CHAPTER 11

THE EFFECT OF PHENOTYPE AND GENOTYPE ON THE PLASMA PROTEOME IN PATIENTS WITH INFLAMMATORY BOWEL DISEASE

Arno R. Bourgonje1,*, Shixian Hu1,2,*, Lieke M. Spekhorst1,*, Daria V. Zhernakova2,3, Arnau Vich Vila1,2, Yanni Li1,2, Michiel D. Voskuil1,2, Lisette A. van Berkel4, Brenda Bley Folly5, Mohammed Charrouit6, Ahmed Mahfouz6,7, Marcel J. T. Reinders6,8, Julia I. P. van Heck6, Leo A. B. Joosten6, Marijn C. Visschedijk1, Hendrik M. van Dullemen1, Klaas Nico Faber1, Janneke N. Samsom1, Eleonora A. M. Festen1,2, Gerard Dijkstra1, Rinse K. Weersma1

1Department of Gastroenterology and Hepatology, University of Groningen, University Medical Centre Groningen, Groningen, the Netherlands
2Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, the Netherlands
3Laboratory of Genomic Diversity, Centre for Computer Technologies, ITMO University, St. Petersburg, Russia
4Department of Paediatrics, Division of Gastroenterology, Erasmus University Medical Centre, Rotterdam, the Netherlands
5Delft Bioinformatics Lab, Delft University of Technology, Delft, the Netherlands
6Leiden Computational Biology Centre, Leiden University Medical Centre, Leiden, the Netherlands
7Department of Human Genetics, Leiden University Medical Centre, Leiden, the Netherlands
8Department of Internal Medicine and Radboud Institute of Molecular Life Sciences, Radboud University Medical Centre, Nijmegen, the Netherlands
*These authors contributed equally.
Abstract

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

Methods
Ninety-two (92) inflammation-related proteins were quantified in plasma of 1,028 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 fecal metagenomic data were used for complementary analyses.

Results
Thirty-two (32) proteins were differentially abundant between IBD and healthy individuals, of which 22 proteins independent of active inflammation. Sixty-nine (69) proteins were associated with 15 demographic and clinical factors. Fibroblast growth factor-19 levels were decreased in patients with CD 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 fecal 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.

Keywords
inflammatory bowel disease; genetics; proteomics.
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. IBD is a heterogeneous disease, impeding the prediction of disease course and therapeutic response. Consequently, clinicians are being challenged in disease management as symptomatology is often non-specific and surrogate disease biomarkers are lacking.

Over the past decades, there has been made a tremendous effort to unravel the role of genetics in IBD susceptibility. To date, more than 200 genomic loci have been associated with IBD disease risk. 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. While protein profiling has been well studied in healthy human populations, it has not yet been systematically performed in patients with IBD. High-throughput proteome research techniques have evolved only recently, and large studies so far have been primarily focused on metabolic markers and cardiovascular disease. 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. 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 ninety-two (92) inflammation-related plasma proteins, which was performed in 1,028 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 fecal 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, assist in drug selection, safety and repurposing, and co-localizing identified pQTLs with known IBD risk loci may expose novel molecular pathways relevant to IBD.
Materials and Methods

Study cohort: 1000IBD
This study was conducted at the University Medical Centre Groningen (UMCG), Groningen, the Netherlands. Patients were included based on their participation in the 1000IBD project. Within the 1000IBD project, detailed phenotypic information and multi-omics profiles have been collected for over 1,000 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 biological 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 last visit to the outpatient clinic. Serum 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 Centre, Nijmegen, the Netherlands) to comparatively identify disease-associated proteins, which were included in the time period from April 2017 to June 2018. 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.

Figure 1(A-D) | 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. (A) Proteomics data were generated for both patients with IBD (n=1,068) 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 normalized 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 1,028 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 to the observed findings from the pQTL analysis by integrating RNA sequencing data (n=280 IBD) and fecal microbiome data (n=435 IBD) and studying the effect of pQTL variants on intestinal gene expressions and microbial abundances, respectively.
Collagen biomarkers and Crohn's disease behavior

**a Data filtering**

UMCG IBD  
\( n=1068 \)

Radboud UMC control  
\( n=148 \)

Proximity extension assay (PEA)

92 plasma inflammation proteins

QC steps

- NPX deviation >0.3
- Detection rate <0.1

567 CD  
461 UC  
+ 83 proteins

148 controls

**b Phenotypic determinates of proteins**

**IBD-control comparison**

| Protein level | CD  | UC  | Control |
|---------------|-----|-----|---------|
| IBD           |     |     |         |
\( n=1028 \)

| Protein level | CD  | UC  | Control |
|---------------|-----|-----|---------|
| Controls      |     |     |         |
\( n=148 \)

Variation explained by 17 phenotypes

**c Genetic determinates of proteins**

Whole exome sequencing  
Global screen array  
~8 million genomic variants

83 plasma proteins

**d pQTL effects on intestinal gene expressions and fecal microbiome**

**pQTL effects on intestinal gene expressions**

IBD biopsies  
\( n=280 \)

Gene level  
pQTL Genotype

**pQTL effects on gut microbiota**

IBD fecal  
\( n=435 \)

Lifelines-DEEP fecal  
\( n=920 \)

Microbiota abundance  
pQTL Genotype
Proximity extension assay (PEA) 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. To reduce technical variation between plates, plasma samples were randomized on different plates using a randomization algorithm, including randomization 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, hybridization 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

Prior to analysis, an inter-plate intensity normalization procedure was performed using the plate median as normalization factor. Data were normalized on a log2-scale where values were derived from inverted Ct-values of real-time qPCR and expressed as normalized 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 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), since including them did not change the obtained results. In the present study, proteomic profiling of 83 proteins was available for 1,028 patients with IBD (567 CD and 461 UC) and 148 healthy controls.

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 was 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 using PLINK v.1.9. Only genetic data from patients clustering with individuals of European descent were included in the analyses.

GSA data was 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), 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 $R^2>0.4$. GSA genotype data was combined with WES data using PLINK 1.9. Variants with a 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).

**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. 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 was available from 280 intestinal biopsies of 165 patients with IBD (ileum biopsies, n=89, colon biopsies, n=191).

**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 cohort (n=435), and a population-based cohort Lifelines-DEEP (n=920). 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.
Statistical analysis

**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-squared test or Fisher’s exact test if \( n \) of observations were <10. \( P \)-values <0.05 were considered significant.

**Associations between demographic and clinical variables and plasma proteins**
All analyses were performed in R (v.3.6.3). Principle component analysis (PCA) was used for dimensionality reduction for all 83 plasma proteins in 1,028 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 generalized linear model containing seventeen 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 categorized as having active disease. Individual protein variation explained by each factor was further assessed by an analysis of variance (ANOVA) on the generalized 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 behavior; UC: disease extension) as well as associations between proteins and the HBI and SCCAI scores. For all analyses, a FDR <0.05 was considered as statistically significant.

**Protein quantitative trait loci (pQTL) mapping**
\( \text{Cis-} \)pQTL variants were defined as genomic variants located within ± 1 Mb of the region of each protein-coding gene center, whereas \( \text{trans-} \)pQTL variants were defined as variants located ± 1 Mb outside of the region of each protein-coding gene center. For both \( \text{cis-} \)pQTL and \( \text{trans-} \)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 \( \text{cis-} \)pQTLs, in subsequent rounds, we regressed out the top statistically significant pQTLs from the last round until no independent signal was present anymore. \( \text{Trans-} \)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 x 10^{-7} (0.05/353,612). For trans-pQTL analysis, the significance threshold was 1.01 x 10^{-11} (0.05/496,624,608).

**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.
Results

Cohort description
Demographic and clinical characteristics of the study population (IBD: $n=1,028$; healthy controls: $n=148$) are presented in Table 1. In total, 567 patients had a diagnosis of CD, while 461 patients were diagnosed with UC. More females had CD, while more men had UC ($P<0.01$), and the proportion of females was higher in patients with IBD compared to healthy individuals ($P<0.01$). Median age at date of plasma sampling was 38 years in patients with CD compared to 43 years in patients with UC ($P<0.01$), whereas healthy controls were younger ($P<0.01$). Patients with CD smoked more often compared to patients with UC ($P<0.01$). Concerning biological use, patients with CD used more anti-TNF-α (e.g., infliximab, adalimumab, certolizumab and golimumab) compared to patients with UC ($P<0.01$). Patients with CD also used thiopurines and methotrexate more often (both $P<0.01$), while 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 ileocecal resection compared to patients with UC ($P<0.01$), while there was no significant difference for (partial) colon resections ($P=0.36$).
Table 1 | Descriptive statistics of the study population (n = 1,028 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%) | < 0.01 |
| Male                                          | 206 (36%)  | 223 (48%)  | 73 (51%)  | -       |
| 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%)  | -         | < 0.01 |
| Yes                                           | 164 (30%)  | 47 (11%)   | -         | -       |
| No                                            | 379 (70%)  | 390 (89%)  | -         | -       |
| Montreal classification                       |     |     |     | < 0.01 |
| Montreal Age (A)                               | 565 (100%) | 460 (100%) | -         | < 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 Behavior (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%) | -         | < 0.01 |
| 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    |
Distinct plasma protein signatures between (quiescent) IBD and healthy individuals

Heterogeneity of plasma protein levels was visualized 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). Fifty-nine (59) proteins were different between healthy controls and IBD (IBD, CD or UC, FDR<0.05, Supplementary Table S2). 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<0.1, Figure 2B, Supplementary Table S3, Supplementary Figures S1-3). 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 (AXIN1), and STAM-binding protein (STAMPB) (Figure 2C). Ten (10) proteins were increased in both CD and UC compared to 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 to 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, Supplementary Table S4). Importantly,
The top significant plasma proteins were fairly similar compared to the previous analysis. Next, we explored plasma proteins that distinguished quiescent CD from quiescent UC. Here, 9 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), tumour necrosis factor superfamily member 9 (TNFRSF9), C-X-C motif chemokine 10 (CXCL10) and 9 (CXCL9)) decreased in CD compared to UC (Supplementary Figure S5, Supplementary Tables S5-6).
Chapter 11

A

B

C

CD vs. control, FDR = 8.33e-08
UC vs. control, FDR = 9.52e-06
IBD vs. control, FDR = 1.10e-07

CD vs. control, FDR = 5.24e-07
UC vs. control, FDR = 8.10e-04
IBD vs. control, FDR = 2.19e-06

CD vs. control, FDR = 3.15e-10
UC vs. control, FDR = 3.46e-03
IBD vs. control, FDR = 2.63e-06

CD vs. control, FDR = 1.68e-04
UC vs. control, FDR = 1.04e-05
IBD vs. control, FDR = 6.50e-06

CD vs. control, FDR = 9.43e-07
UC vs. control, FDR = 1.55e-03
IBD vs. control, FDR = 6.64e-06

CD vs. control, FDR = 1.02e-05
UC vs. control, FDR = 2.42e-03
IBD vs. control, FDR = 2.98e-05

DNER
SIRT2
FGF-19
OSM
AXIN1
STAMBP
4E-BP1
IL7
CXCL1
CASP-8
HGF
LAP TGF–beta-1
MCP-4
IL6
TNFSF14
AXIN1
SIRT2
OSM
MMP-1
STAMBP
MMP-10
IL–18R1
IL–17A
CCL11
IL8
IL18
CDCP1
CXCL11
uPA
FGF–19
TNFRSF9
EN–RAGE
CCL25
CD6
LIF–R
SCF
TWEAK
TNFB
TRANCE
DNER
Figure 2 (A-C) | 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 for healthy controls. Both CD and UC were significantly different from healthy controls 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.

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 S6-7, Supplementary Table S7). Mycophenolate mofetil and methotrexate use were the only factors that were not significantly associated to 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 S6-7). In subsequent analyses, we corrected for protein-specific statistically significant demographic or clinical factors (Supplementary Table S8).
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).
**Plasma FGF-19 levels are decreased in patients with CD having ileal disease or a history of ileocolic resection**

In patients with CD, a history of ileocolic 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 to patients with solely colonic disease (Montreal L2) ($P=4.06\times10^{-11}$, Figure 4A). In addition, FGF-19 level was lower in patients with either stricturing (Montreal B2, $P=1.61\times10^{-7}$) or penetrating (Montreal B3, $P=6.69\times10^{-6}$) disease behaviour compared to non-stricturing, non-penetrating disease behaviour (Montreal B1) (Figure 4B) (FDR<0.05). These differences remained statistically significant when patients with a history of ileocolic 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).

**Figure 4 (A-B)** | Plasma FGF-19 levels are decreased in patients with ileal and stricturing and penetrating Crohn's disease. (A) FGF-19 level is significantly elevated in patients with CD having colonic disease involvement compared to patients with ileal disease involvement. (B) FGF-19 level is decreased in patients with CD having stricturing or penetrating disease behaviour.
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 to 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 to 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).

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; CD: 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).

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) (Figure 5A, Supplementary Table S10). A negative association was observed for DNER (r=-0.18), while 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).
Figure 5 (A-B) | 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 CD. (B) Five proteins significantly correlate with the Simple Clinical Colitis Activity Index (SCCAI) in patients with UC. Abbreviations: FDR, false discovery rate; r, correlation coefficient.

Associations between genetics and protein level

**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 S11-12), followed by a meta-analysis (IBD: n=919). In total, 1,655 cis-pQTLs (Supplementary Table S13), corresponding to 23 independent cis-pQTLs, were found to be statistically significantly associated with the level of 21 different proteins (Bonferroni-adjusted $P=1.41\times10^{-7}$) (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, while levels of the other 19 proteins were associated with one independent cis-pQTL variant. Ten (10) 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, $P=1.97\times10^{-49}$). 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. Four (4) independent cis-pQTL variants also appeared to have an eQTL effect, including variants of CCL25 (P=0.0015, ileum), CXCL5 (P=8.36x10^{-7}, colon), MMP-1 (P=1.68x10^{-4}, colon) and IL10-RB (P=0.0036, colon) (Supplementary Table S14). When checking the overlap with IBD GWAS signals, five (5) 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=5.86x10^{-22}, rs602662) was further identified, which is located in the fucosyltransferase 2 (FUT2) gene, a known IBD-associated risk locus (Supplementary Table S15).
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 |
|---------|-----------|-----|--------|---------------|--------|---------------|--------|--------------|
| CDBA    | rs3020726*| 2   | G      | 2.03x10^{-19} | 0.39   | 3.64x10^{-11} | 0.32   | 1.19x10^{-28}|
| IL-18R1 | rs11377261| 2   | A      | 1.55x10^{-22} | 0.44   | 3.67x10^{-20} | 0.47   | 1.77x10^{-40}|
| CDCP1   | rs62244470| 3   | T      | 1.32x10^{-4}  | -0.20  | 1.60x10^{-3}  | 0.16   | 1.52x10^{-8} |
| CXCL1   | rs3117604*| 4   | C      | 5.49x10^{-8}  | 0.25   | 1.15x10^{-3}  | 0.17   | 1.20x10^{-9} |
| CXCL5   | rs425535* | 4   | T      | 1.20x10^{-10} | 0.28   | 1.19x10^{-7}  | 0.26   | 1.49x10^{-16}|
| CXCL6   | rs16850073| 4   | T      | 3.18x10^{-17} | 0.38   | 1.50x10^{-10} | 0.33   | 6.59x10^{-26}|
| IL-12B  | rs4921484*| 5   | T      | 4.67x10^{-8}  | -0.27  | 2.70x10^{-7}  | -0.27  | 1.30x10^{-14}|
| TNFB    | rs2229092 | 6   | C      | 7.62x10^{-21} | -0.40  | 1.80x10^{-16} | -0.40  | 2.40x10^{-35}|
| VEGFA   | rs5921438*| 6   | A      | 8.63x10^{-23} | -0.44  | 7.62x10^{-14} | -0.38  | 1.07x10^{-34}|
| PD-L1   | rs3574625 | 9   | C      | 6.62x10^{-4}  | 0.16   | 6.03x10^{-6}  | 0.23   | 7.04x10^{-8}  |
| IL-15RA | rs4237402*| 10  | A      | 9.38x10^{-7}  | 0.23   | 1.96x10^{-3}  | 0.16   | 1.69x10^{-8}  |
| CD6     | rs11230563| 11  | T      | 4.84x10^{-17} | -0.37  | 1.74x10^{-12} | -0.34  | 1.17x10^{-27}|
| MMP-1   | rs10791596| 11  | T      | 1.35x10^{-4}  | -0.26  | 8.73x10^{-6}  | -0.22  | 1.08x10^{-12}|
| CCL23   | rs712048* | 17  | A      | 1.82x10^{-10} | -0.29  | 2.70x10^{-3}  | -0.16  | 1.75x10^{-11}|
| CCL3    | rs1719144*| 17  | A      | 1.18x10^{-7}  | 0.24   | 4.56x10^{-6}  | 0.23   | 4.87x10^{-12}|
| MCP-2   | rs1133763*| 17  | C      | 6.52x10^{-33} | -0.50  | 1.29x10^{-18} | -0.42  | 1.97x10^{-49}|
| CCL25   | rs2032887 | 19  | G      | 2.88x10^{-26} | 0.45   | 2.53x10^{-13} | 0.36   | 1.92x10^{-37}|
| CCL25   | rs3745387*| 19  | A      | 1.45x10^{-9}  | 0.25   | 3.95x10^{-5}  | 0.20   | 5.45x10^{-12}|
| ADA     | rs112545364| 20  | G      | 7.35x10^{-17} | 0.36   | 8.69x10^{-11} | 0.32   | 8.60x10^{-26}|
| CD40    | rs1883832 | 20  | T      | 1.04x10^{-7}  | -0.24  | 3.66x10^{-3}  | -0.14  | 5.30x10^{-9}  |
| CST5    | rs34269359| 20  | A      | 4.69x10^{-19} | 0.40   | 1.36x10^{-20} | 0.46   | 3.86x10^{-37}|
| CST5    | rs1799841 | 20  | G      | 4.76x10^{-7}  | 0.20   | 9.86x10^{-5}  | 0.19   | 3.65x10^{-9}  |
| IL-10RB | rs4455239 | 21  | C      | 3.33x10^{-4}  | 0.26   | 5.28x10^{-15} | 0.39   | 6.47x10^{-20}|

Abbreviations: Chr, chromosome; SNP, single nucleotide polymorphism; R, correlation coefficient; *Replicated cis-pQTL variants compared with Sun et al, *Nature* 2018.
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-C, boxplots). Subsequently, we observed an additive effect of both cis-pQTL variants and the trans-pQTL variant rs602662 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 to carriage of both cis-pQTL variants, though without the trans-pQTL variant ($P=2.55\times10^{-15}$). 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 ($P=2.26\times10^{-16}$), however, no trans-eQTL effect on CCL25 was observed (Figure 7A).

Figure 6 (A-D) | 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 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 CD for the rs602662 SNP of the FUT2 gene. 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.
Phenotype, genotype and plasma proteins in IBD

[Diagram A] Chr19

rs3745387  rs2032887  rs602662

TMM14  CCL25  FUT2

[Diagram B] Log10 values (p-value)

rs3745387  rs2032887  rs602662

P = 7.92e-10  P = 2.41e-39

CCL25

GG  AG  AA

rs3745387

[Diagram C] Log10 values (p-value)

rs602662

CCL25

GG  AG  AA

rs602662

rs602662

P = 5.86e-22

[Diagram D] CCL25 protein level

No pQTL variant  One cis-pQTL  trans-pQTL  Two cis-pQTL  Two cis+ trans-pQTL

P = 0.00098  P = 0.0057  P = 2.55e-15
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* (\(P_{\text{meta}}=0.00094\), Figure 7B, Supplementary Table S16), 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* (\(P_{\text{meta}}=0.00397\)), a well-known, commensal butyrate-producing bacterial species in the human gut.

Figure 7 (A-B) | 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, while 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.
Phenotype, genotype and plasma proteins in IBD

- **CCL25 gene expression**
  - rs602662
    - P = 0.54
    - IBD cohort
  - rs602662
    - P = 0.0015
    - LifeLines-DEEP cohort

- **FUT2 gene expression**
  - rs602662
    - P = 2.26e-16
    - IBD cohort
  - rs602662
    - P = 0.90
    - LifeLines-DEEP cohort

- **Relative abundance of Ruminococcus obeum**
  - rs602662
    - P-meta = 0.00094
    - IBD cohort
  - rs602662
    - P-meta = 0.00397
    - LifeLines-DEEP cohort

- **Relative abundance of Faecalibacterium prausnitzii (strain GCF_000166039)**
  - rs602662
    - IBD cohort
  - rs602662
    - LifeLines-DEEP cohort
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 to 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 to patients with CD having colonic disease. Secondly, 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. Thirdly, integration with intestinal mucosal gene expression data showed that the cis-pQTL variant rs3745387 appeared to be a cis-expression quantitative trait locus (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 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 to healthy individuals. In our study, active smoking was strongly associated with lower IL-12B levels in both patients with CD and UC. Previous studies have shown that serum IL-12B levels as well as IL-12B gene expression is decreased upon cigarette smoke exposure compared to 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 biological ustekinumab inhibits the IL-12/23 axis and thereby the activation of Th1/Th17-lymphocytes, one could hypothesize 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 to 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 to 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
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 diarrhea. Patients with CD who underwent ileocecal resection indeed show impaired release of FGF-19 to the circulation. Furthermore, intestinal inflammation disrupts the epithelial barrier integrity, which is accompanied by impairment of transepithelial transport mechanisms, including bile acid reabsorption. Therefore, not only ileocecal resection but also intestinal inflammation, which could lead to deterioration of bile acid homeostasis, is associated with lower plasma FGF-19 levels. 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 patients with CD with active disease or surgical history and concurrent disruption of bile acid metabolism.

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. CCL25 elicits recruitment of peripheral blood lymphocytes expressing the chemokine receptor CCR9 and the adhesion molecule α4β7. Through interaction with CCL25, CCR9-expressing lymphocytes are guided to the intraepithelial 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. 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 to colonic CD. Recent data suggest that colonic CCL25 expression and CCR9+ CD4+ effector T-cell recruitment may also be upregulated in patients with active colitis. 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. 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.

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. 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. 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. Notch-1 signalling is associated with enhanced mucosal barrier function, stimulated by 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 fecal 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 fecal 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, therefore the observed variation in correlations and fold changes may be limited. Still, distinct plasma protein signatures could be identified in case-control analyses accounted 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 to 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 to 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 high-sensitive and high-throughput analysis without significant loss of specificity. Each of the oligonucleotide antibody pairs consists of a unique DNA sequence only allowing hybridization to each other and thus preventing antibody cross-reactivity. It is characterized by relatively high precision compared to other multiplex proteomics techniques with typical intra- and inter-coefficients of variation (CV) values of 8% and 11%, respectively.

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-localization of identified pQTLs with known IBD genetic susceptibility loci may expose novel IBD-associated molecular pathways. In the light of personalized medicine, combining both genomics and proteomics may provide further molecular understanding to improve diagnostics and therapeutics in patients with IBD.

**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 LM.S.], a VENI grant from NWO [grant no. 194.006, to D.V.Z.], and the Junior Scientific Masterclass (JSM) of the University of Groningen [grant no. 17-57, to A.R.B.].

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

**Conflicts 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. All other authors have no conflicts of interest to declare.

**Authors’ 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.

**Data Availability Statement**
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.
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.
5. Jones J, Loftus EV Jr, Panaccione R, Chen LS, Peterson S, McConnell J, 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, Blokzijl T, Buist-Homan M, Heegsma J, 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, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet 2017;49:256-261.
8. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. 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, Ng SC, Alberts R, Takahashi A, et al. 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, Stacey D, Staley J, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature 2018;558:73-79.
11. Zhernakova DV, Le TH, Kurilshikov A, Atanasovska B, Bonder MJ, Sanna S, et al. Individual variations in cardiovascular-disease-related protein levels are driven by genetics and gut microbiome. Nat Genet 2018;50:1524-1532.
12. Suhre K, McCarthy MI, Schwenk JM. Genetics meets proteomics: perspectives for large population-based studies. Nat Rev Genet 2020; DOI:10.1038/s41576-020-0268-2.
13. Folkersen L, Gustafsson S, Wang Q, Hansen DH, Hedman AK, Schork A, et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. Nat Metab 2020;2:1135-1148.
14. Imhann F, van der Velde KJ, Barbieri R, Alberts R, Voskuil MD, Vich Vila A, 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, Matzaraki V, de Bree LCJ, Koeken VACM, et al. The role of Toll-like receptor 10 in modulation of trained immunity. Immunology 2020;159: 289-297.
16. Assarsson E, Lundberg M, Holmquist G, Björksten J, Thorsen SB, Ekman D, 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, Hart MH, Ooijevaar-de Heer P, Bloem K, 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, Colli J, Stevens C, Fu JM, et al. Whole exome sequencing analyses reveal gene-microbiota interactions in the context of IBD. Gut 2021;70:285-296.
20. Hu S, Uniken Venema WT, Westra HJ, Vich Vila A, Barbieri R, Voskuil MD, et al. Inflammation status modulates the effect of host genetic variation on intestinal gene expression in inflammatory bowel disease. Nat Commun 2021;12:1122.
Di Narzo AF, Telesco SE, Brodmerkel C, Argmann C, Peters LA, Li K, et al. High-Throughput Characterization of Blood Serum Proteomics of IBD Patients with Respect to Aging and Genetic Factors. PLoS Genet 2017;13:e1006565.

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.

Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 2013;45:1238-1243.

Opstad TB, Brusletto BS, Arnesen H, Pettersen AA, 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;222:169-175.

Opstad TB, Arnesen H, Pettersen AA, 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 Ddisord 2016;14:242-8.

Glas J, Seiderer J, Wagner J, Olszak T, Fries C, Tillack C, et al. Analysis of IL12B gene variants in inflammatory bowel disease. PLoS One 2012;7:e34349.

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.

Gadaleta RM, Moschetta A. Metabolic messengers: fibroblast growth factor 15/19. Nat Metab 2019;1:588-594.

Walters JRF, Tasleem AM, Orner 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.

Lenicek M, Duricova D, Komarek V, Gabrysova B, Lukas M, Smerhovsky Z, et al. Bile acid malabsorption in inflammatory bowel disease: assessment by serum markers. Inflamm Bowel Dis 2011;17:1322-7.

Nolan JD, Johnston IM, Pattini SS, Dew T, Orchard TR, Walters JRF. Diarrhea in Crohn's disease: investigating the role of the ileal hormone fibroblast growth factor 19. J Crohns Colitis 2015;9:125-31.

Andersson E, Bergemalm D, Kruse R, Neumann G, D'Amato M, Repsilber D, et al. Subphenotypes of inflammatory bowel disease are characterized by specific serum protein profiles. PLoS One 2017;12:e0186142.

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.

Vitek L. Bile acid malabsorption in inflammatory bowel disease. Inflamm Bowel Dis 2015;21:476-83.

Gadaleta RM, van Erpecum KJ, Oldenburg B, Willemsen EC, Renooij W, Murzilli S, et al. Farnesoid X receptor activation inhibits inflammation and preserves the intestinal barrier in inflammatory bowel disease. Gut 2011;60:463-72.

van Schaik FD, Gadaleta RM, Schaap FG, van Mil SWC, Siersema PD, Oldenburg B, 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.

Gadaleta RM, Garcia-Irigoyen O, Cariello M, Scialpi N, Peres C, Vetrano S, et al. Fibroblast Growth Factor 19 modulates intestinal microbiota and inflammation in presence of Farnesoid X Receptor. EBioMedicine 2020;54:102719.

Vicari AP, Figueroa DJ, Hedrick JA, Foster JS, Singh KP, Menon S, 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.
Papadakis KA, Prehn J, Nelson V, Cheng L, Binder SW, Ponath PD, 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.

Eksteen B, Grant AJ, Miles A, Curbishley SM, Lalor PF, Hübscher SG, et al. Hepatic endothelial CCL25 mediates the recruitment of CCR9⁺ gut-homing lymphocytes to the liver in primary sclerosing cholangitis. J Exp Med 2004;200:1511-1517.

Engel DR, Koscielny A, Wehner S, Maurer J, Schiwon M, Franken L, et al. T helper type 1 memory cells disseminate postoperative ileus over the entire intestinal tract. Nat Med 2010;16:1407-13.

Papadakis KA, Prehn J, Moreno ST, Cheng L, Kouroumalis EA, Deem R, et al. CCR9-positive lymphocytes and thymus-expressed chemokine distinguish small bowel from colonic Crohn's disease. Gastroenterology 2001;121:246-54.

Trivedi PJ, Bruns T, Ward S, Mai M, Schmidt C, Hirschfield GM, et al. Intestinal CCL25 expression is increased in colitis and correlates with inflammatory activity. J Autoimmun 2016;68:98-104.

Wendt E, Keshav S. CCR9 antagonism: potential in the treatment of Inflammatory Bowel Disease. Clin Exp Gastroenterol 2015;8:119-130.

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.

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.

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.
Supplementary Methods

**Expression quantitative trait loci (eQTL) and microbial quantitative trait loci (mbQTL) mapping**

Mapping of eQTL variants was performed for ileum and colon biopsies separately. A generalized linear mixed model was used to assess the effects of genetic variants on gene expression after correcting for the first 20 principle components (PCs) that captured the major confounding effects. To reduce the multiple testing burden, we only performed mbQTL mapping between genetic variant rs602662 located in the FUT2 gene and 316 microbial taxa, including 4 phyla, 58 genera and 51 different species. Relative abundances of microbial taxa were normalized through inverse rank transformation. Multivariate linear regression was used to adjust for age, sex, BMI, smoking, disease location, and medication usage (proton pump inhibitors [PPI], antibiotics and laxatives) and Spearman rank correlation tests were applied to determine the relationship between non-zero taxa values and the rs602662 genotype.
Supplementary Results

Age, sex and BMI are the main contributors to protein level variations
Age was the largest contributor to the explained variance on protein levels, as it was associated with 24 out of 83 proteins (28.9%) in patients with CD and with 19 of 83 proteins (22.9%) in patients with UC (FDR<0.05). Most of these proteins were positively associated with age (CD: 23 of 24 proteins, 95.8%; UC: 19 of 19 proteins, 100%), whereas only one protein (TRANCE) showed a negative association with age in patients with CD. Among CD patients, males had higher levels of ten proteins (12.0%) compared to females. Among UC patients, males had higher levels of 17 proteins (20.5%) and lower levels of one protein (CXCL5, 1.2%) compared to females. In patients with UC, BMI was positively associated with IL-6, FGF-21, CCL3 and IL-18R1, and negatively associated with GDNF, CCL11 and FGF-19 (FDR<0.05), while in patients with CD, none of the analysed proteins was associated with BMI.

Commonly used medications in IBD affect the levels of specific proteins
Thiopurine use in patients with CD was associated with increased levels of two proteins, namely FMS-like tyrosine kinase 3 ligand (Flt3L) and MCP-1, and decreased levels of 10 proteins (CD6, IL-12B, CD5, TNF-β, CD244, HGF, VEGF-A, CXCL10 and TGF-α, in order of statistical significance) (FDR<0.05). Fairly similar associations were observed in patients with UC using thiopurines, showing increased levels of Flt3L and OPG, and decreased levels of CD6, CXCL9 and IL-12B (FDR<0.05) (Supplementary Table S7).

Steroid use in patients with CD was associated with decreased levels of TRANCE, CCL19 and CST5 (FDR<0.05). Among patients with UC that used steroids, levels of CXCL11 and OSM were increased, whereas those of TRANCE were decreased (FDR<0.05).

In patients with CD, use of aminosalicylate (5-ASA) drugs was associated with a decreased level of the CCL28 protein, whereas it was not associated with any of the analysed proteins in patients with UC.

In patients with CD, TNF-α-antagonist use was associated with increased levels of IFN-γ, IL-17A and SLAMF1, and with decreased levels of CCL4, CXCL9, TNFRSF9, CCL3, MCP-1, CCL20 and CXCL10 (FDR<0.05). In patients with UC, use of TNF-α-antagonists was associated with decreased levels of CCL4 and CX3CL1 (FDR<0.05).

Among patients with CD, the use of calcineurin inhibitors was associated with increased levels of TGF-α and CX3CL1 (FDR<0.05), while in patients with UC, the use of calcineurin inhibitors was associated with increased levels of GDNF, CCL28, IL-17A, IL10RB, CX3CL1, PD-L1, CCL11, CD8A, SLAMF1 and β-NGF (FDR<0.05).
Oral contraceptive use in female patients influences levels of multiple proteins

Use of oral contraceptives was associated with increased levels of OPG, NT-3 and CCL23, and with decreased levels of uPA, TRAIL, SCF, LIF-R, TRANCE, CD6, CCL28, TWEAK, CCL11, GDNF and CX3CL1 in patients with CD (FDR<0.05). Among patients with UC, use of oral contraceptives was associated with increased levels of OPG and CSF-1, and decreased levels of TWEAK, TRAIL, TRANCE, uPA, MCP-4, SCF, CCL25, CD6, LIF-R, CCL4, FGF-19 and HGF (FDR<0.05).

Plasma storage time influences levels of some plasma proteins

Plasma storage time is an important factor to take into consideration as this could potentially affect the consistency of the results obtained. Plasma storage time was associated with 11 of 83 (13.3%) in patients with CD and with 9 (10.8%) of 83 proteins in patients with UC. In patients with CD, plasma storage time was associated with increased levels of 4E-BP1, SIRT2, STAMBP, AXIN1, CXCL5, TNFSF14, CD40, PD-L1, ADA, and CD244, whereas it was associated with a decreased level of CASP-8 (FDR<0.05). Among patients with UC, plasma storage time was associated with increased levels of 4E-BP1, SIRT2, AXIN1, MMP-1, STAMBP, CXCL5 and TNFSF14, while it was associated with decreased levels of CASP-8 and CX3CL1 (FDR<0.05).
Supplementary Discussion

In this study, thiopurine use was strongly associated with higher Flt3L levels in patients with IBD. Flt3L is a hematopoietic growth factor, stimulating proliferation and differentiation of early hematopoietic progenitor cells through acting on its receptor Flt3 (CD135). Higher levels of Flt3L associate with bone marrow aplasia, as is observed during chemotherapy for malignancy or preparation for bone marrow transplantation. Murine models demonstrate that thiopurines increase plasma Flt3L levels, which directly relate to the time of bone marrow aplasia. As the sensitivity of Flt3L detection is very high in our system, we would anticipate that the association of Flt3L with thiopurine use could reflect the activity of the drug in the bone marrow, without a relation to bone marrow toxicity.

\textit{FUT2} is a known candidate IBD-risk locus and is abundantly expressed in intestinal tissue. \textit{FUT2} encodes alpha-(1,2)-fucosyltransferase that regulates the secretion of ABO blood group antigens by the intestinal epithelium and secretory glands. Homozygosity for non-functional \textit{FUT2} alleles is associated with a “non-secretor status” as these individuals fail to express ABO blood group antigens, and this has been associated with CD in a large GWAS meta-analysis. In addition, “non-secretors” of \textit{FUT2} show alterations in their gut microbiota, including a decrease in microbial diversity, changes in several microbial taxa, and altered microbial energy metabolism. A recent study demonstrated that \textit{FUT2} non-secretors exhibit lower fecal levels of the genus \textit{Blautia} which is consistent with our findings on the rs602662 non-secretor \textit{FUT2} variant. Altogether, \textit{FUT2} non-secretor status is associated with several host-microbe interactions in which fucosylated glycans are involved and these may be driven by the genetic associations between \textit{FUT2} variants and CD. The predominant \textit{FUT2} non-secretor polymorphism (present in approximately 20% of Caucasians) constitutes the W143X allele (rs601338), which is strongly associated with CD, and is observed to be in strong linkage disequilibrium with other \textit{FUT2}-inactivating variants, including our reported trans-pQTL missense variant of \textit{FUT2} (S258G, rs602662). Considering the above, one may hypothesize that activation of mucosal T-lymphocytes is shaped by \textit{FUT2}-mediated host-microbe interactions, as \textit{FUT2} is associated with an increased susceptibility to IBD. However, it remains unclear what interactions could determine this putative disease mechanism.

In our study, we had detailed demographic and clinical information available for all patients, enabling proper covariate adjustment. For instance, it is well-established that factors like age, sex and sample storage time considerably affect plasma protein levels, as we confirmed in our study as well. A previous population-based cohort study that investigated the effect of phenotype and genotype on the plasma proteome demonstrated that age and body weight had strong influences on a broad range of proteins, which is in accordance with our findings. Similarly, medication use has been identified as an important clinical factor that should be taken into account when using protein biomarkers for diagnostic purposes or risk stratification. Notably, the same study identified a strong negative association between smoking and circulating IL-12 levels.
References

1. Ramos MI, Tak PP, Lebre MC. Fms-like tyrosine kinase 3 ligand-dependent dendritic cells in autoimmune inflammation. Autoimmun Rev 2014;13:117-24.
2. Molyneux G, Gibson FM, Chen CM, Marway HK, McKeag S, Mifsud CVJ, et al. The haemotoxicity of azathioprine in repeat dose studies in the female CD-1 mouse. Int J Exp Pathol 2008;89:138-58.
3. Kelly RJ, Rouquier S, Giorgi D, Lennon GG, Lowe JB. Sequence and expression of a candidate for the human secretor blood group alpha(1,2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. J Biol Chem 1995;270:4640-9.
4. McGovern DP, Jones MR, Taylor KD, Marcianente K, Yan X, Dubinsky M, et al. Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn’s disease. Hum Mol Genet 2010;19:1756-65.
5. Tong M, McHardy I, Ruegger P, Goudarzi M, Kashyap PC, Haritunians T, et al. Reprograming of gut microbiome energy metabolism by the FUT2 Crohn’s disease risk polymorphism. ISME J 2014;8:2193-2206.
6. Rausch P, Rehman A, Künzel S, Häslter R, Ott SJ, Schreiber S, et al. Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. Proc Natl Acad Sci U S A 2011;108:19030-5.
7. Wacklin P, Tuimala J, Nikkilä, Tims S, Mäkivuokko H, Alakulppi N, et al. Faecal microbiota composition in adults is associated with the FUT2 gene determining the secretor status. PLoS One 2014;9:e94863.
8. Zhernakova DV, Le TH, Kurihiskov A, Atanassovska B, Bonder MJ, Sanna S, et al. Individual variations in cardiovascular-disease-related protein levels are driven by genetics and gut microbiome. Nat Genet 2018;50:1524-1532.
9. Maroni L, van de Graaf SFJ, Hohenester SD, Oude Elferink RPJ, Beuers U. Fucosyltransferase 2: a genetic risk factor for primary sclerosing cholangitis and Crohn’s disease—a comprehensive review. Clin Rev Allerg Immunol 2015;48:182-91.
10. Li B, Selmi C, Tang R, Gershwin ME, Ma X. The microbiome and autoimmunity: a paradigm from the gut-liver axis. Cell Mol Immunol 2018;15:595-609.
11. Tanaka T, Biancotto A, Moaddel R, Moore AZ, Gonzalez-Freire M, Aon MA, et al. Plasma proteomic signature of age in healthy humans. Aging Cell 2018;17:e12799.
12. Romanov N, Kuhn M, Aebersold R, Ori A, Beck M, Bork P. Disentangling Genetic and Environmental Effects on the Proteotypes of Individuals. Cell 2019;177:1308-1318.e10.
13. Enroth S, Hallmans G, Grankvist K, Gyllensten U. Effects of Long-Term Storage Time and Original Sampling Month on Biobank Plasma Protein Concentrations. EBioMedicine 2016;12:309-314.
14. 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.
Supplementary Data

All supplementary tables as referred to in this chapter can be found at ECCO-JCC online: https://academic.oup.com/ecco-jcc/article/16/3/414/6365883.

Supplementary Tables - Index

Supplementary Table S1. Names, abbreviations, UniProt IDs and detection rates of all 92 biomarkers measured using the Olink® Inflammation panel.

Supplementary Table S2. Differences in plasma protein levels between Crohn’s disease (CD) vs. controls, ulcerative colitis (UC) vs. controls, inflammatory bowel disease (IBD) vs. controls (without age, sex and BMI correction)

Supplementary Table S3. Differences in plasma protein levels between Crohn’s disease (CD) vs. controls, ulcerative colitis (UC) vs. controls, inflammatory bowel disease (IBD) vs. controls (with age, sex and BMI correction)

Supplementary Table S4. Differences in plasma protein levels between Crohn’s disease (CD) in clinical remission vs. controls, ulcerative colitis (UC) in clinical remission vs. controls, and inflammatory bowel disease (IBD) in biochemical remission vs. controls (with age, sex and BMI correction)

Supplementary Table S5. Differences in plasma protein levels between Crohn’s disease (CD) and ulcerative colitis (UC) (without age, sex and BMI correction)

Supplementary Table S6. Differences in plasma protein levels between Crohn’s disease (CD) and ulcerative colitis (UC) (with age, sex and BMI correction)

Supplementary Table S7. Associations between phenotypic factors and protein level in a multivariate general linear model.

Supplementary Table S8. Factors corrected for each individual plasma protein in CD and UC separately.

Supplementary Table S9. Associations between phenotypic factors and protein level in a multivariate general linear model stratified by disease activity status at time of sampling.

Supplementary Table S10. Associations between clinical disease activity scores and plasma protein levels.

Supplementary Table S11. Local independent (cis-)pQTL mapping in patients with CD. 420

Supplementary Table S12. Local independent (cis-)pQTL mapping in patients with UC.
Supplementary Table S13. Summary statistics of cis-pQTL mapping in the full IBD cohort.

Supplementary Table S14. Results from expression quantitative trait loci (eQTL) analysis of the detected independent cis-pQTL variants.

Supplementary Table S15. One independent trans-pQTL variant (rs602662) was observed for the CCL25 protein.

Supplementary Table S16. Microbial quantitative trait loci (mbQTL) analysis showing associations of the FUT2 rs602662 missense variant with the gut microbiota.
Supplementary Figures

Supplementary Figure S1 | Top differentially abundant plasma proteins between IBD and healthy individuals. Violin plots visualizing the top 12 most differentially abundant plasma protein levels between either IBD, CD, or UC and healthy controls (DNER, SIRT2, FGF-19, OSM, AXIN1, STAMPB, IL-8, CCL11, LAP TGF-β1, CD6, IL-6 and CXCL11). The y-axis indicates NPX values. Abbreviations: CD, Crohn’s disease; FDR, false discovery rate; IBD, inflammatory bowel disease; UC, ulcerative colitis.
**Supplementary Figure S2 | Top differentially abundant plasma proteins between IBD and healthy individuals.** Violin plots visualizing the most differentially abundant plasma protein levels (ranking 13 to 24) between either IBD, CD, or UC and healthy controls (TWEAK, CCL25, uPA, IL-17A, HGF, TRANCE, SCF, TNF-B, MMP-1, CASP-8, LIF-R and IL-7). The y-axis indicates NPX values. Abbreviations: CD, Crohn's disease; FDR, false discovery rate; IBD, inflammatory bowel disease; UC, ulcerative colitis.
Supplementary Figure S3 | Top differentially abundant plasma proteins between IBD and healthy individuals. Violin plots visualizing the most differentially abundant plasma protein levels (ranking 25 to 32) between either IBD, CD, or UC and healthy controls (MCP-4, CXCL1, CDCP1, MMP-10, 4E-BP1, IL-18, IL-18R1 and TNFRSF9). The y-axis indicates NPX values. Abbreviations: CD, Crohn’s disease; FDR, false discovery rate; IBD, inflammatory bowel disease; UC, ulcerative colitis.
Supplementary Figure S4 (A-B) | Patients with quiescent IBD show distinct plasma protein signatures compared to healthy individuals. (A) Principal component analysis (PCA) plots demonstrating the first two principal components (PCs) for patients with CD in clinical remission (HBI<5), UC in clinical remission (SCCAI≤2), and IBD in biochemical remission (CRP<5 mg/L). (B) In total, 22 proteins were significantly different between either quiescent IBD, CD or UC and healthy controls. *FDR<0.1; **FDR<0.05. Red colouring indicates higher levels, whereas blue colouring indicates lower levels. Abbreviations: HBI, Harvey-Bradshaw Index; SCCAI, Simple Clinical Colitis Activity Index; CRP, C-reactive protein; CD, Crohn’s disease; UC, ulcerative colitis; IBD, inflammatory bowel disease; PC, principal component.
Supplementary Figure S5 (A-B) | Distinct plasma protein signatures between patients with CD and UC. (A) Volcano plot displaying 9 significantly differentially expressed proteins between CD and UC (in red). The x-axis indicates the log2-fold change in normalized protein expression (NPX) values; the y-axis indicates -log10-transformed \( P \)-values; the blue dashed line indicates the threshold for statistical significance corrected for multiple comparisons (\( n=83, \text{FDR}<0.05 \)); dark-red dots indicate significantly differentially expressed proteins while grey dots represent proteins that were not differentially expressed. (B) Top six most significantly different plasma protein levels between CD and UC (FGF-19, FGF-21, IFN-\( \gamma \), TNFRSF9, IL-10, MMP-10). The x-axis indicates CD (blue) and UC (orange). The y-axis indicates NPX values. Abbreviations: CD, Crohn's disease; UC, ulcerative colitis; FDR, false discovery rate.
Supplementary Figure S6 | Demographic and clinical factors strongly influence the plasma proteome in patients with CD and UC. Associations between demographic and clinical factors and protein levels for patients with CD (upper plot) and UC (lower plot) separately. Bar plots indicate the variance explained by significant demographic and clinical factors (FDR<0.05) for each of the 83 proteins.
Supplementary Figure S7 | Associations between plasma proteins and demographic and clinical factors show marked consistency between CD and UC. Visualization of intersections between (combinations of) demographic and clinical factors and measured proteins among Crohn’s disease (CD, upper panel) and ulcerative colitis (UC, lower panel) patients. The set size of each factor is represented by the left bar plot and represents the number of significant associations to plasma proteins. Associations between proteins and demographic or clinical factors are represented by the (interconnected) dots within the bottom plots, while their occurrence (intersection size) is displayed on the top bar plots.
Supplementary Figure S8 (A-B) | Plasma FGF-19 levels are decreased in patients with ileal and tricturing and penetrating Crohn's disease also in those without having a history of ileocecal resection. (A) FGF-19 level is significantly elevated in patients with CD having colonic disease involvement compared to patients with ileal disease involvement, all without having a history of ileocecal resection. (B) FGF-19 level is decreased in patients with CD having stricturing or penetrating disease behavior and without having a history of ileocecal resection.
Supplementary Figure S9 | Few proteins associate with extension of disease in patients with UC. Differences in protein levels between disease extension categories of the Montreal classification for UC.