

Report

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

Cosmas Giallourakis,^{1,2} Monika Stoll,³ Katie Miller,¹ Jochen Hampe,³ Eric S. Lander,¹ Mark J. Daly,¹ Stefan Schreiber,³ and John D. Rioux¹

¹Center for Genome Research, Whitehead Institute, Cambridge, MA; ²Gastrointestinal Unit, Massachusetts General Hospital, Boston; and ³Department of Internal Medicine, Christian-Albrechts-University, Kiel, Germany

Inflammatory bowel disease (IBD) refers to complex chronic relapsing autoimmune disorders of the gastrointestinal tract that have been traditionally classified into Crohn disease (CD) and ulcerative colitis (UC). We have previously reported that genetic variation within a 250-kb haplotype (IBD5) in the 5q31 cytokine gene cluster confers susceptibility to CD in a Canadian population. In the current study, we first replicated this association by examining 368 German trios with CD and demonstrating, by transmission/disequilibrium testing (TDT), that the same haplotype is associated with CD ($\chi^2 = 5.97$; $P = .007$). Our original association study focused on the role of IBD5 in CD; we next explored the potential contribution of this locus to UC susceptibility in 187 German trios. Given the TDT results in the present cohort with UC, IBD5 may also act as a susceptibility locus for UC ($\chi^2 = 8.10$; $P = .002$). We then examined locus-locus interactions between IBD5 and *CARD15*, a locus reported elsewhere to confer risk exclusively to CD. Our current results indicate that the two loci act independently to confer risk to CD but that these two loci may behave in an epistatic fashion to promote the development of UC. Moreover, IBD5 was not associated with particular clinical manifestations upon phenotypic stratification in the current cohort with CD. Taken together, our results suggest that IBD5 may act as a general risk factor for IBD, with loci such as *CARD15* modifying the clinical characteristics of disease.

Inflammatory bowel disease (IBD) refers to a spectrum of chronic relapsing autoimmune disorders, primarily affecting the gastrointestinal tract, that have been traditionally classified into Crohn disease (CD) and ulcerative colitis (UC). This classification scheme is based on clinical, endoscopic, radiologic, and pathologic criteria, including the distribution of inflammation in the colon and/or small intestine. Patchy transmural inflammation involving any portion of the gastrointestinal tract—but most often affecting the colon and/or the terminal portion of the small intestine, termed the “ileum”—is consistent with CD. In contrast, UC is restricted to the colon, characteristically starting in the rectum and then spread-

ing proximally in a continuous fashion with mucosal ulcerations (Feldman and Sleisenger 1998). Significant variation is observed within the presentation of either phenotypic category, which suggests that multiple subtypes exist. Furthermore, it is clear that the current classification scheme of CD versus UC is insufficient for the subgroup of patients classified as having “indeterminate colitis.” These patients have such a degree of overlap between CD and UC, in terms of their clinical presentation, that it is difficult to confidently assign them to either category, which suggests that IBD represents a continuum of manifestations. Epidemiologic studies have identified a significant genetic contribution to the etiology of IBD. It is important that even though there are distinct phenotypic differences between CD and UC, studies show that relatives of persons with either CD or UC are at increased risk for developing either form of IBD (Bonon and Cho 2003). This suggests that, although there are phenotype-specific susceptibility loci, some genes will be shared by patients with CD and those with UC. The iden-

Received March 5, 2003; accepted for publication April 10, 2003; electronically published May 29, 2003.

Address for correspondence and reprints: Dr. John D. Rioux, Center for Genome Research, Whitehead Institute, One Kendall Square, Building 300, Cambridge, MA 02139. E-mail: rioux@genome.wi.mit.edu

© 2003 by The American Society of Human Genetics. All rights reserved. 0002-9297/2003/7301-0019\$15.00

Table 1**Frequency of Clinical Characteristics in the German Trios with UC and CD**

CHARACTERISTIC	No. (%) OF PATIENTS WITH			
	CD		UC	
	<i>CARD15</i> ^{nonrisk} (<i>n</i> = 205) ^a	<i>CARD15</i> ^{risk} (<i>n</i> = 163) ^b	<i>CARD15</i> ^{nonrisk} (<i>n</i> = 152) ^c	<i>CARD15</i> ^{risk} (<i>n</i> = 35) ^d
Sex (male/female)	58/147	50/113	61/91	18/17
Any history of smoking	103 (50.2)	69 (42.3)	51 (33.6)	9 (25.7)
L colonic ^e	71 (34.6)	32 (19.6)	52 (34.2)	15 (42.9)
R colonic ^f	123 (60.0)	115 (70.6)	17 (11.2)	0 (0)
Colonic localization ^g	41 (20.0)	26 (16.0)	83 (54.6)	20 (57.1)
Resection	102 (49.8)	94 (57.7)	14 (9.2)	5 (14.3)
Fistulization	109 (53.2)	94 (57.7)	7 (4.6)	1 (2.9)
Stenosis	117 (57.1)	100 (61.3)	15 (9.9)	5 (14.3)
Ileal	143 (69.8)	130 (79.8)	9 (5.9)	1 (2.9)

^a Median age at onset 24 years.

^b Median age at onset 22 years.

^c Median age at onset 26 years.

^d Median age at onset 23 years.

^e L colonic = radiological or endoscopic evidence of disease distal to the splenic flexure (descending colon, sigmoid colon, and rectum).

^f R colonic = radiological or endoscopic evidence of disease proximal to splenic flexure (transverse colon, ascending colon, and cecum).

^g Denotes the frequency of individuals who were known to have colonic disease but had endoscopic or radiological procedures done >24 mo from the time of entry into the study or, if repeated procedures were done, there was not complete agreement as to the anatomical localization and the extent of inflammation.

tification of multiple susceptibility loci, coupled with an understanding of how these loci interact, may aid in the development of a molecular taxonomy of IBD to improve treatment and outcome predictions in the disease.

We have reported elsewhere that genetic variation within a common 250-kb haplotype in the 5q31 cytokine gene cluster confers susceptibility to CD (Daly et al. 2001; Rioux et al. 2001). It is estimated that two-thirds of individuals of European-derived ancestry who have CD harbor at least one copy of this risk haplotype. This locus, known as “IBD5,” (MIM 606348) was first identified in a genomewide linkage scan of multiple families with affected sibling pairs from a Canadian population. The locus was of genomewide significance in families with early-onset CD (maximum LOD score 3.9) and of more modest effect when all families with IBD were combined in the analysis (maximum LOD score 2.4) (Rioux et al. 2000). As described in our previous study, a hierarchical linkage disequilibrium (LD)–mapping strategy was employed to search for the causal genetic variation within this 18-cM linked region. Ultimately, this approach led to the discovery of the haplotype structure of the common variation within this locus. Specifically, this region could be parsed into haplotype blocks that were separated by intervals of recurrent recombination events. Moreover, limited diversity within these haplotype blocks (2–4 haplotypes per block accounting for 90%–98% of all chromosomes) was observed. Multilocus analysis and simulations defined a disease risk haplotype (*IBD5*^{risk}) that extended across multiple haplotype blocks (a 250-kb region), with significant association with CD. Eleven

haplotype-tagging SNPs (htSNPs) were identified that had alleles that were unique to this risk haplotype. All 11 SNPs were significantly associated with CD (all *P* < .00023) and had essentially identical information content (as a result of the very strong LD on this risk haplotype) (Daly et al. 2001; Rioux et al. 2001). Any 1 of these 11 SNPs can, therefore, act as a proxy for this risk haplotype for the purposes of association testing (Johnson et al. 2001). Indeed, examination of the htSNPs in a limited number of trios with CD, from an independent Canadian population, confirmed the initial observation of association (Rioux et al. 2001).

In an effort to provide comprehensive evidence of the replication finding for *IBD5* and to gain insights into potential phenotype-genotype relationships, we undertook a study of 555 German trios with IBD. Specifically, we examined a retrospective clinical cohort consisting of 368 patients with CD, 187 patients with UC, and their unaffected parents, who had been recruited through the German Crohn’s and Colitis Foundation from 1996 to 1998 (table 1). The majority of the patients with CD (306) have been described elsewhere, but 62 patients with CD have not (Hampe et al. 2002). The clinical characteristics of the trios with UC have not been previously described.

Prior to commencing this replication study using a single htSNP, we set out to confirm that the haplotype structure in this German cohort is equivalent to that previously reported for the Canadian population. Specifically, 15 SNPs were genotyped in 120 German trios with CD: 4 htSNPs and 5 non-htSNPs (that identified both risk and nonrisk haplotypes in the 250-kb core

region of association) and 6 additional SNPs to examine the structure flanking this core region (3 centromeric SNPs located in the IL-4 region and 3 telomeric SNPs surrounding IL-3). As expected, given the European ancestry of the two populations, we observed an essentially identical haplotype structure in the German chromosomes, as compared with the original Canadian data. There was approximately the same frequency of haplotype variants within each block between the two populations and identical block-to-block correlations (data not shown; analyzed as described in Daly et al. [2001]). As a result, we were able to select a single one of these htSNPs and to type it in the remaining IBD samples (98.6% of the samples typed successfully). This htSNP (known as “IGR2096a_1”) was first identified in our resequencing and typing of the Canadian samples; it uniquely tags the 250-kb risk haplotype and is located between the *IRF-1* and *OCTN2* genes (Rioux et al. 2001).

We first examined the IGR2096a_1 genotype data for evidence of overtransmission of the “A” allele, representing the risk haplotype (IBD5^{risk}) in the 368 trios with CD. Using the transmission/disequilibrium test (TDT) implemented in GENEHUNTER (Daly et al. 1998), we found that this risk allele was significantly overtransmitted from heterozygous parents to their children with CD (192:147, transmitted:untransmitted [T:U]) giving a χ^2 of 5.97 ($P = .007$; table 2). In the original association study of the IBD5-risk haplotype, we had observed a T:U ratio of 2.4:1 in the trios with CD. In the smaller replication set from Quebec, reported in the same study, we observed a ratio of 1.75:1, and, in the current replication set, the German trios with CD have a more modest ratio of 1.3:1. All of these observations are statistically compatible with an intermediate ratio of 1.75:1, and the differences between these studies may be explained by a number of factors, including clinical and genetic heterogeneity of disease, as well as statistical fluctuation (Glazier et al. 2002). Analysis of the German cohort thus provides clear replication of the association of IBD5 with disease susceptibility for CD.

IBD5 was first identified as a CD-susceptibility locus in our linkage study of 158 families with IBD and, subsequently, in an association study that exclusively examined trios with CD (Rioux et al. 2000, 2001; Daly et al. 2001). It should be noted, however, that the linkage study consisted of 124 affected sibling pairs (ASPs) with CD and 20 ASPs with UC, potentially limiting our ability to detect linkage to UC. Considering the increased power of association studies over linkage studies for common alleles (Risch and Merikangas 1996), we wanted to examine the current cohort to determine if there was evidence of an association of IBD5 with UC. Testing of the IGR2096a_1 htSNP in the German trios with UC demonstrated that the IBD5^{risk} allele was significantly over-

Table 2

Association of the IBD5^{risk} Haplotype in the German Cohort with IBD

Phenotype	TDT (T:U) ^a	χ^2	<i>P</i>	Group Comparison ^b
IBD	292:212	12.70	.0002	NS
CD:	192:147	5.97	.0073	
<i>CARD15</i> ^{risk}	80:61	2.56	.055	NS
<i>CARD15</i> ^{nonrisk}	112:86	3.41	.032	
UC:	98:62	8.10	.0022	
<i>CARD15</i> ^{risk}	24:6	10.80	.0005	<.02
<i>CARD15</i> ^{nonrisk}	74:56	2.49	.057	

^a TDT analysis performed for the IBD5 htSNP known as “IGR2096a_1” for the A allele denoting the risk haplotype allele (IBD5^{risk}; T = number of transmitted chromosomes; U = number of untransmitted chromosomes). This htSNP was genotyped in 368 trios with CD and 187 trios with UC. Trios were stratified into groups on the basis of the *CARD15* status and IBD phenotype of the affected child in each trio. *CARD15*^{risk} denotes having at least one mutant *CARD15* allele of SNP 8, 12, or 13. *CARD15*^{nonrisk} denotes being wild type at all three variant positions.

^b No statistical difference (NS) in the T:U ratios (i.e., strength of association) of the IBD5^{risk} haplotype observed between UC and CD, as determined by a 2 × 2 contingency table. There was no significant difference (NS) in terms of the association of IBD5^{risk} haplotype with CD on stratification into *CARD15*^{risk} and *CARD15*^{nonrisk} subgroups by permutation testing, but a significant difference between these two subgroups was observed in trios with UC ($P < .02$).

transmitted from heterozygous parents to affected children, T:U ratio 98:62 ($\chi^2 = 8.10$; $P = .002$) (table 2). It is interesting that this result suggests, that, in fact, IBD5^{risk} may act as susceptibility allele not only for CD, but also for UC. Furthermore, there is no statistical difference in the TDT ratios of the IBD5^{risk} haplotype between the CD and UC trios, when compared by a contingency table ($P = .32$), which indicates that the level of association of IBD5^{risk} haplotype between CD and UC is equivalent in this cohort.

Given the strong evidence of association of IBD5 to CD and UC, we were interested in examining potential locus-locus interactions with *CARD15* (MIM 605956), the other well-characterized CD locus (Hugot et al. 2001; Ogura et al. 2001). *CARD15* (also known as “*NOD2*”) is believed to function in innate immunity-signaling cascades involved in pathogen-associated molecular-recognition pathways. Three polymorphisms in *CARD15*, which occurred independently on a common haplotype, have been associated with susceptibility to CD. These mutations, designated “SNP8” (R702W), “SNP12” (G908R), and “SNP13” (3020insC), yield two nonsynonymous coding changes and a frameshift truncation mutation, respectively. These three genetic variants appear to confer roughly equivalent risk, such that the genotype relative risk for heterozygotes for any one of these mutations ranges from ~2-fold to 6-fold and rises to >20-fold in homozygotes or compound heterozygotes.

The allele frequencies of these three SNPs ranges ~4%–14% in the cohorts with CD who have been examined to date (~30%–40% of individuals with CD are heterozygous for at least one of these three variants) compared with 1%–4% in control subjects and populations with UC (Hampe et al. 2001; Abreu et al. 2002; Ahmad et al. 2002; Cuthbert et al. 2002; Lesage et al. 2002; Vermeire et al. 2002).

Interactions between *IBD5* and *CARD15* were, therefore, examined in the current cohort with IBD by stratifying trios into two groups on the basis of the genotypes of the affected child. The *CARD15*^{risk} group consisted of trios with an affected child who carried at least one copy of any of the three *CARD15* mutations (R702W, G908R, and/or 3020insC), whereas the *CARD15*^{nonrisk} group was composed of trios with an affected child bearing the wild-type allele at all three variants. In trios with CD (as seen in table 2), the *IBD5*^{risk} haplotype was overtransmitted in the *CARD15*^{risk} group (80:61; $\chi^2 = 2.56$; $P = .054$) and in the *CARD15*^{nonrisk} group (112:86; $\chi^2 = 3.41$; $P = .032$). To assess whether there was a meaningful difference in the level of association of the *IBD5*^{risk} haplotype in CD on stratification by *CARD15* genotype, we performed permutation testing by randomly assigning the *CARD15* status to affected individuals in trios. There was no statistical difference in the association of the *IBD5*^{risk} haplotype upon stratification between the *CARD15*^{risk} and *CARD15*^{nonrisk} groups, which strongly suggests that the two loci act in an independent fashion. The finding that *IBD5* and *CARD15* act independently to confer disease susceptibility is consistent with our recent results in a smaller sample of trios with CD from Quebec (Vermeire et al. 2002).

We next stratified trios with UC into *CARD15*^{risk} and *CARD15*^{nonrisk} groups, as above. Of interest, it appeared that the majority of the association signal of *IBD5* arose from individuals who carried the *CARD15*^{risk} alleles (table 2). The TDT ratio for *IBD5*^{risk} in the *CARD15*^{risk} group was 24:6 ($\chi^2 = 10.80$; $P = .0005$), compared with 74:56 ($\chi^2 = 2.49$; $P = .057$) for the *CARD15*^{nonrisk} group. As we had done for CD, we next examined whether a significant difference in association at *IBD5* in trios with UC existed on stratification on *CARD15* alleles by permutation testing. In contrast to the case of CD, permutation testing demonstrated that, indeed, there was a significant difference in association of *IBD5*, on the basis of *CARD15* stratification in the cohort with UC ($P < .02$) (table 2). One interpretation of this result is that *CARD15*^{risk} alleles may act in a synergistic fashion with *IBD5* to promote the development of UC, which suggests a locus-locus interaction.

Having convincingly demonstrated a role for *IBD5* in IBD susceptibility in this population, we next examined the role of this locus in conferring risk to intermediate clinical phenotypes, such as disease behavior (age at on-

set, fistulization, stenosis, and resection) and location (ileal, right colonic, and left colonic) in the presence and absence of *CARD15* alleles. In terms of anatomical landmarks, the colon can be thought of as an inverted U-shaped organ and can be divided into right (proximal) and left (distal), corresponding roughly to the two arms of the U. The terminal portion of the small intestine (ileum) is attached to the colon at its most proximal end, with the colon having a number of anatomically defined subdivisions, moving from right to left: ascending colon→hepatic flexure→transverse colon→splenic flexure→descending colon→sigmoid colon→rectum. Disease localization was included in our analysis only if an endoscopic or radiological study had been performed within 24 mo of the time of entry into the study. If a procedure was done >24 mo before, or if repeated procedures were not in complete agreement in terms of the anatomical localization and the extent of inflammation, the patient was excluded from this part of the multivariate analyses of genotype-phenotype correlations. Eighteen percent of individuals affected with CD and the majority of patients with UC (55%) did not meet our stringent criteria regarding knowledge of stable disease localization (table 1). The latter is not surprising, considering estimates, which suggest that the extent of disease fluctuates in 20%–60% of patients with UC (Farmer et al. 1993; Moum et al. 1996). As a result, formal analyses of genotype-phenotype correlations were performed only in trios with CD. However, given the difference in association of *IBD5* based on *CARD15* stratification in UC, we wanted to examine, in at least a qualitative fashion, whether the group of individuals who comprise the *IBD5*^{risk}/*CARD15*^{risk} group had some unique clinical characteristics that correlated with their genotype status. Only 12 individuals in this subgroup were identified as having stable localization, but it was surprising to find that all 12 had inflammation restricted to the left colon. This observation highlights the need, in general, to examine the role of *IBD5* in a larger number of individuals with UC.

Given the more stable nature of disease localization in CD, a more formal analysis of these patients was possible. Univariate and multivariate analyses demonstrated that *IBD5* did not correlate with any clinical subphenotype of CD we examined, including age at onset, fistulization, stenosis, resection, and disease location (ileal, right colonic, and left colonic) (table 3 and data not shown). The lack of correlation of *IBD5* with any clinical subphenotype that we examined in CD may, therefore, suggest that it is a general risk factor for disease. In contrast, these same analyses revealed a novel correlation between *CARD15* status and disease location. Specifically, multivariate analysis followed by stepwise regression revealed that individuals with CD who carry *CARD15*^{risk} alleles have a decreased chance of having

Table 3**Multivariate Analysis for Disease Localization in Patients with CD**

Genotype Class ^a	Ileum OR ^b (95% CI)	Left Colon OR (95% CI)	Right Colon OR (95% CI)
IBD5 ^{risk} (<i>n</i> = 234)	1.884 (.764–4.645)	1.133 (.601–2.137)	.753 (.348–1.630)
CARD15 ^{risk} (<i>n</i> = 146)	1.130 (.657–1.942)	.345 ^c (.183–.651)	1.425 (.718–2.826)
IBD5 ^{risk} /CARD15 ^{nonrisk} (<i>n</i> = 116)	1.421 (.596–3.387)	2.954 ^c (1.567–5.568)	.644 (.318–1.304)
IBD5 ^{risk} /CARD15 ^{risk} (<i>n</i> = 105)	1.057 (.415–2.992)	.426 ^d (.217–.838)	1.392 (.661–2.935)
IBD5 ^{non-risk} / CARD15 ^{risk} (<i>n</i> = 11)	.772 (.091–6.539)	1.267 (.270–5.945)	.513 (.106–2.471)
IBD5 ^{non-risk} /CARD15 ^{nonrisk} (<i>n</i> = 49)	.479 (.144–1.596)	1.011 (.437–2.336)	1.574 (.535–4.627)

^a IBD5^{risk} denotes having at least one copy of the risk haplotype, as assessed by the IGR2096a_1 htSNP. CARD15^{risk} denotes having at least one mutant CARD15 allele of SNP 8, 12, or 13. CARD15^{nonrisk} denotes being wild type at all three variant positions.

^b The multivariate analysis with logistic regression treats the genotype of the affected individuals as the dependent variable and the three disease localization phenotypes as three independent dichotomous variables. The OR that was calculated compares the relative risk of individuals having, versus not having, the particular disease localization.

^c P ≤ .001.

^d P ≤ .05.

left colonic disease (odds ratio [OR] = 0.345; 95% CI 0.183–0.651; *P* < .001) independent of IBD5 status (see table 3). Conversely, individuals who are CARD15^{nonrisk} have an increased chance of having left-sided disease (OR = 2.89; 95% CI 1.537–5.454; *P* < .001). This effect of CARD15 on localization is independent of IBD5 status. This observation regarding left-sided disease is also supported by univariate analysis (CARD15^{nonrisk}) (OR = 2.21; 95% CI 1.34–3.64; *P* = .002).

Most of the studies to date on genotype-phenotype correlation related to CARD15 have not reported a multivariate analysis where the geographic site of disease within the colon was analyzed as variable but, rather, looked at the colonic phenotype in toto as a single independent variable (Gasche et al. 2000; Hampe et al. 2001; Abreu et al. 2002; Ahmad et al. 2002; Cuthbert et al. 2002; Vermeire et al. 2002). However, our finding is in agreement with the report of Lesage et al., who reported that individuals with CARD15^{risk} have less frequent left-colon involvement at disease onset (Lesage et al. 2002). If, indeed, the effect of CARD15^{risk} alleles is to protect against the development of left-sided colonic disease while contributing to the development of gastrointestinal inflammation, this may be reflected in the previously perceived association of CARD15^{risk} alleles with ileal-right-sided disease. In fact, the study by Hampe et al. reported an ileal association with CARD15 in a retrospective cohort of 447 individuals (of which 306 are included in the current analysis), by use of an expectation maximization algorithm (Hampe et al. 2002). A statistically significant association of CARD15^{risk} alleles with ileal disease was observed only in these same 447 individuals in the univariate analysis (OR = 2.2; CI 0.724–3.867; *P* = .002). The association of CARD15^{risk} with ileal disease did not stand up to testing after multivariate analysis followed by stepwise regression, but the novel effect of CARD15 on left-sided colonic disease

remained significant. It should be noted, however, that even though our evidence is stronger for CARD15 having an impact on colonic disease localization than for its having an impact on ileal localization, either of these observed correlations to specific intermediate disease phenotypes might simply be surrogates for an as-yet-unidentified disease subgroup.

With the current study, we have now demonstrated in two Canadian populations and a large German cohort that genetic variation on the IBD5^{risk} haplotype plays an important role in the susceptibility to IBD. As expected for a complex trait, IBD is believed to be the result of a set of potentially interacting genetic susceptibility loci and environmental factors that together influence the clinical characteristics of the patient. The replication of IBD5, combined with the knowledge of CARD15 genotypes in the current cohort, enabled an exploration of potential locus-locus interactions and correlation of the loci with patient clinical characteristics. Of interest, our results suggest that IBD5 acts independent of CARD15 to confer susceptibility to CD, while appearing to behave as a general risk factor in the sense that it was not associated with any particular clinical intermediate phenotype that we examined in the present cohort with CD. On the other hand, CARD15, independent of IBD5, appears to be associated with a clinical subgroup of patients (i.e., those without left-colonic disease). In contrast to the results in the trios with CD where there was little evidence of a locus-locus interaction, we uncovered a potential genetic interaction between IBD5 and CARD15 in trios with UC. Although this study provides potential insights into IBD5 and CARD15, in terms of locus-locus interactions and their impact on disease subphenotypes, it is clear that additional work will be necessary to completely characterize the interactions and their influence on disease phenotype. Furthermore, identification of additional IBD loci and

characterization of their interactions with IBD5 and *CARD15* should enable a more precise molecular classification of patients with IBD. Such a classification scheme, in conjunction with traditional clinical data, has the potential for identifying patient subgroups that are more related to disease mechanism and, therefore, may be more powerful in diagnosis and treatment than the current clinical classification schemes.

Acknowledgments

The authors thank the members of the Inflammatory Disease Research Group, Whitehead Institute/MIT Center for Genome Research, for their helpful comments. Thanks to Peter Croucher for his helpful discussions.

Electronic-Database Information

The URL for data presented herein is as follows:

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim/> (for IBD5 and *CARD15*)

References

- Abreu MT, Taylor KD, Lin YC, Hang T, Gaiennie J, Landers CJ, Vasiliauskas EA, Kam LY, Rojany M, Papadakis KA, Rotter JI, Targan SR, Yang H. (2002) Mutations in NOD2 are associated with fibrostenosing disease in patients with Crohn's disease. *Gastroenterology* 123:679–688
- Ahmad T, Armuzzi A, Bunce M, Mulcahy-Hawes K, Marshall SE, Orchard TR, Crawshaw J, Large O, de Silva A, Cook JT, Barnardo M, Cullen S, Welsh KI, Jewell DP (2002) The molecular classification of the clinical manifestations of Crohn's disease. *Gastroenterology* 122:854–866
- Bonen DK, Cho JH (2003) The genetics of inflammatory bowel disease. *Gastroenterology* 124:521–536
- Cuthbert AP, Fisher SA, Mirza MM, King K, Hampe J, Croucher PJ, Mascheretti S, Sanderson J, Forbe A, Mansfield J, Schreiber S, Lewis CM, Mathew CG (2002) The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease. *Gastroenterology* 122:867–874
- Daly MJ, Kruglyak L, Pratt S, Houstis N, Reeve MP, Kirby K, Lander ES (1998) GENEHUNTER 2.0—a complete linkage analysis system. *Am J Hum Genet* 63:S286
- Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander, ES (2001) High-resolution haplotype structure in the human genome. *Nat Genet* 29:229–232
- Farmer RG, Easley KA, Rankin GB (1993) Clinical patterns, natural history, and progression of ulcerative colitis: a long-term follow-up of 1116 patients. *Dig Dis Sci* 38:1137–1146
- Feldman M, Sleisenger M (1998) *Gastrointestinal and liver disease*, 6th ed. Vol 2, W. B. Saunders, Philadelphia, pp 1708–1759
- Gasche C, Scholmerich J, Brynkvov J, D'Haens G, Hanauer SB, Irvine EJ, Jewell DP, Rachmilewitz D, Sachar DB, Sandborn WJ, Sutherland LR (2000) A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. *Inflamm Bowel Dis* 6:8–15
- Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. *Science* 298:2345–2349
- Hampe J, Cuthbert A, Croucher PJ, Mirza MM, Mascheretti S, Fisher S, Frenzel H, King K, Hasselmeyer A, MacPherson AJ, Bridger S, van Deventer S, Forbes A, Nikolaus S, Leonard-Jones JE, Foelsch UR, Krawczak M, Lewis C, Schreiber S, Mathew CG (2001) Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 357:1925–1928
- Hampe J, Grebe J, Nikolaus S, Solberg C, Croucher PJ, Mascheretti S, Jahnsen J, Moum B, Klump B, Krawczak M, Mirza MM, Foelsch UR, Vatn M, Schreiber S (2002) Association of NOD2 (*CARD 15*) genotype with clinical course of Crohn's disease: a cohort study. *Lancet* 359:1661–1665
- Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411:599–603
- Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA (2001) Haplotype tagging for the identification of common disease genes. *Nat Genet* 29:233–237
- Lesage S, Zouali H, Cezard J-P and EPWG-IBD Group, Colombel J-F and EPIMAD Group, Belaiche J and GETAID Group, Almer S, Tysk C, O'Morain C, Gassull M, Binder V, Finkel Y, Modigliani R, Gower-Rousseau C, Macry J, Merlin F, Chamaillard M, Jannot A-S, Thomas G, Hugot J-P (2002) *CARD15/NOD2* mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* 70:845–857
- Moum B, Vatn MH, Ekbohm A, Aadland E, Fausa O, Lygren I, Sauar J, Schulz T, Stray N. (1996) Incidence of ulcerative colitis and indeterminate colitis in four counties of south-eastern Norway, 1990–93: a prospective population-based study. The Inflammatory Bowel South-Eastern Norway (IBSEN) Study Group of Gastroenterologists. *Scand J Gastroenterol* 31:362–366
- Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH, Achkar JP, Brant SR, Bayless TM, Kirschner BS, Hanauer SB, Nunez G, Cho JH (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 411:603–606
- Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z, Delmonte T, et al (2001) Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 29:223–228
- Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, McLeod RS, Griffiths AM, Green T, Brettin TS, Stone V, Bull SB, Bitton A, Williams CN, Greenberg GR, Cohen Z, Lander ES, Hudson TJ, Siminovitch KA (2000) Genomewide search

- in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am J Hum Genet* 66: 1863–1870
- Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. *Science* 273:1516–1517
- Vermeire S, Wild G, Kocher K, Cousineau J, Dufresne L, Bitton A, Langelier D, Pare P, Lapointe G, Cohen A, Daly MJ, Rioux JD (2002) *CARD15* genetic variation in a Quebec population: prevalence, genotype-phenotype relationship, and haplotype structure. *Am J Hum Genet* 71:74–83