Published in final edited form as:

Inflamm Bowel Dis. 2011 December ; 17(12): 2488–2496. doi:10.1002/ibd.21661.

Combination of Genetic and Quantitative Serological Immune Markers are Associated with Complicated Crohn's Disease Behavior

Gary R. Lichtenstein, MD^{*}, Stephan R. Targan, MD^{†,‡}, Marla C. Dubinsky, MD^{‡,§}, Jerome I. Rotter, MD^{†,¶}, Derren M. Barken, PhD ^{||}, Fred Princen, PhD ^{||}, Susan Carroll, PhD ^{||}, Michelle Brown, BA ^{||}, Jordan Stachelski, BS ^{||}, Emil Chuang, MD ^{||}, Carol J. Landers, BS^{†,‡}, Joanne M. Stempak, MS^{**}, Sharat Singh, PhD ^{||}, and Mark S. Silverberg, MD, PhD^{**}

^{*}Department of Medicine, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania

[†]Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, California

[‡]Cedars-Sinai Inflammatory Bowel and Immunobiology Research Institute, Los Angeles, California

[§]Department of Pediatrics, Cedars-Sinai Medical Center, Los Angeles, California

[¶]Cedars-Sinai Medical Genetics Institute, Los Angeles, California

^{||} Research and Development, Prometheus Laboratories Inc., San Diego, California

^{**}Mount Sinai Hospital IBD Group, Zane Cohen Centre for Digestive Diseases, University of Toronto, Toronto, Ontario, Canada

Abstract

Background—Treatment of Crohn's disease (CD) with biologics may alter disease progression, leading to fewer disease-related complications, but cost and adverse event profiles often limit their effective use. Tools identifying patients at high risk of complications, who would benefit the most from biologics, would be valuable. Previous studies suggest that biomarkers may aid in determining the course of CD. We aimed to determine if combined serologic immune responses and *NOD2* genetic markers are associated with CD complications.

Methods—In this cross-sectional study, banked blood from well-characterized CD patients ($n = 593$; mean follow-up: 12 years) from tertiary and community centers was analyzed for six serological biomarkers (ASCA-IgA, ASCA-IgG, anti-OmpC, anti-CBir1, anti-I2, pANCA). In a patient subset ($n = 385$), *NOD2* (SNP8, SNP12, SNP13) genotyping was performed.

Complications included stricturing and penetrating disease behaviors. A logistic regression model for the risk of complications over time was constructed and evaluated by cross-validation.

Results—For each serologic marker, complication rates were stratified by quartile. Complication frequency was significantly different across quartiles for each marker (P trend = 0.001). Patients with SNP13 *NOD2* risk alleles experienced increased complications versus patients without *NOD2* mutations (P = 0.001). A calibration plot of modeled versus observed complication rates demonstrated good agreement (R = 0.973). Performance of the model integrating serologic and genetic markers was demonstrated by area under the receiver operating characteristic curve (AUC = 0.801; 95% confidence interval: 0.757–0.846).

Conclusions—This model combining serologic and *NOD2* genetic markers may provide physicians with a tool to assess the probability of patients developing a complication over the course of CD.

Keywords

complications; Crohn's disease; genetics; *NOD2*; serology

Inflammatory bowel disease (IBD) is a chronic inflammatory disorder of the gastrointestinal tract, consisting of ulcerative colitis (UC) and Crohn's disease (CD), which together affect ≈1.6 million individuals in the United States and Canada.^{1,2} There is currently no cure for CD, thus the main goal of treatment is to suppress the inflammatory response and achieve clinical and mucosal remission. Although some patients with CD will experience a benign clinical course, other patients face a progressive disease course which can include the development of complications such as stricturing and penetrating disease behaviors requiring surgery—≈80% of all patients will require surgery during their lifetime.^{3–6} It has been shown that the need for corticosteroids is a marker for debilitating CD.⁷ The use of corticosteroids is associated with an acceleration of the disease course, with ≈35% of patients requiring small bowel surgery within 1 year. In addition, 25%–33% of patients with noncomplicated disease have been reported to transition to stricturing or internal and perianal penetrating disease after 5 years—suggesting that most patients, if undertreated, will transition from a noncomplicated to complicated disease state if followed for sufficient time.⁵ In addition to the need for corticosteroids, the clinical features suggested to be predictors of aggressive CD are disease onset before age 40, smoking, the presence of perianal lesions at diagnosis, and small bowel localization and stricturing disease.^{7–10}

Recent evidence suggests that with appropriate therapy, progression to disease complications can be minimized.¹¹ D'Haens et al¹² demonstrated that newly diagnosed/treatment naive patients treated with a combination of biologics and immunomodulators had significantly higher rates of remission and mucosal healing compared to patients treated with a conventional management approach utilizing corticosteroids as initial therapy. Importantly, mucosal healing has been shown to be the key factor contributing to steroid-free remission.¹³ Together, these data indicate that early effective therapy—referred to as the “top-down” approach—provides significant benefit to patients with CD. Biologics, however, are costly and are associated with rare but serious and sometimes fatal adverse events including opportunistic infections and malignancies.¹¹ Therefore, in order to balance the risks with the benefits of these therapies it is becoming increasingly important for physicians to identify those patients who would be appropriate for early effective intervention.

Evidence suggests that disease progression and the need for surgery are associated with the immune responses to intestinal microorganism antigens.¹⁴ Specifically, an immune response to microbial antigens associated with *Pseudomonas fluorescens* (I2), porin protein C from *Escherichia coli* (OmpC), bacterial flagellin (CBir1), and *Saccharomyces cerevisiae* (ASCA) have been shown to correlate with complicated disease phenotypes such as strictures and fistulas, as well as with the number of small bowel surgeries.^{15–17} Many of these serological markers are already in use in clinical practice as a diagnostic tool to differentiate between CD and UC.¹⁸ Multiple studies have shown that both the presence and the magnitude of individual markers and of marker combinations are correlated with specific phenotypes and with the need for intestinal surgery.^{14–17,19} In a recent prospective pediatric study, the magnitude of immune response against microbial antigens was shown to significantly correlate with complicated CD phenotypes and was associated with rapid disease progression.¹⁹ These observations suggest that responses to microbial antigens are closely associated to clinical disease characteristics and may be associated with disease phenotypes and progression to complicated disease.

Genetic factors have also been demonstrated to play an important role in determining disease phenotype in CD. While numerous loci have been identified as conferring susceptibility to CD overall,²⁰ the innate immunity gene nucleotide oligomeric domain 2 (*NOD2*) has been shown to be associated with disease phenotype.^{18,21–27} Although at least 27 *NOD2* variants have been characterized, three major single nucleotide polymorphisms (SNPs), namely, R702W (SNP8), G908R (SNP12), and 1007fs (SNP13), are responsible for the majority of clinically relevant variation in the population, and are associated with the development of disease complications.^{15,21,22}

Earlier investigations have demonstrated clear, but independent associations of the immune response to microbial antigens and the presence of *NOD2* variants with complicated disease in CD patients. Ippoliti et al²⁸ have recently demonstrated that these two factors are synergistic, and in combination increase the risk of complicated disease. The aim of this study was to develop a sero-genetic model designed to determine the probability of complicated disease behavior occurring by a specific time in patients with CD.

MATERIALS AND METHODS

Study Population

This study used a cross-sectional design employing blood samples obtained from patient sample banks. Our initial cohort consisted of 770 subjects with established CD; 177 subjects were excluded due to inadequate clinical documentation, resulting in a final study population of 593 individuals (51% female and 49% male). Serum samples were obtained from 1) Cedars-Sinai Medical Center, Los Angeles ($n = 274$); 2) Mount Sinai Hospital, Toronto, Canada ($n = 235$); and 3) a multicenter Prometheus study ($n = 84$). In a subgroup of study subjects from Cedars-Sinai Medical Center ($n = 184$) and Mount Sinai Hospital ($n = 201$), *NOD2* genotyping results were available for analysis. Study protocols were Institutional Review Board (IRB)-approved for each site.

Subjects were diagnosed with CD based on a combination of standard criteria that included clinical symptoms, endoscopy, histopathology, video capsule, and/or radiographic studies. This cohort was used because there was extensive and longitudinal medical information available for these patients, including the date of diagnosis, disease location, and disease behavior according to the Montreal Classification.²⁹ Disease behavior of each subject was classified as nonpenetrating/non-stricturing (B1) or stricturing (B2) or internal penetrating (B3) by the physician at the site of patient enrollment. For the purposes of this study, patients who were nonpenetrating/nonstricturing at the time of blood draw or last evaluation were defined as having noncomplicated disease while patients whose behavior evolved to stricturing or penetrating behavior, or both, were defined as having complicated disease.

***NOD2* Genotyping**

NOD2 genotyping was performed at Cedars-Sinai Medical Center for the patients enrolled at that site; Prometheus Laboratories (San Diego, CA) performed *NOD2* genotyping on the additional samples. Genotyping consisted of testing three *NOD2* SNPs: SNP8 is a 2104C→T in exon 4 resulting in a R702W substitution (rs2066844); SNP12 is a 2722G→C in exon 8 resulting in a G908R substitution (rs2066845); and SNP13 is a C insertion in exon 11 (3020InsC) resulting in a frame shift (1007fs) (rs5743293). At Prometheus Laboratories, *NOD2* genotyping consisted of an allelic discrimination polymerase chain reaction (PCR) method including two specific oligonucleotide sequences and two TaqMan probes for each assay (Applied Biosystems, Foster City, CA). The genotyping assays were performed on an ABI 7000 Real-Time PCR system (Applied Biosystems).

Detection of Anti-I2

Enzyme-linked immunosorbent assay (ELISA) was originally developed by Sutton et al,³⁰ and was modified at Prometheus Laboratories to detect concentrations of anti-I2 in the blood. Briefly, the anti-I2 assay utilized a standard 96-well sandwich ELISA format plate. A refolded GST-tagged protein, consisting of 100 amino acids from the I2 sequence, was captured on the plate using a monoclonal anti-GST antibody coated on the well surface (Genscript, Piscataway, NJ). Test human serum samples were diluted 1:100 in order ensure the antibody concentration was within the range of the standard curve. After incubation of the serum samples in the wells, anti-I2 antibodies were detected using an alkaline phosphatase enzyme conjugated to an antihuman IgA reagent (Jackson ImmunoResearch Laboratories, West Grove, PA). The reactions were revealed using a chemiluminescent substrate solution (Applied Biosystems) and expressed as ELISA units that were relative to standards prepared from a pool of sera. A patient was considered positive for anti-I2 if the ELISA result was above the reference range value (368 EU/mL).

Other Serological Analyses

Serum concentrations of anti-CBir1, anti-OmpC, ASCA-IgA, and ASCA-IgG antibodies were measured by ELISA. Testing for perinuclear-staining antineutrophil cytoplasmic antibodies (pANCA) was performed by immunofluorescence staining of neutrophils with the aim of visualizing perinuclear localization and a disrupted staining pattern associated with deoxyribonuclease (DNase) treatment. All assays were performed at Prometheus

Laboratories using a commercial assay (IBD-Serology 7, Prometheus Laboratories). For ELISA, measurements were expressed as ELISA units, relative to standards prepared from a pool of sera. The anti-ASCA ELISA was based on a method designed by Sendid et al.³¹ Two ASCA ELISAs, ASCA-A and ASCA-G, were used to measure IgA and IgG antibodies, respectively. The anti-CBir1 ELISA procedure measured IgG antibodies,³² and the anti-OmpC ELISA procedure measured IgA antibodies.³³ A patient was considered positive for a serology marker if the ELISA result was above the reference range values. The test for IBD-specific pANCA was conducted using indirect immunofluorescence on polymorphonuclear leukocytes (PMNs), either untreated or digested with DNase.³⁴ Treated and untreated PMNs were fixed to glass slides and diluted patient serum added. Following incubation and washing, a fluoresceinated goat antihuman IgG antibody was added to the slides. Epifluorescent microscopy was used to confirm the characteristic perinuclear staining pattern on the untreated cells. If the perinuclear pattern presented, the reactivity on the DNase-digested cells was assessed. Samples were considered positive for pANCA if the perinuclear staining pattern was lost in the DNase digested cells; all other staining patterns were designated as negative for pANCA.

Statistical Methods

The assay results for the serological markers anti-I2, anti-CBir1, anti-OmpC, ASCA-IgA, and ASCA-IgG were converted into a categorical variable, specifically, a quartile. For each marker the increasing trend of complications by quartiles was assessed using the Cochran–Armitage test for trend. Since the pANCA and the genetic variables results were already binary no transformation was necessary and Pearson’s chi-square test was applied.

In order to assess the response of the six combined serology markers, the quartile sum score (QSS) technique was applied.¹⁵ In this study pANCA results were dichotomous in nature and assumed to be negatively correlated with disease complication, in keeping with previous reports.¹⁴ Patients with positive pANCA were assigned a score of 1 and those with negative pANCA were assigned a score of 4. Thus, the minimum QSS of 6 represents a patient with the lowest quartile score of 1 for ASCA-IgG, ASCA-IgA, anti-CBir1, anti-OmpC, and anti-I2, combined with a “positive” pANCA score of 1. The maximum QSS of 24 represents a patient with ASCA-IgG, ASCA-IgA, anti-CBir1, anti-OmpC, and anti-I2 in the highest quartile (individual score 4) combined with a “negative” pANCA score of 4.

The logistic regression model was constructed using a logit link function. The output was the probability of complication within a given time period. This model incorporated QSS, duration of disease from diagnosis to time of blood draw, and the *NOD2* SNPs 8, 12, and 13. For the three SNPs, there were three binary variables indicating the presence of heterozygous SNP risk alleles. There was an additional binary variable indicating the presence of two or more *NOD2* risk alleles across the three SNPs—a value of one for this variable could indicate either homozygous risk alleles, compound heterozygotes, or a combination of homozygous and heterozygous risk alleles, but it always indicated the presence of two or more risk alleles among the three SNPs. An additional input parameter was the duration of disease, a significant factor because by varying the time input, the model could be used to generate associations for a range of possible times. The time input was not

a time to event (i.e., complication), but rather a time interval (duration of disease). This was due to the censored nature of the available data. In this cross-sectional design, the dates of diagnosis and blood draw and the complication status at blood draw were known for each patient. If there was a complication during the interval between diagnosis and blood draw, then the time of complication was effectively interval-censored, (i.e., it was known to fall within the interval between diagnosis and blood draw). To accommodate this limitation in the data, the logistic regression model was trained using the time interval between diagnosis and blood draw (the duration of disease) as an explicit input variable. Consequently, disease duration was taken into account in predictions made by the model, so that longer disease durations were associated with higher probabilities of complications. In this way the model was able to generate associations that explicitly take time into account, despite the fact that time to event data was not available. Using the fitted logistic regression model, sets of probabilities were generated for each individual sample by fixing the QSS score and SNP variables and systematically varying the duration of disease input (1–40 years).

Even though there was a low prevalence of double risk alleles in this cohort, every sample with a double risk allele had a complication (100%). A pseudocount methodology was employed to adjust this rate to 99%, reflecting the possibility that a double risk allele might (rarely) not be associated with complications. Finally, in order to confirm the predictive utility of the quartile sum score, we analyzed a separate logistic regression model, not including disease durations, which was constructed on a subset of patients with a limited disease duration interval of 3–9 years.

The parameters of the multiple logistic regression model were assessed using a Wald test. The associations of the logistic regression model were evaluated using a leave-one-out cross-validation, with two complementary statistical assessments as recommended by Harrell.³⁵ To generate a calibration plot, the output of the logistic regression model was transformed into a categorical variable, through a simple discretization, into 10 categories. Within each category the true complication rate was computed and the agreement of the modeled and observed complication rates was assessed via Pearson's correlation. In the second assessment the discriminatory ability of the model was evaluated using a receiver operating characteristic (ROC) curve. Under this assessment, the overall performance of the test was reported via the area under the curve (AUC) statistic with confidence intervals. All statistical results were computed using the R open source package, v. 2.8.1 (R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>).

RESULTS

Fifty-one percent of the patients were female, and the mean patient age at the time of blood draw was 38 years. Thirty-one percent of the patients were diagnosed at 18 or younger, and 5% were 18 or younger at blood draw. The mean CD duration was 8 years (standard deviation [SD]: 9) and 15 years (SD: 13) (chi-square test, $P = 0.001$) for noncomplicated and complicated patients, respectively, at the time of inclusion. Fifty-seven percent of patients had experienced complications at the time of the blood draw. Clinical characteristics of the patient cohort are shown in Table 1.

Correlation of Serological Immune Responses and Genetic Markers to Disease Behavior

Quartile scores were calculated from the study population of 593 CD patients. As shown in Table 2A, the proportion of patients with disease complications increased with increasing quartile for ASCA-IgA, ASCA-IgG, anti-I2, anti-CBir1, anti-OmpC individually (Cochran-Armitage trend test, $P = 0.001$). Forty-five percent of pANCA-positive patients had complicated disease while 60% of pANCA-negative patients had complicated disease ($P = 0.029$, chi-square test) (Table 2b). Representing the extremes of marker expression, 3% of patients were negative for ASCA-IgA, ASCA-IgG, anti-I2, anti-CBir1, anti-OmpC, but positive for pANCA and 4% were positive for ASCA-IgA, ASCA-IgG, anti-I2, anti-CBir1, anti-OmpC, but negative for pANCA.

Quartile sum scores were used to assess the association of the combined six serology markers (QSS range: 6–24) to complicated and noncomplicated disease. As shown in Figure 1A, the median QSS for patients with complicated disease was 18, compared to a QSS of 14 for patients with noncomplicated disease (Mann–Whitney test, $P < 0.001$).

As noted above, there was a difference between the noncomplicated and complicated patients in the mean disease duration at blood draw. However, the serology and genomic predictors remain statistically significant even when disease duration is included as a possible confounding variable. An additional analysis was carried out for a stratified subset of patients, namely, those patients with disease durations of 3 to 9 years (complicated $n = 70$; non-complicated $n = 112$) where the mean disease duration was 5.9 and 5.4 years for the complicated and noncomplicated groups, respectively. In this subset a logistic regression model was constructed which did not include duration as a predictor (data not shown); the resulting model's coefficient for the QSS predictor was still highly statistically significant (Wald test, $P < 0.001$).

Relationship of *NOD2* Variants to Complicated and Noncomplicated Disease

In 385 patients the three *NOD2* variants were assessed for their relationship to the incidence of disease complications (Table 3). A total of 103 patients (27%) carried one or more *NOD2* variants; this included 43 patients who were heterozygous or homozygous exclusively for SNP8; 22 patients who were heterozygous or homozygous exclusively for SNP12; and 23 patients who were heterozygous or homozygous exclusively for SNP13. An additional 15 patients carried compound risk alleles from among SNPs 8, 12, and 13.

The results show a strong association between disease complications and the presence of a risk allele for SNP13. Moreover, there was a significant difference in the number of patients with the SNP13 variant who experienced complicated disease compared to no risk allele wildtype patients (Fisher's exact test, $P < 0.001$). The total number of patients with a SNP13 variant includes the 23 patients who were heterozygous or homozygous exclusively for SNP13, and an additional 10 patients who were SNP13-containing compound heterozygotes. Thirty-two out of 33 or 97% of patients with the SNP13 variant experienced complicated disease (Table 3). The results also indicated that homozygous risk alleles for SNP8 or SNP12 or compound heterozygotes containing a combination of SNPs 8, 12, and 13 are associated with CD complications. Indeed, among the 20 patients in the cohort with such

risk alleles, all (100%) were observed to have complications (Table 3). In contrast, in this cohort there does not appear to be an effect of heterozygous SNP8 and SNP12 on complication status (Table 3). Therefore, in this model patients with multiple *NOD2* risk alleles, including all homozygous or compound heterozygotes involving SNPs 8, 12, or 13 are assigned a high probability of complications.

Logistic Regression Modeling

The parameters and predictions for the logistic regression model are shown in Table 4. The cumulative probability of complications occurring over time is illustrated in Figure 1B,C. In the model, the complication status was presented as the outcome variable. Figure 1B shows the curves based on QSS score with no *NOD2* risk alleles, representing the predictions for patients who do not carry one of the 3 *NOD2* variants. Each QSS from 6–24 is represented by a single curve. In the first few years after diagnosis the predicted risk difference between the curves representing the low, middle, and high QSS (6, 15, and 24, respectively) is the greatest. For instance, the risk is predicted to be 9%, 34%, and 73% for QSS 6, 15, and 24, respectively, at year 2, and 11%, 39%, and 77% for QSS 6, 15, and 24, respectively, at year 5. The probability of risk remains at a high plateau over time for QSS 24 but continues to rise with increasing time for the mid and low QSS values so that at year 10 the predicted risk is 14%, 47%, 83% for QSS 6, 15, and 24, respectively. The presence of a *NOD2* variant, specifically SNP13, shifts the curves to a greater probability of complications (Figure 1C). This is especially apparent for the curves representing the lower and middle QSSs. The observed risk of complications at 2 years after diagnosis for a QSS 6 is shifted from 9% in the absence of *NOD2* SNP13 variant to 57% if a SNP13 variant is present. When the *NOD2* results are applied, there is a significant transformation to a higher probability of complication in all patients, even early in the disease course and with a lower QSS (Figure 1C). Overall, in the cohort of 385 samples with known genetic results the probabilities of complication for the population as a whole at 5 and 10 years were 57% and 63%.

Use of the Sero-Genetic Regression Model to Determine the Probability of the Development of Complicated CD

In order to assess the accuracy of the overall test a calibration plot was constructed. Patients were divided into 10 groups according to the predicted complication rates (10%–20%, 21%–30%, etc.) and an average predicted complication rate was calculated for each group. In addition, the average rate of observed complications was determined for each group. As shown in Figure 1D, there was a close correlation of between the modeled and observed rates of disease complication in each category (R^2 0.973). The cross-validation probabilities were also evaluated by ROC analysis. Here the area under the ROC curve was 0.801 (95% confidence interval [CI]: 0.757–0.846), thus confirming the accuracy of the model in discriminating complicated and noncomplicated CD (Figure 1E).

DISCUSSION

In this study we demonstrated the utility of a novel test combining quantitative serologic markers and a genetic marker in determining the probability of complicated CD. The test can be utilized to identify those patients who may be at a higher risk of developing a

complicated CD phenotype such as stricturing and penetrating disease behaviors. The test was developed using logistic regression modeling and was validated in a cross-sectional study incorporating the serology results from 593 CD patients and the *NOD2* results from a subset of 385 patients. Modeling based on the data demonstrates increased rates of complications when the serology results are stratified by quartiles, confirming previously published work.^{14,15,19,36} QSS are by themselves informative, but when used as a predictor in a logistic regression model it is possible to more specifically quantify, in probabilistic terms, the expected risk of complications for a range of observation times. Our fitted model incorporated duration of disease as an explicit predictor.

In our cohort the presence of any *NOD2* SNP13 variant, a homozygote for SNP8 or 12, or a compound heterozygote consisting of SNPs 8, 12, and 13 is associated with a significantly increased likelihood of complicated CD behavior above serology data alone.

For a patient with a QSS of 15, in the presence of a *NOD2* SNP13 heterozygous risk allele, there would be a 90% likelihood of a complication within 5 years after diagnosis; this compared to the 39% probability of a complication in that same time frame if the patient did not carry a *NOD2* risk allele. This result confirmed prior research that has demonstrated a strong association between *NOD2* polymorphisms and a complicated disease phenotype.^{18,21–27} As noted previously, these data are supported by recent evidence indicating that there is synergy between *NOD2* and adaptive immunity leading to an increase in susceptibility to a fibrostenosing CD phenotype.²⁸

Two previously published studies reported rates of complications in CD at 5 years after diagnosis of 48%⁵ and 52%,⁶ and at 10 years after diagnosis of 69%⁵ and 70%.⁶ Our model generated rates of complication for the population as a whole of 57% and 63% at 5 and 10 years, respectively. Given inevitable differences in the clinical parameters of these cohorts, the model-generated complication rates for our population are in good agreement with these previously published studies.

There are potential limitations in this study design. The cross-sectional nature of the study limits the interpretation of these data to an association between CD markers and complications, rather than a prediction of future complications. In addition, given that the blood was drawn after a complication had occurred, there may be a question as to the stability of the QSS. Preliminary data suggest that the QSS does not change significantly despite increasing disease duration (data not shown), but it is currently not known if this value changes after a complication has occurred, or in response to therapy. If the serology marker pattern changes dramatically over the course of the disease and treatment, then samples taken later in the disease course may not be representative of samples taken at diagnosis. Although some studies have described an association between marker response level with disease duration,¹⁴ several studies also suggest that there is a basic stability in marker status despite changes in disease activity,³⁷ or duration of disease.^{38,39} Moreover, data from recent prospective studies has shown that the serology markers assessed at or near diagnosis are able to identify patients who are more likely to have complications, thus supporting the conclusions based on cross-sectional data.^{19,36,39} In order to determine the true predictive nature of this novel sero-genetic test, the findings from this study require

further prospective testing in patients who present with noncomplicated disease behaviors and who are then followed longitudinally.

In conclusion, these data demonstrate the value of combining quantitative serologic immune responses and *NOD2* genotype as a test to determine the probability of complicated CD developing over time. Such risk stratification may permit clinicians to develop more individualized treatment plans for the management of CD patients.

Acknowledgments

Supported by Prometheus Laboratories Inc, San Diego, CA.

Disclosures: Dr. Barken, Dr. Princen, Dr. Carroll, Shelly Brown, Jordan Stachelski, Dr. Chuang, and Sharat Singh are employees of Prometheus Laboratories Inc. Carol Landers is a stockholder of Prometheus Laboratories Inc. Dr. Rotter and Joanne Stempak have nothing to disclose. Dr. Lichtenstein has received consulting fees and research support from Prometheus Laboratories Inc. Dr. Targan is a founder and stockholder of Prometheus Laboratories Inc. Dr. Dubinsky is a consultant for Prometheus Laboratories Inc. Dr. Silverberg has received consulting fees and research support from Prometheus Laboratories Inc. Writing assistance: Dr. Anthony Stonehouse provided extensive writing support during the development of this article. Dr. Stonehouse is an employee of Watson & Stonehouse Enterprises, LLC.

REFERENCES

1. Voices of Progress Annual Report 2008 [CCFA Web site]. [Accessed March 8, 2010] 2008 Aug 31. Available at: <http://www.ccfa.org/images/ar09/fullannualrep09.pdf>
2. Bernstein CN, Wajda A, Svenson LW, et al. The epidemiology of inflammatory bowel disease in Canada: a population-based study. *Am J Gastroenterol*. 2006; 101:1559–1568. [PubMed: 16863561]
3. Wolters FL, Russel MG, Sijbrandij J, et al. Disease outcome of inflammatory bowel disease patients: general outline of a Europe-wide population-based 10-year clinical follow-up study. *Scand J Gastroenterol Suppl*. 2006; 243:46–54. [PubMed: 16782622]
4. Loftus EV, Schoenfeld P, Sandborn WJ. The epidemiology and natural history of Crohn's disease in population-based patient cohorts from North America: a systematic review. *Aliment Pharmacol Ther*. 2002; 16:51–60. [PubMed: 11856078]
5. Louis E, Collard A, Oger AF, et al. Behaviour of Crohn's disease according to the Vienna classification: changing pattern over the course of the disease. *Gut*. 2001; 49:777–782. [PubMed: 11709511]
6. Cosnes J, Cattan S, Blain A, et al. Long-term evolution of disease behavior of Crohn's disease. *Inflamm Bowel Dis*. 2002; 8:244–250. [PubMed: 12131607]
7. Beaugerie L, Seksik P, Nion-Larmurier I, et al. Predictors of Crohn's disease. *Gastroenterology*. 2006; 130:650–656. [PubMed: 16530505]
8. Romberg-Camps MJL, Dagnelie PC, Kester ADM, et al. Influence of phenotype at diagnosis and of other potential prognostic factors on the course of inflammatory bowel disease. *Am J Gastroenterol*. 2009; 104:371–383. [PubMed: 19174787]
9. Cottone M, Rosselli M, Orlando A, et al. Smoking habits and recurrence in Crohn's disease. *Gastroenterology*. 1994; 106:643–648. [PubMed: 8119535]
10. Cosnes J, Carbonnel F, Beaugerie L, et al. Effects of cigarette smoking on the long-term course of Crohn's disease. *Gastroenterology*. 1996; 110:424–431. [PubMed: 8566589]
11. Hanauer SB. Positioning biologic agents in the treatment of Crohn's disease. *Inflamm Bowel Dis*. 2009; 15:1570–1582. [PubMed: 19309716]
12. D'Haens G, Baert F, Van Assche G, et al. Early combined immuno-suppression or conventional management in patients with newly diagnosed Crohn's disease: an open randomised trial. *Lancet*. 2008; 371:660–667. [PubMed: 18295023]

13. 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; 138:463–468. [PubMed: 19818785]
14. Arnott ID, Landers CJ, Nimmo EJ, et al. Sero-reactivity to microbial components in Crohn's disease is associated with disease severity and progression, but not NOD2/CARD15 genotype. *Am J Gastroenterol*. 2004; 99:2376–2384. [PubMed: 15571586]
15. Mow WS, Vasilias EA, Lin YC, et al. Association of antibody responses to microbial antigens and complications of small bowel Crohn's disease. *Gastroenterology*. 2004; 126:414–424. [PubMed: 14762777]
16. Vasilias EA, Kam LY, Karp LC, et al. Marker antibody expression stratifies Crohn's disease into immunologically homogeneous subgroups with distinct clinical characteristics. *Gut*. 2000; 47:487–496. [PubMed: 10986208]
17. Targan SR, Landers CJ, Yang H, et al. Antibodies to CBir1 flagellin define a unique response that is associated independently with complicated Crohn's disease. *Gastroenterology*. 2005; 128:2020–2028. [PubMed: 15940634]
18. Lichtenstein GR. Emerging prognostic markers to determine Crohn's disease natural history and improve management strategies: a review of recent literature. *Gastroenterol Hepatol*. 2010; 6:99–107.
19. Dubinsky MC, Lin YC, Dutridge D, et al. Serum immune responses predict rapid disease progression among children with Crohn's disease: immune responses predict disease progression. *Am J Gastroenterol*. 2006; 101:360–367. [PubMed: 16454844]
20. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet*. 2008; 40:955–962. [PubMed: 18587394]
21. Devlin SM, Yang H, Ippoliti A, et al. NOD2 variants and antibody response to microbial antigens in Crohn's disease patients and their unaffected relatives. *Gastroenterology*. 2007; 132:576–586. [PubMed: 17258734]
22. Abreu MT, Taylor KD, Lin YC, et al. Mutations in *NOD2* are associated with fibrostenosing disease in patients with Crohn's disease. *Gastroenterology*. 2002; 123:679–688. [PubMed: 12198692]
23. Cho JH, Abraham C. Inflammatory bowel disease genetics: NOD2. *Annu Rev Med*. 2007; 58:401–416. [PubMed: 16987083]
24. Maeda S, Hsu LC, Liu H, et al. NOD2 mutation in Crohn's disease potentiates NF-kappaB activity and IL-1beta processing. *Science*. 2005; 307:734–738. [PubMed: 15692052]
25. Girardin SE, Boneca IG, Viala J, et al. NOD2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *J Biol Chem*. 2003; 278:8869–8872. [PubMed: 12527755]
26. Annese V, Lombardi G, Perri F, et al. Variants of CARD15 are associated with an aggressive clinical course of Crohn's Disease — an IG-IBD study. *Am J Gastroenterol*. 2005; 100:84–92. [PubMed: 15654786]
27. Helio T, Halme L, Lappalainen M, et al. CARD15/NOD2 gene variants are associated with familiarly occurring and complicated forms of Crohn's disease. *Gut*. 2003; 52:558–562. [PubMed: 12631669]
28. Ippoliti A, Devlin S, Mei L, et al. Combination of innate and adaptive immune alterations increased the likelihood of fibrostenosis in Crohn's disease. *Inflamm Bowel Dis*. 2010; 16:1279–1285. [PubMed: 20027650]
29. Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Can J Gastroenterol*. 2005; 19(suppl A):5–36.
30. Sutton CL, Kim J, Yamane A, et al. Identification of a novel bacterial sequence associated with Crohn's disease. *Gastroenterology*. 2000; 119:23–31. [PubMed: 10889151]
31. Sendid B, Colombel JF, Jacquinet PM, et al. Specific antibody response to oligomannosidic epitopes in Crohn's disease. *Clin Diagn Lab Immunol*. 1996; 3:219–226. [PubMed: 8991640]
32. Lodes MJ, Cong Y, Elson CO, et al. Bacterial flagellin is a dominant antigen in Crohn disease. *J Clin Invest*. 2004; 113:1296–1306. [PubMed: 15124021]

33. Cohavy O, Bruckner D, Gordon LK, et al. Colonic bacteria express an ulcerative colitis pANCA-related protein epitope. *Infect Immun*. 2000; 68:1542–1548. [PubMed: 10678972]
34. Vidrich A, Lee J, James E, et al. Segregation of pANCA antigenic recognition by DNase treatment of neutrophils: ulcerative colitis, type 1 autoimmune hepatitis, and primary sclerosing cholangitis. *J Clin Immunol*. 1995; 15:293–299. [PubMed: 8576315]
35. Harrell, FE, Jr.. *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer; 2001.
36. Dubinsky MC, Kugathasan S, Mei L, et al. Increased immune reactivity predicts aggressive complicating Crohn's disease in children. *Clin Gastroenterol Hepatol*. 2008; 6:1105–1111. [PubMed: 18619921]
37. 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. [PubMed: 12198693]
38. Desir B, Amre DK, Lu S-E, et al. Utility of serum antibodies in determining clinical course in pediatric Crohn's disease. *Clin Gastroenterol Hepatol*. 2004; 2:139–146. [PubMed: 15017619]
39. Rieder F, Schleder S, Wolf A, et al. Serum anti-glycan antibodies predict complicated Crohn's disease behavior: a cohort study. *Inflamm Bowel Dis*. 2010; 16:1367–1375. [PubMed: 20024902]

FIGURE 1.

(A) QSS distributions by complication status: complicated and noncomplicated disease. (B) The curves were generated by logistic regression modeling in the absence of the SNP13 risk allele. (C) The curves were generated by logistic regression modeling with the single SNP13 risk allele. (D) Calibration plot for the comparison of predicted and observed rates of complication by category (decile). Predictions were grouped into categories and compared to observed rates of complications for each category. Number of patients in each category prediction group were: 13 in the 10%–20% category; 29 in the 21%–30% category; 32 in the

31%–40% category; 32 in the 41%–50% category; 59 in the 51%–60% category; 40 in the 61%–70% category; 53 in the 71%–80% category; 65 in the 81%–90% category; 62 in the 91%–100% category. Correlation: 0.973 (E) ROC curve for cross-validation predictions. Probabilities were generated using a leave-one-out cross-validation to repeatedly generate a sero-genetic logistic regression (AUC = 0.801).

TABLE 1

Clinical Characteristics of the Crohn's Disease Cohort

Clinical Characteristics	Noncomplicated <i>n</i> =255	Complicated <i>n</i> =338
Gender	53% female	50% female
Average age at diagnosis	29 years (range 2–68)	24 years (range 2–63) ^a
Average age at blood draw	37 years (range 11–104)	39 years (range 12–91)
Average disease duration	8 years (range 1–59)	15 years (range 0–51) ^a
Disease behavior		
Complicated disease ^b	—	338 (57%)
Strictureing	—	175 (30%)
Penetrating	—	163 (27%)
Non complicated disease (inflammatory)	255 (43%)	—
Disease location		
Ileum	42 (17%)	97 (29%)
Colon	96 (38%)	21 (6%)
Ileum and colon	98 (39%)	176 (52%)
Upper gastrointestinal	17 (7%)	44 (13%)

^a*P* = 0.001 (chi-square test).

^bStrictureing or penetrating phenotypes are defined as complicated Crohn's disease.

TABLE 2

Expression of Serologic Markers in Patients

A. Percentage of complicated disease within each serologic marker quartile.									
Serologic Marker	Disease Behavior	Q1		Q2		Q3		Q4	
		n	% comp.	n	% comp.	n	% comp.	n	% comp.
ASCA-IgA	Complicated	53	36.1	74	49.7	100	67.6	111	74.5 ^a
	Noncomplicated	94		75		48		38	
ASCA-IgG	Complicated	57	38.5	75	50.7	97	65.5	109	73.2 ^a
	Noncomplicated	91		73		51		40	
Anti-CBir1	Complicated	65	44.5	81	54.4	91	61.1	101	67.8 ^a
	Noncomplicated	81		68		58		48	
Anti-OmpC	Complicated	55	37.9	81	54.0	88	59.5	114	76.0 ^a
	Non complicated	90		69		60		36	
Anti-I2	Complicated	64	43.2	82	55.4	101	68.2	91	61.1 ^a
	Noncomplicated	84		66		47		58	

B. Correlation of negative pANCA marker and incidence of complicated disease.

pANCA	Negative		Positive	
	n	% comp.	n	% comp.
Complicated	248	60.2 ^a	54	44.6
Non complicated	188		67	

^a *P* = 0.001 proportion of complicated patients increases with increasing quartile (Cochran-Armitage trend test).

^a *P* = 0.029 comparing complication rate for pANCA-negative vs. pANCA-positive patients (chi-square test).

TABLE 3

Percentage of Complicated and Noncomplicated Disease by *NOD2* Genotypes

Genotypes	Total <i>n</i> (%)	Noncomplicated		Complicated	
		<i>n</i>	%	<i>n</i>	%
No risk alleles	282 (73)	123	44	159	56
SNP8 heterozygotes	40 (10)	15	38	25	62
SNP12 heterozygotes	20 (5)	7	35	13	65
SNP13 heterozygotes	19 (5)	1	5	18	95 ^a
SNP8 homozygotes	3 (1)	0	0	3	100
SNP12 homozygotes	2 (1)	0	0	2	100
SNP13 homozygotes	4 (1)	0	0	4	100
SNP8 + SNP12 heterozygotes	3 (1)	0	0	3	100
SNP8 + SNP13 heterozygotes	5 (1)	0	0	5	100
SNP12 + SNP13 heterozygotes	5 (1)	0	0	5	100
SNP8 + SNP12 hetero/homozygotes	2 (1)	0	0	2	100

^a *P* < 0.001 comparing SNP13 heterozygotes with complicated disease vs. no risk alleles with or without complicated disease (Fisher's exact test).

TABLE 4

Sero-genetic Regression Model: The Risk of Complicated Crohn’s Disease

	Sero-genetic Regression Model (n=593)			
	Estimate	St. Error	z-value	P-value
Intercept	-3.606	0.446	-8.079	<0.001
Duration	0.070	0.011	6.222	<0.001
QSS	0.186	0.027	6.837	<0.001
SNP8 het	0.343	0.368	0.933	0.351
SNP12 het	0.308	0.525	0.587	0.557
SNP13 het	2.611	1.051	2.484	0.013
Two risk alleles	4.474	2.027	2.207	0.027

In the input variables listed, duration is in years (0–40), QSS is in points (6–24), and the SNP heterozygous variables (SNP8, SNP12, SNP13) are coded as binary variables (0–1; 0 is wildtype, 1 is heterozygous mutant). The variable “two risk alleles” is also a binary variable (0–1; 0 is one or fewer risk alleles across all three SNPs, 1 is two or more risk alleles across all three SNPs, whether compound heterozygotes or homozygous).