A genome-wide association study reveals evidence of association with sarcoidosis at 6p12.1

Sylvia Hofmann PhD1,*, Annegret Fischer PhD1,*, Andreas Till PhD1,*, Joachim Müller-Quernheim MD2, Robert Häsler PhD1, Andre Franke PhD1, Karoline I. Gäde PhD3, Heidi Schaarlschmidt MSc1, GenPhenReSa Consortiumª, Philip Rosenstiel MD1, Almut Nebel PhD1, Manfred Schürmann MD3, Michael Nothnagel PhD5, Stefan Schreiber MD1,6

1 Institute for Clinical Molecular Biology; Christian-Albrechts-University; Kiel, Germany
2 Department of Pneumology; University of Freiburg; Freiburg, Germany
3 Department of Clinical Medicine; Research Center Borstel; Borstel, Germany
4 Institute of Human Genetics; University of Lübeck; Lübeck, Germany
5 Institute of Medical Informatics and Statistics; Christian-Albrechts-University; Kiel, Germany
6 Department of General Internal Medicine; University Hospital Schleswig-Holstein; Christian-Albrechts-University; Kiel, Germany

* These authors contributed equally to the work.
ª GenPhenReSa Consortium (see Supplement Material Appendix S1 with the detailed affiliations of the consortium members)

Corresponding authors:
Prof. Sylvia Hofmann, PhD, Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Schittenhelmstrasse 12, D-24105 Kiel, Germany; Tel.: +49-431-597-1306; Fax: +49-431-597-3730; E-mail: s.hofmann@ikmb.uni-kiel.de.
Prof. Stefan Schreiber, MD, Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Schittenhelmstrasse 12, D-24105 Kiel, Germany; Tel.: +49-431-597-2350; Fax: +49-431-597-3730; E-mail: s.schreiber@ikmb.uni-kiel.de.

Experiments were performed at the Institute for Clinical Molecular Biology, Christian-Albrechts University, Kiel, Germany. The study was supported by grants from the Federal Ministry for Education and Research in Germany (BMBF) within the National Genome Research Network (NGFN) [grant numbers 01GS0426 and 01GS0809] and by the German Research Foundation (DFG) through the Cluster of Excellence “Inflammation at Interfaces” as well as through the projects DFG MU 692/8-1 and MU 692/7-1.
No conflict of interest exists for any of the authors.
Abstract

Sarcoidosis is a complex systemic inflammatory disease of unknown aetiology that is influenced by a variety of genetic and environmental factors. To identify further susceptibility loci for sarcoidosis, a genome-wide association study was conducted in 381 patients and 392 control individuals based on Affymetrix 100k GeneChip data. The top 25 SNPs were selected for validation in an independent study panel (1,582 patients vs. 1,783 controls).

Variant rs10484410 on chromosome 6p12.1 was significantly associated, with a Bonferroni-corrected p value of 2.90×10⁻² in the validation sample and a nominal p value of 2.64×10⁻⁴ in the GWAS. Extensive fine mapping of the novel locus narrowed down the signal to a region comprising the genes BAG2, C6orf65, KIAA1586, ZNF451, and RAB23. Verification of the sarcoidosis associated non-synonymous SNP rs1040461 in a further independent case-control sample and quantitative mRNA expression studies point to the RAB23 gene as the most likely risk factor. RAB23 is proposed to be involved in antibacterial defence processes and regulation of the sonic hedgehog signalling pathway.

The identified association of the 6p12.1 locus with sarcoidosis implicates this locus as a further susceptibility factor, and RAB23 as a potential signalling component that may open up new perspectives in the pathophysiology of sarcoidosis.

Keywords

Affymetrix 100k, genome-wide association study, Rab23, Sarcoidosis, susceptibility (alphabetically)
Introduction

Sarcoidosis (MIM 181000) is a multisystem disease with a high risk to become chronic, characterized by noncaseating epitheloid cell granulomas that can manifest in virtually any organ system, most frequently in the lung. It is known as a disease of young adults, with an annual incidence between < 1 and 64 per 100,000, depending on ethnicity and geographic region [1]. The cause of sarcoidosis is still unknown but it is thought to be triggered by a complex combination of yet unknown environmental [2, 3] and genetic factors [4]. Evidence for a strong genetic component is provided by twin and family studies, which have shown a higher concordance rate of sarcoidosis in monozygotic compared to dizygotic twins (0.148 vs. 0.012) and a heritability of 66% [5]. The genetic underpinning of sarcoidosis is further supported by known genetic risk factors located in the human leukocyte antigen (HLA) region (reviewed in [6]) and the discovery of the first confirmed susceptibility gene butyrophilin-like 2 (BTNL2; MIM 606000) on chromosome 6 [7, 8]. Besides these disease loci, numerous association studies of potential candidate regions suggested additional susceptibility genes like the chemokine receptors chemokine receptor 2 (CCR2; MIM 601267) and chemokine receptor 5 (CCR5; MIM 601373), the tumor necrosis factor-α (TNF-α; MIM 191160) and several other HLA-loci (for a review see [4, 6]). However, many of these show conflicting results or await replication. Most recently, we identified annexin A11 (ANXA11; MIM 602572) as a further disease-related gene in sarcoidosis by a genome-wide association study (GWAS) using the Affymetrix 5.0 chip [9]. In addition, we discovered a common susceptibility locus shared by sarcoidosis and Crohn’s disease (CD; MIM266600) on chromosome 10p12.2, based on a joint analysis of the Affymetrix 100k GWA scans for the two diseases [10].

The 100k and the 5.0 chips share only a small proportion of their SNPs (< 15% of the 100k array SNPs are also present on the 5.0 set). Therefore, we analysed our 100k sarcoidosis data set, which was formerly only part of the combined analysis with CD (pooled data) [10], to identify susceptibility loci that could not have been detected by the 5.0 chip due to a difference in marker coverage.
Materials and Methods

Patient recruitment and phenotyping

German sarcoidosis patients of panels A (GWAS), B (validation) and C (fine mapping) were recruited as previously described [7, 9, 11] (for details see Supplement Material Appendix S2). The diagnosis of the participating patients was established on the basis of the International Consensus Statement on Sarcoidosis [12] and by histological demonstration of non-necrotizing granuloma in an involved organ, most commonly the lung. Only patients with classical Löfgren’s syndrome were recruited without histological support. According to the clinical presentation of the disease, patients were classified as having chronic or acute sarcoidosis as described before [13]. Patients (n = 210) that could not be classified unequivocally concerning the course of the disease were excluded from the subphenotype-specific analysis that was performed in the fine mapping stages. German control individuals of panel A, B and C were obtained from the PopGen biobank [14]. Panel D (replication) comprised sarcoidosis patients from different European locations that were recruited within a European network “GenPhenReSa” (for details see Supplement Material Appendix S1) and German controls (n = 2564) that were recruited as reported in detail elsewhere [15]. Information on SNP genotyping and selection, statistical analysis, analysis of tissue-specific expression by RT-PCR, bronchoalveolar lavage cell samples and mRNA isolation and real-time PCR can be found in Supplement Material Appendix S3-S10.
Results

GWAS analyses

After applying conservative and established quality filters to the dataset (see Methods section), 773 German samples (381 cases/392 controls) and 97,088 SNPs were included in the association analysis of the screening stage. At a nominal significance level of 0.05, the experiment had 63% pre hoc power to detect variants with an odds ratio (OR) of 1.5 and higher when assuming a frequency of the disease-associated allele of 20% in control subjects (see Supplement Material Figure S3). The estimated population stratification was small, with a genomic inflation factor of $\lambda_{GC} = 1.076$ (1.0 indicating no inflation). Subsequent association results were corrected according to this inflation in the $\chi^2$ test statistic; the resulting quantile-quantile plot is shown in the Figure 1.

Validation of lead variants

Twenty five SNPs from the screening stage that passed the aforementioned criteria (see Methods section) were genotyped in an independent validation panel B (1,783 German controls, 1,582 German sarcoidosis patients). Association analysis results for these SNPs are shown in Table 1 and full analysis results including genotype counts are listed in Supplement Material Table S1. Only one marker, rs10484410 on chromosome 6p12.1, passed the Bonferroni correction for multiple testing (corrected p value $p_{corr} = 25\times1.17\times10^{-3} = 2.90\times10^{-2}$) and was significantly associated with sarcoidosis (see also genome-wide plots of the chromosomes, Supplement Material Figure S4). The allelic OR for the rare G allele of rs10484410 was 1.26 (95% confidence interval [95% CI]: 1.10-1.44). Only 5.8% of the control individuals were homozygous for the risk allele G of rs10484410, while 7.1% of the sarcoidosis cases were homozygous for it.

The scan did not confirm our previously reported sarcoidosis associations of $BTNL2$ [7] and $ANXA11$ [9] in the screening. Both loci were replicated in independent European populations ($ANXA11$: [16]; $BTNL2$: [17]) and in Americans (BTNL2: [8]) respectively and can be considered as true associations. However, not a single SNP in the $ANXA11$ or in the $BTNL2$ gene was present in the Affymetrix GeneChip®Human Mapping 100k set that shows less than 15% marker overlap with the 500K Array Set. Conversely, in the previously published GWAS using the Affymetrix 5.0 chip [9] the lead SNP rs10484410 itself was not included, but rs1044670 (SNP_A-1824553; p value = $9\times10^{-3}$), which was found to be strongly associated in the fine-mapping stage. However, the SNP ranked only very low in the previous GWAS (rank 6229). The results of the validation stage were verified using TaqMan genotyping as an independent technology (> 99.8% genotype concordance).

Fine-mapping around rs10484410 (6p12.1)

In addition to the lead SNP rs10484410, 44 HapMap tagging SNPs (tagSNPs) were selected for fine-mapping ~800 kb of the 6p12.1 region (see Methods section) carried out in panel C. Forty-one of the 44 SNPs passed the above mentioned quality criteria (see Methods). Nine markers yielded a p value < 0.05 in the analysis; detailed results including genotype counts for the overall sarcoidosis panel are shown in Supplement Material Table S2. Results were not corrected for multiple testing because the aim of the panel C analysis was to identify the likely source of the association signal that was established in panel B. No consistent difference in the significance of the association signal at the markers tested could be observed between the chronic and acute phenotype (see Supplement Material Tables S2 and S3).
Linkage disequilibrium (LD) between rs10484410 and the nine additional SNPs with a significant association from panel C vary greatly (0.41 < r² ≤ 0.99; see Supplement Material Table S2). Of the ten markers, SNP rs1411578 showed the strongest association (p = 6.64×10⁻⁴). Marker rs1411578 (G > C) is located in exon 7 (3'UTR) of the ras-related protein Rab-23 (RAB23; MIM 606144). The minor G allele had a frequency of 17% in affected individuals and 14% in the control subjects, and 3.5% of the patients and 2% of the control individuals were homozygous for it. Figure 2 (A-E) gives an overview of the association signals, the conservation status, the genes and the LD structure at the 6p12.1 locus.

**Extended fine-mapping**

Since the results of the first fine-mapping stage pointed to RAB23, which is sharply delineated by loci with increased recombination rates at positions 57,190 kb and 57,271 kb based on HapMap data (Figure 2B), we aimed to ensure that the association signal is limited to the RAB23 genomic region and does not extend further up- or downstream. We therefore genotyped the samples from panel C for another 46 HapMap CEU-based tagSNPs that covered a ~105 kb region surrounding the lead SNP rs10484410 and the RAB23 genomic region (Supplement Material Table S3).

Looking at the total mapped region (~900 kb), the most significant associated SNP was rs7756421, located in the 3'UTR of the zinc finger protein 451 (ZNF451) gene. The marker is part of a region of high LD (r² ≥ 0.8), which includes eight other associated variants (p ≤ 0.001) that map to RAB23 (rs11398, rs1411578, rs1547226, rs3800018), to ZNF451 (rs6459178, rs17619360, rs10484410) and to a non-genic region (rs12190575; see Figure 3). A non-synonymous SNP (nsSNP; rs1040461) in the RAB23 gene that showed a significant association (p_{nom} = 7.83×10⁻³) with sarcoidosis as well was not part of this set of highly correlated markers.

We used Akaike’s Information Criterion (AIC) [18]-based backward model selection in a logistic regression analysis to determine whether all of these nine significant signals could be attributed, through LD, to one or more underlying causative variants. Model selection confirmed that the observed association signal of all the highly correlated SNPs had a single origin. If the significant associated nsSNP rs1040461 was also included, it remained in the model after selection, thus potentially representing an independent association signal.

**Replication of rs10484410, rs7756421 and rs1040461**

The GWAS lead SNP rs10484410, the SNP with the strongest association in the fine-mapped region (rs7756421) and the nsSNP rs1040461 in the RAB23 gene were genotyped in panel D for replication. We assigned each individual from our panel to the closest subpopulation of a European data set described elsewhere [19] and subsequently included the subpopulation-specific average values of the first six principal components (PC) of the genome-wide data of this reference panel to adjust for population stratification. Without adjustment for sampling location, markers rs10484410 and rs7756421 showed significant association with sarcoidosis on the allelic level (nominal p = 3.30×10⁻³ and p = 5.10×10⁻³, respectively), while marker rs1040461 did not. However, after inclusion of the first six PCs in the model, rs1040461 showed a nominal significance of p = 1.05×10⁻², which remained significant after correction for multiple testing, whereas the other two markers became non-significant (p_{rs10484410} = 1.40×10⁻¹, p_{rs7756421} = 1.80×10⁻¹; see Table 2). Analysis of only German cases and controls yielded a similar result for rs1040461 (p = 1.20×10⁻³ without adjustment for sampling location; p = 1.05×10⁻² with adjustment), but a minor decrease in significance for the other
two markers ($p_{\text{rs10484410}} = 6.90 \times 10^{-2}$, $p_{\text{rs7756421}} = 8.50 \times 10^{-2}$ without adjustment; $p_{\text{rs10484410}} = 1.40 \times 10^{-1}$, $p_{\text{rs7756421}} = 1.80 \times 10^{-1}$ with adjustment).

**Expression analysis of candidate genes**

To narrow down the association signal by plausible biological reasoning, the transcript levels of the five genes (*BAG family molecular chaperone regulator 2* [BAG2; MIM 603882], *BEN domain-containing protein 6* [C6orf65, BEND6], *uncharacterized protein KIAA1586* [KIAA1586], *ZNF451* and *RAB23*) located in the fine-mapped region were first assessed by RT-PCR in a panel of different human tissues. As shown in Figure 4, only RAB23 and ZNF451 showed a high expression in the lung, whereas BAG2 and KIAA1586 expression in this tissue was considerably lower. C6orf65 mRNA could only be detected in brain tissue and in few other tissues at extremely low level. Interestingly, a high expression of RAB23 became also apparent in small intestine and colonic mucosa.

Next, we analyzed the expression levels of the candidate genes in cells derived from bronchoalveolar lavage (BAL) using quantitative Real-time PCR and cDNA from sarcoidosis patients and controls (n = 5 per group). As shown in Figure 5, four of the five candidate genes were expressed in BAL cells, while C6orf65 (BEND6) mRNA was undetectable. Most interestingly, statistical analysis revealed that only RAB23 displayed significant differences in relative expression levels between patients and controls ($p = 2.94 \times 10^{-3}$ based on Mann-Whitney U test for non-parametric data). Compared to controls, BAL cells from sarcoidosis patients exhibited an up to 3-fold increase in RAB23 mRNA levels further emphasizing the potential involvement of this gene in the pathogenesis of sarcoidosis.

For further expression analysis see eQTL results in the online supplemental material (Supplement Material Figures S5A-S7C).
Discussion

We found evidence for a novel sarcoidosis susceptibility locus on chromosome 6p12.1 that harbours the five candidate genes C6orf65 (BEND6), BAG2, KIAA1586, ZNF415 and RAB23. This result was obtained by the analysis of genome-wide case control association scan data with > 97,000 SNP markers and by extensive fine-mapping of the validated region. The scan did not confirm the sarcoidosis locus ANXA11 that has recently been discovered using the Affymetrix 5.0 chip. Not a single SNP in the ANXA11 gene was present in the Affymetrix GeneChip® Human Mapping 100k set that shows less than 15% marker overlap with the 5.0 set. Conversely the 5.0 chip covered the newly discovered risk locus on chromosome 6p12.1 with 16 SNPs, including the highly associated SNP rs1044670. Despite this being a true signal, the SNP ranked below any feasible cut-off for replication in the previous GWAS (rank 6229) and was therefore not validated.

The allelic OR for the rare allele of the detected GWAS lead SNP rs10484410 was moderate (1.26). We have to point out here that the low prevalence of sarcoidosis in the general population strongly limits the number of available samples. In turn, the recruitment of enough samples to achieve genome-wide significance for genetic factors conveying moderate to small risks or with low allele frequencies may be infeasible. In this regard, our study design – with GWAS as a hypothesis-generating step, leading to a small correction factor for multiple testing in the validation stage – at all enabled the detection of such a low-risk factor for sarcoidosis.

Fine-mapping of the novel locus revealed a strong association of sarcoidosis with several SNPs in a small region of high LD spreading from C6orf65 (BEND6), over ZNF451, to RAB23. Among these SNPs are four putative functional variants – rs1044670 (c.*1109G > A) located in the 3'UTR of C6orf65 (BEND6); rs1411578 (c.*416G > C) in the 3'UTR and rs1040461 (c.619G > A) in exon 7 of RAB23; and rs7756421 (c.*742A > G) in the 3'UTR of the ZNF451 gene. Logistic regression analysis suggested that the association signal might be mainly driven by more than one variant in RAB23, including the missense mutation rs1040461. This SNP changes the protein sequence by the amino acid substitution of glycine for the small sized and polar serine (G207/S207) in a non-domain-containing region of the gene. Replication in a further panel revealed that marker rs1040461, but not rs7756421 or the GWAS lead SNP rs10484410, appears to be associated with sarcoidosis in several European populations after adjustment for sampling location. This finding strongly supports the hypothesis of rs1040461 being a true sarcoidosis risk variant.

Analysis of tissue-specific mRNA expression profiles demonstrates high levels of RAB23 and ZNF451 in healthy lung. Moreover, RAB23 was the only candidate gene displaying highly significant differences in relative expression between sarcoidosis patients and controls, indicating a potential involvement of RAB23 in disease pathogenesis. The pathophysiological relevance of our findings is indicated by an eQTL analysis which identified alternations in disease associated processes, e.g. response to external stimulus, cellular defence response and metabolic processes, in response to the presented genetic variation. Although these findings do not prove that the observed over-expression of RAB23 results from variation in rs1040461 or another potentially causative variant in the associated region, they do provide a valuable starting point for future studies elucidating the role of RAB23 and its variants in the etiopathogenesis of sarcoidosis.

The RAB23 gene belongs to the Rab family of 160 small guanosine triphosphatases (GTPases) that regulate intracellular trafficking of membrane-associated proteins [20, 21]. Based on microarray expression data (UCSC: GNF Expression Atlas 2 Data and Affymetrix
All Exon Microarrays; see URL [8]), RAB23 is higher expressed in bronchial epithelial and thyroid cells, in addition to uterus and cerebellum tissue. This appears to be consistent with our expression results in lung tissue (see above). The potential involvement of this locus in sarcoidosis pathogenesis remains to be unravelled and can yet only be proposed as a working hypothesis. Sarcoidosis is a systemic immune disorder in which T-cell-mediated inflammation causes the formation of granulomas, which resemble a delayed hypersensitivity reaction. A delayed hypersensitivity reaction may be caused by the intracellular presence of antigens of a chemical or microbial origin. Many reports describe the presence of microbial cell wall agents in tissues of patients with sarcoidosis, and several clinical studies demonstrate the occurrence of microbes in patients with sarcoidosis [22, 23]. Since RAB23 plays a role in antibacterial activity of the endogenous autophagy machinery [24], it is biologically plausible that dysfunction might lead to an impaired autophagic clearance after exposure of the lung epithelium to bacterial pathogens. The high expression of RAB23 not only in lung but also in small intestine and colonic mucosa may point to a broader potential role of RAB23 for antibacterial defence mechanisms in epithelial barrier organs.

However, the RAB23 protein has also been implicated in facilitating vesicular transport, controlling endocytic progression to lysosomes [24, 25], and particularly in antagonizing sonic hedgehog (SHH) signal transduction in neural systems [26, 27]. Interestingly, the SHH pathway may also play a role in chronic lung fibrosis and immune system communication [28] thereby providing alternative views regarding the involvement of RAB23 in sarcoidosis. Components of the SHH cascade have been identified in the adult immune system, participating in CD4+ T lymphocyte activation, and studies on fibrotic pulmonary disorders have demonstrated SHH in both human and mouse lung restricted to areas of active disease (for review see [29]). It is assumed that SHH signalling may contribute to epithelial repair and may act as an intermediary in cross-talk between damaged epithelium and the immune/inflammatory system. In sarcoidosis, activated pulmonary T-helper type 1 lymphocytes are essential for the inflammatory process and particular CD4+ T cell subsets can be found at dramatically increased levels in the bronchoalveolar lavage fluid of sarcoidosis patients with active disease. It is possible that RAB23 contributes to the disease through dysfunction within the SHH pathway leading to an over-activation of CD4+ T lymphocytes or to an inadequate repair process in the damaged lung in pulmonary sarcoidosis.

Although there is evidence indicating that variations in RAB23 confer susceptibility to sarcoidosis, it is also possible that the causative variants are located in ZNF451, C6orf65 (BEND6) or another, yet undefined genetic element. ZNF451, also known as COASTER, KIAA0576 and KIAA1702, is conserved across vertebrates, ubiquitously expressed and has been suggested to act as a coactivator for steroid receptors; it might also be involved in transcriptional regulation [30]. The putative gene C6orf65 (BEND6) is a complex locus that appears to produce several proteins with no sequence overlap and of yet unknown functions.

To conclude, this is the first report of an association between sarcoidosis and the 6p12.1 locus that comprises several genes and a likely candidate being RAB23. The importance of this observation should be evaluated by further delineating the biological role of RAB23 in sarcoidosis.
Acknowledgments

The authors wish to thank all patients, families and physicians for their cooperation. The support of the Deutsche Sarkoidose-Vereinigung e.V., the PopGen biobank and of the contributing pulmonologists is gratefully acknowledged. Finally, we thank the staff of the Institute for Clinical Molecular Biology (Kiel, Germany) for technical assistance.
References


**Figure Legends**

Figure 1: Quantile-quantile (Q-Q) plot of the test statistic for all SNPs of the GWAS that passed the quality criteria. Over-dispersion of the association test-statistic was estimated as $\lambda = 1.076$ for the sarcoidosis GWAS (panel A). The shaded region reflects the 95% concentration band that is formed by calculating the 2.5th and 97.5th centiles of the distribution of the order statistic under random sampling and the null hypothesis. The plot was created using the function `qq.chisq` of the R-package `snpMatrix` v1.0.33. For details see [31-33].
Figure 2: Regional association plot and structure of LD of the 6p12.1 locus
(A) Plot of the negative natural logarithm of the p values obtained in fine-mapping (FM) of the 800 kb region around the lead variant rs10484410 (marked as green dot) in panel C. The most significantly associated SNP of the FM is highlighted by a blue vertical line. Forty-four tagging SNPs were genotyped in 1,806 controls and 1,837 sarcoidosis patients (FM SA). Nominal p values of the analysis in panel C and of the validated 100k genome-wide association study (GWAS) lead SNP rs10484410, (381 sarcoidosis cases vs. 392 controls; panel A) are shown. The red dotted line corresponds to a significance threshold of 0.05. Positions are in NCBI build-36.1 coordinates. For details, see Supplement Material Tables S1 and S2. (B) shows the plotted recombination rate (in centi-Morgans [cM] per Mb) and the cumulative genetic distance in cM, while (C) shows the sequence conservation score based on 44 different vertebrate species (taken from UCSC Genome Browser, Vertebrate Multiz Alignment & Conservation). (D) The position and intron/exon structure of the underlying genes. (E) shows the pair-wise linkage disequilibrium (LD) measured by $r^2$ in the studied sample for the genotyped variants that passed the quality criteria including the GWAs lead SNP rs10484410 (n = 76).
Figure 3: Overview of the linkage-disequilibrium (LD) structure of the 6p12.1 locus including the most significant sarcoidosis associated SNPs and the non-synonymous marker rs1040461 (c.619G>A) in the RAB23 gene.

Pairwise LD between the most significant associated SNP rs7756421 and SNPs with a nominal p value < 0.001 as well as the non-synonymous SNP rs1040461 in the focused region. LD estimation was based on the genotype data of the fine-mapping stages (panel C). The strength of LD corresponds to shades of gray; the displayed r² values are given as percentage.
Figure 4: Tissue-specific expression profile of genes within the 6p12.1 risk locus. A human cDNA tissue panel was analyzed for expression of the candidate genes using RT-PCR. Housekeeping gene GAPDH was used as internal control.
Figure 5: Relative expression of the candidate genes in cells from bronchoalveolar lavage. Bronchoalveolar lavage (BAL) cell samples were taken from sarcoidosis patients and matched healthy controls (n = 5 per group). cDNA derived from BAL cells was assessed for relative expression of the candidate genes using quantitative real-time PCR and sequence-specific primer sets. Expression levels were normalized to those of housekeeping gene beta-Actin. Relative expression levels are shown as box plots representing the median and minimum/maximum of the five measurements. Significance levels of the differences were calculated using non-parametric Mann-Whitney U test. Only RAB23 expression was highly significantly different between patients and controls displaying a 3-4 fold increase in relative expression (**; p = 3.20×10^{-3}).
Tables

Table 1: Twenty-five hit SNPs that were selected for the validation stage. SNPs are ranked according to their genomic control (GC) adjusted p value (allelic $\chi^2$-test with one degree of freedom) in the screening panel. Minor allele frequencies (MAF) are listed for unaffected controls (U) and sarcoidosis patients (SA). Bonferroni-corrected (Corr.) p values are shown for the analysis of the validation sample B. Only variant rs10484410, highlighted in bold italic type, remains significant after correcting for multiple testing ($n = 25$). Positions are from NCBI’s build 36.1 (see URL [9]). For detailed results including genotype counts see Supplement Material Table S1.

<table>
<thead>
<tr>
<th>#</th>
<th>dbSNP</th>
<th>Chr</th>
<th>Position</th>
<th>MAF U</th>
<th>MAF SA</th>
<th>GC P-value</th>
<th>MAF U</th>
<th>MAF SA</th>
<th>P-value</th>
<th>Corr. P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs1434308</td>
<td>9</td>
<td>28,841,909</td>
<td>0.44</td>
<td>0.54</td>
<td>1.37x10^-7</td>
<td>0.48</td>
<td>0.48</td>
<td>8.06x10^-1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>rs717996</td>
<td>12</td>
<td>82,657,951</td>
<td>0.44</td>
<td>0.35</td>
<td>2.38x10^-6</td>
<td>0.40</td>
<td>0.40</td>
<td>8.28x10^-1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>rs10484410</td>
<td>6</td>
<td>57,104,647</td>
<td>0.12</td>
<td>0.19</td>
<td>2.64x10^-6</td>
<td>0.13</td>
<td>0.15</td>
<td>1.17x10^-3</td>
<td>2.93x10^-2</td>
</tr>
<tr>
<td>4</td>
<td>rs7962346</td>
<td>12</td>
<td>61,952,763</td>
<td>0.14</td>
<td>0.22</td>
<td>3.85x10^-5</td>
<td>0.17</td>
<td>0.16</td>
<td>2.94x10^-4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>rs10495266</td>
<td>1</td>
<td>225,313,639</td>
<td>0.10</td>
<td>0.16</td>
<td>5.19x10^-5</td>
<td>0.13</td>
<td>0.14</td>
<td>2.85x10^-4</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>rs951901</td>
<td>9</td>
<td>56,684,632</td>
<td>0.40</td>
<td>0.31</td>
<td>5.77x10^-5</td>
<td>0.36</td>
<td>0.34</td>
<td>1.34x10^-4</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>rs2383784</td>
<td>9</td>
<td>28,934,277</td>
<td>0.20</td>
<td>0.13</td>
<td>6.58x10^-4</td>
<td>0.17</td>
<td>0.17</td>
<td>7.37x10^-4</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>rs1873942</td>
<td>16</td>
<td>58,553,352</td>
<td>0.18</td>
<td>0.25</td>
<td>7.22x10^-4</td>
<td>0.23</td>
<td>0.20</td>
<td>6.18x10^-3</td>
<td>1.55x10^-2</td>
</tr>
<tr>
<td>9</td>
<td>rs2078539</td>
<td>16</td>
<td>60,141,739</td>
<td>0.35</td>
<td>0.26</td>
<td>7.74x10^-4</td>
<td>0.30</td>
<td>0.30</td>
<td>8.30x10^-3</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>rs10494667</td>
<td>1</td>
<td>164,817,060</td>
<td>0.09</td>
<td>0.04</td>
<td>8.68x10^-4</td>
<td>0.08</td>
<td>0.07</td>
<td>3.88x10^-3</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>rs4856497</td>
<td>3</td>
<td>80,215,335</td>
<td>0.43</td>
<td>0.52</td>
<td>9.35x10^-4</td>
<td>0.50</td>
<td>0.47</td>
<td>6.62x10^-4</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>rs1362941</td>
<td>5</td>
<td>59,037,578</td>
<td>0.30</td>
<td>0.22</td>
<td>1.19x10^-3</td>
<td>0.27</td>
<td>0.25</td>
<td>1.02x10^-4</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>rs10489924</td>
<td>1</td>
<td>99,231,562</td>
<td>0.14</td>
<td>0.09</td>
<td>1.20x10^-3</td>
<td>0.10</td>
<td>0.09</td>
<td>3.65x10^-4</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>rs866632</td>
<td>9</td>
<td>28,845,071</td>
<td>0.35</td>
<td>0.27</td>
<td>1.33x10^-3</td>
<td>0.31</td>
<td>0.33</td>
<td>2.58x10^-2</td>
<td>6.45x10^-3</td>
</tr>
<tr>
<td>15</td>
<td>rs694788</td>
<td>5</td>
<td>113,954,835</td>
<td>0.41</td>
<td>0.49</td>
<td>1.38x10^-3</td>
<td>0.45</td>
<td>0.45</td>
<td>8.49x10^-4</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>rs7132697</td>
<td>12</td>
<td>66,819,108</td>
<td>0.32</td>
<td>0.25</td>
<td>1.56x10^-3</td>
<td>0.30</td>
<td>0.29</td>
<td>5.33x10^-4</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>rs185793</td>
<td>5</td>
<td>45,483,437</td>
<td>0.39</td>
<td>0.31</td>
<td>1.67x10^-3</td>
<td>0.33</td>
<td>0.34</td>
<td>1.65x10^-4</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>rs10483261</td>
<td>14</td>
<td>21,426,391</td>
<td>0.51</td>
<td>0.42</td>
<td>1.73x10^-3</td>
<td>0.48</td>
<td>0.46</td>
<td>5.80x10^-2</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>rs10483437</td>
<td>14</td>
<td>32,905,552</td>
<td>0.34</td>
<td>0.26</td>
<td>1.74x10^-4</td>
<td>0.29</td>
<td>0.30</td>
<td>4.64x10^-1</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>rs2351010</td>
<td>5</td>
<td>96,319,685</td>
<td>0.47</td>
<td>0.39</td>
<td>1.75x10^-4</td>
<td>0.42</td>
<td>0.44</td>
<td>1.09x10^-4</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>rs10487432</td>
<td>7</td>
<td>125,458,829</td>
<td>0.36</td>
<td>0.28</td>
<td>1.90x10^-4</td>
<td>0.32</td>
<td>0.31</td>
<td>4.78x10^-4</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>rs929056</td>
<td>13</td>
<td>92,901,302</td>
<td>0.39</td>
<td>0.31</td>
<td>2.00x10^-4</td>
<td>0.37</td>
<td>0.34</td>
<td>1.33x10^-2</td>
<td>3.33x10^-4</td>
</tr>
<tr>
<td>23</td>
<td>rs7920803</td>
<td>10</td>
<td>28,481,267</td>
<td>0.30</td>
<td>0.23</td>
<td>2.11x10^-4</td>
<td>0.25</td>
<td>0.24</td>
<td>2.37x10^-4</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>rs1439172</td>
<td>14</td>
<td>45,797,214</td>
<td>0.20</td>
<td>0.27</td>
<td>2.17x10^-4</td>
<td>0.23</td>
<td>0.22</td>
<td>6.77x10^-4</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>rs4941920</td>
<td>13</td>
<td>38,740,373</td>
<td>0.33</td>
<td>0.25</td>
<td>2.19x10^-4</td>
<td>0.29</td>
<td>0.30</td>
<td>8.32x10^-3</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2: Association analysis results for the GWAS lead SNP (rs10484410), the fine mapping hit SNP (rs7756421) and rs1040461 ($RAB23$; c.619G>A) on chromosome 6 based on genotyping in the independent replication panel D (935 cases and 2,564 controls). P-values of the likelihood ratio test (LRT) and allelic odds ratios (OR) are presented for the multiplicative model. Similar data were obtained in the genotypic and recessive model (data not shown). Variant rs1040461, highlighted in bold italic type, showed significant association with sarcoidosis after adjustment for population stratification.

<table>
<thead>
<tr>
<th>dbSNP</th>
<th>Position</th>
<th>MAF U</th>
<th>MAF SA</th>
<th>p(LRT)</th>
<th>OR (95 CI)</th>
<th>p(LRT)</th>
<th>OR (95 CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10484410</td>
<td>57,104,647</td>
<td>0.14</td>
<td>0.17</td>
<td>3.33x10^-3</td>
<td>1.24 (1.08-1.44)</td>
<td>1.36x10^-1</td>
<td>1.21 (NA)</td>
</tr>
<tr>
<td>rs7756421</td>
<td>57,141,968</td>
<td>0.14</td>
<td>0.17</td>
<td>5.05x10^-3</td>
<td>1.23 (1.07-1.43)</td>
<td>1.78x10^-1</td>
<td>1.20 (NA)</td>
</tr>
<tr>
<td>rs1040461</td>
<td>57,163,313</td>
<td>0.09</td>
<td>0.08</td>
<td>1.37x10^-4</td>
<td>1.16 (0.95-1.42)</td>
<td>1.05x10^-2</td>
<td>1.68 (NA)</td>
</tr>
</tbody>
</table>