Chapter 3

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


*shared first authors

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 CD patients with ileal disease or a history of ileocecal resection. Thirteen novel cis-pQTLs were identified and 10 replicated from previous studies. One trans-pQTL of the fucosyltransferase 2 (FUT2) gene (rs602662) and two independent cis-pQTLs of C-C motif chemokine 25 (CCL25) affected plasma CCL25 levels. Intestinal gene expression data revealed an overlapping cis-expression (e)QTL-variant (rs3745387) of the CCL25 gene. The FUT2 rs602662 trans-pQTL was associated with reduced abundances of 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.

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 proteomic 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. 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-seq 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.

Methods

Study cohort: 1000IBD

This study was conducted at the University Medical Center Groningen (UMCG), Groningen, the Netherlands. Patients were included based on their participation in the 1000IBD project.14 Within the 1000IBD project, detailed phenotypic information and multi-omics profiles have been collected for over 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.

Patients provided written informed consent (study was approved by the Institutional Review Board [IRB] of the UMCG [registered as no. 08/338]). In addition, we included Dutch non-IBD controls (n=148) from the 300BCG cohort (Radboud University Medical Center, Nijmegen, the Netherlands) to comparatively identify disease-associated proteins, which were included in the time period from April 2017 to June 2018.15 These healthy individuals also provided written informed consent (study approved by the IRB of the Arnhem-Nijmegem 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.

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.11,18 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], leukemia 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 (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), 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 by using PLINK v1.9. 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://glidmdna.org/globalscreening-array.html, GSA-MD). Genotypes were called using OptiCall (ref: optical.bitbucket.io), QC steps were performed using PLINK v1.9 (minor allele frequency (MAF)>5%, call rate <0.99, Hardy-Weinberg equilibrium test P<0.0001). Genotype data were phased using the Eagle algorithm and imputed to the Haplotype Reference Consortium reference panel using the Michigan Imputation Server (https://imputationserver.readthedocs.io/en/latest/pipeline/). After imputation, genetic variants were filtered for R2>0.4. GSA genotype data 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.20 In brief, 26 million paired-end 150-bp reads were generated per sample. The quality of the raw reads was checked using FastQC with default parameters (v0.11.7). The adaptors and low-quality reads were clipped using Trimmomatic (v0.36) with settings length <50 nucleotides, quality <25. Reads were aligned to the human genome (Homo_sapiens_assembly19.fasta) using STAR (v2.7.3). Reads sorting and mapping statistics were obtained using SAMtools (v0.1.19), sambamba (v0.7.0) and picard (v2.20.5). Gene expression was estimated through HTSeq (0.9.1) based on the annotation from GTEx v7 (gencode.v19.annotation.patched_contigs.gtf). After QC, data was available on 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).19 Sequencing reads that mapped to the human genome (version NCBI37) were removed using Kneaddata (v0.5.1). Microbiome taxa and predicted pathways profiling were performed using MetaPhlan (v2.6.0) and HUMAnN2 (v2.20.5). 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, 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

Cis-pQTL variants were defined as genomic variants located within ± 1 Mb of the region of each protein-coding gene center, whereas trans-pQTL variants were defined as variants located ± 1 Mb outside of the region of each protein-coding gene center. For both cis-pQTL and trans-pQTL mapping, we first performed the analysis in patients with CD and UC separately, followed by a weighted z-score meta-analysis within the full IBD cohort. A step-wise conditional analysis was used for pQTL identification.\(^{1,23}\) Briefly, Spearman rank correlation tests were performed to assess the effect of all genetic variants on protein level in the first round, adjusted for covariates using corrected estimates (see above). To identify all independent cis-pQTLs, in subsequent rounds, we regressed out the top statistically significant pQTLs from the last round until no independent signal was present anymore. 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 1.41x10\(^{-7}\) (0.05/353,612). For trans-pQTL analysis, the significance threshold was 1.01 x10\(^{-11}\) (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 analyzed. 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-\(\alpha\) (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 ileocolonic resection compared to patients with UC (P<0.01), while there was no significant difference for (partial) colon resections (P=0.36).
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.
Table 1. Descriptive statistics of the study population (n = 1,028 patients with IBD and n = 148 healthy individuals).

<table>
<thead>
<tr>
<th>Variable</th>
<th>CD</th>
<th>UC</th>
<th>HC</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex, n (%)</td>
<td>567 (100%)</td>
<td>461 (100%)</td>
<td>148 (100%)</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Male</td>
<td>206 (36%)</td>
<td>223 (48%)</td>
<td>73 (51%)</td>
<td>0.001</td>
</tr>
<tr>
<td>Female</td>
<td>361 (64%)</td>
<td>238 (52%)</td>
<td>75 (51%)</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Age (years)</td>
<td>38 [27.53]</td>
<td>43 [30.55]</td>
<td>26 [16.36]</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>24 [21.62, 27.95]</td>
<td>25 [22.39, 28.34]</td>
<td>22 [20.24]</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Plasma storage time (years)</td>
<td>6.7 [4.3, 8.2]</td>
<td>6.2 [4.2, 7.9]</td>
<td>2.6 [2.5, 2.7]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Current smoking, n (%)</td>
<td>543 (96%)</td>
<td>437 (95%)</td>
<td>-</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Yes</td>
<td>164 (30%)</td>
<td>47 (11%)</td>
<td>-</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>No</td>
<td>379 (70%)</td>
<td>390 (89%)</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Montreal classification

Montreal Age (A)  
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)  
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)  
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)  
E1 (proctitis)  450 (98%)  | -        | -  
E2 (left-sided colitis)  144 (32%)  | -        | -  
E3 (pancolitis)  247 (55%)  | -        | -  
Medication use, n (%)  
Aminosalicylates 60 (11%)  | 305 (66%)  | -  | < 0.01 |
Thiopurines 252 (44%)  | 151 (33%)  | -  | < 0.01 |
Steroids 108 (19%)  | 108 (23%)  | -  | 0.09 |
Calcineurin inhibitors 10 (2%)  | 23 (5%)  | -  | < 0.01 |
Methotrexate 51 (9%)  | 4 (0.9%)  | -  | < 0.01 |
Mycophenolate mofetil 5 (0.9%)  | 9 (2%)  | -  | 0.18 |
Oral contraceptives 72 (13%)  | 36 (8%)  | -  | < 0.05 |
Antibiotics 20 (4%)  | 13 (3%)  | -  | 0.52 |
Anti-TNF-κ use included infliximab, adalimumab, certolizumab and golimumab. CD, Crohn’s disease; UC, ulcerative colitis; BMI, body-mass index; TNF-κ, tumor necrosis factor alpha; HBI, Harvey-Bradshaw Index; SCCAI, Simple Clinical Colitis Activity Index.

Data are presented as proportions n with corresponding percentages (%) or median [interquartile range, IQR] in case of continuous variables. *P-values < 0.05 were considered statistically significant. *, actual percentage is 99.6%. #, actual percentage is 99.8%.

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 =3.1x10⁻²⁹; UC vs. controls: P =7.77x10⁻²⁵; 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<0.05, Supplementary Figure S4, 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: SCCAIs ≤ 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, 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), interleukin-17A (IL-17A), and osteoprotegerin (OPG)) being increased and...
six (FGF-19, interleukin-10 (IL-10), matrix metalloproteinase-10 (MMP-10), tumor necrosis factor superfamily member 9 (TNFRSF9), C-X-C motif chemokine 10 (CXCL10) and 9 (CXCL9)) decreased in CD compared to UC (Supplementary Figure S5, Supplementary Tables S5-6).

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 analyzed 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 analyzed 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).

Plasma FGF-19 levels are decreased in patients with CD having ileal disease or a history of ileocecal resection
In patients with CD, a history of ileocecal 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), tumor 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=4.06x10^-11, Figure 4A). In addition, FGF-19 level was lower in patients with either strictureing (Montreal B2, P=1.61x10^-7) or penetrating (Montreal B3, P=6.69x10^-14) disease behavior compared to non-stricturing, non-penetrating disease behavior (Montreal B1) (Figure 4B) (FDR<0.05). When excluding patients from this analysis who had a history of ileocecal resection, these differences remained statistically significant (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 with the extension of the disease (Montreal E) (FDR<0.05, Supplementary Figure S9).

**Smoking is associated with decreased plasma IL-12B levels**

In patients with CD, active smoking behavior was associated with decreased levels of IL-12B, stem cell factor (SCF), CXCL10 and beta-nerve growth factor (β-NGF), 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 (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).

**Figure 4 (A–B).** Plasma FGF-19 levels are decreased in patients with ileal and non-stricturing, non-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 behavior.

**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, tumor necrosis factor ligand superfamily member 12 (TWEAK) and urokinase-type plasminogen activator (uPA). In patients with UC, 25 different plasma proteins were associated with CRP levels, with top significant associations for IL-6, SCF, DNER, CSF-1 and matrix metalloproteinase-1 (MMP-1) (FDR<0.05). Five of these 25 proteins were inversely associated with CRP: SCF, DNER, TWEAK, TNF-related activation-induced cytokine (TRANCE) and T-cell surface glycoprotein CD6 isoform (CD6). Stratified analyses of associations between demographic and clinical factors and plasma proteins for disease activity (CD: n=372 in remission; n=168 active disease; UC: n=306 in remission; n=137 active disease; n=45 with unknown HBI/SCCAI scores) demonstrated a high degree of consistency of main phenotypic-protein associations (Supplementary Table S9).

**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).

**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 $P=1.41 \times 10^{-7}$) (Table 2). Plasma levels of two proteins, cystatin D (CST5) and C-C motif chemokine 25 (CCL25), were found to associate with two independent cis-pQTL variants, 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. For example, monocyte chemotactic protein 2 (MCP-2, CCL8) levels showed the strongest association with a specific missense variant (rs1133763, $P=1.97 \times 10^{-49}$). This study discovered 13 novel independent cis-pQTL variants, including genetic variants of CST5, TNFB, CD6, T-cell surface glycoprotein CD8 alpha chain (CD8A), adenosine deaminase (ADA), C-X-C motif chemokine 6 (CXCL6), interleukin-10 receptor subunit beta (IL10-RB), MMP-1, CD40L receptor (CD40), programmed cell death 1 ligand 1 (PD-L1), CUB domain-containing protein 1 (CDCP1) and one additional cis-pQTL variant of CCL25 (Table 2).

Next, the overlap between all detected cis-pQTLs with protein-coding gene expression quantitative trait loci (eQTLs) was investigated using intestinal mucosal biopsies derived from both ileum and colon of a subset of patients with IBD ($n=280$ biopsies from 165 individual patients), and in the largest public IBD GWAS so far. Four (4) independent cis-pQTL variants also appeared to have an eQTL effect, including variants of CCL25 ($P=0.0015$, ileum), CXCL5 ($P=8.36 \times 10^{-4}$, colon), MMP-1 ($P=1.68 \times 10^{-4}$, 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 ($P=5.86 \times 10^{-22}$, rs602662) was further identified, which is located in the fucosyltransferase 2 (FUT2) gene, a known IBD-associated risk locus (Supplementary Table S15).
Chapter 3 - The Effect of Phenotype and Genotype on the Plasma Proteome

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 alone (G-allele for rs2032887 and A-allele for rs3745387), and the same was observed for the trans-pQTL variant rs602662 (G-allele) (P = 2.55x10^{-15}). To determine whether the genetic regulation can also be observed at gene expression level, these pQTL variants were analyzed 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-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 on CCL25 levels (Figure 6D) of only carriage of both cis-pQTL variants and the trans-pQTL variant rs602662 on CCL25 levels was also observed to be a cis-eQTL variant on FUT2 gene expression (P = 0.22). To determine whether the genetic regulation can also be observed at gene expression level, these pQTL variants were analyzed 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-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 on CCL25 levels (Figure 6D) of only carriage of both cis-pQTL variants and the trans-pQTL variant rs602662 on CCL25 levels was also observed to be a cis-eQTL variant on FUT2 gene expression (P = 0.22). To determine whether the genetic regulation can also be observed at gene expression level, these pQTL variants were analyzed 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-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 on CCL25 levels (Figure 6D) of only carriage of both cis-pQTL variants and the trans-pQTL variant rs602662 on CCL25 levels was also observed to be a cis-eQTL variant on FUT2 gene expression (P = 0.22).  

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

<table>
<thead>
<tr>
<th>Protein</th>
<th>SNP</th>
<th>Chr</th>
<th>Allele</th>
<th>P-value (CD)</th>
<th>R (CD)</th>
<th>P-value (UC)</th>
<th>R (UC)</th>
<th>Meta-P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD8A</td>
<td>rs3020726*</td>
<td>2</td>
<td>G</td>
<td>2.03x10^{-19}</td>
<td>0.39</td>
<td>3.64x10^{-11}</td>
<td>0.32</td>
<td>1.19x10^{-26}</td>
</tr>
<tr>
<td>IL-18R1</td>
<td>rs11377261</td>
<td>2</td>
<td>A</td>
<td>1.55x10^{-22}</td>
<td>0.44</td>
<td>3.67x10^{-20}</td>
<td>0.47</td>
<td>1.77x10^{-40}</td>
</tr>
<tr>
<td>CDCP1</td>
<td>rs6244470</td>
<td>3</td>
<td>T</td>
<td>1.32x10^{-6}</td>
<td>-0.20</td>
<td>1.60x10^{-8}</td>
<td>-0.16</td>
<td>1.52x10^{-8}</td>
</tr>
<tr>
<td>CXCL1</td>
<td>rs317604*</td>
<td>4</td>
<td>C</td>
<td>5.49x10^{-6}</td>
<td>0.25</td>
<td>1.15x10^{-9}</td>
<td>0.17</td>
<td>1.20x10^{-7}</td>
</tr>
<tr>
<td>CXCL5</td>
<td>rs425535*</td>
<td>4</td>
<td>T</td>
<td>1.20x10^{-10}</td>
<td>0.28</td>
<td>1.19x10^{-7}</td>
<td>0.26</td>
<td>1.49x10^{-16}</td>
</tr>
<tr>
<td>CXCL6</td>
<td>rs16850073</td>
<td>4</td>
<td>T</td>
<td>3.18x10^{-11}</td>
<td>0.38</td>
<td>1.50x10^{-10}</td>
<td>0.33</td>
<td>6.59x10^{-26}</td>
</tr>
<tr>
<td>IL-12B</td>
<td>rs4921484*</td>
<td>5</td>
<td>T</td>
<td>4.67x10^{-6}</td>
<td>-0.27</td>
<td>2.70x10^{-9}</td>
<td>-0.27</td>
<td>1.30x10^{-14}</td>
</tr>
<tr>
<td>TNFB</td>
<td>rs2229092</td>
<td>6</td>
<td>C</td>
<td>7.62x10^{-21}</td>
<td>-0.40</td>
<td>1.80x10^{-16}</td>
<td>-0.40</td>
<td>2.40x10^{-35}</td>
</tr>
<tr>
<td>VEGFA</td>
<td>rs6921438*</td>
<td>6</td>
<td>A</td>
<td>8.63x10^{-23}</td>
<td>-0.44</td>
<td>7.62x10^{-14}</td>
<td>-0.38</td>
<td>1.07x10^{-24}</td>
</tr>
<tr>
<td>PD-L1</td>
<td>rs35744625</td>
<td>9</td>
<td>C</td>
<td>6.62x10^{-6}</td>
<td>0.16</td>
<td>6.03x10^{-6}</td>
<td>0.23</td>
<td>7.04x10^{-6}</td>
</tr>
<tr>
<td>IL-15RA</td>
<td>rs4237402*</td>
<td>10</td>
<td>A</td>
<td>9.38x10^{-7}</td>
<td>0.23</td>
<td>1.96x10^{-3}</td>
<td>0.16</td>
<td>1.69x10^{-8}</td>
</tr>
<tr>
<td>CD6</td>
<td>rs11230563</td>
<td>11</td>
<td>T</td>
<td>4.84x10^{-17}</td>
<td>-0.37</td>
<td>1.14x10^{-12}</td>
<td>-0.34</td>
<td>1.17x10^{-27}</td>
</tr>
<tr>
<td>MPP-1</td>
<td>rs10791596</td>
<td>11</td>
<td>T</td>
<td>1.35x10^{-6}</td>
<td>-0.26</td>
<td>8.73x10^{-6}</td>
<td>-0.22</td>
<td>1.08x10^{-12}</td>
</tr>
<tr>
<td>CCL23</td>
<td>rs712048*</td>
<td>17</td>
<td>A</td>
<td>1.82x10^{-10}</td>
<td>-0.29</td>
<td>2.70x10^{-9}</td>
<td>-0.16</td>
<td>1.75x10^{-11}</td>
</tr>
<tr>
<td>CCL3</td>
<td>rs1719144*</td>
<td>17</td>
<td>A</td>
<td>1.18x10^{-6}</td>
<td>0.24</td>
<td>4.56x10^{-4}</td>
<td>0.23</td>
<td>4.87x10^{-12}</td>
</tr>
<tr>
<td>MCP-2</td>
<td>rs1133763</td>
<td>17</td>
<td>C</td>
<td>6.52x10^{-33}</td>
<td>-0.50</td>
<td>1.29x10^{-18}</td>
<td>-0.42</td>
<td>1.97x10^{-49}</td>
</tr>
<tr>
<td>CCL25</td>
<td>rs2032887</td>
<td>19</td>
<td>G</td>
<td>2.88x10^{-26}</td>
<td>0.45</td>
<td>2.53x10^{-13}</td>
<td>0.36</td>
<td>1.92x10^{-37}</td>
</tr>
<tr>
<td>CCL27</td>
<td>rs3745387*</td>
<td>19</td>
<td>A</td>
<td>1.45x10^{-6}</td>
<td>0.25</td>
<td>3.95x10^{-6}</td>
<td>0.20</td>
<td>5.45x10^{-12}</td>
</tr>
<tr>
<td>ADA</td>
<td>rs112545364</td>
<td>20</td>
<td>G</td>
<td>7.35x10^{-17}</td>
<td>0.36</td>
<td>8.69x10^{-11}</td>
<td>0.32</td>
<td>8.60x10^{-26}</td>
</tr>
<tr>
<td>CD40</td>
<td>rs1883832</td>
<td>20</td>
<td>T</td>
<td>1.04x10^{-7}</td>
<td>-0.34</td>
<td>2.68x10^{-6}</td>
<td>-0.14</td>
<td>5.30x10^{-9}</td>
</tr>
<tr>
<td>CST5</td>
<td>rs34269359</td>
<td>20</td>
<td>A</td>
<td>4.69x10^{-19}</td>
<td>0.40</td>
<td>1.36x10^{-20}</td>
<td>0.46</td>
<td>3.86x10^{-37}</td>
</tr>
<tr>
<td>CST5</td>
<td>rs1799841</td>
<td>20</td>
<td>G</td>
<td>4.76x10^{-7}</td>
<td>0.20</td>
<td>9.86x10^{-6}</td>
<td>0.19</td>
<td>3.65x10^{-9}</td>
</tr>
<tr>
<td>IL-10RB</td>
<td>rs4455239</td>
<td>21</td>
<td>C</td>
<td>3.33x10^{-6}</td>
<td>0.26</td>
<td>5.28x10^{-15}</td>
<td>0.39</td>
<td>6.47x10^{-20}</td>
</tr>
</tbody>
</table>

Abbreviations: Chr, chromosome; SNP, single nucleotide polymorphism; R, correlation coefficient; *Replicated cis-pQTL variants compared with Sun et al, Nature 2018.10
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.

**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. Color hues for R² 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.

**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. To further explore the effect on the gut microbiota, we re-analyzed data from a previous genome-wide mbQTL analysis, and found the rs602662 variant to be associated with lower abundance of the species *Ruminococcus obeum* (Pmeta=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). Similarly, the rs602662 variant was observed to be associated with a lower abundance of *Faecalibacterium prausnitzii* (Pmeta=0.00397), a well-known, commensal butyrate-producing bacterial species in the human gut.

**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. Interestingly, IL-12B encodes the p40 subunit of IL-12, which is also part of IL-23, and is a known IBD susceptibility locus. A previous study from our center discovered a complex gene-environment interaction between the IL-12B SNP rs6887695 in non-smoking patients with IBD, but not in smoking patients. 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. Repressed IL-12B levels may be reflective of an impaired immunity and anti-tumor 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 behavior 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.29-35 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 synthesis in the liver.29 Chronic diarrhea 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.30 Patients with CD who underwent ileocecal resection indeed show impaired release of FGF-19 to the circulation.30-32 Furthermore, intestinal inflammation disrupts the epithelial barrier integrity, which is accompanied by impairment of transepithelial transport mechanisms, including bile acid reabsorption.33 Therefore, not only ileocecal resection but also intestinal inflammation, which could lead to deterioration of bile acid homeostasis, is associated with lower plasma FGF-19 levels.34 Based on these findings, plasma FGF-19 levels may be a potential biomarker to identify patients who are most likely to benefit from therapy with bile acid sequestrants or FXR agonists. Furthermore, previous studies using chemically induced models of colitis demonstrated amelioration of intestinal inflammation, barrier integrity and gut microbiota composition upon pharmacological FXR activation, demonstrating therapeutic potential especially in CD patients with active disease or surgical history and concurrent disruption of bile acid metabolism.35-37

In our large-scale pQTL analysis, one independent trans-pQTL variant and two cis-pQTL variants were significantly associated with plasma CCL25 levels. CCL25, also known as thymus-expressed chemokine (TECK) is constitutively expressed in the thymus and small intestinal epithelium, but nearly absent in the colon.36 CCL25 elicits recruitment of peripheral blood lymphocytes expressing the chemokine receptor CCR9 and the adhesion molecule 47.39 Through interaction with CCL25, CCR9-expressing lymphocytes are guided to the intra-epithelial lymphocyte (IEL) compartment, the small intestinal lamina propria and, to a lesser extent, the colonic lamina propria. Increased activity of the CCL25-CCR9 axis and ensuing CD4+ effector memory T-cell migration are implicated in primary sclerosing cholangitis, experimental postoperative ileus and IBD.40-43 In small intestinal CD, CCL25 is strongly expressed in proximity to lymphocytic infiltrates and CCR9+ gut-homing CD4+ effector T-cells are increased in the circulation when compared to colonic CD.42 Recent data suggest that colonic CCL25 expression and CCR9+ CD4+ effector T-cell recruitment may also be upregulated in patients with active colitis.44 As such, the observed association of genotype and plasma CCL25 levels may play a functional role in the immune dysregulation in these patients, and this may be of particular interest for future functional studies. Although CCR9 has been considered a viable therapeutic target for years, clinical trials still have to demonstrate reliably efficacy of CCR9-inhibitors.44 One independent trans-pQTL variant emerged in the FUT2 gene (located at chromosome 19q13.33) affecting plasma CCL25 levels. However, no trans-eQTL effect on CCL25 was observed. Likewise, previous studies demonstrated limited overlap between pQTLs and eQTLs, where, on average, only one-third of pQTLs appears to have a corresponding eQTL effect.13

FUT2 is a known candidate IBD-risk locus and is abundantly expressed in intestinal tissue. FUT2 encodes alpha-(1,2)-fucosyltransferase that regulates the secretion of ABO blood group antigens by the intestinal epithelium and secretory glands.45 Homozygosity for non-functional 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.46 In addition, “non-secretors” of FUT2 show alterations in their gut microbiota, including a decrease in microbial diversity, changes in several microbial taxa, and altered microbial energy metabolism.47-49 A recent study demonstrated that FUT2 non-secretors exhibit lower fecal levels of the genus Blautia which is consistent with our findings on the rs602662 non-secretor FUT2 variant.50 Altogether, 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 FUT2 variants and CD.51 The predominant 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 FUT2-inactivating variants, including our reported trans-pQTL missense variant
of FUT2 (S258G, rs602662). Considering the above, one may hypothesize that activation of mucosal T-lymphocytes is shaped by FUT2-mediated host-microbe interactions, as FUT2 is associated with an increased susceptibility to IBD. However, it remains unclear what interactions could determine this putative disease mechanism.

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. 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. 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. For example, IL-6 was the top significant association to CRP in both patients with CD and UC, which was not surprising given its importance in stimulating the production of acute-phase reactants in the liver. In IBD, IL-6 levels rise according to the level of inflammatory disease activity, and it is strongly associated with (endoscopically confirmed) intestinal inflammation. Similarly, IFN-α was strongly correlated to CRP, especially in patients with CD. Indeed, IFN-α – as a central cytokine within the Th1-driven immune response – plays a pivotal role in CD pathogenesis and has repeatedly been shown to associate with fecal calprotectin levels and endoscopically-proven intestinal inflammation. 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 signaling pathway regulates cellular apoptosis and intercellular interactions within the intestinal epithelium. Notch-1 signaling 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.

References


