nature genetics

Genetic variation in *DLG5* is associated with inflammatory bowel disease

Monika Stoll[1,7], Brit Corneliussen[2], Christine M Costello[1], Georg H Waetzig[3], Bjorn Mellgard[2], W Andreas Koch[1], Philip Rosenstiel[1], Mario Albrecht[4], Peter J P Croucher[1], Dirk Seegert[3], Susanna Nikolaus[1], Jochen Hampe[1,5], Thomas Lengauer[4], Stefan Pierrou[2], Ulrich R Foelsch[1], Christopher G Mathew[6], Maria Lagerstrom-Fermer[2] & Stefan Schreiber[1,5]

Crohn disease and ulcerative colitis are two subphenotypes of inflammatory bowel disease (IBD), a complex disorder resulting from gene-environment interaction. We refined our previously defined linkage region for IBD on chromosome 10q23 and used positional cloning to identify genetic variants in *DLG5* associated with IBD. *DLG5* encodes a scaffolding protein involved in the maintenance of epithelial integrity. We identified two distinct haplotypes with a replicable distortion in transmission (P = 0.000023 and P = 0.004 for association with IBD, P = 0.00012 and P = 0.04 for association with Crohn disease). One of the risk-associated *DLG5* haplotypes is distinguished from the common haplotype by a nonsynonymous single-nucleotide polymorphism 113G→A, resulting in the amino acid substitution R30Q in the DUF622 domain of DLG5. This mutation probably impedes scaffolding of DLG5. We stratified the study sample according to the presence of risk-associated *CARD15* variants to study potential gene-gene interaction. We found a significant difference in association of the 113A *DLG5* variant with Crohn disease in affected individuals carrying the risk-associated *CARD15* alleles versus those carrying non-risk-associated *CARD15* alleles. This is suggestive of a complex pattern of gene-gene interaction between *DLG5* and *CARD15*, reflecting the complex nature of polygenic diseases. Further functional studies will evaluate the biological significance of DLG5 variants.

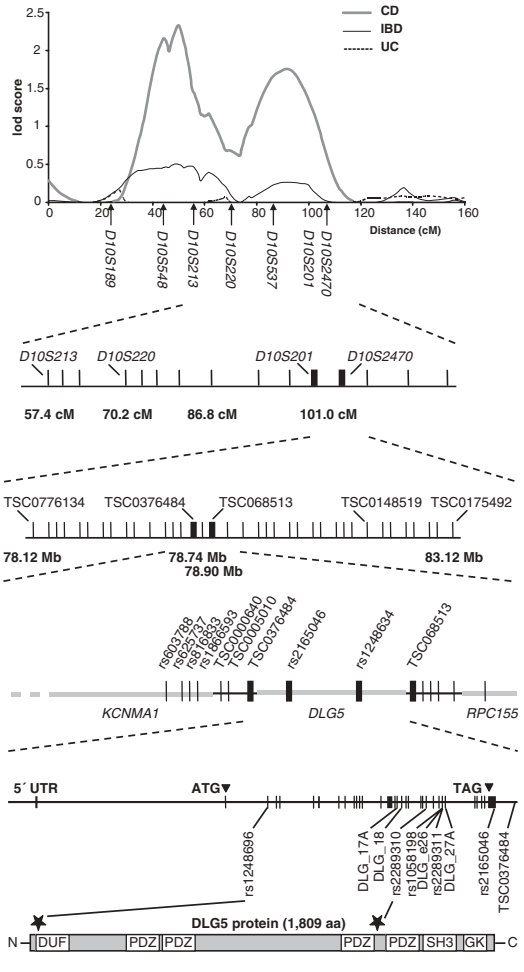

Figure 1 Genetic variants of *DLG5* are associated with IBD. The experimental steps from linkage mapping on the long arm of chromosome 10 to identification of genetic variants in *DLG5* associated with IBD, Crohn disease (CD) and ulcerative colitis (UC) are shown. Annotation of the most important microsatellite and SNP markers that led to the identification of *DLG5* as a susceptibility gene for IBD and Crohn disease is also shown. The markers with strongest association with the IBD phenotype are indicated with bold vertical tick marks. The corresponding association results are presented in detail in **Tables 1** and **2** and **Figure 3**.

[1]First Department of Medicine, University Hospital Schleswig-Holstein, Schittenhelmstr. 12, 24105 Kiel, Germany. [2]AstraZeneca, Experimental Medicine and Department of Molecular Sciences, R&D Molndal, Sweden. [3]Conaris Research Institute AG, Kiel, Germany. [4]Max-Planck-Institute for Informatics, Saarbrücken, Germany. [5]Institute for Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany. [6]Guy's, King's and St. Thomas' School of Medicine, London, UK. [7]Present address: Genetic Epidemiology of Vascular Disorders, Institute for Arteriosclerosis Research, Muenster, Germany. Correspondence should be addressed to S.S. (s.schreiber@mucosa.de).

Published online 11 April 2004; doi:10.1038/ng1345

IBD is a spectrum of chronic relapsing inflammatory disorders affecting the gastrointestinal tract that can be classified into Crohn disease and ulcerative colitis. The identification of *CARD15* (refs.1–3) and several loci associated with susceptibility to IBD in independent linkage studies[4] documents the polygenic etiology of IBD. We previously identified a locus in the pericentromeric region of chromosome 10 that was associated with susceptibility to IBD in a genome-wide linkage scan involving 282 families of European descent[5]. Fine mapping at an average distance of 5 cM using an additional 11 microsatellite markers in an extended linkage cohort (111 additional families including 422 affected sibling pairs) confirmed initial linkage findings and identified a two-peak linkage curve extending from *D10S547* to *D10S192* (multipoint lod score = 2.07, $P = 0.0033$ at *D10S548*) and a second peak at *D10S201* (multipoint lod score = 1.6) for associated with Crohn disease (**Fig. 1**). We used a hierarchical linkage disequilibrium (LD) study to search for the causal variant(s) in the 40-cM interval (**Fig. 1**). Transmission disequilibrium testing (TDT) of trios randomly drawn from each family showed a significant single-point association with Crohn disease at *D10S201* ($P < 0.01$), located in the second linkage peak on 10q22–10q23. We finely mapped the underlying 5-Mb region at an average distance of 75–120 kb using 37 single-nucleotide polymorphisms (SNPs) selected from the TSC allele frequency project in 457 independent trios with IBD (**Supplementary Table 1** online). The marker TSC0376484 (rs1344966) in this panel was significantly associated with Crohn disease ($\chi^2 = 9.00$, $P = 0.002$) and more strongly with IBD (Crohn disease and ulcerative colitis, $\chi^2 = 11.65$, $P = 0.0006$; **Fig. 1**).

TSC0376484 is located near two genes of possible (patho)physiological relevance to chronic intestinal inflammation: *KCNMA1*, encoding a potassium-gated calcium channel[6], and *DLG5*, a member of the membrane-associated guanylate kinase gene family, which is important in the maintenance of epithelial cell integrity[7]. To genetically narrow the association signal to one single candidate, we used LD mapping and genotyped selected publicly available SNPs from each gene in the 457 trios with IBD. The association signal was confined to *DLG5*, and none of the markers in *KCNMA1* was significantly associated with IBD (**Supplementary Table 2** online).

We sequenced coding exons 2–32 and the exon-intron boundaries of *DLG5* in 47 individuals with IBD and identified or verified 33 SNPs (**Supplementary Tables 3** and **4** online). We then tested all these SNPs for disease association. Association with the IBD phenotype was strongest (**Table 1**), with 18 markers in *DLG5* showing significant association. Separate analysis of the Crohn disease and ulcerative colitis subgroups showed a strong association in the Crohn disease subgroup, which is in accordance with the original linkage observation[5]. The weaker signal in the ulcerative colitis subgroup may be due to reduced power in a small sample size. Because the combined group had the strongest association, we suggest that the signal in *DLG5* reflects a factor associated with general susceptibility to IBD rather than to Crohn disease only.

Pairwise LD measures (D′) indicated strong LD across the entire gene, defining a single haplotype block of ~85 kb and D′ values >0.8, except at TSC0000361 (located on the neighboring LD segment). We

Table 1 Summary of TDT results in German trios for single-point association with IBD, Crohn disease and ulcerative colitis

	IBD ($n = 457$)			Crohn disease ($n = 302$)			Ulcerative colitis ($n = 155$)		
Assay name	T:U	χ^2	*P*	T:U	χ^2	*P*	T:U	χ^2	*P*
TSC0000361	24:15	2.077	0.1495	13:12	0.04	0.8415	11:03	4.571	0.0325
rs1248696	92:64	5.026	0.0250	60:42	3.176	0.0747	32:22	1.852	0.1736
rs1248680	151:191	4.678	0.0306	102:127	2.729	0.0985	64:49	1.991	0.1582
rs1248677	203:180	1.381	0.2399	138:123	0.862	0.3532	57:65	0.525	0.4687
rs1248670	172:159	0.511	0.4747	117:109	0.283	0.5947	50:55	0.238	0.6257
rs1270912	201:182	0.943	0.3315	134:124	0.388	0.5334	58:67	0.648	0.4208
DLG5_e13	154:196	5.4	0.0248	105:128	2.24	0.1319	68:49	3.085	0.0790
rs2289308	230:171	8.681	0.0032	155:115	5.926	0.0149	75:56	2.756	0.0969
rs1248634	150:196	6.116	0.0134	104:130	2.889	0.0892	66:46	3.571	0.0588
DLG5_e17	152:192	4.651	0.0310	104:128	2.483	0.1151	64:48	2.286	0.1305
DLG5_e18	236:175	9.054	0.0026	158:119	5.491	0.0191	78:56	3.612	0.0574
rs2289310	40:24	4.0	0.0455	24:17	1.195	0.2743	07:16	3.522	0.0606
rs1261990	147:196	7.0	0.0082	100:131	4.16	0.0414	65:47	2.893	0.0890
DLG5_e25	23:15	1.684	0.1944	13:11	0.167	0.6828	04:10	2.571	0.1088
rs1058198	228:166	9.756	0.0018	151:111	6.107	0.0135	77:55	3.667	0.0555
DLG5_e26	225:162	10.256	0.0014	151:108	7.139	0.0075	74:54	3.125	0.0771
rs2289311	231:171	8.955	0.0028	154:114	5.97	0.0146	77:57	2.985	0.0840
DLG5_e27	235:175	8.72	0.0030	157:118	5.531	0.0187	78:57	3.267	0.0707
rs2241833	152:196	5.563	0.0183	103:132	3.579	0.0585	64:49	1.991	0.1582
rs2579150	146:193	6.516	0.0107	98:130	4.491	0.0341	63:48	2.027	0.1545
rs2812425	201:186	0.581	0.4459	133:129	0.061	0.8049	57:68	0.968	0.3252
rs1058202	176:161	0.668	0.4137	118:107	0.538	0.4633	54:58	0.143	0.7053
rs1058203	155:196	4.789	0.0286	104:130	2.889	0.0892	66:51	1.923	0.1655
rs2165046	241:176	10.132	0.0015	159:120	5.452	0.0195	82:56	4.899	0.0269
rs2165047	143:192	7.167	0.0074	96:129	4.84	0.0278	63:47	2.327	0.1271
rs2579151	148:189	4.988	0.0255	101:130	3.641	0.0564	59:47	1.358	0.2439
TSC0376484	177:123	9.72	0.0018	118:87	4.688	0.0304	59:36	5.568	0.0183

Single-point TDT association for SNP markers located in *DLG5* and flanking regions. Association is shown for IBD (Crohn disease and ulcerative colitis combined) and the subphenotypes Crohn disease and ulcerative colitis separately. T:U, transmitted versus untransmitted allele. The association is most pronounced in the IBD group and the Crohn disease subgroup, whereas only few markers show significance or a trend of association with ulcerative colitis, which is in accord with the original linkage finding[5].

observed a sharp decline in LD at the boundaries of the haplotype block, differentiating *DLG5* from the neighboring genes *KCNMA1* and *RPC155* (**Fig. 2**). Analysis of the extended *DLG5* haplotype identified four common haplotypes (**Fig. 3**), with haplotype A tagged by eight SNPs (haplotype-tagging SNPs or htSNPs) of equivalent genetic information content. Haplotype A was significantly undertransmitted to individuals with IBD and Crohn disease, whereas haplotype D, uniquely tagged by the coding variant 113A, was significantly overtransmitted to individuals with both IBD ($\chi^2 = 8.08$, $P = 0.004$) and Crohn disease ($\chi^2 = 4.15$, $P = 0.04$; **Fig. 3**).

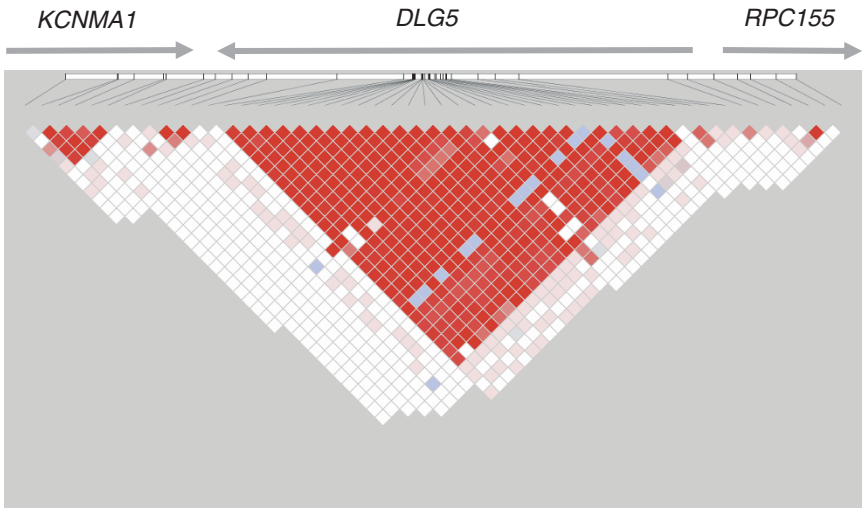

Figure 2 LD across *DLG5* and flanking genomic region. D′ values for pairwise LD between each marker (red = D′ > 0.8) are represented. Top of figure shows spacing of markers across the genomic region. All markers in *DLG5* (large red block) are in strong LD. LD drops off sharply (white blocks) at the boundaries of *DLG5*, indicating that the neighboring genes are located on different (independent) genomic segments.

To corroborate our initial association finding, we genotyped the *DLG5* htSNPs in an independent sample consisting of trios with IBD who had not yet been analyzed ($n = 485$; **Supplementary Table 1** online). The htSNP DLG5_e26 in haplotype A was undertransmitted to the individuals with IBD (transmitted:untransmitted (T:U) ratio of 165:214), replicating the observed association ($P = 0.006$ in a one-tailed test), and rs1058198 had a T:U ratio of 196:237 ($P = 0.024$). 113A was overtransmitted in both IBD (T:U 90:73, $P = 0.09$) and Crohn disease (T:U 58:43, $P = 0.065$) but the distortion was not statistically significant. This can be explained by the smaller proportion of trios with Crohn disease in our replication sample and the reduced power in replication situations. We therefore tested the associated markers in a second independent sample (538 Crohn disease cases and 548 controls) using the case-control study design to estimate the attributable risk in a diverse population of European descent. The 113A variant was significantly associated with the IBD phenotype ($P = 0.001$, odds ratio (OR) = 1.62) as was rs2289310 (4136C→A, resulting in the amino acid substitution P1371Q; $P = 0.01$, $OR = 1.51$). DLG5_e26, tagging haplotype A, was significantly associated with IBD (**Table 2**), providing a second independent replication of association. The combined P values for the repeated, independent associations with IBD ($n = 2$) were 0.029 for 113A and 0.0007 for DLG5_e26, and those for the repeated, independent associations with Crohn disease ($n = 3$) were $P = 0.001$ for 113A and $P = 0.0004$ for DLG5_e26.

Because the 4136A risk allele is not included on the common haplotypes carrying 113A, but instead on a rare haplotype (frequency <1%), we calculated the global differences in genotype combinations for 113A and 4136A to estimate the risk for homozygosity or compound heterozygosity. This analysis identified a significant difference in genotype frequencies (global $\chi^2 = 13.61$, $P = 0.0029$) in individuals with IBD compared with healthy controls. The OR was 1.74 (95% confidence interval = 1.31-2.32) for individuals carrying at least two risk alleles (113A and/or 4136A), suggesting that the overall clinical impact of rare single coding mutations such as 4136A on the IBD phenotype is limited. Our disease model that links 113A and 4136A to the positional signal detected in *DLG5* is further supported by the identification of rare, coding, 'private' variants (resulting in the amino acid substitutions S121G, E514Q, R957H and P979L; frequency <0.5%) through systematic sequence analysis in 47 individuals.

a

Haplotypes	rs1248696 (R30Q)	rs2289308	DLG5_e18	rs2289310 (4136C→A)	rs1058198	DLG5_e26	rs2289311	DLG5_e27	rs2165046	TSC0376484	Frequency
A	2	212221	2	21	2	22	2222	22122	1	22	1	.344
B	2	112211	1	12	1	21	1111	11121	2	11	2	.251
C	2	221122	1	21	1	22	1111	22212	2	22	2	.202
D	1	221122	1	21	1	22	1111	22212	2	22	2	.099

Haplotype D–tagging SNP

Haplotype A–tagging SNPs

b

SNP	Allele	T:U	χ^2	P
113G→A	1	82:54	8.08	0.004

c

SNP	Allele	T:U	χ^2	P
rs2289308	2	137:207	14.0	0.000016
DLG5_e18	2	136:212	16.6	0.000046
rs1058198	2	132:207	16.6	0.000046
DLG5_e26	2	127:204	17.9	0.000023
rs2289311	2	136:208	15.0	0.000104
DLG5_e27	2	139:211	14.8	0.00012
rs2165046	1	140:218	16.99	0.000037
TSC0376484	1	100:162	14.67	0.000128

Figure 3 Multilocus haplotype results. (**a**) The four most common haplotypes for *DLG5*. The boxed alleles refer to the htSNPs. Haplotype D carries the risk-conferring htSNP rs1248696 (113G→A) and is significantly overtransmitted in trios with IBD and Crohn disease (TDT results shown in **b**). Haplotype A is tagged by eight htSNPs that carry equivalent genetic information and is significantly undertransmitted in trios with IBD and Crohn disease (TDT results in shown in **c**).

Table 2 Association of *DLG5* SNPs with IBD in an independent case-control sample

SNP marker and genotypes	Cases	Controls	χ^2	*P*	OR (95% confidence interval)
113G→A					
1-1	7	5	10.18	0.001	1.62 (1.2–2.2)[a]
1-2	125	83			
2-2	393	427			
4136C→A					
1-1	5	0	6.29	0.01	1.51 (0.92–2.57)[a]
1-2	34	26			
2-2	486	490			
DLG5_e26					
1-1	243	207	5.73	0.01	1.35 (1.05–1.72)[b,c]
1-2	209	249			
2-2	62	63			

[a]95% confidence interval for genotypes 1-1 and 1-2 combined (allele 1 as risk allele). [b]95% confidence interval for genotypes 1-2 and 2-2 combined (allele 1 as risk allele). [c]Inverse OR for undertransmitted allele 2 was 0.74.

Finally, we examined potential locus-locus interactions between variants of *DLG5* and variants of *CARD15*, the first susceptibility gene identified for Crohn disease[1–3]. Genetic susceptibility to Crohn disease is mainly conferred by three polymorphisms that induce structural changes in the leucine-rich repeats of CARD15 (R702W, G908R and 3020insC). The allele frequencies of these SNPs range from ~4% to 14% in the cohorts with Crohn disease examined to date, and ~30–40% of individuals with Crohn disease are heterozygous for at least one of the variants, compared with ~10% of control subjects[1–3]. We examined interactions between *DLG5* and *CARD15* by stratifying trios in two groups according to the genotype of the affected child. In trios with IBD and Crohn disease, haplotype A (represented by DLG5_e26) was undertransmitted in the groups carrying both the risk-associated and non-risk-associated variants of *CARD15* (**Table 3**), which suggests that haplotype A reflects genetic variation that acts independently of *CARD15* variants. In trios with Crohn disease, we observed significantly greater transmission of 113A in individuals carrying the risk-associated versus non-risk-associated variants of *CARD15* (**Table 3**). This suggests that the 113A variant is of particular relevance in individuals with Crohn disease and, further, that an interaction may exist between the risk-associated haplotype of *DLG5* and the risk-associated variants of *CARD15*.

We were interested in the hypothetical impact of the associated variants, R30Q and P1371Q, on the function of the DLG5 protein. DLG5 has been implicated in regulating cell growth and maintaining cell shape and polarity[8]. A recent study[9] suggested an epithelial function for DLG5 as a binding partner of vinexin at sites of cell-cell contact, and our preliminary results on expression of *DLG5* mRNA in a variety of tissues confirm the presence of the transcript in the colon, the intestine and isolated intestinal epithelial cells (**Supplementary Fig. 1** online). It is therefore conceivable that DLG5 has a role in maintaining epithelial structure and that genetic variants in *DLG5* interfere with epithelial barrier function in the colon.

DLG5 contains one DUF622 domain, four PDZ domains and one SH3 domain followed by one guanylate kinase domain (**Fig. 1**)[10,11]. All these domains are assumed to be involved in protein-protein interactions, supporting the idea that DLG5 is a multifunctional adapter and scaffold protein. We carried out *in silico* analysis of the potential structural and functional implications of the variants R30Q and P1371Q (**Supplementary Fig. 1** online). The results of this analysis suggested that both variants probably impair the scaffolding functions of DLG5 (**Supplementary Methods** and **Supplementary Fig. 1** online).

We found replicated association between genetic variations in *DLG5* and the risk of developing IBD. The risk-associated *DLG5* haplotype D is uniquely distinguished by the 113A variant and is suggested to be causative, as are rare, private SNPs. The conferred risk is moderate, which is in agreement with a polygenic disease model. Genetic interaction studies suggest interactions between *CARD15* variants and 113A in *DLG5*, but these studies are not yet conclusive and will require large, consolidated efforts by several groups to achieve appropriate statistical power. Future studies in diverse and very large samples are needed to evaluate the population relevance of variants in *DLG5* in this chromosomal region. Functional studies need to define the molecular properties of *DLG5* variants and their influence on the clinical presentation of IBD.

Table 3 Association of *DLG5* haplotypes and interaction with risk-associated *CARD15* alleles

Haplotype	Phenotype	*CARD15* allele[a]	TDT (T:U)[b]	χ^2	*P*	Group comparison[c]
Haplotype A (DLGe_26 allele 2)						
	IBD		162:225	10.25	0.001	
		Risk-associated	58:87	5.8	0.01	
		Non-risk-associated	97:121	2.64	0.1	NS
	Crohn disease		108:151	7.13	0.007	
		Risk-associated	46:73	6.12	0.01	
		Non-risk-associated	56:69	1.35	0.24	NS
Haplotype D (113G→A)	IBD		92:64	5.026	0.02	
		Risk-associated	36:17	6.81	0.009	
		Non-risk-associated	51:43	0.68	0.409	NS
	Crohn disease		60:42	3.17	0.07	
		Risk-associated	30:12	7.71	0.005	
		Non-risk-associated	28:28	0.00	1.00	$P < 0.05$

[a]'Risk-associated' indicates at least one of the mutated *CARD15* alleles of SNP8, SNP12 or SNP13. 'Non-risk-associated' indicates wild-type alleles for all three variants. [b]TDT analysis of the *DLG5* htSNP DLG5_e26 for allele 2 denoting undertransmitted haplotype A, and 113G→A tagging overtransmitted haplotype D (T, number of transmitted chromosomes; U, number of untransmitted chromosomes). This htSNP was genotyped in 457 trios with IBD including 302 trios with Crohn disease and 155 trios with ulcerative colitis. Trios were stratified into groups on the basis of *CARD15* genotype and IBD phenotype of the affected child in each trio. [c]No statistical difference (NS) in the T:U ratios (strength of association) of haplotype A after stratification into groups carrying risk-associated and non-risk-associated *CARD15* variants for haplotype A. The statistical difference for haplotype D in the Crohn disease subgroup was determined by a 2- -2 contingency table and permutation testing.

METHODS

Study samples. Individuals with IBD were recruited by the clinical group through the Charité University Hospital (Berlin, Germany) and at the Department of Internal Medicine I, University Hospital Kiel, Germany. Diagnosis of IBD and subsequent classification into Crohn disease or ulcerative colitis was determined by standard diagnostic criteria[12,13] and has been described previously[3,5,13]. All individuals were of European descent. We carried out LD mapping in trios consisting of father, mother and child affected with IBD, in which one parent or neither parent was affected with Crohn disease or ulcerative colitis. These trios were identified for LD mapping and have been described[3]. For our confirmatory cohort, we extracted trios randomly from the multicase families used in our previous linkage studies[5] and supplemented this group with 92 additional trios recruited for this purpose. For case-control association, we compared 538 additional, independent individuals with IBD (singletons) with age- and sex-matched volunteers from the Kiel University blood donation program. All study participants gave informed, written consent. The recruitment protocols and study procedures were approved by the ethics committees of the Charité University Hospital, Berlin, Germany, and the Schleswig-Holstein University Hospital, Campus Kiel, Germany, respectively.

Microsatellite typing. In the first stage of microsatellite LD mapping, we genotyped 11 microsatellite markers (*D10S547*, *D10S548*, *D10S211*, *D10S611*, *D10S213*, *D10S1780*, *D10S220*, *D10S1790*, *D10S609*, *D10S201* and *D10S2470*) in 393 families with IBD (422 affected sibling pairs).

Information on primer sequence, allele size range, suggested amplification conditions and genetic position can be obtained from the Genethon and Marshfield databases (see URLs). Genotypes were generated at the University of California Los Angeles using PCR and fluorescence-labeled primers on an ABI 377 sequencer.

SNP discovery in *DLG5*. To identify all crucial SNPs in the coding sequence of *DLG5*, as well as exon-intron boundaries and the promoter region, we sequenced 47 individuals with IBD using an ABI 3700 automated sequencer as previously described[3]. The primers and probes for 33 SNPs discovered or verified by resequencing, and the sequences of the new SNPs, are given in **Supplementary Tables 3** and **4** online.

SNP genotyping. We selected SNP markers for the initial fine mapping experiment based on information available from the public databases. For analysis of *DLG5*, we used SNPs generated or verified in-house. We generated SNP genotypes using the TaqMan allelic discrimination method as previously described[3]. Taqman assays were from ABI.

Statistical analysis. We tested each marker for Hardy-Weinberg equilibrium in the control populations using a χ^2 test and then carried out genetic analyses at several levels. To confirm the association with Crohn disease, we first subjected each marker to single-locus tests for linkage and transmission disequilibrium testing (TDT) analysis followed by haplotype analysis as implemented in GENEHUNTER (Vs. 2.1; ref. 14). To assess significance of the TDT results for each marker, we did permutation tests using the same genotype data described previously[15,16]. In 10^5 permutations of the entire data set of 28 analyzed markers for *DLG5*, we observed a single χ^2 value greater than 9.91 4,635 times (empirical $P = 0.046$), and 874 simulations had two markers with a χ^2 value greater than 14.5 (empirical $P = 0.0087$). We calculated pairwise LD between each marker pair and between haplotype blocks as described[15,16]. For case-control analysis, we calculated χ^2 values using Fisher's exact test; we calculated genotype-based ORs using Fisher's contingency tables and tested association similarly. We calculated combined *P* values for determining the overall significance of the observed independent association findings as outlined[17].

Exon 1 identification. Because a BLAST analysis of the sequence from exon 1 as described[10] showed this sequence to be derived from human mitochondrial DNA, we concluded that this sequence probably arose as an artefact of RACE amplification. We used sequence from exon 2 instead to identify expressed-sequence tags from porcine and bovine genomes containing unique 5′ sequences (EMBL IDs BI402246, BM484383 and BI847653). These have high similarity and could be identified within the human contig containing *DLG5*. This new exon of at least 300 nucleotides in the 5′ untranslated region is located ~57 kb upstream of exon 2 of *DLG5*.

URLs. The Marshfield database is available at http://research.marshfieldclinic.org/genetics. The Genethon database is available at http://www.genethon.fr. The National Center for Biotechnology Information's SNP database is available at http://www.ncbi.nlm.nih.gov/SNP. The SNP Consortium website is available at http:///snp.cshl.org. The National Genome Research Network is available at http://www.ngfn.de.

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

ACKNOWLEDGMENTS

We thank all cooperating clinical centers, clinicians, families and individuals with IBD and the German Crohn's and Colitis Foundation (DCCV) for their participation; J. Papp for microsatellite genotyping; A. Andersson, B. Petersen, A. Dellsén, T. Engler, M. van Giezen, Å. Jägervall, T. Kim, N. Tepe and T. Wesse for technical assistance; M. Will and T. Lu for bioinformatics support; F. Friedrichs for assistance in statistical analysis; O. Bengtsson and K. Forsman-Semb for discussions; and M. J. Daly for the Haploview software. This study was supported by the 5th Framework Program of the European Commission and the Federal Ministry of Science and Education through the National Genome Research Network and the Competence Network "Inflammatory Bowel Disease", and by a coordinated research group of the German Research Foundation (DFGFOR423).

COMPETING INTERESTS STATEMENT

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

Received 4 February; accepted 12 March 2004
Published online at http://www.nature.com/naturegenetics/

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. 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).
4. Bonen, D.K. & Cho, J.H. The genetics of inflammatory bowel disease. *Gastroenterology* **124**, 521–536 (2003).
5. Hampe, J. *et al.* A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *Am. J. Hum. Genet.* **64**, 808–816 (1999).
6. Lu, G. *et al.* Inflammatory modulation of calcium-activated potassium channels in canine colonic circular smooth muscle cells. *Gastroenterology* **116**, 884–892 (1999).
7. Nakamura, H. *et al.* Identification of a novel human homolog of the Drosophila dlg, P-dlg, specifically expressed in the gland tissues and interacting with p55. *FEBS Lett.* **433**, 63–67 (1998).
8. Humbert, P., Russell, S. & Richardson, H. Dlg, Scribble and Lgl in cell polarity, cell proliferation and cancer. *Bioessays* **25**, 542–553 (2003).
9. Wakabayashi, M. *et al.* Interaction of lp-dlg/KIAA0583, a membrane-associated guanylate kinase family protein, with vinexin and beta-catenin at sites of cell-cell contact. *J. Biol. Chem.* **278**, 21709–21714 (2003).
10. Shah, G. *et al.* The cloning, genomic organization and tissue expression profile of the human DLG5 gene. *BMC Genomics* **3**, 6 (2002).
11. Purmonen, S. *et al.* HDLG5/KIAA0583, encoding a MAGUK-family protein, is a primary progesterone target gene in breast cancer cells. *Int. J. Cancer* **102**, 1–6 (2002).
12. Lennard-Jones, J.E. Classification of inflammatory bowel disease. *Scand. J. Gastroenterol.* **170** (Suppl), 2–6 (1989).
13. Truelove, S.C. & Pena A.S. Course and prognosis of Crohn's disease. *Gut* **17**, 192–201 (1976).
14. Markianos, K., Daly, M.J. & Kruglyak, L. Efficient multipoint linkage analysis through reduction of inheritance space. *Am. J. Hum. Genet.* **68**, 963–977 (2001).
15. Daly, M.J. *et al.* High-resolution haplotype structure in the human genome. *Nat. Genet.* **29**, 229–232 (2001).
16. Rioux, J.D. *et al.* Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat. Genet.* **29**, 223–228 (2001).
17. Fisher R.A. *Statistical Methods for Research Workers* 10th edn. (Oliver and Boyd, London, 1946).