

RESEARCH

Open Access



Longitudinal DNA methylation profiling of the rectal mucosa identifies cell-specific signatures of disease status, severity and clinical outcomes in ulcerative colitis cell-specific DNA methylation signatures of UC

Suresh Venkateswaran¹, Hari K. Somineni¹, Jason D. Matthews¹, Varun Kilaru², Jeffrey S. Hyams³, Lee A. Denson⁴, Richard Kellamayer⁵, Greg Gibson⁶, David J. Cutler⁷, Karen N. Conneely^{7,8†}, Alicia K. Smith^{2,8,9†} and Subra Kugathasan^{1,7,8*†}

Abstract

Background In peripheral blood, DNA methylation (DNAm) patterns in inflammatory bowel disease patients reflect inflammatory status rather than disease status. Here, we examined DNAm in diseased rectal mucosa from ulcerative colitis (UC) patients, focusing on constituent cell types with the goal of identifying therapeutic targets for UC other than the immune system. We profiled DNAm of rectal mucosal biopsies of pediatric UC at diagnosis ($n = 211$) and non-IBD control ($n = 85$) patients and performed epigenome-wide association studies (EWAS) of specific cell types to understand DNAm changes in epithelial, immune and fibroblast cells across disease states, course, and clinical outcomes. We also examined longitudinal analysis on follow-up samples ($n = 73$), and comparisons were made among patients with clinical outcomes including those undergoing colectomy versus those who did not. Additionally, we included RNA-seq from the same subjects to assess the impact of CpG sites on the transcription of nearby genes during the disease course.

Results At diagnosis, UC rectal mucosa exhibited a lower proportion of epithelial cells and fibroblasts, and higher proportion of immune cells, in conjunction with variation in the DNAm pattern. While treatment had significant effects on the methylation signature of immune cells, its effects on fibroblasts and epithelial cells were attenuated. Individuals who required colectomy exhibited cell composition and DNAm patterns at follow-up more similar to disease onset than patients who did not require colectomy. Combining these results with gene expression profiles, we identify CpG sites whose methylation patterns are most consistent with a contribution to poor disease outcomes and could thus be potential therapeutic targets.

[†]Equal contribution for last authorship: Karen N. Conneely, Alicia K. Smith and Subra Kugathasan.

*Correspondence:

Subra Kugathasan
skugath@emory.edu

Full list of author information is available at the end of the article

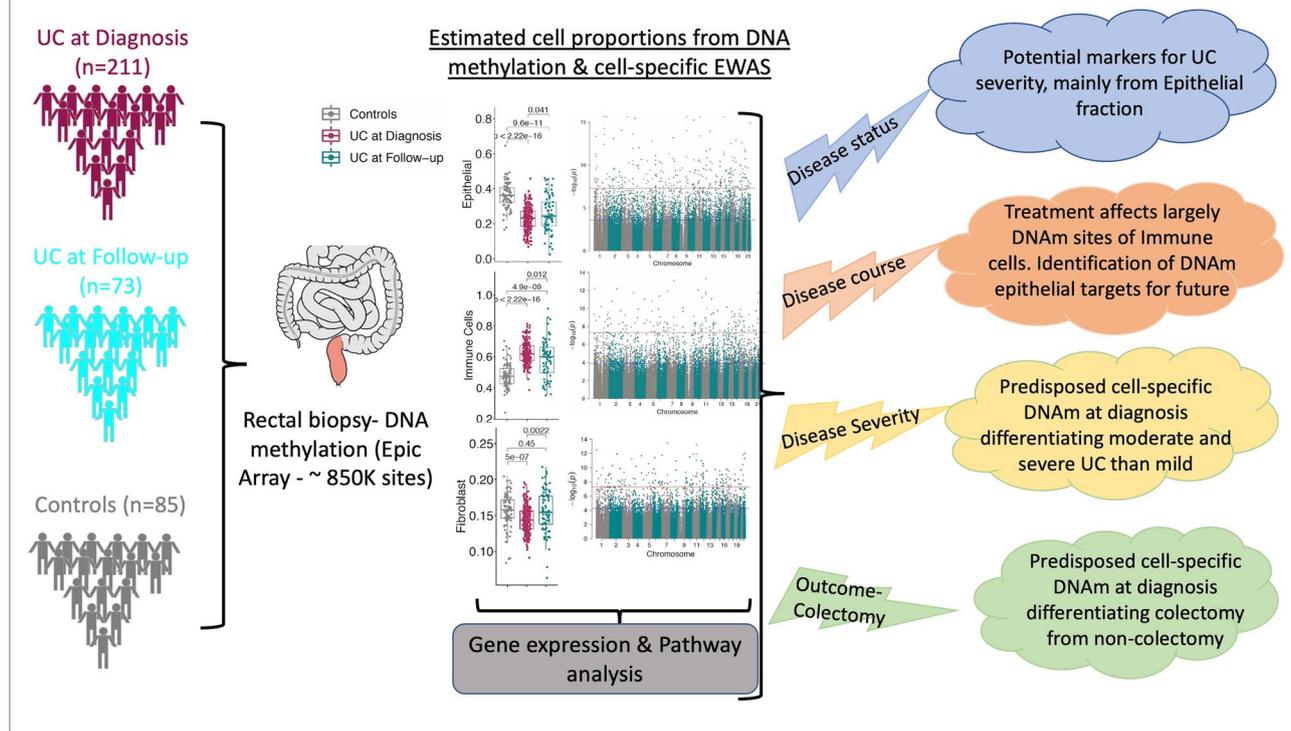


© Crown 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusions Cell-specific epigenetic changes in the rectal mucosa in UC are associated with disease severity and outcome. Current therapeutics may more effectively target the immune than the epithelial and fibroblast compartments.

Keywords Children, Epigenetic changes, Inflammatory bowel disease, Predicting Response to Standardized Pediatric Colitis Therapy (PROTECT) study, Risk Stratification and Identification of Immunogenetic and Microbial Markers of Rapid Disease Progression in Children with Crohn's Disease (RISK) study

Graphical abstract



Background

UC is a form of inflammatory bowel disease that affects an estimated 1% of the population in North America and Europe [1]. The chronic inflammation is limited to colon in UC, and is usually remitting and relapsing in nature, but repeated inflammation of the colon invariably results in progressive tissue damage. By nature, UC subjects exhibit evidence of systemic inflammation during the active disease, and ~10% of all UC subjects also have involvement of extra-intestinal manifestations [2]. Although UC is heterogeneous in disease course and outcome, the need for colectomy shortly after diagnosis is the least favorable outcome. Genome-wide association studies have proven the role of immune-associated genetic variants beyond doubt in UC [3–5], but in total these variants have accounted for less than 10% of the disease susceptibility, and an even smaller proportion of disease outcome. So far, these genetic studies have pointed to few influences of disease beyond the genetic regulation of the immune system itself. Although several

drugs are already available to target the immune system in UC, and more immune targeting therapies are being developed, nearly half of UC patients show insufficient response to these therapies [6, 7]. Thus, we asked if it is possible to identify therapeutic targets for UC other than the immune system itself.

Since genetics itself explained only a small portion of IBD, we turned to other molecular systems implicated in UC pathogenesis [8]. Recent studies in IBD demonstrated that DNAm signatures [9–11] have been associated with CD and UC phenotypes [12–17]. However, most studies have focused on study of peripheral blood rather than the diseased tissue. In our recent study we found that DNAm changes in peripheral blood cells associate primarily with inflammatory status rather than disease status [18], and that methylation patterns in blood largely return to “normal” after anti-inflammatory treatment, regardless of the underlying disease state. Blood-based DNAm studies have underlined the potential role of epigenetics mechanisms in IBD, but to find

cells whose molecular signatures might better reflect the disease state, we move to the location of disease itself. Here we examine genome-wide DNAm of the rectal tissue in an inception cohort of UC at 2 time points, once at diagnosis (treatment naïve) and subsequently at follow-up, to explore how longitudinal DNAm associates with disease onset, disease progression and outcome. Similar studies have been done only in small sets of patients, and in select cellular compartments such as purified epithelial cells [19, 20].

Site-specific DNAm differences have been reported in IBD from peripheral blood [18], and intestinal biopsies [19–21], but the analysis on the source of the cell types that are driving those signals, the temporal relationship between DNAm, disease, and the most unfavorable clinical outcomes (e.g., colectomy status) were not studied for therapeutic benefits. Thus, in this study we conducted a cell type-specific EWAS of DNA methylation (~850,000 sites) changes in the rectal mucosa at diagnosis, follow-up, and across disease phenotypes and clinical outcome trajectories (colectomy and mucosal healing) using rectal mucosal biopsies collected for the RISK and PROTECT pediatric CD and UC inception cohorts [22–24]. Our analysis examined interactions between DNAm-based estimates of proportions of three major cell components of intestinal tissues—epithelial, immune cells and fibroblasts (as a measure of mesenchyme) – and DNAm to identify cell-specific differential DNAm patterns and associated gene expression from the same patients to evaluate disease status, disease course, disease severity, and colectomy status as clinical outcomes [25] with the goal of finding patterns consistent with the cause of disease severity (rather than consequence) to serve as potential targets for molecular therapies.

Results

Altered DNAm in the epithelial, immune and fibroblast compartments are associated with UC at diagnosis

We used rectal mucosal biopsies to profile DNAm changes associated with UC. Principal component (PC) analysis of DNAm levels at ~820 K CpG sites in the mucosa showed separate clusters for UC at diagnosis ($n=211$) and controls ($n=85$) (Additional file 1: Fig. S1). We observed that PC1 and PC2 explained 20.1% and 8% of the variance in DNAm, respectively. Thus, we performed a traditional EWAS and identified 99,989 DNAm sites associated with UC at diagnosis (FDR < 0.05; Additional file 1: Fig. S2).

Identification of this large number of sites may reflect inflammation or unmeasured cellular heterogeneity, but is consistent with the high proportion of variation explained in our PC analysis, and suggests the possibility

of a strong but complex DNAm signature of UC. To better understand what drives these large-scale differences in DNAm, we next decomposed our bulk signatures into constituent cell-type proportions (epithelial, immune, and fibroblast cells) [26].

Figure 1a shows estimated cellular proportions based on DNAm signatures in UC at diagnosis vs. control samples for three primary cell types (epithelial, immune, and fibroblast cells). The estimated cell proportions from mucosal DNAm profiles shows a decrease in the proportion of epithelial cells ($p < 2.2e-16$) as well as fibroblasts ($p = 5e-07$) and increasing proportions of immune cells ($p < 2.2e-16$), changes that are consistent with damaged and inflamed mucosa [27–29]. Cell-specific EWAS between UC at diagnosis ($n=211$) and controls ($n=85$) revealed 3504 (Fig. 1b top panel; Additional file 2: Table S2a), 2279 (Fig. 1b middle panel and Additional file 3: Table S3a) and 910 (Fig. 1c bottom panel and Additional file 4: Table S4a) differentially methylated CpG sites in epithelial, immune and fibroblast cells, respectively at FDR < 0.05. Overall, these data suggest the possibility that cell-type-specific changes in DNAm in the rectal mucosa can be leveraged to differentiate diseased tissue from healthy individuals as well as and behavior during disease.

Regulatory potential of the differentially methylated CpG sites in UC

To gain insight into the regulatory potential of our differentially methylated CpG sites, we identified the nearest genes annotated to these CpGs in the Illumina manifest file. Differentially methylated CpG sites in epithelial ($n=3504$ CpGs), immune ($n=2279$ CpGs) and fibroblast ($n=910$ CpGs) cells were annotated to 1932, 1338 and 561 protein-coding genes, respectively. We first examined the association between the differentially methylated CpG sites and the gene expression levels of their corresponding genes in the rectal mucosa (CpG-gene pairs) in a subset of individuals ($n=119$) with both DNAm and gene expression data, and identified 1046 unique Gene-CpG associations for epithelial cells (Additional file 2: Table S2b), 537 unique Gene-CpG associations for immune cells (Additional file 3: Table S3b) and 272 unique Gene-CpG associations for fibroblasts (Additional file 4: Table S4b) at FDR < 0.05. For these sets of genes, we next performed differential expression analysis comparing expression in rectal mucosa biopsies taken from patients with UC at diagnosis ($n=206$) vs. biopsies taken from non-IBD individuals ($n=20$). 885 (63%) of epithelial-specific differentially methylated CpG-associated genes were associated with UC at diagnosis (FDR < 0.05; (Additional file 2: Table S2c). Similarly, 442 (81%) of immune-specific CpG-associated genes (Additional file 3: Table S3c) and 217 (80%) of fibroblast-specific CpG-associated genes (Additional file 3:

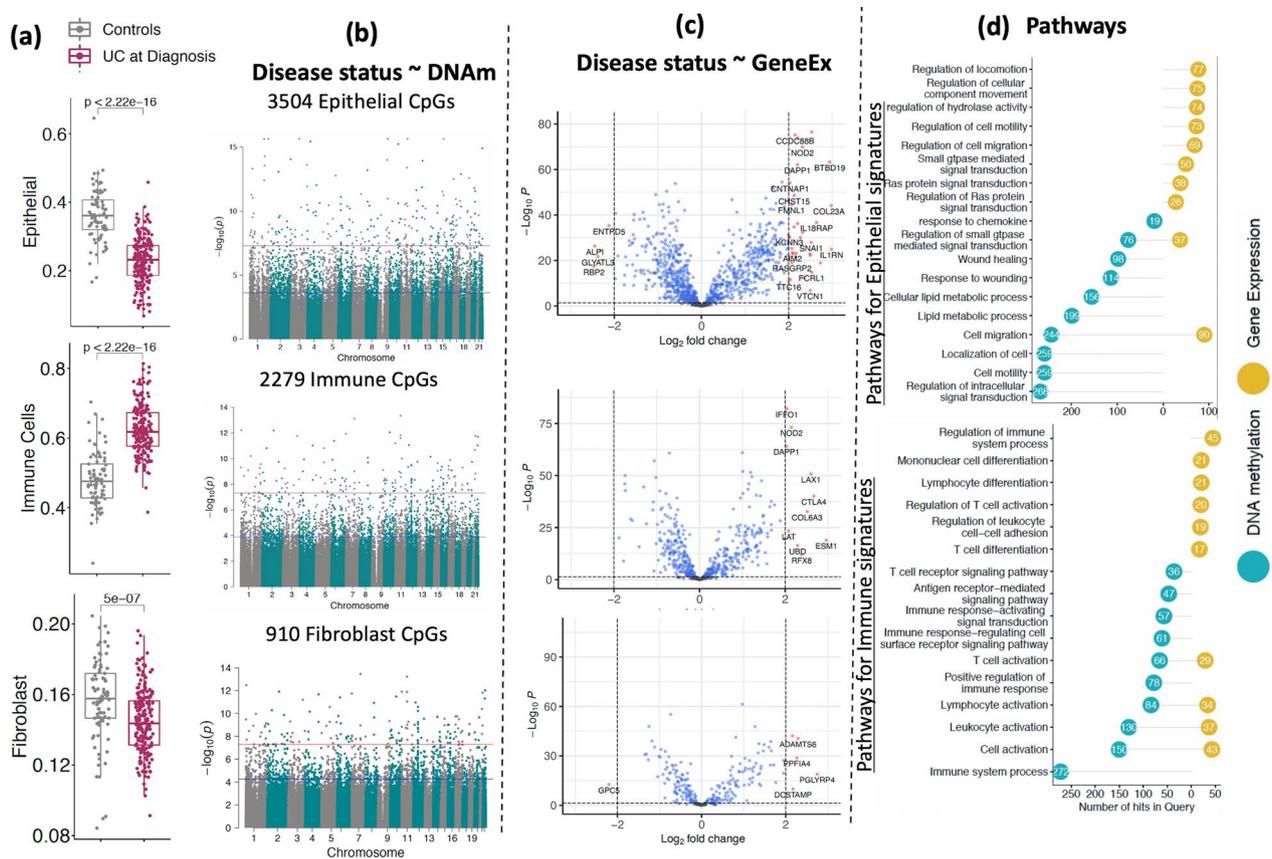


Fig. 1 DNAm and corresponding gene signatures associated with UC at diagnosis: **a** Comparison of estimated cell proportions for epithelial, immune cells and fibroblasts between rectal biopsies from UC patients (at diagnosis) and controls. *p* Values are shown from the Wilcoxon test. **b** Cell-specific epigenome-wide analysis (EWAS) comparing UC patients to controls. The blue line represents significant differential methylation at false discovery rate (FDR) <math>< 0.05</math>, and the red line represents Bonferroni-adjusted genome-wide significance (c The volcano plot shows differentially expressed genes that are associated with cell specific CpG sites in UC for all three cell types. x-axis shows $\log_2\text{FC}$ and y-axis shows the negative log *p*-value detected for each gene in DE analysis. **d** The lollipop diagram shows the top 10 Gene Ontology (GO) biological processes identified as enriched in sets of differentially methylated CpG sites (blue) and differentially expressed genes in UC. Y-axis shows the number of CpGs and genes detected for each GO term

Table S3c) were differentially expressed between patients with UC at diagnosis and non-IBD controls (FDR <math>< 0.05</math>; Fig. 1c). The observation that the majority of the genes found to be associated with UC-specific CpG sites also demonstrated changes in steady state levels of mRNA detectable in the mucosa highlights the regulatory potential of our differentially methylated CpG sites.

Biological relevance of the differentially methylated CpG sites in UC

GO analysis revealed that the set of genes ($n = 488$) associated with the differentially methylated CpG sites in epithelial cells was enriched for numerous biological processes related to epithelial function, including changes in wound response, GTPase signaling and cell migration (Fig. 1d top panel; Additional file 2: Table S2D–E). Similarly, genes associated with immune cell-specific CpGs

($n = 150$) showed an enrichment for immune function related processes, including cell activation and signaling (Fig. 1d bottom panel; Additional file 3: Table S3D–E), all of which are expected to occur in the damaged mucosa [30–33]. Further, a text-mining analysis [34] revealed that 99 of 885 epithelial and 53 of 442 immune genes were associated with UC in the literature (see Additional file 2: Table S2F–G and Additional file 3: Tables S2F–G). However, we did not detect any pathway level enrichment for genes associated with fibroblast-specific CpG sites ($n = 56$) at FDR <math>< 0.05</math>.

DNAm changes in the rectal mucosa of patients with UC post-treatment

To assess the cellular and molecular changes in the rectal mucosa of patients with UC after treatment, we next

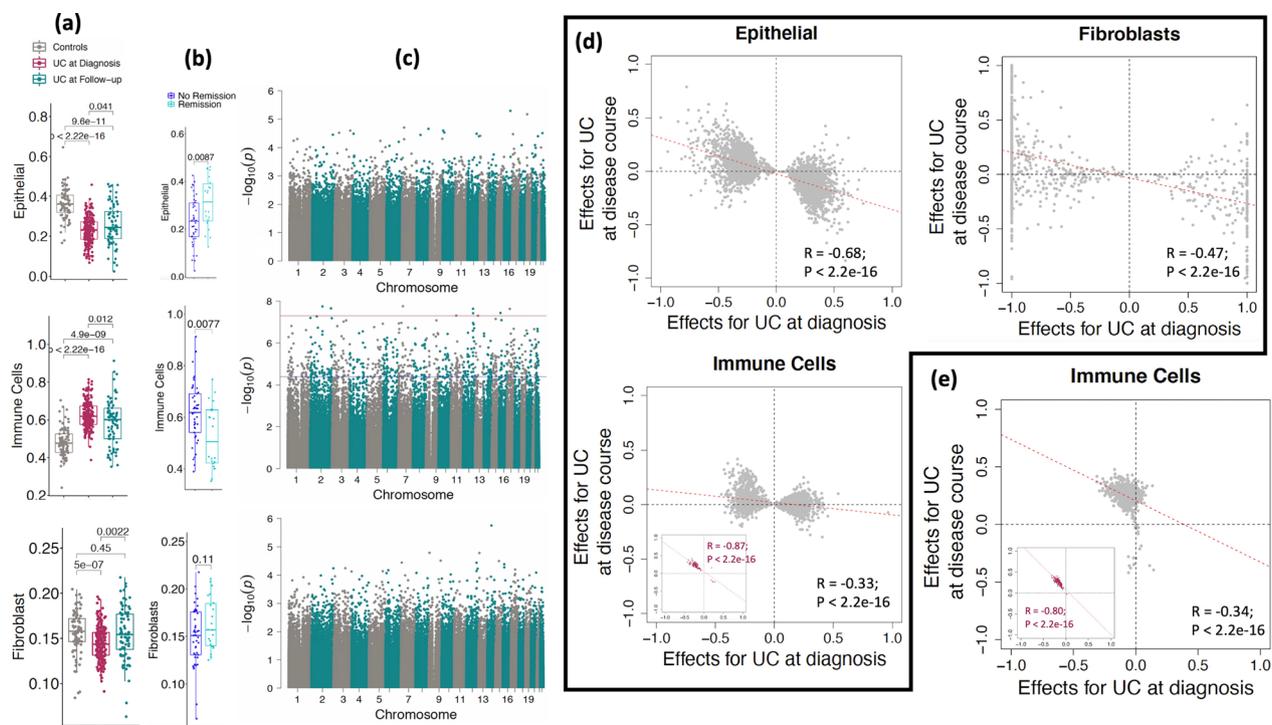


Fig. 2 Longitudinal profiling of DNAm in UC patients at during the disease. **a, b** Boxplots show the rectal biopsy DNA methylation (DNAm) estimated cell proportions for epithelial, immune and fibroblast compared among **a** controls ($n = 85$), UC at diagnosis ($n = 211$) and UC at Follow-up ($n = 73$), and **b** UC patients who underwent clinical remission at week 52 ($n = 22$) vs those who did not ($n = 45$). p Values are from the Wilcoxon test. **c** Cell-specific epigenome-wide analysis (EWAS) analysis shows differentially methylated sites between UC at diagnosis and UC at follow-up (paired samples, $n = 73$) for all three major cell types. The blue line represents the sites significant at FDR < 0.05 , and the red line represents the sites significant at epigenome-wide $p < 1e-08$. **d** The effect sizes for all the UC associated sites from all three types epithelial ($n = 3504$ CpG sites) and fibroblast ($n = 910$) and immune cells ($n = 2279$) of UC at diagnosis i.e., UC at diagnosis vs Controls (x-axis) were compared to UC at disease course i.e., UC at diagnosis vs follow-up (y-axis). In the immune sub-panel, maroon dots represent the 96 CpG sites that reached significance after multiple test corrections on p -values of 2279 CpGs in UC at disease course (FDR < 0.05). **e** The effect sizes for all the CpG sites associated with disease course i.e., UC at diagnosis vs follow-up (y-axis) from only immune cells ($n = 668$) were compared to UC at diagnosis i.e., UC at diagnosis versus Controls (x-axis). In the sub-panel, maroon dots represent the 154 CpG sites that reached significance after multiple test corrections on p values of 668 CpGs in UC at diagnosis (FDR < 0.05)

examined cellular proportions and DNAm profiles at follow-up ($n = 73$; Fig. 2). Clinical activity and/or disease severity in pediatric patients with UC is monitored by a validated activity index known as pediatric UC activity index (PUCAI) [35]. The PUCAI scores of patients at follow-up were significantly lower than those obtained at the time of diagnosis (Additional file 1: Fig. S4a) consistent with patient's disease activity improving, on average, after treatment. Consistent with this, we observed a corresponding change in the estimated proportions of epithelial, immune and fibroblast compartments (Fig. 2a). Interestingly we also observed that the epithelial and immune cell proportions were significantly improved ($p = 0.0087$; $p = 0.0077$) in the follow-up UC patient's rectal biopsies who had undergone clinical remission at week 52 when compared to those who had not (Fig. 2b). We observed a similar trend where the epithelial proportions were significantly increased and immune cells

were significantly depleted (similar to levels in controls) for both corticosteroid-free and calprotectin remissions at week 52 (Additional file 1: Fig. S5), but did not observe such changes in fibroblasts. Compared to UC at diagnosis, we observe an expansion in the proportion of epithelial and fibroblast compartments at follow-up, and depletion of immune cells.

Our cell type-specific EWAS comparing DNAm levels at the time of follow-up to those obtained at the time of diagnosis within the same 73 individuals found 668 CpG sites differentially methylated within the immune compartment (FDR < 0.05 ; Fig. 2c middle panel; Additional file 1: Fig. S4c, Additional file 5: Table S5a). For the majority of significant sites, the direction of the changes moved DNAm levels in cases after treatment closer to levels seen in controls. These patterns are consistent with DNAm levels in the immune compartment reflecting systemic immune status, and thus responding

to systemic anti-inflammatory therapies. On the other hand, EWAS comparing DNAm levels at the time of follow-up to those obtained at the time of diagnosis did not reveal any differentially methylated CpG sites in the epithelial (Fig. 2c, top panel, Additional file 1: Fig. S4b) and fibroblast compartments (Fig. 2c, last panel, Additional file 1: Fig. 4d).

To identify the impact of treatment effects on DNAm sites, we compared the DNAm differences observed for UC at diagnosis vs. controls, and UC at disease course vs. UC at diagnosis. The EWAS in Fig. 1b shows the CpG sites that are differentially methylated between newly diagnosed UC patients and controls, while the EWAS in Fig. 2e shows the CpG sites that are differentially methylated in UC patients before and after treatment. Figure 2d shows that for CpGs significant in each of the three compartments in the UC vs. control EWAS (Fig. 1b), there is a negative correlation in effect sizes from the two EWAS; this is consistent with a pattern where CpGs showing DNAm differences in patients at the time of diagnosis return to control levels upon treatment. Stronger negative correlations were observed for sites associated with UC in epithelial cells ($R = -0.68$) or fibroblasts ($R = -0.47$) vs. those associated in immune cells ($R = -0.33$); however, significant pre- vs. post-treatment DNAm differences were only observed in immune cells (Fig. 2c, middle panel), with 96 CpG sites showing significant DNAm differences in immune cells in both the case-control and pre-post analyses. For these 96 sites we observed a strong negative correlation ($R = -0.86$), with DNAm levels of 95 of these sites moving in the opposite direction from onset and trending towards controls and only one site (cg17642041, located in the transcription start site of the *CCR9* gene) trending away from control levels (maroon dot in Fig. 2C immune cell panels). We observed the same general pattern ($R = -0.34$) at all the 668 CpG sites that differed between diagnosis and follow-up UC: 154 of 668 CpG sites were significantly different at follow-up ($FDR < 0.05$) with a very strong negative correlation ($R = -0.80$) between onset and follow-up (Fig. 2e; Additional file 5: Table S5e).

Extending this analysis to all ~820 K CpG sites analyzed across epigenome-wide, we still observed significant negative correlations between effect sizes ($R = -0.21$ for epithelial cells; $R = -0.16$ for immune cells; $R = -0.21$ for fibroblasts; Additional file 1: Fig. S6), suggesting that at a broad scale, the epigenomic level the DNAm profiles observed at follow-up were trending towards those observed in controls. Of the 2279 sites differing between UC at diagnosis and controls in immune cells (from Fig. 1b, middle panel), 33 showed a significant negative association (epigenome-wide) in the comparison of cases

vs. controls and a significant positive association when comparing UC at diagnosis vs. follow-up in cases, with a negative correlation between the effect sizes from the two analyses ($R = -0.92$; Additional file 1: Fig. S6b).

Biological relevance of the differentially methylated CpG sites in UC at disease course

In support of these patterns showing DNAm profiles at follow-up trending towards control levels, possibly due to inflammation-reducing treatment, gene ontology analysis on the 668 CpG sites differing from UC at diagnosis to follow-up identified numerous biological processes related to immune function, including changes of B-cell/leukocyte activation and differentiation (Fig. 3a, Additional file 5: Table S5g). To gain further biological insight into the pathways related to DNAm changes in the immune compartment during treatment, we analyzed these 668 CpG sites for association between DNAm and nearby gene expression (270 genes). In this analysis we compared changes in gene expression for each gene with DNAm changes at its associated CpG site across a matched cohort of UC patients ($n = 29$) and observed that expression of 39 genes associated with DNAm at 42 CpG sites ($p < 0.05$; Additional file 5: Table S5g). Differential expression analysis at diagnosis vs. follow-up ($n = 29$ matched samples) [25] identified 9 genes as differentially expressed ($FDR < 0.05$; Fig. 3b, Additional file 5: Table S5h). As an example, we highlight two CpG-gene pairs in Fig. 3c, which shows the relative levels of DNAm at CpG sites 1500 base pairs upstream from the transcription start site (TSS) of *FGD2* gene, and their corresponding gene expression level for the same UC subjects at diagnosis and follow-up.

To gain a finer-grained understanding of the immune compartment, we used an immune-subtype reference panel (see methods) to make cell-type composition estimates within this compartment breaking cell-counts into 7 parts (B-, CD8T-, CD4T-cells, monocytes, neutrophils, eosinophils, and nature killer cells). We identified significant changes in the cellular proportions of B-cells and CD4 T-cells between diagnosis and follow-up ($p < 0.05$, Wilcoxon test, Additional file 1: Fig. S7), with decreases in B-cell and increases in T-cell proportions at follow-up. Although neutrophil abundance did not change significantly from diagnosis to follow-up, most of the significant changes in DNAm between diagnosis and follow-up (145 CpG sites) were inferred to be attributable to the neutrophil compartment (Additional file 1: Fig. S8; Additional file 5: Table S5i). We also observed at least 9 CpG sites showing different DNAm patterns in B-cells (Additional file 1: Fig. S8; Additional file 5: Table S5j). Thus,

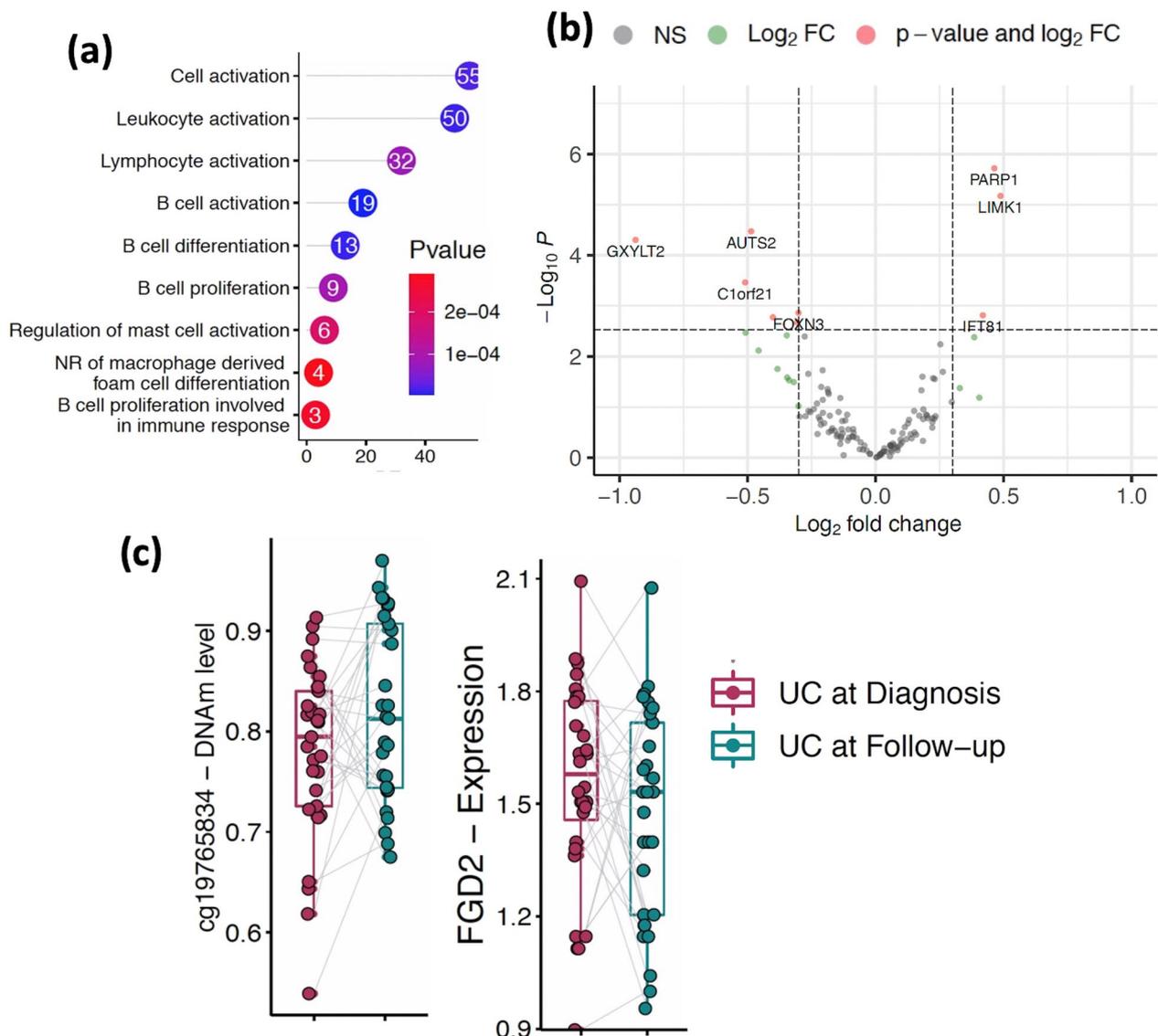


Fig. 3 Regulatory importance of the differentially methylated CpG sites in UC at disease course. **a** The lollipop diagram shows Gene Ontology (GO) biological processes identified for 668 differentially methylated CpG sites (blue) from immune cells. Y-axis shows the number of CpGs detected in each GO term. **b** The volcano plot shows differentially expressed genes that are associated with immune cell specific CpG sites in UC during the course of the disease. x-axis shows log₂FC and y-axis shows the negative log p-value detected for each gene in DE analysis. **c** Boxplots depicting the methylation proportions of UC at diagnosis and during follow-up (matched sample, n = 29, that contains both DNAm and gene expression profiles) at the two CpG sites located TSS1500 to the nearby gene, and the corresponding gene expression values are plotted. Y-axis shows DNAm values and log of read counts representing methylation and gene expression, respectively

neutrophils, and to some extent B-cells, appear to be undergoing the largest amount of epigenetic change in the mucosa during UC treatment, at least as inferred from this model.

UC-specific rectal DNAm signatures at diagnosis correlate with disease severity defined by PUCAI

Next, as a proof-of-concept, we asked if DNAm could distinguish patients with UC based on their disease

severity. A comparison of 44 mild UC (10 < PUCAI < 35), 80 moderate UC (35 < PUCAI < 65) and 87 severe UC patients (PUCAI > 65) across the entire 820 K CpG panel showed significant differences between mild vs moderate, and mild vs severe UC patients in PC1 (Additional file 1: Fig. S9). However, no differences were observed between moderate vs severe UC. The estimated cell proportions from mucosal DNAm profiles show no significant differences for either epithelial cell, fibroblasts, and immune

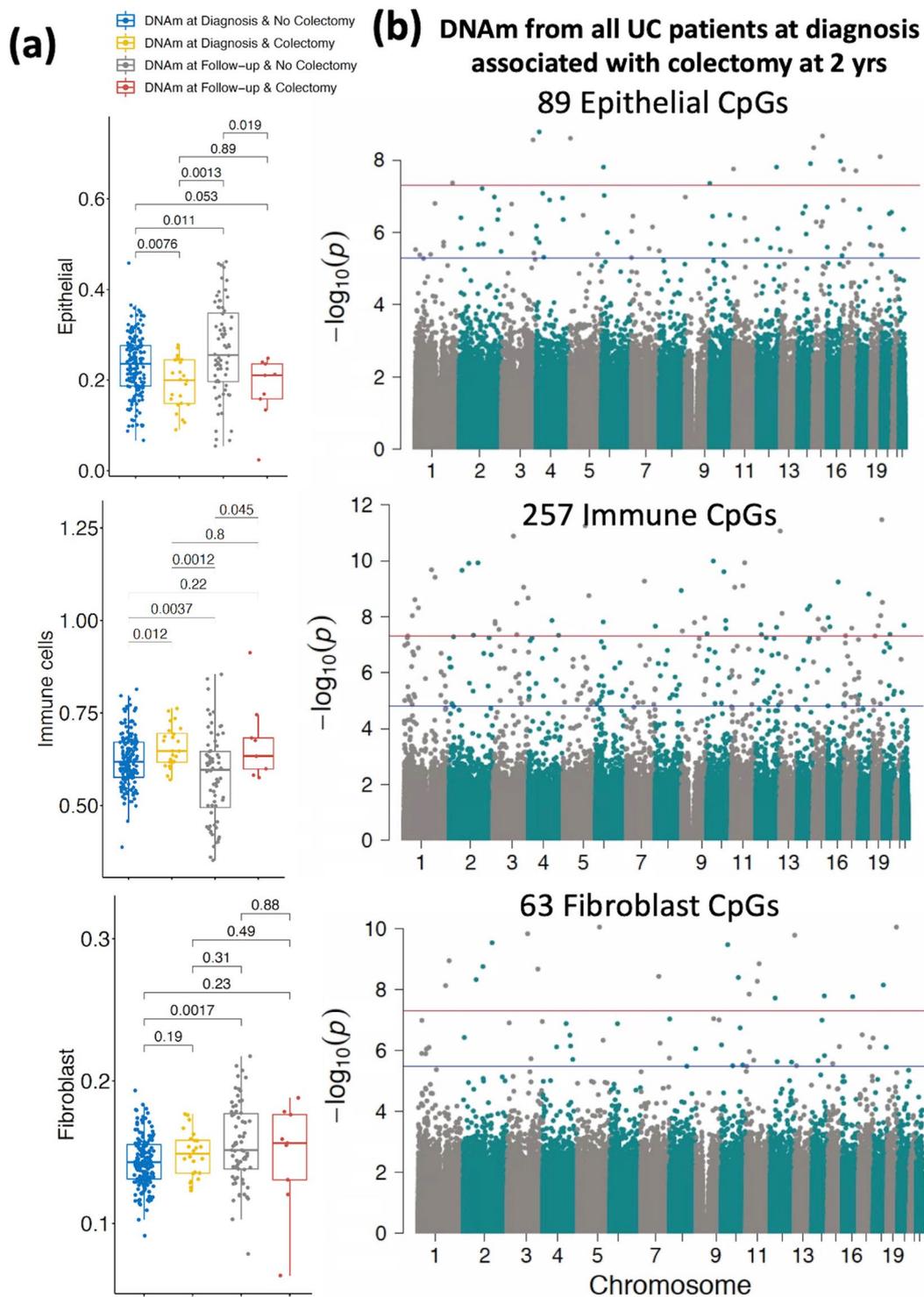


Fig. 4 DNAm signatures at diagnosis associated with colectomy at 2 years. **a** Comparison of estimated cell proportions for epithelial, immune cells and fibroblasts obtained from rectal biopsies at diagnosis compared between colectomy and non-colectomy groups. *P* Values are from the Wilcoxon test. **b** Cell-specific epigenome-wide DNAm analysis (EWAS) comparing UC patients who underwent colectomy at 2 years ($n = 24$) to non-colectomy UC patients ($n = 175$). The blue line represents significant differential methylation at $FDR < 0.05$, and the red line represents Bonferroni-adjusted genome-wide significance ($p < 1e-08$)

cell across disease severity in UC (Additional file 1: Fig. S9a). Consistent with this, our cell type specific EWAS among mild, moderate and severe UC patients did not identify any significant CpGs (data not shown). Therefore, to identify individual CpG sites that capture disease severity within each cell type, we grouped patients with moderate and severe UC ($n=167$) and compared them to those with mild UC ($n=44$), revealing 46, 127 and 13 CpGs in epithelial, immune and fibroblast compartments, respectively (Additional file 1: Fig. S9b, Additional files: Tables S6–8). GO analysis of these sites did not establish strong evidence for enrichment of any biological processes at $FDR < 0.05$.

We tested whether any of the cell-specific CpG sites (as a predictor) are associated with nearby gene expression (as an outcome). We observed 7 Gene-CpG associations for epithelial (Additional file 6: Table S6b) and 12 Gene-CpG associations for immune cells ($FDR < 0.05$; Additional file 7: Table S7b). However, none of these genes were differentially expressed in mild UC when compared to moderate and severe (data not shown).

Rectal DNAm signatures show potential for indicating colectomy risk

The total DNAm signatures were analyzed for signals indicating predisposition towards severe disease resulting in colectomy by year 2 post diagnosis. PC1 of the total 820 K CpG sites showed a significant difference in the DNAm signatures between patients who would eventually undergo colectomy at 2 years post diagnosis with patients who did not (Additional file 1: Fig. S10). Similarly, significant differences in DNAm signatures amongst the three compartments, epithelial, immune and fibroblast were also observed between colectomy and no-colectomy. In total, 24 patients underwent colectomy in the 2 years after diagnosis, and while the “no colectomy” group showed improvements in epithelial ($P=0.01$) and fibroblast ($P=0.002$) proportions along with a decrease in immune cell proportions at follow-up ($P=0.004$) (a sign of mucosal healing and reduction of inflammation), the colectomy group showed no improvement in epithelial ($P=0.89$) and fibroblast proportions ($P=0.49$) and no decrease in the immune proportions (remaining elevated compared to “no colectomy” at follow-up; $P=0.80$) ($n=175$; Fig. 4a).

To identify genes whose DNAm patterns at diagnosis indicate the need for future colectomy, cell-type specific EWAS analysis was performed on non-colectomy UC patients ($n=175$) vs UC patients who eventually underwent colectomy ($n=24$; Fig. 4b; Additional file 1: Fig. S11). We identified 89 CpG sites associated with future colectomy in epithelial cells (Fig. 4b top panel; Additional file 9: Table S9), 257 CpG sites in immune cells (Fig. 4b

middle panel and Additional file 10: Table S10), and 63 CpG sites in fibroblasts (Fig. 4b bottom panel and Additional file 11: Table S11). GO analysis of UC colectomy CpG sites did not establish strong evidence for enrichment of any biological processes at $FDR < 0.05$.

Among CpG sites associated with future colectomy, we observed 15 and 10 Gene-CpG associations for sites associated with colectomy in epithelial cells and immune cells respectively (Additional files 9–10: Tables S9b, S10b). However, none of these genes showed differential expression in UC.

We further analyzed differences at diagnosis in the severe UC group by comparing those patients that eventually underwent colectomy at 2 years post diagnosis ($n=14$) to those that did not ($n=60$). The epithelial cell proportions were nominally decreased ($P=0.05$) at diagnosis in colectomy severe UC patients when compared to non-colectomy severe UC patients (Additional file 1: Fig. S12a, top panel). In contrast, an increase in the immune and fibroblast proportions was observed for the colectomy group, but the differences did not reach statistical significance (Additional file 1: Fig. S12a, middle and bottom panels). Only 2 CpG sites for epithelial and fibroblast cells were found by cell-specific EWAS analysis (Additional file 1: Fig. S12b, top and bottom panels; Additional files: Tables S9,11), and 64 CpG sites for immune cells (Additional file 1: Fig. S12b, middle panel; Additional file 10: Table S10).

Clinical relevance of differentially methylated CpG sites in UC

The distinction between methylation signatures in UC at diagnosis and controls suggests that a diagnostic potential exists in these values that can be leveraged for patient stratification. Supporting this, a random forest analysis showed that CpG sites showing cell-type-specific associations with UC in epithelial, immune cells, and fibroblasts in a training dataset could indeed distinguish UC patients from controls in an independent validation dataset, with 96% accuracy ($AUC=0.96$); Additional file 1: Fig. S13). However, these rectal biopsy DNAm profiles did not show similar predictive power for disease progression or outcome in UC given with the limitation of the sample size in of the testing group (data not shown).

Discussion

In this study, we used a treatment naïve inception UC cohort in which biospecimens underwent multi-omic analysis at baseline and follow-up. Taking advantage of this inception cohort with extensive prospective clinical meta-data, we estimated DNAm-based cell type proportions in the human rectal mucosa – epithelial, immune

and fibroblast compartments—and examined DNAm differences in each of these compartments at the time of diagnosis and during the course of UC.

We first showed that within each compartment, UC-specific epigenetic changes are disturbed at diagnosis. Our first analysis on UC at diagnosis vs. controls DNAm patterns for different cell populations may be reflecting different aspects of the disease. Immune cell-specific responses are consistent with effects of ongoing inflammation. Immune cell DNAm in rectal tissue, as well as in blood, may be more the consequence of disease caused inflammation than it is the underlying cause of UC or disease progression. The epithelial compartment, on the other hand, with observable differences in wound response and cell migration pathways (prior to treatment) might provide insight into why some individuals fail to respond to effective anti-inflammatory treatments.

The next analysis between diagnosis and follow-up showed that the immune compartment undergoes the greatest amount of epigenetic response to treatment, with the epithelial and fibroblast compartments showing fewer changes. These findings are consistent with previous work [19] reporting that DNAm changes in purified epithelial cells from mucosa persisted across two time-points. Especially, overall cell proportions, and many of the cell-specific methylation patterns moved closer to levels seen in controls. This pattern was most pronounced in immune cells and is hardly surprising since current treatment primarily targets the immune system and clearly its effects are most direct and observable in this compartment.

A further the cell-specific disease severity analysis on the rectal biopsy DNAm from UC at diagnosis indicated that epigenetically moderate and severe UC patients group together, whereas mild UC has a distinct epigenetic profile relative to moderate and severe UC as defined by PUCAI. On the other hand, we couldn't establish any evidence that these CpG sites are involved in nearby gene regulation or biological processes related to disease severity in UC. On the other hand, there were widespread methylation differences at follow-up between those with ongoing, severe disease who would soon need colectomy, and those who would not. Remarkably, DNAm signatures observed in the rectal mucosa at diagnosis could also distinguish patients destined to need a colectomy within 2 years and those who did not.

The epithelial barrier is the front-line defense against invading microbes, toxins, and other luminal contents, but also provides selective transport of nutrients and other beneficial substances that maintain homeostasis and carries significant inflammatory consequences to underlying mucosa upon its malfunction and degradation. The lack of epigenetic changes in the epithelial

and fibroblast cells following treatment is of concern to mucosal healing. Notably, we observed that at least 25% of the epithelial derived CpG sites (889 out of 3504) changing during UC are consistent with a previously published study that compared purified epithelial cell methylation patterns between UC patients and controls [19]. Damage and erosion of the intestinal epithelium is a hallmark characteristic of UC, and the large number of epithelial-specific CpG sites affected at diagnosis indicates either a response to external stimuli potentiating the disease or a response to the damage that the body is trying to repair. In fact, in the study of epithelial-derived organoids described in the Introduction, at least 25% of the genes (124 out of 488) associated with epithelial derived CpG sites were also observed to be differentially expressed in UC [19], supporting a relationship between changing DNAm and gene transcription in the epithelium during UC. Our observation of decreased epithelial proportions based on changes in DNAm signals, with a corresponding increase in immune cell abundance, is in line with the well-documented biological consequences of UC. Pathway analysis of the genes influenced by the significantly changing CpG sites showed changes in pathways regulating cell migration, lipid metabolism, small GTPase signaling and wound repair/healing, suggesting that the changes in DNAm in turn influence nearby genes and processes that are critically involved in restitution and repair [36–39].

Fibroblasts originating from the mesenchymal compartment are responsible for maintaining the extracellular milieu that provides structural stability and regulatory growth and differentiation signals for the epithelium [40–42]. The decrease in proportion of this cellular compartment further correlates with symptoms of UC, but this was the one compartment that looked most similar in comparison to control abundance at follow-up. The epigenetic findings here provide new targets that should be considered for correcting UC related epigenetic changes that could promote mucosal healing. For example, *UBE2G1* detected in fibroblasts was found to be hypomethylated and did not revert after treatment, and is a known IBD marker [43]. *UBE2G1* is ubiquitin conjugating enzyme involved in degradation of short-lived proteins and known to have disease implications when its functionality is perturbed [44–46]. Further, whether therapeutic changes in the fibroblasts enhanced their immunosuppressive effects and facilitated a decrease in immune cell proportion was not entirely clear, but a reversion of many of the fibroblasts' CpG signatures to non-IBD status suggested this possibility.

Clearance of the immune response by current IBD treatments (anti-TNF as an example) has long been appreciated as the standard of care for IBD and was

exemplified in our DNAm analysis where healing was associated with a decrease in mucosal immune proportions and reversion of many of the immune-specific CpG sites back to non-IBD signatures. Interestingly, *FOXP1* [47], *IL23R* [48–51], *CCR9* [52, 53] have been documented to play a role in IBD, and here their corresponding CpG loci were found to maintain disease-specific DNAm signatures after treatment. The *CCR9/CCL25* signaling axis is responsible for leukocyte trafficking to the gut and our data may reflect epigenetic changes during UC that are associated with the activation and recruitment of immune cells from the peripheral blood into the intestinal mucosa, consequentially sustaining inflammatory conditions and prolonging or worsening the disease. New targeted therapies that can offset epigenetic effects regulating *CCR9* may show improved resolution of the inflammatory response [53].

Additionally, our study has revealed some important cell- and disease-severity specific findings that were not reported in previous literature. In our previous study, we reported that peripheral blood DNAm signatures in IBD associate with the degree of inflammation, but do not predict/characterize disease in the gut. In the current study, utilizing samples provided by the longitudinal UC patients from PROTECT cohort, we were able to address this question, demonstrating that the DNAm patterns from actual diseased tissues reflect the nature of disease rather than the inflammation status. This is further supported by our comparative analysis of cell-specific EWAS indicating that only 9 sites from both epithelium and immune cells, and 4 sites from fibroblasts, overlap with our previous total peripheral blood DNAm sites [18]. Unlike blood DNAm, the rectal tissue DNAm, particularly epithelial / mesenchymal tissue DNAm, do not revert back to normal levels. While contrasts between mild and moderate UC were largely difficult to interpret, differences between individuals requiring early colectomy and those with less severe disease were more pronounced. Methylation patterns at diagnosis seem to predict which patients are most likely to progress to colectomy after two years post treatment, and methylation and cell-composition patterns which fail to revert towards controls at follow-up are especially good predictors of who will require an early colectomy.

Finally, we note that the model we employed to decompose mucosal data (comprised of numerous cell types) is believed to be well powered to distinguish three main cellular lineages (epithelial, immune and fibroblasts), but the model relies on estimates, which may be less precise than direct measurements obtained through immunophenotyping or single-cell methodologies [19, 54, 55]. The discoveries reported here suggest that future studies performed in purified cells or using single-cell technologies

profiling DNA methylation at the genome-scale will be fruitful in providing further insight into the role of specific cell types in UC. Similarly, because DNAm is influenced by age and environmental factors, it would be interesting to see whether similar cell-type-specific patterns are observed in studies of UC in adults.

Conclusion:

In summary, we show that cell type-specific epigenetic changes that occur in the course of UC are taking place in the diseased tissue (rectal mucosa), and that these changes are associated with disease severity and outcome and can accurately distinguish patients from controls. Based on our findings, we speculate that for individuals who do not sufficiently respond to current therapies, targeting epithelial genes with barrier function may improve the clinical outcomes.

Methods

Overall study participants

The cases used for analysis here are a subset of the PROTECT UC cohort. PROTECT is a multicenter inception cohort with a total of 431 treatment naïve UC patients from 29 centers in the USA and Canada. This cohort was prospectively followed for at least 2 years with rectal biopsy collection at diagnosis (before treatment) and subsequent follow-up (treatment period from 8 weeks to 2 years; follow-up; $n=73$). Details regarding inclusion / exclusion criteria, study protocol, approvals, and other clinical parameters assessed have been reported previously [23, 24]. Based on the availability of the rectal tissue DNA, 211 pediatric UC patients from the PROTECT cohort, aged 4–17 years old, were used. Out of the 211 had baseline rectal DNA availability, 73 of them also had follow-up rectal tissue DNA and were used. All 73 patient follow-up samples were collected as clinically indicated during the follow-up visits between 8 weeks and 2 years. All 73 patients received one or more treatment(s) prior to follow-up, described in detail in Hyams et al. [23, 24]. The UC diagnosis was based on conventional clinical, endoscopic, and histological parameters. For non-IBD controls, we used age- and gender-matched rectal biopsy genomic DNA samples from 85 RISK participants (RISK is described elsewhere [22]) that had no histologic or endoscopic inflammation and remained asymptomatic during the disease course. Both the PROTECT and RISK studies were approved by the Institutional Review Boards at each of the participating RISK and PROTECT sites. The same sites in North America participated in both the PROTECT and RISK studies. All relevant ethical regulations for work on human participants have been met and conducted in accordance with the criteria set by the

Declaration of Helsinki. Informed consent was obtained from the parents of all study participants.

Quantification of genome-wide DNAm

Rectal biopsy genomic DNA was extracted using the All-Prep DNA/RNA Mini Kit (Qiagen, Valencia, CA, USA), and 500 ng of DNA was subjected to bisulfite treatment using EZ DNAm-Gold™ Kits (Zymo Research, Irvine, CA, USA). MethylationEPIC BeadChip (Illumina, San Diego, CA) was used to quantify genome-wide DNAm differences across ~850,000 genome-wide CpG sites [56]. The R package CpGassoc [57] was used to perform the initial quality control (QC). 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. Probes mapping to multiple locations were also removed [58]. After QC, a total of ~820,000 probes and 369 samples (85 non-IBD controls, 211 UC samples at diagnosis, and 73 samples from follow-up) remained. Beta values (β) were calculated for each CpG site as the ratio of methylated (M) to the sum of methylated and unmethylated (U) signals: $\beta = M/(M + U)$. Signal intensities were then normalized using the module beta-mixture quantile dilation (BMIQ) [59] to account for the probe type bias. These normalized signal intensities were used to perform principal component analysis to further identify sample outliers.

Statistical methods

- (i) **Estimating the cell proportions of rectal biopsy tissues.** The R/Bioconductor-package EpiDISH [26] was used to estimate cell type proportions through deconvolution of our bulk DNAm data derived from rectal tissue. Estimation was based on two reference panels: (i) an epithelial, immune and fibroblast -specific reference panel and (ii) another immune subtype cells reference panel that can also map the data to 7 immune cell types; B-cells, CD4+ T-cells, CD8+ T-cells, NK-cells, monocytes, neutrophils, and eosinophils. These estimated cell type proportions were used as covariates in all DNAm analyses to adjust for differences in DNAm due to between-sample differences in cellular composition. Briefly, for a given DNAm data matrix, the CellDMC program in EpiDISH [26] uses a reference panel of DNAm profiles in major cell-types (CTs) to estimate cell-type fractions in each sample. It then fits a statistical model to test for association between DNAm and phenotype adjusted for cell-type fractions. In addition to main effects for cell-type fractions, the model includes interaction terms

between the phenotype and estimated cell-type fractions, enabling identification of DMCs in specific cell-types (DMCTs).

- (ii) **Genotyping and data processing.** Peripheral blood DNA samples of 296 cases and controls with DNAm data were genotyped using the UK Biobank array, and ~850,000 genotypes were called using the Axiom Suite software. All subjects had call rates >95% and consistent gender records with the clinical data. All quality control procedures were performed in PLINK [60]. Principal components were computed based on a pruned version “*-hwe 0.001 -maf 0.2 -geno 0.01 -indep-pairwise 50 5 0.2*” of the data set consisting of 58,237 LD-independent SNPs ($r^2 < 0.1$). We used the first 5 genotype-based principal components to control for population stratification in all analyses.
- (iii) **DNAm association with UC at diagnosis.** To identify CpG sites associated with UC at diagnosis, we performed a case-control EWAS on 211 UC patients at diagnosis compared to 85 non-IBD controls. UC-associated methylation changes in rectal tissue regardless of cell type were first profiled using the R package CpGassoc [57]. Briefly, DNAm was regressed on disease status (0 for control, 1 for UC) with age, gender, epithelial and fibroblast cell proportions, and the first five genotype-based principal components as covariates in the model.
- (iv) **Cell-specific methylation association with UC at diagnosis.** We next used the CellDMC function from the EpiDISH [26] package to identify sets of CpGs that show a cell-type specific association with UC. We performed cell-specific EWAS within the epithelial, immune and fibroblasts to test for association between UC and methylation at the ~820,000 sites that passed QC. DNAm at each CpG was regressed on disease status (0 for control, 1 for UC) and age, gender, and the first five genotype-based principal components were included as covariates in the model, along with covariates for DNAm-based estimates of cellular proportions, and an interaction term between UC and cellular proportion to identify cell-specific signal. We identified significant CpGs within each EWAS using the Benjamini–Hochberg false discovery rate criterion ($FDR < 0.05$).
- (v) **Cell-specific methylation association with UC at diagnosis versus follow-up.** To assess longitudinal changes in DNAm in patients with UC, we compared the methylation levels in rectal biopsy samples obtained at diagnosis and follow-up ($n = 73$; from 8 weeks to 2 years). Age was recalculated for

the follow-up samples based upon their time of visit after diagnosis. In the cell-specific EWAS, DNAm was regressed on disease course (0 for at diagnosis, 1 for at follow-up), and age, gender, and the first five genotype-based principal components were included as covariates in the model, along with the covariates and interactions for cell type proportion. To account for the two time points from each patient, representing DNAm levels at diagnosis and follow-up, we included fixed effect covariates for subject ID. To identify CpG sites associated with UC during treatment, we performed a similar cell-specific EWAS comparing diagnosis vs follow-up within the same samples ($n=73$).

- (vi) **Gene Ontology (GO) biological process enrichment analysis.** Gene ontology (GO) for biological process enrichment analysis was performed for genes annotated to our UC associated CpG sites by the R/Bioconductor package *missMethyl* [61]. Genes with more CpG probes on the MethylationEPIC array are more likely to have differentially methylated CpGs, which could introduce potential bias when performing pathway enrichment analysis. The *gometh* function implemented in *missMethyl* considers the varying number of differentially methylated CpGs by computing a prior probability for each gene based on the gene length and the number of CpGs probed per gene on the array. Similarly, the GO biological process enrichment analysis for UC associated genes that are associated to UC associated CpGs was performed by using *Topgene* [62]. In addition, we used *DisGeNET* [34] to perform a text-mining analysis of whether the cell-specific gene signatures of this study are enriched for genes previously associated with UC. We specifically tested for disease ID C0009324, which consists of 1458 genes associated with UC.
- (vii) **Analysis of DNAm and gene expression.** Differential gene expression analysis was performed on a subset of patients from whom both gene expression and DNAm data was available from the rectal mucosa ($n=119$). Gene expression was previously measured using TruSeq Illumina mRNAseq (20 controls and 211 UC samples) [25] or Lexogen 3'UTR mRNAseq (39 UC at diagnosis and a matched subset of the same 39 UC patients' sampled at 52 week follow-up) [63]. More details on RNA sequencing, data processing and QC are described elsewhere [25, 63].
- (viii) CpG sites associated with disease status, disease course, disease severity, or colectomy status in the EWAS were further tested for association with expression of genes as annotated by Illumina. In

total we used data from 119 patients with UC who had both DNAm and TruSeq Illumina mRNAseq gene expression profiles [25]. Gene expression counts were regressed on DNAm proportions using the DESeq2 R/Bioconductor package, with age, gender and the first five genotype-based principal components included as covariates in the model. A false discovery rate criterion ($FDR < 0.05$) was used to define a set of significant CpG-gene pairs.

- (ix) **Differential gene expression analysis on disease status, disease severity, and colectomy status.** We also performed a targeted differential expression analysis of genes annotated to the CpG sites identified in the EWAS of disease status, disease course, disease severity and/or colectomy status. The TruSeq Illumina mRNAseq dataset [25] ($n=119$) was used to identify genes differentially expressed between disease status, disease severity and/or colectomy status. Similarly, the Lexogen 3'UTR mRNAseq data was analyzed to identify genes differentially expressed for CpGs associated to UC at disease course. The DESeq2 package was used to perform the differential expression analysis. Briefly, gene expression values were regressed on disease status, disease severity, or colectomy status with age and gender adjusted as covariates in the model using the default normalization method. DE genes were identified using either nominal p value < 0.05 and fold change (FC) > 1.2 .
- (x) **Random forest classification** The entire dataset at diagnosis ($n=296$) was divided into a 75% training (64 controls and 158 UC cases) and a 25% validation set (21 controls and 53 UC cases). For this analysis, CellDMC was performed only on training dataset samples and the UC-associated CpG sites across all three cell types were identified. A RandomForest (RF) model was constructed using the UC-associated CpG sites identified in the training dataset as predictors and UC as the outcome, using the *RandomForest* [64] model in R with default parameters. The trained RF model was tested using the test dataset samples to test the prediction performance of the defined model. Accuracies (ACC) and area under the curve (AUC) were calculated by comparing of actual labels to predicted labels of test set class, separately for cases and controls.

Abbreviations

ACC	Accuracies
AUC	Area under the curve
DNAm	DNA methylation

EWAS	Epigenome-Wide Association Studies
FDR	False discovery rate
GO	Gene ontology
IBD	Inflammatory bowel disease
PUCAI	Pediatric Ulcerative Colitis Activity Index
PROTECT	Predicting response to standardized pediatric colitis therapy
PCs	Principle components
RISK	Risk Stratification and Identification of Immunogenetic and Microbial Markers of Rapid Disease Progression in Children with Crohn's Disease
UC	Ulcerative colitis

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13148-023-01462-4>.

Additional file 1. Supplementary Figures.

Additional file 2. List of UC-specific epithelial DNA methylation signatures and pathways.

Additional file 3. List of UC-specific immune cells DNA methylation signatures and pathways.

Additional file 4. List of UC-specific fibroblast DNA methylation signatures and pathways.

Additional file 5. List of DNA methylation signatures and pathways associated in the rectal mucosa of patients with UC post-treatment.

Additional file 6. List of UC-specific epithelial DNA methylation signatures at diagnosis associated with disease severity defined by PUCAI.

Additional file 7. List of UC-specific immune cells DNA methylation signatures at diagnosis associated with disease severity defined by PUCAI.

Additional file 8. List of UC-specific fibroblast DNA methylation signatures at diagnosis associated with disease severity defined by PUCAI.

Additional file 9. List of UC-specific epithelial DNA methylation signatures at diagnosis associated with colectomy at two years.

Additional file 10. List of UC-specific immune cells DNA methylation signatures at diagnosis associated with colectomy through two years.

Additional file 11. List of UC-specific fibroblast DNA methylation signatures at diagnosis associated with colectomy through two years.

Additional file 12. The metadata of the participants used in this study.

Acknowledgements

We are grateful to Anne Dodd, and Jarod Prince for their support and helpful comments.

Author contributions

KNC, AKS, HKS and SK conceived and designed the study. SV performed the analysis with input from DJC, KNC, AKS and GG processed samples for methylation profiling. JSH, LAD and SK participated in the conception and design of the RISK and PROTECT studies. SV, JDM, HKS, RK, GG, DJC, KNC, AKS, and SK interpreted the results and wrote the manuscript. All authors reviewed and approved the manuscript prior to submission.

Funding

This research was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) of the National Institutes of Health (NIH), under grant numbers R21DK119997 (SK) and 5U01DK095745 (LAD and JH). This work was also supported by a research initiative grant (RISK) from the Crohn's and Colitis Foundation, New York, NY.

Availability of data and materials

The DNAm data for all the 369 rectal biopsy samples included in this study have been deposited in the Gene Expression Omnibus (GEO) and are accessible through GEO series accession GSE185061. Metadata details regarding subset of patients used in this study are summarized in Additional file 12: Table S12.

Declarations

Competing interests

The authors declare no competing interests.

Author details

¹Division of Pediatric Gastroenterology, Department of Pediatrics, Emory University School of Medicine and Children's Healthcare of Atlanta, 1760 Haygood Drive, W-427, Atlanta, GA 30322, USA. ²Department of Gynecology and Obstetrics, Emory University School of Medicine, Atlanta, GA, USA. ³Division of Digestive Diseases, Hepatology, and Nutrition, Connecticut Children's Medical Center, Hartford, CT, USA. ⁴Division of Pediatric Gastroenterology, Hepatology, and Nutrition, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. ⁵Section of Pediatric Gastroenterology, Texas Children's Hospital Baylor College of Medicine, Houston, TX, USA. ⁶Center for Integrative Genomics, Georgia Institute of Technology, Atlanta, GA, USA. ⁷Department of Human Genetics, Emory University, Atlanta, GA, USA. ⁸Genetics and Molecular Biology Program, Emory University, Atlanta, GA, USA. ⁹Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA, USA.

Received: 12 May 2022 Accepted: 7 March 2023

Published online: 24 March 2023

References

- Alatab S, Sepanlou SG, Ikuta K, Vahedi H, Bisignano C, Safiri S, Sadeghi A, Nixon MR, Abdoli A, Abolhassani H. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol*. 2020;5(1):17–30.
- Hedin C, Vavricka S, Stagg A, Schoepfer A, Raine T, Puig L, Pleyer U, Navarini A, Van Der Meulen-De JA, Maul J. The pathogenesis of extraintestinal manifestations: implications for IBD research, diagnosis, and therapy. *J Crohns Colitis*. 2019;13(5):541–54.
- Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, Ripke S, Lee JC, Jostins L, Shah T, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet*. 2015;47(9):979–86.
- de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, Jostins L, Rice DL, Gutierrez-Achury J, Ji SG, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet*. 2017;49(2):256–61.
- Venkateswaran S, Prince J, Cutler DJ, Marigorta UM, Okou DT, Prahalad S, Mack D, Boyle B, Walters T, Griffiths A, et al. Enhanced contribution of HLA in pediatric onset ulcerative colitis. *Inflamm Bowel Dis*. 2018;24(4):829–38.
- Feuerstein JD, Isaacs KL, Schneider Y, Siddique SM, Falck-Ytter Y, Singh S. Committee AGAICG: AGA clinical practice guidelines on the management of moderate to severe ulcerative colitis. *Gastroenterology*. 2020;158(5):1450–61.
- Hazel K, O'Connor A. Emerging treatments for inflammatory bowel disease. *Ther Adv Chronic Dis*. 2020;11:2040622319899297.
- Kellermayer R. "Omics" as the filtering gateway between environment and phenotype: the inflammatory bowel diseases example. *Am J Med Genet A*. 2010;152(12):3022–5.
- Saito S, Kato J, Hiraoka S, Horii J, Suzuki H, Higashi R, Kaji E, Kondo Y, Yamamoto K. DNA methylation of colon mucosa in ulcerative colitis patients: correlation with inflammatory status. *Inflamm Bowel Dis*. 2011;17(9):1955–65.
- Taman H, Fenton CG, Hensel IV, Anderssen E, Florholmen J, Paulsen RH. Genome-wide DNA methylation in treatment-naive ulcerative colitis. *J Crohns Colitis*. 2018;12(11):1338–47.
- Barnicle A, Seoighe C, Greally JM, Golden A, Egan LJ. Inflammation-associated DNA methylation patterns in epithelium of ulcerative colitis. *Epigenetics*. 2017;12(8):591–606.
- Ventham NT, Kennedy NA, Adams AT, Kalla R, Heath S, O'Leary KR, Drummond H, consortium IB, Consortium IC, Wilson DC *et al* (2016) Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. *Nat Commun* 7:13507.

13. Karatzas PS, Gazouli M, Safioleas M, Mantzaris GJ. DNA methylation changes in inflammatory bowel disease. *Ann Gastroenterol*. 2014;27(2):125–32.
14. McDermott E, Ryan EJ, Tosetto M, Gibson D, Burrage J, Keegan D, Byrne K, Crowe E, Sexton G, Malone K, et al. DNA methylation profiling in inflammatory bowel disease provides new insights into disease pathogenesis. *J Crohns Colitis*. 2016;10(1):77–86.
15. Li Yim AYF, Duijvis NW, Zhao J, De Jonge WJ, D'Haens G, Mannens M, Mul A, Te Velde AA, Henneman P. Peripheral blood methylation profiling of female Crohn's disease patients. *Clin Epigenetics*. 2016;8:65.
16. Nimmo ER, Prendergast JG, Aldhous MC, Kennedy NA, Henderson P, Drummond HE, Ramsahoye BH, Wilson DC, Semple CA, Satsangi J. 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(5):889–99.
17. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shores N, Whitton H, Ryan RJ, Shishkin AA, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015;518(7539):337–43.
18. Sominen HK, Venkateswaran S, Kilaru V, Marigorta UM, Mo A, Okou DT, Kellermayer R, Mondal K, Cobb D, Walters TD, et al. Blood-derived DNA methylation signatures of Crohn's disease and severity of intestinal inflammation. *Gastroenterology*. 2019;156(8):2254–2265 e2253.
19. Howell KJ, Kraiczky J, Nayak KM, Gasparetto M, Ross A, Lee C, Mak TN, Koo B-K, Kumar N, Lawley T. 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(3):585–98.
20. Harris RA, Nagy-Szakal D, Mir SA, Frank E, Szigeti R, Kaplan JL, Bronsky J, Opekun A, Ferry GD, Winter H. DNA methylation-associated colonic mucosal immune and defense responses in treatment-naïve pediatric ulcerative colitis. *Epigenetics*. 2014;9(8):1131–7.
21. Kraiczky J, Nayak KM, Howell KJ, Ross A, Forbester J, Salvestrini C, Mustata R, Perkins S, Andersson-Rolf A, Leenen E. DNA methylation defines regional identity of human intestinal epithelial organoids and undergoes dynamic changes during development. *Gut*. 2019;68(1):49–61.
22. Kugathasan S, Denson LA, Walters TD, Kim MO, Marigorta UM, Schirmer M, Mondal K, Liu C, Griffiths A, Noe JD, et al. Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. *Lancet*. 2017;389(10080):1710–8.
23. Hyams JS, Davis S, Mack DR, Boyle B, Griffiths AM, LeLeiko NS, Sauer CG, Keljo DJ, Markowitz J, Baker SS, et al. Factors associated with early outcomes following standardised therapy in children with ulcerative colitis (PROTECT): a multicentre inception cohort study. *Lancet Gastroenterol Hepatol*. 2017;2(12):855–68.
24. Hyams JS, Davis Thomas S, Gotman N, Haberman Y, Karns R, Schirmer M, Mo A, Mack DR, Boyle B, Griffiths AM, et al. Clinical and biological predictors of response to standardised paediatric colitis therapy (PROTECT): a multicentre inception cohort study. *Lancet*. 2019;393(10182):1708–20.
25. Haberman Y, Karns R, Dexheimer PJ, Schirmer M, Somekh J, Jurickova I, Braun T, Novak E, Bauman L, Collins MH, et al. Ulcerative colitis mucosal transcriptomes reveal mitochondriopathy and personalized mechanisms underlying disease severity and treatment response. *Nat Commun*. 2019;10(1):38.
26. Zheng SC, Breeze CE, Beck S, Teschendorff AE. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat Methods*. 2018;15(12):1059–66.
27. Strober W, Fuss I, Mannon P. The fundamental basis of inflammatory bowel disease. *J Clin Invest*. 2007;117(3):514–21.
28. Alipour M, Zaidi D, Valcheva R, Jovel J, Martínez I, Sergi C, Walter J, Mason AL, Wong GK-S, Dieleman LA, et al. Mucosal barrier depletion and loss of bacterial diversity are primary abnormalities in paediatric ulcerative colitis. *J Crohn's Colitis*. 2015;10(4):462–71.
29. Fournier BM, Parkos CA. The role of neutrophils during intestinal inflammation. *Mucosal Immunol*. 2012;5(4):354–66.
30. Matthews JD, Sumagin R, Hinrichs B, Nusrat A, Parkos CA, Neish AS. Redox control of Cas phosphorylation requires Abl kinase in regulation of intestinal epithelial cell spreading and migration. *Am J Physiol Gastrointest Liver Physiol*. 2016;311(3):G458–465.
31. Birkl D, Quiros M, García-Hernández V, Zhou DW, Brazil JC, Hilgarth R, Keeney J, Yulis M, Bruewer M, García AJ, et al. TNF α promotes mucosal wound repair through enhanced platelet activating factor receptor signaling in the epithelium. *Mucosal Immunol*. 2019;12(4):909–18.
32. Li Y, de Haar C, Chen M, Deuring J, Gerrits MM, Smits R, Xia B, Kuipers EJ, van der Woude CJ. Disease-related expression of the IL6/STAT3/SOCS3 signalling pathway in ulcerative colitis and ulcerative colitis-related carcinogenesis. *Gut*. 2010;59(2):227–35.
33. Keshavarzian A, Fusunyan RD, Jacyno M, Winship D, MacDermott RP, Sanderson IR. Increased interleukin-8 (IL-8) in rectal dialysate from patients with ulcerative colitis: evidence for a biological role for IL-8 in inflammation of the colon. *Am J Gastroenterol*. 1999;94(3):704–12.
34. Pinerio J, Sauch J, Sanz F, Furlong LI. The DisGeNET cytoscape app: Exploring and visualizing disease genomics data. *Comput Struct Biotechnol J*. 2021;19:2960–7.
35. Turner D, Otley AR, Mack D, Hyams J, De Bruijn J, Uusoue K, Walters TD, Zachos M, Mamula P, Beaton DE. Development, validation, and evaluation of a pediatric ulcerative colitis activity index: a prospective multicenter study. *Gastroenterology*. 2007;133(2):423–32.
36. Matthews JD, Reedy AR, Wu H, Hinrichs BH, Darby TM, Addis C, Robinson BS, Go YM, Jones DP, Jones RM, et al. Proteomic analysis of microbial induced redox-dependent intestinal signaling. *Redox Biol*. 2019;20:526–32.
37. Miyoshi H, VanDussen KL, Malvin NP, Ryu SH, Wang Y, Sonnek NM, Lai CW, Stappenbeck TS. Prostaglandin E2 promotes intestinal repair through an adaptive cellular response of the epithelium. *Embo J*. 2017;36(1):5–24.
38. Leeb SN, Vogl D, Gunckel M, Kiessling S, Falk W, Göke M, Schölmerich J, Gelbmann CM, Rogler G. Reduced migration of fibroblasts in inflammatory bowel disease: role of inflammatory mediators and focal adhesion kinase. *Gastroenterology*. 2003;125(5):1341–54.
39. Sommer K, Wiendl M, Müller TM, Heidebreder K, Voskens C, Neurath MF, Zundler S. Intestinal mucosal wound healing and barrier integrity in IBD—crosstalk and trafficking of cellular players. *Front Med (Lausanne)*. 2021;8:643973.
40. Göke M, Kanai M, Podolsky DK. Intestinal fibroblasts regulate intestinal epithelial cell proliferation via hepatocyte growth factor. *Am J Physiol Gastrointest Liver Physiol*. 1998;274(5):809–18.
41. Simo P, Bouziges F, Lissitzky JC, Sorokin L, Keding M, Simon-Assmann P. Dual and asynchronous deposition of Laminin chains at the epithelial-mesenchymal interface in the gut. *Gastroenterology*. 1992;102(6):1835–45.
42. Pastula A, Marcinkiewicz J. Cellular interactions in the intestinal stem cell niche. *Arch Immunol Ther Exp*. 2019;67(1):19–26.
43. Alsobrook JI, Ma T, Leighton J, Tang L, Doherty P, Zhou F, Williams T, Davis L, Harris C. Novel genomic biomarkers that differentiate between inflammatory bowel disease and normal patients using peripheral blood specimens: 1125. *Off J Am College of Gastroenterol (ACG)*. 2008;103:S439.
44. Filatova EV, Shadrina MI, Alieva A, Kolacheva AA, Slominsky PA, Ugrumov MV. Expression analysis of genes of ubiquitin-proteasome protein degradation system in MPTP-induced mice models of early stages of Parkinson's disease. *Dokl Biochem Biophys*. 2014;456(1):116–8.
45. Lu G, Weng S, Matyskiela M, Zheng X, Fang W, Wood S, Surka C, Mizukoshi R, Lu CC, Mendy D, et al. UBE2G1 governs the destruction of cereblon neomorphic substrates. *Elife* 2018;7.
46. Evangelou M, Smyth DJ, Fortune MD, Burren OS, Walker NM, Guo H, Onengut-Gumuscu S, Chen WM, Concannon P, Rich SS, et al. A method for gene-based pathway analysis using genomewide association study summary statistics reveals nine new type 1 diabetes associations. *Genet Epidemiol*. 2014;38(8):661–70.
47. Ren J, Han L, Tang J, Liu Y, Deng X, Liu Q, Hao P, Feng X, Li B, Hu H, et al. Foxp1 is critical for the maintenance of regulatory T-cell homeostasis and suppressive function. *PLoS Biol*. 2019;17(5): e3000270.
48. Sivanesan D, Beauchamp C, Quinou C, Lee J, Lesage S, Chemtob S, Rioux JD, Michnick SW. IL23R (Interleukin 23 Receptor) variants protective against Inflammatory Bowel Diseases (IBD) display loss of function due to impaired protein stability and intracellular trafficking. *J Biol Chem*. 2016;291(16):8673–85.
49. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science*. 2006;314(5804):1461–3.
50. Sun R, Hedl M, Abraham C. IL23 induces IL23R recycling and amplifies innate receptor-induced signalling and cytokines in human

- macrophages, and the IBD-protective IL23R R381Q variant modulates these outcomes. *Gut*. 2020;69(2):264–73.
51. Kim SW, Kim ES, Moon CM, Park JJ, Kim TI, Kim WH, Cheon JH. Genetic polymorphisms of IL-23R and IL-17A and novel insights into their associations with inflammatory bowel disease. *Gut*. 2011;60(11):1527–36.
 52. Wendt E, Keshav S. CCR9 antagonism: potential in the treatment of Inflammatory Bowel Disease. *Clin Exp Gastroenterol*. 2015;8:119–30.
 53. Koenecke C, Förster R. CCR9 and inflammatory bowel disease. *Expert Opin Ther Targets*. 2009;13(3):297–306.
 54. Lin Z, Hegarty JP, Yu W, Cappel JA, Chen X, Faber PW, Wang Y, Poritz LS, Fan J-B, Koltun WA. Identification of disease-associated DNA methylation in B cells from Crohn's disease and ulcerative colitis patients. *Dig Dis Sci*. 2012;57(12):3145–53.
 55. Li Yim AY, Duijvis NW, Ghiboub M, Sharp C, Ferrero E, Mannens MM, D'Haens GR, de Jonge WJ, Te Velde AA, Henneman P. Whole-genome DNA methylation profiling of CD14+ monocytes reveals disease status and activity differences in Crohn's disease patients. *J Clin Med*. 2020;9(4):1055.
 56. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics*. 2016;8(3):389–99.
 57. Barfield RT, Kilaru V, Smith AK, Conneely KN. CpGassoc: an R function for analysis of DNA methylation microarray data. *Bioinformatics*. 2012;28(9):1280–1.
 58. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genom Data*. 2016;9:22–4.
 59. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013;29(2):189–96.
 60. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75.
 61. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics*. 2016;32(2):286–8.
 62. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res*. 2009;37(suppl_2), W305–W311.
 63. Mo A, Nagpal S, Gettler K, Haritunians T, Giri M, Haberman Y, Karns R, Prince J, Arafat D, Hsu N-Y. Stratification of risk of progression to colectomy in ulcerative colitis using measured and predicted gene expression. *bioRxiv* 2021.
 64. Liaw A, Wiener M. Classification and regression by randomForest.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

