BASIC—ALIMENTARY TRACT

Genome-Wide Association Analysis in Sarcoidosis and Crohn's Disease Unravels a Common Susceptibility Locus on 10p12.2

ANDRE FRANKE,* ANNEGRET FISCHER,* MICHAEL NOTHNAGEL,† CHRISTIAN BECKER,§,‖ NILS GRABE,* ANDREAS TILL,* TIM LU,† JOACHIM MÜLLER–QUERNHEIM,¶ MICHAEL WITTIG,* ALEXANDER HERMANN,* TOBIAS BALSCHUN,* SYLVIA HOFMANN,* REGINA NIEMIEC,§,‖ SABRINA SCHULZ,§,‖ JOCHEN HAMPE,# SUSANNA NIKOLAUS,** PETER NÜRNBERG,§ MICHAEL KRAWCZAK,† MANFRED SCHÜRMANN,†† PHILIP ROSENSTIEL,* ALMUT NEBEL,* and STEFAN SCHREIBER*

*Institute for Clinical Molecular Biology and †Institute of Medical Informatics and Statistics, Christian-Albrechts-University, Kiel, Germany; §Cologne Center for Genomics, University of Cologne, Germany; ‖RZPD Germany Resource Center for Genome Research GmbH, Berlin, Germany; ¶Department of Pneumology, University of Freiburg, Germany; #Department of General Internal Medicine, University Hospital Schleswig-Holstein, Christian-Albrechts-University, Kiel, Germany; **Popgen Biobank, Christian-Albrechts University, Kiel, Germany; ††Institute of Human Genetics, University of Lübeck, Germany

Background & Aims: **Crohn's disease (CD) and sarcoidosis (SA) are chronic inflammatory barrier diseases that share several clinical and immunological features, including the occurrence of granulomas.** ***Methods:*** **A 100k genome-wide association study with 83,360 single-nucleotide polymorphisms (SNPs) was performed on 382 CD patients, 398 SA patients, and 394 control individuals. The 24 SNPs that were most strongly associated in the combined CD/SA phenotype were selected for verification in an independent sample of 1,317 patients (660 CD and 657 SA) and 1,091 controls.** ***Results:*** **The most significant association (Bonferroni corrected $P = .036$) was obtained at SNP rs1398024 on chromosome 10p12.2, with an odds ratio (OR) for both diseases of 0.81 (95% confidence interval [CI], 0.69–0.96) for carriership of the rarer allele A. The P value in the overall combined sample was 4.24×10^{-6}. During further follow-up, a moderate association (OR, 0.83; 95% CI, 0.72–0.96; $P = .015$) was observed between rs1398024 and ulcerative colitis (1,080 patients vs 1,091 controls), the second main subphenotype of inflammatory bowel disease in addition to CD. Extensive fine mapping of the 10p12.2 locus points to yet unidentified variants in the *C10ORF67* gene region as the most likely underlying risk factors.** ***Conclusion:*** **Our study demonstrates that the combined analysis of different, albeit clinically related, phenotypes can lead to the identification of common susceptibility loci.**

The overlapping clinical presentation of different chronic inflammatory disorders is often observed. For example, chronic inflammatory disorders of the lung show coincidences with Crohn's disease (CD) that are higher than expected by chance alone.[1,2] In the genetic etiology, this is reflected by common susceptibility genes (eg, *NOD2* for CD[3–5] and asthma,[6] or *IL23R* for CD[7] and psoriasis[8]). The clinical overlap is demonstrated, for example, by the feature of granuloma formation as a hallmark of pathophysiology shared between CD and sarcoidosis (SA). We therefore tested the hypothesis that a combined systematic genetic study of these 2 diseases may disregard disease-specific genetic causations and directly identify genetic factors that play a role in both diseases.

CD (MIM 266600), 1 of the 2 main subphenotypes of inflammatory bowel disease (IBD; MIM 601458), and SA (MIM 181000) are characterized by noncaseating granulomas, chronic mucosal inflammation including systemic immune activation, and manifestation in young adolescents.[9] Although the primary affection sites are different, the 2 diseases can manifest in almost all organs. For CD, 3 susceptibility loci have been identified by positional cloning, namely *NOD2*, 5q31, and *DLG5* (for review, see Schreiber et al[10]). Most recently, a number of additional susceptibility loci for CD have been detected in genome-wide association studies (GWAS; for review, see Mathew[11]). In SA, a genetic component is best supported by established associations with HLA-DR alleles and the non-HLA susceptibility gene *BTNL2*.[12–14] Additional immune genes have been suggested by candidate gene case-control studies, such as the chemokine receptors *CCR2* and *CCR5*, as well as the cytokine *TNF-α* (for review see Iannuzzi and Rybicki[15]).

Complex diseases are caused by an unprecedented number of disease genes as illustrated by the example of CD.

Abbreviations used in this paper: BMBF, Bundesministerium für Bildung und Forschung; GWAS, genome-wide association analysis; LD, linkage disequilibrium; OR, odds ratio; SNP, single-nucleotide polymorphism; WTCCC, Wellcome Trust Case Control Consortium.

0016-5085/08/$34.00
doi:10.1053/j.gastro.2008.07.017

Moreover, as mentioned genes underlying 1 complex disease often have etiologic relevance for other, seemingly different, phenotypes. However, the concept to first establish a disease gene in a certain disease (most likely through a low-power experiment like a GWAS) and to retest the finding in another disease will result in a large type 2 error neglecting shared disease genes with smaller impact on the phenotype. Therefore, there are only few instances, for example, a Wellcome Trust Case Control Consortium (WTCCC) study,[16] where previously unknown susceptibility loci could be successfully associated in a group of related disorders in a primary study. In the same study, in which 7 common diseases were investigated by GWAS, CD showed the strongest familial aggregation (relative sibling risk λ_s= 17–35), possibly explaining why the disease yielded the highest number of significant associations in that study. No GWAS has been reported yet for SA.

We have explored genome-wide (100k) single-nucleotide polymorphism (SNP) data from a GWAS on CD (that was also evaluated as a stand-alone experiment)[17] and on SA, which is not published up to now. Both studies were of similar size and conducted in parallel using Affymetrix SNP arrays with the predecided plan to conduct a joint analysis. The assumption was that disease-specific findings would be reduced by the combination of 2 case populations and that any common susceptibility loci, if present, would show up stronger. The joint analysis of these data led to the identification of a novel susceptibility locus on chromosome 10p12.2 that is involved in the etiology of both CD and SA.

Subjects and Methods

Patient Recruitment

German CD patients of panels A, B, C, and D (Table 1) and patients with ulcerative colitis (UC) were recruited at the Charité Universitätsmedizin Berlin (Berlin, Germany) and the Department of General Internal Medicine of the Christian-Albrechts-Universität (Kiel, Germany), with the support of the German Crohn and Colitis Foundation and the Bundesministerium für Bildung und Forschung (BMBF) competence network "IBD." Established clinical, radiologic, and endoscopic (ie, type and distribution of lesions) examinations were required to unequivocally confirm the diagnosis of CD or UC.[18–20] CD panels A and B almost completely overlap with the panels (also termed panel A and B) that were used in a recently published IBD association screen of non-synonymous SNPs.[21]

Table 1. Overview of Samples

Panel	Controls	CD	SA
A—Screening panel	394	382	398
B—Verification panel	1,091	660	657
C—Replication panel	2,270	889	—
D—Fine mapping panel (extension)	—	484	267

NOTE. Samples were organized in panels that corresponded to the successive steps of the study. All individuals were of German descent and all panels were independent from each other, except for the CD patients in panel D, which is a subset of panel C.

The German SA patients (Table 1) were contacted as reported previously.[12] Parts of panels A and D and all patients from panel B have been included in previous studies.[22,23] The diagnosis of SA patients was established on the basis of the International Consensus Statement on Sarcoidosis[24] and partly verified by biopsies.

German control individuals were obtained from the popgen biobank, a resource that has been described[25] and whose universal control samples/data sets were used in several previous genetic studies.[17,21,26–29] All controls were drawn from a population-representative sample. Given the low prevalence of both CD and SA (<0.25%) and the fact that all control individuals self-reported to have neither CD nor SA, control individuals were designated "healthy."

Informed written consent was obtained from all study participants and the collection protocols were approved by the institutional ethics committees of the participating centers. Demographic and disease characteristics of healthy controls, CD, and SA patients are shown in Supplementary Table 1 (see Supplementary material online at www.gastrojournal.org).

SNP Genotyping With the Affymetrix 100k Gene Chip Array

Genotyping of 793 patients (393 CD, 400 SA) and 399 controls was carried out using the Affymetrix GeneChip Human Mapping 50K Xba and Hind Arrays (Santa Clara, CA). Genotypes of 116,164 SNPs were called using the GeneChip DNA Analysis Software (GDAS v2.0, Affymetrix). The applied quality controls, which are described in detail in the Supplementary Methods (see Supplementary material online at www.gastrojournal.org), left 382 CD samples, 398 SA samples, and 394 control samples and 83,360 SNPs for inclusion in screening panel A. Experimental details concerning the genotyping of the 100k SNP set are provided in Matsuzaki et al.[30]

Selection Criteria of SNPs for1 Verification

1. P value $<5 \times 10^{-3}$ in the combined analysis of CD and SA in panel A (standard allelic χ^2 test with 1 degree of freedom).
2. Similar P values obtained from separate analyses of CD and SA in panel A, that is, a difference between the negative decadic logarithms of the individual p values of <1.0. This criterion was deemed necessary to exclude SNPs that were highly significant in only 1 of the 2 GWAS, for example SNP rs2076756 in the *NOD2* gene.
3. Correct assignment of genotypes by GDAS and a clear