

# Molecular prediction of disease risk and severity in a large Dutch Crohn's disease cohort

R K Weersma,<sup>1</sup> P C F Stokkers,<sup>2</sup> A A van Bodegraven,<sup>3</sup> R A van Hogezaand,<sup>4</sup> H W Verspaget,<sup>4</sup> D J de Jong,<sup>5</sup> C J van der Woude,<sup>6</sup> B Oldenburg,<sup>7</sup> R K Linskens,<sup>8</sup> E A M Festen,<sup>1</sup> G van der Steege,<sup>9</sup> D W Hommes,<sup>4</sup> J B A Crusius,<sup>10</sup> C Wijmenga,<sup>9</sup> I M Nolte,<sup>11</sup> G Dijkstra,<sup>1</sup> The Dutch Initiative on Crohn and Colitis (ICC)

## See Commentary, p 323

<sup>1</sup> Department of Gastroenterology and Hepatology, University Medical Center Groningen, University of Groningen, The Netherlands;

<sup>2</sup> Department of Gastroenterology and Hepatology, Academic Medical Center, Amsterdam, The Netherlands; <sup>3</sup> Department of Gastroenterology and Hepatology, VU University Medical Center, Amsterdam, The Netherlands; <sup>4</sup> Department of Gastroenterology and Hepatology, Leiden University Medical Center, The Netherlands; <sup>5</sup> Department of Gastroenterology and Hepatology, Radboud University Nijmegen Medical Center, The Netherlands; <sup>6</sup> Department of Gastroenterology and Hepatology, Erasmus Medical Center, Rotterdam, The Netherlands; <sup>7</sup> Department of Gastroenterology and Hepatology, University Medical Center Utrecht, The Netherlands; <sup>8</sup> Department of Gastroenterology and Hepatology, St Anna Zorggroep Hospital, Geldrop, The Netherlands; <sup>9</sup> Department of Medical Genetics, University Medical Center Groningen and University of Groningen, The Netherlands; <sup>10</sup> Department of Immunogenetics, VU University Medical Center, Amsterdam, The Netherlands; <sup>11</sup> Department of Epidemiology, University Medical Center Groningen, University of Groningen, The Netherlands

Correspondence to: Dr R K Weersma, Department of Gastroenterology and Hepatology, University Medical Center Groningen and University of Groningen, P.O. Box 30001, 9700 RB Groningen, The Netherlands; [R.K.Weersma@int.umcg.nl](mailto:R.K.Weersma@int.umcg.nl)

Revised 4 August 2008  
Accepted 19 August 2008  
Published Online First  
29 September 2008

## ABSTRACT

**Background:** Crohn's disease and ulcerative colitis have a complex genetic background. We assessed the risk for both the development and severity of the disease by combining information from genetic variants associated with inflammatory bowel disease (IBD).

**Methods:** We studied 2804 patients (1684 with Crohn's disease and 1120 with ulcerative colitis) and 1350 controls from seven university hospitals. Details of the phenotype were available for 1600 patients with Crohn's disease and for 800 with ulcerative colitis. Genetic association for disease susceptibility was tested for the nucleotide-binding and oligomerisation domain 2 gene (*NOD2*), the IBD5 locus, the Drosophila discs large homologue 5 and autophagy-related 16-like 1 genes (*DLG5* and *ATG16L1*) and the interleukin 23 receptor gene (*IL23R*). Interaction analysis was performed for Crohn's disease using the most associated single nucleotide polymorphism (SNP) for each locus. Odds ratios were calculated in an ordinal regression analysis with the number of risk alleles as an independent variable to analyse disease development and severity.

**Results:** Association with Crohn's disease was confirmed for *NOD2*, IBD5, *DLG5*, *ATG16L1* and *IL23R*. Patients with Crohn's disease carry more risk alleles than controls ( $p = 3.85 \times 10^{-22}$ ). Individuals carrying an increasing number of risk alleles have an increasing risk for Crohn's disease, consistent with an independent effects multiplicative model (trend analysis  $p = 4.25 \times 10^{-23}$ ). Patients with Crohn's disease with a more severe disease course, operations or an age of onset below 40 years have more risk alleles compared to non-stricturing, non-penetrating behaviour ( $p = 0.0008$ ), no operations ( $p = 0.02$ ) or age of onset above 40 years ( $p = 0.028$ ).

**Conclusion:** Crohn's disease is a multigenic disorder. An increase in the number of risk alleles is associated with an increased risk for the development of Crohn's disease and with a more severe disease course. Combining information from the known common risk polymorphisms may enable clinicians to predict the course of Crohn's disease.

Chronic inflammatory bowel diseases (IBDs) comprising Crohn's disease and ulcerative colitis are characterised by chronic recurring inflammation of the gastrointestinal tract. The combined prevalence of Crohn's disease and ulcerative colitis is estimated at 100/100 000 to 200/100 000 in developed countries.<sup>1,2</sup> Concordance rates in twins and siblings suggest that a genetic predisposition contributes to the pathogenesis of IBD, but the

disease is complex and expected to be due to multiple predisposing genes.<sup>3,4</sup>

In 2001 the nucleotide-binding and oligomerisation domain 2 (*NOD2*) gene on chromosome 16 was the first susceptibility gene identified for IBD.<sup>5,6</sup> Three mutations (R702W, G908R and 3020insC) are independently associated with Crohn's disease in Caucasian patients. More recently, several more IBD susceptibility genes have been identified. A widely replicated associated locus is the IBD5 locus on chromosome 5q31–q33, including the genes solute carrier family 22 members 4 and 5 (*SLC22A4* and  $-5$ ) encoding for the organic cation transporters 1 and 2.<sup>7,8</sup> Unfortunately, the causative IBD susceptibility gene from this region has not yet been identified unequivocally. In addition, refining a linkage region previously found on chromosome 10q23 led to the identification of the Drosophila discs large homologue 5 (*DLG5*) gene as a susceptibility gene for Crohn's disease and IBD.<sup>9</sup> Although results have been conflicting, a large meta-analysis showed that carriage of the major allele of the R30Q variant *DLG5* confers a small risk for developing Crohn's disease in women.<sup>10</sup> A recent genome-wide association scan identified a coding single nucleotide polymorphism (SNP), within the autophagy-related 16-like 1 gene (*ATG16L1*) on chromosome 2q37.1, to be strongly associated with Crohn's disease.<sup>11</sup> This association has been confirmed in several independent cohorts.<sup>12,13</sup> Finally, a genome-wide association study found that genetic variation in the interleukin 23 receptor gene (*IL23R*) has a protective effect for developing Crohn's disease.<sup>14</sup> This finding has been confirmed in different study populations.<sup>13,15</sup> In addition to these genes several further genes have been identified very recently by large-scale genome-wide association studies. These include a locus on chromosome 5p13.1, *nel*-like 1 precursor (*NELL-1*) and the autophagy-inducing immunity-related GTP-ase family M (*IRGM*) gene found by the Wellcome Trust Case Control Consortium (WTCCC).<sup>16–18</sup>

IBD shows extensive variation in the individual clinical presentation and outcome, and it is tempting to speculate that this is most likely due to differences in the genetic susceptibility in combination with exposure to environmental factors, such as the commensal bacteria in the intestine, and the intestinal immune system.<sup>1</sup> Since many of the identified polymorphisms only

**Table 1** Clinical characteristics of patients with Crohn's disease

Characteristic	Number	Per cent
Total number	1600	
Sex (male/female)	568/1032	35.5–64.5
Age at diagnosis		
<40 years	1335	83.4
>40 years	265	16.6
Disease localisation		
Ileal	408	25.5
Colon	489	30.6
Ileocolon	703	43.9
Upper gastrointestinal tract*	190	11.9
Disease behaviour		
Non-stricturing, non-penetrating	637	39.8
Stricturing	405	25.3
Penetrating	558	34.9
Perianal disease	433	27.1
Extra-intestinal manifestations	355	22.2
Operation	810	50.6
Family history of inflammatory bowel disease	223	13.9
Mean follow-up (years)	14.2 (SD 10.4)	

\*Patients could be classified as having disease localisation in the upper gastrointestinal tract next to ileal, colonic or ileocolonic localisation.

moderately increase the risk for disease development with allelic odds ratios (ORs) below 1.5, they have no value in predicting an individual's risk for disease development. However, these common, low penetrance risk alleles may be responsible for a large fraction of the genetic component of a complex multifactorial disease like IBD. Previous studies on simulated data, and recently on data from patients with Crohn's disease and type 2 diabetes, have shown the possibility of combining information from multiple risk alleles to predict the risk for disease development.<sup>19–21</sup>

The aim of this study was to confirm IBD-associated genes in the Dutch population and to examine the combined effect of these genetic variants in predicting disease development. We also studied whether an increase in the number of common risk alleles was associated with a more severe disease course. We analysed genetic variants in *NOD2*, the IBD5 locus, *DLG5*, *ATG16L1* and *IL23R* for association with IBD, Crohn's disease, ulcerative colitis and different subsets of Crohn's disease and ulcerative colitis in a large nationwide case-control study,

**Table 2** Clinical characteristics of patients with ulcerative colitis

Characteristic	Number	Per cent
Total number	800	
Sex (male/female)	430/370	53.8–46.2
Age at diagnosis		
<40 years	575	71.9
>40 years	225	28.1
Disease localisation		
Proctitis	130	16.3
Left-sided	304	38.0
Extensive	366	45.7
Extra-intestinal manifestations	106	13.3
Operation	130	16.3
Family history of inflammatory bowel disease	103	12.9
Mean follow-up (years)	11.8 (SD 8.7)	

including well-phenotyped patients with IBD from seven university medical centres (UMCs) in The Netherlands.

## METHODS

### Patients and controls

The study population consisted of 2804 patients of Caucasian ethnicity with IBD (1684 with Crohn's disease, and 1120 with ulcerative colitis) and 1350 Caucasian controls from seven UMCs in The Netherlands: Academic Medical Center, Amsterdam (n = 772), VU University Medical Center, Amsterdam (n = 596), University Medical Center Groningen (n = 516), Leiden University Medical Center (n = 503), Erasmus Medical Center, Rotterdam (n = 183), Radboud University Nijmegen Medical Center (n = 156), and University Medical Center Utrecht (n = 78).<sup>22–26</sup> Controls consisted of healthy volunteers from the UMC Utrecht (n = 784), UMC Groningen (n = 284), UMC Leiden (n = 168) and UMC Nijmegen (n = 114).<sup>22–25</sup> There was no significant number of Jewish patients in this cohort.

Patients were diagnosed according to accepted clinical, endoscopic, radiological and histological findings.<sup>1</sup> All patients and controls gave their informed consent and all DNA samples and data in this study were handled anonymously. For patients with Crohn's disease, phenotypic details were registered according to the Vienna classification. However, perianal disease was also scored as an independent variable. Upper GI localisation was scored separately and was not mutually exclusive with ileal, colonic or ileocolonic localisation.<sup>27</sup> For patients with ulcerative colitis, phenotypes were described according to age of onset, extent of disease (proctitis, left-sided or extensive), necessity of colectomy and the occurrence of extra-intestinal manifestations: arthritis/arthralgia, sacroiliitis, hepatobiliary complications (eg, primary sclerosing cholangitis (PSC)), and cutaneous or ocular manifestations. Phenotyping was performed by the participating gastroenterologist from each UMC by reviewing a patient's chart retrospectively. All participants received a detailed description of classification rules and the details had to be filled out in a structured form. Phenotypic details were available for 1600 of 1684 (95.0%) patients with Crohn's disease and for 800 of 1120 (71.4%) patients with ulcerative colitis and are presented in tables 1 and 2.

### Genotyping and SNP selection

A polymerase chain reaction – restriction fragment – length polymorphism assay was used for genotyping three SNPs in the *NOD2* gene (R702W, G908R and 3020Cins). Full details on primer sequences and assays are available on request. SNP genotyping of all other polymorphisms was performed with TaqMan probes and primers, using assays developed by Applied Biosystems, and an ABI 7900HT system (Applied Biosystems, Nieuwerkerk a/d IJssel, The Netherlands). A set of eight SNPs was analysed for *SLC22A4/5* and the IBD5 region. This set included the two known risk-associated SNPs 1672 C→T (rs1050152) and –207G→C (rs2631367), and SNP rs3792876, which was found to be associated with paediatric Crohn's disease in a cohort from the Academic Medical Center Amsterdam.<sup>8, 28</sup> Three tag SNPs in the region were also analysed (rs272893, rs273900 and rs274551). Two SNPs located at the 3' end (rs2522057) and 5' end (rs7705189) were chosen to analyse linkage disequilibrium in the IBD5 region. Four SNPs were analysed for *DLG5* including the two risk-associated SNPs 113 G→A (R30Q, rs1248696) and 4136 C→A (P1371Q, rs2289310) from the original paper by Stoll *et al.*<sup>9</sup> Two more haplotype

## Inflammatory bowel disease

**Table 3** Marker data for genetic analysis of *NOD2*, *IBD5*, *DLG5*, *ATG16L1* and *IL23R*

Marker	Comment
<b><i>NOD2</i></b>	
rs2066844	R702W
rs2066845	G908R 3020insC
<b><i>IBD5</i></b>	
rs7705189	5' end <i>IBD5</i>
rs3792876	
rs272893	
rs1050152	1672C→T
rs273900	
rs2631367	−207G→C
rs274551	
rs2522057	3' end <i>IBD5</i>
<b><i>DLG5</i></b>	
rs2165047	Haplotype B
rs2289311	Haplotype A
rs2289310	Haplotype C; P1371Q
rs1248696	Haplotype D; R30Q
<b><i>ATG16L1</i></b>	
rs2241880	T300A
<b><i>IL23R</i></b>	
rs11209026	R381Q

*ATG16L1*, autophagy-related 16-like 1 gene; *DLG5*, Drosophila discs large homologue 5 gene; *IBD5*, inflammatory bowel disease 5 gene; *IL23R*, interleukin 23 receptor gene; *NOD2*, nucleotide-binding and oligomerisation domain 2 gene.

tagging SNPs were chosen to determine haplotype A (rs2289311) and haplotype B (rs2165047). For *IL23R*, SNP rs11209026 was analysed. The risk-associated SNP rs2241880 was determined for *ATG16L1*.<sup>11</sup> Marker data for all these SNPs are described in table 3.<sup>14</sup>

**Statistical methods**

As samples from seven centres were combined for analysis, we first tested for homogeneity of the genotype data between centres. Genotype frequencies were compared between the centres using a  $\chi^2$  test separately for cases and controls. SNPs rs272893, rs273900 and rs274551 in the control group of the UMC Groningen and SNP rs2289311 in the cases from the Academic Medical Center, Amsterdam, showed deviating genotype frequencies (homogeneity  $p < 0.05$ ). For these SNPs the groups concerned were discarded from further analyses.

Hardy–Weinberg equilibrium was tested among the unrelated, unaffected individuals using a  $\chi^2$  test with 1 degree of freedom. SNP rs274551 in the control group of the UMC Groningen showed deviation from Hardy–Weinberg equilibrium ( $p < 0.05$ ) but had already been discarded from further analyses because of genotypic heterogeneity.

The allele and genotype frequencies of patients and controls were compared to test for association using the  $\chi^2$  test or Fisher's exact test where appropriate (SPSS 12.0.1). Since all five genes had already been identified as associated genes for IBD, Crohn's disease and ulcerative colitis susceptibility, no multiple testing correction was required. Thus, a multiple testing correction was only performed in the genotype–phenotype subgroup analyses for the number of complementary subgroups of patients by using a Bonferroni correction. Linkage disequilibrium and haplotype analyses were performed with in-house software using an expectation maximisation algorithm. Linkage disequilibrium parameters  $D'$  and  $r^2$  were calculated for all pairs of SNPs in the *IBD5* region and in *DLG5*.

Odds ratios (ORs) and 95% confidence intervals (CIs) were estimated using binary logistic regression (SPSS 12.0.1). Binary logistic regression was also used to determine pairwise and three-way, four-way and five-way gene–gene interaction effects with inclusion of the main effects of each gene in the model. The main

**Table 4** Allelic association analysis for *NOD2*, *IBD5*, *DLG5*, *ATG16L1* and *IL23R* for patients with Crohn's disease (n = 1684) or ulcerative colitis (n = 1120) compared to healthy controls (n = 1350)

Locus	Controls		Crohn's disease		Ulcerative colitis		
	MAF	MAF	p Value	OR (95% CI)	MAF	p Value	OR (95% CI)
<b><i>NOD2</i></b>							
R702W	0.05	0.08	<b><math>1.56 \times 10^{-8}</math></b>	1.92 (1.53 to 2.42)	0.05	0.15	1.22 (0.93 to 1.60)
G908R	0.02	0.04	<b><math>2.47 \times 10^{-10}</math></b>	2.94 (2.07 to 4.18)	0.02	0.65	1.11 (0.71 to 1.73)
3020insC	0.02	0.07	<b><math>3.58 \times 10^{-17}</math></b>	3.26 (2.44 to 4.36)	0.02	0.91	1.02 (0.98 to 1.49)
<b><i>IBD5</i></b>							
rs7705189	0.45	0.48	<b>0.017</b>	1.13 (1.02 to 1.26)	0.48	<b>0.013</b>	1.16 (1.03 to 1.30)
rs3792876	0.09	0.08	0.59	0.95 (0.79 to 1.15)	0.07	0.094	0.83 (0.67 to 1.03)
rs272893	0.37	0.36	0.24	0.94 (0.84 to 1.04)	0.35	0.13	0.91 (0.81 to 1.03)
rs1050152	0.41	0.45	<b>0.015</b>	1.14 (1.03 to 1.26)	0.44	<b>0.036</b>	1.13 (1.01 to 1.27)
rs273900	0.38	0.37	0.21	0.93 (0.84 to 1.04)	0.36	0.22	0.93 (0.82 to 1.05)
rs2631367	0.54	0.49	<b>0.026</b>	0.88 (0.79 to 0.99)	0.50	<b>0.014</b>	0.86 (0.77 to 0.97)
rs274551	0.17	0.15	<b>0.039</b>	0.86 (0.75 to 0.99)	0.17	0.48	0.95 (0.81 to 1.10)
rs2522057	0.40	0.45	<b><math>5.00 \times 10^{-4}</math></b>	1.20 (1.08 to 1.33)	0.44	<b><math>2.72 \times 10^{-03}</math></b>	1.19 (1.06 to 1.33)
<b><i>DLG5</i></b>							
rs1248696	0.13	0.11	0.08	0.87 (0.74 to 1.02)	0.11	<b>0.019</b>	0.81 (0.68 to 0.97)
rs2289310	0.05	0.04	<b>0.024</b>	0.76 (0.59 to 0.97)	0.05	0.26	0.86 (0.66 to 1.12)
rs2289311	0.34	0.34	0.79	0.99 (0.88 to 1.10)	0.34	0.90	0.99 (0.88 to 1.12)
rs2165047	0.25	0.28	<b><math>2.90 \times 10^{-3}</math></b>	1.19 (1.06 to 0.92)	0.28	<b>0.035</b>	1.15 (1.01 to 1.30)
<b><i>ATG16L1</i></b>							
rs2241880	0.44	0.39	<b><math>3.61 \times 10^{-4}</math></b>	0.82 (0.74 to 0.92)	0.45	0.42	1.05 (0.93 to 1.18)
<b><i>IL23R</i></b>							
rs11209026	0.07	0.02	<b><math>2.12 \times 10^{-17}</math></b>	0.31 (0.23 to 0.41)	0.04	<b><math>3.96 \times 10^{-04}</math></b>	0.62 (0.48 to 0.81)

Values in bold type are statistically significant with a p value  $< 0.05$ .

*ATG16L1*, autophagy-related 16-like 1 gene; CI, confidence interval; *DLG5*, Drosophila discs large homologue 5 gene; *IBD5*, inflammatory bowel disease 5 gene; *IL23R*, interleukin 23 receptor gene; MAF, minor allele frequency; *NOD2*, nucleotide-binding and oligomerisation domain 2 gene; OR, odds ratio.

**Table 5** Interaction analysis for *NOD2*, *IBD5*, *DLG5*, *IL23R* and *ATG16L1*

Interaction	p Value	OR (95% CI)
<b>Two-way interaction</b>		
<i>IBD5</i> × <i>DLG5</i>	0.67	0.86 (0.43 to 1.71)
<i>IBD5</i> × <i>ATG16L1</i>	0.58	1.15 (0.70 to 1.88)
<i>IBD5</i> × <i>NOD2</i>	<b>0.02</b>	0.53 (0.31 to 0.92)
<i>IBD5</i> × <i>IL23R</i>	0.34	0.68 (0.31 to 1.51)
<i>DLG5</i> × <i>ATG16L1</i>	0.77	0.88 (0.36 to 2.15)
<i>DLG5</i> × <i>NOD2</i>	<b>0.01</b>	3.02 (1.29 to 7.08)
<i>DLG5</i> × <i>IL23R</i>	0.09	2.71 (0.85 to 8.62)
<i>ATG16L1</i> × <i>NOD2</i>	0.06	0.51 (0.26 to 1.03)
<i>ATG16L1</i> × <i>IL23R</i>	0.74	0.85 (0.32 to 2.25)
<i>NOD2</i> × <i>IL23R</i>	0.77	1.15 (0.45 to 2.89)
<b>Three-way interaction</b>		
<i>IBD5</i> × <i>DLG5</i> × <i>ATG16L1</i>	0.72	1.08 (0.71 to 1.63)
<i>IBD5</i> × <i>DLG5</i> × <i>NOD2</i>	0.17	0.70 (0.42 to 1.17)
<i>IBD5</i> × <i>DLG5</i> × <i>IL23R</i>	0.84	0.95 (0.56 to 1.60)
<i>IBD5</i> × <i>ATG16L1</i> × <i>NOD2</i>	0.46	0.59 (0.35 to 0.99)
<i>IBD5</i> × <i>NOD2</i> × <i>IL23R</i>	<b>0.02</b>	0.54 (0.32 to 0.92)
<i>IBD5</i> × <i>ATG16L1</i> × <i>IL23R</i>	0.69	0.92 (0.59 to 1.41)
<i>DLG5</i> × <i>ATG16L1</i> × <i>NOD2</i>	0.62	0.86 (0.48 to 1.55)
<i>DLG5</i> × <i>ATG16L1</i> × <i>IL23R</i>	0.72	1.12 (0.60 to 2.10)
<i>ATG16L1</i> × <i>NOD2</i> × <i>IL23R</i>	0.22	0.69 (0.38 to 1.25)
<b>Four-way interaction</b>		
<i>IBD5</i> × <i>DLG5</i> × <i>ATG16L1</i> × <i>NOD2</i>	0.16	0.69 (0.41 to 1.15)
<i>IBD5</i> × <i>DLG5</i> × <i>ATG16L1</i> × <i>IL23R</i>	0.83	1.04 (0.71 to 1.53)
<i>IBD5</i> × <i>DLG5</i> × <i>NOD2</i> × <i>IL23R</i>	0.22	0.73 (0.44 to 1.21)
<i>IBD5</i> × <i>ATG16L1</i> × <i>NOD2</i> × <i>IL23R</i>	0.59	0.61 (0.36 to 1.02)
<i>DLG5</i> × <i>ATG16L1</i> × <i>NOD2</i> × <i>IL23R</i>	0.82	0.94 (0.54 to 1.62)
<b>Five-way interaction</b>		
<i>IBD5</i> × <i>DLG5</i> × <i>ATG16L1</i> × <i>NOD2</i> × <i>IL23R</i>	0.74	0.74 (0.44 to 1.23)

Binary logistic regression analyses were performed using both main effects and interaction effects. The results of the interaction effect are presented.

Values in bold type are statistically significant with a p value <0.05.

The polymorphisms carried by the genes are: *NOD2*, carriage of one or more mutations for R702W, G908R or 3020insC; *IBD5*, carriage of the rs2522027\*C allele; *ATG16L1*, carriage of the rs2241880\*G allele; *DLG5*, carriage of the rs2289310\*C allele; *IL23R*, carriage of the rs11209026\*G allele.

*ATG16L1*, autophagy-related 16-like 1 gene; CI, confidence interval; *DLG5*, Drosophila discs large homologue 5 gene; *IBD5*, inflammatory bowel disease 5 gene; *IL23R*, interleukin 23 receptor gene; *NOD2*, nucleotide-binding and oligomerisation domain 2 gene; OR, odds ratio.

effect was determined by the OR after logistic regression analysis for the associated SNPs in each locus. In addition, ORs for Crohn's disease were calculated in a binary logistic regression analysis with the number of risk alleles or risk genotypes at *IBD5* rs2522027, *DLG5* rs2289310 and *ATG16L1* rs2241880, *IL23R* rs11209026 and carrying one of the three risk-associated polymorphisms in *NOD2* (R702W, G908R and 1007fsinsC) as an independent variable. The reference group consisted of all individuals with zero or one risk allele or genotype. Ordinal regression analysis was used to test whether patients with Crohn's disease had more risk alleles than controls. To further test whether the risk of being affected with Crohn's disease increases with increasing number of risk alleles, we performed a  $\chi^2$  test for trend using the ratio of Crohn's disease cases versus controls within categories of number of risk alleles. Since different genetic variants have differing effects on disease susceptibility, a score was determined according to coefficients calculated from separate logistic regression analyses for each SNP associated with Crohn's disease. In this way, the score was weighted for the strength of association for each individual genetic variant. Each patient was then categorised based on the sum of these coefficients and ORs for each category were determined by logistic regression analysis.

Besides disease occurrence we also looked at disease behaviour. For this purpose we again used the number of risk alleles for *IBD5* rs2522027, *DLG5* rs2289310 and *ATG16L1* rs2241880,

*IL23R* rs11209026 and carrying one of the three risk-associated polymorphisms in *NOD2* (R702W, G908R and 1007fsinsC). The reference group consisted of all individuals with zero or one risk allele or genotype. To analyse whether specific disease behaviour of Crohn's disease (early age of onset, familial occurrence, stricturing and/or penetrating behaviour, perianal disease, operation and extra-intestinal manifestations) was associated with a higher number of risk alleles we performed ordinal regression and trend analyses as described above. Since disease behaviour can change over time we also performed these analyses on patients with more than 5 years or more than 10 years of follow-up.

## RESULTS

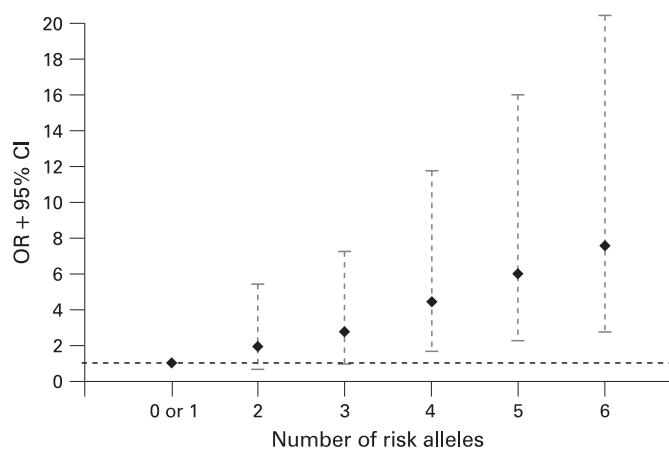
### Association of *NOD2*, *IBD5*, *DLG5*, *ATG16L1* and *IL23R* with IBD

Allelic association results for *NOD2*, *IBD5*, *DLG5*, *ATG16L1*, and *IL23R* with Crohn's disease and ulcerative colitis are summarised in table 4.

### *NOD2*

The polymorphisms R702W (p =  $1.56 \times 10^{-8}$ ; OR, 1.92; CI, 1.53 to 2.42), G908R (p =  $2.47 \times 10^{-8}$ ; OR, 2.77; CI, 1.92 to 4.01;) and 3020insC (p =  $1.64 \times 10^{-17}$ ; OR, 3.26; CI, 2.44 to 4.36) were all strongly associated with Crohn's disease, but did not show any

## Inflammatory bowel disease



**Figure 1** Odds ratios (ORs) and confidence intervals (CIs) for patients carrying increasing numbers of risk alleles for developing Crohn's disease. The risk alleles were: *NOD2* – the presence of one of the R702W, G908R or 3020insC polymorphisms was defined as one risk allele; *IBD5* – the rs2522027\*C allele; *ATG16L1* – the rs2241880\*G allele; *DLG5* – the rs2289310\*C allele; the risk “allele” for *IL23R* is defined as homozygosity for rs11209026\*G. The reference group consists of individuals with zero or one risk allele. The progressive increase in ORs is consistent with an independent effects multiplicative model. *ATG16L1*, autophagy-related 16-like 1 gene; *DLG5*, Drosophila discs large homologue 5 gene; *IBD5*, inflammatory bowel disease 5 gene; *IL23R*, interleukin 23 receptor gene; *NOD2*, nucleotide-binding and oligomerisation domain 2 gene.

association with ulcerative colitis. The 3020insC polymorphism was associated with ileal localisation ( $p = 2.36 \times 10^{-5}$ ; OR, 1.83; CI, 1.37 to 2.41), stricturing disease behaviour ( $p = 0.027$ ; OR, 1.38; CI, 1.04 to 1.83) and the necessity of operation ( $p = 1.72 \times 10^{-4}$ ; OR, 1.69; CI, 1.28 to 2.23). The G908R polymorphism was associated with stricturing behaviour ( $p = 5.91 \times 10^{-3}$ ; OR, 1.65; CI, 1.15 to 2.36).

### IBD5

SNP rs1050152 (1672G→C) was associated with both Crohn's disease ( $p = 0.015$ ; OR, 1.14; CI, 1.03 to 1.26) and ulcerative colitis ( $p = 0.036$ ; OR, 1.13; CI, 1.01 to 1.27). SNP rs2631367 (−207G→C) was also moderately associated with Crohn's disease ( $p = 0.036$ ; OR, 0.88; CI, 0.79 to 0.99) and ulcerative colitis ( $p = 0.011$ ; OR, 0.86; CI, 0.77 to 0.97). The previously described risk-associated TC haplotype for these two SNPs combined was over-represented in Crohn's disease (44.6%) compared to controls (41.1%) ( $p = 0.0075$ ; OR, 1.15 (1.04 to 1.28)). However, the strongest association in the region was observed for SNP rs2522057 at the 3' end of *IBD5* with both Crohn's disease ( $p = 0.0005$ ; OR, 1.20; CI, 1.08 to 1.33) and ulcerative colitis ( $p = 0.0027$ ; OR, 1.19; CI, 1.06 to 1.33). All markers were in strong linkage disequilibrium. No specific associations with the *IBD5* locus were found for any subset of patients with Crohn's disease or ulcerative colitis.

### DLG5

We observed an association of SNPs rs2289310 (P1371Q, tagging haplotype C) with Crohn's disease ( $p = 0.024$ ; OR, 0.76; CI, 0.59 to 0.97) but not with ulcerative colitis. The previously described Crohn's disease-associated SNP rs1248696 (R30Q) was not associated with Crohn's disease ( $p = 0.08$ ) but showed moderate association with ulcerative colitis in our cohort ( $p = 0.018$ ; OR, 0.81; CI, 0.68 to 0.97). SNP rs2165047 was associated with both

Crohn's disease ( $p = 0.0029$ ; OR, 1.19; CI, 1.06 to 1.33) and ulcerative colitis ( $p = 0.035$ ; OR, 1.15; CI, 1.01 to 1.30). Although SNP rs2289311 was not associated with Crohn's disease ( $p = 0.79$ ) we did observe an association with the subgroup with colonic disease localisation ( $p = 0.036$ ; OR, 1.20; CI, 1.01 to 1.43). We did not find any gender specific associations of any of the tested SNPs in *DLG5*.

### ATG16L1

For the *ATG16L1* rs2241880 SNP we observed an increased frequency of the major allele G in Crohn's disease cases (61%) versus controls (56%) ( $p = 0.00036$ ; OR, 1.22; CI, 1.09 to 1.35). On subgroup analysis rs2241880 was associated with stricturing behaviour ( $p = 0.005$ ; OR, 1.27; CI, 1.07 to 1.50) and perianal disease ( $p = 0.035$ ; OR, 1.19; CI, 1.01 to 1.40).

### IL23R

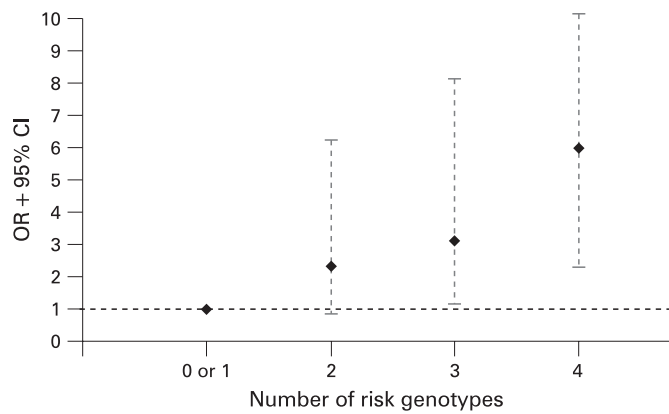
Carriers of allele A for SNP rs11209026 have a decreased risk for developing Crohn's disease ( $p = 2.12 \times 10^{-17}$ , OR, 0.31; CI, 0.23 to 0.40) and ulcerative colitis ( $p = 3.96 \times 10^{-4}$ ; OR, 0.62; CI, 0.48 to 0.81).

### Interaction of *NOD2*, *IBD5*, *DLG5*, *ATG16L1* and *IL23R* for Crohn's disease susceptibility

Since the strongest associations are found with Crohn's disease and not with ulcerative colitis, we performed interaction analysis only for patients with Crohn's disease. For each individual gene we selected the strongest associated genetic variant for further analysis as described above. Interaction analysis was performed for the following five associated risk genotypes: *NOD2* positive status (carrying one or more mutations for R702W, G908R or 3020insC), carrying the rs2522027\*C allele for *IBD5*, carrying the rs2289310\*C allele for *DLG5* and carrying the rs2241880\*G allele for *ATG16L1*. Since the allele rs11209026\*A for *IL23R* has a protective effect for Crohn's disease, and hence the major allele G is associated with an increased risk for Crohn's disease, we used allele G for the interaction analysis. Furthermore, since homozygosity for the minor allele is very rare (0.2%), the risk “allele” for *IL23R* is defined as homozygosity for rs11209026\*G.

First, we evaluated whether there is a specific gene–gene interaction. We performed two-way, three-way, four-way and five-way gene–gene interaction analysis. This showed moderate significant interaction between *IBD5* and *NOD2*, *DLG5* and *NOD2* and *IBD5*, *NOD2* and *IL23R*. For all other combinations this did not show any evidence for statistical interaction in addition to the main effects of the individual genes, indicating that each gene contributes independently to the susceptibility to Crohn's disease (table 5).

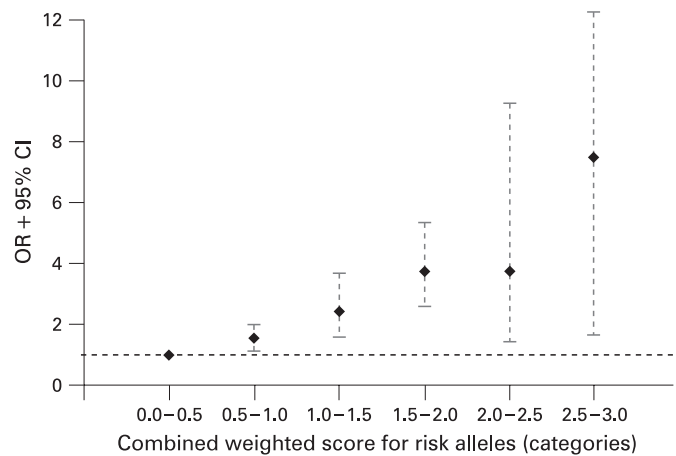
Next, we performed ordinal regression analysis which showed that patients with Crohn's disease carry more risk alleles (mean 4.41, SD 1.23) than controls (mean 3.84, SD 1.21); ordinal regression analysis ( $p = 3.85 \times 10^{-22}$ ). Binary logistic regression analysis showed that individuals carrying an increasing number of risk alleles were at increasing risk for Crohn's disease compared to controls carrying zero or one risk alleles (fig 1). This reference group was chosen because none of the controls carried zero risk alleles. We found that the OR for susceptibility to Crohn's disease increases with an increasing number of risk alleles. Individuals with six risk alleles had an OR of 7.56 (CI, 2.78 to 20.57) for susceptibility to Crohn's disease compared to the reference group with zero or one risk allele. Individuals with seven risk alleles had an OR of 25.6; however, since numbers in



**Figure 2** Odds ratios (ORs) and confidence intervals (CIs) for patients carrying increasing numbers of genotypes for developing Crohn's disease. The genotypes are *NOD2* (carrying R702W, G908R or 3020insC), *IBD5* (carrying the rs2522027\*C allele), *DLG5* (carrying the rs2289310\*C allele), *ATG16L1* (carrying the rs2241880\*G allele) and *IL23R* (carrying the rs11209026\*G allele). The reference group consists of individuals with zero or one risk genotype. *ATG16L1*, autophagy-related 16-like 1 gene; *DLG5*, Drosophila discs large homologue 5 gene; *IBD5*, inflammatory bowel disease 5 gene; *IL23R*, interleukin 23 receptor gene; *NOD2*, nucleotide-binding and oligomerisation domain 2 gene.

this group are becoming small ( $n = 60$ ), the CI is large (6.80 to 96.46). There were too few patients with eight or nine risk alleles to justify further statistical analysis. Trend analysis revealed a highly significant progressive increase in ORs with increasing number of risk alleles ( $p_{\text{trend}} = p = 4.25 \times 10^{-23}$ ). In fig 2 we show that this effect was also seen with an increasing number of risk genotypes ( $p_{\text{trend}} = 2.37 \times 10^{-27}$ ). Since different genetic variants have differing effects on disease susceptibility, we performed the same analysis using a weighted score based on the coefficients calculated from separate logistic regression analyses for each SNP associated with Crohn's disease, as explained in the methods section. A similar progressive increase in ORs for an increasing weighted score is seen (fig 3).

We then tested whether an increasing number of risk alleles was associated with a more severe disease course. For 1353 patients with Crohn's disease we had complete genotypes for all associated genes and complete phenotypic descriptions. First we found that patients with Crohn's disease with a more severe disease course, defined as penetrating or stricturing disease behaviour, had significantly more risk alleles compared to patients with non-stricturing, non-penetrating behaviour (ordinal regression,  $p = 0.001$ ;  $p_{\text{trend}} = 0.0008$ ). Similarly, patients who needed surgical intervention had significantly more risk alleles compared to those without operations (ordinal regression,  $p = 0.03$ ;  $p_{\text{trend}} = 0.02$ ). Patients with an age of onset below 40 years had more risk alleles than patients with an age of onset above 40 years (ordinal regression,  $p = 0.048$ ;  $p_{\text{trend}} = 0.028$ ). There was no association for the number of risk alleles and the occurrence of perianal disease or extra-intestinal manifestations. Since phenotypes can change over time we then performed these analyses on the groups of patients with Crohn's disease with complete genotyping data available and with more than 5 years ( $n = 1072$ ) and more than 10 years ( $n = 763$ ) of follow-up. It has to be emphasised that these groups partly overlap. These results are depicted in table 6. Although it would be expected that the group with more than 10 years of follow-up is enriched with patients with complicated disease behaviour, we could not find any association with



**Figure 3** Odds ratios (ORs) and confidence intervals (CIs) for a combined risk score based on the relative weight for each Crohn's disease risk allele. The coefficient score for the Crohn's disease risk alleles was determined by logistic regression analysis. Each patient was categorised based on the sum of these coefficients. The risk alleles were *NOD2* – the presence of one of the R702W, G908R or 3020insC polymorphisms; *IBD5* – the rs2522027\*C allele; *ATG16L1* – the rs2241880\*G allele; *DLG5* – the rs2289310\*C allele; the risk "allele" for *IL23R* is defined as homozygosity for rs11209026\*G. *ATG16L1*, autophagy-related 16-like 1 gene; *DLG5*, Drosophila discs large homologue 5 gene; *IBD5*, inflammatory bowel disease 5 gene; *IL23R*, interleukin 23 receptor gene; *NOD2*, nucleotide-binding and oligomerisation domain 2 gene.

worse disease behaviour in this particular group. This is probably due to a lack of power in this specific subset.

## DISCUSSION

We have looked at the combined impact of five genetic variants that show association with susceptibility to Crohn's disease in the Dutch IBD population and shown that the relative risk for Crohn's disease increases with an increasing number of risk alleles, consistent with an independent effects multiplicative model. We also showed that an increasing number of risk alleles is associated with a more severe disease course and a younger age of onset.

It is essential that genetic associations found for IBD are confirmed in large, independent cohorts.<sup>29</sup> We first corroborated the association with *NOD2* and then confirmed the association of the *IBD5* haplotype with Crohn's disease. Although the previously described "TC haplotype" in *SLC22A4/5* was over-represented in patients with IBD compared to controls in our study, this effect was not independent of the *IBD5* haplotype.<sup>8</sup> We did find an association for *DLG5* with IBD. However, this association was not with the R30Q variant but with the rs2289310 (P1371Q) and rs2165047 ("haplotype B") SNPs.<sup>9 10</sup> The latter SNP was also associated with ulcerative colitis, which is a novel finding. This indicates that *DLG5* might be a (modest) risk factor for both phenotypes of IBDs.<sup>9</sup> We were also able to confirm the association of the recently identified *ATG16L1* gene with Crohn's disease. On a phenotypic level we found it to be associated with Crohn's disease with stricturing disease behaviour and perianal disease. Although a recent report found a low but significant risk for ulcerative colitis, we did not find an association for *ATG16L1* with ulcerative colitis or any specific subsets of ulcerative colitis.<sup>30</sup> *IL23R* SNP rs11209026 (R381Q) was strongly associated with a protective effect for Crohn's disease in our cohort. We also found a significant but

**Table 6** Results of ordinal regression analysis for the number of risk alleles associated with Crohn's disease, depicting whether an increasing number of risk alleles is associated with specific disease characteristics

Characteristic	All (n = 1353)	Follow-up >5 years (n = 1072)	Follow-up >10 years n = 763
Age at diagnosis <40 years	<b>0.048</b>	0.06	0.58
Familial occurrence	0.12	<b>0.03</b>	0.69
Stricturing and/or penetrating behaviour	<b>0.001</b>	<b>0.006</b>	0.052
Perianal disease	0.57	0.60	0.67
Operation	<b>0.03</b>	0.07	0.13
Extra-intestinal manifestations	0.17	0.42	0.85

Values in bold type are statistically significant with a p value <0.05.

The risk alleles were: *NOD2*, the presence of one of the R702W, G908R or 3020insC polymorphisms was defined as one risk allele; *IBD5*, the rs2522027\*C allele; *ATG16L1*, the rs2241880\*G allele; *DLG5*, the rs2289310\*C allele; the risk "allele" for *IL23R* is defined as homozygosity for rs11209026\*G.

*ATG16L1*, autophagy-related 16-like 1 gene; *DLG5*, Drosophila discs large homologue 5 gene; *IBD5*, inflammatory bowel disease 5 gene; *IL23R*, interleukin 23 receptor gene; *NOD2*, nucleotide-binding and oligomerisation domain 2 gene.

less pronounced protective effect in ulcerative colitis patients. This is consistent with recent reports.<sup>13–15</sup>

We demonstrated that it is possible to combine information from multiple common low-penetrance variants to predict the susceptibility for a complex disease like Crohn's disease. Our findings are consistent with the idea that multiple genes, each with a small overall contribution, are involved in the susceptibility to Crohn's disease. For *NOD2* we used the polymorphisms associated with Crohn's disease combined with the most strongly associated SNPs for the *IBD5* locus, *DLG5*, *ATG16L1* and *IL23R*, for our interaction analysis. These genes (with the exception of *NOD2*) have low ORs (<1.5) for the risk of developing Crohn's disease and are therefore, independently, of limited use in predicting the risk for disease development. However, although there is very little specific gene–gene interaction, combining the information from these common risk alleles increases the predictive value for developing Crohn's disease. There is a significant progressive increase in the OR for Crohn's disease with each added risk allele. Individuals carrying six risk alleles had an OR of 7.56 compared to individuals carrying zero or one risk allele. Since different genetic variants have differing effects on disease susceptibility, the same analysis was performed using categories based on the sum of the weighted scores for each risk allele. The same progressive increase in ORs is observed. Furthermore, since the effect size of the *NOD2* gene might be a dominant factor in the analysis, we performed the same analysis without including data of *NOD2*. Similarly, since *DLG5* is less replicated than the other Crohn's disease susceptibility genes, we performed the same analysis without including *DLG5*. The same significant increase in ORs consistent with an independent effects multiplicative model was observed in both instances (data not shown). In a recent meta-analysis of three genome-wide association scans that have been performed more than 30 distinct susceptibility loci have been identified. When performing additional gene–gene interaction analysis, no strong evidence of deviation from this model of multiplicative effects was found in this large study.<sup>31</sup>

The absolute difference in the number of risk alleles between patients with Crohn's disease and controls is, although highly statistically significant, moderate. This is due to the fact that some of the identified Crohn's disease susceptibility variants, like *ATG16L1*, are common variants with high prevalences in the healthy population. In combination with the low prevalence of Crohn's disease in the population this obviously limits the predictive value for an individual for disease development.<sup>2</sup> Very recently, a number of genome-wide association studies have identified several more genetic associations with Crohn's

disease.<sup>16–18</sup> It can be anticipated that, once these associations are confirmed, they will further increase the risk for developing Crohn's disease. On the other hand, it can also be hypothesised that there will be no further increase in disease risk, but a plateau phase will be reached. In our total group of 4154 individuals studied, only six carried eight risk alleles. Five of them were patients with Crohn's disease and one was a healthy control. Therefore, since the numbers of individuals with more risk alleles will be even smaller, much larger cohorts than ours will be needed to analyse this.

An increasing number of risk alleles is not only associated with the susceptibility to Crohn's disease but also with a more severe course of the disease. It is associated with an age of onset <40 years, fibrostenotic or fistulising disease behaviour and the need for surgical intervention. We did not find any associations for the number of risk alleles with perianal disease or extra-intestinal manifestations. Since disease phenotypes can change over time it can be expected that the patient group with a long-term follow-up would be enriched with patients with complications like fibrostenotic disease and surgery. We therefore analysed patients with a follow-up of more than 5 years and more than 10 years separately for disease behaviour. We found an association with stricturing and/or penetrating behaviour and familial occurrence in the patients with a follow-up of more than 5 years. However, all associations were lost when we analysed patients with a follow-up of more than 10 years. This could be due to a lack of power in this specific subset. We show that molecular testing can help to identify specific subgroups. Since Crohn's disease is a chronic disorder with an unpredictable disease course, it is important for clinicians to identify these patients, since they might benefit from early intervention and a more aggressive treatment of Crohn's disease. On the other hand, patients with a favourable prognosis can be spared the side effects of unnecessary treatment. In the future, it might be possible to classify patients with Crohn's disease based on their genetic background in combination with other factors such as circulating antibodies, and environmental factors such as smoking. This will allow clinicians to identify patient groups and to adjust their treatment. Obviously, large population-based prospective studies will be needed to establish this.

In conclusion, we have shown that if we combine information from the risk-associated genetic variants for *NOD2*, *IBD5*, *DLG5*, *ATG16L1* and *IL23R*, the relative risk for developing Crohn's disease increases with an increasing number of risk alleles or genotypes, consistent with an independent effects multiplicative model. An increase in the number of risk alleles is also associated with a more severe disease course. These findings

emphasise that it might be possible to create a genetic risk profile to predict the disease course for a patient with Crohn's disease and to adjust their treatment accordingly.

**Acknowledgements:** We thank all the healthy individuals, patients and their physicians for participating in this study. We would like to thank E Oosterom, Department of Medical Genetics, University Medical Center Groningen, and E Vogels and S Meisner, Department of Experimental Internal Medicine of the Amsterdam Medical Center, for performing the genotyping. We thank J Senior for critically reading the manuscript.

**Funding:** This study was supported by a Gastrostart grant from the Dutch Society for Gastroenterology.

**Competing interests:** None.

**Ethics approval:** The study was approved by the ethics review committees of the participating hospitals on various dates between January 2000 and December 2003.

## REFERENCES

- Podolsky DK. Inflammatory bowel disease. *New Engl J Med* 2002;**347**:417–29.
- Binder V, Both H, Hansen PK, *et al*. Incidence and prevalence of ulcerative colitis and Crohn's disease in the County of Copenhagen, 1962 to 1978. *Gastroenterology* 1982;**83**:563–68.
- Orholm M, Munkholm P, Langholz E, *et al*. Familial occurrence of inflammatory bowel disease. *New Engl J Med* 1991;**324**:84–8.
- Halfvarson J, Bodin L, Tysk C, *et al*. Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterology* 2003;**124**:1767–73.
- 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.
- Ogura Y, Bonen DK, Inohara N, *et al*. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001;**411**:603–6.
- Rioux JD, Daly MJ, Silverberg MS, *et al*. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 2001;**29**:223–8.
- Peltekova VD, Wintle RF, Rubin LA, *et al*. Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat Genet* 2004;**36**:471–5.
- Stoll M, Corneliusen B, Costello CM, *et al*. Genetic variation in DLG5 is associated with inflammatory bowel disease. *Nat Genet* 2004;**36**:476–80.
- Browning BL, Barclay ML, Bingham SA, *et al*. Gender-stratified analysis of DLG5 R30Q in 4707 Crohn's disease patients and 4973 controls from 12 Caucasian cohorts. *J Med Genet* 2008;**45**:36–42.
- Hampe J, Franke A, Rosenstiel P, *et al*. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* 2007;**39**:207–11.
- Rioux JD, Xavier RJ, Taylor KD, *et al*. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* 2007;**39**:596–604.
- Weersma RK, Zhernakova A, Nolte IM, *et al*. ATG16L1 and IL23R are associated with inflammatory bowel diseases but not with celiac disease in the Netherlands. *Am J Gastroenterol* 2008;**103**:621–7.
- Duerr RH, Taylor KD, Brant SR, *et al*. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 2006;**314**:1461–3.
- Tremelling M, Cummings F, Fisher SA, *et al*. IL23R variation determines susceptibility but not disease phenotype in inflammatory bowel disease. *Gastroenterology* 2007;**132**:1657–64.
- Libioulle C, Louis E, Hansoul S, *et al*. Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet* 2007;**3**:e58.
- Franke A, Hampe J, Rosenstiel P, *et al*. Systematic association mapping identifies NELL1 as a novel IBD disease gene. *PLoS ONE* 2007;**2**:e691.
- Parkes M, Barrett JC, Prescott NJ, *et al*. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet* 2007;**39**:830–2.
- McGovern DP, Butler H, Ahmad T, *et al*. TUCAN (CARD8) genetic variants and inflammatory bowel disease. *Gastroenterology* 2006;**131**:1190–6.
- Yang Q, Khoury MJ, Botto L, *et al*. Improving the prediction of complex diseases by testing for multiple disease-susceptibility genes. *Am J Hum Genet* 2003;**72**:636–49.
- Weedon MN, McCarthy MI, Hitman G, *et al*. Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. *PLoS Med* 2006;**3**:1877–82.
- Stokkers PC, Huibregtse K Jr, Leegwater AC, *et al*. Analysis of a positional candidate gene for inflammatory bowel disease: NRAMP2. *Inflamm Bowel Dis* 2000;**6**:92–8.
- van Bodegraven AA, Curley CR, Hunt KA, *et al*. Genetic variation in myosin IXB is associated with ulcerative colitis. *Gastroenterology* 2006;**131**:1768–74.
- Zhernakova A, Festen EM, Franke L, *et al*. Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP. *Am J Hum Genet* 2008;**82**:1202–10.
- Wagtman MJ, Verspaget HW, Lamers CB, *et al*. Gender-related differences in the clinical course of Crohn's disease. *Am J Gastroenterol* 2001;**96**:1541–6.
- de Jong DJ, van der Logt EM, van Schaik A, *et al*. Genetic polymorphisms in biotransformation enzymes in Crohn's disease: association with microsomal epoxide hydrolase. *Gut* 2003;**52**:547–51.
- Gasche C, Scholmerich J, Brynskov J, *et al*. A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. *Inflamm Bowel Dis* 2000;**6**:8–15.
- de Ridder L, Weersma RK, Dijkstra G, *et al*. Genetic susceptibility has a more important role in pediatric-onset Crohn's disease than in adult-onset Crohn's disease. *Inflamm Bowel Dis* 2007;**13**:1083–92.
- Moonesinghe R, Khoury MJ, Janssens AC. Most published research findings are false – but a little replication goes a long way. *PLoS Med* 2007;**4**:e28.
- Prescott NJ, Fisher SA, Franke A, *et al*. A nonsynonymous SNP in ATG16L1 predisposes to ileal Crohn's disease and is independent of CARD15 and IBD5. *Gastroenterology* 2007;**132**:1665–71.
- Barrett JC, Hansoul S, Nicolae DL, *et al*. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008;**40**:955–62.