

# Phenotype-Stratified Genetic Linkage Study Demonstrates that *IBD2* Is an Extensive Ulcerative Colitis Locus

Jean-Paul Achkar, M.D.,<sup>1</sup> Themistocles Dassopoulos, M.D.,<sup>2</sup> Mark S. Silverberg, M.D., Ph.D.,<sup>3,4</sup> Jeffrey A. Tuvlin, M.D.,<sup>5</sup> Richard H. Duerr, M.D.,<sup>6,7</sup> Steven R. Brant, M.D.,<sup>2,8</sup> Kathy Siminovitch, M.D.,<sup>3,4,9,10</sup> Deepa Reddy, M.D.,<sup>5</sup> Lisa W. Datta, M.S.,<sup>2</sup> Theodore M. Bayless, M.D.,<sup>2</sup> Leilei Zhang,<sup>6</sup> M. Michael Barnada, Ph.D.,<sup>7</sup> John D. Rioux, Ph.D.,<sup>11</sup> A. Hillary Steinhart, M.D.,<sup>3,4</sup> Robin S. McLeod, M.D.,<sup>3</sup> Anne M. Griffiths, M.D.,<sup>12</sup> Zane Cohen, M.D.,<sup>3</sup> Huiying Yang, Ph.D.,<sup>13,14</sup> Gillian P. Bromfield, M.Sc.,<sup>5</sup> Phil Schumm, M.A.,<sup>15</sup> Stephen B. Hanauer, M.D.,<sup>5</sup> Judy H. Cho, M.D.,<sup>5</sup> Dan L. Nicolae, Ph.D.,<sup>16</sup> and the NIDDK IBD Genetics Consortium

<sup>1</sup>Center for Inflammatory Bowel Disease, Department of Gastroenterology, Cleveland Clinic Foundation, Cleveland, <sup>2</sup>The Harvey M. and Lyn P. Meyerhoff Inflammatory Bowel Disease Center, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, <sup>3</sup>Departments of Medicine and Surgery, University of Toronto, <sup>4</sup>Mount Sinai Hospital, Toronto, <sup>5</sup>The Martin Boyer Laboratories, Gastroenterology Section, Departments of Medicine, University of Chicago, Chicago, <sup>6</sup>Department of Medicine, School of Medicine and <sup>7</sup>Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, <sup>8</sup>Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, <sup>9</sup>Samuel Lunenfeld Research Institute, Mount Sinai Hospital, <sup>10</sup>Toronto General Hospital Research Institutes, Toronto, <sup>11</sup>Broad Institute of MIT and Harvard and Harvard Medical School, Brigham and Women's Hospital, Cambridge, <sup>12</sup>The Hospital for Sick Children, University of Toronto, <sup>13</sup>Division of Medical Genetics and Inflammatory Bowel Disease Research Center, Departments of Medicine and Pediatrics, Steven Spielberg Pediatric Research Center, Cedars-Sinai Medical Center, and <sup>14</sup>School of Medicine, University of California at Los Angeles, Los Angeles, <sup>15</sup>Health Studies, and <sup>16</sup>Statistics, University of Chicago, Chicago

- OBJECTIVES:** The complete elucidation of genetic variants that contribute to inflammatory bowel disease (IBD) will likely include variants that increase risk to both Crohn's disease and ulcerative colitis as well as variants that increase risk for particular phenotypic subsets. The purpose of this study was to assess phenotypic subsets that contribute to the major IBD susceptibility loci.
- METHODS:** This linkage study encompassed 904 affected relative pairs, representing the largest combined phenotyped cohort to date, and allowing for meaningful subset analyses. Genetic linkage data were stratified by disease location and age at diagnosis.
- RESULTS:** We establish that some loci, notably the *IBD3* and chromosome 3q linkage regions demonstrate contributions from both small intestine and colon cohorts, whereas others, notably the *IBD1* (*NOD2/CARD15*) and *IBD2* regions increase risk for small intestine or colon inflammation, respectively. The strongest linkage evidence in this study was for the subset of extensive ulcerative colitis in the region of *IBD2* (lod 3.27;  $p < 0.001$ ). Evidence for linkage in the region of *NOD2/CARD15* (*IBD1*) was stronger for the subset of Crohn's patients with ileal disease (lod 2.56;  $p = 0.035$ ) compared to the overall Crohn's group, consistent with previous findings that *NOD2/CARD15* variants are associated with ileal disease.
- CONCLUSIONS:** Analyses incorporating disease location in IBD increase the power and enhance the accuracy of genomic localization. Our data provide strong evidence that extensive ulcerative colitis represents a pathophysiologic subset of IBD.

(Am J Gastroenterol 2006;101:572-580)

## INTRODUCTION

The inflammatory bowel diseases (IBD), Crohn's disease (CD), and ulcerative colitis (UC) comprise multifactorial dis-

orders with a genetic component. Genetic studies in this field have been uniquely successful in identifying well-replicated disease gene associations which, in turn, are fundamentally advancing present concepts of disease pathogenesis (1, 2).

Gene associations at the *IBD1* (chromosome 16) and *IBD5* (chromosome 5) loci were initially implicated through genetic linkage approaches. Linkage evidence at the *IBD1* locus was observed primarily in pure CD-CD affected relative pairs (3, 4). Therefore, it was not surprising that the subsequent discovery of three major NOD2/CARD15 mutations in the region of *IBD1* were solely associated with CD cases, and not UC (5, 6). In addition, NOD2/CARD15 mutations were associated with CD cases involving the ileum, indicating that the minority of CD cases with inflammation confined to the colon represents a pathophysiologically distinct subset (7).

In contrast, the *IBD5* association at chromosome 5q31 has been observed in both CD and UC cohorts, suggesting a fundamentally different pathogenic role compared to the NOD2/CARD15 mutations (8–10). Recently, two polymorphisms within the *IBD5* region, Leu503Phe and -207G/C in the OCTN1/SLC22A4 (organic cation transporter) and OCTN2/SLC22A5 genes have been reported to alter transporter activity and expression, respectively (11). The genetic effects of these functional polymorphisms in adjacent, functionally similar genes cannot be easily separated (11). However, taken together, the present reports would suggest that some combination of altered intestinal OCTN function genetically increases risk to IBD, as opposed to CD or UC uniquely. Consistent with predictions from epidemiologic studies, a present genetic model of IBD would suggest that some susceptibility genes may increase susceptibility to IBD generally (e.g., OCTN associations), whereas others genes (e.g., NOD2/CARD15) increase risk for specific phenotypic subsets. The judicious application of phenotypic information to genetic linkage studies may assist in future gene identification efforts for some genomic regions and elucidation of disease pathogenesis.

While genetic linkage approaches have successfully resulted in identification of disease associations in IBD, linkage studies in general have significant limitations in complex genetic disorders. These limitations include the presence of likely false positive reports of genetic linkage, as well as the relative difficulty of pinpointing the precise gene/mutations that contribute to disease within large genomic regions. The problem of false positives can be addressed partially through strict statistical requirements for initial linkage and independent replication prior to designation of disease locus status (12). By these criteria, there is sufficient evidence for confirmed linkage to IBD in six regions: (1) linkage between CD and a locus on chromosome 16 (the *IBD1* locus), (2) linkage between IBD (especially UC) and a locus on chromosome 12q (the *IBD2* locus), (3) linkage between IBD (especially CD) and a locus on chromosome 6p (the *IBD3* locus), (4) linkage between CD and a locus on chromosome 14q (the *IBD4* locus), (5) linkage between IBD (especially CD) and a locus on chromosome 19p (the *IBD6* locus), and (6) linkage between IBD and chromosome 3p (1). In addition, suggestive linkage for a locus on chromosome 5q (*IBD5*) has been found in two studies with subsequent findings of association between CD and a common haplotype within this locus and identifica-

tion of functional polymorphisms in the OCTN1/SLC22A4 and OCTN2/SLC22A5 genes (8–11). However, even among these loci, the evidence for linkage has been inconsistently observed between individual linkage studies. This inconsistency may relate to numbers of families studied, as well as to pathogenic heterogeneity within and between cohorts.

To address these issues, we performed a phenotype-stratified genome-wide analysis of four large linkage cohorts comprising 904 affected relative pairs, representing the largest combined linkage analysis performed to date. Our hypothesis was that the application of phenotypic information to this genetic information would identify more homogeneous subgroups and thus increase the power and accuracy of localization in IBD genome scans. We stratified on phenotypes that have been shown to define pathophysiologic subsets, namely disease location and early age of onset.

## METHODS

### Genome Scan Datasets

The results of four previously completed genome scans were combined in order to perform this analysis. A brief summary of these genome scans is as follows:

1. The Mammalian Genotyping Service at the Marshfield Center for Medical Genetics completed two independent 10 cM genome scans on samples from IBD multiplex family DNA repositories. The first genome scan was performed on samples from IBD multiplex family DNA repositories at the University of Chicago and Johns Hopkins University and includes genotyping data for 297 affected relative pairs and their available connecting relatives (13). The second genome scan was performed on samples from the University of Pittsburgh, Johns Hopkins University, and the University of Chicago and includes genotyping data at the same genetic markers for a total of 260 affected relative pairs (14).
2. Investigators at the Mount Sinai Hospital in Toronto and the Broad Institute (formerly Whitehead Institute/MIT Center for Genome Research) performed a 12 cM genome scan on 158 families with two or more IBD-affected siblings with DNA available for both parents in 140 of these families (15). The scan contains genotyping data for 312 loci in a total of 183 affected sibling pairs.
3. Investigators at the University of Pittsburgh and the Mammalian Genotyping Service at the Marshfield Center for Medical Genetics performed a 4.6 cM genome scan on 62 families containing 127 CD-affected relative pairs (16).

All DNA samples were collected with full approval from each participating institution's research ethics board. All data shared for this mega-analysis was done so with full approval under the original or alternate approved protocol.

Markers genotyped in the different genome scans were integrated onto a single map that utilized the deCode genetic map as its backbone (17). The markers that were not found on the deCode map were located in the Marshfield or the

physical maps from the current versions of the human genome assembly maps. Their genetic map location was interpolated using the genetic and physical map coordinates of the two closest flanking markers. No markers were deleted. Genotype data from all sources were merged into a single database, but the marker allele frequencies were calculated separately for all the centers involved in the study.

### Phenotypic Data

Phenotypic criteria were defined *a priori* and determined by consensus of the Phenotype Subcommittee of the NIDDK IBD Genetics Consortium.

*Age at diagnosis* was defined as the time of first endoscopic, radiographic, histologic, or surgical documentation of IBD.

*Sites of intestinal disease* were determined by review of primary endoscopy, histology, radiology, and/or surgical records. Nonspecific findings such as mild erythema with no other changes on endoscopy or nonspecific inflammation on histology were not considered as evidence of disease in that location. Crohn's disease can affect any part of the gastrointestinal tract and therefore disease locations were coded as proximal (esophagus, stomach, or duodenum), jejunum, ileum, colon, and perianal (perianal fistula or abscess or anal canal ulcers). For the purposes of linkage analysis, disease location was further classified as "any ileal" or "any colon" if there was disease in the ileum or colon, respectively, regardless of whether there was or was not evidence of disease elsewhere in the gastrointestinal tract. In UC, inflammation is confined to the colon and always starts in the rectum but with variable degrees of extension proximally. Therefore, for UC, disease locations were grouped based on maximal macroscopic extent of disease: extensive (maximal macroscopic disease extending proximal to the splenic flexure), left-sided (maximal macroscopic disease reaching up to or distal to the splenic flexure), or proctitis (macroscopic disease confined to the rectum).

### Statistical Analyses

Statistical analyses for differences in the age of diagnosis between two groups were done using the Wilcoxon rank sum test (18). Differences in the frequency of location or extent were tested using  $\chi^2$  statistics on two-way tables. The significance of the Wilcoxon rank sum and  $\chi^2$  were assessed using simulations. Each simulated dataset was obtained by assigning permuted labels to each family.

Linkage analyses were performed with the program ALLEGRO using affecteds-only allele-sharing methods (19). The evidence for linkage was measured by the excess of alleles shared identical-by-descent by affected individuals within families. We used the exponential allele-sharing model on the scoring function Spairs (20) to generate the lod scores that are directly related to the nonparametric linkage scores but provide more accurate results. The identity-by-descent status was determined using multipoint calculations. Each data set was analyzed separately; the multipoint probabilities were saved and combined to calculate the likelihood at each marker location.

The information on disease location was incorporated in the analyses by defining new phenotypes. For example, in the any ileal CD analysis, we defined the affected individuals to be those having CD with inflammation located in the ileum. To determine whether there was a significant effect of grouping of IBD patients on the basis of disease location on the evidence for linkage, we randomly permuted the clinical information among affected individuals over 1,000 simulations. The number of affected relative pairs within each simulation was kept constant. Multipoint linkage analyses were performed for each simulated dataset and the results were saved for the 1-lod drop support region of the position with the maximum lod score. Empirical *p* values were determined from these simulated data. These *p* values are adjusted for the number of markers in the tested region, but they are not adjusted for the number of regions tested. Similarly, the differences in lod scores between the Jewish and non-Jewish families were assessed by permuting the ethnicity labels of the families in 1,000 simulated datasets.

For age at diagnosis analyses, families were weighted on the basis of the affected individuals' age at diagnosis (the goal being to up-weight families with affecteds having young age of diagnosis). Genome-wide analysis was performed using a scheme that assigned weights inversely proportional to the square root of the median age of diagnosis for each family. We use the median age of diagnosis because it is more robust than the mean age of diagnosis.

## RESULTS

Tables 1 and 2 show data for age of diagnosis and disease location separated by the different centers and by Jewish ethnicity for CD (Table 1) and for UC (Table 2). The age at diagnosis was similar across centers, and median age at diagnosis was lower in CD for patients of Jewish ethnic background (19 yr for Jewish vs 22 yr for non-Jewish;  $p < 0.0001$ ). In addition, a modestly higher fraction of Jewish CD patients had extensive ileocolonic disease compared to non-Jewish CD.

The overall CD location distribution was within the range of the most commonly reported distributions in the literature (21, 22). In contrast, there was an increase in the fraction of extensive UC (56%) compared to population-based estimates of approximately 35% (23, 24). In addition, for CD, age greater than the median age at diagnosis (22 yr) was associated with a higher frequency of colonic disease compared to age at diagnosis less than 22 yr (20% vs 10%;  $p = 0.008$ ). The heterogeneity in disease location among the centers did not affect the interpretation of the linkage results. The distribution of numbers of affected relative pairs by disease location is shown in Table 3.

Overall linkage results of the combined genome scans are shown in Figure 1 for IBD (Fig. 1A), for CD (Fig. 1B), and for UC (Fig. 1C). In this overall analysis, only one locus achieved a lod score of greater than three. This linkage was seen for the IBD phenotype on chromosome 3q near D3S2436 (peak lod = 3.10), with contributions from pure CD and from

**Table 1.** Demographic Features and Disease Location for Crohn’s Disease

	Disease Location—N (%) of Individuals				Median Age at dx (yr)
	Ileal	Colonic	Ileal-colonic	Other	
Chicago	44 (29%)	20 (13%)	88 (57%)	2 (1%)	22
Johns Hopkins	44 (26%)	31 (19%)	90 (54%)	2 (1%)	22
Pittsburgh-CD	70 (49%)	19 (13%)	54 (38%)	0	22
Pittsburgh	36 (50%)	8 (11%)	28 (39%)	0	23
Toronto	90 (33%)	47 (17%)	124 (46%)	10 (4%)	22
Jewish	75 (31%)	39 (16%)	130 (53%)	1 (0%)	19*
Non-Jewish	201 (37%)	84 (16%)	244 (45%)	13 (2%)	22*
Age at diagnosis <22	144 (36%)	41 (10%)	203 (51%)	10 (3%)	–
Age at diagnosis >22	134 (36%)	77 (20%)	163 (43%)	4 (1%)	–

\*  $p < 0.0001$  for Jewish versus Non-Jewish age at diagnosis.

mixed (CD and UC) affected relative pairs (Fig. 1A). The second highest linkage score for the IBD phenotype was on chromosome 2q near D2S1369 (peak lod = 2.49).

For the overall CD cohort (Fig. 1B), the three most significant linkage peaks were observed at *IBD1* (peak lod 1.65), *IBD3* (peak lod 2.08), and chromosome 3q (peak lod 1.8). For the *IBD1* and *IBD3* loci, the CD subset demonstrated modestly greater evidence for linkage compared to the entire IBD cohort, whereas the CD peak for the chromosome 3q peak was less than for all IBD. At *IBD1*, the any ileal CD subset was associated with both an increase in the peak lod score to 2.56 as well as a shift in the peak location to D16S3136 (51 cM), immediately adjacent to the *NOD2/CARD15* gene (Fig. 2A). The significance of the increase in the lod score was assessed by simulations with a simulated  $p$  value of 0.035. In addition, the 1-lod drop support interval for the linkage peak was 38-cM wide in the overall CD group, but only 10-cM wide in the any ileal CD group. The *IBD1* region demonstrated the greatest evidence for linkage in the genome for the any ileal CD cohort, with the region on chromosome 3q demonstrating the second highest evidence for linkage (data not shown). In contrast, the evidence for linkage at *IBD3* and also chromosome 3q was observed for both the any ileal and any colonic CD subsets (Fig. 2B and D). The *IBD5* locus did not demonstrate significant evidence for linkage in all CD,

but some evidence for linkage was seen for colonic CD with a peak lod score of 1.71 ( $p = 0.052$ ) (Fig. 2C).

In the overall UC cohort (Fig. 1C), the greatest evidence for linkage was observed at D2S335 (lod = 2.35). Of note, the *IBD2* region, which has been reported as being primarily a UC cohort, did not demonstrate evidence for linkage in the overall UC cohort. Because extensive UC can be considered as a more severe phenotype, we performed linkage analysis of this phenotype separately and found a peak lod score of 3.27 for extensive UC at marker GATA91H06 at *IBD2* (Fig. 3). This large increase in the peak lod score of the extensive UC subset is highly significant, being very unlikely to have occurred by chance. In 1,000 simulations, randomly assigning the extensive UC diagnosis to 56% of all UC cases never resulted in a lod score greater than 3.27 (simulated  $p < 0.001$ ). Furthermore, the lod score was not driven by one or two large pedigrees as 18 out of the 26 families included in the extensive disease group had positive nonparametric lod scores. Finally, the 1-lod drop support interval for the linkage peak was from D12S345 (55.25 cM on the deCode map) to D12S85 (60.49 cM on the deCode map), importantly refining the localization of the UC risk alleles in this region. In contrast to these findings we found no evidence for linkage at D2S335 for extensive UC (data not shown).

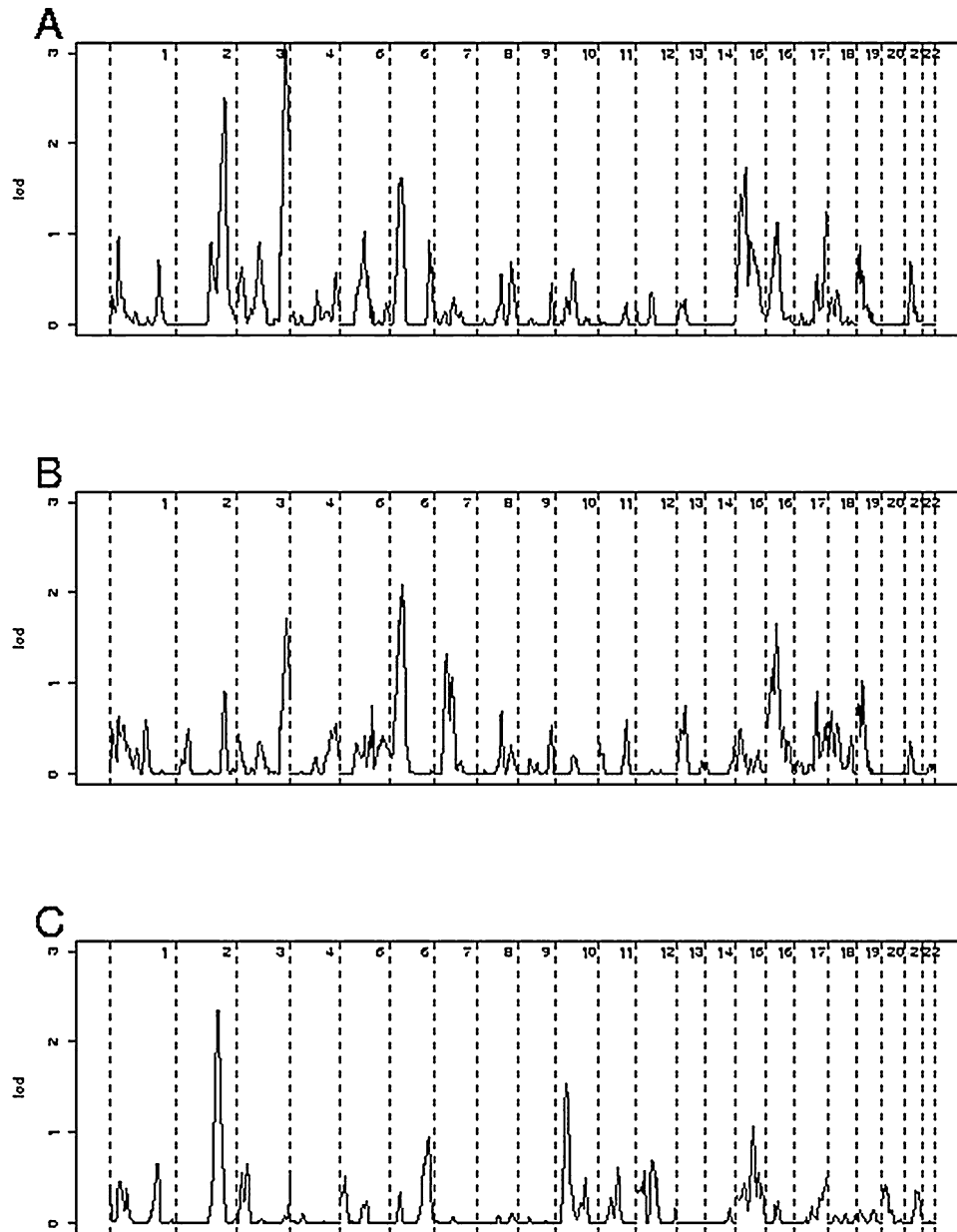
Finally, we stratified the genome scan results by age at diagnosis. There was a slight increase in the peak lod score at *IBD1* for “any ileal” disease when stratified by age (increase from 2.55 to 2.98). However, age at diagnosis did not otherwise significantly affect the linkage results.

**Table 2.** Demographic Features and Disease Location for Ulcerative Colitis

	Disease Location—N (%) of Individuals			Median Age at dx (yr)
	Proctitis	Left sided	Extensive	
Chicago	5 (9%)	14 (25%)	37 (66%)	24
Johns Hopkins	4 (24%)	6 (35%)	7 (41%)	24.5
Pittsburgh	14 (17%)	16 (19%)	53 (64%)	25
Toronto	30 (36%)	15 (18%)	38 (46%)	27.5
Jewish	12 (22%)	14 (26%)	28 (52%)	24
Non-Jewish	41 (22%)	37 (20%)	106 (58%)	25
Age at diagnosis <25	19 (17%)	26 (24%)	64 (59%)	–
Age at diagnosis >25	30 (26%)	22 (19%)	62 (55%)	–

**Table 3.** Number of Affected Relative Pairs by Disease Location

Phenotype	# Affected Relative Pairs
Crohn’s disease	557
Any ileal	374
Any colon	199
Perianal	53
Ulcerative colitis	138
Extensive	29
Left-sided	5
Proctitis	11
Mixed	209

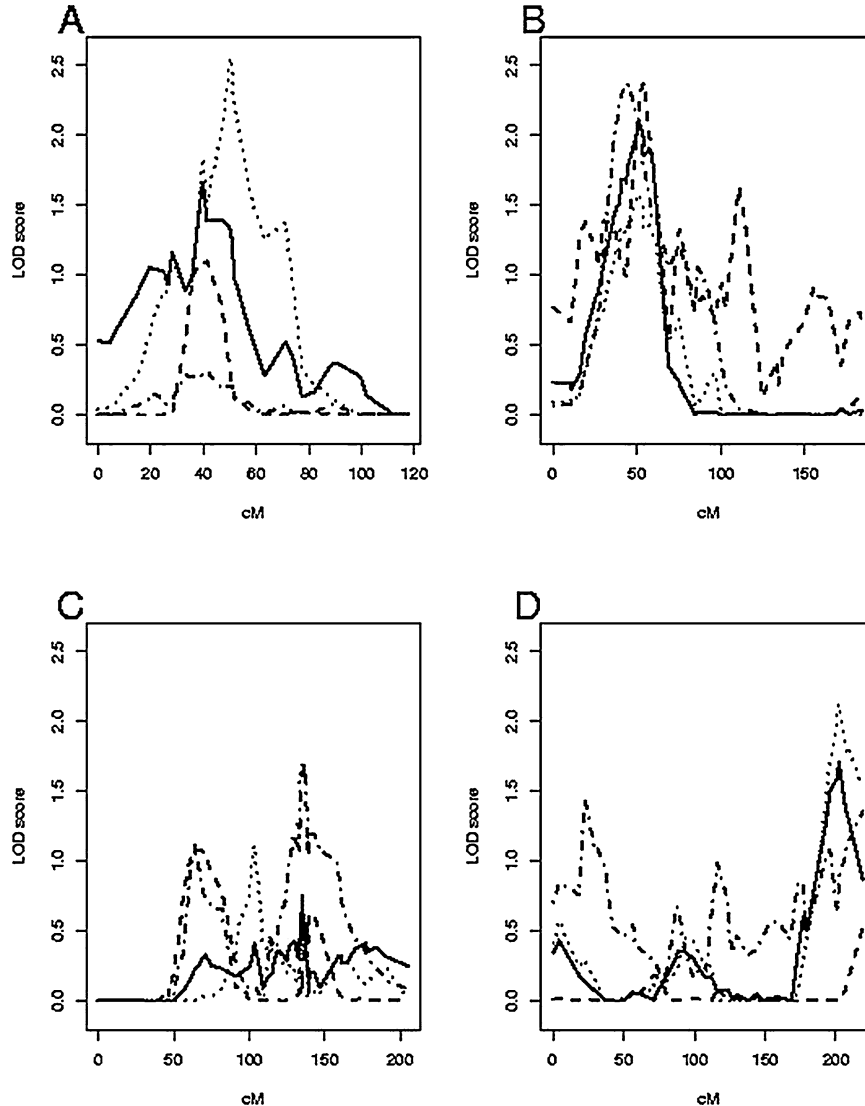


**Figure 1.** Genomewide multipoint nonparametric lod scores are plotted along the length of the entire genome for the IBD phenotype (A), CD phenotype (B), and UC phenotype (C). The lod score is on the y-axis and the distance from the p telomere in centimorgans (cM) is on the x-axis. Vertical dashed lines separate the 22 autosomal chromosomes.

## DISCUSSION

The complete elucidation of genetic variants that contribute to IBD will likely include variants that increase risk generally to both CD and UC, as well as variants that increase risk for a particular phenotypic subset. Application of phenotypic information may assist in gene identification for those regions without established disease gene associations, and provides insight into mechanisms of disease pathogenesis for established, disease-associated variants. The present study represents the first application of phenotype information to a large linkage cohort, thereby providing a broad overview

of the phenotypic subsets that contribute to the major IBD susceptibility loci. We have established that some loci, notably the *IBD3* and chromosome 3q linkage regions, demonstrate contributions from both small intestine and colonic disease location cohorts, whereas others, notably the *IBD1* (NOD2/CARD15) and *IBD2* regions, increase risk for either small intestinal disease location (*IBD1*) or extensive colonic UC (*IBD2*). The present findings have implications for finding and confirming risk alleles especially in the *IBD2* region. These findings would indicate that pathogenic mechanisms for risk alleles at *IBD2* should specifically confer risk for intestinal inflammation throughout the colon.



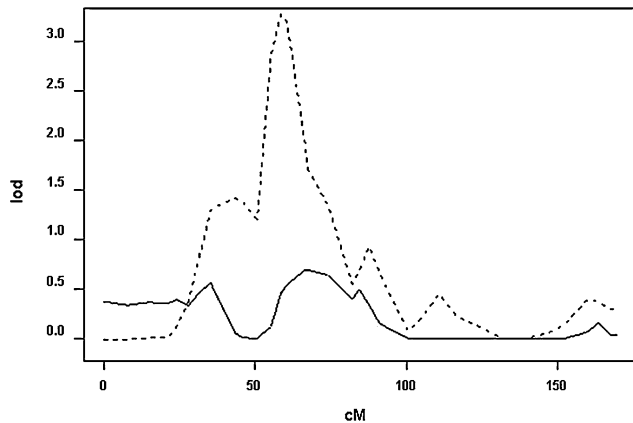
**Figure 2.** Multipoint nonparametric lod score curves for the Crohn's disease (solid curves) and for three phenotypic subsets: "any ileal CD" (dotted curves), "perianal CD" (dashed curves), and "any colon" (dot-dashed curves). (A) Chromosome 16. (B) Chromosome 6. (C) Chromosome 5. (D) Chromosome 3. For all the plots, the lod score is on the y-axis and the distance from the p telomere in centimorgans (cM) is on the x-axis.

The *IBD3* and chromosome 3q linkage regions demonstrate evidence for linkage in both the small intestine and colonic CD subsets. These findings could result either from the risk alleles in these regions increasing risk for IBD generically or could be due to multiple, distinct risk alleles conferring risk for different phenotypic subsets in these general genomic regions. This latter possibility is most likely for the *IBD3* linkage region, given the enormous genetic and immunologic complexity here. For example, the uncommon DRB1\*0103 allele is associated with extensive UC and colonic CD (25, 26). In contrast, the HLA-DRB1\*0701 allele within *IBD3* confers risk for ileal disease specifically (25, 27). These two risk alleles represent only an illustrative subset of the likely genotype-phenotype associations occurring in this region.

Our findings of a significantly greater linkage score in "any ileal" CD compared to all CD cases are consistent with prior

NOD2/CARD15 mutation genotype-phenotype studies and validate our approach of applying disease location for linkage stratification purposes (27, 28). An important part of recent studies with NOD2/CARD15 variants has involved accounting for this small intestinal association. For example, CD and ileitis demonstrate a Th1 cytokine profile in both human and mouse models so it has been hypothesized that the NOD2/CARD15 mutations contribute to Th1 skewing (29–32). In addition, the high expression of NOD2/CARD15 in Paneth cells, granular cells located in the base of small intestinal crypts, could also contribute to the small intestinal association (33–35). Recently, NOD2/CARD15 deficiency has been demonstrated to be required for the expression of a subset of cryptidins, antimicrobial peptides produced by Paneth cells (36).

We did not analyze CD disease behavior in this study because of the concern for interobserver variation and because



**Figure 3.** Multipoint nonparametric lod score curves for chromosome 12 for UC (solid curve) and “extensive UC” (dotted curve). The lod score is on the y-axis and the distance from the *p* telomere in centimorgans (cM) is on the x-axis.

disease behavior has been shown to evolve over time (37–39). Attempting to correct for evolution of disease behavior would greatly limit the linkage analysis. In addition, the data on the association of NOD2/CARD15 genotypes and disease behavior have been somewhat conflicting (7, 27, 28), suggesting that this phenotype may not have as strong of a genetic basis as age at diagnosis and disease location. For these reasons, we did not believe that integrating disease behavior in the context of our linkage study would have been of value.

The contribution of genetic factors in UC broadly speaking is less significant compared to CD as demonstrated by the fact that monozygotic twin concordance is lower in UC than in CD. The relatively smaller numbers of UC-UC affected relative pairs in cohorts ascertained for linkage studies also support this general concept. This paucity of UC-UC relative pairs in prior *individual* linkage studies accounts for the relative lack of significant UC linkage peaks in a recently reported meta-analysis, underscoring the importance of pooling primary genotype data as we did in the present study (40). We found that for the overall UC cohort, the greatest evidence for linkage was observed at D2S335 (lod = 2.35), a region that also scored the highest for UC in the recent meta-analysis (40). Of note, *IBD2*, the only linkage region which has been implicated primarily in UC, being initially reported in a linkage cohort uniquely containing a robust number of UC-UC affected relative pairs (41, 42), did not demonstrate evidence for linkage in the overall UC cohort. However, we found that *IBD2* is specifically linked to the extensive UC phenotype, a highly significant finding. In 1,000 simulations, random assignment of the extensive UC status to 57% of the UC cohort never resulted in a lod score approaching the observed 3.27. Furthermore, the lod score was not driven by one or two large pedigrees as 18 out of the 26 families included in the extensive disease group had positive nonparametric lod scores. This latter point is particularly important as it demonstrates that these results were not driven by any one center’s data or by a limited number of families. Support for con-

sidering extensive UC as a separate phenotype comes from prior studies demonstrating that patients with extensive disease have a more aggressive disease course, are more likely to require surgery, and are more likely to develop pouchitis postoperatively (43, 44). The present findings will assist in localizing and identifying UC risk alleles in the *IBD2* region. Furthermore, statistically associated UC disease susceptibility alleles here would be predicted to have functional effects contributing to more extensive colonic inflammation.

## ACKNOWLEDGMENTS

We express our profound thanks to all of the participating family members and many colleagues of the NIDDK IBD Genetics Consortium (IBDGC) who took part in the sample collection and data preparation. We would also like to acknowledge NIH/NHLBI Mammalian Genotyping Service, Marshfield Center for Medical Genetics (R.H.D. and J.H.C.). Thanks also to Joanne Stempak (University of Toronto), Christina Julian (University of Chicago), Mark Daly (Broad Institute) for his ongoing contributions to this consortium, Jerome I. Rotter (Cedar Sinai Medical Center, Los Angeles), Yin Yao Shugart (Johns Hopkins University Bloomberg School of Public Health), Amir S. Karban (Rambam Hospital, Haifa, Israel), Carolien I.M. Panhuysen (Boston University), Joan E. Bailey-Wilson (NHGRI), and Mary L. Harris (Johns Hopkins University).

This work was supported by grants from NIDDK U01 DK062422, DK062429 (J.H.C.), U01 DK062420 (R.H.D.), U01 DK62423 (M.S.S.), U01 DK62431 (S.R.B.), U01 DK62413 (H.Y.), RO1DK55731 (J.H.C.), NIH R01 DK 55889 (D.L.N), Burroughs Wellcome (J.H.C.), Crohn’s and Colitis Foundation of America (R.H.D., J.H.C.), Crohn’s and Colitis Foundation of Canada (M.S.S., K.S.), the Canadian Institute for Health Research (K.S.), the Gastrointestinal Research Foundation (J.H.C.), Scaife Family Foundation (R.H.D.), NIH General Clinical Research Center grant RR00052 (S.R.B.) and RR00055 (J.H.C.), and Broad Medical Research Program of the Eli and Edythe L. Broad Foundation (J.H.C.).

## STUDY HIGHLIGHTS

### What Is Current Knowledge

- Ulcerative colitis and Crohn’s disease are complex genetic diseases with multiple susceptibility loci, and genetic and phenotypic heterogeneity, complicating gene identification efforts.
- Several studies have demonstrated an association between NOD2/CARD15 mutations and ileal disease location in Crohn’s disease suggesting that IBD-susceptibility genes predispose to particular disease phenotypes.

**What is New Here**

- The present study represents the application of phenotype information to a large linkage cohort and thereby provides a broad overview of the phenotypic subsets that contribute to the major IBD susceptibility loci.
- Some loci, notably the *IBD3* and chromosome 3q linkage regions, demonstrate contributions from both small intestine and colonic disease location cohorts, whereas others, notably the *IBD1* (*NOD2/CARD15*) and *IBD2* regions, increase risk for either small intestinal disease location (*IBD1*) or extensive colonic disease (*IBD2*) respectively.
- The finding that *IBD2* is specifically linked to the extensive ulcerative colitis phenotype is highly significant and should assist in localizing and identifying ulcerative colitis risk alleles in the *IBD2* region.

**Reprint requests and correspondence:** Dr. Dan L. Nicolae, Department of Statistics, The University of Chicago, 5734 South University Avenue, Chicago, IL 60637.

Received June 9, 2005; accepted October 11, 2005.

**REFERENCES**

1. Bonen DK, Cho JH. The genetics of inflammatory bowel disease. *Gastroenterology* 2003;124:521–36.
2. Ahmad T, Tamboli CP, Jewell D, et al. Clinical relevance of advances in genetics and pharmacogenetics of IBD. *Gastroenterology* 2004;126:1533–49.
3. Hugot JP, Laurent-Puig P, Gower-Rousseau C, et al. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 1996;379:821–3.
4. Ohmen JD, Yang HY, Yamamoto KK, et al. Susceptibility locus for inflammatory bowel disease on chromosome 16 has a role in Crohn's disease, but not in ulcerative colitis. *Hum Mol Genet* 1996;5:1679–83.
5. Hugot JP, Chamaillard M, Zouali H, et al. Association of *NOD2* leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;411:599–603.
6. Ogura Y, Bonen DK, Inohara N, et al. A frameshift mutation in *NOD2* associated with susceptibility to Crohn's disease. *Nature* 2001;411:603–6.
7. Lesage S, Zouali H, Cezard JP, et al. *CARD15/NOD2* mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* 2002;70:845–57.
8. Rioux JD, Daly MJ, Silverberg MS, et al. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 2001;29:223–8.
9. Gialourakis C, Stoll M, Miller K, et al. *IBD5* is a general risk factor for inflammatory bowel disease: Replication of association with Crohn disease and identification of a novel association with ulcerative colitis. *Am J Hum Genet* 2003;73:205–11.
10. McGovern DP, Van Heel DA, Negoro K, et al. Further evidence of *IBD5/CARD15 (NOD2)* epistasis in the susceptibility to ulcerative colitis. *Am J Hum Genet* 2003;73:1465–6.
11. Peltekova VD, Wintle RF, Rubin LA, et al. Functional variants of *OCTN* cation transporter genes are associated with Crohn disease. *Nat Genet* 2004;36:471–5.
12. Lander E, Kruglyak L. Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nat Genet* 1995;11:241–7.
13. Cho JH, Nicolae DL, Gold LH, et al. Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: Evidence for epistasis between 1p and *IBD1*. *Proc Natl Acad Sci USA* 1998;95:7502–7.
14. Barmada MM, Brant SR, Nicolae DL, et al. A genome scan in 260 inflammatory bowel disease-affected relative pairs. *Inflamm Bowel Dis* 2004;10:15–22.
15. Rioux JD, Silverberg MS, Daly MJ, et al. Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am J Hum Genet* 2000;66:1863–70.
16. Duerr RH, Barmada MM, Zhang L, et al. High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11–12. *Am J Hum Genet* 2000;66:1857–62.
17. Kong A, Gudbjartsson DF, Sainz J, et al. A high-resolution recombination map of the human genome. *Nat Genet* 2002;31(3):241–7.
18. Hollander M, Wolfe D. Nonparametric statistical inference. New York: John Wiley & Sons, 1973.
19. Gudbjartsson DF, Jonasson K, Frigge ML, et al. Allegro, a new computer program for multipoint linkage analysis. *Nat Genet* 2000;25:12–3.
20. Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 1997;61:1179–88.
21. Farmer RG, Hawk WA, Turnbull WA. Clinical patterns in Crohn's disease: A statistical study of 615 cases. *Gastroenterology* 1975;68:627–35.
22. Veloso FT, Ferreira JT, Barros L, et al. Clinical outcome of Crohn's disease: Analysis according to the Vienna classification and clinical activity. *Inflamm Bowel Dis* 2001;7:306–13.
23. Leijonmarck CE, Persson PG, Hellers G. Factors affecting colectomy rate in ulcerative colitis: An epidemiologic study. *Gut* 1990;31:329–33.
24. Moum B, Ekbohm A, Vatn MH, et al. Change in the extent of colonoscopic and histological involvement in ulcerative colitis over time. *Am J Gastroenterol* 1999;94:1564–9.
25. Newman B, Silverberg M, Gu X, et al. *CARD15* and HLA *DRB1* alleles influence susceptibility and disease localization in Crohn's disease. *Am J Gastroenterol* 2004;99:306–15.
26. Roussomoustakaki M, Satsangi J, Welsh K, et al. Genetic markers may predict disease behavior in patients with ulcerative colitis. *Gastroenterology* 1997;112:1845–53.
27. Ahmad T, Armuzzi A, Bunce M, et al. The molecular classification of the clinical manifestations of Crohn's disease. *Gastroenterology* 2002;122:854–66.
28. Brant SR, Picco MF, Achkar JP, et al. Defining complex contributions of *NOD2/CARD15* gene mutations, age at onset, and tobacco use on Crohn's disease phenotypes. *Inflamm Bowel Dis* 2003;9:281–9.
29. Fuss IJ, Neurath M, Boirivant M, et al. Disparate CD4+ lamina propria (LP) lymphokine secretion profiles in inflammatory bowel disease. Crohn's disease LP cells manifest increased secretion of IFN-gamma, whereas ulcerative colitis LP cells manifest increased secretion of IL-5. *J Immunol* 1996;157:1261–70.
30. Plevy SE, Landers CJ, Prehn J, et al. A role for TNF- $\alpha$  and mucosal T helper-1 cytokines in the pathogenesis of Crohn's disease. *J Immunol* 1997;159:6276–82.
31. Kontoyiannis D, Pasparakis M, Pizarro TT, et al. Impaired on/off regulation of TNF biosynthesis in mice lacking TNF

- AU- rich elements: Implications for joint and gut-associated immunopathologies. *Immunity* 1999;10:387–98.
32. Kosiewicz MM, Nast CC, Krishnan A, et al. Th1-type responses mediate spontaneous ileitis in a novel murine model of Crohn's disease. *J Clin Invest* 2001;107:695–702.
  33. Ogura Y, Lala S, Xin W, et al. Expression of NOD2 in Paneth cells: A possible link to Crohn's ileitis. *Gut* 2003;52:1591–7.
  34. Lala S, Ogura Y, Osborne C, et al. Crohn's disease and the NOD2 gene: A role for paneth cells. *Gastroenterology* 2003;125:47–57.
  35. Wehkamp J, Harder J, Weichenthal M, et al. NOD2 (CARD15) mutations in Crohn's disease are associated with diminished mucosal alpha-defensin expression. *Gut* 2004;53:1658–64.
  36. Kobayashi KS, Chamaillard M, Ogura Y, et al. Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* 2005;307:731–4.
  37. Steinhart AH, Girgrah N, McLeod RS. Reliability of a Crohn's disease clinical classification scheme based on disease behavior. *Inflamm Bowel Dis* 1998;4:228–34.
  38. Louis E, Collard A, Oger AF, et al. Behaviour of Crohn's disease according to the Vienna classification: Changing pattern over the course of disease. *Gut* 2001;49:777–82.
  39. Cosnes J, Cattan S, Blain A, et al. Long-term evolution of disease behavior of Crohn's disease. *Inflamm Bowel Dis* 2002;8:244–50.
  40. Van Heel DA, Udalova IA, De Silva AP, et al. Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. *Hum Mol Genet* 2004;13:763–70.
  41. Parkes M, Barnada MM, Satsangi J, et al. The IBD2 locus shows linkage heterogeneity between ulcerative colitis and Crohn disease. *Am J Hum Genet* 2000;67:1605–10.
  42. Satsangi J, Parkes M, Louis E, et al. Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat Genet* 1996;14:99–202.
  43. Farmer RG, Easley KA, Rankin GB. Clinical patterns, natural history, and progression of ulcerative colitis. A long-term follow-up of 1116 patients. *Dig Dis Sci* 1993;38:1137–46.
  44. Achkar JP, Al-Haddad M, Lashner B, et al. Differentiating risk factors for acute and chronic pouchitis. *Clin Gastroenterol Hepatol* 2005;3:60–6.