Microbiota-sensitive epigenetic signature predicts inflammation in Crohn's disease.

Daniel Kelly, *Center for Inflammation and Tolerance*
Michael Kotliar, *Divisions of Allergy and Immunology and Human Genetics*
Vivienne Woo, *Center for Inflammation and Tolerance*
Sajjeev Jagannathan, *Divisions of Allergy and Immunology and Human Genetics*
Jordan Whitt, *Center for Inflammation and Tolerance*
Jessica Moncivaiz, *Center for Inflammation and Tolerance*
Bruce J. Aronow, *Cincinnati Children’s Hospital Medical Center*
Marla C. Dubinsky, *Mount Sinai Hospital*
Jeffrey S. Hyams, *Connecticut Children’s Medical Center*
James F. Markowitz, *Northwell Health*

*Only first 10 authors above; see publication for full author list.*

---

**Journal Title:** JCI insight  
**Volume:** Volume 3, Number 18  
**Publisher:** American Society for Clinical Investigation | 2018-09-20, Pages e122104-e122104  
**Type of Work:** Article | Final Publisher PDF  
**Publisher DOI:** 10.1172/jci.insight.122104  
**Permanent URL:** https://pid.emory.edu/ark:/25593/th5d1

Final published version: [http://dx.doi.org/10.1172/jci.insight.122104](http://dx.doi.org/10.1172/jci.insight.122104)

**Copyright information:**  
© 2018, American Society for Clinical Investigation  
Accessed September 1, 2019 7:33 PM EDT
Microbiota-sensitive epigenetic signature predicts inflammation in Crohn’s disease

Daniel Kelly,1,2 Michael Kotliar,3 Vivienne Woo,1 Sajjeev Jagannathan,1 Jordan Whitt,1 Jessica Moncivaiz,1 Bruce J. Aronow,4 Marla C. Dubinsky,5 Jeffrey S. Hyams,6 James F. Markowitz,7 Robert N. Baldassano,6 Michael C. Stephens,3 Thomas D. Walters,10 Subra Kugathasan,11 Yael Haberman,2,12 Nambirajan Sundaram,13 Michael J. Rosen,2 Michael Helmrath,11 Rebekah Karns,2 Artem Barski,3 Lee A. Denson,2 and Theresa Alenghat1

1Division of Immunobiology, Center for Inflammation and Tolerance, 2Division of Gastroenterology, Hepatology, and Nutrition, 3Divisions of Allergy and Immunology and Human Genetics, and 4Division of Biomedical Informatics, Cincinnati Children’s Hospital Medical Center (CCHMC) and Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA. 5Department of Pediatrics, Mount Sinai Hospital, New York, New York, USA. 4Division of Digestive Diseases, Hepatology, and Nutrition, Connecticut Children’s Medical Center, Hartford, Connecticut, USA. 6Department of Pediatrics, Northwell Health, New York, New York, USA. 7Department of Pediatrics, University of Pennsylvania, Philadelphia, Pennsylvania, USA. 8Department of Pediatric Gastroenterology, Mayo Clinic, Rochester, Minnesota, USA. 9Division of Pediatric Gastroenterology, Hepatology, and Nutrition, Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, Canada. 10Division of Pediatric Gastroenterology, Emory University School of Medicine, Atlanta, Georgia, USA. 11Sheba Medical Center, Tel Hashomer, affiliated with the Tel-Aviv University, Israel. 12Division of Pediatric General and Thoracic Surgery, CCHMC and Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA.

Introduction

Inflammatory bowel disease (IBD) affects over 1.4 million people in the United States and represents a significant public health and economic challenge (1–3). In pediatric patient populations, specifically, the incidence and prevalence of IBD continues to rise (4). It is clear that the pathogenesis of IBD is multifactorial, involving complex interactions between the environment and genetics of the host (5–8). While abnormal intestinal immune responses are a defining feature of IBD, the mechanisms triggering IBD in genetically susceptible individuals are poorly understood. Current therapeutics focus on controlling excessive inflammation, rather than upstream pathways that predispose to altered intestinal immune response. These treatment options can be effective in ameliorating symptoms in many patients, yet treatment failure is common, side effects of these therapies can be significant, and disease complications occur despite appropriate therapy (9, 10). Therefore, improved understanding of the molecular and cellular mechanisms...
mediating how environmental cues are integrated by intestinal cells to trigger IBD will enable development of new personalized approaches to treat the disease and predict response to therapy.

Commensal bacterial populations in the intestinal environment clearly associate with altered gene expression and the severity of Crohn’s disease (CD), a prevalent form of IBD (7–9, 11–21). Thus, deciphering how the microbiota induces and sustains altered gene regulation in the intestine may reveal new targets for combatting disease. Intestinal epithelial cells (IECs) reside at the direct interface between the mammalian host and intestinal microbiota. These cells play a critical role in responding to commensal bacterial-derived signals to dynamically maintain barrier function through several adaptations, such as tight junction proteins, antimicrobial peptides, pattern recognition receptors, and immune modulatory cytokines (7, 22–30). However, the molecular pathways underlying how IECs integrate diverse microbiota-derived cues from the intestinal environment to regulate host responses remain poorly understood.

Epigenetic modifications can regulate gene expression in response to environmental cues, without altering the genetic sequence (31). Epigenetic modifications of the DNA and histones modify chromatin structure and transcription factor binding in a manner that enables transcriptional priming, activation, or repression. Therefore, epigenetics represent an important, potentially reversible, mechanism that may link genetic predisposition and environmental triggers in the pathogenesis of IBD (8). Sera and affected mucosa of CD patients exhibit changes in DNA methylation and global histone acetylation, although the dependence of these epigenetic modifications on environmental signals is not clear (32–37). In recent work, we identified an epigenetic regulator in IECs that integrates commensal bacterial-derived signals to regulate intestinal homeostasis and susceptibility to intestinal inflammation in the presence of commensal bacteria (38, 39). Thus, microbiota-induced alterations in epigenetic regulation may be a critical initiating trigger in the development and progression of IBD.

Here, we identify an ileal epithelial cell histone methylation signature in newly diagnosed, treatment-naive pediatric CD patients that has not been previously linked with IBD. Further, a significant proportion of the genes that make up this signature exhibit similar histone methylation patterns in the ilea of mice due to exposure to commensal bacteria, suggesting that epigenetic regulation of these targets may reflect a critical mechanism for how microbiota predispose to intestinal inflammation. Remarkably, the profile of several microbiota-sensitive targets correlated with severity of intestinal inflammation and disease in pediatric CD patients, revealing a clinically relevant epigenetic signature in the intestinal epithelium.

Results

Histone methylation profile of epithelium differentiates pediatric CD patients. Previous studies examining DNA methylation indicate that this modification is altered in serum and intestine of pediatric IBD patients (33–35, 37, 40–42); however, evaluation of links between epigenetics in IBD and regulation by the microbiota as well as extension to histone modifications has been limited. Thus, to undertake the initial investigation into whether histone modifications characterize the epithelial landscape in CD, ileal IECs from CD patients at disease onset and healthy controls were processed for ChIP sequencing (ChIP-seq) of the activating histone H3-lysine 4 trimethylation (H3K4me3) modification. IECs were compared between controls and CD patients, representing adolescents with an equal sex distribution (Table 1). Substantial differences in H3K4me3 enrichment were identified within many genes in the ilea of pediatric CD patients, as indicated at representative genes in control and CD patients (Figure 1, A and B). Interestingly, examination of H3K4me3 ChIP-seq data across all 8 patient samples per group highlighted 1,066 shared sites (1,038 genes), with significantly increased H3K4me3 in CD patients, and 539 sites (548 genes) characterized by significantly reduced H3K4me3 in CD patients compared with controls (Figure 1C). This H3K4me3 signature was significantly different across CD patients compared with controls, despite variability in clinical and histologic findings (Figure 1D and Table 1). Substantial differences in H3K4me3 enrichment were identified within many genes in the ilea of pediatric CD patients, as indicated at representative genes in control and CD patients (Figure 1, A and B). Interestingly, examination of H3K4me3 ChIP-seq data across all 8 patient samples per group highlighted 1,066 shared sites (1,038 genes), with significantly increased H3K4me3 in CD patients, and 539 sites (548 genes) characterized by significantly reduced H3K4me3 in CD patients compared with controls (Figure 1C). This H3K4me3 signature was significantly different across CD patients compared with controls, despite variability in clinical and histologic findings (Figure 1D and Table 1). While histologic features tracked with H3K4me3 variability in some of these targets (Supplemental Figure 1; supplemental material available online with this article; https://doi.org/10.1172/jci.insight.122104DS1), it did not account for all variability observed within the CD cohort (Figure 1D). Further examination revealed that CD-specific sites of differential H3K4me3 occurred predominately in promoters, although a significant proportion were also identified within genomic introns (Figure 1E).

Of the CD-specific epigenetically modified genes, 63 have at least 1 reported association with IBD based on the DisGeNET database (Supplemental Table 1) and 14 genes have been linked to CD-associated risk loci (43) (Supplemental Table 2). However, the majority of targets have not yet been examined in relation to IBD...
pathogenesis and prognosis. Clustered pathway analyses of genes that were characterized by increased H3K4me3 signatures in CD strongly enriched for broad immune processes, including genes involved in epithelial response to bacterial factors and cytokine signaling (i.e., ICAM-1, PTGS-2, IL7) (Figure 2, A and B). Interestingly, genes represented by sites of decreased H3K4me3 in CD predominantly regulate intestinal epithelial ion absorption as well as lipid and protein metabolism (Figure 2, C and D). Collectively, H3K4me3 status in IECs defined previously unrecognized genes and pathways that may be epigenetically regulated in treatment-naive pediatric IBD patients and could therefore represent novel targets for diagnostic evaluation and therapeutic intervention.

**Microbiota regulates histone methylation of a unique subset of CD targets.** Altered dynamics between IECs and the intestinal microbiota have been strongly linked to CD; however, whether epigenetic modifications in intestinal cells reflect integration of cues from microbiota-derived environmental signals is not known.

### Table 1. Characteristics of pediatric patients (newly diagnosed CD and controls)

<table>
<thead>
<tr>
<th>Patient demographics</th>
<th>Diagnostics</th>
<th>Histologic</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>Sex</td>
<td>Age (yr)</td>
</tr>
<tr>
<td>Crohn’s Disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Female</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>Male</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>Male</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>Male</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>Female</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>Female</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>Male</td>
<td>17</td>
</tr>
<tr>
<td>8</td>
<td>Female</td>
<td>13</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Male</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>Female</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>Male</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>Male</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>Female</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>Male</td>
<td>13</td>
</tr>
</tbody>
</table>

Crohn’s disease (CD) subgroups are defined by clinically affected (ileal CD [iCD] with or without deep ulcers [DU]) and unaffected (colon-only CD [cCD]) ilea, in comparison with non-IBD controls (Ctl).
Given that the absence and presence of microbiota cannot be manipulated in patients, animal models need to be employed to identify whether IECs exhibit an epigenetic signature that is sensitive to the presence of intestinal microbiota. Therefore, to test whether commensal bacteria that reside in the intestine regulate H3K4me3 in the distal small intestine, H3K4me3 signatures in ileal epithelial cells of conventionally housed (CNV-housed) barrier mice were compared with cells harvested from germ-free (GF) mice that lack commensal microbes. These global analyses demonstrated that the microbiota did, in fact, induce significant alterations in H3K4me3 in IECs (Figure 3A). Notably microbiota-sensitive epigenetic targets included, but were not limited to, genes that have been previously characterized in epithelial antimicrobial responses, such as \( \text{REG3G} \) and \( \text{NOS2} \) (44).

Similar to the distribution observed in IECs from CD patients (Figure 1E), microbiota-sensitive H3K4me3 sites were most commonly located within genomic promoters and introns (Figure 3B). Further, approximately 400 genes that exhibited differential H3K4me3 in response to commensal microbes were regulated similarly at the epigenetic level in IECs of pediatric CD patients (Figure 3, C–H) in a manner more significant than by chance (increased H3K4me3, \( P = 3.8 \times 10^{-6} \); decreased H3K4me3, \( P = 1.9 \times 10^{-16} \)). Of these overlapping targets, 327 genes exhibited significantly enriched H3K4me3 levels in IECs in CD as well as in response to microbes (Figure 3, C–E). Interestingly, \( \text{DUOX2} \), an oxidase whose expression highly correlates with microbial responses in CD (45), demonstrated marked H3K4me3 enrichment in both CD and in response to commensal bacteria (Figure 3D). Conversely, 97 genes that demonstrated loss of H3K4me3 enrichment in IECs from CD patients relative to controls also exhibited decreased H3K4me3 in IECs harvested from CNV mice relative to GF mice (Figure 3, F–H). The significance of most of these CD-associated genes characterized by microbiota-sensitive decreased H3K4me3, such as \( \text{CUBN} \), \( \text{THNSL2} \), and \( \text{PCK1} \), has not yet been evaluated in IBD (Figure 3G).

Epigenetic signatures reveal microbiota-sensitive networks in IBD. To determine whether microbiota-sensitive epigenetic targets in pediatric CD (Figure 3, C and F) form significant functional relationships, clustered
pathway analyses were performed for this refined subset of genes. Dysregulated pathways that were identified based on global epigenetic profiles in CD (Figure 2), such as regulation of the immune response, ion transport, and metabolic pathways, were highly enriched specifically within this microbiota subset (Figure 4, A and B), supporting that the microbiota is a critical regulator of functional epigenetically marked pathways in CD. In fact, targets in IECs of CD patients that were not responsive to the presence of microbiota (Figure 3, C and F) only exhibited enrichment for 76 pathways, whereas the smaller core set of overlapping microbiota-sensitive targets (Figure 3, C [327 genes] and F [97 genes]) enriched for 287 different pathways, highlighting high physiological potential for these microbiota targets in IBD.

Microbiota-sensitive genes that also demonstrate increased H3K4me3 in IECs from CD patients were highly enriched in pathways involving production of reactive oxygen species (DUOX2, RAC1/2; \( P = 4.2 \times 10^{-6} \)) and nitric oxide biosynthesis (ASS1, NOS2; \( P = 1.0 \times 10^{-4} \)) (Figure 4A). Genes regulating the innate mucosal barrier (MUC2, TGF-\( \beta \), IL16, IL2RB) were also enriched in this subset. Notably, targets less commonly evaluated in relation to IBD pathogenesis were also revealed. For instance, ephrin signaling that has been linked to immune signaling and epithelial homeostasis (46–49) was highlighted in this core subset (Figure 4A). Microbiota-sensitive targets characterized by loss of the H3K4me3 modification largely represent regulators of nutrients and metabolites in the intestine. Genes driving enrichment of these pathways included regulators of protein digestion and absorption (XPNPEP2, SLC7A7; \( P = 5.0 \times 10^{-4} \)) and bile transport (SLC10A2, ABCC2; \( P = 5.5 \times 10^{-4} \)) (Figure 4B). Epigenetic analyses also indicated that expression of multiple genes that mediate mineral absorption exhibited microbiota- and CD-specific decreases in H3K4me3 in IECs (Figure 4B). Thus, intestinal absorption deficiencies that occur in IBD may reflect an epithelial-intrinsic component in addition to alterations that occur subsequent to local damage and inflammation (50).

H3K4me3 represents an epigenetic mark of open or transcriptionally active chromatin where transcription factors bind DNA and alter transcriptional output. Therefore, to test whether microbiota-sensitive H3K4me3 sites in IECs of CD patients (Figure 3, C and F) share characteristic binding sites for known transcription factors, extensive global motif characterization was conducted. Examination of microbiota-sensitive sites with differential H3K4me3 enrichment in CD identified AP2 and NF-\( \kappa \)B consensus sequences as enriched in promoter regions with CD-specific H3K4me3 induction (Figure 4, C and D). MEF2A and multiple FOXO transcription factor–binding sites were also enriched in these sites. Collectively, these findings suggest that transcriptional regulation by factors such as AP2 and FOXO may be dysregulated in ileal IECs of IBD patients.

Microbiota-induced epigenetic targets associate with disease state and inflammation in CD. To investigate whether microbiota-sensitive epigenetically modified targets in CD track with transcriptional expression, RNA sequencing (RNA-seq) was performed on the same patient IECs that were analyzed for H3K4me3 by...
ChIP-seq. Analyses of normalized mRNA expression across patients highlighted several microbiota-sensitive epigenetic targets that exhibited significantly altered expression in CD patients (Figure 5, A and B, left). These genes demonstrated significant differential expression ($P \leq 0.05$, fold change $\geq 2$) that strongly paralleled their H3K4me3 epigenetic signature in CD patients (Supplemental Table 3). To address potential concerns about statistical power regarding RNA-seq data in this patient cohort, we expanded evaluation of microbiota-sensitive epigenetic targets to ileal biopsies from the larger pediatric RISK stratification cohort, which includes 179 pediatric CD patients and 38 control patients. Notably, the majority of genes identified in our initial analyses also exhibited marked differences in expression across this larger cohort of pediatric CD patients. Further, these targets also demonstrated a striking gradient in expression across disease subtypes in the RISK cohort, from no inflammation, to no macroscopic ileal inflammation (colon-only CD), to active ileal inflammation (ileo CD) (Figure 5, A and B, right). A similar trend in expression was
also observed for several of these targets in relation to clinical CD subgroups and corresponding histologic severity in the smaller pediatric cohort (Supplemental Table 4). Thus, the level of expression of these relatively few microbiota-sensitive targets in IECs seems to distinguish ileal-defined CD subtypes.

In order to further examine the relationship between expression of these key targets and intestinal inflammation in IBD patients, their expression was examined relative to the inflammatory marker S100 calcium-binding protein A8 (\(S100A8\)) that correlates with CD severity (51). Ileal IEC expression of the majority of these genes (69%) exhibited a significant linear correlation with level of \(S100A8\) expression.

Expression of representative CD enriched candidate genes, such as \(DUOXA2\), correlated strongly with epithelial expression of the IBD inflammatory biomarker (Figure 5C). Further, this positive relationship between microbiota-sensitive epigenetic targets was shared when expression was compared with \(S100A8\) across the larger RISK cohort (Figure 5D). Likewise, epithelial expression of downregulated genes, such as \(CUBN\), negatively correlated with \(S100A8\) levels (Figure 5C), and this relationship was similarly conserved when expanded to compare the full RISK cohort (Figure 5D).

Given that these microbiota-sensitive genes were characterized with significant differences in histone methylation, association between this epigenetic signature and disease was further evaluated. Principle component analysis of H3K4Me3 ChIP-seq data demonstrated more distinct separation between control and CD cohorts relative to the separation observed with RNA-seq data for these targets (Figure 6A). Therefore, to test the relationship between histone methylation at microbiota-sensitive targets and inflammation, correlation between \(S100A8\) and H3K4me3 levels was examined. Several genes exhibited significant correlation between levels of H3K4me3 and \(S100A8\) when examining control and CD samples (Figure 6B). Interestingly, of the
targets characterized as having a significant correlation between H3K4me3 and S100A8, only 10% exhibited significant differential expression in the same patients, whereas the majority (90%) would not have been identified by expression analyses (Figure 6C). In fact, comparing this set of genes across the RISK cohort further confirmed that the majority of these genes (84 of 122; Supplemental Table 5) exhibited minimal differences in expression with disease. As an example of this from this epigenetically modified subset, histone methylation within IL7 and THNSL2 loci correlated with the level of intestinal inflammation in CD patients, despite a lack of significant change in expression (Figure 6D). Thus, microbiota-sensitive targets have been identified that exhibit epigenetic profiles that associate with CD, and, importantly, traditional transcriptomics would be less likely to detect these potential candidates (Supplemental Table 6).

**Discussion**

By identifying sites of H3K4me3 enrichment in patient IECs, this work reveals that the histone methylation profile of the ileal epithelium can effectively distinguish pediatric CD patients from patients that lack intestinal inflammation. Importantly, employing newly diagnosed, treatment-naive pediatric patients circumvented the common and significant challenge of distinguishing CD-associated alterations in the host from effects resulting from therapeutic exposure. Epigenetic dysregulation in IECs of CD patients is strongly enriched in genes that regulate cytokine signaling, metabolic homeostasis, and regulation of reactive oxygen species. Previous epigenetic studies in adult IBD have focused on DNA methylation and
identified different subsets of genes than those observed here for histone methylation (33–37, 40–42). Thus, chromatin alterations in histone methylation form a distinct epigenetic signature in IECs that differs in IBD patients compared with healthy controls at the time of their initial diagnosis. Despite relatively small patient numbers in this newly diagnosed, treatment-naive pediatric cohort, there was remarkable commonality in regions of altered histone methylation relative to controls. Given that samples used for ChIP and histologic analyses were distinct, it is not surprising that histologic interpretation did not always parallel the magnitude of epigenetic variability within this smaller cohort. Increased sample number and complimentary studies using inflamed and noninflamed mucosa from the same patient will enable future discovery of more subtle epigenetic differences and comprehensive histologic comparisons.

Beyond defining a histone methylation profile in the ilea of CD patients, this work goes further to predict which CD-specific epigenetic marks in the epithelium may be sensitive to the microbiota. Most striking is that the histone methylation signature itself often correlated with a molecular marker of inflammation in IBD. Many of these genes, in fact, do not exhibit significant differential expression and would, therefore, not be identified through transcriptomics to be clinically linked to IBD, highlighting that epigenetic analyses represent a complimentary approach for identification of epithelial molecular abnormalities. For example, \( IL7 \), a gene that is characterized as epigenetically modified in CD epithelium, regulates mucosal lymphocytes (52) and is produced by epithelial cells in response to commensal microbes (53), suggesting that the threshold to induce \( IL7 \)-dependent intestinal inflammation may be lower in CD patients or affected by other epigenetic modifications. As such, these types of targets may represent a group that is epigenetically primed by the microbiota during the onset of disease and could more likely exhibit transient alterations in expression or even enhanced response following exposure to additional disease-associated environmental triggers in the course of IBD.

Binding sites for AP2 and FOXO are enriched at microbiota-sensitive H3K4me3 loci in CD ileal epithelium, suggesting that altered regulation by these families of transcription factors may be particularly critical in early stages of IBD pathogenesis and may represent a potential target for therapy. Similarly, NF-\( \kappa B \) motifs are also enriched at regions characterized by increased H3K4me3, highlighting that the CD epigenetic profile identifies critical immunoregulatory regions in the genome that may be primed by the microbiota to drive enhanced inflammatory responses in reaction to subsequent environmental triggers. Progression of
IBD has been linked to the composition of commensal bacteria, and, remarkably, we found that expression and/or H3K4me3 levels of specific microbiota-sensitive targets correlate with those in CD clinical subgroups and with S100A8, a marker of intestinal inflammation. Some of these genes, such as DUOX2 and DUOXA2, have been found to play a role in bacterial sensing and exhibit increased expression in the ilea of CD patients (45). In fact, DUOX2 drives ileitis in transgenic mice lacking glutathione peroxidase 1 and 2 (54). Many aspects of intestinal physiology and immune homeostasis are sensitive to the microbiota. Thus, the microbiota-sensitive alterations characterized using this murine GF comparison reflect a combination of both direct and indirect effects of intestinal colonization and may not fully recapitulate exposure to the CD-associated microbiota compositions in humans. However, identification of microbiota-sensitive sites in these targets with H3K4me3 alterations in CD highlights regions of epigenetic dysregulation that can now be examined in relation to microbiota composition.

Notably, this study highlights targets and pathways that are less commonly considered with regards to IBD pathogenesis. For instance, ephrin signaling genes, which have indeed been linked to immune signaling and epithelial homeostasis (46–49), are now shown to represent exciting targets that are epigenetically modified due to the intestinal microbiota. Thus, microbial-primed ephrin signaling in the intestine may be a critical, but yet unexplored, early trigger in CD. In addition to genes characterized by elevation of the activating H3K4me3 mark, many genes were also identified with frequent reduction of this modification in the ileal epithelium of CD patients. Altered intestinal electrolyte transport has been described in the pathogenesis of IBD, particularly with regards to understanding disease-associated diarrhea (55). Sodium transport, in particular, appears to exhibit significant epigenetic dysregulation, and previous work has, in fact, suggested a potential role for intestinal sodium regulation in IBD (55). As an example, the SLC9A3R1 gene encoding a sodium-hydrogen regulator is characterized in CD with decreased H3K4me3 enrichment and is downregulated in IBD (56). Epigenetic analyses also indicate that CUBN, a gene encoding a vitamin B12 receptor, displays a remarkably consistent decrease in histone methylation and expression in CD. Interestingly, vitamin B12 deficiency is a common deficiency in IBD patients, and these patients may require vitamin B12 supplementation (57). Electrolyte and vitamin imbalances in IBD have been commonly attributed to intestinal damage caused by local inflammation (50, 58). However, our data suggest that the microbiota could trigger epigenetic alterations that reflect downregulation of epithelial absorption in the intestine. In fact, dysbiosis and sodium flux often do not normalize with otherwise successful anti-TNF therapy, even in asymptomatic CD patients (59, 60). Therefore, the microbiota may be a key factor that sustains altered ion and nutrient homeostasis in the intestine through epigenetic restructuring early in the course of IBD development.

Collectively, this microbiota-dependent H3K4me3 signature may identify targets that are epigenetically modified due to microbiota exposure during the onset of disease. This work forms the critical entry point defining how microbiota-derived signals may be integrated by the host, emphasizing a fundamental role of epigenetics in priming epithelium for overt clinical disease with subsequent disease-associated environmental triggers. As progression of IBD can be tightly influenced by a patient’s unique microbiota composition, identification of H3K4me3 sites highlights regions of epigenetic dysregulation that can now be examined mechanistically in relation to other histone modifications as well as human microbiota using gnotobiotic colonization approaches. These studies will direct more specific approaches for individualized prevention and/or treatment of IBD based on modulation of epigenetically dysregulated pathways. In addition, assessing epithelial epigenetic targets in intestinal biopsies may provide a sensitive marker of intestinal inflammation and distinguish patients that could be more responsive to microbiota-based therapies.

**Methods**

**Patients.** Children and adolescents undergoing colonoscopy were recruited prospectively for participation in the IBD biorepository protocol between 2015 and 2017 at CCHMC. Adolescents, between the ages of 12 and 18 years, with treatment-naive CD were selected for this study, along with age- and sex-matched controls. Control patients presented due to intestinal symptoms, and colonoscopy was warranted as part of their evaluation. These patients were found to have no evidence of intestinal disease based on a combination of endoscopy, histology, lab testing, and imaging. Molecular analyses were performed on pooled biopsy samples collected together by endoscopy from sites in the ileum that exhibited representative gross pathology for the patient, ranging from no visible disease to severe inflammation, with or without ulceration. Collecting direct biopsies of ulcers was avoided. Histological analyses were performed using separate biopsies collected from multiple locations throughout both the small and large intestine.
Animals. Mice were housed up to 4 per cage in a ventilated isolator cage system in a 12-hour-light/dark cycle, with free access to water and chow. GF mice were maintained in plastic isolator units in the CCHMC Gnotobiotic Mouse Facility, fed autoclaved feed and water, and routinely monitored to ensure the absence of microbial contamination. All experiments were performed with age- and sex-matched C57BL/6 mice.

IEC isolation. For isolation of IECs, samples were incubated with 1 mM EDTA and 2 mM DTT while shaking at 37°C at a 45° angle for 10 minutes to obtain high epithelial purity with similarly small amounts of CD45+ cells (≤5%) in both control and CD samples. Supernatant with IECs was pipetted into a new 50-ml tube and pelleted at 500 g for 5 minutes at 4°C. A portion of IECs was processed for ChIP-seq, and 100 μl RNAlater (Ambion) was added to the remainder for RNA-seq analyses. Mouse IECs were harvested from 10 cm of the distal small intestine. The intestine was opened longitudinally, washed, and incubated in buffer containing 1 mM EDTA, 2 mM DTT, and 5% FBS in PBS for 10 minutes, while shaking at 37°C at a 45° angle to isolate IECs.

ChIP-seq. Fresh formaldehyde was added to each IEC sample for a final concentration of 1% and mixed end over end at room temperature for 10 minutes to crosslink proteins to DNA. Cells were lysed with a Triton X-100 and Igepal buffer (0.25% Triton C-100, 0.5% Igepal, 10% glycerol, 1 mM EDTA, 140 mM NaCl, and 50 mM HEPES). Nuclei were then washed and chromatin sheared to 150- to 500-bp size in a 0.1% SDS in TE buffer using a S220 Covaris Sonicator. Immunoprecipitation was carried out with a SX-8G IP-STAR robot (Diagenode) and an antibody optimized for H3K4me3 as described previously (61, 62). Sequencing was performed using an Illumina HiSeq 2500. All samples exhibited high-quality sequencing, reads were mapped to human genome hg19 or mouse genome mm9 with Bowtie, and peaks were identified with MACS2 (63, 64). Gene ontology mappings are from Ensembl BioMart (65).

RNA-seq. RNA was isolated using the RNeasy Kit (Qiagen). Directional polyA RNA-seq was performed by the Genomics, Epigenomics, and Sequencing Core at the University of Cincinnati. NEBNext Poly(A) mRNA Magnetic Isolation Module (New England BioLabs) was used for polyA RNA purification with the Apollo 324 system (WaferGen) and PrepX PolyA script for automated polyA RNA isolation. To study differential gene expression, individually indexed and compatible libraries were proportionally pooled (~25 million reads per sample in general) for clustering in the cBot system (Illumina). Libraries at the final concentration of 15 pM were clustered onto a single-read flow cell using the Illumina NEBNext Ultra Directional RNA Library Prep Kit and sequenced to 50 bp on the Illumina HiSeq system. Sequence reads were aligned to the Hg19 genome by using the Illumina sequence analysis pipeline created by the Laboratory for Statistical Genomics and Systems Biology at the University of Cincinnati. Sequence reads were aligned to the reference genome using the TopHat2 aligner, and reads aligning to each known transcript were counted using Bioconductor packages for next-generation sequencing data analysis (66, 67). Differential expression analysis was performed using the negative binomial statistical model of read counts as implemented in the edgeR Bioconductor package (68). Significantly differentially expressed genes (P < 0.05; fold change ≥ 2) were compared with previously published RNA-seq analyses from the Crohn's & Colitis Foundation–sponsored multicenter RISK study that enrolled pediatric control patients (n = 38) and CD patients with and without endoscopic ileal inflammation (n = 179) (18).

Computational analyses. Diagnostic confirmation of whether a subject had CD or not (control) was conducted prior to analyses. Global sequencing data were analyzed with Biowardrobe (69), and differential ChIP-seq data were analyzed with replicates in THOR (70). Sample signal was normalized based on peaks at housekeeping genes (EMC7, GPI, PSMB4, RAB7A, VCP, CHMP2A, VPS29). Differential peaks between CD and control groups with all replicates were detected using a Hidden Markov Model approach. Housekeeping genes were individually inspected to ensure no differential enrichment was detected. Significant differential enrichment was defined as Benjamini-Hochberg–corrected P < 0.05 and fold change ≥ 1.75. For GF versus CNV analyses, to restrict to the most significant targets, a log10(P value) ≤ 4.5 was used. Pathway enrichment analyses with clustering was performed using ClueGo in Cytoscape with a Bonferroni-corrected P < 0.05 (71). Functional groups were created with a x score greater than 0.4. Unclustered pathway analysis using Toppgene was employed with a false discovery rate < 0.05 (72). PscanChIP was used to identify transcription factor–binding site motifs within 150 bp of the center of the differential peaks identified (73). JASPAR and TRANSFAC databases were used to rank based on P value. The DisGeNET database (5/2017) identified individual genes associated with IBD, a specificity score, and the number of NCBI citations. PCA plots were made in GeneSpring GS (Agilent) based on log1-transformed ChIP-seq and RNA-seq data baselined to median of all samples. RNA-seq and ChIP-seq data have been deposited in GEO and are accessible through GEO series accession numbers GSE117875 and GSE117889, respectively.
Statistics. All inclusion/exclusion criteria were preestablished. THOR assigned $P$ values employing biological ChIP-seq replicates, and $P$ values were considered significant at $P < 0.05$. Significance of overlap of genes between groups was calculated based on hypergeometric distribution, with 17,094 possible protein coding orthologs. For linear correlation analyses, normalized and transformed gene expression and methylation data were plotted against $SI00A8$ expression for each patient sample. Pearson correlation was used to determine $r$ coefficient and corresponding $P$ values. All data met the assumptions of the statistical tests used.

Study approval. Studies in humans were performed with approval by the CCHMC Institutional Review Board. Parents and adult patients, age 18 and older, provided consent, and all patients, age 11 and older, provided assent. All murine experiments were performed according to the animal experimental guidelines upon approval of the Institutional Animal Care and Use Committee at CCHMC.

Author contributions
TA, DK, and LAD designed the studies. DK, VW, SJ, JW, and JM carried out experiments. DK, MK, YH, RK, and AB assisted with computational analyses. LAD, BJA, MCD, JSH, JFM, RNB, MCS, TDW, and SK provided patient samples and clinical expertise. NS, MJR, and MH provided critical reagents and tools. TA and DK analyzed the data and wrote the manuscript.

Acknowledgments
We thank the S. Way, D. Haslam, and J. Qualls labs (Infectious Disease, CCHMC) and the H. Deshmukh lab (Pulmonology, CCHMC) for useful discussions and members of the Alenghat lab for critical reading of the manuscript. The RISK committee provided patient samples and clinical expertise. We thank CCHMC Veterinary Services, the CCHMC DNA Sequencing Core, and the University of Cincinnati Genomics, Epigenomics, and Sequencing Core for services. We also thank RISK investigators for patient recruitment and the Crohn’s & Colitis Foundation for RISK support. This research is supported by the National Institutes of Health (DK093784, DK114123, and DK116868 to TA; DK098231 to LAD; and T32 DK007727 to DK), and a Crohn’s & Colitis Foundation/Janssen/AGA award to TA. TA holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund and is a Pew Scholar in the Biomedical Sciences, supported by the Pew Charitable Trust. This project is supported in part by Public Health Service grant P30 DK078392 and the CCHMC Trustee Award and Procter Scholar’s Program.

Address correspondence to: Theresa Alenghat, 3333 Burnet Ave., MLC 7038, Cincinnati, Ohio, USA. Phone: 513.803.7498; Email: theresa.alenghat@cchmc.org.


