Journal of Crohn's and Colitis, 2023, **17**, 170–184
https://doi.org/10.1093/ecco-jcc/jjac127
Advance access publication 27 August 2022
Original Article

Analysis of Systemic Epigenetic Alterations in Inflammatory Bowel Disease: Defining Geographical, Genetic and Immune-Inflammatory influences on the Circulating Methylome

Rahul Kalla,[a,b,*] Alex T. Adams,[a,c,*] Jan K. Nowak,[d] Daniel Bergemalm,[e] Simen Vatn,[f] Nicholas T. Ventham,[a] Nicholas A. Kennedy,[a,g] Petr Ricanek,[f,h] Jonas Lindstrom,[i,h] IBD-Character Consortium, Johan Söderholm,[j] Marie Pierik,[k] Mauro D'Amato,[l] Fernando Gomollón,[m] Christine Olbjørn,[f,h] Rebecca Richmond,[n] Caroline Relton,[n] Jørgen Jahnsen,[f,h] Morten H. Vatn,[h] Jonas Halfvarson,[e] Jack Satsangi[a,c]

[a]Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
[b]MRC Centre for Inflammation Research, Queens Medical Research Institute, University of Edinburgh, Edinburgh, UK
[c]Translational Gastroenterology Unit, Nuffield Department of Medicine, Experimental Medicine Division, University of Oxford, John Radcliffe Hospital, Oxford, UK
[d]Department of Paediatric Gastroenterology and Metabolic Diseases, Poznan University of Medical Sciences, Poznan, Poland
[e]Department of Gastroenterology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden
[f]Department of Gastroenterology, Akershus University Hospital, Lørenskog, Norway
[g]Exeter IBD and Pharmacogenetics group, University of Exeter, Exeter, UK
[h]Institute of Clinical Medicine, Campus Ahus, University of Oslo, Oslo, Norway
[i]Health Services Research Unit, Akershus University Hospital, Lørenskog, Norway
[j]Department of Surgery and Clinical and Experimental Medicine, Linköping University, Linköping, Sweden
[k]Maastricht University Medical Centre (MUMC), Department of Gastroenterology and Hepatology, Maastricht, Netherlands
[l]CIC bioGUNE – BRTA, Derio, Spain and IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
[m]HCU 'Lozano Blesa', IIS Aragón, CIBEREHD, Zaragoza, Spain
[n]Medical Research Council Integrative Epidemiology Unit (MRC IEU), School of Social and Community Medicine, University of Bristol, Bristol, UK

Corresponding author: Dr Rahul Kalla, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. Email: rahul.kalla@ed.ac.uk
*These authors share co-first authorship.

Abstract

Background: Epigenetic alterations may provide valuable insights into gene–environment interactions in the pathogenesis of inflammatory bowel disease [IBD].

Methods: Genome-wide methylation was measured from peripheral blood using the Illumina 450k platform in a case-control study in an inception cohort (295 controls, 154 Crohn's disease [CD], 161 ulcerative colitis [UC], 28 IBD unclassified [IBD-U)] with covariates of age, sex and cell counts, deconvoluted by the Houseman method. Genotyping was performed using Illumina HumanOmniExpressExome-8 BeadChips and gene expression using the Ion AmpliSeq Human Gene Expression Core Panel. Treatment escalation was characterized by the need for biological agents or surgery after initial disease remission.

Results: A total of 137 differentially methylated positions [DMPs] were identified in IBD, including *VMP1/MIR21* [$p = 9.11 \times 10^{-15}$] and *RPS6KA2* [6.43×10^{-13}], with consistency seen across Scandinavia and the UK. Dysregulated loci demonstrate strong genetic influence, notably VMP1 [$p = 1.53 \times 10^{-15}$]. Age acceleration is seen in IBD [coefficient 0.94, $p < 2.2 \times 10^{-16}$]. Several immuno-active genes demonstrated highly significant correlations between methylation and gene expression in IBD, in particular OSM: IBD $r = -0.32$, $p = 3.64 \times 10^{-7}$ vs non-IBD $r = -0.14$, $p = 0.77$]. Multi-omic integration of the methylome, genome and transcriptome also implicated specific pathways that associate with immune activation, response and regulation at disease inception. At follow-up, a signature of three DMPs [*TAP1*, *TESPA1*, *RPTOR*] were associated with treatment escalation to biological agents or surgery (hazard ratio of 5.19 [CI: 2.14–12.56], logrank $p = 9.70 \times 10^{-4}$).

Conclusion: These data demonstrate consistent epigenetic alterations at diagnosis in European patients with IBD, providing insights into the pathogenetic importance and translational potential of epigenetic mapping in complex disease.

Key Words: DNA methylation; genetics; inflammatory bowel diseases [IBD]; prognosis; methylation; quantitative trait loci; gene expression; epigenetic clock; Mendelian randomization

1. Introduction

Inflammatory BOWEL DISEASES [IBD] phenotypically classified into two main entities, Crohn's disease [CD] and ulcerative colitis [UC], represent an important public health concern, with a projected prevalence of 1% in Western populations by 2030. There are significant implications for healthcare planning as costs, particularly of new treatments, increase.[1] These considerations have accelerated global efforts in better understanding the aetiology of IBD. Unequivocal data now implicate the interaction of host susceptibility with the exposome in the development of IBD.[2] These data have inevitably stimulated studies to explore the potential importance of epigenetic mechanisms, including DNA methylation, in pathogenesis. DNA methylation may regulate gene expression through its effect on the chromatin state as well as accessibility of the transcription binding sites[3–5] and can be influenced by many pertinent environmental factors including smoking and age.[6,7]

We have previously identified several alterations in the IBD-associated circulating epigenome. Our initial study in a high-prevalence Scottish population characterized a replicable pattern of DNA alterations in children with CD, with highly significant enrichment of methylation changes around genome-wide association study [GWAS] single nucleotide polymorphisms [SNPs], in particular the HLA region and *VMP1/MIR21*.[8] More recently, our adult epigenome-wide study within the Scottish population identified distinct differential methylation across key IBD genes.[9] Certain signals were highly cell-specific; RPS6KA2 was shown to be a $CD14^{+}$ monocyte-specific signal in IBD.[9] These top differentially expressed methylation signals were independently confirmed and replicated in a treatment-naïve paediatric CD cohort in North America [RISK consortium],[10] providing strong stimulus for further research in this field. Importantly, the extent to which the findings can be generaliszd to other populations is largely unknown. Furthermore, the timing, stability and functional importance of epigenetic alterations in affecting gene transcription have not been fully investigated.

With the advancing therapeutic repertoire in IBD, there is keen interest in risk-stratifying patients at diagnosis in order to allow personalized medicine, with recent optimism that genetic, transcriptional, glycomic or serological markers may predict disease course, and help position new therapies.[11–18] Promising data have emerged that epigenetic alterations are helpful in other diseases, notably colorectal cancer. The potential clinical utility of these epigenetic marks as biomarkers in immune-mediated diseases has yet to be determined.

In our multi-centre study, we aim to extend current understanding of the pathogenetic and translational importance of the circulating epigenome in IBD. We first aim to assess the consistency of DNA methylation alterations in IBD across geographically distinct populations in the UK, Scandinavia and Spain. We investigate environmental modifiers of the methylome including inflammation, smoking and ageing. Furthermore, we define the genetic contribution to these alterations, their association with gene expression and the methylome of disease progression in IBD. Whilst focused on IBD, the findings have potential implications beyond this disease field, to other complex and immune-mediated diseases.

2. Methods

2.1. Study design

Patients were recruited prospectively as part of the IBD-CHARACTER inception cohort [reference 305676] from gastroenterology appointments across seven centres in Europe [Table 1]. All IBD cases met the standard diagnostic criteria for either UC or CD following thorough clinical, microbiological, endoscopic, histological and radiological evaluation. The Lennard–Jones, Montreal and Paris criteria were used for diagnosis and classification of clinical phenotypes. The control group consisted of symptomatic controls attending gastroenterology clinics during the same period with no evidence of IBD after further investigations and at follow-up. Healthy volunteers were also recruited into the study.

We collected patient demographics including sex, age at diagnosis and date of diagnosis. Details of drug therapy and concomitant medications were recorded. Treatment naïvety within the IBD cohort was defined as no exposure to any IBD-related medical therapies such as oral and topical steroids, oral and topical 5-aminosalicylic acid [5-ASA] therapies, biological therapies and immunomodulators. High-sensitivity C-reactive protein [hsCRP] and albumin were re-assayed in a single batch at the end of recruitment. Other routine markers including haemoglobin and white cell count were also recorded.

Patients with IBD were followed prospectively and information on clinical outcomes was collected during follow-up. Treatment escalation was defined as the need for a biologic, ciclosporin or surgery, instituted for disease flare after initial induction therapy and aiming to induce disease remission. In UC, the definition of treatment escalation also included colectomy during index admission.

All patients and controls provided written, informed consent with local ethical approval at each centre.

2.2. Genome-wide methylation profiling

Peripheral blood leukocyte DNA was bisulphite-converted and analysed using the Illumina HumanMethylation450 platform [Illumina].[19] Cases and controls were randomly distributed across chips. Data were processed using the meffil package[20] in R [R Foundation for Statistical Computing]. Samples containing >1% probes with detection p values >0.01 were removed. Probes with bead counts of <3 in 10% of samples, or detection p values >0.01 in 10% of samples were also removed. Sex mismatches were identified by analysing the median intensities of the sex chromosome probes and removed from further analyses. Genotypes were compared with genotyping probes on the methylation array. Probes containing SNPs with a minor allele frequency of ≥0.01 in the European population in the 1000 Genomes Project were also removed.

Three samples with low signal intensities were removed [>1% of probes with detection p values >0.01] and 684 probes were filtered out due to low signal. There were $n = 15$ failed quality controls who were removed from further analyses. Batch correction were performed for slide, array and centre using ComBat.[21]

2.3. Differentially methylated positions and regions

Cell proportions were estimated from methylation data using the Houseman algorithm[22] using the meffil package in

Table 1. Study demographics and study recruitment according to clinical centre

Study demographics		
Variables	**Inflammatory bowel diseases [*n* = 343]**	**Controls [*n* = 295]**
Males [%]	185 [54]	130 [44]
Subtype IBD [CD:UC:IBD-U]	154:161:28	
Subtype controls [HC:non-IBD]		54:241
Smoking status [current:never:ex:missing]	54:147:103:39	53:147:58:37
Mean age, years [range]	34 [7–79]	33 [3–79]
Montreal classification for CD		
L1	48, 1 [31%, 1%]	
L2, +L4	41, 5 [27%, 3%]	
L3, +L4	46, 10 [30%, 6%]	
Isolated L4	3 [2%]	
Montreal behaviour for CD		
B1, B1p [non-stricturing and non-penetrating, +perianal]	120, 6 [78%, 4%]	
B2, B2p [stricturing, +perianal]	12, 0 [8%, 0%]	
B3, B3p [penetrating, +perianal]	6, 6 [4%, 4%]	
Not available	4 [2%]	
Paris extent for UC		
E1	41 [25%]	
E2	52 [32%]	
E3	67 [42%]	
Not available	1 [1%]	

Centre	**CD**	**UC**	**IBD-U**	**Controls**	**Total**
Edinburgh	47	57	10	86	200
Linkoping	11	4	1	41	57
Maastricht	4	1	0	0	5
Orebro	24	21	9	65	119
Oslo	54	68	6	63	191
IACS	14	10	2	40	66
Total	154	161	28	295	638

Numbers denote the recruitment figures per centre within the study.

R. Differentially methylated position [DMP] analysis [single CpG probe] was performed using age, sex and cell proportions as covariates.[23] Statistical significance was set at $p < 0.05$ following adjustment for multiple testing using the Holm correction[24] for whole blood data.

Differentially methylated regions [DMRs] were identified using the Lasso function from the ChaMP pipeline[25,26] [distinct from the lasso function described below in the biomarker discovery section] and defined as three or more contiguous probes within a 2-kb distance threshold based on expected probe density, each achieving a false discovery rate [FDR]-corrected $p < 0.05$ in DMP analysis.

2.4. Epigenetic clock and age acceleration

Age acceleration [*AgeAccel*] is defined as the residual resulting from a linear regression model, regressing the Horvath estimate of epigenetic age [biological age] on chronological age.[7] A positive value for *AgeAccel* indicates that the observed epigenetic age is higher than that predicted, based on chronological age. DiffAge according to Horvath was defined as the difference between predicted biological age and chronological age.[7] A positive DiffAge is seen for individuals with an increased biological age compared to their chronological age.[27]

2.5. Methylation quantitative trait loci analyses and Mendelian randomization

Whole blood leukocyte DNA was extracted using the Nucleon BACC 3 DNA extraction kit [GE Healthcare]. Patients were genotyped using the Illumina HumanOmniExpressExome-8 Bead Chips [Illumina]. A sex-check was performed using PLINK to identify and remove sex-mismatches. Methylation quantitative trait loci [meQTLs] and expression quantitative trait loci (eQTLs) were estimated using the Matrix-eQTL package,[28] with a distance threshold of 1 Mb. SNPs with a minor allele frequency of <5% were filtered from downstream analyses. Age and sex were included as covariates in order to identify specific disease-associated meQTLs. A Holm-corrected $p < 0.05$ was used as the threshold for statistical significance for disease-associated meQTLs.

Mendelian randomization was performed using TwoSampleMR to determine causal inference[29] for the top meQTLs. For this analysis, meQTLs, independent of IBD,

were obtained using the same filtering thresholds as described above. meQTL SNPs were used as the instrumental variable, CpG as the exposure and IBD as the outcome variable.

2.6. Gene expression profiling

Whole blood RNA underwent targeted RNA sequencing performed using an Ion AmpliSeq Human Gene Expression Core Panel, containing 20 802 genes. Quality control was performed using the Ion Library Taqman quantification kit. Sequence reads were aligned using the Torrent Suite Software [TSS] and the number of matches per amplicon was quantified. After filtering, 14 182 transcripts were available for further analysis.

2.7. Cell-of-origin analysis

We used eFORGE v2.0 [http://eforge.cs.ucl.ac.uk/] to identify if the replicated CpGs were enriched in DNase I hypersensitive sites [DHSs] [markers of active regulatory regions] and loci with overlapping histone modifications [H3Kme1, H3Kme4, H3K9me3, H3K27me3 and H3K36me3] across available cell lines and tissues from Roadmap Epigenomics Project, BLUEPRINT Epigenome and ENCODE [Encyclopedia of DNA Elements] consortia data.

2.8. Data integrative and pathway analysis

Integration of genomic, methylomic and transcriptomic layers was done with Multi-Omics Factor Analysis v2 [MOFA+].[30] MOFA+ enables cross-omics, unsupervised dimensionality reduction that identifies major factors underlying data variability. These latent factors found in the ensemble of data may represent features that would remain hidden from view otherwise. Recent applications of MOFA+ revealed its potential to extract knowledge from complex datasets.[31,32] Here, we conduct a targeted MOFA+ analysis that uses genomic, methylomic and transcriptomic data. The analysis is centred around scaled beta values of 137 differentially methylated probes from the IBD vs non-IBD comparison. Genomic data include these of the SNPs described by De Lange *et al.*[33] that were present in the microarray data [$n = 99$]. The mode imputation was used to fill sporadically missing data. Transcriptomes were filtered to include information on the expression of 1021 genes that were identified as IBD-related using a list from the DisGeNET database [ID: C0021390, $n = 1577$] and had mean expression >10. The expression data were transformed logarithmically ($\log_2[x + 4]$).

MOFA+ object was trained using the default settings. Latent factors were analysed to compare the strength of input from various omics layers and correlation with phenotype data. Gene set enrichment analysis was conducted using Molecular Signatures Database v7.2 biological process ontology from the Broad Institute. The R package tidyverse was used for data manipulation and visualization, including custom plots based on data extracted from the MOFA+ model object. Note that the inclusion of top differentially methylated sites focused this analysis on methylation, with the intention of discovering SNPs and transcripts related to IBD-specific methylation effects.

Prioritized features were subject to gene ontology analysis using gene set enrichment analysis from the Broad Institute. Further details are available in the Supplementary Methods.

2.9. Prognostication analysis

For IBD prognostic marker analysis, data were randomly split into a training [2/3rd] and testing set [1/3rd]. Unsupervised principal component analyses [PCAs] were performed to examine correlates of treatment escalation and clinical predictors of disease severity using the PCAtools package [https://github.com/kevinblighe/PCAtools]. Principal components [PCs] that associate with treatment escalation were selected to identify methylation probes that explain the top 5% variance within each PC. These methylation probes were then selected to fit a model using least absolute shrinkage and selection operator [LASSO] penalized Cox regression with age and sex as covariates and 1000 cross-validated iterations. Results of the Cox regression were then validated in the testing cohort by generating a combined model using all significant variables. Hazard ratios [HRs] were calculated from Cox regression coefficients. We then performed survival analyses based on the need for treatment escalation as defined above. Comparison for treatment escalation was performed using a log-rank test [1 d.f.] and graphically represented with a Kaplan–Meier plot. The performance of the methylation model was also assessed in the IBD subtypes. Conventional predictors of treatment escalation including age, sex, hsCRP and albumin in CD, and age, sex, hsCRP, albumin and extent of colitis in UC were also examined.

3. Results

3.1. Study demographics

The IBD Character cohort represents a multi-centre inception cohort in which 247 [72%] of the 343 IBD patients were treatment-naïve at recruitment. Genome-wide methylation was measured in 638 DNA samples extracted from peripheral blood (295 controls, 154 CD, 161 UC and 28 IBD-unclassified [IBD-U] patients). Table 1 summarizes study recruitment and patient demographics. Mean age in patients with IBD [$n = 343$] was 34 years [range 7–79], and 33 years [range 3–79] in controls [$n = 295$]. A total of 27% of CD patients had a colonic disease phenotype at recruitment while 42% of patients with UC exhibited extensive colitis. Two hundred individuals were recruited in the UK, 367 in Scandinavia and 66 in Spain at presentation for investigation of suspected IBD. A total of 33% of those recruited in the UK had confirmed IBD after investigation, while 58% and 39% had IBD in the Scandinavian and Spanish cohort, respectively [Table 1].

3.2. Differentially methylated probes in IBD

Across the entire cohort, 137 probes exhibited Holm significant IBD-associated methylation differences in comparing IBD with controls [Supplementary Table 1]. These include probes mapping at the loci *VMP1/MIR21* [$p = 9.11 \times 10^{-15}$], *SBNO2* [2.70×10^{-14}], *RPS6KA2* [6.43×10^{-13}] and *TNFSF10* [7.72×10^{-8}], thereby replicating and validating our previous findings.[9] Novel findings include differential methylation of *PHOSPHO1* [3.43×10^{-9}] and *SELPLG* [2.54×10^{-7}]. Table 2 summarizes the top DMPs.

Similar analyses were performed to identify UC- and CD-specific DMPs. There were 72 DMPs that differentiated CD from controls [Supplementary Table 2] and 67 DMPs that differentiated UC from controls [Supplementary Table 3]. There were 24 DMPs that demonstrated overlap across UC and CD analyses [Supplementary Figure 1]. There were

no probes that differentiated UC from CD *and no probes that differentiated colonic from isolated ileal CD*. Analysis was performed of DMRs, defined as regions with three of more contiguous probes with FDR-corrected $p < 0.05$ within a distance threshold based on expected probe density. The *VMP1* locus on chromosome 17 was the only DMR identified within our dataset that remained significant using these criteria [four probes, DMR size 1150].

Next, we compared our dataset with previous findings from our group on DMPs in adult and paediatric IBD by correlation analyses of the top 1000 methylation probe beta values in the three cohorts [Supplementary Figure 2].[8,9] Strong correlation was seen across populations [adult BIOM cohort[9] $r = 0.96$, $p < 2.20 \times 10^{-308}$; paediatric cohort $r = 0.81$, $p < 2.20 \times 10^{-308}$].

We then investigated the influence of germline variation and recruitment centre effects on IBD-associated DMPs. Linear modelling using age, gender, recruiting centre, cell proportions and the first ten genetic PCAs identified 162 DMPs that differentiated IBD from controls [Supplementary Table 4]; the top probes remaining significant are shown in Supplementary Table 1.

3.3. Consistency of DMPs across Northern Europe

We then analysed the consistency of methylation across Europe [Figure 1] by splitting our cohort based on geographical area [Scandinavia vs UK vs Spain]. Independent DMP analysis using all probes was performed in Scandinavia,

Table 2. Top ten differentially methylated positions [DMPs] in patients with inflammatory bowel diseases versus controls

Probe ID	Chr	Gene ID	Log FC	*p* value	Holm *p*	Region	Relation to CpG island
cg07573872	19	SBNO2	−0.05	3.07E-23	1.05E-17	Body	S_Shelf
cg09349128	22	CRELD2	−0.03	3.87E-21	1.32E-15		N_Shore
cg17501210	6	RPS6KA2	−0.04	1.25E-19	4.26E-14	Body	OpenSea
cg12054453	17	VMP1	−0.07	4.89E-19	1.67E-13	Body	OpenSea
cg16292768	8	CLU	−0.03	4.85E-18	1.66E-12	Body	OpenSea
cg18608055	19	SBNO2	−0.04	8.67E-17	2.96E-11	Body	OpenSea
cg16936953	17	VMP1	−0.07	2.32E-16	7.91E-11	Body	OpenSea
cg19821297	19	HOOK2	−0.04	5.32E-16	1.81E-10		S_Shore
cg27469606	19	SBNO2	−0.03	1.09E-15	3.73E-10	5′UTR	Island
cg12170787	19	SBNO2	−0.02	4.29E-15	1.47E-09	Body	OpenSea

Positive values indicate increased methylation in cases compared to controls; negative values indicate hypomethylation in IBD cases vs controls. Body: within body of the gene. 5′UTR: 5′ untranslated region.

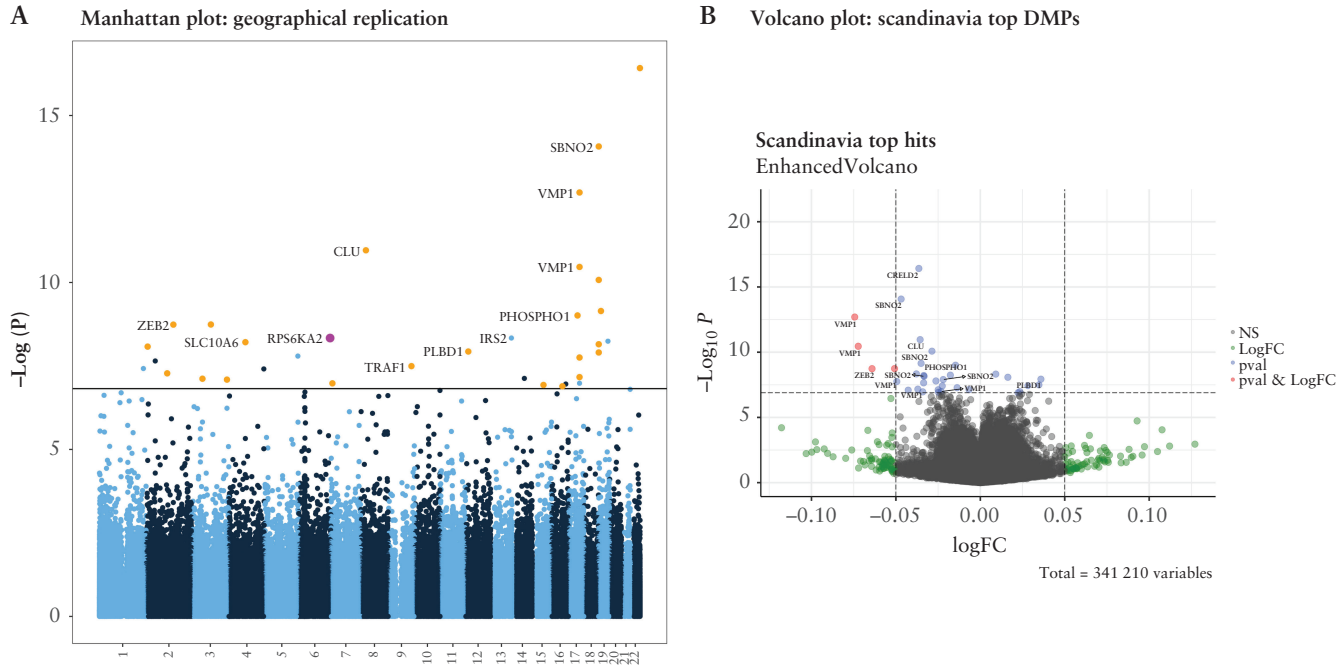

Figure 1. [A] Manhattan plot of top differentially methylated positions [DMPs] in IBD vs controls in Scandinavia. Yellow dots represent hits that replicate in the UK cohort and purple dots represent hits that replicate in Spain. [B] Volcano plot of the top DMPs across Scandinavia. Horizontal dashed line represents Holm significance. The *y*-axis for both plots represents $-\log_{10}$[*p*-values] with the horizontal line representing statistical significance after Holm testing.

Table 3. Top ten differentially methylated positions [DMPs] in patients with inflammatory bowel diseases vs controls that were tested in the Scandinavian cohort [NE Holm represents Holm-corrected *p* value] and replicated across UK [UK Holm represents *p* value] and Spain [SE Holm]. The highlighted probe remains significant across all cohorts

Probe ID	Chr	Gene	Region	Relation to CpG island	Log FC	NE *p* value	NE Holm	UK Holm	SE Holm
cg17501210	6	RPS6KA2	Body	OpenSea	−0.04	4.61E-09	1.57E-03	5.26E-08	2.51E-02
cg12992827	3	ZLPD1		OpenSea	−0.05	1.81E-09	6.18E-04	9.89E-04	2.13E-01
cg02464912	14	SYNE2	TSS200	N_Shore	−0.01	7.46E-08	2.55E-02	1.00E-01	3.28E-01
cg12054453	17	VMP1	Body	OpenSea	−0.07	2.02E-13	6.89E-08	2.37E-04	5.08E-01
cg01101459	1	LINC01132		OpenSea	0.03	3.78E−08	1.29E-02	5.72E-02	5.08E-01
cg02782634	17	VMP1	Body	OpenSea	−0.02	1.04E-07	3.56E-02	1.00E-01	5.55E-01
cg18942579	17	VMP1	Body	OpenSea	−0.05	1.77E-08	6.03E-03	8.90E-05	5.55E-01
cg09349128	22	CRELD2		N_Shore	−0.04	3.80E-17	1.30E-11	3.07E-03	6.88E-01
cg18608055	19	SBNO2	Body	OpenSea	−0.03	7.08E-09	2.42E-03	5.21E-06	6.88E-01
cg16936953	17	VMP1	Body	OpenSea	−0.07	3.43E-11	1.17E-05	2.23E-04	8.08E-01

Positive values indicate increased methylation in cases compared to controls; negative values indicate hypomethylation in IBD cases vs controls. TSS200: 200 bp from transcription start site. Body: within body of the gene.

identifying 34 probes that differentiated IBD from controls with a Holm $p < 0.05$ [Table 3]. These included *SBNO2* [$p = 2.88 \times 10^{-9}$], *VMP1* [$p = 6.89 \times 10^{-8}$] and *RPS6KA2* [$p = 1.57 \times 10^{-3}$]. A total of 26 of these probes were significant in the UK cohort [$n = 200$]. In the Spanish cohort [$n = 66$] only one [*RPS6KA2*] remained significant [Holm $p = 0.03$]. Power calculations were performed based on these findings, taking into consideration the effect sizes noted in the Scandinavian cohort. This post-hoc analysis confirmed that the Spanish cohort was adequately powered to detect significant differences in 11 of the 34 DMPs identified in the Scandinavian dataset [power = 0.8, alpha = 0.05, effect size cut-off = 0.8; Supplementary Table 5]. DMPs consistently replicated across samples from the UK and Scandinavia, notably *ZEB2*, *SBNO2* and *ZPLD1*, were not detected in samples from Spain despite being adequately powered to detect these DMPs.

3.4. Deciphering tissue and immune cell specificity of differentially methylated probes in IBD

Blood cell-type-specific DNase hypersensitive site [DHS] enrichment testing using the eFORGE v2.0 online tool[34] was performed for our top 137 differentially expressed methylation probes. This demonstrated that 62 of the 137 IBD-associated CpGs are enriched in DHSs within monocytes [binomial $p = 2.04 \times 10^{-8}$]. These encompassed the top-most differentially expressed DMPs in disease such as *OSM*, *VMP1*, *SELPLG* [*PSGL-1*] and *AIM2*. No enrichments were seen in other immune cells such as T-cells, B-cells, NK cells or haematopoietic stem cells [Supplementary Figure 3].

We then examined for overlap between the circulating blood DMPs in CD and UC in this study and those differentially methylated in human colonic intraepithelial cells [IECs] in CD and UC cohorts.[35] A total of 38 and five probes overlapped in UC and CD respectively [Supplementary Table 6]. Top IEC signals with overlap in the circulating methylome across both UC and colonic CD included *SBNO2*, *RPS6KA2* and *SELPLG*. Overlapping probes unique to UC included *AIM2*, *MAD1L1*, *MIR4470* and *MIR3679*.

3.5. The association of DNA methylation with inflammation

To better understand the influence of inflammation on the top differentially methylated probes, three distinct analyses were performed. These included correlations with inflammatory markers, DMP analysis using inflammatory markers such as hsCRP as covariates and DMP overlap with published inflammation-associated methylation probes.

We investigated the correlation of the 137 differentially expressed DMPs in IBD with inflammatory markers, i.e. hsCRP and albumin, in individuals with complete data [$n = 591$; Supplementary Figure 4]. A total of 98 probes correlated with concentration of hsCRP and 118 with albumin levels. Most significant correlations with hsCRP included *LIPC* [cg27307975, $r = -0.56$, $p < 2.22 \times 10^{-16}$] and *ZEB2* [cg20995564, $r = -0.48$, $p < 2.22 \times 10^{-16}$]. Top DMPs such as *VMP1* [cg12054453, $r = 0.25$, $p = 3.78 \times 10^{-8}$] and *RPS6KA2* [cg17501210, $r = -0.30$, $p = 7.81 \times 10^{-12}$] showed moderate correlation with hsCRP. There were, however, 39 probes that did not correlate with hsCRP and included another *VMP1* probe [cg02782634, $r = -0.13$, $p = 0.08$] [Supplementary Table 7]. To adjust for inflammatory activity, we also performed DMP analysis with previously described covariates and with the addition of hsCRP. A total of 59 probes remained significant and included *SBNO2* [$p = 5.26 \times 10^{-14}$], *RPS6KA2* [$p = 4.51 \times 10^{-11}$] and *VMP1* [$p = 5.36 \times 10^{-10}$] [Supplementary Table 8].

We then compared our top DMPs to those that have been reported to be associated with CRP levels.[36] A total of 43 of the 218 DMPs identified by Lighthart *et al.* were differentially methylated in our study. There were, however, 94 probes that still associated with IBD and were independent of inflammation-associated probes [Supplementary Table 9]. These included *CLU* [cg16292768], *SBNO2* [cg12170787] and *VMP1* [cg02782634].

Finally, to identify inflammation-independent DMPs, we excluded probes that correlated with hsCRP or reported inflammation-associated DMPs and identified 30 DMPs that showed no overlap with published inflammation-associated DMPs or any correlation with inflammation [Supplementary Table 10].

3.6. The association of DNA methylation with smoking

Smoking can affect DNA methylation at certain CpG sites in blood, and a recent systematic review summarized a total of 1460 smoking-associated CpG sites.[37] None of the 137 probes identified in this study overlap with the reported smoking-associated CpGs. Given the divergent effects of smoking on IBD subtypes, we performed analyses for probes that associate with CD or UC by including smoking, age, sex and cell proportions as covariates. In CD, 107 DMPs remained significant; *SBNO2*, *RPS6KA2* and *VMP1* were the top probes [Supplementary Table 11]. In UC, 59 probes remained significant: *SBNO2*, *VMP1* and *ZEB2* were the most significant probes [Supplementary Table 12].

3.7. 'Epigenetic age' acceleration and its association with IBD

The 'epigenetic age' of the patients was calculated using the methodology described by Horvath *et al.*[7] There was a strong correlation between actual age and 'epigenetic age' in this cohort ($r = 0.94$, confidence interval [CI]: 0.93–0.95, $p < 2.20 \times 10^{-16}$; Supplementary Figure 5). Age acceleration [*AgeAccel*] was seen in IBD [$r = 0.94$, $p < 2.2 \times 10^{-16}$]. DiffAge, defined as the difference between predicted biological age and chronological age, was determined for IBD [including subtypes] and controls.[7,27] There were significant DiffAge values seen between IBD and controls (non-IBD: median 4.34 years [interquartile range, IQR: 3.83–4.70] vs IBD: 5.28 years [IQR: 4.72–5.64]; $p < 2.20 \times 10^{-16}$). Differences were also seen between IBD subtypes compared to non-IBD (vs UC: 5.08 years [IQR: 4.51–5.49]; $p < 2.20 \times 10^{-16}$ and vs CD: 5.53 years [IQR: 4.99–5.79]; $p < 2.20 \times 10^{-16}$; Supplementary Figure 5). There was poor correlation seen between DiffAge and inflammatory markers in the entire cohort [hsCRP: $r = 0.09$, $p = 0.03$; Alb: $r = -0.13$, $p = 2.88 \times 10^{-3}$]. No correlations were seen between DiffAge and treatment exposure in UC or CD [UC $p = 0.89$; CD $p = 0.21$].

3.8. Germline variations show a strong correlation with DNA methylation [meQTLs]

Using paired genetic and methylation data for the entire cohort [$n = 638$] and age and sex as covariates, meQTLs were generated using the top DMPs using the entire cohort [DMPs $n = 137$]. A total of 2991 *cis*-meQTLs were identified. After applying an minor allele frequency (MAF) > 0.05 and Holm adjustment, 341 *cis*-meQTLs remained significant across 21 unique genes, indicating a strong genetic influence on methylation. Several key loci that were significantly differentially methylated in IBD had a strong genetic influence including *ITGB2* [seven *cis*-meQTLs; top $p = 2.83 \times 10^{-16}$], and 143 *VMP1/MIR21* *cis*-meQTLs across six probes [Supplementary Table 13]. This includes meQTLs with a known GWAS SPN rs1292053 and also with its previously reported linkage disequilibrium [LD] SNP [rs8078424, $r^2 = 0.43$, top $p = 1.48 \times 10^{-20}$].[9] Other novel IBD-relevant associations include *AIM2* [16 *cis*-meQTLs; top $p = 2.83 \times 10^{-16}$].

To determine the causal role of DNA methylation in IBD, Mendelian randomization was applied to our dataset using TwoSampleMR.[29] The most significant meQTLs for each CpG [sentinel meQTLs] were generated using all SNPs and methylation probes independent of a diagnosis of IBD [Supplementary Table 14]. Using sentinel meQTLs as the instrument variable, methylation as the exposure variable and

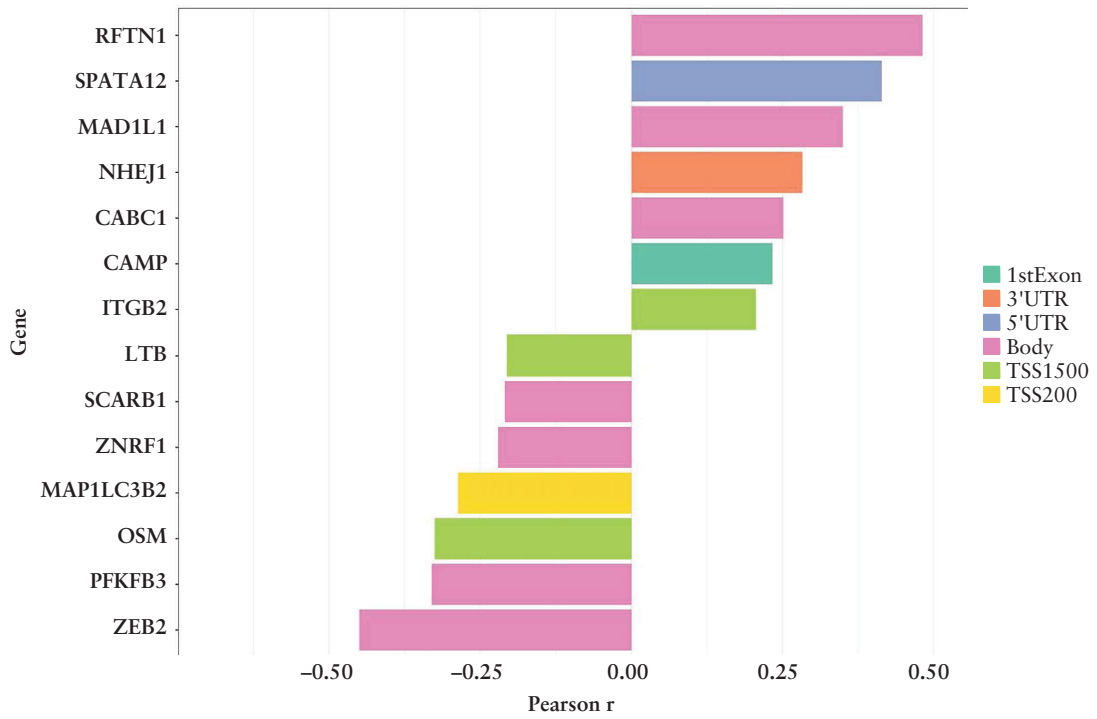

Figure 2. Correlation between DNA methylation probes and gene expression in IBD [Holm $p < 0.05$] where no significant correlation was seen in non-IBD for these probes [Holm $p > 0.05$]. The *y*-axis represents gene names. Bar colours represent CpG probe location within the gene. TSS200 and TSS1500, 200 or 1500 nucleotides upstream of the transcriptional start sites. Within regulatory regions of a gene (5′ untranslated region [5′UTR], 3′ untranslated region [3′UTR]) or body of the gene [Body].

IBD as the outcome variable, no causal associations were identified in our dataset.

There is interest in investigating the relationship of the known IBD GWAS SNP rs1819333 on RPS6KA2 given its causal association with IBD through Mendelian randomization analysis.[10,38] Although rs1819333 was not in our genotype dataset, we identified 123 SNPs in perfect [$R^2 = 1$, $D' = 1$] LD with rs1819333 in UK + Iberian populations according to LDproxy. Of these SNPs, only rs394522 was present within our dataset. Using a linear model including age and gender as covariates, the rs1819333 proxy rs394522 predicted RPS6KA2 expression [$p = 0.03$].

3.9. IBD-associated genes are differentially methylated and correlate with gene expression

In this study, paired whole blood gene expression data were available in 590 patients. Of the 137 DMPs identified, 15 probes were located either 200 or 1500 nucleotides upstream of the transcriptional start sites [TSS200 and TSS1500], within regulatory regions of a gene (*5′* untranslated region [5′UTR], *3′* untranslated region [3′UTR]) or body of the gene. In this study, we discovered 51 highly significant correlations [Supplementary Table 15] with a series of novel genes demonstrating high correlation between methylation alterations and expression including *CD247* [Body, $r = -0.69$, $p = 4.70 \times 10^{-80}$], *SELPLG* [5′UTR, $r = -0.56$, $p = 3.10 \times 10^{-46}$], *OSM* [TSS1500, $r = -0.33$, $p = 2.48 \times 10^{-14}$] and *AIM2* [TSS1500, $r = -0.51$, $p = 6.32 \times 10^{-37}$]. Further sub-analysis was performed investigating the correlation differences in IBD and non-IBD cohorts [Supplementary Table 16]. *SELPLG* expression and methylation demonstrated a relatively higher negative correlation within IBD than non-IBD [IBD $r = -0.53$, $p = 2.24 \times 10^{-22}$ vs -0.37, $p = 6.34 \times 10^{-8}$; 5′UTR]. A total of 14 probes demonstrated significant correlation in IBD patients but no correlation in non-IBD [Figure 2]. Examples include *ZEB2* which demonstrated a negative correlation in IBD [$r = -0.45$, $p = 6.43 \times 10^{-15}$; gene body] and no significant correlation in controls [$r = -0.16$, $p = 0.29$]. Similarly, OSM expression negatively correlated with methylation within IBD cases but no correlation was seen within controls [IBD $r = -0.32$, $p = 3.64 \times 10^{-7}$ vs non-IBD $r = -0.14$, $p = 0.77$; TSS1500]. There were six genes with methylation probes within the TSS200 region, all demonstrating negative correlations with DNA methylation apart from *SYNE2* [IBD $r = 0.31$, $p = 2.89 \times 10^{-6}$ vs non-IBD $r = 0.22$, $p = 0.02$].

Further sub-analyses were performed to determine the correlation between DNA methylation and gene expression for probes within 200 and 1500 nucleotides upstream of the transcriptional start site using age, gender and the first ten genotype PCAs as covariates [Supplementary Table 17]. A total of 15 probes demonstrated Holm significant correlations including *BCL6* [cg06164260, $r = 0.40$, $p = 8.54 \times 10^{-53}$, TSS1500], *AIM2* [cg00490406, $r = 0.31$, $p = 4.70 \times 10^{-36}$, TSS200] and *OSM* [cg25739715, $r = 0.18$, $p = 2.99 \times 10^{-16}$]. Similar analyses were performed to identify IBD-associated differential gene expression. A total of 12 probes within TSS200 or TSS1500 were associated with IBD [Supplementary Table 18]. The top probes included *BCL6* [$r = .25$, $p = 3.48 \times 10^{-27}$, TSS1500], *AIM2* [$r = 0.17$, $p = 2.02 \times 10^{-15}$, TSS200] and *OSM* [$r = 0.18$, $p = 2.85 \times 10^{-17}$, TSS1500].

3.10. Integrative analysis identifies immune cell-related activation in IBD

Multi-omics analysis was performed using MOFA.[30] Integration of IBD-related SNPs and mRNAs around the 137 DMPs produced ten factors, of which the first four explained most of the variability in the dataset [Figure 3A]. Factor 1 and Factor 4 correlated strongly with IBD and hsCRP [Figure 3B and E]. These two factors were mostly influenced by DMPs and mRNAs [Figure 3C] and remained independent of each other [$r = 0.052$, Figure 3D]. Factor 1 was higher in IBD compared with controls regardless of inflammation [Figure 3F; $p < 2.20 \times 10^{-16}$] and appeared specific for IBD. By contrast, at null hsCRP, Factor 4 was similar in IBD patients and controls [$p = 0.374$ when <0.5 mg/L] but was found to be elevated in IBD overall [Figure 3D; $p = 4.21 \times 10^{-13}$] due to higher values relative to controls starting at even marginally increased hsCRP. In UC, Factor 1 reflected the extent of colitis [Figure 3G]. Thus, multi-omics integration revealed DMP-driven, IBD- and hsCRP-associated factors 1 and 4.

Factor 1 was defined by greater methylation of genes such as *CXCR6* and *CD247*, and reduced methylation of *ZEB2* [Figure 3H and I]. The transcriptomic repertoire of Factor 1 prominently featured S100 proteins and matrix metalloproteinases. Only one IBD GWAS SNP contributed to Factor 1: rs7495132 [CRTC3]. In gene ontology analysis, Factor 1 was related to inflammation with activation of undifferentiated leukocytes, and lipid metabolism [Figure 3J].

Reduced methylation at VMP1 [TMEM49] and the interferon-inducible inflammasome trigger AIM2 were driving Factor 4, which also shared lowered ZEB2 methylation with Factor 1 [Figure 3H and I]. Gene expression behind Factor 4 included the immunosuppressive CD274 [PD-L1], along with Fc gamma receptor 1 [FCGR1A, FCGR1B] and AIM2 [overexpressed, consistent with lower methylation; Figure 3J]. The GWAS polymorphism rs1801274 [FCGR2A] was negatively correlated with Factor 4. Pathway analysis uncovered potential roles for Factor 4 in immune regulation, lipid metabolism and cell death [Figure 3K]. The composition of Factor 4 shows that IBD-related differential methylation at VMP1 may relate to opsonization and phagocytosis and testifies to a close coupling of pro- and anti-inflammatory responses.

3.11. DNA methylation associates with disease course in IBD

Follow-up data were available for 291 patients with IBD in order to identify methylation markers that predicted treatment escalation in IBD over a median follow-up period of 526 days [IQR: 223–775] [Table 4]. Thirty-nine patients with a diagnosis of CD, 26 patients with UC and two with IBD-U required escalation after a median follow-up time of 98 days [IQR: 40–229]. The median age in this group was 28 years [range: 18–67] and 58% were male [$n = 39$]. For downstream analysis, data were split into a training [2/3rd] and testing set [1/3rd] for signal validation. To investigate DMPs that associate with treatment escalation, PCA using all methylation probes was performed to identify PCs that associate with treatment escalation. In the training set [$n = 194$; 40 escalations], 11 PCs significantly associated with treatment escalation. Probes that represent the top/bottom 5% of the variance within five of these PCs are shown in Supplementary Figure 6.

Probes that represented the top 5% of the variance across all 11PCs that associate with treatment escalation in the training set were selected [$n = 55$] for further LASSO penalized Cox regression. Three methylation probes remained significant predictors [Table 5] in IBD. These included Transporter associated with Antigen Processing 1

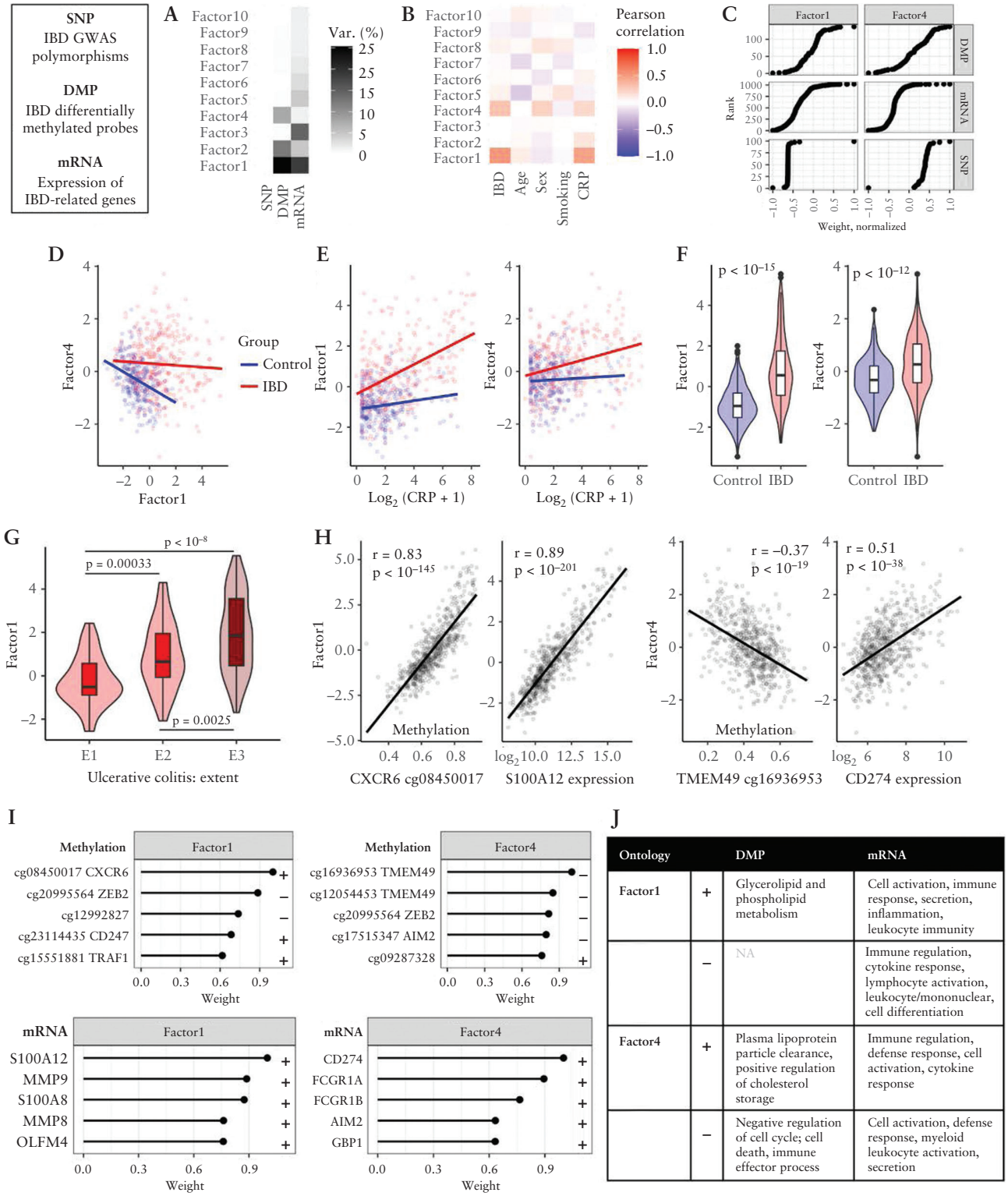

Figure 3. Multi-omic factor analysis [MOFA] integrating IBD genome-wide association analysis [GWAS] polymorphisms, mRNA expression of IBD-related genes and differentially IBD-related methylated positions [DMPs]. [A] MOFA heatmap demonstrating the variance explained by each factor. [B–E] Analyses demonstrating the correlation of the MOFA factors with inflammatory markers such as hsCRP, IBD, age and smoking. [F] Boxplots for Factors 1 and 4 in IBD and controls. [G] Boxplot showing the association of Factor 1 with disease extent in UC. [H, I] Correlation plots of the key methylation and mRNA expression genes that define Factor 1 and Factor 4. [J] Gene ontology analysis of the top MOFA Factor 1 and Factor 4.

[*TAP1*; HR: 12.79], Regulatory protein of MTOR complex 1[*RPTOR*; HR: 1.47] and Thymocyte expressed, positive selection associated 1 [*TESPA1*; HR: 1.29]. A model incorporating these three probes was then tested in an independent testing set [$n = 97$; 27 escalations]. The three-probe combined model remained a significant predictor of treatment escalation in the testing set with an HR of 5.19 [CI: 2.14–12.56, logrank $p = 9.70 \times 10^{-4}$; Figure 4].

Adjusting for treatment naivety did not influence the top differentially methylated probes among patients with IBD. Similar analyses were performed to generate a model that included known clinical predictors: age, sex, hsCRP and albumin in the training set [n = 174] and validated in the testing set [n = 88] where data for covariates were complete. This clinical model had an HR of 4.55 [CI: 2.00–10.35; $p = 1.03 \times 10^{-4}$] and performed at par with the methylation model. Supplementary Table 19 shows correlations of the top three prognostic DMPs with conventional inflammatory markers.

4. Discussion

4.1. Are geographical replicability and variability preferentially driven by germline variation or alterations in the exposome?

In complex diseases such as IBD, DNA methylation potentially represents a mechanism at the interface between genetics and the environment. The strengths of this prospective case-control study include recruitment of individuals with a new diagnosis of IBD, mostly naïve to medical therapy, across multiple clinical centres in Europe. Importantly, our data strongly replicate, validate and extend previous key DMP findings from children with CD and adults in the index Scottish population where our initial data were generated[8,9] to Scandinavia. We demonstrate significant evidence for dysregulation of several previously implicated loci, notably *VMP1*, *RPS6KA2* and *SBNO2* across Scandinavia and the UK, populations with a shared genetic ancestry. Replication was less evident in the smaller Southern European cohort recruited in Spain. Despite being adequately powered to detect significant differences in 11 of the 34 probes tested, significant differential methylation for top probes such as *ZEB2* and *SBNO2* was not seen in the Spanish cohort. *RPS6KA2* remains the only consistent signal across the UK, Scandinavia and Spain. These differences may be in line with the north–south gradient in IBD that have been reported across the USA and Europe.[41–43] Whilst more detailed

Table 4. Patient demographics for predicting disease course in inflammatory bowel disease

Variables	IBD escalation group [n = 67]	Non-escalation group [n = 224]
Subtype IBD [CD:UC:IBD-U]	39:26:2	77:127:20
Edin:Oslo:Orebro:Spain	28:21:14:4	85:77:41:21
Males [%]	39 [58]	118 [53]
Smoking status [current:never:ex:missing]	16:35:15:1	37:104:82:1
Mean age, years [range]	32 [18–67]	36 [18–79]
Montreal classification for CD		
L1	13	27
L2	9	23
L3	17	26
L4	0	1
Not available	0	0
Montreal behaviour for CD		
B1 + B1p	29	66
B2	6	4
B3 + B3p	4	5
Not available	0	2
Paris extent for UC		
E1	0	41
E2	8	41
E3	18	45
Not available	0	0

Table 5. Top three CpGs associated with escalation in treatment [surgery or anti-TNF therapies] in inflammatory bowel disease [IBD]

CpG probe	Gene	Chr	IBD HR training set	Function/relevance in IBD
cg26033526	TAP1	6	12.79	MHC-II encoded genes are essential in HLA class I proteins for generation of cytotoxic T-cell-mediated immune response
cg20502501	RPTOR	17	1.47	Encodes an mTOR binding protein and is part of the mammalian target of rapamycin complex 1 [mTORC1]; mTORC1 plays a role in the pathogenesis of UC[39]
cg06352538	TESPA1	12	1.29	Plays a role in development and maturation of T-cells and T-cell antigen receptor-mediated activation of ERK and NFAT signalling.[40]

HR: hazard ratio; Chr: chromosome.

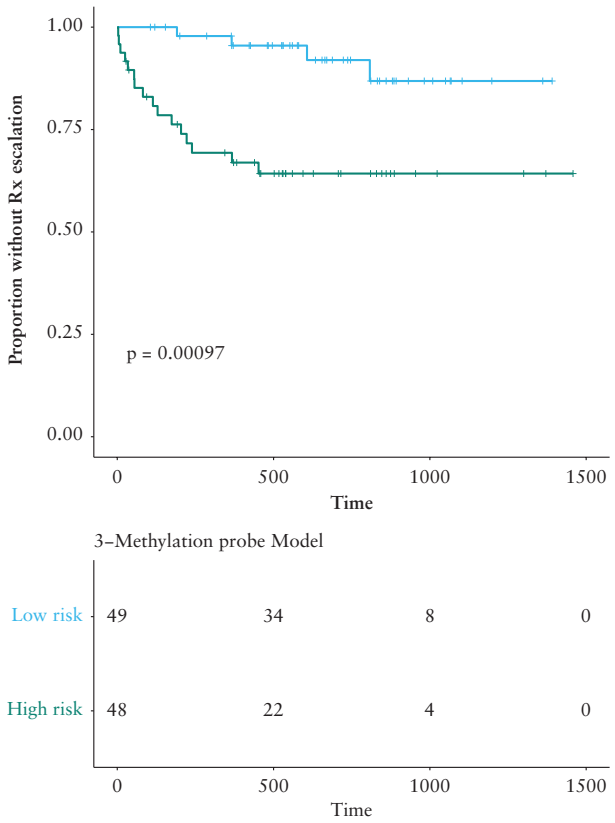

Figure 4. A three-probe methylation panel predicts treatment escalation in inflammatory bowel disease [IBD]. Kaplan–Meier survival curve demonstrating the proportion of IBD patients who did not require a treatment escalation at diagnosis. High [red] and low [blue] risk represents patient groups defined as high or low risk of treatment escalation based on the three-probe panel.

studies in Southern Europe are needed, these data may suggest population specificity of several methylation changes in IBD and heighten the focus on exploring genetic influences and factors within the exposome in these populations.

DMPs implicated in our study such as *VMP1*, *RPS6KA2* and *SBNO2* were also highly significant in the paediatric studies at diagnosis [RISK consortium; USA and Canada].[10] It remains to be seen whether the methylation differences seen within our predominantly Caucasian non-Jewish cohort are replicated in other non-Caucasian populations. Population-specific methylation differences at 439 CpG sites were reported in a study that included Caucasian-American, African-American and Han Chinese-American healthy individuals and associated these changes with distinct phenotypic characteristics such as drug metabolism, disease susceptibility and appearance.[44] Although the majority of these differences were due to underlying genetic variations, up to one-third of the DMPs were independent of germline variations, bringing into focus the role of non-genetic influences such as the exposome on the epivariance across populations. Similar findings were also reported in a Hispanic origin cohort where genetic ancestry explained up to three-quarters of the variance in methylation[45] and an Indonesian cohort where up to 10% of genes showed differential expression and methylation patterns between islands, probably attributable to small-scale environmental differences.[46] It remains to be determined whether the differences which are apparent between Northern Europe and Southern Europe within our study are genetically determined or whether these are related to the exposome. Further multiregional epigenome studies are now needed to explore this further in IBD.

4.2. The influence of genetics on DNA methylation in IBD

The differences between populations focus interest on the relative importance of germline variation on methylation, and in turn on gene expression. Our data also shed new light on the relationship between germline variation at specific loci and epigenetic alterations in complex disease. A key locus, *VMP1*, showed 143 *cis*-meQTLs across six probes and includes meQTLs with a known GWAS SNP rs1292053 and its LD SNP [rs8078424, top $p = 1.48 \times 10^{-20}$].[9] However, causal inference using Mendelian randomization could not be made. A recent study in CD demonstrated that DNA methylation at three CpG sites was causally associated with a GWAS SNP rs1819333 using Mendelian randomization analysis, probably through transcriptional regulation of RPS6KA2 expression.[10,47] This SNP was not on our genotyping platform and further validation could not be performed in this study. *RPS6KA2*-specific hypomethylation has consistently been demonstrated in IBD across several independent cohorts and across all ages.[8–10] This gene codes for a ribosomal kinase, a member of the serine/threonine kinase family, and regulates the autophagy-associated mTOR pathway and particularly relevant in CD.[48] These data provide insight into the complex interaction of genetics and epigenetics in the pathophysiology of IBD.

4.3. Pro-inflammatory pathways

Of particular interest within the field of DNA methylation has been to address whether the changes seen are causal or a consequence of disease initiation. Somineni *et al.* explored this in the paediatric cohort through several methods including longitudinal dynamics of methylation, correlation with known inflammatory markers and Mendelian randomization.[10] Although the vast majority of methylation signals in the Somineni *et al.* study represented a consequence of disease initiation through inflammation, there were ten CpGs that were unchanged at follow-up.[38] In our studies, some of the methylation alterations have been shown to associate with CRP levels.[36] Even though these changes may be a consequence of inflammation, the genes implicated are of great interest as potential targets for future drug discovery, and the mechanisms of how these genes associate with inflammation, disease onset and disease course need further exploration. An example includes *OSM*, a gene that has been recently shown to be upregulated in inflamed intestinal tissue in IBD and is pro-inflammatory. This gene is able to predict anti-TNF response in IBD, making it a useful clinical biomarker and a potential drug target.[49] In our study, a 450K probe close to the TSS [TSS1500, $p = 8.6 \times 10^{-11}$] of *OSM* is differentially methylated in IBD compared to controls and demonstrates correlation with *OSM* gene expression in IBD compared to controls [IBD $r = -0.32$, $p = 3.64 \times 10^{-7}$ vs non-IBD $r = -0.14$, $p = 0.77$; TSS1500]. Other signals linked to CRP-associated DMPs have relevance in IBD pathogenesis. One of the top signals is P-selectin glycoprotein ligand-1 [*PSGL-1* or *SELPLG*] and plays a critical role in immune cell recruitment to sites of tissue inflammation.[50] Anti-PSGL-1 antibody is currently in a phase 2 drug trial for the treatment of CD [NIH project

no. 1R44DK085845-01A1]. Furthermore, differential methylation within this gene correlated with PSGL-1 expression in our study [IBD $r = -0.53$, $p = 2.24 \times 10^{-22}$ vs $r = -0.37$, $p = 6.34 \times 10^{-8}$; 5′UTR]. It is known that a normal level of DNA methylation is required to control differential expression of maternal and paternal alleles of imprinted genes[51] and plays an important role in cell differentiation and embryonic development. Studies have shown that methylation at gene promoter regions can vary depending on cell type, with hypermethylation corresponding to low or no gene transcription.[52,53]

Multi-omic data integration using MOFA revealed two factors [Factors 1 and 4] strongly associated with IBD and pro-inflammatory pathways. While Factor 1 was defined by greater methylation within *CXCR6* and *CD247* genes, reduced methylation of *ZEB2* [Figure 3H and I] and mRNA expression predominantly featuring S100 proteins and matrix metalloproteinases, Factor 4 was defined by reduced methylation at *VMP1* [*TMEM49*] and the interferon-inducible inflammasome trigger *AIM2*. The CpGs and mRNA repertoire that define these two factors highlight the role immune cell activation, cellular response and regulation play in shaping the circulatory methylome and transcriptome in patients with IBD at disease inception. These data provide a repertoire of novel targets to develop future drug therapies that target these pro-inflammatory pathways at disease onset.

4.4. Age acceleration in IBD

There is emerging interest in studying disease-associated age acceleration, defined as the difference between predicted age [determined by DNA methylation patterns] and chronological age. Recent studies have linked epigenetic age acceleration to all-cause mortality[54,55] and cardiovascular disease outcomes.[56] Epigenetic age acceleration has yet to be explored in the context of immune-mediated diseases such as IBD. In our cohort, significant DiffAge was seen between IBD and controls (non-IBD: median 4.34 years [IQR: 3.83-4.70] vs IBD: 5.28 years [IQR: 4.72–5.64]; $p < 2.20 \times 10^{-16}$); this was consistently seen within CD and UC when compared to non-IBD (vs UC: 5.08 years [IQR: 4.51–5.49]; $p < 2.20 \times 10^{-16}$ and vs CD: 5.53 years [IQR: 4.99–5.79]; $p < 2.20 \times 10^{-16}$). Of particular interest was the absence of its association with drug exposure in both CD and UC [UC $p = 0.89$, CD $p = 0.24$] and poor correlation with inflammatory indices at disease inception. Further studies measuring serial methylation profiles may help determine whether these differences persist after treatment, over time.

4.5. The epigenome of treatment escalation

Of clinical interest, DNA methylation alterations were investigated for their performance as prognostic markers. A unique strength of our study lies in the evaluation of methylation in predicting an aggressive disease course in a treatment-naïve unselected inception cohort. We identify a simple molecular signature worthy of further attention as a potential predictive biomarker. The reported prognostic signals have relevance in the pathogenesis of IBD. *TAP1* is essential in HLA class I proteins for the generation of a cytotoxic T-cell-mediated immune response. Differential expression of this gene has been shown to associate with immune escape and poor prognosis in colorectal cancer and is an indicator of an aggressive breast tumour[57,58] [Table 5]. In IBD, *TAP1* differential methylation has been previously reported as a CD-specific signal in IECs from the small intestine and colon and reported to be a regulatory DMR [rDMR a DMR that is located within 10 kb of the transcription start site of a differentially expressed gene].[35] The *RPTOR* gene forms part of the mTORC1 complex, which has been shown to promote UC through COX-2-mediated Th17 responses.[39] Depletion of *RPTOR* inactivates mTORC1 and ameliorates UC.[39] These differentially methylated genes also differ from those implicated in disease susceptibility. Similar trends have been seen in GWAS in CD where unique prognostic genes have been identified that do not overlap with disease susceptibility.[14]

Several studies including GWAS have utilized the $CD8^+$ T-cell-derived transcriptome criteria for escalation, defined as the need for two or more immunosuppressants after initial disease remission.[13,14,59–61] These studies collectively have provided the first lines of evidence for a circulating biomarker profile to define a subset of patients who require treatment escalation and perhaps may benefit from a ‘top-down’ approach to management at an early stage. Our data complement these findings and provide further molecular depth by defining their methylome signature.

Within the paediatric CD group, DNA methylation in separated circulating $CD8^+$ T cells has been shown not to predict outcomes in CD.[62] Given that $CD8^+$ cells only represent a small proportion of the cellular compartment in whole blood, our signals warrant further exploration of the prognostic methylome in other immune cells. It is noteworthy in this context that the e-FORGE analysis in our dataset preferentially implicates monocytes, and further studies exploring differential methylation in intestinal tissue macrophages in IBD are now needed. Furthermore, our prognostic analysis includes UC; the molecular prognostic profile of this subtype has not been examined previously. Further large multi-centre studies are now needed to explore these findings in both adults and children.

4.6. Strengths and limitations

Our study has recruited over 4 years one of the largest multicentre European inception cohorts reported to date, and generated IBD-specific methylation signals, associated with disease onset and progression. We have explored geographical, genetic and non-genetic correlates as mechanistic as well as translational implications. We have previously demonstrated that the most dysregulated areas implicated are cell-specific.[9] There are certain methodological considerations to take into account when interpreting our data. First, our dataset is adequately powered in Scandinavia and the UK, but we acknowledge that the sample size will detect some but not all DMPs in Southern Europe [Spain]. We have demonstrated, in the Spanish cohort, replication for *RPS6KA2*, one of the most robust findings in Northern Europe and North America, whilst excluding a significant effect for ten other markers all highly significant in Scandinavia.

One of the challenges of methylation studies in inflammatory diseases is to disentangle the effect of inflammation on the top differentially expressed DMPs. We have approached this in appropriate detail by several methods including correlation analyses with inflammatory markers such as hsCRP and matching our top probes with publicly available inflammation-associated DMPs.

There are also potential challenges in the prognostic analysis of data from multi-centre studies associated with the concern that decision-making can vary across centres. However, in our study all European centres used similar criteria and guidelines for decisions on escalation in IBD [also known as a 'step-up' approach] where increments are made in therapy based on response to the initial treatment. Although we were able to identify and internally validate IBD-specific prognostic signals in this analysis, IBD subtype analysis and validation was not feasible due to the small number of patients who escalated therapy within each subtype. It is worth noting, however, that certain prognostic signals such as *TAP1* have been shown to be CD-specific in the published literature. Future studies are needed to further validate our findings.

5. Conclusions

This study highlights the stability of the IBD-specific circulating methylome across regions with shared ancestry. We demonstrate a close association of the methylome with inflammation and through integrative multi-omic analyses we identify key pro-inflammatory genes that are upregulated in IBD at inception. Furthermore, differential methylation within certain genes such as *TAP1* associate with disease course over time. These data provide a rich resource for future translational studies investigating the epigenome in IBD, and potentially represent a paradigm for analysis in other complex diseases.

Funding

The study was funded by the following European Commission FP7 grant: IBD-CHARACTER [contract no. 2858546]. NAK was funded by the Wellcome Trust [grant no. WT097943MA]. IBD-Character Consortium: Erik Andersson, Ian D. Arnott, MD, Monica Bayes, PhD, Ferdinando Bonfiglio, PhD, Ray K. Boyapati, MD, Adam Carstens, MD, Christina Casén, MSc, Ewa Ciemniejewska, MSc, Mauro D'Amato, PhD, Fredrik A. Dahl, PhD, Trond Espen Detlie, MD, Hazel E. Drummond, BSc, Gunn S. Ekeland, MSc, Daniel Ekman, MSc, Anna B. Frengen, PhD, Mats Gullberg, PhD, Ivo G. Gut, PhD, Marta Gut, PhD, Simon C. Heath, PhD, Fredrik Hjelm, PhD, Henrik Hjortswang, MD, PhD, Gwo-Tzer Ho, PhD, Daisy Jonkers, PhD, Nicholas A. Kennedy, MBBS, PhD, FRACP, Charles W. Lees, PhD, Torbjørn Lindahl, MSc, Mårten Lindqvist, PhD, Angelika Merkel, PhD, Eddie Modig, BSc, Aina E. F. Moen, PhD, Hilde Nilsen, PhD, Elaine R. Nimmo, PhD, Colin L. Noble, MD, Niklas Nordberg, PhD, Kate R. O'Leary, MSc, Anette Ocklind, PhD, Christine Olbjørn, MD, Erik Pettersson, PhD, Marieke Pierik, MD, PhD, Dominique Poncelet, PhD, Dirk Repsilber, PhD, Céline Sabatel, PhD, Renaud Schoemans, PhD, Alan G. Shand, MD, Johan D. Söderholm, MD, PhD, Janne Sølvernes, MS, Mikael Sundell, BSc, Tone M. Tannæs, PhD, Leif Törkvist, MD, PhD, Anne-Clémence Veillard, PhD, Nicholas T Ventham, MRCS[Eng] PhD, MBBS, David C. Wilson, MD, MRCPCH, Panpan You, MS.

Conflict of Interest

R. Kalla Financial support for research: EC IBD-Character, Lecture fee[s]: Ferring, N. Kennedy Financial support for research: Wellcome Trust, Conflict with: Abbvie, MSD, Warner Chilcott, Ferring speaker fees. Shire Travel bursary, A. Adams: none declared, J. Satsangi Financial support for research: EC grant IBD-BIOM,IBD-Character, Wellcome, CSO, MRC, Conflict with: Consultant for: Takeda, Conflict with: MSD speaker fees. Shire travelling expenses

Author Contributions

Study design RK, JH, MDA, MV, JS. Patient recruitment and sample processing NTV, RK, NAK, DB, SV, ATA CO, FG, JSO, MP, PR. Experimental work RK, NTV, ATA, NAK. Data Analysis RK, ATA, JKN. RK, ATA and JKN wrote the manuscript. All authors were involved in critical review, editing, revision and approval of the final manuscript.

Data Availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Supplementary Data

Supplementary data are available online at *ECCO-JCC* online.

References

1. Kaplan GG, Windsor JW. The four epidemiological stages in the global evolution of inflammatory bowel disease. *Nat Rev Gastroenterol Hepatol* 2021;**18**:56–66.
2. 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**:1913–22. Doi: 10.1038/ajg.2012.335
3. Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* 2009;**10**:295–304.
4. Kraiczy J, Nayak K, Ross A, *et al*. Assessing DNA methylation in the developing human intestinal epithelium: potential link to inflammatory bowel disease. *Mucosal Immunol* 2016;**9**:647–58.
5. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 2012;**13**:484–92. doi:10.1038/nrg3230.
6. Tsaprouni LG, Yang T-P, Bell J, *et al*. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics* 2014;**9**:1382–96.
7. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol* 2013;**14**:R115.
8. Adams AT, Kennedy NA, Hansen R, *et al*. Two-stage genome-wide methylation profiling in childhood-onset Crohn's disease implicates epigenetic alterations at the VMP1/MIR21 and HLA loci. *Inflamm Bowel Dis* 2014;**20**:1784–93.
9. Ventham NT, Kennedy NA, Adams AT, *et al*. Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. *Nat Commun* 2016;**7**:13507.
10. Somineni HK, Venkateswaran S, Kilaru V, *et al*. Blood-derived DNA methylation signatures of Crohn's disease and severity of intestinal inflammation. *Gastroenterology* 2019;**156**:2254–2265.e3. doi:10.1053/j.gastro.2019.01.270
11. Kugathasan S, Denson LA, Walters TD, *et al*. Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. *Lancet* 2017;**389**:1710–8.
12. Marigorta UM, Denson LA, Hyams JS, *et al*. Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's disease. *Nat Genet* 2017;**49**:1517–21.

13. Kalla R, Kennedy NA, Ventham NT, *et al*. Serum calprotectin: a novel diagnostic and prognostic marker in inflammatory bowel diseases. *Am J Gastroenterol* 2016;**111**:1796–805.
14. Lee JC, Biasci D, Roberts R, *et al*. Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. *Nat Genet* 2017;**49**:262–8.
15. Miyahara K, Nouso K, Saito S, *et al*. Serum glycan markers for evaluation of disease activity and prediction of clinical course in patients with ulcerative colitis. *PLoS One* 2013;**8**:e74861e74861.
16. Biasci D, Lee JC, Noor NM, *et al*. A blood-based prognostic biomarker in IBD. *Gut* 2019;**68**:1386–95.
17. Verstockt B, Verstockt S, Dehairs J, *et al*. Low TREM1 expression in whole blood predicts anti-TNF response in inflammatory bowel disease. *EBioMedicine* 2019;**40**:733–42.
18. Gaujoux R, Starosvetsky E, Maimon N, *et al*. Cell-centred meta-analysis reveals baseline predictors of anti-TNFα non-response in biopsy and blood of patients with IBD. *Gut* 2019;**68**:604–14.
19. Bibikova M, Barnes B, Tsan C, *et al*. High density DNA methylation array with single CpG site resolution. *Genomics* 2011;**98**:288–95.
20. Min JL, Hemani G, Davey Smith G, Relton C, Suderman M. Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics* 2018;**34**:3983–3989. doi:10.1093/bioinformatics/bty476
21. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;**28**:882–3.
22. Houseman EA, Accomando WP, Koestler DC, *et al*. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinf* 2012;**13**:86.
23. Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, eds. *Bioinformatics and Computational Biology Solutions Using {R} and Bioconductor*. New York: Springer; 2005: 397–420.
24. Holm SA. simple sequentially rejective multiple test procedure. *Scand J Stat* 1979;**6**:65–70.
25. Morris TJ, Butcher LM, Feber A, *et al*. ChAMP: 450k chip analysis methylation pipeline. *Bioinformatics* 2014;**30**:428–30.
26. Aryee MJ, Jaffe AE, Corrada-Bravo H, *et al*. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014;**30**:1363–9.
27. Lind L, Ingelsson E, Sundström J, Siegbahn A, Lampa E. Methylation-based estimated biological age and cardiovascular disease. *Eur J Clin Invest* 2018;**48**:e12872.
28. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 2012;**28**:1353–8.
29. Hemani G, Zheng J, Elsworth B, *et al*. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018;**7**:e34408. doi:10.7554/eLife.34408
30. Argelaguet R, Arnol D, Bredikhin D, *et al*. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol* 2020;**21**:111. doi:10.1186/s13059-020-02015-1
31. Ha KCH, Sterne-Weiler T, Morris Q, Weatheritt RJ, Blencowe BJ. Differential contribution of transcriptomic regulatory layers in the definition of neuronal identity. *Nat Commun* 2021;**12**:335.
32. Waszak SM, Robinson GW, Gudenas BL, *et al*. Germline Elongator mutations in Sonic Hedgehog medulloblastoma. *Nature* 2020;**580**:396–401.
33. de Lange KM, Moutsianas L, Lee JC, *et al*. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* 2017;**49**:256–61.
34. Breeze CE, Reynolds AP, van Dongen J, *et al*. eFORGE v2.0: updated analysis of cell type-specific signal in epigenomic data. *Bioinformatics* 2019;**35**:4767–9.
35. Howell KJ, Kraiczy J, Nayak KM, *et al*. DNA methylation and transcription patterns in intestinal epithelial cells from pediatric patients with inflammatory bowel diseases differentiate disease subtypes and associate with outcome. *Gastroenterology* 2017;**154**:585–598. doi:10.1053/j.gastro.2017.10.007
36. Ligthart S, Marzi C, Aslibekyan S, *et al*. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol* 2016;**17**:255. doi:10.1186/s13059-016-1119-5
37. Gao X, Jia M, Zhang Y, Breitling LP, Brenner H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. *Clin Epigenetics* 2015;**7**:113.
38. Kalla R, Adams AT, Satsangi J. Blood-based DNA methylation in Crohn's disease and severity of intestinal inflammation. *Transl Gastroenterol Hepatol* 2019;**4**:76–76.
39. Lin X, Sun Q, Zhou L, *et al*. Colonic epithelial mTORC1 promotes ulcerative colitis through COX-2-mediated Th17 responses. *Mucosal Immunol* 2018;**11**:1663–73.
40. Wang D, Zheng M, Lei L, *et al*. Tespa1 is involved in late thymocyte development through the regulation of TCR-mediated signaling. *Nat Immunol* 2012;**13**:560–8.
41. Shivananda S, Lennard-Jones J, Logan R, *et al*. Incidence of inflammatory bowel disease across Europe: is there a difference between north and south? Results of the European Collaborative Study on Inflammatory Bowel Disease (EC-IBD). *Gut* 1996;**39**:690–7.
42. Sonnenberg A, McCarty DJ, Jacobsen SJ. Geographic variation of inflammatory bowel disease within the United States. *Gastroenterology* 1991;**100**:143–9.
43. Armitage EL, Aldhous MC, Anderson N, *et al*. Incidence of juvenile-onset Crohn's disease in Scotland: association with northern latitude and affluence. *Gastroenterology* 2004;**127**:1051–7.
44. Heyn H, Moran S, Hernando-Herraez I, *et al*. DNA methylation contributes to natural human variation. *Genome Res* 2013;**23**:1363–72.
45. Galanter JM, Gignoux CR, Oh SS, *et al*. Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. *Elife* 2017;**6**:e20532. doi:10.7554/eLife.20532
46. Natri HM, Bobowik KS, Kusuma P, *et al*. Genome-wide DNA methylation and gene expression patterns reflect genetic ancestry and environmental differences across the Indonesian archipelago. *PLoS Genet* 2020;**16**:e1008749.
47. Wu Y, Zeng J, Zhang F, *et al*. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat Commun* 2018;**9**:918.
48. Pancholi S, Lykkesfeldt AE, Hilmi C, *et al*. ERBB2 influences the subcellular localization of the estrogen receptor in tamoxifen-resistant MCF-7 cells leading to the activation of AKT and RPS6KA2. *Endocr Relat Cancer* 2008;**15**:985–1002.
49. West NR, Hegazy AN, Owens BMJ, *et al*. Oncostatin M drives intestinal inflammation and predicts response to tumor necrosis factor-neutralizing therapy in patients with inflammatory bowel disease. *Nat Med* 2017;**23**:579–89.
50. Guyer DA, Moore KL, Lynam EB, *et al*. P-selectin glycoprotein ligand-1 (PSGL-1) is a ligand for L-selectin in neutrophil aggregation. *Blood* 1996;**88**:2415–21.
51. Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. *Nature* 1993;**366**:362–5.
52. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 2008;**9**:465–76.
53. Han H, Cortez CC, Yang X, Nichols PW, Jones PA, Liang G. DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. *Hum Mol Genet* 2011;**20**:4299–310.
54. Marioni RE, Shah S, McRae AF, *et al*. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol* 2015;**16**:25.
55. Chen BH, Marioni RE, Colicino E, *et al*. DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging* 2016;**8**:1844–65.
56. Roetker NS, Pankow JS, Bressler J, Morrison AC, Boerwinkle E. Prospective study of epigenetic age acceleration and incidence of cardiovascular disease outcomes in the ARIC study (Atherosclerosis Risk in Communities). *Circ Genomic Precis Med* 2018;**11**:e001937.

57. Ling A, Löfgren-Burström A, Larsson P, *et al*. TAP1 down-regulation elicits immune escape and poor prognosis in colorectal cancer. *Oncoimmunology* 2017;**6**:e1356143e1356143.
58. Henle AM, Nassar A, Puglisi-Knutson D, Youssef B, Knutson KL. Downregulation of TAP1 and TAP2 in early stage breast cancer. *PLoS One* 2017;**12**:e0187323e0187323.
59. Lee JC, Lyons P, McKinney EF, *et al*. Gene expression profiling of CD8+ T cells predicts prognosis in patients with Crohn disease and ulcerative colitis. *J Clin Invest* 2011;**121**:4170–9.
60. Kalla R, Adams AT, Bergemalm D, *et al*. Serum proteomic profiling at diagnosis predicts clinical course, and need for intensification of treatment in inflammatory bowel disease. *J Crohns Colitis* 2021;**15**:699–708. doi:10.1093/ecco-jcc/jjaa230
61. Kalla R, Adams AT, Ventham NT, *et al*. Whole blood profiling of T-cell derived miRNA allows the development of prognostic models in inflammatory bowel disease. *J Crohns Colitis* 2020;**14**:1724–1733. doi:10.1093/ecco-jcc/jjaa134
62. Gasparetto M, Payne F, Nayak K, *et al*. Transcription and DNA methylation patterns of blood derived CD8+ T cells are associated with age and inflammatory bowel disease but do not predict prognosis. *Gastroenterology* 2021;**160**:232–244.e7. doi:10.1053/j.gastro.2020.08.017