Linkage and Association between Inflammatory Bowel Disease and a Locus on Chromosome 12

Richard H. Duerr,[1,3] M. Michael Barmada,[4] Leilei Zhang,[1,3] Sean Davis,[4] Robert A. Preston,[2,3] Lara J. Chensny,[1,3] Jody L. Brown,[1,3] Garth D. Ehrlich,[2,3] Daniel E. Weeks,[4,5] and Christopher E. Aston[3,4]

Departments of [1]Medicine and [2]Pathology and Otolaryngology and [3]Center for Genomic Sciences, University of Pittsburgh School of Medicine, and [4]Department of Human Genetics, University of Pittsburgh School of Public Health, Pittsburgh; and [5]The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford

Summary

Genetic epidemiological studies have shown that genetic factors are important in the pathogenesis of the idiopathic inflammatory bowel diseases (IBD), Crohn disease (CD), and ulcerative colitis (UC). A genome screen in the United Kingdom found linkage of IBD to a 41-cM region of chromosome 12, surrounding D12S83. We aimed to replicate this linkage and to narrow the region of interest. Nonparametric linkage analyses at microsatellites surrounding D12S83 were performed in 122 North American Caucasian families containing 208 genotyped IBD-affected relative pairs. Transmission/disequilibrium tests (TDTs) were also performed. We confirmed that IBD is linked to chromosome 12 (peak GENEHUNTER-PLUS LOD* score 2.76 [P = .00016] between D12S1724 and D12S90). The evidence for linkage is contributed by both the group of CD-affected relative pairs (peak GENEHUNTER-PLUS LOD* score 1.79 [P = .0021] between D12S1724 and D12S90) and the group of UC-affected relative pairs (peak GENEHUNTER-PLUS LOD* score 1.82 [P = .0019] at D12S335). The TDT is positive at the D12S83 locus (global χ^2 = 16.41, 6 df, P = .012). In conclusion, we have independently confirmed linkage of IBD to the chromosome 12 region that we investigated. A positive TDT at D12S83 suggests that we have greatly narrowed the chromosome 12 region that contains an IBD locus.

Received February 12, 1998; accepted for publication April 22, 1998; electronically published May 22, 1998.

Address for correspondence and reprints: Dr. Richard H. Duerr, 565 Scaife Hall, 3550 Terrace Street, Pittsburgh, PA 15261. E-mail: duerr+@pitt.edu

0002-9297/98/6301-0017$02.00

Introduction

Crohn disease (CD) (MIM 266600) and ulcerative colitis (UC) (MIM 191390) are chronic inflammatory disorders of the gastrointestinal tract that can be distinguished on the basis of clinical parameters (Lashner 1995; Singleton 1995). These disorders are collectively known as "idiopathic inflammatory bowel disease" (IBD-II) (MIM 601458). The prevalence of IBD in Europe and North America is ~100–200/100,000 (Calkins and Mendelhoff 1995).

Ethnic (especially Ashkenazi Jewish) aggregation of IBD, familial aggregation of IBD, and a greater concordance for IBD in MZ twins than in DZ twins suggest that genetic susceptibility is important in the pathogenesis (Yang and Rotter 1994; Duerr 1996). However, CD and UC do not exhibit simple Mendelian inheritance attributable to a single gene locus; they are genetically complex traits that involve incomplete penetrance and probably involve genetic heterogeneity and more than one susceptibility locus (Yang and Rotter 1994; Duerr 1996). Since both forms of IBD can coexist in single families, CD and UC may share some susceptibility genes, although there may be additional, CD-specific and UC-specific genes (Yang and Rotter 1994; Duerr 1996).

A genomewide search for IBD-susceptibility loci in a British population provided evidence for linkage between IBD (both CD and UC) and microsatellite loci on chromosomes 3, 7, and 12 (Satsangi et al. 1996). The strongest evidence for linkage of both CD and UC in this study was to a chromosome 12 region spanning 41 cM, in the vicinity of the D12S83 microsatellite locus (Satsangi et al. 1996). We report here the results of nonparametric linkage analyses and transmission/disequilibrium tests (TDTs) in a North American population of IBD families, using genotyping data at microsatellite loci surrounding the D12S83 microsatellite locus.

Subjects and Methods

Study Subjects

The study protocol was approved by the University of Pittsburgh Health Sciences Institutional Review Board. Patients diagnosed with CD or UC and with another CD- or UC-affected family member were ascertained from the population of patients at the University of Pittsburgh Medical Center or from responders to notices in publications by the Crohn's & Colitis Foundation of America, Inc., and the United Ostomy Association. IBD index cases were interviewed to determine the availability of another IBD-affected family member. Affected and connecting unaffected members of families with at least one IBD-affected relative pair as distant as second cousins but other than a parent-child pair were asked to contribute blood specimens for the IBD Family Lymphoblastoid Cell Line, DNA, and Serum Bank at the University of Pittsburgh. All study subjects were Caucasians of European descent. All study subjects were interviewed to complete a demographic and clinical database, and consents for release of medical records to verify diagnoses were obtained from affected study subjects.

Genotyping

Genomic DNA was isolated either from the buffy coat, white-blood-cell fraction of whole blood or from Epstein-Barr-virus–transformed lymphoblastoid cell lines (Neitzel 1986), by a standard method (Miller et al. 1988). The 12 polymorphic microsatellite loci that were genotyped in this study were selected from purchased ABI PRISM linkage-mapping panels (Perkin Elmer, Applied Biosystems Division), or they were selected from the Généthon database (Dib et al. 1996); and fluorescent dye-labeled forward primers and unlabeled reverse primers were custom synthesized (Research Genetics). PCR was used to amplify the microsatellite loci. PCR mixtures consisted of 20 ng of genomic DNA, 0.25 mmol of each dNTP/liter, 1.25 pmol of dye-labeled forward primer, 1.25 pmol of unlabeled reverse primer, 0.3 U of Amplitaq Gold DNA polymerase (Perkin Elmer, Applied Biosystems Division), 50 mmol of KCl/liter, 2.5 mmol of $MgCl_2$/liter, and 10 mmol of Tris-HCl (pH 8.3)/liter, in 7.5-μl reaction volumes. Thermal cycling parameters consisted of an initial 95°C denaturation step; 10 cycles of denaturation at 95°C for 15 s, annealing at 55°C for 30 s, and extension at 72°C for 60 s; 20 cycles of denaturation at 89°C for 15 s, annealing at 55°C for 30 s, and extension at 72°C for 60 s; and a final extension step at 72°C for 10 min. Subsets of the PCR products from each DNA sample were pooled together, heat denatured, loaded into 24-cM long, 5% denaturing polyacrylamide gels, and electrophoretically resolved and detected by use of an ABI 373A GENESCAN system (Perkin Elmer, Applied Biosystems Division). Allele sizes were determined by use of ABI GENESCAN analysis software (version 1.2.2-1) (Perkin Elmer, Applied Biosystems Division).

Statistical Analyses

The consistency of genotyping data was checked by comparing the expected number of heterozygotes (based on the marker's calculated heterozygosity) against the observed number of heterozygotes. Mendelian inheritance was checked by use of the PedCheck program, which uses the individual's genotype information to check for parent-child inconsistencies and too many alleles in a sibship and which then uses genotype elimination to find more-subtle errors (O'Connell and Weeks 1997).

For the nonparametric linkage analyses, allele frequencies for each microsatellite locus were estimated by counting the alleles in unrelated individuals, augmented by the addition of nontransmitted alleles derived from each nuclear-family unit. IBD, CD, and UC analyses were performed by use of the allele-frequency estimates from the entire data set and with affection status for the entire data set coded for IBD, CD, or UC, respectively.

Single-point affected-relative-pair analyses were performed by use of the simIBD program (Davis et al. 1996). This program measures identity-by-descent sharing, at a given marker locus, between all relative pairs, where two alleles are identical by descent when they share the same label and have the same ancestral origin. The probability of the observed degree of sharing is then calculated empirically. When this sharing is significantly different from what is expected under the hypothesis of no linkage, statistical evidence exists for linkage between the marker and the disease.

Multipoint affected-relative-pair analyses were performed by use of the S_{pairs} scoring function in the GENEHUNTER-PLUS modification of the GENEHUNTER package (Kruglyak et al. 1996; Kong and Cox 1997). The S_{pairs} statistic of GENEHUNTER measures identity-by-descent allele sharing between all pairs of affected relatives. Two assumptions are made in the calculation of this statistic and its significance: (1) since the actual inheritance pattern (or "inheritance vector") is never known with absolute surety, the statistic is calculated as the expected value of the conditional inheritance-vector distribution (the "perfect data" approximation); and (2), since the actual distribution of the statistic is unknown, the statistic is assumed to follow a normal distribution (the "large-sample" approximation). However, calculating the statistic under the first assumption and then deriving its *P* value under the second assumption leads to a conservative measure of the true allele sharing. To

address this, Kong and Cox (1997) proposed a modification of the method used to calculate the S_{pairs} statistic. This new statistic, Zlr, can be more accurately calculated, and it is well approximated by a normal distribution. This modification also permits the calculation of a nonparametric LOD score (LOD*) based on the allele sharing.

The TRANSMIT software program was used to perform the TDT on the data from nuclear families (Spielman et al. 1993; Clayton 1997). This test examines the transmission of alleles from parents to affected offspring, looking for a deviation from the expected ratio. Deviations from this expectation suggest linkage disequilibrium between the allele and a putative disease locus. For these analyses, allele-frequency estimates were derived, by maximum-likelihood techniques, for the nuclear families. Rare alleles (frequency < .05) were collapsed. A global χ^2 statistic was computed for each microsatellite locus. Statistics for individual alleles were computed only when the global χ^2 statistic for a microsatellite locus was significant.

Results

Description of Study Subjects, Families, and Affected Relative Pairs

We genotyped 216 CD-affected, 79 UC-affected, 2 indeterminate IBD-affected, and 206 unaffected members of 122 families that contained at least one IBD-affected relative pair other than a parent-child affected pair. Affected relative pairs were as distant as second cousins. Thirty-nine percent of the nonfounder affecteds had both parents genotyped, and 24% had only one parent genotyped. Thirty-four percent of the affecteds who had either only one or neither parent genotyped had at least two of their siblings genotyped. All study subjects were Caucasians of European descent. Jewish ethnicity was reported in 27 of the 122 families. The distributions of families and classes of affected relative pairs for each analysis are shown in table 1.

Table 1

Families and Genotyped Affected Relative Pairs

Category	No. with IBD	CD	UC
Families with each type of affected relative pair	122	80	22
Genotyped affected relative pairs:[a]			
Sibling pairs	115	84	15
Half-sibling pairs	3	2	0
Uncle (aunt)–nephew (niece) pairs	33	22	5
Half-uncle (half-aunt)–nephew (niece) pairs	1	0	0
Grandparent-grandchild pairs	5	0	3
Great uncle (great aunt)–nephew (niece) pairs	1	1	0
First-cousin pairs	27	11	9
First-cousin–once-removed pairs	17	5	9
Second-cousin pairs	6	2	2
Subtotal	208	127	43

[a] All possible combinations of pairs formed by all affecteds within each family (excluding parent-child affected pairs).

Description of Microsatellite Loci

The microsatellite loci that were studied, their positions on the Genetic Location Database map (Collins et al. 1996), the calculated heterozygosity, and the number of alleles observed at each locus are listed in table 2. The Genetic Location Database map was used because it incorporates partial maps of physical, genetic, regional, mouse-homology, and cytogenetic data into an integrated map with an ordering of loci that is likely to be more accurate than the ordering of loci in genetic maps alone (Collins et al. 1996).

All of the study subjects were genotyped at six of the microsatellite loci listed in table 2. An initial analysis of data from the UC-affected relative pairs showed evidence for linkage between UC and the q-telomeric end of the original grid of six microsatellite loci. In contrast, the initial analysis of data from the whole set of IBD-affected relative pairs and the subset of CD-affected relative pairs did not show evidence for linkage to the q-telomeric end of the original grid of six microsatellite loci. Therefore, we extended the grid of genotyped microsatellite loci to include six additional microsatellites for only the UC-affected relative pairs and their family members, but we did not genotype the remainder of the study subjects at these additional six loci.

Linkage of IBD to Chromosome 12

The results of single-point nonparametric linkage analyses using the simIBD program (Davis et al. 1996) are shown in table 2. The results of multipoint nonparametric linkage analyses using the GENEHUNTER-PLUS program (Kong and Cox 1997) are shown in figure 1.

These analyses confirm that there is linkage between IBD and the chromosome 12 region that we investigated (simIBD $P = .012$ at D12S1724; peak GENEHUNTER-PLUS LOD* score 2.79, $P = .00016$, between D12S1724 and D12S90). Analysis of the data, with affection status coded for CD (with UC- and indeterminate IBD-affected individuals coded as unaffected) shows that CD-affected relative pairs contribute to the overall evidence for linkage between IBD and the chromosome 12 region that we investigated (simIBD $P = .045$ at D12S1724; peak GENEHUNTER-PLUS LOD* score 1.79, $P = .0021$, between D12S1724 and D12S90). Analysis of the data with affection status coded for UC (with CD- and indeterminate IBD-affected individuals coded as unaffected) shows that UC-affected relative pairs also contribute to the overall evidence for linkage

Table 2

Microsatellite Loci and simIBD Analyses

LDB Map Position[a] (cM)	Microsatellite Locus	Heterozygosity (%)	No. of Alleles Observed	simIBD *P* for[b]		
				IBD	CD	UC
69.3	D12S359	77.1	11	.26	.23	.41
71.5	D12S1586	84.9	12	.24	.54	.031
74.6	D12S1724	80.2	12	.012	.045	.059
79.2	D12S90	78.6	9	.08	.24	.23
80.8	D12S83	78.6	13	.10	.28	.023
81.2	D12S1585	78.6	7	...	...	.072
81.5	D12S1686	82.3	15	...	...	.21
82.0	D12S1702	85.8	16	...	...	.001
82.1	D12S335	81.7	11	.29	.60	.028
82.9	D12S313	80.9	12	...	...	.005
83.5	D12S1703	91.2	20	...	...	.007
86.6	D12S326	79.3	9	...	...	.095

[a] Source: Collins et al. (1996).
[b] Source: Davis et al. (1996). Values <.05 are underlined.

between IBD and this region (.05 > simIBD P > .01 at D12S1586, D12S83, and D12S335; and simIBD P < .01 at D12S1702, D12S313, and D12S1703; peak GENEHUNTER-PLUS LOD* score 1.82, P = .0019 at D12S335). Although the group of CD-affected relative pairs and the group of UC-affected relative pairs both contribute to the overall evidence for linkage, the GENEHUNTER-PLUS peak LOD* score for the UC-affected relative pairs is located 6.7 cM q-telomeric, on the Genetic Location Database map, from the peak LOD* score for the CD-affected relative pairs.

Distortion of Allele Transmission at the D12S83 Microsatellite Locus

The TDT is positive at the D12S83 microsatellite locus in multiplex IBD nuclear families when alleles with frequencies <.05 are collapsed (global χ^2 = 16.41, 6 df, P = .012). The most significant finding for an individual allele is distortion of transmission of allele 7 at the D12S83 locus (observed transmission 78, expected transmission 68.65; χ^2 = 4.10, 1 df, P = .043). There are no other significant results when the TDT is applied to the IBD, CD, and UC data sets for the 12 microsatellite loci that we studied.

Discussion

We have independently confirmed Satsangi et al.'s (1996) finding of linkage between IBD and chromosome 12. Our confirmation of linkage is statistically significant because the level of evidence necessary to confirm a linkage in a particular region is lower than that needed to initially detect a linkage over the whole genome (Lander and Kruglyak 1995). In our multipoint analysis, the strongest evidence for linkage is to the region between the D12S1724 and D12S90 microsatellite loci.

Stratified analyses suggest that the group of CD-affected relative pairs and the group of UC-affected relative pairs both contribute to the evidence for linkage between IBD and the chromosome 12 region that we investigated. The region with the strongest evidence for linkage in the UC-affected relative pairs is located 6.7 cM q-telomeric, on the Genetic Location Database map, from the region with the strongest evidence for linkage in the CD-affected relative pairs. However, overlap in the 1-LOD*-unit support intervals for the CD- and UC-affected relative pairs, the possibility of misspecification of the precise order of microsatellite loci on currently available maps of this region, the relatively small UC-affected relative-pair sample in our study, and limitations inherent to affected-relative-pair allele-/haplotype-sharing methods in the determination of fine localization of susceptibility loci for complex traits prevent a conclusion that there are distinct CD- and UC-susceptibility loci on chromosome 12 (Kruglyak and Lander 1996).

The distortion of allele transmission from parents to IBD-affected offspring in our study occurs at the very same locus (D12S83) to which IBD was most strongly linked in the report by Satsangi et al. (1996) and is located within 3 cM, on the Genetic Location Database map, from the region with the strongest evidence for linkage in our IBD-affected relative-pair multipoint analysis. Although it would have been more pleasing if the positive TDT had occurred at exactly the same position as the peak evidence for linkage in our study, the TDT and the linkage results are not inconsistent with each other, since the positive TDT is within the 1-LOD*-unit support interval for linkage and since affected-relative-pair allele-/haplotype-sharing methods are known to

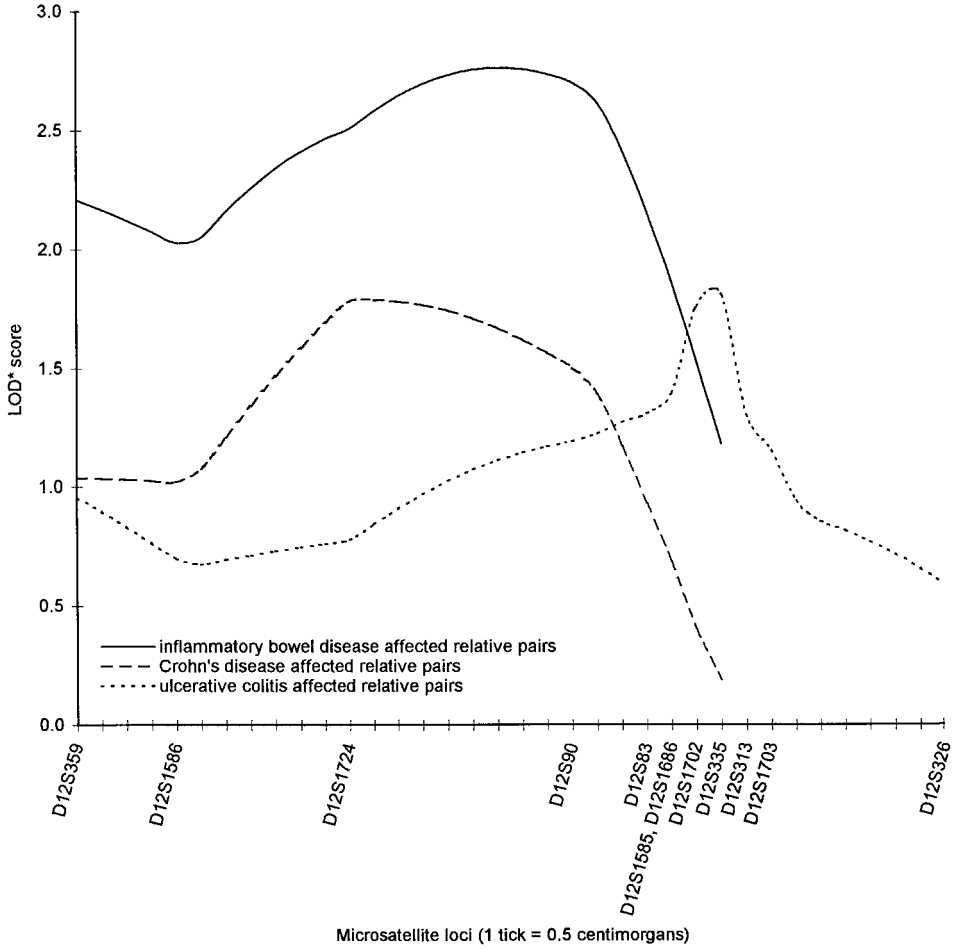

Figure 1 Results of multipoint nonparametric linkage analyses using the S_{pairs} scoring function in the GENEHUNTER-PLUS modification of the GENEHUNTER package (Kruglyak et al. 1996; Kong and Cox 1997), for IBD-, CD-, and UC-affected relative pairs, plotted as the GENEHUNTER-PLUS LOD* score (*Y*-axis) against the genetic distance (in cM) from the first microsatellite locus (*X*-axis).

have limitations in the fine mapping of complex traits (Kruglyak and Lander 1996).

To be confident that our observation of distortion of allele transmission, from parents to IBD-affected offspring, at D12S83 is a significant finding would require confirmation in a larger data set or an independent study, since the *P* value for the TDT at D12S83 is not significant after correction for the multiple loci that we studied. During the time that the manuscript of this article was undergoing the editorial review process, Parkes et al. (1998), the same group that originally reported evidence for linkage between IBD and chromosome 12, reported in an abstract that the TDT is positive at the D12S83 microsatellite locus in their independent set of IBD-affected sib pairs and parents. The finding of a positive TDT at the same locus, within the same region of linkage on chromosome 12, in two independent data sets establishes an IBD-susceptibility locus in this region of chromosome 12. Furthermore, the positive TDT at the D12S83 microsatellite locus in these two independent data sets suggests that the region of interest can be narrowed to the immediate vicinity of D12S83.

Acknowledgments

This research was supported by grants (to R.H.D.) from the Scaife Family Foundation, the National Institutes of Health (grant DK50993), and the Crohn's & Colitis Foundation of America, Inc., and by funds from the University of Pittsburgh School of Medicine Center for Genomic Sciences. Part of this work was reported in abstract form at the 1997 American Gastroenterological Association annual meeting in Washington, DC. We thank Katherine Kim and April Studinski for their assistance in the recruitment of study subjects.

Electronic-Database Information

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

Clayton D (1997) Transmit, http://www.mrc-bsu.cam.ac.uk/pub/methodology/genetics/transmit

Généthon, http://www.genlink.wustl.edu/genethon_frame

Genetic Location Database, http://cedar.genetics.soton.ac.uk/public_html

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for UC [MIM 191390], CD [MIM 266600], and IBD-II [MIM 601458])

References

Calkins BM, Mendelhoff AI (1995) The epidemiology of idiopathic inflammatory bowel disease. In: Kirsner JB, Shorter RG (eds) Inflammatory bowel disease, Williams & Wilkins, Baltimore, pp 31–68

Collins A, Frezal J, Teague J, Morton NE (1996) A metric map of humans: 23,500 loci in 850 bands. Proc Natl Acad Sci USA 93:14771–14775

Davis S, Schroeder M, Goldin LR, Weeks DE (1996) Nonparametric simulation-based statistics for detecting linkage in general pedigrees. Am J Hum Genet 58:867–880

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, et al (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154

Duerr RH (1996) Genetics of inflammatory bowel disease. Inflamm Bowel Dis 2:48–60

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Kruglyak L, Lander ES (1996) Limits on fine mapping of complex traits. Am J Hum Genet 58:1092–1093

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Lashner BA (1995) Clinical features, laboratory findings, and course of Crohn's disease. In: Kirsner JB, Shorter RG (eds) Inflammatory bowel disease. Williams & Wilkins, Baltimore, pp 344–354

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

Neitzel H (1986) A routine method for the establishment of permanent growing lymphoblastoid cell lines. Hum Genet 73:320–326

O'Connell JR, Weeks DE (1997) PedCheck: a program for identifying marker typing incompatibilities in linkage analysis. Am J Hum Genet Suppl 60:A288

Parkes M, Satsangi J, Jewell DP (1998) Transmission disequilibrium of a marker haplotype fine maps chr 12 linkage in IBD. Gut Suppl 42:A49

Satsangi J, Parkes M, Louis E, Hashimoto L, Kato N, Welsh K, Terwilliger JD, et al (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

Singleton JW (1995) Clinical features, course, and laboratory findings in ulcerative colitis. In: Kirsner JB, Shorter RG (eds) Inflammatory bowel disease. Williams & Wilkins, Baltimore, pp 335–343

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506–516

Yang H, Rotter JI (1994) Genetics of inflammatory bowel disease. In: Targan SR, Shanahan F (eds) Inflammatory bowel disease: from bench to bedside. Williams & Wilkins, Baltimore, pp 32–64