

Sarcoidosis is associated with a truncating splice site mutation in *BTNL2*

Ruta Valentonyte^{1,11}, Jochen Hampe^{1,2,11}, Klaus Huse³, Philip Rosenstiel¹, Mario Albrecht⁴, Annette Stenzel^{1,2}, Marion Nagy⁵, Karoline I Gaede^{6,7}, Andre Franke², Robert Haesler¹, Andreas Koch², Thomas Lengauer⁴, Dirk Seeger⁸, Norbert Reiling⁶, Stefan Ehlers⁶, Eberhard Schwinger⁹, Matthias Platzer³, Michael Krawczak¹⁰, Joachim Müller-Quernheim⁷, Manfred Schürmann⁹ & Stefan Schreiber^{1,2}

Sarcoidosis is a polygenic immune disorder with predominant manifestation in the lung. Genome-wide linkage analysis previously indicated that the extended major histocompatibility locus on chromosome 6p was linked to susceptibility to sarcoidosis. Here, we carried out a systematic three-stage SNP scan of 16.4 Mb on chromosome 6p21 in as many as 947 independent cases of familial and sporadic sarcoidosis and found that a 15-kb segment of the gene butyrophilin-like 2 (*BTNL2*) was associated with the disease. The primary disease-associated variant (rs2076530; $P_{TDT} = 3 \times 10^{-6}$, $P_{\text{case-control}} = 1.1 \times 10^{-8}$; replication $P_{TDT} = 0.0018$, $P_{\text{case-control}} = 1.8 \times 10^{-6}$) represents a risk factor that is independent of variation in *HLA-DRB1*. *BTNL2* is a member of the immunoglobulin superfamily and has been implicated as a costimulatory molecule involved in T-cell activation on the basis of its homology to B7-1. The G→A transition constituting rs2076530 leads to the use of a cryptic splice site located 4 bp upstream of the affected wild-type donor site. Transcripts of the risk-associated allele have a premature stop in the spliced mRNA. The resulting protein lacks the C-terminal IgC domain and transmembrane helix, thereby disrupting the membrane localization of the protein, as shown in experiments using green fluorescent protein and V5 fusion proteins.

Sarcoidosis is a multisystemic immune disorder characterized by noncaseating granulomas and an exaggerated cellular immune response due to increased inflammatory activity of macrophages and CD4 helper T cells^{1,2}. Typical sites of disease manifestation include the lung, lymph nodes, eyes and skin. The clinical presentation of the disease varies from subclinical inflammation with slowly progressing pulmonary fibrosis to an acute inflammatory syndrome (Loefgren syndrome). Although spontaneous resolution is observed in more than 50% of cases, the disease can also take a recalcitrant course with ultimate chronic respiratory failure. The prevalence of sarcoidosis ranges from 10 to 14 per 100,000 in central Europe and European Americans to 64 per 100,000 in Sweden. Familial clustering of the disease has been observed in several populations; estimates of the relative risk to first-degree relatives vary between 2.8 and 18 (ref. 3).

In a recent genome screen of 63 families with sarcoidosis⁴, we identified linkage of the disease to chromosome 6p21. Several HLA markers have been investigated for association with the disease^{5–8} but initial results conflicted; *HLA-DRB1* was recently identified as a

consistent risk factor for sarcoidosis^{6,8}. Nevertheless, the population attributable risk of mutations in this gene is small (~10% for individuals of European descent) and, in view of the strong linkage signal obtained for 6p21, additional risk factors probably exist in the region. Some candidate genes have already been investigated^{9,10}, but to search systematically for additional disease susceptibility variants, we carried out systematic three-stage SNP fine-mapping of 6p21. We recruited a combination of extended families, trios and single individuals with sarcoidosis and carried out family-based (transmission disequilibrium test, TDT) and population-based (case-control) association analyses in parallel. We used consistency between TDT results and case-control results as a pragmatic guide to the trustworthiness of individual association signals.

RESULTS

Stage-one SNP scan

In the initial stage-one SNP scan, we genotyped individuals from the 'basic' sample (Table 1 and Supplementary Table 1 online) for 69 SNPs

¹Institute for Clinical Molecular Biology and ²Kiel Center of the German National Genotyping Platform, Christian-Albrechts-University Kiel, Universitätsklinikum Schleswig-Holstein, Campus Kiel, Schittenhelmstr. 12, 24105 Kiel, Germany. ³Institute for Molecular Biotechnology, Beutenbergstraße 11, 07745 Jena, Germany. ⁴Max-Planck-Institute for Informatics, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany. ⁵Institute of Forensic Medicine, Charite University Hospital, Hannoversche Straße 6, 10115 Berlin, Germany. ⁶Research Center Borstel, Parkallee 1-40, 23845 Borstel, Germany. ⁷Department of Pneumology, University of Freiburg, Killian-Str.5, 79106 Freiburg, Germany. ⁸Conaris Research Institute AG, Schauenburgerstr. 116, 24118 Kiel, Germany. ⁹Institute of Human Genetics, University of Lübeck; Ratzeburger Allee 160, 23538 Lübeck, Germany. ¹⁰Institute of Medical Statistics and Informatics, Christian-Albrechts-University Kiel, Universitätsklinikum Schleswig-Holstein, Campus Kiel, 24105 Kiel, Germany. ¹¹These authors contributed equally to this work. Correspondence should be addressed to J.H. (j.hampe@mucosa.de).

Table 1 Individuals with sarcoidosis used in association analysis

Sampling unit	Sample			
	Basic	Extension	Replication	Full
Affected sibling pairs (2 siblings)	70	9	10	89
Affected sibling pairs (3 siblings)	8	–	–	8
Complex families	11	5	–	16
Trios (complete)	153	72	203	428
Trios (one parent missing)	10	79	62	151
Single cases	13	65	177	255
Total number of independent cases	265	230	452	947
Total number of patients (male/female)	372 (163/209)	248 (123/125)	462 (175/287)	1,082 (461/621)
Average age at diagnosis (s.d.) in years	34.7 (10.3)	37.2 (11.4)	37.3 (11.3)	36.4 (11.1)
Experiment	Initial SNP screen SNP fine mapping		Replication	Differentiation of <i>HLA-DRB1</i> – <i>BTNL2</i> effects

The subphenotype composition of these samples is given in **Supplementary Table 1**. For case-control analysis, one case individual was extracted randomly from each family. The total number of independent cases equals the sum of the sampling units in the upper part. A group of 447 regionally, age- and sex-matched healthy individuals of German descent was used as a control in the initial SNP screen and SNP fine mapping experiments. For the replication experiment, an independent sample of 879 age- and sex-matched control individuals was used.

in an interval surrounding the position with the maximum nonparametric lod score in the original genome screen⁴ (**Fig. 1a**; National Center for Biotechnology Information (NCBI) release 32 map coordinates 27.8–44.1 Mb). Defined by a decrease in nonparametric lod score by 1.5 units or less, this interval represented an approximate 99% confidence region for linkage. To achieve sufficient power to identify loci associated with sarcoidosis, we adopted a *P* value of 0.001 as a threshold for significance in both the TDT and the case-control analyses of three-locus SNP haplotypes (**Fig. 1b,c**). This thresh-

old corresponds to a global significance level of 5% following Bonferroni correction for the number of SNPs investigated in the region. Applying this criterion, only the *BTNL2* subregion gave a consistent association signal. A second peak in the *MICB* subregion yielded a *P* value <0.001 in the case-control analysis but not in the TDT analysis (*P* = 0.3). We investigated this second putative region of association further by genotyping markers rs1063635, rs3134900, rs3130062, rs1041981 and rs1799964 in the ‘extension’ sample (**Table 1**), together with *BTNL2* markers rs2076523, hCV2455668, hCV2455646 and rs7192. Logistic regression analysis of genotypes, applying likelihood ratio-based forward inclusion at the 5% significance level, resulted in a model that included only two markers from the *BTNL2* region (rs2076523 and rs7192) and no markers from the *MICB* region. Furthermore, we observed substantial long-range linkage disequilibrium (LD) between markers in the *BTNL2* and *MICB* regions, with *D'* values as high as 0.94. Therefore, we confined all subsequent mapping efforts to the *BTNL2* region.

BTNL2 gene region and stage-two SNP scan

In stage two, we genotyped 48 SNPs from the 440-kb region of association around *BTNL2* (NCBI release 32 map coordinates 32.177–32.616 Mb) in the ‘basic’ and ‘extension’ samples. We identified SNPs from the Applied Biosystems myScience database, the Applied Biosystems SNP-Browser, dbSNP or by direct genomic sequencing of 47 individuals with sarcoidosis. Mutation screening of these individuals covered all exon and adjacent intron sequences of the revised *BTNL2* model. We also identified SNPs in the genes *C6orf10* and *HLA-DRA* (including the promoter, which had not previously been analyzed for SNPs) and in some noncoding sequence telomeric of

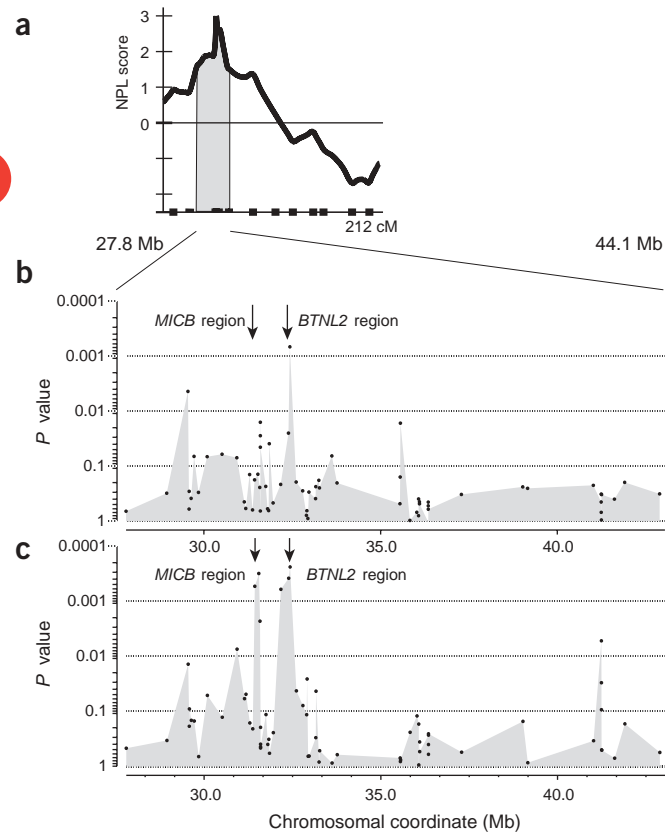


Figure 1 Graphical representation of the stage I SNP screen on chromosome 6p21. The genomic interval corresponds to the genes from *RFP* to *CAPN11*. (a) Nonparametric lod (NPL) score curve from the original genome scan⁴. An approximate 99% c.i. for linkage is marked in gray. (b) Results of a three-locus haplotype TDT using TRANSMIT³⁸. (c) Results of a three-locus haplotype case-control analysis using HAPMAX⁴⁰. In both panels, the *P* value is plotted against the physical map position (NCBI coordinates) of the most telomeric marker of each haplotype. Using a cut-off of *P* = 0.001, two potential lead regions (*MICB* and *BTNL2* regions) were identified (marked by arrows).

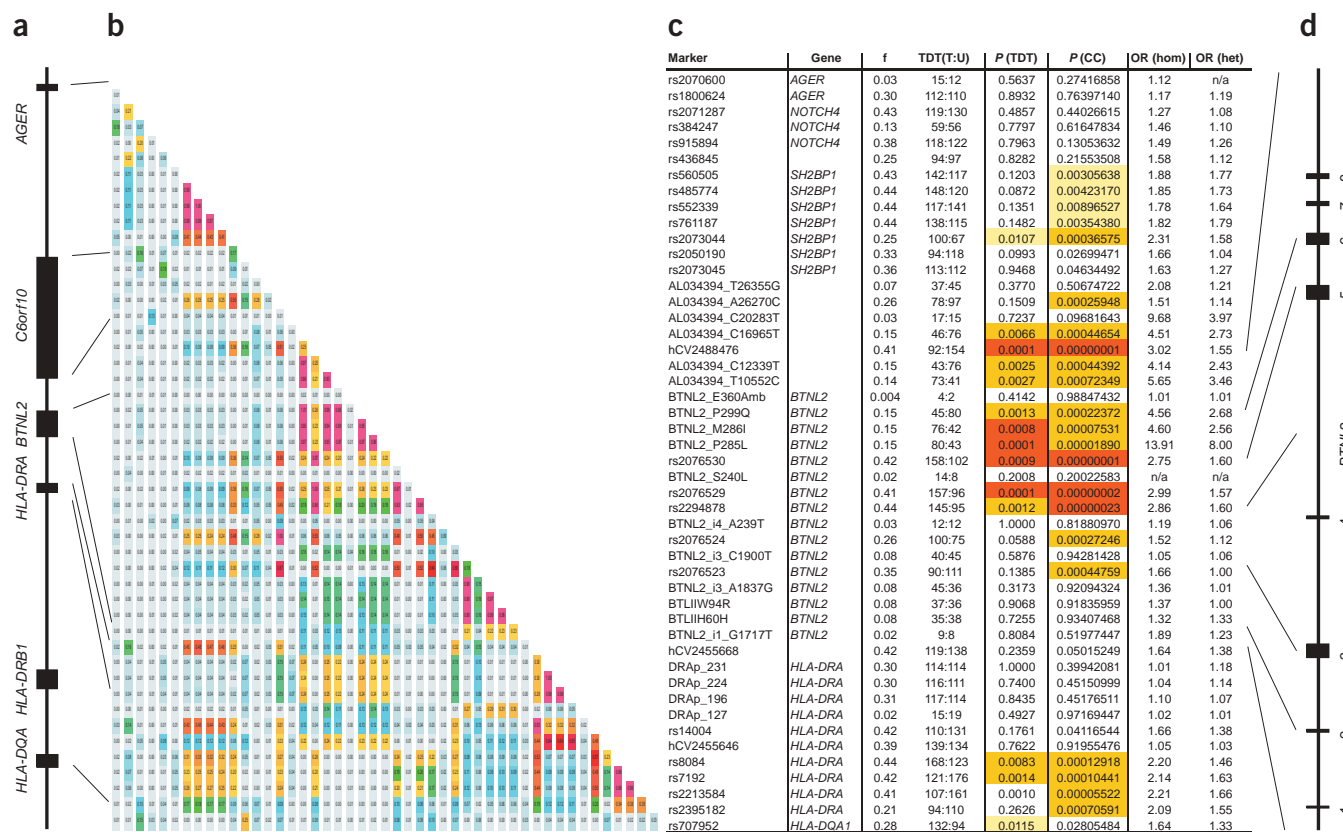


Figure 2 Graphical representation of the stage II SNP screen of the combined 'basic' and 'extension' samples. **(a)** Gene map of the analyzed 440-kb region on human chromosome 6p21. **(b)** Equidistance plot of pairwise LD (r^2) between all markers in the analyzed region. Fields are color-coded according to LD strength (blue, $r^2 = 0.0$; red, $r^2 = 1.0$). **(c)** Results of disease-association tests. f, allele frequency of the minor allele in controls; TDT (T:U), numbers of transmitted versus untransmitted alleles; P (TDT), single-point P value reported by GENEHUNTER 2.1, color-coded as white ($P \geq 0.01$), yellow ($0.01 \leq P < 0.05$), orange ($0.001 \leq P < 0.01$) or red ($P < 0.001$); P (CC), P value from a genotype-based χ^2 test of case-control data, color-coded as white ($P \geq 0.01$), yellow ($0.001 \leq P < 0.01$), orange ($10^{-6} \leq P < 0.001$) or red ($P < 10^{-6}$); OR (hom) and OR (het), OR for homozygous and heterozygous risk allele carriers, respectively. **(d)** Revised exon-intron structure of *BTNL2*.

BTNL2. We obtained consistent results in the two tests for a region of ~15 kb at the 3' end of *BTNL2* (markers hCV2488476–rs2294878). P values for this region were two to three orders of magnitude smaller than those for the other regions (Fig. 2).

Functional and structural effects of *BTNL2* mutations

We examined the structure of *BTNL2*, as logged in public sequence databases, by cDNA cloning. When the gene was first reported¹¹, no contiguous cDNA amplification was achieved for exons 1–4 and 5–6, suggesting that two different genes exist at this locus. We identified three additional exons (one 5' and two 3') and excluded original exon 3 from the gene model. Overlapping RT-PCR confirmed the presence of a single transcript and, thereby, corroborated the validity of the revised gene model.

The 15-kb segment most strongly associated with sarcoidosis contains four functional SNPs (amino acid substitutions P299Q, M286I and P285L and rs2076530), which together yielded P values of 1.9×10^{-7} in a four-locus TDT and 8.2×10^{-7} in a case-control haplotype analysis (single-point results are given in Fig. 2c). Most of the association signal was due to rs2076530, with a P value of 0.04 for the inclusion of the other three markers in a genotype-based logistic regression analysis of the case-control sample. The neighboring non-functional SNPs rs2076529 and hCV2488476 are in strong pairwise LD with rs2076530 ($D' = 1.0$) and, therefore, provide efficient tagging

SNPs for the risk haplotype, which may contain additional, albeit rare, susceptibility variants. We verified the strong association between sarcoidosis and rs2076530 in an independent stage-three 'replication' sample (Table 1) of German cases and 876 independent German controls (TDT: $\chi^2 = 9.7$, degrees of freedom (d.f.) = 1, $P = 0.0018$; case-control $\chi^2 = 26.4$, d.f. = 1, $P = 1.8 \times 10^{-6}$).

According to the previously reported gene model¹¹, rs2076530 would be predicted to cause an amino acid exchange, but in reality, the SNP is located at position -1 of a donor splice site (Fig. 3). A 4-bp cDNA deletion observed in cloned RT-PCR products from the gene model verification can be explained by the fact that rs2076530 alters the anatomy of the splice site. Guanine is present at position -1 of 77% of donor sites^{12,13}, and so the G→A transition of rs2076530 results in the recruitment of an alternative splice site located 4 bp upstream. We confirmed genotype-specific splicing by RT-PCR of DNA-cDNA pairs from lymphoblastoid cell lines and peripheral blood samples (Table 2) and in bronchoalveolar lavage (BAL) samples from affected individuals (Fig. 4). The loss of four bases from the cDNA transcribed from the A allele causes a frameshift and a premature stop in the downstream exon. In the corresponding protein product, the 118 C-terminal residues in the nontruncated protein are replaced by five different amino acids.

We investigated the effects of the disease-associated variant on the structure of the *BTNL2* protein *in silico*. The *BTN* (butyrophilin-like)

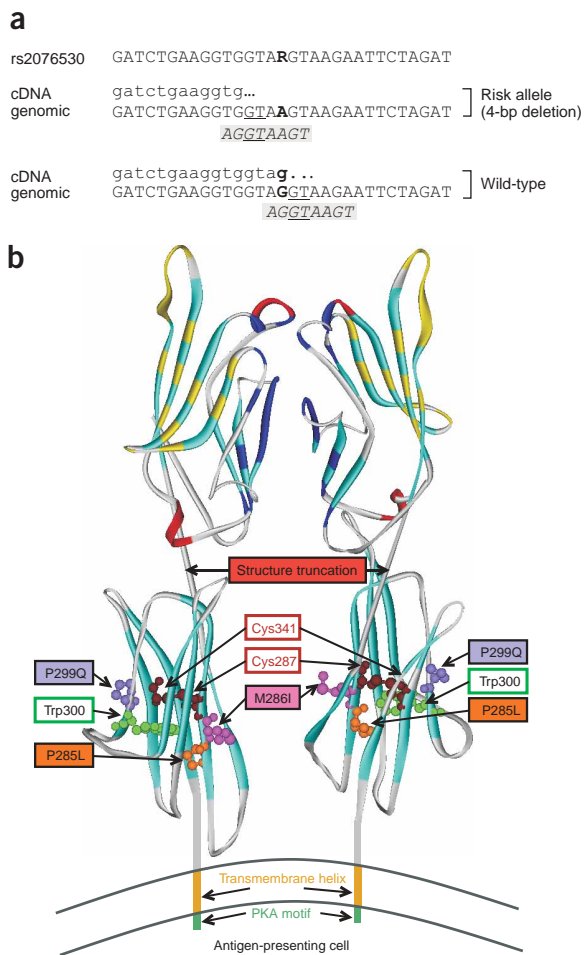


Figure 3 Effect of rs2076530 on splicing and protein structure. **(a)** Splicing at SNP rs2076530 in exon 5 of *BTNL2*. Top, genomic sequence around rs2076530. Center, splicing in the presence of the A allele. Bottom, splicing in the presence of the G allele. The GT of the active donor site is underlined. The consensus splice sequence^{12,13} is highlighted in gray. **(b)** Computationally derived 3D structure model of the second IgV domain (upper half) and the following IgC domain (lower half) of the *BTNL2* homodimer. The model was created using the B7-1 structure as a template. The location of the C-terminal truncation and of sequence variants P285L, M286I and P299Q are noted. A stabilizing disulfide bond between residues Cys287 and Cys341, which are located adjacent to variants P285L and M286I, is highlighted in brown. A highly conserved Trp300 located close to variant P285L and the disulfide bond are highlighted in light green. The putative binding site of *BTNL2* to an unknown receptor in the IgV domain is marked in yellow. This site was deduced from the location of residues interacting with the B7-1 receptor CTLA-4. Residues participating in the dimerization interface of the IgV domain are shown in dark blue. The predicted transmembrane helix and a potential PKA phosphorylation site at the C terminus of *BTNL2* are also indicated.

shows a truncation of the protein structure at the site of the premature stop codon introduced by rs2076530, just before the start of the IgC domain and the attached transmembrane helix. Substitutions P285L, M286I and P299Q occur in β strands on the protein surface of the IgC domain²¹ and map to residues Ile126, Ile127 and Ser140 of B7-1. Notably, all variants represent changes to amino acids whose side chains are physicochemically more similar to those of the corresponding wild-type B7-1 residues in its IgC domain (**Supplementary Note** online). Several residues involved in binding of the IgV domain of B7-1 to its receptor, CTLA-4 (ref. 19), are conserved in *BTNL2* (**Fig. 3b** and **Supplementary Fig. 1** online). This observation suggests that *BTNL2* may have a receptor-binding site similar to that of B7-1.

To evaluate the expression of the *BTNL2* transcript in target cells, we carried out RT-PCR analysis in a tissue panel, in monocyte-derived macrophages²² and in BAL cells from individuals with sarcoidosis and from controls. Because *BTNL2* expression was generally low, we used nested RT-PCR to detect sufficient amounts of transcript in tissue samples and samples from individuals. We observed expression in various tissues including leucocytes and thymus (**Fig. 4a**). We also detected transcripts in BAL cell samples from individuals with sarcoidosis of all *BTNL2* rs2076530 genotypes but not in BAL cells from control individuals or in normal lung (**Fig. 4b**). The expression pattern observed was not due to a different cellular composition of BAL samples between individuals with sarcoidosis and controls. We confirmed the genotype-specific splice pattern by cloning and sequencing the PCR products obtained from the BAL samples (**Fig. 4b**). Expression of *BTNL2* could be induced by the prototypic inflammatory cytokines TNF- α and IL1- β in the myelomonocytic cell line

genes are members of the immunoglobulin superfamily. The *BTN* genes encode a small, variable number of consecutive immunoglobulin-like ectodomains, with an N-terminal signal peptide and a C-terminal transmembrane helix anchoring the protein to the membrane of antigen-presenting cells^{14,15}. The protein product of the revised *BTNL2* gene contains two N-terminal homologous immunoglobulin-like variable (IgV) domains with a sequence identity of 46% and one C-terminal immunoglobulin-like constant (IgC) domain (**Supplementary Fig. 1** online). The cleavage site of the putative N-terminal signal peptide^{14,15} is predicted to be located between residues 26 and 27, and a hydrophathy index plot (**Supplementary Fig. 2** online), together with the results of transmembrane prediction servers¹⁶, suggests that there is a single C-terminal transmembrane helix following the IgC domain. Such a membrane-anchoring helix is also found at the C termini of IgC domains of major histocompatibility complex (MHC) antigens and other costimulatory receptors. A distant homology of butyrophilin-like proteins to MHC antigens and B7-1/B7-2 (CD80/CD86) costimulatory receptors^{17–20} has previously been reported¹⁴. The IgV and IgC domain sequence of the crystal structure of the B7-1 homodimer^{17,19,21} is detected, with a significant *E* value of 3×10^{-7} , by a BLAST search of PDB using the second IgV and the adjacent IgC domain of *BTNL2*. We modeled the three-dimensional (3D) structure of both *BTNL2* domains based on a manually curated sequence-structure alignment of *BTNL2* and the B7-1 template, characterized by an amino acid sequence identity of 24%. The resulting 3D model of *BTNL2* (**Fig. 3b**)

Table 2 Verification of the splicing effect of rs2076530

Source	rs2076530 genotype	Samples (#)	Clones (#)	BTNL2-L	BTNL2-S
Cell lines	GG	4	64	64	0
	AG	5	98	22	76
	AA	3	131	0	131
Leukocytes	GG	1	36	36	0
	AG	3	36	9	27
	AA	3	35	0	35

Matching pairs of cDNA and DNA from lymphoblastoid cell lines (upper part) and from the peripheral blood of normal controls (lower part) were genotyped and the *BTNL2* splice forms were evaluated by RT-PCR, cloning and sequencing of PCR products.

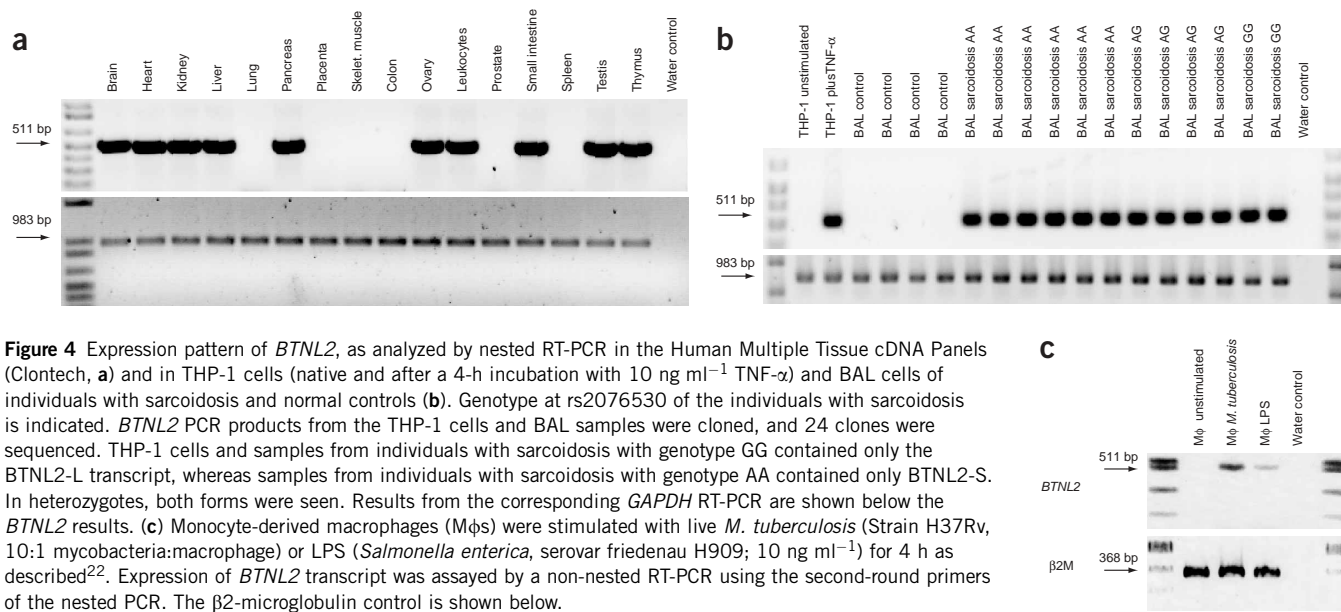


Figure 4 Expression pattern of *BTNL2*, as analyzed by nested RT-PCR in the Human Multiple Tissue cDNA Panels (Clontech, **a**) and in THP-1 cells (native and after a 4-h incubation with 10 ng ml⁻¹ TNF- α) and BAL cells of individuals with sarcoidosis and normal controls (**b**). Genotype at rs2076530 of the individuals with sarcoidosis is indicated. *BTNL2* PCR products from the THP-1 cells and BAL samples were cloned, and 24 clones were sequenced. THP-1 cells and samples from individuals with sarcoidosis with genotype AA contained only the *BTNL2*-L transcript, whereas samples from individuals with sarcoidosis with genotype GG contained only the *BTNL2*-S transcript. In heterozygotes, both forms were seen. Results from the corresponding *GAPDH* RT-PCR are shown below the *BTNL2* results. (**c**) Monocyte-derived macrophages (M ϕ s) were stimulated with live *M. tuberculosis* (Strain H37Rv, 10:1 mycobacteria:macrophage) or LPS (*Salmonella enterica*, serovar friedenaue H909; 10 ng ml⁻¹) for 4 h as described²². Expression of *BTNL2* transcript was assayed by a non-nested RT-PCR using the second-round primers of the nested PCR. The β 2-microglobulin control is shown below.

THP-1 (**Fig. 4b** and real-time PCR data in **Supplementary Fig. 3** online). *Mycobacterium tuberculosis* and lipopolysaccharide (LPS) induced *BTNL2* in monocyte-derived macrophages, as shown by non-nested PCR analysis (**Fig. 4c**)²².

We cloned *BTNL2* transcripts from both the deletion allele (*BTNL2*-S) and the wild-type allele (*BTNL2*-L) into a mammalian expression vector with a C-terminal fusion green fluorescent protein. In transfection experiments of HeLa and HEK cells (data not shown), these constructs showed different subcellular localization. In contrast to *BTNL2*-L, which showed membrane localization, the mutant *BTNL2*-S protein was contained in cytoplasmic vesicular structures (**Fig. 5a,b**). The high expression levels characteristic of transfection experiments do not reflect the physiological concentration of the protein in immunoregulatory cells, where nested PCR was necessary to detect the transcript. Therefore, the abundant vesicle formation and high expression level of *BTNL2*-L in all membranes probably resulted from *in vitro* overexpression. We confirmed the differential localization of the allelic protein variants by western blotting of subcellular fractions of HeLa cells transiently transfected with V5-tagged *BTNL2*-L and *BTNL2*-S constructs (**Fig. 5c**).

Independence of *BTNL2* and *HLA* effects

The truncating allele, A, of rs2076530 predisposes for sarcoidosis (**Table 3**). The marginal odds ratio (OR) estimates were 2.75 (95% confidence interval (c.i.) = 1.86–4.05) for AA versus GG genotypes and 1.60 (95% c.i. = 1.10–2.32) for AG versus GG genotypes. The difference between the two values was highly significant ($\chi^2 = 16.886$, d.f. = 1, $P = 3.9 \times 10^{-5}$), arguing in favor of a multiplicative effect on risk by the truncating allele. Nevertheless, *BTNL2* is located close (~200 kb) to *HLA-DRB1*, which, together with other *HLA* loci (*HLA-DQB1* and *HLA-DPB1*), is implicated in the etiology of sarcoidosis^{5,6,8}. Group typing of *HLA-DRB1* alleles in our three samples²³ confirmed the disease association of allele *HLA-DRB1**11 (**Supplementary Table 2** online). When *HLA-DRB1* alleles with ORs > 1.3 were jointly coded as R (risk) and alleles with ORs < 1.3 were coded as N (nonsignificant), marginal OR estimates were 2.19 (95% c.i. = 1.43–3.36) for RR versus NN genotypes and 1.70 (95% c.i. = 1.32–2.19) for NR versus NN genotypes. Because the difference between the two ORs was not statistically significant ($\chi^2 = 1.192$, d.f. = 1, $P = 0.27$), we considered the R allele to be dominant and grouped genotypes RR and NR together as high risk genotypes in subsequent analyses.

Table 3 Association analysis between sarcoidosis and joint *HLA-DRB1 BTNL2* genotype

<i>BTNL2</i>	<i>HLA-DRB1</i>							
	NN		NR		RR		Total	
	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases
AA	43	99	66	223	25	97	134	419
AG	127	188	84	190	10	25	221	403
GG	59	68	12	14	1	0	72	82
Total	229	355	162	427	36	122	427	904
OR ^a	1.56 (1.00–2.43)		1.54 (1.08–2.18)				1.54 ^b (1.19–2.01)	

^aOR of *BTNL2* genotypes AA versus AG, stratified by *HLA-DRB1* genotype, with 95% c.i. given in parentheses. ORs of genotypes AA versus GG are 2.00 (95% c.i. = 1.18–3.40) for NN and 3.27 (95% c.i. = 1.39–7.68) for NR or RR. ORs of genotypes AG versus GG are 1.28 (95% c.i. = 0.83–1.99) for NN and 2.12 (95% c.i. = 0.90–5.02) for NR or RR. ^bMantel-Haenszel estimate of common OR.

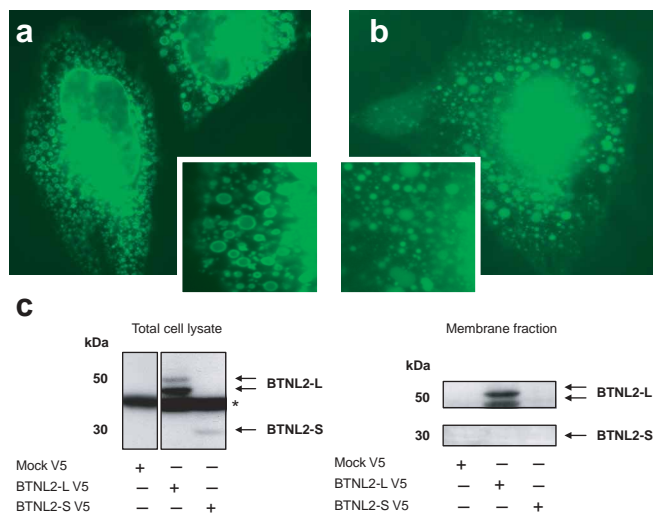


Figure 5 Subcellular localization of the long and truncated forms of BTNL2 protein. (a, b) Fluorescence images of HeLa cells transiently transfected with allelic BTNL2-GFP fusion proteins. (a) BTNL-L: long, untruncated protein (G allele at rs2076530). (b) BTNL-S: short, truncated protein (A allele with 4-bp deletion). The truncated protein is not localized in membranes but rather in vesicular structures in the cytoplasm. The long, untruncated protein shows a ubiquitous membrane distribution in the nuclear envelope, in vesicular membranes and a fine granular pattern at the plasma membrane. (c) Western blots (using antibody to Flag-V5) of total cell lysates (right top panel) and Triton-soluble membrane fractions (left bottom panel) of HeLa cells transiently transfected with BTNL2-V5 constructs. Transfection with vectors containing mock V5 tag, BTNL2-L and BTNL2-S is indicated below each blot. The asterisk indicates an unspecific band detected by the V5 antibody.

In a single-locus, allele-based test of association, *HLA-DRB1*, *HLA-DQB1* and *HLA-DPB1* yielded *P* values of 2.4×10^{-5} , 1.3×10^{-3} and 2.9×10^{-5} , respectively, compared with 1.1×10^{-8} for rs2076530. To assess the extent of possible confounding of the rs2076530 association by the *HLA* loci, we repeated the disease association analysis, stratifying the data each time according to genotype at one of the *HLA* loci. When stratified by *HLA-DRB1* genotype, the OR estimates of rs2076530 genotypes AA versus AG were virtually identical (1.54 for high-risk, 1.56 for NN). The number of individuals carrying *HLA-DRB1* high-risk genotypes but lacking rs2076530 risk allele A was too small to allow for a statistically robust comparison of *HLA-DRB1*-stratified ORs involving *BTNL2* genotype GG (i.e., AA versus GG and AG versus GG). Although the respective estimates seemed to be higher for high-risk than for NN genotypes, the confidence intervals associated with these values overlapped widely (Table 3).

DISCUSSION

We carried out a three-stage SNP mapping experiment on chromosome 6p21 to localize genetic variants that predispose to sarcoidosis. To interpret the association results, we used logistic regression analysis as a formal tool to differentiate primary from secondary effects. This analysis resulted in a model that included only two markers from the *BTNL2* region (rs2076523 and rs7192) and no markers from the *MICB* region. In addition, we used the consistency of TDT and case-control results as a pragmatic guide to judge the credibility of association signals, a procedure that also implicated the *BTNL2* region. The observed discordance between the significance levels of TDT and case-control association tests probably reflects the different statistical

power of the two methods. Their efficiency depends on the specific regional LD structure, the population of origin, the sample composition and the mode of inheritance of the risk factor²⁴. Long-range LD has also been a matter of concern in mapping studies of chromosome 6p. But integration of our data with those in a recently published study²⁵ uncovered no relevant long-range LD between the *BTNL2*-associated haplotype block and other *HLA* subregions (Supplementary Fig. 4 online).

We identified a disease-associated SNP (rs2076530) in the 3' region of *BTNL2* that exerted most, if not all, of its predisposing influence independently of the neighboring *HLA* loci. ORs of homozygosity versus heterozygosity for the truncating allele (A) were the same, irrespective of *HLA-DRB1* genotype, and differences involving homozygosity with respect to the functional allele (G) were insignificant throughout. Furthermore, the minor discrepancies observed were mainly due to an unusually small OR (0.93) of the high-risk *HLA-DRB1* genotype (NR and RR) versus genotype NN among *BTNL2* GG individuals. This genotype-specific lack of association may imply that *HLA-DRB1* is a risk factor for sarcoidosis only in the presence of the truncating *BTNL2* allele, a result that would be particularly important given the functional relatedness of costimulation and antigen presentation. Answering this and other questions about the possible interaction of specific *HLA-DRB1* alleles on different *BTNL2* backgrounds requires larger number of observations than are currently available, especially in individuals carrying genotype GG plus NR or RR. Alternatively, the small *HLA-DRB1*-related OR could be an underestimate, implying that the apparent modification of the *BTNL2* risk by *HLA-DRB1* genotype was merely due to sampling variation. Accordingly, no statistically significant confounding of the *BTNL2* risk by *HLA-DRB1* became apparent when we analyzed joint haplotypes for disease association (Supplementary Table 2 online) or when we assessed disease risks by genotype-based logistic regression. For the latter type of analysis, we modeled the presence or absence of the high-risk *HLA-DRB1* genotype (NR or RR) as a dichotomous variable and for *BTNL2*, considered the dosage of the A allele (i.e., 0, 1 or 2). Both genes were significant risk factors at the 1% significance level, but their interaction was not significant (*P* = 0.44). Finally, we carried out group genotyping for the *HLA-DPB1* and *HLA-DQB1* loci, resolving alleles with frequencies >5% in the German population (Supplementary Table 3 online). Joint analyses with *BTNL2* yielded results similar to *HLA-DRB1*.

The susceptibility allele of rs2076530 is characterized by ORs of 1.60 in heterozygotes and 2.75 in homozygotes. The SNP therefore has moderate influence on individual disease risk but makes a substantial contribution at the population level (population attributable risk = 23% for heterozygotes and homozygotes). This is notable given the profound impact of the SNP on BTNL2 protein structure and function, an apparent discrepancy that may be explained by a highly redundant system of costimulatory molecules, including the products of the butyrophilin gene cluster, B7-1 and other, as yet unknown members of the same functional class of proteins. The human *BTN* gene cluster was recently localized to chromosome 6p22 (ref. 14), ~6 Mb telomeric of *BTNL2*. For completeness, we also genotyped six markers from the *BTN* cluster in the combined 'basic' and 'extended' samples but found no evidence for disease association (data not shown).

One of the principal costimulatory pathways for naive T cells is initiated when CTLA-4 and CD28 engage B7 class molecules in the context of a primary signal, delivered through the TCR²⁶. The B7-1 protein (encoded by *CD80*) is a costimulatory molecule with known anti-inflammatory activity mediated through interaction with CTLA-4 (refs. 27,28). Deletion of the IgC domain in *CD80* had a

substantial proinflammatory effect in a mouse plasmid vaccination model²⁹. On the basis of amino acid homology and domain structure, we modeled BTNL2 on B7-1 (Fig. 4). Therefore, a potential T-cell downregulatory function of BTNL2 could be impaired by the truncating mutation rs2076530 and the concomitant loss of the IgC domain and transmembrane helix, possibly by the lack of membrane localization (Fig. 5). Inappropriate T-cell activation would also be compatible with the clinical immunology of sarcoidosis, characterized by a dysregulated T-helper cell activation³⁰. The *BTNL2* transcript is expressed in diseased tissue and is inducible by LPS (Fig. 4) in macrophages, supporting the idea that BTNL2 has a putative role as a costimulatory molecule, because its activation pattern resembles those of CD80 and CD86 (ref. 31).

The importance of costimulation and its dysfunction in many autoimmune disorders is increasingly recognized³². Our findings also highlight the importance of this pathway, as CTLA-4, a prototype of a putative BTNL2 receptor, is a risk factor for a wide range of autoimmune disorders, including type I diabetes, autoimmune hypothyroidism and Graves disease³². The exact functional role of BTNL2 in the costimulatory system remains to be elucidated. The identification of a major variant in *BTNL2* as a risk factor for sarcoidosis will give new impetus to the investigation of immunological costimulation and may provide a perspective for future therapeutic targeting of this disorder.

METHODS

Sample recruitment. Index individuals were contacted between 2000 and 2003 either through specialized hospitals and general practitioners or through the German Sarcoidosis Patient Organization (Deutsche Sarkoidose-Vereinigung e. V.). All diagnoses were made on the basis of the International Consensus Statement on Sarcoidosis^{33,34}. An overview of the individuals is given in Table 1 and Supplementary Table 1 online. We interviewed individuals and families with sarcoidosis by telephone about their individual and familial sarcoidosis history. All individuals (families and singletons) completed a questionnaire about the course of their disease. We contacted physicians to confirm the diagnosis and also verified the presence of sarcoidosis by biopsy in 78% of cases from the 'basic' sample, 90% of cases from the 'extension' sample and 88% of cases from the 'replication' sample. For the remaining cases, the clinical course and radiology and laboratory data were sufficiently consistent with the diagnosis of sarcoidosis^{33,34}. All participants were of German origin and gave written informed consent to participate in the study. The protocols were approved in writing by the institutional ethics (Ethikkommission des Universitätsklinikums Schleswig-Holstein, Kiel und Lübeck) and data protection authorities (Landesdatenschutzbeauftragter Schleswig-Holstein).

Genotyping and sequencing. We genotyped SNPs as previously described, using TaqMan (Applied Biosystems) technology on an automated platform and a LIMS system running on an IBM x-server^{35,36}. We sequenced genomic DNA using Applied Biosystems BigDye chemistry in accordance with the supplier's recommendations. We genotyped *HLA-DRB1* using group-specific primers in PCRs on 2.5 ng of genomic DNA as described²³. We subtyped the *HLA-DRB1* locus and group typed the *HLA-DQB1* and *HLA-DPB1* loci using the same methodology²³. All markers were tested for Hardy-Weinberg equilibrium in controls before inclusion in the association statistics.

cDNA cloning and RT-PCR. To verify the published *BTNL2* gene model, we searched for conserved exons in mouse and human genomic sequences (Twinscan gene predictions) and related them to predicted splice sites. We verified the predicted exons by RT-PCR in a pool of cDNA tissue samples (Human Multiple Tissue cDNA Panels I and II, Clontech) and by subsequent cloning and sequencing of PCR products.

We obtained BAL cell cDNA samples from individuals with sarcoidosis with overt disease and from controls in accordance with standard clinical and molecular procedures³⁷. BAL samples from patients and controls were matched for their cell count differentials (Supplementary Note online). For expression

analysis, we carried out nested PCR. All primer sequences are available on request. In both amplification rounds, we used the following thermoprofile: an initial denaturation step (3 min at 95 °C); 35 cycles of 30 s at 93 °C, 30 s at 59 °C and 30 s at 72 °C; and a final hold at 72 °C for 10 min. We amplified *GAPDH* in a non-nested PCR using the same thermoprofile.

Statistical analysis. We carried out single-locus and haplotype TDTs using the TRANSMIT³⁸ and GENEHUNTER³⁹ programs. In families with multiple affected individuals, one trio was randomly extracted for TDT analysis. We obtained haplotype frequency estimates among single cases using an implementation of the EM algorithm (HAPMAX)⁴⁰. We also used HAPMAX for significance testing of haplotype frequency differences, making use of the fact that twice the log-likelihood ratio between two nested data models approximately follows a χ^2 distribution with k degrees of freedom, where k is the difference in parameter number between the two models. We assessed the significance of associations with or between single-locus genotypes using χ^2 or Fisher's exact tests for 2×3 contingency tables. We carried out genotype-based logistic regression analysis with SPSS and SAS, coding individual SNP genotypes as categorical variables.

In silico protein analysis. We retrieved protein domain architectures from the Pfam database and examined them through the NCBI conserved domain search website. To predict the 3D structure of the *BTNL2* gene product, we explored all fold recognition servers available through the meta-server BioInfo.PL. On the basis of the prediction results, we constructed a sequence-structure alignment of BTNL2 to B7-1 (Supplementary Fig. 1 online) for the 3D modeling server WHAT IF, which returned a full-atom structure model of the second IgV and the following IgC domain of BTNL2. Additional details are given in Supplementary Methods online.

Generation of *BTNL2* expression constructs. We constructed full-length allelic variants of the *BTNL2* transcript (BTNL2-L and BTNL2-S) from clones generated during the gene model verification. We cloned fragments into the expression vector pEGFP-N2 (Clontech) and transfected HeLa cells with 1 µg of the respective constructs, in accordance with the manufacturer's protocol (Fugene 6, Roche). Empty pEGFP-N2 vector served as the positive control. We fixed cells in 4% paraformaldehyde in phosphate-buffered saline (pH 7.6) after 24 h and mounted them onto glass slides in 1:1 glycerol in phosphate-buffered saline. Fluorescence was visualized with an epifluorescence microscope (Axiohot, Zeiss).

We ligated full-length coding sequences for the long and short isoforms of BTNL2 into pCDNA3.1TOPO V5 vector (Invitrogen) to generate expression constructs for BTNL2-L and -S with a C-terminal V5 tag. We carried out cell fractionation as described⁴¹. We lysed transfected cells in hypotonic buffer and added sucrose to the homogenate at a final concentration of 250 mM. Centrifugation at 1,000g for 10 min at 4 °C removed nonhomogenated debris. We centrifuged the supernatant at 60,000g for 2 h at 4 °C. The supernatant from this step was considered the cytosolic fraction. We resuspended the pellet in lysis buffer containing 2% Triton X-100 and pelleted the Triton-insoluble fraction by centrifugation at 16,000g for 45 min at 4 °C. The remaining supernatant contained the Triton-soluble membrane protein fraction. To prepare total cell lysate, we pelleted cells and lysed them in denaturing lysis buffer as described⁴². We detected BTNL-2 V5 variants using a monoclonal antibody to Flag-V5 (Invitrogen) at 1:5,000 dilution.

URLS. All primer and probe sequences are available at <http://www.mucosa.de/sarcoidosis/btnl2/>. Assay onDemand genotyping assays are available from <http://myscience.appliedbiosystems.com/>. For splice site prediction, we used http://www.fruitfly.org/seq_tools/splice.html. HAPMAX is available at <http://www.uni-kiel.de/medinfo/mitarbeiter/krawczak/download/> and the WHAT IF server is available at <http://www.cmbi.kun.nl/WIWWWI/>.

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

ACKNOWLEDGMENTS

We thank all affected individuals, families, physicians, the German Sarcoidosis Patient Organization (Deutsche Sarkoidose-Vereinigung e. V.) and the contributing pulmonary specialist physicians for their cooperation; M. Albrecht, T. Wesse, P. Petersen, T. Henke, S. Kröger and I. Gorlich for technical help;

P. Croucher and S. Jenisch for discussions and providing reagents; C. Manaster for database support; and H. Brade for providing the LPS. This study was supported by the German National Genome Research Network, the German Human Genome Project, the German Research Council, the BioSapiens NoE of the EU, the POPGEN population project, Mucosaimmunologie Forschungsgesellschaft, Applied Biosystems and a SUR grant from IBM.

AUTHORS' CONTRIBUTIONS

R.V. helped to develop the SNP, *HLA-DRB1*, *HLA-DQB1* and *HLA-DPBI* assays and carried out genotyping, mutation detection and basic data analysis. J.H. prepared the manuscript and performed the data analysis, experimental planning, strategy and method development. K.H. and M.P. carried out cDNA cloning and splice assays. P.R. and D.S. carried out expression analysis and transfection experiments. A.S. and A.K. helped to develop the SNP assay. M.N.: helped to design the *HLA-DRB1*, *HLA-DQB1* and *HLA-DPBI* assays. A.F. carried out genotyping of *HLA-DRB1*, *HLA-DQB1* and *HLA-DPBI*. K.I.G. carried out BAL cDNA work. R.H. carried out expression analysis. M.K. helped with data analysis and preparation of the manuscript. M.A. and T.L. helped with preparation of the manuscript and *in silico* analysis. E.S., M.S. and J.M.Q. helped with experimental design, patient recruitment, clinical characterization and preparation of the manuscript. N.R. and S.E. carried out induction experiments in monocytes. S.S. helped with experimental strategy and preparation of the manuscript.

COMPETING INTERESTS STATEMENT

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

Received 18 June 2004; accepted 12 January 2005

Published online at <http://www.nature.com/naturegenetics/>

- Newman, L.S., Rose, C.S. & Maier, L.A. Sarcoidosis. *N. Engl. J. Med.* **336**, 1224–1234 (1997).
- Ziegenhagen, M. & Müller-Quernheim, J. The cytokine network in sarcoidosis and its clinical relevance. *J. Intern. Med.* **253**, 18–30 (2003).
- Rybicki, B.A. *et al.* Familial aggregation of sarcoidosis. A case-control etiologic study of sarcoidosis (ACCESS). *Am. J. Respir. Crit. Care Med.* **164**, 2085–2091 (2001).
- Schurmann, M. *et al.* Results from a genome-wide search for predisposing genes in sarcoidosis. *Am. J. Respir. Crit. Care Med.* **164**, 840–846 (2001).
- Sato, H. *et al.* HLA-DQB1*0201: a marker for good prognosis in British and Dutch patients with sarcoidosis. *Am. J. Respir. Cell. Mol. Biol.* **27**, 406–412 (2002).
- Foley, P.J. *et al.* Human leukocyte antigen-DRB1 position 11 residues are a common protective marker for sarcoidosis. *Am. J. Respir. Cell. Mol. Biol.* **25**, 272–277 (2001).
- Rybicki, B.A. *et al.* The major histocompatibility complex gene region and sarcoidosis susceptibility in african americans. *Am. J. Respir. Crit. Care Med.* **167**, 444–449 (2003).
- Rossmann, M.D. *et al.* HLA-DRB1*1101: a significant risk factor for sarcoidosis in blacks and whites. *Am. J. Hum. Genet.* **73**, 720–735 (2003).
- Grutters, J.C. *et al.* Increased frequency of the uncommon tumor necrosis factor -857T allele in British and Dutch patients with sarcoidosis. *Am. J. Respir. Crit. Care Med.* **165**, 1119–1124 (2002).
- Abdallah, A. *et al.* Inhibitor kappa B-alpha (IkappaB-alpha) promoter polymorphisms in UK and Dutch sarcoidosis. *Genes Immunity* **4**, 450–454 (2003).
- Stammers, M., Rowen, L., Rhodes, D., Trowsdale, J. & Beck, S. BTL-II: a polymorphic locus with homology to the butyrophilin gene family, located at the border of the major histocompatibility complex class II and class III regions in human and mouse. *Immunogenetics* **51**, 373–382 (2000).
- Krawczak, M., Reiss, J. & Cooper, D.N. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet.* **90**, 41–54 (1992).
- Long, M. & Deutsch, M. Association of intron phases with conservation at splice site sequences and evolution of spliceosomal introns. *Mol. Biol. Evol.* **16**, 1528–1534 (1999).
- Rhodes, D.A., Stammers, M., Malcherek, G., Beck, S. & Trowsdale, J. The cluster of BTN genes in the extended major histocompatibility complex. *Genomics* **71**, 351–362 (2001).
- Sharpe, A.H. & Freeman, G.J. The B7-CD28 superfamily. *Nat. Rev. Immunol.* **2**, 116–126 (2002).
- Albrecht, M., Domingues, F.S., Schreiber, S. & Lengauer, T. Identification of mammalian orthologs associates PYPAF5 with distinct functional roles. *FEBS Lett.* **538**, 173–177 (2003).
- Ikemizu, S. *et al.* Structure and dimerization of a soluble form of B7-1. *Immunity* **12**, 51–60 (2000).
- Schwartz, J.C., Zhang, X., Fedorov, A.A., Nathenson, S.G. & Almo, S.C. Structural basis for co-stimulation by the human CTLA-4/B7-2 complex. *Nature* **410**, 604–608 (2001).
- Stamper, C.C. *et al.* Crystal structure of the B7-1/CTLA-4 complex that inhibits human immune responses. *Nature* **410**, 608–611 (2001).
- Zhang, X., Schwartz, J.C., Almo, S.C. & Nathenson, S.G. Crystal structure of the receptor-binding domain of human B7-2: insights into organization and signaling. *Proc. Natl. Acad. Sci. USA* **100**, 2586–2591 (2003).
- Bajorath, J. Structural biology of T-cell costimulatory proteins: New insights, more surprises. *J. Mol. Graph. Model.* **19**, 619–623 (2001).
- Reiling, N., Blumenthal, A., Flad, H.D., Ernst, M. & Ehlers, S. Mycobacteria-induced TNF-alpha and IL-10 formation by human macrophages is differentially regulated at the level of mitogen-activated protein kinase activity. *J. Immunol.* **167**, 3339–3345 (2001).
- Hampe, J. *et al.* A non-electrophoretic method for high-throughput HLA-DRB1 group genotyping. *Biotechniques* **36**, 148–151 (2004).
- Morton, N.E. & Collins, A. Tests and estimates of allelic association in complex inheritance. *Proc. Natl. Acad. Sci. USA* **95**, 11389–11393 (1998).
- Stenzel, A. *et al.* Patterns of linkage disequilibrium in the MHC region on human chromosome 6p. *Hum. Genet.* **114**, 377–385 (2004).
- Walunas, T.L., Bakker, C.Y. & Bluestone, J.A. CTLA-4 ligation blocks CD28-dependent T cell activation. *J. Exp. Med.* **183**, 2541–2550 (1996).
- Borriello, F. *et al.* B7-1 and B7-2 have overlapping, critical roles in immunoglobulin class switching and germinal center formation. *Immunity* **6**, 303–313 (1997).
- Collins, A.V. *et al.* The interaction properties of costimulatory molecules revisited. *Immunity* **17**, 201–210 (2002).
- Agadjanyan, M.G. *et al.* Costimulatory molecule immune enhancement in a plasmid vaccine model is regulated in part through the Ig constant-like domain of CD80/86. *J. Immunol.* **171**, 4311–4319 (2003).
- Zissel, G. *et al.* Human alveolar epithelial cells type II are capable of regulating T-cell activity. *J. Invest. Med.* **48**, 66–75 (2000).
- Hoebe, K. *et al.* Upregulation of costimulatory molecules induced by lipopolysaccharide and double-stranded RNA occurs by Trif-dependent and Trif-independent pathways. *Nat. Immunol.* **4**, 1223–1229 (2003).
- Ueda, H. *et al.* Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* **423**, 506–511 (2003).
- Costabel, U. & Hunninghake, G.W. ATS/ERS/WASOG statement on sarcoidosis. Sarcoidosis Statement Committee. American Thoracic Society. European Respiratory Society. World Association for Sarcoidosis and Other Granulomatous Disorders. *Eur. Respir. J.* **14**, 735–737 (1999).
- Statement on sarcoidosis. Joint Statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG) adopted by the ATS Board of Directors and by the ERS Executive Committee. *Am. J. Respir. Crit. Care Med.* **160**, 736–755 (1999).
- 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).
- Hampe, J. *et al.* An integrated system for high throughput TaqMan based SNP genotyping. *Bioinformatics* **17**, 654–655 (2001).
- Hunninghake, G.W., Gadek, J.E., Kawanami, O., Ferrans, V.J. & Crystal, R.G. Inflammatory and immune processes in the human lung in health and disease: evaluation by bronchoalveolar lavage. *Am. J. Pathol.* **97**, 149–206 (1979).
- Clayton, D. & Jones, H. Transmission/disequilibrium tests for extended marker haplotypes. *Am. J. Hum. Genet.* **65**, 1161–1169 (1999).
- Kruglyak, L., Daly, M.J., Reeve Daly, M.P. & Lander, E.S. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.* **58**, 1347–1363 (1996).
- Krawczak, M. *et al.* Allelic association of the cystic fibrosis locus and two DNA markers, XV2c and KM19, in 55 German families. *Hum. Genet.* **80**, 78–80 (1988).
- Legouis, R. *et al.* Basolateral targeting by leucine-rich repeat domains in epithelial cells. *EMBO Rep.* **4**, 1096–1102 (2003).
- Rosenstiel, P. *et al.* TNF-alpha and IFN-gamma regulate the expression of the NOD2 (CARD15) gene in human intestinal epithelial cells. *Gastroenterology* **124**, 1001–1009 (2003).

Corrigendum: Sarcoidosis is associated with a truncating splice site mutation in the gene *BTNL2*

R Valentonyte, J Hampe, K Huse, P Rosenstiel, M Albrecht, A Stenzel, M Nagy, K I Gaede, A Franke, R Haesler, A Koch, T Lengauer, D Seegert, N Reiling, S Ehlers, E Schwinger, M Platzer, M Krawczak, J Müller-Quernheim, M Schürmann & S Schreiber
Nat. Genet. 37, 357–364 (2005).

In the version of Supplementary Table 3 initially published online, the nomenclature of DQB and DPB alleles was partly incorrect. The errors have now been corrected and Supplementary Table 3 has been replaced. Neither the stratified analyses that highlighted the independence of the *BTNL2* effect from the two HLA loci (Supplementary Table 3 online) nor any other conclusions of the manuscript were affected by these mistakes.