



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

S. Hofmann^{*,##}, A. Fischer^{*,##}, A. Till^{*,##}, J. Müller-Quernheim[#], R. Häslér^{*}, A. Franke^{*}, K.I. Gäde[†], H. Schaarschmidt^{*}, P. Rosenstiel^{*}, A. Nebel^{*}, M. Schürmann⁺, M. Nothnagel[§], S. Schreiber^{*,f} the GenPhenReSa Consortium^{**}

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 (GWAS) was conducted in 381 patients and 392 control individuals based on Affymetrix 100k GeneChip data. The top 25 single-nucleotide polymorphisms (SNPs) were selected for validation in an independent study panel (1,582 patients versus 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 nonsynonymous 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

Sarcoidosis (Mendelian Inheritance in Man (MIM) 181000) is a multisystem disease with a high risk of becoming chronic that is characterised by noncaseating epithelioid 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 cases per 100,000 persons, 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 with dizygotic twins (0.148 versus 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, such as the chemokine receptors *chemokine receptor 2 (CCR2)*; MIM 601267) and *chemokine receptor 5 (CCR5)*; MIM 601373), the *tumour necrosis factor- α (TNFA)*; 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 *annexinA11 (ANXA11)*; MIM 602572) as a further disease-related gene in sarcoidosis in a genome-wide association study (GWAS) using the Affymetrix 5.0 chip (Affymetrix, Santa Clara, CA, USA) [9]. In addition, we discovered a common susceptibility

AFFILIATIONS

^{*}Institute of Clinical Molecular Biology,
[†]Institute of Medical Informatics and Statistics,
[‡]Dept of General Internal Medicine, University Hospital Schleswig-Holstein, Christian-Albrechts-University, Kiel,
[§]Dept of Pneumology, University of Freiburg, Freiburg,
[¶]Dept of Clinical Medicine, Research Center Borstel, Borstel, and
^{||}Institute of Human Genetics, University of Lübeck, Lübeck, Germany.
^{**}For a full list of the GenPhenReSa Consortium members and their affiliations, see the Acknowledgements section.
^{##}These authors contributed equally to the study.

CORRESPONDENCE

S. Hofmann
Institute of Clinical Molecular Biology
Christian-Albrechts-University Kiel
Schittenhelmstrasse 12
D-24105 Kiel
Germany
E-mail: s.hofmann@ikmb.uni-kiel.de

Received:
Jan 05 2011
Accepted after revision:
April 02 2011
First published online:
May 03 2011

European Respiratory Journal
Print ISSN 0903-1936
Online ISSN 1399-3003

This article has supplementary material available from www.erj.ersjournals.com

locus shared by sarcoidosis and Crohn's disease (MIM266600) on chromosome *10p12.2*, based on a joint analysis of the Affymetrix 100k GWAS scans for the two diseases [10].

The 100k and 5.0 chips share only a small proportion of their single-nucleotide polymorphisms (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 Crohn's disease (pooled data) [10], to identify susceptibility loci that could not have been detected by the 5.0 chip due to the difference in marker coverage.

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 Appendix S2 in the online supplementary material). 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-necrotising 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 previously [13]. Patients (n=210) who 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 panels A, B and C were obtained from the PopGen Biobank (Montreal, QC, Canada) [14]. Panel D (replication) comprised sarcoidosis patients from different European locations that were recruited within the European network Genotype-Phenotype Relationship in Sarcoidosis (GenPhenReSa) (for details see the Acknowledgements section and Appendix S1 in the online supplementary material) and German controls (n=2,564) that were recruited as reported in detail previously [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 Appendices S3–S10 in the online supplementary material.

RESULTS

GWAS analyses

After applying conservative and established quality filters to the dataset (see Methods section), 773 German samples (381 cases and 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 when assuming a frequency of the disease-associated allele of 20% in control subjects (see fig. S3 in the online supplementary material). 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-squared test statistic; the resulting quantile-quantile plot is shown in figure 1.

Validation of lead variants

25 SNPs from the screening stage that passed the aforementioned criteria (see Methods section) were genotyped in the independent validation panel B (1,783 German controls and 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 table S1 in the online supplementary material. Only one marker, rs10484410 on chromosome *6p12.1*, passed the Bonferroni correction for multiple testing (corrected $p=25 \times 1.17 \times 10^{-3}=2.90 \times 10^{-2}$) and was significantly associated with sarcoidosis (see also genome-wide plots of the chromosomes; fig. S4 in the online supplementary material). The allelic OR for the rare G allele of rs10484410 was 1.26 (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 with *BTNL2* [7] and *ANXA11* [9] in the screening. Both loci were replicated in independent European populations (*ANXA11* [19] and *BTNL2* [20]) and in Americans (*BTNL2* [8]) 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, which shows <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; $p=9 \times 10^{-3}$) was found to be strongly associated in the fine-mapping stage. However, the SNP ranked only very low in the previous GWAS (rank 6,229). The results of the validation stage were verified using TaqMan genotyping as an independent technology (>99.8% genotype concordance).

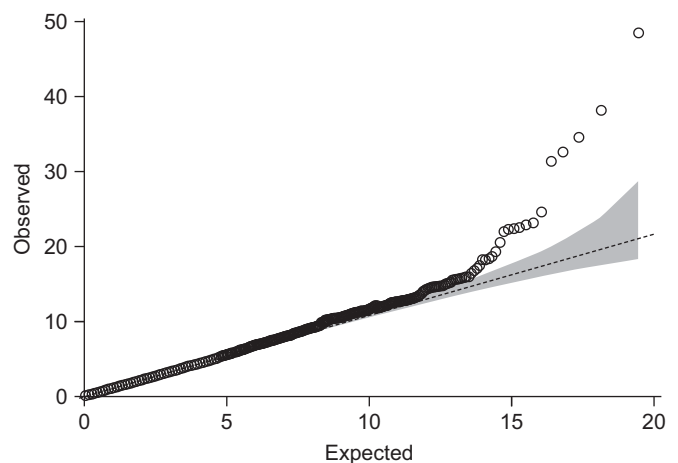


FIGURE 1. Quantile-quantile plot of the test statistic for all single-nucleotide polymorphisms of the genome-wide association study (GWAS) that passed the quality criteria. ----: over-dispersion of the association test statistic was estimated as $\lambda=1.076$ for the sarcoidosis GWAS (panel A); ■: 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` version 1.0.33 (Bioconductor, Seattle, WA, USA). The expected distribution is based on a Chi-squared test with one degree of freedom. For details see [16–18].

TABLE 1 25 hit single-nucleotide polymorphisms (SNPs) that were selected for the validation stage

SNP	dbSNP	Chr	Position	Panel A [#]		Panel B [†]				
				MAF		GC p-value	MAF		p-value	Corr p-value
				U	SA		U	SA		
1	rs1434308	9	28841909	0.44	0.54	1.37 × 10 ⁻⁴	0.48	0.48	8.06 × 10 ⁻¹	1
2	rs717996	12	82657951	0.44	0.35	2.38 × 10 ⁻⁴	0.40	0.40	8.28 × 10 ⁻¹	1
3	rs10484410	6	57104647	0.12	0.19	2.64 × 10 ⁻⁴	0.13	0.15	1.17 × 10 ⁻³	2.93 × 10 ⁻²
4	rs7962346	12	61952763	0.14	0.22	3.85 × 10 ⁻⁴	0.17	0.16	2.94 × 10 ⁻¹	1
5	rs10495266	1	225313639	0.10	0.16	5.19 × 10 ⁻⁴	0.13	0.14	2.85 × 10 ⁻¹	1
6	rs951901	12	56684,632	0.40	0.31	5.77 × 10 ⁻⁴	0.36	0.34	1.34 × 10 ⁻¹	1
7	rs2383784	9	28934277	0.20	0.13	6.58 × 10 ⁻⁴	0.17	0.17	7.37 × 10 ⁻¹	1
8	rs1873942	16	58553352	0.18	0.25	7.22 × 10 ⁻⁴	0.23	0.20	6.18 × 10 ⁻³	1.55 × 10 ⁻¹
9	rs2078539	16	60141739	0.35	0.26	7.74 × 10 ⁻⁴	0.30	0.30	8.30 × 10 ⁻¹	1
10	rs10494467	1	164817060	0.09	0.04	8.68 × 10 ⁻⁴	0.08	0.07	3.88 × 10 ⁻¹	1
11	rs4856497	3	80215335	0.43	0.52	9.35 × 10 ⁻⁴	0.50	0.47	6.62 × 10 ⁻²	1
12	rs1362941	5	59037578	0.30	0.22	1.19 × 10 ⁻³	0.27	0.25	1.02 × 10 ⁻¹	1
13	rs10489924	1	99231562	0.14	0.09	1.20 × 10 ⁻³	0.10	0.09	3.65 × 10 ⁻¹	1
14	rs866632	9	28845071	0.35	0.27	1.33 × 10 ⁻³	0.31	0.33	2.58 × 10 ⁻²	6.45 × 10 ⁻¹
15	rs694788	5	113954835	0.41	0.49	1.38 × 10 ⁻³	0.45	0.45	8.49 × 10 ⁻¹	1
16	rs7132697	12	66819108	0.32	0.25	1.56 × 10 ⁻³	0.30	0.29	5.33 × 10 ⁻¹	1
17	rs185793	15	45483437	0.39	0.31	1.67 × 10 ⁻³	0.33	0.34	1.65 × 10 ⁻¹	1
18	rs10483261	14	21426391	0.51	0.42	1.73 × 10 ⁻³	0.48	0.46	5.80 × 10 ⁻²	1
19	rs10483437	14	32905552	0.34	0.26	1.74 × 10 ⁻³	0.29	0.30	4.64 × 10 ⁻¹	1
20	rs2351010	5	96319685	0.47	0.39	1.75 × 10 ⁻³	0.42	0.44	1.09 × 10 ⁻¹	1
21	rs10487432	7	125458829	0.36	0.28	1.90 × 10 ⁻³	0.32	0.31	4.78 × 10 ⁻¹	1
22	rs920956	13	92901302	0.39	0.31	2.00 × 10 ⁻³	0.37	0.34	1.33 × 10 ⁻²	3.33 × 10 ⁻¹
23	rs7920803	10	28481267	0.30	0.23	2.11 × 10 ⁻³	0.25	0.24	2.37 × 10 ⁻¹	1
24	rs1439172	14	45797214	0.20	0.27	2.17 × 10 ⁻³	0.23	0.22	6.77 × 10 ⁻¹	1
25	rs4941920	13	38740373	0.33	0.25	2.19 × 10 ⁻³	0.29	0.30	8.32 × 10 ⁻¹	1

SNPs are ranked according to their genomic control (GC)-adjusted p-value (allelic Chi-squared test with one degree of freedom) in the screening panel. Minor allele frequencies (MAFs) are listed for unaffected controls (U) and sarcoidosis patients (SA). Bonferroni-corrected (Corr) p-values are shown for the analysis of the validation sample (panel B). Only variant rs10484410 remained significant after correcting for multiple testing (n=25). Positions are from National Center for Biotechnology Information build 36.1 [9]. For detailed results including genotype counts see table S1 in the online supplementary material. Chr: chromosome. #: 392 controls and 381 cases; †: 1,783 controls and 1,582 cases.

Fine mapping around rs10484410 (6p12.1)

In addition to the lead SNP rs10484410, 44 HapMap tagging SNPs (tagSNPs) were selected for the fine mapping of ~800 kb of the 6p12.1 region (see Methods section) carried out in panel C. 41 out of the 44 SNPs passed the abovementioned quality criteria (see Methods section). Nine markers yielded a p-value <0.05 in the analysis; detailed results, including genotype counts for the overall sarcoidosis panel, are shown in table S2 in the online supplementary material. The results were not corrected for multiple testing because the aim of the panel C analysis was to identify the probable 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 tables S2 and S3 in the online supplementary material).

Linkage disequilibrium (LD) between rs10484410 and the nine additional SNPs with a significant association from panel C varied greatly ($r^2=0.41-0.99$; see table S2 in the online

supplementary material). Of the 10 markers, SNP rs1411578 showed the strongest association ($p=6.64 \times 10^{-4}$). Marker rs1411578 (G>C) is located in exon 7 (3' untranslated region (UTR)) of the Ras-related protein Rab23 (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 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 (fig. 2b), we aimed to ensure that the association signal was limited to the RAB23 genomic region and did not extend further up- or downstream. We therefore genotyped the samples from panel C for another 46 HapMap

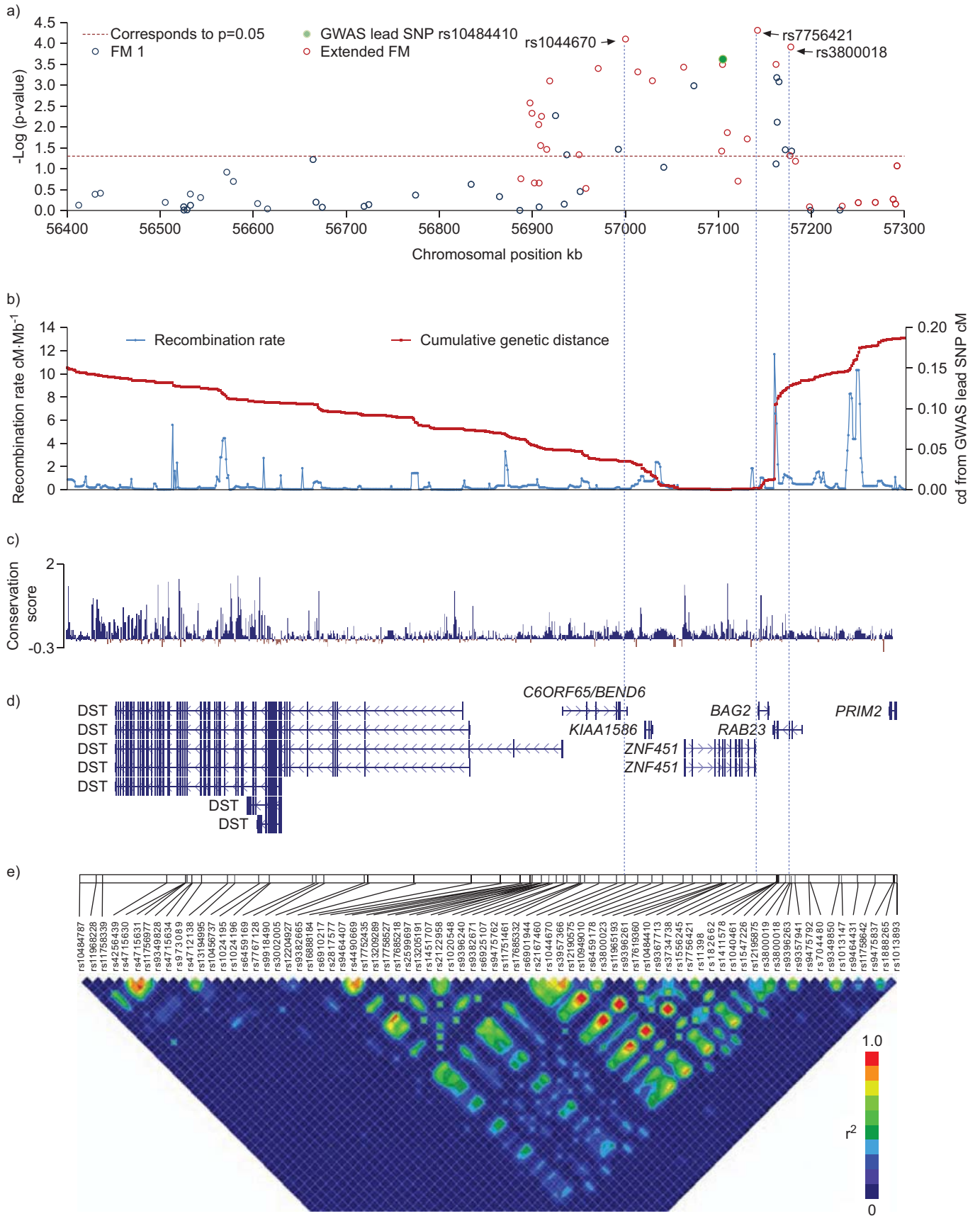


FIGURE 2. Legend is on the following page.

FIGURE 2. Regional association plot and structure of linkage disequilibrium (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 (green circle) in panel C. The most significantly associated single-nucleotide polymorphisms (SNPs) of the FM is highlighted by a blue vertical line. 44 tagging SNPs were genotyped in 1,806 controls and 1,837 sarcoidosis patients. 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 versus 392 controls; panel A) are shown. The red dotted line corresponds to a significance threshold of 0.05. Positions are in National Center for Biotechnology Information build 36.1 coordinates. For details, see tables S1 and S2 in the online supplementary material. b) Plotted recombination rate (in centimorgans (cM) per megabase (Mb)) and the cumulative genetic distance (cd) in cM. c) Sequence conservation score based on 44 different vertebrate species (taken from University of California at Santa Cruz (Santa Cruz, CA, USA) Genome Browser, Vertebrate Multiz Alignment and Conservation). d) The position and intron/exon structure of the underlying genes. *DST*: dystonin. e) Pair-wise LD measured as r^2 in the studied sample for the genotyped variants that passed the quality criteria, including the GWAS lead SNP rs10484410 ($n=76$).

CEU-based tagSNPs that covered a ~ 105 -kb region surrounding the lead SNP rs10484410 and the *RAB23* genomic region (table S3 in the online supplementary material).

Looking at the total mapped region (~ 900 kb), the most significantly 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 ($r^2 \geq 0.8$), which includes eight other associated variants ($p \leq 0.001$) that map to *RAB23* (rs11398, rs1411578, rs1547226 and rs3800018), *ZNF451* (rs6459178, rs17619360 and rs10484410) and to a noncoding region (rs12190575) (fig. 3). A nonsynonymous SNP (nsSNP) (rs1040461) in the *RAB23* gene that showed a significant association (nominal p-value (p_{nom})= 7.83×10^{-3}) with sarcoidosis as well was not part of this set of highly correlated markers.

We used the information criterion-based backward model selection of AIKAIKE [21] in a logistic regression analysis to

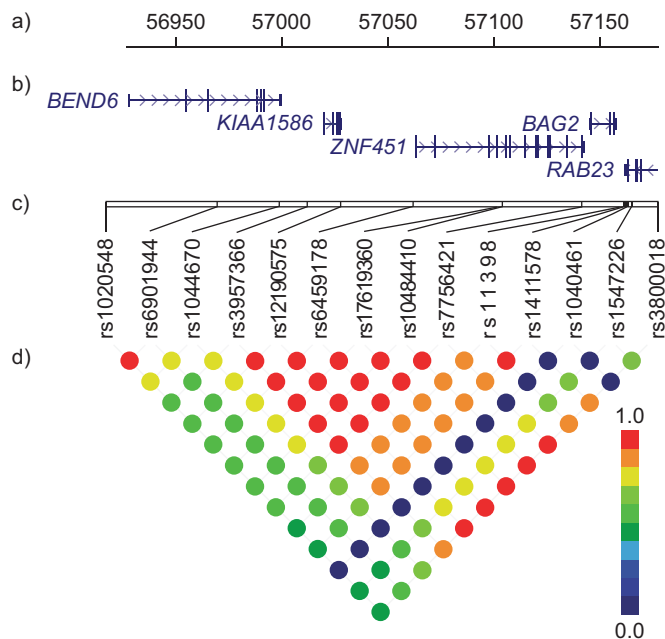


FIGURE 3. Overview of the linkage disequilibrium (LD) structure of the *6p12.1* locus, including the most significant sarcoidosis-associated single-nucleotide polymorphisms (SNPs) and the nonsynonymous marker rs1040461 (c.619G>A) in the *RAB23* gene. Pairwise LD between the most significant associated SNP rs7756421 and SNPs with a nominal p-value of <0.001 as well as the nonsynonymous 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 colour.

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 significantly 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 [22] and subsequently included the subpopulation-specific average values of the first six principal components (PCs) 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 ($p_{\text{nom}}=3.30 \times 10^{-3}$ and $p_{\text{nom}}=5.10 \times 10^{-3}$, respectively), while marker rs1040461 did not. However, after inclusion of the first six PCs in the model, rs1040461 showed a nominal significance of $p_{\text{nom}}=1.05 \times 10^{-2}$, which remained significant after correction for multiple testing, whereas the other two markers became nonsignificant ($p_{\text{rs10484410}}=1.40 \times 10^{-1}$ and $p_{\text{rs7756421}}=1.80 \times 10^{-1}$; table 2). Analysis of only German cases and controls yielded a similar result for rs1040461 ($p=1.20 \times 10^{-3}$ without adjustment for sampling location and $p=1.05 \times 10^{-2}$ with adjustment), but a minor decrease in significance for the other two markers ($p_{\text{rs10484410}}=6.90 \times 10^{-2}$ and $p_{\text{rs7756421}}=8.50 \times 10^{-2}$ without adjustment, and $p_{\text{rs10484410}}=1.40 \times 10^{-1}$ and $p_{\text{rs7756421}}=1.80 \times 10^{-1}$ 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*), uncharacterised protein *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 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 a few other tissues at an extremely low level. Interestingly, a high expression of *RAB23* also became apparent in small intestine and colonic mucosa.

Next, we analysed the expression levels of the candidate genes in cells derived from bronchoalveolar lavage (BAL) using

TABLE 2 Association analysis results

dbSNP	Position	MAF		No adjustment		After adjustment	
		U	SA	pLRT	OR (95% CI)	pLRT	OR (95% CI)
rs10484410	57104647	0.14	0.17	3.33×10^{-3}	1.24 (1.08–1.44)	1.36×10^{-1}	1.21 (NA)
rs7756421	57141968	0.14	0.17	5.05×10^{-3}	1.23 (1.07–1.43)	1.78×10^{-1}	1.20 (NA)
rs1040461	57163313	0.09	0.08	1.37×10^{-1}	1.16 (0.95–1.42)	1.05×10^{-2}	1.68 (NA)

Results are shown for the genome-wide association study lead single-nucleotide polymorphism (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 (ORs) are presented for the multiplicative model. Similar data were obtained in the genotypic and recessive model (data not shown). Variant rs1040461 showed a significant association with sarcoidosis after adjustment for population stratification. MAF: minor allele frequency; U: unaffected controls; SA: sarcoidosis patients; NA: not applicable.

quantitative real-time PCR and cDNA from sarcoidosis patients and controls (n=5 per group). As shown in figure 5, four out 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 nonparametric data). Compared with controls, BAL cells from sarcoidosis patients exhibited up to a three-fold increase in *RAB23* mRNA levels, further emphasising the potential involvement of this gene in the pathogenesis of sarcoidosis.

For further expression analysis see the expression quantitative trait loci (eQTL) results in the online supplementary material (fig. S5A–S7C).

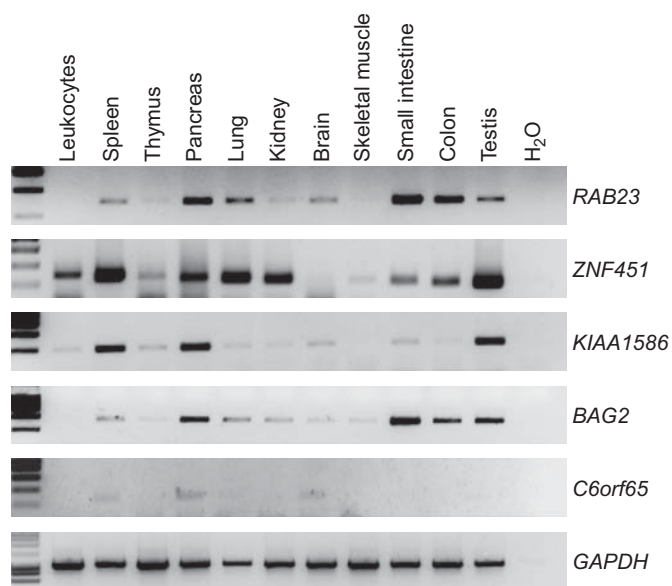


FIGURE 4. Tissue-specific expression profile of genes within the *6p12.1* risk locus. A human cDNA tissue panel was analysed for expression of the candidate genes using RT-PCR. Housekeeping gene *GAPDH* was used as an internal control.

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 <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 6,229) 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) 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, polar amino acid serine (G^{207}/S^{207}) in a non-domain-containing

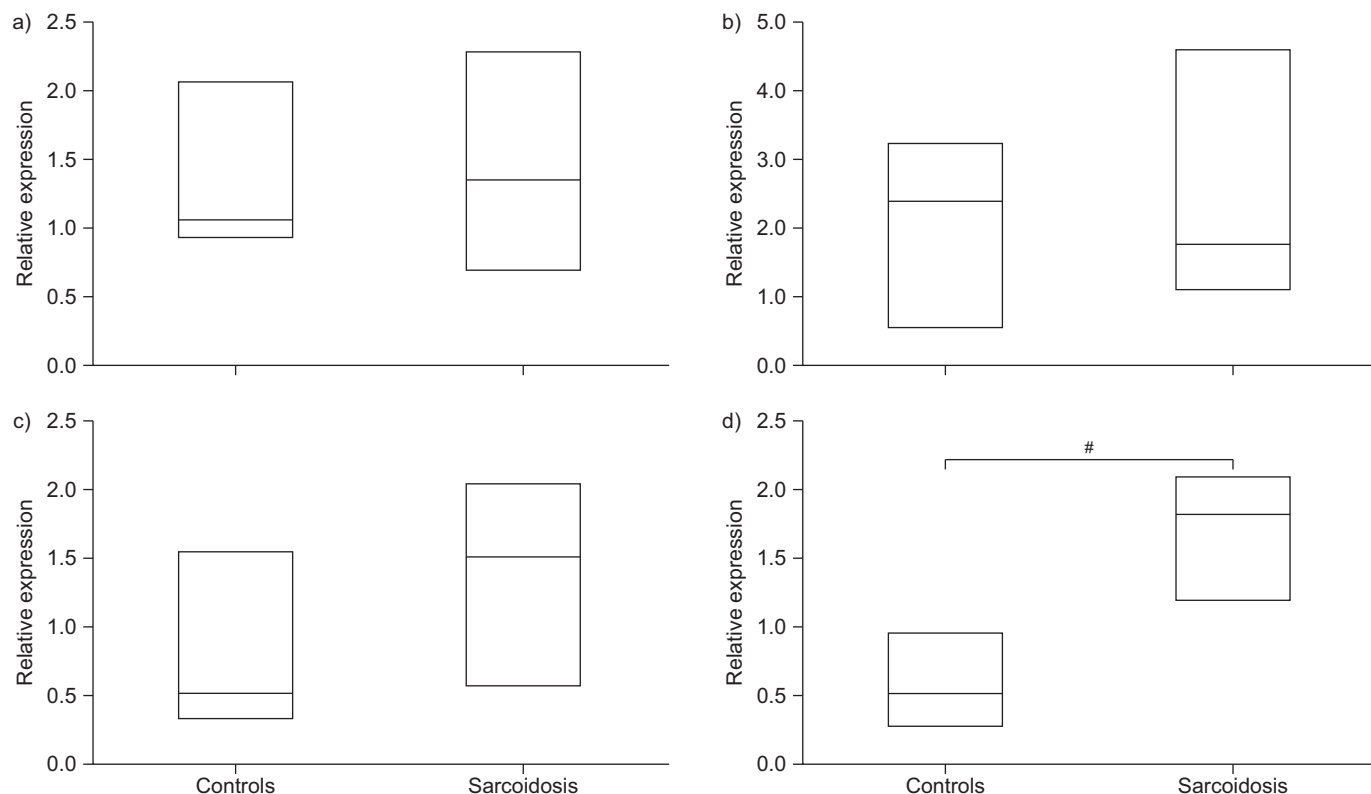


FIGURE 5. Relative expression of a) *BAG2*, b) *KIAA1586*, c) *ZNF451* and d) *RAB23* in cells from bronchoalveolar lavage (BAL). BAL cell samples were taken from sarcoidosis patients and matched healthy controls ($n=5$ per group). cDNA derived from BAL cells was assessed for relative expression of the candidate genes using quantitative real-time PCR and sequence-specific primer sets. Expression levels were normalised to those of housekeeping gene *beta-actin* (*ACTB*). Relative expression levels are presented as box plots representing the median and range of five measurements. Significance levels of the differences were calculated using the nonparametric Mann–Whitney U-test. Only *RAB23* expression was highly significantly different between patients and controls, displaying a three- to four-fold increase in relative expression. #: $p=3.20 \times 10^{-3}$.

region of the gene. Replication in a further panel revealed that marker rs1040461, but not rs7756421 or the GWAS lead SNP rs10484410, appeared 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 that 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 overexpression 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 aetiopathogenesis of sarcoidosis.

The *RAB23* gene belongs to the Rab family of 160 small guanosine triphosphatases that regulate intracellular trafficking of membrane-associated proteins [23, 24]. Based on

microarray expression data (UCSC GNF Expression Atlas 2 Data and Affymetrix All Exon Microarrays; see URL 8 in the online supplementary material), *RAB23* is more highly 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 earlier).

The potential involvement of this locus in sarcoidosis pathogenesis remains to be unravelled and can, as 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 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 [25, 26]. Since *RAB23* plays a role in the antibacterial activity of the endogenous autophagy machinery [27], it is biologically plausible that dysfunction might lead to 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 [27, 28] and, particularly, in antagonising sonic hedgehog (Shh) signal transduction in neural systems [29, 30]. Interestingly, the Shh pathway may also play a role in chronic lung fibrosis and immune system communication [31], 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-cell 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 [32]). 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 cells are essential for the inflammatory process and particular CD4+ T-cell subsets can be found at dramatically increased levels in BAL 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-cells 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, as yet undefined, genetic element. ZNF451, also known as COASTER, KIAA0576 and KIAA1702, is conserved across vertebrates, is ubiquitously expressed and has been suggested to act as a coactivator for steroid receptors; it might also be involved in transcriptional regulation [33]. The putative gene C6orf65 (BEND6) is a complex locus that appears to produce several proteins with no sequence overlap and of as yet unknown functions.

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

SUPPORT STATEMENT

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.

STATEMENT OF INTEREST

None declared.

ACKNOWLEDGEMENTS

The members of the GenPhenReSa Consortium are A. Günther (Lung Center, University of Giessen, Giessen, and Lung Clinic Waldhof-Elgershausen, Greifenstein, Germany), A. Dubaniewicz (Dept of Pneumology, Medical University of Gdansk, Gdansk, Poland), S. Pabst (Medical Clinic II, Dept of Pneumology, University of Bonn, Bonn, Germany), D. Bumbacea (Marius Nasta Institute of Pulmonology, Bucharest, Romania), A. Prasse and A. Grubanic

(Dept of Pneumology, University Medical Center, Albrecht Ludwigs University, Freiburg, Germany), P. Rottoli and E. Bargagli (Respiratory Diseases Section, Clinical Medicine and Immunology, University of Siena, Siena, Italy), P. Bresser (Dept of Pulmonology, Academic Medical Center, Amsterdam, the Netherlands), V. Poletti (Ospedale GB Morgagni, Dipartimento di Malattie dell'Apparato Respiratorio e del Torace UO Endoscopia Toracica, Forlì, Italy), M. Luisetti (Clinica Malattie Apparato Respiratorio, IRCCS Policlinico San Matteo-Pavia, University of Pavia, Pavia, Italy), and V. Vucinic and J. Videnovic-Ivanov (Medical School Belgrade, Belgrade, Serbia).

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

REFERENCES

- Bresnitz EA, Strom BL. Epidemiology of sarcoidosis. *Epidemiol Rev* 1983; 5: 124–156.
- Ezzie ME, Crouser ED. Considering an infectious etiology of sarcoidosis. *Clin Dermatol* 2007; 25: 259–266.
- McGrath DS, Goh N, Foley PJ, et al. Sarcoidosis: genes and microbes – soil or seed? *Sarcoidosis Vasc Diffuse Lung Dis* 2001; 18: 149–164.
- Muller-Quernheim J, Schurmann M, Hofmann S, et al. Genetics of sarcoidosis. *Clin Chest Med* 2008; 29: 391–414.
- Sverrild A, Backer V, Kyvik KO, et al. Heredity in sarcoidosis: a registry-based twin study. *Thorax* 2008; 63: 894–896.
- Iannuzzi MC, Rybicki BA. Genetics of sarcoidosis: candidate genes and genome scans. *Proc Am Thorac Soc* 2007; 4: 108–116.
- Valentonyte R, Hampe J, Huse K, et al. Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nat Genet* 2005; 37: 357–364.
- Rybicki BA, Walewski JL, Maliarik MJ, et al. The BTNL2 gene and sarcoidosis susceptibility in African Americans and Whites. *Am J Hum Genet* 2005; 77: 491–499.
- Hofmann S, Franke A, Fischer A, et al. Genome-wide association study identifies ANXA11 as a new susceptibility gene for sarcoidosis. *Nat Genet* 2008; 40: 1103–1106.
- Franke A, Fischer A, Nothnagel M, et al. Genome-wide association analysis in sarcoidosis and Crohn's disease unravels a common susceptibility locus on 10p12.2. *Gastroenterology* 2008; 135: 1207–1215.
- Valentonyte R, Hampe J, Croucher PJ, et al. Study of C–C chemokine receptor 2 alleles in sarcoidosis, with emphasis on family-based analysis. *Am J Respir Crit Care Med* 2005; 171: 1136–1141.
- 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, February 1999. *Am J Respir Crit Care Med* 1999; 160: 736–755.
- Fischer A, Nothnagel M, Franke A, et al. Association of inflammatory bowel disease risk loci with sarcoidosis, and its acute and chronic subphenotypes. *Eur Respir J* 2011; 37: 610–616.
- Krawczak M, Nikolaus S, von Eberstein H, et al. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet* 2006; 9: 55–61.
- Nebel A, Croucher PJP, Stiegeler R, et al. No association between microsomal triglyceride transfer protein (MTP) haplotype and longevity in humans. *Proc Natl Acad Sci USA* 2005; 102: 7906–7909.
- Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; 38: 904–909.

- 17 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; 447: 661–678.
- 18 Clayton D, Leung HT. An R package for analysis of whole-genome association studies. *Hum Hered* 2007; 64: 45–51.
- 19 Li Y, Pabst S, Kubisch C, et al. First independent replication study confirms the strong genetic association of ANXA11 with sarcoidosis. *Thorax* 2010; 65: 939–940.
- 20 Li Y, Wollnik B, Pabst S, et al. BTNL2 gene variant and sarcoidosis. *Thorax* 2006; 61: 273–274.
- 21 Akaike, H. Information theory and the extension of the maximum likelihood principle. In: Petrov VN, Csaki F, eds. Second International Symposium on Information Theory. Budapest, Akailseoniai-Kiudo, 1973; pp. 267–281.
- 22 Lao O, Lu TT, Nothnagel M, et al. Correlation between genetic and geographic structure in Europe. *Curr Biol* 2008; 18: 1241–1248.
- 23 Eathiraj S, Pan X, Ritacco C, et al. Structural basis of family-wide Rab GTPase recognition by rabenosyn-5. *Nature* 2005; 436: 415–419.
- 24 Pereira-Leal JB, Seabra MC. Evolution of the Rab family of small GTP-binding proteins. *J Mol Biol* 2001; 313: 889–901.
- 25 Eishi Y, Suga M, Ishige I, et al. Quantitative analysis of mycobacterial and propionibacterial DNA in lymph nodes of Japanese and European patients with sarcoidosis. *J Clin Microbiol* 2002; 40: 198–204.
- 26 Song Z, Marzilli L, Greenlee BM, et al. Mycobacterial catalase-peroxidase is a tissue antigen and target of the adaptive immune response in systemic sarcoidosis. *J Exp Med* 2005; 201: 755–767.
- 27 Smith AC, Heo WD, Braun V, et al. A network of Rab GTPases controls phagosome maturation and is modulated by *Salmonella enterica* serovar Typhimurium. *J Cell Biol* 2007; 176: 263–268.
- 28 Guo A, Wang T, Ng EL, et al. Open brain gene product Rab23: expression pattern in the adult mouse brain and functional characterization. *J Neurosci Res* 2006; 83: 1118–1127.
- 29 Evans TM, Simpson F, Parton RG, et al. Characterization of Rab23, a negative regulator of sonic hedgehog signaling. *Methods Enzymol* 2005; 403: 759–777.
- 30 Eggenschwiler JT, Espinoza E, Anderson KV. Rab23 is an essential negative regulator of the mouse Sonic hedgehog signalling pathway. *Nature* 2001; 412: 194–198.
- 31 Stewart GA, Hoyne GF, Ahmad SA, et al. Expression of the developmental Sonic hedgehog (Shh) signalling pathway is up-regulated in chronic lung fibrosis and the Shh receptor patched 1 is present in circulating T lymphocytes. *J Pathol* 2003; 199: 488–495.
- 32 Fitch PM, Wakelin SJ, Lowrey JA, et al. Shh expression in pulmonary injury and disease. In: Ruiz I Altaba A, ed. Hedgehog-Gli Signaling in Human Disease. New York, Springer, 2006; pp. 119–128.
- 33 Karvonen U, Jaaskelainen T, Rytinki M, et al. ZNF451 is a novel PML body- and SUMO-associated transcriptional coregulator. *J Mol Biol* 2008; 382: 585–600.