

ARTICLE

Haplotype-based association analysis of 56 functional candidate genes in the *IBD6* locus on chromosome 19

Marcela Karey Tello-Ruiz¹, Christine Curley¹, Terrye DelMonte¹, Cosmas Giallourakis¹, Andrew Kirby¹, Kathryn Miller¹, Gary Wild², Albert Cohen², Diane Langelier³, Anna Latiano⁴, Neils Wedemeyer⁵, Eric Lander¹, Stefan Schreiber⁵, Vito Annese⁴, Mark J Daly¹ and John D Rioux^{*,1,6,7}

¹The Broad Institute of MIT and Harvard, Cambridge, MA, USA; ²McGill University Health Center, Montreal, Canada; ³University of Sherbrooke, Sherbrooke, Canada; ⁴Gastroenterology Unit, CSS-IRCCS Hospital, San Giovanni Rotondo, Italy; ⁵The Christian-Albrechts-University, Kiel, Germany; ⁶Montreal Heart Institute, Montreal, Canada; ⁷Université de Montréal, Montreal, Canada

Evidence from four independent linkage studies and two meta-analyses of genome-wide data support the existence of a locus conferring susceptibility to inflammatory bowel diseases (IBD) in chromosomal region 19p. Identification of a susceptibility allele in this ~28.5 Mb region with over 600 genes is a formidable task. To tackle this problem, we undertook two approaches: (1) haplotype-based candidate-gene screen, and (2) evaluation of previously reported associations. For the former, we selected genes with potential implication in IBD pathogenesis based on published functional and expression data, typed SNPs, constructed haplotypes, screened for association in 180 IBD trios, and followed up preliminary associations in 343 IBD patients and 207 control individuals. Overall, we analyzed 465 SNPs, and 260 haplotypes distributed across 56 candidate genes. We found suggestive evidence of association (nominal $P < 0.01$) with four genes (*C3*, *FCER2*, *IL12RB1*, and *CRLF1*) in a screening stage, but were unable to confirm these preliminary observations at follow-up. In the second approach, we typed four nonsynonymous polymorphisms in genes *C3* (R102G and L314P) and *ICAM1* (G241R and K469E) in four independent cohorts totaling 2178 IBD cases. We evaluated these data together with previously published reports for three of these variants (*C3*-Gly102, *ICAM1*-Arg241, and *ICAM1*-Glu469), in a meta-analysis. Our pooled meta-analysis provides compelling evidence against association of these variants with disease. Overall, we performed the most comprehensive candidate-gene association study for IBD to date. The information hereby generated constitutes a valuable resource to investigate other common genetic immune diseases, such as celiac disease.

European Journal of Human Genetics (2006) 14, 780–790. doi:10.1038/sj.ejhg.5201612; published online 29 March 2006

Keywords: inflammatory bowel disease; *IBD6*; candidate-gene association

Introduction

The inflammatory bowel diseases (IBD), Crohn's disease (CD) and ulcerative colitis (UC), are complex traits

involving both environmental and genetic risk factors. These diseases are characterized by chronic relapsing inflammation of the gastrointestinal tract, and symptoms include diarrhea, constipation, and abdominal pain. The inflammation in CD is transmural and discontinuous, and most often localized to the terminal ileum and colon, and results in fistulas and obstructions.¹ Inflammation in UC, on the other hand is continuous and limited to mucosal layer of the colon and rectum.² Linkage studies have

*Correspondence: Dr JD Rioux, Montreal Heart Institute, Université de Montréal, 5000 Belanger Street, Montreal, Quebec, Canada H1T 1C8. Tel: +1 514 376 3330 ext. 2181; Fax: +1 514 376 1355; E-mail: rioux@broad.mit.edu

Received 1 July 2005; revised 7 February 2006; accepted 15 February 2006; published online 29 March 2006

revealed at least eight chromosomal regions (ie, *IBD1*–*IBD8*) harboring disease predisposition genes.^{3–7} Given the fact that in many multiplex families there is a mixture of CD and UC patients, it is believed that some susceptibility genes will be specific for CD or UC, whereas others will be in common to both. In a linkage scan of 158 Canadian affected sib-pair families, we identified a linkage peak of genome-wide significance on chromosome 19p (LOD score = 4.6) that appeared to confer susceptibility to both CD and UC.³ Two independent genome-wide linkage studies also provided supporting evidence for linkage to this region.^{4,5} Moreover, two recent meta-analyses of all published genome-wide scans^{6,7} identified linkage to IBD on 19p. These results suggest the existence of IBD susceptibility allele(s) in this region also known as the *IBD6* locus (OMIM 606674).

Data from genetic association studies and animal models, as well as gene expression studies, support a view where multiple pathways including apoptosis, immune-homing mechanisms, and bacterial antigens may all act together to produce the IBD phenotypes.^{8–10} Based on this paradigm, examination of the linkage interval on chromosome 19p, which encompasses ~28.5 Mb and contains 638 known genes,¹¹ reveals an especially rich array of candidate genes for susceptibility to IBD. Rather than testing all 638 genes, we conceived a strategy to systematically test the genetic variation characteristic of the most likely candidates for IBD susceptibility, based on functional and expression information. We selected 68 candidate genes in the 19p region by means of literature searches, gene ontology (GO), protein domain comparisons, and gene expression information. Among the functional groups represented in this gene set are: (a) receptors or ligands involved in cell migration/homing, such as *CD209* and *FCER2*, (b) transcriptional regulators of immune function, such as *PIASY*, (c) cell surface receptors involved in the modulation of inflammatory secondary messengers, such as *EDG6*, (d) costimulatory molecules for T cells or dendritic cells, such as the cluster of TNF ligands, (e) cytokine receptors, such as *IL12RB1*, and (f) essential elements of the antimicrobial immune defense system, such as *TRIF*. Given that these genes were chosen within a region with multiple sources of evidence for linkage to IBD, and based on our current understanding of IBD etiopathology, they are both positional and functional candidates.

To comprehensively evaluate the underlying genetic variation for each one of these genes, we constructed haplotypes using SNPs and statistical inference methods that exploit linkage disequilibrium metrics. Haplotypes capture common genetic patterns throughout the genome, thus serving as excellent surrogates for all common genetic variation.¹² Overall, we performed a step-wise association analysis consisting of a screening and a follow-up stage. During the screening stage, we typed SNPs across the selected candidate-gene regions in a family-based cohort.

We then determined the underlying haplotype structures for these candidate regions, and tested individual SNPs and multi-SNP haplotypes for association with disease. Thus, using a haplotype-based procedure we were able to assess association even at sites that were not necessarily typed. Of note, this study precedes the existence of the International HapMap Project, so that SNP selection was done based on the limited amount of information publicly available in dbSNP build 106 (September 2002). In the later stage of this candidate-gene study, alleles and haplotypes with suggestive evidence of association with IBD were followed up in a population-based cohort.

In addition to the candidate-gene screen, we directly examined four nonsynonymous coding polymorphisms on the *C3* (R102G and L314P) and *ICAM1* (G241R and K469E) genes, in four independent cohorts that included 2178 IBD cases. Three of these polymorphisms (*C3*-R102G, *ICAM1*-G241R, and *ICAM1*-K469E) had previously published evidence of association with IBD or subsets of IBD.^{13–18} Therefore, we specifically tested for replication of those variants with prior evidence of association, and combined all data in a pooled meta-analysis.

Materials and methods

DNA samples

We analyzed four cohorts of Caucasian origin in this study (cohorts A–D; Table 1). In the candidate-gene analysis, we used cohort A (Canadian family trios) for the screening stage, and cohort B (Canadian and Italian cases and controls), for the follow-up stage. None of these samples overlap with those of the previous study that identified this locus³ as these were no longer available. The Canadian samples in cohorts A and B come from multiple sites in the province of Quebec, and the Italian samples in cohort B were centrally collected at S. Giovanni Rotondo 'C.S.S.' Hospital. The Canadian and Italian case–control populations were combined for analysis to constitute cohort B because of their modest sizes, joint genotyping, and no statistically significant differences between allele frequencies in control individuals ($P=0.37$). In addition, we evaluated four nonsynonymous polymorphisms on *C3* and *ICAM1*, in all four cohorts. Cohorts C (extended families) and D (family trios and case–controls) are constituted by German patients. The extended families in cohort C were recruited consecutively during 1993–1998 through the Charité University Hospital in Berlin, and the Department of General Internal Medicine, Christian-Albrechts University in Kiel, while the family trios and cases in cohort D were invited through the German Crohn's and Colitis Foundation and collected through two rounds of recruitment during 2001–2003. For the German case–control study, we used sex- and age-matched healthy volunteers from the Kiel University Blood

Table 1 Description of DNA samples

Cohort	Study design	Experiment ^a	Sample description (no. of cases)	Country	Typing ^b
A	Trios	Screening stage of candidate-gene analysis and <i>C3</i> and <i>ICAM1</i> SNPs	180 IBD trios (150 CD, 30 UC)	Canada	SQNM
B	Case-control	Replication stage of candidate-gene analysis and <i>C3</i> and <i>ICAM1</i> SNPs	166 CD cases; 104 controls 177 IBD cases (85 CD, 92 UC); 103 controls	Canada Italy	SQNM
C	Family pedigrees	<i>C3</i> and <i>ICAM1</i> SNPs	851 extended families: 412 CD and 319 UC trios; 63 CD and 36 UC DSP	Germany	TaqMan
D	Trios Case-control	<i>C3</i> and <i>ICAM1</i> SNPs	288 IBD trios (177 CD, 111 UC); 537 IBD cases (178 CD, 359 UC); 885 controls	Germany	TaqMan

^aExperiment no. 1 is the two-stage candidate-gene analysis: Stage (1) screening of variants for functional candidate genes along chromosome 19p in cohort A, and Stage (2) replication of preliminary associations from stage 1 in cohort B. Experiment no. 2 is the evaluation of putative associations with nonsynonymous coding polymorphisms in the *C3* and *ICAM1* genes across cohorts.

^bGenotyping was performed using two platforms: SQNM, Sequenom at The Broad Institute (Cambridge, MA, USA); TaqMan, at The Christian-Albrechts University (Kiel, Germany).

Donation Program. Protocols for sample collection for all populations were approved by the corresponding local Institutional Review Boards. In all populations studied, the diagnosis of IBD and classification as CD or UC was confirmed by established clinical criteria, including radiological, histological, and endoscopic analysis.^{13,14} Phenotypic heterogeneity within the patient groups of individuals classified as having CD or UC undoubtedly exists as in all complex traits. This underlying phenotypic heterogeneity within disease categories is difficult to assess, and therefore may potentially be different across different patient cohorts. In order to formally test whether genetic heterogeneity existed between the Quebec and Italian samples (given that these were combined in our analyses) we compared allele frequencies for 14 unrelated markers (separated by a distance greater than 500 kb) typed in all control individuals. Specifically, no genetic heterogeneity was detected ($P=0.37$) when control allele frequencies were compared (frequencies for each marker were compared using the χ^2 test and $P=0.37$ corresponds to the sum of these χ^2 comparisons with 14 degrees of freedom).

Selection of candidate genes and SNPs

The selection of the candidate genes across chromosome 19p was carried out using publicly available annotation tools that included PubMed literature surveys, GO information, protein domain algorithms, and mining of microarray expression and tissue distribution data sets, such as the mammalian transcriptome collection by Su *et al.*¹⁵ Examples of criteria we used to select candidate genes are: using the terms 'gene' and 'inflammation' in a PubMed query followed by manual review of the resulting papers, 'defense response to pathogen' as a process query in GO, proteins with Toll and interleukin 1 receptor domains to query ProSite, or 'chromosome 19' and 'intestine' (tissue expression) to query Unigene.

For SNP selection, each gene locus – including 10–20 kb beyond the first and last exons – was mined for publicly available SNPs (dbSNP build 105; June 2002). We aimed for a target density of one SNP every 5 kb, and favored SNPs with both alleles present at least twice in different DNAs (ie, double-hit validated). This resulted in the selection of 789 SNPs for genotyping.

Genotyping and assessment of genotyping quality

Samples from cohorts A and B were typed using the Sequenom MassArray system as previously described,¹² and samples from cohorts C and D with the TaqMan allelic discrimination method.¹⁶ Assays were considered successful and genotype data were included in our analyses if they passed the following criteria: a minimum of 75% of all genotype calls were obtained, genotypes did not deviate from Hardy-Weinberg equilibrium (HWE; $P>0.01$), and no more than 0.2% mendelian errors (ME) were observed in family data (Supplementary Table A). Although the genotyping cutoff was 75%, the average call rate for this data set is 96.8% and the median corresponds to 98% (Supplementary Table A). SNPs were considered monomorphic and excluded from analysis if <1% of all individuals typed were heterozygous. The observed ME did not cluster in families or gene regions. Of 789 SNPs genotyped, 549 had successful genotyping assays, but only 465 were polymorphic in our populations, and so, were included in this analysis.

Construction of haplotypes and gene-based analysis

We constructed multimarker haplotypes using the expectation maximization¹⁷ statistical algorithm modified to process large data sets using the partition-ligation approach.¹⁸ Of the 465 polymorphic SNPs successfully genotyped, 396 SNPs mapped to gene regions for which at least 75% of the genomic extent (ie, the region that exons and introns span in the genome) of the correspond-

ing candidate gene was covered by SNPs, at a minimum SNP density of 1 per 15 kb. Therefore, we analyzed these SNPs as a candidate-gene SNP set. Of 68 candidates examined, 56 genes met the above-mentioned SNP coverage criteria and were analyzed on a per gene basis (Supplementary Table B). Genotypes are available upon request.

Genetic analysis

Association testing Association in family trios was assessed by the transmission disequilibrium test (TDT¹⁹), in extended pedigrees by the pedigree disequilibrium test (PDT²⁰), and in case-control groups by a standard χ^2 test carried out on a 2×2 contingency table. To combine data from trios and case-control populations, the number of risk alleles transmitted in trios or found in cases was reformatted as a mean or expected count (E_T and E_C), observed count (O_T and O_C), and variance (var_T and var_C). A combined Z -score was then calculated by summation as $Z = [(O_T + O_C) - (E_T + E_C)] / \sqrt{(\text{var}_T + \text{var}_C)}$. For analyses with prior evidence of association (ie, evaluation of *C3* and *ICAM1* polymorphisms), we obtained one-tailed P -values, as replication testing is limited to a single allele or haplotype, and therefore rejection of the null hypothesis (ie, no association) occurs only in one tail of the sampling distribution. Nominal one-tailed P -values were derived as the P -value corresponding to the standard normal cumulative distribution of a Z -score or the square root of a χ^2 score. To determine empirical significance of the results, we performed 1000 permutations per gene of the genotype data and compared our association results with the distribution of statistical test results from these permutations. Although none of the putative associations identified in the screen were replicated in our replication cohorts, we performed a directed analysis looking for epistasis with the *IBD5* haplotype. This was performed because van Heel *et al*²¹ have reported an epistatic interaction between the *CARD15*, *IBD5*, and *IBD6* loci, therefore suggesting the potential for greater evidence of association by examining these two together. No interactions between *IBD5* and *IBD6*, or between *CARD15* and *IBD6*, however, were detected in the current data set. It should be noted that van Heel and colleagues examined this question by partitioning linkage data whereas here we looked at the specific alleles in *IBD6* region that had putative association with the known *IBD5* risk allele/haplotype.

Meta-analysis We examined evidence for association of four nonsynonymous polymorphisms: C3-R102G, C3-L314P, *ICAM1*-G241R, and *ICAM1*-K469E, in a meta-analysis. In addition to our genotyping data, we evaluated allele frequencies for C3-R102G from one published study,²² *ICAM1*-G241R from three,^{23–25} and *ICAM1*-K469E from five independent reports.^{23–27} We computed

a pooled estimate of the odds ratio (OR) using the logit method, and then derived a confidence interval for the OR, as previously described for a series of unmatched case-control studies.²⁸ For trios, we used the relative risk as an estimate of the OR. For extended families constituted by trios and discordant sibling pairs (DSP), and analyzed by the PDT, we formatted the data (ie, parental contributions (transmitted or not transmitted) for trios, and number of siblings (affected or unaffected) for DSP, into 2×2 tables, and then proceeded as in case-control studies.

Results

Candidate genes and polymorphisms

Evidence for linkage to *IBD* on chromosome 19 is very compelling and indicates the presence of at least one *IBD* susceptibility gene on the p-arm. The 19p region is ~28.5 Mb and contains 638 genes (UCSC Browser, May 2004). Therefore, we designed a strategy to systematically test for association of genes with potential involvement in *IBD* pathogenesis based on known function and/or expression information. This strategy consisted of two sequential stages: a screening stage to identify potential variants associated with *IBD*, and a subsequent replication stage to corroborate preliminary findings (Figure 1). For the screening stage, we selected 68 candidate genes involved in processes such as inflammation, cell migration/homing, and antimicrobial immune defense response (see Supplementary Table B for gene descriptions). We selected 789 SNPs spanning these genes at an average density of 1 SNP every 5 kb and genotyped them in cohort A consisting of 180 *IBD* trios. Evaluation of genotyping data resulted in 475 SNP assays (61%) that passed quality metrics and were not extremely rare or monomorphic in this population, while 314 SNPs fell in any of the following overlapping groups: had <75% of genotypes ($n=138$; 17%), had excessive ME ($n=98$; 12%), were not in HWE ($n=77$; 10%), did not map unambiguously to the current genome assembly ($n=9$; 1%), had <1% heterozygous individuals ($n=166$; 21%). Common underlying genetic variation has to be accounted for to thoroughly assess whether a candidate gene is associated with *IBD*. Thus, we determined that a gene had to meet minimum SNP coverage to be included in this analysis. Thus, minimum coverage was established as reliable SNP assays covering at least 75% of the gene's transcript with a minimum SNP density of 1 per 15 Kb. The 56 candidate genes that met these criteria are listed in Supplementary Table B.

Screening stage

We constructed haplotypes using the genotypes of individuals in cohort A (example in Figure 2). On average, we observed 1.3 haplotype blocks (as defined in Gabriel *et al*¹²) per gene; each block with 3.4 SNPs and ~3.6 common haplotypes (ie, frequency >1%). We tested for association

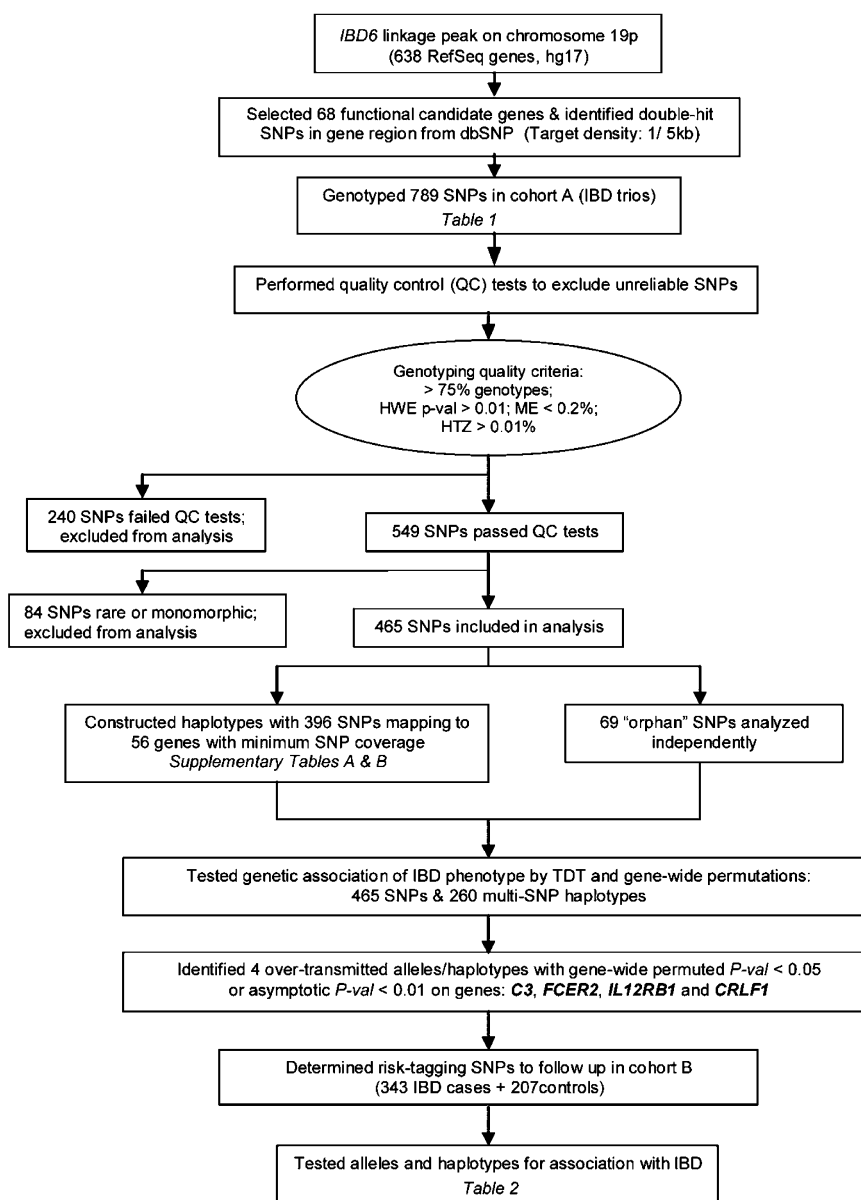


Figure 1 Overview for candidate-gene screen. The screen consisted of two stages: (1) screening of SNPs within and around loci bearing genes potentially involved in the pathogenesis of IBD for diseases association in IBD-affected family trios; and (2) replication of preliminary associations in IBD patients and controls. Orphan SNPs refers to those SNPs that fall within a gene locus that did not satisfy minimum SNP coverage criteria (see text), and therefore were analyzed separately, and not as part of a candidate gene. Abbreviations: HWE, Hardy-Weinberg Equilibrium; ME, mendelian errors; HTZ, heterozygous individuals; see Materials and methods for a complete definition.

to IBD on both individual alleles and multiallelic haplotypes, using the TDT and permutation analysis. Since the strength of association for a potential causal allele was unknown, we set relaxed thresholds for suggestive evidence of association: permuted $P < 0.05$ or asymptotic $P < 0.01$. Although a number of false positives were expected from this preliminary assessment, the built-in replication in our study design aimed to eliminate spurious associations. One haplotype on *CRLF1* (permuted $P = 0.014$; asymptotic $P = 0.001$), and three alleles on

FCER2 (permuted $P = 0.046$; asymptotic $P = 0.005$), *IL12RB1* (asymptotic $P = 0.006$), and *C3* (asymptotic $P = 0.007$), respectively, showed excessive transmission to affected probands (Table 2), which constituted suggestive evidence for association as defined by our thresholds.

Replication stage

We were therefore interested in testing whether we could replicate the putative association of the four genes identified in the initial screen, by testing for

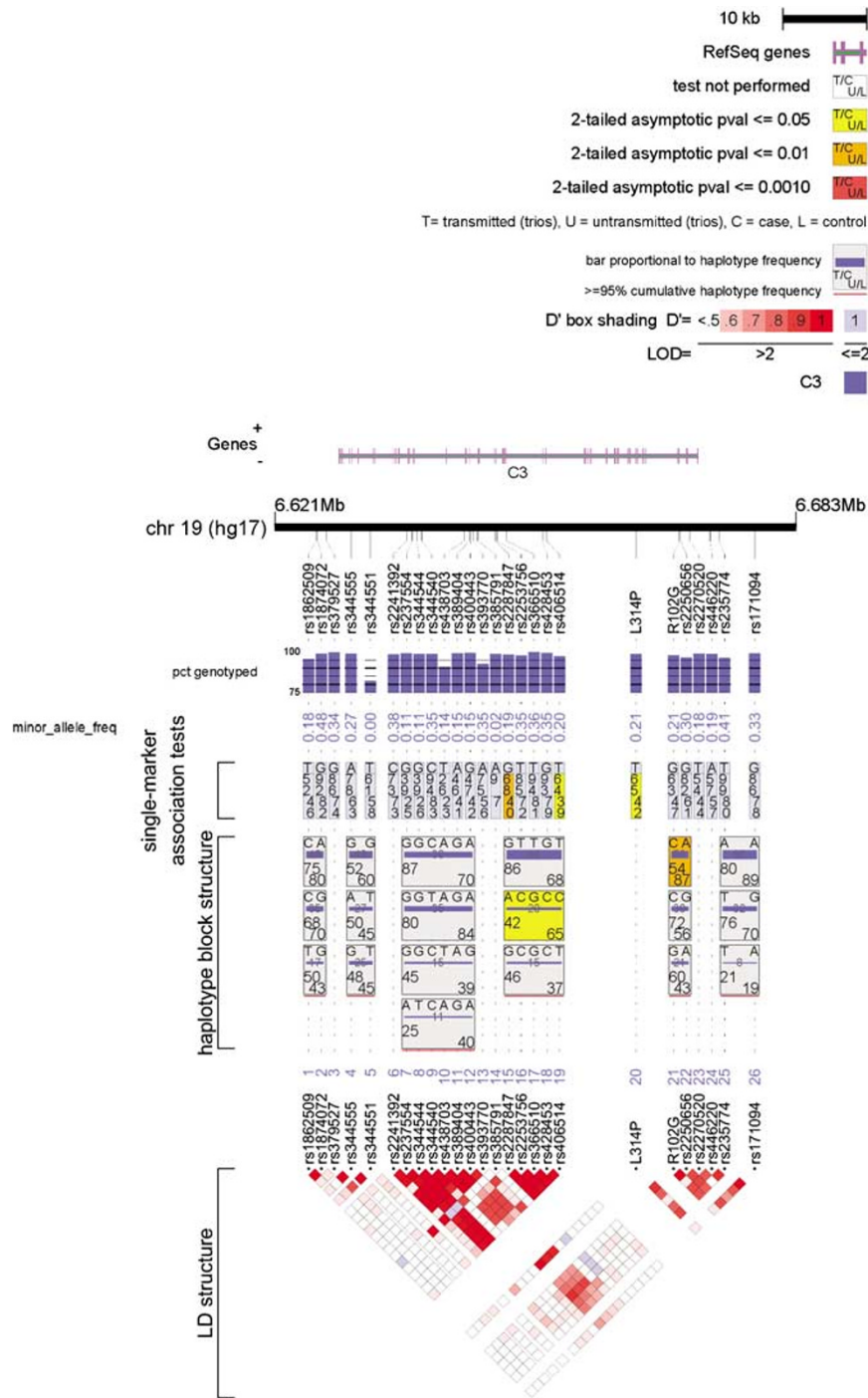


Figure 2 Haplotype structure and association analysis of *C3* variants. Haplotypes constructed with 26 polymorphisms (including R102G and L314P) typed in 180 Canadian IBD family trios. Image generated with LocusView software package.

association in an independent set of IBD samples. Specifically, we typed the five SNPs representing the alleles and/or haplotypes putatively associated in our screen (Table 2) in an independent set of 343 IBD cases and 207 controls (cohort B). The SNPs were chosen as they

were the actual SNPs with putative evidence for association in the screening stage either in a single- or multiple-marker test. Although we tested the exact same SNP or haplotype, none had significant evidence of replication (Table 2).

Table 2 Association results for best-scoring candidate genes

Gene	SNP(s)	Putative risk allele/haplotype	T-allele counts (cohort A)	U-allele counts (cohort A)	Two-tailed asymptotic P-value	Gene-wide permuted P-value	Allele frequency in cases (cohort B)	Allele frequency in controls (cohort B)	Overall genetic association
<i>C3</i>	rs2287847	G	68	40	0.007	0.208	0.80	0.83	Not replicated
<i>FCER2</i>	rs2277989	G	105	68	0.005	0.046	0.54	0.54	Not replicated
<i>IL12RB1</i>	rs2045387	G	78	47	0.006	0.119	0.79	0.81	Not replicated
<i>CRLF1</i>	rs2238649 /rs2238650	CA	73	39	0.001	0.014	0.44	0.47	Not replicated

T = transmitted; U = untransmitted.

Allele refers to SNP allele or multiallelic haplotype. During the screening stage of this study, 475 SNPs were tested for association in cohort A (180 IBD trios). The variants presented in this table had the highest uncorrected nominal and permuted *P*-values and were subsequently genotyped in cohort B (343 cases and 207 controls).

Table 3 Association analysis of *C3* and *ICAM1* SNPs

Gene	Marker	Allele	Disease	Cohort A		Cohort B		Cohort C		Cohort D	
				TDT	P-value (T:U)	Chi2	P-value (C:L)	PDT P-value ^a (PT:PU and A:N)	Z-score	P-value (T:U and C:L)	
<i>C3</i>	R102G ^b	Gly-102	IBD	0.063	(63:47)	0.258	(0.18:0.16)	0.016	(339:288 and 32:19)	0.326	(74:105 and 0.20:0.19)
			CD	0.074	(54:40)	0.057	(0.21:0.16)	0.034	(193:163 and 20:14)	0.039	(46:67 and 0.17:0.19) ^c
			UC	0.439	(9:7)	0.477	(0.11:0.16)	0.126	(146:125 and 12:6)	0.149	(28:38 and 0.22:0.19)
	L314P ^d	Pro-314	IBD	0.026	(65:42)	0.148	(0.17:0.14)	0.274	(319:290 and 31:19)	0.239	(73:97 and 0.18:0.19)
			CD	0.037	(56:36)	0.034	(0.19:0.14)	0.333	(182:165 and 21:15)	0.011	(41:64 and 0.15:0.19) ^c
			UC	0.439	(9:6)	0.477	(0.12:0.14)	0.540	(137:125 and 10:5)	0.572	(32:33 and 0.20:0.19)
<i>ICAM1</i>	G241R ^b	Arg-241	IBD	0.305	(19:16)	0.444	(0.06:0.06)	0.193	(182:179 and 25:13)	0.206	(64:71 and 0.13:0.14)
			CD	0.289	(16:13)	0.349	(0.05:0.06) ^c	0.022	(113:96 and 12:8)	0.024	(34:49 and 0.11:0.14) ^c
			UC	0.500	(3:3)	0.145	(0.08:0.06)	0.138	(69:83 and 13:6)	0.312	(30:22 and 0.14:0.14)
	K469E ^b	Glu-469	IBD	0.444	(88:90)	0.121	(0.53:0.57)	NT	NT	NT	NT
			CD	0.403	(72:75)	0.242	(0.55:0.57)	NT	NT	NT	NT
			UC	0.431	(16:15)	0.050	(0.49:0.57) ^c	NT	NT	NT	NT

For trios: T = transmitted allele counts; U = untransmitted allele counts; for case–controls: C = allele frequency in cases; L = allele frequency in controls; for extended pedigrees: PT = parental contributions transmitted (trios); PU = parental contributions untransmitted (trios); A = allele counts in affected sib of DSP; N = allele counts in unaffected sib of DSP; NT = Not tested.

^aFor PDT, only the *P*-value corresponding to the PDT-avg statistic (Martin *et al* 2001) is shown.

^bUncorrected one-tailed *P*-values provided prior evidence for association of the allele tested.

^cThe directionality of these potential associations is reversed (ie, allele in question is undertransmitted or more frequent in controls/unaffected individuals).

^dUncorrected two-tailed *P*-values provided for this variant as no prior evidence for association of this allele existed before this study.

Evaluation of published putative associations

In a parallel effort to identify IBD susceptibility alleles in the 19p region, we evaluated two genes (*C3* and *ICAM1*) in this region for which putative associations had previously been reported in the literature.

C3 An association between *C3* and CD was reported over two decades ago,²² but was never confirmed through replication. Specifically, a phenotypic *C3* variant characterized by a faster migration rate in gel electrophoresis compared to its alternative allelic form was reported to be associated with CD. This phenotypic variant was later found to be encoded by the Gly-102 allele of the nonsynonymous R102G polymorphism.²⁹ We noted that a second nonsynonymous polymorphism, L314P, was located 5 kb from R102G. We determined that these two SNPs are in strong LD with one another ($D' = 0.9$ and

$r^2 = 0.77$ between Gly-102 and Pro-314) and therefore concluded that L314P was an equally likely candidate. We typed both SNPs in four independent cohorts: A (180 family trios), B (347 cases and 207 controls), C (851 extended families), and D (288 trios; 537 cases and 837 controls).

We observed modest but significant association between Gly-102 and IBD (one-tailed $P = 0.016$) or CD (one-tailed $P = 0.034$) in cohort C. In addition, association between Pro-314 and IBD was found in cohort A (asymptotic $P = 0.026$), and between Pro-314 and CD in cohorts A (asymptotic $P = 0.037$), and B (asymptotic $P = 0.034$). Importantly, although not reaching statistical significance, similar trends of association for both variants with both IBD and CD were observed in cohorts A–C. This trend was not observed in cohort D, and in fact, association was observed for the alternate alleles of these variants (Table 3).

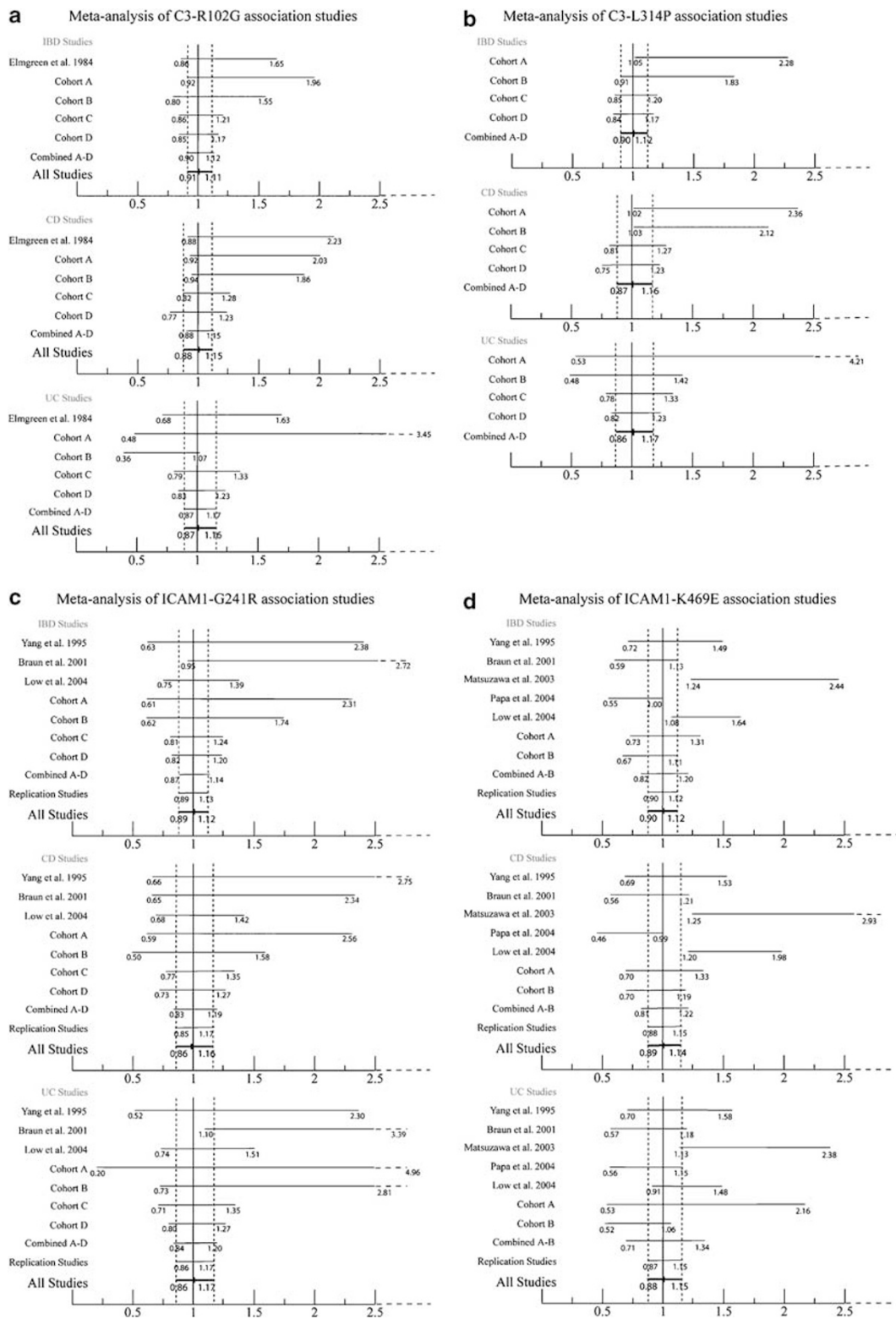


Figure 3 Meta-analysis of *C3* and *ICAM1* association studies. Polymorphisms tested: (a) C3-R102G, (b) C3-L314P, (c) ICAM1-G241R, and (d) ICAM1-K469E. Horizontal bars and delimiting numbers represent 95% confidence intervals for the odds ratio (OR) of a particular study. The dashed vertical lines correspond to the confidence interval for the pooled estimate of the OR for all studies. *Replication studies* refers to all studies subsequent to the original study (including the current study, which involves cohorts A–D).

Overall, meta-analysis of all four cohorts together did not support association of either *C3* variant with IBD, CD, or UC (Figure 3).

ICAM1 Three studies reported putative association of the Gly-241 allele of the *ICAM1*-R241G polymorphism, with IBD or subsets of IBD patients, at various degrees of statistical significance^{23–25}. Similarly, claims of association were made for *ICAM*-K469E: two studies found association of Glu-469,^{25,26} a third study pointed Lys-469 as the associated allele,²⁷ and two more studies found no association with either allele.^{23,24} To evaluate whether these two polymorphisms are associated with IBD, we typed them in our cohorts: G241R in cohorts A–D totaling 2178 IBD cases, and K469E in cohorts A and B, totaling 523 IBD cases.

We found modest evidence for association of the Arg-241 allele with CD in 475 cases from cohort C (one-tailed $P=0.022$; Table 3); however, we also found evidence favoring association of the alternate Gly-241 allele with CD in 355 cases from cohort D (one-tailed $P=0.024$; Table 3). Meta-analysis of all four cohorts did not support association of the G241R polymorphism with IBD, CD, or UC (Figure 3).

As for the K469E polymorphism, we did not find evidence of association of the Glu-469 with IBD or CD, and found only marginal evidence for association with UC in 92 cases from cohort B (one-tailed $P=0.05$; Table 3). The combined analysis of cohorts A and B did not support association of this polymorphism with IBD, CD, or UC (Figure 3).

Meta-analysis To increase the statistical power of our association analysis for the *C3* and *ICAM1* polymorphisms, we combined our genotyping data with all the published reports in a pooled meta-analysis (Figure 3). Meta-analysis of 2303 IBD cases obtained after combining our four cohorts (Table 1) with the cohort studied by Elmgreen *et al*,²² did not support association of the *C3*-R102G polymorphism with IBD, CD, or UC. Similarly, meta-analysis of the *ICAM1*-G241R polymorphism in 3113 IBD cases resulting from combining our four cohorts with those analyzed in three published studies,^{23–25} did not support association of this variant. Finally, meta-analysis of 1308 IBD cases, including two of our cohorts (A and B) and five prior studies,^{23–27} did not provide evidence for association of the *ICAM1*-K469E polymorphism with IBD, CD, or UC.

Discussion

The genetic analysis of common, multifactorial diseases, such as IBD, has proven to be a challenging task. Eight chromosomal regions or linkage peaks potentially conferring susceptibility to IBD have been identified over the past

decade. Our challenge was to identify the genetic variation responsible for IBD susceptibility in the *IBD6* locus, a ~28.5Mb linkage peak on chromosome 19p containing over 600 genes. Viable methods to achieve our goal included comprehensive association mapping across the entire region and candidate-gene association.³⁰ Given our finite sample sizes (Table 1), and technical (hence cost) limitations to cover such large chromosomal region, we chose a candidate-gene approach. We exploited the characteristics of our cohorts without sacrificing statistical power by implementing a two-stage study design. First, we used a family-based cohort to screen the candidate-gene variants, thus protecting against genetic stratification. Second, we used a case–control cohort with larger size to confirm preliminary associations. The screening stage of this approach primarily produced the most likely candidates for association with IBD among 56 functional candidate genes analyzed. However, our screening cohort had relatively modest power to confidently discard false negatives (for instance, the power to rule out association for a variant such as *C3*-L314P, which is present in ~15% of the general population, was only about 50%). To accommodate this limitation, we relaxed the statistical threshold for suggestive association in our primary screen (ie, asymptotic $P<0.01$). We anticipated that this measure would result in a higher number of false positives than if we had used a more stringent cutoff. Accordingly, we included a built-in replication in our study design to discard false positives as spurious associations with at least 95% certainty, based on power calculations for the size of our replication cohort.³¹ In the screening stage we obtained preliminary evidence for association of four independent genetic variants (ie, three alleles and one multimarker haplotype). We attempted to replicate these suggestive findings in an independent cohort of 343 cases and 207 controls; however, we did not find any supporting evidence of association and therefore concluded that these are likely to have been false-positive results.

In addition, we attempted to specifically replicate the two previously published putative associations to genes in this region: *C3* and *ICAM1*. This was carried out by testing for association in a total of 2178 IBD cases, however, no convincing evidence of association was observed. Combining our data with the published results also failed to provide evidence of association of either *C3* or *ICAM1* with CD, UC, or IBD. Several reasons that may explain the failure to replicate genetic association studies, including allelic heterogeneity, true variation in disease association between populations, modifying genetic and/or environmental factors, statistically underpowered small sample sizes, and a low prior probability of obtaining a true result. The most parsimonious explanation, given our experimental design, for the lack of replication of our initial putative association results would be fortuitous

initial false-positive results. Specifically, the experimental design that we chose for the current study was based on power calculations for alleles of various frequency (5–30%) and risk ratios (2–4) that would be consistent with the magnitude of linkage ($\lambda_s=1.35$) seen in our original scan of 158 affected sibling pairs (LOD 4.6) and what probably constitutes a more accurate, modest picture of significant linkage ($\lambda_s=1.13$) in two large meta-analyses of IBD genome-wide linkage scans.^{3,6,7} These models justified the use of a cost efficient, modestly sized, multi-stage association study – certainly not an approach recommended for most susceptibility alleles that have typically given more modest evidence of linkage. The failure of the current study to identify significant association in this region of strong linkage would therefore suggest that our candidate-gene selection strategy failed to identify the appropriate gene and/or the causal allele(s) are significantly more modest than predicted and were not detected in the initial screen. Given these results, as well as the recent advances in our knowledge of the haplotype patterns across most of the genome and powerful approaches to select informative SNPs^{32–34} for powerful association studies, the next step in our examination of the IBD6 locus will likely entail screening a much larger set of samples with a dense set of SNPs across the entire locus. Finally, although the current effort did not identify genetic variation explaining IBD6, we hope that the data presented here (alone and in conjunction with HapMap), will be useful in the study of other related genetic diseases, such as celiac disease, another inflammatory disease with a linkage peak also mapping to the 19p region.³⁵

Databases and links to software

PubMed (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>)

Gene Ontology (<http://www.geneontology.org/>)

dbSNP (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=SNP>)

UCSC Genome Browser (<http://genome.ucsc.edu/>)

LocusView (<http://www.broad.mit.edu/mpg/locusview/>)

Acknowledgements

We thank the IBD patients and their families for their collaboration. We also extend our thanks to Tanja Wesse (Christian-Albrechts-University, Kiel, Germany) for technical assistance, and members of the Inflammatory Disease Research Group (The Broad Institute, Cambridge, USA) for helpful discussions. This study was supported by a research fellowship from the Crohn's and Colitis Foundation of America to MKT, the German Federal Department of Research and Education through the German National Genome Research Network (NGFN2) and the German Research Council (research group polygenic diseases; DFGFOR423) to SS, the Italian Health Minister (RC0503GA22) to VA, and grants from the NIH/NIDDK and The Broad Foundation to JDR.

References

- 1 Kornbluth A, Sachar DB, Salomon P: Crohn's disease; in Feldman M, Scharcchmidt BF, Sleisenger MH (eds): *Sleisenger and Fordtrans's Gastrointestinal and Liver Disease*. London: WB Saunders, (1998), vol 2, pp 1708–1734.
- 2 Jewell DP: Ulcerative colitis; in Feldman M, Scharcchmidt BF, Sleisenger MH (eds): *Sleisenger and Fordtrans's Gastrointestinal and Liver Disease*. London: WB Saunders, (1998), vol 2, pp 1735–1761.
- 3 Rioux JD, Silverberg MS, Daly MJ *et al*: Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am J Hum Genet* 2000; **66**: 1863–1870.
- 4 Cho JH, Nicolae DL, Gold LH *et al*: Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: evidence for epistasis between 1p and IBD1. *Proc Natl Acad Sci USA* 1998; **95**: 7502–7507.
- 5 Ma Y, Ohmen JD, Li Z *et al*: A genome-wide search identifies potential new susceptibility loci for Crohn's disease. *Inflamm Bowel Dis* 1999; **5**: 271–278.
- 6 van Heel DA, Fisher SA, Kirby A, Daly MJ, Rioux JD, Lewis CM: Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. *Hum Mol Genet* 2004; **13**: 763–770.
- 7 Williams CN, Kocher K, Lander ES, Daly MJ, Rioux JD: Using a genome-wide scan and meta-analysis to identify a novel IBD locus and confirm previously identified IBD loci. *Inflamm Bowel Dis* 2002; **8**: 375–381.
- 8 Fox JG, Rogers AB, Whary MT *et al*: Gastroenteritis in NF-kappaB-deficient mice is produced with wild-type *Campylobacter jejuni* but not with *C. jejuni* lacking cytolethal distending toxin despite persistent colonization with both strains. *Infect Immun* 2004; **72**: 1116–1125.
- 9 Fiocchi C: Inflammatory bowel disease: etiology and pathogenesis. *Gastroenterology* 1998; **115**: 182–205.
- 10 Blumberg RS, Saubermann LJ, Strober W: Animal models of mucosal inflammation and their relation to human inflammatory bowel disease. *Curr Opin Immunol* 1999; **11**: 648–656.
- 11 Grimwood J, Gordon LA, Olsen A *et al*: The DNA sequence and biology of human chromosome 19. *Nature* 2004; **428**: 529–535.
- 12 Gabriel SB, Schaffner SF, Nguyen H *et al*: The structure of haplotype blocks in the human genome. *Science* 2002; **296**: 2225–2229.
- 13 Lennard-Jones JE: Classification of inflammatory bowel disease. *Scand J Gastroenterol Suppl* 1989; **170**: 2–6; discussion 16–19.
- 14 Podolsky DK: Inflammatory bowel disease (1). *N Engl J Med* 1991; **325**: 928–937.
- 15 Su AI, Cooke MP, Ching KA *et al*: Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci USA* 2002; **99**: 4465–4470.
- 16 Hampe J, Cuthbert A, Croucher PJ *et al*: Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 2001; **357**: 1925–1928.
- 17 Excoffier L, Slatkin M: Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995; **12**: 921–927.
- 18 Qin ZS, Niu T, Liu JS: Partition–ligation–expectation–maximization algorithm for haplotype inference with single-nucleotide polymorphisms. *Am J Hum Genet* 2002; **71**: 1242–1247.
- 19 Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; **52**: 506–516.
- 20 Martin ER, Monks SA, Warren LL, Kaplan NL: A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am J Hum Genet* 2000; **67**: 146–154.
- 21 Van Heel DA, Dechairo BM, Dawson G *et al*: The IBD6 Crohn's disease locus demonstrates complex interactions with CARD15 and IBD5 disease-associated variants. *Hum Mol Genet* 2003; **12**: 2569–2575.
- 22 Elmgreen J, Sorensen H, Berkowicz A: Polymorphism of complement C3 in chronic inflammatory bowel disease. Predominance

- of the C3F gene in Crohn's disease. *Acta Med Scand* 1984; **215**: 375–378.
- 23 Yang H, Vora DK, Targan SR, Toyoda H, Beaudet AL, Rotter JI: Intercellular adhesion molecule 1 gene associations with immunologic subsets of inflammatory bowel disease. *Gastroenterology* 1995; **109**: 440–448.
- 24 Braun C, Zahn R, Martin K, Albert E, Folwaczny C: Polymorphisms of the ICAM-1 gene are associated with inflammatory bowel disease, regardless of the p-ANCA status. *Clin Immunol* 2001; **101**: 357–360.
- 25 Low JH, Williams FA, Yang X *et al*: Inflammatory bowel disease is linked to 19p13 and associated with ICAM-1. *Inflamm Bowel Dis* 2004; **10**: 173–181.
- 26 Matsuzawa J, Sugimura K, Matsuda Y *et al*: Association between K469E allele of intercellular adhesion molecule 1 gene and inflammatory bowel disease in a Japanese population. *Gut* 2003; **52**: 75–78.
- 27 Papa A, Pola R, Flex A *et al*: Prevalence of the K469E polymorphism of intercellular adhesion molecule 1 gene in Italian patients with inflammatory bowel disease. *Dig Liver Dis* 2004; **36**: 528–532.
- 28 Morris JA, Gardner MJ: Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. *Br Med J (Clin Res Ed)* 1988; **296**: 1313–1316.
- 29 Botto M, Fong KY, So AK, Koch C, Walport MJ: Molecular basis of polymorphisms of human complement component C3. *J Exp Med* 1990; **172**: 1011–1017.
- 30 Hirschhorn JN, Daly MJ: Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005; **6**: 95–108.
- 31 Purcell S, Cherny SS, Sham PC: Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003; **19**: 149–150.
- 32 Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P: A haplotype map of the human genome. *Nature* 2005; **437**: 1299–1320.
- 33 Johnson GC, Esposito L, Barratt BJ *et al*: Haplotype tagging for the identification of common disease genes. *Nat Genet* 2001; **29**: 233–237.
- 34 de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D: Efficiency and power in genetic association studies. *Nat Genet* 2005; **37**: 1217–1223.
- 35 Van Belzen MJ, Meijer JW, Sandkuijl LA *et al*: A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterology* 2003; **125**: 1032–1041.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)