The Effect of Phenotype and Genotype on the Plasma Proteome in Patients with Inflammatory Bowel Disease

Bourgonje, Arno R.; Hu, Shixian; Spekhorst, Lieke M.; Zhernakova, Daria V.; Vila, Arnau Vich; Li, Yanni; Charrout, Mohammed; Mahfouz, Ahmed; Reinders, Marcel J.T.; More Authors

DOI
10.1093/ecco-jcc/jjab157

Publication date
2022

Document Version
Final published version

Published in
Journal of Crohn's and Colitis

Citation (APA)
Bourgonje, A. R., Hu, S., Spekhorst, L. M., Zhernakova, D. V., Vila, A. V., Li, Y., Charrout, M., Mahfouz, A., Reinders, M. J. T., & More Authors (2022). The Effect of Phenotype and Genotype on the Plasma Proteome in Patients with Inflammatory Bowel Disease. *Journal of Crohn's and Colitis*, *16*(3), 414-429. https://doi.org/10.1093/ecco-jcc/jjab157

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Journal of Crohn's and Colitis, 2022, 414–429
https://doi.org/10.1093/ecco-jcc/jjab157
Advance Access publication September 7, 2021
Original Article

Original Article

The Effect of Phenotype and Genotype on the Plasma Proteome in Patients with Inflammatory Bowel Disease

Arno R. Bourgonje,[a*] Shixian Hu,[a,b*] Lieke M. Spekhorst,[a*] Daria V. Zhernakova,[b,c] Arnau Vich Vila,[a,b] Yanni Li,[a,b] Michiel D. Voskuil,[a,b] Lisette A. van Berkel,[d] Brenda Bley Folly,[d] Mohammed Charrout,[e,f] Ahmed Mahfouz,[e,f,g] Marcel J. T. Reinders,[e,f] Julia I. P. van Heck,[h] Leo A. B. Joosten,[h] Marijn C. Visschedijk,[a] Hendrik M. van Dullemen,[a] Klaas Nico Faber,[a] Janneke N. Samsom,[d] Eleonora A. M. Festen,[a,b] Gerard Dijkstra,[a] Rinse K. Weersma[a]

[a]Department of Gastroenterology and Hepatology, University Medical Center Groningen, Groningen, The Netherlands [b]Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands [c]Laboratory of Genomic Diversity, Center for Computer Technologies, ITMO University, St Petersburg, Russia [d]Department of Pediatrics, Division of Gastroenterology, Erasmus University Medical Center, Rotterdam, The Netherlands [e]Delft Bioinformatics Lab, Delft University of Technology, Delft, The Netherlands [f]Leiden Computational Biology Center, Leiden University Medical Center, Leiden, The Netherlands [g]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands [h]Department of Internal Medicine and Radboud Institute of Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands

*Joint first authors.

Corresponding author: Prof. Rinse K. Weersma, MD, PhD, Department of Gastroenterology and Hepatology, University of Groningen and University Medical Center Groningen, P.O. Box 30.001, 9700 RB Groningen, The Netherlands. Tel.: +31 50 361 26 20; fax: +31 50 361 93 06; email: r.k.weersma@umcg.nl

Abstract

Background and Aims: Protein profiling in patients with inflammatory bowel diseases [IBD] for diagnostic and therapeutic purposes is underexplored. This study analysed the association between phenotype, genotype, and the plasma proteome in IBD.

Methods: A total of 92 inflammation-related proteins were quantified in plasma of 1028 patients with IBD (567 Crohn's disease [CD]; 461 ulcerative colitis [UC]) and 148 healthy individuals to assess protein-phenotype associations. Corresponding whole-exome sequencing and global screening array data of 919 patients with IBD were included to analyse the effect of genetics on protein levels (protein quantitative trait loci [pQTL] analysis). Intestinal mucosal RNA sequencing and faecal metagenomic data were used for complementary analyses.

Results: Thirty-two proteins were differentially abundant between IBD and healthy individuals, of which 22 proteins were independent of active inflammation; 69 proteins were associated with 15 demographic and clinical factors. Fibroblast growth factor-19 levels were decreased in CD patients with ileal disease or a history of ileocecal resection. Thirteen novel *cis*-pQTLs were identified and 10 replicated from previous studies. One *trans*-pQTL of the fucosyltransferase 2 [*FUT2*] gene [rs602662] and two independent *cis*-pQTLs of C-C motif chemokine 25 [*CCL25*] affected plasma CCL25 levels. Intestinal gene expression data revealed an overlapping *cis*-expression [e]QTL-variant [rs3745387]

of the *CCL25* gene. The *FUT2* rs602662 *trans*-pQTL was associated with reduced abundances of faecal butyrate-producing bacteria.

Conclusions: This study shows that genotype and multiple disease phenotypes strongly associate with the plasma inflammatory proteome in IBD, and identifies disease-associated pathways that may help to improve disease management in the future.

Key Words: Inflammatory bowel disease; genetics; proteomics

1. Introduction

Inflammatory bowel diseases [IBD], encompassing Crohn's disease [CD] and ulcerative colitis [UC], are complex immune-mediated diseases of the gastrointestinal [GI] tract. Although the aetiology of IBD remains unclear, it involves a complex interplay between host genetics, the gut microbiome, a dysregulated immunological response, and environmental triggers.[1,2] IBD is a heterogeneous disease, impeding the prediction of disease course and therapeutic response.[3,4] Consequently, clinicians are being challenged in disease management as symptomatology is often non-specific and surrogate disease biomarkers are lacking.[5,6]

Over the past decades, there has been made a tremendous effort to unravel the role of genetics in IBD susceptibility.[7] To date, more than 200 genomic loci have been associated with IBD disease risk.[8,9] The field of genetics has shifted from genome-wide association studies [GWAS] towards gene expression studies in relevant tissues.

More recently, there is a growing interest in the function and variability of circulating proteins, as this has the potential to improve our understanding of biological pathways involved in IBD. Proteins can be regarded as intermediate phenotypes, connecting genetic variation to clinical traits by perturbation of their levels. Whereas protein profiling has been well studied in healthy human populations, it has not yet been systematically performed in patients with IBD.[10,11] High-throughput proteomic research techniques have evolved only recently, and large studies so far have been primarily focused on metabolic markers and cardiovascular disease.[12,13] Protein profiling in patients with IBD could be of importance for diagnostic purposes, monitoring disease activity, identification of therapeutic targets, and predicting response to treatment.[6,10,13] For instance, assessment of interactions between genotype and the plasma proteome could lead to the identification of associations that overlap with known genetic risk variants, potentially exposing disease-associated pathways, and accelerating the discovery of potential drug targets and translational biomarkers. Such a study should be accompanied by detailed phenotypic patient information to enable stratification for adequate estimation of inter-individual protein variability.

In this study, we quantified the plasma protein profile of IBD by performing a proximity extension assay [PEA] [Olink Proteomics®], a large-scale high-throughput proteomics screening technology. For this study, we used an assay including 92 inflammation-related plasma proteins, which was performed in 1028 patients with IBD and in 148 healthy individuals. First, we aimed to study the associations between demographic and clinical factors and plasma protein levels to uncover their contributions in shaping the plasma proteome. Second, we aimed to assess the effect of genotype on protein levels [protein quantitative trait loci, pQTL] in the context of IBD. Subsequently, we performed further *in silico* downstream analysis by integrating RNA-sequencing data from intestinal biopsies and gut microbiome data from faecal samples from the same patient cohort, to provide more insight into the observed pQTLs. Here, we present the largest pQTL study performed thus far in patients with IBD, with integration of multiple biological data layers. Identification of blood-based pQTLs is important, because in future studies it may help to stratify patients according to treatment response, and assist in drug selection, safety, and repurposing; and co-localising identified pQTLs with known IBD risk loci may expose novel molecular pathways relevant to IBD.

2. Methods

2.1. Study cohort: 1000IBD

This study was conducted at the University Medical Center Groningen [UMCG], Groningen, The Netherlands. Patients were included based on their participation in the 1000IBD project.[14] Within the 1000IBD project, detailed phenotypic information and multi-omics profiles have been collected for over 1000 patients with IBD. Patients included in this study were enrolled in the 1000IBD project from November 2009 to April 2019. Upon inclusion in the 1000IBD cohort, plasma was drawn from the patients for protein profiling. For each patient, detailed demographic and clinical information was available, including age, sex, body mass index [BMI], smoking status, medication use [including biologic use], history of bowel surgery, and disease activity, all of which was assessed at time of plasma sampling. Clinical disease activity was assessed at time of plasma sampling using the Harvey-Bradshaw Index [HBI] for patients with CD and the Simple Clinical Colitis Activity Index [SCCAI] for patients with UC. The Montreal disease classification was recorded from the latest visit to the outpatient clinic. Serum C-reactive protein [CRP] levels were routinely measured by nephelometry at the exact same day of plasma sampling. Patients provided written informed consent (study was approved by the Institutional Review Board [IRB] of the UMCG [registered as no. 08/338]). In addition, we included Dutch non-IBD controls [n = 148] from the 300BCG cohort [Radboud University Medical Center, Nijmegen, The Netherlands] to comparatively identify disease-associated proteins, which were included in the time period from April 2017 to June 2018.[15] These healthy individuals also provided written informed consent (study approved by the IRB of the Arnhem-Nijmegen Medical Ethical Committee [registered as NL58553.091.16]). The study has been performed in accordance with the principles of the Declaration of Helsinki [2013]. A methodological workflow of this study is presented in Figure 1.

2.2. Proximity extension assay technology

Plasma concentrations of 92—mainly inflammation-related—proteins were measured using the proximity extension immunoassay [PEA] technology [Olink Proteomics®, Uppsala, Sweden] with the ProSeek Multiplex Inflammation panel. A complete list of all 92 proteins with their full names, abbreviations, and corresponding UniProt IDs can be found in Supplementary Table S1, available as Supplementary data at *ECCO-JCC* online. To reduce technical

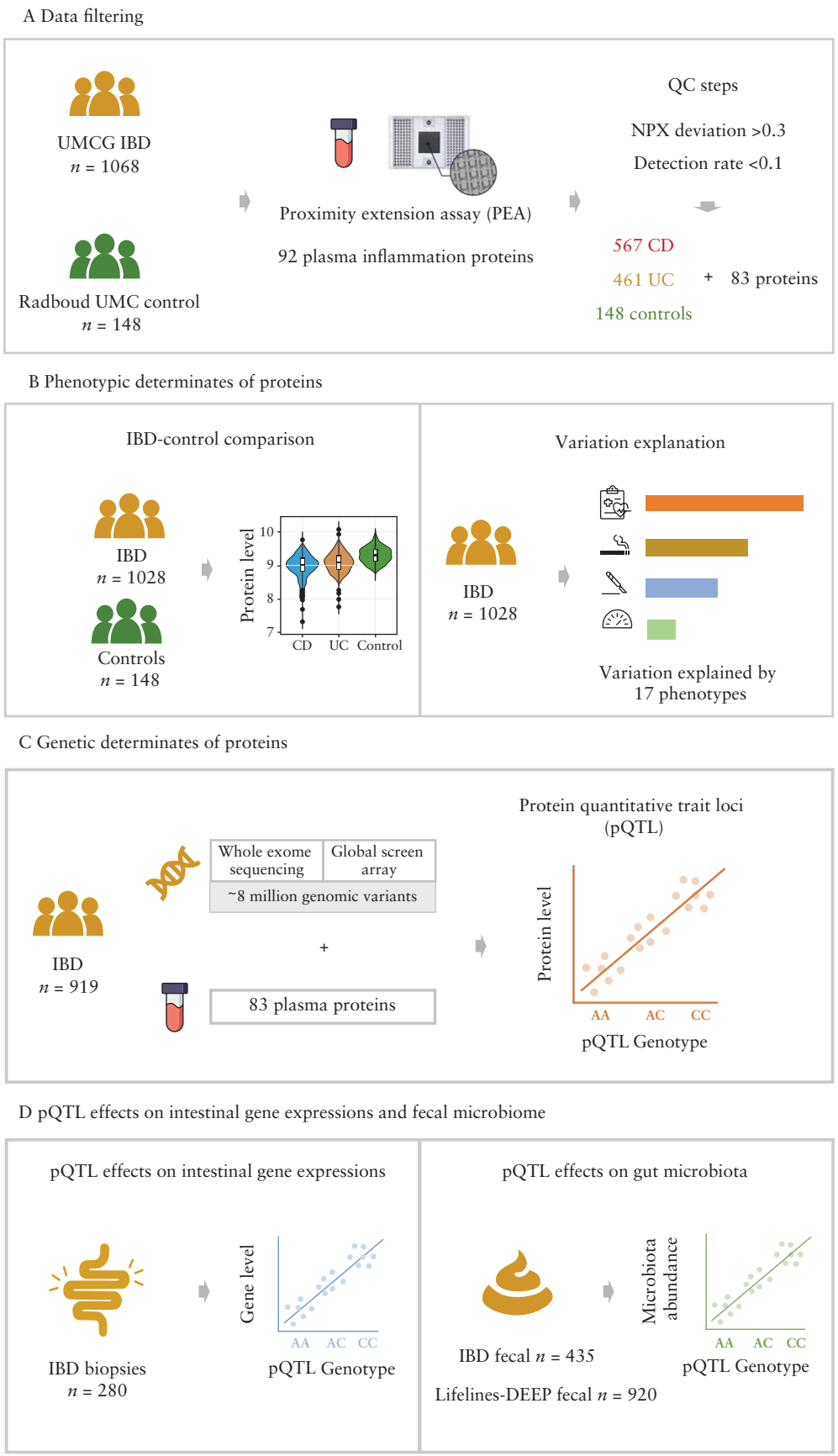

Figure 1. Schematic overview of the methodological study workflow. In this study, we aimed to analyse the effect of phenotype and genotype on the plasma proteome in patients with IBD by integrating information from multiple biological data layers, permitting comprehensive assessment of the observed findings.

variation between plates, plasma samples were randomised on different plates using a randomisation algorithm, including randomisation over age, sex, and IBD subtypes. Samples were measured in the Olink® testing facility in Uppsala, Sweden. Using PEA technology, 92 matched oligonucleotide-labelled antibody pairs [probes] were added to the samples and allowed to pair-wise bind to the target protein biomarkers present in the samples. When two probes of the same type are brought in close proximity, hybridisation occurs, followed by DNA polymerase extension. Subsequently, the resulting DNA sequence is detected and amplified by real-time microfluidic quantitative polymerase chain reaction [qPCR] [Biomark HD Instrument, Fluidigm®, San Francisco, CA, USA].[16]

Before analysis, an inter-plate intensity normalisation procedure was performed using the plate median as normalisation factor. Data were normalised on a log2-scale where values were derived from inverted Ct-values of real-time qPCR and expressed as normalised protein expression [NPX] values. NPX values are arbitrary units and only represent relative quantification, meaning that values can be compared for the same protein across samples, though no comparison can be made for absolute levels between different proteins. Samples that deviated >0.3 NPX from the median of the internal controls did not pass quality control [QC] and were excluded [n = 40]. The tumour necrosis factor alpha [TNF-α] protein [UniProt ID: P01375] was excluded from the analysis, as the Olink TNF-α assay [no. 95302] used for this study revealed suboptimal results as it is excessively influenced by anti-TNF-α antibodies [e.g., infliximab, adalimumab bound TNF-α]. The assay employs polyclonal antibodies, which also allows for the detection of the monomeric TNF-α form, meaning that biologically inactive forms were also detected, as we observed in our data.[17,18] In addition, eight proteins (fibroblast growth factor-5 [FGF-5], interleukin-1 alpha [IL-1α], interleukin-2 [IL-2], interleukin-20 [IL-20], interleukin-22 receptor subunit alpha-1 [IL-22RA1], interleukin-33 [IL-33], leukaemia inhibitory factor [LIF], and thymic stromal lymphopoietin [TSLP]) with a very low detection rate [<10%] in both healthy individuals and patients with IBD were removed across all samples, whereas one protein [neurturin, NRTN] only had a detection rate <10% in healthy individuals and was retained. Proteins with NPX values below the detection limit were treated as missing values [for detection rates, see Supplementary Table S1], and including them did not change the obtained results. In the present study, proteomic profiling of 83 proteins was available for 1028 patients with IBD [567 CD and 461 UC] and 148 healthy controls.

2.3. Whole-exome sequencing [WES] and global screening array [GSA]

Patients were genotyped using both whole-exome sequencing and a genome-wide genotyping array, performed with DNA derived from blood samples.[19] WES data were obtained from 840 patients with IBD. Library preparation and sequencing were done at the Broad Institute of the Massachusetts Institute of Technology [MIT] and Harvard University. On average, 86.06 million high-quality reads were generated per sample, and 98.85% of reads were aligned to a human reference genome [hg19]. Moreover, 81% of the exonic regions were covered with a read depth >30x. Next, the Genome Analysis Toolkit was used for variant calling. Variants with a call rate <0.99 or Hardy-Weinberg equilibrium test with p <0.0001 were excluded by using PLINK v.1.9. Only genetic data from patients clustering with individuals of European descent were included in the analyses.

GSA data were generated for 936 patients with IBD, using the Infinium GSA-24 v1.0 BeadChip combined with the optional Multi-Disease drop-in panel [http://glimdna.org/globalscreening-array.html, GSA-MD]. Genotypes were called using OptiCall [ref opticall.bitbucket.io], and QC steps were performed using PLINK v.1.9 (minor allele frequency [MAF] >5%, call rate < 0.99, Hardy-Weinberg equilibrium test p <0.0001). Genotype data were phased using the Eagle algorithm and imputed to the Haplotype Reference Consortium reference panel using the Michigan Imputation Server [https://imputationserver.readthedocs.io/en/latest/pipeline/]. After imputation, genetic variants were filtered for R2 >0.4. GSA genotype data were combined with WES data using PLINK 1.9. Variants with an MAF <5% were removed. The combined WES-GSA genetic dataset covered a total of 8 142 054 variants for 919 patients with IBD [517 CD and 402 UC].

2.4. Data processing of RNA-sequencing of intestinal biopsies

Intestinal mucosal bulk RNA sequencing was performed on 299 intestinal biopsies of 171 patients with IBD and has been described before.[20] In brief, 26 million paired-end 150-bp reads were generated per sample. The quality of the raw reads was checked using FastQC with default parameters [v0.11.7]. The adaptors and low-quality reads were clipped using Trimmomatic [v0.36] with settings length <50 nucleotides, quality <25. Reads were aligned to the human genome [Homo_sapiens_assembly19.fasta] using STAR [v2.7.3]. Reads sorting and mapping statistics were obtained using SAMtools [v0.1.19], sambamba [v0.7.0], and picard [v2.20.5]. Gene expression was estimated through HTSeq [0.9.1] based on the annotation from GTEx v7 [gencode.v19.annotation.patched_contigs.gtf]. After QC, data were available on 280 intestinal biopsies of 165 patients with IBD [ileum biopsies, n = 89, colon biopsies, n = 191].

2.5. Data obtained for microbial quantitative trait loci [mbQTL] analysis

To determine the effect of the *FUT2* gene on the gut microbiota composition [mbQTL], we obtained shotgun metagenomics sequencing data from a previous study, including a subset of the present IBD

[A] Proteomics data were generated for both patients with IBD [n = 1068] and healthy controls [n = 148] by proximity extension assay technology [PEA, Olink® Proteomics]. Filtering of proteomics data was performed by performing quality control [QC] steps: patients were excluded when their samples did not pass QC, i.e., deviation of >0.3 in normalised protein expression [NPX] value from the median of internal controls, or with sample detection rates <10%, resulting in plasma levels of 83 different proteins for 1028 patients with IBD and 148 healthy controls. [B] As first analysis step, case-control analyses were performed for all 83 proteins, comparing their levels between patients with IBD and healthy controls. Subsequently, in all patients with IBD, and separately in patients with CD [n = 567] and UC [n = 461], the contribution of 17 different demographic and clinical factors in explaining the variance of plasma protein levels was determined. [C] Genetic determinates of plasma proteins were established by integrating whole-exome sequencing [WES] and genome-wide genotyping array [GSA] data of 919 patients with IBD, and performing association analyses between >8 million genetic variants and plasma protein levels (protein quantitative trait loci [pQTL] analysis). [D] Downstream complementary analyses were performed on the observed findings from the pQTL analysis by integrating RNA sequencing data [n = 280 IBD] and faecal microbiome data [n = 435 IBD] and studying the effect of pQTL variants on intestinal gene expressions and microbial abundances, respectively. IBD, inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis.

cohort [n = 435], and a population-based cohort Lifelines-DEEP [n = 920].[19] Sequencing reads that mapped to the human genome [version NCBI37] were removed using Kneaddata [v0.5.1]. Microbiome taxa and predicted pathways profiling were performed using MetaPhlan [v2.6.0] and HUMAnN2 [v0.6.1]. The mbQTL effect of the *FUT2* gene was assessed in 435 patients with IBD and 920 population-based individuals.

2.6. Statistical analysis

2.6.1. Descriptive statistics

Data were presented as medians [interquartile range, IQR] or as proportions n with corresponding percentages [%]. Descriptive variables were compared between groups using Mann-Whitney U tests or Kruskal-Wallis tests. Group comparisons were performed by Pearson's chi square test or Fisher's exact test if n of observations were <10; p-values <0.05 were considered significant.

2.6.2. Associations between demographic and clinical variables and plasma proteins

All analyses were performed in R [v.3.6.3]. Principal component analysis [PCA] was used for dimensionality reduction for all 83 plasma proteins in 1028 patients with IBD and 148 healthy controls. Each protein was compared between groups, including IBD [full cohort], CD, and UC vs. healthy controls, and CD vs UC using Mann-Whitney U tests. This analysis was repeated with covariate adjustment for age, sex, and BMI using a general linear model, except for geographical location of which we could not exclude a possible confounding effect. To assess the associations between demographic and clinical factors and plasma proteins in patients with IBD or separately within remissive and active disease groups [CD and UC separately], a multivariate generalised linear model containing 17 different demographic and clinical variables and all 83 proteins was performed.[21,22] Stratified analysis of this model was performed for disease activity, where patients with C-reactive protein [CRP] levels <5 mg/L and low clinical disease activity scores (Harvey-Bradshaw Index [HBI] <5 in case of CD and Simple Clinical Colitis Activity Index [SCCAI] ≤ 2 in case of UC) were considered to be in remission, as opposed to the remainder of patients who were categorised as having 'active disease'. Individual protein variation explained by each factor was further assessed by an analysis of variance [ANOVA] on the generalised linear model. Protein levels were individually corrected for statistically significant demographic and clinical variables, and corrected estimates for each plasma protein were incorporated into further analyses. Differential protein level analysis was performed between categories of Montreal classifications [CD: disease location and disease behaviour; UC: disease extension] as well as associations between proteins and the HBI and SCCAI scores. For all analyses, a false discovery rate [FDR] <0.05 was considered as statistically significant.

2.6.3. Protein quantitative trait loci mapping

Cis-pQTL variants were defined as genomic variants located within ± 1 Mb of the region of each protein-coding gene centre, whereas *trans*-pQTL variants were defined as variants located ± 1 Mb outside the region of each protein-coding gene centre. For both *cis*-pQTL and *trans*-pQTL mapping, we first performed the analysis in patients with CD and UC separately, followed by a weighted z-score meta-analysis within the full IBD cohort. A step-wise conditional analysis was used for pQTL identification.[11,23] Briefly, Spearman rank correlation tests were performed to assess the effect of all genetic variants on protein level in the first round, adjusted for covariates using corrected estimates [see above]. To identify all independent *cis*-pQTLs, in subsequent rounds we regressed out the top statistically significant pQTLs from the last round until no independent signal was present any more. *Trans*-pQTL mapping was performed while correcting for all independent statistically significant *cis*-pQTLs. The Bonferroni method was used to correct for multiple comparisons, accounting for the test numbers of all variant-protein combinations. For *cis*-pQTL analysis, the threshold for statistical significance was 1.41×10^{-7} [0.05/353 612]. For *trans*-pQTL analysis, the significance threshold was 1.01×10^{-11} [0.05/496 624 608].

2.6.4. Expression quantitative trait loci [eQTL] and microbial quantitative trait loci [mbQTL] mapping

Statistically significant pQTL variants were selected and their effects on intestinal mucosal gene expression [eQTL mapping] and associations with the gut microbiota [mbQTL mapping] were analysed. This analysis was performed in the 1000IBD and LifeLines-DEEP cohorts separately, followed by a weighted z-score meta-analysis. Details of these analyses are provided in the Supplementary Methods, available as Supplementary data at *ECCO-JCC* online.

3. Results

3.1. Cohort description

Demographic and clinical characteristics of the study population [IBD: n = 1028; healthy controls: n = 148] are presented in Table 1. In total, 567 patients had a diagnosis of CD and 461 patients were diagnosed with UC. More females had CD and more men had UC [p <0.01], and the proportion of females was higher in patients with IBD compared with healthy individuals [p <0.01]. Median age at date of plasma sampling was 38 years in patients with CD compared with 43 years in patients with UC [p <0.01], whereas healthy controls were younger [p <0.01]. Patients with CD smoked more often compared with patients with UC [p <0.01]. Concerning biologic use, patients with CD used more anti-TNF-α [e.g., infliximab, adalimumab, certolizumab, and golimumab] compared with patients with UC [p <0.01]. Patients with CD also used thiopurines and methotrexate more often [both p <0.01], whereas patients with UC used aminosalicylates and calcineurin inhibitors more often [both p <0.01]. Oral contraceptives were more frequently used by patients with CD [p <0.05], which might partially be explained by the higher percentage of females in this group [64%]. Few patients [CD: 0.9%; UC: 2%] used mycophenolate mofetil due to severe CD or because they were liver transplant recipients. Patients with CD more often underwent an ileocaecal resection compared with patients with UC [p <0.01], whereas there was no significant difference for [partial] colon resections [p = 0.36].

3.2. Distinct plasma protein signatures between [quiescent] IBD and healthy individuals

Heterogeneity of plasma protein levels was visualised by PCA, where healthy controls were different from patients with IBD by the first PC [PC1 comparison, CD vs controls: $p = 3.10 \times 10^{-29}$; UC vs controls: $p = 7.77 \times 10^{-25}$; Figure 2A]; 59 proteins were different between healthy controls and IBD [IBD, CD, or UC, FDR <0.05, Supplementary Table S2, available as Supplementary data at *ECCO-JCC* online]. After adjustment for age, sex, and BMI, 32 proteins were differentially abundant, of which 24 proteins were specifically different in CD, 20 in UC, and 26 in the full IBD cohort [FDR

Table 1. Descriptive statistics of the study population [*n* = 1028 patients with IBD and *n* = 148 healthy individuals].

Variable	CD	UC	HC	p-value
	n = 567	*n* = 461	*n* = 148	
Sex, *n* [%]	567 [100%]	461 [100%]	148 [100%]	
Male	206 [36%]	223 [48%]	73 [51%]	<0.01
Female	361 [64%]	238 [52%]	75 [51%]	
Age [years]	38 [27;53]	43 [30;55]	26 [16;36]	<0.01
BMI, kg/m²	24 [21.62;27.95]	25 [22.39;28.34]	22 [20;24]	<0.01
Plasma storage time [years]	6.7 [4.3;8.2]	6.2 [4.2;7.9]	2.6 [2.5;2.7]	<0.001
Current smoking, *n* [%]	543 [96%]	437 [95%]	-	
Yes	164 [30%]	47 [11%]	-	<0.01
No	379 [70%]	390 [89%]	-	
Montreal classification				
Montreal age [A]	565 [100%][a]	460 [100%][b]		<0.01
A1 [≤16 years]	87 [15%]	55 [12%]	-	
A2 [17–40 years]	371 [66%]	278 [60%]	-	
A3 [>40 years]	107 [19%]	127 [28%]	-	
Montreal location [L]	567 [100%]	-		
L1 [ileal disease]	201 [35%]	-	-	
L2 [colonic disease]	111 [20%]	-	-	
L3 [ileocolonic disease]	200 [35%]	-	-	
L4 [upper GI disease]	10 [2%]	-	-	
L1 + L4	18 [3%]	-	-	
L2 + L4	11 [2%]	-	-	
L3 + L4	16 [3%]	-	-	
Montreal behaviour [B]	567 [100%]	-		
B1 [nonstricturing, nonpenetrating]	230 [40%]	-	-	
B2 [stricturing]	114 [20%]	-	-	
B3 [penetrating]	56 [10%]	-	-	
B1 + P [perianal disease]	58 [10%]	-	-	
B2 + P [perianal disease]	66 [12%]	-	-	
B3 + P [perianal disease]	43 [8%]	-	-	
Montreal extension [E]	-	450 [98%]		
E1 [proctitis]	-	59 [13%]	-	
E2 [left-sided colitis]	-	144 [32%]	-	
E3 [pancolitis]	-	247 [55%]	-	
Medication use, *n* [%]	567 [100%]	461 [100%]		
Aminosalicylates	60 [11%]	305 [66%]	-	<0.01
Thiopurines	252 [44%]	151 [33%]	-	<0.01
Steroids	108 [19%]	108 [23%]	-	0.09
Calcineurin inhibitors	10 [2%]	23 [5%]	-	<0.01
Methotrexate	51 [9%]	4 [0.9%]	-	<0.01
Mycophenolate mofetil	5 [0.9%]	9 [2%]	-	0.18
Oral contraceptives	72 [13%]	36 [8%]	-	<0.05
Antibiotics	20 [4%]	13 [3%]	-	0.52
Anti-TNF-α[c]	192 [34%]	39 [8%]	-	<0.01
Disease activity score	540 [95%]	443 [96%]		
HBI score [CD]			-	
Remission <5	372 [69%]			
Mild disease 5–7	88 [16%]			
Moderate disease 8–16	72 [13%]			
Severe disease >16	8 [2%]		-	
SCCAI score [UC]				
Remission ≤2		306 [69%]		
Active disease >2		137 [31%]		
CRP	469 [83%]	373 [81%]		
<5 mg/L	329 [70%]	283 [76%]	-	0.06
≥5 mg/L	140 [30%]	90 [24%]	-	
Surgical history	567 [100%]	461 [100%]		
Ileocaecal resection, *n* [%]	168 [30%]	3 [0.7%]	-	<0.01
Colon resection [or partial], *n* [%]	84 [15%]	78 [17%]	-	0.36

Data are presented as proportions *n* with corresponding percentages [%] or median [interquartile range, IQR] in case of continuous variables. *p*-values <0.05 were considered statistically significant.
IBD, inflammatory bowel disease; CD, Crohn's disease; HC, healthy controls; GI, gastrointestinal; UC, ulcerative colitis; BMI, body-mass index; TNF-α, tumour necrosis factor alpha; HBI, Harvey-Bradshaw Index; SCCAI, Simple Clinical Colitis Activity Index.

[a]Actual percentage is 99.6%.

[b]Actual percentage is 99.8%.

[c]Anti-TNF-α use included infliximab, adalimumab, certolizumab, and golimumab.

<0.1, Figure 2B; Supplementary Table S3 and Supplementary Figures S1–3, available as Supplementary data at *ECCO-JCC* online]. Top significantly different plasma protein levels between either IBD, CD, or UC and healthy controls were Delta and Notch-like epidermal growth factor-related receptor [DNER], SIR2-like protein 2 [SIRT2], fibroblast growth factor-19 [FGF-19], oncostatin-M [OSM], axin-1

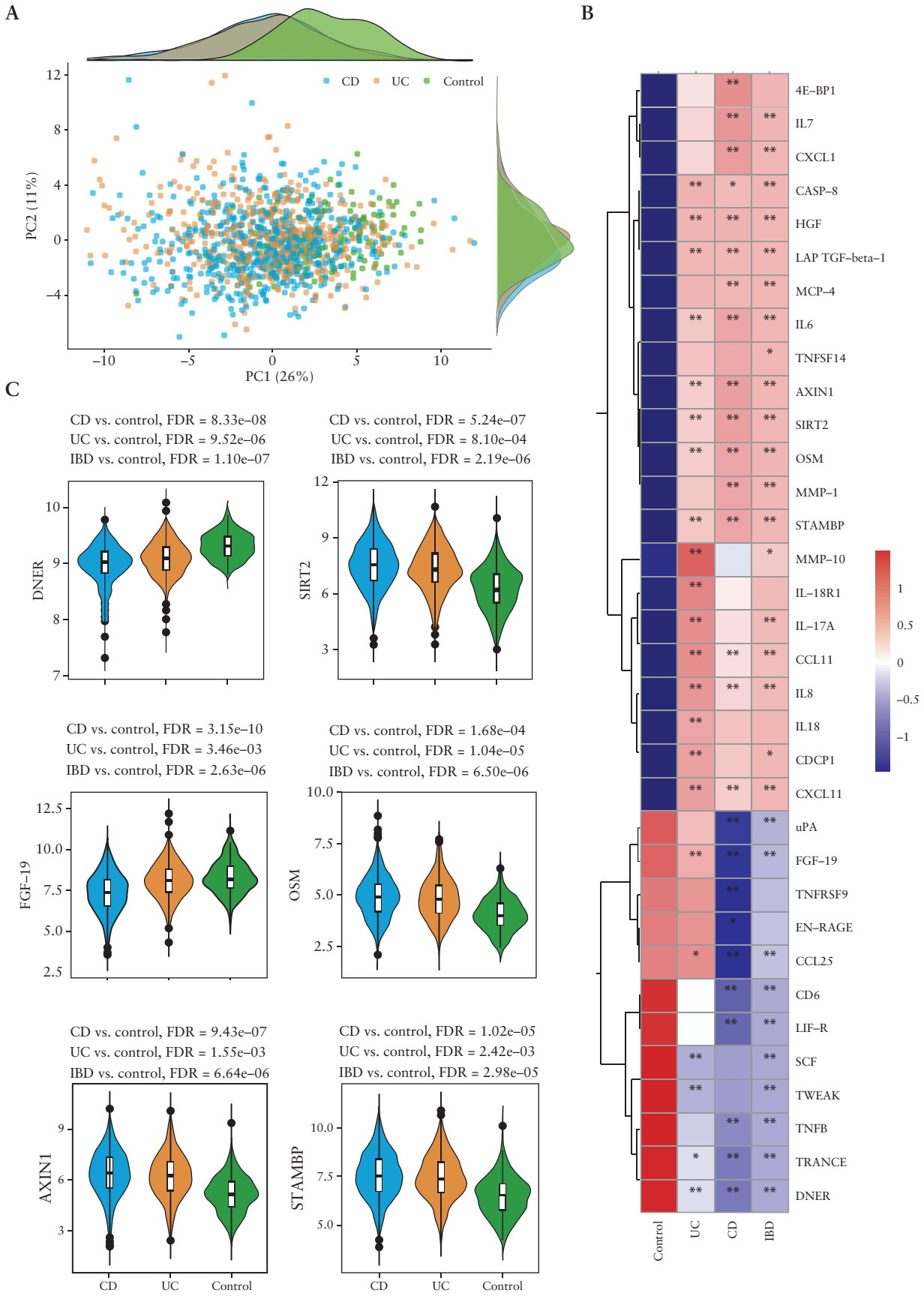

Figure 2. Distinct plasma protein signatures between patients with IBD and healthy individuals. [A] Principal component analysis [PCA] plot demonstrating the first two principal components [PCs] for patients with CD, UC, and healthy controls. Both CD and UC were significantly different from healthy controls

[AXIN1], and STAM-binding protein [STAMPB] [Figure 2C]. Ten proteins were increased in both CD and UC compared with controls, including OSM, interleukin-8 [IL-8], and interleukin-6 [IL-6], and two proteins were both decreased in CD and UC, namely DNER and FGF-19, which was confirmed when comparing the full IBD cohort against healthy controls. Of note, interleukin-17A [IL-17A] level was specifically elevated in UC, but not in CD, compared with controls.

Subsequently, to assess whether inflammatory proteins are still differentially abundant in quiescent IBD, we compared patients with IBD in clinical remission with healthy controls to account for disease activity [CD: HBI <5; UC: SCCAI ≤2; IBD: CRP <5 mg/l]. Here, 22 proteins were still different after adjustment for age, sex, and BMI, of which 15 specifically in CD, 12 in UC, and 16 in the full IBD cohort [FDR < 0.05, Supplementary Figure S4 and Supplementary Table S4, available as Supplementary data at *ECCO-JCC* online]. Importantly, top significant plasma proteins were fairly similar compared with the previous analysis. Next, we explored plasma proteins that distinguished quiescent CD from quiescent UC. Here, nine proteins were different, with three proteins (fibroblast growth factor-21 [FGF-21], interferon-gamma [IFN-γ] and osteoprotegerin [OPG]) being increased and six [FGF-19, interleukin-10 [IL-10], matrix metalloproteinase-10 [MMP-10], tumor necrosis factor superfamily member 9 [TNFRSF9], C-X-C motif chemokine 10 [CXCL10], and 9 [CXCL9]) decreased in CD compared with UC [Supplementary Figure S5 and Supplementary Tables S5 and S6, available as Supplementary data at *ECCO-JCC* online].

3.3. Demographic and clinical factors associated with plasma proteins in the context of IBD

Next, associations between 17 patient demographic and clinical factors [including IBD diagnosis] and levels of 83 plasma proteins were analysed in all patients with IBD [Figure 3]. Overall, all demographic and clinical factors were associated with at least one of the proteins. Considering the heterogeneity between CD and UC, we performed the same analysis in CD and UC separately. Here, 15 demographic or clinical factors were associated with the level of at least one out of 69 proteins in one or both IBD subtypes [CD and UC] [FDR <0.05] [Supplementary Figures S6 and 7 and Supplementary Table S7, available as Supplementary data at *ECCO-JCC* online]. Mycofenolate mofetil and methotrexate use were the only factors that were not significantly associated with any of the proteins [FDR >0.05]. Conversely, 14 proteins were not associated with any of the analysed demographic and clinical factors [FDR >0.05]. Most significant associations between proteins and demographic or clinical factors were consistent between patients with CD and UC [Supplementary Figures S6 and S7]. In subsequent analyses, we corrected for protein-specific statistically significant demographic or clinical factors [Supplementary Table S8, available as Supplementary data at *ECCO-JCC* online].

3.3.1. Plasma FGF-19 levels are decreased in patients with CD having ileal disease or a history of ileocaecal resection

In patients with CD, a history of ileocaecal resection was associated with a decreased level of FGF-19 [FDR <0.05]. In patients with UC, a history of colectomy was associated with a decreased level of FGF-19 and increased levels of interleukin-12 subunit beta [IL-12B], C-C motif chemokine 23 [CCL23], tumour necrosis factor-beta [TNF-B], and IL-6 [FDR <0.05]. In patients with CD, FGF-19 level was decreased in patients with solely ileal disease [Montreal L1] compared with patients with solely colonic disease [Montreal L2] [$p = 4.06 \times 10^{-11}$] [Figure 4A]. In addition, FGF-19 level was lower in patients with either stricturing [Montreal B2, $p = 1.61 \times 10^{-7}$] or penetrating [Montreal B3, $p = 6.69 \times 10^{-6}$] disease behaviour compared with non-stricturing, non-penetrating disease behaviour [Montreal B1] [Figure 4B] [FDR <0.05]. These differences remained statistically significant when patients with a history of ileocaecal resection were excluded from the analysis [Supplementary Figure S8]. In patients with UC, levels of IL-6, interleukin-15 receptor subunit alpha [IL-15RA], C-C motif chemokine 19 [CCL19], C-C motif chemokine 20 [CCL20], CXCL9, and CXCL10 were increased together with disease extent [Montreal E] [FDR <0.05] [Supplementary Figure S9, available as Supplementary data at *ECCO-JCC* online].

3.3.2. Smoking is associated with decreased plasma IL-12B levels

In patients with CD, active smoking behaviour was associated with decreased levels of IL-12B, stem cell factor [SCF], CXCL10, and beta-nerve growth factor [β-NGF] compared with non-smoking patients, and with increased levels of OSM, hepatocyte growth factor [HGF], C-C motif chemokine 11 [CCL11], MMP-10, monocyte chemotactic protein 1 [MCP-1], FGF-21, neurotrophin-3 [NT-3], and CCL20 as compared with non-smokers [FDR <0.05]. In patients with UC, active smoking was not associated [though near to significantly] with decreased levels of IL-12B [FDR = 0.06].

3.3.3. Disease activity affects plasma protein levels, but does not affect main phenotype-protein associations

Serum CRP levels were associated with 33 different plasma proteins in patients with CD, where top significant associations were observed for IL-6, macrophage colony-stimulating factor 1 [CSF-1], IFN-γ, CXCL9, and DNER [FDR <0.05, Supplementary Table S7]. Four of these 33 proteins were inversely associated with CRP levels: DNER, SCF, tumour necrosis factor [ligand] superfamily member 12 [TWEAK], and urokinase-type plasminogen activator [uPA]. In patients with UC, 25 different plasma proteins were associated with CRP levels, with top significant associations for IL-6, SCF, DNER, CSF-1, and matrix metalloproteinase-1 [MMP-1] [FDR <0.05]. Five of these 25 proteins were inversely associated with CRP: SCF, DNER, TWEAK, TNF-related activation-induced cytokine [TRANCE], and T-cell surface glycoprotein CD6 isoform [CD6]. Stratified analyses of associations between demographic and clinical factors and plasma proteins for disease activity [CD: $n = 372$ in remission; $n = 168$ active disease; UC: $n = 306$ in remission; $n = 137$ active disease; $n = 45$ with unknown HBI/SCCAI scores] demonstrated a high degree of consistency of main phenotypic-protein associations [Supplementary Table S9, available as Supplementary data at *ECCO-JCC* online].

3.3.4. Clinical disease activity modestly associates with plasma protein levels

In patients with CD, four proteins were associated [FDR <0.05] with clinical disease activity as measured by the Harvey-Bradshaw Index [HBI]

by the first PC [p <0.001]. [B] In total, 32 proteins were significantly different between either IBD [26 proteins], CD [24 proteins], or UC [20 proteins] and healthy controls, *FDR <0.1; **FDR <0.05. Red colouring indicates higher levels, whereas blue colouring indicates lower levels. [C] Top six most significantly different plasma protein levels between either IBD, CD, or UC and healthy controls [DNER, SIRT2, FGF-19, OSM, AXIN1, and STAMPB]. The y-axis indicates NPX values. Abbreviations: CD, Crohn's disease; UC, ulcerative colitis; IBD, inflammatory bowel disease; PC, principal component; FDR, false discovery rate; NPX, normalised protein expression.

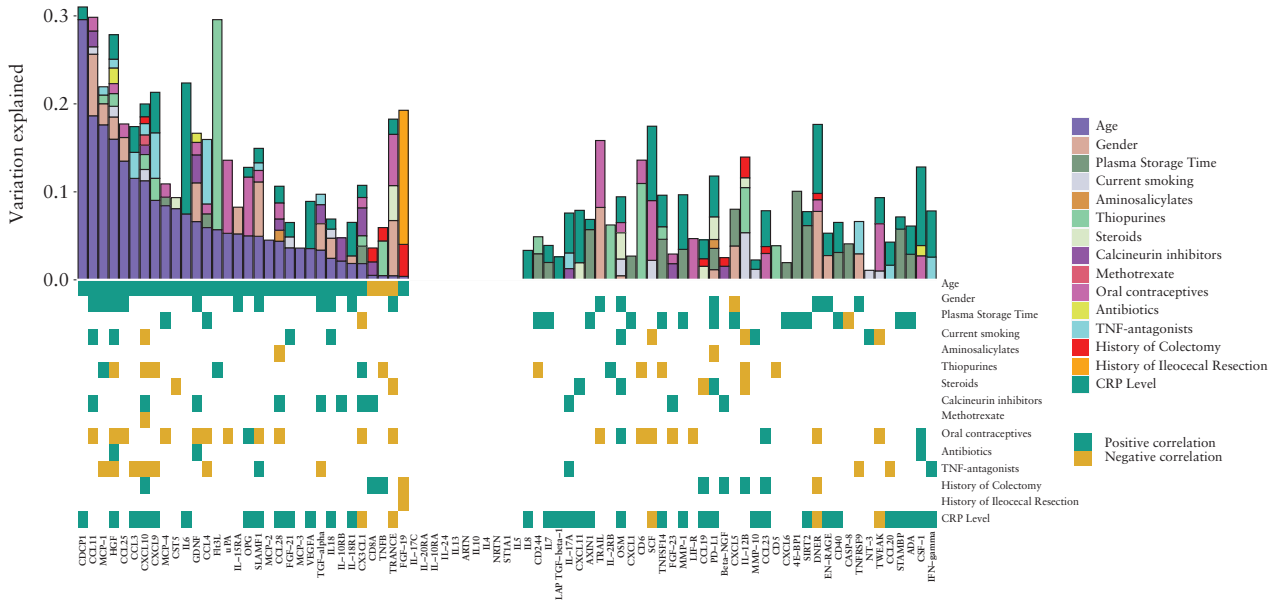

Figure 3. Demographic and clinical factors strongly influence the plasma proteome in patients with IBD. Associations between demographic and clinical factors and protein level in the full IBD cohort [with subtype of IBD included as covariate]. Bar plots indicate the variance explained by significant demographic and clinical factors [FDR <0.05] for each of the 83 proteins. The heatmap below indicates the directions of associations for each individual protein [positive associations, gender: female-to-male; non-medication users to medication users; non-surgery to surgery]. IBD, inflammatory bowel disease; FDR, false discovery rate.

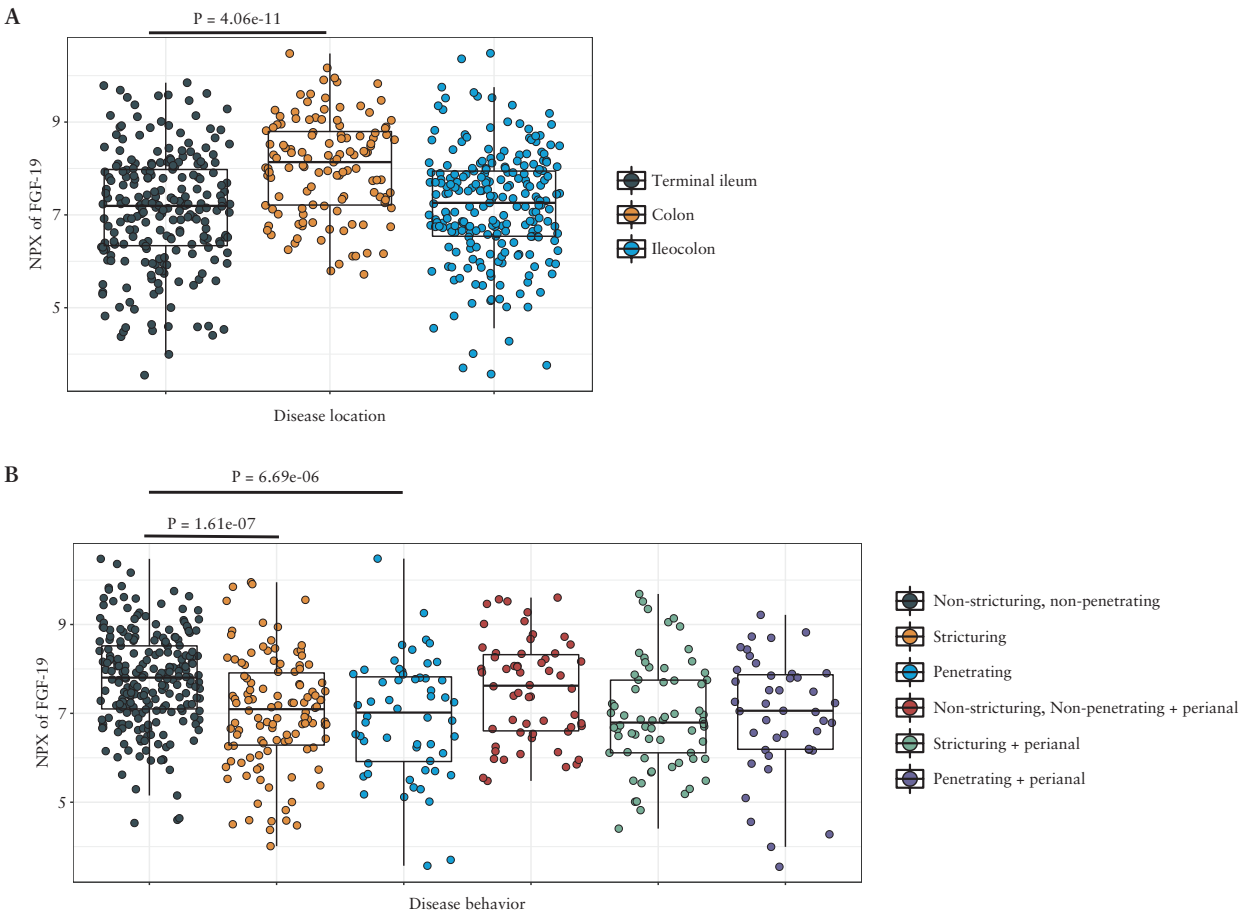

Figure 4. Plasma FGF-19 levels are decreased in patients with ileal and non-stricturing, non-penetrating Crohn's disease [CD]. [A] FGF-19 level is significantly elevated in patients with CD having colonic disease involvement compared with patients with ileal disease involvement. [B] FGF-19 level is decreased in patients with CD having stricturing or penetrating disease behaviour.

[Figure 5A; Supplementary Table S10, available as Supplementary data at *ECCO-JCC* online]. A negative association was observed for DNER [r = -0.18], whereas positive associations were found for IL-6 [r = 0.17], C-C motif chemokine 3 [CCL3] [r = 0.16], and OSM [r = 0.15] [FDR <0.05]. In patients with UC, five proteins were associated with clinical disease activity as measured by the Simple Clinical Colitis Activity Index [SCCAI] [Figure 5B]: IL-17A [r = 0.22], IL-8 [r = 0.22], transforming growth factor alpha [TGF-α] [r = 0.21], HGF [r = 0.20], and C-C motif chemokine 28 [CCL28] [r = 0.17] [FDR <0.05].

3.4. Associations between genetics and protein level

3.4.1. Protein quantitative trait loci [pQTL] analysis

To assess genetic associations with the targeted plasma proteome of inflammation-related proteins, both *cis*- and *trans*-pQTL mapping was performed separately for CD [n = 517] and UC [n = 402] [Supplementary Tables S11 and S12, available as Supplementary data at *ECCO-JCC* online], followed by a meta-analysis [IBD: n = 919]. In total, 1655 *cis*-pQTLs [Supplementary Table S13, available as Supplementary data at *ECCO-JCC* online], corresponding to 23 independent *cis*-pQTLs, were found to be statistically significantly associated with the level of 21 different proteins [Bonferroni-adjusted $p = 1.41 \times 10^{-7}$] [Table 2]. Plasma levels of two proteins, cystatin D [CST5] and C-C motif chemokine 25 [CCL25], were found to associate with two independent *cis*-pQTL variants, whereas levels of the other 19 proteins were associated with one independent *cis*-pQTL variant. Ten out of the 23 *cis*-pQTLs were reported in a previous population cohort-based pQTL analysis, and these associations were in the same direction.[10] For example, monocyte chemotactic protein 2 [MCP-2, CCL8] levels showed the strongest association with a specific missense variant [rs1133763, $p = 1.97 \times 10^{-49}$]. This study discovered 13 novel independent *cis*-pQTL variants, including genetic variants of CST5, TNFB, CD6, T-cell surface glycoprotein CD8 alpha chain [CD8A], adenosine deaminase [ADA], C-X-C motif chemokine 6 [CXCL6], interleukin-10 receptor subunit beta [IL10-RB], MMP-1, CD40L receptor [CD40], programmed cell death 1 ligand 1 [PD-L1], CUB domain-containing protein 1 [CDCP1], and one additional *cis*-pQTL variant of CCL25 [Table 2].

Next, the overlap between all detected *cis*-pQTLs with protein-coding gene expression quantitative trait loci [eQTLs] was investigated using intestinal mucosal biopsies derived from both ileum and colon of a subset of patients with IBD [n = 280 biopsies from 165 individual patients], and in the largest public IBD GWAS so far.[7] Four independent *cis*-pQTL variants also appeared to have an eQTL effect, including variants of *CCL25* [p = 0.0015, ileum], *CXCL5* [$p = 8.36 \times 10^{-7}$, colon], *MMP-1* [$p = 1.68 \times 10^{-4}$, colon], and *IL10-RB* [p = 0.0036, colon] [Supplementary Table S14, available as Supplementary data at *ECCO-JCC* online]. When checking the overlap with IBD GWAS signals, five *cis*-pQTL variants were located in known IBD genetic susceptibility loci, including CD40, CD6, IL15-RA, interleukin-18 receptor 1 [IL18-R1], and TNFB.

In a subsequent *trans*-pQTL analysis which was corrected for all statistically significant *cis*-pQTL variants, one independent *trans*-pQTL variant for the CCL25 protein [$p = 5.86 \times 10^{-22}$, rs602662] was further identified, which is located in the fucosyltransferase 2 [*FUT2*] gene, a known IBD-associated risk locus [Supplementary Table S15, available as Supplementary data at *ECCO-JCC* online].

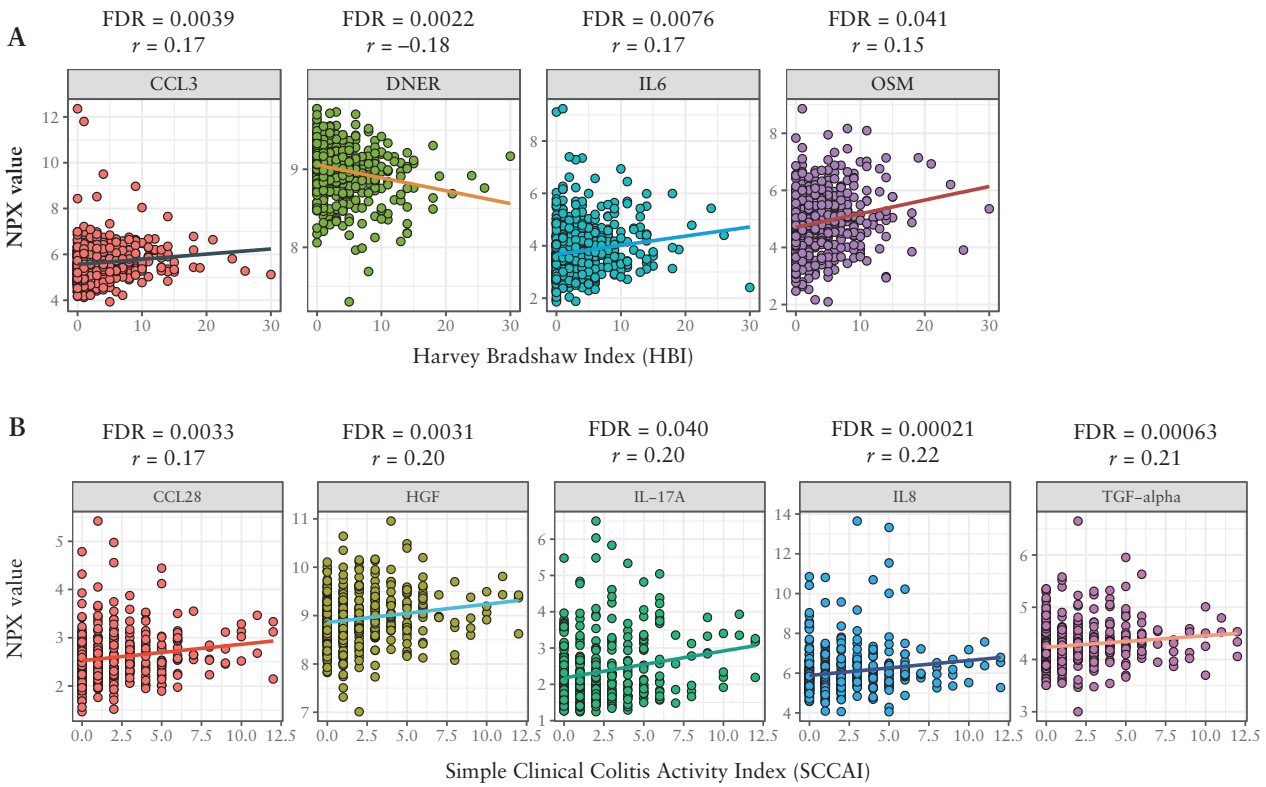

Figure 5. Associations between plasma proteins and clinical [HBI, SCCAI] disease activity. Top significant correlations [FDR <0.05] between plasma protein levels and measures of disease activity, calculated by Spearman's rank correlation coefficients. [A] Four proteins significantly correlate with the Harvey-Bradshaw Index [HBI] in patients with Crohn's disease [CD]. [B] Five proteins significantly correlate with the Simple Clinical Colitis Activity Index [SCCAI] in patients with ulcerative colitis [UC]. FDR, false discovery rate; *r*, correlation coefficient.

Table 2. Local [*cis*-]pQTL meta-analysis revealed 23 independent *cis*-pQTL variants for 21 different plasma proteins in patients with IBD.

Protein	SNP	Chr	Allele	p-value [CD]	r [CD]	p-value [UC]	r [UC]	Meta-p-value
CD8A	rs3020726[a]	2	G	2.03×10^{-19}	0.39	3.64×10^{-11}	0.32	1.19×10^{-28}
IL-18R1	rs11377261	2	A	1.55×10^{-22}	0.44	3.67×10^{-20}	0.47	1.77×10^{-40}
CDCP1	rs62244470	3	T	1.32×10^{-6}	-0.20	1.60×10^{-3}	-0.16	1.52×10^{-8}
CXCL1	rs3117604[a]	4	C	5.49×10^{-8}	0.25	1.15×10^{-3}	0.17	1.20×10^{-9}
CXCL5	rs425535[a]	4	T	1.20×10^{-10}	0.28	1.19×10^{-7}	0.26	1.49×10^{-16}
CXCL6	rs16850073	4	T	3.18×10^{-17}	0.38	1.50×10^{-10}	0.33	6.59×10^{-26}
IL-12B	rs4921484[a]	5	T	4.67×10^{-9}	-0.27	2.70×10^{-7}	-0.27	1.30×10^{-14}
TNFB	rs2229092	6	C	7.62×10^{-21}	-0.40	1.80×10^{-16}	-0.40	2.40×10^{-35}
VEGFA	rs6921438[a]	6	A	8.63×10^{-23}	-0.44	7.62×10^{-14}	-0.38	1.07×10^{-34}
PD-L1	rs35744625	9	C	6.62×10^{-4}	0.16	6.03×10^{-6}	0.23	7.04×10^{-8}
IL-15RA	rs4237402[a]	10	A	9.38×10^{-7}	0.23	1.96×10^{-3}	0.16	1.69×10^{-8}
CD6	rs11230563	11	T	4.84×10^{-17}	-0.37	1.74×10^{-12}	-0.34	1.17×10^{-27}
MMP-1	rs10791596	11	T	1.35×10^{-8}	-0.26	8.73×10^{-6}	-0.22	1.08×10^{-12}
CCL23	rs712048[a]	17	A	1.82×10^{-10}	-0.29	2.70×10^{-3}	-0.16	1.75×10^{-11}
CCL3	rs1719144[a]	17	A	1.18×10^{-7}	0.24	4.56×10^{-6}	0.23	4.87×10^{-12}
MCP-2	rs1133763[a]	17	C	6.52×10^{-33}	-0.50	1.29×10^{-18}	-0.42	1.97×10^{-49}
CCL25	rs2032887	19	G	2.88×10^{-26}	0.45	2.53×10^{-13}	0.36	1.92×10^{-37}
CCL25	rs3745387[a]	19	A	1.45×10^{-8}	0.25	3.95×10^{-5}	0.20	5.45×10^{-12}
ADA	rs112545364	20	G	7.35×10^{-17}	0.36	8.69×10^{-11}	0.32	8.60×10^{-26}
CD40	rs1883832	20	T	1.04×10^{-7}	-0.24	3.66×10^{-3}	-0.14	5.30×10^{-9}
CST5	rs34269359	20	A	4.69×10^{-19}	0.40	1.36×10^{-20}	0.46	3.86×10^{-37}
CST5	rs1799841	20	G	4.76×10^{-7}	0.20	9.86×10^{-5}	0.19	3.65×10^{-9}
IL-10RB	rs4455239	21	C	3.33×10^{-8}	0.26	5.28×10^{-15}	0.39	6.47×10^{-20}

IBD, inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis; Chr, chromosome; SNP, single nucleotide polymorphism; r, correlation coefficient.
[a]Replicated *cis*-pQTL variants compared with Sun *et al.*, *Nature* 2018.[10]

3.4.2. *Cis*- and *trans*-pQTL co-regulation effect on CCL25 plasma levels

We examined whether the observed independent *cis*-pQTL variants [rs2032887 and rs3745387] and the *trans*-pQTL variant [rs602662] for CCL25 could have a synergistic effect on plasma levels of CCL25 [Figure 6A]. Regarding the two *cis*-pQTL variants, CCL25 levels statistically significantly increased upon allele carrier status [G-allele for rs2032887 and A-allele for rs3745387], and the same was observed for the *trans*-pQTL variant rs602662 [G-allele] [Figure 6B and C, boxplots]. Subsequently, we observed an additive effect of both *cis*-pQTL variants and the *trans*-pQTL variant rs602622 on CCL25 levels [Figure 6D]. Of note, carriage of both *cis*-pQTL variants combined with the rs602662 *trans*-pQTL variant was associated with significantly higher CCL25 plasma levels, compared with carriage of both *cis*-pQTL variants, though without the *trans*-pQTL variant [$p = 2.55 \times 10^{-15}$]. To determine whether the genetic regulation can also be observed at gene expression level, these pQTL variants were analysed in relation to intestinal mucosal RNA sequencing data from a subset [$n = 280$] of the present IBD cohort. Here, we observed the *cis*-pQTL variant rs3745387 to be a *cis*-expression quantitative trait locus [eQTL]-variant for *CCL25* gene expression in ileal tissue [Figure 7A; $p = 0.0015$]. In contrast, the rs2032887 *cis*-pQTL variant for CCL25 did not show a significant eQTL effect [$p = 0.54$]. In addition, the *trans*-pQTL variant rs602662 for CCL25 was also observed to be a *cis*-eQTL variant on *FUT2* gene expression [$p = 2.26 \times 10^{-16}$]; however, no *trans*-eQTL effect on CCL25 was observed [Figure 7A].

3.4.3. Associations of the FUT2 rs602662 variant with gut microbial species

The *FUT2* gene is involved in intestinal mucosal barrier integrity and interacts with the gut microbiota.[11] To further explore the effect on the gut microbiota, we re-analysed data from a previous genome-wide mbQTL analysis, and found the rs602662 variant to be associated with lower abundance of the species *Ruminococcus obeum* [$p_{meta} = 0.00094$, Figure 7B; Supplementary Table S16, available as Supplementary data at *ECCO-JCC* online], belonging to the genus *Blautia*, of which many bacterial species are involved in the production of short-chain fatty acids [SCFAs].[19] Similarly, the rs602662 variant was observed to be associated with a lower abundance of *Faecalibacterium prausnitzii* [$p_{meta} = 0.00397$], a well-known, commensal, butyrate-producing bacterial species in the human gut.

4. Discussion

In this study, a multitude of statistically significant associations between demographic or clinical factors and plasma protein levels were uncovered. First, distinct plasma protein signatures were identified for both CD and UC, of which most proteins remained differentially abundant when quiescent IBD was compared with healthy individuals. In phenotype association analyses, active smoking was associated with decreased plasma IL-12B levels in patients with CD and UC. Furthermore, patients with ileal CD and a history of bowel surgery had decreased plasma levels of FGF-19 compared with patients with CD having colonic disease. Second, we identified 13 novel pQTL variants in the context of IBD and replicated 10 previously reported pQTL variants, together affecting levels of 21 [mainly inflammatory] proteins. Among these, we observed two independent *cis*-pQTL variants followed by a single *trans*-pQTL variant [rs602662] that were associated with plasma levels of the CCL25 chemokine. Even more important, we observed an additive effect of allele carrier status of both *cis*-pQTL variants and the *trans*-pQTL variant on CCL25 plasma levels. Third, integration with intestinal mucosal gene expression data showed that the *cis*-pQTL variant rs3745387 appeared to be a *cis*-expression quantitative trait locus

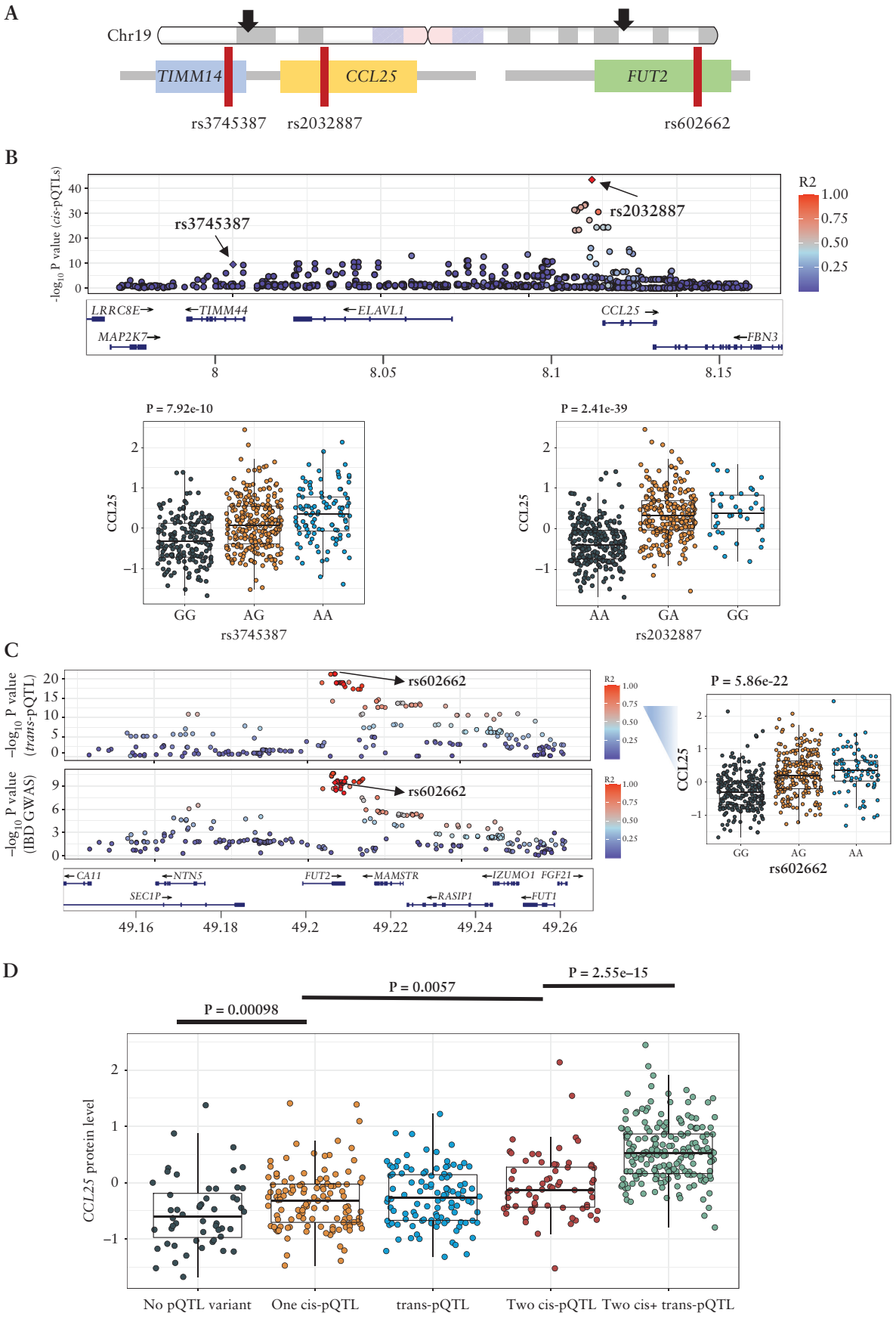

Figure 6. Co-regulation effect of *cis*- and *trans*-pQTL variants on plasma CCL25 levels. [A] Two independent *cis*-pQTL variants [rs2032887, located in the CCL25 protein-coding region, and rs3745387, located upstream of the CCL25 protein-coding region] and one *trans*-pQTL variant [rs602662, missense variant of the *FUT2*

[eQTL] variant for *CCL25* gene expression level. Complementary analysis of the *FUT2* rs602662 variant in relation to the gut microbiota [mbQTL analysis] showed significantly lower abundances of butyrate-producing bacterial species, including those belonging to the genus *Blautia* and genus *Faecalibacterium*.

In the present study, many plasma proteins remained differentially abundant in IBD vs healthy control analyses, indicating persistent systemic inflammation in patients with clinically quiescent IBD. Furthermore, it seems fundamental to focus on distinguished mechanisms and pathways active in each individual disease entity in comparison with healthy individuals. In our study, active smoking was strongly associated with lower IL-12B levels in both patients with CD and those with UC. Previous studies have shown that serum IL-12B levels, as well as *IL-12B* gene expression, are decreased upon cigarette smoke exposure compared with non-exposure.[24,25] Interestingly, *IL-12B* encodes the p40 subunit of IL-12, which is also part of IL-23, and is a known IBD susceptibility locus.[26] A previous study from our centre discovered a complex gene-environment interaction between the *IL-12B* SNP rs6887695 in non-smoking patients with IBD, but not in smoking patients.[27] Similarly, a more recent study identified a specific IL-12p40 genetic variant that was associated with increased circulating IL-12 levels in patients with coronary artery disease, but only in non-smoking patients.[24] Repressed IL-12B levels may be reflective of an impaired immunity and anti-tumour activity in smokers, but in the context of IBD, it may have additional implications.[24] For instance, as the biologic ustekinumab inhibits the IL-12/23 axis and thereby the activation of Th1/Th17-lymphocytes, one could hypothesise that smoking behaviour and its associated genetic background may modulate the response to this therapy in patients with IBD.

Strikingly, plasma levels of FGF-19 were significantly decreased in patients with CD compared with both UC and healthy individuals, in line with results from previous studies that showed an impaired Farnesoid X Receptor [FXR]-FGF-19 axis in CD.[28–35] Patients with CD having solely ileal disease or with a history of bowel surgery particularly exhibited decreased levels of plasma FGF-19 compared with patients with colonic disease. FGF-19 is mainly produced in the ileum upon activation of the FXR bile acid sensor and acts as a gut-derived hormone to inhibit bile acid synthesis in the liver.[28] Chronic diarrhoea due to bile acid malabsorption leads to decreased FGF-19 production, which in turn stimulates hepatic bile acid synthesis. Excessive bile acid production further aggravates bile acid malabsorption diarrhoea.[29] Patients with CD who underwent ileocaecal resection indeed showed impaired release of FGF-19 to the circulation.[30–32] Furthermore, intestinal inflammation disrupts the epithelial barrier integrity, which is accompanied by impairment of transepithelial transport mechanisms including bile acid reabsorption.[33] Therefore, not only ileocaecal resection but also intestinal inflammation, which could lead to deterioration of bile acid homeostasis, are associated with lower plasma FGF-19 levels.[34] Based on these findings, plasma FGF-19 levels may be a potential biomarker to identify patients who are most likely to benefit from therapy with bile acid sequestrants or FXR agonists. Furthermore, previous studies using chemically induced models of colitis demonstrated amelioration of intestinal inflammation, barrier integrity, and gut microbiota composition upon pharmacological FXR activation, demonstrating therapeutic potential especially in CD patients with active disease or surgical history and concurrent disruption of bile acid metabolism.[35–37]

In our large-scale pQTL analysis, one independent *trans*-pQTL variant and two *cis*-pQTL variants were significantly associated with plasma CCL25 levels. CCL25, also known as thymus-expressed chemokine [TECK], is constitutively expressed in the thymus and small intestinal epithelium, but nearly absent in the colon.[38] CCL25 elicits recruitment of peripheral blood lymphocytes expressing the chemokine receptor CCR9 and the adhesion molecule α4β7.[39] Through interaction with CCL25, CCR9-expressing lymphocytes are guided to the intra-epithelial lymphocyte [IEL] compartment, the small intestinal lamina propria and, to a lesser extent, the colonic lamina propria. Increased activity of the CCL25-CCR9 axis and ensuing $CD4^+$ effector memory T cell migration are implicated in primary sclerosing cholangitis, experimental postoperative ileus, and IBD.[40–43] In small intestinal CD, CCL25 is strongly expressed in proximity to lymphocytic infiltrates, and $CCR9^+$ gut-homing $CD4^+$ effector T cells are increased in the circulation when compared with colonic CD.[42] Recent data suggest that colonic CCL25 expression and $CCR9^+$ $CD4^+$ effector T cell recruitment may also be upregulated in patients with active colitis.[43] As such, the observed association of genotype and plasma CCL25 levels may play a functional role in the immune dysregulation in these patients, and this may be of particular interest for future functional studies. Although CCR9 has been considered a viable therapeutic target for years, clinical trials still have to demonstrate reliably efficacy of CCR9-inhibitors.[44] One independent *trans*-pQTL variant emerged in the *FUT2* gene [located at chromosome 19q13.33] affecting plasma CCL25 levels. However, no *trans*-eQTL effect on CCL25 was observed. Likewise, previous studies demonstrated limited overlap between pQTLs and eQTLs where, on average, only one-third of pQTLs appears to have a corresponding eQTL effect.[13]

At present, to our best knowledge, our study is the largest pQTL analysis performed in patients with IBD, using high-resolution genotype data. In a previous study, 41 pQTLs were identified in 51 patients with CD, in which—except for age and sex—no other clinical phenotypes were considered.[21] Furthermore in the present study, disease activity [e.g., as represented by serum CRP as indicator of systemic inflammation] affected levels of many [inflammatory] plasma proteins, including circulating levels of proteins that are well studied in the context of IBD [e.g., IL-6 that was the top significant association in both CD and UC, or IFN-γ which was strongly associated with disease activity in CD, confirming previous findings].[6,45] Finally however, some plasma proteins that associated with CRP are less well-reported in literature. For instance, the Delta and Notch-like epidermal growth factor-related receptor [DNER] protein, which is an activator of the Notch-1 pathway, was strongly inversely associated with CRP levels. Among various known functions, the Notch-1 signalling pathway regulates cellular apoptosis and intercellular interactions within the intestinal epithelium.[46] Notch-1 signalling is associated with enhanced mucosal barrier function, stimulated by

gene] for the CCL25 protein, all located on chromosome 19. [B] Upper panel shows a regional association plot of the two detected independent *cis*-pQTL variants rs2032887 and rs3745387 of the *CCL25* gene. Lower panel shows boxplots of *cis*-pQTL effects of these variants. [C] Upper panel shows a regional association plot of one independent *trans*-pQTL variant for CCL25 [rs602662] constituting a missense variant of the *FUT2* gene; the lower panel displays a regional association plot from a genome-wide association study [GWAS] in Crohn's disease [CD] for the rs602662 SNP of the *FUT2* gene.[10] Colour hues for r^2 indicate the degree of linkage disequilibrium for each associated SNP. [D] An additive effect of both *cis*- and *trans*-pQTL variants of *CCL25* is observed on CCL25 protein levels. SNP, single nucleotide polymorphism.

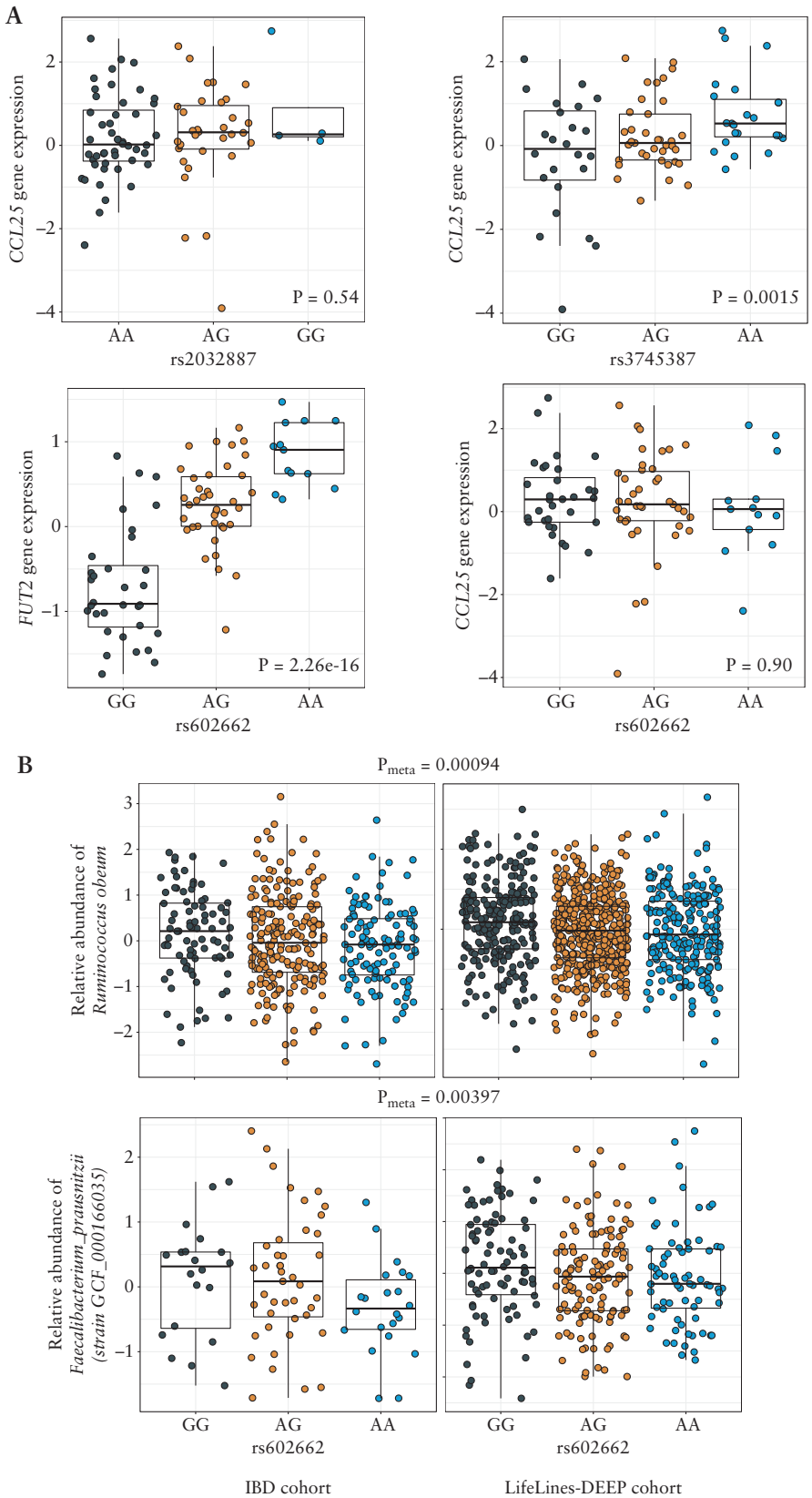

Figure 7. Effects of eQTL- and mbQTL-variants of *CCL25* and *FUT2*. [A] The *cis*-pQTL variant rs2032887 for CCL25 did not significantly affect *CCL25* gene expression, whereas the *cis*-pQTL variant rs3745387 appeared as *cis*-eQTL-variant for *CCL25* gene expression levels, and the *trans*-pQTL variant rs602662 for CCL25 also appeared to be a *cis*-eQTL variant on *FUT2* gene expression levels. However, no *trans*-eQTL effect of the *FUT2* rs602662 variant on *CCL25* could be observed. [B] The AA allele carriers of variant rs602662 show lower relative abundance of *Ruminococcus obeum* and a strain of *Faecalibacterium prausnitzii*, which shows mbQTL effects on gut microbiota.

intestinal lamina propria lymphocytes [LPLs], which may explain the negative association between its circulating levels and systemic inflammation.

A highlight of the present study included the overlap with multiple data layers, such as bulk RNA sequencing data of intestinal biopsies and faecal metagenomics data, which permitted integrative assessment of the discovered associations. However, several limitations also have to be considered. For example, we were only able to explain a limited amount of variation in protein levels, because some information was missing that potentially affects plasma protein levels, e.g., information on dietary intake, lifestyle habits, gut microbiota composition, and other environmental factors. Additionally, this study had to rely on clinical and serological assessment of disease activity, as data of faecal calprotectin levels or endoscopic investigations were not sufficiently recorded at time of sampling, and thus our results relating to disease activity necessitate cautious interpretation. Furthermore, the majority of our cohort [~70%] was in disease remission, and therefore the observed variation in correlations and fold changes may be limited. Still, distinct plasma protein signatures could be identified in case-control analyses accounting for disease activity status. Further, our study design did not permit the assessment of potential causality between genetic variants and protein levels, but rather associations between these data entities. In this respect, functional studies are required to gain more in-depth knowledge and to provide biological explanation of our observations, as well as independent replication to validate the findings. Finally, protein levels are expressed in relative units derived from the PEA technology, which precludes the comparison of absolute concentrations between different proteins, and limits comparability with other studies that conducted proteomic profiling by using more traditional methods such as enzyme-linked immunosorbent assays [ELISAs]. Nevertheless, PEA technology has the great advantage of highly sensitive and high-throughput analysis without significant loss of specificity. Each of the oligonucleotide antibody pairs consists of a unique DNA sequence only allowing hybridisation to each other and thus preventing antibody cross-reactivity. It is characterised by relatively high precision compared with other multiplex proteomics techniques with typical intra- and inter-coefficients of variation [CV] values of 8% and 11%, respectively.[47]

Our results demonstrate a complex and rich interplay between genotype and the human plasma proteome, with significant involvement of demographic and clinical traits in a large cohort of patients with IBD. This study highlights many associations between genotype, phenotype, and circulating proteins that are known to modulate a variety of inflammatory pathways in the context of IBD. These, in turn, may provide a foundation for future mechanistic research that is required to disentangle the relevant pathophysiological pathways. Furthermore, in future clinical studies, identification of plasma-based pQTLs may help to stratify patients according to their response to treatment as they support drug selection and validation, as well as drug safety and repurposing. In addition, co-localisation of identified pQTLs with known IBD genetic susceptibility loci may expose novel IBD-associated molecular pathways. In the light of personalised medicine, combining both genomics and proteomics may provide further molecular understanding in order to improve diagnostics and therapeutics in patients with IBD.

Data Availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request. The data for the Groningen IBD cohort can be requested with the accession number EGAS00001002702. All code used for analyses in this study can be found at the following link: [https://github.com/WeersmaLabIBD/Proteomics_project].

Funding

This work was supported by the collaborative T cell-driven Immune Mediated Inflammatory Diseases [TIMID] project [LSHM18057-SGF] financed by the PPP allowance made available by Top Sector Life Sciences & Health to Samenwerkende Gezondheidsfondsen [SGF] to stimulate public-private partnerships and co-financing by health foundations that are part of the SGF. This work was further supported by Takeda [no grant number, to LMS], a VENI grant from Nederlandse Organisatie voor Wetenschappelijk Onderzoek [NWO] [grant no. 194.006, to DVZ], and the Junior Scientific Masterclass [JSM] of the University of Groningen [grant no. 17–57, to ARB].

Conflict of Interest

GD received an unrestricted research grant from Takeda, and received speaker fees from Pfizer and Janssen Pharmaceuticals. RKW acted as consultant for Takeda, received unrestricted research grants from Takeda, Johnson & Johnson, Tramedico, and Ferring, and received speaker fees from MSD, Abbvie, and Janssen Pharmaceuticals.

Author Contributions

RKW designed the study. ARB, SH, LMS, DVZ, AVV, YL, MDV, LAvB, BBF, MC, AM, MJTR, JIPvH, LABJ, JNS gathered and prepared the data. ARB, SH, and LMS analysed the data. ARB, SH, and LMS wrote the manuscript. AVV and DVZ provided statistical advice. All authors were involved in either sample collection, data generation, or critically reviewing the manuscript.

Acknowledgements

The authors would like to thank all participants of the 1000IBD cohort.

Supplementary Data

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

References

1. Ananthakrishnan AN. Epidemiology and risk factors for IBD. *Nat Rev Gastroenterol Hepatol* 2015;**12**:205–17.
2. Abraham C, Cho JH. Inflammatory bowel disease. *N Engl J Med* 2009;**361**:2066–78.
3. Cosnes J, Gower-Rousseau C, Seksik P, Cortot A. Epidemiology and natural history of inflammatory bowel diseases. *Gastroenterology* 2011;**140**:1785–94.
4. Colombel JF, Narula N, Peyrin-Biroulet L. Management strategies to improve outcomes of patients with inflammatory bowel diseases. *Gastroenterology* 2017;**152**:351–61.e5.
5. Jones J, Loftus EV Jr, Panaccione R, *et al*. Relationships between disease activity and serum and fecal biomarkers in patients with Crohn's disease. *Clin Gastroenterol Hepatol* 2008;**6**:1218–24.
6. Bourgonje AR, von Martels JZH, Gabriëls RY, *et al*. A combined set of four serum inflammatory biomarkers reliably predicts endoscopic disease activity in inflammatory bowel disease. *Front Med (Lausanne)* 2019;**6**:251.
7. 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.
8. Jostins L, Ripke S, Weersma RK, *et al*.; International IBD Genetics Consortium [IIBDGC]. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;**491**:119–24.
9. Liu JZ, van Sommeren S, Huang H, *et al*.; International Multiple Sclerosis Genetics Consortium; International IBD Genetics Consortium. Association analyses identify 38 susceptibility loci for inflammatory bowel

disease and highlight shared genetic risk across populations. *Nat Genet* 2015;**47**:979–86.
10. Sun BB, Maranville JC, Peters JE, *et al*. Genomic atlas of the human plasma proteome. *Nature* 2018;**558**:73–9.
11. Zhernakova DV, Le TH, Kurilshikov A, *et al*.; LifeLines cohort study; BIOS consortium. Individual variations in cardiovascular-disease-related protein levels are driven by genetics and gut microbiome. *Nat Genet* 2018;**50**:1524–32.
12. Suhre K, McCarthy MI, Schwenk JM. Genetics meets proteomics: perspectives for large population-based studies. *Nat Rev Genet* 2021;**22**:19–37.
13. Folkersen L, Gustafsson S, Wang Q, *et al*. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat Metab* 2020;**2**:1135–48.
14. Imhann F, Van der Velde KJ, Barbieri R, *et al*. The 1000IBD project: multi-omics data of 1000 inflammatory bowel disease patients; data release 1. *BMC Gastroenterol* 2019;**19**:5.
15. Mourits VP, Arts RJW, Novakovic B, *et al*. The role of Toll-like receptor 10 in modulation of trained immunity. *Immunology* 2020;**159**:289–97.
16. Assarsson E, Lundberg M, Holmquist G, *et al*. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One* 2014;**9**:e95192.
17. Berkhout LC, l'Ami MJ, Ruwaard J, *et al*. Dynamics of circulating TNF during adalimumab treatment using a drug-tolerant TNF assay. *Sci Transl Med* 2019;**11**:eaat3356.
18. van Schie KA, Ooijevaar-de Heer P, Dijk L, Kruithof S, Wolbink G, Rispens T. Therapeutic TNF inhibitors can differentially stabilize trimeric TNF by inhibiting monomer exchange. *Sci Rep* 2016;**6**:32747.
19. Hu S, Vich Vila A, Gacesa R, *et al*. Whole exome sequencing analyses reveal gene-microbiota interactions in the context of IBD. *Gut* 2021;**70**:285–96.
20. Hu S, Uniken Venema WT, Westra HJ, *et al*. Inflammation status modulates the effect of host genetic variation on intestinal gene expression in inflammatory bowel disease. *Nat Commun* 2021;**12**:1122.
21. Di Narzo AF, Telesco SE, Brodmerkel C, *et al*. High-throughput characterization of blood serum proteomics of IBD patients with respect to aging and genetic factors. *PLoS Genet* 2017;**13**:e1006565.
22. Enroth S, Johansson A, Enroth SB, Gyllensten U. Strong effects of genetic and lifestyle factors on biomarker variation and use of personalized cutoffs. *Nat Commun* 2014;**5**:4684.
23. Westra HJ, Peters MJ, Esko T, *et al*. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013;**45**:1238–43.
24. Opstad TB, Brusletto BS, Arnesen H, Pettersen AÅ, Seljeflot I. Cigarette smoking represses expression of cytokine IL-12 and its regulator miR-21 - An observational study in patients with coronary artery disease. *Immunobiology* 2017;**222**:169–75.
25. Opstad TB, Arnesen H, Pettersen AÅ, Seljeflot I. Combined elevated levels of the proinflammatory cytokines IL-18 and IL-12 are associated with clinical events in patients with coronary artery disease: an observational study. *Metab Syndr Relat Disord* 2016;**14**:242–8.
26. Glas J, Seiderer J, Wagner J, *et al*. Analysis of IL12B gene variants in inflammatory bowel disease. *PLoS One* 2012;**7**:e34349.
27. van der Heide F, Nolte IM, Kleibeuker JH, Wijmenga C, Dijkstra G, Weersma RK. Differences in genetic background between active smokers, passive smokers, and non-smokers with Crohn's disease. *Am J Gastroenterol* 2010;**105**:1165–72.
28. Gadaleta RM, Moschetta A. Metabolic messengers: fibroblast growth factor 15/19. *Nat Metab* 2019;**1**:588–94.
29. Walters JR, Tasleem AM, Omer OS, Brydon WG, Dew T, le Roux CW. A new mechanism for bile acid diarrhea: defective feedback inhibition of bile acid biosynthesis. *Clin Gastroenterol Hepatol* 2009;**7**:1189–94.
30. Lenicek M, Duricova D, Komarek V, *et al*. Bile acid malabsorption in inflammatory bowel disease: assessment by serum markers. *Inflamm Bowel Dis* 2011;**17**:1322–7.
31. Nolan JD, Johnston IM, Pattni SS, Dew T, Orchard TR, Walters JR. Diarrhoea in Crohn's disease: investigating the role of the ileal hormone fibroblast growth factor 19. *J Crohns Colitis* 2015;**9**:125–31.
32. Andersson E, Bergemalm D, Kruse R, *et al*. Subphenotypes of inflammatory bowel disease are characterized by specific serum protein profiles. *PLoS One* 2017;**12**:e0186142.
33. Pavlidis P, Powell N, Vincent RP, Ehrlich D, Bjarnason I, Hayee B. Systematic review: bile acids and intestinal inflammation-luminal aggressors or regulators of mucosal defence? *Aliment Pharmacol Ther* 2015;**42**:802–17.
34. Vítek L. Bile acid malabsorption in inflammatory bowel disease. *Inflamm Bowel Dis* 2015;**21**:476–83.
35. Gadaleta RM, van Erpecum KJ, Oldenburg B, *et al*. Farnesoid X receptor activation inhibits inflammation and preserves the intestinal barrier in inflammatory bowel disease. *Gut* 2011;**60**:463–72.
36. van Schaik FD, Gadaleta RM, Schaap FG, *et al*. Pharmacological activation of the bile acid nuclear farnesoid X receptor is feasible in patients with quiescent Crohn's colitis. *PLoS One* 2012;**7**:e49706.
37. Gadaleta RM, Garcia-Irigoyen O, Cariello M, *et al*. Fibroblast Growth Factor 19 modulates intestinal microbiota and inflammation in presence of Farnesoid X Receptor. *EBioMedicine* 2020;**54**:102719.
38. Vicari AP, Figueroa DJ, Hedrick JA, *et al*. TECK: a novel CC chemokine specifically expressed by thymic dendritic cells and potentially involved in T cell development. *Immunity* 1997;**7**:291–301.
39. Papadakis KA, Prehn J, Nelson V, *et al*. The role of thymus-expressed chemokine and its receptor CCR9 on lymphocytes in the regional specialization of the mucosal immune system. *J Immunol* 2000;**165**:5069–76.
40. Eksteen B, Grant AJ, Miles A, *et al*. Hepatic endothelial CCL25 mediates the recruitment of CCR9+ gut-homing lymphocytes to the liver in primary sclerosing cholangitis. *J Exp Med* 2004;**200**:1511–7.
41. Engel DR, Koscielny A, Wehner S, *et al*. T helper type 1 memory cells disseminate postoperative ileus over the entire intestinal tract. *Nat Med* 2010;**16**:1407–13.
42. Papadakis KA, Prehn J, Moreno ST, *et al*. CCR9-positive lymphocytes and thymus-expressed chemokine distinguish small bowel from colonic Crohn's disease. *Gastroenterology* 2001;**121**:246–54.
43. Trivedi PJ, Bruns T, Ward S, *et al*. Intestinal CCL25 expression is increased in colitis and correlates with inflammatory activity. *J Autoimmun* 2016;**68**:98–104.
44. Wendt E, Keshav S. CCR9 antagonism: potential in the treatment of inflammatory bowel disease. *Clin Exp Gastroenterol* 2015;**8**:119–30.
45. Bourgonje AR, von Martels JZH, de Vos P, Faber KN, Dijkstra G. Increased fecal calprotectin levels in Crohn's disease correlate with elevated serum Th1- and Th17-associated cytokines. *PLoS One* 2018;**13**:e0193202.
46. Dahan S, Rabinowitz KM, Martin AP, Berin MC, Unkeless JC, Mayer L. Notch-1 signaling regulates intestinal epithelial barrier function, through interaction with CD4+ T cells, in mice and humans. *Gastroenterology* 2011;**140**:550–9.
47. Lundberg M, Eriksson A, Tran B, Assarsson E, Fredriksson S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res* 2011;**39**:e102.