

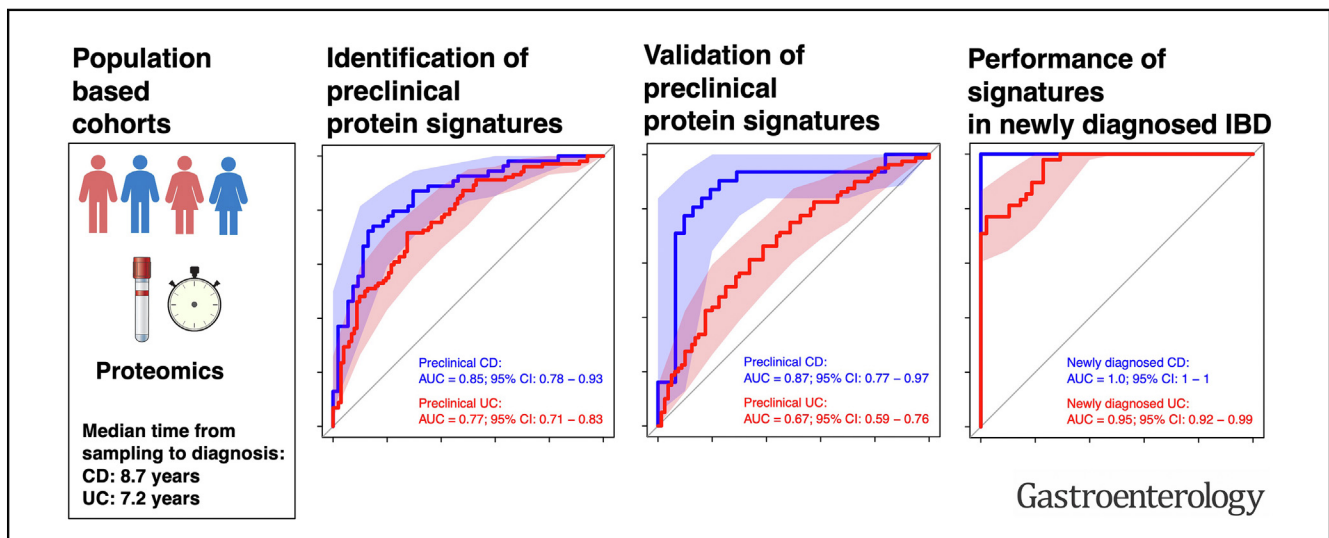
INFLAMMATORY BOWEL DISEASE

Preclinical Protein Signatures of Crohn's Disease and Ulcerative Colitis: A Nested Case-Control Study Within Large Population-Based Cohorts



Olle Grännö,¹ Daniel Bergemalm,² Benita Salomon,³ Carl Mårten Lindqvist,³ Charlotte R. H. Hedin,^{4,5} Marie Carlson,⁶ Katharina Dannenberg,² Erik Andersson,² Åsa V. Keita,⁷ Maria K. Magnusson,⁸ Carl Eriksson,^{2,9} Vivekananda Lanka,¹⁰ BIOIBD consortium, Patrik K. E. Magnusson,¹⁰ Mauro D'Amato,^{11,12,13} Lena Öhman,⁸ Johan D. Söderholm,^{7,14} Johan Hultdin,¹⁵ Robert Kruse,¹⁶ Yang Cao,^{17,18} Dirk Repsilber,³ Olof Grip,¹⁹ Pontus Karling,²⁰ and Jonas Halfvarson²

¹Department of Laboratory Medicine, Clinical Microbiology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden; ²Department of Gastroenterology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden; ³Faculty of Medicine and Health, School of Medical Sciences, Örebro University, Örebro, Sweden; ⁴Gastroenterology Unit, Department of Gastroenterology, Dermatovenereology and Rheumatology, Karolinska University Hospital, Stockholm, Sweden; ⁵Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden; ⁶Department of Medical Sciences, Gastroenterology Research Group, Uppsala University, Uppsala, Sweden; ⁷Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden; ⁸Department of Microbiology and Immunology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Göteborg, Sweden; ⁹Clinical Epidemiology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden; ¹⁰Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; ¹¹Department of Medicine and Surgery, LUM University, Casamassima, Italy; ¹²Gastrointestinal Genetics Lab, CIC BioGUNE-BRTA, Derio, Spain; ¹³Ikerbasque, Basque Foundation for Science, Bilbao, Spain; ¹⁴Department of Surgery, Linköping University, Linköping, Sweden; ¹⁵Department of Medical Biosciences, Clinical Chemistry, Umeå University, Umeå, Sweden; ¹⁶Department of Clinical Research Laboratory, Faculty of Medicine and Health, Örebro University, Örebro, Sweden; ¹⁷Clinical Epidemiology and Biostatistics, School of Medical Sciences, Faculty of Medicine and Health, Örebro University, Örebro, Sweden; ¹⁸Unit of Integrative Epidemiology, Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden; ¹⁹Department of Gastroenterology, Skåne University Hospital, Malmö, Sweden; and ²⁰Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden



BACKGROUND & AIMS: Biomarkers are needed to identify individuals at elevated risk of inflammatory bowel disease. This study aimed to identify protein signatures predictive of inflammatory bowel disease. **METHODS:** Using large population-based cohorts ($n \geq 180,000$), blood samples were obtained from individuals who later in life were diagnosed with inflammatory bowel disease and compared with age and sex-

matched controls, free from inflammatory bowel disease during follow-up. A total of 178 proteins were measured on Olink platforms. We used machine-learning methods to identify protein signatures of preclinical disease in the discovery cohort ($n = 312$). Their performance was validated in an external preclinical cohort ($n = 222$) and assessed in an inception cohort ($n = 144$) and a preclinical twin cohort ($n = 102$).

RESULTS: In the discovery cohort, a signature of 29 proteins differentiated preclinical Crohn's disease (CD) cases from controls, with an area under the curve (AUC) of 0.85. Its performance was confirmed in the preclinical validation (AUC = 0.87) and the inception cohort (AUC = 1.0). In preclinical samples, downregulated (but not upregulated) proteins related to gut barrier integrity and macrophage functionality correlated with time to diagnosis of CD. The preclinical ulcerative colitis signature had a significant, albeit lower, predictive ability in the discovery (AUC = 0.77), validation (AUC = 0.67), and inception cohorts (AUC = 0.95). The preclinical signature for CD demonstrated an AUC of 0.89 when comparing twins with preclinical CD with matched external healthy twins, but its predictive ability was lower (AUC = 0.58; $P = .04$) when comparing them with their healthy twin siblings, that is, when accounting for genetic and shared environmental factors. **CONCLUSION:** We identified protein signatures for predicting a future diagnosis of CD and ulcerative colitis, validated across independent cohorts. In the context of CD, the signature offers potential for early prediction.

Keywords: Preclinical Disease; Inflammatory Bowel Disease; Crohn's Disease; Ulcerative Colitis; Proteomics.

Inflammatory bowel disease (IBD), including Crohn's disease (CD) and ulcerative colitis (UC), is caused by complex interactions between a range of genetic and environmental risk factors, resulting in a dysregulated immune response. Recent studies have shown that a diagnosis of IBD is preceded by a preclinical period, characterized by systemic subclinical inflammation,^{1,2} disrupted barrier function,³ development of antibodies to microbial antigens, and potentially autoantibodies.^{4,5}

Overall, 30% to 40% of patients with CD have developed disease complications (eg, intestinal fistulas and strictures) already at diagnosis.⁶ Moreover, some patients with UC require colectomy during the index flare, demonstrating the need for earlier diagnosis and intervention.⁷ Many patients with IBD have an incomplete response to medication, and despite recent advances in the drug armamentarium, none of the existing therapies can reverse the progressive nature of IBD. Predictive biomarkers of future clinical onset of active IBD could detect the disease during "a window of opportunity" when the immune dysregulation is potentially reversible.^{8,9} Although no interventional studies have targeted the preclinical disease phase of IBD, such initiatives have been conducted in other immune-mediated diseases.^{10,11} Notably, a recent study of type 1 diabetes demonstrated the clinical utility of robust preclinical biomarkers in complex diseases, using autoantibodies for risk stratification in a cohort of first-degree relatives. Treatment of participants with ≥ 2 diabetes-related autoantibodies with teplizumab postponed disease onset by 2 years compared with placebo.¹²

Recent studies analyzing preclinical disease samples assembled from a US Army biorepository indicate that screening for future disease onset may also be achievable for IBD.^{4,13} A panel of autoantibodies and serum proteins

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

There is a need for validated biosignatures capable of predicting a future diagnosis of inflammatory bowel disease.

NEW FINDINGS

We identified a blood protein signature that predicted Crohn's disease as early as 16 years before its diagnosis and validated its high predictive capacity across multiple independent cohorts. The corresponding signature for preclinical ulcerative colitis had a lower, albeit significant, predictive ability.

LIMITATIONS

A limitation of this study is the case-control design, which calls for careful attention to matching, protein normalization, and statistical methods.

CLINICAL RESEARCH RELEVANCE

Our findings support the possibility of prognosticating inflammatory bowel disease. The long preclinical period in Crohn's disease endorses the adoption of early preventive strategies (eg, diet alterations and medication) to potentially attenuate disease progression and improve the natural history of Crohn's disease.

BASIC RESEARCH RELEVANCE

In preclinical samples, downregulated proteins (but not upregulated) related to gut barrier integrity and macrophage functionality correlated with time to diagnosis of Crohn's disease. These findings could suggest that a decline in the expression of seemingly protective proteins may serve as a characteristic feature of the preclinical phase of Crohn's disease. Mechanistic studies are needed to decipher the precise role of these proteins in preclinical disease.

was identified that accurately predicted a future diagnosis of CD, whereas high expression of anti- $\alpha v \beta 6$ autoantibodies was highly predictive of UC. However, the predictive ability of these biomarkers, defined as the area under the receiver operating characteristic curve (AUC), was not validated in external cohorts. Experiences from other disease areas have consistently shown that external validation is crucial for accurately assessing predictive ability.¹⁴

Based on these considerations, we used a rigorous study design to examine the potential of identifying highly generalizable predictive biosignatures of a future diagnosis of CD and UC through protein profiling. By analyzing preclinically

Abbreviations used in this paper: AUC, area under the receiver operating characteristic curve; CD, Crohn's disease; IBD, inflammatory bowel disease; ICD, International Classification of Diseases; IQR, interquartile range; LR, likelihood ratio; MDC, Malmö diet and cancer; NSHDS, Northern Sweden Health and Disease Study register cohort; SIC IBD, Swedish Inception Cohort in IBD; UC, ulcerative colitis.

 Most current article

© 2025 The Author(s). Published by Elsevier Inc. on behalf of the AGA Institute. 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.2024.11.006>

biobanked blood samples from large population-based cohorts, we aimed first to identify predictive protein signatures in a discovery cohort and then validate their performance in an external independent cohort. In addition, we sought to examine the relevance of these preclinical signatures at diagnosis and assess the influence of genetic and shared environmental risk factors on their expression.

Materials and Methods

Study Design

This is a case-control study, nested within large population-based cohorts, in which we identified preclinical protein signatures in blood from healthy individuals who later in life were diagnosed with CD and UC (preclinical cases) and compared them with individuals from the general population who remained free from IBD at follow-up (controls), as outlined in [Figure 1](#). Next, we validated our findings by assessing the predictive accuracy of these signatures to distinguish preclinical cases from controls in an external population-based cohort. We analyzed samples from a nationwide inception cohort to examine the relevance of the preclinically dysregulated proteins at diagnosis and assess the ability of these signatures to differentiate incident patients with CD and UC from healthy individuals. Finally, we controlled for the influence of genetic and shared environmental risk factors on the predictive protein signatures by examining their predictive ability in twins with preclinical disease within a population-based twin cohort.

The study was approved by the Swedish Ethical Review Authority (dnr 2020-07065). This study followed the TRIPOD guideline for the development and validation of prediction models.

Data Sources

The National Patient Register. Sweden's National Patient Register has held information about hospital admissions since 1964 and has been covered nationally since 1987. It also contains data on nonprimary outpatient care from 2001.¹⁵ The main and contributory diagnoses are coded according to the International Classification of Diseases (ICD) codes ([Supplementary Tables 1–4](#)).

Population-based cohorts. The "Malmö diet and cancer" (MDC, $n = 28,000$) cohort served as the discovery cohort for preclinical IBD and findings were validated in the Northern Sweden Health and Disease Study register cohort (NSHDS, $n = 143,000$). In addition, we used the TwinGene study ($n = 12,591$), which is nested within the Swedish Twin Registry, to assemble a twin cohort of preclinical IBD. The details of these data sources are provided in the [Supplementary Materials](#) and [Supplementary Table 5](#).

Clinical cohorts. *Discovery cohort of preclinical IBD.* Individuals in the discovery cohort (MDC) diagnosed with IBD later in life were identified using the unique personal identification number assigned to all Swedish residents and linking the dataset with the Swedish National Patient Register. The diagnosis of CD or UC was defined as having ≥ 2 IBD-associated ICD codes, corresponding to a positive predictive value of 93%, according to Jakobsson et al. ([Supplementary Figure 1](#) and [Supplementary Tables 1–4](#)).^{16,17} For 94 of 156 (60%) preclinical cases in the discovery cohort, an experienced

gastroenterologist also scrutinized the medical records. Patients with IBD-unclassified were excluded to reduce the risk of misclassification between CD and UC. For each preclinical case, we identified 1 control (free from IBD during follow-up), matched by birth year, sex, and calendar period at inclusion, and the serum samples were retrieved from the MDC biobank.

Validation cohort of preclinical IBD. Individuals with preclinical IBD in the validation cohort (NSHDS) were identified by applying a similar approach, that is, linking the Västerbotten Intervention Project and the Mammography Screening Project datasets with the ICD Code register of Region of Västerbotten, Sweden. Copies of the medical records for all individuals with at least 1 inpatient or nonprimary outpatient care visit listing a diagnosis of IBD were manually scrutinized by an experienced gastroenterologist to confirm or reject a diagnosis of IBD.¹⁸ Each preclinical IBD case was matched to a healthy control, using the same matching criteria as in the discovery cohort. Finally, plasma samples from preclinical cases and controls were retrieved from the NSHDS biobank.

Inception cohort for assessment of the relevance of preclinically dysregulated proteins at diagnosis. To ascertain the relevance of the identified proteins at the time of IBD diagnosis, we conducted a proof-of-concept analysis using serum samples from the Swedish Inception Cohort in IBD (SIC IBD). The SIC IBD is a multicenter study comprising patients aged ≥ 18 years who presented with gastrointestinal symptoms and were referred to gastroenterology units for suspected IBD at 7 Swedish hospitals between 2011 and 2021. In addition, healthy individuals without any history of chronic gastrointestinal symptoms or disease were recruited, and samples were collected and processed according to the same protocol. The diagnosis of IBD was established based on internationally accepted criteria, following clinical, microbiological, endoscopic, histologic, and radiologic evaluation.¹⁹ To maintain consistency with the methodology used in the other cohorts, we used a down-sampling technique, matched each case of CD and UC to a healthy control by sex and age, and conducted a case-control analysis. Similarly, we matched 100 patients with newly diagnosed CD to 100 patients with newly diagnosed UC to enable direct comparisons of these 2 subtypes of IBD.

Twin cohort for assessment of the influence of genetic and shared environmental risk factors of preclinically dysregulated proteins. Twin pairs discordant for preclinical CD or UC within the TwinGene study were identified using ICD codes and using the same validated register-based definition of IBD as in the preclinical discovery cohort. Each twin pair discordant for preclinical IBD was also matched to an unrelated twin (external twin) from the general population by sex and zygosity. To specifically examine the influence of genetic predisposition on each protein marker associated with preclinical IBD, we assessed the intraclass correlation coefficients and calculated the heritability using twin pairs without IBD in the TwinGene study.

Protein Measurements

Some 178 proteins were analyzed across 2 protein panels (inflammation and oncology II) using proximity extension assay technology (Olink Proteomics, Uppsala, Sweden), as outlined in the [Supplementary Materials](#) and [Supplementary Figures 2 and 3](#). [Supplementary Table 6](#) presents the overlap of proteins from the inflammatory panel in the current study and our previous publication.²⁰

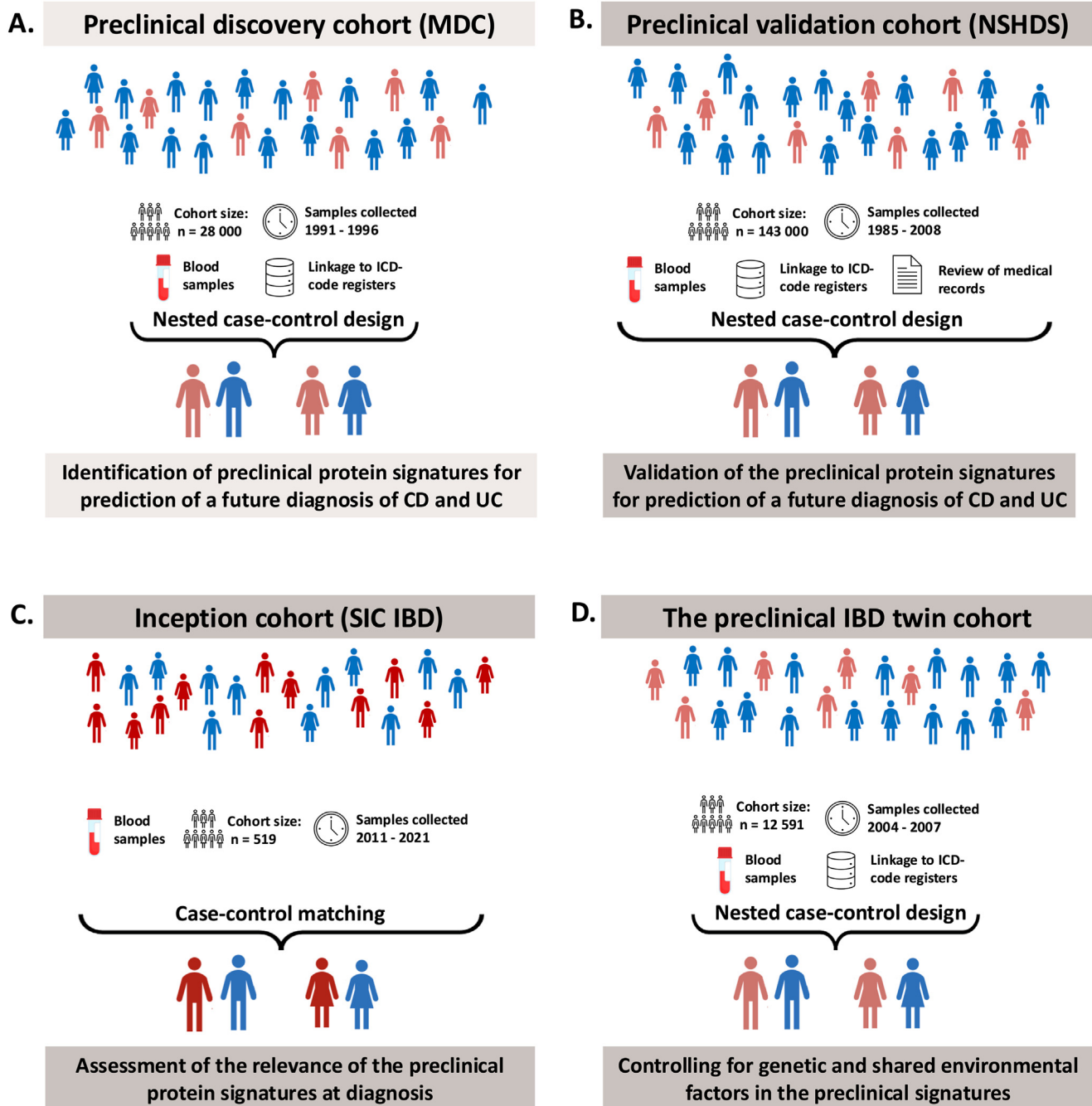


Figure 1. Two population-based cohorts with biorepositories were used to identify and validate preclinical protein signatures for the prediction of a future diagnosis of CD and UC. (A) We assembled a discovery cohort of preclinical IBD by identifying individuals who had provided a blood sample in the MDC study and later in life were diagnosed with CD or UC. Each of these individuals was matched to a healthy control. Proteomic profiling and various machine-learning algorithms were used to identify protein signatures of preclinical CD and UC. (B) An external cohort of preclinical CD and UC was assembled from the NSHDS register cohort and used to validate the performance of the identified preclinical biosignatures. (C) Next, the SIC IBD was examined to verify the relevance of the preclinical signatures at diagnosis of IBD. (D) Finally, we assessed the protein signatures in twin pairs discordant for preclinical IBD to control for the influence of genetic and shared environmental factors on the preclinical signatures.

Statistical Analysis

Data pre-processing. Details on how protein values below the lower limit of detection and missing values were managed are provided in the [Supplementary Materials](#). To enable the use of common machine-learning methods on the

case-control datasets, we applied pairwise mean normalization as described by Stanfill et al,²¹ as detailed in the [Supplementary Materials](#).

Single protein analyses. We used logistic regression to assess associations of single proteins with a future diagnosis of

CD and UC and visualized the results using volcano plots. To account for multiple testing, P values underwent adjustment using the Benjamini-Hochberg procedure. Proteins demonstrating a false discovery rate (P_{FDR}) $<.10$ in the preclinical discovery cohort (MDC) were validated in the preclinical validation cohort (NSHDS). We used a stringent P_{FDR} of $<.05$ in the validation cohort to minimize the probability of false positive findings. Proteins remaining statistically significant in the validation cohort were tested within an inception cohort (SIC IBD) of patients with newly diagnosed IBD and their matched healthy controls, adhering to the same P_{FDR} threshold of $<.05$.

For proteins implicated in preclinical IBD cohorts, Pearson's correlation coefficients were calculated to investigate the temporal correlation between protein levels and the time to diagnosis.

Next, we aimed to assess the influence of genetic predisposition at a healthy stage on each protein marker associated with preclinical IBD by assessing their intraclass correlations and computing their heritability, as described in the [Supplementary Materials](#). Heritability, which ranges from 0 to 1, represents the proportion of phenotypic variation (eg, protein levels) that can be explained by variation in genetic factors. This analysis helps determine the extent to which genetic influences contribute to variation in protein levels related to preclinical IBD.

Preclinical biosignatures. We used a regularized logistic regression model and several additional machine-learning methods (ie, L1-penalized regression, elastic net regression, support vector machines, and random forests) to identify protein signatures capable of predicting the future onset of IBD. L2- and smoothly clipped absolute deviation were used to penalize the logistic regression model, chosen for their ability to mitigate overfitting and handle multicollinearity while generating sparse and interpretable models.

The preclinical CD and UC models were first trained on data from the preclinical discovery cohort (MDC). The models were then validated on unseen data from the preclinical validation cohort (NSHDS). A detailed description of the training and evaluation of the models is provided in the [Supplementary Materials](#).

Next, we investigated whether the performance of the logistic regression models for preclinical CD and UC was influenced by the timing of sampling relative to diagnosis. We divided the time interval between sampling and diagnosis into quartiles and computed AUC estimates for each quartile.

We continued with a proof-of-concept analysis, comparing patients with newly diagnosed IBD with healthy controls within the inception cohort. By applying the models, we evaluated their diagnostic capacity at the date of IBD diagnosis. In addition, we examined the ability of the preclinical protein signatures to differentiate between CD and UC in newly diagnosed patients.

The AUC was used to evaluate model performance, and the Youden index was applied to define the optimal cutoff value. This value was used to determine sensitivity, specificity, the F1 score, and the positive and negative likelihood ratio (LR). The cutoff was then applied to the preclinical validation and inception cohorts. The Brier score was used to assess model performance. In addition, stratified analyses for sex and age were performed.

Next, we tested whether the model from the discovery cohort (MDC) could differentiate twin individuals who later in life were

diagnosed with IBD from external twin controls from the general population, matched for age, sex, and zygosity. Thereafter, these models were applied to the same preclinical twins, but this time compared with their healthy twin siblings, as outlined in the [Supplementary Materials](#). DeLong's test was used to compute P values for the comparison of AUC estimates.²²

Software. Statistical analyses, including data pre-processing, were conducted in R versions 4.03 and 4.05 (R Foundation for Statistical Computing).²³

Results

Study Population

Demographics of individuals with preclinical and newly diagnosed IBD, along with their respective sex- and age-matched controls, are shown in [Table 1](#) and [Supplementary Tables 7 to 10](#). In the preclinical cohorts, the median time from sampling to a diagnosis of IBD was 8.7 (interquartile range [IQR]: 14.1–3.2) years for CD and 7.2 (IQR: 14.2–3.4) years for UC.

Proteins Associated With a Future Diagnosis of IBD

Crohn's disease. Through logistic regression and applying a P_{FDR} criterion of $<.10$, we identified 34 proteins associated with preclinical CD in the discovery cohort (MDC) during the single protein analyses ([Figure 2](#)). Subsequent analyses of the validation cohort (NSHDS) corroborated the associations for 9 of these proteins ($P_{FDR} <.05$). Six of these protein markers were upregulated (CXCL9, IL6, MMP-10, CCL20, MDK, and CXCL17) and 3 were downregulated (DNER, GPNMB, and CX3CL1) in preclinical CD. All downstream univariate analyses used the 9 proteins validated in the NSHDS cohort.

Next, we conducted a proof-of-concept analysis to examine the relevance of these proteins at the time of diagnosis by comparing newly diagnosed CD with healthy controls in the SIC IBD. Consistent with the results in the preclinical cohorts, all protein markers, except MDK, CXCL17, and GPNMB, were significantly differentially regulated between the 2 groups in the inception cohort ($P_{FDR} <.05$).

Last, we assessed correlations between protein levels and time to diagnosis by applying linear correlation analysis for cases with preclinical CD from the MDC and NSHDS cohorts. We observed significant negative correlations ($P <.05$) for the 3 proteins (DNER, GPNMB, and CX3CL1) downregulated in preclinical disease ([Supplementary Figure 4](#)). In contrast, no significant temporal correlations were observed for any of the 6 proteins upregulated in preclinical disease.

Ulcerative colitis. In the corresponding analysis of preclinical UC, we identified 45 proteins as differentially regulated in the preclinical discovery cohort MDC ([Figure 2](#)). The associations were confirmed for 12 upregulated proteins (STAMBP, SIRT2, SCAMP3, ADAM 8, GZMB, ABL1, CCL11, CD5, TNFRSF6B, MMP-10, CDKN1A, and CXCL9), but none of the downregulated proteins could be

Table 1. Demographics of Individuals in the Study Cohorts

	Individuals, n	Median (IQR) age at sampling, y	Median (IQR) age at diagnosis, y	Female sex, n (%)
Discovery cohort (MDC)				
Preclinical CD cases	54	53.8 (43.8–63.9)	67.7 (54.1–81.4)	35 (65)
Preclinical UC cases	102	55.5 (44.2–66.8)	69.1 (57.5–80.7)	52 (51)
Healthy controls	156	54.8 (43.1–66.5)	-	87 (56)
Validation cohort (NSHDS)				
Preclinical CD cases	31	50.3 (30.6–70)	53.6 (38.7–68.5)	16 (52)
Preclinical UC cases	80	50.1 (30.4–69.8)	54.8 (38.5–71.2)	42 (52)
Healthy controls	111	50.1 (30.5–69.8)	-	58 (52)
Inception cohort (SIC IBD)				
Newly diagnosed CD	48	25.5 (19.0–32.0)	25.5 (19.0–32.0)	28 (58)
Newly diagnosed UC	48	25.5 (19.2–31.8)	25.5 (19.2–31.8)	28 (58)
Healthy controls	48	25.5 (19.0–32.0)	-	28 (58)
Preclinical IBD twin cohort				
Preclinical CD twins	10	61.2 (52.8–69.6)	65.9 (56.1–75.7)	8 (80)
Preclinical UC twins	24	66.9 (59.1–74.7)	71.2 (61.7–80.8)	15 (62)
Healthy twin siblings	34	65.1 (55.4–74.9)	-	17 (50)
Healthy external twins	34	65.3 (54.7–75.8)	-	17 (50)

confirmed when comparing preclinical UC cases with matched controls in the validation cohort (NSHDS). These 12 protein markers were also significantly upregulated in patients with newly diagnosed UC from the inception cohort (SIC IBD); however, within preclinical cases, only MMP-10 demonstrated a correlation with time to diagnosis of UC (Supplementary Figure 5).

Sensitivity analysis. Sensitivity analysis was performed by omitting measurements within a 2-year period before diagnosis, excluding 11 individuals with preclinical CD and 25 individuals with preclinical UC. The validated proteins of preclinical IBD only showed minor differences in fold changes and P_{FDR} values when the analyses were restricted to samples obtained more than 2 years before the diagnosis of CD or UC (Supplementary Tables 11 and 12).

Protein Signatures Predictive of a Future Diagnosis of IBD

Crohn's disease. Within the preclinical discovery cohort, we used regularized logistic regression to identify a biosignature of 29 proteins, which demonstrated a high ability to differentiate individuals who later in life were diagnosed with CD from healthy controls (AUC = 0.85; 95% CI, 0.78–0.93), assessed via leave-one-out cross-validation (Figure 3 and Supplementary Figure 6). The predictive capacity remained high (AUC = 0.87; 95% CI, 0.77–0.97) when the model was applied to the preclinical validation cohort, with a sensitivity of 77% and a specificity of 87% at the optimal cutoff, corresponding to an LR(+) of 6.0 and LR(–) of 0.26 (Table 2 and Supplementary Table 13). The other machine-learning models consistently differentiated individuals with preclinical CD from healthy controls in the validation cohort. The AUC estimates of the alternative

algorithms were comparable or slightly lower than the logistic regression model (Supplementary Table 14), with the elastic net model showing a numerically higher sensitivity at its optimal cutoff (Supplementary Table 15).

In the preclinical cohorts, the performance of the logistic regression model increased toward diagnosis. However, a high predictive capacity was also observed when restricting the analysis of the MDC cohort to samples taken more than 16 years before the diagnosis (AUC = 0.82, Supplementary Figures 7 and 8). The model for preclinical CD seemed to perform better for male (AUC = 0.99; 95% CI, 0.98–1.00) compared with female participants (AUC = 0.76; 95% CI, 0.57–0.94), whereas no significant differences were observed when stratifying the analyses for age at inclusion (Supplementary Tables 16 and 17).

Next, we evaluated model performance in the inception cohort of patients with newly diagnosed CD (SIC IBD), resulting in a high discriminatory capacity (AUC = 1.00). Following the application of the Youden index-derived cutoffs, the sensitivity, specificity, and likelihood ratios across all models and cohorts are presented, with their corresponding 95% CIs in Table 2 and Supplementary Table 15. However, the model demonstrated a poor ability to differentiate between CD and UC in newly diagnosed patients (AUC = 0.56; 95% CI, 0.48–0.64).

Ulcerative colitis. For UC, the predictive capacity of the logistic regression signature was numerically higher within the preclinical discovery cohort (AUC = 0.77; 95% CI, 0.71–0.83, Figure 3 and Supplementary Figure 9) compared with when applied to the preclinical validation cohort (AUC = 0.67; 95% CI, 0.59–0.76), corresponding to a modest sensitivity and specificity at 61% for both estimates, with an LR(+) and LR(–) at 1.58 and 0.63, respectively (Table 2). The AUC estimates of the various machine-learning algorithms

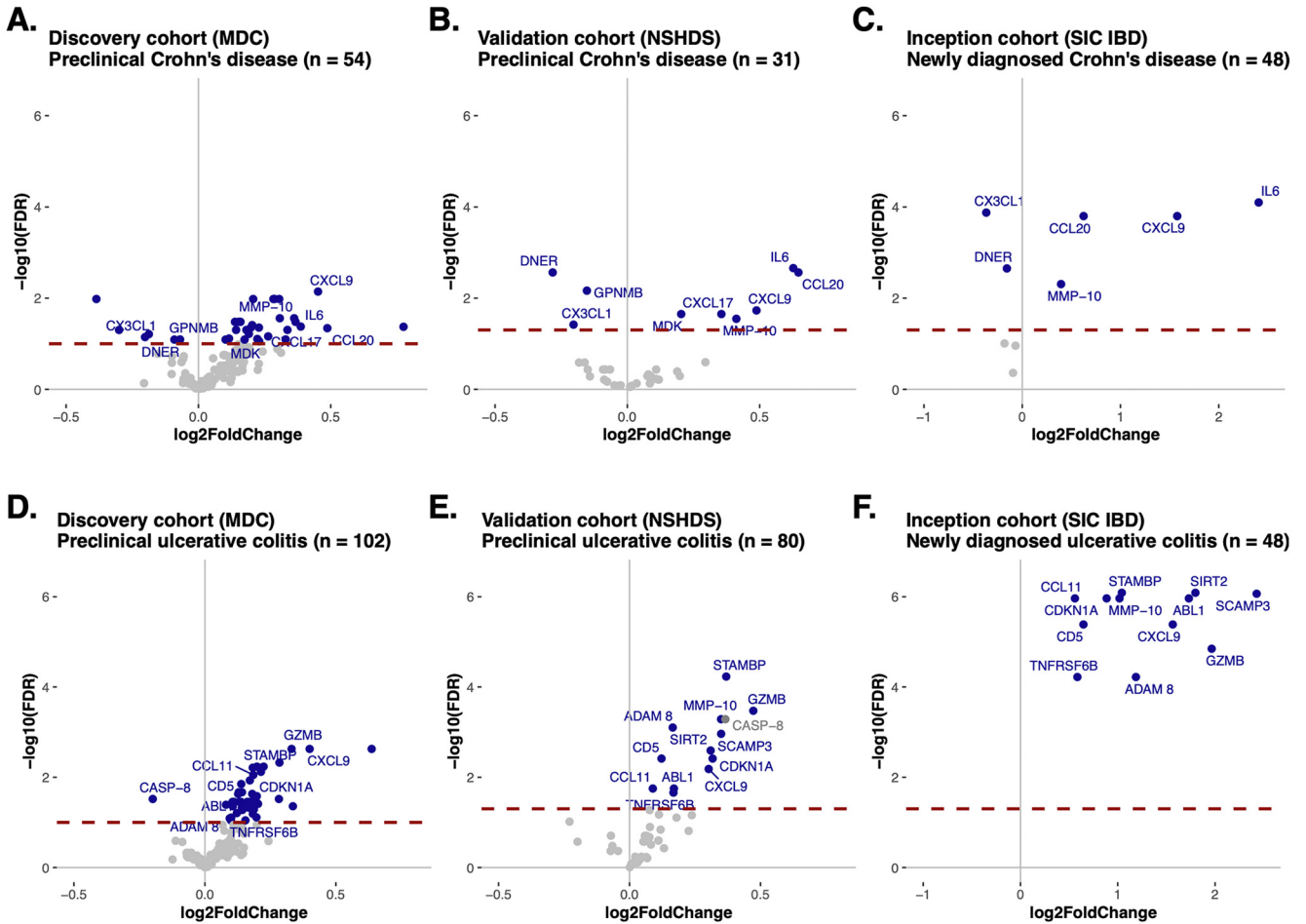


Figure 2. Volcano plots of (A) Serum protein markers in individuals who later in life were diagnosed with CD (cases with preclinical CD) compared with individuals who remained free from IBD during follow-up (healthy controls) in the discovery cohort (MDC). Thirty-four protein markers were differentially regulated ($P_{FDR} < .10$). (B) Subsequent analyses of plasma samples from cases with preclinical CD and healthy controls in the validation cohort (NSHDS) corroborated the associations for 9 of these proteins ($P_{FDR} < .05$). (C) The 9 validated protein markers were compared between patients with newly diagnosed CD and healthy controls in the inception cohort (SIC IBD), and all protein markers were differentially regulated, except MDK, CXCL17, and GPNMB ($P_{FDR} < .05$). (D) The corresponding analyses of preclinical UC in the discovery cohort identified 45 proteins as differentially regulated ($P_{FDR} < .10$). (E) Twelve of these protein markers were significantly associated ($P_{FDR} < .05$) with preclinical UC in the validation cohort. (F) All 12 proteins were also upregulated in patients with newly diagnosed UC from the inception cohort ($P_{FDR} < .05$).

used were similar or marginally higher than the logistic regression model (Supplementary Tables 14 and 18). The performance of the logistic regression model did not increase toward the date of diagnosis of UC (Supplementary Figures 10 and 11). The stratified analyses indicated that the preclinical UC signature performed better for older (AUC = 0.79; 95% CI, 0.69–0.90) than younger participants (AUC = 0.55; 95% CI, 0.42–0.68) in the preclinical validation cohort (Supplementary Tables 16 and 17).

Similar to CD, the logistic regression model demonstrated a high discriminatory capacity for newly diagnosed UC cases in the inception cohort (AUC = 0.95; 95% CI, 0.92–0.99) (Figure 3). However, its capacity to differentiate between CD and UC in newly diagnosed patients was lower (AUC = 0.67; 95% CI, 0.59–0.74). The overlap of proteins included in the preclinical signatures of CD and UC is shown in Supplementary Figure 12.

Sensitivity analysis. We conducted a sensitivity analysis by evaluating the predictive performance of the logistic regression model when excluding samples obtained within 2 years before diagnosis. For both CD and UC, only minor alterations in AUC estimates were observed (Supplementary Figure 13).

Heritability of Proteins Associated With Preclinical CD and UC

A cohort of 111 healthy twin pairs (monozygotic pairs, n = 35; dizygotic pairs, n = 76) from the TwinGene study was used to calculate intraclass correlation coefficients and heritability estimates for each of the protein markers associated with preclinical CD or UC (Supplementary Table 19). IL6, GZMB, ADAM 8, and CCL11 showed high estimates of

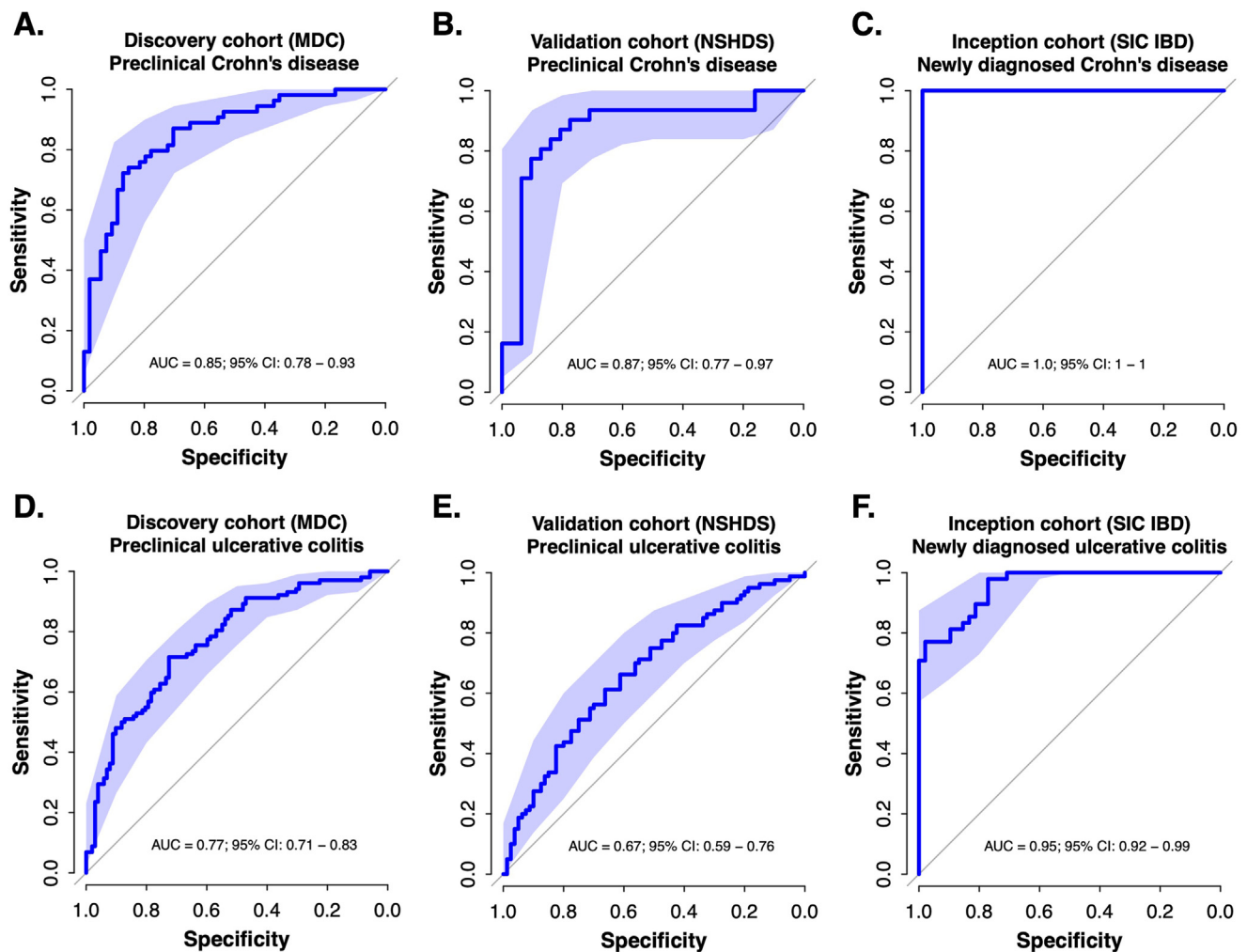


Figure 3. Receiver operating curves and AUC estimates of regularized logistic regression models showing the predictive performance of protein signatures in preclinical and newly diagnosed CD and UC. (A) The logistic regression model achieved an AUC of 0.85 in differentiating individuals with a future diagnosis of CD from healthy controls in the discovery cohort (MDC). (B) The performance of the protein signature was confirmed (AUC = 0.87) by independent validation in an external cohort of preclinical IBD (NSHDS). (C) The model was applied to patients with newly diagnosed CD in the inception cohort (SIC IBD), showing excellent performance. (D–F) The corresponding protein signature for UC resulted in lower discriminatory capabilities in both preclinical and newly diagnosed UC compared with the corresponding estimates of CD.

Table 2. Performance Metrics for the Logistic Regression Model When Youden's Index Was Employed to Select the Optimal Cutoff Point in the Discovery Cohort, Which Was Then Applied to the Validation and Inception Cohorts

	Sensitivity (95% CI)	Specificity (95% CI)	LR(+)	LR(–)	F1 Score	Median Brier score (IQR)
CD						
Discovery cohort (MDC)	0.74 (0.60–0.85)	0.85 (0.73–0.93)	5.0 (2.59–9.66)	0.30 (0.19–0.48)	0.78	0.14 (0.08–0.23)
Validation cohort (NSHDS)	0.77 (0.59–0.90)	0.87 (0.70–0.96)	6.0 (2.36–15.3)	0.26 (0.13–0.50)	0.81	0.15 (0.09–0.21)
Inception cohort (SIC IBD)	0.92 (0.80–0.98)	1.00 (0.93–1.00)	45.0 (6.44–313) ^a	0.08 (0.03–0.21)	0.96	0.11 (0.05–0.17)
UC						
Discovery cohort (MDC)	0.72 (0.62–0.80)	0.72 (0.62–0.81)	2.58 (1.84–3.62)	0.39 (0.28–0.55)	0.72	0.18 (0.03–0.34)
Validation cohort (NSHDS)	0.61 (0.50–0.72)	0.61 (0.50–0.72)	1.58 (1.14–2.19)	0.63 (0.46–0.88)	0.61	0.2 (0.05–0.36)
Inception cohort (SIC IBD)	0.83 (0.70–0.93)	0.83 (0.70–0.93)	5.00 (2.62–9.53)	0.20 (0.10–0.38)	0.83	0.15 (0.04–0.25)

CI, confidence interval; IQR, interquartile range; LR(+), positive likelihood ratio; LR(–), negative likelihood ratio.

^aTo avoid an infinite estimate, a constant of 1 was added to the false positive group before computing the positive likelihood ratio.

Table 3. Predictive Performance of the Logistic Regression Models From the Discovery Cohort When Applied to the Preclinical IBD Twin Cohort

	Number of pairs	AUC, external twin analysis (95% CI)	AUC, co-twin analysis (95% CI)	AUC difference	P value
CD					
Preclinical cases	10	0.89 (0.73–1.0)	0.58 (0.32–0.84)	0.31	0.04
UC					
Preclinical cases	24	0.74 (0.59–0.88)	0.58 (0.42–0.75)	0.16	0.17

NOTE. The AUC estimates refer to the capacity of the models to differentiate between twins with preclinical CD and UC from the controls.

heritability (>0.60), indicating a large degree of influence from genetic factors on regulating these protein markers.

Impact of Genetic and Shared Environmental Factors on Predictive Biosignatures

Next, we tested whether the logistic regression from the discovery cohort (MDC) could differentiate twin individuals who later in life were diagnosed with IBD from matched healthy, unrelated twin controls. Thereafter, these models were applied to the same preclinical twins, but this time compared with their healthy twin siblings.

Crohn's disease. In the comparative analysis of twins with preclinical CD vs external twin controls, the logistic regression model yielded an AUC of 0.89 (Table 3). The predictive ability diminished to an AUC of 0.58 when accounting for genetic and shared environmental factors, namely when twins with preclinical CD were matched with their healthy twin siblings ($P = .04$). Given the limited sample size of just 10 twins with preclinical CD, it is important to exercise caution when interpreting these findings; however, they do indicate that genetic and shared environmental factors may have a predominant influence on the predictive logistic regression protein signature

Ulcerative colitis. In the analysis of preclinical UC, we observed only a minor difference in predictive performance when contrasting preclinical UC against unrelated twin controls (AUC = 0.74) and against their healthy twin siblings (AUC = 0.58) (Table 3). This finding indicates a limited impact of genetic and shared environmental factors on the logistic regression protein signature of UC.

Discussion

Previous studies have reported preclinical signatures of IBD using omics data from single cohorts.^{3,13,20,24–26} Here, we advance the field by developing a protein model of preclinical CD capable of predicting a future diagnosis of CD, also when applied to an independent preclinical validation cohort (AUC = 0.87). The generalizability of our protein signature of preclinical CD across multiple large population-based cohorts underscores the novelty and significance of our findings. In comparison, the corresponding signature for UC had a lower predictive performance in the preclinical validation cohort (AUC = 0.67). However, when applied to

the inception cohort, both signatures had an excellent capacity to differentiate patients with new-onset CD and UC from matched healthy controls (AUC = 1.00 and 0.95, respectively). This proof-of-concept analysis demonstrates the relevance of these proteins at diagnosis. Results from the twin cohort indicate that exposures to genetic and shared environmental risk factors may contribute to the predictive signature of future CD, whereas this did not seem to be the case for UC. Moreover, for temporal correlations, discrepancies were identified between the 2 subtypes of IBD. In preclinical samples, downregulated (but not upregulated) proteins related to gut barrier integrity and macrophage functionality correlated with time to diagnosis of CD. Contrarily, all proteins associated with preclinical UC were upregulated, and only one protein marker correlated with the time to diagnosis.

Proteomic profiling has emerged as a potential tool for delineating varied stages of IBD, including predicting future IBD.^{3,13,20,27} In the PREDICTS (Proteomic Evaluation and Discovery in an IBD Cohort of Tri-service Subjects) study, Torres et al¹³ used an aptamer-based assay (SomaScan) to examine serum samples from young, predominantly male, active-duty military staff. Through a cross-validation method, the investigators identified a signature predictive of a CD diagnosis within 5 years, with an AUC of 0.76, whereas their attempts to decipher a predictive signature for future UC failed. In contrast, Lochead et al. examined the Nurses' Health Study and reported increased preclinical levels of high-sensitive C-reactive protein and interleukin 6 in both CD and UC.²⁸ In addition, 25 proteins were recently associated with preclinical CD in the GEM (Genetics Environment Microbial Project) cohort, and we have previously reported on 6 inflammatory proteins related to a future diagnosis of UC, also using samples from the NSHDS.^{3,20} However, interpreting these studies is challenged by the lack of external cohorts for independent validation and a relatively short period between blood collection and diagnosis of IBD. The present study addressed these limitations and showed aberrations in the serum proteome many years before the clinical manifestations of CD. Intriguingly, our protein model demonstrated a high predictive ability more than 16 years before the diagnosis in the discovery cohort (AUC = 0.82), even though its performance increased toward diagnosis, with an AUC of 0.93 within a 6-year period

pre-diagnosis. Furthermore, we applied a stringent study design and confirmed their ability to predict a future IBD diagnosis in an external validation cohort. A recent Danish registry study supports a long preclinical period before the diagnosis of CD.¹ Vestergaard et al¹ observed aberrations of standard biochemical and hematological tests up to 8 years before diagnosis, even though these deviations were relatively subtle.

Differences in patient populations and proteomic platforms may explain the discrepancy between previous studies. Most patients in our cohorts were aged >40 years at the diagnosis of IBD, and evenly distributed with respect to sex, whereas most individuals in the PREDICTS study were male patients diagnosed during their second or third decade of life.¹³ Compared with our cohorts, participants in the GEM study were also younger, as only first-degree relatives aged 6 to 35 years were included.³ Moreover, we applied the antibody-based proximity extension assay technique, whereas the aptamer-based SomaScan platform was used in the PREDICTS study. Pietzner et al²⁹ noted the high variability between the 2 platforms (median correlation coefficient 0.38), which may account for the disparities in results.

In addition to the existing literature on preclinical proteomic profiles, Livanos et al⁴ recently reported that an increased expression of the $\alpha\nu\beta 6$ autoantibody in the PREDICTS cohort predicts a future diagnosis of UC with an AUC of 0.79 up to a decade before diagnosis. The authors also confirmed an increased expression of the anti-integrin autoantibody in the GEM cohort, but without reporting its precise predictive performance in the validation cohort. Hence, there is a need for additional validation in external cohorts, and future research should examine the potential of combining proteomic profiling and anti- $\alpha\nu\beta 6$ autoantibodies to enhance the predictive accuracy of UC.

To investigate how genetic and shared environmental factors influenced the performance of our protein signatures, we continued by analyzing a twin cohort with preclinical IBD cases. The preclinical protein signature for CD seemed to have a higher predictive ability when preclinical twin cases ($n = 10$) were matched to unrelated controls compared with their twin siblings (who remained free from IBD during follow-up). Conversely, no meaningful differences were observed in the analysis of UC based on 24 cases with preclinical disease. These observations suggest that the influence of genetic and shared environmental factors on the predictive protein signatures was more pronounced in preclinical CD compared with preclinical UC. Although we believe this study design is a promising model for investigating the causes of protein dysregulation in preclinical IBD, interpreting the results from the analysis of twins with preclinical disease was hampered by limited statistical power. It is important to consider that twins are more tightly matched for environmental exposures than first-degree relatives in general, including siblings. Hence, our findings do not necessarily negate the utility of the preclinical protein signatures in other first-degree relative settings, such as assessing ordinary siblings to patients with IBD.

In line with the previous analyses of the GEM and NSHDS cohorts, we observed a preclinical dysregulation of CXCL9, MMP-10, IL6, and CCL-11 in the univariate analyses.^{3,20} Beyond the reiteration of these previously reported protein findings, our study also elucidated several novel associations. In preclinical CD, we observed a downregulation of DNER, GPNMB, and CX3CL1, whereas CCL-20, MDK, and CXCL17, all chemokines, were upregulated. Intriguingly, the 3 downregulated proteins also exhibited temporal correlations, with diminishing concentrations toward diagnosis. DNER is intricately linked to the Notch-1 signaling pathway and has been proposed to hold considerable relevance in maintaining gastrointestinal mucosal barrier integrity.³⁰ Bourgonje et al³¹ recently also reported an inverse relationship between DNER levels and clinical disease activity in patients with CD. Of note, mechanistic studies showed that both GPNMB and CX3CL1 are important for the functionality of macrophages to intestinal inflammation.^{32,33} GPNMB is highly expressed in macrophages and seems to suppress the production of pro-inflammatory cytokines.³² In mice, GPNMB is also involved in the response to intestinal damage.³⁴ The soluble form of CX3CL1 acts as a chemoattractant, whereas the membrane-bound form is expressed by endothelial cells and induces adhesion of circulating monocytes by binding to the C-X3-C motif chemokine receptor 1 (CX3CR1).³⁵ Using mouse models, Medina-Contreras et al³³ found that deficiency of either CX3CL1 or CX3CR1 results in a decreased frequency of macrophages in the intestinal mucosa, increased translocation of commensal gut bacteria, and increased severity of dextran sulfate sodium (DSS)-induced colitis.

In UC, our analyses revealed the upregulation of 9 proteins hitherto not associated with its preclinical phase. Of these, increased mucosal expression of GZMB and TNFRSF6B has been reported in studies of patients with prevalent UC, providing additional support to their roles in the underlying pathogenesis of UC. GZMB, situated in the granule of cytotoxic T-lymphocytes, functions as a potent inducer of apoptosis.³⁶⁻³⁸ The observed upregulation of TNFRSF6B could be ascribed to feedback mechanisms designed to mitigate an increasing inflammatory cascade in preclinical UC. TNFRSF6B acts by sequestering and neutralizing pro-inflammatory cytokines within the tumor necrosis superfamily. Moreover, the observed upregulation of SIRT2 and CD5 may suggest other feedback loops, as mice deficient in SIRT2 or CD5 demonstrate exacerbated colitis.^{39,40} However, mechanistic studies are needed to decipher the precise mechanisms underlying the observed upregulation of these proteins and of the remaining novel associations (ie, ABL-1, ADAM-8, and SCAMP-3) in preclinical UC.

Unlike the temporal correlations observed between preclinical levels of downregulated proteins and time to diagnosis of CD, none of the upregulated proteins showed a temporal correlation with the period to diagnosis of CD or UC, except for MMP-10 in UC. These findings could suggest that a decline in the expression of seemingly protective proteins may serve as a characteristic feature of the

preclinical phase of CD. In accordance with the findings from recent comprehensive analyses of protein quantitative trait locus variants,^{41–43} our analysis of healthy twin pairs showed a significant influence of genetic predisposition on protein levels for several of the preclinical markers. It is worth noting that none of the downregulated proteins demonstrated high heritability; however, our broad CIs challenged the interpretation of these estimates.

This study has several strengths. First, we confirmed the predictive performance of the preclinical protein signatures by external validation in an independent cohort. Second, our analysis encompassed demographically heterogeneous cohorts, setting it apart from previous studies that primarily focused on samples composed of individuals with a particular profession (eg, army personnel and nurses)^{13,28} or first-degree relatives of patients with CD.³ Third, the diagnosis of IBD in the preclinical validation cohort was confirmed by reviewing medical records.

A limitation of this study and previous studies on preclinical IBD is the reliance on a case-control design, which necessitates considerations in the context of alternative study designs. Notably, our study's pairwise mean normalization procedure is tailored for matched data, which restricts its applicability in non-matched settings. Although inflammation is a hallmark of both UC and CD, differences in immune pathway involvement and cellular responses between the 2 subtypes of IBD may result in varying expression levels of specific markers. Hence, the preselection of proteins could potentially result in a more robust signature for CD, as the chosen protein markers may have been more closely aligned with its pathophysiological processes. Nevertheless, when applied to the inception cohort, the preclinical protein signatures for CD and UC showed limited ability to differentiate the 2 subtypes of IBD.

Another limitation is the relatively high median age at diagnosis in our preclinical cohorts. Thus, the findings should be interpreted cautiously when considering their applicability to younger populations. This is especially relevant for the preclinical UC signature, where the performance seemed to be inferior for younger individuals in the validation cohort.

Plasma was used in the preclinical validation cohort, in contrast to serum in all other cohorts. This and differences in sample size across the biorepositories may account for the non-replication of some dysregulated proteins in the preclinical discovery cohort. In addition, the absence of longitudinally collected samples impeded our investigation into the potential of individual protein trajectories to augment the predictive ability of preclinical biosignatures. Last, we did not perform matching on the date of sample collection between cases and controls in the inception cohort, and the controls had to be free from chronic gastrointestinal symptoms and were not recruited from a population-based sample. Therefore, they likely represent a healthier population than the background population.

In summary, we identified a protein signature capable of predicting CD many years before diagnosis and demonstrated its generalizability in an independent cohort using unrelated individuals from the general population as

controls. However, the results from our twin cohort suggest that genetic and shared environmental factors partly influence the dysregulation of the proteins within the signature. The corresponding signature for preclinical UC had a lower, albeit significant, predictive ability. In preclinical samples, downregulated (but not upregulated) proteins related to gut barrier integrity and macrophage functionality correlated with time to diagnosis of CD. Collectively, these findings support the possibility of prognosticating IBD. The long preclinical period in CD endorses the adoption of early preventive strategies (eg, dietary modifications and medication) to potentially attenuate disease progression and improve the natural history of CD.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at <https://doi.org/10.1053/j.gastro.2024.11.006>.

References

1. Vestergaard MV, Allin KH, Poulsen GJ, et al. Characterizing the pre-clinical phase of inflammatory bowel disease. *Cell Rep Med* 2023;4:101263.
2. Torres J, Halfvarson J, Rodríguez-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;15:1443–1454.
3. Leibovitz H, Lee S-H, Raygoza Garay JA, et al. Immune response and barrier dysfunction-related proteomic signatures in preclinical phase of Crohn's disease highlight earliest events of pathogenesis. *Gut* 2023;72:1462–1471.
4. Livanos AE, Dunn A, Fischer J, et al. Anti-integrin $\alpha\text{v}\beta\text{6}$ autoantibodies are a novel biomarker that antedate ulcerative colitis. *Gastroenterology* 2023;164:619–629.
5. Halfvarson J, Standaert-Vitse A, Järnerot G, et al. Anti-*Saccharomyces cerevisiae* antibodies in twins with inflammatory bowel disease. *Gut* 2005;54:1237–1243.
6. Fiorino G, Morin M, Bonovas S, et al. Prevalence of bowel damage assessed by cross-sectional imaging in early Crohn's disease and its impact on disease outcome. *J Crohns Colitis* 2017;11:274–280.
7. Frolkis AD, Dykeman J, Negrón ME, et al. Risk of surgery for inflammatory bowel diseases has decreased over time: a systematic review and meta-analysis of population-based studies. *Gastroenterology* 2013;145:996–1006.
8. Torres J, Burisch J, Riddle M, et al. Preclinical disease and preventive strategies in IBD: perspectives, challenges and opportunities. *Gut* 2016;65:1061–1069.
9. Nguyen VQ, Jiang D, Hoffman SN, et al. Impact of diagnostic delay and associated factors on clinical outcomes in a U.S. inflammatory bowel disease cohort. *Inflammatory Bowel Diseases* 2017;23:1825–1831.
10. Gerlag DM, Safy M, Majjer KI, et al. Effects of B-cell directed therapy on the preclinical stage of rheumatoid

- arthritis: the PRAIRI study. *Ann Rheum Dis* 2019; 78:179–185.
11. Writing Group for the TRIGR Study Group. Effect of hydrolyzed infant formula vs conventional formula on risk of type 1 diabetes: the TRIGR randomized clinical trial. *JAMA* 2018;319:38–48.
 12. Herold KC, Bundy BN, Long SA, et al. An anti-CD3 antibody, teplizumab, in relatives at risk for type 1 diabetes. *N Engl J Med* 2019;381:603–613.
 13. 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.
 14. Siontis GCM, Tzoulaki I, Castaldi PJ, et al. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol* 2015;68:25–34.
 15. Ludvigsson JF, Andersson E, Ekbohm A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011;11:450.
 16. Olén O, Askling J, Sachs MC, et al. Mortality in adult-onset and elderly-onset IBD: a nationwide register-based cohort study 1964–2014. *Gut* 2020;69:453–461.
 17. Jakobsson GL, Sternegård E, Olén O, et al. Validating inflammatory bowel disease (IBD) in the Swedish National Patient Register and the Swedish Quality Register for IBD (SWIBREG). *Scand J Gastroenterol* 2017; 52:216–221.
 18. **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.
 19. Maaser C, Sturm A, Vavricka SR, et al. ECCO-ESGAR guideline for diagnostic assessment in IBD part 1: initial diagnosis, monitoring of known IBD, detection of complications. *J Crohns Colitis* 2019;13:144–164K.
 20. Bergemalm D, Andersson E, Hultdin J, et al. Systemic inflammation in preclinical ulcerative colitis. *Gastroenterology* 2021;161:1526–1539.e9.
 21. Stanfill B, Reehl S, Bramer L, et al. Extending classification algorithms to case-control studies. *Biomed Eng Comput Biol* 2019;10:1179597219858954.
 22. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–845.
 23. R Core Team. R: A Language and Environment for Statistical Computing, 2021. Available at: <https://www.R-project.org/>.
 24. **Hua X, Ungaro RC**, Petrick LM, et al. Inflammatory bowel disease is associated with prediagnostic perturbances in metabolic pathways. *Gastroenterology* 2023; 164:147–150.e2.
 25. Raygoza Garay JA, Turpin W, Lee S-H, et al. Gut microbiome composition is associated with future onset of Crohn's disease in healthy first-degree relatives. *Gastroenterology* 2023;165:670–681.
 26. Gaifem J, Rodrigues CS, Petralia F, et al. A unique serum IgG glycosylation signature predicts development of Crohn's disease and is associated with pathogenic antibodies to mannose glycan. *Nat Immunol* 2024; 25:1692–1703.
 27. 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.
 28. 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.
 29. Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nat Commun* 2021;12:6822.
 30. Dahan S, Rabinowitz KM, Martin AP, et al. Notch-1 signaling regulates intestinal epithelial barrier function, through interaction with CD4+ T cells, in mice and humans. *Gastroenterology* 2011;140:550–559.
 31. Bourgonje AR, Hu S, Spekhorst LM, et al. The effect of phenotype and genotype on the plasma proteome in patients with inflammatory bowel disease. *J Crohns Colitis* 2022;16:414–429.
 32. Saade M, Araujo de Souza G, Scavone C, et al. The role of GPNMB in inflammation. *Front Immunol* 2021;12:674739.
 33. Medina-Contreras O, Geem D, Laur O, et al. CX3CR1 regulates intestinal macrophage homeostasis, bacterial translocation, and colitogenic Th17 responses in mice. *J Clin Invest* 2011;121:4787–4795.
 34. Sasaki F, Kumagai K, Uto H, et al. Expression of glycoprotein nonmetastatic melanoma protein B in macrophages infiltrating injured mucosa is associated with the severity of experimental colitis in mice. *Mol Med Rep* 2015;12:7503–7511.
 35. Schwaeble WJ, Stover CM, Schall TJ, et al. Neuronal expression of fractalkine in the presence and absence of inflammation. *FEBS Lett* 1998;439:203–207.
 36. Russell JH, Ley TJ. Lymphocyte-mediated cytotoxicity. *Annu Rev Immunol* 2002;20:323–370.
 37. Cupi ML, Sarra M, Marafini I, et al. Plasma cells in the mucosa of patients with inflammatory bowel disease produce granzyme B and possess cytotoxic activities. *J Immunol* 2014;192:6083–6091.
 38. Kugathasan S, Baldassano RN, Bradfield JP, et al. Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. *Nat Genet* 2008;40:1211–1215.
 39. Schuster C, Kief B, Hatzihristidis T, et al. CD5 Controls gut immunity by shaping the cytokine profile of intestinal T cells. *Front Immunol* 2022;13:906499.
 40. Lo Sasso G, Menzies KJ, Mottis A, et al. SIRT2 deficiency modulates macrophage polarization and susceptibility to experimental colitis. *PLoS One* 2014;9:e103573.
 41. **Sun BB, Chiou J, Traylor M**, et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* 2023;622:329–338.
 42. **Eldjarn GH, Ferkingstad E**, Lund SH, et al. Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* 2023;622:348–358.
 43. Zhao JH, Stacey D, Eriksson N, et al. Genetics of circulating inflammatory proteins identifies drivers of immune-mediated disease risk and therapeutic targets. *Nat Immunol* 2023;24:1540–1551.

Author names in bold designate shared co-first authorship.

Received January 17, 2024. Accepted November 7, 2024.

Correspondence

Address correspondence to: Olle Grännö, Department of Laboratory Medicine, Clinical Microbiology, Faculty of Medicine and Health, Campus USÖ, Örebro University, Fakultetsgatan 1, SE 701 82 Örebro, Sweden. e-mail: olle.granno@regionorebrolan.se.

Acknowledgments

The BIOIBD consortium includes Sven Almer,^{1,3} André Blomberg,² Francesca Bresso,³ Adam Carstens,⁴ Henrik Hjortswang,⁵ Jóhann Páll Hreinsson,⁶ Maria Ling Lundström,⁷ Jan Marsal,⁸ and Hans Strid³; from the ¹Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden; ²Department of Medicine, Geriatrics and Emergencies, Gastroenterology Section, Sahlgrenska University Hospital, Östra, Göteborg, Sweden; ³Gastroenterology Unit, Department of Gastroenterology, Dermatovenereology and Rheumatology, Karolinska University Hospital, Stockholm, Sweden; ⁴Department of Internal Medicine, Ersta Hospital, Stockholm, Sweden; ⁵Department of Health, Medicine, and Caring Sciences, Linköping University, Linköping, Sweden; ⁶Department of Molecular and Clinical Medicine, Sahlgrenska Academy, Göteborg, Sweden; ⁷Department of Medical Sciences, Uppsala University, Uppsala, Sweden; and ⁸Department of Gastroenterology, Skåne University Hospital, Lund/Malmö, Sweden.

The authors thank the personnel at the Örebro biobank for their contribution to this study, including project coordination (Elisabeth Tina) regulatory work (Karin Johansson), and sample processing (Line Bergman and Maja Edvinsson). The authors thank Ralf Kujala-Halkola for statistical advice regarding the twin analysis. The authors acknowledge the Malmö diet and cancer cohort and The Swedish Twin Registry for access to data. The Swedish Twin Registry is managed by Karolinska Institutet and receives funding through the Swedish Research Council under grant no. 2017-00641. We acknowledge support from Lund University Infrastructure grant "Malmö population-based cohorts" (STYR 2019/2046). We thank the participants of the Northern Sweden Health and Disease Study register cohort (NSHDS) for their participation and acknowledge our gratitude to the Department of Biobank Research at Umeå University (<https://www.umu.se/en/biobank-research-unit/>), NSHDS Cohort, and Västerbotten County Council for delivering data and blood samples. In addition, we thank the laboratory staff at the Department of Biobank Research.

CRedit Authorship Contributions

Olle Grännö, MD (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Funding acquisition: Equal; Investigation: Lead; Software: Lead; Validation: Lead; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Equal)

Daniel Bergemalm, MD, PhD (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Funding acquisition: Equal; Investigation: Lead; Validation: Equal; Writing – review & editing: Equal)

Benita Salomon, MS (Data curation: Equal; Formal analysis: Lead; Investigation: Equal; Software: Lead; Validation: Lead; Visualization: Equal; Writing – review & editing: Equal)

Carl Mårten Lindqvist, PhD (Data curation: Lead; Formal analysis: Equal; Investigation: Equal; Software: Lead; Supervision: Equal; Validation: Lead; Writing – original draft: Equal)

Charlotte R. H. Hedin, MD, PhD (Data curation: Equal; Formal analysis: Equal; Funding acquisition: Equal; Investigation: Equal; Writing – review & editing: Equal)

Marie Carlsson, MD, PhD (Data curation: Equal; Formal analysis: Equal; Funding acquisition: Equal; Investigation: Equal; Writing – review & editing: Equal)

Katharina Dannenberg, PhD (Data curation: Equal; Formal analysis: Lead; Investigation: Equal; Software: Lead; Validation: Lead; Visualization: Equal; Writing – review & editing: Equal)

Erik Andersson, MD (Data curation: Equal; Formal analysis: Equal; Investigation: Equal; Writing – review & editing: Equal)

Åsa V. Keita, PhD (Data curation: Equal; Formal analysis: Equal; Funding acquisition: Equal; Investigation: Equal; Writing – review & editing: Equal)

Maria K. Magnusson, PhD (Data curation: Equal; Funding acquisition: Equal; Investigation: Equal; Writing – review & editing: Equal)

Carl Eriksson, MD, PhD (Data curation: Equal; Formal analysis: Equal; Investigation: Equal; Writing – review & editing: Equal)

Vivekananda Lanka, MS (Data curation: Equal; Formal analysis: Equal; Investigation: Equal; Writing – review & editing: Equal)

Patrik K.E. Magnusson, PhD (Formal analysis: Equal; Investigation: Equal; Writing – review & editing: Equal)

Mauro D'Amato, PhD (Data curation: Equal; Formal analysis: Equal; Funding acquisition: Equal; Investigation: Equal; Writing – review & editing: Equal)

Lena Öhman, PhD (Data curation: Equal; Formal analysis: Equal; Funding acquisition: Equal; Investigation: Equal; Writing – review & editing: Equal)

Johan D Söderholm, MD, PhD (Data curation: Equal; Formal analysis: Equal; Funding acquisition: Equal; Investigation: Equal; Writing – review & editing: Equal)

Johan Hultdin, MD, PhD (Data curation: Equal; Formal analysis: Equal; Investigation: Equal; Writing – review & editing: Equal)

Robert Kruse, PhD (Data curation: Equal; Formal analysis: Equal; Funding acquisition: Equal; Investigation: Equal; Writing – review & editing: Equal)

Yang Cao, PhD (Formal analysis: Lead; Software: Equal; Writing – review & editing: Equal)

Dirk Reptsilber, PhD (Data curation: Equal; Formal analysis: Lead; Funding acquisition: Equal; Investigation: Equal; Software: Equal; Validation: Equal; Writing – review & editing: Equal)

Olof Grip, PhD (Data curation: Lead; Formal analysis: Equal; Investigation: Lead; Validation: Equal; Writing – review & editing: Equal)

Pontus Karling, MD, PhD (Data curation: Equal; Formal analysis: Equal; Investigation: Equal; Validation: Equal; Writing – review & editing: Equal)

Jonas Halfvarson, MD, PhD (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Funding acquisition: Lead; Investigation: Lead; Validation: Lead; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Equal)

Conflict of interest

These authors disclose the following: Charlotte R. H. Hedin has received speaker fees from Takeda, Ferring, AbbVie, and Janssen, as well as consultancy fees from Pfizer. She has also acted as the local principal investigator for clinical trials for Janssen and GlaxoSmithKline and received project grants from Takeda and Tillotts. Carl Eriksson received grant support, lecture fees, and served on an advisory board for BMS, Takeda, Janssen Cilag, Pfizer, and AbbVie. Jonas Halfvarson served as a speaker and/or advisory board member for AbbVie, AlfaSigma, BMS, Celgene, Celltrion, Dr Falk Pharma and the Falk Foundation, Eli Lilly, Ferring, Galapagos, Gilead, Hospira, Index Pharma, Janssen, Johnson & Johnson, MEDA, Medivir, Medtronic, Merck, MSD, Novartis, Pfizer, Prometheus Laboratories Inc., Sandoz, Shire, STADA, Takeda, Thermo Fisher Scientific, Tillotts Pharma, Vifor Pharma, and UCB. He received grant support from Janssen, MSD, and Takeda. The remaining authors disclose no conflicts.

Funding

This study was supported by the Swedish Research Council (grant number 2020-02021 to Jonas Halfvarson), the Swedish Foundation For Strategic Research (grant number RB13-016 to Jonas Halfvarson), the Örebro University Hospital Research Foundation (grant numbers OLL-986849, OLL-974710, OLL-936004, OLL-890291, OLL-790011, and OLL-962042 to Jonas Halfvarson), the Swedish Foundation for Gastrointestinal Research (Gunilla Falk award 2021 to Jonas Halfvarson), the Swedish state under the agreement between the Swedish government and the county councils, the ALF agreement (grant number OLL-961742 to Olle Grännö and OLL-685051 to Daniel Bergemalm), and the Bengt Ihre research foundation to Daniel Bergemalm. Västerbotten County Council funded the Västerbotten Intervention Program and Biobank Sweden was supported by the Swedish Research Council (grant number VR 2017-00650). This work was funded by the European Union under the Horizon Europe grant 101095470, project miGut-Health, Personalised Blueprint of Intestinal Health; however, the views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the HaDEA can be held responsible for them. The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

Data Availability

The data sets used for this paper, including proteomic data, cannot be shared directly under current data protection legislation and must be requested directly from the respective registry holders after approval by the Swedish Ethical Review Authority. The R code used for the statistical analysis is available at <https://git.oru.se/mh/ibd/preclinical-protein-signatures-in-ibd>.