to identify a set of DNA methylation markers useful for early diagnosis of CRC starting from a comprehensive analysis of the DNA methylation profile differences between tumors and paired normal mucosa in known cancer-related genes. The best predictor was validated in an independent set of samples, prior to the assessment of its performance in stool DNA. For comparisons purposes, two previously described and promising methylation-based markers were evaluated in a subset of the same samples: VIM (15, 18), and SEPT9 (19, 20, 21). Additionally, we implemented our panel to evaluate their performance as markers for CRC in non-neoplastic mucosa from inflammatory bowel disease (IBD) patients.
MATERIAL AND METHODS

Patients and samples

All patients gave informed written consent to participate and to have their biological specimens analyzed. The study was cleared by the Ethical Committee of the University Hospital of Bellvitge, Barcelona, Spain.

For the discovery phase we assessed DNA methylation levels using the Illumina® GoldenGate Methylation Cancer Panel I microarray in 92 sporadic CRC patients randomly selected from participants in a previous case-control study in our Institution (22). The original study included consecutive new diagnosed and pathology-verified sporadic CRC cases that had undergone surgical removal of the tumor. Samples from resected tumors and paired normal adjacent mucosa (> 10 cm away from tumor) were collected and immediately frozen at -80ºC until they were processed. After microarray pre-processing and quality control, 88 patients remained suitable for further analysis (Table 1, see Supplementary Methods for details of microarray pre-processing). These data were submitted in NCBI's Gene Expression Omnibus (23, 24) and are accessible through GEO series accession number GSE43369.

An in silico validation was performed for the selected probes using public datasets GSE17648 (25) and GSE29490 (26) downloaded from Gene Expression Omnibus (GEO) genomics data repository. These datasets provided methylation from sporadic CRC tissue samples and paired adjacent normal mucosa (22 and 24 individuals, respectively), hybridized in the Illumina Infinium HumanMethylation27 BeadChip array (27). Pre-processed series matrixes
originally provided by the authors were used for analysis. Given that this platform includes different CpG loci than those used in our study, the nearest loci to a given marker was consider as a good proxy for validation purposes.

The biological validation of the selected methylation biomarkers was performed by pyrosequencing in three independent groups of samples: 1) 62 tumors, 32 normal-appearing adjacent mucosa and 87 stool samples were tested from 126 sporadic CRC patients. Paired tumor and normal mucosa was available for 26 patients and paired tumor and stool for 27 patients. Twelve patients contributed to the study with the three types of samples. Tumor and normal mucosa were obtained at surgery and immediately stored at -80ºC until use; 2) 25 patients with IBD-associated neoplasia (17 cancers and 8 mild dysplasia) stored as formalin-fixed paraffin-embedded blocks and obtained from surgical resections (partial or complete colectomy). Tumor, adjacent normal-appearing mucosa samples, and, if available, actively inflamed mucosa, were collected; 3) colonic biopsy specimens from 39 non-IBD patients without abnormalities at colonoscopy and histories of gastrointestinal disease served as controls. Methylation levels of VIM and SEPT9 were evaluated in a subset of these samples comprising 26 mucosa from healthy individuals, and 14 neoplastic tissue and 35 stool samples from sporadic CRC patients.

The main clinicopathological characteristics of the patients are described in Table 1. Diagnoses were verified after examination of sections stained with hematoxylin and eosin. Carcinomas were classified according to the Union for International Cancer Control (UICC) TNM classification system.
DNA extraction and Bisulfite Treatment

DNA was extracted from paraffin embedded samples using the QIAamp DNA Mini Kit® (Qiagen, Valencia, CA) according to the manufacturer’s instructions. DNA from fresh-frozen tissues and stools was extracted using fenol-clorophorm (28). For stool samples, a final purification step was added using Wizard DNA Clean-Up System® (Promega) to remove PCR inhibitors. In all cases 500 ng of DNA was chemically modified to convert all unmethylated cytosine to uracil by the EZ DNA Methylation-Gold Kit® (Zymo Research, Orange, CA) according to the manufacturer’s protocol.

Microarray DNA Methylation assay

Illumina GoldenGate Methylation Cancer Panel I was developed to assay 1,505 CpG sites selected from 807 genes, including oncogenes and tumor suppressor genes, previously reported differentially methylated or differentially expressed genes, imprinted genes, genes involved in various signaling pathways and those responsible for DNA repair, cell cycle control, metastasis, cell migration and invasion, differentiation and apoptosis. After bisulfite treatment of DNA, the GoldenGate genotyping assay protocol was followed using Illumina-supplied reagents and conditions (29).

The methylation status at a CpG locus was estimated by Beta values computed as the ratio between the fluorescent signal from the methylated allele and the sum of the fluorescent signals from both methylated and unmethylated alleles. Beta values provide a continuous measure of DNA methylation and are interpreted as the proportion of methylated cells at the locus being interrogated;
zero indicates unmethylation while one represents a totally methylated state. Beta values are averaged over 30 perfect replicates to reduce experimental variability (see Supplementary Methods for details of microarray pre-processing).

**Pyrosequencing assay**

Pyrosequencing assays were designed to analyze and validate the results obtained from the array. Primers were designed for 28 CpGs located in the promoter region of the selected candidate genes. Additional primers were designed to interrogate four and five CpG sites within the promoter region of VIM and SEPT9 respectively. Bisulfite-treated DNA was used as a template for a PCR. The primers for PCR amplification and sequencing were designed with PyroMark assay design software version 2.0.01.15. Primer sequences (Supplementary Table S1) were designed, when possible, to hybridize with CpG-free sites to ensure methylation-independent amplification. PCR was performed with primers biotinylated to convert the PCR product to single-stranded DNA templates. We used the Vacuum Prep Tool (Biotage) to prepare single-stranded PCR products according to the manufacturer’s instructions. Pyrosequencing reactions and quantification of methylation were performed in a PyroMark Q24 System version 2.0.6 (QIAGEN). In order to rule out the presence of technical bias in the quantification of DNA methylation values, internal sequence-specific and bisulfite-conversion controls were considered in the interpretation of the results. Representative pyrograms for the markers on a healthy colonic mucosa are represented on Supplementary Figure S1.
Statistical analysis

For exploratory purposes and in order to identify global methylation patterns, array Beta values were analyzed using a Principal Component Analysis (PCA) and a hierarchical clustering with euclidean distance and Ward agglomeration method. Clustering analysis was also performed on paired differences Tumor-Normal within samples. Association between clusters and demographic and clinical covariates was assessed using an exact Fisher's test.

To maximize the likelihood of finding useful markers for stool testing, candidates in the discovery phase were required to be hypermethylated in tumor and to show methylation mean Beta values difference larger than 0.2. In order to further increase their potential sensitivity, a special attention was paid to loci that were consistently unmethylated in adjacent normal mucosa (low mean and low variability, see Supplementary Methods). Candidates were prioritized according to their p-value derived from a t-test for paired data. To assess the discriminant value of multiple loci, simple unweighted mean Beta values were used as summary. The area under the ROC curve (AUC) and sensitivity at 80% specificity were used to assess the discriminant ability.

For pyrosequencing data, the 99th percentile was computed on all methylation values obtained in healthy mucosa for all CpG sites being interrogated. This threshold value was found at 18% methylation and was used as decision cutoff for all markers. A CpG was considered methylated when its value exceeded this threshold. At gene level, a gene was considered methylated when at least one of its CpG sites was methylated. When multiple genes were jointly tested, a
sample was predicted as positive (altered) when at least one gene was methylated. Statistical association was assessed by Fisher’s exact test applied to the contingency table of predicted and actual outcome. In each case, sensitivity was computed in order to measure the accuracy of the markers for CRC detection. An estimation of specificity was obtained using the colonic mucosa from the healthy patients. In this analysis, only samples with complete data and those showing at least one methylated marker were included (see Supplementary Methods).

For each of accuracy measures used, 95% confidence intervals were calculated using the Bias Corrected and Accelerated bootstrap (BCa) method with 1,000 re-samples and outcome stratification. Empirical influence values were estimated by the usual jackknife method. These calculations were performed using packages boot (30) and ROCR (31) within the R language and environment for statistical computing (32).
RESULTS

Methylation patterns in tumor and normal colorectal mucosa

Clustering results on paired differences of our sample set showed large methylation changes between tumors and normal mucosa for the CpGs analyzed in the GodenGate array (Supplementary Figure S2). Hypermethylation in tumor was predominant in most of the patients being analyzed. The first principal component derived from a PCA on the array Beta values clearly discriminated between normal and tumor samples (37% variability explained, Supplementary Figure S3). A hierarchical cluster analysis of samples identified two main groups, which generally separated normal and tumors, with the exception of 12 misclassified tumor samples (Supplementary Figure S4). These results indicated a large degree of methylation differences between tumor and normal mucosa that was captured easily by these unsupervised classification techniques. Clustering results of CpG loci in the vertical axis also showed that methylation changes are typically consistent among samples for large sets of genes. No association was found among normal or tumor samples between age, sex, or tumor location and the main clusters derived from this analysis.

Selection of candidate methylation biomarkers

An unbiased selection of candidate-genes resulted from the analysis of the methylation data provided by the methylation array (Supplementary Table S2). Out of the 873 loci being analyzed, 147 were found to be highly hypermethylated in tumor samples according to the required criteria (difference > 0.20 in mean Beta values), being thereby informative indicators of aberrant methylation associated to tumorigenesis. The differences ranged from 0.20 to
0.65 and were highly significant as highlighted by their paired t-test (p-values <4.1e-10, Supplementary Table S3). Out of them, 33 CpGs were identified as consistently unmethylated in most of the normal samples as described in the methods section, a trait which could potentially increase sensitivity if included in a cancer panel for population screening (Supplementary Table S4 and Supplementary Methods). Candidates for subsequent validation were selected from these two ranked lists of markers and the selection process took into account statistical criteria (t-test p-value) and technical criteria as primer design for validation. The final selection included the two most significant probes from the 147 markers list, which were located in genes EYA4 and SLIT2, and three probes from the 33 markers list located in genes AGTR1, WNT2 and EPHA7 which ranked in positions 8, 11 and 25 respectively (Supplementary Table S2).

Figure 1 gives a first insight into the promising discrimination potential contained in the selected loci. When evaluated formally, all markers showed AUCs that ranged between 0.85 (EPHA7) and 0.97 (EYA4). Sensitivity at 80% specificity for tumor discrimination was at least 78% (EPHA7) for all loci and reached their maximum at 96% (EYA4 and SLIT2). A combination of all markers using their mean as summary achieved an AUC of 0.98 and nearly 98% sensitivity at 80% specificity (Table 2).

The GoldenGate array included four additional loci in genes SLIT2, EPHA7 and two probes in EYA4. These loci were characterized and showed a similar methylation pattern to the loci that prompted the selection of the genes (Supplementary Figure S5).
In silico validation of selected genes

Selected markers were externally validated in public independent datasets. Evaluation at the locus level resulted in AUCs which ranged from 0.71 to 1.00, while loci combined together achieved a perfect classification of all samples in GSE17648 and an excellent performance in GSE29490 (AUCs 1.00 and 0.96, respectively; Table 2). These datasets used the Infinium HumanMethylation27 array, that has good overlap with the GoldenGate cancer panel I but not perfect, adding extra value to the markers identified in our analysis. Distances from the original marker to the nearest Infinium loci in which markers were evaluated ranged from 14 (EYA4) to 316 (EPHA7) pairs of bases (Reference Sequence 36.1), and methylation values displayed similar methylation patterns to those observed in our platform (Supplementary Figure S6).

Validation of selected genes by pyrosequencing

To further test the prediction ability of the panel, we conducted a validation step in an independent cohort using a different technique, aiming to confirm the differential methylation of the biomarkers in tumorigenesis. Primers were designed for 28 CpGs located in the promoter region of the five selected genes, and their methylation level was quantitatively estimated in 267 normal mucosa, primary tumor and stool samples from a total of 190 individuals. As expected, in normal biopsies of healthy controls (n=39) gene promoters showed methylation in a minority of cases (0 to 7% of samples). An increased methylation was observed in cancer-associated mucosa, primary tumors, stool and inflammatory bowel disease samples.
Furthermore and for comparisons purposes, methylation levels of *VIM* and *SEPT9* were also estimated in a subset of these samples. Four and five CpG sites were interrogated within the promoter region of *VIM* and *SEPT9* respectively, which showed a similar methylation pattern to that observed in the panel markers. Figure 2 shows the distribution of the methylation values for all the CpG sites evaluated in the pyrosequencing assay.

**Methylation status in stool DNA of sporadic CRC**

Aiming to define the minimal yet best-informative methylation signature, we used the three best-performing markers available for this analysis (*AGTR1*, *WNT2* and *SLIT2*). Pyrosequencing of stool DNA detected 17 of 24 (71%) of carcinomas that scored positive for at least one methylation marker in tissue. No stool methylation was detected in 7 of 24 tumors (29%) that were hypermethylated.

When the total stool DNA samples (n=87) was considered, the panel succeeded to detect 50 out the 64 samples with evaluable data, which resulted in sensitivity for colorectal cancer detection of 78% (95% CI: 56% - 89%). The relative contribution of every marker to the detection of carcinoma was: 21% (14 of 68) for *AGTR1*, 40% (21 of 52) for *WNT2*, 52% (37 of 71) for *SLIT2* (Table 3). Whenever methylation was detected in stool, it was also detected in the corresponding tumor tissue except in 2 cases. In tissue DNA, the sensitivity of the panel of 3 markers for sporadic CRC detection was 95% (95% CI: 73% - 100%) and the specificity was 89% (95% CI: 82% - 92%; Table 3).
As a reference, *VIM* and *SEPT9* showed similar performances when evaluated in a subset of these samples: *SEPT9* reached a sensitivity of 93% in tumor tissue (13 of 14; 95% CI: 50% - 100%) and 20% in stool samples (7 of 35; 95% CI: 6% - 31%); *VIM* obtained sensitivities of 83% in tumor samples (10 of 12; 95% CI: 33% - 92%) and 55% in stools (18 of 33; 95% CI: 33% - 70%). Estimations of specificity based on healthy mucosa were 100% for *SEPT9* and 86% for *VIM* (Table 3). When evaluated in this subset of samples, the 3-gene panel showed similar performance to that obtained in the full dataset: 74% sensitivity (20 of 27; 95% CI: 50% - 90%) and 88% specificity (22 of 25; 95% CI: 77% - 93%).

**Methylation status in IBD-associated neoplasia**

Methylation of the panel of genes analyzed was a common phenomenon observed in IBD-associated neoplasia (93%). *AGTR1*, *WNT2* and *SLIT2* markers were methylated between 38 and 86% for every single marker. The non-neoplastic mucosa of patients with IBD-CRC showed higher methylation levels in *AGTR1*, *WNT2* and *SLIT2* than colorectal mucosa of healthy controls (11% vs 82%, p=3.6E-07), indicating the existence of underlying molecular alterations in pre-altered mucosa as a prior step to malignant transformation. The proportion of samples methylated in non-neoplastic mucosa in IBD-associated neoplasia was more than three times the observed in sporadic cancer-adjacent mucosa (82 vs 22% respectively; Table 3).
DISCUSSION

Early detection of CRC is a major challenge to improve patient's survival and widen the window of therapeutic intervention. In this study we followed an unbiased approach for candidate selection starting from a large candidate gene set and performed several validation steps to define a consistent panel of DNA methylation biomarkers useful for colorectal cancer early detection. The analysis of this panel in stool samples identified the presence of colorectal carcinomas with 78% sensitivity. An estimate of specificity was 89%, derived from the analysis of normal tissue, because no stool samples from healthy subjects were available. These estimates are encouraging about the potential utility of these markers in clinical practice.

Several studies have focused in the detection of aberrant DNA methylation of specific loci in body fluids that can be informative of tumor initiation and progression and easily performed by friendly and inexpensive PCR-based methods. A good example is the aberrant methylation of SEPT9 in colorectal cancer, that was first reported by Lofton-Day et al. in 2008 (19), and later identified as an informative screening marker in blood samples from affected patients (20, 21), allowing the detection of CRC at diverse stages and colonic locations. Another remarkable example is VIM (15, 18) which is the cornerstone of ColoSure™, the only currently commercially or clinically available fecal DNA test marketed for CRC screening in the U.S (33). Other studies have attempted to define comprehensive panels of DNA methylation biomarkers to test in stool samples, being frequently based on candidate-gene approaches investigating frequent targets of aberrant hypermethylation in cancer (34). A compiler meta-
analysis evaluating several studies reporting aberrantly methylated genes as biomarkers for colorectal cancer diagnosis was reporting an overall sensitivity and specificity of 0.54 and 0.88, respectively. The conclusion derived from this study inferred that panels reported up to date did not provide additional accuracy for the detection of colorectal neoplasia and underscored the need for additional biomarkers with better performance (35).

The panel of biomarkers identified in the present study equals or improves the detection power of previously reported genes, including PHACTR3 (55-66%) (36) and VIM. About the later, several works have been published reporting quite diverse sensitivity estimations of VIM for CRC detection in stool samples, ranging from 38% to 81% (reviewed in 33). In the present work, VIM showed an intermediate value in relation to this range which represented, however, the highest sensitivity obtained among all the markers when evaluated individually (55%). Specificity, though, was lower than those observed in the rest of markers (86%), which could be related to the relatively higher methylation levels observed for VIM in healthy mucosa. The better performance achieved by the 3-gene panel underscores the advantages of using a combination of markers identified from an agnostic procedure in front of the candidate-gene approach.

This panel also increases the sensitivity provided by the mSEPT9 assay in plasma, which is capable of identifying around 70% of CRC patients at high specificity (21). Furthermore, we aimed to evaluate the diagnostic power of SEPT9 in stool-based CRC diagnosis, since it had never been evaluated in this
particular scenario. The sensitivity of this marker alone in stool-DNA, however, was not as high as that reported in plasma (20%).

Our panel has similar diagnostic characteristics for colorectal cancer than others previously published, but we have proven that it may help in the early identification of colorectal dysplasia or cancer in high risk IBD patients. Patients with IBD, both ulcerative colitis (UC) and Crohn’s disease (CD) of the colon are at an increased risk of developing CRC (37-39). Disease duration, extent and severity of the inflammation are associated with an increased risk of CRC. Therefore, biomarkers that anticipate the development of full-blown tumors are needed to improve the follow-up and prognosis for these patients.

In sporadic CRC the precursor lesion is the adenoma that is typically removed by simple endoscopic polypectomy. In contrast, dysplasia in patients with IBD can be polypoid or flat, localized, diffuse, or multifocal invisible for the endoscopist and difficult to detect for the pathologist, especially in areas of active inflammation (40, 41). In UC, promoter methylation seems to precede dysplasia and occurs throughout the mucosa of colitis (42-44) reinforcing the link between chronic inflammation and DNA methylation (45, 46) and could be a potential biomarker for testing to detect pre-cancerous changes in routine surveillance biopsies. In line with these observations we have been able to show that our panel of markers is hypermethylated in non-cancerous mucosa of patients with IBD-associated neoplasia being more than three times of that observed in sporadic cancer-adjacent mucosa. Surveillance colonoscopy (including careful inspection of the mucosa and multiple random biopsy
sampling) in intervals of 1 to 2 years is recommended for these patients and histological assessment of dysplasia in mucosal biopsy specimens is considered the best marker of CRC risk in IBD patients. However, its interpretation is subject to a high level of inter-observer variability. Evaluation of our panel in these non-neoplastic biopsy specimens may contribute in a better follow up and early diagnosis of neoplasia in these high risk patients.

Among the five genes in our panel, abnormal methylation has been previously tested for \textit{SLIT2} in the detection of neoplasia in high-risk ulcerative colitis and Crohn's colitis patients by our group (47). \textit{EYA4} has been proposed as biomarker for pancreatic cancer (48). The other three genes have been previously investigated in relation to several types of cancer, but not in this particular scenario of diagnosis biomarker search.

One limitation of this study is that the series of patients analyzed did not include colorectal adenomas. We acknowledge that it is very important to analyze this group of patients in order to use the panel of biomarkers for the early detection of colorectal neoplasia. Also in our study, solid stools samples were not subject to a strict protocol regarding immediate storage temperature and collection buffer which may have affected the sensitivity of our approach. In spite of that, our data provides evidence showing that the analysis of methylation in stool DNA may be useful for the diagnosis of colorectal carcinomas.

This panel of biomarkers should be tested in other studies, and extended to larger cohorts of patients to confirm the robustness of the findings. In addition,
given the availability of platforms able to investigate DNA methylation in a genome-wide manner, a similar approach to the one followed here should be undertaken. The consecution of similar studies with increasing resolution and well-annotated clinical specimens would allow to define an optimal DNA methylation signature that refines the early detection of colorectal cancer, improving the management of the disease in the clinic and increasing the survival of the patients.

In conclusion, we provide a novel panel of specific methylation markers that can be assessed in stools and may complement currently applied protocols for the early detection of sporadic CRC. Moreover, our panel may contribute to improve the follow up and early diagnosis of high risk IBD patients when assessed in non-neoplastic tissues obtained by surveillance colonoscopy.
ACKNOWLEDGEMENTS

We thank Cristina Mangas and Rocio G Urdinguio for technical support and helpful discussion.

GRANT SUPPORT

This study was supported by the Catalan Institute of Oncology, the Private Foundation of the Biomedical Research Institute of Bellvitge (IDIBELL), the Instituto de Salud Carlos III grants (FIS PI08-1635, PI08-1359, PI09-01037, PI11-01439, PI12-01420), CIBERESP CB07/02/2005, the Spanish Association Against Cancer (AECC) Scientific Foundation, the Catalan Government DURSI grant 2009SGR1489, and the European Commission FP7-COOP-Health-2007-B HiPerDART.
REFERENCES


## Table 1. Characteristics of patients.

<table>
<thead>
<tr>
<th></th>
<th><strong>Discovery Phase</strong></th>
<th><strong>Validation Phase</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sporadic CRC (N=88)</td>
<td>Healthy Controls (N=39)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>51 (58%)</td>
<td>20 (51%)</td>
</tr>
<tr>
<td>Female</td>
<td>37 (42%)</td>
<td>19 (41%)</td>
</tr>
<tr>
<td><strong>Mean age (yrs) ± Sd</strong></td>
<td>63.2 ± 10.3</td>
<td>58 ± 7</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distal</td>
<td>71 (81%)</td>
<td>84 (67%)</td>
</tr>
<tr>
<td>Proximal</td>
<td>16 (10%)</td>
<td>42 (33%)</td>
</tr>
<tr>
<td>Missing data</td>
<td>1 (1%)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-II</td>
<td>23 (26%)</td>
<td>65 (52%)</td>
</tr>
<tr>
<td>III-IV</td>
<td>75 (85%)</td>
<td>61 (48%)</td>
</tr>
<tr>
<td><strong>Extent of IBD</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extensive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left-sided</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing data</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Duration of IBD (yrs)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Histologic grade</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGD</td>
<td>0 (0%)</td>
<td></td>
</tr>
<tr>
<td>LGD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Histologic grade.
Table 2. Prediction results in microarray data.

<table>
<thead>
<tr>
<th>Discovery phase</th>
<th>In silico validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GoldenGate data</td>
</tr>
<tr>
<td></td>
<td>AUC</td>
</tr>
<tr>
<td>AGTR1</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>89</td>
</tr>
<tr>
<td>EPHA7</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>78</td>
</tr>
<tr>
<td>WNT2</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>80</td>
</tr>
<tr>
<td>EYA4</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>96</td>
</tr>
<tr>
<td>SLIT2</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>96</td>
</tr>
<tr>
<td>All loci</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>98</td>
</tr>
</tbody>
</table>
### Table 3. Prediction results in pyrosequencing data.

<table>
<thead>
<tr>
<th>PANEL MARKERS</th>
<th>Healthy controls (N = 39)</th>
<th>Sporadic CRC (N = 32)</th>
<th>Fisher's p-value</th>
<th>Sporadic CRC (N = 62)</th>
<th>Fisher's p-value</th>
<th>IBD - CRC (N = 27)</th>
<th>Fisher's p-value</th>
<th>IBD - CRC (N = 20)</th>
<th>Fisher's p-value</th>
<th>stool (N = 87)</th>
<th>Fisher's p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NON NEOPLASTIC TISSUE</td>
<td></td>
<td></td>
<td>NEOPLASTIC TISSUE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Healthy controls</td>
<td></td>
<td></td>
<td>Sporadic CRC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(N = 26)</td>
<td></td>
<td></td>
<td>(N = 14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>VIM</td>
<td>methylated - no. (%)</td>
<td>3/22 (14%)</td>
<td>10/12 (83%)</td>
<td>0.0001</td>
<td>18/33 (55%)</td>
<td>0.0040</td>
<td>15/33 (45%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>unmethylated - no. (%)</td>
<td>19/22 (86%)</td>
<td></td>
<td>2/12 (17%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SEPT9</td>
<td>methylated - no. (%)</td>
<td>0/26 (0%)</td>
<td>13/14 (93%)</td>
<td>1E-09</td>
<td>7/35 (20%)</td>
<td>0.0169</td>
<td>3/35 (80%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>unmethylated - no. (%)</td>
<td>26/26 (100%)</td>
<td></td>
<td>1/14 (7%)</td>
<td></td>
<td>19/35 (54%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|               | NEOPLASTIC TISSUE         |                       |                  | STOOL                |                  |                  |                  |                  |                  |                  |                  |
|               | Reference markers         |                       |                  | (N = 35)             |                  |                  |                  |                  |                  |                  |                  |
|               | Healthy controls          |                       |                  | Sporadic CRC          |                  |                  |                  |                  |                  |                  |                  |
|               | (N = 26)                  |                       |                  | (N = 14)              |                  |                  |                  |                  |                  |                  |                  |
|               | WNT2                      | methylated - no. (%)  | 1/39 (3%)        | 46/56 (82%)           | 0.0013           | 21/52 (40%)      | 0.005           | 31/52 (60%)       |                  |                  |                  |
|               | unmethylated - no. (%)    | 38/39 (97%)           |                  | 10/56 (18%)           |                  | 8/19 (42%)       |                  |                  |                  |                  |                  |
|               | SLIT2                     | methylated - no. (%)  | 2/37 (5%)        | 48/56 (86%)           | 0.0012           | 37/71 (52%)      | 0.005           | 34/71 (48%)       |                  |                  |                  |
|               | unmethylated - no. (%)    | 35/37 (95%)           |                  | 8/56 (14%)            |                  | 1/7 (14%)        |                  |                  |                  |                  |                  |
|               | >=1 methylated gene       | yes                    | 4/38 (11%)       | 57/60 (95%)           | 0.0001           | 50/64 (78%)      | 0.005           | 14/64 (22%)       |                  |                  |                  |
|               |                           | no                     | 34/38 (89%)       | 14/18 (22%)          | 0.2544           | 4/22 (18%)       |                  |                  |                  |                  |                  |

|               | REFERENCE MARKERS         |                       |                  | Sporadic CRC          |                  |                  |                  | Sporadic CRC          |                  |                  |                  |
|               | (N = 35 )                 |                       |                  | (N = 14 )             |                  |                  |                  | (N = 35)             |                  |                  |                  |
|               | VIM                       | methylated - no. (%)  | 3/22 (14%)       | 10/12 (83%)           | 0.0001           | 18/33 (55%)      |                  |                  |                  |                  |                  |
|               | unmethylated - no. (%)    | 19/22 (86%)           |                  | 2/12 (17%)            |                  | 15/33 (45%)      |                  |                  |                  |                  |                  |
|               | SEPT9                     | methylated - no. (%)  | 0/26 (0%)        | 13/14 (93%)           | 1E-09            | 7/35 (20%)       | 0.0169           | 3/35 (80%)        |                  |                  |                  |
|               | unmethylated - no. (%)    | 26/26 (100%)          |                  | 1/14 (7%)             |                  | 19/35 (54%)      |                  |                  |                  |                  |                  |
TABLE LEGENDS

Table 1. Characteristics of patients. HGD: High-Grade Dysplasia; LGD: Low-Grade Dysplasia. No significant differences were observed comparing IBD-associated neoplasias (Dysplasias and Cancers).

Table 2. Prediction results in microarray data: for each loci and their combination, results from Area Under the Roc Curve (AUC) and %sensitivity at 80% specificity (SENS80) computed in the microarray data used for markers selection (GoldenGate) and public datasets used for in silico validation. Brackets contain the corresponding intervals at 95% confidence.

Table 3. Prediction results in pyrosequencing data: performance of individual genes and the 3-gene panel using methylation estimates from pyrosequencing data.
FIGURE LEGENDS

Figure 1. Methylation values in microarray experiments: microarray Beta values for selected markers and their mean values in adjacent-normal (blue) and tumor (red) samples. Segments inside the scatter plots indicate median and percentiles 25 and 75. Dot lines join paired samples from the same patient. In the graphics, marker’s ability to discriminate between tissue types become apparent and it notably seems to increase when they are combined using a simple mean as summary.

Figure 2. Methylation values in pyrosequencing experiments: percentage of methylation in genes analyzed by pyrosequencing in healthy mucosa (HM), tumor (SN), normal-appearing adjacent mucosa (SNN) and stool (SF) from sporadic CRC patients, and tumor (IN) and morphologically normal mucosa (INN) from IBD associated CRC patients. Segments inside the scatter plots indicate median and percentiles 25 and 75. Methylation patterns are similar to those obtained in microarray analysis although stool and IBD differences are not as high as in tumor analysis.
Cancer Prevention Research

DNA methylation biomarkers for non-invasive diagnostic of Colorectal Cancer


Cancer Prev Res  Published OnlineFirst May 21, 2013.

Updated version  Access the most recent version of this article at:
doi:10.1158/1940-6207.CAPR-12-0501

Supplementary Material  Access the most recent supplemental material at:
http://cancerpreventionresearch.aacrjournals.org/content/suppl/2013/05/22/1940-6207.CAPR-12-0501.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.