Journal Pre-proof

DEVELOPMENT AND VALIDATION OF A TEST TO MONITOR ENDOSCOPIC ACTIVITY IN PATIENTS WITH CROHN'S DISEASE BASED ON SERUM LEVELS OF PROTEINS

Geert D'Haens, Orlaith Kelly, Robert Battat, Mark S. Silverberg, David Laharie, Edouard Louis, Edoardo Savarino, Giorgia Bodini, Andres Yarur, Brigid S. Boland, Waqqas Afif, Xiao-jun Li, Michael Hale, Jessica Ho, Venkateswarlu Kondragunta, Benjamin Huang, Crystal Kuy, Lauren Okada, Kelly D. Hester, Kurtis R. Bray, Larry Mimms, Anjali Jain, Siddharth Singh, Angelina Collins, Mark A. Valasek, William J. Sandborn, Severine Vermeire, Parambir S. Dulai

PII: S0016-5085(19)41525-4
DOI: https://doi.org/10.1053/j.gastro.2019.10.034
Reference: YGAST 62990

To appear in: *Gastroenterology*
Accepted Date: 26 October 2019

Please cite this article as: D'Haens G, Kelly O, Battat R, Silverberg MS, Laharie D, Louis E, Savarino E, Bodini G, Yarur A, Boland BS, Afif W, Li X-j, Hale M, Ho J, Kondragunta V, Huang B, Kuy C, Okada L, Hester KD, Bray KR, Mimms L, Jain A, Singh S, Collins A, Valasek MA, Sandborn WJ, Vermeire S, Dulai PS, DEVELOPMENT AND VALIDATION OF A TEST TO MONITOR ENDOSCOPIC ACTIVITY IN PATIENTS WITH CROHN'S DISEASE BASED ON SERUM LEVELS OF PROTEINS, *Gastroenterology* (2019), doi: https://doi.org/10.1053/j.gastro.2019.10.034.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 by the AGA Institute

Journal Pre-proof

DEVELOPMENT AND VALIDATION OF A TEST TO MONITOR ENDOSCOPIC ACTIVITY IN PATIENTS WITH CROHN'S DISEASE BASED ON SERUM LEVELS OF PROTEINS

Short Title: Development and Validation of EHI for Crohn's Disease

Geert D'Haens[1*], Orlaith Kelly[2*], Robert Battat[3*], Mark S. Silverberg[5], David Laharie[6], Edouard Louis[7], Edoardo Savarino[8], Giorgia Bodini[9], Andres Yarur[10], Brigid S. Boland[2], Waqqas Afif[11], Xiao-jun Li[12], Michael Hale[12], Jessica Ho[12], Venkateswarlu Kondragunta[12], Benjamin Huang[12], Crystal Kuy[12], Lauren Okada[12], Kelly D. Hester[12], Kurtis R. Bray[12], Larry Mimms[12], Anjali Jain[12], Siddharth Singh[3], Angelina Collins[3], Mark A. Valasek[4], William J. Sandborn[3], Severine Vermeire[13#] and Parambir S. Dulai [3#]

[1] Department of Gastroenterology, Academic Medical Centre, Amsterdam, Netherlands; [2] Connolly Hospital Blanchardstown, Dublin, Ireland; [3] Division of Gastroenterology, University of California San Diego, La Jolla, California, USA; [4] Department of Pathology, University of California San Diego, La Jolla, California, USA; [5]Mount Sinai Hospital, Toronto, Canada; [6] CHU Bordeaux, Bordeaux, France; [7] University Hospital CHU Liege, Liege, Belgium; [8] University of Padua, Padua, Italy; [9] University of Genoa, Genoa, Italy; [10] Medical College of Wisconsin, Milwaukee, WI, USA; [11] McGill University, Montreal, Canada; [12] Prometheus Laboratories Inc., San Diego, CA, USA; [13] University Hospitals Leuven, Leuven, Belgium.

*Co-First Authors
[#]Equal contribution as senior author

Corresponding Author:
Geert D'Haens MD, PhD
Dept of Gastroenterology
Amsterdam University Medical Centers
Meibergdreef 9
1100 DZ Amsterdam, The Netherlands
Phone +3120 56 63534, Fax +3120 69 17033
g.dhaens@amc.uva.nl

Funding Sources: The UCSD cohorts were recruited with support from a grant to Drs. Sandborn and Dulai through the Litwin Pioneers Program Crohn's and Colitis Foundation, and an unrestricted collaboration grant from Prometheus. The TAILORIX cohort was recruited in a prospective clinical trial by the GETAID, sponsored by Janssen Biologics and Merck Sharp and Dome.

Role of Sponsor(s): Prometheus was responsible for running assays and analyses. Primary clinical data and reference assessments for disease activity for the training and validation cohorts were collected independent of Prometheus by local coordinating sites.

Word Count: Abstract: 478; Manuscript body: 6996; Methods and Materials: 1680. Tables: 4; Figures: 3; Supplementary Tables: 8; Supplementary Figures: 7

Abbreviations: Active disease (AD), Adenosine deaminase (ADA), Activated leukocyte cell adhesion molecule (ALCAM), Analytical Method Validation (AMV), Angiopoietin-1 (ANG1), Angiopoietin-2 (ANG2), Annexin A13 (ANXA13), Area under the receiver operating curve (AUROC), Amphiregulin (AREG), Betacellulin (BTC), Effect of tight control management on Crohn's disease (CALM), Crohn's disease (CD), Crohn's disease endoscopic index of severity (CDEIS), Carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1), Confidence interval (CI), C-reactive protein (CRP), Epidermal growth factor (EGF), Endohistopathologic healing (EHPH), Effect size (ES), Endoscopic healing index (EHI), Extracellular matrix metalloproteinase inducer (EMMPRIN), Endoscopic remission (ER), Epiregulin (EREG), Fecal calprotectin (FC), Food and Drug Administration (FDA), Fibroblast growth factor 2 (FGF2), Global Histologic Disease Activity score (GHAS), Granulocyte-macrophage colony stimulating factor (GM-CSF), Heparin binding EGF like growth factor (HB-EGF), HGF - Hepatocyte growth factor (HGF), Histologic remission (HR), Interferon gamma (IFN-γ), Intercellular adhesion molecule 1 (ICAM1), Inflammatory bowel diseases (IBD), Interleukin-1 beta (IL1-β), Interleukin-2 (IL2), Interleukin-5 (IL5), Interluekin-6 (IL6), Interluekin-7 (IL7), Interleukin-10 (IL10), p40 subunit of Interleukins 12 and 23

(IL12/23 p40), Interleukin-13 (IL13), Interleukin-15 (IL15), Interleukin-17A (IL17A), Interleukin-17F (IL17F), Interleukin-22 (IL22), Interleukin-23 (IL23), Interleukin-31 (IL31), Interleukin-33 (IL33), Interquartile ranges (IQR), Positive likelihood ratio (PLR), positive predictive value (PPV), Mucosal vascular addressin cell adhesion molecule 1 (MADCAM1), Mucosal healing (MH), Mixed logistic regression (MLG), Matrix metalloproteinase-1 (MMP1), Matrix metalloproteinase-2 (MMP2), Matrix metalloproteinase-3 (MMP3), Matrix metalloproteinase-9 (MMP9), Negative likelihood ratio (NLR), negative predictive value (NPV), Patient-reported Outcomes-2 (PRO2), PE – Phycoerythrin, Prospective-specimen collection retrospective-blinded-evaluation (PRoBE), Receiver–operating characteristic (ROC), Serum amyloid A1 (SAA1), Simple endoscopic score for Crohn's disease (SES-CD), Stem cell factor (SCF), Standard deviation (SD), Standards for Reporting of Diagnostic Accuracy Studies (STARD), Stop infliximab in patients with Crohn's disease (STORI), Tailored treatment with infliximab for active Crohn's disease (TAILORIX), Transforming growth factor alpha (TGF-α), Transforming growth factor beta (TGF-β), True positive (TP), True negative (TN), TNF-related weak inducer of apoptosis (TNF receptor superfamily member 12A) (TWEAK), University of California San Diego (UCSD), Vascular cell adhesion molecule 1 (VCAM1), Vascular endothelial growth factor A (VEGF-α).

Conflicts of Interests and Disclosures: GD'H: Consulting and/or lecture fees from Abbvie, Ablynx, Allergan, Alphabiomics, Amakem, Amgen, AM Pharma, Arena Pharmaceuticals, Biogen, Bristol Meiers Squibb, Boehringer Ingelheim, Celgene/Receptos, Celltrion, Echo Pharmaceuticals, Eli Lilly, Engene, Ferring, DrFALK Pharma, Galapagos, Genentech/Roche, Gilead, Glaxo Smith Kline, Gossamerbio, Pfizer, Immunic, Johnson and Johnson, Kintai Therapeutics, Millenium/Takeda, Medtronics, Mitsubishi Pharma, Merck Sharp Dome, Mundipharma, Nextbiotics, Novonordisk, Otsuka, Pfizer/Hospira, Photopill, Prodigest, Prometheus laboratories/Nestle, Progenity, Protagonist, RedHill, Robarts Clinical Trials, Samsung Bioepis, Sandoz, Seres/Nestle, Setpoint, Shire, Takeda, Teva, Tigenix, Tillotts, Topivert, Versant and Vifor. RB: none. MSS: speaker: Abbvie, Janssen, Prometheus, Takeda, Shire, Pfizer/Hospira, Ferring, Novartis, Lilly; advisory board: Abbvie, Allergan, Janssen, Prometheus, Takeda, Shire, Pfizer/Hospira, Ferring; research support: Abbvie, Janssen, Prometheus, Takeda, Pfizer/Hospira; consultant: Abbvie, Janssen, Prometheus, Takeda, Pfizer/Hospira. OK: Tofacinib advisory board Ireland 2018, Abbvie CIHR/ CAG Advanced IBD Fellowship Bursary award, Canada 2014 and 2015. DL: board or lectures fees from Abbvie, Celgene, Ferring, Janssen, MSD, Novartis, Pfizer, Roche and Takeda. EL: Research Grant: Takeda, Pfizer; educational grant: Abbvie, MSD, Takeda; speaker fees: Abbvie, Ferring, MSD, Chiesi, Falk, Takeda, Hospira, Janssen, Pfizer; Advisory Board: Abbvie, Ferring, MSD, Mitsubishi Pharma, Takeda, Celltrion, Celgene, Hospira, Janssen ; Consultant: Abbvie. ES: consulting/lecture fees from Abbvie, MSD, Takeda, Janssen, Sofar, Malesci. GB: invited speaker for Abbvie, MSD, Takeda. AY: consulting fees from Takeda Pharmaceuticals and Prometheus Laboratories; speaker bureau, Abbvie, Takeda Pharmaceuticals and Prometheus Laboratories. BSB: consulting fees from Abbvie, Prometheus Laboratories outside of the submitted work. WA: consulting Abbvie, Janssen, Pfizer, Merck, Takeda, Shire, Allergan; research support Abbvie, Janssen, Theradiag, Prometheus, Ferring. SV: grant support from MSD, Abbvie, Pfizer, J&J and Takeda, lecture fees from Abbvie, MSD, Ferring Pharmaceuticals, Takeda, Hospira, and consultancy fees from Abbvie, Takeda, Pfizer, Ferring Pharmaceuticals, Shire Pharmaceuticals Group, Prometheus, MSD, Hospira, Mundipharma, Celgene, Galapagos and Genentech/Roche. SS: received consulting fees from AbbVie, Pfizer, Takeda and AMAG Pharmaceuticals. AC: consulting fees, honorarium from AbbVie and Janssen. MAV: research support Prometheus. WJS: research grants from Atlantic Healthcare Limited, Amgen, Genentech, Gilead Sciences, Abbvie, Janssen, Takeda, Lilly, Celgene/Receptos; consulting fees from Abbvie, Allergan, Amgen, Boehringer Ingelheim, Celgene, Conatus, Cosmo, Escalier Biosciences, Ferring, Genentech, Gilead, Janssen, Lilly, Miraca Life Sciences, Nivalis Therapeutics, Novartis Nutrition Science Partners, Oppilan Pharma, Otsuka, Paul Hastings, Pfizer, Precision IBD, Progenity, Prometheus Laboratories, Ritter Pharmaceuticals, Robarts Clinical Trials (owned by Health Academic Research Trust or HART), Salix, Shire, Seres Therapeutics, Sigmoid Biotechnologies, Takeda, Tigenix, Tillotts Pharma, UCB Pharma, Vivelix; and stock options from Ritter Pharmaceuticals, Oppilan Pharma, Escalier Biosciences, Precision IBD, Progenity. PSD: research support, consulting fees, honorarium from Takeda; research support from Pfizer, research support and consulting fees Janssen. MH, JH, VK, BH, CK, LO, XL, KDH, KRB, LM, and AJ are all employees of Prometheus Laboratories Inc.

Author Contributions: All authors have approved the final draft submitted.

1. Study concept, design, and supervision: AJ, KRB, GD'H, WJS, MSS, MH, LM, PSD
2. Provision of specimens: GD'H, MSS, DL, SV, WJS, EL, ES, GB, AY, WA
3. Acquisition of clinical data: GD'H, RB, OK, MSS, DL, SV, WJS, EL, ES, GB, AY, BSB, WA, PSD
4. Analytical biomarker development and data acquisition: MH, AJ, JH, BH, CK, KH, LO
5. Statistical analysis: VK, AJ, XL, MH, LO, KRB, PSD
6. Drafting of manuscript: AJ, XL, KRB, GD'H, WJS, KH, PSD
7. Interpretation, critical revisions, and final approval: All authors

ABSTRACT

Background & Aims: Non-invasive tests to measure endoscopic activity in patients with Crohn's disease (CD) have limitations. We aimed to develop a test to identify patients in remission, based on endoscopic analysis, and monitor CD activity based on serum levels of proteins.

Methods: We developed a test to measure 13 proteins in blood (ANG1, ANG2, CRP, SAA1, IL7, EMMPRIN, MMP1, MMP2, MMP3, MMP9, TGFA, CEACAM1, and VCAM1), called the endoscopic healing index [EHI] using samples from 278 patients with CD from multi-national training cohort. We validated the test using 2 independent cohorts of patients with CD: 116 biologic-naïve patients with early-stage CD (validation cohort 1) and 195 biologic-exposed patients with chronic CD (validation cohort 2). The ability of the test to identify patients with active disease vs patients in remission (defined as a simple endoscopic score for CD of ≤ 2 and ≤ 1 in each segment, or a total CD endoscopic index of severity score < 3) was assessed using area under receiver operating characteristic curve (AUROC) analysis. The diagnostic accuracy of the test was compared with that of measurement of serum CRP and fecal calprotectin (FC).

Results: The EHI scores range from 0 to 100 units; higher scores indicate more severe CD activity, based on endoscopy findings. The EHI identified patients in remission with an AUROC of 0.962 in validation cohort 1 (95% CI, 0.942–0.982) and an AUROC of 0.693 in validation cohort 2 (95% CI, 0.619–0.767), regardless of CD location or phenotype. A cut-off value of 20 points identified patients in remission with the highest level of sensitivity (97.1% in validation cohort 1 and 83.2% in validation cohort 2), with specificity values of 69.0% and 36.6%, respectively. A cut-off value of 50 points identified patients in remission with the highest level of specificity (100% in validation cohort 1 and 87.8% in validation cohort 2), with sensitivity values of 37.3% and 30.0%, respectively. The EHI identified patients in remission with a significantly higher AUROC value than the test for CRP (0.876, *P*<.001 in validation cohort 1 and 0.624, *P*=.109 in validation cohort 2). In analysis of patients with available FC measurements, the AUROC value for the EHI did not differ significantly from that of measurement of FC (AUROC, 0.950 for EHI vs AUROC, 0.923 for FC, *P*=.147 in validation cohort 1 and AUROC, 0.803 for EHI vs AUROC, 0.854 for FC, *P*=.298 in validation cohort 2).

Conclusions: We developed an index to identify patients with CD in endoscopic remission based on blood levels of 13 proteins, called the EHI. The EHI identified patients with resolution of endoscopic disease activity, with good overall accuracy, although with variation between the 2 cohorts assessed. The EHI AUROC values were comparable to measurement of FC and higher than measurement of serum CRP. The test might be used in practice for assessing endoscopic activity in patients with CD.

Key Words: Monitr; IBD; response to treatment; resolution

Journal Pre-proof

Journal Pre-proof

INTRODUCTION

Crohn's disease (CD) is a chronic condition characterized by mucosal ulcerations and transmural inflammation anywhere along the gastrointestinal tract. Approximately one-third of patients have stricturing or penetrating disease complications at the time of diagnosis, and half of the remaining patients experience an intestinal complication within 20 years of diagnosis.[1] Achieving endoscopic healing, also traditionally referred to as mucosal healing (MH), has consistently been associated with reductions in disease-related complications, including corticosteroid use, hospitalization, and surgery.[2-5] For this reason, achieving MH is now considered to be a primary treatment target in CD and it is recommended that all CD patients initiating immunosuppressive and/or biologic therapy have a follow-up assessment for MH within 6-9 months of treatment initiation.[6]

The optimal approach for assessing MH has traditionally been through the use of endoscopy. Alternative methods such as cross-sectional imaging, abdominal ultrasound and video capsule imaging are under development but scoring systems with these techniques have been poorly developed or validated. Therefore, many centers worldwide have adopted endoscopy based treat-to-target monitoring algorithms, in which treatment is optimized, modified or switched based on serial endoscopic monitoring. Although this strategy has been associated with achieving higher rates of MH,[7, 8] it is not without cost, risk, or burden and endoscopy is ranked as the least acceptable tool for this purpose by CD patients.[9] These limitations are likely to explain why the majority of CD patients treated with biologic therapy have no follow-up endoscopy within the first 24 months after treatment initiation.[10] This suggests that endoscopy based treat-to-target strategies may be difficult to implement in current health care landscapes.

The effect of tight control management on Crohn's disease (CALM) trial recently demonstrated that treatment escalation based on symptoms combined with elevated serum C-reactive protein (CRP) and/or fecal calprotectin (FC) resulted in higher rates of MH as compared to symptom-based escalation alone.[11] Despite an overall favorable accuracy,[12, 13] the utilization of FC in clinical practice is somewhat impractical and is currently only being done in less than 2% of CD patients in the US.[10, 14] If given an option, CD patients strongly prefer blood based biomarkers over fecal biomarkers. However, all prior blood based biomarkers have had poor accuracy and therefore limited utility.[13] A need therefore exists for blood-based biomarkers that accurately quantify mucosal disease activity in CD.

The objective of the current study was to develop and validate a multi-marker, serologic, algorithm-based diagnostic test that reliably reflects the severity of endoscopic inflammation in CD. Through a multi-center international collaboration we derived the Endoscopic Mucosal Healing Index (EHI, Prometheus Laboratories Inc., San Diego, CA) and subsequently validated it in two independent cohorts, demonstrating that it is associated with endoscopic inflammation. We further explored the comparative performance of EHI against CRP and FC, the responsiveness of EHI to changes in endoscopic disease state, and, the diagnostic performance of EHI for histologic inflammation on a selected sub-cohort with limited sample size.

Journal Pre-proof

METHODS

We followed the PRoBE (prospective-specimen collection retrospective-blinded-evaluation) study design for evaluating the accuracy of a biomarker used for classification of an outcome (i.e. mucosal healing).[15] The results are reported in accordance with the Standards for Reporting of Diagnostic Accuracy Studies (STARD) guidelines.[16] (**Supplementary Form 1**)

Patient Selection

Adult CD patients (≥18 years) were included if they had: (A) a confirmed diagnosis of CD based on clinical, endoscopic, and histologic data; (B) documented endoscopic disease activity; and (C) sufficient volume of serum sample available for testing. For the validation cohorts we required samples to be available within +/- 45 days of endoscopy. Patients selected for this study were not excluded based on current or prior therapies, prior bowel surgeries, or the presence of an ostomy (ileostomy or colostomy).

Cohorts

The study consisted of 3 independent cohorts of prospectively collected, retrospectively analyzed samples for training and validation. The training cohort included samples obtained from prospectively recruited convenience sampling biobanks between June 2006 - August 2015 at University of Padua (U Padua), Italy (Jul 2011-Mar 2014); Mount Sinai Hospital, Toronto, Canada (Oct 2008 – Aug 2015); University of California San Diego (UCSD, Jun 2014 – May 2015), USA; and the STORI clinical trial (GETAID, France, Jun 2006 – Jan 2007).[17] Validation cohort 1 included samples collected during the prospective TAILORIX clinical trial (July 2012 – September 2015). These included baseline, week 12, and week 54 samples from 116 biologic naïve CD patients recruited from 27 centers in Belgium, France and the Netherlands.[18] Validation cohort 2 included samples collected prospectively from a tertiary referral center in San Diego, USA (UCSD, June 2014 – January 2018), which were distinct from those included from this institution in the training cohort.

Clinical Data Variables

Data on available variables of interest included patient characteristics (age, gender; ethnicity), disease characteristics (prior surgeries, disease-related complications, Montreal phenotype classification), current and prior treatments (corticosteroids, immunosuppressives, biologics), and clinical disease activity (patient reported outcomes (PRO2) or Crohn's Disease Activity Index (CDAI)).

Endoscopic Healing Definitions

Endoscopic remission (ER) was defined as a total simple endoscopic score for CD (SES-CD) of ≤ 2 and ≤ 1 in each segment (in the two validation cohorts) or a total Crohn's Disease Endoscopic Index of Severity (CDEIS) <3 (in the training cohort).[19] Consequently, active disease (AD) was defined as CDEIS≥3 or SES-CD>2 or SES-CD=2 if only one segment had a score of 2 with a score of 0 in the remaining segments. Endohistopathologic healing (EHPH) was defined as achieving both ER and histologic remission (Global Histologic Disease Activity (GHAS) ≤ 2).[19-21] Endoscopic scores were derived using either the SES-CD or the CDEIS.[20, 21] Scoring was done locally by site investigators at the time of endoscopy in all datasets except TAILORIX where endoscopies were scored by blinded central readers. In the derivation-training cohort all SES-CD scores were converted to CDEIS scores for consistency during training.(**Supplementary Figure 1**)

Journal Pre-proof

Histologic disease activity assessments were available in validation cohort 2 and were done by a pathologist with expertise in gastrointestinal pathology and IBD (M.V.) who was blinded to endoscopy scores. Four biopsies were taken from each intestinal segment (using segments identical to those used to calculate SES-CD score) with matching endoscopic scores. Biopsies were taken from the most active endoscopic area, and if no active inflammation was observed then random biopsies were taken.

Sample Storage and Testing

All serum samples and fecal samples were frozen within 24 hours of collection to avoid degradation or loss in biomarkers,[22] and kept frozen at -80 degrees Celsius until testing. Thawed serum and stool samples were tested for all biomarkers in a randomized manner with clinical data blinded to the operator. Further details can be found in the Supplementary Methods.

EHI Development

Preliminary serum biomarker candidates were identified from literature review and assessed by the strength of the corresponding evidence, the relevance of their biological functions to CD, and the involvement of their signal pathways in CD pathogenesis. (**Supplementary Table 1**) Assays were then developed for the selected biomarker candidates and evaluated for their analytical performance. Biomarker candidates whose assays showed poor analytical reproducibility, low detection rate in serum specimens, and/or lack of correlation to disease severity in preliminary studies (data not shown) were eliminated from further consideration as training progressed to validation. As candidate biomarkers were eliminated from consideration, new panels were developed with progressively fewer target biomarkers such that training panels targeted 47- and 38-biomarkers, and subsequent validation was performed on a 24-analyte panel. Regardless of panel configuration the EHI algorithm was trained on the same 13 analytes whose assays were robust and which demonstrated correlation with clinical disease. These combined training and validation cohorts provided the foundation for validation of a 13-biomarker panel currently reporting EHI to physicians and patients (Monitr™, Prometheus Laboratories Inc., San Diego, California.). An analytical method validation (AMV) was also performed on the final 13-biomarker panel. Multiple logistic regression method was used to predict endoscopic activity as a function of serum biomarker concentrations proposed as continuous predictors after logarithmic transformation and combined through backward elimination with Akaike information criterion (AIC). Biomarkers were removed one by one by sequentially reducing the AIC value until a minimum of AIC was reached, using standard settings in JMP (version 12.0; SAS Institute, Cary, NC). 'EHI' was obtained by transforming the logistic function in terms of probability to be in active disease. A small fraction of patients (11.5%) contributed more than one sample in the training cohort. Samples from same patients were treated as independent samples, an imperfection limited to the training cohort.

Endpoints

Our primary aim was to assess the sensitivity (proportion of patients above a specific EHI limit among patients with active disease (SES-CD>2 or SES-CD=2 if only one segment has a score of 2 with a score of 0 in the remaining segments)) and specificity (proportion of patients below a specific EHI limit among patients in endoscopic remission (SES-CD) of ≤ 2 and ≤ 1 in each segment)) of the EHI at various cut-offs for identifying

Journal Pre-proof

the presence of endoscopic inflammation. Secondary aims were to explore the diagnostic accuracy of the EHI at various cut-offs for identifying the presence of endohistopathologic inflammation, and to compare the diagnostic accuracy of EHI to CRP and FC. Positive likelihood ratio (PLR), negative likelihood ratio (NLR), positive predictive value (PPV), negative predictive value (NPV) and area under the receiver operating characteristic (ROC) curve (AUROC) were used as secondary measures to assess the performance of EHI. Finally, we assessed the responsiveness of EHI as compared to endoscopy, CRP, and FC, to assess its utility as a tool for monitoring endoscopic disease activity in patients with CD.

Statistical Analysis

The Delong method was used for computing the 95% confidence interval (CI) of AUROC and for comparing AUROCs of different biomarkers on paired samples.[23] Exact binomial confidence limits were used for the 95% CIs of sensitivity and specificity. The 95% CIs of PLR and NLR were computed using formulae provided by Simel et al.[24] Pairwise Wilcoxon rank sum test was used for comparing effect size of different variables. A p value (two-sided) of 0.05 or lower was considered as significant. All data analysis was carried out using JMP (version 12.0; SAS Institute, Cary, NC) or R (version 3.3.2). A mixed-effect logistic regression modeling was utilized for Validation cohort 1 to assess the performance of EHI, CRP and FC. effect sizes[25] of SES-CD, CDEIS, EHI, FC and CRP were calculated between baseline and week 12 and between baseline and week 54 to assess the responsiveness of those variables. Further statistical details can be found in the Supplementary Methods.

Sample Size Calculation

No sample size calculation was performed on the training cohort. EHI had a sensitivity of 90% at the threshold of 20 and a specificity of 95% at the threshold of 50 in the training cohort. We aimed to validate such performance in the validation cohorts with precisions such that the corresponding one-sided lower 95% confidence limits of sensitivity and specificity were ≥80% and ≥85%, respectively. Based on exact binomial confidence limits, a minimal of 57 samples from CD patients with active disease (AD) and a minimal of 40 samples from CD patients in ER were needed for the validation study.

Ethics

All authors had access to study results, reviewed and approved the final manuscript. Informed consent was obtained from all patients by the study sites according to institutional review board–approved clinical protocols and local regulatory guidance.

RESULTS

Patient Demographics

Serum samples from a total of 589 patients were used (**Supplementary Table 2**) distributed across the training and two distinct validation cohorts without any sample overlap among the training and validation cohorts. The flow chart describing the patients and samples used in the two validation cohorts is shown in **Supplementary Figure 2**. All validation samples used in this study were obtained ± 45 days of endoscopy with ~66% (311/470) of samples collected on the same day as the endoscopy: Validation cohort 1: 123/275 (44.7%), Validation cohort 2: 188/195 (96.4%). 90.1% of the training samples were also obtained ± 45 days of

Journal Pre-proof

endoscopy with 43.9% (147/335) samples collected on the same day as the endoscopy. Median elapse from endoscopy to sample collection was 0 (day, IQR: 0-14.5) in the training cohort.

The training cohort included 278 patients with a total of 335 endoscopy visits. (**Supplementary Table 3**) Median age was 30.0 years (IQR: 24.9-40.0) with 46.0% females. Median disease duration was 4 years (IQR: 3.0-12.5). Validation cohort 1 included 116 patients with a median age of 30.2 years (IQR: 22.4-45.2) and 59.5% females. The cohort included a total of 275 endoscopy visits distributed at baseline (102 visits), weeks 12 (98 visits) and 54 (75 visits). FC was available at the same time points. (**Supplementary Table 4**) Validation cohort 2 included 195 patients with one endoscopy visit per patient. Median age of the cohort was 38.5 years (IQR: 28.0-52.0) with 49.7% females. A sub-cohort of validation cohort 2 (N=81) patients also had paired fecal calprotectin values obtained from stool samples collected within 45 days of endoscopy. (**Supplementary Table 5**)

The two validation cohorts differed significantly in baseline age (30.2 vs. 38.5 years, p<0.001), disease duration (0.7 vs. 11.0 years, p<0.001), disease phenotype (Non-stricturing, non-penetrating 73% vs. 58%, p=0.006), prior IBD related bowel surgery (10% vs. 46%, p<0.001) and prior biologic exposure (0% vs. 77%, p<0.001). The median SES-CD (6.0 vs. 3.0, p<0.001) and FC (336 vs. 55, p<0.001) were significantly higher in validation cohort 1, with comparable CRP values (2.5 vs. 2.6, p=0.460). All disease locations were represented in the training and both validation cohorts.

Training of the Mucosal Healing Index (EHI)

The 47 markers evaluated for the development of EHI are listed in **Supplementary Table 1**. Initially, 23 out of 47 biomarkers were eliminated due to poor assay lot-to-lot reproducibility, poor analytical sensitivity, absence of detectable concentrations in serum, or lack of correlation with endoscopic disease severity. Eleven markers were further eliminated during logistic regression analyses as they did not enhance the performance of EHI, and the final EHI model included serum concentrations of 13 biomarkers: angiopoietin 1 (ANG1) and 2 (ANG2), carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1), C-reactive protein (CRP), serum amyloid A1 (SAA1), Interleukin-7 (IL7), transforming growth factor alpha (TGFα), vascular cell adhesion molecule 1 (VCAM1), extracellular matrix metalloproteinase inducer (EMMPRIN), and matrix metalloproteinase-1 (MMP1), -2 (MMP2), -3 (MMP3), and -9 (MMP9). EHI was constructed as a scale of 0 – 100 arbitrary units of EHI activity, a higher score indicating more severe disease activity. Analytical reproducibility of EHI was established using Deming regression (**Supplementary Figure 3**) with a slope of 1.005 (95% CI: 0.926 – 1.087) and an intercept of -0.298 (95% CI: -2.790 – 2.144).

AUROC of EHI for distinguishing AD from ER in the training cohort was 0.748 (95% CI: 0.696 – 0.800) (**data not shown**). Sensitivity and specificity of EHI was evaluated at increasing cut-offs from 20-50 (**Supplementary Table 6**), covering the clinically relevant region (from high sensitivity to high specificity) as observed in the training cohort. EHI demonstrated a sensitivity and specificity of 90.3% (95% CI: 85.0 – 94.3) and 95.0% (90.3 – 97.8) at cut-offs 20 and 50, respectively, in the training cohort.

Validation Cohort 1

Journal Pre-proof

The median global endoscopic SES-CD score with corresponding median EHI values were 15.0 (IQR 9.0-22.0) and 55.5 (IQR 42.0-72.8) at baseline, 4.0 (IQR: 2.0-8.0) and 33.5 (IQR 23.0-41.8) at week 12, 1.0 (IQR: 0.0-3.0) and 25.0 (IQR 19.0-37.5) at week 52, respectively. The prevalence of ER among patients with paired endoscopy and serum samples in validation cohort 1 was 0% at baseline, 26.5% at week 12, and 60% at week 52. AUROC for distinguishing AD from ER was 0.962 (95% CI: 0.942 – 0.982) (**Figure 1A**). Sensitivity for ruling out endoscopic inflammation at an EHI cut-off ≤20 was 97.1% (95% CI: 93.7 – 98.9). (**Table 1**) At EHI cut-offs 40 and 50, the specificity for ruling in AD was 100% (95% CI: 94.9 – 100.0). At a cut-off of 30, sensitivity and specificity for endoscopic inflammation was 84.8% (95% CI: 79.1-89.4) and 91.5% (95% CI: 82.5-96.8), respectively. The positive likelihood ratios (PLR) for detecting AD demonstrated a steady increase from 3.132 (95% CI: 2.212-4.436) to infinity with increasing EHI cut-offs while the negative likelihood ratios (NLR) were no greater than 0.627 (95%CI: 0.564-0.697). Prevalence of AD at EHI ≥50 was 100% and the prevalence of ER at EHI<20 was 89.1% (**Supplementary Figure 2**).

Validation Cohort 2

The median global SES-CD score was 3.0 (IQR 0.0-6.5) and the median EHI value was 32 (IQR 19.5-46.5). Prevalence of ER was 42.1% (82/195). AUROC of EHI for distinguishing AD from ER was 0.693 (0.619 – 0.767) (**Figure 1C**). Sensitivity was the highest at an EHI cut-off 20 at 83.2% (95% CI: 75.0 – 89.6). (**Table 1**) The specificity of the test progressively increased with increasing EHI cut-offs, and at an EHI cut-off of 50 the specificity was observed to be 87.8% (95% CI: 78.7 – 94.0). Prevalence of AD at EHI ≥50 was 77.3% and the prevalence of ER at EHI<20 was 61.2% (**Supplementary Figure 2**). PPV and NPV in Validation cohorts 1 and 2 at EHI cut-offs 20 and 50 were calculated at assumed AD prevalence ranging from 5-75% (**Table 2**). Presence or absence of prior IBD-related surgeries did not impact the performance of EHI (AUROC in patients with prior surgery (**Supplementary Figure 4**): 0.699 (95% CI: 0.588 – 0.811); AUROC in patients without prior surgery: 0.680 (95% CI: 0.578 – 0.782), p=0.801). Paired histology data was available for a subset of Validation cohort 2 (N = 79) patients. The AUROC estimate of EHI for distinguishing endohistopathologic healing (defined as endoscopic remission + histologic remission) from active endoscopic or histologic disease was 0.666 (95% CI: 0.536 – 0.797). (**Supplementary Figure 5**)

Diagnostic Performance of EHI by Disease Location and Phenotype

The AUROC of EHI for distinguishing AD vs ER was not significantly different across disease locations in both validation cohorts (**Supplementary Figure 6A and 6B**; pairwise P≥0.171 and P≥0.292, respectively). EHI performance was also comparable across disease behaviors B1, B2, B3 (**Supplementary Figure 7**).

Sensitivity and specificity in each location were evaluated at cut-offs that had a high performance in both validation cohort 1 (**Supplementary Table 7**) and validation cohort 2 (**Supplementary Table 8**). In validation cohort 1, an EHI cut-off of 20 demonstrated a high sensitivity when the cohort was limited to L1 disease (98.1%; 95% CI: 89.7 – 100.0%), L2 disease (100%, 95% CI: 88.4 – 100.0%) or L3 disease (95.7%; 95% CI: 90.1 – 98.6%). The specificity at EHI cut-offs 40 and 50 was 100% regardless of the disease location. Similarly, in validation cohort 2, sensitivity and specificity at cut-offs 20 and 50, respectively, by disease

Journal Pre-proof

location was as follows: L1 (sensitivity = 84.6%, specificity = 100.0%), L2 (sensitivity = 78.9%, specificity = 79.3%) and L3 (sensitivity = 85.1%, specificity = 86.2%).

Comparison of EHI to CRP

AUROC of EHI to distinguish active endoscopic disease from ER was significantly higher than that of CRP alone in Validation cohort 1 (EHI = 0.962; 95% CI: 0.942 – 0.982, CRP = 0.876; 95% CI: 0.835 – 0.916, p<0.001) (**Figures 1A and 1B**). In Validation cohort 2, AUROC of EHI was numerically better than CRP but did not reach significance (EHI = 0.693; 95% CI: 0.619 – 0.767, CRP = 0.624; 95% CI: 0.544 – 0.704, p=0.109) (**Figures 1C and 1D**). Diagnostic performance of EHI was also significantly better than the corresponding AUROC of CRP in the training cohort (EHI vs CRP: 0.748 vs. 0.604; p<0.001, data not shown). CRP cut-off of 5 mg/L had a sensitivity of 41.7 – 44.3% in both validation cohorts 1 and 2 (**Tables 3 and 4**). At a cut-off of 20, EHI had a sensitivity of 91.7 – 96.2%. PPV and NPV in Validation cohorts 1 and 2 at CRP cut-offs 3 and 5mg/L were calculated at assumed AD prevalence ranging from 5-75% (**Table 4**) and compared to that of EHI.

Comparison of EHI to FC

A total of 247 FC assessments were available in Validation cohort 1 (**Figures 2A and 2B**), and 81 paired stool samples were available in Validation cohort 2 (**Figures 2C and 2D**), for comparison between EHI and FC. The sub-cohorts with and without FC available from validation cohort 2 were comparable for all baseline characteristics (**Supplementary Table 5**) although a hidden bias cannot be ruled out. The diagnostic accuracy of EHI was not significantly different from that of FC in either of the two validation cohorts; in Validation cohort 1 it was numerically superior to FC (EHI vs FC: AUROC 0.950 vs 0.923, p=0.147) but numerically inferior in Validation cohort 2 (EHI vs FC: AUROC 0.803 vs 0.854, p=0.298). A FC cut-off of 50 µg/g had 100% sensitivity in validation cohort 1, and 75% sensitivity in validation cohort 2. Corresponding sensitivity for EHI at a cut-off of 20 was 96% in validation cohort 1 and 92% in validation cohort 2. A FC cut-off of 250 µg/g had 89% specificity in validation cohort 1, and 100% specificity in validation cohort 2. Corresponding specificity for EHI at a cut-off of 50 was 100% in validation cohort 1 and 91% in validation cohort 2. (**Tables 3 and 4**) PPV and NPV in Validation cohorts 1 and 2 at FC cut-offs 50 and 250 µg/g were calculated at assumed AD prevalence ranging from 5-75% (**Table 2**) and compared to that of EHI.

Responsiveness of EHI

Endoscopy paired, longitudinal serum samples were available from 97 patients in Validation cohort 1. Effect sizes (ES) were calculated between baseline and week 12 (n = 70 patients; **Figure 3A**) and between baseline and week 54 (n = 59 patients; **Figure 3B**) for the 2 endoscopic indices (SES-CD and CDEIS) and the 3 biomarkers (EHI, CRP, FC). Between baseline and week 12, median ES of EHI (1.10, IQR 0.52 – 1.83) was numerically better than that of FC (0.96, IQR 0.43 – 1.96, p=0.423) and significantly better than that of CRP (0.26, IQR 0.11 – 0.51, p<0.001). Similar results were noted between baseline and week 54 where median ES of EHI (1.64, IQR 0.65 – 2.29) was numerically better than that of FC (1.16, IQR 0.51 – 2.32, p=0.574) and significantly better than that of CRP (0.21, IQR 0.09 – 0.56, p<0.001). Between both time intervals, median ES of EHI was on par with those of the endoscopic scores and mirrored changes in SES-CD and CDEIS (SES-CD between weeks 0-12: 1.53, IQR 0.67 – 2.23, p=0.077; SES-CD between weeks 0-54: 1.87, IQR 0.93 – 2.67,

Journal Pre-proof

p=0.069; CDEIS between weeks 0-12: 1.29, IQR 0.80 – 2.25, p=0.182; CDEIS between weeks 0-54: 1.50, IQR 0.81 – 2.17, p=0.997).

DISCUSSION

Symptoms are often not representative of disease activity in CD, and endoscopy represents the current gold standard for objective disease assessment. Despite this, the majority of CD patients starting biologic therapy have no follow-up endoscopy within 24 months of treatment initiation.[10] The exact reason for this gap is unknown, but cost, risk, and burden are likely to be drivers of these practice patterns. FC is a stool-based biomarker that is widely available and approved by regulatory agencies to aid in the diagnosis of IBD. Its routine availability lends itself to use as a monitoring tool, particularly considering the recent emerging evidence supporting biomarker based adjustments in therapy to optimize the achievement of MH.[11] Although its use in Europe is well established and home FC monitoring kits are available,[26] in the US less than 2% of patients with established IBD undergo FC testing.[10] A significant gap remains in monitoring CD patients for mucosal inflammation to guide treatment decisions.

In the current study we have developed and validated a novel 13-biomarker panel serum-based assay (EHI, Prometheus Laboratories Inc. San Diego, CA) that detects mucosal inflammation in CD. All the selected 13 markers have documented roles in CD pathophysiology. SAA, an acute phase reactant like CRP, correlates with lack of mucosal healing and serves as a surrogate marker of disease activity even in those patients where CRP is not upregulated.[27] Compromised barrier function observed in CD may be due to altered barrier permeability due to MMP family members, recruitment of pro-inflammatory cytokines and altered angiogenesis. MMP3 is downregulated in CD patients with fibrostenotic phenotype[28] and MMP9 has been implicated in the pathogenesis of CD[29, 30] with anti-MMP9 antibodies considered for targeting active CD.[31] EMMPRIN, an inducer of MMPs, plays a role in wound healing, nutrient transport inflammation and has been suggested as a protein interaction partner of NOD2.[32, 33] TGFα is a growth factor and EGFR pathway ligand. EGFR pathway ligands have been implicated in immunity, inflammation and tissue repair.[34] The angiopoietin system including Ang1 and Ang2 have been proposed as factors to maintain pathological angiogenesis during the development of IBD.[35, 36] IL7 is a critical survival factor for lymphocytes and its availability determines the size and proliferative state of resting T cell pool.[37] Adhesion molecules ICAM-1, VCAM-1 and CEACAM family members are known to be upregulated in IBD. Zundler et al.[38] have suggested that the a4b1-VCAM1 axis is involved in mechanisms controlling the homing of T effector cells to the inflamed gut in Crohn's disease.

EHI assay was subsequently validated in 2 independent cohorts representing both early disease and biologic naive CD patients and longer duration CD patients with prior bowel surgeries, disease-related complications, and multiple biologic exposures. Across both cohorts EHI was observed to have an overall favorable diagnostic accuracy for identifying endoscopic inflammation, and in the second validation cohort an early exploratory analysis observed it to have a reasonable diagnostic accuracy for identifying histologic inflammation. Based on these data a cut-off of 20 was observed to have a high sensitivity for ruling out endoscopic inflammation and a cut-off of 50 was observed to have a high specificity for ruling in endoscopic inflammation. The sensitivity of CRP was consistently poor across both validation cohorts indicating that CRP is an unreliable serum marker

for ruling out endoscopic inflammation. Most notably, the diagnostic accuracy of EHI was observed to be consistent across disease locations and disease phenotypes, and its performance was comparable to that of FC and superior to CRP.

CD patients strongly prefer blood based testing over fecal testing,[13] but to date there have been no routinely available blood based tests with a diagnostic performance comparable to that of FC. A recent study examined the diagnostic accuracy of serum calprotectin for differentiating IBD from healthy controls, and although serum calprotectin had a favorable diagnostic accuracy for identifying IBD (AUC 0.87, 95% CI 0.78-0.97), it was still less accurate than FC (AUC 0.99, 95% CI 0.87-1.00, p=0.01).[39] Our study is therefore novel and substantially additive to the current tests available for monitoring CD patients in routine practice.

When comparing EHI to FC there are several important observations from our study worth noting. First, the sensitivity and specificity of EHI at cut offs of 20 and 50 respectively remained consistent across both validation cohorts (92-96% and 91-100%). Although the specificity of FC at a cut-off of 250 µg/g remained stable between validation cohorts (89-100%), the sensitivity of FC at a cut-off of 50 µg/g was quite different between validation cohort 1 (an early disease biologic naive population; 100%) and validation cohort 2 (routine practice, longer duration, biologic exposed population; 75%). Second, the diagnostic accuracy of EHI was consistent across disease locations and phenotypes. Prior literature has demonstrated that the diagnostic performance of FC varies by disease location, and even in the presence of very large ulcers ileal CD patients may not have markedly elevated FC levels.[40] Furthermore, the presence of perianal fistulas even in the absence of colonic inflammation leads to elevated FC.[41] Third, one of the major limitations of FC is the variability across platforms, collection techniques, and timing of sample collection, which have downstream implications on the diagnostic performance of FC.[22, 26, 42-47] EHI was built to ensure reproducibility and consistency in performance, which was observed throughout the validation process. Although no power analysis was performed for the comparison of EHI and FC, we note that the sample size for this comparison was somewhat limited with 247 samples in Validation cohort 1 and 81 samples in Validation cohort 2.

EHI performance was consistent as compared to other serum markers across various endoscopic active disease prevalence in patients with established CD that ranged from 5-75%. At a threshold of 20, EHI had a high NPV (84.9-99.7%) in validation cohort 1 across disease prevalence of 5-75% and an NPV of 63-99% in validation cohort 2. In contrast NPV of CRP decreased with increasing disease prevalence and was as low as 29% at a cut-off of 5mg/l in validation cohort 2 indicating that CRP is a poor marker to rule out active disease. Performance of FC was better than CRP with a better NPV at a cut-off of 50µg/g which was comparable to or lower than that of EHI 20.

Our study has several strengths, including the multi-center multi-national collaboration with varying patient populations and disease characteristics, availability of both endoscopic and histologic disease activity assessments, comparative accuracy assessments against both FC and CRP, and longitudinal comparisons for responsiveness in a prospective clinical trial. Some limitations, however, remain. First, the observed diagnostic accuracy of EHI for identifying histologic inflammation was exploratory and only available in a sub-set of patients from validation cohort 2. Given the test was not trained on a cohort with histologic assessments, the

current test is associated with endoscopic mucosal healing and further work is needed to understand how EHI performs for identifying the evolving definition of MH which encompasses both endoscopic and histologic activity. Second, we did not have routine cross sectional imaging assessments and further analyses are required to understand how EHI compares against cross-sectional imaging based assessments of disease activity, particularly for isolated small bowel CD. Third, although we observed a similar performance of EHI in patients with and without prior IBD surgeries, we were unable to assess the prognostic value of EHI for predicting future endoscopic recurrence or disease relapse, particularly in post-operative CD patients where FC has a clearly established role.[48] Fourth, reasons for differences in performance of EHI in various CD populations need to be further explored. Performance of EHI in Validation cohort 2 was lower than the corresponding performance in Validation cohort 1 but so was of FC and CRP too, likely reflecting the fact that subjects in the cohort were mainly chronic CD patients with altered cellular signaling pathways. Finally, although the second validation cohort encompassed multiple biologic exposures, as treatment paradigms shift to include additional pathway targeted therapies, continued validation of the EHI will be needed to ensure generalizability across all populations. In addition, future studies should assess the cost effectiveness of EHI, relative to both colonoscopy, and to other available biomarker tests such as CRP and FC. Such efforts will need to take into account the accuracy of the various biomarker tests in patient populations with different prevalence of endoscopic disease activity.

In conclusion, we have developed and validated a serum-based assay with a favorable diagnostic accuracy for identifying mucosal inflammation, which is comparable to FC. The serum-based assay was observed to be responsive to changes in endoscopic disease activity and accuracy was consistent across sub-groups. This test could help to bridge current gaps in monitoring patients with CD.

Journal Pre-proof

Journal Pre-proof

References

1. Peyrin-Biroulet L, Loftus EV, Jr., Colombel JF, et al. The natural history of adult Crohn's disease in population-based cohorts. Am J Gastroenterol 2010;105:289-97.
2. Baert F, Moortgat L, Van Assche G, et al. Mucosal healing predicts sustained clinical remission in patients with early-stage Crohn's disease. Gastroenterology 2010;138:463-8; quiz e10-1.
3. Froslie KF, Jahnsen J, Moum BA, et al. Mucosal healing in inflammatory bowel disease: results from a Norwegian population-based cohort. Gastroenterology 2007;133:412-22.
4. af Bjorkesten CG, Nieminen U, Sipponen T, et al. Mucosal healing at 3 months predicts long-term endoscopic remission in anti-TNF-treated luminal Crohn's disease. Scand J Gastroenterol 2013;48:543-51.
5. Shah SC, Colombel JF, Sands BE, et al. Systematic review with meta-analysis: mucosal healing is associated with improved long-term outcomes in Crohn's disease. Aliment Pharmacol Ther 2016;43:317-33.
6. Peyrin-Biroulet L, Sandborn W, Sands BE, et al. Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE): Determining Therapeutic Goals for Treat-to-Target. Am J Gastroenterol 2015;110:1324-38.
7. Bouguen G, Levesque BG, Feagan BG, et al. Treat to target: a proposed new paradigm for the management of Crohn's disease. Clin Gastroenterol Hepatol 2015;13:1042-50 e2.
8. Bouguen G, Levesque BG, Pola S, et al. Endoscopic assessment and treating to target increase the likelihood of mucosal healing in patients with Crohn's disease. Clin Gastroenterol Hepatol 2014;12:978-85.
9. Buisson A, Gonzalez F, Poullenot F, et al. Comparative Acceptability and Perceived Clinical Utility of Monitoring Tools: A Nationwide Survey of Patients with Inflammatory Bowel Disease. Inflamm Bowel Dis 2017;23:1425-1433.
10. Limketkai BN, Singh S, Jairath V, et al. US Practice Patterns and Impact of Monitoring for Mucosal Inflammation after Biologic Initiation in Inflammatory Bowel Disease IBD 2019.
11. Colombel JF, Panaccione R, Bossuyt P, et al. Effect of tight control management on Crohn's disease (CALM): a multicentre, randomised, controlled phase 3 trial. Lancet 2018;390:2779-2789.
12. Lin JF, Chen JM, Zuo JH, et al. Meta-analysis: Fecal Calprotectin for Assessment of Inflammatory Bowel Disease Activity. Inflamm Bowel Dis 2014;20:1407-15.
13. Mosli MH, Zou G, Garg SK, 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 2015;110:802-19; quiz 820.
14. van Deen WK, van Oijen MG, Myers KD, et al. A nationwide 2010-2012 analysis of U.S. health care utilization in inflammatory bowel diseases. Inflamm Bowel Dis 2014;20:1747-53.
15. Pepe MS, Feng Z, Janes H, et al. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. J Natl Cancer Inst 2008;100:1432-8.
16. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015;351:h5527.
17. Louis E, Mary JY, Vernier-Massouille G, et al. Maintenance of remission among patients with Crohn's disease on antimetabolite therapy after infliximab therapy is stopped. Gastroenterology 2012;142:63-70 e5; quiz e31.
18. D'Haens G, Vermeire S, Lambrecht G, et al. Increasing Infliximab Dose Based on Symptoms, Biomarkers, and Serum Drug Concentrations Does Not Increase Clinical, Endoscopic, and Corticosteroid-Free Remission in Patients With Active Luminal Crohn's Disease. Gastroenterology 2018;154:1343-1351 e1.
19. Dulai PS, Levesque BG, Feagan BG, et al. Assessment of mucosal healing in inflammatory bowel disease: review. Gastrointest Endosc 2015;82:246-55.

20. Daperno M, D'Haens G, Van Assche G, et al. Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD. Gastrointest Endosc 2004;60:505-12.
21. Mary JY, Modigliani R. Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective multicentre study. Groupe d'Etudes Therapeutiques des Affections Inflammatoires du Tube Digestif (GETAID). Gut 1989;30:983-9.
22. Padoan A, D'Inca R, Scapellato ML, et al. Improving IBD diagnosis and monitoring by understanding preanalytical, analytical and biological fecal calprotectin variability. Clin Chem Lab Med 2018.
23. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-45.
24. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol 1991;44:763-70.
25. Cohen J. Statistical Power Analysis for the Behavioral Sciences: Lawrence Erlbaum Associates, 1988.
26. Heida A, Knol M, Kobold AM, et al. Agreement Between Home-Based Measurement of Stool Calprotectin and ELISA Results for Monitoring Inflammatory Bowel Disease Activity. Clin Gastroenterol Hepatol 2017;15:1742-1749.e2.
27. Yarur AJ, Quintero MA, Jain A, et al. Serum Amyloid A as a Surrogate Marker for Mucosal and Histologic Inflammation in Patients with Crohn's Disease. Inflamm Bowel Dis 2017;23:158-164.
28. de Bruyn JR, van den Brink GR, Steenkamer J, et al. Fibrostenotic Phenotype of Myofibroblasts in Crohn's Disease is Dependent on Tissue Stiffness and Reversed by LOX Inhibition. J Crohns Colitis 2018;12:849-859.
29. Baugh MD, Perry MJ, Hollander AP, et al. Matrix metalloproteinase levels are elevated in inflammatory bowel disease. Gastroenterology 1999;117:814-22.
30. Faubion WA, Jr., Fletcher JG, O'Byrne S, et al. EMerging BiomARKers in Inflammatory Bowel Disease (EMBARK) study identifies fecal calprotectin, serum MMP9, and serum IL-22 as a novel combination of biomarkers for Crohn's disease activity: role of cross-sectional imaging. Am J Gastroenterol 2013;108:1891-900.
31. Schreiber S, Siegel CA, Friedenberg KA, et al. A phase 2, Randomized, Placebo-controlled Study Evaluating Matrix Metalloproteinase-9 Inhibitor, Andecaliximab, in Patients with Moderately to Severely Active Crohn's Disease. J Crohns Colitis 2018.
32. Kumar D, Vetrivel U, Parameswaran S, et al. Structural insights on druggable hotspots in CD147: A bull's eye view. Life Sci 2019;224:76-87.
33. Till A, Rosenstiel P, Brautigam K, et al. A role for membrane-bound CD147 in NOD2-mediated recognition of bacterial cytoinvasion. J Cell Sci 2008;121:487-95.
34. Zaiss DMW, Gause WC, Osborne LC, et al. Emerging functions of amphiregulin in orchestrating immunity, inflammation, and tissue repair. Immunity 2015;42:216-226.
35. Algaba A, Linares PM, Encarnacion Fernandez-Contreras M, et al. The effects of infliximab or adalimumab on vascular endothelial growth factor and angiopoietin 1 angiogenic factor levels in inflammatory bowel disease: serial observations in 37 patients. Inflamm Bowel Dis 2014;20:695-702.
36. Linares PM, Chaparro M, Gisbert JP. Angiopoietins in inflammation and their implication in the development of inflammatory bowel disease. A review. J Crohns Colitis 2014;8:183-90.
37. Martin CE, Spasova DS, Frimpong-Boateng K, et al. Interleukin-7 Availability Is Maintained by a Hematopoietic Cytokine Sink Comprising Innate Lymphoid Cells and T Cells. Immunity 2017;47:171-182 e4.
38. Zundler S, Fischer A, Schillinger D, et al. The alpha4beta1 Homing Pathway Is Essential for Ileal Homing of Crohn's Disease Effector T Cells In Vivo. Inflamm Bowel Dis 2017;23:379-391.
39. Kalla R, Kennedy NA, Ventham NT, et al. Serum Calprotectin: A Novel Diagnostic and Prognostic Marker in Inflammatory Bowel Diseases. Am J Gastroenterol 2016;111:1796-1805.

40. Gecse KB, Brandse JF, van Wilpe S, et al. Impact of disease location on fecal calprotectin levels in Crohn's disease. Scand J Gastroenterol 2015;50:841-7.
41. Stevens T, D'Haens G R, Buskens CJ, et al. P150 Low diagnostic accuracy of faecal calprotectin to predict mucosal inflammation in Crohn's disease patients with active perianal fistulas. European Crohn's and Colitis Organization Congress, Vienna, Austria 2018.
42. Whitehead SJ, French J, Brookes MJ, et al. Between-assay variability of faecal calprotectin enzyme-linked immunosorbent assay kits. Ann Clin Biochem 2013;50:53-61.
43. Oyaert M, Boel A, Jacobs J, et al. Analytical performance and diagnostic accuracy of six different faecal calprotectin assays in inflammatory bowel disease. Clin Chem Lab Med 2017;55:1564-1573.
44. Rodriguez A, Yokomizo L, Christofferson M, et al. Correlation of rapid point-of-care vs send-out fecal calprotectin monitoring in pediatric inflammatory bowel disease. World J Gastrointest Pharmacol Ther 2017;8:127-130.
45. Hejl J, Theede K, Mollgren B, et al. Point of care testing of fecal calprotectin as a substitute for routine laboratory analysis. Pract Lab Med 2018;10:10-14.
46. Whitehead SJ, Ford C, Gama RM, et al. Effect of faecal calprotectin assay variability on the management of inflammatory bowel disease and potential role of faecal S100A12. J Clin Pathol 2017;70:1049-1056.
47. Jang HW, Kim HS, Park SJ, et al. Accuracy of three different fecal calprotectin tests in the diagnosis of inflammatory bowel disease. Intest Res 2016;14:305-313.
48. Wright EK, Kamm MA, De Cruz P, et al. Measurement of fecal calprotectin improves monitoring and detection of recurrence of Crohn's disease after surgery. Gastroenterology 2015;148:938-947. e1.

Author names in bold designate shared co-first authorship

Journal Pre-proof

Journal Pre-proof

Table and Figure Legends:

Table 1: Diagnostic Accuracy of EHI
True Positives (TPs), True Negatives (TNs), Sensitivity, Specificity, Positive Likelihood Ratio (PLR) and Negative Likelihood Ratio (NLR) of endoscopic Mucosal Healing Index (EHI) in Distinguishing AD vs ER in Validation Cohort 1 and Validation Cohort 2

Table 2: Comparative PPV and NPV of EHI, CRP and FC
Sensitivity, Specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) of endoscopic Mucosal Healing Index (EHI), CRP and FC in Distinguishing Endoscopic Remission versus Active Disease in the Two Validation Cohorts under Different Possible Prevalence of Active Disease.

Table 3: Comparative Diagnostic Accuracy of EHI to CRP and FC in Validation Cohort 1
True Positives (TPs), True Negatives (TNs), Sensitivity, Specificity, Positive Likelihood Ratio (PLR) and Negative Likelihood Ratio (NLR) of endoscopic Mucosal Healing Index (EHI), C-Reactive Protein (CRP) and Fecal Calprotectin (FC) in Distinguishing AD (n=183) vs ER (n=64) in the FC Sub-Cohort of Validation 1.

Table 4: Comparative Diagnostic Accuracy of EHI to CRP and FC in Validation Cohort 2
True Positives (TPs), True Negatives (TNs), Sensitivity, Specificity, Positive Likelihood Ratio (PLR) and Negative Likelihood Ratio (NLR) of endoscopic Mucosal Healing Index (EHI), C-Reactive Protein (CRP) and Fecal Calprotectin (FC) in Distinguishing AD (n=48) vs ER (n=33) in the FC Sub-Cohort of Validation 2.

Figure 1: Receiver operating characteristic curves of EHI and CRP
ROC curves of EHI and CRP for distinguishing active disease vs endoscopic remission in Validation Cohort 1 (A) and in Validation Cohort 2 (B). In Validation Cohort 1, mixed logistic regression models with random intercepts for individual subjects were used to combine multiple samples of same subjects.

Figure 2: Receiver operating characteristic curves of EHI and FC
ROC curves of EHI and FC for distinguishing active disease vs endoscopic remission in Validation Cohort 1 (A) and in Validation Cohort 2 (B). In Validation Cohort 1, mixed logistic regression models with random intercepts for individual subjects were used to combine multiple samples of same subjects.

Figure 3: Use of EHI for Monitoring
Boxplots of effect size (ES) of SES-CD, CDEIS, endoscopic Mucosal Healing Index (EHI), fecal calprotectin and C-reactive protein (CRP) in monitoring disease changes in Validation Cohort 1. (A) Between baseline and Week 12 (n=70). The median ES of EHI (1.10, IQR 0.52 - 1.83) was on par with those of SES-CD (1.53, IQR 0.67 - 2.23, P=0.077) and CDEIS (1.29, IQR 0.80 - 2.25, P=0.182), slightly better than that of fecal calprotectin (0.96, IQR 0.43 - 1.96, P=0.423) and significantly better than that of CRP (0.26, IQR 0.11 - 0.51, P<0.001). (B) Between baseline and Week 54 (n=59). The median ES of EHI (1.64, IQR 0.65 – 2.29) was at par with that of SES-CD (1.87, IQR 0.93 - 2.67, P=0.069), slightly better than those of CDEIS (1.50, IQR 0.81 - 2.17, P=0.997) and fecal calprotectin (1.16, IQR 0.51 – 2.32, P=0.574) and significantly better than that of CRP (0.21, IQR 0.09 - 0.56, P<0.001).

Journal Pre-proof

Journal Pre-proof

Table 1. True Positives (TPs), True Negatives (TNs), Sensitivity, Specificity, Positive Likelihood Ratio (PLR) and Negative Likelihood Ratio (NLR) of endoscopic Healing Index (EHI) in Distinguishing Active Disease (AD) vs Endoscopic Remission (ER) in Validation Cohort 1 and Validation Cohort 2

Cohort	EHI Threshold	MLG Probability[a]	TPs (n)	TNs (n)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PLR (95% CI)	NLR (95% CI)
Validation 1 ER (n=71) AD (n=204)	20	0.550	198	49	97.1 (93.7-98.9)	69.0 (56.9-79.5)	3.13 (2.21-4.44)	0.04 (0.02-0.10)
	30	0.746	173	65	84.8 (79.1-89.4)	91.5 (82.5-96.8)	10.04 (4.66-21.63)	0.17 (0.12-0.23)
	40	0.876	118	71	57.8 (50.7-64.7)	100.0 (94.9-100.0)	*infinity*	0.42 (0.36-0.50)
	50	0.945	76	71	37.3 (30.6-44.3)	100.0 (94.9-100.0)	*infinity*	0.63 (0.56-0.70)
Validation 2 ER (n=82) AD (n=113)	20	-	94	30	83.2 (75.0-89.6)	36.6 (26.2-48.0)	1.31 (1.10-1.58)	0.46 (0.28-0.76)
	30	-	74	49	65.5 (56.0-74.2)	59.8 (48.3-70.4)	1.63 (1.21-2.19)	0.58 (0.42-0.79)
	40	-	54	65	47.8 (38.3-57.4)	79.3 (68.9-87.4)	2.31 (1.45-3.67)	0.66 (0.54-0.81)
	50	-	34	72	30.1 (21.8-39.4)	87.8 (78.7-94.0)	2.47 (1.29-4.70)	0.80 (0.69-0.92)

[a]The population-averaged probability from the mixed logistic regression (MLG) model.

Journal Pre-proof

Journal Pre-proof

Table 2. Sensitivity, Specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) of endoscopic Healing Index (EHI), CRP and FC in Distinguishing Endoscopic Remission versus Active Disease in the Two Validation Cohorts under Different Possible Prevalence of Active Disease

	EHI Threshold 20				**EHI Threshold 50**			
	Validation 1		Validation 2		Validation 1		Validation 2	
Sensitivity	0.962		0.917		0.355		0.354	
Specificity	0.641		0.424		1.000		0.909	
Prevalence	PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV
0.05	0.124	0.997	0.077	0.990	1.000	0.967	0.170	0.964
0.25	0.472	0.981	0.347	0.939	1.000	0.823	0.565	0.808
0.40	0.641	0.962	0.515	0.885	1.000	0.699	0.722	0.679
0.60	0.801	0.918	0.705	0.773	1.000	0.508	0.854	0.484
0.75	0.889	0.849	0.827	0.630	1.000	0.341	0.921	0.319
	CRP Threshold 3				**CRP Threshold 5**			
	Validation 1		Validation 2		Validation 1		Validation 2	
Sensitivity	0.596		0.625		0.443		0.417	
Specificity	0.938		0.636		0.969		0.727	
Prevalence	PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV
0.05	0.336	0.978	0.083	0.970	0.429	0.971	0.074	0.960
0.25	0.762	0.874	0.364	0.836	0.826	0.839	0.337	0.789
0.40	0.865	0.777	0.534	0.718	0.905	0.723	0.505	0.652
0.60	0.935	0.608	0.720	0.531	0.955	0.537	0.696	0.454
0.75	0.966	0.436	0.837	0.361	0.977	0.367	0.821	0.294

	FC Threshold 50				**FC Threshold 250**			
	Validation 1		Validation 2		Validation 1		Validation 2	
Sensitivity	1.000		0.750		0.683		0.438	
Specificity	0.063		0.788		0.891		1.000	
Prevalence	PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV
0.05	0.053	1.000	0.157	0.984	0.248	0.982	1.000	0.971
0.25	0.262	1.000	0.541	0.904	0.676	0.894	1.000	0.842
0.40	0.416	1.000	0.702	0.825	0.807	0.808	1.000	0.727
0.60	0.616	1.000	0.841	0.678	0.904	0.652	1.000	0.543
0.75	0.762	1.000	0.914	0.512	0.949	0.484	1.000	0.372

Journal Pre-proof

Journal Pre-proof

Journal Pre-proof

Table 3. True Positives (TPs), True Negatives (TNs), Sensitivity, Specificity, Positive Likelihood Ratio (PLR) and Negative Likelihood Ratio (NLR) of endoscopic Healing Index (EHI), C-Reactive Protein (CRP) and Fecal Calprotectin (FC) in Distinguishing Active Disease (n=183) vs Endoscopic Remission (n=64) in the FC Sub-Cohort of Validation 1

Test	Threshold	MLG Probability[a]	TPs (n)	TNs (n)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PLR (95% CI)	NLR (95% CI)
EHI	20	0.542	176	41	96.2 (92.3-98.4)	64.1 (51.1-75.7)	2.68 (1.93-3.72)	0.06 (0.03-0.13)
	30	0.728	153	59	83.6 (77.4-88.7)	92.2 (82.7-97.4)	10.70 (4.60-24.89)	0.18 (0.13-0.25)
	40	0.858	109	64	59.6 (52.1-66.7)	100.0 (94.4-100.0)	*infinity*	0.40 (0.34-0.48)
	50	0.932	65	64	35.5 (28.6-42.9)	100.0 (94.4-100.0)	*infinity*	0.65 (0.58-0.72)
CRP (mg/L)	3	0.830	109	60	59.6 (52.1-66.7)	93.8 (84.8-98.3)	9.53 (3.66-24.80)	0.43 (0.36-0.52)
	5	0.868	81	62	44.3 (36.9-51.8)	96.9 (89.2-99.6)	14.16 (3.59-55.95)	0.58 (0.50-0.66)
	10	0.908	49	64	26.8 (20.5-33.8)	100.0 (94.4-100.0)	*infinity*	0.73 (0.67-0.80)
FC (µg/g)	50	0.230	183	4	100.0 (98.0-100.0)	6.2 (1.7-15.2)	1.07 (1.00-1.14)	0.00 (0.00 -)
	150	0.624	144	53	78.7 (72.0-84.4)	82.8 (71.3-91.1)	4.58 (2.66-7.88)	0.26 (0.19-0.35)
	250	0.787	125	57	68.3 (61.0-75.0)	89.1 (78.8-95.5)	6.25 (3.08-12.65)	0.36 (0.28-0.45)

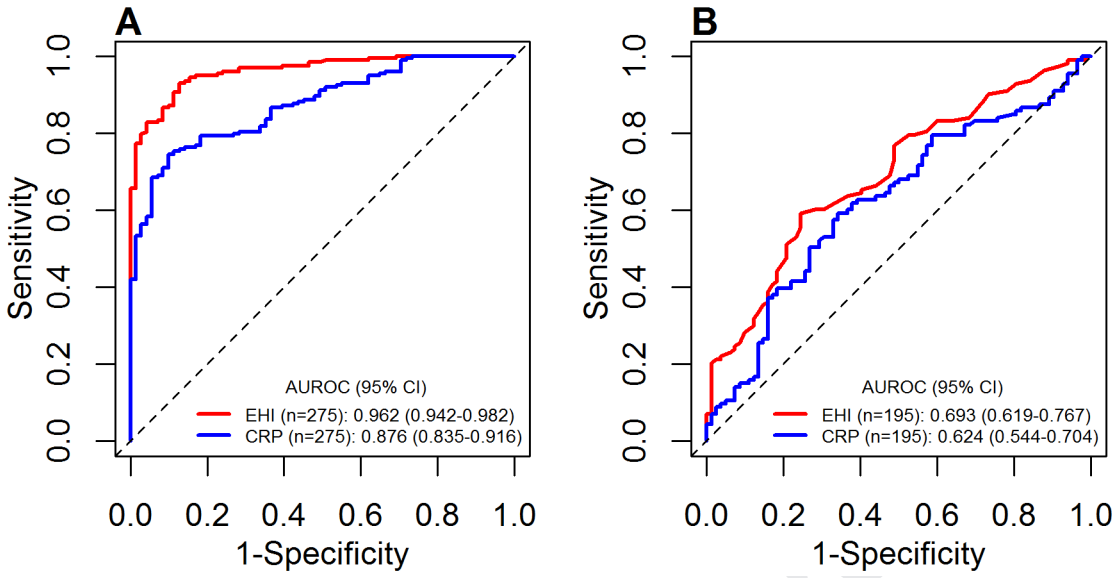
[a]The population-averaged probability from the mixed logistic regression (MLG) models.

Journal Pre-proof

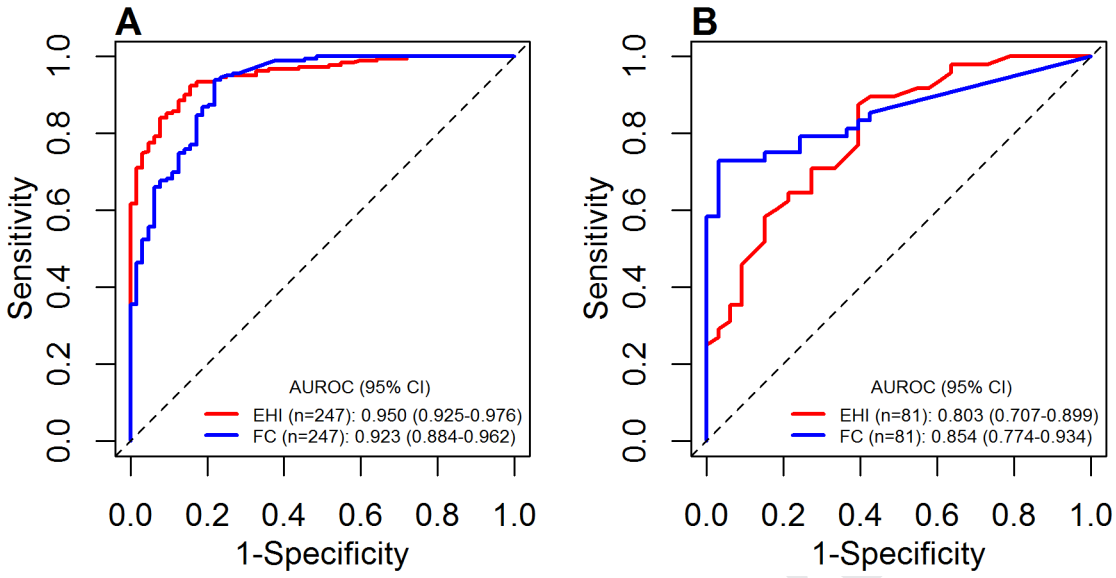
Journal Pre-proof

Table 4. True Positives (TPs), True Negatives (TNs), Sensitivity, Specificity, Positive Likelihood Ratio (PLR) and Negative Likelihood Ratio (NLR) of endoscopic Healing Index (EHI), C-Reactive Protein (CRP) and Fecal Calprotectin (FC) in Distinguishing Active Disease (n=48) vs Endoscopic Remission (n=33) in the FC Sub-Cohort of Validation 2

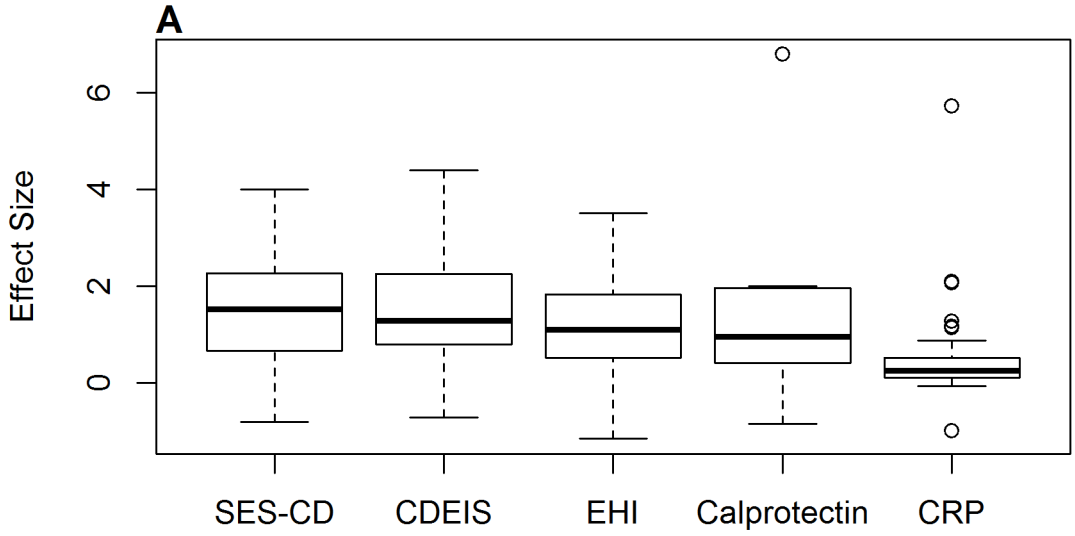
Test	Threshold	TPs (n)	TNs (n)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PLR (95% CI)	NLR (95% CI)
EHI	20	44	14	91.7 (80.0-97.7)	42.4 (25.5-60.8)	1.59 (1.17-2.16)	0.20 (0.07-0.54)
	30	34	24	70.8 (55.9-83.0)	72.7 (54.5-86.7)	2.60 (1.45-4.67)	0.40 (0.25-0.65)
	40	25	28	52.1 (37.2-66.7)	84.8 (68.1-94.9)	3.44 (1.47-8.06)	0.57 (0.41-0.78)
	50	17	30	35.4 (22.2-50.5)	90.9 (75.7-98.1)	3.90 (1.24-12.24)	0.71 (0.56-0.90)
CRP (mg/L)	3	30	21	62.5 (47.4-76.0)	63.6 (45.1-79.6)	1.72 (1.04-2.84)	0.59 (0.38-0.92)
	5	20	24	41.7 (27.6-56.8)	72.7 (54.5-86.7)	1.53 (0.80-2.93)	0.80 (0.58-1.10)
	10	11	26	22.9 (12.0-37.3)	78.8 (61.1-91.0)	1.08 (0.47-2.50)	0.98 (0.77-1.24)
FC (µg/g)	50	36	26	75.0 (60.4-86.4)	78.8 (61.1-91.0)	3.54 (1.80-6.97)	0.32 (0.19-0.53)
	150	28	33	58.3 (43.2-72.4)	100.0 (89.4-100.0)	*infinity*	0.42 (0.30-0.58)
	250	21	33	43.8 (29.5-58.8)	100.0 (89.4-100.0)	*infinity*	0.56 (0.44-0.72)

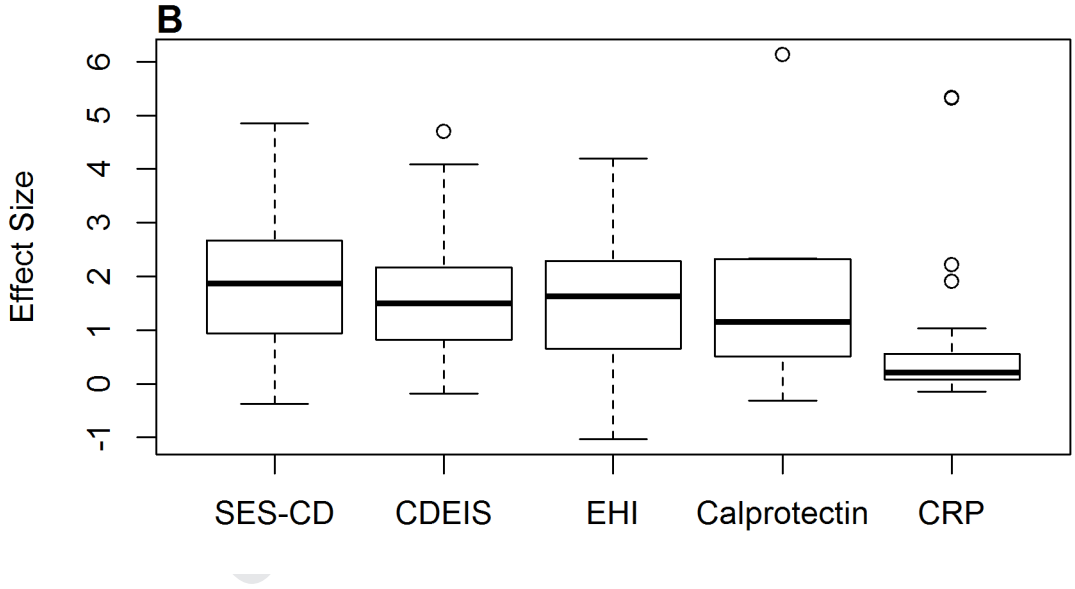


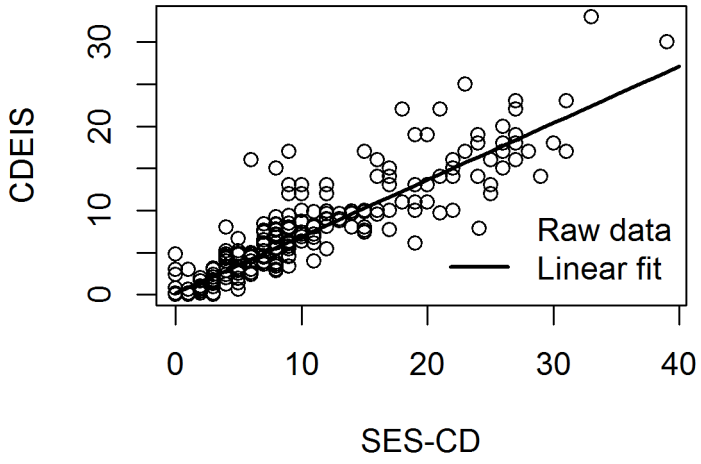
Journal Pre-proof



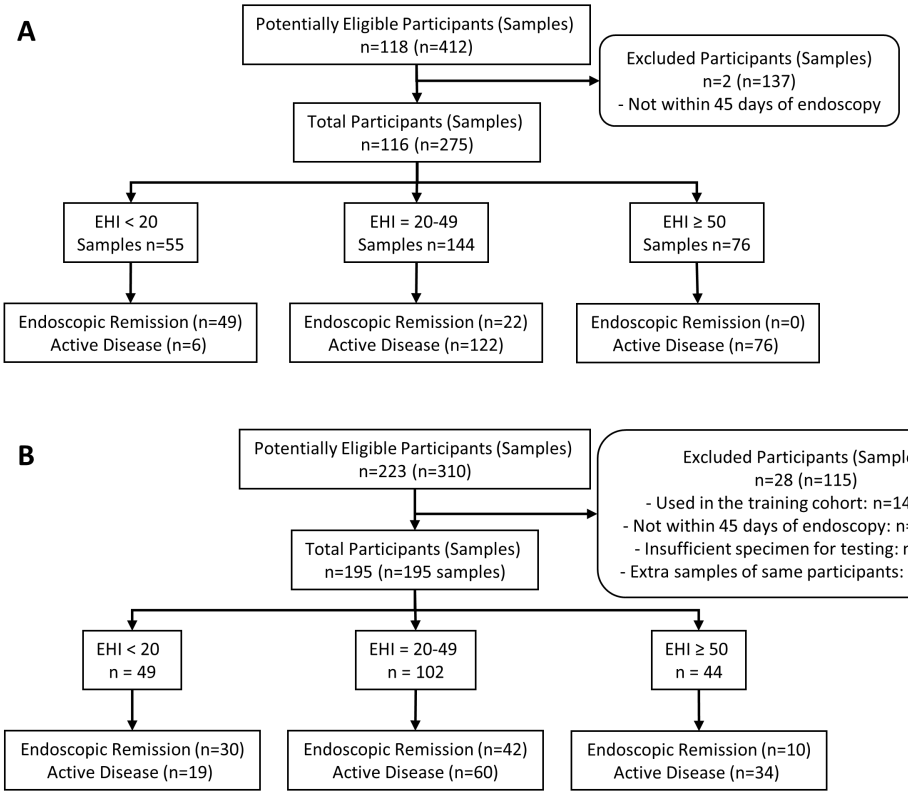
Journal Pre-proof

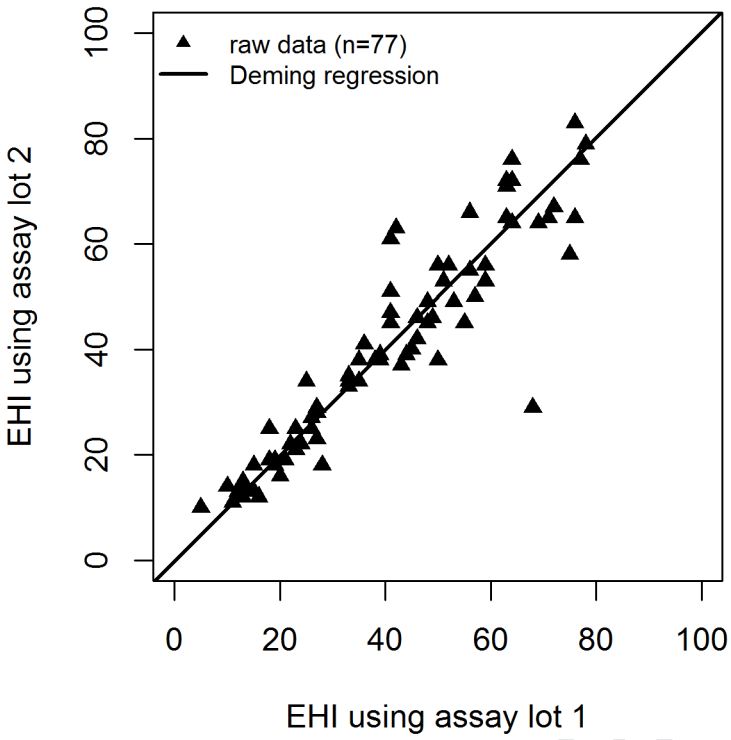

A
Effect Size
0
2
4
6
SES-CD
CDEIS
EHI
Calprotectin
CRP


B
Effect Size
-1
0
1
2
3
4
5
6
SES-CD
CDEIS
EHI
Calprotectin
CRP

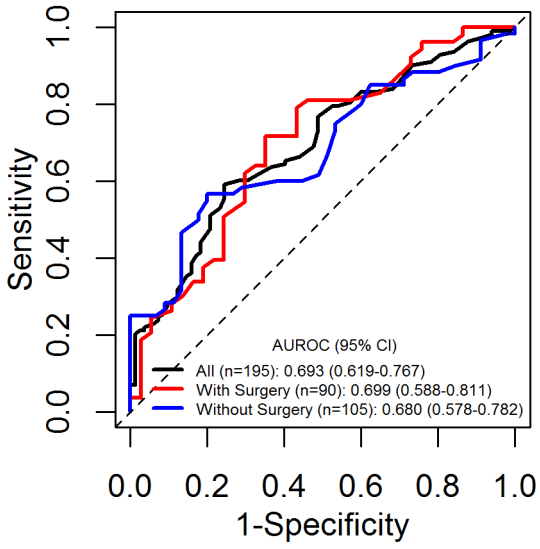


Journal Pre-proof

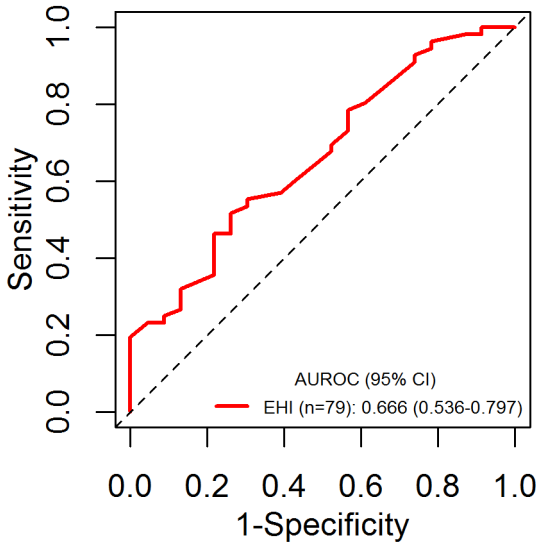

A
Potentially Eligible Participants (Samples)
n=118 (n=412)
Excluded Participants (Samples)
n=2 (n=137)
- Not within 45 days of endoscopy
Total Participants (Samples)
n=116 (n=275)
EHI < 20
Samples n=55
EHI = 20-49
Samples n=144
EHI ≥ 50
Samples n=76
Endoscopic Remission (n=49)
Active Disease (n=6)
Endoscopic Remission (n=22)
Active Disease (n=122)
Endoscopic Remission (n=0)
Active Disease (n=76)
B
Potentially Eligible Participants (Samples)
n=223 (n=310)
Excluded Participants (Samples)
n=28 (n=115)
- Used in the training cohort: n=14 (n=45)
- Not within 45 days of endoscopy: n=13 (n=18)
- Insufficient specimen for testing: n=1 (n=1)
- Extra samples of same participants: n=0 (n=51)
Total Participants (Samples)
n=195 (n=195 samples)
EHI < 20
n = 49
EHI = 20-49
n = 102
EHI ≥ 50
n = 44
Endoscopic Remission (n=30)
Active Disease (n=19)
Endoscopic Remission (n=42)
Active Disease (n=60)
Endoscopic Remission (n=10)
Active Disease (n=34)



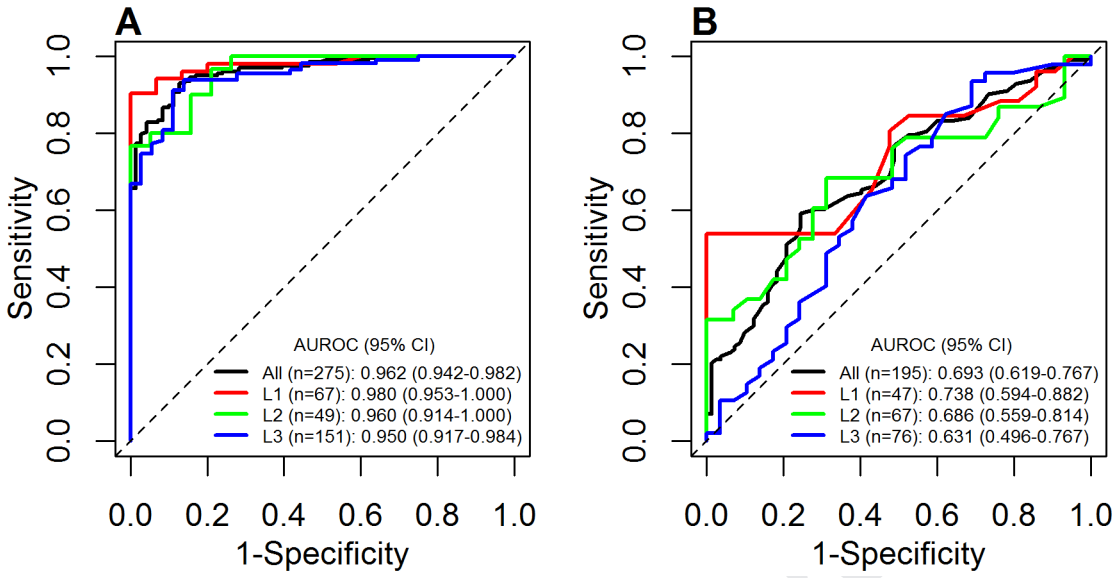
Journal Pre-proof



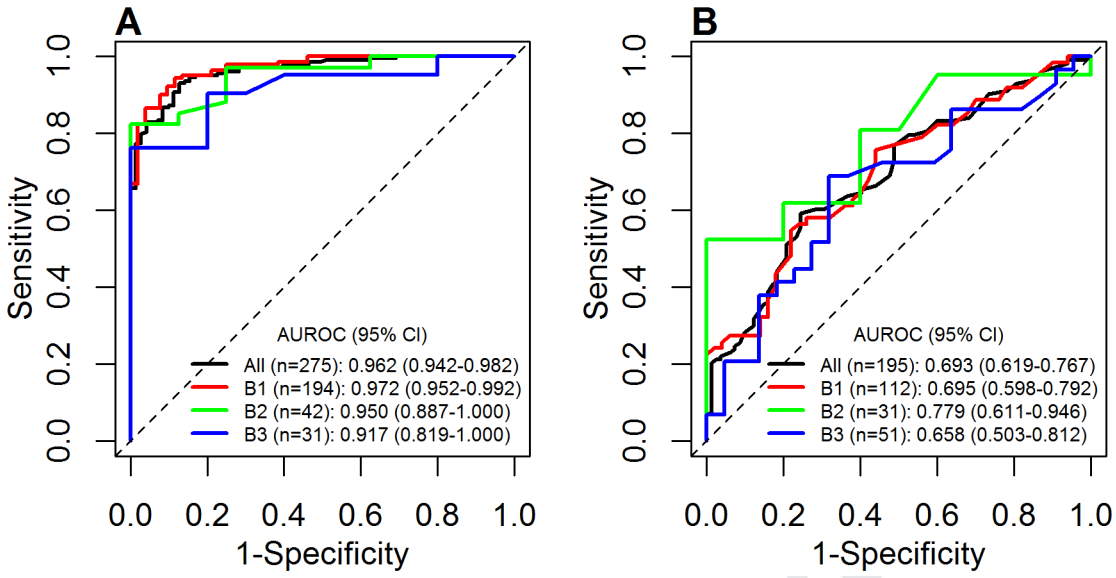
Journal Pre-proof



Journal Pre-proof



Journal Pre-proof



Journal Pre-proof

Journal Pre-proof

Supplementary Figure Legends:

Supplementary Figure 1: SES-CD and CDEIS scores in Validation Cohort 1
The linear fit (CDEIS = 0.1569 + 0.6744*SES-CD) was used to convert SES-CD scores to CDEIS scores for samples that had only SES-CD scores in the training cohort.

Supplementary Figure 2: Patient and sample flowchart
(A) Validation Cohort 1 and (B) Validation Cohort 2.

Supplementary Figure 3: Reproducibility of EHI
Reproducibility of EHI when same samples were analyzed using two different lots of reagents. Serum samples from clinically diagnosed CD patients (n = 77) were used to study reproducibility of EHI. The Deming regression had a slope of 1.005 (95% CI: 0.926 to 1.087) and an intercept of -0.298 (95% CI: -2.790 to 2.144).

Supplementary Figure 4: Diagnostic Accuracy of EHI in Patients with Surgery
The receiver operating characteristic (ROC) curves of endoscopic Mucosal Healing Index (EHI) in distinguishing AD vs ER in patient sub-cohorts with (red) or without (blue) a history of IBD-related surgery in Validation Cohort 2. The area under the ROC curve (AUROC) of EHI in the two sub-cohorts was not significantly different (p=0.801).

Supplementary Figure 5: Diagnostic Accuracy of EHI in Endohistopathologic Healing
ROC curve of EHI in distinguishing EHPH versus non-EHPH in Validation Cohort 2.

Supplementary Figure 6: Comparative Accuracy of EHI Across Disease Locations
The receiver operating characteristic (ROC) curves of endoscopic Mucosal Healing Index (EHI) in distinguishing AD vs ER by disease location in (A) Validation Cohort 1 and (B) Validation Cohort 2. A mixed logistic regression model with random intercepts for individual subjects was used to combine multiple samples of same subjects in Validation Cohort 1. The minimum pairwise P values comparing the area under the ROC curve (AUROC) of EHI on patients with different disease locations were (A) 0.171 and (B) 0.292, respectively, indicating that EHI performance was consistent across disease locations.

Supplementary Figure 7: Comparative Accuracy of EHI Across Disease Behaviors
ROC curves of EHI in distinguishing active disease vs endoscopic remission by disease behavior in (A) Validation Cohort 1 and (B) Validation Cohort 2. A mixed logistic regression model with random intercepts for individual subjects was used to combine multiple samples of same subjects in Validation Cohort 1. The pairwise P values comparing the area under the ROC curve (AUROC) of EHI on different groups of patients were (A) $P \geq 0.290$, and (B) $P \geq 0.300$, respectively.

Journal Pre-proof

SUPPLEMENTARY METHODS

Sample Storage and Testing

Serum CRP was tested using a turbidity assay (hsCRP, Beckman Coulter, Brea, CA). Other serum biomarkers were measured via multiplexed fluorescent immunoassays. Serum specimens were added to a mixture of color-coded beads, which were pre-coated with analyte-specific capture antibodies. Biotinylated detection antibodies specific to the target analytes were then added to form an antibody-antigen sandwich. Afterwards phycoerythrin (PE)-conjugated streptavidin was added, which bound to the biotinylated detection antibodies. The magnitude of the PE-derived signal, which was directly proportional to the amount of target analyte in the sample, was then detected in a flow-based instrument. Values of FC from the STORI trial and the TAILORIX trial were used directly without retesting (Buhlmann Laboratories, Schonenbuch, Switzerland). Values of FC in Validation cohort 2 were measured using the QUANTA Lite® Calprotectin Extended Range assay (Inova Diagnostics, San Diego, CA).

Statistical Analysis

Continuous variables were reported as medians with interquartile ranges (IQR), and compared between groups using the Mann-Whitney test. Categorical variables are reported as numbers (n) and percentages (%), and compared between groups using the Fisher's exact test. The Delong method was used for computing the 95% confidence interval (CI) of AUROC and for comparing AUROCs of different biomarkers on paired samples.[23] Exact binomial confidence limits were used for the 95% CIs of sensitivity and specificity. The 95% CIs of PLR and NLR were computed using formulae provided by Simel et al.[24] Pairwise Wilcoxon rank sum test was used for comparing effect size of different variables. A p value (two-sided) of 0.05 or lower was considered as significant. All data analysis was carried out using JMP (version 12.0; SAS Institute, Cary, NC) or R (version 3.3.2).

For Validation cohort 1, where longitudinal samples were available, a mixed-effect logistic regression modeling was utilized to assess the performance of EHI, CRP and FC (response: disease status of endoscopic remission or active disease; fixed effect: EHI, CRP or FC; random effect: random intercepts for study subjects). The values of CRP and FC were first logarithmic transformed in the modeling. For samples whose FC test result was 0, the corresponding FC values were set to 10 µg/g, which was well below the minimal nonzero FC value of 31 µg/g observed in the cohort. Subsequently, Effect sizes[25] of SES-CD, CDEIS, EHI, FC and CRP were calculated between baseline and week 12 and between baseline and week 54 to assess the responsiveness of those variables. Since the underlying data were mostly not normally distributed, the corresponding median and inter quartile range (IQR) were reported instead of the mean.

Journal Pre-proof

Supplementary Table 1. The 47 biomarkers evaluated for the development of endoscopic Healing Index (EHI) and their corresponding signaling pathways, from which 13 biomarkers were selected for the EHI assay from logistic regression modeling

Angiogenesis	Cell Adhesion	Growth Factors	Immune Modulation	Inflammation	Matrix Remodeling
k = 4	k = 6	k = 10	k = 16	k = 5	k = 6
Ang1***	Ceacam1*	TGFα*	IL7***	CRP*	EMMPRIN***
Ang2[a]	VCAM-1**	BTC[1]	GM-CSF[2]	SAA1*	MMP-1*
VEGFα[1]	Alcam[1]	EGF[1]	IL1β[2]	ADA[1]	MMP-2[b]
FGF2[2]	α4β7[1]	SCF[1]	IL2[2]	TWEAK[1]	MMP-3**
	ICAM-1[1]	AREG[2]	IL5[2]	IFN-γ[2]	MMP-9***
	MAdCAM[1]	ANXA13[2]	IL6[2]		Fibronectin[1]
		EREG[2]	IL10[2]		
		HB-EGF[2]	IL12/23p40[2]		
		HGF[2]	IL13[2]		
		TGFβ[2]	IL15[2]		
			IL17a[2]		
			IL17f[2]		
			IL22[2]		
			IL23[2]		
			IL31[2]		
			IL33[2]		

[1]These 11 biomarkers did not enhance the performance of EHI and were not included in the EHI assay.
[2]These 23 biomarkers were eliminated due to poor analytical reproducibility, low detection rate in serum specimens, and/or lack of correlation to disease severity in preliminary studies.
p-values of the 13 selected biomarkers were as follows: ***$p<0.001$, **$0.001\leq p<0.01$, *$0.01\leq p<0.05$, [a]$p=0.094$, [b]$p=0.121$

Journal Pre-proof

Journal Pre-proof

Supplementary Table 2. Subject and Sample Characteristics of the Study Cohorts[a]

Characteristics	Training	Validation 1 (TAILORIX)	Validation 2 (UCSD)	P value[b] (Training vs Validation 1)	P value[b] (Training vs Validation 2)	P value[b] (Validation 1 vs 2)
Subjects						
n	278	116	195			
Age (years)	30.0 (24.9-40.0)[2]	30.2 (22.4-45.2)	38.5 (28.0-52.0)[17]	0.842	<0.001	<0.001
Female sex	128 (46.0)	69 (59.5)	92 (49.7)[10]	0.020	0.449	0.123
Race/Ethnicity				-	0.039	-
African American	5 (1.8)	-	5 (2.6)			
Asian	14 (5.0)	-	3 (1.5)			
White	160 (57.6)	-	125 (64.1)			
Hispanic	5 (1.8)	-	8 (4.1)			
Other	5 (1.8)	-	0 (0.0)			
Unknown	89 (32.0)	-	54 (27.7)			
CD duration (years)	4.0 (3.0-12.5)[260]	0.7 (0.1-6.9)	11.0 (5.0-19.0)[16]	0.002	0.018	<0.001
Age at diagnosis (years)	[100]		[12]	<0.001	0.017	<0.001
A1: ≤16	55 (30.9)	6 (5.2)	49 (26.8)			
A2: 17-40	106 (59.6)	86 (74.1)	97 (53.0)			
A3: >40	17 (9.6)	24 (20.7)	37 (20.2)			
CD location	[100]	[5]	[5]	0.738	<0.001	0.002
L1: Ileal	43 (24.2)	27 (24.3)	47 (24.7)			
L2: Colonic	26 (14.6)	20 (18.0)	67 (35.3)			
L3: Ileocolonic	109 (61.2)	64 (57.7)	76 (40.0)			
CD behavior	[186]	[5]	[1]	<0.001	<0.001	0.006
B1: Non-stricturing, non-penetrating	26 (28.3)	81 (73.0)	112 (57.7)			
B2: Stricturing	57 (62.0)	17 (15.3)	31 (16.0)			
B3: Penetrating	9 (9.8)	13 (11.7)	51 (26.3)			
Perianal disease modifier	29 (17.5)[112]	31 (27.9) [5]	21 (10.8)	0.052	0.069	<0.001
Biologic medication use	-	0 (0)	139 (77.2)[15]	-	-	<0.001
History of IBD related surgery	-	12 (10.3)	90 (46.2)	-	-	<0.001
Samples						
n	335	275	195			
Endoscopic remission[c]	159 (47.5)	71 (25.8)	82 (42.1)	<0.001	0.241	<0.001
CDEIS	2.8 (0.2-6.0)[202]	4.4 (0.8-9.1)	-	0.016	-	-
SES-CD	6.0 (1.0-12.0)[133]	6.0 (2.0-12.0)	3.0 (0.0-6.5)	0.321	<0.001	<0.001
CRP (mg/L)	2.0 (0.7-6.5)	2.5 (0.5-7.2)	2.6 (0.7-7.1)	0.586	0.172	0.460
Fecal calprotectin (μg/g)	50.8 (30.1-270.3)[273]	336.0 (100.0-1197.5)[28]	55.0 (0.0-251.1)[114]	<0.001	0.086	<0.001
EHI	32 (20-44)	38 (25-53)	32 (19.5-46.5)	0.001	0.752	0.006

[a]Continuous variables are reported as median (inter-quartile range), categorical variables are reported as n (%), and numbers of missing data, if any, are listed inside brackets ([n]).

[b]Based on Mann-Whitney test for continuous variables and Fisher's exact test for categorical variables.

[c]On the training cohort, SES-CD scores were first converted to CDEIS scores by: CDEIS = 0.1569 + 0.6744*SES-CD (see Supplementary Figure 1). Endoscopic remission was then defined as either original or converted CDEIS score < 3. On validation cohorts, endoscopic remission was defined as a total SES-CD of ≤ 2 and ≤ 1 in each segment.

Journal Pre-proof

Journal Pre-proof

Supplementary Table 3. Subject and Sample Characteristics of the Training Cohort[a]

Characteristics	Total	U Padua, Italy	MSH, Toronto	STORI	UCSD
Subjects					
n	278	18	146	83	31
Age (years)	30.0 (24.9-40.0)[2]	34.5 (26.5-51.5)	29.0 (23.0-39.2)[2]	31.6 (25.6-39.2)	30.0 (23.5-45.0)
Female sex	128 (46.0)	5 (27.8)	63 (43.2)	45 (54.2)	15 (48.4)
Race/Ethnicity					
African American	5 (1.8)	0 (0.0)	5 (3.4)	-	0 (0.0)
Asian	14 (5.0)	0 (0.0)	13 (8.9)	-	1 (3.2)
White	160 (57.6)	18 (100.0)	115 (78.8)	-	27 (87.1)
Hispanic	5 (1.8)	0 (0.0)	3 (2.1)	-	2 (6.5)
Other	5 (1.8)	0 (0.0)	4 (2.7)	-	1 (3.2)
Unknown	89 (28.7)	0 (0.0)	6 (4.1)	-	0 (0.0)
Disease duration (years)	4.0 (3.0-12.5)[260]	4.0 (3.0-12.5)	-	-	-
Age at diagnosis (years)	[100]		[17]	[83]	
A1: ≤16	55 (30.9)	0 (0.0)	47 (36.4)	-	8 (25.8)
A2: 17-40	106 (59.6)	13 (72.2)	74 (57.4)	-	19 (61.3)
A3: >40	17 (9.6)	5 (27.8)	8 (6.2)	-	4 (12.9)
CD location	[100]		[17]	[83]	
L1: Ileal	43 (24.2)	9 (50.0)	27 (20.9)	-	7 (22.6)
L2: Colonic	26 (14.6)	3 (16.7)	11 (8.5)	-	12 (38.7)
L3: Ileocolonic	109 (61.2)	6 (33.3)	91 (70.5)	-	12 (38.7)
CD behavior	[186]		[102]	[83]	[1]
B1: Non-stricturing, non-penetrating	26 (28.3)	7 (38.9)	0 (0.0)	-	19 (63.3)
B2: Stricturing	57 (62.0)	6 (33.3)	44 (100.0)	-	7 (23.3)
B3: Penetrating	9 (9.8)	5 (27.8)	0 (0.0)	-	4 (13.3)
Perianal disease modifier	29 (17.5)[112]	2 (11.1)	25 (17.1)	-	2 (100) [29]
Samples					
n	335	50	157	83	45
Endoscopic Remission[b]	159 (47.5)	5 (10.0)	67 (42.7)	63 (75.9)	24 (53.3)
CDEIS	2.8 (0.2-6.0)[202]	6.6 (4.0-17.6)	-	0.7 (0.0-2.8)	-
SES-CD	6.0 (1.0-12.0)[133]	-	6.0 (2.0-12.0)	-	4.0 (0.0-9.0)
CDAI	59.5 (24.2-123.8)[205]	147.0 (120.0-240.0)[3]	-	36.5 (17.2-61.1)	-
CRP (mg/L)	2.0 (0.7-6.5)	1.3 (0.6-4.4)	3.6 (1.0-6.5)	1.4 (0.6-2.7)	2.0 (0.6-3.9)
Fecal calprotectin (μg/g)	50.8 (30.1-270.3)[273]	-	-	50.8 (30.1-270.3)[21]	-
EHI	32 (20-44)	42 (33-59)	33 (22-47)	21 (13-29)	39 (30-50)

[a]Continuous variables are reported as median (inter-quartile range), categorical variables are reported as n (%), and numbers of missing data, if any, are listed inside brackets ([n]).

[b]SES-CD scores were converted to CDEIS scores by: CDEIS = 0.1569 + 0.6744*SES-CD (see Supplementary Figure 1). Disease activity status was defined by either original or converted CDEIS scores as remission (CDEIS < 3) or active (CDEIS ≥ 3).

Journal Pre-proof

Supplementary Table 4. Subject and Sample Characteristics of Validation Cohort 1[a]

Characteristics	Total	Baseline	Week 12	Week 54
Subjects[b]				
n	116	102	98	75
Age (years)	30.2 (22.4-45.2)	30.9 (24.1-45.6)	30.9 (22.9-45.7)	30.4 (23.1-44.5)
Female sex	69 (59.5)	61 (59.8)	58 (59.2)	43 (57.3)
CD duration (years)	0.7 (0.1-6.9)	0.9 (0.1-7.3)	0.5 (0.1-6.1)	0.5 (0.0-5.6)
Age at diagnosis (years)				
A1: ≤16	6 (5.2)	5 (4.9)	3 (3.1)	3 (4.0)
A2: 17-40	86 (74.1)	77 (75.5)	72 (73.5)	55 (73.3)
A3: >40	24 (20.7)	20 (19.6)	23 (23.5)	17 (22.7)
CD location	[5]	[3]	[4]	[1]
L1: Ileal	27 (24.3)	26 (26.3)	23 (24.5)	18 (24.3)
L2: Colonic	20 (18.0)	16 (16.2)	17 (18.1)	16 (21.6)
L3: Ileocolonic	64 (57.7)	57 (57.6)	54 (57.4)	40 (54.1)
CD behavior	[5]	[3]	[4]	[1]
B1: Non-stricturing, non-penetrating	81 (73.0)	73 (73.7)	69 (73.4)	52 (70.3)
B2: Stricturing	17 (15.3)	16 (16.2)	13 (13.8)	13 (17.6)
B3: Penetrating	13 (11.7)	10 (10.1)	12 (12.8)	9 (12.2)
Perianal disease modifier	31 (27.9) [5]	26 (26.3) [3]	25 (26.6) [4]	21 (28.4) [1]
History of IBD related surgery	12 (10.3)	12 (11.8)	8 (8.2)	4 (5.3)
Samples				
n	275	102	98	75
Endoscopic remission[c]	71 (25.8)	0 (0.0)	26 (26.5)	45 (60.0)
CDEIS	4.4 (0.8-9.1)	10.0 (7.4-16.0)	3.0 (0.7-5.2)	0.1 (0.0-2.6)
SES-CD	6.0 (2.0-12.0)	15.0 (9.0-22.0)	4.0 (2.0-8.0)	1.0 (0.0-3.0)
CDAI	166.0 (72.5-261.0) [4]	279.5 (233.0-321.8)	112.0 (58.5-181.5) [3]	66.5 (39.8-115.2)[1]
CRP (mg/L)	2.5 (0.5-7.2)	8.2 (3.5-14.3)	0.9 (0.3-3.5)	0.8 (0.3-2.5)
Fecal calprotectin (µg/g)	336.0 (100.0-1197.5) [28]	1462.5 (709.4-1800.0) [6]	122.0 (100.0-430.5) [17]	105.2 (100.0-215.8) [5]
EHI	38.0 (25.0-53.0)	55.5 (42.0-72.8)	33.5 (23.0-41.8)	25.0 (19.0-37.5)

[a]Continuous variables are reported as median (inter-quartile range), categorical variables are reported as n (%), and numbers of missing data, if any, are listed inside brackets ([n]).

[b]Subjects in the three time points were subsets of the full cohort that contributed the corresponding samples.

[c]Endoscopic remission was defined as a total SES-CD of ≤ 2 and ≤ 1 in each segment.

Journal Pre-proof

Supplementary Table 5. Subject and Sample Characteristics of Validation Cohort 2 and Sub-Cohorts With or Without Fecal Calprotectin[a]

Characteristics	Full Cohort	Sub-Cohort (with Calprotectin)	Sub-Cohort (without Calprotectin)	P Value[b]
n	195	81	114	
Age (years)	38.5 (28.0-52.0)[17]	37.0 (28.0-47.0)[9]	39.0 (28.0-53.8)[8]	0.603
Female sex	92 (49.7)[10]	36 (48.0)[6]	56 (50.9)[4]	0.765
Race/Ethnicity				<0.001
African American	5 (2.6)	1 (1.2)	4 (3.5)	
Asian	3 (1.5)	2 (2.5)	1 (0.9)	
White	125 (64.1)	33 (40.7)	92 (80.7)	
Hispanic	8 (4.1)	3 (3.7)	5 (4.4)	
Unknown	54 (27.7)	42 (51.9)	12 (10.5)	
CD duration (years)	11.0 (5.0-19.0)[16]	11.0 (5.0-20.0)[8]	11.0 (5.0-18.0)[8]	0.684
Age at diagnosis (years)	[12]	[8]	[4]	0.530
A1: ≤16	49 (26.8)	22 (30.1)	27 (24.5)	
A2: 17-40	97 (53.0)	39 (53.4)	58 (52.7)	
A3: >40	37 (20.2)	12 (16.4)	25 (22.7)	
CD location	[5]	[3]	[2]	0.566
L1: Ileal	47 (24.7)	22 (28.2)	25 (22.3)	
L2: Colonic	67 (35.3)	28 (35.9)	39 (34.8)	
L3: Ileocolonic	76 (40.0)	28 (35.9)	48 (42.9)	
CD behavior	[1]		[1]	0.444
B1: Non-stricturing, non-penetrating	112 (57.7)	51 (63.0)	61 (54.0)	
B2: Stricturing	31 (16.0)	12 (14.8)	19 (16.8)	
B3: Penetrating	51 (26.3)	18 (22.2)	33 (29.2)	
Perianal disease modifier	21 (10.8)	11 (13.6)	10 (8.8)	0.350
Biologic medication use	139 (77.2)[15]	59 (76.6)[4]	80 (77.7)[11]	1.000
History of IBD related surgery	90 (46.2)	33 (40.7)	57 (50.0)	0.244
Endoscopic remission[c]	82 (42.1)	33 (40.7)	49 (43.0)	0.771
Endohistopathologic healing[d]	23 (29.1)[116]	11 (28.9)[43]	12 (29.3)[73]	1.000
SES-CD	3.0 (0.0-6.5)	3.0 (0.0-7.0)	3.0 (0.0-6.0)	0.609
CDAI PRO2	7.7 (2.9-15.5)[39]	6.9 (2.4-15.4)[10]	8.3 (3.7-15.4)[29]	0.457
GHAS	3.0 (1.0-6.0)[116]	3.0 (1.0-6.0)[43]	4.0 (1.0-6.0)[73]	0.886
CRP (mg/L)	2.6 (0.7-7.1)	3.2 (0.7-7.5)	2.4 (0.8-6.2)	0.476
Fecal calprotectin (µg/g)	55.0 (0.0-251.1)[114]	55.0 (0.0-251.1)	-	-
EHI	32 (19.5-46.5)	32 (21-47)	32.0 (19.0-46.0)	0.576

[a]Continuous variables are reported as median (inter-quartile range), categorical variables are reported as n (%), and numbers of missing data, if any, are listed inside brackets ([n]).
[b]Based on Mann-Whitney test for continuous variables and Fisher's exact test for categorical variables.

[c]Endoscopic remission was defined as a total SES-CD of ≤ 2 and ≤ 1 in each segment.

[d]Endohistopathologic healing was defined as achieving both endoscopic remission and histologic remission (GHAS ≤ 2).

Journal Pre-proof

Journal Pre-proof

Supplementary Table 6. True Positives (TPs), True Negatives (TNs), Sensitivity, Specificity, Positive Likelihood Ratio (PLR) and Negative Likelihood Ratio (NLR) of endoscopic Healing Index (EHI) in Distinguishing Active Disease (n=176) vs Endoscopic Remission (n=159) versus in the Training Cohort

'EHI' Threshold	TPs (n)	TNs (n)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PLR (95% CI)	NLR (95% CI)
20	159	62	90.3 (85.0-94.3)	39.0 (31.4-47.0)	1.48 (1.30-1.69)	0.25 (0.15-0.41)
30	129	100	73.3 (66.1-79.7)	62.9 (54.9-70.4)	1.98 (1.58-2.46)	0.43 (0.32-0.56)
40	86	131	48.9 (41.3-56.5)	82.4 (75.6-88.0)	2.78 (1.92-4.01)	0.62 (0.53-0.73)
50	54	151	30.7 (24.0-38.1)	95.0 (90.3-97.8)	6.10 (3.00-12.41)	0.73 (0.66-0.81)

Journal Pre-proof

Journal Pre-proof

Supplementary Table 7. True Positives (TPs), True Negatives (TNs), Sensitivity, Specificity, Positive Likelihood Ratio (PLR) and Negative Likelihood Ratio (NLR) of endoscopic Healing Index (EHI) in Distinguishing Active Disease (AD) vs Endoscopic Remission (ER) by Disease Location in Validation Cohort 1

CD Location	EHI Threshold	MLG Probability[a]	TPs (n)	TNs (n)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PLR (95% CI)	NLR (95% CI)
All	20	0.550	198	49	97.1	69.0	3.13	0.04
ER (n=71)					(93.7-98.9)	(56.9-79.5)	(2.21-4.44)	(0.02-0.10)
AD (n=204)	30	0.746	173	65	84.8	91.5	10.04	0.17
					(79.1-89.4)	(82.5-96.8)	(4.66-21.63)	(0.12-0.23)
	40	0.876	118	71	57.8	100.0	*infinity*	0.42
					(50.7-64.7)	(94.9-100.0)		(0.36-0.50)
	50	0.945	76	71	37.3	100.0	*infinity*	0.63
					(30.6-44.3)	(94.9-100.0)		(0.56-0.70)
L1	20	0.550	51	12	98.1	80.0	4.90	0.02
ER (n=15)					(89.7-100.0)	(51.9-95.7)	(1.78-13.50)	(0.00-0.17)
AD (n=52)	30	0.746	48	14	92.3	93.3	13.85	0.08
					(81.5-97.9)	(68.1-99.8)	(2.08-92.12)	(0.03-0.21)
	40	0.876	29	15	55.8	100.0	*infinity*	0.44
					(41.3-69.5)	(78.2-100.0)		(0.33-0.60)
	50	0.945	17	15	32.7	100.0	*infinity*	0.67
					(20.3-47.1)	(78.2-100.0)		(0.56-0.81)
L2	20	0.550	30	12	100.0	63.2	2.71	0.00
ER (n=19)					(88.4-100.0)	(38.4-83.7)	(1.51-4.89)	(0.00-)
AD (n=30)	30	0.746	23	18	76.7	94.7	14.57	0.25
					(57.7-90.1)	(74.0-99.9)	(2.14-99.15)	(0.13-0.48)
	40	0.876	17	19	56.7	100.0	*infinity*	0.43
					(37.4-74.5)	(82.4-100.0)		(0.29-0.65)
	50	0.945	13	19	43.3	100.0	*infinity*	0.57
					(25.5-62.6)	(82.4-100.0)		(0.41-0.78)
L3	20	0.550	110	25	95.7	69.4	3.13	0.06
ER (n=36)					(90.1-98.6)	(51.9-83.7)	(1.91-5.13)	(0.03-0.15)
AD (n=115)	30	0.746	96	32	83.5	88.9	7.51	0.19
					(75.4-89.7)	(73.9-96.9)	(2.97-18.99)	(0.12-0.29)
	40	0.876	67	36	58.3	100.0	*infinity*	0.42
					(48.7-67.4)	(90.3-100.0)		(0.34-0.52)
	50	0.945	43	36	37.4	100.0	*infinity*	0.63
					(28.5-46.9)	(90.3-100.0)		(0.54-0.72)

[a]The population-averaged probability from the mixed logistic regression (MLG) model.

Journal Pre-proof

Journal Pre-proof

Supplementary Table 8. True Positives (TPs), True Negatives (TNs), Sensitivity, Specificity, Positive Likelihood Ratio (PLR) and Negative Likelihood Ratio (NLR) of endoscopic Healing Index (EHI) in Distinguishing Active Disease (AD) vs Endoscopic Remission (ER) by Disease Location in Validation Cohort 2

CD Location	EHI Threshold	TPs (n)	TNs (n)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PLR (95% CI)	NLR (95% CI)
All	20	94	30	83.2	36.6	1.31	0.46
ER (n=82)				(75.0-89.6)	(26.2-48.0)	(1.09-1.58)	(0.28-0.76)
AD (n=113)	30	74	49	65.5	59.8	1.63	0.58
				(56.0-74.2)	(48.3-70.4)	(1.21-2.19)	(0.42-0.79)
	40	54	65	47.8	79.3	2.31	0.66
				(38.3-57.4)	(68.9-87.4)	(1.45-3.67)	(0.54-0.81)
	50	34	72	30.1	87.8	2.47	0.80
				(21.8-39.4)	(78.7-94.0)	(1.29-4.70)	(0.69-0.92)
L1	20	22	8	84.6	38.1	1.37	0.40
ER (n=21)				(65.1-95.6)	(18.1-61.6)	(0.94-1.99)	(0.14-1.16)
AD (n=26)	30	14	17	53.8	81.0	2.83	0.57
				(33.4-73.4)	(58.1-94.6)	(1.09-7.32)	(0.36-0.91)
	40	10	21	38.5	100.0	*infinity*	0.62
				(20.2-59.4)	(83.9-100.0)		(0.45-0.83)
	50	8	21	30.8	100.0	*infinity*	0.69
				(14.3-51.8)	(83.9-100.0)		(0.54-0.90)
L2	20	30	9	78.9	31.0	1.15	0.68
ER (n=29)				(62.7-90.4)	(15.3-50.8)	(0.85-1.54)	(0.30-1.54)
AD (n=38)	30	26	15	68.4	51.7	1.42	0.61
				(51.3-82.5)	(32.5-70.6)	(0.92-2.19)	(0.34-1.10)
	40	23	21	60.5	72.4	2.19	0.55
				(43.4-76.0)	(52.8-87.3)	(1.15-4.17)	(0.35-0.86)
	50	16	23	42.1	79.3	2.04	0.73
				(26.3-59.2)	(60.3-92.0)	(0.91-4.55)	(0.53-1.01)
L3	20	40	11	85.1	37.9	1.37	0.39
ER (n=29)				(71.7-93.8)	(20.7-57.7)	(1.01-1.87)	(0.17-0.90)
AD (n=47)	30	32	15	68.1	51.7	1.41	0.62
				(52.9-80.9)	(32.5-70.6)	(0.92-2.16)	(0.36-1.07)
	40	19	20	40.4	69.0	1.30	0.87
				(26.4-55.7)	(49.2-84.7)	(0.68-2.48)	(0.62-1.21)
	50	8	25	17.0	86.2	1.23	0.96
				(7.6-30.8)	(68.3-96.1)	(0.41-3.74)	(0.79-1.17)

Journal Pre-proof

Journal Pre-proof

What You Need to Know

Background and Context: We aimed to develop a test to identify patients in remission from Crohn's disease (CD), based on endoscopic analysis, and monitor CD activity based on serum levels of proteins.

New Findings: We developed an index to identify patients with CD in endoscopic remission based on blood levels of 13 proteins. The test (called the EHI) identified patients with resolution of mucosal inflammation, based on endoscopic analysis, with good overall accuracy.

Limitations: We analyzed data from 2 cohorts; and there was some variation in sensitivity and specificity of detection between the cohorts. We did not validate the test in pediatric patients.

Impact: The test might be used in practice to monitor mucosal inflammation, usually evaluated by endoscopy, in patients with CD.

Short Summary

We developed a serum test for monitoring Crohn's disease activity, based on endoscopic factors. We show that it accurately monitors Crohn's disease activity in all regions of the intestine.