2012

Genome-wide Peripheral Blood Leukocyte DNA Methylation Microarrays Identified a Single Association with Inflammatory Bowel Diseases

R. Alan Harris
Baylor College of Medicine, rharris1@bcm.edu

Dorottya Nagy-Szakal
Baylor College of Medicine

Natalia Pedersen
Herlev University Hospital, natalia.pedersen@zeniavej.dk

Antone Opekun
Baylor College of Medicine, aopekun@bcm.tmc.edu

Jiri Bronsky
Charles University and University Hospital Motol, bronsky@email.cz

See next page for additional authors

Authors

R. Alan Harris, Dorottya Nagy-Szakal, Natalia Pedersen, Antone Opekun, Jiri Bronsky, Pia Munkholm, Cathrine Jespersgaard, Paal Skytt Andersen, Bela Melegh, George Ferry, Tine Jess, and Richard Kellermayer

ORIGINAL ARTICLE

Genome-wide Peripheral Blood Leukocyte DNA Methylation Microarrays Identified a Single Association with Inflammatory Bowel Diseases

R. Alan Harris, PhD,[1]* Dorottya Nagy-Szakal, MD,[2,3]* Natalia Pedersen, MD,[4]* Antone Opekun, MS,[2,5] Jiri Bronsky, MD,[6] Pia Munkholm, MD, Dr. Med. Sci,[4] Cathrine Jespersgaard, PhD,[7] PaalSkytt Andersen, PhD,[8] Bela Melegh, MD, PhD,[9] George Ferry, MD,[2] Tine Jess, M.D, Dr. Med. Sci,[10] and Richard Kellermayer, MM, PhD[2,3]

Background: Crohn's disease (CD) and ulcerative colitis (UC) are common forms of inflammatory bowel disease (IBD). Monozygotic (MZ) twin discordance rates and epidemiologic data implicate that environmental changes and epigenetic factors may play a pathogenic role in IBD. DNA methylation (the methylation of cytosines within CpG dinucleotides) is an epigenetic modification, which can respond to environmental influences. We investigated whether DNA methylation might be connected with IBD in peripheral blood leukocyte (PBL) DNA by utilizing genome-wide microarrays.

Methods: Two different high-throughput microarray-based methods for genome-wide DNA methylation analysis were employed. First, DNA isolated from MZ twin pairs concordant (CD: 4; UC: 3) and discordant (CD: 4; UC: 7) for IBD was interrogated by a custom-made methylation-specific amplification microarray (MSAM). Second, the recently developed Illumina Infinium HumanMethylation450 BeadChip arrays were used on 48 samples of PBL DNA from discordant MZ twin pairs (CD: 3; UC: 3) and treatment-naive pediatric cases of IBD (CD: 14; UC: 8), as well as controls ($n = 14$). The microarrays were validated with bisulfite pyrosequencing.

Results: The MSAMs did not yield significant IBD associations. The Methylation BeadChip approach identified a single DNA methylation association of IBD at *TEPP* (testis, prostate and placenta-expressed protein) when DNA isolated selectively from peripheral blood mononuclear cells was analyzed (8.6% increase in methylation between CD and control, FDR = 0.0065).

Conclusions: Microarray interrogation of IBD-dependent DNA methylation from PBLs appears to have limited ability to detect significant disease associations. More detailed and/or selective approaches may be useful for the elucidation of connections between the DNA methylome and IBD in the future.

(Inflamm Bowel Dis 2012;18:2334–2341)

Key Words: inflammatory bowel disease, DNA methylation, peripheral blood, twin, TEPP

The inflammatory bowel diseases (IBDs), including Crohn's disease (CD) and ulcerative colitis (UC), are common human disorders with autoimmune characteristics that affect more than four million people worldwide, about a third of whom reside in the United States.[1] The disorders are associated with high morbidity, leading to a significant

healthcare burden.[2] The incidence of IBD peaks in young adulthood and its prevalence is on the rise, especially in the pediatric population.[3–6] Recent observations implicate that the increase in the number of pediatric patients with IBD is a consequence of the overall amplified incidence rather than a shift toward disease onset at younger age.[7] Dietary and environmental factors of the industrialized world have been proposed as etiologic culprits for these shifting demographics by imposing critical changes during prenatal and/or pediatric development in the key physiologic components of IBD pathogenesis.[8–11]

One group of molecular processes that has been shown to dynamically respond to environmental/nutritional influences is designated as epigenetic. Epigenetic mechanisms mediate mitotically heritable changes in gene expression that are not associated with DNA sequence variation.[12] The most stable epigenetic alteration is the methylation of cytosines at CpG dinucleotides (DNA methylation). This molecular modification at gene promoters generally correlates with transcriptional downregulation. DNA methylation is critical for development and differentiation.[13] It is catalyzed by DNA methyltransferases and is ultimately dependent on dietary substrates and cofactors.[14] Modification of the prenatal environment can shift levels of DNA methylation at certain genomic loci and can associate with permanent phenotype changes in insects[15] and mammals as well.[16] Therefore, it has been proposed that epigenetic changes, such as DNA methylation, may be an important factor in the developmental origins of common human diseases,[17,18] including IBD.[19] The high rate of discordance for IBD within monozygotic (MZ) twin pairs supports the possible role of epigenetic processes in their etiologies.[20,21] Discordant MZ twins have been recognized to provide an opportunity for unraveling the epigenetic aspects of disease etiology in autoimmune diseases, such as systemic lupus erythematosus (SLE).[22] Peripheral blood leukocyte (PBL) DNA samples from MZ twins discordant and concordant for SLE have been recently used successfully to identify DNA methylation associations of disease.[23] Similar high-throughput observations on PBL DNA methylation are lacking in IBD.

In this study we set out to identify PBL DNA methylation associations of IBD by two independent methods of genome-wide microarray interrogation. The study was based on samples collected through the Danish Twin Registry,[24,25] the gene bank of the University of Pecs, Hungary,[26] the pediatric gastroenterology tissue bank of Charles University, Prague, Czech Republic, and tissue bank of the Pediatric Inflammatory Bowel Disease Consortium Registry of the Baylor College of Medicine, Houston, Texas.

MATERIALS AND METHODS

PBL DNA Samples

PBL DNA was isolated by different methodologies depending on the tissue bank protocols at each site. The prediction of this work was that uniform DNA methylation changes associate with IBD, affecting each human cell lineage equally. Therefore, we did not control for DNA isolation methodology during our initial approach (see Results section on TEPP).

Deidentified PBL DNA samples were obtained through the Danish Twin Registry,[24,25] the gene bank of the University of Pecs, Hungary,[26] the pediatric gastroenterology tissue bank of Charles University, Prague, Czech Republic, and the Pediatric Inflammatory Bowel Disease Consortium Registry of the Baylor College of Medicine, which were established in agreement with local and federal regulations. DNA isolation from the twin samples was performed with the Maxwell 16 system from Promega (Madison, WI) with the Blood DNA purification kit. In order to increase the concentration of the extracted DNA, Millipore microcon filters were used (http://www.millipore.com/userguides/tech1/99394). PBL DNA at the University of Pecs was extracted with standard salting-out methods. For the PBL DNA of the pediatric cohorts (used on the Illumina arrays, other than the twins) either Gentra Puregene (Qiagen, Valencia, CA) for whole blood DNA isolation was used according to manufacturer's recommendation on the samples from the Charles University, or the QuiAmp DNA mini kit (Qiagen) was utilized on isolated peripheral blood mononuclear cells (PBMCs) from the cohort of the Baylor College of Medicine. BD Vacutainer CPT Cell Preparation Tube with Sodium Citrate was used in the latter case to isolate PBMCs from PBL. Only age, gender, and disease status were collected with the samples (except for the pediatric cohort, where pharmacologic therapy was also recorded), the details of which are shown in Table 1. Adult controls were healthy, free from any chronic diseases. Pediatric controls included children who underwent colonoscopic evaluation for diagnoses of hematochezia, diarrhea, or abdominal pain, but whose endoscopy was grossly and histologically normal.

Methylation Specific Amplification Microarray (MSAM)

MSAM was carried out as previously described.[18,27] MSAM is based on serial digestion of the DNA with methylation-sensitive and insensitive restriction endonuclease isochizomers: *SmaI* and *XmaI*, respectively. The endonuclease digestion is followed by polymerase chain reaction (PCR)-mediated amplification and processing toward differentially labeled microarray hybridization (cohybridization) to compare the relative difference in DNA methylation between two samples at a large number of *SmaI/XmaI* genomic intervals. Eighteen independent MZ twin–twin cohybridizations (see Table 1) were performed. We designed a custom microarray (Agilent Technologies, Santa Clara, CA). *SmaI/XmaI* intervals between

TABLE 1. Summary of PBL DNA Samples

Cohorts	Diagnosis	Age (years)			Gender (M/F)
		Mean	Median	Min to Max	
Danish twin registry (*n*=number of twin pairs)	CD discordant (*n*=4)	44.75	43.5	43–49	2/2
	UC discordant (*n*=7)	45.71	49	33–56	2/5
	CD concordant (*n*=4)	46.5	46	44–50	2/2
	UC concordant (*n*=3)	41	43	36–44	0/3
Hungarian cohort	C-CD	36.85	37	24–52	10/10
(*n*=80)	C-UC	37.6	37.5	22–56	10/10
	CD	36.85	37	24–52	10/10
	UC	36.85	37	24–52	10/10
Illumina cohort	H	12.14	13.25	3.5–17.5	6/8
(*n*=36)	CD	14.43	15.5	9–17	8/6
	UC	12.88	13.5	9–17	5/3
BCM - Crohn's disease cohort (PBMC DNA)	H	11.94	13	0.1–17.5	6/15
(*n*=55)	CD new	14.50	15.05	9–18	13/11
	CD Rx	14.55	14.9	9–18	4/6

BCM, Baylor College of Medicine; PBMC, peripheral blood mononuclear cell. C, healthy control; H, control where no intestinal inflammation was detected.

100–2000 bp, along with negative control intervals in the range 10–15 kbp, were identified based on the human genome assembly (NCBI36/UCSC hg18) excluding chromosome X and Y. Intervals containing single nucleotide polymorphisms, based on dbSNP 129 (http://www.ncbi.nlm.nih.gov/projects/SNP/), which disrupted or introduced *SmaI/XmaI* cut sites were removed from consideration. The intervals were uploaded to Agilent eArray (https://earray.chem.agilent.com/earray/) in order to identify high-density probes contained within the intervals. eArray filtering for probes in known copy number variation regions was also applied. The final array design consisted of 99,071 probes interrogating 41,079 (100–2000 bp length) intervals and 100 probes interrogating 10–15 kbp negative controls in addition to the standard Agilent array controls. Therefore, the custom array covered 45.31% (41,079 out of 90,657 in the human genome) of the potentially informative 100–2000 bp *SmaI/XmaI* intervals. To exclude falsely significant findings arising from the massive numbers of loci interrogated, multiple testing correction of the *P*-values was performed using Benjamini and Hochberg's method[28] (false discovery rate: FDR). Based on our previous experience,[18] we first examined intergroup differences exceeding 1.8-fold with $P < 0.0002$ for significant association. A secondary criteria of significance was determined at FDR < 0.5 at intervals where intergroup differences at select *SmaI/XmaI* intervals were >1.2-fold.

Infinium Methylation Assay Microarrays

In all, 48 PBL DNA (3 MZ twin discordant for CD, 3 MZ twin discordant for UC; and pediatric samples: 14 control, 14 CD and 8 UC collected at the time of diagnostic endos-

copy in untreated patients) samples following quality control with PicoGreen (http://probes.invitrogen.com/media/pis/mp07581.pdf) were processed by Infinium HumanMethylation450 BeadChip Kits (Illumina San Diego, CA; http://www.illumina.com/products/methylation_450_beadchip_kits.ilmn) according to the manufacturer's recommendations through automated processes in the Core Laboratory for Translational Genomics of the Baylor College of Medicine. Arrays were imaged with BeadArray Reader using standard recommended Illumina scanner settings. GenomeStudio software v. 2010.3.0.30128 was used to generate beta values normalized to internal control probes. Internal controls determined the array processing to be of good quality. Only the 482,421 CpG probes on the array were used for subsequent analysis. The R LIMMA package[29] was employed to compare beta values between normal controls and CD or UC affected individuals. Limma fitted a linear model to beta values for each probe in the compared samples and then calculated *t*-statistic using an empirical Bayesian model that moderates the standard errors across probes. *P*-values were calculated from the moderated *t*-statistics and FDR was determined to identify differentially methylated probes. The criterion for significant association was set at FDR < 0.2. The raw data of the microarrays was uploaded to Gene Expression Omnibus (GEO; Series GSE32148) and is accessible at: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=pjszvekkmmaeyzu&acc=GSE32148.

Bisulfite Pyrosequencing

Analytical validation using the same cohort as the array studies was performed by bisulfite pyrosequencing at select

loci. This analysis was followed by validation of differential methylation status using expanded or independent cohorts (see PBL DNA Samples paragraph above). PBL DNA was bisulfite converted with EZ DNA Methylation-Gold Kit (D5006, Zymo Research, Orange, CA). DNA was amplified with traditional PCR methodology following bisulfite conversion and including a biotinylated primer in the sets (Supporting Table 1). A quantitative bisulfite pyrosequencing protocol was used for all methylation analyses with the utilization of the Pyro Q CpG program (Qiagen, Hilden, Germany). Methylation measurements at (*ANO1*-associated *SmaI/XmaI*: chr11:69597807-69598085, *EHD1*-associated *SmaI/XmaI*: chr11:64377161-64377539) or in the vicinity (*MGRN1*-associated *SmaI/XmaI*: chr16:4654296-4655675) of both of the *SmaI/XmaI* sites of the gene-associated intervals were performed in the case of the validation measurements of the MSAMs. An algorithm was used to calculate and compare methylation ratios at these intervals.[27] Illumina array validation was performed at three candidate loci: *FAM53B*-associated CpG site (chr10:126360669-126360669); *SLC6A9*-associated CpG site (chr1:44466601-44466601); *TEPP*-associated CpG site (chr16: 58019367-58020365). Only the results of assays that passed quality control were used. The criterion for significant association for bisulfite pyrosequencing results was $P < 0.05$ with nonparametric *U*-test. The primer sequences for the bisulfite pyrosequencing measurements are described in Supporting Table 1.

Statistical Analyses

Statistical analysis of the Illumina arrays is described above. Pearson's correlation analysis was used in the correlation studies. The nonparametric Mann–Whitney *U*-test was employed in the intergroup comparisons. Significance was set at $P < 0.05$.

RESULTS

MSAM in Discordant MZ Twins Found No Significant IBD Association

None of the discordant MZ twin comparisons yielded average disease to healthy ratio comparisons that met our criteria of significance (see Materials and Methods). Nevertheless, we wished to test whether relaxing the selection criteria may still provide significant findings. The selection criteria were relaxed by different approaches and top candidates were selected for bisulfite pyrosequencing. *MRGN1* showed over 1.4 average fold difference between disease and healthy ($P < 0.05$); *ANO1* showed two out of four comparisons exceeding 1.95-fold difference between disease and healthy; and *EHD1* showed three out of four comparisons exceeding 1.5-fold difference between disease and healthy. Pyrosequencing failed to confirm the significant DNA methylation differences between MZ twins discordant for IBD indicated by the arrays (correlation at *MGRN*: $r = 0.28$, $P = 0.72$, at *ANO1*: $r = -0.68$, $P = 0.32$; and at
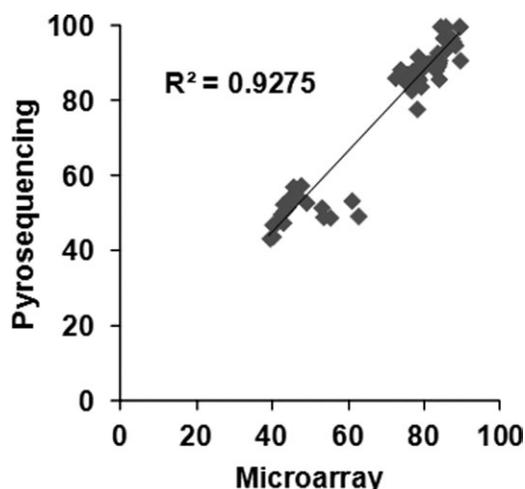

FIGURE 1. Scatterplot of methylation levels obtained by Infinium HumanMethylation450 Microarrays (Microarray) versus bisulfite pyrosequencing at the *SCL6A9*, *FAM53B*, and *TEPP*-associated CpG sites. Significant correlation was observed. $r = 0.963$; $P < <0.00001$.

EHD1: $r = 0.3$, $P = 0.7$, respectively). Therefore, MSAM did not yield significant results.

With the assumption that disease-specific epigenetic differences between MZ twins may be suppressed,[30] and the MSAM might have been more sensitive than our validation method, we measured the methylation of candidate loci in an independent cohort of adults with IBD and healthy controls. None of our measurements showed disease-specific DNA methylation in the PBL DNA samples from this independent cohort (see Supporting Fig. 1 as an example).

Infinium Methylation Assay Microarrays Did Not Show Significant IBD Association

We first examined the genetically unrelated pediatric samples on the Illumina arrays. After correcting for multiple testing, none of the CD to control comparisons was significant (FDR > 0.2 for all). Similar relaxation of the FDR was done with success when assessing PBL DNA methylation associations of SLE.[23]

We found a total of six CpG sites in the human genome, where UC to control comparisons resulted in over 4% mean methylation difference (this value approaches the reliability of the bisulfite pyrosequencing validation method[19]) with FDR < 0.2 (Table 2) in the pediatric groups of the Illumina arrays. Bisulfite pyrosequencing validation at select loci from this UC-specific group gave excellent correlation with the microarrays (Fig. 1). However, group comparisons between UC and control were not significant (Supporting Fig. 2). This latter finding is likely the result of the low (less than 15% difference in five out of the six candidates) intergroup methylation differences

TABLE 2. Genomic Loci

Coordinates	Associated Gene	C	UC	UC-C	FDR
chr13:20771319-20771319		56.19	27.33	−28.86	0.07
chr20:60915009-60915009	*LAMA5*	65.17	51.74	−13.43	0.16
chr6:31803832-31803832	*C6orf48,SNORD52*	41.81	46.46	4.65	0.08
chr2:34719565-34719565		68.06	73.05	4.99	0.19
chr1:44466601-44466601	*SLC6A9*	75.83	81.34	5.51	0.16
chr10:126360669-126360669	*FAM53B*	42.40	55.00	12.60	0.07

Coordinates of UC-associated CpG candidates where disease-specific DNA methylation occurred according to the Illumina arrays from the pediatric cohort (loci meeting criteria for >4% mean methylation difference between UC and control: C; and Benjamini-Hochberg FDR < 0.2). Numbers represent percent methylation. Gene associations are based on the Infinium HumanMethylation450 v1.1 manifest supplied by Illumina.

detected by the microarrays in addition to relatively large intragroup variation that was predicted by the overall high FDRs.

Selective PBMC DNA Analysis Identifies TEPP CpG Methylation Association with IBD

The DNA samples of this study originated from different tissue banks utilizing various extraction methods. To address the possibility that the sample variation originating from technical and geographic disparity masked disease-specific DNA methylation, we performed a selective bioinformatic analysis of the samples from the Baylor College of Medicine cohort. In this case, the DNA was extracted from PBMCs excluding neutrophil granulocytes with uniform methodology. This analysis showed a single CD-specific DNA methylation change associated with *TEPP* with FDR < 0.2 (chr16:58019866-58019866, CD-C average methylation difference: 8.6%; FDR = 0.0065). Bisulfite pyrosequencing validated the results (included in Fig. 1) and showed a significant ($P = 0.011$) CD-associated increase in methylation at this locus. However, this increase was not specific to CD, but was present in untreated UC patients as well ($P = 0.013$; Fig. 2A). When the control and treatment-naïve CD groups were expanded, CpG methylation at this locus showed an even more significant association with CD ($P < 0.0001$; Fig. 2B). In the meantime, the amplified DNA methylation in treatment-naïve CD was absent in treated children (Fig. 2C).

DISCUSSION

Identification of disease-specific epigenetic (including DNA methylation) changes is a very extensive area of current biomedical research. Epigenetic associations could support the unraveling of disease etiology, provide novel biomarkers, and aid the development of potential preventative and/or therapeutic measures for disorders such as IBD.[31,32] Based on MZ twin discordance rates, the relative genetic contribution to CD and UC is low,[33] supporting the importance of nongenetic (including epigenetic and/or other

molecular processes within the large biological components of gastrointestinal disease[10]) processes in their etiology, as discussed in the introduction.

Peripheral blood (PB) is the most commonly obtained and stored biological specimen and PBL DNA constitutes the vast majority of gene bank samples. Therefore, identification of disease-specific molecular characteristics from PB or PBL is highly feasible. Recent findings indicate that PB microRNAs (small RNA molecules modulating gene expression through epigenetic mechanisms) may serve as useful biomarkers of CD.[36]

In case a DNA methylation modification at a disease susceptibility locus occurs before gastrulation, it can be detected in all tissues of the body, as has been observed at metastable epialleles in mice.[16] Therefore, if such loci exist in humans for certain diseases, then the disease-specific DNA methylation change could be detected from any type of cell, including leukocytes obtained from PB specimens. Our approach to DNA methylation associations of IBD in PBL DNA was based on this prediction.

We first used the MSAM technique on MZ twin samples, but could not identify any significant IBD association. These results indicated that either the MZ twin-based approach or our MSAM technique is insufficient to identify IBD-specific DNA methylation from PBL DNA, or that uniform epigenetic changes may not exist in association with the disease group in peripheral white blood cells. One limitation of the MSAM technique is that it does not provide information on the actual levels of DNA methylation, just the relative difference in methylation between two different (disease-healthy, for example) samples at the same genomic region (*SmaI/XmaI* interval in this case). Therefore, at genomic loci where DNA methylation is low, even subtle changes may show up as highly significant (for example, 1% methylation in healthy individuals and 3% methylation in disease would result in a 3-fold difference on the array). Subtle methylation differences are of questionable physiologic relevance and are very difficult to validate with sequencing methodologies. Therefore, we turned
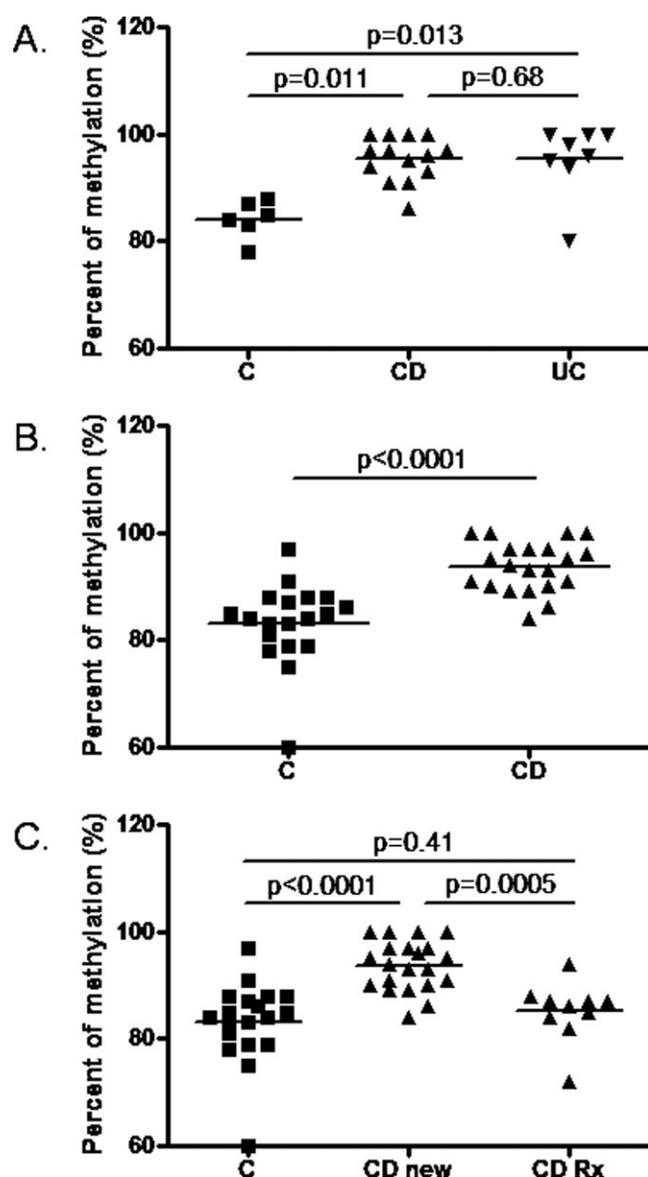
FIGURE 2. Methylation levels at the *TEPP*-associated CpG site in the select pediatric PBMC cohort of the study (see Table 2). (A) Bisulfite pyrosequencing results from the PBMC samples of the Infinium HumanMethylation450 Microarrays. The significant increase in methylation was confirmed by bisulfite pyrosequencing in the CD and UC cohorts compared to controls (*n* = 6–14). (B) Extension of the measurements to larger groups of PBMC DNA from controls and treatment-naïve CD patients further supported the validity of our findings (*n* = 19–22). (C) The increased level of methylation in treatment-naïve (CD) patients was lost when compared to CD patients receiving therapy (CD-rx; three biologic monotherapy; two biologic+immunomodulator, two mesalamine alone, one steroid alone, one immunomodulator alone, one steroid+immunomodulator) (*n* = 10–22).

to the Infinum methylation assay based Infinium Human-Methylation450 BeadChip Kits.

Secondary to our negative findings with the MSAM approach in the twin cohort, only a subset of MZ twins discordant for CD and UC were examined by the Illumina arrays. One reason for this approach was the prediction based on healthy twin observations[37] that a twin with a disease in discordant MZ pairs may be epigenetically less different from the healthy twin sibling at critical pathogenic loci, as compared to unrelated healthy individuals. However, at least according to one publication, SLE-associated DNA methylation was not influenced by relatedness (i.e., disease-specific DNA methylation difference was as high between unrelated patients and controls, as between affected twins and healthy counterparts in MZ twin pairs). Nevertheless, we decided to first study unrelated individuals with IBD where epigenetic differences could be higher than in MZ twins. Therefore, PBL DNA samples collected at diagnosis from untreated pediatric cases of IBD were tested and compared to controls. The untreated pediatric cases eliminated the possible confounding factor of therapeutic interventions inducing nonspecific DNA methylation changes.[38] However, because the patients were unrelated their comparisons carried the potential for genetic variation-induced DNA methylation differences being identified. Therefore, our plan was to overlap candidate loci determined in the pediatric cohort with loci identified through the discordant MZ twin comparisons.

We could not identify IBD-specific PBL DNA methylation association in unrelated children where even IBD-linked genetic variation induced DNA methylation changes could have been present. Consequently, we did not proceed with overlapping the pediatric cohort results with the independent adult MZ twin samples on the Illumina arrays.

Studies with high-throughput methodologies applied on clinical samples frequently bare limitations. Some of the confounding factors of our analyses may have been: the low power secondary to small sample sizes, the limited coverage and/or resolution of the microarrays employed, and the clinical heterogeneity of both the control and IBD groups (nonhealthy control children, healthy adults, twins, adult treated IBD patients likely with different disease subtypes) studied. However, as stated earlier, the original hypothesis of our work was that uniform DNA methylation changes may exist in humans, which predispose to the development of IBD independently from disease subtypes within CD and UC.

Our largely negative findings with MSAMs and Infinum Methylation Assays implicate that microarray profiling of DNA methylation from PBL is less likely to yield positive results in IBD. Subfractionation of PBL into lymphocyte subsets and performing microarray comparisons with such DNA samples may be valuable in the future since DNA methylation is cell type-specific. The one positive IBD association at *TEPP* when PBMCs were selectively analyzed would support this conclusion. Indeed,

interferon-gamma (*IFNG*) methylation in isolated peripheral T cells correlated with IBD, for example.[39] However, *IFNG* methylation difference between IBD and controls was less than 4% in this case, which would have been likely missed by the resolution of the currently available high-throughput interrogation methods.

TEPP encodes for testis, prostate, and placenta-expressed protein with uncertain function.[40] It is poorly expressed in bone marrow, whole blood, thymus, monocytes, and T-cell subsets, as well as in the appendix (expression atlas 2 data from U133A and GNF1H chips), implicating that the physiologic relevance of the small average DNA methylation increase (from 83% in controls to 94% in CD, Fig. 2A) detected from PBMCs is questionable. Furthermore, the CD dependent methylation increase was lost in treated patients. One explanation for this finding is that PBMC subset modifications (altered ratios of macrophages, T- and B-cell subsets, etc., with differing *TEPP* methylation levels) characteristic for IBD at onset are lost upon treatment. However, there are no flow cytometric data available for our patients, which could specifically address this prediction. Nevertheless, we can conclude that IBD-associated DNA methylation increase in PBMC DNA at *TEPP*, a gene poorly expressed in peripheral blood leukocytes, is small and is not disease-specific. Additionally, only treatment-naïve patients showed this difference. Therefore, the functional relevance and the potential diagnostic value of this finding appear limited.

Numerous observations on links between gene and cell type-specific DNA methylation and IBD have been made, where either the low fraction of the specific cell type (regulatory T cells, for example[41]) or the unique location of the DNA methylation correlate (such as in the interleukin 17A: *IL17A*[42]) provide limitations to detectability from PBL DNA by microarray methodology. Increasing the depth of interrogation with methods such as genomewide direct sequencing[43] may increase the chances of determining DNA methylation associations of IBD in PBL, and PBL subtypes. Unfortunately, the financial aspects of such interrogation in reliable sample sizes provide significant constraints to such studies currently.

Intestinal biopsies may offer more readily identifiable DNA methylation correlates of IBD through genome-wide microarray analysis. Both targeted gene methylation[44] and DNA methylation microarray[45] assessments have detected numerous colonic mucosal associates of inflammation in IBD. Techniques with higher resolution may offer even more insight into the intestinal mucosal DNA methylation aspects of the disease group.

In summary, our largely negative study can significantly guide epigenetic investigations in IBD regarding technical planning and tissue selection. Our findings on *TEPP*-associated DNA methylation may also prove of significance in the future.

REFERENCES

1. Loftus EV Jr. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterology*. 2004;126:1504–1517.
2. Park KT, Bass D. Inflammatory bowel disease-attributable costs and cost-effective strategies in the United States: a review. *Inflamm Bowel Dis*. 2011;17:1603–1609.
3. Benchimol EI, Guttmann A, Griffiths AM, et al. Increasing incidence of paediatric inflammatory bowel disease in Ontario, Canada: evidence from health administrative data. *Gut*. 2009;58:1490–1497.
4. Abramson O, Durant M, Mow W, et al. Incidence, prevalence, and time trends of pediatric inflammatory bowel disease in northern California, 1996 to 2006. *J Pediatr*. 2010;157:233–239.e1.
5. Gupta N, Bostrom AG, Kirschner BS, et al. Incidence of stricturing and penetrating complications of Crohn's disease diagnosed in pediatric patients. *Inflamm Bowel Dis*. 2010;16:638–644.
6. Kunde S, Prasad M, Kugathasan S. Rising incidence of inflammatory bowel disease in young children: what does the future hold? *J Pediatr Gastroenterol Nutr*. 2011;53:128.
7. Braegger CP, Ballabeni P, Rogler D, et al. Epidemiology of inflammatory bowel disease: is there a shift towards onset at a younger age? *J Pediatr Gastroenterol Nutr*. 2011;53:141–144.
8. Barnett M, Bermingham E, McNabb W, et al. Investigating micronutrients and epigenetic mechanisms in relation to inflammatory bowel disease. *Mutat Res*. 2010;690:71–80.
9. Hou JK, Abraham B, El-Serag H. Dietary intake and risk of developing inflammatory bowel disease: a systematic review of the literature. *Am J Gastroenterol*. 2010;106:563–573.
10. Kellermayer R. Genetic drift. "Omics" as the filtering gateway between environment and phenotype: The inflammatory bowel diseases example. *Am J Med Genet A*. 2010;152A:3022–3025.
11. Kugathasan S, Amre D. Inflammatory bowel disease—environmental modification and genetic determinants. *Pediatr Clin North Am*. 2006;53:727–749.
12. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*. 2003;33(Suppl):245–254.
13. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*. 1992;69:915–926.
14. Waterland RA, Michels KB. Epigenetic epidemiology of the developmental origins hypothesis. *Annu Rev Nutr*. 2007;27:363–388.
15. Lyko F, Foret S, Kucharski R, et al. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol*. 2010;8:e1000506.
16. Dolinoy DC, Das R, Weidman JR, et al. Metastable epialleles, imprinting, and the fetal origins of adult diseases. *Pediatr Res*. 2007;61:30R–37R.
17. Hochberg Z, Feil R, Constancia M, et al. Child health, developmental plasticity, and epigenetic programming. *Endocr Rev*. 2011;32:159–224.
18. Waterland RA, Kellermayer R, Laritsky E, et al. Season of conception in rural gambia affects DNA methylation at putative human metastable epialleles. *PLoS Genet*. 2010;6:e1001252.
19. Balasa A, Gathungu G, Kisfali P, et al. Assessment of DNA methylation at the interferon regulatory factor 5 (IRF5) promoter region in inflammatory bowel diseases. *Int J Colorectal Dis*. 2010;25:553–556.
20. Petronis A, Petroniene R. Epigenetics of inflammatory bowel disease. *Gut*. 2000;47:302–306.
21. Petronis A. Epigenetics and twins: three variations on the theme. *Trends Genet*. 2006;22:347–350.
22. Ballestar E. Epigenetic alterations in autoimmune rheumatic diseases. *Nat Rev Rheumatol*. 2011;7:263–271.

23. Javierre BM, Fernandez AF, Richter J, et al. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res.* 2010;20:170–179.

24. Jess T, Riis L, Jespersgaard C, et al. Disease concordance, zygosity, and NOD2/CARD15 status: follow-up of a population-based cohort of Danish twins with inflammatory bowel disease. *Am J Gastroenterol.* 2005;100:2486–2492.

25. Halfvarson J, Jess T, Bodin L, et al. Longitudinal concordance for clinical characteristics in a Swedish-Danish twin population with inflammatory bowel disease. *Inflamm Bowel Dis.* 2007;13:1536–1544.

26. Csongei V, Jaromi L, Safrany E, et al. Interaction of the major inflammatory bowel disease susceptibility alleles in Crohn's disease patients. *World J Gastroenterol.* 2010;16:176–183.

27. Waterland RA, Kellermayer R, Rached MT, et al. Epigenomic profiling indicates a role for DNA methylation in early postnatal liver development. *Hum Mol Genet.* 2009;18:3026–3038.

28. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B.* 1995;57:289–300.

29. Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Dudoit S, Irizarry R, et al., eds. Bioinformatics and Computational Biology Solutions using R and Bioconductor. 2005 ed.: New York: Springer; 2005:397–420.

30. Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature.* 2010;465:721–727.

31. de Zoeten EF, Wang L, Sai H, et al. Inhibition of HDAC9 increases T regulatory cell function and prevents colitis in mice. *Gastroenterology.* 2010;138:583–594.

32. Schaible TD, Harris RA, Dowd SE, et al. Maternal methyl-donor supplementation induces prolonged murine offspring colitis susceptibility in association with mucosal epigenetic and microbiomic changes. *Hum Mol Genet.* 2011;20:1687–1696.

33. Halfvarson J. Genetics in twins with Crohn's disease: less pronounced than previously believed? *Inflamm Bowel Dis.* 2011;17:6–12.

34. Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet.* 2010;42:1118–1125.

35. Anderson CA, Boucher G, Lees CW, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet.* 2011;43:246–252.

36. Zahm AM, Thayu M, Hand NJ, et al. Circulating MicroRNA is a biomarker of pediatric Crohn disease. *J Pediatr Gastroenterol Nutr.* 2011;53:26–33.

37. Kaminsky ZA, Tang T, Wang SC, et al. DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet.* 2009;41:240–245.

38. Yuan B, O'Connor TR, Wang Y. 6-Thioguanine and S-methylthioguanine are mutagenic in human cells. *ACS Chem Biol.* 2010;5:1021–1027.

39. Gonsky R, Deem RL, Landers CJ, et al. Distinct IFNG methylation in a subset of ulcerative colitis patients based on reactivity to microbial antigens. *Inflamm Bowel Dis.* 2011;17:171–178.

40. Bera TK, Hahn Y, Lee B, et al. TEPP, a new gene specifically expressed in testis, prostate, and placenta and well conserved in chordates. *Biochem Biophys Res Commun.* 2003;312:1209–1215.

41. Tao R, de Zoeten EF, Ozkaynak E, et al. Deacetylase inhibition promotes the generation and function of regulatory T cells. *Nat Med.* 2007;13:1299–1307.

42. Kim SW, Kim ES, Moon CM, et al. Genetic polymorphisms of IL-23R and IL-17A and novel insights into their associations with inflammatory bowel disease. *Gut.* 2011;60:1527–1536.

43. Flusberg BA, Webster DR, Lee JH, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods.* 2010;7:461–465.

44. Saito S, Kato J, Hiraoka S, et al. DNA methylation of colon mucosa in ulcerative colitis patients: correlation with inflammatory status. *Inflamm Bowel Dis.* 2011;17:1955–1965.

45. Lin Z, Hegarty JP, Cappel JA, et al. Identification of disease-associated DNA methylation in intestinal tissues from patients with inflammatory bowel disease. *Clin Genet.* 2011;80:59–67.
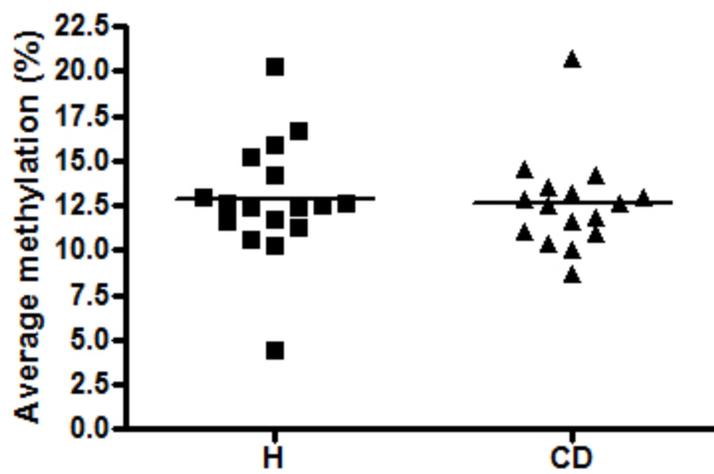
Supplementary Figure 1. Bisulfite pyrosequencing results on average methylation at 2 upstream (-52, and -58 bases) CpG sites from the 3' *SmaI/XmaI* CpG of the *MGRN1* interval in the Hungarian cohort of healthy (H) and Crohn's disease (CD) patients. p=0.84
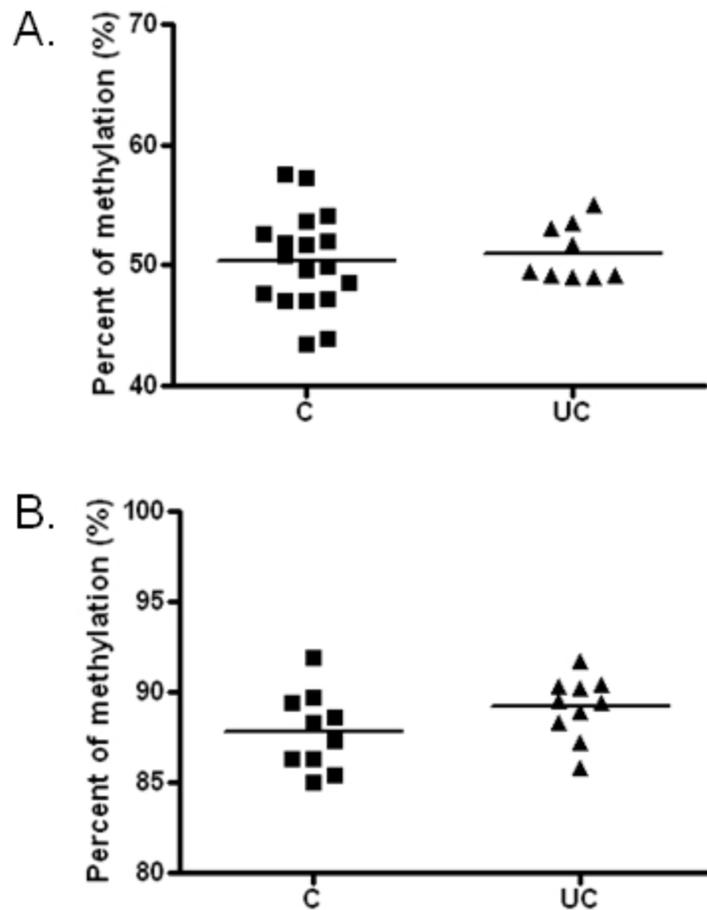
Supplementary Figure 2. Absence of significant DNA methylation difference between controls (C) and ulcerative colitis patients (UC) in the Illumina cohort of the study. A. DNA methylation results at the *FAM53B* associated CpG site (n=18-9; p=0.654). B. DNA methylation results at the *SCL6A9* associated CpG site (n=10-10; p=0.143)