Preclinical Serological Signatures are Associated With Complicated Crohn's Disease Phenotype at Diagnosis

Rok Seon Choung,[1,*] Francesca Petralia,[2,*] Joana Torres,[3,4] Ryan C. Ungaro,[3] Chad Porter,[5] Takahiro Sato,[6] Shannon Telesco,[6] Richard S. Strauss,[6] Scott Plevy,[7] Fred Princen,[8] Mark S. Riddle,[5,9] Joseph A. Murray,[1,§] and Jean Frederic Colombel[2,§]

[1]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota; [2]Department of Genetics and Genomics Sciences, Icahn School of Medicine at Mount Sinai, New York, New York; [3]Division of Gastroenterology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York; [4]Division of Gastroenterology, Hospital Beatriz Ângelo, Loures, Portugal; [5]Naval Medical Research Center, Silver Spring, Maryland; [6]Janssen R&D, Spring House, Pennsylvania; [7]Protagonist Therapeutics, Newark, California; [8]Prometheus Laboratories Inc, San Diego, California; and [9]Department of Internal Medicine, University of Nevada, Reno, School of Medicine, Reno, Nevada

BACKGROUND: **At diagnosis, up to one-third of patients with Crohn's disease (CD) have a complicated phenotype with stricturing (B2) or penetrating (B3) behavior or require early surgery. We evaluated protein biomarkers and antimicrobial antibodies in serum archived years before CD diagnosis to assess whether complicated diagnoses were associated with a specific serological signature.**

METHODS: **Prediagnosis serum was obtained from 201 patients with CD and 201 healthy controls. Samples were evaluated with a comprehensive panel of 1129 proteomic markers (SomaLogic) and antimicrobial antibodies. CD diagnosis and complications were defined by the International Classification of Diseases–Ninth Revision and Current Procedural Terminology codes. Cox regression models were utilized to assess the association between markers and the subsequent risk of being diagnosed with complicated CD. In addition, biological pathway and network analyses were performed.**

RESULTS: **Forty-seven CD subjects (24%) had a B2 (n = 36) or B3 (n = 9) phenotype or CD-related surgery (n = 2) at diagnosis. Subjects presenting with complicated CD at diagnosis had higher levels of antimicrobial antibodies six years before diagnosis as compared with those diagnosed with noncomplicated CD. Twenty-two protein biomarkers (reflecting inflammatory, fibrosis, and tissue protection markers) were found to be associated with complicated CD. Pathway analysis of the altered protein biomarkers identified higher activation of the innate immune system and complement or coagulation cascades up to six years before diagnosis in complicated CD.**

CONCLUSIONS: **Proteins and antimicrobial antibodies associated with dysregulated innate immunity, excessive adaptive response to microbial antigens, and fibrosis precede and predict a complicated phenotype at the time of diagnosis in CD patients.**

Keywords: Preclinical Period; Crohn's Disease; Complications; Serologic Biomarkers.

Crohn's disease (CD) is a chronic inflammatory condition of the gastrointestinal tract with peak onset between the ages of 15 and 30 years.[1,2] CD is a progressive disease that can lead to complications including strictures, fistulae, or abscesses that require aggressive medical and/or surgical treatment.[3,4] Up to one-third of patients present with a complicated phenotype at disease onset.[5]

Mounting evidence suggests that the diagnosis of CD is preceded by a lengthy asymptomatic preclinical period.[6–10] Gaining insight into this phase may allow a better understanding of the primary events that lead to its

*Authors share co-first authorship. §Authors share co-senior Authorship.

Abbreviations used in this paper: AIC, Akaike information criterion; ASCA, anti-*Saccharomyces cerevisiae* antibodies; CI, confidence interval; CRP, C-reactive protein; CD, Crohn's disease; DoDSR, Department of Defense Serum Repository; FAP, fibroblast activation protein α; HC, healthy control; HR, hazard ratio; IBD, inflammatory bowel disease; ICD-9, International Classification of Diseases–Ninth Revision; IL, interleukin; KM, Kaplan-Meier; LBP, lipopolysaccharide-binding protein; NTRK2, neurotrophic tyrosine kinase receptor type 2; SERPINA4, serpina family A member 4.

Most current article

1542-3565/$36.00
https://doi.org/10.1016/j.cgh.2023.01.033

development and offer potential strategies to predict and prevent the disease including its complications. The PREDICTS (Proteomic Evaluation and Discovery in an IBD Cohort of Tri-service Subjects) cohort was initiated to identify biomarkers and altered biological pathways that precede inflammatory bowel disease (IBD) onset using preclinical sera from the U.S. Department of Defense Serum Repository (DoDSR).[11] In our prior studies, we have observed that antibodies against several microbial antigens and specific proteins can be detected years before diagnosis and are highly predictive of CD.[9,10] We have also demonstrated that a higher prevalence of antimicrobial antibodies in preclinical CD serum samples has shown to be associated with complications at diagnosis.[9] Thus, we postulated that aberrant innate and adaptive immunity against gut microbiota that occurs in the preclinical stage of CD is further amplified and distinct in those patients with complications at diagnosis. Using the PREDICTS cohort, we evaluated antimicrobial antibodies and protein biomarkers in longitudinal serum samples before diagnosis to assess whether CD patients with complications at diagnosis had a specific preclinical antimicrobial or proteomic profile.

What You Need to Know

Background

Crohn's disease (CD) has a preclinical period, similar to other autoimmune diseases like type 1 diabetes, rheumatoid arthritis, and systemic lupus erythematosus. However, it is unclear whether early pathophysiologic specific changes are found in preclinical serum samples of patients with complicated CD.

Findings

A specific preclinical serological signature, including biomarkers of innate and adaptive immunity, fibrosis, tissue damage, and amplified antibody response to commensal microorganisms, is strongly associated with complicated (stricturing or penetrating) behavior at diagnosis.

Implications for patient care

The findings of the current study suggest that these specific preclinical signatures could predict the development of complications even many years before diagnosis, which could result in improving prevention strategies and uncovering pathways important in the development of disease progression.

Materials and Methods

Study Design and Study Population

We conducted a nested case-control study using the previously described PREDICTS study.[9–11] Briefly, patients with an incident diagnosis of CD were identified between 1998 and 2013 in the Department of Defense Medical Surveillance System.[11,12] Linked serum samples from each subject were obtained through the DoDSR.[11,12]

Incident CD was defined based on procedural and International Classification of Diseases–Ninth Revision (ICD-9) codes. The date of CD diagnosis was based on the first ICD-9 code for CD. For each subject, 4 serum samples were obtained from the DoDSR: sample A was the closest sample available to the date of CD diagnosis, and sample D was the earliest serum sample available in the repository before clinical diagnosis; samples B and C were approximately two and four years before diagnosis, respectively. For sample D, the SomaLogic proteomic panel and antimicrobial antibodies were tested for 201 CD, while for sample B and sample C, the number of samples was 116 and 166, respectively. For this reason, the sample B and sample C groups were aggregated into one group. Samples of healthy control (HC) subjects were also obtained from the DoDSR. As previously described,[10] control subjects were matched on age, sex, race, and timing of the diagnostic sample and were required to have no medical encounter with evidence of IBD, rheumatoid arthritis, celiac disease, or colorectal cancer (based on ICD-9 codes). From each subject, 3–4 serum samples were retrieved. Sample A from HC subjects was matched to sample A from IBD cases based on the year of collection ($\pm$1 year), whose serum samples were available and stored from the 3-preceding biennial (e.g., every two years) HIV test.

Phenotype Classification

Disease phenotype (ie, behavior) was categorized according to the Montreal classification.[13,14] Complicated CD was defined by the presence of penetrating (B3 behavior), stricturing (B2 behavior), or surgical history of intestinal resection using ICD-9 and Current Procedural Terminology codes from the time of diagnosis (index ICD-9 code). Detailed information for phenotype classification is provided in Supplementary Table 1.

Serum Testing

To evaluate protein abundance, serum was tested using the SomaLogic (Boulder, CO) assay, a multiplex platform profiling 1129 protein biomarkers, representing a range of biological functions including innate immune response and inflammatory signals. Samples were also tested for a panel of antimicrobial antibodies, including anti-*Saccharomyces cerevisiae* antibodies (anti-ASCA) IgA and IgG, anti-CBir1, anti-OmpC (anti-outer membrane protein C precursor), anti-Flagellin 2, and anti-Flagellin X. These antibodies were measured by a standardized enzyme-linked immunosorbent assay using a Freedom EVO 200 liquid-handling robot (Tecan) at Prometheus Laboratories (San Diego, CA).[15]

Abbreviations of serologic and protein biomarkers are in the Supplementary Table 2.

Statistical Analyses

The association between each marker (SomaLogic proteomic biomarker and serum antibodies against microbiota) and complication was assessed via Cox regression after adjusting by age and sex. *P* values were adjusted for multiple comparisons via Benjamini-Hochberg[16] and markers passing a false discovery rate of 10% were reported as significant. Cox regression models were estimated for different times before diagnosis (ie, 2–4 and 6 years before diagnosis). We repeated this analysis adjusting by disease location in the Cox regression model considering a subset of 167 patients for which disease location was measured. We also performed differential analysis via the Wilcoxon rank sum test between patients with ileal involvement and colonic involvement (L1/L3 vs L2). *P* values were adjusted for multiple comparisons via Benjamini-Hochberg[16] adjustment, and only markers with adjusted *P* value <10% were reported as significant. To visualize the association between each marker abundance and complication, Kaplan-Meier (KM) survival curves were utilized. To display the association of a set of markers and time of complication, the estimated mean of the Cox regression model was utilized in order to stratify patients into high- and low-risk groups and derive KM curves. Different multivariate Cox regression models were compared using the Akaike information criterion (AIC).[17] This index is used to determine how well the model fits the data while accounting for the total number of parameters in the model. Although models having more markers result in a better fit, they are usually less suitable to predict other datasets. For this reason, finding a good balance between model fit and parsimony is essential to select the best model. One of the ways to compare the goodness of fit among models with the different number of parameters is the AIC, which is a function of the log-likelihood of the estimated model and the number of parameters utilized. The model with the lowest AIC is the preferred model. The log-likelihood of the Cox regression was computed via the logLik function available in the Survival R package.[18,19]

Biological pathways enriched in the set of proteins associated with complication were identified at each time point via Fisher's exact test. For the analysis, pathways are from the KEGG (Kyoto Encyclopedia of Genes and Genomes)[20] and Reactome databases.[21] were utilized. For each patient, pathway scores were computed as the average abundance of protein mapping to each pathway after the *z* score across patients (mean of 0 and SD of 1). Besides pathway analysis, we performed coexpression network analysis to identify the association across proteomic markers and antimicrobial antibodies for patients with CD and HC subjects using joint random forest.[22] Permutation-based techniques were used to find association significant at a 10% false discovery rate.[22] Network modules based on the CD network were identified based on the cluster-edge-betweenness function available in the iGraph R package.[23] All analyses were performed by using R statistical software (version 3.6.3; R Foundation for Statistical Computing, Vienna, Austria).

Results

A total of 201 patients with CD and 201 HC subjects were included in the study. Among CD patients, 23% presented with complications at diagnosis with 36 B2, 9 B3, and 2 CD-related surgery (1 colectomy and 1 small bowel resection). The detailed study population characteristics are provided in Supplementary Table 3.

Antimicrobial Antibody Markers Associated With Complicated CD at Diagnosis

The mean concentration of all antimicrobial antibodies (except for OmpC) at all serum sample time points before diagnosis was significantly higher in those diagnosed with complicated CD than in those diagnosed with noncomplicated CD or HC (Figure 1*A* and Supplementary Figure 1). High antimicrobial antibodies (>75th percentile) such as ASCA IgA were associated with an increased risk of developing complications (at 2–4 years before diagnosis: hazard ratio [HR], 1.33; 95% confidence interval [CI], 1.13–1.55; and at six years before diagnosis: HR, 1.30; 95% CI, 1.10–1.51) as compared with those with a low abundance of antimicrobial antibodies (<25th percentile) (Figure 1*B*). Antimicrobial antibodies were significantly higher in complicated cases, compared with noncomplicated cases or HC subjects (Figure 1*C*).

Proteomic Biomarkers Associated With Complicated CD at Diagnosis

Overall, after adjusting for age and sex, 38 protein biomarkers (measured six or 2–4 years before diagnosis) were positively associated with the risk of developing CD complications, while 26 protein biomarkers were negatively associated with the risk of developing complications (adjusted *P* value <10%) (Figure 2). Among these 38 protein biomarkers with a positive association, 17 were found significant at different times before diagnosis (ie, six years and 2–4 years before diagnosis) (Figure 2*A*). Conversely, among the 26 negatively associated protein biomarkers, five biomarkers were found significantly different times before diagnosis (ie, six years and 2–4 years before diagnosis) (Figure 2*A*). An additional analysis was performed adjusting for disease location (L1/L3 vs L2) in addition to sex and age, considering a subset of 167 CD patients (ie, L1/L3: 128 samples, L2 disease location: 39 samples) for which disease location was available. After adjusting by age, sex, and disease location, 18 biomarkers (of 38 protein biomarkers) were significantly associated with complicated CD (Supplementary Table 4). Association analysis results can be found in Supplementary Tables 4

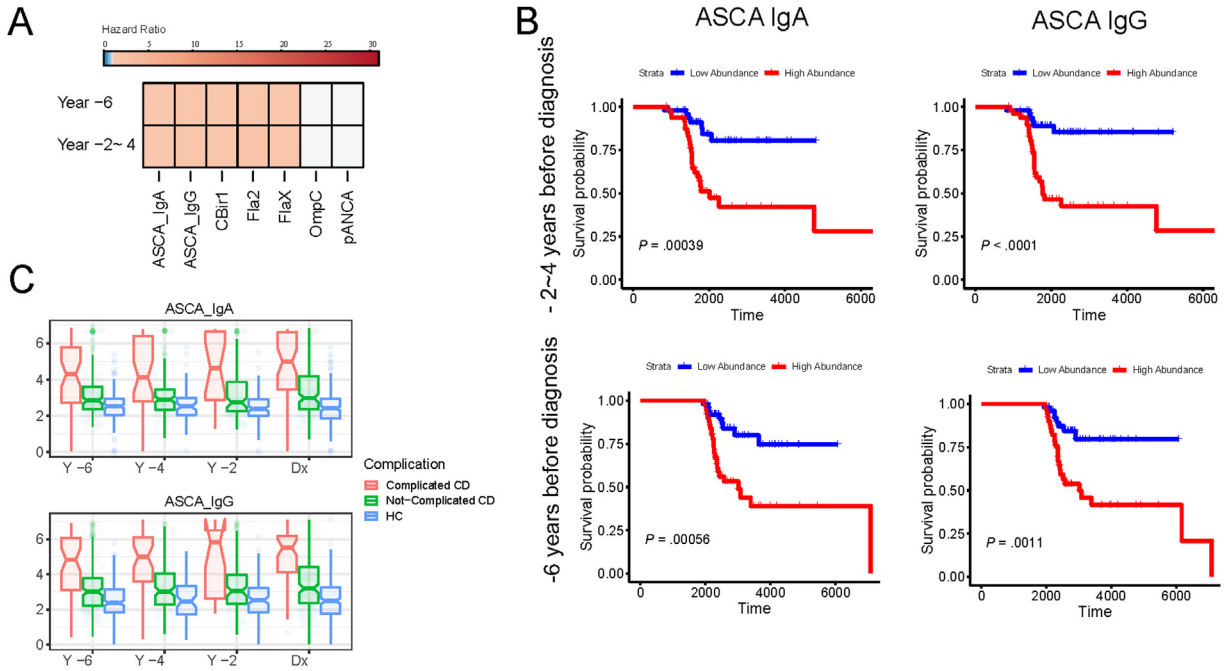


Figure 1. (*A*) Heatmap of HRs for serologic markers significantly associated with complications for different years before diagnosis. (*B*) KM curve of ASCA IgA and ASCA IgG for years 2–4 and year 6 from diagnosis. (*C*) Boxplot of marker abundance for complicated CD, uncomplicated CD, and HC subjects for different years before diagnosis.

and 5. The 95% CIs of estimated coefficients from Cox proportional hazards regression for protein biomarkers, which remained significantly associated with complications, are provided in Supplementary Figures 2 and 3. In addition, we also performed differential analysis between L1/L3 and L2 to find markers associated with disease location. No markers were found significantly associated with disease location six years before diagnosis (Adjusted *P* value <10%). At 2–4 years before diagnosis, only six protein biomarkers (complement factor I, C-reactive protein [CRP], C9, complement factor B, serum amyloid P, stromal cell-derived factor 1) were significantly different between patients with ileal involvement and colonic involvement (L1/L3 vs L2).

Figure 2*B* shows the KM curves of the risk of complicated CD with high and low abundance of representative protein biomarkers (CRP, lipopolysaccharide-binding protein [LBP], and serpina family A member 4 [SERPINA4]). CRP abundance was associated with an increased risk of developing complications at 2–4 years and 6 years before diagnosis (HR, 2.15; 95% CI, 1.50–3.08; and HR, 2.04; 95% CI, 1.46–2.88, respectively), while the abundance of SERPINA4 was associated with decreased risk of complications (HR, 0.16; 95% CI, 0.08–0.35; and HR, 0.16; 95% CI, 0.08–0.34, respectively). All other KM curves of each protein biomarker showing the cumulative incidence of complications are shown in Supplementary Figure 4. Figure 2*C* shows the box plots of CRP, LBP, and SERPINA4 as examples of protein biomarkers, significantly distinct in patients with complications at diagnosis, compared with patients without complications or HC subjects. Additional box plots of each protein biomarker are presented in Supplementary Figure 5.

Combination of Protein Biomarkers and Antimicrobial Antibodies Associated With Complicated CD at Diagnosis

In this section, we present results based on multivariate analysis of antimicrobial antibodies and 22 protein biomarkers that were found to be associated with the development of complications at diagnosis for all time points. Figure 3 shows the KM curves based on a multivariate Cox regression model considering (1) serum antimicrobial antibodies, (2) 22 protein biomarkers, and (3) the integration of the two sets of markers. As shown in Figure 3, proteomic markers result in better separation between KM curves than antimicrobial antibodies. A comparison of the three models based on the AIC revealed that the best explanatory model of CD complications was the model based on protein biomarkers alone (Supplementary Table 6). Note that although the serologic model is the most parsimonious, it does not result in the lowest AIC because it fails to adequately model the data.

Prediagnostic Pathways and Network-Based Approach

Complement and coagulation cascades and innate immune system pathways were significantly enriched in the set of proteins associated with complicated CD at different time points before diagnosis (Figure 4*A*). Figure 4*B* shows the KM curves for the group of samples with high-pathway activity and low-pathway activity (*P* values from log-rank test = .0066 and .014 at year 6

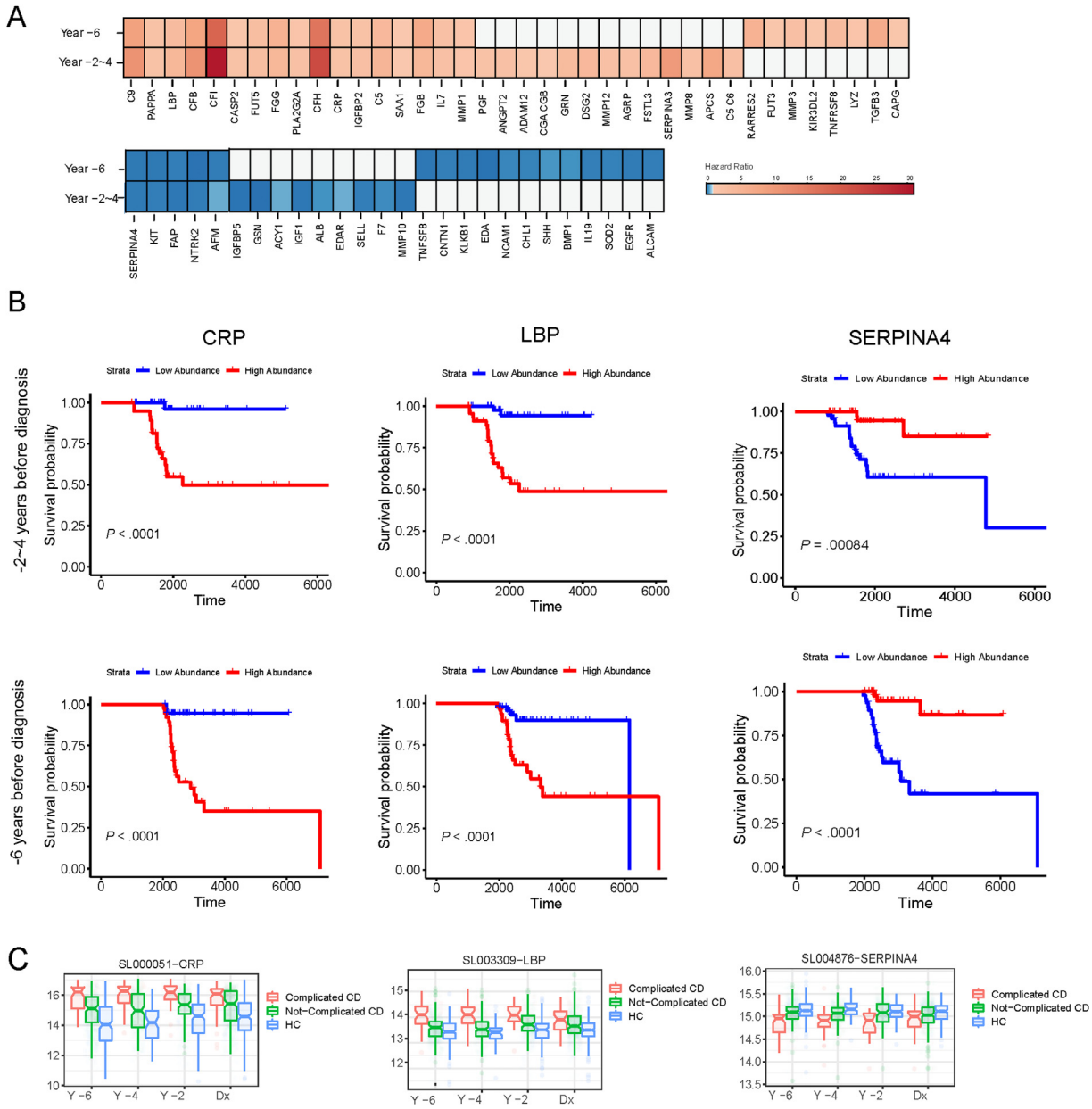


Figure 2. (*A*) Heatmap of HRs from Cox regression model for proteomic biomarkers significantly associated with CD complications (adjusted *P* value <10%) at different years before diagnosis. (*B*) KM curve of CRP, LBP, and complement factor B for years 2–4 and year 6 from diagnosis. (*C*) Boxplot of marker abundance for complicated CD, uncomplicated CD, and HC subjects for different years before diagnosis. Among 38 protein biomarkers positively associated with complications (adjusted *P* value <10%), 17 biomarkers were significantly higher in serum samples prediagnosis (both at 2–4 years and 6 years before diagnosis) in patients with complicated CD (panel *A*). Among 26 protein biomarkers negatively associated with complications, 5 biomarkers were significantly and consistently lower before diagnosis in patients diagnosed with complicated CD, compared with those diagnosed with noncomplicated CD. KM curves or the box plots (panels *B* or *C*, respectively) on the risk of complicated CD among CD cases with high and low abundance of CRP, LBP, and SERPINA4. Dx, diagnosis; Y, year.

before diagnosis for innate immune response and complement and coagulation cascade, respectively). Next, we employed a network-based approach to study the coexpression pattern across different proteins in CD patients and HC subjects considering samples collected at the furthest year from diagnosis. For this network analysis, we integrated both SomaLogic and antimicrobial antibodies to identify associations between the 2 sets of markers. Figure 4*C* shows the Pearson's correlation of markers in this network module for CD patients and HC subjects. As shown, this network module contains proteins that are positively associated with the development of complications such as complement factor B, C9, and CRP, and that are negatively associated with the development of complications such as SERPINA4. ASCA IgA and IgG were strongly correlated with protein biomarkers in patients with CD. This correlation structure was not captured in the HC subjects (Figure 4*C*).

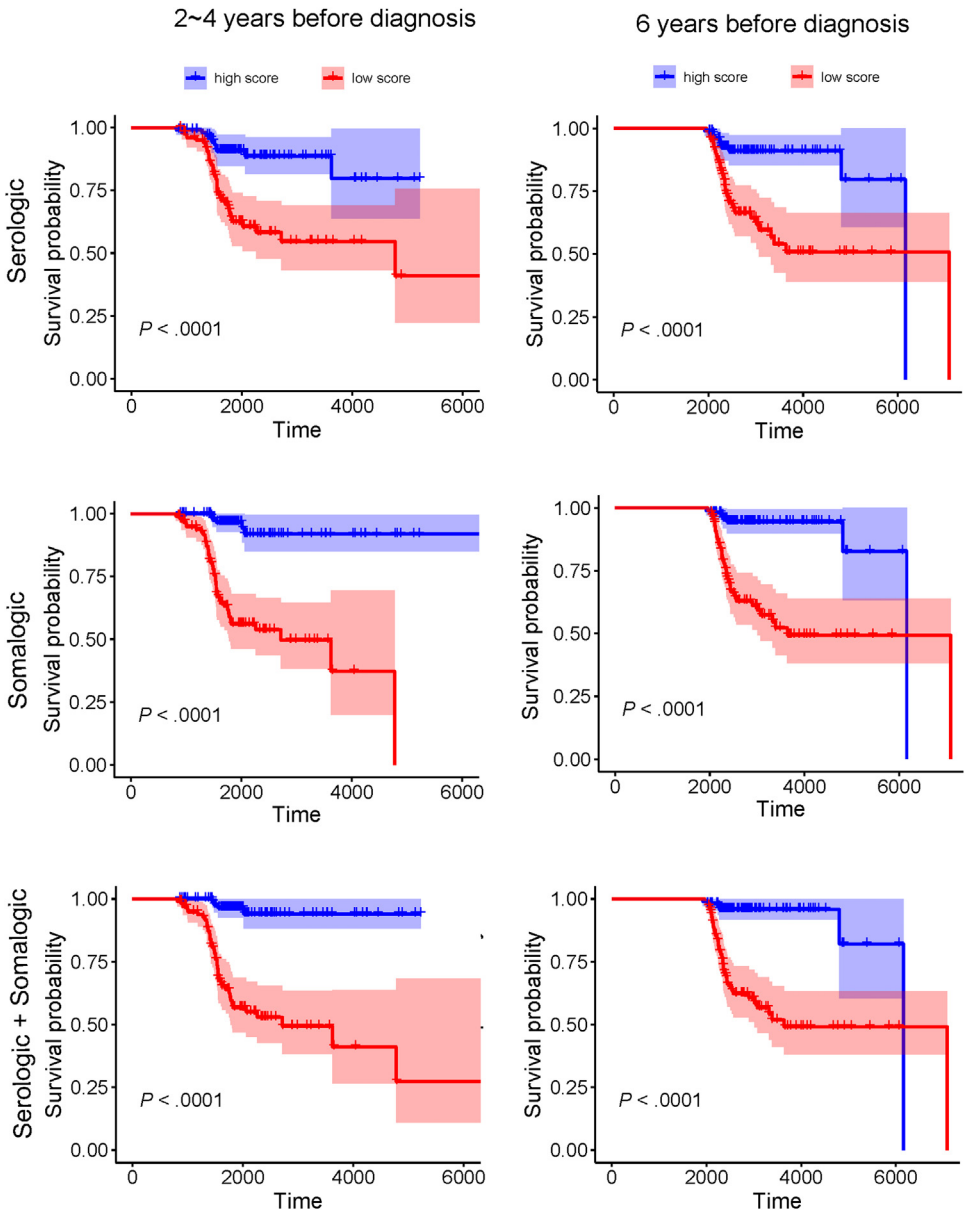


Figure 3. KM curves and corresponding 95% CIs for the antimicrobial antibody-based model, SomaLogic protein biomarker-based model, and the model integrating antimicrobial antibodies and protein biomarkers. For each model, the estimated mean from Cox regression was utilized to stratify patients into high- and low-risk individuals considering the median of the fitted mean as threshold.

Discussion

In this study, we demonstrate that complicated CD at diagnosis is associated with a distinct serologic profile compared with uncomplicated CD years before diagnosis. This profile is characterized by increased levels of ASCA IgA and IgG; anti-Flagellin antibodies; a high abundance of protein biomarkers associated with innate immunity, fibrosis, and adaptive immunity; and a low abundance of protein biomarkers related to protection against tissue damage or fibrosis (SERPINA4, fibroblast activation protein α [FAP], KIT, neurotrophic tyrosine kinase receptor type 2 [NTRK2], and afamin). Moreover, using network-based analysis, we found a significant correlation between ASCA IgA/IgG with protein biomarkers related to innate immunity and lack of tissue-protective factors.

Consistent with previous findings, we confirmed the presence of higher levels of antimicrobial antibodies years before diagnosis in patients with complicated CD at diagnosis as compared with an uncomplicated diagnosis.[9] Alexander et al[13] also showed that patients with CD displayed a strong adaptive immune response to flagellin antigens, with a subset of CD patients having multiflagellin reactivity, which was related to a high frequency of CD complications. Together with these findings, our data suggest a preclinical aberrant adaptive immune response against gut microbiota many years before CD diagnosis, which is amplified in patients with a complicated phenotype at diagnosis.

Using the SomaLogic platform, we identified a unique set of 22 protein biomarkers associated with complicated CD at diagnosis. Among these 22 protein biomarkers associated with disease complication, 15 protein biomarkers were not found associated with disease onset in

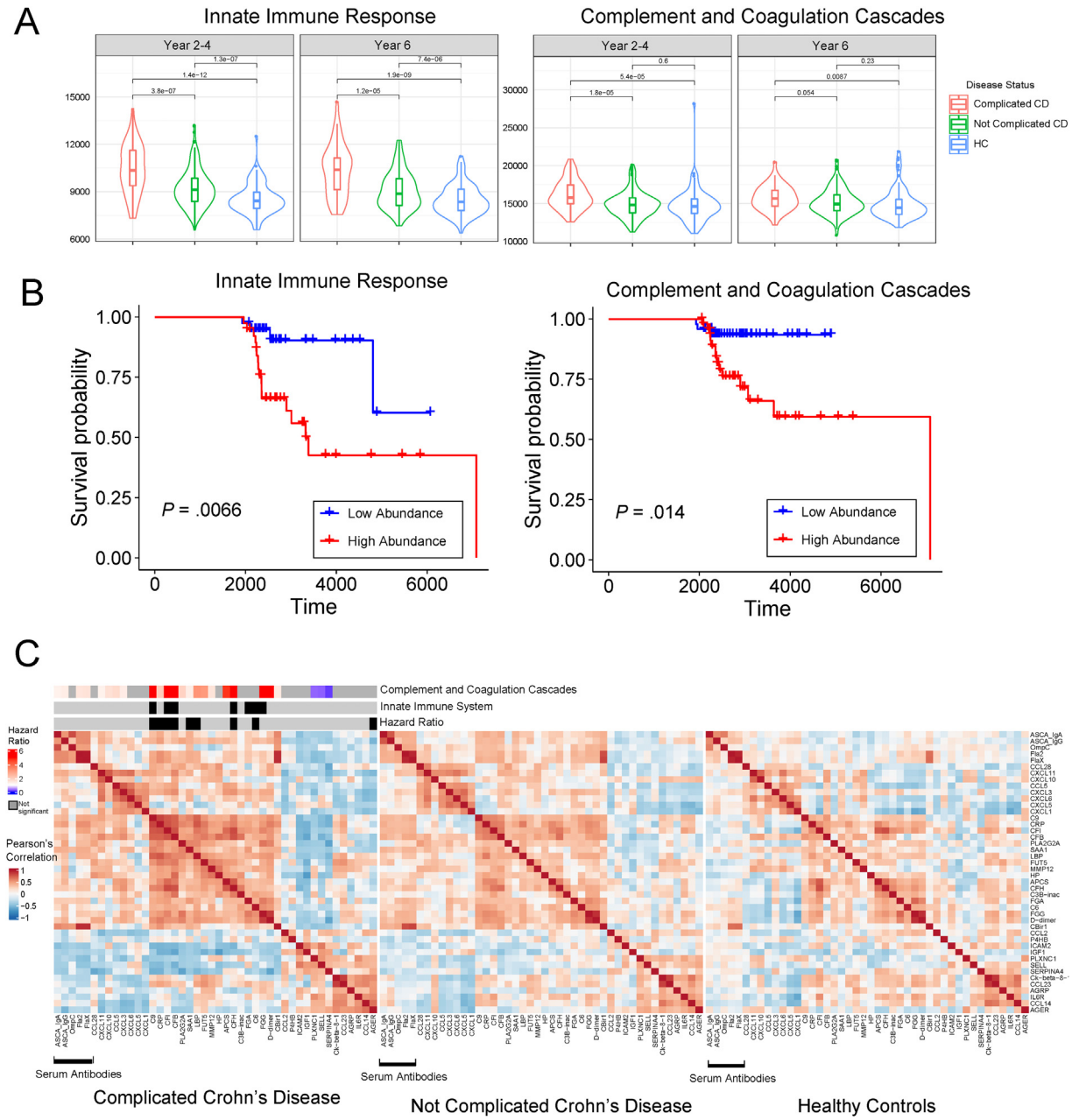


Figure 4. (*A*) Pathway score for the innate immune response and complement and coagulation cascades pathways stratified by different groups of patients corresponding to complicated CD, noncomplicated CD, and HC subjects. (*B*) KM curves for samples at the furthest year of diagnosis based on the score of the complement and coagulation pathway and innate immune response. (*C*) Heatmap of Pearson's correlation of proteins and serum antibodies contained in the network cluster identified based on coexpression network analysis of CD patients. Pearson's correlation of protein and serum antibodies markers is shown for different groups of patients corresponding to complicated CD, noncomplicated CD, and HC subjects.

our previous study (Supplementary Figure 6).[10] Additional analysis confirmed that the association between proteomic markers with CD complications was, for the most part, not confounded with disease location. These biomarkers have biological plausibility. Twelve of the 22 biomarkers, like CRP, LBP, and complement proteins, are associated with the innate immune response and inflammation. Another two biomarkers (IL-7 and fucosyltransferase 5) were also significantly increased in complicated CD. IL-7 plays a central role in B and T cell development and modulates T cell homeostasis.[24] Few studies have shown the overexpression of the IL-7 or IL-7 receptor signaling pathway in IBD patients with an aggressive course.[25,26] Fucosyltransferase 5 is involved in host-commensal interactions with certain bacteria, causing the upregulation of fucosylation in the intestine.[27] Among the 22 protein biomarkers, six were related to fibrosis (3 markers [matrix metalloproteinase-1, pregnancy-associated plasma protein A, insulin-like growth factor binding protein 2] increased and another three markers [KIT, FAP, and NTRK2] decreased). Matrix metalloproteinase-1 has shown to be upregulated in the

areas of intestinal stenosis in patients with CD.[28] Pregnancy-associated plasma protein A is a metalloproteinase, working as an interactive cellular mechanism promoting pulmonary fibrosis.[29] In addition, insulin-like growth factor binding protein 2, which is a transport protein for insulin-like growth factors, is also increased in patients with pulmonary fibrosis or systemic sclerosis.[30] Regarding FAP, Corsi et al[31] also showed that circulating FAP concentration was reduced in patients with IBD, especially those undergoing surgery. The decreased levels of circulating FAP have also been shown to be related to organ damage and fibrosis in other diseases.[32,33] Serum protein KIT is significantly reduced in patients with hypertrophic cardiomyopathy, which is another fibrosis-related condition.[34] NTRK2 is a receptor brain-derived neurotrophic factor, and brain-derived neurotrophic factor/TrKB axis activation has shown an association with lung fibrosis.[35]

Two biomarkers (SERPINA4 and afamin), known to be linked to protection from tissue injury or anti-inflammatory properties,[36–38] were significantly decreased in the preclinical serum of patients with complications at diagnosis, compared with the uncomplicated CD or HC subjects. SERPINA4 is a unique protein, showing protective roles against tissue damage by preventing apoptosis, oxidative stress, and inflammation in several conditions like sepsis and cardiovascular diseases.[36,39] Stadnicki et al[40] showed that intestinal tissue SERPINA4 was significantly decreased in the inflamed intestine in patients with active IBD. Afamin, a vitamin E-binding protein, also plays a protective role in conditions of oxidative and inflammatory stress.[37,38]

Finally, we identified two pathways linked to innate immunity and coagulation and complement cascade to be significantly upregulated in the preclinical serum samples of patients with complications at diagnosis, compared with the noncomplicated CD group. The network analysis showed the clustering of antimicrobial antibodies and protein biomarkers together in serum samples collected six years before the diagnosis of CD. Overall, our findings may suggest that perturbations in innate immune response against gut microbiota may induce the overproduction of inflammatory proteins and stimulate adaptive immunity, leading to the production of antimicrobial antibodies in complicated phenotypes. Figure 5 summarizes the potential mechanism of complications in the preclinical stage of the disease.

Our study has several strengths. We used preclinical samples from a well-characterized cohort, which allowed us to explore the early changes of the innate and adaptive immune response against microbiota and to discover protein biomarkers of the early host response. The unique availability of preclinical samples collected at multiple time points allowed us to examine the sequence of immunological changes and protein biomarkers that occurred before diagnosis. We also evaluated a wide array of protein biomarkers, utilizing a novel proteomic platform, and applied novel rigorous statistical approaches, which allowed us to discover the potential biomarkers and biologic pathways for the complicated phenotypes even before diagnosis.

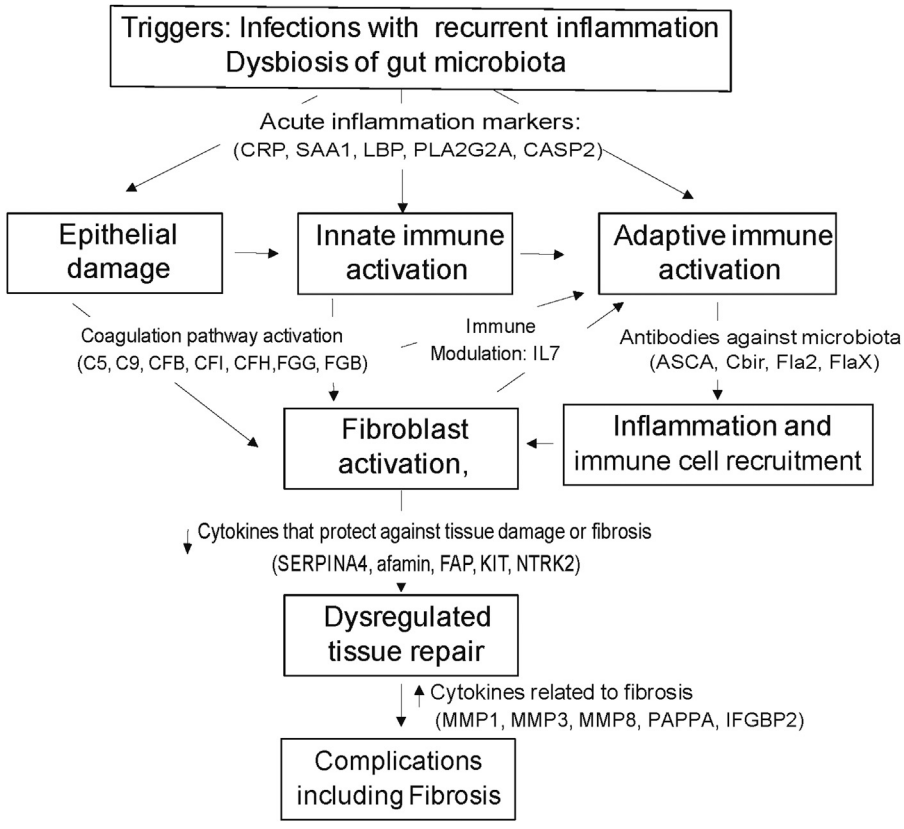


Figure 5. Overview of immune dysregulation and tissue destruction in CD. After infections or dysbiosis of gut microbiota trigger disease development, induced acute and chronic inflammation could cause epithelial damage, which consequently activates the innate immune response and complement system. Then, several cytokines are released by intestinal epithelium and innate immune cells, which subsequently activate the adaptive immune response against gut microbiota. Lack of protective cytokines against tissue damage enhances dysregulated tissue repair. Finally, chronic inflammatory reactions due to persistent interactions between host and environment could result in developing fibrosis in CD.

Our study also has limitations. First, the study's findings may not be generalizable because the study population was mainly White and male. Second, the cases and complicated phenotypes were identified based on the diagnostic and procedural codes and were not validated using medical records review. However, the case ascertainment for IBD, utilizing the diagnostic codes, has shown a high level of accuracy in similar populations.[11] Third, it is possible that our findings are confounded by a diagnostic delay in CD. However, this is unlikely because significant changes in biomarkers were observed six or more years before diagnosis.[5] Finally, data on other potentially important risk factors for complications (eg, smoking history, genetics) were not available in this cohort.

In conclusion, a complicated CD phenotype at diagnosis is associated with a specific serological profile years before diagnosis including biomarkers of innate and adaptive immunity, fibrosis, tissue damage, and amplified antibody response to commensal microorganisms. Altogether, the hypothesis can be proposed that the combination of increasing levels of inflammatory cytokines, loss of anti-inflammatory proteins, and production of antimicrobial antibodies could accelerate and magnify tissue destruction and fibrosis driving complications at diagnosis in CD. These data support the concept that complicated CD may not always be the result of the progression of an uncontrolled inflammatory disease but may also be the consequence of a unique pathophysiological process. The serological signature that we identified could help to further select subjects at risk of developing complicated CD who could be preferential candidates for preventative strategies.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Clinical Gastroenterology and Hepatology* at www.cghjournal.org, and at https://doi.org/10.1016/j.cgh.2023.01.033.

References

1. Loftus EV Jr, Silverstein MD, Sandborn WJ, et al. Crohn's disease in Olmsted County, Minnesota, 1940-1993: incidence, prevalence, and survival. Gastroenterology 1998;114:1161–1168.
2. Aniwan S, Park SH, Loftus EV Jr. Epidemiology, natural history, and risk stratification of Crohn's disease. Gastroenterol Clin North Am 2017;46:463–480.
3. 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;105:289–297.
4. Thia KT, Sandborn WJ, Harmsen WS, et al. Risk factors associated with progression to intestinal complications of Crohn's disease in a population-based cohort. Gastroenterology 2010;139:1147–1155.
5. Rodriguez-Lago I, Zabana Y, Barreiro-de Acosta M. Diagnosis and natural history of preclinical and early inflammatory bowel disease. Ann Gastroenterol 2020;33:443–452.
6. Israeli E, Grotto I, Gilburd B, et al. Anti-Saccharomyces cerevisiae and antineutrophil cytoplasmic antibodies as predictors of inflammatory bowel disease. Gut 2005;54:1232–1236.
7. van Schaik FD, Oldenburg B, Hart AR, et al. Serological markers predict inflammatory bowel disease years before the diagnosis. Gut 2013;62:683–688.
8. Lochhead P, Khalili H, Ananthakrishnan AN, et al. Association between circulating levels of C-reactive protein and interleukin-6 and risk of inflammatory bowel disease. Clin Gastroenterol Hepatol 2016;14:818–824.e6.
9. Choung RS, Princen F, Stockfisch TP, et al. Serologic microbial associated markers can predict Crohn's disease behaviour years before disease diagnosis. Aliment Pharmacol Ther 2016;43:1300–1310.
10. Torres J, Petralia F, Sato T, et al. Serum biomarkers identify patients who will develop inflammatory bowel diseases up to 5 years before diagnosis. Gastroenterology 2020;159:96–104.
11. Porter CK, Riddle MS, Gutierrez RL, et al. Cohort profile of the PRoteomic Evaluation and Discovery in an IBD Cohort of Tri-service Subjects (PREDICTS) study: rationale, organization, design, and baseline characteristics. Contemp Clin Trials Commun 2019;14:100345.
12. Rubertone MV, Brundage JF. The Defense Medical Surveillance System and the Department of Defense serum repository: glimpses of the future of public health surveillance. Am J Public Health 2002;92:1900–1904.
13. Alexander KL, Zhao Q, Reif M, et al. Human microbiota flagellins drive adaptive immune responses in Crohn's disease. Gastroenterology 2021;161:522–535.e6.
14. Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. Can J Gastroenterol 2005;19:5A–36A.
15. Landers CJ, Cohavy O, Misra R, et al. Selected loss of tolerance evidenced by Crohn's disease-associated immune responses to auto- and microbial antigens. Gastroenterology 2002;123:689–699.
16. Ferreira JA, Zwinderman AH. On the Benjamini-Hochberg method. Ann Stat 2006;34:1827–1849.
17. Bozdogan H. Akaike's information criterion and recent developments in information complexity. J Math Psychol 2000;44:62–91.
18. Therneau TM. A package for survival analysis in R. R package version 3.5-0. Available at: https://CRAN.R-project.org/package=survival. Accessed January 26, 2023.
19. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. New York, NY: Springer, 2000.
20. Kanehisa M. The KEGG database. Novartis Found Symp 2002;247:91–101, discussion 101–103, 119–128, 244–252.
21. Fabregat A, Jupe S, Matthews L, et al. The Reactome Pathway Knowledgebase. Nucleic Acids Res 2018;46:D649–D655.
22. Petralia F, Song WM, Tu Z, et al. New method for joint network analysis reveals common and different coexpression patterns among genes and proteins in breast cancer. J Proteome Res 2016;15:743–754.
23. Csárdi G, Nepusz T. The igraph software package for complex network research, 2006. InterJournal. Complex Systems. Available at: https://igraph.org. Accessed March 30, 2023.
24. Dooms H. Interleukin-7: Fuel for the autoimmune attack. J Autoimmun 2013;45:40–48.

25. Belarif L, Danger R, Kermarrec L, et al. IL-7 receptor influences anti-TNF responsiveness and T cell gut homing in inflammatory bowel disease. J Clin Invest 2019;129:1910–1925.
26. Lee JC, Lyons PA, McKinney EF, et al. Gene expression profiling of CD8+ T cells predicts prognosis in patients with Crohn disease and ulcerative colitis. J Clin Invest 2011;121:4170–4179.
27. Li J, Hsu HC, Mountz JD, et al. Unmasking fucosylation: from cell adhesion to immune system regulation and diseases. Cell Chem Biol 2018;25:499–512.
28. Warnaar N, Hofker HS, Maathuis MH, et al. Matrix metalloproteinases as profibrotic factors in terminal ileum in Crohn's disease. Inflamm Bowel Dis 2006;12:863–869.
29. Bale LK, Schafer MJ, Atkinson EJ, et al. Pregnancy-associated plasma protein-A (PAPP-A) is a key component of an interactive cellular mechanism promoting pulmonary fibrosis. J Cell Physiol 2022;237:2220–2229.
30. Guiot J, Bondue B, Henket M, et al. Raised serum levels of IGFBP-1 and IGFBP-2 in idiopathic pulmonary fibrosis. BMC Pulm Med 2016;16:86.
31. Corsi F, Sorrentino L, Albasini S, et al. Circulating fibroblast activation protein as potential biomarker in patients with inflammatory bowel disease. Front Med (Lausanne) 2021;8:725726.
32. Uitte de Willige S, Keane FM, Bowen DG, et al. Circulating fibroblast activation protein activity and antigen levels correlate strongly when measured in liver disease and coronary heart disease. PLoS One 2017;12:e0178987.
33. Tillmanns J, Fraccarollo D, Galuppo P, et al. Changes in concentrations of circulating fibroblast activation protein alpha are associated with myocardial damage in patients with acute ST-elevation MI. Int J Cardiol 2017;232:155–159.
34. Sonnenschein K, Fiedler J, de Gonzalo-Calvo D, et al. Blood-based protein profiling identifies serum protein c-KIT as a novel biomarker for hypertrophic cardiomyopathy. Sci Rep 2021;11:1755.
35. Cherubini E, Mariotta S, Scozzi D, et al. BDNF/TrkB axis activation promotes epithelial-mesenchymal transition in idiopathic pulmonary fibrosis. J Transl Med 2017;15:196.
36. Lin WC, Chen CW, Huang YW, et al. Kallistatin protects against sepsis-related acute lung injury via inhibiting inflammation and apoptosis. Sci Rep 2015;5:12463.
37. Koninger A, Mathan A, Mach P, et al. Is Afamin a novel biomarker for gestational diabetes mellitus? A pilot study. Reprod Biol Endocrinol 2018;16:30.
38. Polkowska A, Pasierowska IE, Paslawska M, et al. Assessment of serum concentrations of adropin, afamin, and neudesin in children with type 1 diabetes. Biomed Res Int 2019;2019:6128410.
39. Li PF, Bledsoe G, Yang ZR, et al. Human kallistatin administration reduces organ injury and improves survival in a mouse model of polymicrobial sepsis. Immunology 2014;142:216–226.
40. Stadnicki A, Mazurek U, Plewka D, et al. Intestinal tissue kallikrein-kallistatin profile in inflammatory bowel disease. Int Immunopharmacol 2003;3:939–944.

Correspondence
Address correspondence to: Joseph A. Murray, MD, Division of Gastroenterology and Hepatology, Mayo Clinic, 200 First Street SW, Rochester, MN 55905. e-mail: murray.joseph@mayo.edu.

Acknowledgments
This work was prepared as part of the official duties of U.S. government personnel and service members (Chad Porter and Mark Riddle). Title 17 U.S.C. §105 provides that "Copyright protection under this title is not available for any work of the United States Government." Title 17 U.S.C. §101 defines a U.S. government work as work prepared by a military service member or employee of the U.S. government as part of that person's official duties. The opinions and assertions expressed herein are those of the author(s) and do not necessarily reflect the official policy or position of the Uniformed Services University, U.S. Navy, or the Department of Defense. The study protocols were approved by the Naval Medical Research Center Institutional Review Board in compliance with all applicable federal regulations governing the protection of human subjects.

CRediT Authorship Contributions
Rok Seon Choung (Conceptualization: Equal; Data curation: Equal; Formal analysis: Equal; Writing – original draft: Equal; Writing – review & editing: Equal)
Joseph A. Murray, MD (Conceptualization: Equal; Formal analysis: Equal; Writing – original draft: Equal; Writing – review & editing: Equal)
Francesca Petralia (Conceptualization: Equal; Data curation: Equal; Formal analysis: Equal; Writing – original draft: Equal; Writing – review & editing: Equal)
Joana Torres (Conceptualization: Equal; Data curation: Equal; Formal analysis: Equal; Writing – review & editing: Equal)
Ryan Ungaro (Formal analysis: Equal; Writing – review & editing: Equal)
Chad Porter (Conceptualization: Equal; Data curation: Equal; Formal analysis: Equal; Writing – review & editing: Equal)
Takahiro Sato (Conceptualization: Equal; Data curation: Equal; Writing – review & editing: Equal)
Shannon Telesco (Conceptualization: Equal; Data curation: Equal; Writing – review & editing: Equal)
Richard Strauss (Conceptualization: Equal; Writing – review & editing: Equal)
Scott Plevy (Conceptualization: Equal; Writing – review & editing: Equal)
Fred Princen (Conceptualization: Equal; Data curation: Equal; Writing – review & editing: Equal)
Mark Riddle (Conceptualization: Equal; Writing – review & editing: Equal)
Jean-Frederic Colombel (Conceptualization: Equal; Formal analysis: Equal; Writing – original draft: Equal; Writing – review & editing: Equal)

Conflicts of interest
These authors disclose the following: Joana Torres has received speaker fees from Janssen, AbbVie, Galapagos, and Pfizer; and grant support from AbbVie and Janssen. Ryan C. Ungaro has served as a consultant for AbbVie, Bristol Myers Squibb, Janssen, Pfizer, and Takeda. Takahiro Sato, Shannon Telesco, and Richard Strauss are employees of Janssen Research and Development. Fred Princen is an employee of Prometheus. Scott Plevy is an employee of Protagonist Therapeutics, a past employee of Janssen, and a shareholder of Johnson & Johnson. Joseph A. Murray has received research grants from Nexpep/ImmusanT, the National Institutes of Health, Immunogenix, Takeda Pharmaceutical, Allakos, ProventionBio, the Oberkotter Foundation, and 9Meters, Inc; held a contract (to his institution) from Kanyos Bio (a wholly owned subsidiary of Anokion); received consultancy fees from Johnson and Johnson, Bristol Myers Squibb, Intrexon Corporation, Dren Bio, Neoleukin, Reistone Pharma, Immunic Therapeutics, Senda Biosciences, Brightseed Bio, Chugai Pharma, Alimentiv, Equillium, Ukko, and Vial Health Technologies; and received royalties from Torax Medical and Evelo. Jean-Frederic Colombel has received research grants from AbbVie, Janssen Pharmaceuticals, and Takeda; has received payment for lectures from AbbVie, Amgen, Allergan Inc, Ferring Pharmaceuticals, Shire, and Takeda; has received consulting fees from AbbVie, Amgen, Arena Pharmaceuticals, Boehringer Ingelheim, Celgene Corporation, Celltrion, Eli Lilly, Enterome, Ferring Pharmaceuticals, Genentech, Janssen Pharmaceuticals, Landos, Ipsen, Medimmune, Merck, Novartis, Pfizer, Shire, Takeda, and TiGenix; and owns stock options in Intestinal Biotech Development and Genfit. A patent application named "A panel of biomarkers for inflammatory bowel disease and uses thereof," has been filed between NRMC, Mount Sinai, Janssen Research and Development, and Prometheus (JBI5154USPSP). The remaining authors disclose no conflicts.

Funding
Support for this study was provided by Janssen Pharmaceuticals and Prometheus Laboratories under a Cooperative Research and Development Agreement entitled "Antimicrobial Antibodies as Predictors of Inflammatory Bowel Diseases" (CRADA number NMR-11-3920). Joseph A. Murray and Rok Seon Choung received a grant from the Crohn's and Colitis Foundation (BMRP Grant No. 342367) and Joana Torres received support from the Sandford J. Grossman Charitable Trust. Ryan C. Ungaro was supported by an National Institutes of Health K23 Career Development Award (5K23DK111995). Francesca Petralia and Jean-Frederic Colombel were supported by the Kenneth Rainin Foundation (Grant #20210021).

Statistical Analysis

For the Kaplan-Meier (KM) survival curves, the time to event was measured as the time when complications developed since the serum sample collection. For most patients, complications developed within one year from diagnosis (complication time; 25th quantile = at diagnosis, 50th quantile = 27 days after diagnosis, and 75th quantile = 340 days after diagnosis); while the last follow-up, time for the noncomplicated Crohn's disease (CD) was for most patients within 600 days after diagnosis (last follow-up time; 25th quantile = 227 days, 50th quantile = 585 days, and 75th quantile = 1227 days after diagnosis). To visualize KM curves, for each group of samples, CD cases were allocated to 2 groups: high (abundance of each biomarker greater than the 75th quantile) or low (abundance of each biomarker lower than the 25th quantile) abundance. The median time to complication from marker measurement was 1508 days (fifth quantile = 936 days and 95th quantile = 3350 days) and 2278 days (fifth quantile = 2019 days, 95th quantile = 4456 days) for samples B/C and D, respectively, while the last follow-up median time for noncomplicated CD was 1948 days (fifth quantile = 973 and 95th quantile = 4191 days) and 2750 days (fifth quantile = 2116 days and 95th quantile = 4906 days), respectively.

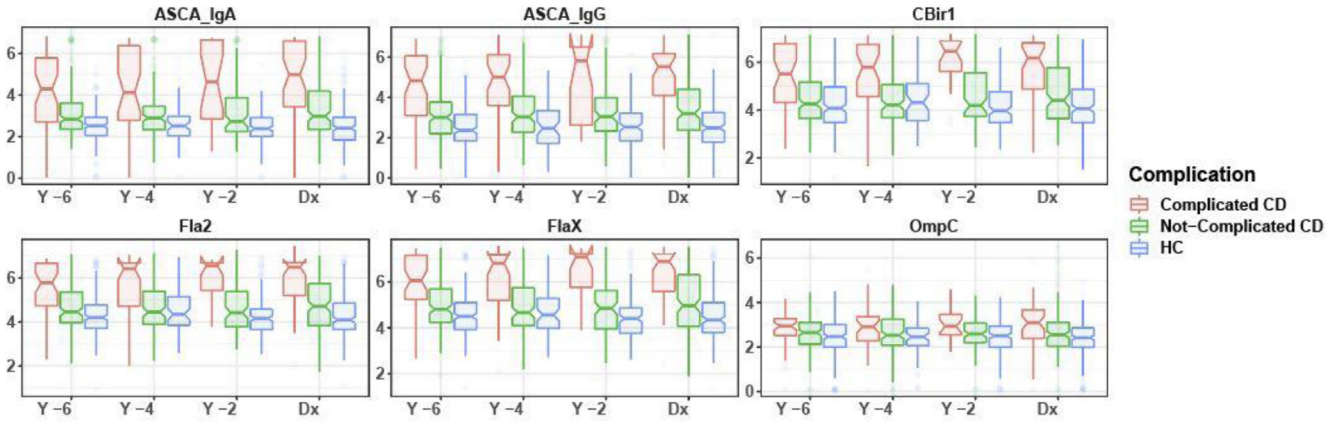
Furthermore, to visualize this association via KM analysis, for each sample, we measured the pathway activity by taking the average between proteins mapping to a particular pathway after standardizing each protein to a z score (mean of 0 and SD of 1). Samples were then divided into 2 groups, high (pathway activity greater than the 75th quantile) or low (pathway activity lower than the 25th quantile) activity, and KM curves were visualized for these 2 groups of samples.

Study Population Characteristics

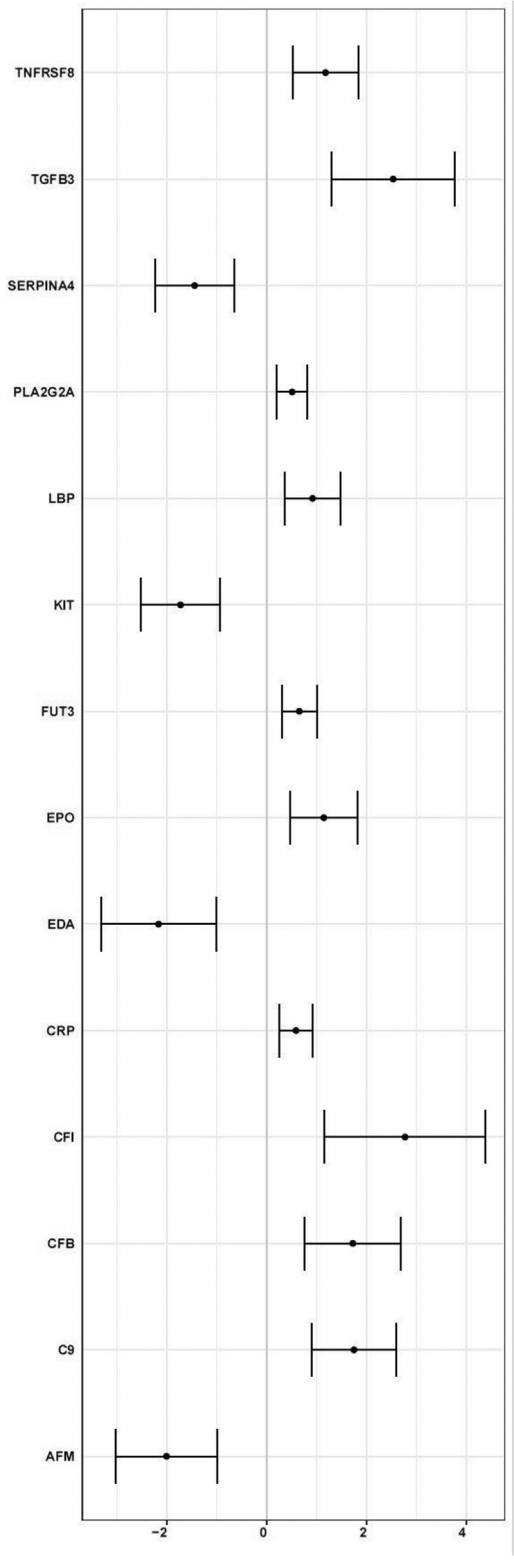
Supplementary Table 3 shows the demographics of cases and healthy control (CD) subjects. The median interval between the earliest sample and the date of diagnosis or timing of sample D was –5.9 (interquartile range [IQR], −6.2, −5.7) years for complicated CD, –5.9 (IQR, −6.1, −5.7) years for noncomplicated CD, and –6.4 (IQR, −7.8, −5.4) years for HC subjects. The mean age at diagnosis of patients with complicated CD was 31.4 ± 6.6 years, of patients with noncomplicated CD was 28.9 ± 5.2 years, and of HC subjects was 28.5 ± 4.8 years. The study population was predominantly male and White (Supplementary Table 3). The majority of patients with a complication at diagnosis had ileal involvement (94% [n = 44 of 47].

Supplementary Reference

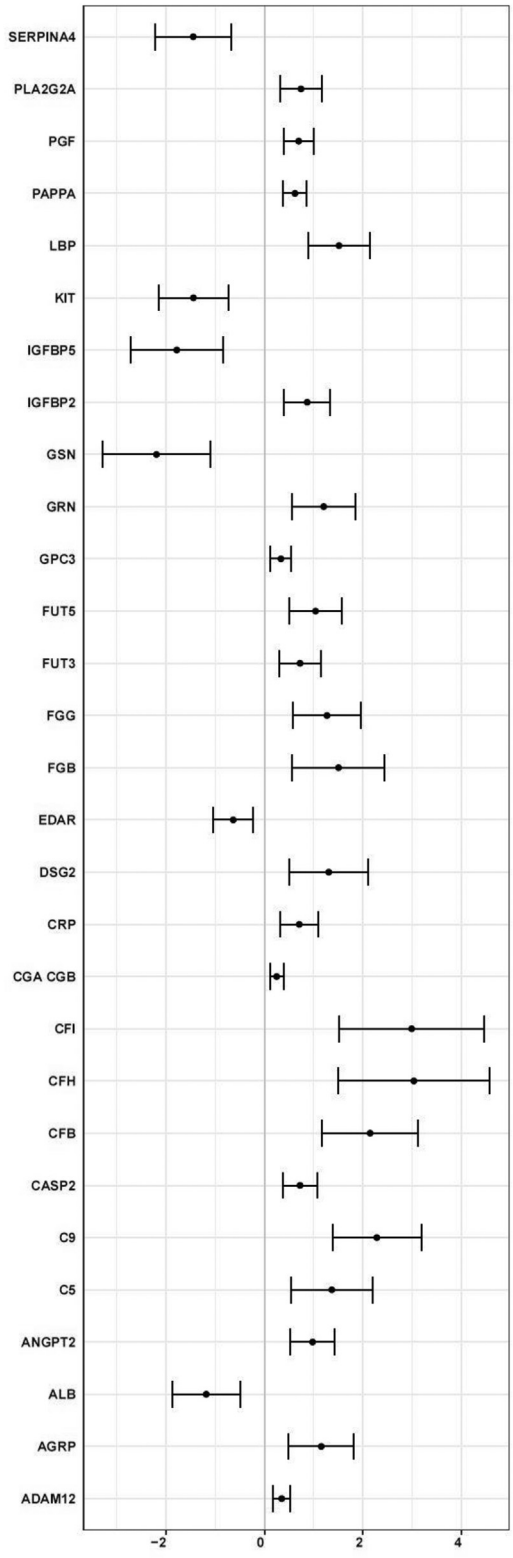
1. Torres J, Petralia F, Sato T, et al. Serum biomarkers identify patients who will develop inflammatory bowel diseases up to 5 years before diagnosis. Gastroenterology 2020;159:96–104.



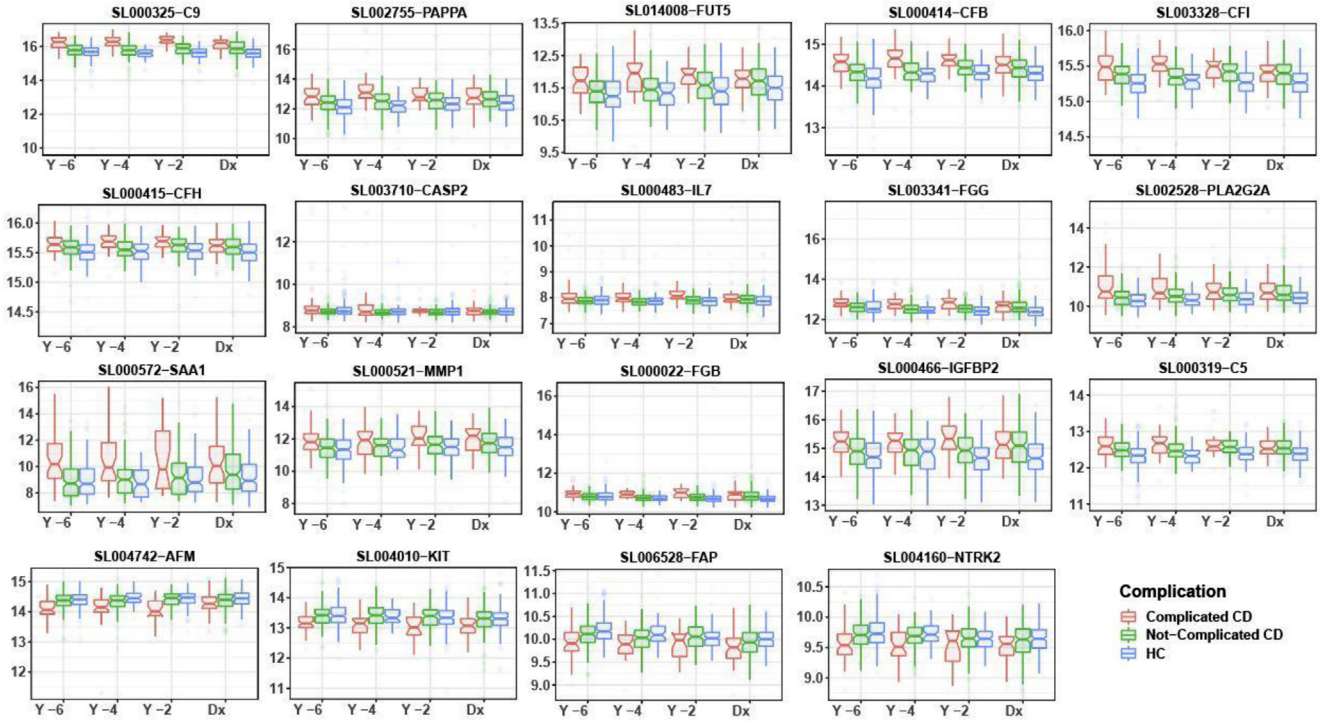
Supplementary Figure 1. Box plots of antimicrobial antibodies according to disease status. Dx, diagnosis; Y, year.



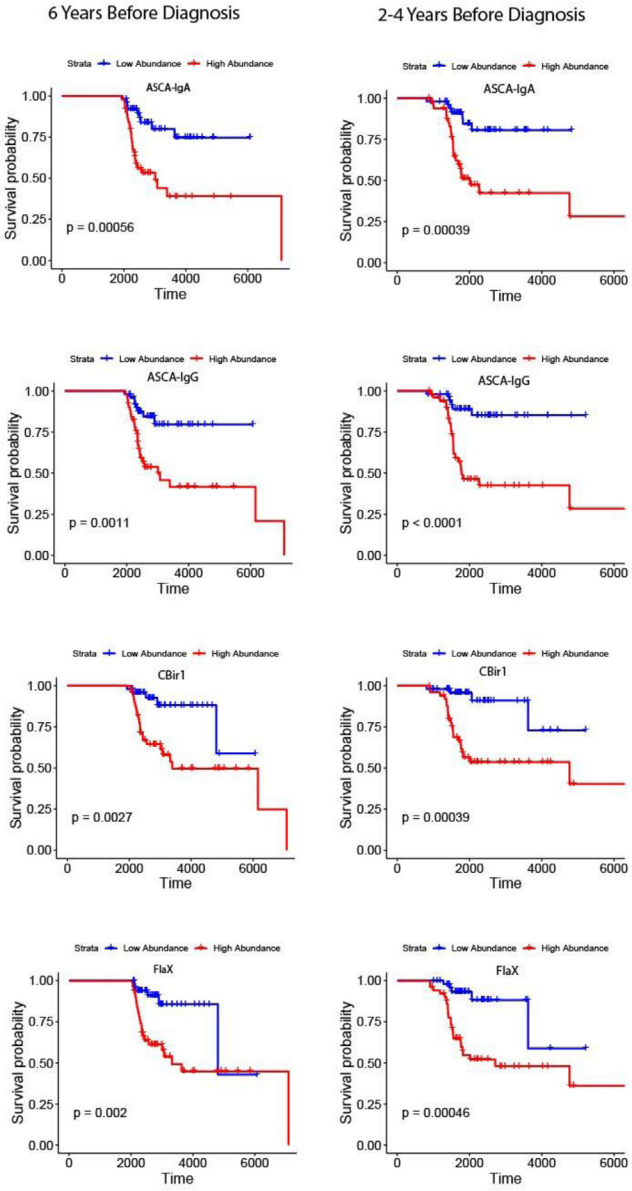
Supplementary Figure 2. Plots of Cox regression coefficients and confidence intervals of significant markers (10% false discovery rate) at 6 years before diagnosis, after adjusting for age, sex, and disease location.



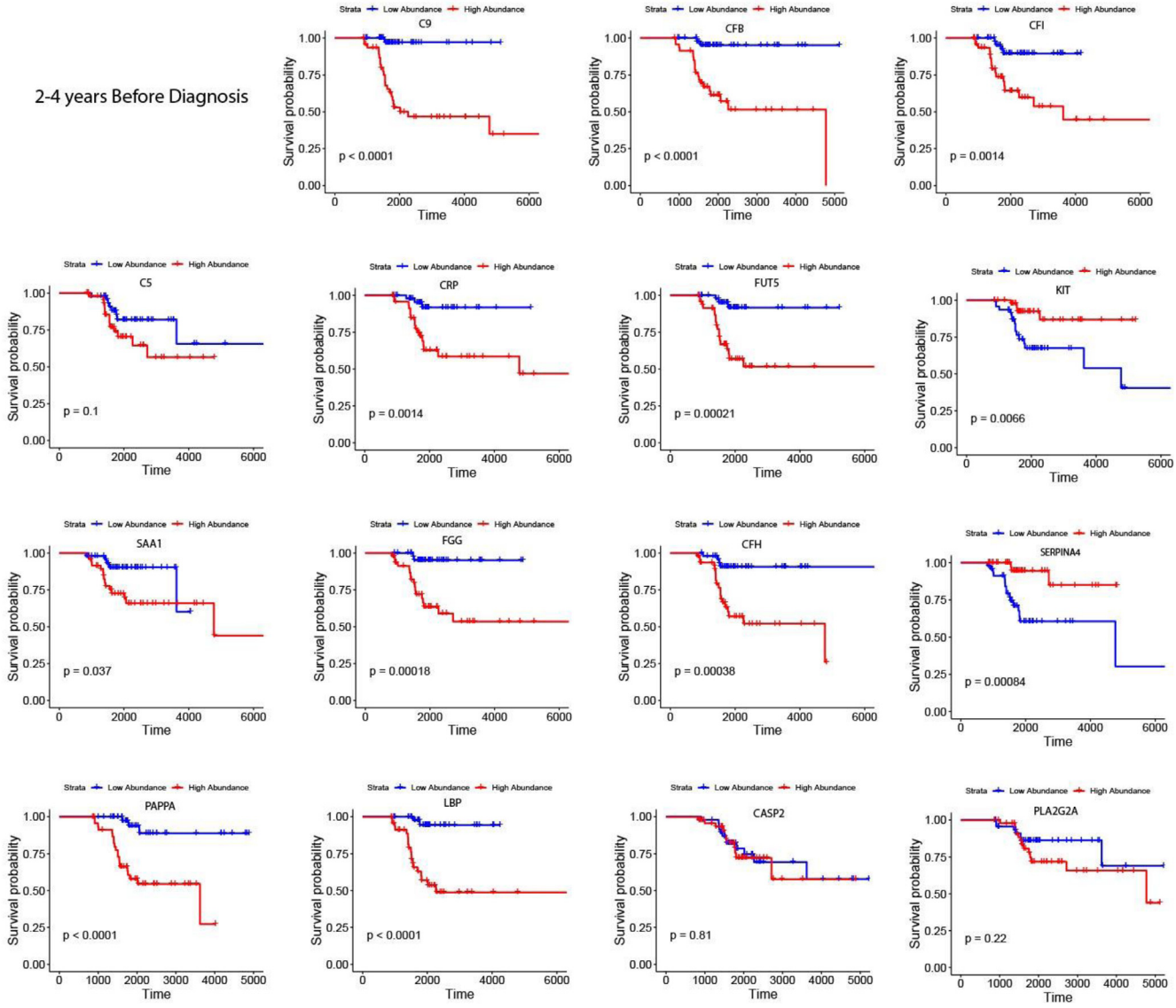
Supplementary Figure 3. Plots of Cox regression coefficients, confidence intervals of significant markers (10% false discovery rate) at 2–4 years before diagnosis, after adjusting for age, sex, and disease location.



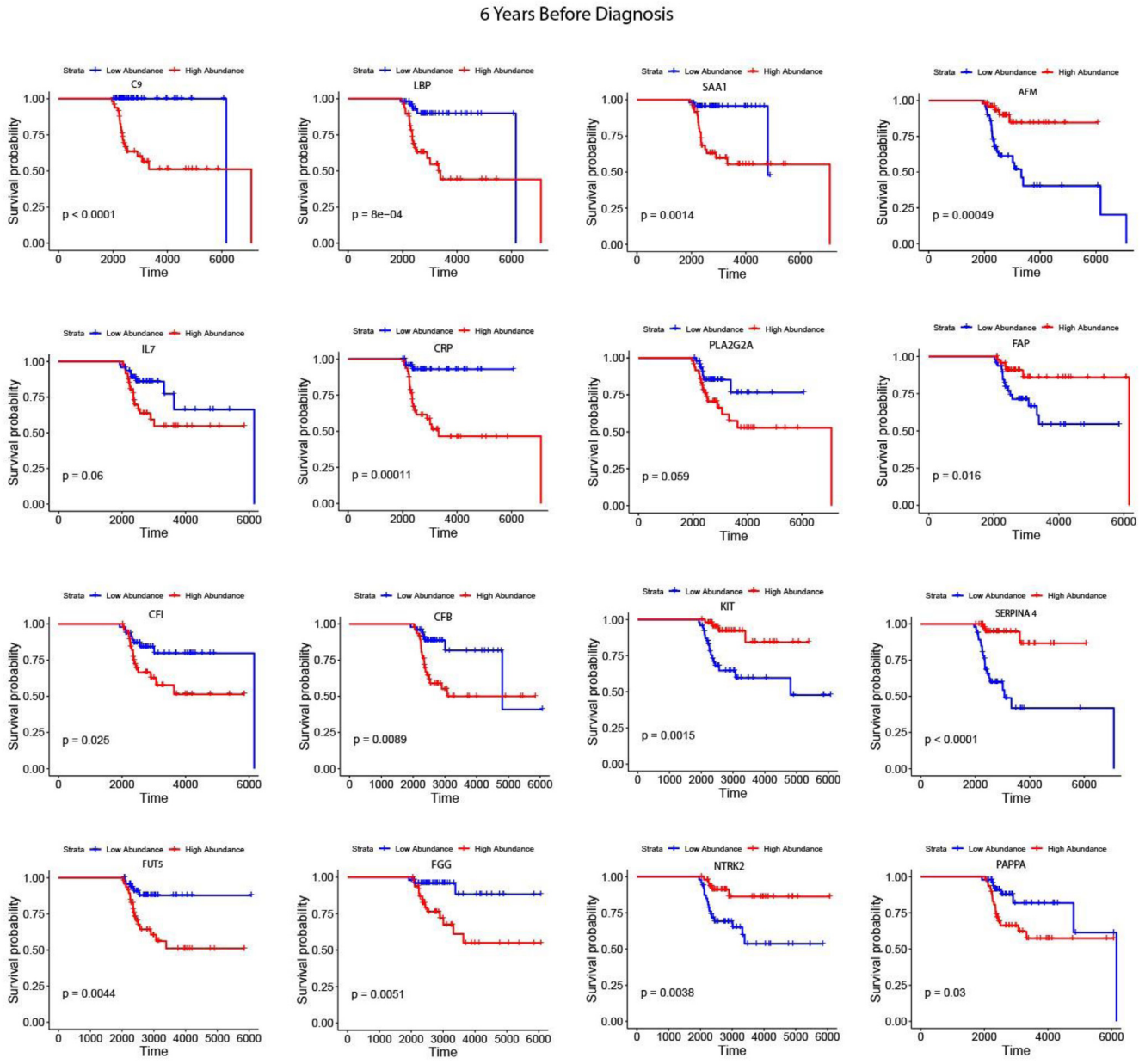
Supplementary Figure 4. Box plots of protein biomarkers according to disease status. Dx, diagnosis; Y, year.



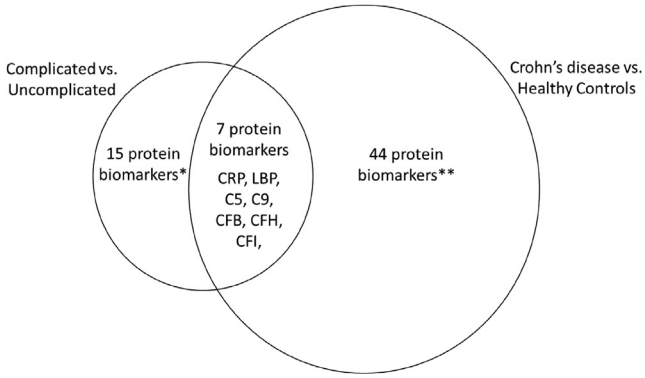
Supplementary Figure 5. KM curves of each antimicrobial antibody and each SomaLogic protein biomarker according to years before diagnosis.



Supplementary Figure 5. (*Continued*)



Supplementary Figure 5. (*Continued*)



Supplementary Figure 6. Distinct protein biomarkers significantly associated with CD onset[1] or complicated phenotypes. Among the 22 protein biomarkers significantly associated with complicated phenotypes at diagnosis, only 7 were associated with CD onset. *, AFM, CASP2, FAP, FGG, FGB, FUT5, IGFBP2, IL-7, KIT, MMP1, NTRK2, PLA2G2A, PAPPA, SERPINA4, SAA1. **, PRSS2, APCS, OMD, ACAN, IL1RA, MRC1, SET, EPHA5, TNFRSF1A, NRCAM, MB, GNS, IGFBP5, IL12BIL23A, BRF1, IL10, C5 C6, EPOR, PON1, HAPLN1, RET, UNC5D, PRTN3, CFC1, GHR, SERPIND1, PLG, TNFRSF1B, Human-virus, HCK, IFGBP6, RARRES2, IL18RAP, PLAUR, CTSL2, AGT, CSK, WIF1, MMP13, ABL1, ECM1, MMP3, UBE21, CSNK2A1.

Supplementary Table 1. ICD-9 or CPT Codes for Complicated Cases

Complication Type	ICD-9-CM Code or CPT Code
Obstructing/stricturing disease (B2)	560, 560.8, 560.89, 560.9, 537.3
Fistulizing/internal penetrating disease (B3)	537.4, 567.22, 567.21, 567.2, 569.5, 569.81, 569.83, 593.82, 596.1, 619.1
Intestinal resection	44160, 44205, 44207, 44120

CPT, Current Procedural Terminology; ICD-9, International Classification of Diseases–Ninth Revision; ICD-9-CM, International Classification of Diseases–Ninth Revision–Clinical Modification.

Supplementary Table 2. Abbreviations of Serologic and Protein Biomarkers

Abbreviation	Full Name
Serologic markers	
ASCA	anti-*Saccharomyces cerevisiae*
OmpC	anti-outer membrane protein C precursor
Protein biomarkers	
AFM	Afamin
CASP2	Caspase-2
CRP	C-reactive protein
CFB	Complement factor B
CFI	Complement factor I
CFH	Complement factor H
C5	Complement C5a
C9	Complement 9
FAP	Fibroblast activation protein α
FGG	Fibrinogen gamma chain dimer
FGB	d-dimer
FUT5	Fucosyltransferase 5
IGFBP2	Insulin like growth factor binding protein 2
IL-7	Interleukin-7
KIT	Stem cell factor receptor/CD117/c-Kit
LBP	Lipopolysaccharide-binding protein
MMP1	Matrix metalloproteinase-1
NTRK2	Neurotrophic tyrosine kinase receptor type 2
PLA2G2A	Phospholipase A2 Group IIA
PAPPA	Pregnancy-associated plasma protein-A
SERPINA4	Serpina family A member 4
SAA1	Serum amyloid A

Supplementary Table 3. Study Population Characteristics

	Crohn's Disease, Complicated at Diagnosis (n = 47)	Crohn's Disease, Noncomplicated at Diagnosis (n = 154)	Healthy Control Subjects (n = 201)
Samples tested at 2–4 y before diagnosis	47	154	200
Samples tested at 6 y before diagnosis	47	154	200
Age, y[a]	30.1 ± 5.9	31.8 ± 6.8	28.5 ± 4.7
Male, %	77%	81%	91%
White, %	75%	68%	89%
Disease behavior at diagnosis, %			
Obstruction	77%	—	—
Penetration	19%		
Surgery	4%		
Disease location at diagnosis[b]			
L1	20 (43%)	37 (24%)	
L2	2 (4%)	38 (25%)	
L3	24 (51%)	46 (30%)	
Unknown	1 (2%)	33 (21%)	

Values are n, mean ± SD, or n (%), unless otherwise indicated.
[a]Crohn's disease cases at diagnosis and healthy control subjects (sample A).
[b]L1, ileal disease; L2, colonic disease; L3, ileocolonic disease.

Supplementary Table 4. The Association Between Complication Phenotypes and Protein Biomarkers After Adjusting by Disease Location, Age, and Sex (Cox Regression Model)

	6 y Before Diagnosis (Time D)			2–4 y Before Diagnosis (Time B/C)		
Gene	*P* Value	Adjusted *P* Value[a]	Coefficient in a Cox Regression[b]	*P* Value	Adjusted *P* Value[a]	Coefficient in a Cox Regression[b]
C9	.00006	.02330[c]	1.75048	.0000007	.00040[c]	2.28231
CFB	.00046	.05744[c]	1.72744	.00002	.00369[c]	2.14613
CRP	.00072	.07812[c]	0.58737	.00039	.02329[c]	0.70789
CFI	.00076	.07812[c]	2.77520	.00007	.00765[c]	2.99112
PLA2G2A	.00097	.08400[c]	0.50989	.00056	.03014[c]	0.74238
LBP	.00112	.09058[c]	0.91962	.00000	.00078[c]	1.51495
SAA1	.00138	.10393	0.23852	.00339	.12346[c]	0.20811
PAPPA	.00447	.16422	0.52767	.00000	.00040[c]	0.62019
CASP2	.00384	.16422	0.39675	.00004	.00631[c]	0.72224
CFH	.00451	.16422	2.59233	.00012	.01016[c]	3.03548
FUT5	.00216	.14325	0.95883	.00012	.01016[c]	1.03816
IGFBP2	.00573	.17970	0.70953	.00031	.02047[c]	0.86780
FGG	.01185	.23658	1.18003	.00034	.02148[c]	1.27079
C5	.01298	.24418	1.22994	.00133	.06018[c]	1.36878
FGB	.01563	.27571	1.40844	.00173	.07214[c]	1.50506
IL7	.01484	.26754	0.63688	.00480	.14875	0.72409
MMP1	.00229	.14337	0.52647	.00835	.20873	0.51570
KIT	.00002	.02000[c]	–1.72917	.00007	.00765[c]	–1.44441
AFM	.00012	.03422[c]	–2.01037	.00960	.22116	–0.56528
SERPINA4	.00037	.05738	–1.44678	.00026	.01948[c]	–1.44749
FAP	.00606	.18493	–1.34379	.01490	.26976	–1.37278
NTRK2	.00305	.16247	–1.71492	.00671	.17619	–1.48064

For abbreviation expansions, please see Supplementary Table 2.

[a]Adjusted *P* values were calculated based on the adjustment of age, gender, and disease locations.

[b]The positive coefficient in a Cox regression indicates a worse prognosis (complications) and a negative coefficient indicates a protective effect of the variables on developing the complications at diagnosis.

[c]Adjusted *P*-value $< 10\%$.

Supplementary Table 5. Complication Phenotypes and Protein Biomarkers Before Diagnosis, Adjusted by Age and Gender (Cox Regression Model)

6 y Before Diagnosis (Time D)				2–4 y Before Diagnosis (Time D)			
Protein	*P* Value	Adjusted *P* Value[a]	Coefficient From Cox Regression[b]	Protein	*P* Value	Adjusted *P* Value[a]	Coefficient From Cox Regression[b]
C9	.000000443	.0002	2.160409	C9	.0000000058	.0000131	2.42725087
SERPINA4	.000002100	.000659	–1.83492	PAPPA	.0000000175	.0000197	0.701759115
PLA2G2A	.000006830	.001285	0.658492	LBP	.0000000792	.0000596	1.503698892
KIT	.000008180	.001319	–1.83507	CFB	.0000002410	.000135783	2.410964398
CFB	.000024700	.002672	2.101364	CFI	.0000007650	.000287902	3.432223201
LBP	.000026800	.002752	1.157265	SERPINA4	.0000023300	.000658968	–1.812038413
AFM	.000028900	.002835	–1.68646	PGF	.0000026900	.000674717	0.744562501
SAA1	.000035400	.003137	0.303273	CASP2	.0000058900	.001284693	0.794488205
CRP	.000036100	.003137	0.715824	FUT5	.0000068300	.001284693	1.138661022
FGG	.00010865	.008178	1.694267	FGG	.0000077300	.001318862	1.487298163
FGB	.00017218	.0119	1.951875	IGFBP5	.0000113000	.001702077	–2.073680362
NTRK2	.00017391	.0119	–2.20944	GSN	.0000162000	.002284455	–2.299462286
EDA	.00022185	.01335	–2.12055	PLA2G2A	.0000207000	.00267244	0.89178537
PAPPA	.00022468	.01335	0.709576	CFH	.0000231000	.00267244	3.137133832
CASP2	.00023973	.01388	0.494113	ANGPT2	.0000241000	.00267244	0.869831802
CFI	.00028185	.015523	2.927202	KIT	.0000249000	.00267244	–1.630929664
FUT5	.00029053	.015615	1.124963	CRP	.0000314000	.002958224	0.766426005
FAP	.00037712	.018923	–1.637	IGFBP2	.0000503000	.004203845	0.976461237
MMP3	.00045914	.022058	1.035944	ADAM12	.0000647000	.005214727	0.340196445
CNTN1	.0006858	.031603	–1.56826	C5	.0000773000	.006016109	1.556588656
IL7	.00078368	.035391	0.835993	SAA1	.000153087	.011150674	0.262222963
LYZ	.00087214	.037871	0.518574	CGA CGB	.000196012	.013017516	0.271883622
TNFRSF8	.00099431	.040363	1.201836	FGB	.000202291	.013050646	1.518847699
NCAM1	.00100001	.040363	–1.34181	GRN	.00022355	.013350401	1.265609
BMP1	.00105976	.041258	–1.02382	DSG2	.000257646	.014544137	1.417792937
CHL1	.00125528	.048041	–1.39902	IL7	.000297357	.015614676	0.921168875
RARRES2	.00134071	.050127	1.699627	FAP	.000315551	.016193512	–1.896628465
TGFB3	.0013764	.050127	1.897887	MMP12	.000455567	.022058259	0.882149767
TNFSF8	.00167785	.058435	–1.88318	AGRP	.000614686	.02891587	0.881751562
IGFBP2	.00170803	.058435	0.775074	ACY1	.000871073	.037871037	–0.608462591
MMP1	.00175916	.059286	0.529295	FSTL3	.000921923	.039277404	1.431282978
KLKB1	.00179756	.05969	–1.71598	SERPINA3	.001001031	.040363013	2.313769749
EGFR	.00192972	.063149	–1.71297	MMP8	.001055714	.04125752	0.842733277
IL19	.00205023	.065203	–1.5741	APCS	.001359678	.050127415	2.024181957
KIR3DL2	.00224954	.069582	0.37067	IGF1	.001398975	.050141038	–1.921406358
ALCAM	.00247377	.074477	–1.63952	ALB	.001690062	.058435159	–1.032066309
CFH	.00262218	.076894	2.767531	C5 C6	.001991609	.064243611	1.007296736

Supplementary Table 5. Continued

6 y Before Diagnosis (Time D)				2–4 y Before Diagnosis (Time D)			
Protein	*P* Value	Adjusted *P* Value[a]	Coefficient From Cox Regression[b]	Protein	*P* Value	Adjusted *P* Value[a]	Coefficient From Cox Regression[b]
SHH	.00272354	.077845	–0.9417	EDAR	.002224038	.069581526	–0.599574033
CAPG	.00288794	.081512	0.468189	SELL	.002396611	.073129034	–1.623055255
FUT3	.00331225	.091208	0.468359	F7	.002507996	.074513877	–1.166047121
SOD2	.00351625	.094899	–1.49092	NTRK2	.002689868	.077844861	–1.675119113
C5	.00353035	.094899	1.417452	MMP10	.003101506	.086459267	–1.932268602
				MMP1	.003699553	.098277525	0.557021939
				AFM	.003802853	.099847006	–0.617827933

For abbreviation expansions, please see Supplementary Table 2.
[a]Adjusted *P* values were calculated based on the adjustment of age and sex.
[b]The positive coefficient in a Cox regression indicates a worse prognosis (complications) and a negative coefficient indicates a protective effect of the variables on developing the complications at diagnosis.

Supplementary Table 6. The AIC for the SomaLogic-Based Model, Serologic-Based Model, and the Model Integrating Both SomaLogic and Serologic Markers

	SomaLogic (k = 22)	Serologic (k = 5)	SomaLogic + Serologic (k = 27)
B/C[a]	398.0359	423.7102	401.3239
D[b]	422.425	427.3763	429.2994

The AIC index is used to determine how well the model fits the data while adjusting for the total number of parameters in the model. This score is calculated as a function of the log likelihood and the parameters estimated in the model (k), ie, AIC = –2*(partial log likelihood)+ 2*k. The model based on SomaLogic protein biomarkers achieves a lower AIC than the serologic + SomaLogic markers and the serologic-based models. This result demonstrates that the SomaLogic-based model can fit the data better than the competitive models.
AIC, Akaike Information Criterion.
[a]B/C= 2–4 years before diagnosis
[b]D = 6 years before diagnosis.