Am. J. Hum. Genet. 71:74–83, 2002

CARD15 Genetic Variation in a Quebec Population: Prevalence, Genotype-Phenotype Relationship, and Haplotype Structure

Severine Vermeire,[1] Gary Wild,[1] Kerry Kocher,[2] Josee Cousineau,[1] Line Dufresne,[1] Alain Bitton,[1] Diane Langelier,[3] Pierre Pare,[4] Gilles Lapointe,[5] Albert Cohen,[1] Mark J. Daly,[2] and John D. Rioux[2]

[1]Department of Gastroenterology, McGill University Health Centre, McGill University, Montreal; [2]Whitehead Institute/Massachusetts Institute of Technology, Center for Genome Research, Cambridge, MA; [3]Department of Gastroenterology, Centre Hospitalier de Sherbrooke, Sherbrooke, Canada; [4]Department of Gastroenterology, Centre Hospitalier Universitaire de Quebec, Hopital l'Hotel-Dieu de Quebec, Quebec, Canada; and [5]Department of Gastroenterology, Centre Hospitalier de la Sagamie, Chicoutimi, Quebec, Canada

The caspase recruitment domain gene (*CARD15*) was recently identified as the underlying gene associated with the *IBD1* locus that confers susceptibility to Crohn disease (CD). *CARD15* is related to the NOD1/Apaf-1 family of apoptosis regulators, and three sequence variants (Arg702Trp, Gly908Arg, and Leu1007fsinsC) in the gene were demonstrated to be associated with CD. We collected a cohort of 231 patients with CD and 71 healthy control individuals from the Canadian province of Quebec, to determine the prevalence of these sequence variants in an independent population. Clinical records of all patients were systematically reviewed, and detailed phenotypic information was obtained. All patient DNA samples were genotyped for the three variants, thus enabling an analysis of genotype-phenotype correlations. In this cohort, 45.0% of patients with CD carried at least one variant in the *CARD15* gene, compared with 9.0% of control individuals ($P < 10^{-7}$). Allele frequencies of Arg702Trp, Gly908Arg, and Leu1007fsinsC were 12.9%, 5.2%, and 10.3% in patients with CD, compared with 4.2%, 0.7%, and 0.7% in control individuals, respectively. Importantly, *CARD15* mutants were seen with equal frequency in patients with familial and sporadic CD. Analysis of the relationship between genotype and phenotype convincingly demonstrates that *CARD15* variants are significantly associated with ileal disease involvement, as opposed to strictly colonic disease ($P < .001$). Moreover, we were able to determine the haplotype structure surrounding this disease gene by genotyping 45 single-nucleotide polymorphisms (SNPs) in a 177-kb region that contained the *CARD15* gene. This structure helps clarify the history of these causal mutations. Finally, this analysis shows that *CARD15* involvement with CD is detectable by use of publicly available SNPs alone.

Introduction

Crohn disease (CD [MIM 266600]) is a chronic, relapsing inflammatory bowel disease (IBD) of unknown etiology. Estimates of prevalence in the northern half of the western hemisphere are 40–250/100,000 individuals in the population (Andres and Friedman 1999). In the northern hemisphere, a north-south prevalence gradient exists, with the highest reported rates in Scandinavia and Canada (Pinchbeck et al. 1988; Ekbom et al. 1991; Bernstein et al. 1999). Multiple lines of evidence suggest that the underlying etiology of IBD is a dysregulated immune response to microbial organisms in the gastrointestinal tract and/or a breakdown in the epithelial barrier, which leads to inflammatory cell damage. Specifically, this appears to involve a Th1-predominant (i.e., interleukin-2, interferon-γ, and tumor necrosis factor–α) proinflammatory mucosal immune response to the presence of normal intestinal bacterial flora (Elson 2002; Shanahan 2002). A strong genetic contribution to susceptibility to CD is supported by a relative risk to siblings of affected individuals (λ_s) of 25 (Ahmad et al. 2001) and a concordance rate among MZ twins of ~40%.

The complexity of the genetic susceptibility in CD is underscored by the identification, by genomewide scanning, of >10 putative susceptibility regions with suggestive or significant evidence of linkage (Hugot et al. 1996; Satsangi et al. 1996; Cho et al. 1998; Hampe et al. 1999; Ma et al. 1999; Duerr et al. 2000; Rioux et al. 2000). The first identified susceptibility locus was the pericentromeric region on chromosome 16 (*IBD1*), which conferred susceptibility to CD only (Hugot et al. 1996). The importance of this locus has been confirmed by various groups, including the International IBD Genetics Consortium (2001).

The combined strategies of positional cloning and can-

Received February 25, 2002; accepted for publication April 8, 2002; electronically published May 17, 2002.

Address for correspondence and reprints: Dr. John D. Rioux, Human Medical and Population Genetics, Whitehead Institute/MIT Center for Genome Research, One Kendall Square, Building 300, Cambridge, MA 02139-1561. E-mail: rioux@genome.wi.mit.edu

0002-9297/2002/7101-0009$15.00

didate gene analysis recently led two independent teams to identify the gene that underlies the *IBD1* locus (Hugot et al. 2001; Ogura et al. 2001*a*). *IBD1* encodes the caspase recruitment domain gene protein (CARD15 [previously known as "NOD2"; MIM 605956]), a member of the NOD1/Apaf-1 family of apoptosis regulators (Inohara et al. 2001; Ogura et al. 2001*b*). CARD15 is expressed intracellularly in monocytes and macrophages, where it appears to function as a sensor for bacterial products, such as lipopolysaccharides (LPS), through the leucine-rich repeat (LRR) domain located in the C-terminal region of the gene. It has been hypothesized that the sensing of bacterial components would, under normal circumstances, result in the activation of nuclear factor–κB (NF-κB) and in apoptosis.

Hugot and coworkers (2001) identified a number of single-nucleotide polymorphisms (SNPs) within the *CARD15* gene and demonstrated that three SNPs, including one frameshift mutation (Leu1007fsinsC) and two missense mutations (Arg702Trp and Gly908Arg), were associated with CD. To date, functional data are available only for the frameshift mutation. The presence of Leu1007fsinsC leads to the truncation of the C-terminal 33 amino acids in the LRR region. An intact LRR region appears to be necessary for normal bacterial/LPS recognition, since partial deletion of the LRR, an event characteristic of the frameshift mutation, was shown to be associated with hyporesponsiveness to LPS (Ogura et al. 2001*a*). This deficit in sensing bacteria would then result in an exaggerated inflammatory response of the adaptive immune system; this is supported by the fivefold increase in NF-κB activation seen in the presence of the frameshift mutation. Although functional data are lacking for the other variants, it is tempting to speculate that these three variants would alter the structure of the LRR region, resulting in similar abnormalities in bacterial recognition.

These recent discoveries provide a rational framework in which to examine genotype-phenotype correlations in a complex human disease. To this end, we collected a cohort of 231 well-characterized patients with CD from the province of Quebec, Canada. Because the prevalence of the Leu1007fsinsC variant has been examined only in European and North American populations (Hampe et al. 2001; Hugot et al. 2001; Ogura et al. 2001*a*) and because the Arg702Trp and Gly908Arg variants were examined exclusively in a European population (Hugot et al. 2001), the present study provides the first opportunity to examine the prevalence of all three sequence variants in an independent population.

In addition, a number of recent reports indicate that common genetic variation in the human genome exists as discrete haplotype blocks, each with limited diversity, separated by intervals of multiple, independent, historical recombination events (Jeffreys et al. 2000; Daly et al. 2001). Knowledge of this haplotype structure allows for common variation in a gene to be tested exhaustively for association with disease, even if the causal variants have not been identified (Daly et al. 2001; Goldstein et al. 2001).

A recent application of the haplotype structure of the human genome to the study of disease association reveals that genetic variation in the cytokine gene cluster on chromosome 5q31, a locus known as *IBD5* (MIM 606348), confers susceptibility to CD (Rioux et al. 2001). We therefore sought to determine the haplotype structure in a 177-kb region containing the *CARD15* gene, and we discuss the implications of this structure for the history of the three causal variants.

Subjects and Methods

Patient Characteristics

A cohort of 231 well-characterized patients with CD from the province of Quebec, Canada, were recruited (179 from Montreal, 32 from Sherbrooke, 15 from Chicoutimi, and 5 from Quebec). All patients were seen by IBD specialists, and all clinical charts (including the physicians' notes, the radiology reports, endoscopy and histopathology results, and surgical specimens) were reviewed in detail by two investigators (G.W. and S.V.) who were blinded for the genotype status of each patient. The patients' records were systematically reviewed for the following demographic and clinical characteristics: age, sex, smoking habits, age at diagnosis, disease localization (ileal, colonic, or ileocolonic), disease behavior (inflammatory, stricturing, or fistulizing), history of abdominal surgery, presence of extraintestinal clinical manifestations (e.g., arthritis, uveitis, erythema nodosum), and familial IBD (Gasche et al. 2000) (table 1). Localization of disease was defined as the maximal extent of the disease during the period between diagnosis and latest follow-up. Full clinical data were obtained for all clinical variables except for smoking habits, which were incomplete for 29 patients. Of the total study cohort, the majority of patients, 133 (58%), were of French Canadian origin, and 14% were of Ashkenazi Jewish origin. The remaining patients had predominantly European ancestry (19 Italian, 16 English, 12 Sephardic Jewish, 1 Algerian, 4 Armenian, 2 Greek, 1 Haitian, 1 Irish, 3 Polish, 1 Spanish, and 1 Ukrainian).

In a subgroup of 84 patients, both parents were available, and parental data were collected (i.e., this set of 84 CD trios mostly overlaps with the samples described as "set D" in our study of the *IBD5* locus [Rioux et al. 2001]). A group of 71 healthy individuals served as controls for the 147 patients for whom parental genotypes were not available. Approval for the study was obtained from the local institutional review boards, and all par-

Table 1

Baseline Clinical Characteristics of the Study Population

Characteristic	No. (%) of Patients
History of smoking	52 (26.0)
Localization:	
Ileitis	68 (29.6)
Colitis	76 (33.2)
Ileocolitis	85 (37.2)
Behavior:	
Inflammatory	110 (48.1)
Structuring	42 (18.3)
Fistulizing	77 (33.6)
Previous abdominal surgery	115 (50.2)
Extraintestinal manifestations	94 (41.0)
Familial IBD	78 (34.1)
Age at diagnosis (years):	
<20	94 (41.1)
20–40	124 (54.1)
>40	11 (4.8)

NOTE.—Among patients, the mean age ± SD was 34 ± 10.3 years; 125 (55%) were female, and 104 (44%) were male. Patients were matched with healthy control individuals, who had a mean age ± SD of 33.4 ± 11.5 years (range 16–74 years) and included 43 females (60.6%) and 28 males (39.4%).

ticipants gave informed consent before being included in this study.

Sequencing

To identify common SNPs, PCR assays were designed to cover the 12 exons of the *CARD15* gene. The assays were designed using Primer 3.0 software, which yielded a target sequence of ~700 bp that included a 100-bp overlap with adjacent assays. The −21 M13 forward and the −28 M13 reverse sequences were added to each of the forward and reverse PCR primers, respectively. These PCR primers were used to amplify 50 ng of genomic DNA from eight independent individuals: six patients with CD, one patient with ulcerative colitis, and one DNA sample from CEPH that served as a control. We purified the PCR products, using the solid-phase reversible immobilization (SPRI) method (Hawkins et al. 1994), and sequenced them, using the appropriate ABI PRISM BigDye Primer −21 M13 or M13REV Cycle Sequencing Kit (Applied BioSystems). All sequencing reactions were run on ABI 377 automated sequencers (Applied BioSystems). The gel files were subsequently processed using BASS software, available on the Whitehead Institute/MIT Center for Genome Research FTP site. The forward and reverse reads were then aligned to a GenBank reference sequence (AC007728), using GAP4 software (Bonfield et al. 1998), and all traces were visually inspected.

Genotyping of Arg702Trp, Gly908Arg, and Leu1007fsinsC

DNA was extracted from peripheral blood lymphocytes, using a salting-out procedure (Miller et al. 1988). Patients and control subjects were genotyped for the Arg702Trp, Gly908Arg, and Leu1007fsinsC variants in the *CARD15* gene (GenBank accession number 64127), as defined by Hugot et al. (2001). The missense mutation Arg702Trp (also referred to as "Hugot SNP8"; GenBank accession number G67950) was genotyped by amplification refractory mutation system (ARMS) with primers 5′-ATCTGAGAAGGCCCTGCTCC-3′ (wild type, forward), 5′-ATCTGAGAAGGCCCTGCTCT-3′ (mutated, forward), and 5′-CCCACACTTAGCCTTGATG-3′ (reverse). The missense mutation Gly908Arg (Hugot SNP12; GenBank accession number G67951) creates a restriction site for *Hha*I and was genotyped by an RFLP-PCR technique (5′-CCCAGCTCCTCCCTCTTC-3′ and 5′-AAGTCTGTAATGTAAAGCCAC-3′). The presence of a wild-type allele results in an intact 380-bp band, whereas the profile of the Gly908Arg variant is characterized by two bands of 138 bp and 242 bp on 2% agarose gel. The ARMS technique, as described by Ogura et al. (2001*a*), was used to genotype the Leu1007fsinsC (SNP13; GenBank accession number G67955). The investigators who performed the *CARD15* genotyping were blinded with respect to the clinical phenotype of each patient.

Genotyping 45 SNPs for CARD15 *Haplotype Analyses*

In a subgroup of 84 patients, we collected DNA from both parents, and we genotyped the parent-offspring trios with a dense set of SNPs. Specifically, the set included two novel SNPs identified in the present study, 10 of the 13 SNPs reported by Hugot et al. (2001) (SNP1, SNP7, and SNP9 were not examined), and an additional 33 SNPs (rs1362390, rs746702, rs933568, rs1558663, rs745230, rs749986, rs749985, rs1990624, rs752301, rs10193, rs1420685, rs1981760, rs1362632, rs751271, rs748855, rs1861758, rs1861757, rs1861756, rs749910, rs1077861, rs718226, rs751919, rs1362698, rs1548990, rs1990752, rs1477176, rs1420873, rs1548989, rs1420871, rs1861762, rs2032688, rs2032687, and rs1861760) selected from the UCSC Human Genome Project Working Draft database. The SNPs were genotyped by time-of-flight mass spectrometry (Ross et al. 1998), using the Sequenom platform. In brief, PCR primer pairs needed to amplify the region around the SNP, as well as extension primers, were designed using Sequenom SpectroDesigner software. The PCR primers were used to amplify 5 ng of genomic DNA in a multiplex reaction (as many as five assays per reaction). Unincorporated nucleotides were removed from the products of the PCR, using shrimp alkaline phosphatase treatment. The homogeneous mass

extend reaction was then performed, adding one or two bases to the extension primer, to include one variant of the SNP or to include the alternate variant plus one base beyond the SNP. The extension reaction was purified using Sequenom SpectroClean resin, and the purified products were placed on a 384-spot DNA chip. The chip was run through a mass spectrometry workstation (Bruker), and the resulting spectra were analyzed using the Sequenom Spectro TYPER-RT software.

Seven of the SNPs (rs752301, rs1420685, rs1362632, rs751271, rs1548990, rs1990752, and rs1477176) from the UCSC database did not genotype well (<75% of the individuals were successfully genotyped), five of the SNPs (rs10193, rs1362390, rs933568, rs1352598, and rs1548989) were monomorphic or nearly so (0 or 1 copy of the minor allele observed) in the 84 trios, and one SNP (rs1861756) gave inaccurate genotyping results; hence, these SNPs were removed from further analysis.

Haplotype Construction

Of the 32 well-genotyped, polymorphic SNPs, 7 (Arg702Trp, Gly908Arg, Leu1007fsinsC, rs746702, rs2032688, rs2032687, and rs1861760) had minor allele frequencies (MAF) <5%. Haplotype construction and counting were performed with the remaining 25 polymorphic SNPs, as described elsewhere (Daly et al. 2001). Specifically, haplotype percentages in figure 1 were computed using haplotypes generated by the transmission/disequilibrium test (TDT) implementation in GENEHUNTER 2.0 (Kruglyak et al. 1996), followed by the use of an expectation-maximization algorithm, to include the minority of chromosomes that had one or more markers with ambiguous phase (i.e., both parents and offspring were heterozygous) or had one marker that was missing genotype data. The SNPs with <5% MAF were then added to the common haplotype patterns. Blocks are defined as regions in which none of the common haplotypes (>3%) show evidence of recombination.

Statistics

Logistic regression (SPSS 10.0) was performed to assess whether *CARD15* mutations were correlated with a particular clinical phenotype ($\alpha = .05$). This analysis was performed on the group of patients (229/231) for whom complete phenotype information was obtained. Given the significant results of the correlation between *CARD15* genotypes and disease location, we specifically examined (using Fisher exact test) whether the *IBD5* haplotype also correlated to disease location.

The association between *CARD15* genotypes and CD was examined by combining the case-control data (147 cases and 71 controls) and the trio data (84 trios), which was accomplished by calculating the observed and expected values and the variance of the number of mutant alleles among cases and transmitted chromosomes, respectively. These numbers were combined to create a studywide Z score for each mutation.

Results

Prevalence of CARD15 *Sequence Variants in a Cohort of Patients with CD from Quebec*

To date, the prevalence of all three causal variants in the *CARD15* gene has only been reported for a European population with CD (Hugot et al. 2001). We therefore collected a cohort of 231 patients with CD from the Canadian province of Quebec and thoroughly characterized them for a number of demographic and clinical characteristics (table 1). We then genotyped all patient and healthy control samples for the three causal *CARD15* variants and observed that the overall prevalence of *CARD15* mutations in the Quebec patients with CD (45.0%; 103/229) was significantly higher than in healthy control individuals (9.0%; 7/71) ($P < .001$). This observation was true for all three mutations when patients were compared with control individuals (Arg702Trp: 12.9% vs. 4.2%, $P = .004$; Gly908Arg: 5.2% vs. 0.7%, $P = .02$; Leu3020fsinsC: 10.3% vs. 0.7%, $P < .001$) (table 2). No significant difference was seen in *CARD15* variant frequencies between patients with sporadic CD (only one person affected in the family) (12.3%, 5.3%, and 11.7% for Arg702Trp, Gly908Arg, and Leu3020fsinsC, respectively) and patients with familial CD (minimum of two persons affected in the family) (13.9%, 5.1%, and 7.7%, respectively) (table 2). Of the 231 patients, 12 (5.2%) were homozygotes for a particular mutant, and 15 (6.5%) were compound heterozygotes. Given an overall frequency of mutations in cases of 27.9%, an excess of patients with two mutations, when compared with the expectation (27 observed vs. 18 expected; $P < .05$), confirms previous reports of a significantly greater risk attributed to double-mutant individuals. The estimated risk associated with carrying one or more *CARD15* mutations is odds ratio (OR) = 7.2 (95% CI 3.2–16.5).

Genotype-Phenotype Correlation

Knowing the genotype status for the three *CARD15* variants in a well-characterized, independent population with CD, we can examine the genotype-phenotype correlations in the disease process. We therefore performed multivariate logistic regression and found that disease localization is strongly correlated with the presence of the *CARD15* variants. Specifically, patients with ileal disease (either isolated ileitis or ileocolitis) had a 2.3–2.9 increased risk of carrying *CARD15* variants, compared with patients with solely colonic in-

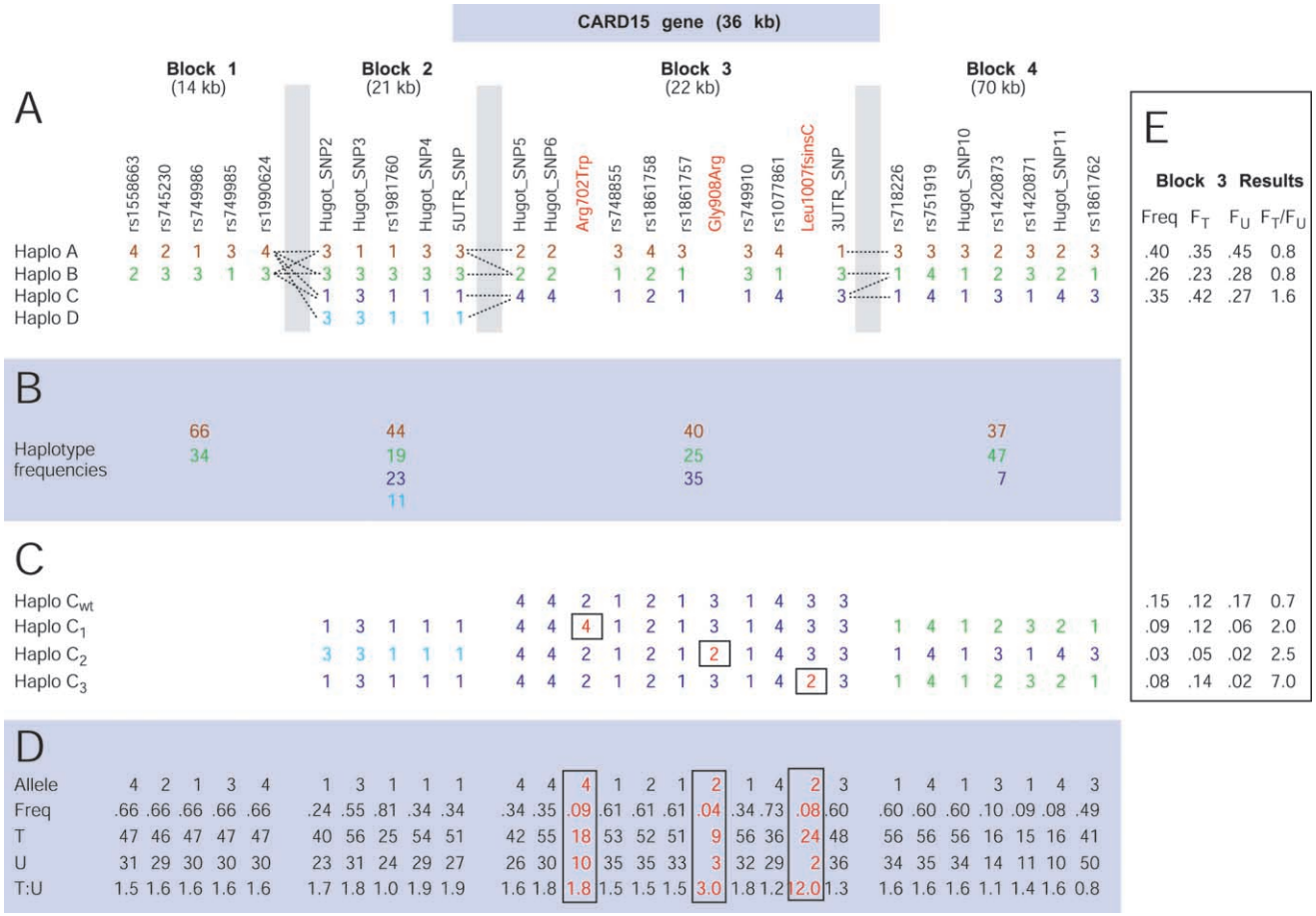

Figure 1 Blocklike haplotype structure of the *CARD15* gene. *A,* Common haplotype patterns in the four blocks of low diversity encompassing the *CARD15* gene. Dashed lines indicate locations where >5% of all chromosomes were observed to connect one common haplotype to another. Alleles at each SNP are indicated as numbers (1 = A, 2 = C, 3 = G, and 4 = T). All SNPs included in the haplotype structure have an MAF >5%. The *CARD15* gene (*topmost purple box*) represents the location of the gene in relation to the blocks. The gray vertical bars represent intervals of historical recombination between the blocks. Distances between Blocks 1 and 2, 2 and 3, and 3 and 4 are 33 kb, 14 kb, and 3 kb, respectively. *B,* Percentage of each of the common patterns among all observed chromosomes (transmitted and untransmitted). *C,* Haplotype structure of the wild-type haplotype C (C_{wt}), which contains no risk alleles, and the three subvariants of haplotype C (C_1, C_2, C_3), which confer risk of CD. For the three subvariants, blocks 2, 3, and 4 are in tight LD; therefore, all three blocks are displayed. Greater recombination is seen between blocks 2, 3, and 4 for the wild-type haplotype C, so only block 3 is displayed. SNPs that confer risk of CD are highlighted in red and boxed. *D,* Frequencies and transmitted:untransmitted ratios of the specified allele calculated for each individual SNP. The SNPs that confer risk of CD are highlighted in red and boxed. *E,* Overall frequencies and frequency among transmitted (F_T) and untransmitted (F_U) chromosomes for the common block 3 haplotypes (A, B, and C [*top*]) and for the C_{wt} and C subvariant block 3 haplotypes (C_1, C_2, and C_3) (*bottom*).

volvement ($P < .001$) (table 3). No other demographic or clinical variables were associated with the presence or absence of the *CARD15* variants, in a multivariate model.

Given the significant results of the correlation between *CARD15* genotypes and disease localization, we were interested to know whether the *IBD5* haplotype was also correlated with disease localization. Examination of the 84 CD trios, however, revealed no correlation between *IBD5* genotypes and disease localization. Furthermore, we found that 74% of all *CARD15*-positive cases had at least one copy of the *IBD5* risk haplotype, which was not significantly different from the frequency (76%) in the total population of cases. Finally, stratification of the data on the basis of *CARD15* or *IBD5* genotypes revealed no interaction between these two loci.

Analysis of the Genetic Variation Surrounding the CARD15 *Gene*

To identify common SNPs in the *CARD15* gene, the gene was resequenced in DNA samples obtained from six patients with CD, one patient with ulcerative colitis, and one control individual. Specifically, we re-

Table 2

Allele Frequencies for Arg702Trp, Gly908Arg, and Leu1007fsinsC in the Study Population

	Frequency in				
Mutation	Patients with CD (n = 231)	Patients with Sporadic IBD (n = 135)	Patients with Familial IBD (n = 96)	Control Individuals (n = 71)	P[a]
Arg702Trp	12.9	12.3	13.9	4.2	.0012
Gly908Arg	5.2	5.3	5.1	.7	.0022
Leu1007fsinsC	10.3	11.7	7.7	.7	5.6×10^{-8}
Overall	45.0	45.3	44.3	9.0	2.9×10^{-8}

Note.—P values were computed by combining the results of the comparison of 147 patients and 71 control subjects with the TDT results from the remaining 84 patients who had both parents genotyped, as described in "Subjects and Methods." No significant differences were found between patients with sporadic and familial disease.

[a] Patients with CD vs. control subjects.

sequenced all coding regions, all intron-exon boundaries, and both UTRs. This resulted in the identification of two novel SNPs: one in the 5′ UTR in exon 1 (−59 bp from start codon GTAGACAGATCCAGGCTCACCAGTCCTGTGCCACTGGGCTTTTGGC-[G/A]TTCTGCACAAGGCCTACCCGCAGATGCCATGCCTGCTCCCCCAGCC) and one in the 3′ UTR in exon 12 (+4,279 from start codon TGTTATTATTTTAAACATTATGATGTGTGAAAACTGGTTAATATTTATAG[A/G]TCACTTTGTTTTACTGTCTTAAGTTTATACTCTTATAGACAACATGGCCGTG).

These two novel SNPs, as well as 10 SNPs reported by Hugot et al. (2001) and 33 SNPs selected from the UCSC Human Genome Working Draft database, were genotyped in the 84 CD trios. Thirty-two of the 45 SNPs were polymorphic in this group and were successfully typed in at least 93% of all samples. These 32 SNPs span a genomic region of 177 kb, encompassing the entire genomic extent (36 kb) of the *CARD15* gene, as well as a 68.3-kb and a 72.7-kb region that flank this gene (to the left and right, respectively). Using the TDT, we examined the *CARD15* mutations (Arg702Trp, Gly908Arg, and Leu1007fsinsC) reported by Hugot et al. (2001) for association with CD. All three were overtransmitted from parents to affected children. Specifically, the transmitted:untransmitted (T:U) ratios for these three variants were 1.8:1 (Arg702Trp), 3.0:1 (Gly908Arg), and 12.0:1 (Leu1007fsinsC) (see fig. 1*D*). These results were combined with the case-control data by calculating a Z score. This demonstrated that all three variants were significantly associated with CD in this Quebec population: Arg702Trp (Z = 3.05; P = .0012), Gly908Arg (Z = 2.85; P = .0022), and Leu1007fsinsC (Z = 5.31; $P = 5.6 \times 10^{-8}$). Examination of the remaining common SNPs in the 84 CD trios by TDT analysis revealed five additional SNPs with T:U ratios ⩾1.8, resulting from linkage disequilibrium (LD) with the three reported mutations. Importantly, public SNP rs749910 (located between Gly908Arg and Leu1007fsinsC) is one of the SNPs associated with CD, suggesting that publicly available SNPs were adequate to detect *CARD15* involvement in disease susceptibility.

Because recent work by our group and others has indicated that the genetic variation in the human genome has a discrete haplotype block structure (Jeffreys et al. 2000; Daly et al. 2001), we sought to identify the pattern of common variation surrounding the *CARD15* gene. This was achieved by examining the 25 SNPs that had MAF >5%. We observed that the entire 177-kb region could be parsed into four discrete blocks (each spanning 14–70 kb [fig. 1*A*] and having two to four common ancestral haplotypes) that account for ⩾91% of all chromosomes. Among these common haplotypes, no evidence exists of recombination within each block. Block 3 contains most, if not all, of the coding sequence of the *CARD15* gene, and the genetic variation in this block falls into three common haplotype patterns (labeled in fig. 1A as "Haplo A," "Haplo B," and "Haplo C"). Examination of the T:U ratio in this block indicates that the "C" haplotype is overtransmitted from parents to affected offspring (T:U = 1.6–1.8:1), whereas the other two haplotypes are undertransmitted (fig. 1*D* and 1*E*).

The inclusion of the three *CARD15* causal variants in this analysis demonstrates that, intriguingly, the disease alleles are all located on the "C" haplotype and create three subvariants (labeled "C_1," "C_2," and "C_3"; fig. 1C). All three subvariant haplotypes are overtransmitted from parents to affected offspring, whereas the wild-type C haplotype that carries the nonrisk alleles for these three variants is undertransmitted. Examination of the longer-range haplotype patterns on chromosomes carrying the *CARD15* risk alleles demonstrates that the C_1, C_2, and C_3 subhaplotypes exist as different combinations of block 3 with its adjacent haplotype blocks. These data strongly support the theory that the three

Table 3

Relationship between *CARD15* Variants and Disease Phenotypes

			Univariate Analysis		Multivariate Analysis	
Characteristic	CARD15−[a]	CARD15+[a]	OR	95% CI	OR	95% CI
Mean age (years)	34.6	33.6	.991	.966–1.017	.989	.962–1.018
Female/male	56.0/44.0	52.4/47.6	1.155	.684–1.950	1.040	.593–1.826
Smoking	26.5	25.3	.936	.494–1.774	…	…
Localization:						
Ileitis	26.0	34.0	2.812	1.403–5.639	2.277	1.064–4.871
Colitis	43.9	20.4	1.00	…	1.00	…
Ileocolitis	30.1	45.6	3.266	1.683–6.339	2.858	1.412–5.785
Behavior:						
Inflammatory	52.5	43.7	1.00	…	1.00	…
Stricturing	14.7	22.3	1.817	.880–3.753	1.150	.498–2.660
Fistulizing	32.8	34.0	1.244	.688–2.251	.857	.416–1.767
Surgery	44.0	58.0	1.776	1.048–3.009	1.478	.758–2.880
Extraintestinal manifestations	41.6	40.8	.967	.567–1.642	.990	.556–1.764
Familial IBD	34.4	34.0	.982	.556–1.701	.875	.487–1.571
Age at diagnosis (years):						
<20	39.5	42.7	1.00	…	1.00	…
20–40	57.3	50.5	.816	.474–1.402	1.135	.608–2.120
>40	3.2	6.8	1.949	.534–7.109	4.940	.964–25.324

[a] Data are percentages, except as otherwise noted.

risk alleles arose independently on different chromosomes that, coincidentally, were locally identical.

Discussion

A susceptibility locus for CD on the pericentromeric region of chromosome 16 was initially described by Hugot et al. (1996); since then, the gene underlying the *IBD1* locus has been identified as *CARD15* (Hugot et al. 2001; Ogura et al. 2001*a*). The identification of the first gene that confers susceptibility to CD is a pivotal step toward an understanding of the causation of this debilitating disease. Specifically, the report by Hugot et al. identifying three *CARD15* sequence variants conferring susceptibility to CD (Hugot et al. 2001), the independent identification by Ogura et al. (2001*a*) of one of these variants (Leu3020insC), and the confirmation of this same variant in an independent European cohort (Hampe et al. 2001) provide an opportunity to ask specific questions regarding the role of this gene in disease susceptibility. The questions include the following: (1) What is the contribution of these variants in geographically diverse patient populations? (2) What is the spectrum of risk-conferring alleles for this gene? (3) How does the presence of these alleles influence the clinical phenotype? (4) How do these variants interact with other susceptibility loci? (5) By what specific mechanism(s) do these variants confer disease susceptibility?

In an attempt to answer some of these important questions, we collected a cohort of 231 patients with CD from the Canadian province of Quebec, and we conducted rigorous clinical phenotyping of all patients, including verification of diagnosis according to well-defined criteria (Gasche et al. 2000). We found that the prevalence of *CARD15* mutations in the Quebec patients with CD (Arg702Trp, 12.9%; Gly908Arg, 5.2%; and Leu1007fsinsC, 10.3%) was very similar to what was reported by Hugot et al. (2001). The prevalences of homozygotes (5.2%) and compound heterozygotes (6.5%) observed in the present study were also similar to those reported by Hugot et al. (2001) (6.0% and 8.5%, respectively). Taken together, these data identify a significant association of all three *CARD15* sequence variants with CD.

To study the genetic variation surrounding the *CARD15* gene, we performed SNP discovery in a limited number of individuals and genotyped a subset of our cases (those for whom we had parental DNA) for 45 sequence variants identified in the present study, the study by Hugot et al., and the public SNP database. The specific aim was to determine the haplotype structure of this genomic region, to determine the extent of LD, and to address the question of whether other common variants exist that could explain the association of *CARD15* with CD.

Recent attempts to characterize the haplotype structure of specific genomic regions have indicated that the genetic variation in humans largely follows simple patterns (Jeffreys et al. 2000; Daly et al. 2001; Patil et al. 2001). This pattern of common variation has been described as blocklike, because regions exist in which there is little or no evidence of recombination and because intervals between these regions show evidence of multiple historical recombination events. The variation between individual chromosomes is therefore a result of

the specific combinations of the limited number of haplotypes that exist in each block. This pattern suggests that examination of the haplotype-block structure surrounding a gene allows for a comprehensive analysis of the association of the gene with disease, even if the causal variants have not been identified (Daly et al. 2001; Goldstein et al. 2001).

We therefore examined the haplotype structure for a 177-kb region that encompassed the entire 36-kb *CARD15* gene. We demonstrated that this region can be parsed into discrete haplotype blocks, with a limited number of common haplotypes per block, as the study of the other genomic regions suggested. Importantly, we observed that the use of common SNPs could identify a haplotype (haplotype C) that had evidence of association with CD (fig. 1). When the three causal alleles were added to the analysis, it was observed that all were unique to the C haplotype, thus creating three independent subvariants of this haplotype, which is consistent with the initial observation made in a European population (Hugot et al. 2001). Moreover, these three variants are uniquely found on chromosomes with two different mosaic patterns of blocks 2, 3, and 4. This observation is consistent with these variants having arisen on individual chromosomes that had previously undergone recombination between the ancestral haplotypes, and it indicates a more recent history for these variants. The observation that the C_{wt} haplotype (i.e., the C haplotype without any of the three causal variants) is not associated with CD, together with our knowledge of the block structure of this region, strongly suggests that there are no additional common causal variants to be identified on the A, B, or C_{wt} haplotypes. However, the work of Hugot et al. (2001) suggests the existence of some rare variants in *CARD15* that contribute to disease risk.

Although grouped under the same unifying term, patients with CD clinically present with heterogeneous disease characteristics, including large differences in disease behavior, localization, and severity. Defining the relationship between *CARD15* variants and phenotypic variation in disease presentation is not only central to probing the clinical diversity in disease presentation and behavior but may also assist in defining rational treatment strategies. In the present study we found a significant association between ileal disease localization and *CARD15* variants. This finding is consistent with the recent report by Lesage et al. (2002), which associated *CARD15* variants with less frequent colonic involvement.

In addition, some studies have suggested that familial and sporadic IBD are two distinct entities, because of the differences in disease localization and behavior between the two groups (Peeters et al. 2000). The *CARD15* variants have thus far primarily been examined in multiply affected families. In the present study, however, we were able to examine the frequency of *CARD15* variants in both familial and nonfamilial cases, and we observed no differences, regardless of familial disease history (table 2).

We recently reported that genetic variation in the cytokine gene cluster on chromosome 5q31 confers risk of CD (Rioux et al. 2001). This *IBD5* risk haplotype is found in ~75% of patients with CD examined in a number of independent patient collections from the Canadian provinces of Ontario and Quebec, including the group of trios examined in the present study. Having the *CARD15* and *IBD5* variants genotyped in the same population allows for the examination of potential locus-locus interactions. In the present study there was no evidence of these two loci interacting, which is consistent with the preliminary observation that we made in samples from Ontario (Rioux et al. 2001). It is nevertheless possible that the number of trios studied did not have sufficient power to detect such an interaction and that larger cohorts of trios will be necessary to further characterize any possible interactions.

The underlying mechanism by which *CARD15* variants confer susceptibility to CD remains incompletely understood. Initial work by Ogura et al. (2001*a*) on the Leu1007fsinsC variant suggests that this susceptibility potentially relates to a deficit in the ability to sense bacteria in the gastrointestinal environment. A deficit in sensing bacteria would then result in an exaggerated inflammatory response of the adaptive immune system. Our observation that these sequence variants are associated with ileal disease may provide additional clues to the mechanisms of disease. Although the idea that there are differences in the bacterial ecosystem between the colon and the ileum is speculative, such differences could certainly account for differences in disease presentation (Shanahan 2002). It is possible that the beneficial effect of antibiotics, seen predominantly in patients who have CD with isolated colonic involvement, reflects a subgroup of patients who have gastrointestinal systems that are more efficient in recognizing and subsequently defending against bacterial invasion than are those in patients with primarily ileal disease (Greenbloom et al. 1998). However, well-designed trials would be necessary to prove this hypothesis.

In conclusion, the results of our study confirm that three sequence variants in the *CARD15* gene confer risk of CD. Haplotype analysis also suggests that these variants are likely to have arisen recently on individual chromosomes that originate from a common ancestral haplotype. Finally, our observation that these sequence variants are correlated with a specific disease subphenotype is another step toward understanding how sequence variation in *CARD15* influences disease susceptibility in CD.

Acknowledgments

We acknowledge the patients with IBD and their families for their collaboration, and nurses N. Pellerin, M. L. Bernier, and S. Bazinet for the collection of the blood samples. The authors would also like to thank L. Gaffney for her help in the preparation of this manuscript. This work was supported by an operating grant from the Canadian Institute for Health Research, by a senior clinician scientist award (to G.W.) and a junior clinician scientist award (to A.B.) from the Fonds de la Recherche en Santé du Quebec (to G.W.), and by research grants from Bristol-Myers Squibb, Affymetrix, and Millennium Pharmaceuticals (to J.D.R.).

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for *CARD15* genomic sequence information [accession number AC007728])

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for CD [MIM 266600], *CARD15* [MIM 605956], and *IBD5* [MIM 606348])

UCSC Human Genome Project Working Draft Database, http://genome.ucsc.edu/ (for identification of SNPs surrounding the *CARD15* gene)

Whitehead Institute/MIT Center for Genome Research FTP site, http://www-genome.wi.mit.edu/ (for BASS software)

References

Ahmad T, Satsangi J, McGovern D, Bunce M, Jewell DP (2001) The genetics of inflammatory bowel disease. Aliment Pharmacol Ther 15:731–748

Andres PG, Friedman LS (1999) Epidemiology and the natural course of inflammatory bowel disease. Gastroenterol Clin North Am 28:255–281

Bernstein CN, Blanchard JF, Rawsthorne P, Wajda A (1999) Epidemiology of Crohn's disease and ulcerative colitis in a central Canadian province: a population-based study. Am J Epidemiol 149:916–924

Bonfield JK, Rada C, Staden R (1998) Automated detection of point mutations using fluorescent sequence trace subtraction. Nucleic Acids Res 26:3404–3409

Cho JH, Nicolae DL, Gold LH, Fields CT, Labuda MC, Rohal PM, Pickles MR, Qin L, Fu Y, Mann JS, Kirschner BS, Jabs EW, Weber J, Hanauer SB, Bayless TM, Brant SR (1998) 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 95: 7502–7507

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. Nat Genet 29:229–232

Duerr RH, Barmada MM, Zhang L, Pfützer R, Weeks DE (2000) High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11-12. Am J Hum Genet 66:1857–1862

Ekbom A, Helmick C, Zack M, Adami HO (1991) The epidemiology of inflammatory bowel disease: a large, population-based study in Sweden. Gastroenterology 100:350–358

Elson CO (2002) Genes, microbes, and T cells: new therapeutic targets in Crohn's disease. N Engl J Med 346:614–616

Gasche C, Scholmerich J, Brynskov 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

Goldstein AM, Liu L, Shennan MG, Hogg D, Tucker MA, Struewing JP (2001) A common founder for the V126D CDKN2A mutation in seven North American melanoma-prone families. Br J Cancer 85:527–530

Greenbloom SL, Steinhart AH, Greenberg GR (1998) Combination ciprofloxacin and metronidazole for active Crohn's disease. Can J Gastroenterol 12:53–56

Hampe J, Cuthbert A, Croucher PJP, Mirza MM, Mascheretti S, Fisher S, Frenzel H, King K, Hasselmeyer A, MacPherson AJS, Bridger S, van Deventer S, Forbers A, Nikolaus S, Lennard-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, Schreiber S, Shaw SH, Lau KF, Bridger S, MacPherson AJS, Cardon LR, Sakul H, Harris TJR, Buckler A, Hall J, Stokkers P, van Deventer SJH, Nürnberg P, Mirza MM, Lee JCW, Lennard-Jones JE, Mathew CG, Curran M (1999) A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. Am J Hum Genet 64:808–816

Hawkins TL, O'Connor-Morin T, Roy A, Santillan C (1994) DNA purification and isolation using a solid-phase. Nucleic Acids Res 22:4543–4544

Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain C, 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

Hugot JP, Laurent-Puig P, Gower-Rousseau C, Olson JM, Lee JC, Beaugerie L, Naom I, Dupas JL, Van Gossum A, Orholm M, Bonaiti-Pellie C, Weissenbach J, Mathew CG, Lennard-Jones JE, Cortot A, Colombel JF, Thomas G (1996) Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature 379:821–823

IBD International Genetics Consortium (2001) International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn disease and chromosome 16. Am J Hum Genet 68: 1165–1171

Inohara N, Ogura Y, Chen FF, Muto A, Nunez G (2001) Human Nod1 confers responsiveness to bacterial lipopolysaccharides. J Biol Chem 276:2551–2554

Jeffreys AJ, Ritchie A, Neumann R (2000) High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. Hum Mol Genet 9: 725–733

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Para-

metric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Lesage S, Zouali H, Cezard JP and the EPWG-IBD group, Colombel JF and the EPIMAD group, Belaiche J and the 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 AS, Thomas G, Hugot JP (2002) CARD15/CARD15 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. Am J Hum Genet 70:845–857

Ma Y, Ohmen JD, Li Z, Bentley LG, McElree C, Pressman S, Targan SR, Fischel-Ghodsian N, Rotter JI, Yang H (1999) A genome-wide search identifies potential new susceptibility loci for Crohn's disease. Inflamm Bowel Dis 5:271–278

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215

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*) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 411: 603–606

Ogura Y, Inohara N, Benito A, Chen FF, Yamaoka S, Nunez G (2001*b*) Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-κB. J Biol Chem 276:4812–4818

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP, Cox DR (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294:1719–1723

Peeters M, Cortot A, Vermeire S, Colombel JF (2000) Familial and sporadic inflammatory bowel disease: different entities? Inflamm Bowel Dis 6:314–320

Pinchbeck BR, Kirdeikis J, Thomson AB (1988) Inflammatory bowel disease in northern Alberta. An epidemiologic study. J Clin Gastroenterol 10:505–515

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

Ross P, Hall L, Smirnov I, Haff L (1998) High level multiplex genotyping by MALDI-TOF mass spectrometry. Nat Biotechnol 16:1347–1351

Satsangi J, Parkes M, Louis E, Hashimoto L, Kato N, Welsh K, Terwilliger JD, Lathrop GM, Bell JI, Jewell DP (1996) Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. Nat Genet 14:199–202

Shanahan F (2002) Crohn's disease. Lancet 359:62–69