Replication Study of Ulcerative Colitis Risk Loci in a Lithuanian–Latvian Case–Control Sample

Jurgita Skieceviciene, PhD, Gediminas Kiudelis, MD, PhD,† Eva Ellinghaus, PhD,‡ Tobias Balschun, PhD,‡ Laimas V. Jonaitis, MD, PhD,† Aida Zvirbliene, MD, PhD,*,† Goda Denapiene, MD, PhD,§ Marcis Leja, MD, PhD,¶ Gitana Pranculiene, MD,‖ Vytenis Kalibatas, MD, PhD,** Hamidreza Saadati, MSc,‡ David Ellinghaus, PhD,‡ Vibeke Andersen, MD, PhD,††,‡‡ Jonas Valantinas, MD, PhD,§ Algimantas Irnius, MD, PhD,§ Aleksejs Derovs, MD,§§ Algimantas Tamelis, MD, PhD,¶¶ Stefan Schreiber, MD, PhD,‡ Limas Kupcinskas, MD, PhD,*,† and Andre Franke, PhD‡*

Background: Differences between populations might be reflected in their different genetic risk maps to complex diseases, for example, inflammatory bowel disease. We here investigated the role of known inflammatory bowel disease–associated single nucleotide polymorphisms (SNPs) in a subset of patients with ulcerative colitis (UC) from the Northeastern European countries Lithuania and Latvia and evaluated possible epistatic interactions between these genetic variants.

Methods: We investigated 77 SNPs derived from 5 previously published genome-wide association studies for Crohn's disease and UC. Our study panel comprised 444 Lithuanian and Latvian patients with UC and 1154 healthy controls. Single marker case–control association and SNP-SNP epistasis analyses were performed.

Results: We found 14 SNPs tagging 9 loci, including 21q21.1, *NKX2-3*, *MST1*, the HLA region, 1p36.13, *IL10*, *JAK2*, *ORMDL3*, and *IL23R*, to be associated with UC. Interestingly, the association of UC with previously identified variants in the HLA region was not the strongest association in our study ($P = 4.34 \times 10^{-3}$, odds ratio [OR] = 1.25), which is in contrast to all previously published studies. No association with any disease subphenotype was found. SNP-SNP interaction analysis showed significant epistasis between SNPs in the *PTPN22* (rs2476601) and *C13orf31* (rs3764147) genes and increased risk for UC ($P = 1.64 \times 10^{-6}$, OR = 2.44). The association has been confirmed in the Danish study group ($P = 0.04$, OR = 3.25).

Conclusions: We confirmed the association of the 9 loci (21q21.1, 1p36.13, *NKX2-3*, *MST1*, the HLA region, *IL10*, *JAK2*, *ORMDL3*, and *IL23R*) with UC in the Lithuanian–Latvian population. SNP-SNP interaction analyses showed that the combination of SNPs in the *PTPN22* (rs2476601) and *C13orf31* (rs3764147) genes increase the risk for UC.

(Inflamm Bowel Dis 2013;19:2349–2355)

Key Words: Lithuanian–Latvian, ulcerative colitis, single nucleotide polymorphisms, case–control

Ulcerative colitis (UC) is a chronic, relapsing inflammatory condition of the colon. It represents 1 of the 2 main subforms of inflammatory bowel disease (IBD), the other being Crohn's disease (CD). IBD is a major burden for health systems in Western countries, with prevalence rates in North America and Europe ranging from 6 to 243 cases per 100,000 inhabitants for UC and of 3.6 to 214 cases per 100,000 inhabitants for CD.[1] The very limited data from the Northeastern European countries demonstrate low incidence rates of IBD:

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.ibdjournal.org).

Received for publication June 18, 2013; Accepted July 7, 2013.

From the *Institute for Digestive Research and †Department of Gastroenterology, Academy of Medicine, Lithuanian University of Health Sciences, Kaunas, Lithuania; ‡Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany; §Center of Hepatology, Gastroenterology and Dietetics, Vilnius University, Vilnius, Lithuania; ¶Digestive Diseases Centre GASTRO, Riga, Latvia; Departments of ‖Children Diseases and **Health Management, Academy of Medicine, Lithuanian University of Health Sciences, Kaunas, Lithuania; ††Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark; ‡‡Medical Department, Hospital of Southern Jutland, Aabenraa, Denmark; §§Internal Disease Department, Riga Stradins University, Riga, Latvia; and ¶¶Department of Surgery, Academy of Medicine, Lithuanian University of Health Science, Kaunas, Lithuania.

Supported by the German Ministry of Education and Research through the National Genome Research Network (NGFN), the PopGen Biobank, and received infrastructure support through the Deutsche Forschungsgemeinschaft (DFG) Cluster of Excellence "Inflammation at Interfaces"; grant of the Research Council of Lithuania (MIP-078/2011) and the European Social Fund (009/0220/1DP/1.1.1.2.0/09/APIA/VIAA/016).

The authors have no conflicts of interest to disclose.

Limas Kupcinskas and Andre Franke contributed equally to this work.

Reprints: Jurgita Skieceviciene, PhD, Institute for Digestive Research, Academy of Medicine, Lithuanian University of Health Sciences, A. Mickeviciaus Street 9, Kaunas LT-44307, Lithuania (e-mail: j.sventoraityte@gmail.com).

DOI 10.1097/MIB.0b013e3182a3eaeb

Published online 22 August 2013.

Lithuania, 2.6 CD cases and 6.5 UC cases per 100,000 inhabitants; Estonia, 5.4 CD cases and 5.7 UC cases per 100,000 inhabitants.[2] Although the precise etiology of IBD is unknown, current data support the hypothesis of a complex interplay between genetic and environmental factors, the latter being subject of drastic changes during the last century.[3]

The past decade has witnessed remarkable success in the identification of low-penetrance high-frequency susceptibility variants in IBD. It has been estimated that the so far identified risk loci explain approximately 13.6% disease variance in CD and 7.5% disease variance in UC[4] indicating that a large part of the genetic risk is still unaccounted for. One of the possible reasons is that part of the genetic risk in complex human diseases may be due to the poorly understood systematic epistatic interactions of genetic variants.[5,6] The inherited combinations of functional and/or disease-linked common single nucleotide polymorphisms (SNPs) may additively or synergistically shape specific biological processes that increase disease susceptibility.[7] Therefore, the effect might be missed if a gene that functions primarily through a complex mechanism is examined separately without regard to its potential interactions with other genes and/or environmental factors.[6] Interactions between SNPs in common human diseases (including IBD[8–11]) have been investigated in a hypothesis-free way using genome-wide association studies (GWAS) data sets. However, only a few well-replicated instances have been reported so far.[12–14]

Thus, we aimed to further dissect the genetic background of UC by confirming and expanding the findings of 5 previously published IBD GWAS in a UC cohort from Lithuania and Latvia. In total, 77 SNPs were genotyped in 447 UC cases and 1154 healthy controls. The selected SNP set partially overlaps with the hits reported in the 2 recent UC meta-analyses (10 overlapping loci with McGovern et al study[8]; 23 loci with Anderson et al study[15]).

MATERIALS AND METHODS

Subjects

The study included 447 unrelated patients with UC (average age ± SD = 44.4 ± 16.5) and 1154 healthy controls (average age ± SD = 40.2 ± 12.7) of Caucasian ethnicity. The recruitment of study individuals was performed in 7 hospitals from Lithuania (Hospital of Lithuanian University of Health Sciences Kaunas Clinics, Vilnius University Hospital Santariskiu Clinics, M. Marcinkevicius Hospital in Vilnius, Klaipeda University Hospital, Klaipeda Seamen's Hospital, Republican Panevezys Hospital, and Republican Siauliai Hospital) and 3 Latvian hospitals in Riga (P. Stradina Clinical University Hospital, Latvian Maritime Medicine Center in Riga, and Riga Eastern Clinical University Hospital), during the years 2003 through 2009. The Lithuanian control individuals were recruited from the National Blood Center, whereas recruitment of the Latvian healthy individuals was performed at the participating clinical centers from Latvia. Although blood bank donors are a selected group compared with population controls, they are regularly and confidently being used in genetic association studies.[10] Principal component plot also indicated a tight overlap between cases and controls (Fig., Supplemental Digital Content 1, http://links.lww.com/IBD/A255). The diagnosis of UC was based on standard clinical, endoscopic, radiological, and histological criteria.[16] A subgroup of patients with UC and controls used in this study has been used and characterized in previous studies.[15,17] The full demographic and phenotypic descriptions of both study groups are summarized in Table, Supplemental Digital Content 2, http://links.lww.com/IBD/A256. The clinical characteristics provided in the table are given according to the Montreal classification.[18] All phenotypic data were collected blind to the results of the genotypic data. Written informed consent from all participants and approval of the ethics committees was obtained.

For replication of the SNP-SNP interaction between *PTPN22* (rs2476601) and *C13orf31* (rs3764147), Danish and German study groups were tested for an association. The Danish study population consisted of 124 patients with UC and 92 healthy blood donors of Caucasian ethnicity overlapping with samples that have been used in a previous study.[19] The German study panel comprised 892 patients with UC and 1901 controls of Caucasian ethnicity. This panel has been used and described in previous publications.[20,21]

Genotyping and SNP Selection

SNPs from 5 original CD or UC GWAS (CD: Franke et al,[22] WTCCC/Parkes et al,[10,23] and Barrett et al[24]; UC: Franke et al[20] and Silverberg et al[25]), along with SNPs in previously IBD-associated genes (*NOD2*,[26,27] *DLG5*,[28] *IL23R*,[29] *SLC22A4* and *SLC22A5*,[30] *TNFSF15*,[31] *PTGER4*,[32] *MST1*,[33] and *ATG16L1*[34]), were selected for genotyping. Three SNP genotyping assays (rs11805303,[20] rs287094616,[25] and rs10801047[10,23]) failed to design; therefore, the final genotyping panel included 77 SNPs from 46 different loci (Table, Supplemental Digital Content 3, http://links.lww.com/IBD/A257). Of these 46 loci, 36 have meanwhile been reported as genome-wide significant in the most recent meta-analyses (marked bold in the Table, Supplemental Digital Content 3, http://links.lww.com/IBD/A257).[8,15]

Genotyping was performed at the Institute for Clinical Molecular Biology (Christian-Albrechts University, Kiel, Germany) by ligation-based SNPlex and TaqMan technologies from Life Technologies (formerly Applied Biosystems, Foster City, CA) according to the manufacturer's recommendations. Genotype assignments were manually confirmed by visual inspection with the Genemapper 4.0 compatible with the SNPlex system and SDS 2.3 software compatible with the TaqMan system.

Statistical Analysis

The combined screening panel of 447 cases and 1154 controls had 80% power to detect an odds ratio (OR) of ≥1.44 at the 5% significance level, assuming a frequency of the disease-associated allele of ≥20% in the controls (Fig., Supplemental Digital Content 4, http://links.lww.com/IBD/A258). Quality assessments and statistical analysis of the genotyping data were performed using PLINK software version 1.07.[35] Individuals with

>10% missing genotypes and SNPs with a call rate <90% or deviation from Hardy–Weinberg equilibrium in the controls ($P < 0.05$) were excluded from further analysis. In total, 99.6% of all cases and controls were successfully genotyped. The average genotyping rate across all samples was 98%. One SNP (rs2289310) failed genotyping, resulting in a total of 76 SNPs that were analyzed in 444 UC cases and 1154 controls.

Differences in allele frequencies between cases and controls were calculated in the combined Lithuanian and Latvian study sample, using the Cochrane–Mantel–Haenszel test together with the Breslow–Day test for heterogeneity of ORs. Six of the 76 SNPs (rs2925757, rs10077785, rs2076756, rs7868736, rs10974944, and rs2066847) showed heterogeneity of ORs between the Lithuanian and Latvian UC study groups ($P_{BD} < 0.05$) and were therefore excluded from further analyses. Nevertheless, the recent study by Nelis et al[36] showed that the 3 Baltic countries (Lithuania, Latvia, and Estonia), Poland, and Western Russia together form a genetic cluster (inflation factor $\lambda = 1.23$), thereby indicating that our 2 study populations can be combined in an association analysis without risking too many false-negative and false-positive results due to population stratification. In the case–control association analysis, two significance criteria were applied: (1) replication of the SNPs that were associated with UC at genome-wide significance ($P < 5 \times 10^{-8}$) in previous studies was defined as an association at a level of $P < 0.05$ with the same risk allele identified in the index studies; (2) new association to UC—a case–control association at a level of $P < 5 \times 10^{-8}$ was required for a SNP to be considered as a new UC locus.

Subphenotypes of UC (disease localization [defined macroscopically as extensive, left-sided, or proctitis only] and extraintestinal manifestations), disease modifiers (age at diagnosis and family history of disease), and disease outcome measures (surgery and treatment using biological therapy) were inspected in within-case analyses. The χ^2 test was used to detect association between each binary phenotype and the genotyped SNPs. P values were adjusted for multiple comparisons based on the Bonferroni procedure (correction was applied for the number of complementary subgroups of patients).

SNP-SNP epistasis for our case–control population-based sample was calculated using the –fast-epistasis command as implemented in PLINK. The latter provides a logistic regression test for interaction that assumes an allelic model for both the main effects and the interactions. This test is based on a Z-score for the difference in SNP-SNP association (OR) between cases and controls and follows a standard normal distribution under the multiplicative model of no interaction. All pairwise combinations of SNPs were tested. Again, P values were adjusted for multiple testing by Bonferroni correction (2404 independent tests, corrected $P = 2 \times 10^{-5}$).

RESULTS

Allelic Association Analysis in UC

Combined analysis of the 2 study groups revealed 14 SNPs from 9 independent loci (21q21.1, 1p36.13, *NKX2-3*, *MST1*, the HLA region, *ORMDL3*, *IL10*, *JAK2*, and *IL23R*) to be associated with UC (Table 1 and Table, Supplemental Digital Content 5, http://links.lww.com/IBD/A259). All identified allelic association signals were in the same direction as reported in previous IBD studies.[10,20,22–25] All 9 loci have been reported as genome-wide significant, by the recent UC meta-analyses.[15]

The statistical power of our study to detect a disease association (assuming the same effect size as documented in original studies) is given in Table 1 for each SNP. Estimated power varied widely between SNPs, with the highest power calculated for the 2 known IBD variants in *NKX2-3* and *BSN*; for the known UC loci, the highest calculated power had HLA and *IL10*. For 21 SNPs, we had a power of >80% out of which 6 SNPs were successfully replicated (Table 1 and Table, Supplemental Digital Content 5, http://links.lww.com/IBD/A259).

Genetic Association with UC Subphenotypes

We tested all UC-associated SNPs with phenotypic characteristics such as age of onset, gender, disease localization, extraintestinal manifestations, family history of IBD, surgery treatment, and treatment using biological therapy. No significant associations were found for the subphenotypes under study following correction for multiple testing.

SNP-SNP Interaction Analysis

SNP-SNP interaction, known as epistasis, has been investigated among all candidate SNPs that passed quality control, using a logistic regression test (Table, Supplemental Digital Content 6, http://links.lww.com/IBD/A260). We found a significant association that withstood Bonferroni correction between rs2476601 and rs3764147 ($P = 1.64 \times 10^{-6}$, OR = 2.44) under the assumption of a nonadditive genetic model. The interacting SNPs were located in genes *PTPN22* (rs2476601) and *C13orf31* (rs3764147), respectively (for the interaction pattern of this SNP pair, see Table 2).

The 2 SNPs had a minor allele frequency of 14.3% (rs2476601, allele A) and 27.5% (rs3764147, allele G) in the control group, and the affected individuals showed an excess of genotype pairs AA–AG, GA–GG, and GA–AG. Risks, relative to the most common homozygous genotype GG–AA, are reported in Table 2. For genotypes GA–GG and GA–AG, the relative risks were significantly larger than 1, with OR = 2.24 (95% CI: 1.09–4.64, $P = 0.03$) and OR = 1.48 (95% CI: 1.02–2.16, $P = 0.04$), respectively. The combined OR for the 2 genotypes (GA–GG and GA–AG) was 1.60 (95% CI: 1.13–2.26, $P = 8.13 \times 10^{-3}$). Although genotype AA–AG did not reach the level of significance ($P = 0.15$, OR = 2.10, 95% CI: 0.75–5.87), possibly due to its low frequency, the OR was still larger than 1, indicating an increased risk. The joint OR of the 3 risk genotypes was 1.63 (95% CI: 1.16–2.29, $P = 4.32 \times 10^{-3}$), demonstrating that the 3 specific allele combinations of the SNPs rs2476601 and rs3764147 confer an elevated risk for UC in the Lithuanian–Latvian sample set.

The evidence for replication of the interaction was sought in the German (892 UC cases, 1901 controls) and Danish (124 UC cases, 92 controls) study panels. The 2 SNPs had a minor allele

TABLE 1. Loci Showing Significant Association With UC in a Case–Control Analysis

Position (bp)	dbSNP ID	Gene of Interest	A1 A2	AF_{A1} Control Case	P_{BD}	P_{CMH}	OR (95% CI)	Power
1p36.13 (20,142,866)	rs3806308[25]	*RNF18*	T C	0.45 0.38	0.47	2.40×10^{-4}	0.74 (0.63–0.87)	84.93%
1p36.13 (20,171,860)	rs6426833[25]	—	A G	0.49 0.54	0.85	6.01×10^{-3}	1.25 (1.07–1.46)	92.48%
1p31.3 (67,702,526)	rs11465804[24,25,29]	*IL23R*	G T	0.05 0.03	0.49	0.012	0.58 (0.38–0.89)	51.74%
1p31.3 (67,705,958)	rs11209026[25,29]	*IL23R*	A G	0.07 0.04	0.51	2.16×10^{-3}	0.55 (0.38–0.81)	63.14%
1q32.1 (206,939,904)	rs3024505[20]	*IL10*	A G	0.13 0.17	0.17	1.04×10^{-3}	1.43 (1.16–1.77)	80.83%
3p21.31 (49,721,532)	rs3197999[20,33]	*MST1*	T C	0.23 0.28	0.65	3.21×10^{-3}	1.32 (1.10–1.58)	50.56%
6p21.32 (32,363,844)	rs9268480[20]	*BTNL2*	T C	0.21 0.17	0.17	0.025	0.79 (0.65–0.97)	47.05%
6p21.32 (32,429,758)	rs9268858[20]	*HLA-DRA*	C T	0.24 0.19	0.49	0.011	0.78 (0.64–0.94)	91.93%
6p21.32 (32,431,147)	rs9268877[20]	*HLA-DRA*	T C	0.51 0.56	0.98	4.34×10^{-3}	1.25 (1.07–1.47)	96.13%
6p21.32 (32,433,167)	rs2395185[25]	*HLA-DRA*	T G	0.24 0.19	0.41	6.44×10^{-3}	0.76 (0.63–0.93)	76.01%
9p24.1 (4,981,602)	rs10758669[24]	*JAK*	C A	0.36 0.43	0.09	8.08×10^{-5}	1.38 (1.17–1.62)	16.71%
10q24.2 (101,291,593)	rs11190140[10,23,24]	*NKX2-3*	T C	0.46 0.52	1.00	7.27×10^{-3}	1.25 (1.06–1.48)	82.85%
17q12 (38,040,763)	rs2872507[24]	*ORMDL3*	A G	0.41 0.49	0.56	1.24×10^{-4}	1.36 (1.16–1.59)	17.42%
21q21.1 (16,805,220)	rs1736135[24]	—	C T	0.43 0.34	0.89	8.01×10^{-6}	0.69 (0.59–0.81)	29.68%

Detailed results for all 70 SNPs can be found in Table, Supplemental Digital Content 5, http://links.lww.com/IBD/A259. Nucleotide positions refer to NCBI build 37. A1 denotes the minor allele and A2 is the common allele. The respective allele frequencies are shown for allele A1 (AF_{A1}). *P* values were obtained from the Breslow–Day test (P_{BD}) and the Cochran–Mantel–Haenzsel test (P_{CMH}). Odds ratios and 95% confidence intervals for carriership of allele A1 are shown as OR (95% CI). The power of this study to replicate the association at 0.05 significance level is shown (AF and OR presented in the original studies were used for calculations).

frequency of 11.1% and 13% (rs2476601, allele A) and 25.1% and 22.8% (rs3764147, allele G) in the German and Danish control groups, respectively. The *PTPN22* and *C13orf31* SNP pair was confirmed to increase UC risk in the Danish study group ($P = 0.04$, OR = 3.25, Table, Supplemental Digital Content 7, http://links.lww.com/IBD/A261), whereas in the German study group, no association was found ($P = 0.45$, OR = 0.89, Table, Supplemental Digital Content 8, http://links.lww.com/IBD/A262). However, possibly due to the small sample size, none of the genotype pairs was overrepresented in the Danish UC group (Table, Supplemental Digital Content 7, http://links.lww.com/IBD/A261). Although the tendency for an increase in frequency of genotype pairs GA–AG and GG–GG in UC group could be observed. The risk-increasing genotype pair GA–GG was not detected in the Danish control group but was present in 1.1% of the UC cases, whereas AA–AG was not determined neither in the cases nor in the controls.

DISCUSSION

In the past 5 years, GWAS in CD and UC have identified many novel susceptibility loci. Our study represents the first comprehensive analysis of the contribution of previously reported genetic risk factors[10,20,22–25] to UC susceptibility in a low-incidence population of Northeastern Europe—Lithuania and Latvia.[2] The collaboration between these 2 small Northeastern European countries (combined study panel of 1601 individuals) was initiated to form a large IBD study sample with enough study power for the

TABLE 2. Genotype Counts for the SNP Pair (rs2476601, rs3764147) in Patients with UC and the Calculated ORs Relative to the Most Common Double Homozygous Genotype (rs3764147 = AA, rs2476601 = GG)

		rs3764147 (*C13orf31*)		
	rs2476601 (*PTPN22*)	AA	AG	GG
Controls	GG	407 (36.3%)	357 (31.9%)	63 (5.6%)
	GA	161 (14.4%)	92 (8.2%)	16 (1.4%)
	AA	17 (1.5%)	8 (0.7%)	0
Ulcerative colitis	GG	170 (39%)	119 (27.3%)	19 (4.4%)
	GA	45 (10.3%)	57 (13.1%)	15 (3.4%)
	AA	3 (0.7%)	7 (1.6%)	1 (0.2%)
OR (95% CI)	GG	1	0.80 (0.61–1.05)	0.72 (0.42–1.24)
	GA	**0.67 (0.46–0.97)**	**1.48 (1.02–2.16)**	**2.24 (1.09–4.64)**
	AA	0.42 (0.12–1.46)	2.10 (0.75–5.87)	NA

ORs for the genotypes reaching the level of significance ($P < 0.05$) are presented in bold.

analysis of moderate effect size susceptibility variants. A recent study on the genetic differences in European populations, investigating the detailed structure of the Baltic countries and other Northeastern European populations, showed that the 3 Baltic countries (Lithuania, Latvia, and Estonia), Poland, and Western Russia together form a genetic cluster, and hence negligible population structure exists, which proves that our 2 study populations can be combined in association analyses.[36]

In our Lithuanian–Latvian case–control study of 444 UC cases and 1154 controls, we found 9 robust genetic risk loci to be associated with UC. The allele frequencies and contributed risks of the respective SNPs in the risk loci were similar to previous reports in other Caucasian populations.[9,20,25,37–42] The reported risk loci are implemented in the pathogenesis of other immune-related diseases as well, e.g., *ORMDL3* has been previously shown to be associated with asthma, type 1 diabetes, psoriasis, CD and *IL10* with type 1 diabetes, CD, systemic lupus erythematosus, Behçet's disease. Despite of the long known and very strong association between UC and the HLA region, and the sufficient power of our study for the respective SNPs, the HLA variants were not between the strongest associated SNPs in our Lithuanian–Latvian cohort. However, a significant difference of effect between our study and previously published reports could not be detected as the 95% confidence intervals (CIs) of the ORs overlap (our study: OR = 1.25 and 95% CI = 1.07–1.47 versus Franke et al[20]: OR = 1.45 and 95% CI = 1.33–1.58). Finally, several previously reported risk loci, including *STAT3*, *IL12B*, *PTPN2*, *NELL1*, and *ARP2C* were not replicated in our study, which of course does not necessarily mean that these are not UC-associated genes in the Baltic population but may only reflect a lack of statistical power or confounding factors such as different environmental factors.

A number of studies have demonstrated the presence of gene–gene interactions in complex human diseases, including IBD. The epistasis analyses in IBD have been performed between pathway-related genes,[43,44] genes that were individually associated with IBD,[45,46] and in a hypothesis-free way using GWAS and GWAS meta-analyses data.[8–11] However, none of the findings were replicated in the independent sample sets. In the frames of this study, we investigated the impact of all possible studied SNP pairs, including those that were not associated with UC in our single marker case–control analysis. We were able to reveal a statistically significant interaction between 2 SNPs (rs2476601 and rs3764147) that were not associated with UC in the single marker association analysis. The association was only confirmed in the small Danish study group but not in the large German cohort. This indicates a possible role of population differences or of a false-positive association due to the small study and replication panel size. Therefore, further replication in larger future studies is warranted before this finding can be considered confirmed. The interacting SNPs are located in the genes *PTPN22* (rs2476601) and *C13orf31* (rs3764147) and are both coding mutations (rs2476601: synonymous; rs3764147: missense). Both interacting regions have been related to the development of autoimmune diseases (*PTPN22*: CD,[24] type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus, Graves' disease, autoimmune thyroid disease, alopecia areata, juvenile idiopathic arthritis, and Hashimoto's thyroiditis[47]; *C13orf31*: CD[24] and leprosy[48]).

PTPN22 encodes a lymphoid-specific protein tyrosine phosphatase (LYP), a member of a family of proteins involved in suppressing spontaneous T-cell activation.[49] *PTPN22* is expressed in many hematopoietic cell types, notably T cells. The rs2476601 risk allele is a gain-of-function mutation that results in a higher catalytic activity of the phosphatase and more potent negative regulation of T-cell activation.[50,51] By contrast, knockout mice (*Lyp* is the mouse ortholog of *PTPN22*) have an increased T-cell activation in combination with an increased production of antibodies.[52] The biological functions of the *C13orf31* gene product are not known. However,

because *C13orf31* polymorphisms were previously found to be associated with CD and leprosy, it has been suggested that this locus might be involved in mycobacteria clearance.[48]

To further investigate the possible interaction between *PTPN22* and *C13orf31*, we performed an in silico pathway analysis ("*In silico* prediction analysis," Supplemental Digital Content 9, http://links.lww.com/IBD/A263), which indicated that both genes could be connected through coexpression and protein–protein interaction (i.e., physical interaction) pathways (Fig., Supplemental Digital Content 10, http://links.lww.com/IBD/A263). The most significant molecular processes predicted by the program were the regulation of positive T-cell activation and segregation of the TCR complex, which mainly affect immune-mediated regulatory processes and cell activation (lymphocyte, T cell). These processes are mainly linked to the query gene *PTPN22*.

Taken together, our study results support the previously proposed functional implications of the genetic associations in the pathogenesis of UC, i.e., the importance of gene sets that influence barrier function, transcriptional regulation, cell-specific innate responses, and regulate adaptive immunity. Although we did not find an association between IBD variants and phenotypic characteristics, including age of onset, gender, and family history of IBD, we could show that SNPs that were not associated with disease in a single marker analysis contribute to the overall disease risk by epistatic effects. However, functional and further replication studies are needed to confirm the potential protein–protein interactions because statistically significant SNP-SNP epistasis does not necessarily mirror biological interactions. Finally, the contribution of the HLA region to genetic susceptibility to UC will require more comprehensive analyses in a larger cohort from Northeastern Europe by using a dense HLA marker set.

ACKNOWLEDGMENTS

The authors wish to thank all patients, families, and physicians for their cooperation. The authors also thank the technical staff at the Institute for Clinical Molecular Biology (Kiel, Germany) for their expert help. The authors gratefully acknowledge following gastroenterologists V. Svalbonas, R. Kucinskiene, D. Krukas, G. Simulionis, Z. Sukys, A. Alisauskas, I. Vilcinskaite, L. Panina, and J. Derova for their cooperation and support during this study.

REFERENCES

1. Loftus EV. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterology*. 2004; 126:1504–1517.
2. Burisch J, Pedersen N, Cukovic-Cavka S, et al. East-West gradient in the incidence of inflammatory bowel disease in Europe: the ECCO-EpiCom inception cohort. *Gut*. [published online ahead of print April 20, 2013]. doi: 10.1136/gutjnl-2013-304636.
3. Cooney R, Jewell D. The genetic basis of inflammatory bowel disease. *Dig Dis*. 2009;27:428–442.
4. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491:119–124.
5. Moore JH. A global view of epistasis. *Nat Genet.* 2005;37:13–14.
6. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet.* 2009;10:392–404.
7. Moore JH. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum Hered.* 2003;56:73–82.
8. Franke A, McGovern DPB, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet.* 2010;42:1118–1125.
9. McGovern DPB, Gardet A, Törkvist L, et al. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat Genet.* 2010; 42:332–337.
10. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661–678.
11. Barrett JC, Lee JC, Lees CW, et al. Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat Genet.* 2009;41:1330–1334.
12. Evans DM, Spencer CC, Pointon JJ, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat Genet.* 2011;43:761–767.
13. Strange A, Capon F, Spencer CC, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat Genet.* 2010;42:985–990.
14. Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet.* 2009;41:703–707.
15. Anderson CA, Boucher G, Lees CW, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet.* 2011;43:246–252.
16. Podolsky DK. Inflammatory bowel disease (1). *N Engl J Med.* 1991;325: 928–937.
17. Sventoraityte J, Zvirbliene A, Franke A, et al. NOD2, IL23R and ATG16L1 polymorphisms in Lithuanian patients with inflammatory bowel disease. *World J Gastroenterol.* 2010;16:359–364.
18. Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a working party of the 2005 montreal world congress of gastroenterology. *Can J Gastroenterol.* 2005;19(suppl A):5–36.
19. Andersen V, Ernst A, Sventoraityte J, et al. Assessment of heterogeneity between European Populations: a Baltic and Danish replication case-control study of SNPs from a recent European ulcerative colitis genome wide association study. *BMC Med Genet.* 2011;12:139.
20. 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;40:1319–1323.
21. Franke A, Balschun T, Sina C, et al. Genome-wide association study for ulcerative colitis identifies risk loci at 7q22 and 22q13 (IL17REL). *Nat Genet.* 2010;42:292–294.
22. Franke A, Hampe J, Rosenstiel P, et al. Systematic association mapping identifies NELL1 as a novel IBD disease gene. *PLoS One.* 2007;2:e691.
23. 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;39:830–832.
24. 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;40:955–962.
25. 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;41:216–220.
26. Ogura Y, Bonen DK, Inohara N, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*. 2001;411:603–606.
27. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*. 2001; 411:599–603.
28. Stoll M, Corneliussen B, Costello CM, et al. Genetic variation in DLG5 is associated with inflammatory bowel disease. *Nat Genet.* 2004;36:476–480.
29. Duerr RH, Taylor KD, Brant SR, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science*. 2006; 314:1461–1463.
30. Peltekova VD, Wintle RF, Rubin LA, et al. Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat Genet.* 2004;36:471–475.

31. Yamazaki K, McGovern D, Ragoussis J, et al. Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum Mol Genet.* 2005;14:3499–3506.
32. 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.
33. Raelson J V, Little RD, Ruether A, et al. Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. *Proc Natl Acad Sci U S A.* 2007;104:14747–14752.
34. 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;39:207–211.
35. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–575.
36. Nelis M, Esko T, Mägi R, et al. Genetic structure of Europeans: a view from the North-East. *PLoS One.* 2009;4:e5472.
37. 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;40:710–712.
38. 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;40:713–715.
39. Törkvist L, Halfvarson J, Ong RTH, et al. Analysis of 39 Crohn's disease risk loci in Swedish inflammatory bowel disease patients. *Inflamm Bowel Dis.* 2010;16:907–909.
40. Wang K, Baldassano R, Zhang H, et al. Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects. *Hum Mol Genet.* 2010;19:2059–2067.
41. Festen EAM, Stokkers PCF, van Diemen CC, et al. Genetic analysis in a Dutch study sample identifies more ulcerative colitis susceptibility loci and shows their additive role in disease risk. *Am J Gastroenterol.* 2010;105:395–402.
42. Andersen V, Ernst A, Christensen J, et al. The polymorphism rs3024505 proximal to IL-10 is associated with risk of ulcerative colitis and Crohns disease in a Danish case-control study. *BMC Med Genet.* 2010;11:82.
43. Glas J, Seiderer J, Wagner J, et al. Analysis of IL12B gene variants in inflammatory bowel disease. *PLoS One.* 2012;7:e34349.
44. Torok HP, Glas J, Endres I, et al. Epistasis between Toll-like receptor-9 polymorphisms and variants in NOD2 and IL23R modulates susceptibility to Crohn's disease. *Am J Gastroenterol.* 2009;104:1723–1733.
45. Glas J, Seiderer J, Wetzke M, et al. rs1004819 is the main disease-associated IL23R variant in German Crohn's disease patients: combined analysis of IL23R, CARD15, and OCTN1/2 variants. *PLoS One.* 2007;2:e819.
46. Glas J, Konrad A, Schmechel S, et al. The ATG16L1 gene variants rs2241879 and rs2241880 (T300A) are strongly associated with susceptibility to Crohn's disease in the German population. *Am J Gastroenterol.* 2008;103:682–691.
47. Lee YH, Rho YH, Choi SJ, et al. The PTPN22 C1858T functional polymorphism and autoimmune diseases—a meta-analysis. *Rheumatology (Oxford).* 2007;46:49–56.
48. Zhang FR, Huang W, Chen SM, et al. Genomewide association study of leprosy. *N Engl J Med.* 2009;361:2609–2618.
49. Stanford SM, Mustelin TM, Bottini N. Lymphoid tyrosine phosphatase and autoimmunity: human genetics rediscovers tyrosine phosphatases. *Semin Immunopathol.* 2010;32:127–136.
50. Rieck M, Arechiga A, Onengut-Gumuscu S, et al. Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes. *J Immunol.* 2007;179:4704–4710.
51. Vang T, Congia M, Macis MD, et al. Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. *Nat Genet.* 2005;37:1317–1319.
52. Hasegawa K, Martin F, Huang G, et al. PEST domain-enriched tyrosine phosphatase (PEP) regulation of effector/memory T cells. *Science.* 2004;303:685–689.