Epigenetically associated colonic mucosal immune- and defense responses in treatment naïve pediatric ulcerative colitis

Short title: Epigenetics of untreated UC

R. Alan Harris[1#], Dorottya Nagy-Szakal[2#], Sabina A. V. Mir[2], Eibe Frank[3], Reka Szigeti[4], Jess L. Kaplan[5], Jiri Bronsky[6], Antone Opekun[7], George D. Ferry[2], Harland Winter[5], and Richard Kellermayer[2]*

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; [2]Department of Pediatrics, Baylor College of Medicine, USDA/ARS Children's Nutrition Research Center, Texas Children's Hospital, Houston, TX; [3]Department of Computer Science, University of Waikato, Hamilton, New Zealand; [4]Department of Pathology, Baylor College of Medicine; [5]Department of Pediatrics, MassGeneral Hospital *for* Children, Boston, MA; [6]Department of Pediatrics, Charles University and University Hospital Motol, Prague, Czech Republic; [7]Department of Gastroenterology, Baylor College of Medicine, Houston, TX; [#]contributed equally

[*]Correspondence:
Richard Kellermayer
Section of Pediatric Gastroenterology, Hepatology & Nutrition
Baylor College of Medicine
6621 Fannin St., CC1010.00
Houston, TX 77030-2399
Voice: 713-798-0319
Fax: 832-825-3633
Email: kellerma@bcm.edu

ABSTRACT

Inflammatory bowel diseases (IBD) are emerging globally, indicating that environmental factors may be important in their pathogenesis. Colonic mucosal epigenetic processes, such as DNA methylation, can respond to environment, and have been implicated in IBD pathology. However, mucosal DNA methylation has not been examined in treatment naïve patients. We studied DNA methylation from untreated, left sided colonic biopsy specimens by Infinium HumanMethylation450 BeadChip Kits in 22 control [C], 15 untreated Crohn's disease [CD], and 9 untreated ulcerative colitis [UC] patients from two cohorts. Samples at the time of clinical remission from two of the treatment naïve UC patients were also included into the analysis. UC specific gene expression was interrogated in a subset (5 C and 5 UC) of adjacent samples by Affymetrix GeneChip PrimeView Human Gene Expression Arrays. Only treatment naïve UC separated from control. One hundred and twenty genes with significant expression (>2 fold, $p<0.05$) change in UC associated with differentially methylated regions (DMR's). The epigenetically associated gene expression changes (including *IFITM1*, *ITGB2*, *S100A9*, *SLPI, SAA1*, and *STAT3*) linked to colonic mucosal immune and defense responses. These findings underscore the epigenetic relationships of inflammation in pediatric treatment naïve UC and may have potential etiologic, diagnostic, and therapeutic relevance for IBD.

Keywords: Pediatric inflammatory bowel disease; ulcerative colitis; treatment naïve; epigenetics; DNA methylation; DNA expression; DNA methylome

Abbreviations: AUC, area under the curve; CD, Crohn disease; CpG, cytosine-guanine dinucleotide; DMR: differentially methylated region; FDR, false discovery rate; IBDs, inflammatory bowel diseases; PCA, principal component analysis; RMA, robust multi-array average; SNP, single nucleotide polymorphism; UC, ulcerative colitis.

INTRODUCTION

Inflammatory bowel diseases (IBD), including Crohn disease (CD) and ulcerative colitis (UC), are emerging globally.[1] IBD is associated with high morbidity, leading to a significant health care burden.[2] The incidence of IBD peaks in young adulthood, and cases are becoming more common, especially in the pediatric population.[3-5] The increased number of pediatric cases appear to result from an overall amplified incidence rather than a shift toward disease onset at younger age.[6] About 20% of IBD presents in children,[7-9] where the diseases usually carry a more aggressive phenotype than in adults.[10]

While the cause or causes of IBD remain to be identified, dietary and environmental changes in the industrialized world have been proposed as possible etiologic factors. Such factors may impose critical alterations in the key physiologic components of IBD pathogenesis during prenatal and/or pediatric development.[11, 12] One molecular process that can dynamically respond to nutritional changes in the gut mucosa is DNA methylation.[13] DNA methylation is an epigenetic change occurring at cytosines in CpG dinucleotides and plays an important role in gene expression regulation. Epigenetic maturation relevant for intestinal immune regulation continues beyond infancy in mammals, indicating that even postnatally occurring DNA methylation changes may be important in IBD pathogenesis.[14] Furthermore, maternal micronutrient supplementation can induce colonic mucosal DNA methylation modification relevant for murine colitis susceptibility, implicating this epigenetic process in the developmental origins of IBD.[13, 15]

Differentially methylated CpG sites in inflamed intestinal tissues of IBD patients have been detected with limited (1,505 methylation sites interrogated) molecular screens[16] supporting this epigenetic change as a potential etiologic component in IBD. Metabolite profiling also supported the role of DNA methylation in mammalian mucosal inflammation.[17] A functional methylome map of UC was recently laid out based on colonic mucosal samples from adult

discordant monozygotic twin pairs.[18] However, none of the prior studies examined treatment naïve, or pediatric samples. In the meantime, immuno-therapy and long standing disease may complicate the identification of pathogenic DNA methylome changes in IBD. Additionally, secondary to its more aggressive course, pediatric IBD may carry unique molecular characteristics compared to adult. Therefore, we set out to explore DNA methylation and gene expression associations of pediatric IBD in the colonic mucosa of untreated children.

RESULTS

Mucosal DNA Methylation Separates in Treatment Naïve Pediatric Ulcerative Colitis

Overall methylation variation between combined treatment naïve samples and 2 treated UC samples was examined by correlation based principal component analysis (PCA). There was no consistent separation within the discovery and validation samples when CD cases were compared to controls (**Figure 1**). On the contrary, all treatment naïve UC mucosal samples segregated distinctly from controls by DNA methylation driven PCA in one dimension of the 3 dimensional PCA (**Figure 1, upper panel**). Interestingly, in the same dimension, there was a gender based separation between the samples as well. Nevertheless, treatment naïve UC mucosal DNA methylomes separated from controls both in the female and male groups. Importantly, this separation was lost in the 2 males examined (**grey arrows, Figure 1, upper panel**) upon treatment with infliximab or 6-mercaptopurine, respectively, which induced clinical remission in both patients according to the pediatric ulcerative colitis activity index (PUCAI[19]).

Differentially Methylated Regions in Pediatric Colonic IBD Mucosa

Regional DNA methylation variation may be more important in the regulation of gene expression than single CpG site specific ones. Therefore, we performed bump-hunter analysis[20] to identify differentially methylated regions (DMRs) in pediatric IBD colonic mucosa compared to

controls. We included the treated samples into the analysis to decrease the inevitable bias induced by inflammation as shown in the section above. We chose an arbitrary p value cut-off of $p<10^{-4}$. By this means, we found 182 CD associated (**Table S2**), and 3,365 UC associated DMRs (**Table S3**). These DMRs linked to 108 CD associated, and 2243 UC associated genes, respectively. Fifty-eight (53.7%) of the CD associated genes with DMR links overlapped with the UC associated genes with DMRs. This finding indicated that a large portion of the IBD associated mucosal epigenetic changes overlap and that those may arise from non-specific inflammation, which may be more pronounced in face of the confluent colitis characteristic of UC.

Mucosal Epigenetic Changes in UC Associate with Immune and Defense Responses

Gene expression was interrogated by microarrays in a subset (5 C and 5 UC) of the patients from adjacent biopsy samples to examine the direct functional relevance of the UC specific regional DNA methylation changes. There were 809 genes with increased (>2 fold, $p<0.05$) expression (**Table S4**), and 353 with decreased expression in UC mucosa compared to control (**Table S5**). Ninety five (11.7%) of the genes with increased expression had UC linked DMR associations (**Table S6**). Twenty five (7.01%) of the genes with decreased expression had UC linked DMR associations (**Table S7**).

There was no significant enrichment of biological processes among the genes with decreased expression that had DMR associations, compared to the rest of the genome. On the contrary, immune response, defense response, antigen processing and presentation of peptide antigen, antigen processing and presentation, MHC class I peptide loading complex, and MHC protein complex were significantly (corrected $p<0.05$) enriched among the genes with increased expression that had DMR association.

DISCUSSION

This study includes the most in depth Infinium Methylation Array based interrogation of intestinal mucosal samples in IBD and is the first to examine pediatric cases. Yet, the peak incidence of IBD in young adulthood suggests that epigenetic changes occurring during childhood may be important in the etiology of these disorders.[14] Furthermore, the limited length of disease and the usual absence of co-morbid conditions at the onset of pediatric IBD present a unique opportunity to examine dynamic molecular processes, such as epigenetic changes, in respect to the developmental origins of the disease group.[21]

Our findings indicate that DNA methylation changes in the transverse colonic mucosa are more prominent and therefore lead to a more significant separation in pediatric UC than in CD, compared to controls.

The recently developed methylation microarray employed herein has been validated in normal adult colonic mucosa and colorectal cancer cell lines.[22] Our validation (**Figure S1**) also supports the reliability of this microarray in human colonic mucosal studies. Although we examined small cohorts, the independent interrogation of discovery and validation sets with in-depth genomic microarrays (482,421 CpGs interrogated) and the combined analysis of those provided significant strength to our results. Additionally, the treatment naïve nature of the samples examined is unique compared to any other epigenomic study on mucosal DNA in relationship to IBD thus far. Geographical bias can also complicate multi-center studies on flexible biological systems involved in IBD pathogenesis.[23] Our validation cohort contained samples from a single center to address this potential confounding factor as well.

There was a remarkable colonic mucosal epigenetic separation of UC compared to CD (**Figure 1**). This separation may in part be linked to the clinically less diverse nature of UC. Furthermore, the more prominent UC clustering may have resulted from a more intense and confluent inflammation that frequently characterizes the disease, compared to CD. This

predicament is supported by the fact that the two biopsy samples from treated patients (in clinical remission) clustered together with controls, as opposed to their treatment naïve counterparts (**Figure 1 grey arrows**). The finding for loss of colonic mucosal DNA methylome separation upon conventional treatment of UC is the most important observation in this work. This result indicates that the vast majority of the DNA methylation changes identified were secondary to the inflammatory process of UC and that those are not persistent when the severity of the disease declines.

Over 50% of the CD associated DMRs were identified in UC mucosa as well. This finding also supports our conclusion that a large portion of the colonic mucosal epigenetic modifications in pediatric IBD arise from non-specific inflammatory changes. Our study was not powered to enable the determination whether disease specific DNA methylation differences exist in treatment naïve colonic mucosa, or not. Whether the DNA methylome can differentiate UC from CD will require the examination of much larger sample sizes of the two major IBD classes. Although obtaining mucosal biopsies is more invasive than stool collection or phlebotomy, colonoscopy is generally performed to confirm the diagnosis of IBD. Sampling for DNA methylation based differential diagnostic purposes could take place during this procedure. DNA methylation in more readily available cells such as peripheral blood leukocytes (PBL) did not differentiate between UC and CD in treatment naïve children with the same arrays as utilized in this study.[24] This may be secondary to the need for cell type based fractionation of PBLs prior to DNA methylation interrogation,[25, 26] since epigenetic patterns are cell and tissue specific. Similarly, mucosal cell sub-fractionation from colonic biopsy specimens[27] may further our understanding of IBD-epigenome relationships. However, sub-fractionation required pooled biopsy samples to generate sufficient DNA material for high-throughput analyses, limiting its clinical feasibility. Cell sub-fractionation may also result in the potential loss of critical epigenetic information, which question requires further analyses in the future.

The associations between DNA methylation and gene expression delineated in this study highlight the functional relevance of the UC specific epigenetic changes identified. There was a significantly increased expression of genes involved in immune and defense responses in association with the DNA methylome of treatment naïve UC. This result highlights the importance of epigenetic mechanisms in the regulation of the inflammatory cascade characteristic for UC.

Among the genes with UC specific increased expression in association with DNA methylation changes, S100 calcium binding protein A9 (*S100A9*) has already been found with increased expression in both CD and UC colonic mucosa.[28] In agreement with our results, other investigations outside of the GI tract on *S100A9* have shown its increased expression upon treatment with 5-aza-2'-deoxycitidine, and inhibitor of DNA methyltransferase 1 (DNMT1).[29] Similar to our findings, secretory leukocyte peptidase inhibitor (*SLPI*) expression was increased selectively in UC mucosa (not in CD) that appeared to originate from mucosal inflammatory cells.[30] This later work indicates that some of the gene expression signals detected in this study may have derived from mucosal immune cells rather than epithelial, albeit specific to UC. Another important result is the increased expression of *STAT3* in the colonic mucosa of treatment naïve UC patients in association with decreased gene body methylation overlapping 2 exons (Supplementary Table 6). The importance of the STAT3 signaling pathway has been highlighted in pediatric onset UC, where increased expression of the gene was noted in the colonic mucosa,[31] similar to our findings. Inverse correlation between gene methylation and expression of *STAT3* has been recently demonstrated,[32] supporting our results, which indicate that epigenetic changes may be an important element in the regulation of the STAT3 signaling pathway. Further publications demonstrate the reliability and value of our findings on inflammation associated increased expression of *SAA1*,[33] *IFITM1* and *ITGB2*[34] in UC, for example.

This work includes the first high-throughput DNA methylation and gene expression association study on IBD from treatment naïve colonic mucosal samples. The results indicate the potential importance of epigenetic mechanisms in the modulation of the inflammatory changes characteristic of UC. We also conclude that the large majority of the DNA methylation and gene expression changes at the diagnosis of the disease are transient and respond to treatments/disease remission. Further work towards deciphering the contribution of different cell types to the findings of this study performed on whole mucosal biopsies is needed. Our findings may promote the establishment of DNA methylation based differential diagnostic and therapeutic modalities for pediatric UC.

PATIENTS AND METHODS

Patients and Samples. Secondary to the dynamic nature of epigenetic changes, the following factors were controlled: age, gender, colonic location,[35, 36] and treatment status.[37] Hence, only treatment-naïve patients and controls were examined. Discovery (10 controls [C], 10 Crohn disease [CD], and 4 ulcerative colitis [UC]) and validation cohorts (12 C, 5 CD, 5 UC) of patients were included (**Table 1 and 2**). Control patients in the discovery cohort were recruited prior to endoscopy through the pediatric gastroenterology tissue bank of the Pediatric Inflammatory Bowel Disease Consortium Registry of the Baylor College of Medicine (BCM), Charles University, Prague, Czech Republic, and MassGeneral Hospital *for* Children, which were established in agreement with local and federal regulations. Only patients with grossly and histologically normal mucosa at colonoscopy were designated as controls. Treatment naïve IBD cases in the discovery cohort were recruited prior to their first diagnostic colonoscopy in the latter two locations and their disease was determined based upon clinical, biochemical and histological characteristics. The validation cohort patients were all enrolled at BCM to eliminate potential center bias. Left sided (from mid transverse to rectal) colonic mucosal samples were

snap frozen on dry ice or in liquid nitrogen immediately after biopsy and stored at -80^{o}C until further analysis. The inflammatory cell infiltration in the adjacent transverse colon biopsies of UC patients was rated between mild to severe on a numeric scale by a blinded pathologist. Treated samples from one patient each (**Table 1***) from the two cohorts were also included into the analyses to examine and decrease inflammation bias.

DNA Extraction. After thawing, the colonic mucosal biopsies of the discovery cohort were centrifuged at 14,000 rpm for 30 seconds and resuspended in 500µl RLT buffer (Qiagen, Valencia, CA) (with β- mercaptoethanol). Sterile 5mm steel beads (Qiagen, Valencia, CA) and 500µl sterile 0.1mm glass beads (Scientific Industries, Inc., NY, USA) were added for complete bacterial lyses in a Qiagen TissueLyser (Qiagen, Valencia, CA), run at 30Hz for 5min. Samples were centrifuged briefly and 100µl of 100% ethanol added to a 100µl aliquot of the sample supernatant. This mixture was added to a DNA spin column, and DNA recovery protocols were followed as instructed in the QIAamp DNA Mini Kit (Qiagen, Valencia, CA) starting at step 5 of the Tissue Protocol. DNA was eluted from the column with 30µl water and samples were diluted accordingly to a final concentration of 20ng/µl. As for the discovery cohort, genomic DNA was isolated by standard proteinase-k digestion and phenol-chloroform extraction and bisulfite converted as described previously.[38] This approach eliminated potential DNA extraction method based bias. DNA concentrations were determined by a Nanodrop spectrophotometer (Nyxor Biotech, Paris, France).

RNA Extraction and Processing for Microarrays. Colonic mucosal RNA was isolated by Qiangen-Qiazol - miRNA Isolation Kit. cDNA amplification and labeling was performed with Ovation Pico WTA System V2 and Encore Biotin Module (NuGEN), respectively. Array hybridization was performed according to Affymetrix FS450_0002 Hybridization Protocol for

gene expression. The Affymetrix GeneChip® PrimeView™ Human Gene Expression Arrays were scanned with Affymetrix Genechip Scanner 7G.

Infinium Methylation Assay Microarrays. The transverse colonic mucosal DNA sample quality was examined with PicoGreen (http://probes.invitrogen.com/media/pis/mp07581.pdf) before processing towards the microarrays. The samples that passed quality control were processed by Infinium HumanMethylation450 BeadChip Kits (Illumina San Diego, CA, USA; 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 Illumina scanner settings. The R Bioconductor minifi package[39] was used to generate beta values normalized to internal control probes. Internal controls determined the array processing to be of good quality. After removal of probes containing SNPs (http://www.rforge.net/IMA/snpsites.txt), 390,433 CpG probes on the array were used for subsequent analysis. The R Linear Models for Microarray Data (Limma) package[40] was used to compare beta values and identify differentially methylated probes between 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 a t-statistic using an empirical Bayesian model that moderates the standard errors across probes. P-values were calculated from the moderated t-statistics and multiple testing correction of the p-values was performed using Benjamini and Hochberg's method[41] (false discovery rate: FDR) to identify differentially methylated probes. The R Bioconductor bumphunter package was used for identifying differentially methylated regions. The R prcomp function was used to perform principal component analysis (PCA) based on beta values for the samples. The raw data of the microarrays was uploaded to Gene Expression

Omnibus (GEO; Series GSE32146) and is accessible at: Discovery cohort: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=frutfkgmamamkvi&acc=GSE32146. Validation cohort: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=pvojjiqcqiuqori&acc=GSE42921.

Bisulfite Pyrosequencing Validation. Colonic mucosal DNA from the discovery cohort was bisulfite converted with EZ DNA Methylation-Gold Kit (D5006, Zymo Research, Orange, CA, USA). DNA was amplified with traditional primer biotinylation following bisulfite conversion. A quantitative bisulfite pyrosequencing protocol was used for all methylation analyses with the utilization of the Pyro Q CpG program (QIAGEN GmbH, QIAGEN Strasse 1. 40724 Hilden, Germany). Infinum array validation was performed at 5 candidate loci with corrected $p<0.01$ in the discovery cohort: (*SMAD3* associated: chr15:67442893-67442893; *C8orf74* associated: chr8:10555466-10555466; *FLT1* associated: chr13:28975690-28975690; "Chr1": chr1:120289923-120289923 and "Achr1": chr1:150535935-150535935). **Table S1** shows the pyrosequencing primers utilized for the validation.

Bisulfite pyrosequencing at 5 independent loci correlated significantly with the microarrays of the discovery cohort ($r=0.87$, $p<0.0001$; **Figure S1**) supporting the reliability of the results. **Table S1** shows the pyrosequencing primers utilized for the validation of the discovery microarrays.

Gene Expression Microarray Analysis. The R Bioconductor affy package[42]_ENREF_43 RMA (robust multi-array average) function was used to compute expression. The CEL files and RMA normalized expression values can be downloaded from http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=pzyffweqggmsgnk&acc=GSE42911.

Affymetrix GeneChip PrimeView Human Gene Expression Array probe annotations were associated with Infinium HumanMethylation450 v1.1 probe annotations based on gene symbols. The beta values of Infinium probes showing differential methylation between UC and control samples were compared to the averaged RMA values for genes with multiple probe sets based on the gene symbol association.

Statistical and Bioinformatic Analysis. Unpaired, two tailed, t-test; was used in the group comparisons. Statistical significance was declared at $p<0.05$. DAVID (http://david.abcc.ncifcrf.gov/) was utilized to examine gene ontology enrichment in the select gene lists; significance was declared at P values adjusted for multiple testing $p<0.05$.

Acknowledgements: The authors would like to acknowledge the patients for providing samples, and all of our colleagues who contributed to the successful banking of the tissues.

Data Access:
DNA methylation: Gene Expression Omnibus (GEO; Series GSE32146) and is accessible at: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=frutfkgmamamkvi&acc=GSE32146.
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=pvojjiqcqiuqori&acc=GSE42921
Gene expression: The CEL files and RMA normalized expression values can be downloaded from http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=pzyffweqggmsgnk&acc=GSE42911.

Disclosures: The authors have no conflict of interests to declare.

Grant Support: R.K. was supported in part by the Broad Medical Research Program, the Broad Foundation (IBD-0252); the Child Health Research Career Development Agency of the Baylor College of Medicine (NIH # 5K12 HD041648); a Public Health Service grant DK56338, funding the Texas Medical Center Digestive Diseases Center; and the Gutsy Kids Fund including philanthropic donation from the Karen and Brock Wagner family. H.S.W. was supported in part by a philanthropic donation from the Martin Schlaff Family. This project was also supported by the Pediatric Inflammatory Bowel Disease Foundation.

FIGURE LEGENDS

Figure 1 Principal component analysis (PCoA) of DNA methylation. CD: Crohn disease; Ctrl: Control; UC: ulcerative colitis. The ulcerative colitis DNA methylome separated in one dimension from controls both in males and females (**upper panel**). Interestingly, males and females separated from each other regardless of disease status in this dimension. The UC specific separation was lost upon treatment induced clinical remission in 2 of the male patients studied (grey arrows). There was no obvious disease specific separation in the other dimensions of the analysis (**middle and lower panels**).

REFERENCES

1. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, Benchimol EI, Panaccione R, Ghosh S, Barkema HW, Kaplan GG. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology 2012;142:46-54 e42; quiz e30.
2. Park KT, Bass D. Inflammatory bowel disease-attributable costs and cost-effective strategies in the United States: a review. Inflamm Bowel Dis 2011;17:1603-9.
3. Benchimol EI, Guttmann A, Griffiths AM, Rabeneck L, Mack DR, Brill H, Howard J, Guan J, To T. Increasing incidence of paediatric inflammatory bowel disease in Ontario, Canada: evidence from health administrative data. Gut 2009;58:1490-7.
4. Abramson O, Durant M, Mow W, Finley A, Kodali P, Wong A, Tavares V, McCroskey E, Liu L, Lewis JD, Allison JE, Flowers N, Hutfless S, Velayos FS, Perry GS, Cannon R, Herrinton LJ. Incidence, Prevalence, and Time Trends of Pediatric Inflammatory Bowel Disease in Northern California, 1996 to 2006. J Pediatr.
5. Gupta N, Bostrom AG, Kirschner BS, Ferry GD, Gold BD, Cohen SA, Winter HS, Baldassano RN, Abramson O, Smith T, Heyman MB. Incidence of stricturing and penetrating complications of Crohn's disease diagnosed in pediatric patients. Inflamm Bowel Dis;16:638-44.
6. Braegger CP, Ballabeni P, Rogler D, Vavricka SR, Friedt M, Pittet V. Epidemiology of Inflammatory Bowel Disease: Is There a Shift Towards Onset at a Younger Age? J Pediatr Gastroenterol Nutr 2011;53:141-144.
7. Griffiths AM. Specificities of inflammatory bowel disease in childhood. Best Pract Res Clin Gastroenterol 2004;18:509-23.
8. Kappelman MD, Moore KR, Allen JK, Cook SF. Recent Trends in the Prevalence of Crohn's Disease and Ulcerative Colitis in a Commercially Insured US Population. Dig Dis Sci 2013;58:519-25.
9. Rocchi A, Benchimol EI, Bernstein CN, Bitton A, Feagan B, Panaccione R, Glasgow KW, Fernandes A, Ghosh S. Inflammatory bowel disease: a Canadian burden of illness review. Can J Gastroenterol 2012;26:811-7.
10. Van Limbergen J, Russell RK, Drummond HE, Aldhous MC, Round NK, Nimmo ER, Smith L, Gillett PM, McGrogan P, Weaver LT, Bisset WM, Mahdi G, Arnott ID, Satsangi J, Wilson DC. Definition of phenotypic characteristics of childhood-onset inflammatory bowel disease. Gastroenterology 2008;135:1114-22.
11. Barnett M, Bermingham E, McNabb W, Bassett S, Armstrong K, Rounce J, Roy N. Investigating micronutrients and epigenetic mechanisms in relation to inflammatory bowel disease. Mutat Res 2010;690:71-80.
12. 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 2011;106:563-73.
13. Schaible TD, Harris RA, Dowd SE, Smith CW, Kellermayer R. Maternal methyl-donor supplementation induces prolonged murine offspring colitis susceptibility in association with mucosal epigenetic and microbiomic changes. Hum Mol Genet 2011;20:1687-96.
14. Kellermayer R, Balasa A, Zhang W, Lee S, Mirza S, Chakravarty A, Szigeti R, Laritsky E, Tatevian N, Smith CW, Shen L, Waterland RA. Epigenetic maturation in colonic mucosa continues beyond infancy in mice. Hum Mol Genet 2010;19:2168-76.
15. Mir SA, Nagy-Szakal D, Dowd SE, Szigeti RG, Smith CW, Kellermayer R. Prenatal methyl-donor supplementation augments colitis in young adult mice. PLoS ONE 2013;8:e73162.

16. Lin Z, Hegarty J, Cappel J, Yu W, Chen X, Faber P, Wang Y, Kelly A, Poritz L, Peterson B, Schreiber S, Fan JB, Koltun W. Identification of disease-associated DNA methylation in intestinal tissues from patients with inflammatory bowel disease. Clin Genet 2010.
17. Kominsky DJ, Keely S, MacManus CF, Glover LE, Scully M, Collins CB, Bowers BE, Campbell EL, Colgan SP. An endogenously anti-inflammatory role for methylation in mucosal inflammation identified through metabolite profiling. J Immunol 2011;186:6505-14.
18. Hasler R, Feng Z, Backdahl L, Spehlmann ME, Franke A, Teschendorff A, Rakyan VK, Down TA, Wilson GA, Feber A, Beck S, Schreiber S, Rosenstiel P. A functional methylome map of ulcerative colitis. Genome Res 2012;22:2130-7.
19. Turner D, Otley AR, Mack D, Hyams J, de Bruijne J, Uusoue K, Walters TD, Zachos M, Mamula P, Beaton DE, Steinhart AH, Griffiths AM. Development, validation, and evaluation of a pediatric ulcerative colitis activity index: a prospective multicenter study. Gastroenterology 2007;133:423-32.
20. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, Irizarry RA. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int J Epidemiol 2012;41:200-9.
21. Kellermayer R. Epigenetics and the developmental origins of inflammatory bowel diseases. Can J Gastroenterol 2012;26:909-15.
22. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics 2011;6:692-702.
23. Kellermayer R, Mir SA, Nagy-Szakal D, Cox SB, Dowd SE, Kaplan JL, Sun Y, Reddy S, Bronsky J, Winter HS. Microbiota separation and C-reactive protein elevation in treatment-naive pediatric granulomatous Crohn disease. J Pediatr Gastroenterol Nutr 2012;55:243-50.
24. Harris RA, Nagy-Szakal D, Pedersen N, Opekun A, Bronsky J, Munkholm P, Jespersgaard C, Andersen P, Melegh B, Ferry G, Jess T, Kellermayer R. Genome-wide peripheral blood leukocyte DNA methylation microarrays identified a single association with inflammatory bowel diseases. Inflamm Bowel Dis 2012;18:2334-41.
25. Lin Z, Hegarty JP, Yu W, Cappel JA, Chen X, Faber PW, Wang Y, Poritz LS, Fan JB, Koltun WA. Identification of disease-associated DNA methylation in B cells from Crohn's disease and ulcerative colitis patients. Dig Dis Sci 2012;57:3145-53.
26. Kellermayer R. Hurdles for epigenetic disease associations from peripheral blood leukocytes. Inflamm Bowel Dis 2013;19:E66-7.
27. Jenke AC, Postberg J, Raine T, Nayak KM, Molitor M, Wirth S, Kaser A, Parkes M, Heuschkel RB, Orth V, Zilbauer M. DNA methylation analysis in the intestinal epithelium-effect of cell separation on gene expression and methylation profile. PLoS ONE 2013;8:e55636.
28. Lawrance IC, Fiocchi C, Chakravarti S. Ulcerative colitis and Crohn's disease: distinctive gene expression profiles and novel susceptibility candidate genes. Hum Mol Genet 2001;10:445-56.
29. Xiao YH, Yi H, Tan T, Liang T, Chen ZC, Xiao ZQ. [Analysis of in vitro anti-leukemia effect of 5-aza-2'-deoxycitydine]. Zhong Nan Da Xue Xue Bao Yi Xue Ban 2008;33:344-52.
30. Schmid M, Fellermann K, Fritz P, Wiedow O, Stange EF, Wehkamp J. Attenuated induction of epithelial and leukocyte serine antiproteases elafin and secretory leukocyte protease inhibitor in Crohn's disease. J Leukoc Biol 2007;81:907-15.
31. Carey R, Jurickova I, Ballard E, Bonkowski E, Han X, Xu H, Denson LA. Activation of an IL-6:STAT3-dependent transcriptome in pediatric-onset inflammatory bowel disease. Inflamm Bowel Dis 2008;14:446-57.

32. Yang J, Bai W, Niu P, Tian L, Gao A. Aberrant hypomethylated STAT3 was identified as a biomarker of chronic benzene poisoning through integrating DNA methylation and mRNA expression data. Exp Mol Pathol 2014;96:346-353.
33. Okahara S, Arimura Y, Yabana T, Kobayashi K, Gotoh A, Motoya S, Imamura A, Endo T, Imai K. Inflammatory gene signature in ulcerative colitis with cDNA macroarray analysis. Aliment Pharmacol Ther 2005;21:1091-7.
34. Roman J, Planell N, Lozano JJ, Aceituno M, Esteller M, Pontes C, Balsa D, Merlos M, Panes J, Salas A. Evaluation of responsive gene expression as a sensitive and specific biomarker in patients with ulcerative colitis. Inflamm Bowel Dis 2013;19:221-9.
35. Horii J, Hiraoka S, Kato J, Harada K, Kuwaki K, Fujita H, Toyooka S, Yamamoto K. Age-related methylation in normal colon mucosa differs between the proximal and distal colon in patients who underwent colonoscopy. Clin Biochem 2008;41:1440-8.
36. Wang Y, Devkota S, Musch MW, Jabri B, Nagler C, Antonopoulos DA, Chervonsky A, Chang EB. Regional mucosa-associated microbiota determine physiological expression of TLR2 and TLR4 in murine colon. PLoS ONE 2010;5:e13607.
37. Lepage P, Hasler R, Spehlmann ME, Rehman A, Zvirbliene A, Begun A, Ott S, Kupcinskas L, Dore J, Raedler A, Schreiber S. Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. Gastroenterology 2011;141:227-36.
38. Waterland RA, Jirtle RL. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. Mol Cell Biol 2003;23:5293-300.
39. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30:1363-9.
40. Smyth GK. Limma: linear models for microarray data. In: R. Gentleman VC, S. Dudoit, R. Irizarry, W. Huber ed. Bioinformatics and Computational Biology Solutions using R and Bioconductor. 2005 ed: Springer, New York, 2005:397-420.
41. Benjamini YaH, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. B 1995;57:289-300.
42. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 2004;20:307-15.

Cohorts		Age (y)	G	R	Diagnosis		
Discovery cohort	C	3.5	M	W	Juvenile polyp		
	C	15	F	W	Hematochezia		
	C	17	M	W	Perianal fissure		
	C	17	F	W	Abdominal pain		
	C	17	F	W	IBS		
	C	17.5	F	W	Healthy		
	C	8	F	W	Hematochezia		
	C	12	M	W	Diarrhea		
	C	13	M	W	Abdominal pain		
	C	10	M	W	Abdominal pain		
					Montreal classification		
					Age	**Location**	**Behavior**

Cohorts		Age (y)	G	R	Diagnosis
Validation cohort	C	11	F	H	Abdominal pain
	C	17	M	H	Diarrhea, weight loss
	C	15	M	W	Abdominal pain
	C	17	M	AA	Abdominal pain
	C	14	F	W	Abdominal pain
	C	15	F	W	Abdominal pain
	C	16	F	W	Diarrhea
	C	16	F	W	Abdominal pain
	C	15	F	W	Abdominal pain
	C	16	F	W	Hematochezia
	C	15	M	W	Abdominal pain, hematochezia
	C	9	F	W	Abdominal pain, hematochezia

CD	16	M	W	A2	L3	B1
CD	17	M	W	A2	L3	B1
CD	15	M	W	A1	L3 L4	B1
CD	8	F	W	A1	L3 L4	B1
CD	16.5	F	W	A2	L3 L4	B1
CD	13	F	W	A1	L3 L4	B1
CD	8	F	W	A1	L2 L4	B1
CD	15.5	F	AA	A2	L3 L4	B1
CD	17.5	M	W	A2	L3 L4	B1
CD	13	F	na	A1	L2	B2
				Montreal classification of extent of UC		
UC	10	F	W	E3		
UC	15	F	W	E3		
UC	19	M	na	E2		
*UC	13	M	W	E3		

				Montreal classification		
				Age	**Location**	**Behavior**
CD	6	F	W	A1	L3 L4	B1
CD	11	M	A	A1	L2	B1
CD	11	M	W	A1	L2 L4	B2
CD	19	F	AA	A2	L2	B2
CD	16	F	H	A2	L2	B1
				Montreal classification of extent of UC		
UC	12	F	W	E3		
UC	11	M	AA	E3		
*UC	16	M	W	E3		
UC	13	F	AA	E3		
UC	5	F	H	E3		

Table 1. Demographics and disease characteristics of the discovery and validation cohorts. A (y):Age (years); A:Asian; AA:African American; B:Behavior, C:Control; CD:Crohn Disease; E:Extent; F:Female; G:Gender; H:Hispanic; L:Location; M:Male; na:non-applicable; R:Race; UC: Ulcerative Colitis; W:White. * indicates patients whose post-treatment samples were also interrogated

Cohorts	Diagnosis	Age (years)			Gender
		Mean	Median	Min to Max	(M;F)
Discovery Cohort (n=24)	C	13.0	14.0	3.5-17.5	5;5
	CD	14	15	8-17.5	4;6
	UC	14	14	10-19	2;2
Validation Cohort (n=22)	C	14.67	15.0	9-17	4;8
	CD	12.6	11.0	6-19	2;3
	UC	11.4	12.0	5-16	2;3

Table 2. Summary of cohort groups including the diagnosis, age and gender of the patients in this study. C:Control; CD:Crohn Disease; F:Female; M:Male; UC: Ulcerative Colitis.