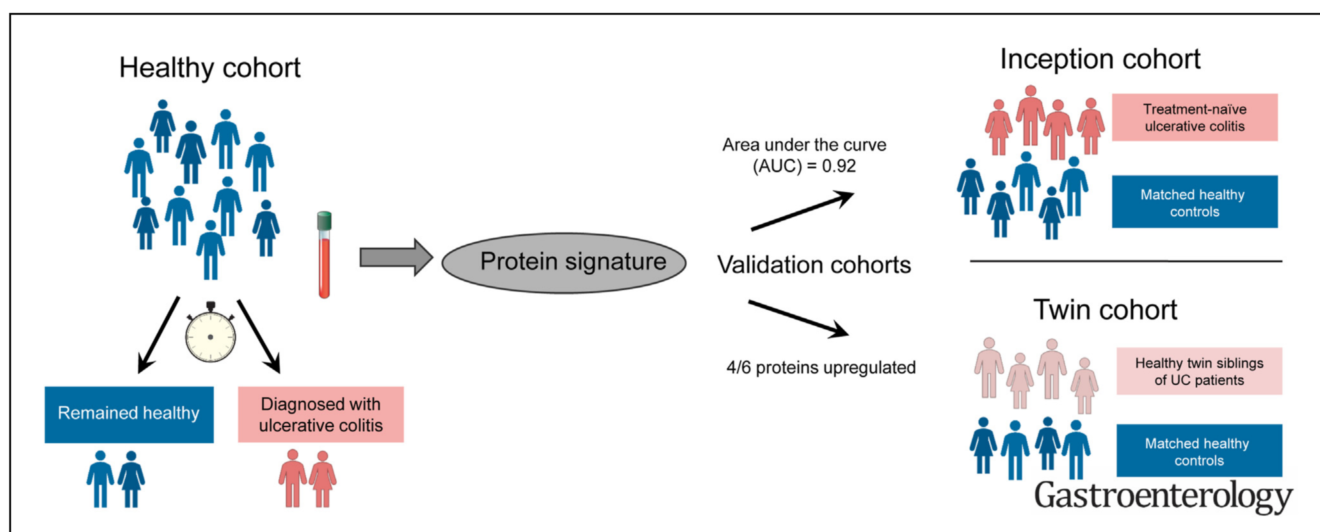
Gastroenterology 2021;161:1526–1539

BASIC AND TRANSLATIONAL—ALIMENTARY TRACT

Systemic Inflammation in Preclinical Ulcerative Colitis

Daniel Bergemalm,[1] Erik Andersson,[1] Johan Hultdin,[2] Carl Eriksson,[1] Stephen T. Rush,[3] Rahul Kalla,[4] Alex T. Adams,[5] Åsa V. Keita,[6] Mauro D'Amato,[7,8] Fernando Gomollon,[9] Jørgen Jahnsen,[10,11] IBD Character Consortium, Petr Ricanek,[10] Jack Satsangi,[5,12] Dirk Repsilber,[3] Pontus Karling,[13] and Jonas Halfvarson[1]

[1]Department of Gastroenterology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden; [2]Department of Medical Biosciences, Division of Clinical Chemistry, Umeå University, Umeå, Sweden; [3]School of Medical Sciences, Faculty of Medicine and Health, Örebro University, Örebro, Sweden; [4]Medical Research Council Centre for Inflammation Research, Queens Medical Research Institute, University of Edinburgh, Edinburgh, United Kingdom; [5]Translational Gastroenterology Unit, Nuffield Department of Medicine, Experimental Medicine Division, University of Oxford, John Radcliffe Hospital, Oxford, United Kingdom; [6]Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden; [7]CIC bioGUNE Basque Research and Technology Alliance and Basque Science Foundation, Bilbao, Spain; [8]Division of Clinical Epidemiology, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden; [9]Hospital Clinico Universitario Lozano Blesa, IIS Aragón, Zaragoza, Spain; [10]Department of Gastroenterology, Akershus University Hospital, Lørenskog, Norway; [11]Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [12]Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom; and [13]Department of Public Health and Clinical Medicine, Division of Medicine, Umeå University, Umeå, Sweden



BACKGROUND & AIMS: Preclinical ulcerative colitis is poorly defined. We aimed to characterize the preclinical systemic inflammation in ulcerative colitis, using a comprehensive set of proteins. **METHODS:** We obtained plasma samples biobanked from individuals who developed ulcerative colitis later in life (n = 72) and matched healthy controls (n = 140) within a population-based screening cohort. We measured 92 proteins related to inflammation using a proximity extension assay. The biologic relevance of these findings was validated in an inception cohort of patients with ulcerative colitis (n = 101) and healthy controls (n = 50). To examine the influence of genetic and environmental factors on these markers, a cohort of healthy twin siblings of patients with ulcerative colitis (n = 41) and matched healthy controls (n = 37) were explored. **RESULTS:** Six proteins (MMP10, CXCL9, CCL11, SLAMF1, CXCL11 and MCP-1) were up-regulated ($P < .05$) in preclinical ulcerative colitis compared with controls based on both univariate and multivariable models. Ingenuity Pathway Analyses identified several potential key regulators, including interleukin-1β, tumor necrosis factor, interferon-gamma, oncostatin M, nuclear factor-κB, interleukin-6, and interleukin-4. For validation, we built a multivariable model to predict disease in the inception cohort. The model discriminated treatment-naïve patients with ulcerative colitis from controls with leave-one-out cross-validation (area under the curve = 0.92). Consistently, MMP10, CXCL9, CXCL11, and MCP-1, but not CCL11 and SLAMF1, were significantly up-regulated among the healthy twin siblings, even though their relative abundances seemed higher in incident ulcerative colitis. **CONCLUSIONS:** A set of inflammatory proteins are up-regulated several years before a diagnosis of ulcerative colitis. These proteins were highly predictive of an ulcerative colitis diagnosis, and some seemed to be up-regulated already at exposure to genetic and environmental risk factors.

Keywords: Inflammatory Bowel Disease; Preclinical Disease; Proximity Extension Assay; MMP10; CXCL9.

The inflammatory bowel diseases (IBDs), including the 2 major forms—ulcerative colitis and Crohn's disease—are chronic inflammatory diseases of the gastrointestinal tract. As with many other chronic immune-mediated diseases, inflammation seems to result from exposure to environmental risk factors in genetically predisposed individuals. Our current understanding of early key drivers and initiating triggers of the immune dysregulation that ultimately leads to an irreversible inflammation is poor.[1] Nevertheless, an increasing amount of data indicates that the onset of symptoms and the diagnosis of an overt disease is preceded by a preclinical disease phase. During this phase, complex interactions between genetic and environmental risk factors lead to microbial shifts, loss of epithelial integrity, and initiation and propagation of a dysregulated immune response.[1,2] These early disease processes eventually culminate in subclinical inflammation, with activation of both the innate and adaptive immune systems.[1] This hypothesis is supported by analyses of individuals who are at high risk of developing IBD, such as healthy twin siblings of patients with IBD. Zhulina et al[3] demonstrated that exposure to genetic and environmental risk factors translates into a subclinical mucosal inflammation, defined by increased nuclear factor-κB (NFκB) activity, and activation of neutrophils among 73%–75% of the healthy twin siblings among pairs discordant for IBD.

In contrast to many other chronic diseases in which the initiation of the preclinical phase is defined by development of circulating autoantibodies,[4,5] disease onset is poorly defined in IBD. Even though no autoantigen has been identified, antibodies to microbial antigens have been demonstrated in 19%–61% of individuals who develop Crohn's disease later in life,[6–8] and perinuclear anti-neutrophil cytoplasmic antibodies (pANCAs) can be present several years before diagnosis of ulcerative colitis.[7,8] Activation of the adaptive immune response seems to be associated with systemic inflammation. Lochhead et al[9] reported elevated levels of high-sensitive C-reactive protein and interleukin (IL)-6 in serum samples from individuals with preclinical Crohn's disease and ulcerative colitis. Recent data from the PREDICTS (Proteomic Evaluation and Discovery in an IBD Cohort of Tri-service Subjects) study demonstrate that Crohn's disease can be predicted up to 5 years before diagnosis by analyzing a panel of antibodies and SOMAmers, that is, protein biomarkers.[10] However, in individuals with preclinical ulcerative colitis, only 2 SOMAmers—IL-11 receptor subunit alpha and lymphocyte activation gene 3 protein— were identified as differentially regulated across all time points and their overall predictive performance was low in the multivariate analysis. Taken together, these studies indicate that no protein signature has yet been identified in the serum of individuals who are diagnosed with ulcerative colitis later in life.

Gaining insight into the early phases of ulcerative colitis could advance our knowledge of disease pathogenesis and treatment targets because the pattern of early key inflammatory pathways can be masked by nonspecific inflammation once patients develop symptoms and are diagnosed with established disease. We designed a study with the aim to characterize the systemic preclinical inflammatory immune response of ulcerative colitis. Plasma samples, collected up to 15 years before the diagnosis of ulcerative colitis, were analyzed using a highly specific protein panel of inflammatory markers. To examine the biological relevance, we assessed the identified dysregulated protein markers in treatment-naïve, newly diagnosed patients with ulcerative colitis. Ultimately, we examined whether the dysregulated proteins were triggered by exposure to genetic and environmental risk factors in a twin cohort with ulcerative colitis.

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

Preclinical assessment of inflammatory proteins could provide insight into the early phases of ulcerative colitis, identify individuals at increased risk for developing the disease, and work as a diagnostic tool.

NEW FINDINGS

We identified an inflammatory protein signature in plasma from individuals who developed ulcerative colitis later in life. This signature had a high diagnostic capacity in an independent inception cohort. According to analyses of healthy twin siblings of patients with ulcerative colitis, some proteins were already up-regulated at exposure to genetic and environmental risk factors.

LIMITATIONS

These findings must be replicated in independent cohorts.

IMPACT

This panel can be used as a diagnostic tool and provides important insights into the preclinical stage of ulcerative colitis.

Methods

Study Design

We performed a case–control study, nested within a cohort study, and compared prediagnostic plasma samples from patients who developed ulcerative colitis later in life (cases) with those from individuals who remained free from IBD during follow-up (controls). Next, we validated our findings in an inception cohort of treatment-naïve patients with ulcerative

Abbreviations used in this paper: **AUC, area under the curve; DSS, dextran sulfate sodium; IBD, inflammatory bowel disease; IFN, interferon; IL, interleukin; LOD, limit of detection; LOO, leave-one-out; MMP, matrix metalloproteinase; NFκB, nuclear factor-κB; OSM, oncostatin-M; pANCA, perinuclear anti-neutrophil cytoplasmic antibody; PEA, proximity extension assay; TNF, tumor necrosis factor.**

© 2021 by the AGA Institute. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

0016-5085
https://doi.org/10.1053/j.gastro.2021.07.026

BASIC AND TRANSLATIONAL AT

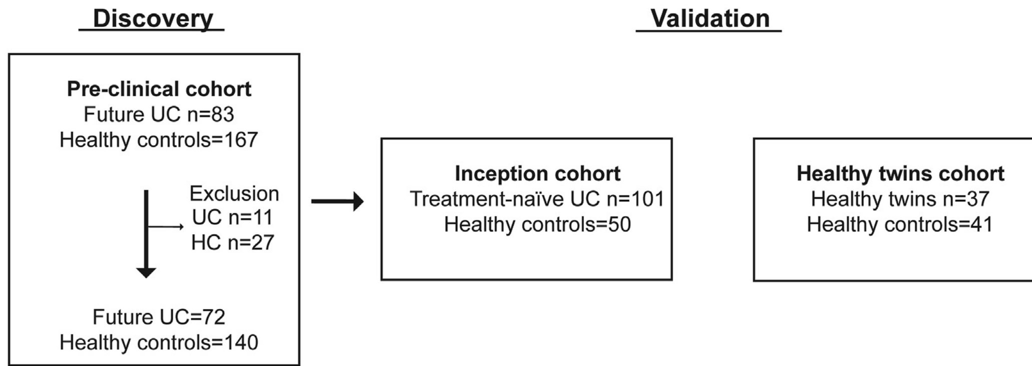


Figure 1. Flowchart showing cohorts included in the study.

colitis. To examine the influence of genetic and environmental risk factors, we examined twin pairs discordant for ulcerative colitis and compared healthy twin siblings with external healthy controls, as outlined in Figure 1.

Register Sources

The unique personal identity number, issued to all Swedish residents,[11] was used to link records from the following population-based registers.

The National Patient Registry. The Swedish National Patient Registry comprises data on hospital admissions since 1964, with national coverage since 1987. From 2001, it also includes information on nonprimary outpatient care. Main and contributory diagnoses are coded according to the International Classification of Diseases codes and are assigned by the treating physician.

The Northern Sweden Health and Disease Study register. The Northern Sweden Health and Disease Study register cohort contains 3 subcohorts, from which individuals within the Västerbotten Intervention Program and the Mammography Screening Project (MA) were eligible for the current study.[12] The Västerbotten Intervention Program has been described in detail previously.[13] In the MA cohort, blood samples and survey data were collected at mammography. In total, 54,000 blood samples were taken from women aged 18–82 years, of whom 95% were aged 48–70 years at sample collection.[14]

The Swedish Twin Registry. The Swedish Twin Registry was established in the 1960s and has information about some 85,000 twin pairs for which zygosity is known. Zygosity is determined by DNA analyses or a questionnaire on intrapair similarities in childhood and being of the opposite sex. The questionnaire has been validated previously and intrapair similarities have ≥98% accuracy compared with DNA analyses.[15]

Patients and Cohorts

The diagnosis of ulcerative colitis was based on internationally accepted criteria, following thorough clinical, microbiological, endoscopic, histologic, and radiologic evaluation in all 3 cohorts.[16] The date of diagnosis was defined as the date of the first endoscopy demonstrating macroscopic and histologic evidence of ulcerative colitis. Detailed information on each cohort is provided in the Supplementary Material.

Cohort of preclinical ulcerative colitis. Individuals who developed ulcerative colitis later in life were identified by linking the Northern Sweden Health and Disease Study register dataset with the International Classification of Diseases code register of Region of Västerbotten, Sweden. Copies of the medical notes from the departments of medicine, surgery, and pathology for all individuals with at least 1 inpatient or nonprimary outpatient care visit listing a diagnosis of IBD, according to the International Classification of Diseases, Tenth Revision codes (K50.1-9 or K51.1-9), were manually scrutinized by an experienced gastroenterologist to confirm or reject a diagnosis of ulcerative colitis and to classify the disease according to the Montreal classification.[17]

Inception cohort of treatment-naïve ulcerative colitis. To validate the biological relevance of the findings in the preclinical cohort, the identified proteins were measured among patients with ulcerative colitis within the IBD Character (Inflammatory Bowel Disease Characterization by a Multi-Modal Integrated Biomarker Study) cohort. The IBD Character cohort represents a multicenter IBD inception cohort, recruited at 6 European centers (Edinburgh, UK; Oslo, Norway; Örebro, Sweden; Linköping, Sweden; Zaragoza, Spain; and Maastricht, Netherlands) between 2012 and 2015 (EU ref. no. 305676).[18]

Cohort of twin pairs discordant for ulcerative colitis. Twin pairs discordant for ulcerative colitis were identified from a previously described nationwide, population-based cohort of twins with IBD in Sweden.[19,20] For each twin pair, an external nonrelated healthy control, matched by sex and age ±5 years was randomly identified from a previously described cohort of healthy blood donors with no history of chronic gastrointestinal disease, recruited at Örebro University Hospital, Sweden.[21,22]

Protein Analysis

The relative concentrations of 92 different proteins were analyzed using the Proseek Multiplex Inflammation I Probe kit 96x96 (Olink Bioscience, Uppsala, Sweden) and reported as arbitrary units, that is, normalized protein expression on a log2 scale, as described previously.[23] Further details can be found in the Supplementary Material.

Data Analysis

Continuous data, such as age, and data on an ordinal scale, including body mass index, are presented as median and range

BASIC AND TRANSLATIONAL AT

or interquartile range, and differences were tested with the Mann–Whitney U test, 2-sided, with significance level 95%, and nominal *P* values were reported. Categorical data are presented as frequencies and were compared using the Pearson χ^2 test or Fisher exact test when appropriate. Principal component analyses were performed, and score plots were visually inspected for identification of possible outliers and gross separation of cases and controls.

Preclinical cohort. Potential associations between inflammatory proteins and future diagnosis of ulcerative colitis were assessed by 2 types of models. First, we used logistic regression models, with age, sex, and smoking status as covariates, to investigate associations between individual proteins and risk for future ulcerative colitis. Next, we fitted multiple-protein logistic regression models based on all 65 proteins above limit of detection (LOD) in at least 20% of individuals who developed ulcerative colitis later in life or controls. We used the same covariates as in the single-protein regression models. Hereto, minimax concave penalty–regularized regression models were employed,[24] using a default gamma = 3 and a lambda obtained by optimization in internal cross-validations. We did not use any further constraint to numbers of included proteins. The selection of predictive proteins was based on 1000 model fits, and all proteins with nonzero coefficients in any model are reported. Correlations between individual markers or combinations of markers and time period to a diagnosis of ulcerative colitis were assessed using Spearman ρ or Pearson correlation test. To examine potential heterogeneity within the group of individuals with prediagnostic ulcerative colitis and increase the likelihood of identifying subgroups of individuals, each protein marker was assigned a quartile score of 1, 2, 3, and 4, respectively.[25] Quartile sums for significant protein markers were calculated and tested against time to diagnosis, age, sex, and extent of inflammation at the diagnosis of ulcerative colitis.

Pathway analyses. We used Ingenuity Pathway Analysis software (Ingenuity Systems Inc, Qiagen, Redwood City, CA) to identify canonical pathways and upstream regulators of dysregulated preclinical protein markers. A detailed description of the analysis can be found in the Supplementary Material.

Inception cohort and twin cohort. Using the proteins derived from the preclinical cohort, biosignature models were created by implementing logistic regression for the inception cohort and twin cohort. The predictive ability of the models was validated through leave-one-out (LOO) cross-validation and visualized by receiver operating characteristic curves and by assessing the area under the receiver operating characteristic curve (AUC).

Information on statistical comparisons across cohorts is provided in the Supplementary Material. Statistical analyses and data processing were performed in IBM SPSS Statistics for Windows, version 25.0 (IBM Corp, Armonk, NY) and R 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria) with the packages caret 6.0-84, ncvreg 3.12.0 and pROC 1.15.3.[24,26,27]

Ethics Statement

The study was approved by the regional ethical boards (Dnr 06-024M, 2010-284-31M, Dnr 2010/313, Dnr 167/03), and all centers were granted local ethics approval. All patients gave written and informed consent before participating in this study.

Results

Cohort With Preclinical Ulcerative Colitis

Prediagnostic plasma samples from 83 individuals who developed ulcerative colitis later in life were identified from the prospectively collected Northern Sweden Health and Disease Study register cohort. To minimize the potential effects of active inflammation because of a likely diagnostic delay, 8 individuals who were diagnosed with ulcerative colitis within 1 year from blood sampling were excluded. During experimental analyses, 3 plasma samples failed proximity extension assay (PEA) for technical reasons. The final cohort used for the characterization of preclinical inflammation comprised 72 cases with preclinical ulcerative colitis and 140 matched healthy controls. Demographics and clinical characteristics of patients with ulcerative colitis are reported in Table 1. The median period from when a prediagnostic sample was obtained to the diagnosis of ulcerative colitis was 4.8 years (interquartile range, 2.2–7.2 years). A numerically higher but not statistically different proportion of former smokers ($P = .08$) was observed among individuals who developed ulcerative colitis later in life compared with controls (Table 1), but no difference in body mass index was observed ($P = 0.84$).

Differentially Regulated Proteins in Individuals Who Develop Ulcerative Colitis Later in Life

To characterize the preclinical inflammatory profile, plasma samples were subjected to PEA using a panel of 92 predefined cytokines, chemokines, and other proteins involved in inflammation (Supplementary Table 1). Twenty-seven proteins had quantified levels below the limit of detection in >80% of samples from preclinical cases and controls and were excluded from further analyses. The principal component analysis did not reveal any clear separation between preclinical cases and controls with respect to the overall inflammatory signature (Supplementary Figure 1). However, single-protein logistic regression models identified 6 specific proteins (MCP-1, CXCL9, SLAMF1, CCL11, MMP10, and CXCL11 that were differentially regulated (nominal *P* value <.05) between preclinical ulcerative colitis and controls (Table 2). All markers remained significant after including sex, age, and smoking habits as covariates. Among the markers, MMP10, CXCL9, and CCL11 were most strongly associated with a future diagnosis of ulcerative colitis. The measured value of SLAMF1 was above LOD for only 26% of preclinical cases and even less for controls.

The minimax concave penalty penalized multiple-protein models with all 65 PEA-assessed proteins identified 9 proteins (MCP-1, CXCL9, SLAMF1, CCL11, MMP10, PDL1, CXCL5, CXCL6, and STAM-binding protein), that is, these models included all but 1 (CXCL11) of those proteins identified by the single-protein models. Strongest correlations were observed between CXCL11 and CXCL9 (Spearman $\rho = 0.74$; $P < .01$), as well as with CCL11 (Spearman $\rho = 0.33$; $P < .01$) (Supplementary Table 2). When excluding CXCL9 and CCL11 from the analyses,

BASIC AND TRANSLATIONAL AT

Table 1. Demographic and Clinical Characteristics of the Study Cohorts

	Preclinical cohort		Inception cohort		Twin cohort	
Clinical characteristics	Ulcerative colitis (n = 72)	Controls (n = 140)	Ulcerative colitis (n = 101)	Controls (n = 50)	Healthy twin (n = 37)	Controls (n = 41)
Sex, *male*, n (%)	34 (47)	64 (46)	63 (62)**	20 (40)	14 (38)	18 (44)
Body mass index, kg/m^2, median (IQR)	25 (23–28)	26 (23–28)	NA	NA	NA	NA
Smoking status, n (%)						
Current	22 (31)	32 (23)	5 (5)	3 (6)	8 (22)	NA
Former	26 (36)	39 (28)	44 (44)**	10 (20)	7 (19)	NA
Never	24 (33)	66 (47)	48 (48)	37 (74)	17 (46)	NA
Missing	0 (0)	3 (2)	4 (4)	0 (0)	5 (13)	41 (100)
Age at sample, *y*, median (range)	50 (30–70)	50 (30–70)	37 (18–77)**	26 (19–65)	59 (29–80)	60 (29–69)
Age at diagnosis, *y*, median (range)	54 (31–74)	—	37 (18–77)	—	—	—
Disease extent, n (%)						
Proctitis (E1)	16 (22)	—	25 (25)	—	—	—
Left-sided colitis (E2)	28 (39)	—	33 (33)	—	—	—
Extensive colitis (E3)	28 (39)	—	43 (43)	—	—	—

NOTE. Cases and controls in the preclinical cohort were matched by sex, age, area of residence, and time of sampling. Data on sex, age, body mass index, and smoking status were collected at the time of plasma sampling. There were no statistically significant differences between cases and controls with regard to the demographic and clinical variables. The cohort of treatment-naïve patients with incident ulcerative colitis and healthy controls were collected at 6 European IBD centers. Patients with ulcerative colitis were significantly older and a larger proportion were men and former smokers compared with healthy control subjects. In the cohort of healthy twin siblings among pairs discordant for ulcerative colitis, healthy blood donors, matched for age and sex, were included as controls.
IQR, interquartile range; NA, not available.
$^{**}P < .01$.

Table 2. Differentially Expressed Plasma Protein Markers in the Preclinical Cohort

Protein	Symbol	Uniprot ID	*P* value	AUC (95% CI)
Stromelysin-2	MMP10	P09238	.004	0.65 (0.57–0.73)
C-X-C motif chemokine 9	CXCL9	Q07325	.004	0.64 (0.56–0.72)
Eotaxin	CCL11	P51671	.008	0.63 (0.55–0.71)
Signaling lymphocytic activation molecule	SLAMF1	Q13291	.025	0.60 (0.52–0.68)
C-X-C motif chemokine 11	CXCL11	O14625	.028	0.60 (0.52–0.68)
Monocyte chemotactic protein 1	MCP-1	P13500	.033	0.61 (0.53–0.69)
All 6 markers combined	—	—	.001	0.71 (0.63–0.78)

NOTE. The relative abundance of plasma proteins among individuals with preclinical ulcerative colitis and healthy controls was compared using logistic regression and adjusting for age, sex, and smoking status.
CI, confidence interval.

CXCL11 was selected in the minimax concave penalty penalized multivariable models.

The 6 proteins that were identified by both the single-protein and multiple-protein models were used in all downstream analyses. To examine potential correlations with the period from when a sample was obtained to until the date of diagnosis of ulcerative colitis, we first analyzed measured values of each of the 6 prediagnostic markers alone. We did not observe any significant correlations between the individual markers and period to diagnosis of ulcerative colitis (Supplementary Figure 2). Next, we combined data on all 6 differentially regulated markers by dividing the protein levels of each of the markers in quartiles and calculating the quartile sum score for each individual (ranging from 6 to 24) (Supplementary Figure 3). There was no significant correlation between each individual's quartile sum score and period to the diagnosis of ulcerative colitis (Supplementary Figure 4). Finally, we examined possible correlations between the individual markers or combinations of these markers, based on quartile sum scores, and clinical variables, that is, age, sex, and disease extent at diagnosis. All prediagnostic markers, except MMP10, and the quartile sum scores correlated with age (Supplementary Figure 5), but not with sex or extent of ulcerative colitis (Supplementary Figures 6and 7).

Inflammation Pathway Analyses

We used the Ingenuity Pathway Analysis to explore common inflammatory pathways and potential upstream regulators of the identified prediagnostic proteins. To increase identification of dysregulated pathways in the preclinical cohort, we included all proteins that were selected from the multiprotein models and also added proteins that were showing a trend toward dysregulation ($P < .1$) by single-protein models, that is, STAM-binding protein, caspase 8, and T-cell surface glycoprotein CD5. Identified common upstream regulators of the protein signature are displayed in Figure 2 and included IL-4, IL-1β, TNF, IFN-gamma, NFκB, oncostatin M (OSM), and IL-6 ($z > +2.0$; $P < 1.0 \times 10^{-5}$).

Validation of the Biological Relevance of the Preclinical Protein Markers

The biological relevance of the identified proteins was validated by comparing treatment-naïve patients included at the diagnosis of ulcerative colitis and healthy controls within the IBD Character inception cohort; baseline characteristics are shown in Table 1. Serum protein levels were analyzed using the same PEA platform. Consistent with the findings in the preclinical cohort, all identified proteins, except MCP-1, were found to be differentially regulated when the patients with incident ulcerative colitis were compared with the healthy controls (Supplementary Table 3).

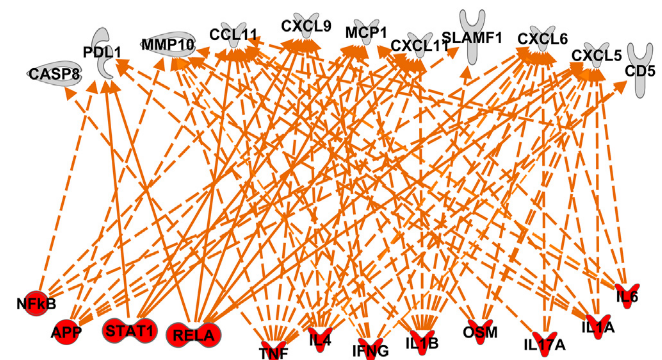


Figure 2. Pathway analysis of prediagnostic protein regulations. Proteins that were significantly dysregulated ($P < .05$) in the preclinical cohort were imported into the Ingenuity Pathway Analysis (IPA) software. The analysis was extended by addition of proteins showing a trend toward dysregulation ($P < .1$) from the univariate analyses or selected as predictive from the multivariable analysis. Using upstream analyses, the database identifies possible regulators of protein alterations in the dataset. The results were filtered to show upstream regulators with corrected $P < .05$ and absolute z score $> \pm 2$. All upstream regulators identified were predicted as up-regulated. *Arrows* indicate interactions based on available information in the IPA database, *orange arrows* represent a predicted up-regulation. APP, amyloid beta precursor protein; IFNG, interferon-gamma; RELA; RELA proto-oncogene; STAT1; signal transducer and activator of transcription 1.

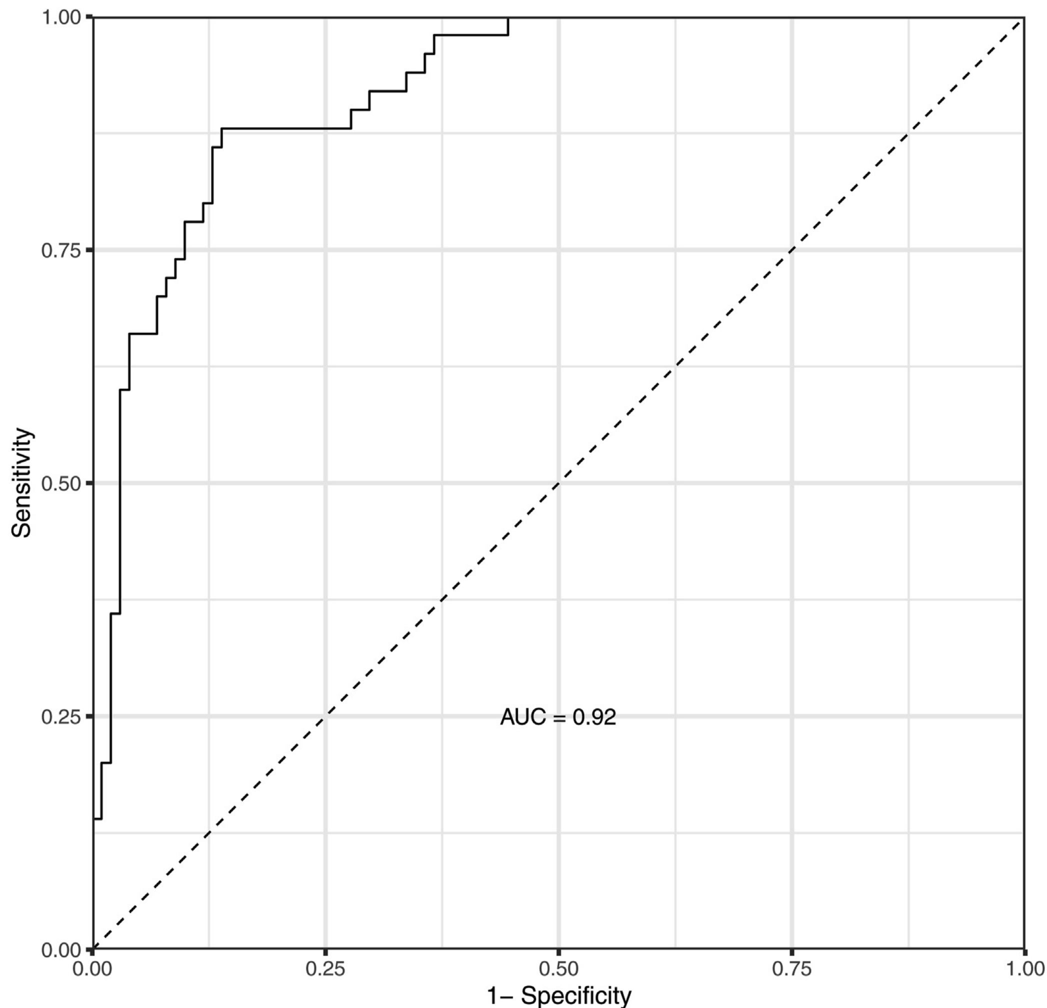


Figure 3. Receiver operating curve (ROC) analysis discriminating treatment-naïve patients with incident ulcerative colitis from healthy controls in the inception cohort. The prediction model, built on the inception cohort of patients with ulcerative colitis with protein selection derived from the dysregulated proteins (n = 5) in the preclinical cohort, was evaluated using LOO cross-validation. The model easily separated patients from healthy controls (LOO AUC = 0.92).

We examined the predictive capacity of these markers by building a multivariable prediction model based on the 5 inflammatory proteins that were replicated and validated its predictive capacity in the inception cohort by performing LOO cross-validation analyses (Figure 3). The model had a high predictive capacity in terms of separating treatment naïve, newly diagnosed patients with ulcerative colitis from healthy controls (LOO AUC = 0.92).

Impact of Shared Genetic and Environmental Risk Factors on Preclinical Protein Markers

We examined the effect of shared genetic and environmental risk factors on the preclinical protein markers by comparing healthy twin siblings of patients with ulcerative colitis and healthy blood donors (baseline characteristics are presented in Table 1). Consistent with the findings in the preclinical cohort, an up-regulation of MMP10, CXCL9, CXCL11, and MCP-1 was seen among the healthy twin siblings (Supplementary Table 4). In addition, CCL11 seemed to be up-regulated (P = .08), but SLAMF1 could not be replicated (P = .22). When stratifying for zygosity and disease status, levels of these protein markers did not significantly differ between monozygotic and dizygotic healthy twin siblings (Supplementary Figure 8). The healthy twin siblings could, to some extent, also be separated from the healthy blood donors in a LOO cross-validation prediction model built from the 4 significantly replicated biomarker proteins (LOO AUC = 0.62, Figure 4).

Exploration of Protein Levels in Different Cohorts

Protein regulations were further compared between the 3 different cohorts (Figure 5*A*). Plasma samples had been collected in the preclinical cohort, and serum samples had been obtained in the other 2 cohorts. When the relative protein levels were normalized against the mean of the healthy controls of each of the cohorts and compared by visual inspection, a stepwise increase was seen across the cohorts with respect to the relative abundance of each of the 6 prediagnostic markers, except for MCP-1. Lowest levels were seen among healthy controls (adjusted to 1), healthy

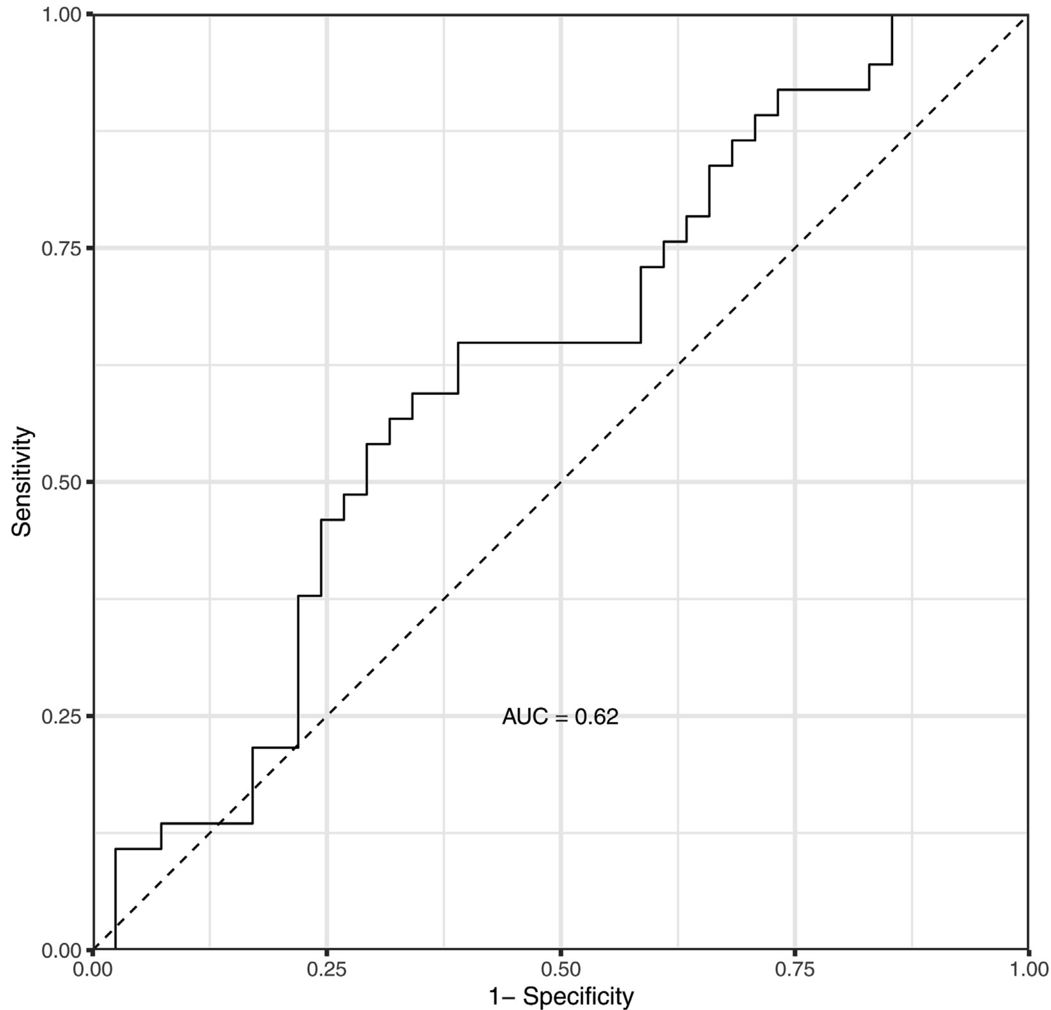


Figure 4. Receiver operating curve (ROC) analysis discriminating healthy twin siblings among pairs discordant for ulcerative colitis from healthy blood donors. The prediction model, built on the twin cohort with protein selection derived from the dysregulated proteins (n = 4) in the preclinical cohort, was evaluated using LOO cross-validation. The model separated healthy twin siblings among pairs discordant for ulcerative colitis from healthy blood donors (AUC = 0.62).

twin siblings and preclinical patients showed intermediate levels, and patients with incident ulcerative colitis demonstrated the highest relative abundances (Figure 5*B*–*G*).

Discussion

The preclinical inflammatory response of ulcerative colitis preceding symptoms is yet to be defined, and our current knowledge about how genetic and early environmental risk factors initiate and propagate an immune response that can lead to a subclinical inflammation is poor. Here, we provide a comprehensive characterization of the preclinical inflammatory profile by analyzing plasma samples from individuals who developed ulcerative colitis later in life and identify a prediagnostic protein signature consisting of MMP10, CXCL9, CCL11, SLAMF1, CXCL11, and MCP-1. The biological validity of the protein signature is demonstrated by the capacity of these proteins to separate ulcerative colitis from healthy subjects in an independent inception cohort (LOO AUC = 0.92). Based on analyses of twin pairs discordant for ulcerative colitis and comparisons of the healthy twin siblings with matched healthy blood donors, we have shown that the up-regulation of 4 of 6 proteins from the prediagnostic signature (MCP-1, CXCL9, CXCL11, and MMP10) is triggered by exposure to genetic and early environmental risk factors.

The concept of preclinical disease has been comprehensively explored in many other chronic immune-mediated diseases, including type 1 diabetes, rheumatoid arthritis, and systemic lupus erythematosus.[1] However, our current understanding of the initial triggers that ultimately lead to irreversible inflammation with tissue destruction and overt clinical IBD is still very preliminary, especially for ulcerative colitis. Consistent with the identification of autoantibodies among other chronic complex diseases,[28–30] previous work on preclinical IBD has focused mostly on the identification of serologic markers, including pANCA in ulcerative colitis.[7,8] These studies demonstrate that the adaptive immune response can be activated many years before the diagnosis of Crohn's disease or ulcerative colitis. However, the

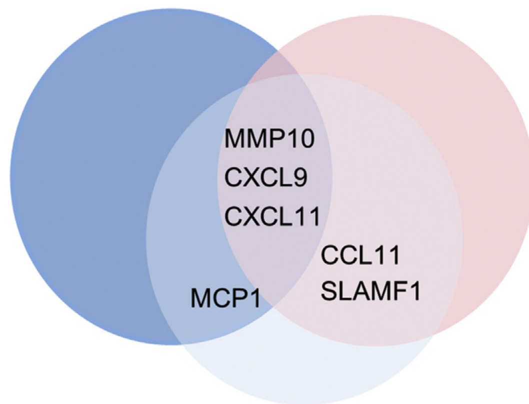


Figure 5. (*A*) Venn diagram showing the overlap of the significant protein regulations from the 3 different cohorts for the 6 proteins derived from the preclinical cohort. (*B–G*) Relative levels of the dysregulated proteins in cases with preclinical ulcerative colitis, patients with incident ulcerative colitis and healthy twin siblings of patients with ulcerative colitis vs healthy controls. The abundance of proteins among cases was normalized against the controls within each cohort. The level of each protein among controls was set to 1, to allow indirect comparisons. Mean relative levels were compared using *t* test and level of significance was annotated as $^*P < .05$; $^{**}P < .01$.

absence of an autoantibody and the fact that these serologic markers are only seen in approximately 19%–61% of cases challenges our attempts to define this preclinical phase of IBD, and this indicates the importance of a more detailed characterization of the subclinical inflammatory profile.

Recently, Lochhead et al[9] reported elevated levels of high-sensitive C-reactive protein and IL-6 in prediagnostic serum samples from individuals who were diagnosed with ulcerative colitis later in life. In our cohort, plasma IL-6 levels were below LOD for almost all cases, which might be caused by a less sensitive IL-6 assay. However, the results of our Ingenuity Pathway Analysis support the report by Lochhead et al because the analysis identified IL-6 among potential key inflammatory regulators. In the present study, we went beyond the previous report on IL-6 and C-reactive protein and performed comprehensive analyses of 92 key inflammatory proteins.

Our findings indicate that the up-regulation of MMP10 represents an early event in the pathogenesis of ulcerative colitis. Increased levels of serum MMP10 have recently been reported in prevalent ulcerative colitis.[31] Matrix metalloproteinases (MMPs) are peptidases that are categorized based on their specificity for extracellular matrix components, and MMP10 is classified as a stromelysin.[32] Up-regulated levels of stromelysin have been detected in inflamed segments of the colon from patients with ulcerative colitis.[33,34] On the contrary, increased levels of inflammatory cytokines and dysplastic lesions have been observed in the *MMP10* knockout mouse model when exposed to dextran sulfate sodium (DSS).[35] This disease-counteracting effect of MMP10 could seem surprising, as MMPs are traditionally implicated in an increase tissue breakdown and inflammation. Instead, the $MMP10^{-/-}$ mouse model might suggest that MMP10 promotes wound healing in the colon and resolution of disease. The observed preclinical up-regulation of MMP10 in plasma might indicate that endogenous pathways for wound healing are up-regulated several years before clinically overt ulcerative colitis to counteract disease progression and maintain mucosal homeostasis.

Eotaxin (CCL11) is a potent chemoattractant of monocytes, Th2 T cells, and polymorphonuclear leukocytes, predominantly eosinophils.[36] Several studies have shown elevated levels of eotaxin in both serum and tissue specimens from patients with established ulcerative colitis,[31,37–42] and treatment with anti-eotaxin antibodies has been shown to ameliorate disease in the DSS mouse model.[37] Our findings suggest that eosinophilic-driven inflammation represents an early element in the pathogenesis of ulcerative colitis and support the recent report on eosinophilic infiltration in prediagnostic colonic biopsies from individuals participating in the colorectal cancer screening program of the Basque Country (Spain) who were diagnosed with ulcerative colitis later in life.[43]

Up-regulation of CXCL9 and CXCL11 has been observed previously in inflamed colonic tissue specimens and blood from patients with ulcerative colitis, and *CXCL11* gene variants have also been associated with the disease.[31,44] Both chemokines are regulated by IFN-gamma and attract CXCR3-positive $CD4^{+}$ T cells and natural killer cells to the inflammatory site.[45] DSS-induced inflammation has also been shown to be abolished in the IFN-gamma knockout mice model.[46,47]

Our pathway analyses identified 12 upstream regulators, including IL-1β, TNF, IFN-gamma, IL-6, OSM, NFκB, and IL-4, as possible activators of the dysregulated inflammatory protein profile that we observed among patients with preclinical ulcerative colitis. These potential upstream regulators are all known activators of major inflammatory pathways and have previously been implicated in the immunopathogenesis of IBD,[2] even though they have never been associated with the preclinical phase of ulcerative colitis. IFN-gamma and IL-1β represent key cytokines of a Th1 immune response and have historically been associated with Crohn's disease, but recent studies have shown that the Th1/Th2 paradigm is an oversimplistic model of IBD.[48] Similar to IL-6, IFN-gamma, IL-4, and TNF were all part of our protein panel, but their levels were below LOD for most individuals in our study. This might indicate that the activation of the dysregulated proteins takes place in the gut mucosa and that the plasma levels of these upstream regulators are less relevant.[2] OSM has recently been shown to be up-regulated in inflamed intestinal mucosa from patients with IBD. Mucosal expression of OSM and its abundance in serum have been shown to predict response to anti-TNF treatment.[49,50] The transcription factor NFκB is a well-recognized central regulator of inflammation and reported to be dysregulated in IBD.[51] The up-regulation of NFκB is also in accordance with our previous twin study in which we observed an increased activation of NFκB in mucosal biopsies from healthy twin siblings of patients with IBD.[3] Jointly, these results indicate that NFκB activation represents an early step in the pathogenesis of ulcerative colitis.

A major advantage of this study, compared with most previous studies of preclinical disease,[6,7,9] is the use of prediagnostic samples from a population-based cohort, which increases the generalizability of the findings. The results are strengthened by our validation of the identified markers in an independent inception cohort and because we examined the influence of genetic and environmental risk factors on the dysregulated protein markers in the twin cohort. The study design, in which controls to cases with preclinical ulcerative colitis were matched by sex, age, and time of sampling, reduces the influence of bias due to potential differences in demographics. The inclusion criteria were similar to those used in the European Prospective Investigation Into Cancer and Nutrition study and Nurses' Health Study cohorts,[8,9] resulting in a study population with relatively high ages at diagnosis, which can potentially limit the possibility to generalize the findings to all patients with ulcerative colitis. Old-onset IBD has been associated with a longer doctor and patient delay, that is, presence of symptoms long before diagnosis compared with younger peers.[52] To minimize the risk that individuals with symptoms indicative of ulcerative colitis were included erroneously, we applied a 1-year washout period and excluded all individuals who had been in contact with health care because of gastrointestinal symptoms within 1 year before a prediagnostic plasma sample was obtained. Given the

indisputable involvement of the rectum, we believe that the risk of including individuals with already established but not yet recognized disease is less in a preclinical cohort of ulcerative colitis than Crohn's disease. The use of voluntary population-based cohorts and screening programs such as the Mammography Screening Project cohort might have introduced selection bias, due to the inclusion of individuals who are more activated toward seeking health care. However, the risk of differential bias was minimized, considering the close matching and the fact that this applies equally to individuals with preclinical ulcerative colitis and controls. The lack of multiple prediagnostic samples represents a major limitation and might have hampered our possibilities to identify significant correlations between individual markers or combination of markers and the period to diagnosis of ulcerative colitis. When investigating the relationship between protein markers and specific phenotypes of ulcerative colitis or clinical variables, we only observed a correlation with age. This finding is supported by previously reported correlations between age and 4 (CCL11, CXCL11, CXCL9, and MMP10) of the 6 markers.[53] The lack of significant correlations with other phenotypes of ulcerative colitis, including disease extent, should be interpreted with caution because the number of individuals within each category was low. The study is also limited by the lack of data on anti-neutrophil cytoplasmic antibodies (pANCA), the most specific marker of ulcerative colitis. Van Schaik et al[8] reported an increased presence of pANCA several years before the diagnosis of ulcerative colitis, but the predictive capacity of these antibodies was questioned in the recent PREDICTS study.[10]

BASIC AND TRANSLATIONAL AT

In contrast to our findings, Torres et al[10] were not able to identify a prediagnostic protein signature of ulcerative colitis in the PREDICTS study. This discrepancy might reflect differences in study populations and technologies. Serum samples from young, predominantly male, active-duty military personnel were examined in the PREDICTS study, and we analyzed plasma samples from a population-based screening cohort in which the mean age was 50 years and more than half of the participants were women. As opposed to the aptamer-based assay (SomaScan) that was used in the PREDICT study, we assessed proteins by using the PEA methodology. The fact that different protein-profiling platforms were used, hampers the possibility of comparing results across the 2 preclinical cohorts. Pietzner et al[54] recently reported a median correlations coefficient of 0.38 (range, –0.61 to 0.96) when examining 871 protein targets measured using the 2 different platforms. Several factors, including the fraction of measurement values below the detection limit of each assay, presence of transmembrane domains, glycosylation, protein length, and mass, might have contributed to the poor level of agreement. To our knowledge, there are no data in the public domain that provide information on the accuracy of the PEA- vs the aptamer-based assay, with respect to the 6 dysregulated protein markers that we identified. Among the dysregulated proteins, information on SOMAmers in the PREDICTS study was only available for MCP-1 (P13500). The observed AUC of MCP-1 (0.61) in our preclinical cohort seemed similar to the AUC (0.67–0.68) in the PREDICTS study. The fact that we were only able to analyze samples obtained before the diagnosis of ulcerative colitis in the preclinical cohort limits our study. In addition, we did not analyze follow-up samples after the diagnosis of ulcerative colitis in the inception cohort. Future studies should aim at longitudinal assessments of biological material that have been collected during all various phases of ulcerative colitis, that is, during the preclinical stage, at diagnosis, and also after the diagnosis. Another limitation of our study is the lack of data on potential concomitant immune-mediated diseases and family history of IBD. Also, we measured systemic levels of inflammatory proteins in a disease that primarily involves the colonic mucosa. Our findings might therefore reflect the “tip of the iceberg,” as most pathogenic mechanisms are probably restricted to the colonic mucosa. To change the natural history of ulcerative colitis, an intervention should ideally be initiated already during the preclinical stage and target the primary processes that lead to a clinical stage of the disease.[55] The need to start treatment at an earlier stage underscores the importance of identifying predictive signatures of ulcerative colitis. However, 4 of the 6 prediagnostic protein markers were also up-regulated in the twin cohort, that is, when comparing the healthy twin siblings with healthy blood donors. This observation and the modest capacity of our signature in predicting preclinical ulcerative colitis (AUC = 0.71) indicate that additional markers and risk factors need to be integrated before a predictive signature can be developed for future prospective prevention trials.

In conclusion, we identified an up-regulation of 6 plasma proteins indicating activation of both pro-inflammatory and tissue-repairing pathways several years before clinically overt ulcerative colitis. Our findings provide novel data on the early sequence of inflammatory events that eventually cause ulcerative colitis because activation of several of the proteins was triggered by exposure to genetic and environmental risk factors. Further knowledge about early disease mechanisms might disclose novel therapeutic targets and open avenues for disease prediction and interventions to delay or even stop progression to clinically manifest disease.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at http://doi.org/10.1053/j.gastro.2021.07.026.

References

1. Torres J, Burisch J, Riddle M, et al. Preclinical disease and preventive strategies in IBD: perspectives, challenges and opportunities. Gut 2016;65:1061–1069.
2. de Souza HS, Fiocchi C. Immunopathogenesis of IBD: current state of the art. Nat Rev Gastroenterol Hepatol 2016;13:13–27.

3. Zhulina Y, Hahn-Stromberg V, Shamikh A, et al. Subclinical inflammation with increased neutrophil activity in healthy twin siblings reflect environmental influence in the pathogenesis of inflammatory bowel disease. Inflamm Bowel Dis 2013;19:1725–1731.
4. Barker JM, Barriga KJ, Yu L, et al. Prediction of autoantibody positivity and progression to type 1 diabetes: Diabetes Autoimmunity Study in the Young (DAISY). J Clin Endocrinol Metab 2004;89:3896–3902.
5. del Puente A, Knowler WC, Pettitt DJ, et al. The incidence of rheumatoid arthritis is predicted by rheumatoid factor titer in a longitudinal population study. Arthritis Rheum 1988;31:1239–1244.
6. Choung RS, Princen F, Stockfisch TP, et al. Serologic microbial associated markers can predict Crohn's disease behaviour years before disease diagnosis. Aliment Pharmacol Ther 2016;43:1300–1310.
7. Israeli E, Grotto I, Gilburd B, et al. Anti-*Saccharomyces cerevisiae* and antineutrophil cytoplasmic antibodies as predictors of inflammatory bowel disease. Gut 2005; 54:1232–1236.
8. van Schaik FD, Oldenburg B, Hart AR, et al. Serological markers predict inflammatory bowel disease years before the diagnosis. Gut 2013;62:683–688.
9. Lochhead P, Khalili H, Ananthakrishnan AN, et al. Association between circulating levels of C-reactive protein and interleukin-6 and risk of inflammatory bowel disease. Clin Gastroenterol Hepatol 2016;14:818–824 e6.
10. Torres J, Petralia F, Sato T, et al. Serum biomarkers identify patients who will develop inflammatory bowel diseases up to 5 years before diagnosis. Gastroenterology 2020;159:96–104.
11. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, et al. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. Eur J Epidemiol 2009;24:659–667.
12. Salih A, Widbom L, Hultdin J, et al. Smoking is associated with risk for developing inflammatory bowel disease including late onset ulcerative colitis: a prospective study. Scand J Gastroenterol 2018;53:173–178.
13. Norberg M, Wall S, Boman K, et al. The Vasterbotten Intervention Programme: background, design and implications. Glob Health Action 2010;3.
14. Widbom L, Schneede J, Midttun O, et al. Elevated plasma cotinine is associated with an increased risk of developing IBD, especially among users of combusted tobacco. PLoS One 2020;15:e0235536.
15. Lichtenstein P, De Faire U, Floderus B, et al. The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. J Intern Med 2002; 252:184–205.
16. **Dignass A, Eliakim R**, Magro F, et al. Second European evidence-based consensus on the diagnosis and management of ulcerative colitis part 1: definitions and diagnosis. J Crohns Colitis 2012;6:965–990.
17. Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. Can J Gastroenterol 2005;19(Suppl A): 5A–36A.
18. Kalla R, Adams AT, Bergemalm D, et al. Serum proteomic profiling at diagnosis predicts clinical course, and need for intensification of treatment in inflammatory bowel disease. J Crohns Colitis 2021;15:699–708.
19. Halfvarson J. Genetics in twins with Crohn's disease: less pronounced than previously believed? Inflamm Bowel Dis 2011;17:6–12.
20. Halfvarson J, Bodin L, Tysk C, et al. Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. Gastroenterology 2003;124:1767–1773.
21. Drobin K, Assadi G, Hong MG, et al. Targeted analysis of serum proteins encoded at known inflammatory bowel disease risk loci. Inflamm Bowel Dis 2019;25:306–316.
22. Fransen K, Franzen P, Magnuson A, et al. Polymorphism in the retinoic acid metabolizing enzyme CYP26B1 and the development of Crohn's Disease. PLoS One 2013;8: e72739.
23. **Assarsson E, Lundberg M**, Holmquist G, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. PLoS One 2014;9:e95192.
24. Breheny P, Huang J. Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. Ann Appl Stat 2011;5:232–253.
25. Mow WS, Vasiliauskas EA, Lin YC, et al. Association of antibody responses to microbial antigens and complications of small bowel Crohn's disease. Gastroenterology 2004;126:414–424.
26. Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008;28:26; 2008.
27. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77.
28. Deane KD, O'Donnell CI, Hueber W, et al. The number of elevated cytokines and chemokines in preclinical seropositive rheumatoid arthritis predicts time to diagnosis in an age-dependent manner. Arthritis Rheum 2010; 62:3161–3172.
29. van de Sande MG, de Hair MJ, van der Leij C, et al. Different stages of rheumatoid arthritis: features of the synovium in the preclinical phase. Ann Rheum Dis 2011; 70:772–777.
30. Verge CF, Gianani R, Kawasaki E, et al. Prediction of type I diabetes in first-degree relatives using a combination of insulin, GAD, and ICA512bdc/IA-2 autoantibodies. Diabetes 1996;45:926–933.
31. Andersson E, Bergemalm D, Kruse R, et al. Subphenotypes of inflammatory bowel disease are characterized by specific serum protein profiles. PLoS One 2017;12:e0186142.
32. Egeblad M, Werb Z. New functions for the matrix metalloproteinases in cancer progression. Nat Rev Cancer 2002;2:161–174.
33. Makitalo L, Piekkala M, Ashorn M, et al. Matrix metalloproteinases in the restorative proctocolectomy pouch

of pediatric ulcerative colitis. World J Gastroenterol 2012;18:4028–4036.
34. O'Sullivan S, Gilmer JF, Medina C. Matrix metalloproteinases in inflammatory bowel disease: an update. Mediators Inflamm 2015;2015:964131.
35. Koller FL, Dozier EA, Nam KT, et al. Lack of MMP10 exacerbates experimental colitis and promotes development of inflammation-associated colonic dysplasia. Lab Invest 2012;92:1749–1759.
36. Filippone RT, Sahakian L, Apostolopoulos V, et al. Eosinophils in inflammatory bowel disease. Inflamm Bowel Dis 2019;25:1140–1151.
37. **Adar T, Shteingart S**, Ben-Ya'acov A, et al. The importance of intestinal eotaxin-1 in inflammatory bowel disease: new insights and possible therapeutic implications. Dig Dis Sci 2016;61:1915–1924.
38. Chen W, Paulus B, Shu D, et al. Increased serum levels of eotaxin in patients with inflammatory bowel disease. Scand J Gastroenterol 2001;36:515–520.
39. Coburn LA, Horst SN, Chaturvedi R, et al. High-throughput multi-analyte Luminex profiling implicates eotaxin-1 in ulcerative colitis. PLoS One 2013;8:e82300.
40. Korolkova OY, Myers JN, Pellom ST, et al. Characterization of serum cytokine profile in predominantly colonic inflammatory bowel disease to delineate ulcerative and Crohn's colitides. Clin Med Insights Gastroenterol 2015; 8:29–44.
41. Manousou P, Kolios G, Valatas V, et al. Increased expression of chemokine receptor CCR3 and its ligands in ulcerative colitis: the role of colonic epithelial cells in in vitro studies. Clin Exp Immunol 2010;162:337–347.
42. Mir A, Minguez M, Tatay J, et al. Elevated serum eotaxin levels in patients with inflammatory bowel disease. Am J Gastroenterol 2002;97:1452–1457.
43. Rodriguez-Lago I, Ramirez C, Merino O, et al. Early microscopic findings in preclinical inflammatory bowel disease. Dig Liver Dis 2020;52:1467–1472.
44. Schroepf S, Kappler R, Brand S, et al. Strong overexpression of CXCR3 axis components in childhood inflammatory bowel disease. Inflamm Bowel Dis 2010; 16:1882–1890.
45. Singh UP, Singh R, Singh S, et al. CXCL10+ T cells and NK cells assist in the recruitment and activation of CXCR3+ and CXCL11+ leukocytes during Mycobacteria-enhanced colitis. BMC Immunol 2008; 9:25.
46. Chami B, Yeung AW, van Vreden C, et al. The role of CXCR3 in DSS-induced colitis. PLoS One 2014;9: e101622.
47. Ito R, Shin-Ya M, Kishida T, et al. Interferon-gamma is causatively involved in experimental inflammatory bowel disease in mice. Clin Exp Immunol 2006;146:330–338.
48. **Mavroudis G, Magnusson MK**, Isaksson S, et al. Mucosal and systemic immune profiles differ during early and late phases of the disease in patients with active ulcerative colitis. J Crohns Colitis 2019;13:1450–1458.
49. **West NR, Hegazy AN**, Owens BMJ, et al. Oncostatin M drives intestinal inflammation and predicts response to tumor necrosis factor-neutralizing therapy in patients with inflammatory bowel disease. Nat Med 2017;23:579–589.
50. Bertani L, Fornai M, Fornili M, et al. Serum oncostatin M at baseline predicts mucosal healing in Crohn's disease patients treated with infliximab. Aliment Pharmacol Ther 2020;52:284–291.
51. McDaniel DK, Eden K, Ringel VM, et al. Emerging roles for noncanonical NF-kappaB signaling in the modulation of inflammatory bowel disease pathobiology. Inflamm Bowel Dis 2016;22:2265–2279.
52. Arnott I, Rogler G, Halfvarson J. The management of inflammatory bowel disease in elderly: current evidence and future perspectives. Inflamm Intest Dis 2018;2:189–199.
53. Larsson A, Carlsson L, Gordh T, et al. The effects of age and gender on plasma levels of 63 cytokines. J Immunol Methods 2015;425:58–61.
54. Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Cross-platform proteomics to advance genetic prioritisation strategies. Preprint. Posted Online March 19, 2021. bioRxiv 2021.03.18.435919. https://doi.org/10.1101/2021.03.18.435919
55. Torres J, Halfvarson J, Rodriguez-Lago I, et al. Results of the Seventh Scientific Workshop of ECCO: Precision medicine in IBD- prediction and prevention of inflammatory bowel disease. J Crohns Colitis 2021;Mar 17jjab048.

Author names in bold designate shared co-first authorship.

Received December 3, 2020. Accepted July 12, 2021.

Correspondence
Address correspondence to: Daniel Bergemalm, MD, PhD, Örebro University Hospital, 70185 Örebro, Sweden. e-mail: daniel.bergemalm@regionorebrolan.se.

Acknowledgments
The authors thank the Västerbotten Intervention Program participants, Department of Biobank Research at Umeå University (https://www.umu.se/en/biobank-research-unit/), Northern Sweden Health and Disease Cohort, and Västerbotten County Council for delivering data and blood samples, and acknowledge the contribution of Biobank Sweden.

CRediT Authorship Contributions
Daniel Bergemalm, MD, PhD (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Resources: Supporting; Writing – original draft: Lead). Erik Andersson, M.D (Conceptualization: Supporting; Data curation: Supporting; Formal analysis: Equal; Writing – review & editing: Equal). Johan Hultdin, MD, PhD (Conceptualization: Equal; Funding acquisition: Equal; Resources: Lead; Writing – review & editing: Equal). Carl Eriksson, MD, PhD (Data curation: Equal; Formal analysis: Equal; Writing – review & editing: Equal). Stephen Rush, PhD (Formal analysis: Equal; Writing – review & editing: Supporting). Rahul Kalla, MD, PhD (Resources: Equal; Writing – review & editing: Equal). Alex T Adams, PhD (Resources: Equal; Writing – review & editing: Equal). Åsa V Keita, PhD (Resources: Equal; Writing – review & editing: Supporting). Mauro D'Amato, PhD (Writing – review & editing: Equal). Fernando Gomollon, MD, PhD (Resources: Equal; Writing – review & editing: Equal). Jørgen Jahnsen, MD, PhD (Resources: Equal; Writing – review & editing: Equal). Petr Ricanek, MD, PhD (Resources: Equal; Writing – review & editing: Equal). Jack Satsangi, MD, PhD (Resources: Equal; Writing – review & editing: Equal). Dirk Repsilber, PhD (Data curation: Equal; Formal analysis: Equal; Resources: Equal; Writing – review & editing: Equal). Pontus Karling, MD, PhD (Conceptualization: Equal; Data curation: Equal; Resources: Equal; Writing – review & editing: Equal). Jonas Halfvarson, MD, PhD (Conceptualization: Equal; Data curation: Equal; Formal analysis: Equal; Funding acquisition: Lead; Resources: Equal; Writing – review & editing: Lead).

IBD Character Consortium
Ian D. Arnott, MD, Monica Bayes, PhD, Ferdinando Bonfiglio, PhD, Ray K. Boyapati, MD, Adam Carstens, MD, Christina Casén, MSc, Ewa Ciemniejewska, MSc, Fredrik A. Dahl, PhD, Trond Espen Detlie, MD, Hazel E. Drummond, BSc, Gunn S. Ekeland, MSc, Daniel Ekman, MSc, Anna B. Frengen, PhD, Mats Gullberg, PhD, Ivo G. Gut, PhD, Marta Gut, PhD, Simon C. Heath, PhD, Fredrik Hjelm, PhD, Henrik Hjortswang, MD, PhD, Gwo-Tzer Ho, PhD, Daisy Jonkers, PhD, Johan Söderholm, MD, PhD, Nicholas A. Kennedy, MBBS, PhD, FRACP, Charles W. Lees, PhD, Torbjørn Lindahl, MSc, Mårten Lindqvist, PhD, Angelika Merkel, PhD, Eddie Modig, Bsc, Aina E. F. Moen, PhD, Hilde Nilsen, PhD, Elaine R. Nimmo, PhD, Colin L. Noble, MD, Niklas Nordberg, PhD, Kate R. O'Leary, MSc, Anette Ocklind, PhD, Christine Olbjørn, MD, Erik Pettersson, PhD, Marieke Pierik, MD, PhD, Dominique Poncelet, PhD, Céline Sabatel, PhD, Renaud Schoemans, PhD, Alan G. Shand, MD, Janne Sølvernes, MS, Mikael Sundell, BSc, Tone M. Tannæs, PhD, Leif Törkvist, MD, PhD, Anne-Clémence Veillard, PhD, Nicholas T. Ventham, MRCS(Eng) PhD, MBBS, David C. Wilson, MD, MRCPCH, Panpan You, MS.

Pontus Karling and Jonas Halfvarson contributed equally to this work.

Conflicts of interest
These authors disclose the following: From the IBD Character Consortium: Fredrik Hjelm, Niklas Nordberg, and Anette Ocklind were employees at Olink Proteomics, Uppsala, Sweden at the time of data collection. The remaining authors disclose no conflicts.

Funding
Västerbotten County Council funded the Västerbotten Intervention Program. Biobank Sweden was supported by the Swedish Research Council (VR 2017-00650). The study has been funded by the following EU FP7 grant: IBD-CHARACTER (contract 2858546), The Bengt Ihre research foundation to Daniel Bergemalm, the Örebro University Hospital Research Foundation, grant number OLL-709831 to Daniel Bergemalm; numbers OLL-936004, OLL-890291, OLL-790011 and OLL-723021 to Jonas Halfvarson; the Swedish Foundation For Strategic Research, grant number RB13-016 to Jonas Halfvarson; and Swedish Research Council, grant number 2020-02021 to Jonas Halfvarson. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Patients and Cohorts

Cohort of preclinical ulcerative colitis. In the Västerbotten Intervention Program cohort, all residents of Västerbotten County have been invited to participate at the ages of 30, 40, 50, and 60 years since 1985. The participation rate is approximately 75%, and 66,000 individuals were included by 2011, when cases were identified. At enrollment, the participants completed a self-administered questionnaire to collect demographic, medical, and lifestyle information, as well as a separate food frequency questionnaire. During a medical examination, blood samples, including heparinized plasma samples, were collected and processed by centrifugation and separation and frozen in aliquots at –80°C within 1 hour of collection.

To ensure that no individual had manifest ulcerative colitis at inclusion in the Northern Sweden Health and Disease Study register, when a plasma sample was obtained, we strictly applied a 1-year washout period and excluded all individuals who had been in contact with health care because of gastrointestinal symptoms within 1 year before a prediagnostic plasma sample was obtained. For each individual with preclinical ulcerative colitis, 2 control subjects, matched by age, sex, area of residence, and date of inclusion (plasma sampling), were randomly assigned from the Northern Sweden Health and Disease Study register. To ensure an equal follow-up period, the date for collection of plasma samples was not allowed to differ more than 8 weeks between cases with preclinical ulcerative colitis and controls.

Inception cohort of treatment-naïve ulcerative colitis. Besides patients with IBD, healthy individuals with no history of chronic gastrointestinal disease were also included as part of the cohort. In the current study, we only included patients who were treatment-naïve at the diagnosis of ulcerative colitis. Blood samples were collected before initiation of therapy. The serum was separated after centrifugation at room temperature within 2 hours at $2000 \times g$ for 10 minutes. All serum samples were stored as aliquots and frozen within 2 hours at −80°C until being shipped.

Cohort of twin pairs discordant for ulcerative colitis. To explore the impact of genetic and shared environmental risk factors on the proteins associated with preclinical ulcerative colitis, we compared the abundance of these proteins among healthy twin siblings of patients with ulcerative colitis and healthy external controls. All twins with a minimum of 1 inpatient visit listing a diagnosis of IBD, according to the International Classification of Diseases codes, were identified by linking the Swedish Twin Registry with the Swedish National Patient Registry. Identified twins and their twin siblings were asked to respond to a questionnaire regarding general gastrointestinal symptoms, including a possible diagnosis of IBD. All medical notes were reviewed for verification of diagnosis of IBD and for determination of disease subtype, according to the Montreal Classification.[16,17] Twin pairs with a confirmed diagnosis of IBD were invited to take part, provide biological samples, and undergo an endoscopy. Blood samples were collected and the serum was separated after centrifugation at $2400 \times g$ for 6 minutes at room temperature. All serum samples were stored as aliquots and frozen at −80°C. In the current study, we only included mono- and dizygotic twin pairs of the same sex and discordant for ulcerative colitis.

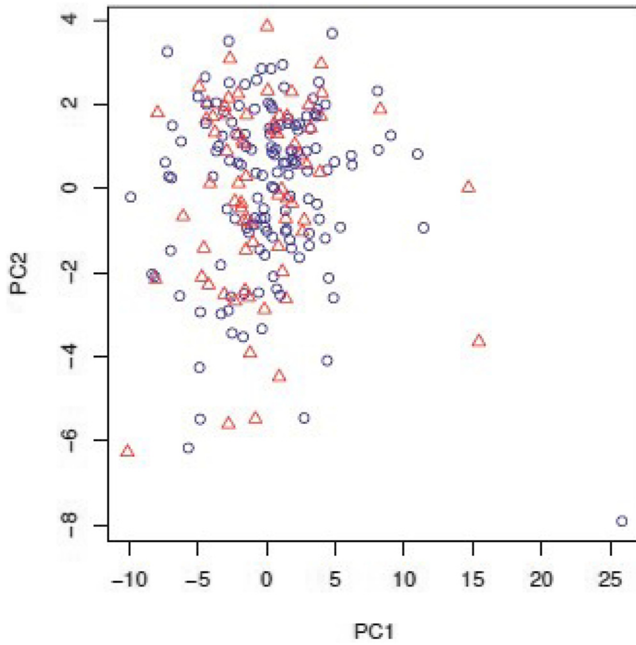
Protein Analysis

A PEA was performed in which pairs of different antibodies toward the same antigen are used. When both antibodies bind to the same antigen in close proximity, attached oligonucleotides hybridize. The oligonucleotide templates are extended and amplified using polymerase chain reaction (96.96, Dynamic Array IFC, Fluidigm Biomark) on a Biomark HD Instrument. The analyses were performed at the Clinical Biomarkers Facility, Science for Life Laboratory, Uppsala, Sweden. Data were preprocessed with the Olink Wizard for GenEx (Multid Analyses, Sweden) to generate normalized log2 values corresponding to relative protein quantities. These relative concentrations are presented as normalized protein expressions (NPX) on log2 scale. Because samples were randomly distributed over several chip runs, NPX values for each test were normalized against intra- and interplate controls. Proteins were excluded if signals were below the LOD in >80% of the samples in both cases and controls. In the remaining proteins, values below LOD were substituted with a fixed value set to $LOD/\sqrt{2}$.

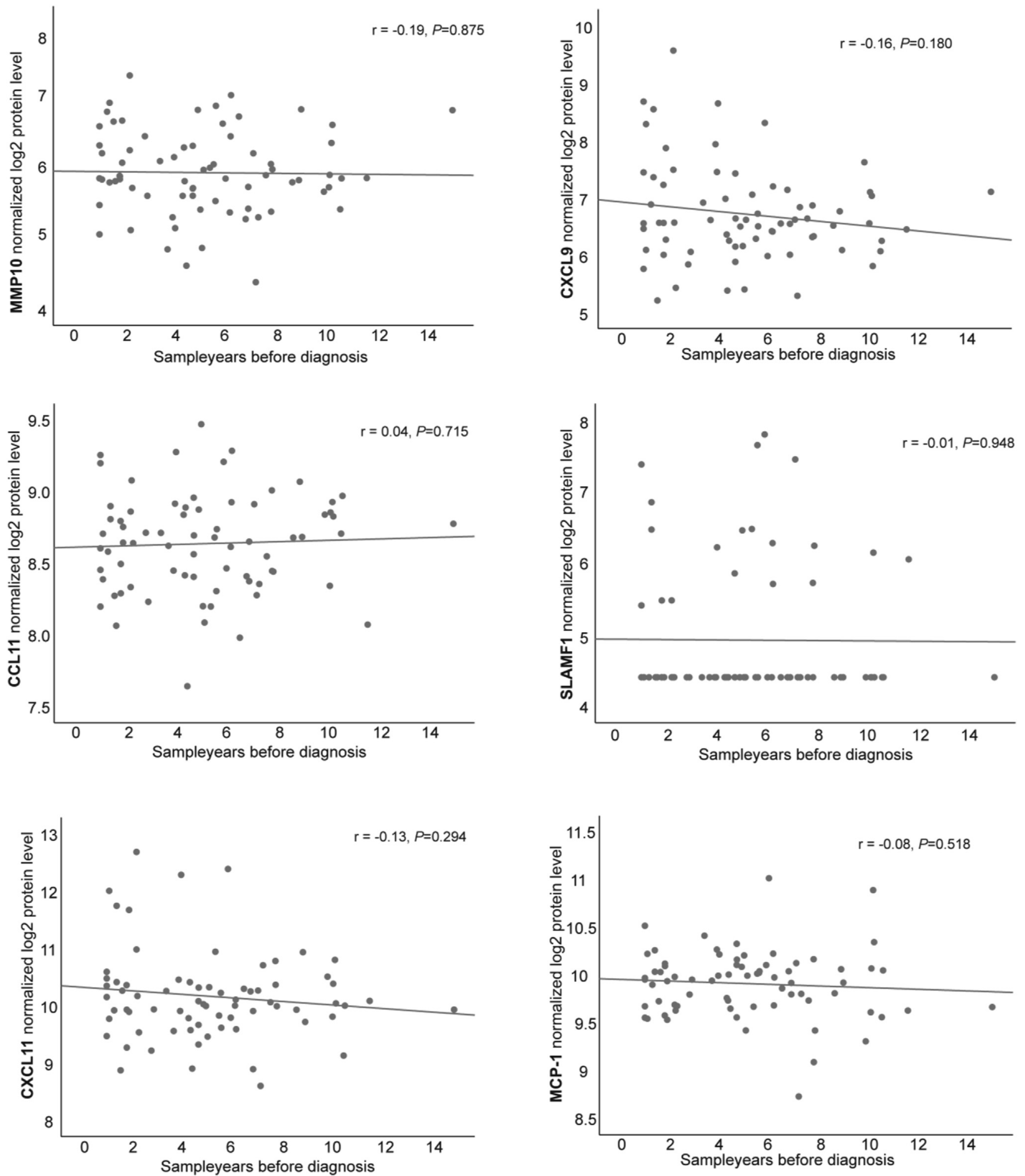
Data Analysis

Comparison of protein levels across cohorts. To compare relative protein levels between different cohorts, we adjusted the relative levels of the significant protein markers against the mean for each individual protein in the matched healthy controls within each cohort. The differences in means were then converted from the log2 scale (2^{NPX}) to levels of relative concentrations, that is, to fold-changes of relative concentrations and compared by *t* tests.

Pathway analyses. We used the Ingenuity Pathway Analysis software to identify canonical pathways and upstream regulators of dysregulated preclinical protein markers (21/04/2021). The *P* value (calculated by right-sided Fisher exact test) of the model indicates whether there is a significant overlap between the dataset of dysregulated proteins and proteins or genes that are known to be regulated by 1 or more specific upstream regulators. The *z* score is calculated to indicate the activation state, where an absolute *z* score of ≥2.0 is considered relevant. A positive *z* score predicts activation of the regulator.



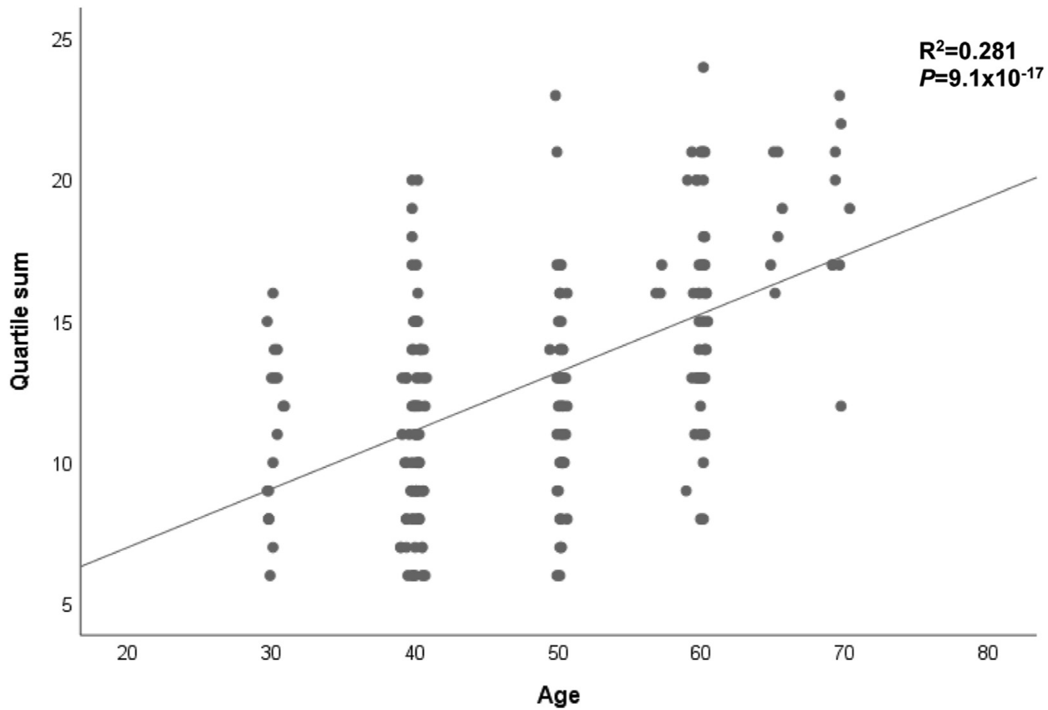
Supplementary Figure 1. Principal component analysis of the preclinical cohort. Little separation could be seen between samples obtained from individuals with preclinical ulcerative colitis (*red triangles*) and healthy controls (*blue circles*) on visual assessment.



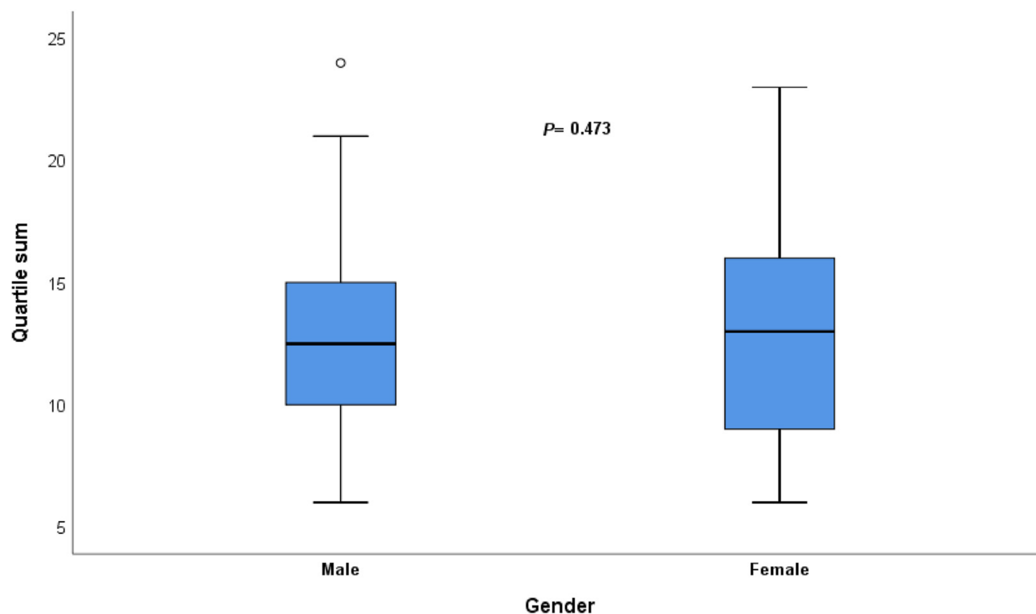
Supplementary Figure 2. Correlation with time to diagnosis. The 6 prediagnostic protein markers were plotted against time to diagnosis. There were no significant correlations using Pearson correlation test.

Supplementary Figure 3. Distribution of quartile sum for the 6 prediagnostic protein markers among the 72 preclinical cases. Individual protein levels of the 6 significant prediagnostic protein markers were divided in quartiles and substituted with scores 1–4. Quartile sum was calculated for each individual of the cohort where the lowest possible protein levels for all markers result in a score of 6 and the highest in 24. The *histogram* shows distribution of quartile sums for the preclinical cases.

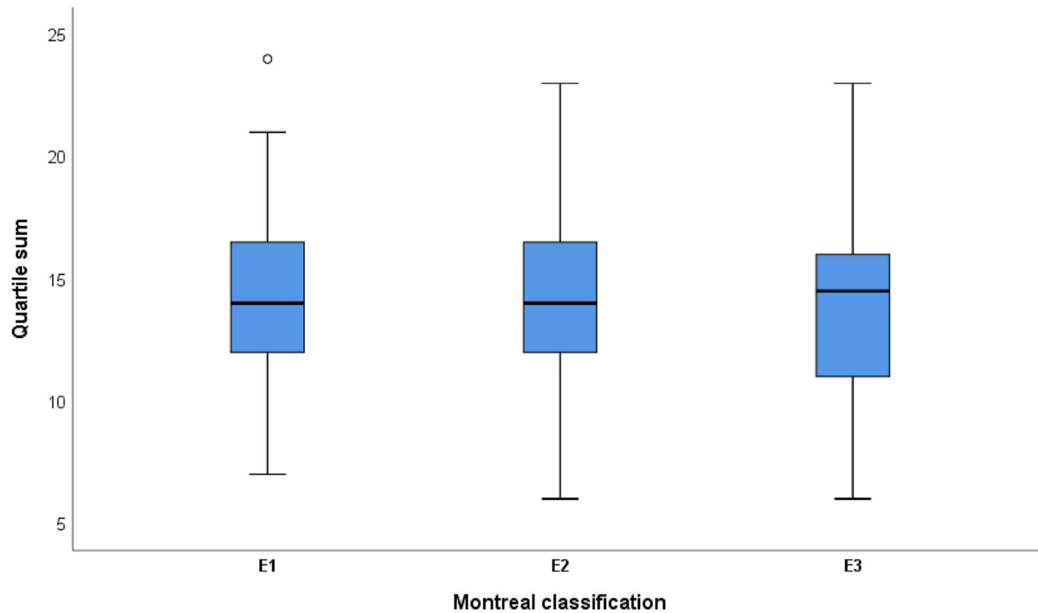
Supplementary Figure 4. Quartile sum, correlation with time to diagnosis. The quartile sum of the 6 prediagnostic protein markers was plotted against time to diagnosis for all preclinical ulcerative colitis cases. There was no significant correlation for quartile sums, similar to findings from individual markers.



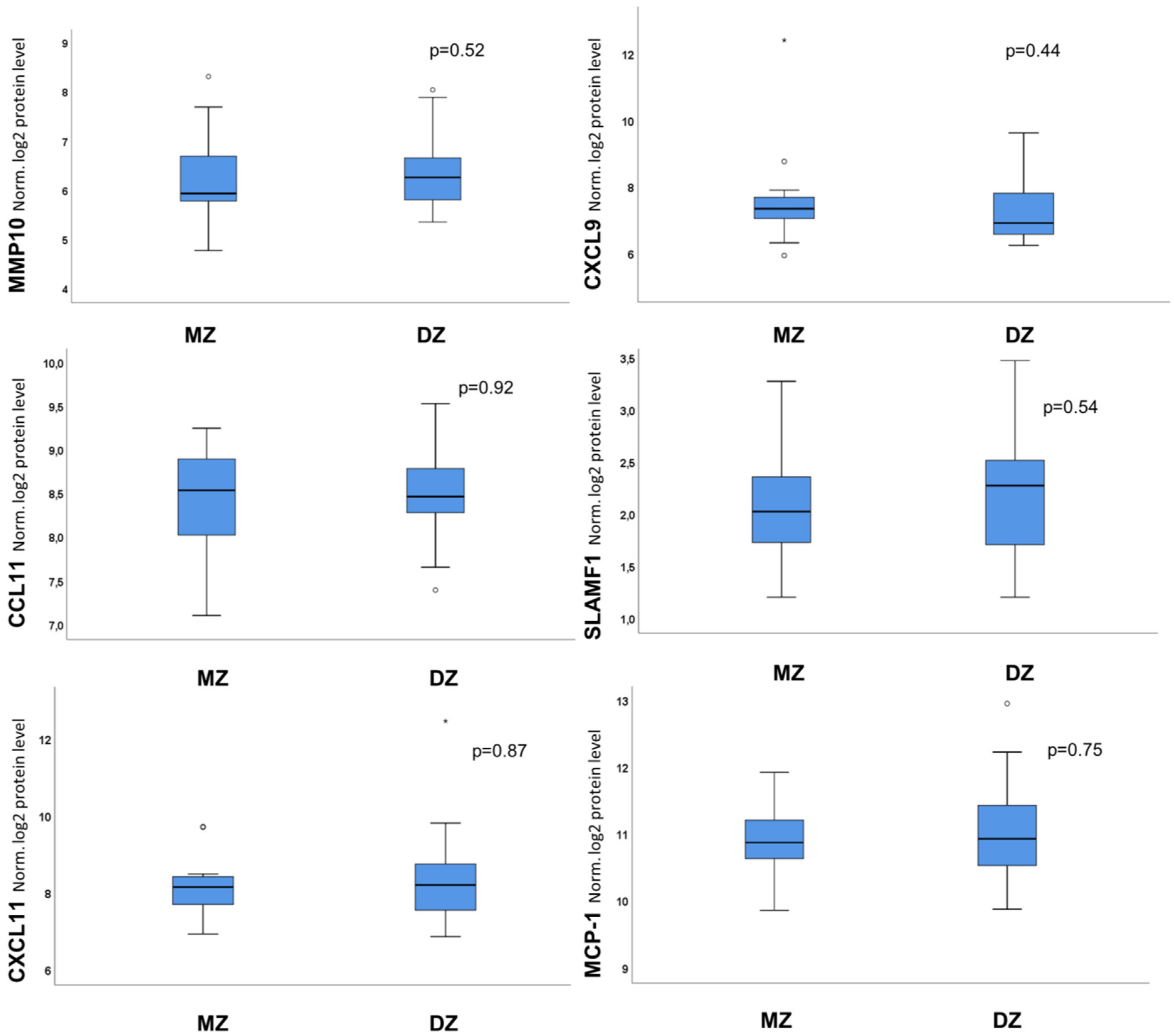
Supplementary Figure 5. Correlation with age and quartile sum. Quartile sums of the 6 prediagnostic markers were plotted against age for all individuals in the preclinical cohort. There was a strong correlation with age as expected for some of these markers. Individually, all markers except MMP10 were significantly increasing with age in both patients and controls (not shown).



Supplementary Figure 6. Quartile sum by sex. There was no significant difference in quartile sum of the 6 prediagnostic protein markers and sex.



Supplementary Figure 7. Quartile sum and disease extent at diagnosis. When analyzed individually, none of the 6 prediagnostic markers were differently regulated dependent on disease extent (not shown). The quartile sums of the 6 protein markers were homogenously distributed in patients with different disease extent at diagnosis (P = .954; E3; E1/2).



Supplementary Figure 8. *Box plots* of levels of the prediagnostic markers in monozygotic vs dizygotic healthy twin siblings of patients with ulcerative colitis. There were no significant differences in any of the protein markers (Mann-Whitney U test).

Supplementary Table 2. Correlations (Spearman ρ) Between Significant Protein Markers

Spearman	CXCL9	CXCL11	CCL11	MMP10	MCP-1	SLAMF1
CXCL9	—	.739[a]	.287[a]	.249[a]	.245[a]	.257[a]
CXCL11	.739[a]	—	.326[a]	.200[a]	.283[a]	.293[a]
CCL11	.287[a]	.326[a]	—	.127	.396[a]	.270[a]
MMP10	.249[a]	.200[a]	.127	—	.132	.110
MCP-1	.245[a]	.283[a]	.396[a]	.132	—	.238[a]
SLAMF1	.257[a]	.293[a]	.270[a]	.110	.238[a]	—

[a]Correlation is significant at the .01 level.

Supplementary Table 3. Comparison of the Relative Abundance of Prediagnostic Proteins Among Treatment-Naïve Patients With Incident Ulcerative Colitis and Healthy Controls in the IBD Character Cohort

Protein	Symbol	Uniprot ID	*P* value
Stromelysin-2	MMP10	P09238	2.5×10^{-7}
C-X-C motif chemokine 9	CXCL9	Q07325	4.3×10^{-7}
Eotaxin	CCL11	P51671	.0028
Signaling lymphocytic activation molecule	SLAMF1	Q13291	.011
C-X-C motif chemokine 11	CXCL11	O14625	7.6×10^{-4}
Monocyte chemotactic protein 1	MCP-1	P13500	.787

NOTE. The relative abundance of the 6 prediagnostic proteins was measured in serum samples from treatment-naïve patients with incident ulcerative colitis and healthy controls. Comparisons were performed by logistic regression, adjusting for age, sex, and smoking status.

Supplementary Table 4. Comparison of the Relative Abundance of Prediagnostic Proteins Among Healthy Twin Siblings of Patients With Ulcerative Colitis and Blood Donors

Protein	Symbol	Uniprot ID	*P* value
Stromelysin-2	MMP10	P09238	.045
C-X-C motif chemokine 9	CXCL9	Q07325	.013
Eotaxin	CCL11	P51671	.078
Signaling lymphocytic activation molecule	SLAMF1	Q13291	.222
C-X-C motif chemokine 11	CXCL11	O14625	.007
Monocyte chemotactic protein 1	MCP-1	P13500	.022

NOTE. The relative abundance of the 6 prediagnostic proteins was measured in serum samples from healthy twin siblings among pairs discordant for ulcerative colitis and matched healthy blood donors. Comparisons were performed by logistic regression, adjusting for age, sex, and smoking status. As data on smoking were missing for the healthy blood donors, smoking habits were randomly assigned to these controls, according to frequencies in the healthy twin siblings.