nature genetics

A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in *ATG16L1*

Jochen Hampe[1,2,10], Andre Franke[1,10], Philip Rosenstiel[1,9], Andreas Till[1], Markus Teuber[1], Klaus Huse[3], Mario Albrecht[4], Gabriele Mayr[4], Francisco M De La Vega[5], Jason Briggs[5], Simone Günther[5], Natalie J Prescott[6], Clive M Onnie[6], Robert Häsler[1], Bence Sipos[7], Ulrich R Fölsch[2], Thomas Lengauer[4], Matthias Platzer[3], Christopher G Mathew[6], Michael Krawczak[8] & Stefan Schreiber[1,2]

We performed a genome-wide association study of 19,779 nonsynonymous SNPs in 735 individuals with Crohn disease and 368 controls. A total of 7,159 of these SNPs were informative. We followed up on all 72 SNPs with $P \leq 0.01$ with an allele-based disease association test in 380 independent Crohn disease trios, 498 Crohn disease singleton cases and 1,032 controls. Disease association of rs2241880 in the autophagy-related 16-like 1 gene (*ATG16L1*) was replicated in these samples ($P = 4.0 \times 10^{-8}$) and confirmed in a UK case-control sample ($P = 0.0004$). By haplotype and regression analysis, we found that marker rs2241880, a coding SNP (T300A), carries virtually all the disease risk exerted by the *ATG16L1* locus. The *ATG16L1* gene encodes a protein in the autophagosome pathway that processes intracellular bacteria. We found a statistically significant interaction with respect to Crohn disease risk between rs2241880 and the established *CARD15* susceptibility variants ($P = 0.039$). Together with the lack of association between rs2241880 and ulcerative colitis ($P > 0.4$), these data suggest that the underlying biological process may be specific to Crohn disease.

Sequence variations associated with Crohn disease have been reported for several genes, including *CARD15* (also known as *NOD2*; refs. 1,2), *SLC22A4* and *SLC22A5* within the 5q31 haplotype[3,4]; *DLG5* (ref. 5); *ABCB1* (also known as *MDR1*; refs. 6–8); *CARD4* (also known as *NOD1*; ref. 9); *TNFSF15* (ref. 10) and *IL23R* (ref. 11). The most consistent replication and the clearest functional data are available for *CARD15* (refs. 12–16). If the cumulated relative risk exerted by all known Crohn disease susceptibility alleles is compared with the relative sibling risk (λ_s) as estimated in individuals of western European origin (15–35), less than 30% of the sibling recurrence risk seems to be explicable by known mutations.

We genotyped a total of 19,779 coding SNPs in panel A (735 individuals with Crohn disease and 368 controls from northern Germany; **Table 1**) using the SNPlex Genotyping System. Genotyping was successful for 16,360 assays, with 'success' defined as a mean fluorescence reading >500 units on the ABI 3730xl sequencer. Of the successfully genotyped SNPs, 7,159 had a minor allele frequency ≥1% and were included in the subsequent analyses. Next, we evaluated those markers with $P \leq 0.01$ in the allele-based test for disease association ($N = 72$) in panel B (380 German Crohn disease trios, 498 singleton affected individuals and 1,032 independent controls; **Table 2** and **Supplementary Table 1** online). When we used $P \leq 0.05$ as

Table 1 Samples used for association analysis

Panel	Number of affected individuals	Number of controls	Number of trios
Crohn disease (Germany) (panel A)	735	368	–
Crohn disease (Germany) (panel B)	498	1,032	380
Crohn disease (UK) (panel C)	509	656	–
Ulcerative colitis (Germany)	788	1,032[a]	439

The samples are organized into panels that make up successive steps of the study. Index cases from trios were also used in the case-control analyses so that, for example, a total of 878 affected individuals (498 + 380) were available for the case-control comparison in panel B. [a]The controls from Crohn disease panel B were also used for the analysis of ulcerative colitis.

[1]Institute for Clinical Molecular Biology and [2]First Department of Medicine, Christian-Albrechts University Kiel, University Hospital Schleswig-Holstein, 24105 Kiel, Germany. [3]Genome Analysis Group, Leibniz Institute for Age Research, Fritz Lipmann Institute, Beutenbergstrasse 11, 07745 Jena, Germany. [4]Max Planck Institute for Informatics, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany. [5]Applied Biosystems, 850 Lincoln Center Drive, Foster City, California 94404, USA. [6]Department of Medical and Molecular Genetics, King's College London School of Medicine, London SE1 9RT, UK. [7]Department of Pathology and [8]Institute of Medical Informatics and Statistics, Christian-Albrechts University Kiel, University Hospital Schleswig-Holstein, 24105 Kiel, Germany. [9]Max-Planck Institute for Molecular Genetics, Ihnestr. 63, 14195 Berlin, Germany. [10]These authors contributed equally to this work. Correspondence should be addressed to J.H. (jhampe@1med.uni-kiel.de) or S.S. (s.schreiber@mucosa.de).

Received 31 July; accepted 5 December; published online 31 December 2006; doi:10.1038/ng1954

Table 2 Top 35 Crohn disease–associated SNPs, ranked with respect to the *P* value obtained in an allele-based case-control comparison in panel A

	Screening (panel A)							Replication (panel B)		
#	Gene	Celera ID	dbSNP ID	Chr.	Position	P_{CCA}	P_{CCG}	P_{CCA}	P_{CCG}	P_{TDT}
1	*DCP1B*	hCV2194128	rs12423058	12	1,934,927	5.8×10^{-14}	3.6×10^{-13}	0.92	0.54	0.07
2	*TINAG*	hCV12027972	rs1058768	6	54,232,983	1.7×10^{-12}	7.9×10^{-11}	0.27	0.15	0.21
3	*OR8H1*	hCV25770775	rs17613241	11	55,839,523	2.2×10^{-9}	2.8×10^{-9}	0.27	0.41	0.15
4	*TTN*	hCV25626488	rs10497517	2	179,646,084	3.4×10^{-7}	2.7×10^{-6}	0.17	0.37	0.70
5	*OR10A4*	hCV15895352	rs2595453	11	6,862,804	0.00005	0.0003	0.18	0.21	0.43
6	*hCG1744077*	hCV3111449	rs211716	1	75,529,932	0.0001	0.0006	0.05	0.16	0.14
7	*hCG1744077*	hCV928121	rs211715	1	75,530,066	0.0002	0.001	0.07	0.21	0.19
8	**S100Z**	**hCV8796177**	**rs1320308**	**5**	**76,255,325**	**0.0003**	**0.0010**	**0.05**	**0.11**	**0.40**
9	*IL7R*	hCV2025977	rs6897932	5	35,920,076	0.0004	0.0002	0.91	0.99	0.95
10	***APG16L1***	***hCV9095577***	***rs2241880***	***2***	***234,470,182***	***0.0004***	***0.002***	***0.00001***	***0.00007***	***0.00001***
11	*FLJ23577*	hCV25770123	–	5	35,715,804	0.0004	0.004	0.21	0.41	0.81
12	*U2*	hCV25637975	rs6730351	2	223,793,960	0.0007	0.003	0.55	0.81	0.50
13	***APBB2***	**hCV1558531**	**rs4861358**	**4**	**40,931,441**	**0.0009**	**0.004**	**0.04**	**0.06**	**0.57**
14	*SLC17A3*	hCV1911085	rs1165165	6	25,970,445	0.0009	0.004	0.58	0.26	0.90
15	***hCG1789632***	**hCV25929364**	**rs10948733**	**6**	**52,867,218**	**0.0009**	**0.004**	**0.05**	**0.04**	**0.50**
16	*NALP13*	hCV2092168	rs303997	19	61,116,255	0.001	0.005	0.73	0.86	0.82
17	*hCG1812162*	hCV25994942	rs10483261	14	20,346,679	0.001	0.005	0.79	0.76	0.08
18	*hCG1646471*	hCV15965545	rs2291479	3	179,495,857	0.001	0.006	0.49	0.72	0.60
19	*HS6ST3*	hCV3118872	rs2282135	13	95,187,906	0.001	0.003	0.12	0.25	0.86
20	***PKD1L2***	**hCV8443426**	**rs1869348**	**16**	**80,921,788**	**0.002**	**0.003**	**0.0004**	**0.002**	**0.52**
21	*VGF*	hCV25649609	–	7	100,378,082	0.002	0.006	0.27	0.17	0.13
22	***TXNDC11***	**hCV1388401**	**rs3190321**	**16**	**11,740,094**	**0.002**	**0.002**	**0.09**	**0.04**	**0.74**
23	*PLSCR4*	hCV25647383	rs3762685	3	147,259,528	0.002	0.005	0.71	0.77	0.30
24	*OR5U1*	hCV2519378	rs9257694	6	29,382,496	0.002	0.008	0.52	0.77	1.00
25	***UBQLN4***	**hCV16187524**	**rs2297792**	**1**	**153,228,236**	**0.002**	**0.006**	**0.003**	**0.009**	**0.65**
26	***CARD15***	**hCV11717466**	**rs2066845**	**16**	**50,543,573**	**0.002**	**0.008**	**8.6×10^{-8}**	**7.1×10^{-7}**	**0.002**
27	*FUCA1*	hCV12023629	rs11549094	1	23,650,437	0.002	0.008	0.48	0.54	0.69
28	*hCG1999532*	hCV2481084	rs3129096	6	29,291,365	0.002	0.008	0.45	0.67	0.83
29	*OR2J2*	hCV11194783	rs3116817	6	29,257,553	0.002	0.010	0.57	0.84	0.61
30	*FLJ25660*	hCV2537241	rs541169	19	40,410,860	0.003	0.01	0.77	0.39	0.32
31	*KUB3*	hCV25770320	rs3751325	12	56,621,893	0.003	0.001	0.46	0.75	0.25
32	*SLC16A4*	hCV15961275	rs2271885	1	110,220,442	0.003	0.01	0.31	0.09	0.10
33	***U1***	**hCV2475291**	**rs2157453**	**1**	**170,103,324**	**0.003**	**0.0008**	**0.002**	**0.006**	**0.29**
34	***SLC22A4***	**hCV3170459**	**rs1050152**	**5**	**131,752,536**	**0.003**	**0.003**	**2.6×10^{-6}**	**1.5×10^{-6}**	**0.02**
35	***AQP9***	**hCV11669234**	**rs1867380**	**15**	**56,192,337**	**0.003**	**0.01**	**0.02**	**0.03**	**0.67**

CCA: case-control comparison, panel A. The full list of SNPs fulfilling the follow-up criterion is provided in **Supplementary Table 1**. Also included above are the *P* values for the genotype-based case-control comparison (CCG) and the TDT. Nucleotide positions refer to NCBI build 34. Markers with $P \leq 0.05$ in either the case-control or the TDT analysis in replication panel B are highlighted in boldface. SNPs with a significant result in both panel B tests are marked by underlining. In addition to rs2241880, only SNP rs1050152 (L503F) in the *SLC22A4* gene reported earlier[4] and the known *CARD15* SNP rs2066845 ('SNP12'; ref. 1) yielded consistent replication.

indication of formal replication in both the transmission disequilibrium test (TDT) and the case-control comparison of panel B, only (i) rs2241880 (T300A) in the *ATG16L1* gene, (ii) the two previously reported variants rs1050152 (L503F) in the *SLC22A4* gene[4] and (iii) rs2066845 ('SNP12'; ref. 1) in *CARD15* matched this criterion. *CARD15* variant rs2066844 (also known as 'SNP8') failed genotyping; rs2066847 ('SNP13') was not part of the coding SNP (cSNP) panel because the necessary insertion or deletion (indel) design was not feasible on the SNPlex v1.0 platform. Association of the G allele of rs2241880 with Crohn disease was significant in both the allele-based case-control comparison ($P = 1.6 \times 10^{-5}$) and the TDT ($P = 2.7 \times 10^{-5}$). We confirmed genotyping results obtained with SNPlex using TaqMan (99.8% genotype concordance), thereby excluding a technical artifact. We did not observe any significant differences in terms of age of onset, frequency of ileal disease, stenoses or fistulae between different genotypes ($P > 0.5$; **Supplementary Table 2** online).

We searched for additional mutations in the *ATG16L1* gene by resequencing all exons, splice sites and the promoter region in the genomic DNA of 47 individuals with Crohn disease. We did not identify any other coding or splice site variants besides rs2241880. In panel B, we genotyped 28 tagging SNPs ($r^2 > 0.8$, minor allele frequency $\geq 1\%$; **Supplementary Fig. 1** online) selected from the CEU HapMap sample (which comprises Utah residents with ancestry from northern and western Europe) (**Supplementary Table 3** online). We found that the intronic SNP rs2289472 had the same minor allele frequency (0.47) as coding SNP rs2241880, and it yielded a slightly more significant disease association ($P = 1.4 \times 10^{-5}$). Marker rs2289472 is localized 1,082 bp upstream of exon 9 and is not part of any recognized regulatory motif. In a logistic regression analysis, none of the 28 tagging SNPs significantly improved the model fit in the presence of rs2241880 (all $P > 0.05$). Taken together with the results of a subsequent haplotype analysis (**Table 3** and

Table 3 Results of a haplotype analysis of 12 SNPs at the *ATG16L1* locus

	Haplotype	f_{cases}	$f_{controls}$	OR	$P_{COCAPHASE}$	$f_{transmitted}$	$f_{non\text{-}transmitted}$	$P_{TDTPHASE}$
1	ACAGCAA**G**TGCG	0.603	0.533	1.00	0.00002	0.515	0.305	0.002
2	ACGACTG**A**CACG	0.254	0.285	0.79	0.04	0.280	0.406	0.04
3	GGAACAG**A**CATG	0.053	0.067	0.70	0.07	0.088	0.130	0.19
4	AGAATAG**A**CATG	0.046	0.066	0.62	0.008	0.080	0.084	0.87
5	ACGACAG**A**CGTA	0.044	0.048	0.81	0.57	0.038	0.075	0.10

SNPs included in the analysis are marked by asterisks in **Supplementary Figure 1** to illustrate their block assignment. All analyses were carried out using COCAPHASE or TDTPHASE[30]. *P* values refer to the null hypothesis of either equal haplotype frequencies (COCAPHASE) or of a transmission–to–non-transmission ratio equal to 0.5 (TDTPHASE). f_{cases} and $f_{controls}$ represent haplotype frequency among affected individuals and controls, respectively. $f_{transmitted}$ and $f_{non\text{-}transmitted}$ represent haplotype frequency on transmitted chromosomes and nontransmitted chromosomes, respectively. Nonsynonymous SNP rs2241880 is highlighted in bold, and the risk allele is underlined. Obviously, the sole risk haplotype (ACAGCAA**G**TGCG) is fully signified by rs2241880 allele G; all other haplotypes carry allele A (note that the case-control–based ORs were normalized so that OR = 1.0 for haplotype 1). This haplotype pattern strongly suggests that rs2241880 is indeed the major risk variant at the *ATG16L1* locus.

Supplementary Fig. 1), these findings suggest that the Crohn disease risk conferred by *ATG16L1* gene variation is probably confined to individuals carrying susceptibility allele G at rs2241880. In the combined analysis of panels A and B, odds ratios (ORs) were 1.45 (95% confidence interval (c.i.): 1.21–1.74) for carrying allele G (population attributable risk (PAR) = 0.26) and 1.77 (95% c.i.: 1.43–2.18) for homozygosity (PAR = 0.17). The frequency of the G allele was 0.60 in affected individuals and 0.53 in controls.

We also replicated the Crohn disease association of rs2242880 in a sample from the UK (panel C) using an independent TaqMan assay. The British data yielded $P = 0.0004$ in the allele-based test (OR: 1.35; 95% c.i.: 1.14–1.59) and $P = 0.0002$ in the genotype-based test (OR for homozygosity for allele G: 1.71; 95% c.i.: 1.23–2.39). The frequency of the G allele (59% among affected individuals, 52% among controls) was similar to that in the German sample.

Some of the variants listed in **Table 2** or **Supplementary Table 1** may still represent important candidates for disease susceptibility. The requirement of a significant replication ($P \leq 0.05$) in both the TDT and the case-control comparison may have led to the exclusion of relevant variants owing to a lack of power. Although the present study still fell short of comprehensively assessing all putatively functional variation in the human genome, our results, together with those of other recent studies[17], demonstrate that direct association analysis using cSNPs is a meaningful complement for genome-wide linkage disequilibrium–based association studies like the ones reported for Crohn disease[10,11]. The coding variant rs11209026 in *IL23R* (ref. 11) was not part of the cSNP panel and thus was not detected in this study. Weak linkage evidence to the *ATG16L1* (2q37.1) region has been found in some genome scans: LOD = 1.0 was reported in ref. 18 at the 190 cM position, and LOD = 1.2 was obtained at 180 cM in the German and UK genome scan[19]. Notably, in the published reanalysis[20] of the German and UK scan, the 2q LOD peak increased to 1.5 in families with wild-type *CARD15*.

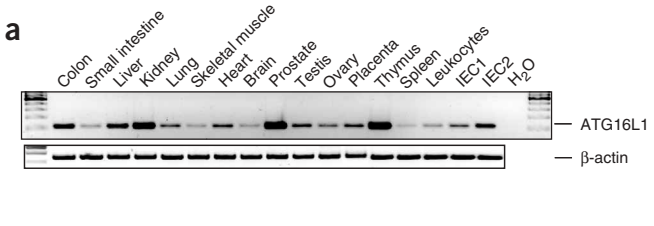
Frequencies of rs2241880 allele G in individuals with ulcerative colitis (0.54) and in controls (0.53) were virtually identical. Thus, we did not obtain evidence for a disease association either from the case-control comparison ($P > 0.4$) or the TDT ($P > 0.9$) for ulcerative colitis.

We investigated the disease-associated variants in *ATG16L1* and *CARD15* for statistical interaction with respect to Crohn disease risk. We classified individuals in the German panel B either homozygous wild-type (dd), heterozygous carrier (Dd) or homozygous carrier (DD, which included compound heterozygotes) for the three main causative *CARD15* SNPs (rs2066844, rs2066845 and rs2066847), as suggested[21]. The frequencies and ORs of individual *CARD15* risk genotypes, stratified by rs2241880 genotype, are shown in **Table 4**. A statistical interaction became apparent between rs2241880 and the *CARD15* low-risk genotypes dd and Dd. The OR difference was significant for rs2241880 genotype GG (2.03 on dd background versus 1.04 for Dd; Breslow-Day $\chi^2 = 4.267$, 1 degree of freedom, $P = 0.039$) but not for AG. Thus, the *ATG16L1* variant rs2241880 is a risk factor even in the absence of *CARD15* mutations. On the background of *CARD15* high-risk genotype DD, the risk conferred by carrying rs2241880 allele G seemed to be higher than in the presence of dd or Dd, but the confidence intervals of the respective ORs were still wide owing to the small number of DD controls (**Table 4**). Nevertheless, when we combined rs2241880 genotypes GG and AG, the joint OR of 5.89 (95% c.i.: 1.23–29.21) was significantly larger than unity (Fisher's exact

Table 4 Analysis of the statistical interaction between *ATG16L1* SNP rs2241880 and *CARD15* genotype

		CARD15 genotype		
	ATG16L1	dd	dD	DD
Control	GG	219	62	2
	AG	435	87	2
	AA	185	35	5
Crohn disease	GG	175	92	42
	AG	232	136	57
	AA	73	50	21
OR relative to AA (95% c.i.)[a]	GG	2.03 (1.43–2.88)	1.04 (0.59–1.84)	5.00 (0.76–41.05)
	AG	1.35 (0.98–1.87)	1.09 (0.64–1.88)	6.79 (1.04–55.16)
	AA	1.00	1.00	1.00
OR relative to AA/dd (95% c.i.)[b]	GG	2.03 (1.43–2.88)	3.76 (2.42–5.86)	53.22 (12.15–325.70)
	AG	1.35 (0.98–1.87)	3.96 (2.66–5.91)	72.23 (16.68–440.10)
	AA	1.00	3.62 (2.11–6.23)	10.64 (3.62–33.60)

Abbreviations representing *CARD15* mutation status are as described in Results.
[a]Odds ratio relative to rs2241880 genotype AA in each *CARD15* genotype stratum. [b]Odds ratio relative to the joint *ATG16L1*/*CARD15* low-risk genotype AA/dd.



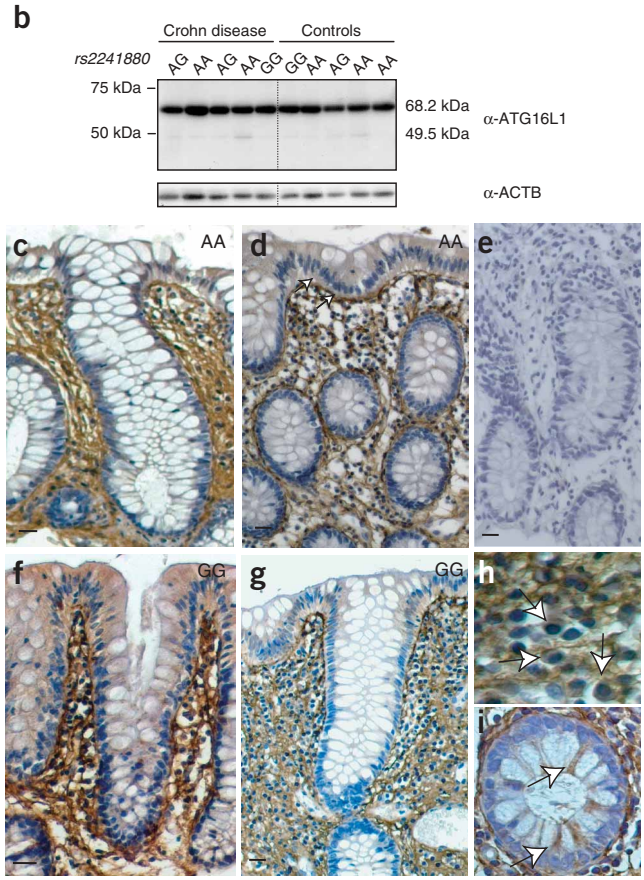


Figure 1 Expression and cellular localization of the *ATG16L1* gene product. Amounts of ATG16L1 protein are influenced neither by the rs2241880 genotype nor by intestinal inflammation. (**a**) Expression of *ATG16L1* in a set of different tissues as detected by RT-PCR (IEC, intestinal epithelial cells). The corresponding β-actin control (with a 518-bp amplicon size) is given below. (**b**) Amounts of ATG16L1 in colonic mucosal biopsy specimens. Proteins (15 μg) from rectal mucosal biopsies of Crohn disease patients and normal controls (N) were separated by denaturing SDS-PAGE, transferred onto polyvinylidine fluoride (PVDF) membranes and probed for the presence of ATG16L1 using a specific primary antibody and horseradish peroxidase–coupled secondary antibody. The abundant 68.2-kDa band corresponds to the predicted molecular weight of the isoform encoded by GenBank AY398617 (UniProt accession number Q676U5). This major protein form was used for three-dimensional modeling of the ATG16L1 protein (**Fig. 2**). The 48.5-kDa band corresponds to GenBank EF079890, which contains exon 9 (harboring rs2241880) in the same reading frame as the major form. (**c**–**i**) Representative immunohistochemistry results (out of six normal controls and six individuals with Crohn disease) from colonic mucosal tissue samples from normal controls for genotype AA (**c**) and GG (**f**) and individuals with Crohn disease for genotype AA (**d**) and GG (**g**) are shown to demonstrate the cellular localization of the ATG16L1 protein. Cellular expression patterns in mononuclear cells (**h**) and intestinal epithelial cells (**i**) of a Crohn disease sample (*rs2241880AA*) are shown at higher magnification. **e** shows a Crohn disease sample without the primary antibody, used as a control. Bar represents 10 μm in **c**–**g**.

two-sided $P = 0.016$), thereby confirming that rs2241880 allele G is a risk factor on a high-risk *CARD15* genotype background as well. Larger studies will be needed to determine the exact nature of the statistical interaction between the *ATG16L1* and *CARD15* variants and to obtain narrower confidence intervals, particularly for the DD stratum.

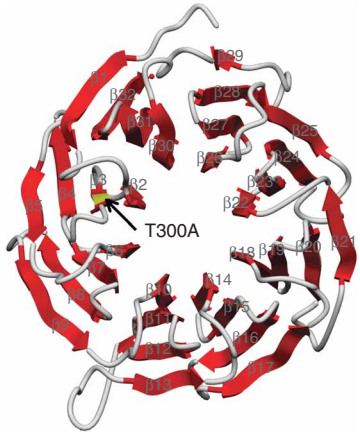


Figure 2 Three-dimensional structural model of the WD-repeat domain of human ATG16L1. The 32 β-strands forming an eight-bladed β-propeller are numbered as in **Supplementary Note**. The location of the variant amino acid T300A in strand β3, corresponding to rs2241880, is marked in yellow.

We investigated expression of the *ATG16L1* gene by RT-PCR in a tissue panel, confirming expression in colon, small bowel, intestinal epithelial cells and immune tissues like spleen and leukocytes (**Fig. 1a**). Recently, multiple splice variants of *ATG16L1* have been reported[22]. In all annotated and reported splice variants, exon 9, which contains the Crohn disease susceptibility variant rs2241880, is translated in the same reading frame, thereby consistently leading to a threonine to alanine substitution by the SNP (**Fig. 1b**). We detected expression of ATG16L1 in the intestinal epithelium by immunohistochemistry (**Fig. 1c–i**) and did not find any significant difference in the amount or pattern of expression in tissue from controls and affected individuals. Expression of ATG16L1 was also independent of the rs2241880 genotype at both the protein and the cDNA level (**Fig. 1b**).

The exact functional consequences of rs2241880 remain tentative at this point. The human ATG16L1 protein has an N-terminal APG16 domain consisting of coiled coils and eight C-terminal WD repeats (see three-dimensional model in **Fig. 2** and *in silico* protein analysis in the **Supplementary Note** online). The variant leads to an amino acid exchange (polar threonine to nonpolar alanine) at the evolutionarily conserved position 300 of the N terminus of the WD-repeat domain in ATG16L1 (**Fig. 2**). The interaction partner of this domain has not yet been identified experimentally[23]. However, it is clear that ATG16L1 is part of the autophagosome pathway. The established role of the autophagosome in the processing of intracellular bacteria would further support the emerging concept of Crohn disease as an inflammatory barrier disorder. The lack of association with ulcerative colitis and the statistical interaction with CARD15 suggest that the underlying biological processes may be disease specific.

METHODS

Participants. Recruitment details are given in **Supplementary Methods** online. All participants gave written, informed consent, and the recruitment protocols were approved by the ethics committees at the respective recruiting institutions. Genotyping using SNPlex and TaqMan technologies (Applied Biosystems) was performed as described, using an automated laboratory setup[24,25]. For the construction of the panel of 19,779 nonsynonymous SNPs, data from dbSNP (build 117) were combined with polymorphisms discovered by the Applera exon resequencing and Celera shotgun sequencing projects[26] (see **Supplementary Methods**, **Supplementary Table 4** online and the URL noted below). Sequencing of genomic DNA was performed using ABI BigDye chemistry (primer sequences are given in **Supplementary Table 5** online). Traces were inspected for SNPs and indels using InSNP[27] and Sequencher. Standard protocols were used for the isolation of epithelial cells, RT-PCR, protein blot and immunohistochemistry (**Supplementary Table 6** and **Supplementary Methods** online).

Statistical analysis. Markers were tested for Hardy-Weinberg equilibrium in controls before inclusion in the analyses ($P > 0.05$). Single-marker case-control analyses and TDT were performed using Haploview[28] and GENOMIZER[29]. Haplotype frequency estimates were obtained from singletons using COCAPHASE[30]. Significance testing of haplotype frequency differences was performed with COCAPHASE and TDTPHASE[30]. Genotype-based logistic regression analysis was performed with R, coding individual SNP genotypes as binary indicator variables. Analysis of statistical interaction was carried out using the FREQ procedure of SAS.

Accession codes. The GenBank accession codes of the major splice variants of ATG16L1 are AY398617 and EF079890.

URLs. The R suite can be found at http://www.r-project.org/. Detailed information on the cSNP panel can be found at http://cSNP.applied-biosystems.com/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank all affected individuals, families and physicians for their cooperation. We acknowledge the cooperation of the German Crohn and Colitis Patient Association (Deutsche Morbus Crohn und Colitis Vereinigung) and the contributing gastroenterologists. The authors thank T. Wesse, B. Petersen, L. Bossen, T. Henke, S. Ehlers, A. Dietsch, T. Kaacksteen and D. Soars for technical help. The ongoing technical and logistic support and helpful discussions with A. Toeppel, C.R. Scafe, A. Kejariwal, H.M. Wenz, M. Rhodes, S. Short, T. Woodage and D.A. Gilbert from Applied Biosystems is especially acknowledged. We thank C.-C. Chiang for preparing the submission of new SNPs to dbSNP. T.H. Karlsen (Rikshospitalet) is acknowledged for helpful discussions. We thank A. Forbes, J. Sanderson and S. Fisher for case ascertainment and establishment of the UK Crohn disease database. We acknowledge use of genotype data from the British 1958 Birth Cohort DNA collection, funded by Medical Research Council grant G0000934 and Wellcome Trust grant 068545/Z/02. This study was supported by the German Ministry of Education and Research (BMBF) through the National Genome Research Network (environmental diseases network and SMP-GEM), the POPGEN biobank, the analysis infrastructure and methods of the MediGrid project and the German Research Council (Ha 3091/1-1, 2-1), Applied Biosystems, TECAN, the Wellcome Trust and CORE (UK).

AUTHOR CONTRIBUTIONS

J.H. and A.F. established the genotyping and analysis methodology; A.F. performed the genotyping and association analysis and contributed to the writing of the manuscript; S.S. supervised the patient recruitment; J.H. recruited the German patients and drafted the manuscript; P.R., A.T., A.F., K.H., R.H., B.S. and M.P. performed the protein, immunohistochemistry and cDNA experiments; M.T. provided LIMS programming support; M.A., G.M. and T.L. performed *in silico* protein analysis and contributed to writing the manuscript; F.D.L.V. designed the cSNP panel and genotyping assays and contributed to the manuscript; J.B. and S.G. helped establish the SNPlex automation system; N.P., C.O. and C.M. performed the replication experiment in the UK samples; U.F. contributed to the design and the writing of the paper; M.K. provided genetic epidemiology consulting, performed the interaction analysis and helped draft the manuscript; and J.H. and S.S. jointly designed and supervised the experiment.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Hugot, J.P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).
2. Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606 (2001).
3. Rioux, J.D. *et al.* Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat. Genet.* **29**, 223–228 (2001).
4. Peltekova, V.D. *et al.* Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat. Genet.* **36**, 471–475 (2004).
5. Stoll, M. *et al.* Genetic variation in DLG5 is associated with inflammatory bowel disease. *Nat. Genet.* **36**, 476–480 (2004).
6. Brant, S.R. *et al.* MDR1 Ala893 polymorphism is associated with inflammatory bowel disease. *Am. J. Hum. Genet.* **73**, 1282–1292 (2003).
7. Ho, G.T. *et al.* ABCB1/MDR1 gene determines susceptibility and phenotype in ulcerative colitis: discrimination of critical variants using a gene-wide haplotype tagging approach. *Hum. Mol. Genet.* **15**, 797–805 (2006).
8. Schwab, M. *et al.* Association between the C3435T MDR1 gene polymorphism and susceptibility for ulcerative colitis. *Gastroenterology* **124**, 26–33 (2003).
9. McGovern, D.P. *et al.* Association between a complex insertion/deletion polymorphism in NOD1 (CARD4) and susceptibility to inflammatory bowel disease. *Hum. Mol. Genet.* **14**, 1245–1250 (2005).
10. Yamazaki, K. *et al.* Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum. Mol. Genet.* **14**, 3499–3506 (2005).
11. Duerr, R.H. *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* **314**, 1461–1463 (2006).
12. Maeda, S. *et al.* Nod2 mutation in Crohn's disease potentiates NF-kappaB activity and IL-1beta processing. *Science* **307**, 734–738 (2005).
13. Kobayashi, K.S. *et al.* Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* **307**, 731–734 (2005).
14. Girardin, S.E. *et al.* Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *J. Biol. Chem.* **278**, 8869–8872 (2003).
15. Hampe, J. *et al.* Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* **357**, 1925–1928 (2001).
16. Watanabe, T., Kitani, A., Murray, P.J. & Strober, W. NOD2 is a negative regulator of Toll-like receptor 2-mediated T helper type 1 responses. *Nat. Immunol.* **5**, 800–808 (2004).
17. Smyth, D.J. *et al.* A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat. Genet.* **38**, 617–619 (2006).
18. Duerr, R.H., Barmada, M.M., Zhang, L., Pfutzer, R. & Weeks, D.E. High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11–12. *Am. J. Hum. Genet.* **66**, 1857–1862 (2000).
19. Hampe, J. *et al.* A genome-wide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *Am. J. Hum. Genet.* **64**, 808–816 (1999).
20. Shaw, S.H. *et al.* Stratification by CARD15 variant genotype in a genome-wide search for inflammatory bowel disease susceptibility loci. *Hum. Genet.* **113**, 514–521 (2003).
21. Croucher, P.J.P. *et al.* Haplotype structure and association to Crohn's disease of CARD15 mutations in two ethnically divergent populations. *Eur. J. Hum. Genet.* **11**, 6–16 (2003).
22. Zheng, H. *et al.* Cloning and analysis of human Apg16L. *DNA Seq.* **15**, 303–305 (2004).
23. Mizushima, N. *et al.* Mouse Apg16L, a novel WD-repeat protein, targets to the autophagic isolation membrane with the Apg12-Apg5 conjugate. *J. Cell Sci.* **116**, 1679–1688 (2003).
24. Hampe, J. *et al.* An integrated system for high throughput TaqMan based SNP genotyping. *Bioinformatics* **17**, 654–655 (2001).
25. Hampe, J. *et al.* Evidence for a NOD2-independent susceptibility locus for inflammatory bowel disease on chromosome 16p. *Proc. Natl. Acad. Sci. USA* **99**, 321–326 (2002).
26. Venter, J.C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
27. Manaster, C. *et al.* InSNP: a tool for automated detection and visualization of SNPs and InDels. *Hum. Mutat.* **26**, 11–19 (2005).
28. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
29. Franke, A. *et al.* GENOMIZER: an integrated analysis system for genome-wide association data. *Hum. Mutat.* **27**, 583–588 (2006).
30. Dudbridge, F. Pedigree disequilibrium tests for multilocus haplotypes. *Genet. Epidemiol.* **25**, 115–121 (2003).