Genetic Analysis in A Dutch Study Sample Identifies More Ulcerative Colitis Susceptibility Loci and Shows Their Additive Role in Disease Risk

Eleonora A.M. Festen, MD[1,2], Pieter C.F. Stokkers, MD, PhD[3], Cleo C. van Diemen, PhD[2], Adriaan A. van Bodegraven, MD, PhD[4], H. Marieke Boezen, PhD[5], Bart J.A. Crusius, PhD[6], Daniel W. Hommes, MD, PhD[7], Janneke C. van der Woude, MD, PhD[8], Tobias Balschun, MSc[9], Hein W. Verspaget, PhD[7], Stephan Schreiber, MD, PhD[9], Dirk J. de Jong, MD, PhD[10], Andre Franke, PhD[9], Gerard Dijkstra, MD, PhD[1], Cisca Wijmenga, PhD[2,11] and Rinse K. Weersma, MD, PhD[1,11]

OBJECTIVES: **Genetic susceptibility is known to make a major contribution to the pathogenesis of ulcerative colitis (UC). Recently, three studies, including a genome-wide association study (GWAS), reported novel UC risk loci.**

METHODS: **The top-20 single-nucleotide polymorphisms (SNPs) from the first UC-GWAS were genotyped, as part of the study's replication phase, in 561 UC cases and 728 controls from our Dutch UC study sample. We genotyped eight SNPs identified in two more studies, in these individuals, and replicated all significantly associated SNPs in an additional 894 UC cases and 1,174 controls from our Dutch UC study sample. A combined analysis for all patients (*n*=1,455) and controls (*n*=1,902) was performed. Dose–response models were constructed with the associated risk alleles.**

RESULTS: **We found 12 SNPs tagging ten loci, including *HLA-DRA*, *IL10, IL23R*, *JAK2, S100Z, ARPC2,* and *ECM1,* to be associated with UC. We identified 10q26, flagged by the UC-GWAS but not confirmed in its replication phase, as a UC locus and found a trend toward association for *GAS7*. No association with disease localization or severity was found. The dose–response models show that individuals carrying 11 or more risk alleles have an odds ratio of 8.2 (confidence interval 3.0–22.8) for UC susceptibility.**

CONCLUSIONS: **We confirmed the association of multiple loci with UC in the Dutch population and found evidence for association of 10q26 and a trend suggesting association for *GAS7*. Genetic models show that multiple risk loci in an individual increase the risk for developing UC.**

SUPPLEMENTARY MATERIAL is linked to the online version of the paper at http://www.nature.com/ajg

Am J Gastroenterol 2010; 105:395–402; doi:10.1038/ajg.2009.576; published online 27 October 2009

INTRODUCTION

Ulcerative colitis (UC) is an inflammatory disease of the colon, which along with Crohn's disease (CD) comprises the major part of the inflammatory bowel diseases (IBD). After asthma and rheumatoid arthritis, IBD is the most common form of chronic inflammatory disease in the world, with a prevalence of 6–243/100,000 for UC and of 3.6–214/100,000 for CD (1). IBD is a complex disease, which means that both environmental and genetic factors determine the development and course of the disease. In UC, the involvement of genetic factors is evident from the higher concordance of disease in monozygotic twins (10%) compared with dizygotic twins (3%) (2).

[1]Department of Gastroenterology and Hepatology, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands; [2]Department of Genetics, University Medical Centre Groningen and University of Groningen, Groningen, The Netherlands; [3]Department of Gastroenterology and Hepatology, Academic Medical Centre, Amsterdam, The Netherlands; [4]Department of Gastroenterology and Hepatology, VU University Medical Centre, Amsterdam, The Netherlands; [5]Department of Epidemiology, University Medical Centre Groningen and University of Groningen, Groningen, The Netherlands; [6]Department of Immunogenetics, VU University Medical Centre, Amsterdam, The Netherlands; [7]Department of Gastroenterology and Hepatology, Leiden University Medical Centre, Leiden, The Netherlands; [8]Department of Gastroenterology and Hepatology, Erasmus Medical Centre, Rotterdam, The Netherlands; [9]Institute for Clinical Molecular Biology, Christian-Albrechts University, Kiel, Germany; [10]Department of Gastroenterology and Hepatology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; [11]These authors contributed equally to this work. **Correspondence:** Rinse K. Weersma, MD, PhD, Department of Gastroenterology and Hepatology, University Medical Centre Groningen, PO Box 30001, Groningen, 9700 RB, The Netherlands. E-mail: R.K.Weersma@int.umcg.nl
Received 17 March 2009; accepted 3 September 2009

Crohn's disease has been the subject of many genetic studies and more than 30 CD risk loci have been identified in the past decade (3). Identification of these loci has led to a better understanding of the disease mechanisms and, in the future, diagnosing CD or predicting the course of disease on the basis of genetic tests will become feasible (4). The search for genetic risk factors involved in UC has been less fruitful but this has recently changed with the publication of several papers focusing on the genetics of UC. Both Franke *et al.* (5) and Fisher *et al.* (6) identified new UC risk loci by testing previously known CD risk loci in UC patients. Fisher *et al.* (6) also performed a nonsynonymous single-nucleotide polymorphism (SNP) scan, which led to the identification of another UC locus. In November 2008, the first genome-wide association study (GWAS) for UC was performed and identified six new risk loci for this disease (7). In February 2009, a second GWAS for UC identified two new risk loci (8). This paper was published after our study had been started and was not taken into consideration. New UC loci will increase our understanding of UC pathogenesis and may eventually become instrumental in disease diagnosis and choice of therapy.

We aimed to further study the genetic background of UC, by confirming and expanding the recent findings of the first UC GWAS and the results of Fisher *et al.* (6) and Franke *et al.* (7) in the full UC study sample from the Dutch Initiative on Crohn and Colitis (ICC). The top-20 SNPs from the first UC GWAS were tested in the first part of our ICC UC study sample as part of the replication phase, and these results were published in combination with a Belgian study sample. We set out to test these and two additional SNPs that were significant in the first part of our study sample in the entire ICC UC study sample of 1,455 cases and 1,902 controls. Once we knew which loci showed statistically significant association in the full Dutch ICC UC study sample, we tested for genotype–phenotype association and constructed genetic dose–response models based on these loci.

METHODS

Subjects

All UC cases in this study are part of the IBD study sample of the Dutch ICC. The study sample consists of two parts. Part 1 is comprised of the 561 UC patients from the Academic Medical Centre, Amsterdam ($n = 334$) and the University Medical Centre Groningen ($n = 227$). These cases were used in the confirmation phase of the original GWAS of Franke *et al.* (7). The controls for this study were 728 blood bank donors collected by the University Medical Centre Utrecht and the VU University Medical Centre, Amsterdam. Part 2 of the study sample is comprised of 894 UC patients collected from the outpatient clinics of the Departments of Gastroenterology and Hepatology at the VU University Medical Centre Amsterdam ($n = 341$), the University Medical Centre Leiden ($n = 213$), the Erasmus University Medical Centre, Rotterdam ($n = 73$), and the Radboud University Medical Centre, Nijmegen, the Netherlands ($n = 267$). Controls for part 2 came from an independent study sample of 1,174 controls from the UMC Leiden ($n = 169$), the VUMC Amsterdam ($n = 180$), and a general population-based cohort collected by the UMC Groningen from the Dutch towns Vlagtwedde and Vlaardingen ($n = 825$). These study samples have all been described in previous publications (4,7,9,10). **Table 1** shows a comparison of cases and controls on the basis of age and gender. Cases were on an average 3 years younger than controls; this difference was significant in an independent samples *t*-test with a *P* value of 0.001. The case group contained relatively more males than the control group, but this difference was not significant with a *P* value of 0.4 in a χ^2-test.

Table 1. Comparison of gender and age for cases and controls

	Cases	Controls
Age (years)		
Median	48	51
Min	8	22
Max	97	84
Gender		
Male	52.0%	50.4%
Female	48.0%	49.6%

The diagnosis of UC required (a) one or more symptoms of diarrhea, rectal bleeding, abdominal pain, or fever, (b) occurrence of symptoms on two or more occasions separated by at least 8 weeks or ongoing symptoms of at least 6 weeks duration, and (c) objective evidence of inflammation from radiologic, endoscopic, and histopathologic evaluation. All affected subjects fulfill the clinical criteria for UC. Information on sex and date of birth was available for all patients and controls. Complete phenotypic descriptions were available for 1,291 UC patients and are presented in **Table 2**. Phenotypes of the UC patients were described according to age of onset, maximum extent of disease (proctitis, left-sided, or extensive), necessity of colectomy, and the occurrence of malignancy and extraintestinal manifestations. In all cases, informed consent was obtained using protocols approved by the local institutional review board in all participating institutions. DNA extraction was successful in >99% of the cases, and all DNA samples and data in this study were handled anonymously.

Genotyping and SNP selection

For the first part of the ICC UC study sample, genotypes of the 20 statistically most significantly associated SNPs from the recently published UC-GWAS were obtained and published as part of the confirmation phase for that study. These included the nine associated SNPs that were confirmed in the combined analysis and 11 more SNPs that were associated in the GWAS but not confirmed in the combined analysis (7). We included these data in our study to give a comprehensive view

Table 2. Detailed phenotypic characteristics of ulcerative colitis patients in the Dutch ICC study sample

	Part 1		Part 2	
		%		%
Total number	561		894	
Detailed phenotypes available	426	76.2%	876	98%
Sex (M/F)	290/271	51.7–48.3%	470/424	52.6–47.4%
Age at diagnosis (years)				
Mean (s.d.)	31.6 (13.1)		33.9 (13.9)	
Min/max	4/73		6/86	
Disease localization				
Proctitis	62	13.1%	121	14.8%
Left sided	100	21.2%	318	38.8%
Extensive	198	41.9%	364	44.4%
Extra-intestinal manifestations	65	13.8%	51	6.2%
Colectomy	97	20.6%	89	10.9%
Malignancy	2	0.4%	2	0.2%
Family history of IBD	51	10.8%	78	9.5%
Mean follow-up (years)	15.8 (s.d. 9.5)		10.0 (s.d. 8.9)	

IBD, inflammatory bowel disease; ICC, Initiative on Crohn's and Colitis. Total number = all patients included in the study; detailed phenotypes available = number of patients for whom detailed phenotypes were available; % = percentage of all the patients included in the study.

of the association of known UC loci in the Dutch UC population, so that we could determine which loci to include in our dose–response modeling and genotype–phenotype association analysis, as well as to increase our power for performing these analyses. The data for the top-20 SNPs from the UC-GWAS in part 1 of the Dutch UC study sample were obtained by SNPlex technologies (Applied Biosystems, Foster City, CA) performed at the Department of Genetics, Christian-Albrechts University, Kiel, Germany, from September to October 2007. We genotyped and analyzed these 20 SNPs in part 2 of the ICC UC study sample and in 1,174 controls. The assay for SNP rs9268858 in the HLA-DRA locus could not be designed and was replaced by rs9268853, which is a perfect proxy (D' = 1.0, $r^2 = 1.0$) of rs9268858. A combined analysis of the top-20 GWAS SNPs in the entire ICC study sample was carried out. We also genotyped eight SNPs for eight new UC loci (5,6) in part 1 of the ICC study sample. Only two of these eight SNPs were significant ($P < 0.05$) in part 1 of the study sample: rs13085791 in macrophage stimulating 1 (*MST1*) and rs13294 in extra cellular matrix protein 1 (*ECM1*); these were then further genotyped in the rest of the ICC study sample. Hence, genotyping data were available for 22 SNPs in the full ICC study sample. However, the *MST1* SNP rs13085791 failed in many of our controls and to solve this, we replaced this SNP by rs3197999, which is a perfect proxy (D' = 1, $r^2 = 0.97$) of rs13085791 (6). All the SNPs tested are described in **Supplementary Table 1** online.

Genotyping was performed at the Department of Genetics, UMC Groningen, the Netherlands, using TaqMan technology and SNP genotyping assays for PCR obtained from Applied Biosystems (Nieuwerkerk a/d IJssel, the Netherlands) from June to July 2008. The patient and control DNA samples were processed in 384-well plates and each plate also contained 16 genotyping controls (four duplicates of the Centre d'Etudes du Polymorphisme Humain (CEPH) DNA samples 123002, 102405, 090203, and 081505). For all SNPs we obtained >99.8% concordance between our CEPH genotype data and the CEU (European ancestry) data available from HapMap. All samples of patients and controls were analyzed in the same batch.

Statistical analysis

All genotypes obtained were tested for Hardy–Weinberg equilibrium by χ^2-testing. Controls showed no deviation from Hardy–Weinberg equilibrium at $P < 0.05$. Differences in allele and genotype distribution in the cases and controls in each phase of the study were tested for significance by χ^2-test. The significance threshold was determined at $P < 0.05$. As all UC samples were taken from the Dutch ICC study sample, a combined analysis of parts 1 and 2 of this study sample was performed using χ^2-test and because all loci were specifically chosen as likely candidates for UC susceptibility, a Bonferroni multiple testing correction was only applied to correct for the two phases of testing. Odds ratios (ORs) were calculated and the confidence intervals were approximated using Woolf's method with Haldane's correction.

Two risk models for UC susceptibility were constructed based on all the SNPs that showed association ($P < 0.05$) in the full ICC UC study sample. We tested for the independence of the SNPs with a pair wise logistic regression, apart from the

SNPs in the *HLA-DRA* region, all significantly associated SNPs appeared to be independent. In model 1, we calculated the cumulative number of risk alleles per individual and summed them; these could be either the minor or major alleles (if the minor allele had a protective effect on UC development). We made categories of similarly sized groups of individuals with a specific number of risk alleles and then calculated ORs for UC in a binary logistic regression analysis for each category with a reference group consisting of all the individuals with zero or only one risk allele or genotype. We chose this reference group because the group containing zero risk alleles was too small to be used as a reference. Student's *t*-test was used to see whether UC patients had more risk alleles than controls. Model 2 was a weighted score for the number of risk alleles, which was calculated per individual, as different genetic variants have different effect sizes on disease susceptibility. We calculated β-coefficients per genotype obtained from separate binary logistic regression analyses for each UC-associated SNP and summed these coefficients per individual to obtain a weighted genetic load (referred to as weighted score). In this way, the sum of risk alleles was adjusted for the strength of association for each genetic variant. All patients and controls were then categorized on the basis of weighted score, and ORs for each category were determined by logistic regression analysis, with the category with the lowest weighted score being used as a reference group.

Genotype–phenotype associations were calculated with χ^2-tests combining SNPs with phenotype subgroups of interest in 2×2 tables. A Bonferroni (multiple testing) correction was applied for the number of complementary subgroups of patients. All the statistical analyses were performed using SPSS version 14.0 (SPSS Inc., Chicago, IL).

RESULTS

Follow-up of signals from the UC-GWAS

Our results for the significantly associated SNPs from the top-20 SNPs in the UC-GWAS are shown in **Table 3**. The results for these SNPs from the first part of the ICC UC study sample have only been published in combination with a Belgian study sample. In the combined analysis of the complete ICC UC study sample we could confirm the association of eight SNPs tagging six loci that were identified and replicated in the UC-GWAS study by Franke *et al.* (7). The most strongly associated SNP was rs9268853 in the DR-α (*HLA-DRA*) gene in the class II major histocompatibility complex region. Two other SNPs from the same region, rs9268877 and rs9268480, were also replicated. We confirmed SNP rs3024505 in the interleukin-10 (*IL10*) locus and found SNP rs11805303 in the interleukin-23 receptor precursor (*IL23R*) gene to be strongly associated with UC. This latter SNP is not in linkage disequilibrium with the previously described *IL23R* SNP R381Q-rs11209026 (D' = 0.57, $r^2 = 0.01$), which has been reported to have a protective effect in IBD in contrast to this new *IL23R* risk conveying polymorphism (11). SNP rs10974944 in the janus kinase 2 (*JAK2*) locus, which was moderately associated in the original study, was found to be more strongly associated in our study. We also confirmed associations with SNP rs12612347 in *ARPC2* and SNP rs7712957 in *S100Z*.

From the remaining 12 SNPs that were flagged by the UC-GWAS but which could not be confirmed in the replication phases of that specific study, we found an association with an intergenic region on chromosome 10q26. We also found an association with SNP rs11078827 in the growth arrest-specific 7 (*GAS7*) gene in our combined analysis, but this association did not withstand the multiple testing correction. In part 2 of the UC study sample, we observed association with rs7831616 in the general transcription factor IIE, polypeptide 2 (*GTF2E2*) and rs527313 in an intergenic region on 4p15.33. However, these two SNPs did not show association in our combined analysis. The remaining eight SNPs that were significantly associated in the primary phase of the UC GWAS were not replicated in our study.

Replication of other susceptibility loci

From the other UC loci described by Franke *et al.* (7) and Fisher *et al.,* (6) two out of eight SNPs were found to be associated in part 1 of our study; (5,6) the results are shown in **Table 2**. SNP rs13085791 in the *MST1* locus was associated with UC in part 1, but in the full ICC study sample this association did not withstand multiple testing correction. SNP rs13294 in the *ECM1* locus was associated with UC in part 1 and even more strongly in the full ICC study sample. The remaining six SNPs were not found to be associated in the first part of our study and we did not pursue them in the remainder of the ICC study sample.

Analysis of the number and weight of susceptibility loci on UC risk

SNPs from the 10 associated UC loci were used to construct genetic dose–response risk models. For the *HLA-DRA* locus, we took only the most strongly associated SNP from the three associated ones, which resulted in 10 SNPs to be included. We tested for epistasis between the other loci with pair wise logistic regression and found all loci to be independent. As some SNPs did not follow a recessive but a dominant association, the maximum number of risk alleles was 18 instead of 20. An independent samples *t*-test on the number of risk alleles in UC patients and controls showed a significant difference (*P* value 8.9×10^{-24}) in the mean number of risk alleles carried by UC patients (mean 7.5, s.d. 2.1) and controls (mean 6.5, s.d. 2.2). This difference in the mean number of risk alleles is caused by a shift in the distribution of risk alleles between the two groups (**Figure 1**). Binary logistic regression showed that individuals with a higher number of risk alleles are at higher risk for UC compared with individuals carrying zero or one risk allele (**Figure 2a**). For example, individuals carrying 11 or more risk alleles have an OR of 8.2 (confidence interval 3.0–22.8) for UC susceptibility compared with the reference group.

Table 3. Allelic Association results for both phases of the study and the combined analysis

			Part 1				Part 2				Combined analysis			
			561 UC cases, 728 controls				894 UC cases, 1,174 controls				1,455 UC cases,1,902 controls			
Marker	Gene	A	AF controls	AF cases	*P* value	OR (95% CI)	AF controls	AF cases	*P* value	OR (95% CI)	AF controls	AF cases	Corrected *P* value	OR (95% CI)
rs9268858	*HLA-DRA*	G	0.26	0.21	2.4E-03[c]	0.75 (0.62–0.90)	0.30	0.23	3.0E-06	0.71 (0.62–0.82)	0.28	0.22	4.0E-08	0.72 (0.65–0.81)
Rs3024505	*IL10*	T	0.15	0.20	3.0E-04[c]	1.46 (1.18–1.79)	0.16	0.20	2.1E-04	1.36 (1.16–1.60)	0.15	0.20	7.8E-07	1.39 (1.23–1.58)
Rs11805303	*IL23R*	T	0.27	0.33	2.1E-03[c]	1.31 (1.10–1.55)	0.27	0.32	7.5E-05	1.32 (1.14–1.51)	0.27	0.32	1.1E-06	1.31 (1.18–1.46)
Rs9268877	*HLA-DRA*	T	0.44	0.50	1.0E-03[c]	1.30 (1.11–1.52)	0.44	0.49	5.4E-04	1.25 (1.10–1.42)	0.44	0.50	3.4E-06	1.27 (1.15–1.40)
Rs10974944	*JAK2*	G	0.26	0.30	0.0193[c]	1.23 (1.02–1.46)	0.25	0.29	1.7E-03	1.25 (1.09–1.44)	0.25	0.30	2.0E-04	1.24 (1.11–1.38)
Rs10886580	Intergenic	G	0.46	0.49	0.0922[c]	0.87 (0.75–1.02)	0.45	0.50	2.7E-03	1.21 (1.07–1.37)	0.45	0.50	1.2E-03	1.19 (1.08–1.31)
Rs9268480	*BTNL2*	A	0.24	0.20	8.9E-03[c]	0.78 (0.64–0.94)	0.26	0.23	0.0287	0.85 (0.73–0.98)	0.25	0.22	1.4E-03	0.82 (0.73–0.92)
Rs7712957	*S100Z*	G	0.08	0.11	0.0104[c]	1.30 (1.00–1.71)	0.08	0.10	0.0493	1.24 (1.00–1.54)	0.08	0.10	4.2E-03	1.30 (1.10–1.54)
Rs12612347	*ARPC2*	T	0.44	0.48	0.0894[c]	1.15 (0.98–1.34)	0.46	0.50	0.0328	1.15 (1.01–1.30)	0.45	0.49	0.0138	1.15 (1.04–1.26)
Rs11078827	*GAS7/GLP2R*	C	0.48	0.49	0.6111[c]	1.04 (0.89–1.22)	0.48	0.51	0.0145	1.17 (1.03–1.33)	0.48	0.51	0.0528	1.12 (1.01–1.23)
Rs527313	Intergenic	A	0.18	0.19	0.5038[c]	1.07 (0.88–1.31)	0.18	0.21	0.0294	1.21 (1.02–1.43)	0.18	0.20	0.3020	1.10 (0.97–1.24)
Rs7831616	*GTF2E2*	A	0.33	0.34	0.5226[c]	1.06 (0.89–1.25)	0.32	0.29	0.0261	0.86 (0.75–0.98)	0.33	0.31	0.4036	1.07 (0.96–1.19)
Rs13294[c]	*ECM1*	A	0.35	0.43	0.0019	1.36 (1.12–1.66)	0.37	0.42	0.005189	1.23 (1.06–1.43)	0.36	0.41	4.6E-04	1.24 (1.11–1.39)
Rs13085791[a] Rs3197999[b]	*MST1*	A	0.37	0.31	0.0065	0.76 (0.62–0.93)	0.26	0.29	0.209401	1.11 (0.94–1.30)	0.27	0.29	0.0844	1.14 (1.00–1.30)

A, minor allele; AF, minor allele frequency; CI, confidence interval; GWAS, genome-wide association study; OR, odds ratio; SNP, single-nucleotide polymorphism; UC, ulcerative colitis.
The top-20 SNPs from the UC-GWAS were genotyped in both parts of the study sample, and a combined analysis was performed. The other new loci were screened in part 1 and, if positive, replicated in part 2; only the positive SNPs are shown.
[a]Positive SNPs from the new loci (negative SNPs not shown).
[b]Due to technical problems rs13085791 was replaced by its proxy rs3197999 in the second part of the study sample. Corrected *P* value: the combined *P* value was corrected for the two phases of testing.
[c]Genotype data previously published as part of the replication phase of the UC GWAS by Franke *et al.* (5).

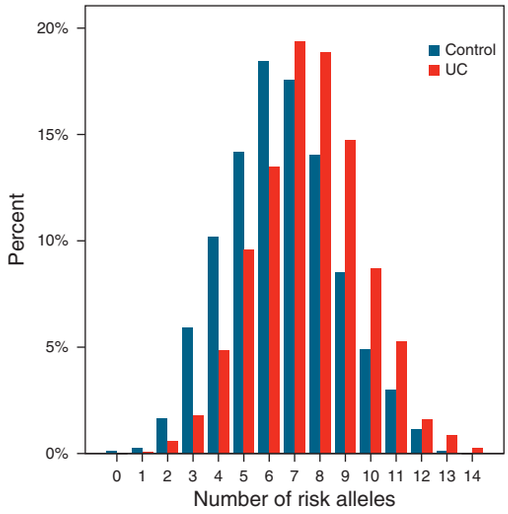

Figure 1. Graph showing the distribution of the number of risk alleles per individual for controls (blue bars) and cases (red bars). Both in cases and controls, the number of risk alleles per individual follows a normal distribution, but in cases this normal distribution is shifted to the right.

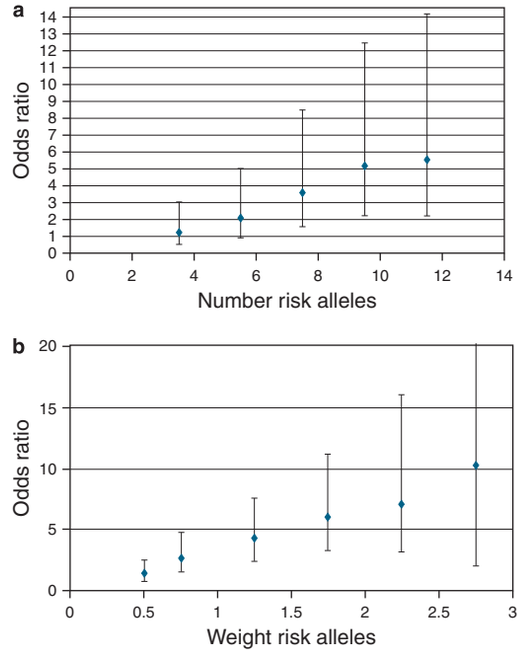

Figure 2. Effect of number of risk alleles on disease risk. (**a**) Graph showing the increase of the odds ratio for developing ulcerative colitis (UC) with an increasing number of risk alleles. Dots represent the mean odds in a group of patients with a certain range in number of risk alleles. The error bars indicate the 95% confidence interval of each mean. This confidence interval increases in size toward the higher number of risk alleles because the groups of individuals become smaller. (**b**) Graph showing the increase of the odds ratio for developing UC with the increasing weight carried in risk alleles. Dots represent the mean odds ratio in a group of patients with a certain range in number of risk alleles. The error bars indicate the 95% confidence interval of each mean. This confidence interval increases in size toward the higher number of risk alleles because the groups of individuals become smaller.

As different genetic variants have different effects on disease susceptibility, we decided to perform the same analysis on the basis of the β-coefficients calculated from separate binary logistic regression for each UC-associated SNP, as explained in the Methods section. This model showed an even larger increase of the OR for disease susceptibility with an increase of the weighted score in risk alleles (**Figure 2b**). As reference, we used a group of controls with a weighted score in risk alleles of 0.50 or less. Individuals with a weighted score in risk alleles over 2.50 had an OR for UC susceptibility of 10.3 (confidence interval 4.3–24.2) compared with this reference group.

Subsequently, we hypothesized that an increase in weighted score in risk alleles would cause a more severe disease course. We chose pancolitis and colectomy as measures of severe UC disease, but binary logistic regression showed no significant association between the risk allele load and these measures of severity.

Testing for genotype–phenotype association

We tested whether age of onset and gender changed our observed associations by adding these as covariates in a logistic regression analysis. Neither age of onset nor gender had an effect on the associations we observed. Using χ^2-tests we tested all associated SNPs for association with phenotypic characteristics such as gender, disease localization, and age of onset. We found associations for the *HLA-DRA* SNP rs9268853 with proctitis (P value = 0.014) and the *HLA-DRA* SNP rs9268877 with colectomy (P value = 0.037), but these associations disappeared when we applied a Bonferroni correction for multiple testing. No other genotype–phenotype associations were found.

DISCUSSION

We have identified 12 SNPs tagging 10 genetic risk loci as associated with UC in a Dutch study sample of 1,455 cases and 1,902 controls, and we have shown that the risk for UC increases with an increasing number of risk alleles per individual. The previously reported risk loci *HLA-DRA*, *IL10*, *IL23R*, *ARPC2*, and *ECM1* were confirmed in our Dutch UC study sample, whereas loci reported to be only moderately associated with UC, such as *JAK2* and *S100Z*, proved to be strongly associated in our study sample. We also found evidence for association of one additional locus that was flagged in the recent UC-GWAS but which could not be confirmed in its replication phase: an intergenic region on 10q26. For a locus on 17p13 harboring the growth arrest-specific 7 (*GAS7*) and glucagon-like peptide-2 receptor (*GLP2R*) genes, we found a trend toward association that did not withstand multiple testing correction (7). The newly identified regions bring the number of known UC loci to more than 20.

There are no genes in the direct vicinity of the associated SNP rs10886580 in the 10q26 region; however, about 413 kb upstream of this SNP is the gene *PPAPDC1A* (phosphatidic

acid phosphatase type-2 domain containing 1A) and 102 kb downstream is *SEC23IP* (Sec23 interacting protein), which is involved in the early secretory pathway. Whether one of these genes is responsible for the association cannot be resolved until further analysis of the region has been performed. The SNP rs11078827 for which we observed a trend toward association with UC is located in the *GAS7* gene on the minus strand. This gene is expressed in brain cells and is involved in neuronal development, which does not make it a likely candidate gene for association with UC. On the plus strand, SNP rs11078827 is located downstream of the *GLP2R* gene, which encodes the receptor for a peptide produced by intestinal enteroendocrine cells. GLP2 stimulates intestinal growth and upregulates villus height in the small intestine, by regulating increased crypt cell proliferation and decreased enterocyte apoptosis. *GLP2R* might thus be a more likely UC candidate gene to explain the association with rs11078827, although fine mapping of this locus is needed to identify the true causative variant(s).

A number of previously reported risk loci, including *STAT3*, *CADM2, IL12B*, *NKX2-3,* and *PTPN2,* were not replicated in our Dutch UC study sample. This does not necessarily mean that these are not truly UC-associated genes; it may merely reflect a lack of statistical power, even in our relatively large study sample.

A second UC-GWAS, from the USA, reported association with both the *HLA-DRA/BTNL2* locus and the *IL23R* region (8). This GWAS also identified two new UC loci: one containing the *IL26* gene and one containing phospholipase-A2 group IIE *PLA2G2E*. Another recently published study focused mainly on overlapping susceptibility loci between UC and CD (12) and showed that there is a considerable overlap in genetic background between UC and CD, but that there are also genes specifically associated to only one of the two diseases, such as *NOD2* to CD and *ECM1* to UC. In our study we also found an association with UC for several genes that are already known to be associated with CD, including *IL23R* and *JAK2*. This overlap of risk loci between the diseases suggests a partly shared pathogenesis for UC and CD, whereas the loci that are more specific for CD or UC may point to disease mechanisms specific for each disease. Contrary to our own findings, this other study was able to confirm *IL12B* and *STAT3* as UC-specific risk loci (12).

Although each of the replicated loci has only a very small individual effect on UC disease risk, our risk models clearly show that an individual with more risk alleles has an increased risk for UC. We observed a similar increase in risk with each additional risk allele in CD in a previous study (4). Although each individual risk allele only conveys an OR of about 1.2, an individual with six or more risk loci already has a strongly increased risk for developing UC, with an OR of 8.2. If we sum the weight of the individual effect of each risk locus, the risk for UC increases even more with an OR of 10.2 for the same number of risk loci. We would like to test our dose–response model in an independent cohort. We now used the same cases and controls for both replication of the loci and building the dose–response model, which might add a bias to this model. Unfortunately, a second cohort of sufficient size was currently unavailable. Predictive testing is not yet feasible as the difference in absolute number and weight of risk alleles between UC cases and control individuals is significant but small. This small difference is due to the fact that many disease-associated variants are common, i.e., highly prevalent in the general population. The identification of additional UC risk loci in GWAS that are now underway is expected to improve the sensitivity and specificity of the genetic dose–response risk model. Noninvasive genetic testing would be especially valuable in the differential diagnosis of IBD as it often remains difficult to differentiate between CD and UC with current diagnostic methods. A correct diagnosis is essential for the correct treatment of patients, as many drugs that are effective in one form of IBD have insufficient beneficial therapeutic effect in the other. The same holds for surgical therapy: proctocolectomy is a beneficial remedy for UC, but can induce severe complications in CD.

We have confirmed the association of eight loci (*HLA-DRA*, *IL10, IL23R*, *JAK2, S100Z, ARPC2,* and *ECM1*) with UC and identified a novel UC locus at 10q26. We have also shown that multiple risk loci in an individual increase the risk for UC development.

ACKNOWLEDGMENTS

We thank all the patients and controls who participated in this study and Jackie Senior for checking the final text.

CONFLICT OF INTEREST

Guarantor of the article: Rinse K. Weersma, MD, PhD.

Specific author contributions: E.A.M.F. performed the genotyping carried out at the UMCG and statistical analyses for the confirmation phase and risk models, and she wrote the article.; P.C.S. provided DNA samples and phenotyping data from the AMC; C.C.D. assisted with the statistical analyses for the complete study and participated in helpful discussions about the article; A.A.B. and J.B.A.C. provided DNA samples and phenotyping data from the VUMC; H.M.B. provided control samples from the Vlagtwedde-Vlaardingen cohort; D.W.H. and H.W.V. provided DNA samples and phenotyping data from the LUMC; C.J.W. provided DNA samples and phenotyping data from the EUMC; T.B. performed the genotyping carried out at the Christian-Albrechts University in Kiel and was supervised by S.S. and A.F.; G.D. and R.K.W. provided DNA samples and phenotyping data from the UMCG; D.J.J. provided DNA samples and phenotyping data from the NUMC; C.W. and R.K.W. supervised the project and helped to write the article.

Financial support: This study was made possible by a VICI grant to C.W. (918.66.620), a KF grant to R.W. (90700281), and an AGIKO grant to E.F. (92.003.533), both from the Netherlands Organization for Scientific Research.

Potential competing interest: None.

Study Highlights

WHAT IS CURRENT KNOWLEDGE

- ✓ Ulcerative colitis (UC) is a chronic inflammatory disease of the colon.
- ✓ Genetic susceptibility contributes to the pathogenesis of UC.

WHAT IS NEW HERE

- ✓ Several previously identified loci were confirmed with consistent association with UC.
- ✓ 10q26 is a susceptibility locus for UC.
- ✓ Multiple genetic risk loci in an individual each add to the risk for developing UC.

REFERENCES

1. Loftus EV Jr. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. Gastroenterology 2004;126:1504–17.
2. Baumgart DC, Carding SR. Inflammatory bowel disease: cause and immunobiology. Lancet 2007;369:1627–40.
3. 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.
4. Weersma RK, Stokkers PC, van Bodegraven AA *et al.* Molecular prediction of disease risk and severity in a large Dutch Crohn's disease cohort. Gut 2009;58:388–95.
5. Franke A, Balschun T, Karlsen TH *et al.* Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. Nat Genet 2008;40:713–5.
6. Fisher SA, Tremelling M, Anderson CA *et al.* Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. Nat Genet 2008;40:710–2.
7. Franke A, Balschun T, Karlsen TH *et al.* Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. Nat Genet 2008;40:1319–23.
8. Silverberg MS, Cho JH, Rioux JD *et al.* Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. Nat Genet 2009;41:216–20.
9. van Diemen CC, Postma DS, Vonk JM *et al.* A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. Am J Respir Crit Care Med 2005;172:329–33.
10. 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.
11. 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.
12. Anderson CA, Massey DC, Barrett JC *et al.* Investigation of Crohn's disease risk loci in ulcerative colitis further defines their molecular relationship. Gastroenterology 2009;136:523–9.