

Two-stage Genome-wide Methylation Profiling in Childhood-onset Crohn's Disease Implicates Epigenetic Alterations at the VMPI/MIR21 and HLA Loci

Alex T. Adams, MBChB,* Nicholas A. Kennedy, MBBS,* Richard Hansen, PhD,[†] Nicholas T. Ventham, MBBS,* Kate R. O'Leary, BSc,* Hazel E. Drummond, BSc,* Colin L. Noble, MD,* Emad El-Omar, PhD,[‡] Richard K. Russell, PhD,[‡] David C. Wilson, MD,[§] Elaine R. Nimmo, PhD,* Georgina L. Hold, PhD,[†] and Jack Satsangi, DPhil*

Background: As a result of technological and analytical advances, genome-wide characterization of key epigenetic alterations is now feasible in complex diseases. We hypothesized that this may provide important insights into gene-environmental interactions in Crohn's disease (CD) and is especially pertinent to early onset disease.

Methods: The Illumina 450K platform was applied to assess epigenome-wide methylation profiles in circulating leukocyte DNA in discovery and replication pediatric CD cohorts and controls. Data were corrected for differential leukocyte proportions. Targeted replication was performed in adults using pyrosequencing. Methylation changes were correlated with gene expression in blood and intestinal mucosa.

Results: We identified 65 individual CpG sites with methylation alterations achieving epigenome-wide significance after Bonferroni correction ($P < 1.1 \times 10^{-7}$), and 19 differently methylated regions displaying unidirectional methylation change. There was a highly significant enrichment of methylation changes around GWAS single nucleotide polymorphisms ($P = 3.7 \times 10^{-7}$), notably the HLA region and *MIR21*. Two-locus discriminant analysis in the discovery cohort predicted disease in the pediatric replication cohort with high accuracy (area under the curve, 0.98). The findings strongly implicate the transcriptional start site of *MIR21* as a region of extended epigenetic alteration, containing the most significant individual probes ($P = 1.97 \times 10^{-15}$) within a GWAS risk locus. In extension studies, we confirmed hypomethylation of *MIR21* in adults ($P = 6.6 \times 10^{-5}$, $n = 172$) and show increased mRNA expression in leukocytes ($P < 0.005$, $n = 66$) and in the inflamed intestine ($P = 1.4 \times 10^{-6}$, $n = 99$).

Conclusions: We demonstrate highly significant and replicable differences in DNA methylation in CD, defining the disease-associated epigenome. The data strongly implicate known GWAS loci, with compelling evidence implicating *MIR21* and the HLA region.

(*Inflamm Bowel Dis* 2014;20:1784–1793)

Key Words: Crohn's disease, Epigenetics, *MIR21*

The last decade has seen tremendous success in identifying genetic loci associated with inherited susceptibility to Crohn's disease (CD), with 140 loci identified in the most recent GWAS

meta-analysis.¹ However, these determinants collectively explain only an estimated 13.6%¹ of disease variability, and the biological variation each confers is unclear. The importance of noninherited factors in pathogenesis has been highlighted by studies on the increasing incidence of CD, especially in children,^{2,3} and in the developing world,^{4,5} and by a greater understanding of the effects of gut microbiota and diet on risk.^{6,7} A critical objective for CD research is to characterize the interaction between genetic and environmental factors.

Epigenetic alteration has emerged as a potential mechanism through which these interactions may occur.⁸ Developments allowing rapid assaying of cytosine methylation at nearly 5×10^5 positions, and insights into confounding effects in study design⁹ have provided the impetus to build on promising pilot data from previous generation technology to allow epigenome-wide association studies (EWAS) to become a valuable complement to the more mature GWAS.¹⁰ Epigenome mapping has been used to identify DNA regulatory elements, explore cancer biology, and provide a growing body of findings in complex diseases, such as rheumatoid arthritis,¹¹ multiple sclerosis,¹² type 2 diabetes, and obesity.^{13,14}

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.ibdjournal.org).

Received for publication July 8, 2014; Accepted July 8, 2014.

From the *Gastrointestinal Unit, Centre for Genetics and Experimental Medicine, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, United Kingdom; [†]Gastrointestinal Research Group, Division of Applied Medicine, University of Aberdeen, Aberdeen, United Kingdom; [‡]Department of Paediatric Gastroenterology, Royal Hospital for Sick Children, Glasgow, United Kingdom; and [§]Paediatric Gastroenterology and Nutrition, Child Life and Health, University of Edinburgh, Royal Hospital for Sick Children, Edinburgh, United Kingdom.

The authors have no conflicts of interest to disclose.

A.T. Adams, N.A. Kennedy, and R. Hansen are co-first authors. G.L. Hold and J. Satsangi as co-senior authors.

Reprints: Jack Satsangi, DPhil, Gastrointestinal Unit, Centre for Genetics and Experimental Medicine, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, EH4 2XU, United Kingdom (e-mail: j.satsangi@ed.ac.uk).

Copyright © 2014 Crohn's & Colitis Foundation of America, Inc.

DOI 10.1097/MIB.0000000000000179

Published online 20 August 2014.

Although the relationship between methylation and gene expression and function is incompletely understood, relevant modifying influences include age, ethnicity, smoking, gut microbiota, and diet.¹⁵⁻¹⁸ DNA-binding factors can directly influence methylation, and in turn, altering methylation can directly influence expression.^{19,20} We hypothesized that identification of altered levels of methylation, which are significantly associated with disease state, whether predating or following disease, offers the potential for discovering new pathways integral to the disease process and for predicting disease status.

As the 2 pediatric stages were of similar design and showed strong replication, they were amenable to a single joint analysis, an approach recommended for suitable data sets⁹ because of the increase in power it gives in discovering CpG sites with disease-associated methylation changes. Seven of the most significant CpG sites implicated by the combined pediatric analysis were assessed by pyrosequencing in an adult cohort with established CD, and disease-associated expression changes were explored (Fig. 1A).

MATERIALS AND METHODS

Study Design

Illumina 450k DNA methylation analysis was performed separately in 2 pediatric cohorts (see Table, Supplemental Digital Content 1, <http://links.lww.com/IBD/A557> that contains demographic data for pediatric cohorts). Patients in the first cohort were treatment-naive, newly diagnosed CD cases. Patients in the second cohort were children with an established diagnosis of CD. Both cohorts used symptomatic controls, which were investigated by colonoscopy, but in whom no pathology or abnormality was discovered at initial investigation or subsequently. Linear discriminant analysis (LDA) was applied to data generated in the discovery cohort to identify biomarker candidates, which were then tested in the replication cohort.

Pediatric Patient Selection and Ethics

Pediatric samples were collected from centers across Scotland. The Bacteria in Inflammatory bowel disease (IBD) in Scottish Children Undergoing Investigation before Treatment (BISCUIT) study provided peripheral blood leukocyte DNA for the discovery cohort from 18 treatment-naive newly diagnosed patients and 18 matched nondiseased controls from Aberdeen, Glasgow, and Dundee. Controls had been rigorously investigated for gastrointestinal symptoms but did not have or subsequently develop any organic gastrointestinal pathology, including IBD.²¹ The replication cohort comprised DNA samples from 18 children with established CD supplied by the Paediatric-onset IBD Cohort and Treatment Study (PICTS),²² analyzed against a second set of 18 controls from the BISCUIT study. Within both cohorts, patients and controls were matched for age and gender. The BISCUIT study was approved by the North of Scotland Research Ethics Committee (09/S0802/24) and PICTS by ethics committees

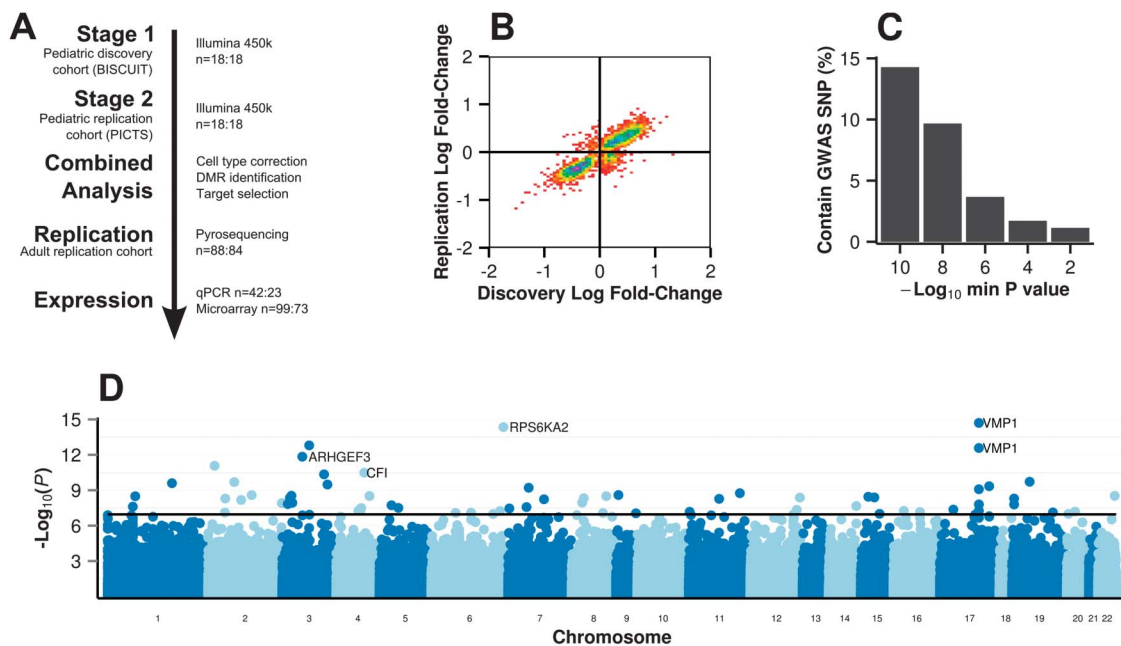


FIGURE 1. A, Schematic representation of study design, n = ratio of CD samples to control samples. B, Log fold-change (\log_2 mean methylation in CD/mean methylation in controls) for all probes with nominally significant (uncorrected $P < 0.05$, $n = 3620$) methylation changes in both pediatric cohorts. Data are binned; colors of shorter wavelength indicate higher frequency. C, A 50 kb of genomic regions are more likely to contain GWAS SNPs for IBD or CD if they also contain more significant disease-associated methylation changes ($P = 3.66 \times 10^{-7}$). D, Manhattan plot of disease-associated methylation changes; horizontal line corresponds to significance after Bonferroni correction.

at participating centers (Edinburgh, Glasgow, Aberdeen, and Dundee—LREC 2002/6/18). Written informed consent was obtained from the parents of all participating children. Informed assent was also obtained from older children capable of understanding the nature of the study. Demographics for both cohorts are in Table, Supplemental Digital Content 1, <http://links.lww.com/IBD/A557>.

Genome-wide Methylation Profiling

Peripheral blood leukocyte DNA was bisulfite converted and analyzed using the Illumina Human Methylation 450k platform (Illumina, San Diego, CA)²³ with cases and controls distributed across chips. Probes were filtered to remove any with a detection *P* value of ≥ 0.01 , those from sex chromosomes, and those that had single nucleotide polymorphisms (SNPs) with a minor allele frequency of ≥ 0.01 in the European population in the 1000 Genomes Project²⁴ within CpGs assayed by the array. Samples were removed if there was a gender mismatch or if more than 5% of probes failed.

Data were corrected using background removal and quantile normalization in the lumi R package^{25,26} followed by beta mixture quantile dilation.²⁷ Batch effects were controlled for using ComBAT.²⁸

Differential leukocyte counts from the same day that DNA samples taken were available for 24 patients and 19 controls. Linear models were created for all Illumina 450k probes and disease status in these samples. Probes were selected, which had *F* test with *P* values $< 10^{-8}$ but a *P* value for disease association of > 0.05 . Combinations of 100 probes were tested, and the best probe set was used to predict the differential cell counts for samples without measured differential leukocyte counts. This model is similar to that described by Houseman et al.²⁹

Analysis of the methylation status of cases versus controls was performed using limma³⁰ in R using linear modeling of beta values with measured or predicted neutrophil, other granulocyte, lymphocyte, and monocyte numbers as covariates.

The Benjamini–Hochberg false discovery rate (FDR)³¹ was calculated for each probe, with a FDR corrected *P* < 0.05 used to define significance in analysis of broader methylation patterns, such as identifying differentially methylated regions. For significance of individual probes, the more conservative Bonferroni correction was used.

Differently Methylated Regions

Differently methylated regions (DMRs) were defined using a reimplementation of the probe lasso DMR-calling technique used by the ChAMP pipeline³² in R. This defined a DMR as 3 or more sequential probes with significant (FDR adjusted *P* < 0.05) unidirectional methylation changes falling within the lasso distance threshold. The distance threshold uses a base size of 2 kb, modified based on the methylation patterns of local genomic features.³³

Replication of Methylation Findings in Adults

Twenty whole-blood DNA samples from adults with CD were recruited from gastroenterology clinic while their disease was quiescent, with 20 healthy controls collected during the same time period (LREC 2000/4/192). In addition, methylation changes in

VMP1/MIR21 were replicated in an extended adult replication cohort of 87 adults with CD and 85 healthy controls (see Table, Supplemental Digital Content 2, <http://links.lww.com/IBD/A558> that contains demographic data), of which the smaller group was a subset.

DNA was bisulfite converted with EZ-96 DNA Methylation Kits (Zymo Research, Irvine, CA), assays were designed using PyroMark Assay Design Software (version 2.0.1.15; Qiagen, Dusseldorf, Germany) and primers (see Table, Supplemental Digital Content 3, <http://links.lww.com/IBD/A559> that contains primer sequences) were ordered from Sigma-Aldrich (St. Louis). Sequencing was performed on a PyroMark Q96 ID machine (Qiagen) and analyzed in R v3.0.1.

VMP1 and MIR21 Expression

MIR21 primary transcript (pri-miR-21, primer details in Table, Supplemental Digital Content 3, <http://links.lww.com/IBD/A559>) was assayed by qPCR, with all patients and controls giving written, informed consent (LREC 06/S1101/16, LREC 2000/4/192). Suitable patients with CD were prospectively recruited from gastroenterology clinic and endoscopy lists, and healthy controls were recruited from volunteers. Blood samples were taken using a 21-gauge butterfly needle and 9 mL K3 EDTA vacuette (Greiner, Germany) and stored at 4°C for up to 2 hours. Total RNA was then extracted from 1.5 mL whole blood using QIAamp RNA blood mini kit (Qiagen) and stored at -80°C . cDNA was converted using SuperScript Vilo cDNA synthesis kits (Invitrogen, Carlsbad) and analyzed on a Corbett Rotor-Gene 6000 (Qiagen) with DyNAmo Flash SYBR green reagent (Thermo Scientific, Waltham). Expression of pri-miR-21 was normalized to reference gene *TBP*, (primer details in Table, Supplemental Digital Content 3, <http://links.lww.com/IBD/A559>) after initial optimization against 4 reference genes (*GAPDH*, *TBP*, *SDHA*, and *ACTB*) and analyzed by the $\Delta\Delta\text{Ct}$ method in R. Expression in inflamed and uninfamed intestinal biopsies from patients with CD, ulcerative colitis, and healthy controls was assessed in previously reported microarray data.³⁴

Linear Discriminant Analysis

LDA of methylation beta values in the pediatric discovery cohort was used to create biomarkers for the presence of CD, using the LDA function in the R package “MASS.”³⁵ All probes with FDR adjusted *P* values < 0.05 in the discovery cohort (see Table, Supplemental Digital Content 4, <http://links.lww.com/IBD/A560> that lists pediatric discovery results with FDR adjusted *P* < 0.05) were used as covariates, regardless of performance in the replication cohort, with each model including 2 probes. Models were tested using the pediatric replication cohort methylation beta values.

GWAS Colocalization

For range thresholds between 25 kb and 4 Mb, the lowest *P* value within range of each GWAS SNP for CD or IBD¹ was compared by Wilcoxon rank sum test to 1000 randomly selected bins of the same genomic size, matched for probe density.

GO Term Enrichment Analysis

The R package Goseq³⁶ was used as described by Goeleher et al³⁷ to correct for bias introduced by the variation in Illumina 450k probes per gene and analyze gene ontology (GO) term enrichment. The number of probes sharing each gene symbol annotation and whether that annotation covers at least one differentially methylated probe (FDR corrected $P < 0.05$) was used to create a probability weighting function, used in the GO term enrichment analysis.

RESULTS

Pediatric Illumina 450k

Nine probes with FDR corrected $P < 0.05$ were identified in the discovery cohort, 8 of which achieved nominal significance in the replication cohort (see Table, Supplemental Digital Content 4, <http://links.lww.com/IBD/A560> that lists pediatric discovery results with FDR adjusted $P < 0.05$). Correlation between cohorts was high, with 89% of probes reaching nominal significance in both cohorts showing the same direction of change (Fig. 1B).

Combined analysis of the 2 cohorts identified 1319 probes with significant FDR adjusted P values. Of these, 65 CpGs (Table 1, Fig. 1D) retained epigenome-wide significance after the more stringent Bonferroni correction for multiple testing. At these probes, there were absolute differences in mean methylation between CD and control groups of up to 16% (mean 6%), with 89% of probes showing hypomethylation in CD. Methylation variance in CD at these probes was greater than controls in 65%, 53% of which

differences were statistically significant (17% for probes where variance was greater in controls). The mean ratio of variances between CD and control were 2.4 and 1.4 for probes where CD or control, respectively had the greater variance.

Nineteen DMRs were identified (Table 2) from the 1319 probes with significant FDR adjusted P values. These regions involve several genes in pathways relevant to CD including *TNF* within the HLA region, *MIR21*, Toll-like receptor signaling (*TOLLIP*), and apoptosis (*VMP1*, *PRF1* and *DIABLO*).

Colocalization of significant Illumina 450k methylation changes with GWAS SNPs was found across distance thresholds between 25 kb and 4 Mb, with peak correlation between 50 kb ($P = 3.66 \times 10^{-7}$) and 100 kb ($P = 2.41 \times 10^{-7}$) in line with previously published work,³⁸ (Fig. 1C). This relationship remained significant if *VMP1/MIR21* was excluded from analysis.

GO term analysis found 170 significantly enriched terms (see Table, Supplemental Digital Content 6, <http://links.lww.com/IBD/A562> that lists all enriched GO terms) in the combined pediatric data, including terms related to NF- κ B signaling, apoptosis, and the JAK-STAT cascade (FDR corrected P values 6.32×10^{-6} , 2.31×10^{-4} and 3.23×10^{-3}).

Adult Replication with Pyrosequencing

Pyrosequencing assays were designed for a series of 7 regions corresponding to significant disease-associated methylation changes in the combined pediatric Illumina 450k data. Methylation changes were assayed in a group of 20 adults with CD and 20 controls with resultant P values between 0.004 and 2×10^{-5} (Fig. 2B). As with

TABLE 1. The 15 Most Significant Individual Probes from the Combined Analysis of the Combined Pediatric Data

Probe ID	Probe Details			Discovery Cohort		Replication Cohort		Combined Analysis		
	Chr	Coordinate ^a	Gene	$\Delta\beta^b$	P	$\Delta\beta$	P	$\Delta\beta$	P	Bonferroni
cg12054453	17	57915717	<i>VMP1</i>	-0.15	2.19×10^{-10}	-0.12	1.55×10^{-5}	-0.13	1.97×10^{-15}	8.86×10^{-10}
cg17501210	6	166970252	<i>RPS6KA2</i>	-0.11	1.17×10^{-8}	-0.10	3.73×10^{-6}	-0.11	4.47×10^{-15}	2.01×10^{-9}
cg12992827	3	101901234	—	-0.12	3.82×10^{-9}	-0.08	5.14×10^{-5}	-0.10	1.59×10^{-13}	7.13×10^{-8}
cg16936953	17	57915665	<i>VMP1</i>	-0.17	9.11×10^{-10}	-0.14	9.25×10^{-5}	-0.16	2.65×10^{-13}	1.19×10^{-7}
cg04389058	3	57041402	<i>ARHGEF3</i>	-0.08	3.32×10^{-8}	-0.06	4.69×10^{-5}	-0.07	1.44×10^{-12}	6.48×10^{-7}
cg02560388	2	11969958	—	-0.04	1.93×10^{-6}	-0.05	1.60×10^{-5}	-0.04	8.45×10^{-12}	3.80×10^{-6}
cg00382138	4	110723299	<i>CFI</i>	-0.08	1.88×10^{-6}	-0.08	2.60×10^{-5}	-0.07	3.27×10^{-11}	1.47×10^{-5}
cg01671681	3	155421735	<i>PLCHI</i>	-0.05	4.43×10^{-6}	-0.05	5.23×10^{-5}	-0.05	4.52×10^{-11}	2.03×10^{-5}
cg19821297	19	12890029	—	-0.08	3.35×10^{-6}	-0.08	7.09×10^{-5}	-0.08	1.91×10^{-10}	8.60×10^{-5}
cg27049094	2	75067716	<i>HK2</i>	0.07	9.03×10^{-6}	0.07	8.62×10^{-6}	0.07	2.04×10^{-10}	9.18×10^{-5}
cg27310092	1	151945663	—	-0.05	2.67×10^{-6}	-0.07	2.56×10^{-5}	-0.06	2.54×10^{-10}	0.0001
cg01059398	3	172235808	<i>TNFSF10</i>	-0.07	9.77×10^{-6}	-0.07	1.63×10^{-4}	-0.07	3.28×10^{-10}	0.0001
cg18181703	17	76354621	<i>SOCS3</i>	-0.09	3.05×10^{-5}	-0.09	2.43×10^{-5}	-0.09	4.56×10^{-10}	0.0002
cg01726890	7	38370874	—	0.03	2.22×10^{-5}	0.03	1.50×10^{-4}	0.03	6.14×10^{-10}	0.0003
cg18942579	17	57915773	<i>VMP1</i>	-0.10	1.69×10^{-5}	-0.09	2.99×10^{-4}	-0.10	8.23×10^{-10}	0.0004

All 65 probes significant after Bonferroni correction are listed in Table, Supplemental Digital Content 5, <http://links.lww.com/IBD/A561>.

^aGRCh37/hg19.

^bMean beta value for CD relative to the mean beta value for controls.

TABLE 2. Differently Methylated Regions

Direction ^a	Genes	P (Minimum)	Chr	Coordinate ^b	Length	Probes
↓	<i>VMP1</i>	1.97×10^{-15}	17	57915150	1138	4
↑	<i>ZBTB16</i>	1.78×10^{-9}	11	113946633	1241	3
↑	<i>SLC15A4</i>	4.58×10^{-8}	12	129281291	408	3
↑	<i>TNF</i>	8.31×10^{-8}	6	31543764	2126	4
↑	<i>DIABLO</i>	1.17×10^{-7}	12	122711835	306	6
↑	<i>ZFYVE28</i>	1.76×10^{-7}	4	2321442	1151	3
↑	<i>RUNX3</i>	2.19×10^{-7}	1	25291794	351	5
↓	<i>HGF</i>	2.44×10^{-7}	7	81399131	536	3
↓	<i>GPR56</i>	6.43×10^{-7}	16	57662273	536	3
↑	<i>TOLLIP</i>	8.18×10^{-7}	11	1296881	554	3
↑	<i>ITGB2</i>	1.21×10^{-6}	21	46341036	536	3
↑	<i>PIEZO1</i>	1.19×10^{-6}	16	88831961	1086	3
↑	<i>PRF1</i>	1.47×10^{-6}	10	72362714	190	6
↑	<i>ZBTB12</i>	2.23×10^{-6}	6	31867623	330	9
↑	<i>PRF1</i>	6.32×10^{-6}	10	72360292	160	6
↑	<i>CD247</i>	7.59×10^{-6}	1	167486463	1685	4
↓	<i>CETP</i>	1.39×10^{-5}	16	56995572	536	3
↑	<i>UBASH3A</i>	1.45×10^{-5}	21	43823747	1030	3
↑	<i>WRAP73</i>	3.21×10^{-5}	1	3563024	1937	3

^aDirection of methylation change.

^bCoordinates refer to GRCh37/hg19.

the pediatric data, the commonest finding was of hypomethylation with increased variance and combining methylation results at 2 probes achieved strong separation between control and disease groups (see Fig., Supplemental Digital Content 7, <http://links.lww.com/IBD/A563> that shows 2-locus methylation plots for the adult replication cohort).

Linear Discriminant Analysis

Using the pediatric discovery cohort methylation beta values as learning set for LDA, models were created for each possible combination of 2 probes to predict the presence of CD, which were then tested using the beta values from the pediatric replication cohort. Area under the curve values for the performance of these models in the replication cohort ranged from 0.79 to 0.98 (median 0.93). Figure 2A shows the separation in 2-dimensional beta values by diagnosis in 10 two-probe combinations.

Interpretation and Selection of Genes for Further Study

To highlight genomic regions for further study, we considered 3 criteria: the significance of individual CpGs, clustering of CpGs into DMRs, and colocalization of methylation changes with risk loci identified by GWAS. Genes that scored highly in multiple categories were given the highest priority for further investigation (Fig. 3), with *VMP1/MIR21* emerging as the strongest candidate. Similarly, the

TNF locus within the HLA region was enriched for highly significant methylation changes within a DMR, in a region of established interest in CD. Other genes of interest include *SBNO2* and *IL18RAP*, where highly significant CpGs are found within risk loci established by GWAS studies and *ZBTB16* and *RUNX3*, where DMRs contain or neighbor highly significant individual CpGs.

VMP1/MIR21

Five probes within the *VMP1/MIR21* locus, 4 of which lie within a DMR (Fig. 4), had disease-associated changes in methylation surviving Bonferroni correction. These probes are clustered at the 3' end of *VMP1*, around the 11th exon, within 50 kb of a GWAS SNP (rs1292053). The DMR is directly adjacent to the transcription start site and promoter region for the primary transcript of *MIR21* (pri-miR-21).

We confirmed CD-associated hypomethylation of this region in blood by pyrosequencing in 172 adults ($P = 6.6 \times 10^{-5}$, Fig. 5A). The qPCR for pri-miR-21 in 43 adults with CD and 23 healthy controls demonstrated an increase of expression in CD ($P < 0.005$, Fig. 5B). Analysis of previously published data³⁴ demonstrated an increased expression of *MIR21* ($P = 1.4 \times 10^{-6}$) and *VMP1* ($P = 2.6 \times 10^{-3}$) in biopsies from inflamed versus uninfamed mucosa in CD, which was not observed in controls (Fig. 5C). *MIR21* showed significantly increased expression in inflamed versus uninfamed UC biopsies (5.1×10^{-7}), but there was no inflammation-related increase in *VMP1* expression, unlike that seen in CD.

DISCUSSION

Principal Findings

This study establishes a significant and highly replicable pattern of DNA methylation associated with pediatric CD, with further replication in adults for many of the most significant pediatric results. We show a significant enrichment of methylation changes in proximity to GWAS risk loci, offering a novel approach in exploring the biological variation associated with common genetic variants and have derived biomarkers, which show remarkable accuracy in determining the presence of CD. As such, this study provides an important confirmation of the validity and feasibility of methylation screening in complex disease and complements the emerging evidence implicating epigenetic alterations in IBD and other immune-mediated diseases and complex traits, such as rheumatoid arthritis,¹¹ obesity,¹⁴ and diabetes.¹³

VMP1 and MIR21

The discovery of methylation alterations within the *VMP1/MIR21* locus emerges as the strongest individual result. Further confirmation of altered methylation of this region in CD by pyrosequencing in adults is augmented with data showing increased expression of *MIR21* in blood in CD and increased expression of *MIR21* in inflamed intestinal biopsies in CD but not controls.

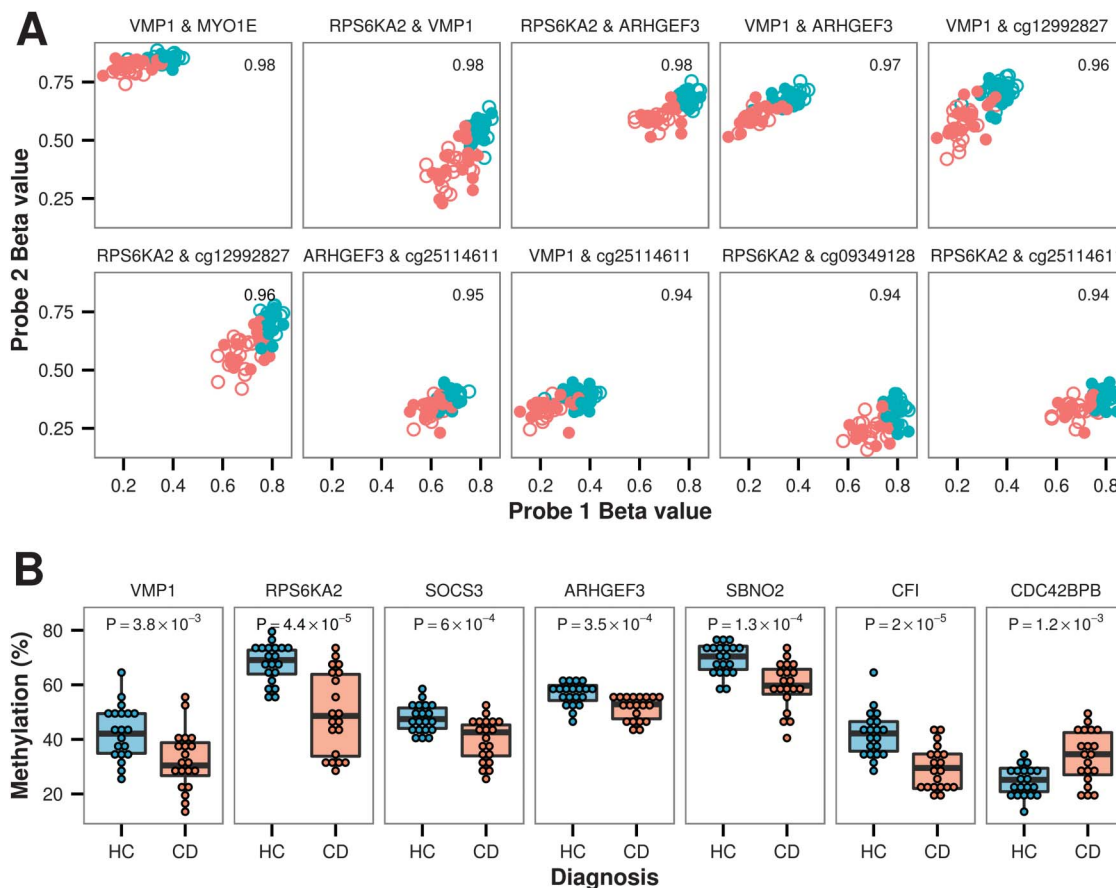


FIGURE 2. A, Separation by diagnosis is achieved by plotting beta values for combinations of 2 Illumina 450k probes with significant disease-associated methylation changes in the discovery cohort. Beta values for both discovery (open) and replication (filled) cohorts are shown. Area under the curve for the model based on each probe combination shown in the top right of each panel. B, Replication of disease-associated methylation changes in 7 significant probes by pyrosequencing in 40 adults. HC, healthy controls.

VMP1 encodes a transmembrane protein located in the Golgi apparatus, endoplasmic reticulum, and vacuoles with high degrees of expression in the intestine, kidney, ovary, and placenta.³⁹ There is high transspecies conservation of *VMP1*, and it is noteworthy that expression induces autophagy through interactions with *BECN1*.^{40,41}

MIR21 was one of the earliest described microRNAs and has been implicated in numerous cancers, including IBD-associated colorectal cancer.⁴² The mature sequence is produced from a precursor overlapping with the 3' end of *VMP1*. This region is highly conserved, exhibits DNase I hypersensitivity and is associated with the promoter-associated histone marks H3K4Me1 and H3K4Me3.⁴³

MIR21 has a known role in T-cell differentiation and development.⁴⁴⁻⁴⁷ Increased expression of *MIR21* in active IBD and IBD-associated dysplasia has been described elsewhere,^{48,49} and *MIR21* knockout mice have been shown to be protected from DSS-induced colitis.⁵⁰

There is a growing body of evidence for numerous microRNAs being involved in CD such as the regulation of *NOD2* by microRNAs^{51,52} and *NOD2* genotype influencing IL-23 production in dendritic cells by regulation of *MIR29* production.⁵³ Recent work

has shown *ATG16L1* can be regulated by multiple microRNAs, with resulting effects on autophagy,^{54,55} particularly interesting about our data as *ATG16L1* contains an *MIR21* target motif.⁵⁶

Other than the *VMP1/MIR21* discovery, a number of the other loci implicated by our study are noteworthy in the context of disease pathogenesis and will bear further investigation. The data implicating the HLA region, and, in particular, the *TNF* locus complement the genetic data implicating this region in determining IBD susceptibility and phenotype and the body of evidence implicating *TNF* in disease.

Other regions that are of great interest to intestinal immune regulation showing highly significant replicable alterations in methylation in our study include *SOCS3*, a suppressor of cytokine signaling to the JAK/STAT pathway,^{57,58} *TOLLIP*,⁵⁹ and *RPS6KA2*,⁶⁰ a ribosomal S6 kinase interacting with MAPkinase1/3.

Linear Discriminant Analysis

DNA methylation of specific loci has found use as a biomarker in diagnosis and prognosis of cancer,^{61,62} such as methylation at a tumor suppressor CpG Island. The results of our LDA serve as a proof of concept for the development of methylation-based

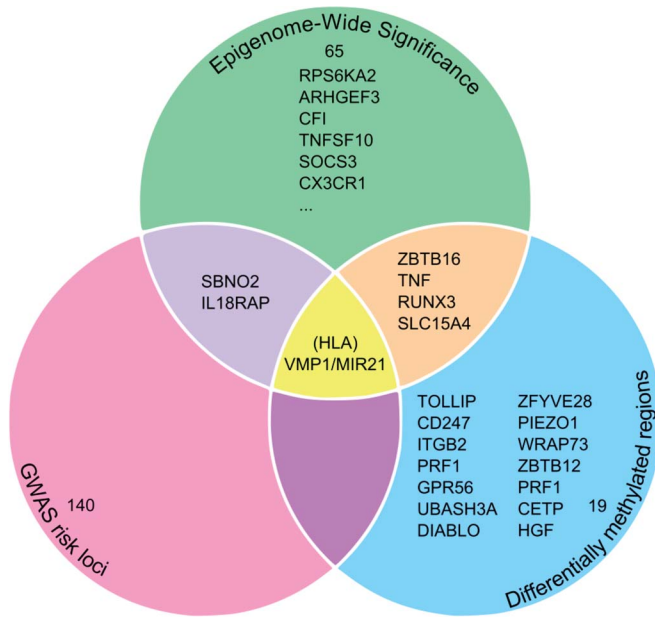


FIGURE 3. Schema for the selection of targets for further study, showing examples and total numbers for each set. GWAS risk loci correspond to all CD and IBD results from recent GWAS meta-analysis,¹ inclusion within epigenome-wide significance and differentially methylated regions sets based on individual probe significance surviving Bonferroni correction, and being identified as a DMR by the modified ChAMP algorithm, respectively (Methods). *VMP1/MIR21* shown at the intersection of all 3 sets.

diagnostic biomarkers in complex diseases. Future work should also seek to establish prospective links with other clinical outcomes, such as response to treatment and disease course.

The use of children who required colonoscopy to rule out IBD as controls precisely models the clinical scenario in which a diagnostic biomarker would find use. CD-specific methylation patterns weakened with increased age and were absent in the elderly (data not shown), possibly due to the accumulation of confounding factors, such as environmental exposure, comorbidity and polypharmacy, or inherent effects of aging on methylation.¹⁸ It remains to be determined if this approach is equally pertinent for conditions with a later age of onset.

Strengths and Limitations

This study provides an impetus for further analysis of alterations of leukocyte DNA methylation in IBD and other complex diseases, with many targets emerging for further study. In comparison with GWAS data, we show highly reproducible and significant disease-associated methylation changes using a modest number of samples. Indeed, the strength and reproducibility of our findings compare favorably with epigenetic data generated to date in IBD and other complex diseases^{10,11,63} and also with the results of theoretical modeling based on predicted disease-associated methylation patterns. In particular, the magnitude and variance of observed methylation changes in whole blood contrasts with models used to predict required group sizes (see Fig., Supplemental Digital Content 8, <http://links.lww.com/IBD/A564>, shows power to detect methylation changes similar to *VMP1/MIR21*).⁹ These data may inform future study design in IBD and other complex diseases.

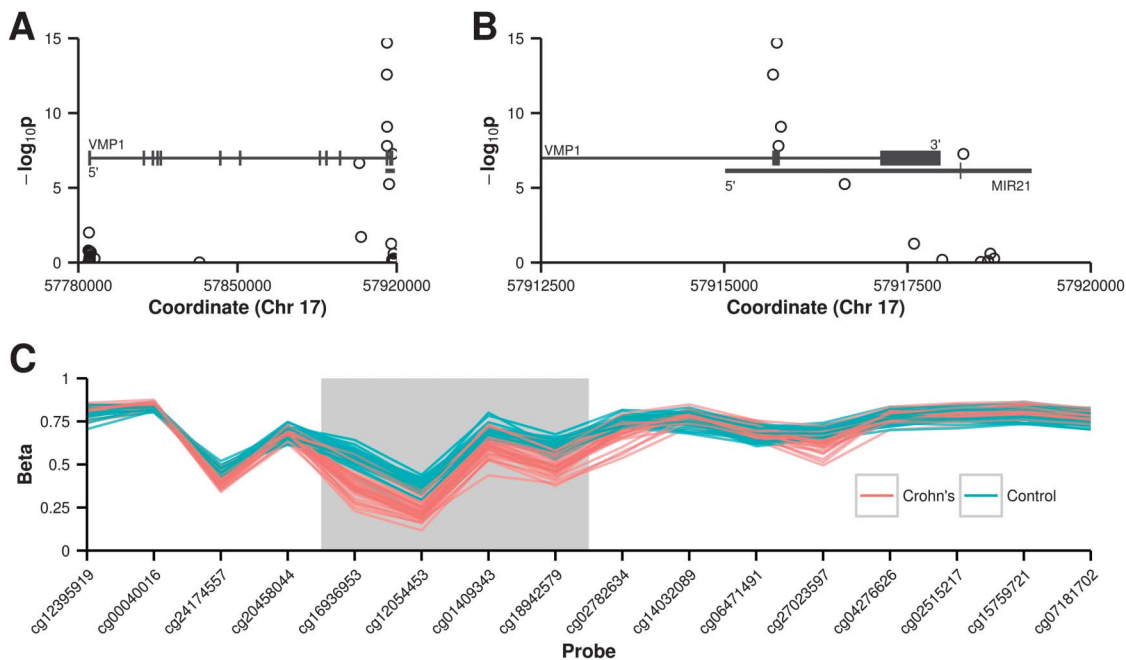


FIGURE 4. A, The significance of disease-associated methylation changes in all Illumina 450k probes in *VMP1*. A schematic representation of the gene is overlaid (bars represent exons, lines represent introns), the height of which corresponds to a Bonferroni corrected $P < 0.05$. B, Expanded view of the 3' end of *VMP1* in (A), with *MIR21* primary transcript (line) and mature *MIR21* (bar) plotted below. C, Beta values for each sample, colored by diagnosis, at all Illumina 450k probes across *VMP1* (excluding the unmethylated 5' CpG island). Background shading highlights probes contained in the DMR.

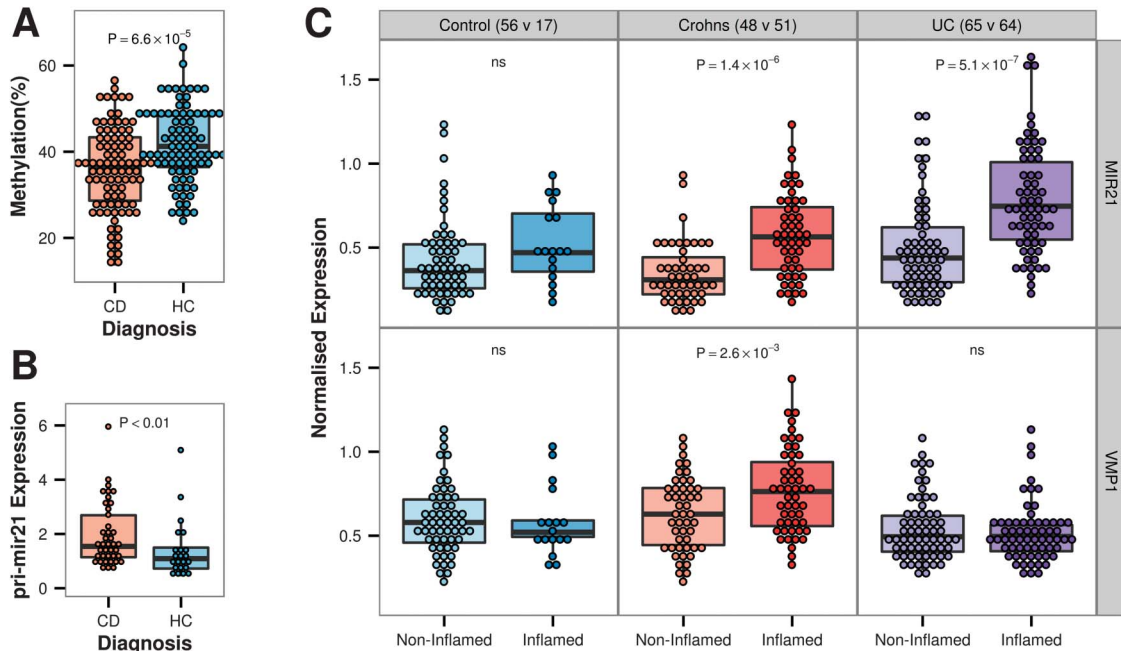


FIGURE 5. A, Replication of *VMP1/MIR21* hypomethylation at cg16936953 by pyrosequencing of the same region in 172 adults. B, Increased leukocyte *MIR21* primary transcript in CD measured by qPCR (n = 66). C, Microarray data³⁴ showing significantly increased pri-miR-21 mRNA in response to inflammation in CD and UC. *VMP1* increased in CD, but not UC or control.

In designing this study, we addressed the key confounding issues relevant to epigenome analysis, which are currently subject to intensive scientific debate. Our approach of basing the study initially in pediatric disease has been successful in generating data replicable in children and adults. Studies in children have the advantage of reducing the influence of age, comorbidity, polypharmacy, smoking, and environmental factors, which could confound epigenetic changes. The focus on circulating leukocytes in IBD rather than intestinal mucosa in this study is strongly supported by scientific evidence of immune dysregulation, the well-recognized clinical extraintestinal manifestations, and indeed the recent evidence of an encouraging response to autologous bone marrow transplant in refractory disease.⁶⁴ Methylation at numerous sites has also been shown to influence PBMC response to stimulation of toll-like receptors ex-vivo with multiple ligands.¹⁶ Ease of access to blood is clearly advantageous in biomarker discovery.

The heterogeneity of studied tissues is a commonly cited concern in epigenome-wide analysis.⁶⁵ We demonstrated the ability to use genome-wide methylation data with contemporaneous clinical full blood count data to correct for whole-blood heterogeneity. If such data are not available, comparison with reference methylation data sets^{29,66} from separated cells and alternative techniques⁶⁷ has been demonstrated effective and accurate. These strategies are feasible for translational studies, especially high throughput clinical investigations, where cell separation adds considerable processing and expense.

Our data are highly significant even after applying Bonferroni correction for multiple testing—this correction that is widely applied in GWAS studies is likely to emerge as overly conservative in the context of EWAS because it ignores the correlation of

methylation between neighboring probes. The establishment of a consensus on the limit of epigenome-wide significance for DNA methylation arrays remains a priority for future reporting of epigenetic findings in complex diseases.

Although the combined factors of moderate study size and conservative correction for multiple testing may well contribute to false negatives (type II error), the reproducibility in 3 independent cohorts and level of statistical significance provide a high degree of confidence in our positive findings. Parallels may be drawn with the early linkage and association studies in IBD, which allowed modeling of the genetic architecture and delivered “low-hanging fruit” in terms of *NOD2* and HLA associations,^{68–70} findings that have subsequently been unequivocally replicated in large scale experiments.

The emerging evidence of a role for *MIR21* in IBD from other approaches enhances the biological plausibility of this finding and strengthens the case for using EWAS in CD and other complex diseases to discover novel biologically significant genes. Moreover, the enrichment of methylation differences near to genetic risk loci variants seen in these data and previous work^{38,71} raises the possibility that epigenetic modifications may help identify specific points within large genetic susceptibility loci where genetic and biological variation overlap.

There is at present intense interest in the application of epigenomic analyses in complex diseases, and technologies and analytic approaches are evolving rapidly. The strengths and limitations of the approach are becoming better understood, leading to the very real hope that EWAS will now evolve to complement GWAS in understanding pathogenesis. To date, methylation profiling has been limited by the application of

analytic approaches developed for genetic rather than epigenomic analysis. In cases, this has led to overestimating the strength of results, such as failing to appreciate the bias introduced by the wide variety of probe numbers per gene in pathway analysis.³⁷ However, failing to appreciate the difference between a SNP of limited possible states weakly correlated with disease risk, and the bounded continuous variable of DNA methylation has led to underestimations of the power of moderate-scale epigenetic studies.

CONCLUSIONS

Overall, these observations serve to highlight the need to integrate methylation, genetic, and expression data in future studies of the pathogenesis of complex diseases and provide insight into potential mechanisms involved in gene-environmental interaction. There are exciting and immediate implications for early clinical translation; the discovery of easily accessible biomarkers in peripheral blood to predict disease susceptibility, progression or response to therapy, and the potential for new therapeutic targets.

Future studies should evaluate altered methylation and expression at these sites, including *MIR21*, both in whole blood and specific cell types, before initiation of disease and in association with environmental factors to better understand causality.

ACKNOWLEDGMENTS

The Wellcome Trust Clinical Research Facility, Edinburgh performed the Illumina 450k experiments. A. T. Adams was funded by CICRA (Crohn's In Childhood Research Association), N. A. Kennedy by the Wellcome Trust [grant number 097943], N. T. Ventham and H. Drummond by the European Commission IBD-BIOM, K. R. O'Leary by the Cunningham Trust. The Scottish Government Chief Scientist Office funded BISCUIT through a Clinical Academic Training Fellowship for R. Hansen (CAF/08/01), the Medical Research Council funded PICTS (grant number G0800675). R. K. Russell is supported by an NHS Research Scotland career fellowship Award. The IBD team at Yorkhill Hospital, Glasgow is supported by the Catherine McEwan Foundation and Yorkhill IBD fund. The authors wish to thank Dr. Johan Van Limbergen and Dr. Paul Henderson for their contributions to the PICTS cohort.

Author contributions: *Experimental work*: A. T. Adams, K. R. O'Leary; *statistical analysis*: N. A. Kennedy, A. T. Adams; *supply of samples and data*: R. Hansen, N. T. Ventham, H. E. Drummond, C. L. Noble, E. El-Omar, R. K. Russell, D. C. Wilson, G. L. Hold; *study design*: E. R. Nimmo, G. L. Hold, J. Satsangi; *drafting of manuscript*: A. T. Adams, N. A. Kennedy, N. T. Ventham, E. R. Nimmo, and J. Satsangi; *substantial contributions to the final text*: All authors. E. R. Nimmo, G. L. Hold, and J. Satsangi contributed equally.

REFERENCES

- Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491:119–124.
- Henderson P, Wilson DC. The rising incidence of paediatric-onset inflammatory bowel disease. *Arch Dis Child*. 2012;97:585–586.
- Benchimol EI, Fortinsky KJ, Gozdyra P, et al. Epidemiology of pediatric inflammatory bowel disease: a systematic review of international trends. *Inflamm Bowel Dis*. 2011;17:423–439.
- Molodecky NA, Soon IS, Rabi DM, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*. 2012;142:46.e42–54.e42; quiz e30.
- Burisch J, Munkholm P. Inflammatory bowel disease epidemiology. *Curr Opin Gastroenterol*. 2013;29:357–362.
- Hold GL. Western lifestyle: a “master” manipulator of the intestinal microbiota? *Gut*. 2014;63:5–6.
- Hansen R, Russell RK, Reiff C, et al. Microbiota of de-novo pediatric IBD: increased Faecalibacterium prausnitzii and reduced bacterial diversity in Crohn's but not in ulcerative colitis. *Am J Gastroenterol*. 2012;107:1913–1922.
- Relton CL, Davey Smith G. Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. *PLoS Med*. 2010;7:e1000356.
- Rakyan VK, Down TA, Balding DJ, et al. Epigenome-wide association studies for common human diseases. *Nat Rev Genet*. 2011;12:529–541.
- Ventham NT, Kennedy NA, Nimmo ER, et al. Beyond gene discovery in inflammatory bowel disease: the emerging role of epigenetics. *Gastroenterology*. 2013;145:293–308.
- Liu Y, Aryee MJ, Padyukov L, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol*. 2013;31:142–147.
- Huynh JL, Garg P, Thin TH, et al. Epigenome-wide differences in pathology-free regions of multiple sclerosis-affected brains. *Nat Neurosci*. 2014;17:121–130.
- Bell CG, Finer S, Lindgren CM, et al. Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. *PLoS One*. 2010;5:e14040.
- Dick KJ, Nelson CP, Tsaprouni L, et al. DNA methylation and body-mass index: a genome-wide analysis. *Lancet*. 2014;383:1990–1998.
- Breitling LP, Yang R, Korn B, et al. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet*. 2011;88:450–457.
- Lam LL, Emberly E, Fraser HB, et al. Factors underlying variable DNA methylation in a human community cohort. *Proc Natl Acad Sci U S A*. 2012;109(suppl 2):17253–17260.
- Kellermayer R. Epigenetics and the developmental origins of inflammatory bowel diseases. *Can J Gastroenterol*. 2012;26:909–915.
- Bell JT, Tsai PC, Yang TP, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet*. 2012;8:e1002629.
- Stadler MB, Murr R, Burger L, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*. 2011;480:490–495.
- Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife (Cambridge)*. 2013;2:e00523.
- Hansen R, Berry SH, Mukhopadhyay I, et al. The microaerophilic microbiota of de-novo paediatric inflammatory bowel disease: the BISCUIT study. *PLoS One*. 2013;8:e58825.
- Van Limbergen J, Russell RK, Drummond HE, et al. Definition of phenotypic characteristics of childhood-onset inflammatory bowel disease. *Gastroenterology*. 2008;135:1114–1122.
- Bibikova M, Barnes B, Tsan C, et al. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011;98:288–295.
- Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56–65.
- Du P, Kibbe WA, Lin SM. Lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. 2008;24:1547–1548.
- Marabita F, Almgren M, Lindholm ME, et al. An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. *Epigenetics*. 2013;8:333–346.
- Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013;29:189–196.

28. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* 2007;3:1724–1735.
29. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* 2012;13:86.
30. Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, et al, eds. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor.* New York, NY: Springer; 2005: 397–420.
31. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol.* 1995;57:289–300.
32. Morris TJ, Butcher LM, Feber A, et al. ChAMP: 450k chip analysis methylation pipeline. *Bioinformatics.* 2014;30:428–430.
33. Li Y, Zhu J, Tian G, et al. The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol.* 2010;8:e1000533.
34. Noble CL, Abbas AR, Cornelius J, et al. Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis. *Gut.* 2008;57:1398–1405.
35. Venables WN, Ripley BD. *Modern Applied Statistics with S.* 4th ed. New York, NY: Springer; 2002.
36. Young MD, Wakefield MJ, Smyth GK, et al. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 2010;11:R14.
37. Geelheer P, Hartnett L, Egan LJ, et al. Gene-set analysis is severely biased when applied to genome-wide methylation data. *Bioinformatics.* 2013;29: 1851–1857.
38. Nimmo ER, Prendergast JG, Aldhous MC, et al. Genome-wide methylation profiling in Crohn's disease identifies altered epigenetic regulation of key host defense mechanisms including the Th17 pathway. *Inflamm Bowel Dis.* 2012;18:889–899.
39. Calvo-Garrido J, Carilla-Latorre S, Escalante R. Vacuole membrane protein 1, autophagy and much more. *Autophagy.* 2008;4:835–837.
40. Kang R, Zeh HJ, Lotze MT, et al. The Beclin 1 network regulates autophagy and apoptosis. *Cell Death Differ.* 2011;18:571–580.
41. Molejon MI, Ropolo A, Re AL, et al. The VMP1-Beclin 1 interaction regulates autophagy induction. *Sci Rep.* 2013;3:1055.
42. Kanaan Z, Rai SN, Eichenberger MR, et al. Plasma miR-21: a potential diagnostic marker of colorectal cancer. *Ann Surg.* 2012;256:544–551.
43. Good PJ, Guyer MS, Kamholz S, et al. The ENCODE (ENCyclopedia of DNA elements) project. *Science.* 2004;306:636–640.
44. Lu TX, Hartner J, Lim EJ, et al. MicroRNA-21 limits in vivo immune response-mediated activation of the IL-12/IFN-gamma pathway, Th1 polarization, and the severity of delayed-type hypersensitivity. *J Immunol.* 2011;187:3362–3373.
45. Chang CC, Zhang QY, Liu Z, et al. Downregulation of inflammatory microRNAs by Ig-like transcript 3 is essential for the differentiation of human CD8(+) T suppressor cells. *J Immunol.* 2012;188:3042–3052.
46. Sawant DV, Wu H, Kaplan MH, et al. The Bcl6 target gene microRNA-21 promotes Th2 differentiation by a T cell intrinsic pathway. *Mol Immunol.* 2013;54:435–442.
47. Ludwig K, Fassin M, Mescoli C, et al. PDCD4/miR-21 dysregulation in inflammatory bowel disease-associated carcinogenesis. *Virchows Arch.* 2013;462:57–63.
48. Wu F, Zikusoka M, Trindade A, et al. MicroRNAs are differentially expressed in ulcerative colitis and alter expression of macrophage inflammatory peptide-2 alpha. *Gastroenterology.* 2008;135:1624–1635.e24.
49. Wu F, Zhang S, Dassopoulos T, et al. Identification of microRNAs associated with ileal and colonic Crohn's disease. *Inflamm Bowel Dis.* 2010; 16:1729–1738.
50. Shi C, Liang Y, Yang J, et al. MicroRNA-21 knockout improve the survival rate in DSS induced fatal colitis through protecting against inflammation and tissue injury. *PLoS One.* 2013;8:e66814.
51. Chuang AY, Chuang JC, Zhai Z, et al. NOD2 expression is regulated by microRNAs in Colonic epithelial HCT116 cells. *Inflamm Bowel Dis.* 2014;20:126–135.
52. Chen Y, Wang C, Liu Y, et al. miR-122 targets NOD2 to decrease intestinal epithelial cell injury in Crohn's disease. *Biochem Biophys Res Commun.* 2013;438:133–139.
53. Brain O, Owens BMJ, Pichulik T, et al. The intracellular sensor NOD2 induces microRNA-29 expression in human dendritic cells to limit IL-23 release. *Immunity.* 2013;39:521–536.
54. Nguyen HTT, Dalmasso G, Müller S, et al. Crohn's disease-associated adherent invasive escherichia coli modulate levels of microRNAs in intestinal epithelial cells to reduce autophagy. *Gastroenterology.* 2014;146:508–519.
55. Lu C, Chen J, Xu HG, et al. MIR106B and MIR93 prevent removal of bacteria from epithelial cells by disrupting ATG16L1-mediated autophagy. *Gastroenterology.* 2014;146:188–199.
56. Griffiths-Jones S, Saini HK, van Dongen S, et al. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 2008;36:D154–D158.
57. Suzuki a, Hanada T, Mitsuyama K, et al. CIS3/SOCS3/SSI3 plays a negative regulatory role in STAT3 activation and intestinal inflammation. *J Exp Med.* 2001;193:471–481.
58. Carow B, Rottenberg ME. SOCS3, a major regulator of infection and inflammation. *Front Immunol.* 2014;5:58.
59. Maillard MH, Bega H, Uhlig HH, et al. Toll-interacting protein modulates colitis susceptibility in mice. *Inflamm Bowel Dis.* 2014;20:660–670.
60. Zhao Y, Bjørbaek C, Wermowicz S, et al. RSK3 encodes a novel pp90orsk isoform with a unique N-terminal sequence: growth factor-stimulated kinase function and nuclear translocation. *Mol Cell Biol.* 1995;15:4353–4363.
61. Grützmann R, Molnar B, Pilarsky C, et al. Sensitive detection of colorectal cancer in peripheral blood by septin 9 DNA methylation assay. *PLoS One.* 2008;3:e3759.
62. Church TR, Wandell M, Lofton-Day C, et al. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut.* 2014;63:317–325.
63. Harris RA, Nagy-Szakal D, Pedersen N, et al. Genome-wide peripheral blood leukocyte DNA methylation microarrays identified a single association with inflammatory bowel diseases. *Inflamm Bowel Dis.* 2012;2399:1–8.
64. Hawkey CJ. Stem cells as treatment in inflammatory bowel disease. *Dig Dis.* 2012;30(suppl 3):134–139.
65. Callaway E. Epigenomics starts to make its mark. *Nature.* 2014;508:22.
66. Reinius LE, Acevedo N, Joerink M, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One.* 2012;7:e41361.
67. Houseman EA, Molitor J, Marsit CJ. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics.* 2014;30: 1431–1439.
68. Hugot JP, Laurent-Puig P, Gower-Rousseau C, et al. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature.* 1996; 379:821–823.
69. Satsangi J, Parkes M, Louis E, et al. Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat Genet.* 1996;14:199–202.
70. Satsangi J, Welsh KI, Bunce M, et al. Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. *Lancet.* 1996;347:1212–1217.
71. Cooke J, Zhang H, Greger L, et al. Mucosal genome-wide methylation changes in inflammatory bowel disease. *Inflamm Bowel Dis.* 2012;18: 2128–2137.