

## INFLAMMATORY BOWEL DISEASE

## Genotype-phenotype analysis of the Crohn's disease susceptibility haplotype on chromosome 5q31

A Armuzzi, T Ahmad, K-L Ling, A de Silva, S Cullen, D van Heel, T R Orchard, K I Welsh, S E Marshall, D P Jewell

Gut 2003;52:1133-1139

See end of article for authors' affiliations

Correspondence to:  
A Armuzzi,  
Gastroenterology Unit,  
Gibson Laboratories,  
University of Oxford,  
Radcliffe Infirmary,  
Woodstock Rd, Oxford  
OX2 6QX, UK;  
clearmuzzi@yahoo.com

Accepted for publication  
6 March 2003

**Background and aims:** Recent molecular data suggest that genetic factors may underlie the disease heterogeneity observed in both ulcerative colitis (UC) and Crohn's disease (CD). A locus on chromosome 5q has been implicated in susceptibility to CD, and recently refined by linkage disequilibrium mapping to a conserved 250 kb haplotype (5q31). No data regarding the contribution of this locus to clinical phenotype exist. In this case control study, we investigated the contribution of this haplotype to both susceptibility and phenotype of CD and UC.

**Patients and methods:** We studied 330 Caucasian CD and 457 UC patients recruited from a single UK centre. Association with disease susceptibility and phenotype was analysed with haplotypes reconstructed from three single nucleotide polymorphisms chosen to span this susceptibility region. Evidence for possible genetic epistasis between *IBD5* and *NOD2/CARD15* was sought.

**Results:** Linkage disequilibrium across this region was confirmed, with two haplotypes comprising 88% of all chromosomes. Susceptibility to CD, but not to UC, was associated with homozygosity for a common haplotype, H2 ( $p_c=0.002$ ; relative risk (RR) 2.0). Genotype-phenotype analyses demonstrated that this association was particularly strong in patients with perianal disease ( $p_c=0.0005$ ; RR 1.7), especially in individuals homozygous for this haplotype ( $p_c=0.0005$ ; RR 3.0). Importantly, no association with H2 was found in 186 patients without perianal disease. No evidence of epistasis between *IBD5* and *NOD2/CARD15* was demonstrated.

**Conclusions:** The *IBD5* risk haplotype is associated with CD only. Genotype-phenotype analysis reveals that the strongest association is observed in patients with perianal CD. While the precise gene involved is unclear, these data provide further molecular evidence for a genetic basis of the clinical heterogeneity of CD.

The chronic idiopathic inflammatory bowel diseases (IBD; MIM 601458) Crohn's disease (CD; MIM 266600) and ulcerative colitis (UC; MIM 191390) are a significant cause of morbidity in the Western world, with a combined prevalence of 200/100 000 in the UK.<sup>1</sup> Epidemiological, clinical, and molecular studies have provided strong evidence for the role of genetics in determining susceptibility to disease. Genome wide scans have implicated a number of susceptibility loci: some influence susceptibility to IBD overall while others confer susceptibility to either CD or UC.<sup>2-10</sup> Furthermore, epidemiological evidence suggests that genetic factors also determine the clinical heterogeneity characteristic of both CD and UC, with high concordance rates for age at presentation, location, and behaviour of disease in multiply affected families.<sup>11-14</sup> This is supported by a growing body of molecular data. With reference to CD, six independent groups including our own have now reported an association between variants in the *NOD2/caspase recruitment domain containing protein 15* (*CARD15*) susceptibility gene on chromosome 16 (*IBD1*) and ileal disease only.<sup>15-20</sup> We have additionally shown that specific extended haplotypes across the HLA region (*IBD3*) are associated with colonic and perianal disease.<sup>15</sup> These molecular data increasingly suggest that CD is not a single disorder but a limited heterogeneous group of oligogenic disorders manifesting similar but specific clinical features. It is likely that the precise phenotype that an individual manifests depends upon the specific interaction of one or more of a few genetic variants with an unidentified luminal antigen.

Genome wide scans have identified a number of other putative susceptibility loci for CD but the contribution of these areas to disease phenotype is unknown. Linkage to an area on chromosome 5q31-33 has been demonstrated in two studies

for CD only,<sup>6,8</sup> although few UC affected families were studied. Fine mapping of this locus, designated *IBD5*, defined an area that achieved genome wide significance (LOD score 3.9) in families characterised by early age of onset of CD.<sup>8</sup> Further characterisation of this area by linkage disequilibrium mapping identified a single highly conserved 250 kb haplotype associated with CD.<sup>21</sup> This susceptibility haplotype includes 11 single nucleotide polymorphic variants found exclusively on this haplotype, all of which demonstrated significance in a transmission disequilibrium test. Identifying the causal mutation on this haplotype has been hampered by both the degree of linkage disequilibrium and the density of immunoregulatory genes encoded, including the cytokine gene cluster (interleukin (IL)-4, IL-5, and IL-13).

The contribution of this locus to disease phenotype remains unknown. In this case control study, we determine the association of the *IBD5* risk haplotype with specific disease phenotypes in a large British cohort of accurately defined CD and UC patients.

## METHODS

## Patients and controls

This study was approved by the Central Oxford Research Ethics Committee (COREC 00.083) and written informed consent was obtained from all participants. A total of 330 Caucasoid

**Abbreviations:** CD, Crohn's disease; *CARD15*, caspase recruitment domain containing protein 15; IBD, inflammatory bowel disease; IL, interleukin; PAR, population attributable risk; RR, relative risk; SNP, single nucleotide polymorphism; UC, ulcerative colitis.

**Table 1** Primers used for sequence specific polymerase chain reaction

Polymorphism	Specificity	Amplicon size (bp)	Sense primer	Antisense primer
5q31	IGR2060a_1	C	ggTgCAgCCTggAgCTC	AgCTCAgTCATTACCTTgCg
		G	ggTgCAgCCTggAgCTC	AgCTCAgTCATTACCTTgCC
	IGR2198a_1	G	ATCTCTCAgTAGACgAACg	CATgAAgAAgggACATAACC
		C	ATCTCTCAgTAGACgAACg	CATgAAgAAgggACATAACC
	IGR3096a_1	C	CCCAAACATCTggAgAAC	gCgCCTTCTAATTCATCC
		T	CCCAAACATCTggAgAAT	gCgCCTTCTAATTCATCC
NOD2/CARD15	R702W	C	TgAgAAggCCCTgCTCC	AgAgTTgTAgTCCAgCTgCAg
		T	CTgAgAAggCCCTgCTCT	AgAgTTgTAgTCCAgCTgCAg
	G908R	G	ggCCTTTTCAGATTCTggg	gACATTTCCAAGTCACCCAg
		C	ggCCTTTTCAGATTCTggC	gACATTTCCAAGTCACCCAg
	1007fsinsC	—	CCCTCTgCAggCCCT	AACCgCAgAAggTCTgATC
		Cins	CCCTCTgCAggCCCC	AACCgCAgAAggTCTgATC

patients with CD and 457 with UC were recruited from a single tertiary referral centre. Of the CD patients, 240 had previously been included in a study of IBD1 and IBD3.<sup>15</sup> Diagnosis of IBD was determined by conventional clinical, radiological, endoscopic, and histological criteria.<sup>22</sup> Phenotypic details were obtained by retrospective case note review for all patients between January 2001 and March 2002 by two investigators (TA, AA). Duration of follow up was defined as the interval between diagnosis and case note review. Details regarding ethnicity, family history, and smoking history were further supplemented by a patient completed postal questionnaire. Ethnicity was defined as Jewish if three or more grandparents were Jewish. Only one member from multiply affected families was included.

#### CD phenotype

CD phenotype was classified by age of diagnosis, location, and behaviour of disease, as previously described.<sup>15</sup> Briefly, location of disease was classified by current or past history of disease at three sites (ileal, colonic, perianal). Perianal disease was defined by the presence of perianal abscesses, fistulae, or ulcers, but not by the presence of skin tags. Disease behaviour was classified by current or past behavioural types. The presence of stenotic disease was defined by surgical history or by small bowel enema. Fistulating disease was defined by the presence of an abnormal communication between two epithelial surfaces. This included the presence of ischio-rectal, inter-sphincteric, and pelvic abscesses if they connected with an epithelial surface, either spontaneously or following surgical intervention. However, simple perianal abscesses were excluded from this definition.

#### UC phenotype

UC phenotype was classified by disease extent and disease severity. Disease extent was defined by the most proximal point of inflammation identified at colonoscopy or barium enema. The splenic flexure was used as a landmark to distinguish limited from extensive disease. If macroscopic and microscopic extent at the same time point were discordant, the microscopic extent was recorded. For individuals who had undergone a colectomy, disease extent was defined by microscopic extent in the resection specimen. Disease severity was judged by the need for colectomy for failed medical therapy and not by the use of colitis activity indices or the requirement for immunosuppressant drugs. Patients with indeterminate colitis or colorectal cancer complicating UC were not eligible for inclusion.

The control group of 870 individuals was randomly selected from anonymised blood samples of two independent populations collected from general practitioner health screening

clinics in Bedfordshire (n=516, OXCHECK study<sup>23</sup>) and Oxfordshire (n=354). These two groups were identical with respect to age, sex, smoking history, and genotype (data not shown) and were therefore combined for all analyses. Although none of the control subjects had IBD, no information regarding their family history was available.

#### Genotyping

We studied three single nucleotide polymorphisms (SNPs) (IGR2060a\_1 C/G, IGR2198a\_1 G/C, and IGR3096a\_1 C/T) previously shown by a Canadian group to be found uniquely on the *IBD5* risk haplotype (web site: <http://www-genome.wi.mit.edu/humgen/IBD5>). In order to confirm haplotype structure and extent of linkage disequilibrium in our British population, these SNPs were selected by their position at both ends and the middle of the 250 kb risk haplotype. These SNPs, which are not thought to be functionally relevant, were used here as markers of the CD associated risk haplotype.

All genotyping was carried out using polymerase chain reaction-sequence specific primers.<sup>24</sup> Primers were designed using published genomic sequences (GenBank accession No 18562169). CD patients and controls were also typed for polymorphisms in the NOD2/CARD15 gene (R702W, G908R, and 1007fsinsC) using previously described primers (table 1).<sup>15</sup> All reactions contained control primers to verify appropriate amplification and electrophoresis. A single thermocycling programme was used for all reactions. The products were electrophoresed on 1% agarose gels with ethidium bromide, viewed under ultraviolet light, and an image recorded digitally.

#### Data analysis

In the absence of parental genotype data, the three SNPs spanning the IBD5 haplotype were constructed into haplotypes using a statistical haplotype reconstruction method, implemented in the computer software PHASE.<sup>25</sup> This software was recompiled to run under the Macintosh OS10 operating system. For the construction haplotypes, 30 000 iterations of the PHASE algorithm were used, each comprising 100 runs through the Markov chain. Each haplotype was then analysed for association with CD and UC susceptibility and phenotype.

Phenotype-genotype associations were analysed by the  $\chi^2$  test, using the Knowledge Seeker program (Angoss, Guildford, Surrey, UK). A Bonferroni correction factor was applied where appropriate, and when applied is indicated by  $p_c$ . The population attributable risk percentage (PAR%) was calculated according to the method of Schlesselman,<sup>26</sup> assuming that the frequency of all variants in the control population reflects that of the general population.

**Table 2** Demographic and clinical characteristics of 330 Crohn's disease patients

Total No	330
Sex (M/F)	137/193
Age at diagnosis (y)	
Mean	28.1
Median	24.6
Range	3.0–82.2
Interquartile range	20.2; 31.9
Follow up (y)	
Mean	16.9
Median	14.3
Range	0.42–56.8
Interquartile range	8.5; 24.0
Jewish (%)	3 (0.9)
Family history of IBD in 1st or 2nd degree relative (%)	118 (35.7)
Smoking habit (%)	
Never	131 (39.7)
Ex	112 (33.9)
Current	87 (26.4)
Disease location (%)	
Ileal	255 (77.3)
Colonic	201 (60.9)
Perianal	144 (43.6)
No perianal	186 (56.4)
Disease behaviour (%)	
Stenotic disease	237 (71.8)
Fistulating	118 (35.8)
Severe inflammatory colonic disease	65 (19.7)
Extraintestinal manifestations (%)	
Type 1 peripheral arthropathy	45 (13.6)
Type 2 peripheral arthropathy	15 (4.5)
Ankylosing spondylitis	7 (2.1)
Erythema nodosum	31 (9.4)
Pyoderma gangrenosum	5 (1.5)
Uveitis	16 (4.8)
Primary sclerosing cholangitis	6 (1.8)

## RESULTS

A total of 330 patients with CD (table 2), 457 patients with UC (212 men and 245 women), and 870 controls were studied. Mean age at diagnosis for CD and UC patients was 28.1 years (range 3.0–82.2) and 33.8 years (range 0.8–80.3), respectively. Median duration of follow up for the entire cohort of patients was 14.5 years (range 0.9–59.9). There was no significant difference between cases and controls with respect to age or sex. A family history of IBD (UC or CD) was reported by 118 (35.7%) of CD patients and 102 (22.3%) of UC patients.

### Disease susceptibility

We studied three common SNPs on the previously reported *IBD5* CD associated haplotype. As previously described, each of these three SNPs was significantly associated with susceptibility to CD only (IGR2060a\_1 C:  $p_c=0.0003$ , relative risk (RR) 1.4; IGR2198a\_1 G:  $p_c=0.0003$ , RR 1.4; IGR3096a\_1 C:  $p_c=0.0001$ , RR 1.5) (table 3).

Haplotype analysis is potentially more informative than analysis of single polymorphisms as it facilitates surveillance of variation within a region of linkage disequilibrium.<sup>27</sup> Thus haplotypes were constructed from these three SNPs. In the absence of parental data, the association of one SNP with another (phase) was inferred using statistical computer software.<sup>25</sup> The probability of accurate haplotype designation was >90% for 99% of chromosomes. Eight haplotypes were identified, of which four occurred at a frequency of more than 2% in the overall population. Two haplotypes (designated H1 and H2, table 4) accounted for 87.7% of all chromosomes, similar to the previous findings in a Canadian cohort.<sup>21</sup> Only haplotype 2 (H2) was significantly associated with CD overall (46.8% v 38.9%;  $p=0.0004$ ,  $p_c=0.003$ ) (table 4). Conversely, H1 (comprising alleles absent on the risk haplotype) was

negatively associated with CD (cases 39.8% v controls 48.8%;  $p=0.00007$ ,  $p_c=0.0006$ ). No difference in haplotype frequencies was observed between UC patients and controls (table 4).

In order to investigate the existence of a haplotype dose effect, further analysis was performed (table 5). In CD patients overall, the risk associated with the H2 haplotype was only seen in homozygotes (homozygotes 23.0% v 14.7% ( $p_c=0.001$ , RR=2.0, 95% confidence interval (CI) 1.4–2.8); heterozygotes 47.6% v 48.4% (NS)). The PAR% of the risk haplotype H2 for CD was estimated to be 20.3%.

### CD phenotype

#### Disease location

Epidemiological studies suggest that anatomical location of disease is a stable defining characteristic in an individual with CD.<sup>28</sup> In this study, 63.3% of patients had disease at more than one site. Disease was located to the ileum, colon, and perianal regions in 77.3%, 60.9%, and 43.6% of patients, respectively (table 2).

The highest frequency of haplotype H2 (52.1%) was found in 144 patients with perianal disease. Compared with controls, haplotype H2 was associated with two overlapping phenotypic subgroups only, one defined by the presence of perianal disease ( $p_c=0.0005$ , RR 1.7) and the other by ileal disease ( $p_c=0.02$ , RR 1.4) (table 6). Importantly, the association between ileal disease and H2 was not seen in the absence of perianal disease (either in the ileal, no perianal group, or in the pure ileal group), suggesting that the association with ileal disease overall is due to the coexistence of perianal disease in 46 of these patients. However, when patients with perianal disease were compared with those without perianal disease, statistical significance did not withstand strict Bonferroni correction (perianal v non-perianal:  $p=0.01$ ,  $p_c=NS$ , RR=1.5).

Again, a haplotype dose effect was observed (table 7). Patients homozygous for the H2 haplotype possessed a three-fold increased risk of perianal disease compared with controls (H2/H2: 25.7% v 14.7%;  $p_c=0.0005$ , RR 3.0 (1.8–5.0)).

#### IBD5 and disease location stratified by NOD2/CARD15 status

NOD2/CARD15 mutations have recently been reported to be associated with ileal CD.<sup>15–20</sup> To determine any evidence of genetic epistasis, the association with the *IBD5* risk haplotype H2 was reanalysed following stratification by NOD2/CARD15 status (table 8). NOD2/CARD15 positive patients were defined by possession of at least one of the three common CD associated variant alleles. Stratification analyses, carried out both in the entire CD cohort and in the subgroup of patients with ileal disease, demonstrated that the frequency of haplotype H2 did not differ between NOD2/CARD15 positive and NOD2/CARD15 negative patients. This suggests that these genetic loci may act independently to determine disease at two different anatomical locations. Thus in the 107 patients with both ileal and perianal disease, association was found with both the NOD2/CARD15 variants (minimum one variant NOD2/CARD15 allele: CD 42.1% v controls 15.4%;  $p=2 \times 10^{-10}$ ) and the *IBD5* risk haplotype (*IBD5* H2 frequency: CD 53.3% v controls 38.9%;  $p=0.00005$ ;  $p_c=0.0004$ ).

#### Disease behaviour

The *IBD5* risk haplotype was not associated with specific disease behaviour, nor was it associated with the presence of extraintestinal manifestations. Whether perianal disease can justifiably be classified as "fistulating"<sup>29</sup> is not yet clear and so depends on the study definition. Importantly, no association was found with the *IBD5* risk haplotype H2 and fistulating disease overall (including perianal fistulae) (H2 frequency: fistulating disease (n=118) 51.3% v no fistulating disease (n=212) 44.3%). In addition, no association with the *IBD5* haplotype was found in the 6.7% of patients who had fistulae

**Table 3** *IBD5* allele frequencies for inflammatory bowel disease (IBD) patients and controls

SNP	Allele frequency %			Significance CD v controls	
	CD	UC	Controls	p Value	RR (95% CI)
IGR2060a_1 C	52.1	43.7	43.3	<b>0.0001</b>	1.4 (1.2–1.7)
IGR2198a_1 G	50.8	42.9	42.0	<b>0.0001</b>	1.4 (1.2–1.7)
IGR3096a_1 C	58.0	49.1	48.7	<b>&lt;0.0001</b>	1.5 (1.2–1.7)

IBD, inflammatory bowel disease; SNP, single nucleotide polymorphism; CD, Crohn's disease; UC, ulcerative colitis; RR, relative risk.  
p values in bold are significant ( $p < 0.05$ ) after Bonferroni correction factor of three (number of SNPs tested for association).

**Table 4** Definition and frequency of *IBD5* haplotypes in Crohn's disease patients and controls

Haplotype definition			Haplotype frequency			Significance CD v controls	
IGR2060a_1	IGR2198a_1	IGR3096a_1	CD (n = 330)	UC (n=457)	Controls (n = 870)	p Value	RR (95% CI)
H1 G	C	T	39.8	48.7	48.8	<b>&lt;0.0001</b>	0.7 (0.6–0.8)
H2 C	G	C	46.8	39.6	38.9	<b>0.0004</b>	1.4 (1.2–1.7)
H3 G	C	C	5.5	6.2	6.8	—	—
H4 C	C	C	3.5	2.1	2.0	—	—
H5 C	G	T	1.4	1.7	1.9	—	—
H6 G	G	C	2.3	1.3	1.0	—	—
H7 C	C	T	0.4	0.2	0.5	—	—
H8 G	G	T	0.3	0.2	0.1	—	—

p values in bold are significant ( $p < 0.05$ ) after Bonferroni correction factor of 8 (number of haplotypes tested for association).  
CD, Crohn's disease; IBD, inflammatory bowel disease; RR, relative risk; UC, ulcerative colitis.

**Table 5** Genotype frequencies of *IBD5* risk haplotype in Crohn's disease patients and controls

	Crohn's disease (n=330)	Controls (n=870)	p Value	RR (95% CI)
H2/H2	76 (23.0)	128 (14.7)	<b>0.0002</b>	2.0 (1.4–2.8)
H2/non-H2	157 (47.6)	421 (48.4)	NS	—
Non-H2/non-H2	97 (29.4)	321 (36.9)	—	—

p values in bold are significant ( $p < 0.05$ ) after Bonferroni correction factor of 8 (number of haplotypes tested for association).  
IBD, inflammatory bowel disease; RR, relative risk.

**Table 6** *IBD5* risk haplotype frequency stratified by Crohn's disease location

Group	No of individuals	<i>IBD5</i> risk haplotype (H2) frequency	Significance CD subgroup v controls	
			p Value	RR (95%CI)
All CD	330	46.8	<b>0.0004</b>	1.4(1.2–1.7)
Perianal CD	144	52.1	<b>&lt;0.0001</b>	1.7(1.3–2.2)
No perianal CD	186	42.2	—	—
Perianal CD only	9	50.0	—	—
Ileal CD	255	47.3	<b>0.0007</b>	1.4(1.2–1.7)
Ileal and no perianal CD	148	42.9	—	—
Ileal CD only	74	38.5	—	—
Colonic CD	201	44.3	—	—
Colonic and no perianal CD	112	39.3	—	—
Colonic CD only	23	48.7	—	—
Controls	870	38.9	—	—

p values in bold are significant ( $p < 0.05$ ) after application of Bonferroni correction factor of 24 (3 phenotypes $\times$ 8 haplotypes tested for association).  
CD, Crohn's disease; IBD, inflammatory bowel disease; RR, relative risk.

at other sites (H2 frequency: fistulating disease and no perianal disease (n=22) 50.0% v no fistulating disease 44.3%). These data suggest that the *IBD5* association is with perianal disease alone, rather than with fistulating disease, but the number of patients in this latter group was too low to allow firm conclusions to be drawn.

#### Age at diagnosis

Early age of diagnosis was associated with ileal (mean (SEM) age: ileal 26.4 (0.7) v no ileal 34.0 (1.8) years;  $p_c = 0.00001$ ), perianal (perianal 26.0 (1.0) v no perianal 29.8 (1.0) years;  $p_c = 0.02$ ), stenotic (stenotic 26.8 (0.7) v non-stenotic 31.5 (1.5) years;  $p_c = 0.009$ ), and fistulating disease (25.3 (1.0) v no

**Table 7** Genotype and haplotype frequencies of *IBD5* risk haplotype in perianal disease

	Genotype frequencies			H2 haplotype frequency (%)
	H2/H2 (%)	H2/non-H2 (%)	Non-H2/non-H2 (%)	
Perianal (n=144)	37 (25.7)	76 (52.8)	31 (21.5)	52.1
No perianal (n=186)	39 (21.0)	81 (43.5)	66 (35.5)	42.2
Controls (n=870)	128 (14.7)	421 (48.4)	321 (36.9)	38.9
Significance p/RR (95% CI)				
Perianal v no perianal	0.02/2.0 (1.1–3.7)	0.01/2.0 (1.2–3.4)	—	0.01/1.5 (1.1–2.0)
Perianal v controls	<b>&lt;0.0001</b> /3.0 (1.8–5.0)	0.005/1.9 (1.2–2.9)	—	<b>&lt;0.0001</b> /1.7 (1.3–2.2)

p values in bold are significant (p<0.05) after Bonferroni correction factor of 24 (3 phenotypes×8 haplotypes tested for association). RR, relative risk.

**Table 8** *IBD5* risk haplotype frequency, stratified by NOD2/CARD15 status

	Genetic stratification	Subjects	H2 haplotype frequency	Significance subgroup v controls	RR (95% CI)
CD overall	All	330	46.8	<b>0.0004</b>	1.4 (1.2–1.7)
	NOD2/CARD15 pos	127	47.6	0.008	1.4 (1.1–1.9)
	NOD2/CARD15 neg	203	46.3	<b>0.006</b>	1.4 (1.1–1.7)
Ileal disease	All	255	47.3	<b>0.0007</b>	1.4 (1.2–1.7)
	NOD2/CARD15 pos	108	46.3	0.04	1.4 (1.0–1.8)
	NOD2/CARD15 neg	147	48.0	<b>0.003</b>	1.5 (1.1–1.8)
Controls	All	870	38.9	—	—

p values in bold are significant (p<0.05) after Bonferroni correction factor of 8 (number of haplotypes tested for association). CD, Crohn's disease; IBD, inflammatory bowel disease; RR, relative risk.

fulminating 29.7 (0.9) years;  $p_c=0.01$ ). In the overall group of CD patients, no association was found between age of diagnosis (mean age of diagnosis: H2 positive, 27.4 years v H2 negative, 29.8 years) or the presence of a family history (H2 frequency: family history 45.8% v no family history 47.4%) and either possession or homozygosity of the *IBD5* haplotype H2.

**UC disease phenotype**

The most proximal point of microscopic inflammation defined disease extent. In this study, 242 (52.9%) had extensive disease defined as inflammation proximal to the splenic flexure. No association was found with disease extent and either possession or homozygosity of the *IBD5* risk haplotype H2. A total of 124 (27.1%) patients had severe disease, defined by the need for colectomy. The *IBD5* risk haplotype H2 was not associated with disease severity (colectomy 41.9% v no colectomy 38.7%). Early age of diagnosis was associated with extensive disease (mean (SEM) age: extensive disease 32.2 (0.8) v distal disease 35.6 (0.9) years;  $p=0.009$ ) and need for colectomy (need for colectomy 31.0 (1.0) v no colectomy 35.0 (1.0) years;  $p=0.005$ ). Consistent with the findings in CD patients, no association was found with the presence of extraintestinal manifestations. In addition, neither age of diagnosis (mean age of diagnosis: possession of H2 32.9 v absence of H2 35.3 years) or the presence of a family history of IBD (H2 frequency: family history 38.5% v no family history 39.9%) was associated with the *IBD5* risk haplotype H2.

**DISCUSSION**

In CD, it is increasingly believed that not only susceptibility but also clinical phenotype are genetically determined. The *IBD5* locus on chromosome 5 has been implicated in CD susceptibility by both linkage and transmission disequilibrium tests.<sup>6, 8, 21</sup> However, the role of this genetic region in determining clinical characteristics is unknown. In this case control study of 330 well defined Caucasian UK CD patients, 457 UC patients, and 870 healthy controls, we sought to identify the contribution of *IBD5* firstly to CD and UC susceptibility and then to phenotype, classified by age of onset, location, extent, and behaviour of disease.

The *IBD5* locus has recently been fine mapped to a 250 kb haplotype.<sup>21</sup> We genotyped three SNPs chosen not for functional relevance but to span this susceptibility region. The presence of tight linkage disequilibrium across this region was confirmed, and two common three-locus haplotypes identified, comprising 88% of all chromosomes. Association by haplotypes was carried out as this is potentially more informative, especially where the actual disease causing mutation has not been identified, as in this case.<sup>30</sup> However, the highly conserved haplotype structure of this region, combined with the high prevalence of two common haplotypes, meant that individual SNPs were almost as informative as haplotype analysis.

Susceptibility to CD only was associated with each of the three polymorphisms studied, and with the common reconstructed *IBD5* haplotype H2. Analysis by genotype revealed that this association was found exclusively in homozygous individuals, consistent with the previously reported dose effect.<sup>21</sup> Interestingly, the relative risk associated with haplotype H2 in our CD population was less than that reported previously in the Canadian cohort<sup>21</sup> but consistent with the risk determined by a recent UK multicentre transmission disequilibrium test of 294 CD trios.<sup>31</sup> This apparent discrepancy between Canadian and UK cohorts is perhaps not a surprising finding when one considers that linkage to this region has been identified by this former group,<sup>8</sup> but not in a UK cohort.<sup>3</sup> There are a number of possible explanations for these differences, including both genetic and disease heterogeneity between these two populations. Details of ethnicity and disease phenotype of patients recruited to the Canadian study may in the future permit dissection of these differences.

In our ethnically homogenous and rigorously phenotyped single centre cohort, we demonstrated that the frequency of haplotype H2 was particularly high in patients with perianal disease. Compared with controls, carriage of a single copy of the *IBD5* risk haplotype was associated with a relative risk of perianal disease of 1.9. This increased to 3.0 in individuals homozygous for this haplotype, consistent with the findings from the overall cohort. Importantly, in the absence of perianal disease, the *IBD5* risk haplotype was not associated with CD. However, in comparison with the association

between NOD2/CARD15 variant alleles and ileal CD, this represents a modest association. Thus the difference in haplotype H2 frequency between patients with and without perianal disease is insufficient to withstand strict Bonferroni correction.

The ideal genotype-phenotype study categorises patients using clearly defined phenotypic variables that demonstrate good inter- and intraobserver variability. Classification should preferably be applicable at diagnosis, and stable with time. These standards are difficult to meet in IBD studies for a number of reasons. Firstly, existing internationally agreed classification systems are insufficiently detailed to prevent ambiguous assignment of disease subgroup. Not surprisingly therefore, poor inter- and intraobserver variability, even among experts, may result.<sup>32</sup> This may explain why the reported prevalence of perianal involvement varies widely.<sup>33–35</sup> In this study, strict definitions of clinical subgroups were implemented to minimise the risk of ambiguous subgroup assignment. Secondly, classifying clinical subgroups may be difficult with phenotypes, particularly CD behaviour, which are not stable with time.<sup>28</sup> To overcome this limitation disease subtype was not classified at diagnosis, or at a single point in time, but by retrospective review of case notes detailing disease course over a median follow up time of 14 years.

Other polymorphic genes have also been associated with clinical phenotype of CD. Most importantly, variants of the recently identified NOD2/CARD15 gene<sup>36–38</sup> have been associated with ileal disease.<sup>15–20</sup> In this study, stratification of patients with ileal disease by NOD2/CARD15 status revealed no evidence of epistasis between the *IBD5* risk haplotype H2 and NOD2/CARD15 variants. This is not surprising as these two genetic regions may be associated with intestinal inflammation at different sites. Indeed, in the 107 individuals in whom disease affected both the perianal region and the ileum, associations with both the H2 haplotype and NOD2/CARD15 variants were identified. This suggests that location of disease is determined by possession of more than one phenotype determining allele or haplotype.

In contrast, no association was found either with UC overall susceptibility or disease phenotype, which is consistent with the absent linkage of the UC phenotype to this chromosomal region.<sup>8</sup>

The *IBD5* risk haplotype H2 spans 250 kb, encompassing a number of immunoregulatory genes. The highly conserved haplotypes across this region impede the identification of the true disease causing variant but there are a number of possible candidates. These include IL-4, IL-5, and IL-13, all of which are crucially involved in Th1-Th2 immune regulation and the inflammatory response.<sup>39</sup> Other candidates are those involved in dendritic cell maturation, including CSF2, which encodes granulocyte macrophage-colony stimulating factor, and IL-3.<sup>40</sup> Interestingly, this human locus is syntenic to a mouse region on chromosome 11 which has been implicated by linkage studies in the dextran sulphate model of colitis.<sup>41</sup> Why the *IBD5* locus should only be associated with perianal CD is not clear but these data raise the intriguing possibility that perianal disease results from dysregulated immune responses to site specific bacterial colonisation.

In summary, we have confirmed the association of the *IBD5* risk haplotype H2 with CD only, and detailed genotype/phenotype analysis has revealed that this association is most marked in patients with perianal CD. While the precise gene involved is unclear, these data provide further evidence for the genetic basis of the clinical heterogeneity of CD.

## ACKNOWLEDGEMENTS

This work was supported by a grant from the Catholic University of Rome (to AA) and the National Association of Crohn's and Colitis (NACC, UK) (to TA). The authors thank the patients who participated in this study. The authors acknowledge the contributions of Sue

Goldthorpe for organisation and collection of blood specimens and Robert Walton for allowing us to use the OXCHECK samples.

.....

## Authors' affiliations

**A Armuzzi, T Ahmad, K-L Ling, A de Silva, S Cullen, D van Heel, T R Orchard, D P Jewell**, Gastroenterology Unit, University of Oxford, Gibson Laboratories, Radcliffe Infirmary, Oxford, UK  
**K I Welsh**, Clinical Genomics, National Heart and Lung Institute, Imperial College, London, UK  
**S E Marshall**, Department of Immunology, Wright-Fleming Institute, Imperial College, London, UK

## REFERENCES

- 1 **Farrokhyar F**, Swarbrick E, Irvine E. A critical review of epidemiological studies in inflammatory bowel disease. *Scand J Gastroenterol* 2001;**36**:2–15.
- 2 **Hugot JP**, Laurent-Puig P, Gower-Rousseau C, et al. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 1996;**379**:821–3.
- 3 **Satsangi J**, Parkes M, Louis E, et al. Two-stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat Genet* 1996;**14**:199–202.
- 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 U S A* 1998;**95**:7502–7.
- 5 **Hampe J**, Schreiber S, Shaw S, et al. A genome-wide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *Am J Hum Genet* 1999;**64**:808–16.
- 6 **Ma Y**, Ohmen J, Li Z, et al. A genome wide search identifies potential new susceptibility loci for Crohn's disease. *Inflamm Bowel Dis* 1999;**5**:271–8.
- 7 **Duerr R**, Barmada M, Zhang L, et al. High-density genome scan in Crohn's disease shows confirmed linkage to chromosome 14q11–12. *Am J Hum Genet* 2000;**66**:1857–62.
- 8 **Rioux J**, Silverberg M, Daly M, et al. Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am J Hum Genet* 2000;**66**:1863–70.
- 9 **Vermeire S**, Satsangi J, Peeters M, et al. Evidence for inflammatory bowel disease of a susceptibility locus on the X chromosome. *Gastroenterology* 2001;**120**:834–40.
- 10 **Cavanaugh J**. The IBD International Genetic Consortium. 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* 2001;**68**:1165–71.
- 11 **Bayless TM**, Tokayer AZ, Polito II JM, et al. Crohn's disease: concordance for site and clinical type in affected family members—potential hereditary influences. *Gastroenterology* 1996;**111**:573–9.
- 12 **Polito II JM**, Childs B, Mellits D, et al. Crohn's disease: influence of age at diagnosis on site and clinical type of disease. *Gastroenterology* 1996;**111**:580–6.
- 13 **Peeters M**, Nevens H, Baert F, et al. Familial aggregation in Crohn's disease: increased age adjusted risk and concordance in clinical characteristics. *Gastroenterology* 1996;**111**:597–603.
- 14 **Colombel J-F**, Grandbastien B, Gower-Rousseau C, et al. Clinical characteristics of Crohn's disease in 72 families. *Gastroenterology* 1996;**111**:604–7.
- 15 **Ahmad T**, Armuzzi A, Bunce M, et al. The molecular classification of the clinical manifestations of Crohn's disease. *Gastroenterology* 2002;**122**:854–66.
- 16 **Cuthbert A**, Fisher S, Mirza M, et al. The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease. *Gastroenterology* 2002;**122**:867–74.
- 17 **Lesage S**, Zouali A, Cézard JP, et al. CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* 2002;**70**:845–57.
- 18 **Vavassori P**, Borgiani P, D'Apice MR, et al. 3020insC mutation within the NOD2 gene in Crohn's disease: frequency and association with clinical pattern in an Italian population. *Dig Liver Dis* 2002;**34**:153.
- 19 **Hampe J**, Grebe J, Nikolaus S, et al. Association of NOD2 (CARD15) genotype with clinical course of Crohn's disease: a cohort study. *Lancet* 2002;**359**:1661–5.
- 20 **Vermeire S**, Wild G, Kocher K, et al. CARD15 genetic variation in a Quebec population: prevalence, genotype-phenotype relationship, and haplotype structure. *Am J Hum Genet* 2002;**71**:74–83.
- 21 **Rioux J**, Daly M, Silverberg M, et al. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 2001;**29**:223–8.
- 22 **Lennard-Jones J**. Classification of inflammatory bowel disease. *Scand J Gastroenterol* 1989;**24**:2–6.
- 23 **Langham S**, Thorogood M, Normand C, et al. Costs and cost effectiveness of health checks conducted by nurses in primary care: the Oxcheck study. *BMJ* 1996;**312**:1265–8.
- 24 **Welsh K**, Bunce M. Molecular typing for the MHC with PCR-SSP. *Rev Immunogenet* 1999;**1**:157–76.
- 25 **Stephens M**, Smith N, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;**68**:978–89.

- 26 **Schlesselman J.** *Case-Control Studies*. Oxford: Oxford University Press, 1982.
- 27 **Goldstein D, Weale M.** Population genomics: Linkage disequilibrium holds the key. *Curr Biol* 2001;**11**:R576-9.
- 28 **Louis E, Collard A, Oger A, et al.** Behaviour of Crohn's disease according to the Vienna classification: changing pattern over the course of the disease. *Gut* 2001;**49**:777-82.
- 29 **Gasche C, Scholmerich J, Brynskov J, et al.** A simple classification of Crohn's disease: Report of the Working Party for the World Congress of Gastroenterology, Vienna 1998. *Inflamm Bowel Dis* 2000;**6**:8-15.
- 30 **Johnson G, Esposito L, Barratt B, et al.** Haplotype tagging for the identification of common disease genes. *Nat Genet* 2001;**29**:233-7.
- 31 **Negoro K, McGovern DPB, Kinouchi Y, et al.** Analysis of the IBD5 locus and potential gene-gene interactions in Crohn's disease. *Gut* 2003;**52**:541-6.
- 32 **Steinhart AH, Girgrah N, McLeod RS.** Reliability of a Crohn's disease clinical classification scheme based on disease behavior. *Inflamm Bowel Dis* 1998;**4**:228-34.
- 33 **Rankin G, Watts H, Melnyk C, et al.** National Cooperative Crohn's disease Study: extraintestinal manifestations and perianal complications. *Gastroenterology* 1979;**77**:914-20.
- 34 **Cohen Z, McLeod R.** Perianal Crohn's disease. *Gastroenterol Clin North Am* 1987;**16**:175-89.
- 35 **McClane S, Rombeau J.** Anorectal Crohn's disease. *Surg Clin North Am* 2001;**81**:169-83.
- 36 **Hugot J-P, Chamaillard M, Zouali H, et al.** Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;**411**:599-603.
- 37 **Ogura Y, Bonen D, Inohara N, et al.** A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001;**411**:603-6.
- 38 **Hampe J, Cuthbert A, Croucher P, et al.** Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 2001;**357**:1925-8.
- 39 **Romagnani S.** Th1/Th2 cells. *Inflamm Bowel Dis* 1999;**5**:285-94.
- 40 **Banchereau J, Steinmann R.** Dendritic cells and the control of immunity. *Nature* 1998;**392**:245-52.
- 41 **Mahler M, Bristol I, Sundberg J, et al.** Genetic analysis of susceptibility to dextran sulfate sodium-induced colitis in mice. *Genomics* 1999;**55**:147-56.



Want to know more?

**Data supplements**

Limited space in printed journals means that interesting data and other material are often edited out of articles; however, limitless cyberspace means that we can include this information online. Look out for additional tables, references, illustrations.

[www.gutjnl.com](http://www.gutjnl.com)