Title

Blood-Derived DNA Methylation Signatures of Crohn's Disease and Severity of Intestinal Inflammation

Authors

Somineni, Hari K
Venkateswaran, Suresh
Kilaru, Varun
et al.

Blood-derived DNA Methylation Signatures of Crohn's Disease and Severity of Intestinal Inflammation

Hari K Somineni[1,2], **Suresh Venkateswaran**[2], **Varun Kilaru**[3], **Urko M Marigorta**[4], **Angela Mo**[4], **David T Okou**[2], **Richard Kellermayer**[5], **Kajari Mondal**[2], **Dawayland Cobb**[3], **Thomas D Walters**[6], **Anne Griffiths**[6], **Joshua D Noe**[7], **Wallace V Crandall**[8], **Joel R Rosh**[9], **David R Mack**[10], **Melvin B Heyman**[11], **Susan S Baker**[12], **Michael C Stephens**[13], **Robert N Baldassano**[14], **James F Markowitz**[15], **Marla C Dubinsky**[16], **Judy Cho**[16], **Jeffrey S Hyams**[17], **Lee A Denson**[18], **Greg Gibson**[4], **David J Cutler**[19], **Karen N Conneely**[1,19,$], **Alicia K Smith**[1,3,20,$], and **Subra Kugathasan**[1,2,19,$,*]

[1]Genetics and Molecular Biology Program, Emory University, Atlanta, GA, USA

[2]Division of Pediatric Gastroenterology, Department of Pediatrics, Emory University School of Medicine & Children's Healthcare of Atlanta, Atlanta, GA, USA

[3]Department of Gynecology & Obstetrics, Emory University School of Medicine, Atlanta, GA, USA

[4]Center for Integrative Genomics, Georgia Institute of Technology, Atlanta, GA, USA

[5]Section of Pediatric Gastroenterology, Baylor College of Medicine, Texas Children's Hospital, Houston, TX, USA

[6]Division of Pediatric Gastroenterology, Hepatology and Nutrition, Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada

[7]Department of Pediatric Gastroenterology, Hepatology and Nutrition, Medical College of Wisconsin, Milwaukee, WI, USA

*Corresponding author: Subra Kugathasan, MD, Division of Pediatric Gastroenterology, Emory University School of Medicine &, Children's Healthcare of Atlanta, 1760 Haygood Drive, W-427, Atlanta, GA 30322, Tel: 404 727 1316, Fax: 404 727 4069, skugath@emory.edu.
$Equal contribution for last authorship

[8]Division of Pediatric Gastroenterology, Nationwide Children's Hospital, Ohio State University College of Medicine, Columbus, OH, USA

[9]Department of Pediatrics, Goryeb Children's Hospital, Morristown, NJ, USA

[10]Department of Pediatrics, Children's Hospital of Eastern Ontario IBD Centre and University of Ottawa, Ottawa, ON, Canada

[11]Department of Pediatrics, University of California, San Francisco, San Francisco, CA, USA

[12]Department of Digestive Diseases and Nutrition Center, University at Buffalo, Buffalo, NY, USA

[13]Department of Pediatric Gastroenterology, Mayo Clinic, Rochester, MN, USA

[14]Department of Pediatrics, University of Pennsylvania, Philadelphia, PA, USA

[15]Department of Pediatrics, Northwell Health, New York, NY, USA

[16]Department of Pediatrics, Mount Sinai Hospital, New York, NY, USA

[17]Division of Digestive Diseases, Hepatology, and Nutrition, Connecticut Children's Medical Center, Hartford, CT, USA

[18]Division of Pediatric Gastroenterology, Hepatology, and Nutrition, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

[19]Department of Human Genetics, Emory University, Atlanta, GA, USA

[20]Department of Psychiatry & Behavioral Sciences, Emory University, Atlanta, GA, USA

Abstract

Background & Aims: Crohn's disease is a relapsing and remitting inflammatory disorder with a variable clinical course. Although most patients present with an inflammatory phenotype (B1), approximately 20% of patients rapidly progresses to complicated disease, which includes stricturing (B2) within 5 years. We analyzed DNA methylation patterns in blood samples of pediatric patients with Crohn's disease at diagnosis and later time points to identify changes that associate with and might contribute to disease development and progression.

Methods: We obtained blood samples from 164 pediatric patients (ages, 1–17 years) with Crohn's disease (B1 or B2) who participated in a North American study and were followed for 5 years. Participants without intestinal inflammation or symptoms were used as controls (n=74). DNA methylation patterns were analyzed at approximately 850,000 sites, in samples collected at time of diagnosis and 1–3 years later. We used genetic association and the concept of Mendelian randomization to identify changes in DNA methylation patterns that might contribute to development of or result from Crohn's disease.

Results: We identified 1189 CpGs that were differentially methylated between patients with Crohn's disease (at diagnosis) and controls. Methylation changes at these sites correlated with plasma levels of C-reactive protein. A comparison of methylation profiles of DNA collected at diagnosis of Crohn's disease vs during the follow-up period revealed that, during treatment, alterations identified in methylation profiles at the time of diagnosis of Crohn's disease more closely resembled patterns observed in controls, irrespective of disease progression to B2. We

identified methylation changes at 3 CpGs that might contribute to the development of Crohn's disease. Most CpG methylation changes associated with Crohn's disease disappeared with treatment of inflammation, and might be a result of Crohn's disease.

Conclusions: Methylation patterns observed in blood samples from patients with Crohn's disease accompany acute inflammation; with treatment, they change to resemble methylation patterns observed in patients without intestinal inflammation. These findings indicate that Crohn's disease-associated patterns of DNA methylation observed in blood samples are a result of the inflammatory features of the disease and are less likely to contribute to disease development or progression.

Graphical Abstract

Gastroenterology
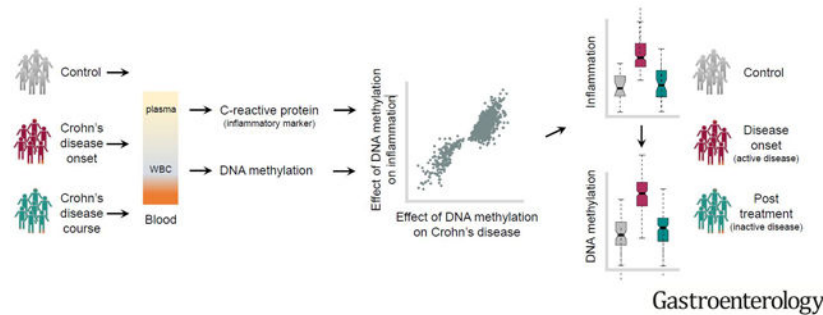
Keywords

IBD; children; epigenetic alteration; RISK study

Inflammatory Bowel Diseases (IBD) encompassing Crohn's disease and ulcerative colitis arise in the context of complex interactions between genetic and environmental factors. While these diseases can manifest at any age, pediatric-onset Crohn's disease has a higher incidence than ulcerative colitis[1, 2], and patients diagnosed with Crohn's disease in childhood are more likely to suffer from an aggressive and severe disease course[2].

DNA methylation, occuring predominantly in the cytosine-guanine (CpG) dinucleotide context, is a key epigenetic mechanism that can regulate gene expression and thereby influence the development and progression of complex diseases. Cross-sectional studies of DNA methylation have begun to reveal epigenetic associations with IBD in both pediatric and adult populations; across a range of cell and tissue types[3-11]. For instance, site-specific DNA methylation differences in peripheral blood[3] and blood-derived mononuclear cells[5] of adult patients with IBD have been reported. Similarly, studies of mixed or purified cells from blood and intestinal mucosa of pediatric populations revealed distinct methylation profiles in relevance to IBD[8, 9]. Howell *et al.*, recently reported a gut segment-specific methylation signature in pediatric patients with IBD in the purified intestinal epithelial cells, and its persistence during the course of the disease[12]. However, due to the relapsing-remitting behavior of Crohn's disease, and the dynamic nature of DNA methylation and its resulting vulnerability to confounding and reverse causation, delineating the causal role of methylation in Crohn's disease requires longitudinal studies along with the application of integrative analytical approaches. Understanding how the methylome changes during the

course of the disease, as a result of varying clinical characteristics, and how disease complications evolve may aid in the identification of potentially causal epigenetic targets, which could subsequently be leveraged for therapeutic benefits.

Here, we performed an epigenome-wide association analysis of DNA methylation in peripheral blood at ~850,000 sites and Crohn's disease 1) at diagnosis and 2) at later stages (1 to 3 years after diagnosis) during which time ~33% of the patients progressed from an initial stage of B1 inflammatory behavior to B2 stricturing behavior. Study participants (summarized in Table 1) were sampled from the RISK cohort[13], a pediatric prospective inception Crohn's disease cohort. Since the current Crohn's disease therapeutics systematically targets the peripheral immune system, and considerable genetic and cell biological evidence including previous epigenetic studies implicates the immune system in the etiology of Crohn's disease[3, 14], we investigated methylation changes in peripheral blood with respect to their potential causal versus consequential roles in disease.

METHODS

Study Population.

We utilized a subset of pediatric subjects recruited under the Risk Stratification and Identification of Immunogenetic and Microbial Markers of Rapid Disease Progression in Children with Crohn's Disease (RISK) study[13]. The RISK inception cohort study is thus far, the largest pediatric Crohn's disease cohort recruited at 28 sites in the USA and Canada to identify genetic, clinical, microbial and immunologic factors that predispose Crohn's disease patients (B1) to a complicated disease course (B2 or B3). Briefly, the RISK study recruited children with ages 1-17 who presented to gastroenterology clinics with suspected inflammatory bowel disease and followed them for a period of 5 years at regular intervals to determine the incidence of inflammatory bowel disorders or complications of an established disorder. The RISK study design, recruitment details, inclusion-exclusion criteria, disease behaviors, and data collection have been described in detail elsewhere[13].

Study design.

The initial recruitment and follow-up have been previously described[13]. A subset of age-, sex-, and ethnicity-matched non-IBD control subjects and Crohn's disease patients with B1 inflammatory behavior and B2 stricturing behavior were drawn from the RISK cohort[13], based on the availability of patient samples at two time points – at diagnosis and at a follow-up visit 1 to 3 years after diagnosis (Table 1). Subjects who were negative for gut inflammation and depicted no bowel pathology on endoscopy, and remained IBD symptom-free during the course of the follow-up period served as non-IBD controls. Peripheral blood DNA samples from 164 newly diagnosed, treatment-naïve pediatric patients with Crohn's disease (cases) and 74 non-IBD controls (controls) were considered for baseline analysis to identify Crohn's disease associated CpGs. Of these, 150 cases presented purely with an inflammatory phenotype (non-complicated Crohn's disease; B1) while the remaining 14 presented with stricturing phenotype (B2) at diagnosis. However, sensitivity analysis comparing 150 B1 cases or 14 B2 cases to 74 non-IBD controls versus 164 cases (150 B1, 14 B2) to 74 non-IBD controls showed that our findings are robust to disease behavior states

(B1 or B2), allowing grouping of all cases (both B1 and B2) at diagnosis into a single large cohort.

The longitudinal analysis relied on follow-up samples taken from established cases ($n = 164$) as part of a longitudinal follow-up in the RISK study, which was 1 to 3 years from diagnosis. Exact details are provided in Table 1. Of the 150 B1 cases at diagnosis, 55 of them progressed to B2 (progressors) during the course of the follow-up period, while the rest ($n = 95$) remained as B1 at the time of the follow-up sampling (non-progressors). We note that in order to increase statistical power to define (if any) methylomic changes involved in disease progression, we purposefully inflated the number of progressors by selecting more pediatric cases who experienced B2 complication during the course of their prospective follow-up in the original RISK study[13]. With the 14 at-diagnosis-B2 patients who also remained as B2 at the time of follow-up sampling, we had a total of 95 B1 and 69 B2 at the follow-up. More details about phenotype classification are available in Kugathasan *et al.*[13].

Quantification of genome-wide DNA methylation and data processing.

Peripheral blood genomic DNA was extracted using the AllPrep DNA/RNA Mini Kit (Qiagen, Valencia, CA). 500ng of extracted DNA from each sample was subjected to bisulfite treatment using EZ DNA Methylation-GoldTM Kit (Zymo Research, Irvine, CA). Genome-wide DNA methylation was quantified in bisulfite-converted genomic DNA at single-base resolution using the MethylationEPIC BeadChip (Illumina, San Diego, CA). The initial quality control for the data set was performed with the R package CpGassoc[15]. CpG sites called with low signal or low confidence (detection $P > 0.05$) or with data missing for greater than 10% of samples were removed, and samples with data missing or called with low confidence for greater than 10% of CpG sites were removed. In addition, probes mapping to multiple locations were removed[16]. After the above steps, a total of 807,511 probes and 402 samples (74 non-IBD controls and 2 samples from each of the 164 cases) remained. Beta values (β) were calculated for each CpG site as the ratio of methylated (M) to methylated and unmethylated (U) signal: $\beta = M/(M+U)$. Signal intensities were then normalized using the module beta-mixture quantile dilation (BMIQ)[17] to account for the probe design bias in the EPIC array data. These normalized signal intensities were used to perform principal component analysis to further identify sample outliers (Supplementary Fig. 1). Differential cell counts of the constituent cell types, $CD4^+$ T cells, $CD8^+$ T cells, NK cells, B cells, monocytes, and granulocytes were estimated for each individual from the methylation data using the Houseman algorithm[18].

Genotyping and data processing.

Peripheral blood DNA samples from the 238 subjects with methylation data were genotyped using the Infinium Multi-Ethnic Global-8 Kit (Illumina, San Diego, CA) and genotypes were called using the GenomeStudio software. All these subjects had call rates >95% and inferred gender consistent with the clinical records. We tested for relatedness among the subjects by calculating pairwise identity by descent based on 59,889 LD-independent SNPs ($r^2 < 0.1$), which confirmed no relatedness among the subjects. The Multi-Ethnic array contained 1,762,905 variants before quality control. Removal of (i) SNPs with low call rate (< 95%), (ii) SNPs not in Hardy-Weinberg equilibrium ($P < 1.0 \times 10^{-3}$), and (iii) SNPs with minor

allele frequency (MAF) < 5%, resulted in the retention of 1,751,369 SNPs, 1,736,281 SNPs and 651,370 SNPs, respectively. We further removed non-autosomal SNPs and SNPs mapping to multiple locations. This resulted in a data set consisting of 636,006 high quality SNPs. All quality control procedures were performed in PLINK[19].

Genotype-based principal components.

Principal components were computed based on a pruned version of the data set consisting of 59,889 LD-independent SNPs ($r^2 < 0.1$) and MAF > 0.05. Unless stated otherwise, the first 3 genotype-based principal components were used to control for population stratification in all analyses (Supplementary Fig. 1).

Genetic risk scores.

We used the *score* function available in PLINK to compute weighted genetic risk scores. These scores were calculated based on the observed genotypes at 93 of the genotyped Crohn's disease risk SNPs and their corresponding effect sizes reported for Caucasian population in Liu *et al.*[20].

Statistical and Bioinformatical analyses.

A detailed description of the statistical and bioinformatical methods used in this study can be found in the Supplementary Methods. Briefly, case-control DNA methylation association analysis was performed using the R package, CATE[21], which implements a state-of-the-art batch correction method to remove inflation and test-statistic bias in association tests. Linear regression was used to analyze the association of each CpG site in blood with Crohn's disease at diagnosis. Linear mixed effects models were used to analyze longitudinal changes in DNA methylation of Crohn's disease patients, and to identify methylation changes associated with CRP and PCDAI. KEGG pathway enrichment analysis was performed using missMethyl[22], an R/Bioconductor package. *Cis*- methylation quantitative trait loci (mQTL) analysis was performed using a linear mixed model implemented in GEMMA[23]. Genetic association and the concept of Mendelian randomization was used to clarify the causal versus consequential roles of methylation changes that are associated with Crohn's disease. To ascertain if peripheral blood methylation could distinguish Crohn's disease patients from non-IBD controls, we divided the baseline methylation dataset consisting of 238 subjects (164 cases and 74 controls) at random into equally weighted (cases and controls) training and testing datasets with 70% of the samples going into the training dataset. The training dataset was fit with a logistic regression model using the R package, glmnet[24], and the fitted model was used to predict the case status for the test dataset. Diagnostic accuracy was assessed via area under the receiver operator characteristic curve.

Results

Differentially methylated CpGs associated with Crohn's disease at diagnosis.

Epigenome-wide association analysis of 164 newly diagnosed, treatment naive pediatric patients with Crohn's disease (150 B1, 14 B2; cases) and 74 non-IBD controls identified 1189 CpG sites associated with Crohn's disease in blood at diagnosis (FDR < 0.05; Fig. 1 and Supplementary Table 1). Of these, 976 CpG sites (82%) had increased methylation in

cases relative to controls and 213 (18%) had decreased methylation. Because disease-associated inflammation can influence expression within a cell population and create differences in total cell composition (Supplementary Fig. 2), our analysis included covariates to adjust for estimated proportions[18] of the 6 dominant cell types (CD4$^+$ T cells, CD8$^+$ T cells, B cells, NK cells, monocytes and granulocytes) in blood (see Methods; Supplementary Fig. 3). Sensitivity analyses demonstrated that our findings are robust to disease behavior states (B1 or B2; Supplementary Fig. 4 and Supplementary Table 1) and manifestation in the bowel (L1, L2 or L3; Supplementary Fig. 5), suggesting that baseline methylomic contributions to Crohn's disease do not vary strongly by disease behavior or location. The strongest association signals were found on chromosomes 16, 17 and 19 (Fig. 1), with CpGs in a long non-coding RNA, *LOC100996291* (*LINCO1993*), showing the peak association with Crohn's disease at diagnosis (Supplementary Table 1). Apart from identifying novel CpG sites, we replicated several findings that were previously associated with Crohn's disease (including *TMEM49* (*VMP1*), *SBNO2, RPS6KA2, ITGB2*, and *TXK*)[3] (Supplementary Table 2). CpGs annotated to prominent IBD therapeutic targets such as Tumor Necrosis Factor (*TNF*), Janus Kinase 3 *(JAK3)*, Interleukin 12B *(IL12B)*, Interleukin 23 Subunit Alpha *(IL23A)*, and Interleukin 1 Receptor Type1 (*IL1R1*) were amongst the disease-associated CpGs (Supplementary Table 1). Notably, enrichment analysis of our Crohn's disease associated CpGs indicated that they are more likely to occur in gene bodies (OR = 1.67, $P < 2.2 \times 10^{-16}$) and CpG shelves (OR = 1.42, $P = 5.0 \times 10^{-4}$), and are less likely to be in gene promoters (OR = 0.35, $P = 8.2 \times 10^{-13}$ for less than 200 base pairs from transcription start site (TSS); and OR = 0.57, $P = 5.9 \times 10^{-08}$ for less than 1500 base pairs from TSS), and CpG islands (OR = 0.14, $P < 2.2 \times 10^{-16}$) and shores (OR = 0.59, $P = 9.7 \times 10^{-10}$). Exact details of the distribution of Crohn's disease associated CpGs in relation to gene regions and CpG islands are provided in Supplementary Tables 3, 4.

Gene expression profiles of differentially methylated genes in Crohn's disease.

The 1189 Crohn's disease associated CpGs mapped to 717 unique genes. To better understand how these CpGs might reflect functional processes that are perturbed during the diagnosis of Crohn's disease, we examined the expression profiles of our differentially methylated genes in blood RNA-Seq data available from an independent data set consisting of 60 newly diagnosed pediatric patients with Crohn's disease and 12 non-IBD controls[25]. Of the 585 (of 717) genes available for analysis after quality control, 162 (28%) of those were differentially expressed at FDR < 0.05 and 233 (40%) at the less stringent threshold of *P* < 0.05 (Supplementary Fig. 6 and Supplementary Table 5). Overlapped with these 233 differentially expressed genes were 295 of the Crohn's disease associated CpGs with an average of 1.3 CpG sites associated per gene (range = 1-4). As shown in Supplementary Fig. 7, the direction of effects between DNA methylation and gene expression changes in relation to Crohn's disease appears to be context dependent with some CpG methylation-gene expression probes demonstrating negative association while others showed positive relationship, irrespective of the position of the CpG site in the associated gene (Fisher's test, *P* > 0.05; Supplementary Table 6). Collectively, these observations suggest the integrative involvement of methylomic and transcriptomic processes underlying Crohn's disease pathogenesis.

Biological processes enriched in Crohn's disease associated CpGs.

Next, we evaluated whether the disease-associated CpGs in blood were enriched for biological processes relevant to Crohn's disease. Our pathway enrichment analysis identified 164 KEGG pathways that were more likely to occur in the Crohn's disease associated CpGs than would be expected by chance (FDR < 0.05; Supplementary Table 7). Among these were pathways relevant to immune function including TNF-alpha, Jak-STAT, Rap1 and PI3K-Akt signaling; and inflammation such as the IL-17 signaling pathway, cytokine-cytokine receptor interaction and chemokine signaling.

Relationship between DNA methylation signatures of Crohn's disease and inflammation.

To further evaluate the relationship between the disease-associated methylation signatures and inflammation, we tested the 1189 CpG sites for association with plasma C-reactive protein (CRP) levels, a marker of inflammation, and compared the estimated effects of methylation changes on CRP versus Crohn's disease at diagnosis. The relationship was extremely strong (R = 0.91, $P < 2.2 \times 10^{-16}$) suggesting that the methylation signatures of Crohn's disease either cause the inflammatory status of the patient, or directly result from it (Fig. 2a). 1155 (97%) of the 1189 Crohn's disease CpGs exhibited directional consistency, and 872 (73%) showed statistically significant association with CRP ($P < 0.05$; Fig. 2a and Supplementary Table 8).

Next to assess the relevance of these methylation signatures to Crohn's disease related inflammation, we compared the effect sizes of Crohn's disease associated CpGs on Crohn's disease and CRP in our dataset to a recently published meta-analysis of epigenome-wide association of CRP in subjects that were not selected for any particular disorder[26]. These meta-analyses comprised of 8863 participants that were sampled from 9 different prospective cohort studies with a wide-range of focus from cardiometabolic phenotypes to physical activity, intelligence, and aging. Surprisingly, we noted an extremely strong correlation between the estimated effects of Crohn's disease associated CpGs on Crohn's disease and chronic low-grade inflammation that is associated with a broad range of complex diseases, including diabetes and cardiovascular disease (Fig. 2b,c). To validate this inference, we examined the overlap and directional consistency of previously reported Crohn's disease CpGs[3] with the CRP meta-analysis[26], obtaining consistent results (Supplementary Table 2).

Longitudinal dynamics of DNA methylation in Crohn's disease.

In order to establish the direction of causality of this strong association, we next examined the longitudinal dynamics of inflammation and disease-associated methylation profiles during the course of the disease. Since frontline treatment of IBD attempts to lower inflammation in the patients, and as expected, CRP levels in patients at follow-up 1 to 3 years after diagnosis were dramatically lower than at diagnosis ($P = 8.4 \times 10^{-9}$; Supplementary Fig. 8), the direction of causality seems obvious. The patients received treatment known to lower inflammation, and the primary marker for inflammation was much lower. Next to assess the dynamics of DNA methylation pre- and post-treatment, we compared the methylation profiles at follow-up to the profiles at diagnosis. Here the effects at diagnosis reflect differences between newly diagnosed patients and controls, and the

effects at follow-up reflect differences in patients before and after treatment. At 1179 (99.2%) of the 1189 sites associated with Crohn's disease at diagnosis, the sign of the effect had reversed, while the magnitude of the change remained the same, generating a strong negative correlation (R = −0.93, $P < 2.2 \times 10^{-16}$; Fig. 3a and Supplementary Table 8). In fact, after treatment, methylation at these sites is largely indistinguishable in patients versus controls (Supplementary Figs. 9,10). We noted similar results even after stratifying patients based on disease progression to B2 (R = −0.91, $P < 2.2 \times 10^{-16}$ for progressors; R = −0.90, $P < 2.2 \times 10^{-16}$ for non-progressors; Fig. 3b,c). Collectively, our data establish that during the course of the disease, methylation patterns that are disrupted at the diagnosis of Crohn's disease revert back to the levels seen in non-IBD controls, irrespective of the disease behavior states (B1 or B2). Only 10 (0.8%) CpGs had the same sign of effect during diagnosis versus follow-up; these CpGs corresponded to 8 unique genes (*RORC*, *CXXC5*, *GMNN*, *GPR183*, *DIDO1*, *SMARCD3*, *ESPNL*, and *EPS8L3*; Supplementary Table 8). Interestingly, genes such as *RORC*, *SMARCD3* and *EPS8L3* have previously been linked with IBD, including in genome-wide association studies and gene expression studies[27-31]. For instance, *RORC* encodes a key transcription factor for the Th17 pathway involved in transcriptional regulation of the effector cytokines *IL17A*, *IL17F*, *IL21*, *IL22*, *IL26* and *CCL20*[32] and was previously reported to be differentially expressed in peripheral blood and intestinal Crohn's disease samples compared to healthy controls[28].

Relationship between DNA methylation signatures of Crohn's disease and disease activity.

Next, to test whether our finding of methylomic and inflammatory reversion extends to other clinical and laboratory measurements, we examined the measures of the pediatric Crohn's disease activity index (PCDAI), a multi-item index which incorporates clinical symptoms, laboratory parameters, and endoscopic findings[33] and noted higher PCDAI scores (median score of 30; *n* = 159) during diagnosis which were significantly lower during the follow-up (median score of 5; *n* = 149; $P < 2.2 \times 10^{-16}$; Supplementary Fig. 11). Following association analysis of PCDAI with the 1189 sites (*n* = 308 samples; see Methods), estimated effect sizes demonstrated a strong correlation with their estimated effects on Crohn's disease (R = 0.91, $P < 2.2 \times 10^{-16}$), suggesting a potential relationship between disease activity (based on PCDAI) and DNA methylation in blood (Supplementary Table 8 and Supplementary Fig. 12).

Role of medication in DNA methylation reversal.

To evaluate the potential impact of therapy on the reversal of Crohn's disease associated methylation signatures, we stratified patients based on the class of medications they were taking at the time of the follow-up sampling (Table 1). Comparative analysis of methylation levels in blood at the time of the follow-up did not reveal any genome-wide significant differences between subsets of patients who received biologics, immunomodulators, biologics plus immunomodulators, or other drugs, except for one CpG, cg24052338 (in the 3'UTR region of *ZNF837*), that showed significant association with other drugs (FDR < 0.05; Supplementary Tables 9-12), indicating that the medication is probably not the primary contributor to the methylomic reversion during the course of the disease. Boxplots depicting methylation beta values of follow-up patients' samples stratified based on the class of medications at the top 5 disease-associated CpGs were shown in Fig. 4. However, it is

possible that the medication-induced reductions in inflammation and consequently disease activity may account for the reversal of the disrupted methylomic signatures in blood. Consistent with our interpretation, a study of site-specific methylation differences in peripheral blood mononuclear cells[10] and a different study of 2 colonic mucosa samples[9], both showed methylomic reversion in response to treatment and/or disease remission via modulation of the disease-specific inflammatory characteristics. In contrast, another study reported stable methylation differences in patients with newly diagnosed (treatment-naïve) versus established IBD (exposed to IBD medications)[3]. However, our finding that the reversion of disease-associated methylation patterns associates with clinical characteristics of the disease (CRP, PCDAI) rather than medication underscores the importance of having prospectively followed inception cohorts with well-documented disease measures.

Understanding the causal versus consequential roles of DNA methylation in Crohn's disease.

Given the methylomic reversion occurring during the course of the disease, and its strong relationship with plasma CRP levels, it appears that Crohn's disease associated methylation signatures are tightly linked to inflammation rather than the disease development itself. However, if methylation at specific sites plays a role in disease development, their identification would provide valuable therapeutic targets. To distinguish sites that may have causal versus consequential roles in Crohn's disease, we employed the concept of Mendelian randomization as operationalized by Wahl *et al.*[34]. As shown in Supplementary Fig. 13, CpGs that emerge on the path between the instrumental variable and the outcome (Crohn's disease; Model 1), where methylation appears to mediate genetic risk of Crohn's disease, are interpreted to be causal rather than being the consequence (Model 2) of the disease. 194 out of the 1189 Crohn's disease CpGs at diagnosis associated with DNA sequence variation in a *cis* methylation quantitative trait loci (mQTL) analysis (FDR < 0.05; Supplementary Table 13). Of these, 174 CpGs with genetic data available for the associated mQTL SNPs from a large meta-analysis of genome-wide association studies (GWAS) of Crohn's disease[20] were evaluated for potential causal relationships between methylation in blood and Crohn's disease. For each CpG, we identified the most significantly associated SNP (sentinel mQTL) and applied the concept of Mendelian randomization using the sentinel mQTL SNP as the instrumental variable, CpG as a mediator, and Crohn's disease as the outcome for methylation cause of Crohn's disease (Model 1, Supplementary Fig. 13).

Causal role of DNA methylation in Crohn's disease.

Using this set of sentinel SNP-CpG pairs, we first investigated SNP to DNA methylation (β coefficient; β_1) and DNA methylation to Crohn's disease (β_2) relationships to obtain predicted effects ($\beta_1 \times \beta_2$) of the corresponding SNPs on Crohn's disease via DNA methylation (Fig. 5a). Subsequently, genetic effect sizes of SNPs on Crohn's disease were obtained from large GWAS meta-analyses of Crohn's disease[20] to assess the observed effects of genotypes at these SNPs on Crohn's disease (β_3). If methylation contributes causally to Crohn's disease, we would expect the observed effect of SNP on phenotype to be consistent, if not equivalent to its predicted effect mediated through methylation. Notably, methylation changes at 3 CpGs (cg15706657, cg23216724: near *GPR31*; and cg20406979: near *RNASET2*) showed significant causal associations with Crohn's disease at diagnosis

(FDR < 0.05; Fig. 5a and Supplementary Table 14). Consistent with the potentially causal effect, methylation levels at cg23216724 and cg20406979 became even more pronounced or remained about the same without exhibiting signs of reversion during the follow-up (Supplementary Fig. 14), supporting our inference regarding causality. Further support for their potentially causal influence is provided by the observation that all 3 CpGs are influenced by the known IBD-associated SNP, rs1819333, identified through large GWAS[20, 35]. The IBD-risk locus containing the SNP rs1819333 harbors (within 1 Mb flanking rs1819333) key genes *RPS6KA2*, *RNASET2* and *CCR6* that have previously been implicated in IBD pathology at both genomic and/or molecular levels, including in transcriptomic and epigenomic studies[3, 20, 35-37]. Although genetic variation at rs1819333 has been associated with significant risk for Crohn's disease susceptibility, underlying causal gene(s) and molecular mechanisms of this strong GWAS association are yet to be elucidated. Remarkably, all 3 identified potentially causal CpGs that are associated with rs1819333 were recently shown to causally regulate transcriptional levels of *RPS6KA2* in peripheral blood using a summary data-based Mendelian randomization (SMR) approach[38]. Taken together, these findings suggest DNA methylation as a potential mediator of genetic effects of rs1819333 on Crohn's disease, possibly through transcriptional regulation of *RPS6KA2*.

Consequential role of DNA methylation in Crohn's disease.

Conversely, to identify Crohn's disease associated sites where changes in methylation are more likely to be the consequence of the disease, we used a weighted Crohn's disease genetic risk score (see Methods) as an instrumental variable, Crohn's disease as the mediator and methylation as the outcome (Model 2, Supplementary Fig. 13). An extremely strong correlation (R = 0.86; $P < 2.2 \times 10^{-16}$) between the observed effect of the weighted genetic risk score on methylation and its predicted effect through Crohn's disease was seen (Fig. 5b). In particular, we identified 8 CpGs corresponding to 7 genes that showed significant consequential associations with Crohn's disease at diagnosis (FDR < 0.05; Fig. 5b and Supplementary Table 15). In keeping with their consequential role, methylation levels at these CpG sites demonstrated drastic changes approaching levels seen in non-IBD controls during the follow-up (Supplementary Fig. 15). Differential methylation at cg18942579: *TMEM49* and cg17501210: *RPS6KA2*, CpGs that have been consistently found to be associated with Crohn's disease[3, 8], appears to be a consequence of the disease rather than exerting causal effects. For instance, cg17501210 has previously been reported to be the top-most differentially methylated CpG site in peripheral blood of IBD patients, whose effects were (i) even more pronounced in purified CD14+ monocytes; (ii) strongly correlated with disease-relevant markers, including CRP, albumin and hemoglobin; and (iii) not influenced by treatment status. Overall, the strong correlation between the observed and predicted effects in Fig. 5b suggests that most disease-associated methylation changes are triggered by the onset of Crohn's disease. This finding is consistent with findings from other complex diseases[34, 39, 40], suggesting that only a minority of the trait-associated methylation changes are likely to exert causal effects.

Role of DNA methylation in diagnosis and prognosis of Crohn's disease.

Biological data that enable accurate diagnosis and/or prognosis of IBD has always been of considerable interest from the point of view of clinical application. In line with previous

studies[3], we noted that peripheral blood DNA methylation data could indeed distinguish patients with Crohn's disease from non-IBD controls (AUC = 0.91; Supplementary Fig. 16). However, given the non-inflammatory nature of the sampled control subjects, supplemented by our finding that the signatures of methylation observed at diagnosis of Crohn's disease capture general inflammation rather than Crohn's disease-specific, we are hesitant to propose peripheral methylation as a diagnostic biomarker for Crohn's disease based on the prevailing evidence. Future studies of side-by-side evaluation of methylation data from patients with different immune-mediated inflammatory diseases along with the disease-relevant tissue-specific inflammatory characterization are required to definitively establish the diagnostic potential conferred by methylation signatures of complex diseases.

Next, to assess the utility of methylation in prognosis, by stratifying Crohn's disease patients based on subsequent progression to complicated disease (see Methods), we asked if methylation signatures at diagnosis could predict who would in time progress to complicated Crohn's disease. In keeping with our finding from Supplementary Fig. 4 and Supplementary Table 1, we did not find any CpGs showing significant differences when the baseline methylation profiles of subsequent progressors were compared to non-progressors. Taken together, our data suggests that peripheral blood methylation profiles do not predict or change in relevance to the evolution or presence of Crohn's disease complications.

Conclusion

In conclusion, we characterized temporal relationships connecting methylation changes in blood with varying inflammatory characteristics at diagnosis and during treatment for Crohn's disease in children. Systemic inflammation has long been understood to be a pathogenetic hallmark of Crohn's disease, and medication to relieve the burden of inflammation has been part of all frontline treatment strategies for managing the disease. Our results provide convincing evidence that the signatures of methylation observed at diagnosis accompany acute inflammation that declines with treatment, but revert toward the levels seen in non-IBD controls despite ongoing bowel disease, arguing that they are primarily a symptom of the disease rather than a cause. If so, treatment of inflammation signatures may fundamentally be treating the symptoms of Crohn's disease rather than the etiology, partially explaining why IBD often remains a life-long remitting and relapsing disorder, despite effective treatment of the inflammation symptoms.

A caveat to this interpretation is that we measured circulating immune cells whereas the inflammation is manifest in the bowel. Data for gut-resident immune cells will be required to establish whether the methylomic reversion we describe is also observed in the gut and affected by the diverse treatment regimens independent of the epithelial signature. By contrast, a recent study[12] of methylation in intestinal epithelial cells described a distinct IBD profile more related to disease than inflammation, which was stable over time in the 23 patients examined, further suggesting the need for new drugs that treat the cells in which, persistent molecular changes underlie disease pathogenesis. Future epigenetic studies of IBD could profile circulating, gut-resident immune, and gut epithelial cells in parallel, using our framework to definitively identify causal CpGs, and leverage the epigenome for the development of targeted therapeutics. This, in combination with the current armamentarium

of IBD medications that hold promise for successfully managing the disease by targeting the immune system, may put us one step closer to sustained remission and mucosal healing.

One of the long-term complications of IBD is inflammation-related cancers[41-44], and our finding that aberrant DNA methylation of Crohn's disease is predominantly a consequence of inflammation provides a strong rationale for the molecular link between IBD and cancer, as both chronic inflammation and aberrant DNA methylation having a known role in malignancy development. It remains to be seen whether these methylation signatures detected in blood as a consequence of inflammation may in part predict new onset, incident cancer, a major clinical consequence associated with IBD.

Our study has certain limitations. Although, it was well powered to detect CpGs that are different between non-IBD controls and newly-diagnosed Crohn's disease patients, and examine how they change during the course of the disease, we were limited in terms of power to apply a Mendelian randomization framework to infer causal associations, and hence, we may have missed detecting some CpGs with a potential causal influence. Despite this limitation, we identified differential methylation of several CpG sites associated with an IBD-associated SNP, rs1819333, to be potentially causal to Crohn's disease. This result, however, should be interpreted with caution given the lack of replication. Nevertheless, Mendelian randomization revealed results consistent with findings from our longitudinal framework that the peripheral blood methylation changes associated with Crohn's disease in children are predominantly a consequence of disease. Causal versus consequential analyses of adult cohorts should confirm the potential impact of blood-derived DNA methylation or the lack thereof in adult patients diagnosed with Crohn's disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Van Limbergen J, Russell RK, Drummond HE, et al. Definition of phenotypic characteristics of childhood-onset inflammatory bowel disease. Gastroenterology 2008; 135:1114–22. [PubMed: 18725221]

2. Ruel J, Ruane D, Mehandru S, et al. IBD across the age spectrum: is it the same disease? Nat Rev Gastroenterol Hepatol 2014;11:88–98. [PubMed: 24345891]

3. Ventham NT, Kennedy NA, Adams AT, et al. Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. Nat Commun 2016;7:13507. [PubMed: 27886173]

4. Karatzas PS, Gazouli M, Safioleas M, et al. DNA methylation changes in inflammatory bowel disease. Ann Gastroenterol 2014;27:125–132. [PubMed: 24733658]

5. McDermott E, Ryan EJ, Tosetto M, et al. DNA Methylation Profiling in Inflammatory Bowel Disease Provides New Insights into Disease Pathogenesis. J Crohns Colitis 2016;10:77–86. [PubMed: 26419460]

6. Li Yim AYF, Duijvis NW, Zhao J, et al. Peripheral blood methylation profiling of female Crohn's disease patients. Clin Epigenetics 2016;8:65. [PubMed: 27279921]

7. Nimmo ER, Prendergast JG, Aldhous MC, et al. Genome-wide methylation profiling in Crohn's disease identifies altered epigenetic regulation of key host defense mechanisms including the Th17 pathway. Inflamm Bowel Dis 2012;18:889–99. [PubMed: 22021194]

8. Adams AT, Kennedy NA, Hansen R, et al. Two-stage genome-wide methylation profiling in childhood-onset Crohn's Disease implicates epigenetic alterations at the VMP1/MIR21 and HLA loci. Inflamm Bowel Dis 2014;20:1784–93. [PubMed: 25144570]

9. Harris RA, Nagy-Szakal D, Mir SA, et al. DNA methylation-associated colonic mucosal immune and defense responses in treatment-naive pediatric ulcerative colitis. Epigenetics 2014;9:1131–7. [PubMed: 24937444]

10. Harris RA, Nagy-Szakal D, Pedersen N, et al. Genome-wide peripheral blood leukocyte DNA methylation microarrays identified a single association with inflammatory bowel diseases. Inflamm Bowel Dis 2012;18:2334–41. [PubMed: 22467598]

11. Taman H, Fenton CG, Hensel IV, et al. Genome-wide DNA Methylation in Treatment-naive Ulcerative Colitis. J Crohns Colitis 2018;12:1338–1347. [PubMed: 30137272]

12. Howell KJ, Kraiczy J, Nayak KM, et al. DNA Methylation and Transcription Patterns in Intestinal Epithelial Cells From Pediatric Patients With Inflammatory Bowel Diseases Differentiate Disease Subtypes and Associate With Outcome. Gastroenterology 2018;154:585–598. [PubMed: 29031501]

13. Kugathasan S, Denson LA, Walters TD, et al. Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. Lancet 2017;389:1710–1718. [PubMed: 28259484]

14. Farh KK, Marson A, Zhu J, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 2015;518:337–43. [PubMed: 25363779]

15. Barfield RT, Kilaru V, Smith AK, et al. CpGassoc: an R function for analysis of DNA methylation microarray data. Bioinformatics 2012;28:1280–1. [PubMed: 22451269]

16. McCartney DL, Walker RM, Morris SW, et al. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. Genom Data 2016;9:22–4. [PubMed: 27330998]

17. Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics 2013;29:189–96. [PubMed: 23175756]

18. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 2012; 13:86. [PubMed: 22568884]

19. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75. [PubMed: 17701901]

20. Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet 2015;47:979–986. [PubMed: 26192919]

21. Wang J, Zhao Q, Hastie T, et al. Confounder adjustment in multiple hypothesis testing. arXiv 2015.

22. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics 2016;32:286–8. [PubMed: 26424855]

23. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet 2012;44:821–4. [PubMed: 22706312]

24. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33:1–22. [PubMed: 20808728]

25. Mo A, Marigorta UM, Arafat D, et al. Disease-specific regulation of gene expression in a comparative analysis of juvenile idiopathic arthritis and inflammatory bowel disease. Genome Med 2018;10:48. [PubMed: 29950172]

26. Ligthart S, Marzi C, Aslibekyan S, et al. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. Genome Biol 2016; 17:255. [PubMed: 27955697]

27. Bogaert S, Laukens D, Peeters H, et al. Differential mucosal expression of Th17-related genes between the inflamed colon and ileum of patients with inflammatory bowel disease. BMC Immunol 2010; 11:61. [PubMed: 21144017]

28. Granlund A, Flatberg A, Ostvik AE, et al. Whole genome gene expression meta-analysis of inflammatory bowel disease colon mucosa demonstrates lack of major differences between Crohn's disease and ulcerative colitis. PLoS One 2013;8:e56818. [PubMed: 23468882]

29. Sipos F, Galamb O, Wichmann B, et al. Peripheral blood based discrimination of ulcerative colitis and Crohn's disease from non-IBD colitis by genome-wide gene expression profiling. Dis Markers 2011;30:1–17. [PubMed: 21508504]

30. Kang J, Kugathasan S, Georges M, et al. Improved risk prediction for Crohn's disease with a multi-locus approach. Hum Mol Genet 2011;20:2435–42. [PubMed: 21427131]

31. de Lange KM, Moutsianas L, Lee JC, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet 2017;49:256–261. [PubMed: 28067908]

32. Maddur MS, Miossec P, Kaveri SV, et al. Th17 cells: biology, pathogenesis of autoimmune and inflammatory diseases, and therapeutic strategies. Am J Pathol 2012;181:8–18. [PubMed: 22640807]

33. Hyams J, Markowitz J, Otley A, et al. Evaluation of the pediatric crohn disease activity index: a prospective multicenter experience. J Pediatr Gastroenterol Nutr 2005;41:416–21. [PubMed: 16205508]

34. Wahl S, Drong A, Lehne B, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature 2017;541:81–86. [PubMed: 28002404]

35. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012;491: 119–24. [PubMed: 23128233]

36. Marigorta UM, Denson LA, Hyams JS, et al. Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's disease. Nat Genet 2017;49:1517–1521. [PubMed: 28805827]

37. Hannon E, Weedon M, Bray N, et al. Pleiotropic Effects of Trait-Associated Genetic Variation on DNA Methylation: Utility for Refining GWAS Loci. Am J Hum Genet 2017;100:954–959. [PubMed: 28528868]

38. Wu Y, Zeng J, Zhang F, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. Nat Commun 2018;9:918. [PubMed: 29500431]

39. Mendelson MM, Marioni RE, Joehanes R, et al. Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease: A Mendelian Randomization Approach. PLoS Med 2017;14:e1002215. [PubMed: 28095459]

40. Elliott HR, Shihab HA, Lockett GA, et al. Role of DNA Methylation in Type 2 Diabetes Etiology: Using Genotype as a Causal Anchor. Diabetes 2017;66:1713–1722. [PubMed: 28246294]

41. Scharl S, Barthel C, Rossel JB, et al. Malignancies in Inflammatory Bowel Disease: Frequency, Incidence and Risk Factors-Results from the Swiss IBD Cohort Study. Am J Gastroenterol 2018.

42. Choi CR, Bakir IA, Hart AL, et al. Clonal evolution of colorectal cancer in IBD. Nat Rev Gastroenterol Hepatol 2017;14:218–229. [PubMed: 28174420]

43. Peneau A, Savoye G, Turck D, et al. Mortality and cancer in pediatric-onset inflammatory bowel disease: a population-based study. Am J Gastroenterol 2013;108:1647–53. [PubMed: 23939626]

44. Aardoom MA, Linda Joosse ME, de Vries ACH, et al. Malignancy and Mortality in Pediatric-onset Inflammatory Bowel Disease: A Systematic Review. Inflamm Bowel Dis 2018;24:732–741. [PubMed: 29522170]

Need to Know

Background and Context: Site-specific changes in DNA methylation in blood and intestinal mucosal samples have been associated with inflammatory bowel diseases such as Crohn's disease. However, it was not clear if these are markers of inflammation or disease development or progression.

New Findings: In an analysis of DNA methylation patterns in blood samples from pediatric patients with Crohn's, we found most methylation changes to associate with inflammation, and resolve with treatment of inflammation, irrespective of disease progression or treatment.

Limitations: We analyzed DNA methylation patterns of circulating immune cells, rather than intestinal tissues.

Impact: Changes in methylation patterns of DNA from blood samples from children with Crohn's disease associate with inflammation, and are less likely to contribute to disease development or progression. Studies of mucosal biopsy-derived immune and epithelial samples are needed to support these findings.
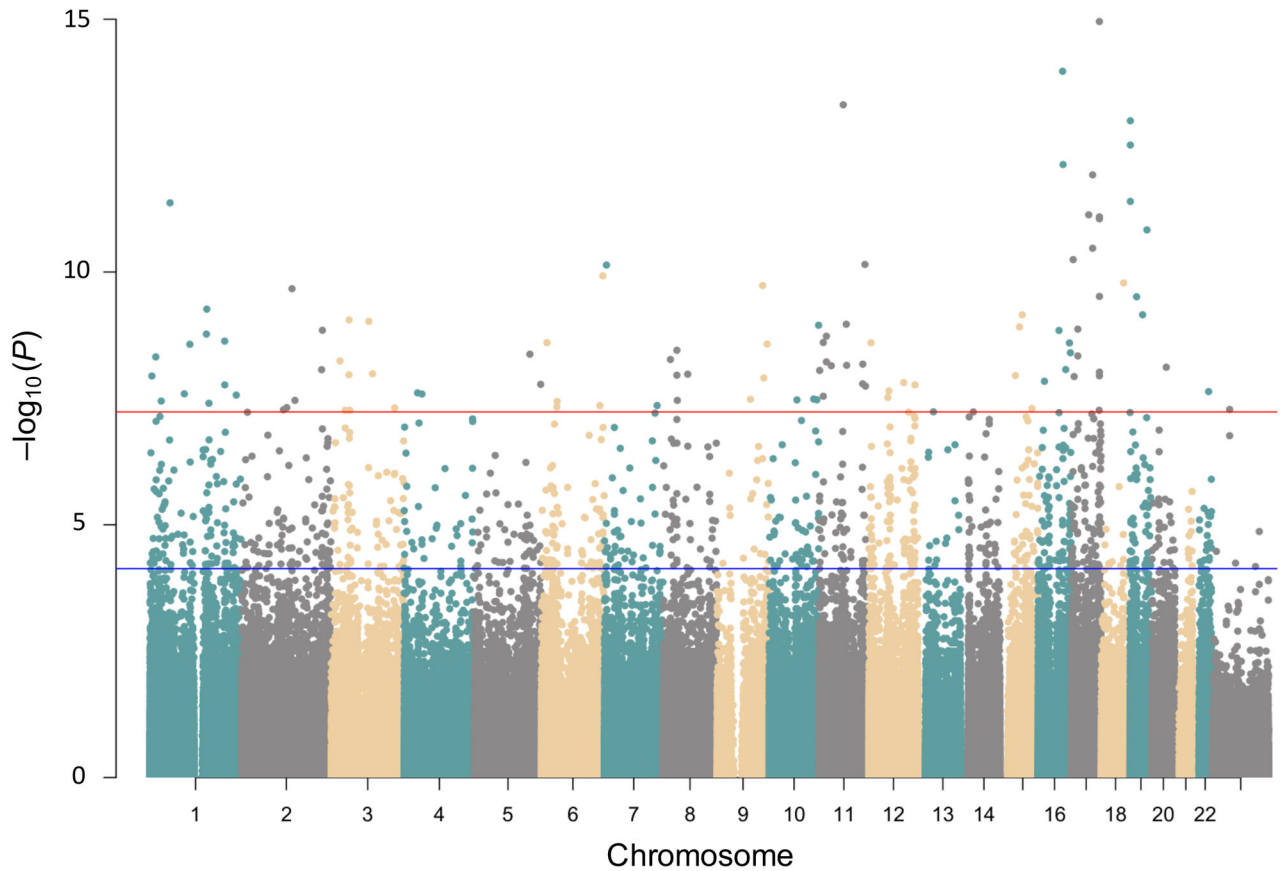
Figure 1:

Crohn's disease at diagnosis is associated with methylation changes at 1189 CpG sites in blood. All the ~850 K CpG sites represented by dots are ordered by genomic position per chromosome (x axis). P values (-$\log_{10}$) of site-specific association with Crohn's disease is shown on y axis. Dots above the blue line represent CpGs reaching epigenome-wide significance (FDR < 0.05). Dots above the red line represent CpGs reaching epigenome-wide significance after Bonferroni correction ($n = 114$ CpGs).
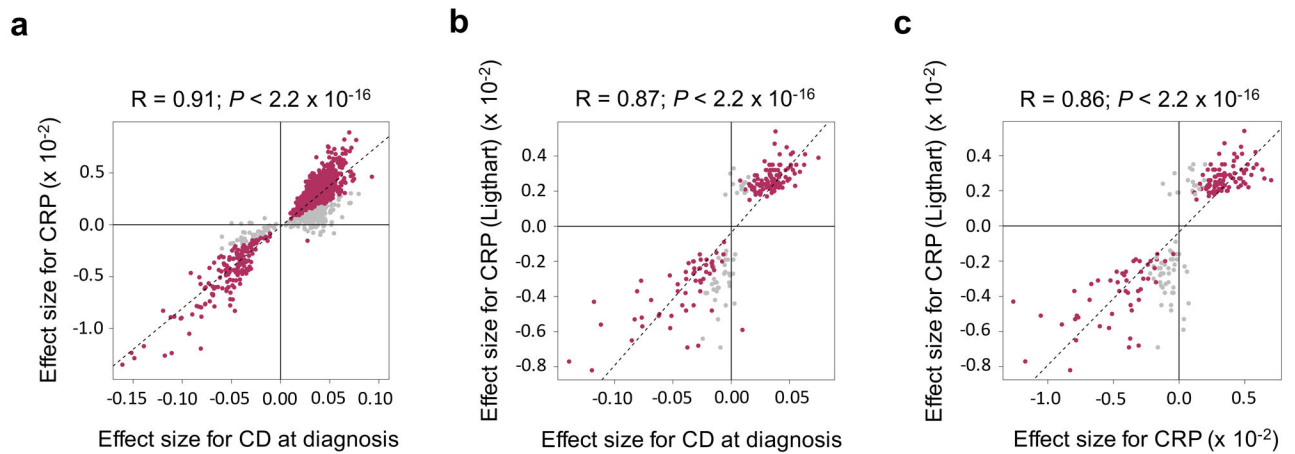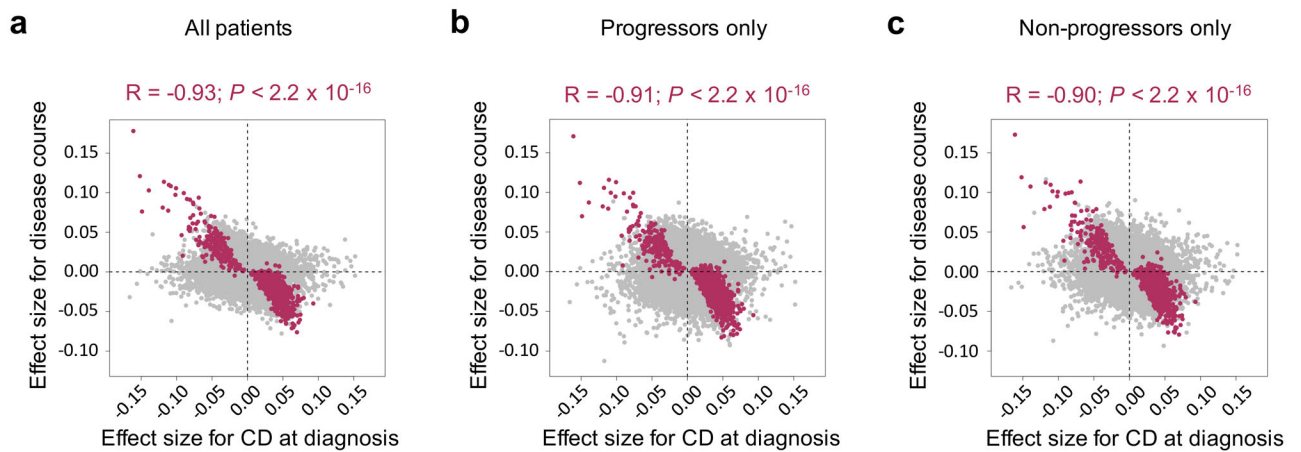
Figure 2:

Methylation signatures of Crohn's disease reflect inflammatory status of the patient. (**a**) For the 1189 Crohn's disease associated CpGs, estimated effects (n = 164 cases and 74 controls) on Crohn's disease at diagnosis (x axis) are strongly correlated with the estimated effects (n = 272: 45 non-IBD, 132 at diagnosis and 95 follow-up samples) on plasma CRP levels within the same subjects (y axis). Maroon dots represent Crohn's disease CpGs that showed significance with plasma CRP (P < 0.05). (**b,c**) At the 206 (of 218) CRP-associated CpGs in the latest meta-analysis (n = 8863) by the Ligthart *et al.*[26] (y axis), (**b**) 199 had effects on Crohn's disease at diagnosis, and (**c**) 196 had effects on CRP, in the same direction in our data. Maroon dots are CpGs from[26] that showed significance with (**b**) CD and (**c**) CRP in our data (P < 0.05).

Figure 3:
Disrupted methylation patterns during the diagnosis of Crohn's disease revert back during the course of the disease. (**a**) Estimated methylation differences between diagnosis ($n = 164$) and follow-up ($n = 164$; y axis) were of similar magnitude to baseline differences between cases ($n = 164$) and controls ($n = 164$; x axis). All ~850K CpG sites are shown; the correlation coefficient and P value is for the 1189 Crohn's disease CpGs that are colored maroon. (**b**) Same comparison for patients ($n = 55$) who received an initial diagnosis of B1 at the time of diagnosis and progressed to B2 during the course of the follow-up period. (**c**) Same comparison for patients ($n = 95$) who started and remained as B1 during diagnosis and follow-up.
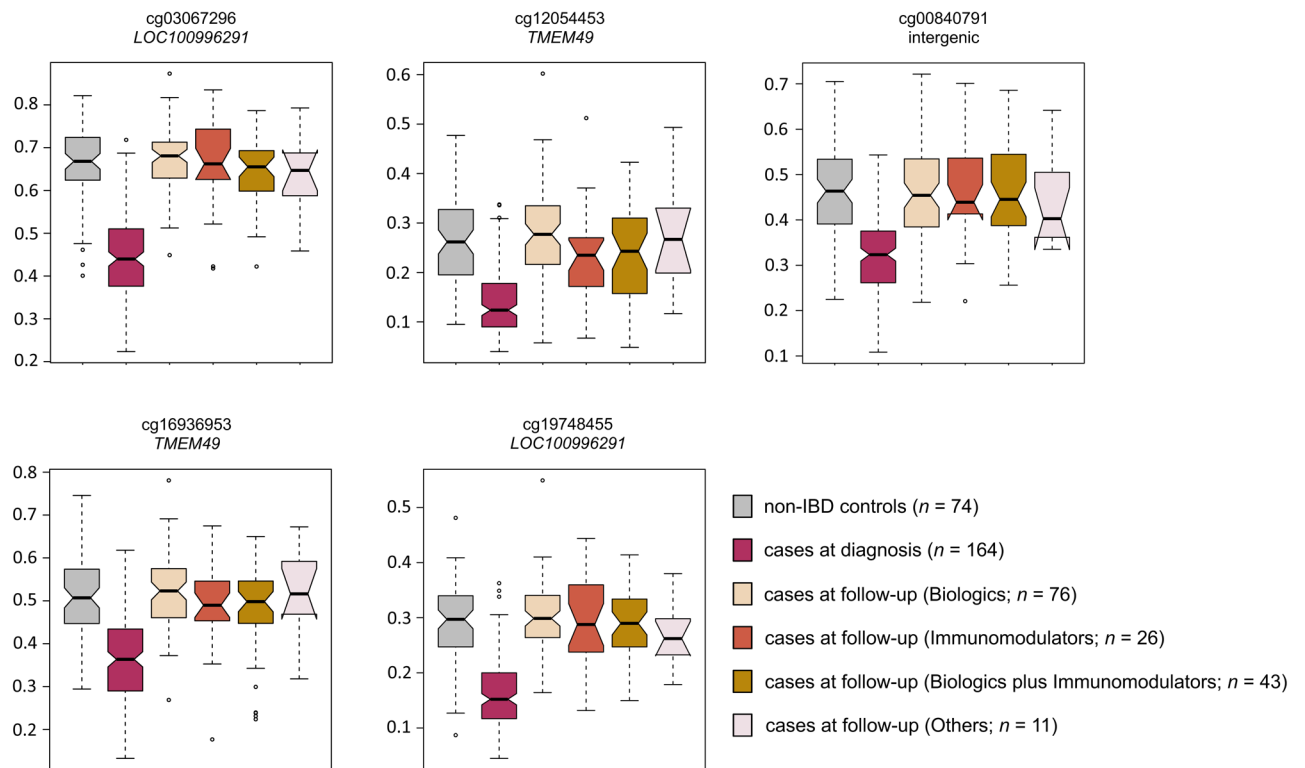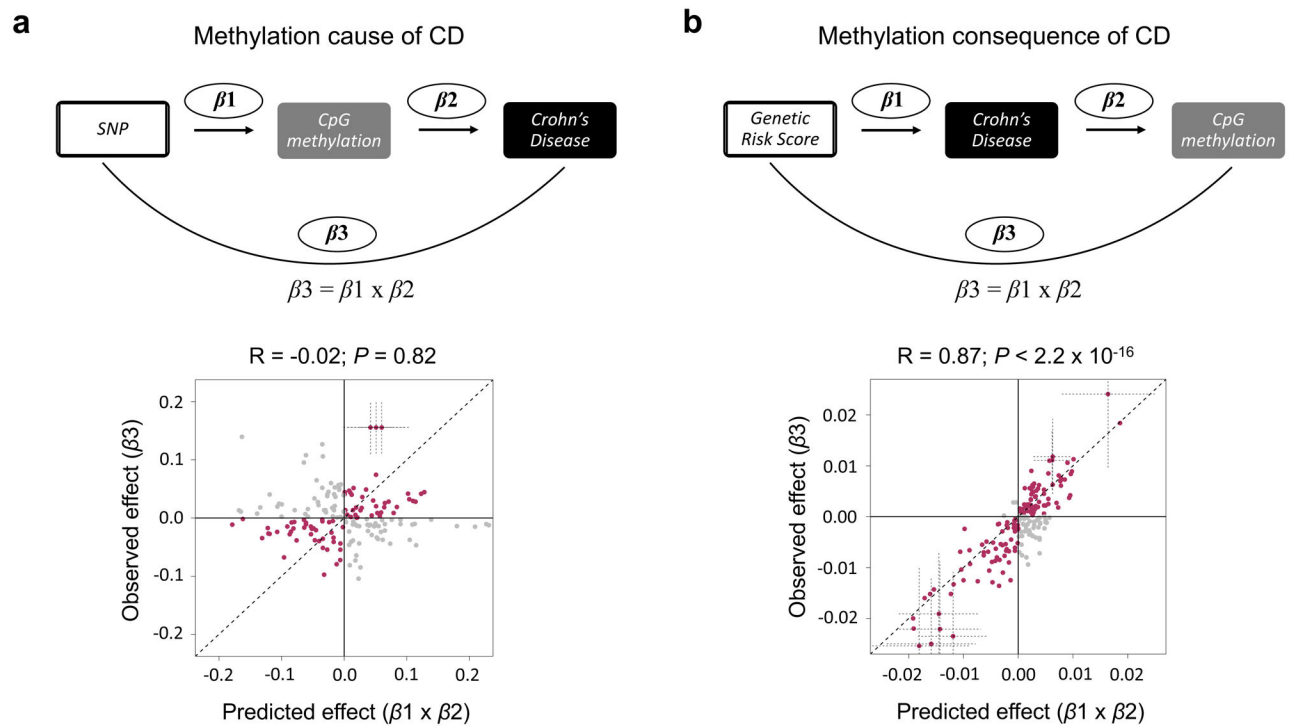
Figure 4:
Boxplots demonstrating the impact of the class of medications on methylation beta values at the top-five Crohn's disease associated CpGs during follow-up. Methylation levels for the same CpGs in non-IBD controls and Crohn's disease patients at diagnosis are also shown.

a Methylation cause of CD

$\beta 3 = \beta 1 \times \beta 2$

R = -0.02; *P* = 0.82

b Methylation consequence of CD

$\beta 3 = \beta 1 \times \beta 2$

R = 0.87; *P* < 2.2 x 10^{-16}

Figure 5:

Evaluation of directionality among Crohn's disease associated CpG sites. Schematic diagram and results of genetic association and the concept of Mendelian randomization framework implemented to clarify the causal versus consequential role of Crohn's disease associated methylation changes in blood. (**a**) Overall strength of causality of the 174 CpGs tested for methylation cause of Crohn's disease is inferred from the correlation between the observed (*y* axis) and predicted effects (*x* axis) of SNP on Crohn's disease. To infer causality of individual CpG sites, the association of the sentinel mQTL with Crohn's disease should be significant (FDR < 0.05). 95% CI error bars are shown for the 3 CpG sites with an associated mQTL that also associated with Crohn's disease. 82 of the 174 CpGs that demonstrated directional consistency are shown in maroon; CpGs that are directionally inconsistent with the observed versus predicted effects are shown in grey. (**b**) Observed effect of Crohn's disease genetic risk score on methylation (*y* axis) is highly correlated with its predicted effect (*x* axis) through Crohn's disease, suggesting a strong consequential signal at the 194 CpG sites investigated. 95% CI error bars are shown for 8 CpGs demonstrating statistically significant consequential association (FDR < 0.05). 142 of the 194 CpGs with directionally consistent effects are shown in maroon; CpGs that are directionally inconsistent are shown in grey.

Table 1:

Summary of patient characteristics

	non-IBD controls	Crohn's disease at diagnosis	Crohn's disease at follow-up	P value; Crohn's disease Vs non-IBD	P value; Crohn's disease at diagnosis Vs follow-up
Number of samples	74	164	164		
Age in yrs [median (IQR)]	12.6 (9.8 to 14.3)	12.6 (10.6 to 15.0)	15.2 (13.2 to 17.3)	0.237	2.5×10^{-14}
Female sex (%)	34 (46%)	68 (41%)	68 (41%)	0.523	
Disease state [B1/B2] *	0/0	150/14	95/69		
Disease location [L1/L2/L3/missing] *		39/39/69/17	31/21/95/17		
Number of samples with CRP data	45	132	95		
CRP$_{mgPerL}$ [median (IQR)]	0.5 (0.1 to 3.4)	4.4 (1.5 to 14.5)	1.0 (0.5 to 3.0)	0.037	0.0001
Number of samples with CRP < 1/CRP > 1	30/15	25/107	52/43		
Number of samples with PCDAI data	NA	159	149		
PCDAI [median (IQR)]	NA	30.0 (20.0 to 42.5)	5.0 (0.0 to 15.0)		2.2×10^{-16}
Number of samples with PCDAI 10/PCDAI > 10	NA	19/140	99/50		
Treatment Naïve (number of subjects)	74	164	0		
Biologics	NA	NA	76		
Immunomodulators	NA	NA	26		
Biologics plus immunomodulators	NA	NA	43		
Others *	NA	NA	11		
Medication data missing	NA	NA	8		

*
In the Montreal classification of Crohn's disease behavior, B1 corresponds to inflammatory behavior with no stricturing or luminal penetrating complications and B2 to stricturing behavior with no luminal penetrating complications. Similarly, for disease location, L1 corresponds to disease located in the ileum, L2 in the colon, and L3 ileocolon. Others include patients who received 5-ASA, Steroids, and/or Antibiotics.