scientific reports

Check for updates

OPEN

Plasma interleukin-22 concentration and disease activity in inflammatory bowel disease

Claudia Cabrera[1]✉, Daniel Fernández-Llaneza[2,10], Marita Olsson[3], Charles Sopp[2], Tara Fehlmann[4], Elizabeth A. Duncan[5,11], Emon Khan[6], Junmei Cairns[7], Jessica Neisen[8], Ulf Gehrmann[7,12] & Nick Powell[9]

The relationship between interleukin-22 and clinical characteristics of patients with inflammatory bowel disease is uncertain. We sought to determine whether plasma interleukin-22 concentrations are associated with disease activity in a large population of patients with inflammatory bowel disease. This was an observational study of patients with Crohn's disease or ulcerative colitis in the Study of a Prospective Adult Research Cohort with IBD (SPARC IBD) from the IBD Plexus registry of the Crohn's & Colitis Foundation in the United States. Interleukin-22 concentrations were measured using an R&D Systems human interleukin-22 enzyme-linked immunosorbent assay. Of 3993 patients included, 1604 with Crohn's disease and 836 with ulcerative colitis had non-missing interleukin-22 baseline samples, of which 33% and 25%, respectively, were above the lower limit of quantification (7.8 pg/mL). Significantly more patients with moderate/severe disease activity had high interleukin-22 concentrations (above lower limit of quantification) versus those with remission/mild disease activity (Crohn's disease: 44.7% vs. 30.0%, $P < 0.00001$; ulcerative colitis: 39.8% vs. 23.0%, $P < 0.001$). A multinomial logit model found approximately twice the odds of moderate/severe disease activity in those with high interleukin-22 versus low interleukin-22 (below lower limit of quantification). The strongest predictors of high interleukin-22 concentrations identified by a machine-learning algorithm were higher Short Crohn's Disease Activity Index score in patients with Crohn's disease and prescription of biologics in patients with ulcerative colitis. In summary, high interleukin-22 concentrations tended to be associated with higher disease activity in this large, real-world population of patients with inflammatory bowel disease.

Keywords Crohn's disease, Ulcerative colitis, Interleukin-22, Inflammatory bowel disease

Inflammatory bowel disease (IBD) is a collection of immune-mediated inflammatory diseases including Crohn's disease (CD) and ulcerative colitis (UC). Both CD and UC are associated with serious complications such as impaired gastrointestinal function, strictures, increased risk of colorectal cancer, extraintestinal complications, and impaired health-related quality of life, including anxiety, depression, fatigue, and work impairment[1–5]. In line with the traditional view of CD, UC is increasingly being recognized as progressive in nature[5]. Thus, early treatment with disease-modifying drugs is important in controlling inflammation and slowing the development of complications.

There is overlap in the pathophysiology of CD and UC[6,7]. The interleukin (IL)-23 pathway is a major driver of IBD[8–11]. IL-22 is induced by IL-23 and is a key cytokine in the pathogenesis of IBD[12,13]. IL-22 is produced by

[1]BioPharmaceutical Medical, Real World Science & Analytics, AstraZeneca, Pepparedsleden 1, 431 83 Gothenburg, Sweden. [2]BioPharmaceutical Medical, Real World Science & Analytics, AstraZeneca, Cambridge, UK. [3]Biometrics and Statistical Innovation, Late-Stage Development Respiratory & Immunology, AstraZeneca, Gothenburg, Sweden. [4]Crohn's & Colitis Foundation, New York, NY, USA. [5]Late Respiratory and Immunology (R&I), BioPharmaceuticals R&D, AstraZeneca, Durham, NC, USA. [6]Late Respiratory and Immunology (R&I), BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK. [7]Translational Science & Experimental Medicine, Research and Early Development, Respiratory and Immunology (R&I), BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden. [8]Translational Science & Experimental Medicine, Research and Early Development, Respiratory and Immunology (R&I), BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK. [9]Faculty of Medicine, Imperial College London, London, UK. [10]Present address: Amsterdam UMC, Amsterdam, Netherlands. [11]Present address: GSK, Durham, NC, USA. [12]Present address: Biomedical Research, Novartis, Basel, Switzerland. ✉email: claudia.s.cabrera@astrazeneca.com

different subsets of immune cells and has been shown to have protective effects on intestinal barrier function[14]. While IL-22 may be protective in acute tissue injury, chronic expression of IL-22 may promote inflammation and has been shown to exacerbate colitis[13,15–18].

Studies have suggested a relationship between IL-22 levels and IBD. For example, patients with CD and UC have increased IL-22 expression in colonic tissue[15,19]. Furthermore, IL-22 concentrations in colonic tissue in patients with UC and serum IL-22 concentrations in patients with CD have been found to correlate with disease activity in small populations of patients from a variety of clinical settings[12,18,20]. Further supporting a link between IL-22 and disease activity, one study showed that among patients receiving brazikumab, an IL-23 inhibitor, remission rates were higher in patients with high serum IL-22 than those with low serum IL-22 at baseline[21]. Thus, circulating IL-22 concentrations may have clinical relevance as a biomarker in patients with IBD.

Given the potential relevance of IL-22 as a biomarker, it is important to understand which diseases and clinical characteristics impact IL-22 concentrations. Real-world studies are critical to confirm this association. IBD Plexus is a US patient registry owned by the Crohn's & Colitis Foundation that consists of several cohorts, including the Study of a Prospective Adult Research Cohort with IBD (SPARC IBD) cohort[22]. The SPARC IBD cohort enrolls patients diagnosed with CD, UC, or unclassified IBD. The aim of this observational study was to evaluate the association of IL-22 concentrations and the severity of CD and UC, with higher IL-22 concentrations expected to be associated with greater disease severity based on the literature described above.

Methods

Study design

This is a cross-sectional study of patients in the SPARC IBD cohort from the IBD Plexus registry, a data ecosystem owned by the Crohn's & Colitis Foundation. This study used patient data entered in SPARC IBD from initiation in 2016 to January 2022. Patients from SPARC IBD were divided into two cohorts: those with CD and those with UC.

The use of SPARC IBD and blood plasma samples for this study was approved by the Crohn's & Colitis Foundation in 2021. This study was conducted according to US and international standards of Good Clinical Practice (FDA Title 21 part 312 and International Conference on Harmonization guidelines), applicable government regulations and institutional research policies and procedures. The study protocol was approved by the Institutional Review Board (IRB) at the University of Pennsylvania, which is the single IRB for the SPARC IBD study.

Patients

Adult patients (age 18 years or older) with either CD or UC diagnosis and an identifiable index date were included in the study. Patients were excluded if they had indeterminate colitis (e.g., having a diagnosis of both CD and UC, or unspecified IBD). All patients enrolled in the study provided informed consent to participate through a universal consent form authorizing use of their personal information for future IBD studies. These data include information about health and health-related issues, information collected from the medical record, insurance claims data, and biosamples collected as part of the research study. SPARC has allowed for both paper and e-consent since 2020 (due to the COVID-19 pandemic), prior to which consent was obtained via a paper-based form.

Assessments

Data were collected from patient- and provider-reported surveys along with individual electronic medical records and structured electronic case report forms. Patient blood plasma samples stored in SPARC IBD were used to measure IL-22 concentrations. IL-22 concentrations were measured using an R&D Systems (Minneapolis, MN, USA) human IL-22 enzyme-linked immunosorbent assay kit according to the manufacturer's instructions. In brief, samples were thawed to room temperature, diluted 1:1 in assay diluent RD1-88 and incubated for 2 h at room temperature in pre-coated 96-well microplate strips. Samples were analyzed in duplicates. Following incubation, wells were washed 4 times using wash buffer, 200 µL human anti-IL-22 conjugate was added and incubated for 2 h at room temperature, wells were again washed 4 times, and 200 µL substrate solution was added and incubated for 30 min at room temperature before 50 µL stop solution was added. Within 30 min, optical density of the final solution was determined using a microplate reader set to 450 nm, with wavelength correction at 540 nm. Final IL-22 concentrations were extrapolated using optical density measurements and the standard curve measurements, run in parallel on each plate, and fitted with a 4-parameter curve with inverse response weighting (1/Y).

IL-22 concentrations were categorized as low (at or below the lower limit of quantitation [LLOQ]) or high (> LLOQ). Baseline demographics were assessed descriptively in the overall population and disease characteristics were assessed descriptively both in the overall population and stratified by IL-22 level. IL-17A was measured by high-sensitivity single-molecule array (Simoa) assay (Rules-Based Medicine, Austin TX, USA).

All clinical data considered for any analysis were assessed during the baseline period, which is defined as within 30 days of the index date (the index date was set as the enrollment date as identified through date of consent). If more than one value of a variable was registered during the baseline period, the value closest in time to the index date was defined as the baseline value.

Statistical analyses

This study is a hypothesis-generating, exploratory study; no formal power calculations were performed. All available patients in the CD and UC cohorts were included.

Cluster analysis

An unsupervised cluster analysis was performed for each IL-22 cohort (IL-22 all, IL-22 high, and IL-22 low) to identify groups of patients with similar characteristics. Input variables included demographics (age, sex, race), disease characteristics (disease duration, 6-point Ulcerative Colitis Disease Activity Score [UCDAI] or short Crohn's Disease Activity Index [sCDAI] score, phenotype, extraintestinal manifestations), symptom variables, treatments, and comorbidities. Only variables with ≤ 30% missing values were included. Categorical variables included had a minor category proportion of at least 5%. A dissimilarity matrix was created by averaging 5 separate dissimilarity matrices. These matrices were generated using a random forest model with 2000 decision trees. In building each tree, a standard approach of selecting a random subset of variables was followed, specifically the square root of the total number of variables. The centroid clustering model algorithm partitioning around the medoid (PAM) was used to select clusters.

Results from the cluster analysis were visualized by plotting in a 2-dimensional space using uniform manifold approximation and projection (UMAP) as a dimensionality reduction technique.

Association model

The association between IL-22 concentration and disease activity was assessed using a multinomial logit model. The response variable was disease activity, which had 3 categories: remission (sCDAI < 150 or 6-point UCDAI ≤ 1), mild (sCDAI 150–219 or 6-point UCDAI 2–3), and moderate/severe (sCDAI ≥ 220 or 6-point UCDAI ≥ 4). The remission category was used as reference. The exposure was IL-22 level, categorized as IL-22 low (< LLOQ) and high (≥ LLOQ). Odds ratios (ORs) for having mild or moderate/severe disease activity for those with higher IL-22 were compared with the odds for those with low IL-22. The following covariates were identified as potential confounders or risk factors and were adjusted for in the model: age, sex, disease duration, comorbidities, and main treatment classes. The variable selection for the association model was based on a combination of evidence from the literature, discussion with medical experts, and statistical considerations (e.g., amount of missing data and model robustness considerations). Only variables with ≤ 10% missingness were included.

Relevance of variables predictive of IL-22 level: eXtreme gradient boosting (XGBoost) and SHapley additive explanations (SHAP)

XGBoost[23] was utilized to classify IL-22 concentration levels as high or low based on available demographic and clinical features (variables). The resulting model was used to interpret IL-22 level predictions using SHAP values[24]. A stratified train-test split with a 70:30 ratio was employed, which preserves the proportion of the 2 (binary) classes (i.e., high IL-22, low IL-22) in the training and test set. The training set was used for the XGBoost model to learn patterns in the data by utilizing the available variables per patient and their corresponding IL-22 concentration label (high or low; category). The test set was used for inference and assessing the performance of the model with area under the receiver operating curve (ROC-AUC) (i.e., extent to which the model correctly discriminates between the 2 labels). To improve the performance of the XGBoost model, we carried out hyperparameter tuning. A detailed account of the scanned hyperparameters and the selected values for attaining optimal performance are provided in Table S1, Online Resource 1. Finally, SHAP values were estimated using the trained XGBoost model. SHAP values allow interpretation of why the model predicts a specific IL-22 concentration category per patient based on the available variables. For a given variable, a high positive SHAP value is predictive of high IL-22 and a negative SHAP value is predictive of low IL-22.

Software

Python version 3.9.7 and R version 4.2.1 were used for the cluster analysis. For clustering, the randomForest version 4.7–1.1 and diceR version 1.2.2 packages were used. The Python UMAP package version 0.5.3 was used to produce UMAP plots for each cohort. Python version 3.9.7 and package XGBoost version 1.7.1 with Scikit learn version 1.1.3 were used for implementing the XGBoost model, and SHAP values were estimated and visualized with SHAP package version 0.41.0. The whole analysis was carried out with the random seed set at 42.

Results

Population characteristics

Of 2830 patients with CD and 1431 patients with UC in the registry, 2667 patients with CD and 1326 with UC had clinical data available (Figure S1, Online Resource 1). For patients with CD, median age was 39.0 years, 56.7% were female, and median disease duration was 13.0 years (Table 1). For patients with UC, median age was 40.0 years, 50.5% were female, and median disease duration was 9.0 years (Table 1).

The final study population only included patients with plasma available for quantifying IL-22 and a non-missing IL-22 value (Figure S1, Online Resource 1). Plasma IL-22 concentrations were analyzed in 1604 patients with CD and 836 patients with UC (Figure S1, Online Resource 1). Among patients with IL-22 values, 32.9% of the CD cohort and 25.2% of the UC cohort were above the LLOQ (7.8 pg/mL). For patients with CD, median (Q1, Q3) IL-22 concentrations were < LLOQ (< LLOQ, 10.8 pg/mL). For patients with UC, median (Q1, Q3) IL-22 concentrations were < LLOQ (< LLOQ, 7.9 pg/mL). Most patients were in clinical remission based on sCDAI or 6-point UCDAI (CD: 59.8%; UC: 63.8%) (Table 2). IBD-associated arthropathy was the most common extraintestinal manifestation. Rates of extraintestinal manifestations were similar between patients with low and high IL-22 concentrations. Patients with CD were most commonly receiving treatment with anti–tumor necrosis factor-α (TNFα) therapy, immunomodulators, and corticosteroids at baseline. Patients with UC were most commonly receiving treatment with aminosalicylates, anti-TNFα therapy, and corticosteroids at baseline. Greater proportions of patients with high IL-22 concentrations were receiving treatment with biologics than patients with low IL-22 (Table 2). Anti-TNFα therapy in particular was more common in those with high IL-22 concentrations (CD: 65.5%, UC: 58.8%) than those with low IL-22 concentrations (CD: 54.2%, UC: 39.8%).

	CD ($n=2667$)	UC ($n=1326$)
Age, median (Q1, Q3), years	39.0 (29.0, 52.0)	40.0 (30.0, 54.0)
Female, n (%)	1513 (56.7)	670 (50.5)
Disease duration, median (Q1, Q3), years	13.0 (7.0, 21.0)	9.0 (4.0, 16.0)
Race and ethnicity, n (%)		
White not Hispanic	1961 (73.5)	1001 (75.5)
Black not Hispanic	201 (7.5)	56 (4.2)
Hispanic	37 (1.4)	33 (2.5)
Asian not Hispanic	26 (1.0)	33 (2.5)
Other not Hispanic	16 (0.6)	17 (1.3)
Missing	426 (16.0)	186 (14.0)

Table 1. Patient demographics. *CD* Crohn's disease, *Q* quartile, *UC* ulcerative colitis.

Association of plasma IL-22 concentration and disease severity/activity

We evaluated plasma IL-22 concentrations in CD and UC patients stratified by disease activity. In patients with CD, significantly ($P<0.00001$) more patients had low IL-22 concentrations among those in remission or with mild disease activity (70.0%) compared with those with moderate/severe disease activity (55.3%), and conversely, significantly ($P<0.00001$) more patients with moderate/severe disease activity had high IL-22 concentrations (44.7%) than those in remission or with mild disease activity (30.0%) (Fig. 1A). A similar pattern was seen in patients with UC ($P<0.001$; Fig. 1B). Furthermore, among patients with moderate/severe disease activity, more patients had IL-22 concentrations above the third quartile compared with those in remission/low disease activity, both in patients with CD and those with UC (Fig. 2). Comparison of subgroups of patients with low and high IL-22 concentrations showed a consistent pattern, with a lower proportion of patients in remission and a greater proportion of patients with moderate/severe disease in the high IL-22 group than in the low IL-22 group (Table S2, Online Resource 1).

Cluster analysis

An unsupervised cluster analysis was performed to determine whether any demographic or disease characteristics cluster together in IBD patients and whether this varies by IL-22 concentration. Three distinct clusters were identified for patients with CD in the IL-22 all cohort (Fig. 3A). The first cluster consisted mainly of patients with moderate and severe disease activity (Table S3, Online Resource 1). In the second cluster, 99.3% of patients were in remission, and in the last cluster, 62.0% of patients were in remission and 24.1% had mild disease activity. Among patients with IL-22 high, two clusters were identified (Fig. 3A), characterized mainly by differences in disease activity (Table S3, Online Resource 1). In the population of patients with IL-22 low, four clusters were identified, which seemed to be mainly characterized by differences both in age and disease activity (Fig. 3A, Table S3, Online Resource 1).

Among patients with UC, three clusters were identified in the IL-22 all cohort (Fig. 3B). The first cluster consisted of 100% of patients in remission, the second consisted of mostly patients in remission and mild disease, and the third consisted of a combination of patients with mild (26.9%), moderate (51.3%), and severe (18.5%) disease activity (Table S4, Online Resource 1). Similar clusters were seen in patients with high IL-22 (Fig. 3B, Table S4, Online Resource 1). In patients with low IL-22, however, there were only two clusters consisting mostly of patients in remission and another of patients in remission and mild disease activity (Fig. 3B, Table S4, Online Resource 1). Median age, disease duration, and treatment varied across clusters.

Association model

IL-22 concentrations were found to be associated with disease activity when comparing ORs for risk of moderate/severe disease activity or mild disease activity using a multinomial logit model. The risk of moderate/severe disease activity versus remission was 2 times higher in patients with CD (OR: 2.02; 95% confidence interval [CI]: 1.52, 2.69) and 2.7 times higher in patients with UC (OR: 2.71; 95% CI: 1.70, 4.32) among those with high IL-22 concentrations versus those with low IL-22 concentrations (Fig. 4). The ORs (95% CI) for risk of mild disease versus remission for patients with high IL-22 versus low IL-22 were 1.30 (0.98, 1.72) for CD and 1.66 (1.12, 2.46) for UC.

XGBoost and SHAP value analysis

XGBoost was implemented to learn how to classify patients' IL-22 concentration into low or high based on available features. The XGBoost model performance in the test set was evaluated with ROC-AUC. ROC-AUC was 0.63 for the CD cohort and 0.62 for the UC cohort. Based on subsequent SHAP analysis, sCDAI, IL-17A, and biologic therapy (anti-TNFα) were the 3 most relevant features for predicting IL-22 concentration (Fig. 5A). Directional assessment revealed that higher sCDAI and IL-17A, as well as a prescription for biologic therapy (anti-TNFα), were predictive of high IL-22 in patients with CD (Fig. 5A). In patients with UC, the 3 most relevant features for predicting IL-22 concentration were biologic therapy (anti-TNFα), 6-point UCDAI score, and disease activity (Fig. 5B). When introducing directional assessment of SHAP values, there was a clear higher likelihood of predicting high IL-22 concentration when a patient had a prescription for biologic therapy (anti-

	CD			UC		
	IL-22 low[a] (*n* = 1077)	IL-22 high[b] (*n* = 527)	Total (*n* = 1604)	IL-22 low[a] (*n* = 625)	IL-22 high[b] (*n* = 211)	Total (*n* = 836)
sCDAI/6-point UCDAI category, n (%)						
Remission (sCDAI < 150/6-point UCDAI ≤ 1)	686 (63.7)	273 (51.8)	959 (59.8)	426 (68.2)	107 (50.7)	533 (63.8)
Mild (sCDAI 150–219/6-point UCDAI 2–3)	216 (20.1)	114 (21.6)	330 (20.6)	133 (21.3)	60 (28.4)	193 (23.1)
Moderate (sCDAI 220–450/6-point UCDAI 4–5)	165 (15.3)	125 (23.7)	290 (18.1)	55 (8.8)	30 (14.2)	85 (10.2)
Severe (sCDAI > 450/6-point UCDAI 6)	6 (0.6)	13 (2.5)	19 (1.2)	10 (1.6)	13 (6.2)	23 (2.8)
Missing	4 (0.4)	2 (0.4)	6 (0.4)	1 (0.2)	1 (0.5)	2 (0.2)
CD phenotype						
Inflammatory, non-penetrating, non-stricturing	329 (30.5)	130 (24.7)	459 (28.6)	–	–	–
Stricturing	176 (16.3)	96 (18.2)	272 (17.0)	–	–	–
Penetrating	138 (12.8)	84 (15.9)	222 (13.8)	–	–	–
Both stricturing and penetrating	174 (16.2)	101 (19.2)	275 (17.1)	–	–	–
Missing	260 (24.1)	116 (22.0)	376 (23.4)	–	–	–
CD location						
Colonic	106 (9.8)	50 (9.5)	156 (9.7)	–	–	–
Ileal	148 (13.7)	51 (9.7)	199 (12.4)	–	–	–
Ileocolonic	396 (36.8)	223 (42.3)	619 (38.6)	–	–	–
Upper GI only	N/A	1 (0.2)	1 (0.1)	–	–	–
Missing	420 (39.0)	199 (37.8)	619 (38.6)	–	–	–
UC phenotype						
Ulcerative proctitis	–	–	–	61 (9.8)	21 (10.0)	82 (9.8)
Left-sided	–	–	–	121 (19.4)	32 (15.2)	153 (18.3)
Extensive	–	–	–	45 (7.2)	20 (9.5)	65 (7.8)
Pancolitis	–	–	–	253 (40.5)	73 (34.6)	326 (39.0)
Missing	–	–	–	145 (23.2)	65 (30.8)	210 (25.1)
Extraintestinal manifestations, n (%)						
IBD-associated arthropathy	148 (13.7)	90 (17.1)	238 (14.8)	75 (12.0)	23 (10.9)	98 (11.7)
Aphthous ulcer	50 (4.6)	33 (6.3)	83 (5.2)	23 (3.7)	13 (6.2)	36 (4.3)
Erythema nodosum	25 (2.3)	14 (2.7)	39 (2.4)	6 (1.0)	2 (0.9)	8 (1.0)
Pyoderma gangrenosum	5 (0.5)	8 (1.5)	13 (0.8)	3 (0.5)	N/A	3 (0.4)
Primary sclerosing cholangitis	22 (2.0)	2 (0.4)	24 (1.5)	32 (5.1)	5 (2.4)	37 (4.4)
Thrombotic complications	11 (1.0)	5 (0.9)	16 (1.0)	3 (0.5)	7 (3.3)	10 (1.2)
None reported	782 (72.6)	356 (67.6)	1138 (70.9)	477 (76.3)	157 (74.4)	634 (75.8)
Treatment received at baseline[c], n (%)						
Aminosalicylates	176 (16.3)	93 (17.6)	269 (16.8)	349 (55.8)	105 (49.8)	454 (54.3)
Antibiotics[d]	30 (2.8)	27 (5.1)	57 (3.6)	–	–	–
Biologics						
Anti-TNFα	584 (54.2)	345 (65.5)	929 (57.9)	249 (39.8)	124 (58.8)	373 (44.6)
Anti-integrins	175 (16.2)	96 (18.2)	271 (16.9)	172 (27.5)	49 (23.2)	221 (26.4)
Ustekinumab	322 (29.9)	179 (34.0)	501 (31.2)	76 (12.2)	36 (17.1)	112 (13.4)
Corticosteroids	313 (29.1)	190 (36.1)	503 (31.4)	215 (34.4)	82 (38.9)	297 (35.5)
Immunomodulators[e]	409 (38.0)	230 (43.6)	639 (39.8)	209 (33.4)	83 (39.3)	292 (34.9)
JAK inhibitor	10 (0.9)	9 (1.7)	19 (1.2)	50 (8.0)	29 (13.7)	79 (9.4)
Number of IBD surgeries						
0	381 (35.4)	170 (32.3)	551 (34.4)	344 (55.0)	125 (59.2)	469 (56.1)
1	75 (7.0)	42 (8.0)	117 (7.3)	8 (1.3)	2 (0.9)	10 (1.2)
2+	78 (7.2)	25 (4.7)	103 (6.4)	10 (1.6)	1 (0.5)	11 (1.3)
Missing	543 (50.4)	290 (55.0)	833 (51.9)	263 (42.1)	83 (39.3)	346 (41.4)

Table 2. Disease characteristics in the IL-22 cohort. [a]IL-22 ≤ LLOQ. [b]IL-22 > LLOQ. [c]Patients may have received ≥ 1 type of treatment. [d]Including ciprofloxacin, clarithromycin, metronidazole, and rifaximin. [e]Including azathioprine, cyclosporine, mercaptopurine, methotrexate, and tacrolimus. *CD* Crohn's disease, *IBD* inflammatory bowel disease, *IL-22* interleukin-22, *JAK* Janus kinase, *TNF* tumor necrosis factor, *sCDAI* short Crohn's disease activity index measuring disease severity, *UC* ulcerative colitis, *UCDAI* ulcerative colitis disease activity index measuring disease severity.

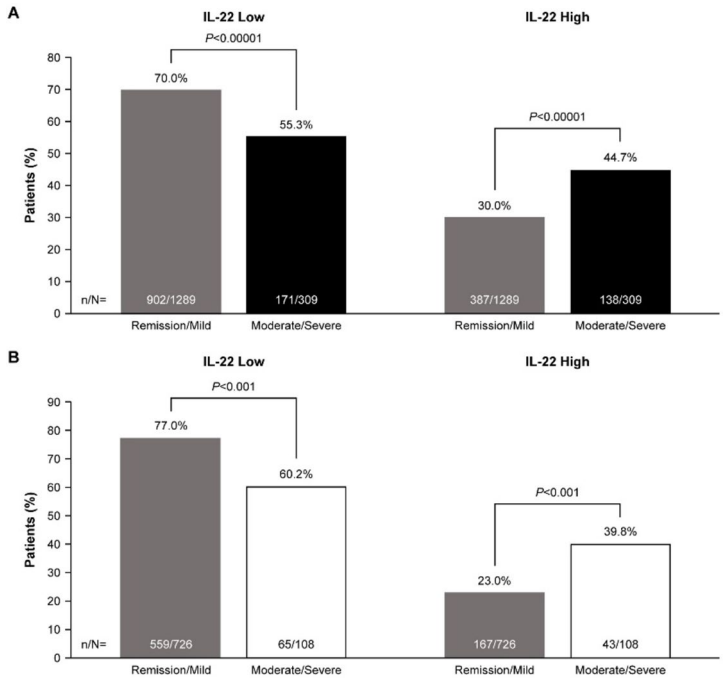

Fig. 1. Proportions of patients with low[a] and high[b] IL-22 by disease activity. [a]IL-22 ≤ LLOQ. [b]IL-22 > LLOQ. LLOQ = 7.8 pg/mL. (**A**) CD. (**B**) UC. *P*-values were calculated by chi-square test. *CD* Crohn's disease, *IL-22* interleukin-22, *LLOQ* lower limit of quantification, *Q3* third quartile, *UC* ulcerative colitis.

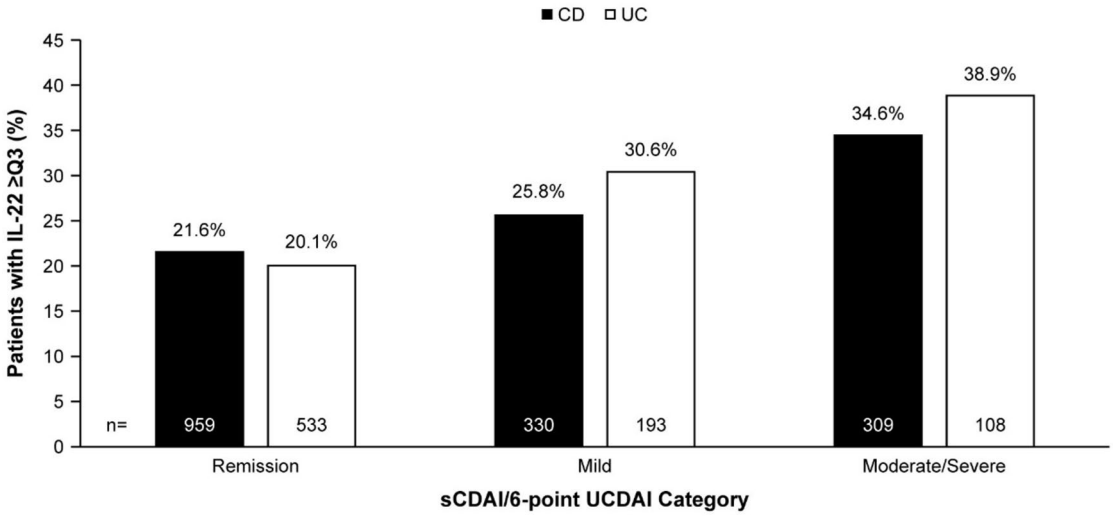

Fig. 2. Proportions of patients with plasma IL-22 concentrations ≥ Q3[a] by disease activity. [a]Q3 = 10.8 pg/mL in CD, 7.9 pg/mL in UC. *CD* Crohn's disease, *IL-22* interleukin-22, *Q3* third quartile, *sCDAI* short Crohn's disease activity index, *UC* ulcerative colitis, *UCDAI* ulcerative colitis disease activity index.

TNFα), a high 6-point UCDAI score, or greater disease activity (Fig. 5B). Patient-level data showing an example of SHAP value analysis to provide local (per-patient) interpretability is shown in Figure S2, Online Resource 1.

Discussion

This is the first study to analyze plasma IL-22 samples in a large population of real-world patients with IBD with detailed clinical characterization. Plasma IL-22 concentrations were broadly similar in patients with CD and UC, with low concentrations more commonly observed in patients in remission, and higher levels observed in patients with moderate or severe disease activity. In an unsupervised cluster analysis, three clusters were observed based on disease activity for both CD and UC: a cluster with mixed moderate/high disease activity patients, a cluster with mixed mild/remission disease activity patients, and a cluster with patients in remission.

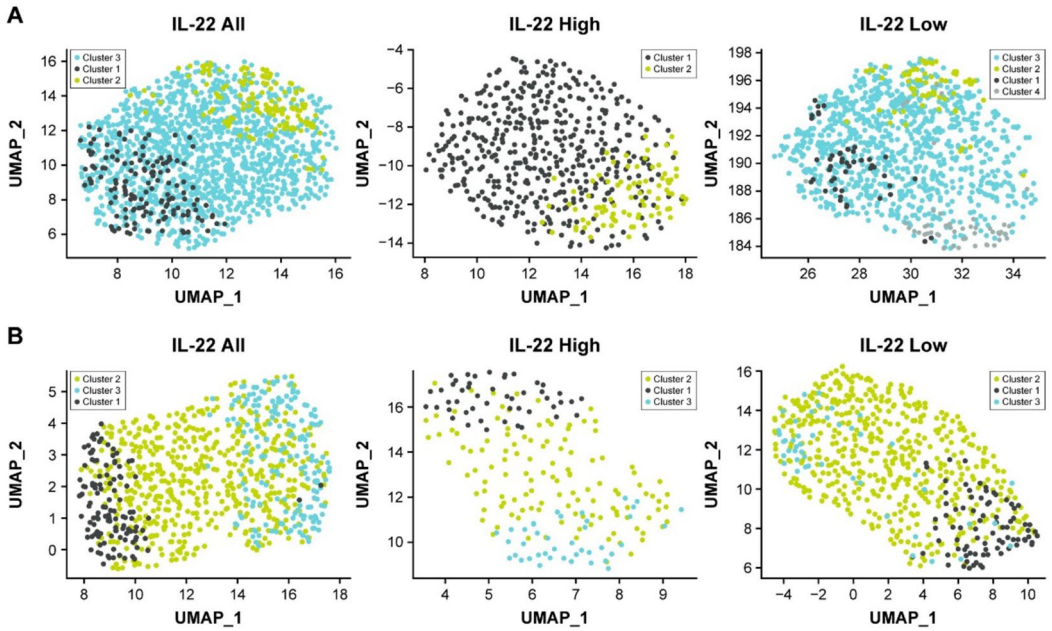

Fig. 3. UMAP plot of clustering in the IL-22 all, IL-22 high[a], and IL-22 low[b] cohorts. (**A**) CD. (**B**) UC. Each data point corresponds to 1 patient. The high-dimensional variables for each patient have been mapped to a 2-dimensional setting, allowing for visualization and interpretation of the clustering analysis. Each cluster is shown with a different color. [a]IL-22 > LLOQ. [b]IL-22 ≤ LLOQ. *CD* Crohn's disease, *IL-22* interleukin-22, *UMAP* uniform manifold approximation and projection, *UC* ulcerative colitis.

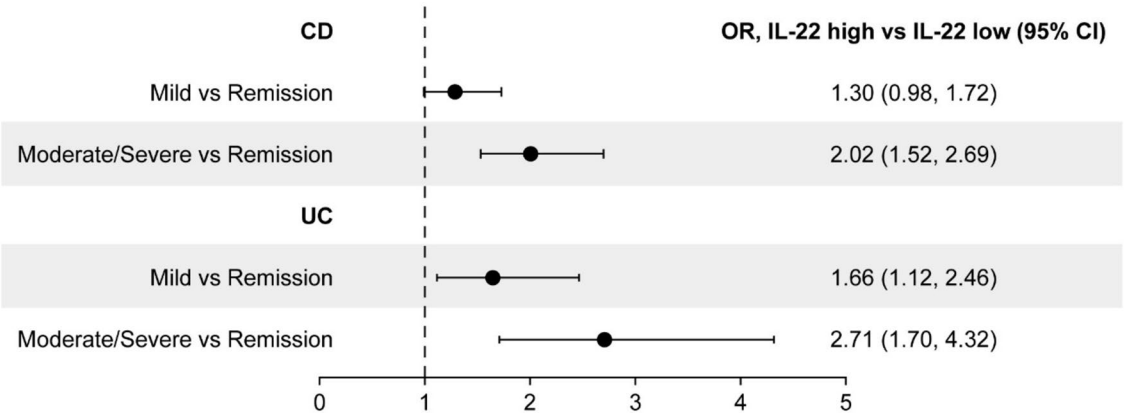

Fig. 4. Odds ratios for disease activity among patients with high IL-22 vs. low IL-22. Based on a multinomial logit model adjusted for age, sex, disease duration, comorbidities (any/none), and main treatment classes (aminosalicylates, immunomodulators, anti-TNFα, anti-integrins, ustekinumab, corticosteroids). An OR of 1 indicates no difference between IL-22 high and low groups. An OR > 1 indicates greater risk in the IL-22 high group relative to the IL-22 low group. *CD* Crohn's disease, *CI* confidence interval, *IL-22* interleukin-22, *OR* odds ratio, *TNF* tumor necrosis factor, *UC* ulcerative colitis.

Cohorts with high concentrations of plasma IL-22 clustered according to disease activity, while in the cohort with low concentrations of plasma IL-22, disease activity did not differentiate between clusters. Overall, the results of the cluster analysis indicated that disease activity was a major differentiating factor in patients overall and among those with high IL-22.

A multinomial logit model confirmed the association between greater disease activity and higher IL-22 concentration, showing ≥ 2 times greater odds of moderate/severe disease versus remission with high concentrations of plasma IL-22 versus low IL-22 concentrations for both CD and UC. A similar pattern was seen for mild disease, with greater odds of mild disease versus remission among those with high IL-22 versus low IL-22. Our results are in agreement with an earlier observational analysis of blood and intestinal biopsies from the Mount Sinai Crohn's and Colitis registry of ~ 1200 patients with IBD and transcriptomic data from 7 other smaller independent IBD datasets. This study identified a circulating molecular inflammation score that

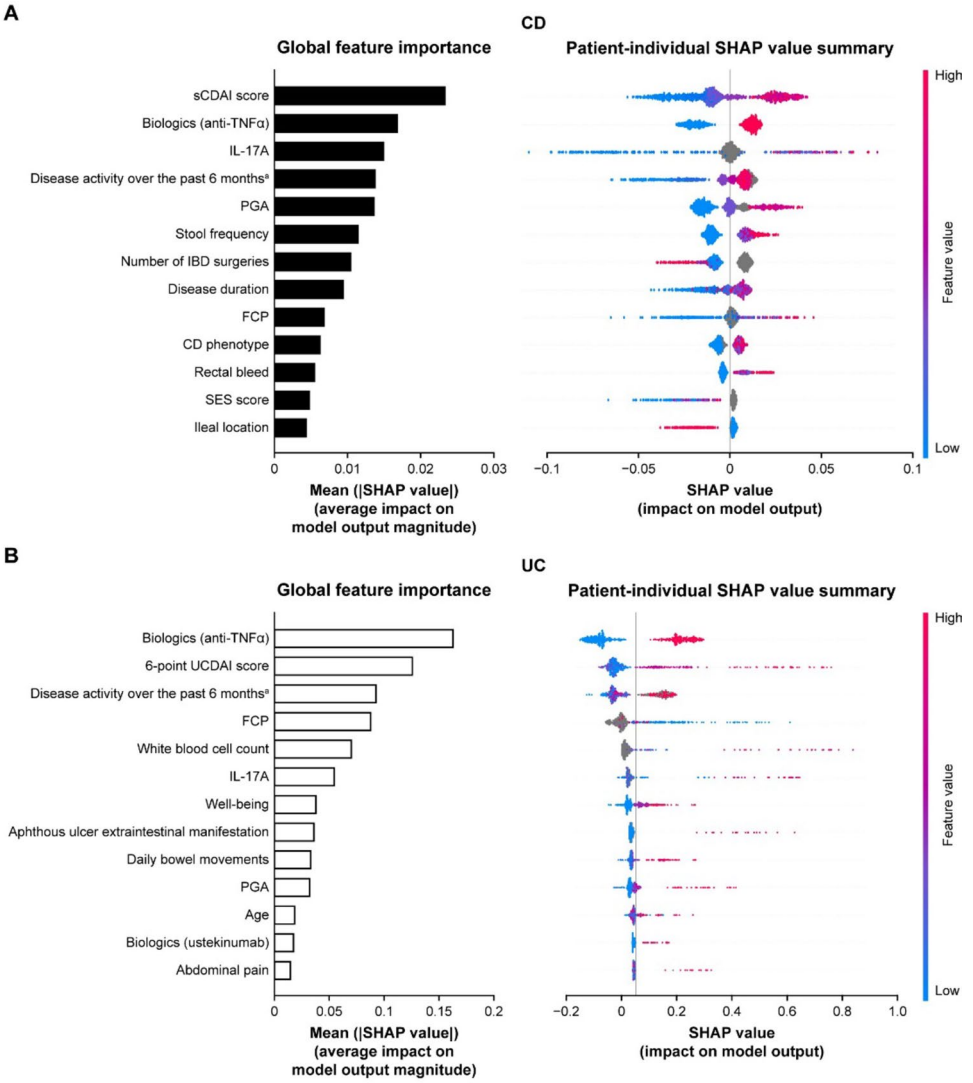

Fig. 5. Relevance of features predictive of IL-22 concentration based on SHAP values. (**A**) CD. (**B**) UC. The left-hand graph provides a general assessment of feature (variable) importance for the model to generate a prediction on IL-22 concentration level. The higher the SHAP value, the higher the contribution of a feature toward predicting the label. Thus, the most influential features are presented at the top. The right-hand graph provides additional detail by showing the directional impact of each feature. Each data point corresponds to 1 patient and is color-coded based on the value of the feature: red indicates high values for a given feature (e.g., a high sCDAI score), blue indicates low values (e.g., a low sCDAI score), and gray indicates a missing value (e.g., no recorded sCDAI score). The y-axis lists the features in order of importance, and the x-axis shows the magnitude of the SHAP value, indicating the extent to which each feature influences the model's prediction. The greater the magnitude of the SHAP value – whether positive or negative – the stronger the impact of that feature on the prediction. A positive SHAP value for a data point indicates that the feature pushes the prediction toward the high IL-22 concentration label (category), while a negative SHAP value for a data point pushes it toward the low IL-22 concentration label. [a]Included the categories "absence of symptoms," "rarely active," "sometimes active," "occasionally active," "often active," and "constantly active." *CD* Crohn's disease, *FCP* fecal calprotectin, *IBD* inflammatory bowel disease, *IFN* interferon, *IL* interleukin, *PGA* physician's general assessment, *sCDAI* short Crohn's disease activity index, *SES* simple endoscopic score, *TNF* tumor necrosis factor, *UC* ulcerative colitis, *UCDAI* ulcerative colitis disease activity index.

correlated with disease activity and showed a high level of similarity between CD and UC[25]. Based on our results, IL-22 may be one factor that makes up this molecular signature.

Machine learning has been recognized as a tool that has the potential to aid clinicians in clinical practice by helping to determine diagnoses, inform treatment decisions, and predict patient prognosis[26]. Machine learning can also help clinicians to better understand complex data. This study introduced SHAP values to provide interpretability to the results obtained by a novel machine learning-based approach. Thus, we identify and explain how factors are associated with high or low plasma IL-22 concentration, which has not yet been reported

in IBD. In agreement with the findings of the previous analyses, greater disease activity (as measured by sCDAI and 6-point UCDAI score) was shown to be an important predictor of high IL-22 plasma concentrations for CD and UC, as indicated by the high SHAP values in patients with high sCDAI/UCDAI scores. There were some similarities in features most strongly predicting high IL-22 plasma concentrations observed between CD and UC. Notably, 3 of the top 4 predictors of high IL-22 concentrations were the same for CD and UC: prescription of biologics (anti-TNFα) therapy, greater daily disease activity, and higher plasma concentrations of IL-17A. The prescription of anti-TNFα therapy and higher IL-17A levels are associated with increased severity of disease, which may explain this association; anti-TNFα therapy is reserved for patients with more active disease that have not responded to conventional immunomodulatory therapy[27], and IL-17 is a proinflammatory cytokine secreted by the same family of immune cells as IL-22, including Th17 cells and group 3 innate lymphoid cells, with its production similarly induced by IL-23[28]. The moderate discriminating power of the model (ROC-AUC = 0.6) suggests a fair ability to distinguish between high and low IL-22. However, this may still limit well-grounded explanations of additional features.

Current IBD biomarkers such as C-reactive protein (CRP) may present limitations in terms of sensitivity or specificity[29]. Approximately 55% of patients with mild UC and 10–20% with moderate to severe UC can have a CRP value within the normal range[30]. Thus, the use of additional biomarkers such as IL-22 or existing clinical markers such as fecal calprotectin (FCP) may help to provide a more complete picture. Our study suggests there is potential to incorporate use of IL-22 in IBD monitoring. Further study is needed to fully understand how IL-22 relates to IBD disease severity and how its use as a biomarker may be implemented in clinical practice. Assessment of differences between CD and UC, subgroup analyses by prior treatment type, and longitudinal analyses to assess the relationship between IL-22 and disease progression may be areas for future analysis.

Limitations

As these data come from a real-world registry, not every patient will have a measurement for every variable. The number of missing values was high for some variables, such as FCP and CRP. Furthermore, the generalizability of the SPARC IBD CD and UC cohorts may be compromised by the lack of representativeness across patient racial and ethnicity groups in the US, and potentially also by disease severity, as most patients were mild or in remission status. Due to the high proportion of plasma IL-22 concentrations below the LLOQ, the discriminating power of the XGBoost classification task in these cohorts was biased toward low IL-22. Although we could not assess the generalizability of the trained XGBoost model because of the limited amount of data, we tried to mitigate this by tuning regularization hyperparameters, which helps to improve the model's ability to make predictions on new, unseen patients.

Conclusion

This is one of the largest studies evaluating the distribution of plasma IL-22 concentrations in a real-world population of patients with CD and UC with granular phenotypic data. Across multiple assessments, a consistent association was found between more severe disease activity and high concentrations of IL-22 in the plasma in patients with CD and UC. These data indicate that high concentrations of circulating IL-22 are related to severe IBD, and are consistent with IL-22 being a suitable biomarker for IBD disease activity.

Data availability

The SPARC IBD data are available upon approved application to Crohn's & Colitis Foundation IBD Plexus (https://www.crohnscolitisfoundation.org/ibd-plexus).

Received: 12 November 2024; Accepted: 9 May 2025
Published online: 19 August 2025

References

1. Nowakowski, J., Chrobak, A. A. & Dudek, D. Psychiatric illnesses in inflammatory bowel diseases - psychiatric comorbidity and biological underpinnings. *Psychiatr. Pol.* **50**, 1157–1166 (2016).
2. Panara, A. J. et al. The incidence and risk factors for developing depression after being diagnosed with inflammatory bowel disease: a cohort study. *Aliment. Pharmacol. Ther.* **39**, 802–810 (2014).
3. Mikocka-Walus, A., Knowles, S. R., Keefer, L. & Graff, L. Controversies revisited: a systematic review of the comorbidity of depression and anxiety with inflammatory bowel diseases. *Inflamm. Bowel Dis.* **22**, 752–762 (2016).
4. van Gennep, S. et al. Impaired quality of working life in inflammatory bowel disease patients. *Dig. Dis. Sci.* **66**, 2916–2924 (2021).
5. Le Berre, C., Ananthakrishnan, A. N., Danese, S., Singh, S. & Peyrin-Biroulet, L. Ulcerative colitis and Crohn's disease have similar burden and goals for treatment. *Clin. Gastroenterol. Hepatol.* **18**, 14–23 (2020).
6. Petagna, L. et al. Pathophysiology of Crohn's disease inflammation and recurrence. *Biol. Direct*. **15**, 23 (2020).
7. Porter, R. J., Kalla, R. & Ho, G. T. Ulcerative colitis: recent advances in the Understanding of disease pathogenesis. *F1000Research* **9**, 294 (2020).
8. Holtta, V. et al. IL-23/IL-17 immunity as a hallmark of Crohn's disease. *Inflamm. Bowel Dis.* **14**, 1175–1184 (2008).
9. Geremia, A. et al. IL-23-responsive innate lymphoid cells are increased in inflammatory bowel disease. *J. Exp. Med.* **208**, 1127–1133 (2011).
10. Croxford, A. L., Mair, F. & Becher, B. IL-23: one cytokine in control of autoimmunity. *Eur. J. Immunol.* **42**, 2263–2273 (2012).
11. Feagan, B. G. et al. Induction therapy with the selective interleukin-23 inhibitor Risankizumab in patients with moderate-to-severe Crohn's disease: a randomised, double-blind, placebo-controlled phase 2 study. *Lancet* **389**, 1699–1709 (2017).
12. Schmechel, S. et al. Linking genetic susceptibility to Crohn's disease with Th17 cell function: IL-22 serum levels are increased in Crohn's disease and correlate with disease activity and IL23R genotype status. *Inflamm. Bowel Dis.* **14**, 204–212 (2008).
13. Pavlidis, P. et al. Interleukin-22 regulates neutrophil recruitment in ulcerative colitis and is associated with resistance to ustekinumab therapy. *Nat. Comm.* **13**, 5820 (2022).
14. Keir, M., Yi, Y., Lu, T. & Ghilardi, N. The role of IL-22 in intestinal health and disease. *J. Exp. Med.* **217**, e20192195 (2020).

15. Andoh, A. et al. Interleukin-22, a member of the IL-10 subfamily, induces inflammatory responses in colonic subepithelial myofibroblasts. *Gastroenterology* **129**, 969–984 (2005).
16. Zhao, N. et al. Role of Interleukin-22 in ulcerative colitis. *Biomed. Pharmacother.* **159**, 114273 (2023).
17. Wei, H. X., Wang, B. & Li, B. IL-10 and IL-22 in mucosal immunity: driving protection and pathology. *Front. Immunol.* **11**, 1315 (2020).
18. Powell, N. et al. Interleukin-22 orchestrates a pathological Endoplasmic reticulum stress response transcriptional programme in colonic epithelial cells. *Gut* **69**, 578–590 (2020).
19. Brand, S. et al. IL-22 is increased in active Crohn's disease and promotes Proinflammatory gene expression and intestinal epithelial cell migration. *Am. J. Physiol. Gastrointest. Liver Physiol.* **290**, G827–838 (2006).
20. Yu, L. Z. et al. Expression of interleukin-22/STAT3 signaling pathway in ulcerative colitis and related carcinogenesis. *World J. Gastroenterol.* **19**, 2638–2649 (2013).
21. Sands, B. E. et al. Efficacy and safety of MEDI2070, an antibody against Interleukin 23, in patients with moderate to severe Crohn's disease: a phase 2a study. *Gastroenterology* **153**, 77–86e76 (2017).
22. Raffals, L. E. et al. The development and initial findings of a study of a prospective adult research cohort with inflammatory bowel disease (SPARC IBD). *Inflamm. Bowel Dis.* **28**, 192–199 (2022).
23. Chen, T., Guestrin, C. & XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. City; Association for Computing Machinery*. 785–794. (2016).
24. Curran Associates Inc., International Conference on Neural Information Processing Systems. A unified approach to interpreting model predictions. https://arxiv.org/abs/1705.07874. (Accessed 27 Sep 2023). (2017).
25. Argmann, C. et al. Biopsy and blood-based molecular biomarker of inflammation in IBD. *Gut* **72**, 1271–1287 (2023).
26. Rajkomar, A., Dean, J. & Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **380**, 1347–1358 (2019).
27. Lamb, C. A. et al. British society of gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. *Gut* **68**, s1–s106 (2019).
28. Aggarwal, S., Ghilardi, N., Xie, M. H., de Sauvage, F. J. & Gurney, A. L. Interleukin-23 promotes a distinct CD4 T cell activation state characterized by the production of interleukin-17. *J. Biol. Chem.* **278**, 1910–1914 (2003).
29. Mosli, M. H. et al. C-reactive protein, fecal calprotectin, and stool lactoferrin for detection of endoscopic activity in symptomatic inflammatory bowel disease patients: A systematic review and meta-analysis. *Am. J. Gastroenterol.* **110**, 802–819 (2015).
30. Turner, D. et al. C-reactive protein (CRP), erythrocyte sedimentation rate (ESR) or both? A systematic evaluation in pediatric ulcerative colitis. *J. Crohns Colitis*. **5**, 423–429 (2011).

Acknowledgements
Writing assistance was provided by Rebecca Lane, PhD, and Kristin Carlin, BSPharm, of Peloton Advantage, an OPEN Health company, and funded by AstraZeneca.

Author contributions
Study design: CC, NP, UG. Data analysis/interpretation: MO, DF, CS, JN, CC. Critical revision and review of the manuscript: All authors. Project/data management: CC, TF, JN, UG. Statistical analyses: MO, DF. Approval of final draft for submission: All authors.

Declarations

Competing interests
EK, JC, JN, MO, and CC: Employees and stock shareholders of AstraZeneca. DFL, EAD, and UG: Former employees of AstraZeneca. CS: Employee of the Kubrick Group, which was contracted by AstraZeneca to perform these analyses. TF: Employee of the Crohn's & Colitis Foundation. NP: Speaker, advisory consultant and/or research grants from AbbVie, Allergan, AstraZeneca, Bristol Myers Squibb, Celgene, Celltrion, Dr Falk Pharma UK Ltd, Ferring, Galapagos, Janssen, Roche, Pfizer, Sobi, Takeda, Tillotts, and Vifor.

Ethics approval
This is a cross-sectional study of patients in the SPARC IBD cohort from the IBD Plexus registry of the Crohn's & Colitis Foundation. This study used patient data entered in SPARC IBD from initiation in 2016 to January 2022. The use of SPARC IBD and blood plasma samples for this study was approved by the Crohn's & Colitis Foundation in 2021.

Consent to participate
All patients enrolled in the study provided informed consent to participate through a universal consent form authorizing use of their personal information for future IBD studies.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-01939-7.

Correspondence and requests for materials should be addressed to C.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.