

Genome-wide association analysis reveals 12q13.3 q14.1 as new risk locus for sarcoidosis

Sylvia Hofmann*, Annegret Fischer*, Michael Nothnagel, Gunnar Jacobs, Benjamin Schmid, Michael Wittig, Andre Franke, Karoline I. Gaede, Manfred Schürmann, Martin Petrek, Frantisek Mrazek, Stefan Pabst, Martin Grohé, Johan Grunewald, Marcus Ronninger, Anders Eklund, Philip Rosenstiel, Kerstin Höhne, Gernot Zissel, Joachim Müller-Quernheim, Stefan Schreiber

* Both authors contributed equally to this work.

Online Data Supplement

Contents

1. Methods	3
Patient and control subjects	3
Genotyping and quality control	4
SNP selection and statistical analysis	5
mRNA isolation from BAL cells and RT-PCR (BAL panel I)	5
2. Results	7
Known risk loci in screening panel A (GWAS)	7
Cell composition in BAL panel II	7
3. Figures	8
Figure E1: Study design	8
Figure E2: Density plot of “outlier detection“-analysis	9
Figure E3: Signal intensity plots in the GWAS for the 99 SNPs selected for validation	10
Figure E4: Quantile-quantile (Q-Q) plot of the association test statistic	11
Figure E5: Disease association of SNPs in the chromosome 6p22.33–p21.32 region	12
Figure E6: Regional plot of the fine-mapped region around the lead SNP	13
Figure E7: Electropherogram for the two novel SNPs detected by Sanger sequencing	14
Figure E8: Expression of the eight candidate genes	15
Figure E9: Quantitative analysis of OS9 expression rs1050045	17
Figure E10: Percentage of alveolar macrophages in BALs	18
Figure E11: CD4+/CD8+ T lymphocytes in the BAL.	19
4. Tables	20
Table E1: SNPs used in the backward model selection in a logistic regression model.	20
Table E2: Primer used for the detection of variants in the OS9 regions	21
Table E3: RT-PCR primer used for tissue-specific expression and BAL panel I	22
Table E4: Primer used for quantification of OS9 mRNA expression (BAL panel II)	22
Table E5: Results of the GWAS and of the validation stage	23
Table E6: Results for the 53 SNPs genotyped to fine-map 12q13.3-q14.1 in panel C	27
Table E7: Results of the haplotype analysis	29
Table E8: Prediction of SNP function	29
5. References	31

1. Methods

Patient and control subjects

German patients of the screening panel A (n = 640) and of the validation panel B (n = 1,664) were actively contacted through the German Sarcoidosis Patients Organization, specialized hospitals, practitioners, and by calls for study participation that were published via health insurance institutions. In the latter case, histological confirmation of the diagnosis was necessary for inclusion of individuals. All patients completed a questionnaire on the course of disease. Patients' physicians were contacted to provide clinical radiology and laboratory data required to confirm the diagnosis of sarcoidosis. Patients were classified as having either chronic or acute sarcoidosis or Löfgren's syndrome. All the recruitments have been done in a retrospective setting and the clinical presentation and course of disease was known according to standard definitions as previously described [1, 2]. All centers followed that recruitment strategy later described in detail in our publication on sarcoidosis phenotypes for scientific purposes [3]. Only patients who could be unequivocally categorized to acute or chronic were recruited to those subphenotypes.

Panel A included 191 "acute" and 401 "chronic" patients (mean age 50.8 years, ± 11.7 , 48% male), while panel B comprised 563 individuals with "acute" and 947 individuals with "chronic" sarcoidosis (mean age 60.8 years, ± 11.7 , 38% male); both panels comprised a total of 123 patients with Löfgren's syndrome. The remaining sarcoidosis patients could not be classified according to the course of the disease, but showed other phenotypes, e.g. solely cutaneous sarcoidosis, or were detected by radiography for other reasons and had no specific complaints.

Healthy German control individuals with a median age of 57.7 years (± 15.1 years) in panel A (n=1,256; 56% male, 44% female) and 52.5 years (± 15.8) in panel B (n= 2,932; 45% male, 55% female) were obtained from the population-derived pool of controls individuals in the PopGen project, identified on the basis of the local population registry. Detailed information on recruitment is given in the respective publication from the PopGen biobank [4]. For both cases and controls, the study was restricted to probands of German ancestry, i. e. only individuals whose parents were born in Germany. Panel A almost completely overlaps with the panel that was used in the previously published sarcoidosis 5.0 screen and to some extent with the panel used in the SNP-scan in 2005 respectively [2, 5].

Swedish patients of panel C-III were referred to the outpatient clinic at the Pulmonary Division at the Karolinska University Hospital, Solna, Sweden, where they were investigated

by clinicians and radiologists with long experience of interstitial pulmonary disease. All patients were diagnosed with sarcoidosis through typical clinical and radiological manifestations, findings at bronchoscopy with bronchoalveolar lavage (BAL) including an elevated CD4/CD8 BAL cell ratio, and positive biopsies, using the criteria as outlined by the World Association of Sarcoidosis and other Granulomatous disorders (WASOG) [6]. Swedish controls were contributed by the Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA) study [7].

Fine-mapping was carried out in 1,829 sarcoidosis patients, comprising all patients from panel B and parts of panel A, including 597 “acute” and 1,055 chronically affected patients, and in 1,465 controls from panel B. Informed written consent was obtained from all study participants and all collection protocols were approved by the institutional review committees of the participating centers.

Genotyping and quality control

Panel A was genotyped by an Affymetrix service facility for a total of 934,968 SNPs using the Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA, USA) and genotypes were generated using the BRLMM-p algorithm. Samples with either more than 10% missing genotypes ($n = 0$), or showing cryptic relatedness to other study participants ($n = 23$), or showing excess of genetic dissimilarity to the other subjects as evaluated by the identical-by-state (IBS) analysis and a multiple scaling analysis using the European HapMap reference data (CEU; CEPH Utah residents with ancestry from northern and western Europe; phase II, build 36; $n = 3$) [8] were removed from the data set. The plot for outlier detection and the quantile-quantile plot (data supplement Figures E2 and E4) were created using R v2.8.0 [9]. All gender assignments could be verified by reference to the proportion of heterozygous SNPs on the X chromosome using the sex check option implemented in PLINK. The following procedures were conducted to assure marker quality using PLINK v.1.06 [10]: SNPs that had a low genotyping rate ($< 95\%$ in cases or controls), that were monomorphic or rare (minor allele frequency $< 2\%$ in cases or controls), or that deviated from Hardy-Weinberg equilibrium (HWE) in the control sample ($p_{\text{HWE}} \leq 0.01$) were excluded ($n = 257,349$; 27.5%). Tests for HWE were performed using an exact test [11]. The number of excluded SNPs might be mainly due to genotyping errors and is in line with that of other published Affymetrix-based GWAS [12-15]. In addition, cluster plots for all SNPs for validation in panel B, were visually inspected for correct genotype assignment. Genotyping of panel B was carried out in-house using ligation-based SNPlex™ technology (Applied Biosystems, Foster City, CA, USA) as previously described [16, 17]. Genotype assignments were visually inspected using

the Genemapper 4.0 software (Applied Biosystems). TaqMan[®] SNP Genotyping (Applied Biosystems) was used to genotype marker rs1050045 in panel B for technical validation. All process data were written to and administered by a previously described database-driven laboratory information management system (LIMS) [18]. Quality control procedures for panels C, D and E were identical, except no evaluation of the IBS status was performed in these panels.

SNP selection and statistical analysis

Statistical analysis of genotype data was carried out using PLINK v.1.06 [10]. In the entire experiment, single-marker case-control analyses were performed using χ^2 tests with one degree of freedom (multiplicative model). Logistic regression was carried out also in R v2.10.1, using the glm function from the stats package. Backward model selection using Akaike's Information Criterion (AIC) [19] was applied using the step function in R from the same package. Linkage disequilibrium (LD) plots were generated using GOLD [20]. Haplotype analysis was conducted by performing a score test assuming additive effect using the haplo.stats R package [21].

The SNPs for validation were selected upon:

- LD ($r^2 > 0.5$) with at least one neighboring SNP (distance < 250 kb) with a p value $< 1.0 \times 10^{-3}$ in the case-control analysis using the “clumping” option in PLINK.
- Fulfillment of the quality criteria mentioned above, including correct genotype assignment regarding the index as well as the neighboring SNP(s).
- Selection of the 99 top-ranking SNPs by uncorrected p value in the case-control analysis ($p \leq 9.05 \times 10^{-4}$) based on power estimation (power ≥ 61 %) using PS power and sample size calculation v2.1.30 [22].
- SNPs for fine-mapping were selected from the European HapMap reference (CEU) via pairwise tagging using Haploview [23] by using the following criteria: minor allele frequency $> 1\%$, tagging threshold $r^2 > 0.8$, $p_{HWE} > 0.01$ and less than two Mendelian errors.

mRNA isolation from BAL cells and quantitative real-time (qRT)-PCR (BAL panel I and II)

Bronchoalveolar lavage (BAL) cell samples were taken from 4 patients with active, non-acute sarcoidosis without any medical treatment at the time of BAL and 4 unmedicated, unaffected individuals without infection and non-malignant at time of lavage. The diagnosis of non-acute

sarcoidosis was established retrospectively in accordance with previously defined criteria [6] including detection of noncaseating granuloma by transbronchial biopsies. All sarcoidosis patients showed clinical signs for an active disease. All BAL cell samples included in this study were matched by their portion of alveolar macrophages. Thus, alterations merely associated with different cell composition could be excluded. All patients and matched controls gave their informed consent to the study. BAL was performed as previously described [24]. The cell suspensions used in this study contained >92% alveolar macrophages. The viability of the cells was 95%.

Total RNA was isolated from snap-frozen BAL cell using a commercial kit (RNeasy, Qiagen, Hilden, Germany). For detection of mRNA levels of *OS9*, *AGAP2*, *TSPAN31*, *CDK4*, *MARCH9*, *CYP27B1*, *METTL1* und *FAM119B*, cDNA was synthesized from 500 ng of total RNA using the Advantage RT-for-PCR kit (Clontech Laboratories, Palo Alto, CA, USA) according to the manufacturer's protocol. SYBR Green PCR Master Mix (Applied Biosystems) and the 7900HT Fast Real Time PCR system was used for qRT-PCR using the same target-specific primers as described above. Transcript amounts were normalized to *GAPDH* mRNA levels. Relative expression levels of the target genes were checked for significant differences between sarcoidosis patients and unaffected individuals (n = 4 each) using non-parametric Mann-Whitney U test using the Graphpad software (Graphpad, Inc.; La Jolla, CA).

Total RNA from BAL cells of panel II was isolated using Trizol (Invitrogen, Carlsbad, CA) and reverse transcribed using StrataScript reverse transcriptase (Stratagene, La Jolla, CA) with oligo-dT primers (Biomers, Ulm, Germany) for 1 h at 50°C. Specific primers for human OS9 and GAPDH were designed using Primer3 software [25], Amplify1.2 software (University of Wisconsin, USA; <http://engels.genetics.wisc.edu/amplify>) using LocusLink and GenBank databases (National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/LocusLink/index.html>). Primer sequences used for panel II are depicted in table E7. All primers were intron-spanning and synthesized by biomers (Biomers.net, Ulm, Germany). Real time PCR was performed with the iQ SYBR Green SuperMix, iCycler thermocycler, and iCycler iQ 3.0 software (Bio-Rad Laboratories GmbH, Germany) according to the manufacturer's protocol. To control for specificity of the amplification products, a melting curve analysis was performed. No amplification of nonspecific products was observed in any of the reactions. A threshold cycle value (Ct) was calculated and used to compute the relative level of specific mRNA by the following formula: "relative expression = $2^{(Ct_{GAPDH}-Ct_{CCL18})}$ x 10,000" for each cDNA sample.

2. Results

Known risk loci in screening panel A (GWAS)

The known risk genes *BTNL2* and *ANXA11* (represented by markers rs2076530 and rs1953600, respectively; [2, 5]) were associated with acute as well as with chronic sarcoidosis in panel A (acute: $p = 1.39 \times 10^{-4}$ and $p = 6.07 \times 10^{-5}$; chronic: $p = 2.75 \times 10^{-11}$ and $p = 9.55 \times 10^{-7}$) in the single-marker case-control analysis. The *Rab23* region (rs3957366), which was identified using the Affymetrix 100k Gene Chip, as well as the common susceptibility locus for sarcoidosis and Crohn's disease on chromosome *10p12.2* (rs1398024) yielded association results, which are consistent with those obtained in the previous studies ($p = 1.36 \times 10^{-3}$ and $p = 0.04$, respectively; [26, 27]).

Cell composition in BAL panel II

The BAL cell composition of *OS9*-positive patients disclosed a higher percentage of lymphocytes compared to controls ($34 \pm 23\%$ versus $9 \pm 4\%$, $p < 2 \times 10^{-2}$), however, the lymphocyte percentage of *OS9*-negative patients was even higher ($52 \pm 21\%$, $p < 2 \times 10^{-3}$). Vice versa, the percentage of alveolar macrophages was highest in controls ($88 \pm 7\%$), lower in *OS9*-positive patients ($63 \pm 23\%$; $p < 2 \times 10^{-2}$) and lowest in *OS9*-negative patients ($44 \pm 22\%$; $p < 2 \times 10^{-3}$, Figure E9). Most interestingly, compared with controls the *OS9*-negative group disclosed a significantly increased percentage of HLA-DR-positive lymphocytes ($5 \pm 3\%$ versus $38 \pm 25\%$; $p < 5 \times 10^{-2}$, Figure E10) indicating an exaggerated T-cell activation in this group. No significant differences could be detected between the *OS9*-positive patient group ($12 \pm 11\%$) and controls.

Figures

Figure E1: Study design. The study followed a multi-stage design, which includes the analysis of the genome-wide SNP data in a screening phase (stage 1) and the validation of most promising candidates in an independent sample (stage 2). Replication in further independent populations confirmed rs1050045 at chromosome 12q13.3-q14.1 as a risk factor for sarcoidosis. Fine-mapping, mutation detection and expression studies point to *OS9* as the likely underlying risk gene. The method used for genotyping is given in italics, and the number of cases and controls is stated in brackets.

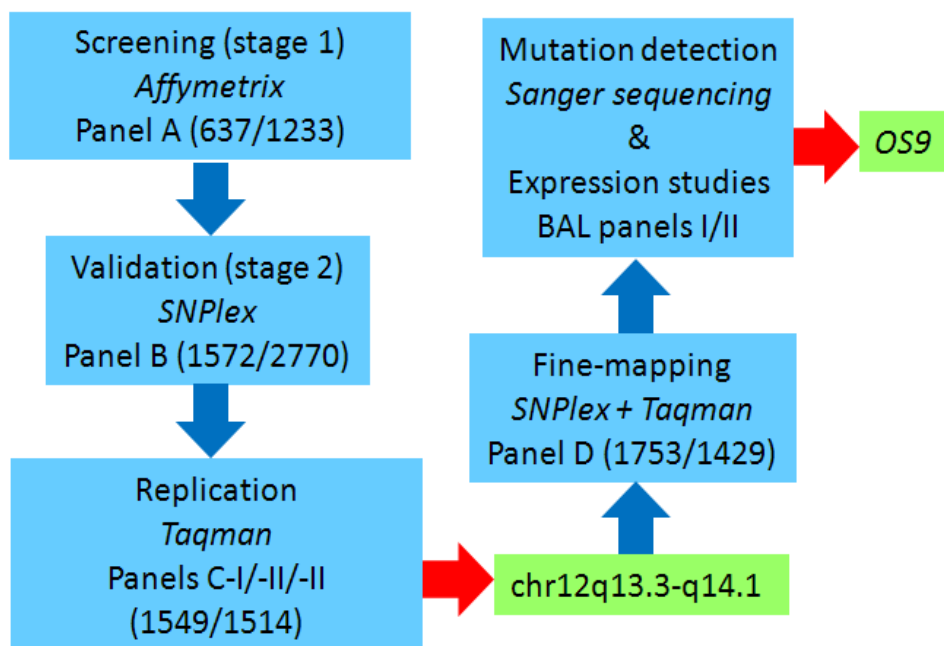


Figure E2: Density plot of pairwise IBS values for screening panel A samples after removal of samples showing cryptic relatedness to other samples. Identical by state (IBS) values were computed between all pairs of samples over all markers that met eligibility criteria on Hardy-Weinberg equilibrium and genotype call rate using PLINK v1.06. Sample outliers were identified by comparing the distribution of IBS values from each individual with the combined IBS distribution of the entire population. 23 individuals were excluded for cryptic relatedness ($IBS > 0.8$) and three individuals were removed due to clear genetic dissimilarity to the majority of the study participants ($IBS < 0.76$).

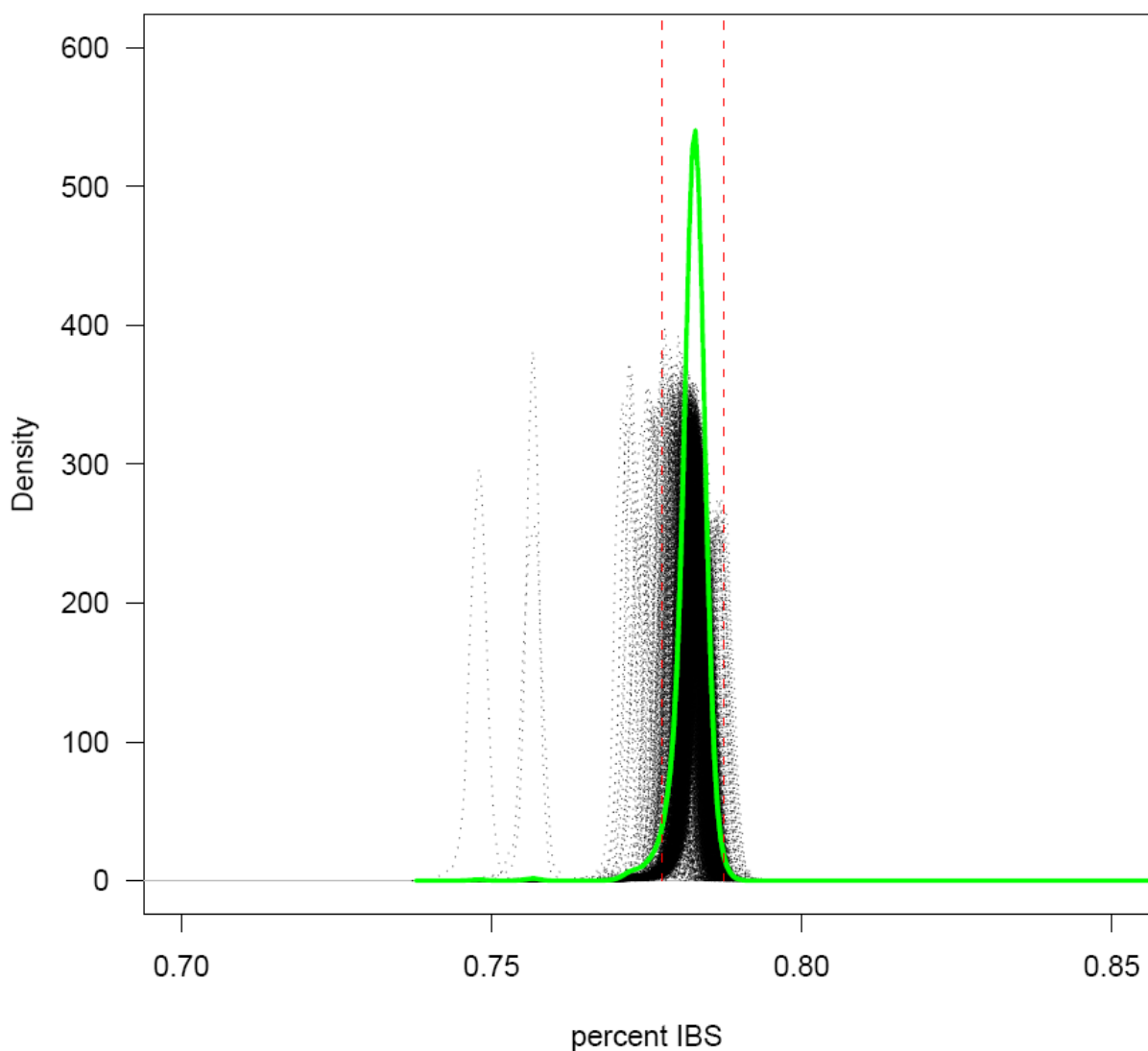


Figure E3: Signal intensity plots in the GWAS for the 98 SNPs selected for validation. Scatter plots of normalized summary probe intensities for the lead SNP rs1050045 (= SNP_A-4304028). Each point represents one individual and is colored according to the genotype assignment by the calling algorithm (blue or red: homozygous for one of the two alleles; green: heterozygous; black circle: 'null' or missing call). Alleles are given according to the Affymetrix annotation table.

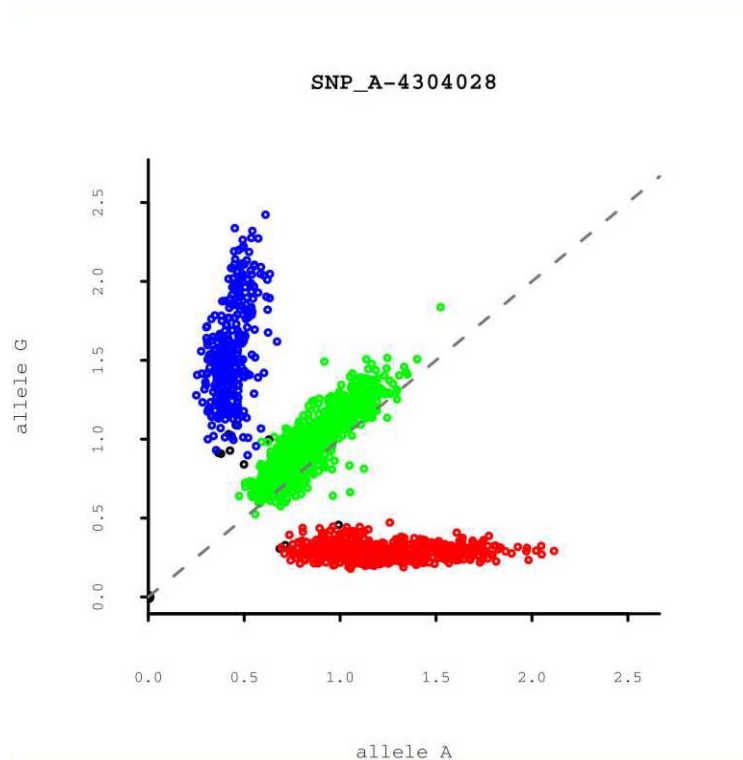


Figure E4: Quantile-quantile (Q-Q) plot of the association test statistic in panel A. The figure is based on all 677,619 SNPs that passed the quality control. Blue dots: whole marker set. Red dots: after exclusion of 2,480 SNPs located in the extended MHC region. The overdispersion of the association test statistic was estimated to be $\lambda_{GC} = 1.15$. The plot was created using the function `qq.chisq` of the R-package `snpMatrix` v1.0.33 (<http://www-gene.cimr.cam.ac.uk/clayton/software/>).

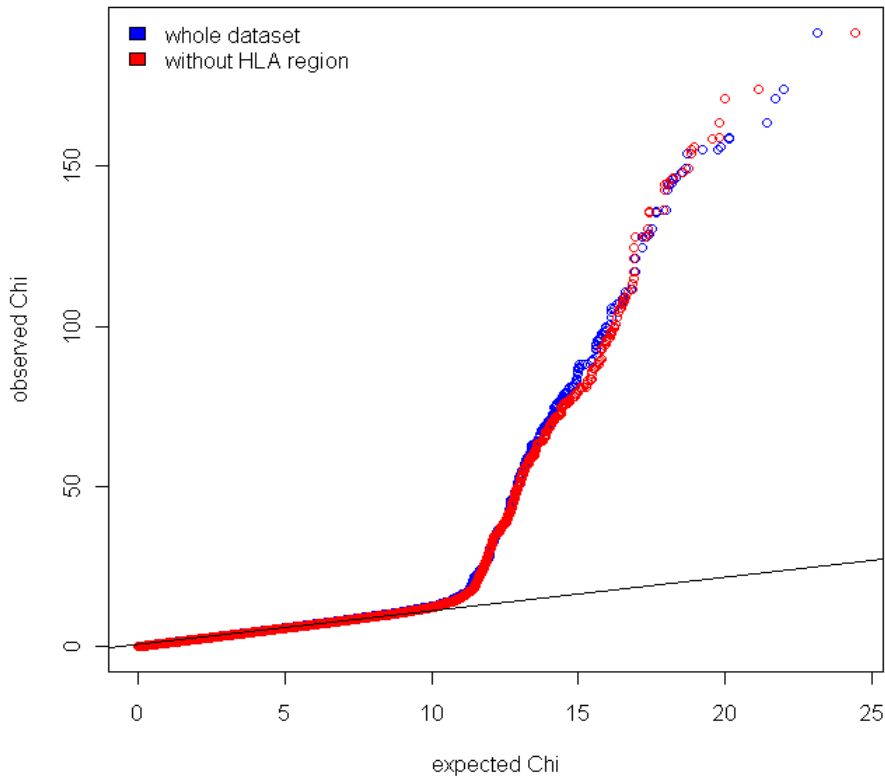


Figure E5: Disease association of SNPs in the chromosome 6p22.33–p21.32 region, obtained from the GWAS, and overview of the genetic structure of the region around the most significantly associated SNPs.

(A) Negative natural logarithm of the P values (allele-based χ^2 test) from the GWAS obtained in the region between positions 24,000 kb and 36,000 kb on chromosome 6. (B) Known genes in the region between positions 32,000 kb and 34,000 kb, including *BTNL2*, *HLA-DRA* and *HLA-DPB1*. The lower panel gives an overview of the linkage disequilibrium (LD) structure of the region, as measured by D' . The blocks were derived from the Caucasian HapMap genotypes. The coordinates refer to the NCBI NCBI36/hg18.

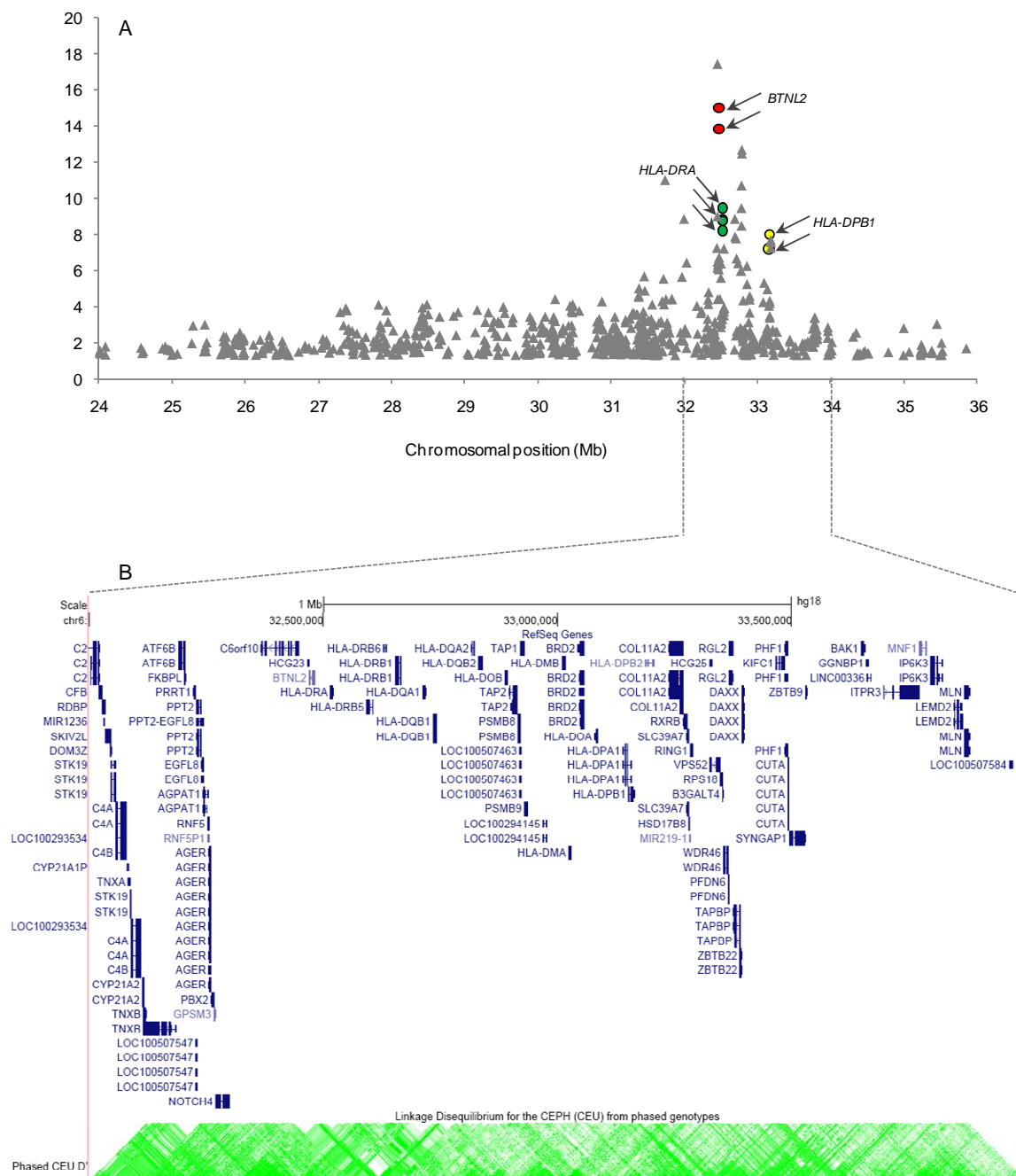


Figure E6: Regional plots of the fine-mapped region around the lead SNP rs1050045.

(A): Nominal $-\log_{10} p$ values obtained in fine-mapping panel D. Association signal are shown for the sarcoidosis and the acute and chronic subphenotypes. (B): Recombination rate intensity (cM/Mb) and cumulative genetic distance (cM) from the lead SNP according to HapMap (Phase I + II, build 36). (C): Annotated Refseq genes. (D): Linkage disequilibrium structure (r^2 in panel D controls).

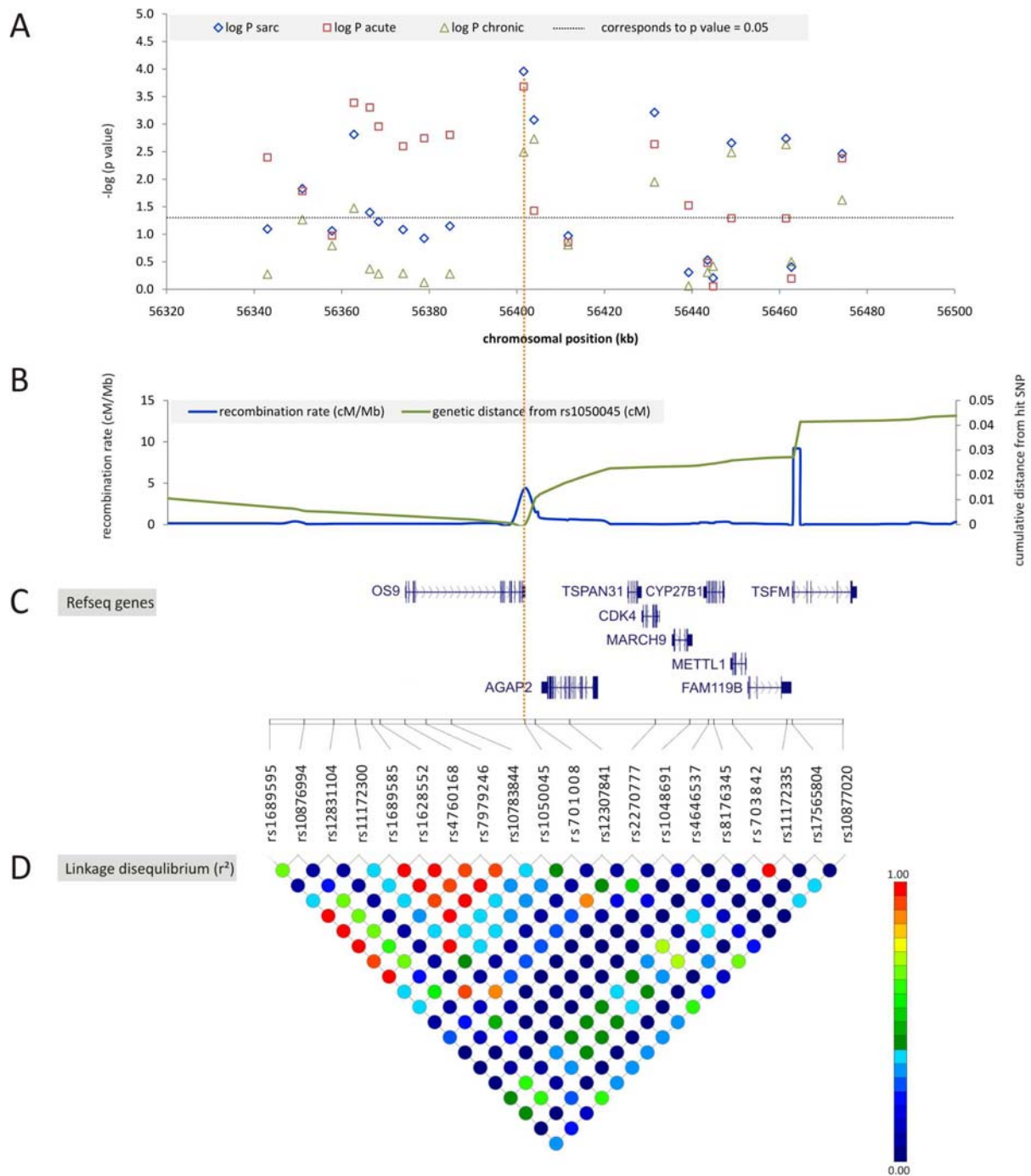
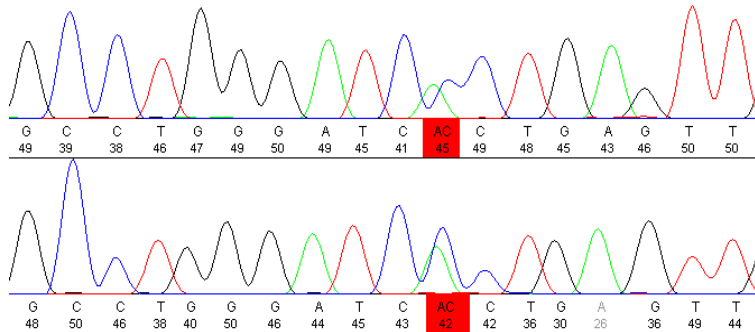


Figure E7: Electropherogram for the two novel SNPs detected by Sanger sequencing.

Plots are shown for the forward and reverse sequencing read.

(a) OS9-SNP1; chr12: 58,088,665 (hg18)



(b) OS9-SNP2; chr12: 58,114,081 (hg18)

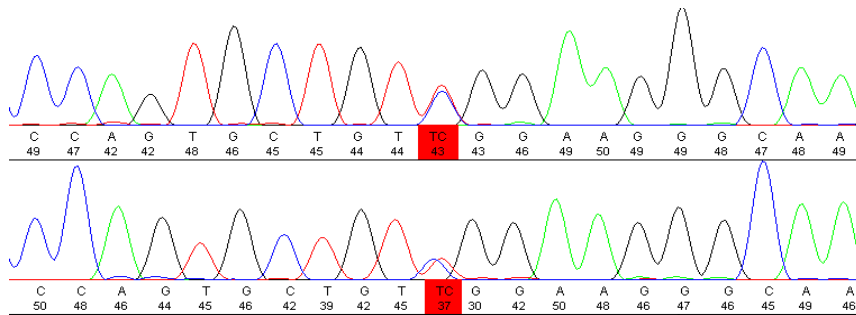
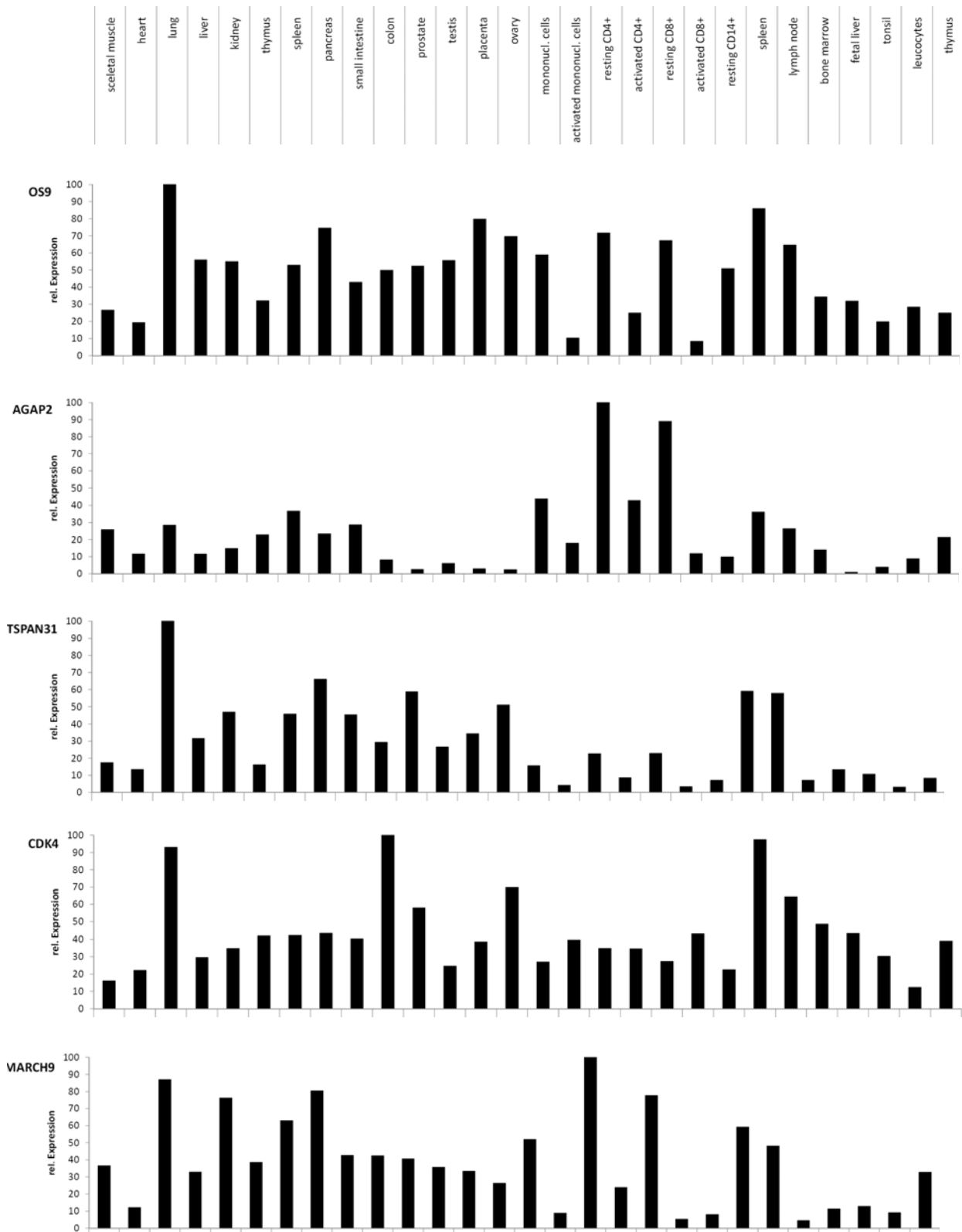


Figure E8: Expression of eight candidate genes in various healthy tissues and immune cells. Transcription levels of *OS9*, *AGAP2*, *TSPAN31*, *CDK4*, *MARCH9*, *CYP27B1*, *METTL1* and *FAM119B* were assessed using semi-quantitative PCR and were normalized on *GAPDH* mRNA levels.



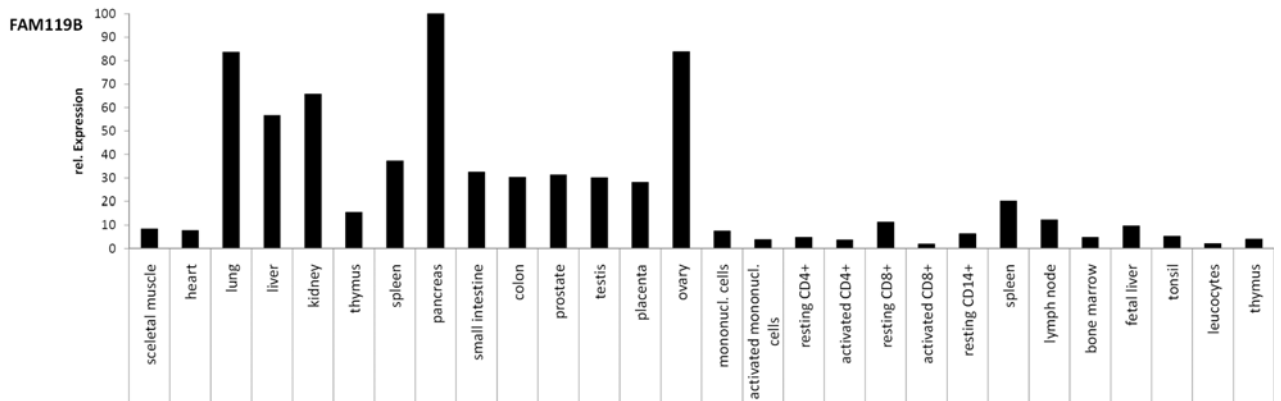
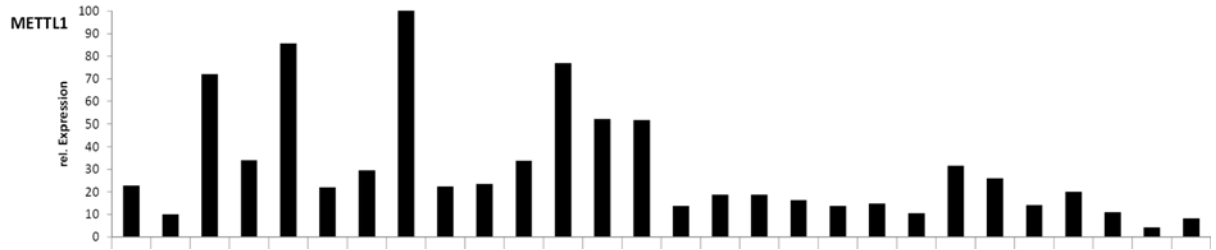
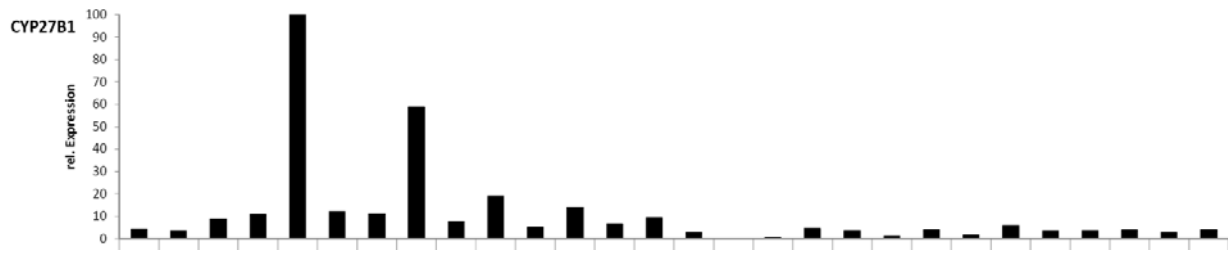


Figure E9: Quantitative analysis of *OS9* expression in BAL cells. Transcription levels of *OS9* in BAL cells were obtained for controls (C) and sarcoidosis patients from BAL panel II (see Material and Method section). In 8 out of 46 patients no *OS9* mRNA could be detected (negative). Differences in the *OS9* expression level were significant between *OS9*-positive (n = 38) compared to *OS9*-negative patients (n = 8) and compared to controls (n = 8).

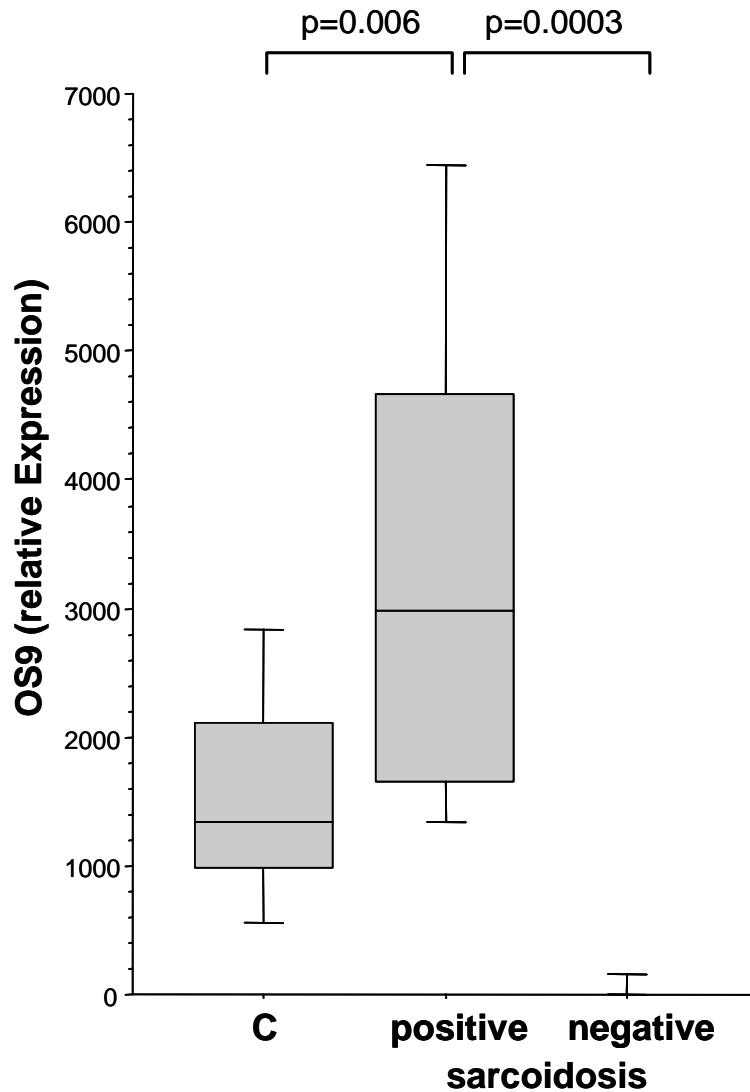


Figure E10: Percentage of alveolar macrophages in BALs of *OS9*-positive and *OS9*-negative patients with sarcoidosis and controls (C) in BAL panel II. BAL cell cells were fixed on stained by May-Grünwald-Giemsa and 300 cells were counted using a light microscope.

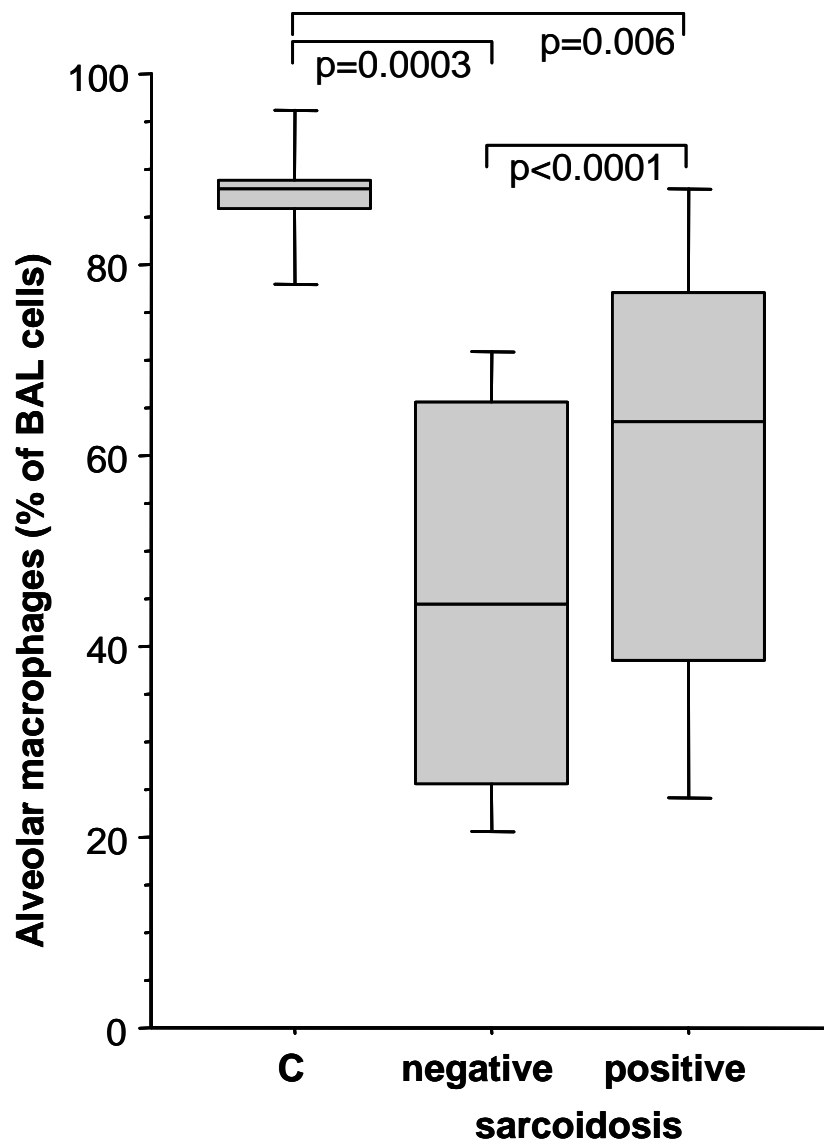
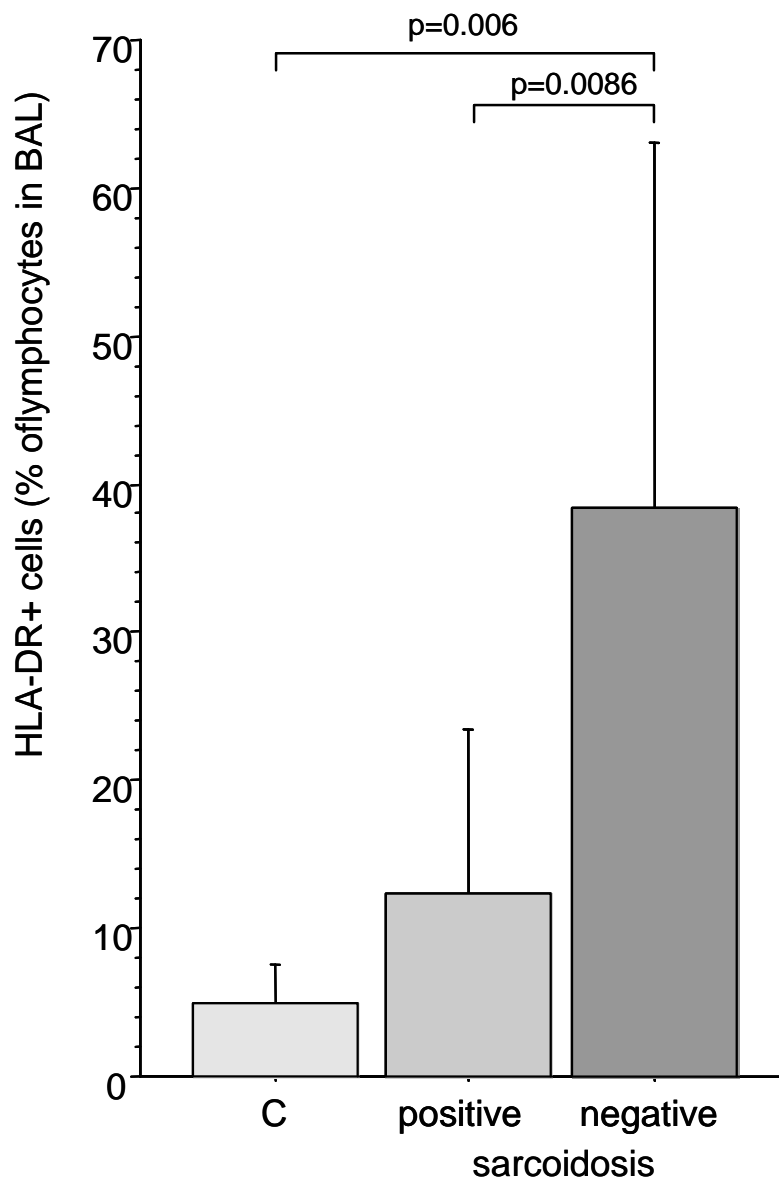


Figure E11: HLA-DR⁺ T lymphocytes in the BAL of OS9-positive and OS9-negative patients with sarcoidosis and controls in BAL panel II. Isolated RNA of BAL cells from controls (C; n=4) and sarcoidosis patients either positive (n=33) or negative (n=8) was transcribed in cDNA and analysed by quantitative real-time PCR. BAL cells were allowed to sediment on poly-L-lysine-coated glass slides and HLA-DR expression was determined using peroxidase-labelled anti-HLA-DR antibodies. Approximately 300 cell were counted using a light microscope. The bars depict the percentage of HLA-DR⁺ T lymphocytes in the BAL of controls (C) and OS9 positive and negative sarcoidosis patients.



Tables

Table E1: SNPs used in the backward model selection in a logistic regression model.

Additional to marker rs1050045, 21 markers of gene regions that have been previously identified as sarcoidosis susceptibility loci (*IL23R*, *BTNL2*, *HLA*, *Rab23* and *ANXA11*) entered the model. Positions are from NCBI build 36.

dbSNP ID	Chr.	Position	Gene	Reference	
1	rs6664119	67428483	<i>IL23R</i>	*	
6	rs644045	31991936	—		SNPs with genome-wide significance in the this GWA study
6	rs9268402	32449331	<i>BTNL2</i>		
6	rs9391858	32449376	<i>BTNL2</i>		
6	rs2076533	32471505	<i>BTNL2</i>		
6	rs2076530	32471794	<i>BTNL2</i>	[28]	
6	rs3177928	32520413	<i>HLA-DRA</i>		
6	rs7194	32520458	<i>HLA-DRA</i>		
6	rs7195	32520517	<i>HLA-DRA</i>		
6	rs502771	32686948	<i>HLA-DRB1/DQA1 region</i>		
6	rs4530903	32689867	<i>HLA-DRB1/DQA1 region</i>		
6	rs9275371	32776274	<i>HLA-DQ region</i>		
6	rs9275418	32778222	<i>HLA-DQ region</i>		
6	rs2856717	32778286	<i>HLA-DQ region</i>		
6	rs9275522	32782948	<i>HLA-DQ region</i>		
6	rs9275523	32782972	<i>HLA-DQ region</i>		
6	rs9277550	33163465	<i>HLA-DPB1</i>		
6	rs3117242	33177871	<i>HLA-DPB1/DPB2 region</i>		
6	rs3128923	33179300	<i>HLA-DPB1/DPB2 region</i>		
6	rs3957366	57013233	<i>BEND6 (Rab23)</i>	[29]	
10	rs1953600	81901705	<i>ANXA11</i>	[2]	
12	rs1050045	56401538	<i>OS9</i>		

*The Genome-Wide Human SNP Array 6.0 does not contain the SNP rs11209026 (*IL23R*) which has been reported previously to be associated with sarcoidosis [30]. Therefore, the most significantly associated SNP within the *IL23R* region was selected from the GWAS data set.

Table E2: Primer used for the detection of variants in the *OS9* exonic, exon-flanking and regulatory regions.

#	Name	Sequence 5'→3'	Usage
1	OS9_P1_L	GCCCGCATTATTTCTCTAGG	PCR + Seq
	OS9_P1_R	CTCGTTCCCAATTCAGAG	PCR + Seq
2	OS9_P2_L	TAGACCACGCCTCTGAAATTG	PCR + Seq
	OS9_P2_R	AAACTGGACAGCAGCGTTTC	PCR + Seq
3	OS9_P3_L	CTGGAGGTCAACTGCAGGA	PCR + Seq
	OS9_P3_R	CGTTCTATCCCTCCCACGTA	PCR + Seq
4	OS9_E1_L	GACCTTGAGCCACGTTTAC	PCR + Seq
	OS9_E1_R	GCTGCCCTTCTTCTCTATC	PCR + Seq
5	OS9_E2_L	AGGGATTTGTTTGTCTCCC	PCR + Seq
	OS9_E2_R	ACAGGTCCCCTAGATCCTGC	PCR + Seq
6	OS9_E3-4_L	AGCCAGGAGACTCTGGTTCC	PCR + Seq
	OS9_E3-4_R	AGCTAATCTTGACCCTCCCC	PCR + Seq
7	OS9_E5_L	TTGTCTCTTTGGGGACGC	PCR + Seq
	OS9_E5_R	TAAGATGCAACCCTTCTCCC	PCR + Seq
8	OS9_E6-7.2_L	AACCAATATCGCCTCACTGG	PCR + Seq
	OS9_E6-7.2_R	CCTTCCCCTCTCTATCACC	PCR + Seq
9	OS9_E6-7_L	TGCTTAGTAAAAGACCTGCTC	PCR + Seq
	OS9_E6-7_R	CACCCACCAGTCAGTCAG	PCR + Seq
10	OS9_E8-9_L	AAGGGACCCCATCCTATTTG	PCR + Seq
	OS9_E8-9_R	CTTAGCTCTCCCCTCCATCC	PCR + Seq
11	OS9_E11_L	TGCTACTGTTTGTCTGCCCC	PCR + Seq
	OS9_E11_R	TTCTTTTAGCCCGTCAGCC	PCR + Seq
12	OS9_E12_L	CTCCTACAGAGCCAGCCC	PCR + Seq
	OS9_E12_R	AGAAGGGAAGGAACAGGGTC	PCR + Seq
13	OS9_E13-14_L	AGATGGAAGAGCTGCCAAAG	PCR + Seq
	OS9_E13-14_R	AAATGGAAGCTCAGCAATGG	PCR + Seq
14	OS9_E15_long_L	GACATGGGTGCTCTATTCC	PCR
	OS9_E15_long_R	TGGTCTTCTCTGCATACCC	PCR
15	OS9_E15_1_L	CATTCCCTGCCTTCCCC	Seq
	OS9_E15_1_R	AATCTCCAGGTAGGCAGCAG	Seq
16	OS9_E15_2_L	TTGCTCTCCTGAACTCTCACTC	Seq
	OS9_E15_2_R	GGAATGGGGAGCAAAGAAC	Seq
17	OS9_E15_3_L	GACCTCTCGGGCAACTCTG	Seq
	OS9_E15_3_R	AAACATCACTAAGGGCAGGTG	Seq

Table E3: RT-PCR primer used for tissue-specific expression and BAL panel I.

Primer	Sequence (5'-3')
OS9-F	gaggagctgagtgagatgc
OS9-R	cccagagatactgcacc
AGAP2-F	gagagacaacaggagtg
AGAP2-R	ccgagccttcagtcttg
TSPAN31-F	caagactcgggatgaactgg
TSPAN31-R	tatgcactcacagaagcttc
CDK4-F	cctgattgggctgcctcc
CDK4-R	aatgggaaggagaaggagaagc
MARCH9-F	tcgctcagcagtagcctgg
MARCH9-R	tggttcactgcctgccagc
CYP27B1-F	ctttctggccgaactttctgc
CYP27B1-R	gtggcgcagccagtgg
METTL1-F	gcgctactaccggcaacg
METTL1-R	aataagtgtctctgggaacagc
FAM119B-F	tgaccatcacgcagaacttgg
FAM119B-R	gggaaggtgggtccagg

Table E4: Primer pairs used for quantification of *OS9* mRNA expression (BAL panel II)

Primer were synthesized by Biomers.

Gene	Forward primer	Reversed primer
OS9	gctgactgatgaggacacga	cggtaatgctctcagctc
GAPDH	caccagggtgcttttaact	gatctcgtcctggaagatg

Table E5: Association results from panels A and B for those 99 SNPs that were investigated in the validation stage. Positions are in NCBI's build 36. A1 denotes the minor allele. Allele frequencies of A1 (AF) are shown for healthy control individuals (Co), sarcoidosis patients (SA) and the chronic and acute sarcoidosis subphenotypes. Genotyping counts are listed as numbers of homozygotes for the minor allele, heterozygotes and homozygotes for the major allele (A1A1/A1A2/A2A2). The nominal p values of the χ^2 -test for alleles (one degree of freedom) are shown. Significant p values ($p < 0.05$) are highlighted in blue.

chr	Position (bp)	dbSNP ID	SNP	A1	rank	Screening panel A					Validation panel B										
						AF Co	SA			AF	SA			acute		chronic					
							A1A1/A1A2/A2A2 Co	AF	A1A1/A1A2/A2A2		P	A1A1/A1A2/A2A2	AF	A1A1/A1A2/A2A2	P	AF	A1A1/A1A2/A2A2	P	AF	A1A1/A1A2/A2A2	P
1	18,032,082	rs685987	SNP_A-2302259	A	45	0.37	166/583/482	0.31	64/266/306	1.72×10 ⁻⁴	0.35	350/1240/1160	0.35	185/691/647	6.83×10 ⁻¹	0.33	50/244/220	2.68×10 ⁻¹	0.35	112/383/371	8.72×10 ⁻¹
1	71,190,213	rs7548299	SNP_A-2307198	T	70	0.08	10/166/1057	0.11	7/126/504	3.98×10 ⁻⁴	0.09	18/460/2288	0.09	11/261/1297	9.35×10 ⁻¹	0.08	3/82/444	5.05×10 ⁻¹	0.09	7/155/732	5.23×10 ⁻¹
1	73,896,164	rs1497848	SNP_A-8686347	G	65	0.23	59/458/711	0.29	50/268/319	2.96×10 ⁻⁴	0.26	202/1048/1507	0.26	99/610/856	5.98×10 ⁻¹	0.27	30/222/276	8.10×10 ⁻¹	0.26	61/335/495	5.57×10 ⁻¹
1	96,244,563	rs321248	SNP_A-2294898	T	61	0.42	215/602/414	0.48	143/328/166	2.49×10 ⁻⁴	0.43	526/1334/892	0.45	316/784/469	1.10×10 ⁻¹	0.45	96/283/150	3.45×10 ⁻¹	0.45	190/428/276	1.67×10 ⁻¹
1	148,862,161	rs11204682	SNP_A-1808322	T	76	0.20	41/409/783	0.25	33/251/353	4.63×10 ⁻⁴	0.21	123/887/1726	0.21	80/487/952	5.20×10 ⁻¹	0.22	27/176/309	2.09×10 ⁻¹	0.21	50/270/547	5.81×10 ⁻¹
1	209,920,803	rs701924	SNP_A-2108310	A	25	0.29	98/520/614	0.35	70/311/256	7.31×10 ⁻⁵	0.31	299/1106/1340	0.32	165/666/732	4.28×10 ⁻¹	0.33	53/243/231	1.82×10 ⁻¹	0.31	98/364/428	7.33×10 ⁻¹
1	223,595,288	rs12728785	SNP_A-8669527	A	81	0.09	8/200/1025	0.06	3/65/569	5.22×10 ⁻⁴	0.08	17/426/2275	0.08	10/215/1323	1.57×10 ⁻¹	0.06	3/58/462	1.13×10 ⁻²	0.08	4/133/745	5.47×10 ⁻¹
2	39,136,188	rs2060988	SNP_A-8572874	A	91	0.07	3/174/1045	0.11	11/113/508	5.99×10 ⁻⁴	0.08	14/420/2324	0.08	11/238/1318	7.76×10 ⁻¹	0.08	4/81/445	7.55×10 ⁻¹	0.09	6/141/744	5.25×10 ⁻¹
2	45,348,647	rs17033293	SNP_A-8697812	C	28	0.14	15/307/910	0.09	6/106/525	9.12×10 ⁻⁵	0.12	34/599/2079	0.10	21/277/1228	1.10×10 ⁻²	0.11	6/101/411	2.08×10 ⁻¹	0.10	13/151/705	1.74×10 ⁻²
2	50,570,272	rs10183541	SNP_A-2208413	G	90	0.11	11/256/964	0.15	23/148/466	5.98×10 ⁻⁴	0.12	35/612/2107	0.14	38/352/1177	8.86×10 ⁻²	0.14	9/129/392	1.82×10 ⁻¹	0.14	24/193/674	2.08×10 ⁻¹
2	61,778,577	rs6736243	SNP_A-2069217	G	64	0.38	178/567/479	0.44	119/307/195	2.92×10 ⁻⁴	0.40	419/1355/975	0.41	262/753/547	3.68×10 ⁻¹	0.41	91/251/184	4.39×10 ⁻¹	0.40	145/430/315	6.72×10 ⁻¹
2	113,633,383	rs12475781	SNP_A-8553121	G	84	0.38	184/559/488	0.32	71/265/301	5.60×10 ⁻⁴	0.37	394/1242/1092	0.34	166/705/660	2.06×10 ⁻³	0.34	53/245/215	6.81×10 ⁻²	0.34	101/401/374	3.52×10 ⁻²
2	140,932,087	rs16844001	SNP_A-1923632	G	39	0.04	2/91/1134	0.07	5/75/555	1.41×10 ⁻⁴	0.05	7/261/2486	0.05	9/142/1409	7.82×10 ⁻¹	0.05	6/41/482	9.74×10 ⁻¹	0.05	3/83/800	9.50×10 ⁻¹
2	157,060,664	rs297581	SNP_A-2120508	C	55	0.41	200/606/427	0.47	133/334/170	2.25×10 ⁻⁴	0.41	501/1252/978	0.42	280/715/535	7.19×10 ⁻¹	0.42	92/253/172	5.62×10 ⁻¹	0.42	169/400/302	4.30×10 ⁻¹
2	195,374,603	rs2191509	SNP_A-1966341	A	35	0.09	14/197/1022	0.13	11/146/480	1.23×10 ⁻⁴	0.10	31/484/2246	0.11	18/307/1241	1.17×10 ⁻¹	0.11	8/103/418	1.80×10 ⁻¹	0.11	9/170/713	4.30×10 ⁻¹
2	203,264,771	rs6748088	SNP_A-8419961	C	89	0.30	100/548/585	0.36	92/273/272	5.92×10 ⁻⁴	0.32	280/1200/1258	0.35	169/733/629	7.58×10 ⁻³	0.33	49/244/222	4.99×10 ⁻¹	0.36	104/423/346	1.93E-03
2	207,453,139	rs13429985	SNP_A-8493426	A	29	0.33	134/541/555	0.39	91/318/227	9.75×10 ⁻⁵	0.35	335/1236/1183	0.36	194/730/636	2.50×10 ⁻¹	0.36	68/244/214	3.45×10 ⁻¹	0.36	109/419/360	3.33×10 ⁻¹
2	215,584,616	rs10198064	SNP_A-1961084	C	5	0.06	5/146/1081	0.10	6/119/512	1.70×10 ⁻⁵	0.08	17/404/2332	0.09	13/256/1273	5.64×10 ⁻²	0.08	5/77/436	6.36×10 ⁻¹	0.10	8/157/713	1.30×10 ⁻²
2	217,520,622	rs1921998	SNP_A-8505333	C	92	0.47	267/625/339	0.41	106/311/218	6.06×10 ⁻⁴	0.46	542/1391/771	0.43	285/761/487	3.63×10 ⁻²	0.45	99/269/148	7.84×10 ⁻¹	0.43	161/432/285	4.20×10 ⁻²
3	21,884,297	rs2620534	SNP_A-4219591	T	58	0.47	271/616/346	0.41	102/314/221	2.41×10 ⁻⁴	0.44	516/1361/871	0.42	278/769/520	2.55×10 ⁻¹	0.40	83/258/187	4.21×10 ⁻²	0.44	174/441/278	6.33×10 ⁻¹
3	47,535,867	rs7628631	SNP_A-8582631	T	63	0.37	173/573/487	0.43	129/295/213	2.70×10 ⁻⁴	0.38	390/1318/1050	0.40	233/781/546	7.65×10 ⁻²	0.40	72/276/178	2.48×10 ⁻¹	0.40	137/438/313	1.21×10 ⁻¹
3	52,119,987	rs9847073	SNP_A-8530530	G	83	0.05	5/120/1107	0.08	3/98/536	5.52×10 ⁻⁴	0.06	10/299/2447	0.06	3/191/1362	3.07×10 ⁻¹	0.07	1/72/453	1.21×10 ⁻¹	0.06	2/99/783	9.64×10 ⁻¹
3	69,788,306	rs6549245	SNP_A-8618182	C	50	0.37	172/562/499	0.31	53/283/300	1.78×10 ⁻⁴	0.36	343/1290/1114	0.33	173/668/716	1.44×10 ⁻³	0.34	62/228/231	1.80×10 ⁻¹	0.32	95/378/417	1.89×10 ⁻³
4	72,334,532	rs1563045	SNP_A-1977540	T	51	0.12	22/254/955	0.17	21/168/446	1.84×10 ⁻⁴	0.13	61/619/2083	0.14	39/361/1166	4.29×10 ⁻¹	0.13	11/115/402	6.92×10 ⁻¹	0.14	24/206/662	3.85×10 ⁻¹

19	37,958,000	rs6510287	SNP_A-4268916	C	21	0.02	0/38/1192	0.04	2/42/592	5.25×10^{-5}	0.02	0/103/2658	0.02	1/56/1504	9.80×10^{-1}	0.02	0/21/508	7.89×10^{-1}	0.02	1/28/858	6.38×10^{-1}
19	56,846,717	rs10412972	SNP_A-1828486	A	52	0.18	40/354/838	0.13	16/132/488	1.92×10^{-4}	0.15	68/714/1977	0.15	23/407/1126	2.91×10^{-1}	0.13	7/126/391	8.51×10^{-2}	0.15	12/241/634	6.12×10^{-1}
20	15,471,295	rs1233729	SNP_A-1950277	A	79	0.30	107/534/592	0.36	73/312/252	4.96×10^{-4}	0.32	274/1207/1278	0.32	150/705/714	8.32×10^{-1}	0.32	37/268/225	7.73×10^{-1}	0.32	94/377/422	8.88×10^{-1}
20	45,011,328	rs1206754	SNP_A-8577489	T	15	0.02	1/39/1193	0.04	3/43/591	3.63×10^{-5}	0.02	1/102/2663	0.02	3/61/1505	4.12×10^{-1}	0.03	2/27/501	2.73×10^{-2}	0.02	1/34/859	7.15×10^{-1}
20	58,478,034	rs6128829	SNP_A-2224064	A	46	0.39	195/562/476	0.32	63/285/287	1.73×10^{-4}	0.35	358/1223/1176	0.35	195/709/653	9.05×10^{-1}	0.34	61/236/229	4.71×10^{-1}	0.36	118/396/372	7.13×10^{-1}
20	59,267,891	rs6101283	SNP_A-2140517	A	82	0.14	20/299/913	0.10	7/111/519	5.23×10^{-4}	0.11	42/529/2155	0.12	25/314/1204	4.42×10^{-1}	0.12	7/108/406	6.66×10^{-1}	0.11	13/176/691	7.89×10^{-1}
22	27,657,907	rs5762900	SNP_A-1805590	A	93	0.06	3/131/1096	0.03	1/37/599	6.07×10^{-4}	0.06	6/316/2422	0.05	2/156/1387	1.25×10^{-1}	0.05	1/51/469	2.65×10^{-1}	0.05	1/81/799	4.69×10^{-2}

Table E6: Association results for those 53 SNPs genotyped in fine-mapping panel D

Positions are in NCBI's build 36. The minor allele is denoted by A1. Allele frequencies for A1 (AF) are given in healthy controls (Co) and sarcoidosis patients (SA). The numbers of homozygotes for the minor allele, heterozygotes and homozygotes for the major allele (A1A1/A1A2/A2A2) are given. Nominal p values of a χ^2 -test for alleles (one degree of freedom) are shown. Significant p values ($p < 0.05$) are highlighted in brown and blue, respectively. Odds ratios (OR) and 95% confidence intervals (95% CI) are listed for of the rare A1 allele.

dbSNP ID	Position (bp)	A1	Co		SA				acute				chronic			
			AF	A1A1/A1A2/A2A2	AF	A1A1/A1A2/A2A2	P	OR [CI 95%]	AF	A1A1/A1A2/A2A2	P	OR [CI 95%]	AF	A1A1/A1A2/A2A2	PA	OR [CI 95%]
rs12831733	56,122,056	C	0.03	2/82/1339	0.03	2/100/1638	9.39×10 ⁻¹	0.99 [0.74-1.32]	0.02	1/24/541	2.14×10 ⁻¹	0.76 [0.48-1.18]	0.03	1/64/948	6.38×10 ⁻¹	1.08 [0.78-1.50]
rs2229357	56,129,978	A	0.25	96/511/816	0.25	113/625/988	9.64×10 ⁻¹	1.00 [0.89-1.12]	0.27	49/212/303	6.79×10 ⁻²	1.16 [0.99-1.35]	0.23	58/351/594	2.60×10 ⁻¹	0.93 [0.81-1.06]
rs507562	56,136,035	C	0.24	75/534/815	0.23	77/665/1007	5.74×10 ⁻¹	0.97 [0.86-1.09]	0.22	21/210/340	1.93×10 ⁻¹	0.90 [0.76-1.06]	0.24	49/380/587	7.00×10 ⁻¹	0.97 [0.85-1.11]
rs2242578	56,139,420	G	0.33	140/646/632	0.32	166/768/807	3.69×10 ⁻¹	0.95 [0.86-1.06]	0.31	51/247/272	2.10×10 ⁻¹	0.91 [0.78-1.06]	0.32	99/445/467	5.26×10 ⁻¹	0.96 [0.85-1.09]
rs1678537	56,186,608	A	0.11	14/281/1111	0.12	23/357/1355	4.37×10 ⁻¹	1.06 [0.91-1.25]	0.12	8/117/438	4.55×10 ⁻¹	1.09 [0.87-1.35]	0.12	15/205/793	5.02×10 ⁻¹	1.06 [0.89-1.27]
rs775241	56,222,594	G	0.12	17/320/1086	0.13	28/387/1323	7.15×10 ⁻¹	1.03 [0.89-1.19]	0.13	12/125/433	5.82×10 ⁻¹	1.06 [0.86-1.30]	0.13	16/222/770	8.60×10 ⁻¹	1.02 [0.85-1.21]
rs730560	56,224,832	C	0.44	274/689/455	0.44	343/860/539	5.47×10 ⁻¹	1.03 [0.93-1.14]	0.42	104/273/190	4.88×10 ⁻¹	0.95 [0.83-1.09]	0.46	211/506/297	1.39×10 ⁻¹	1.09 [0.97-1.22]
rs11172247	56,232,777	G	0.37	174/702/549	0.34	202/781/758	1.99×10 ⁻²	0.88 [0.80-0.98]	0.34	75/240/254	1.31×10 ⁻¹	0.89 [0.77-1.03]	0.34	112/461/439	3.27×10 ⁻²	0.88 [0.78-0.99]
rs1678542	56,254,982	C	0.37	170/696/552	0.34	192/778/753	2.01×10 ⁻²	0.88 [0.80-0.98]	0.34	68/243/252	9.17×10 ⁻²	0.88 [0.76-1.02]	0.34	108/458/435	4.19×10 ⁻²	0.88 [0.78-1.00]
rs2888334	56,262,914	C	0.29	103/624/689	0.27	123/688/916	4.66×10 ⁻²	0.89 [0.80-1.00]	0.27	41/223/305	1.17×10 ⁻¹	0.88 [0.76-1.03]	0.27	71/396/531	7.64×10 ⁻²	0.89 [0.78-1.01]
rs775250	56,263,307	A	0.27	99/559/761	0.29	141/712/862	4.28×10 ⁻²	1.12 [1.00-1.25]	0.31	58/230/273	8.70×10 ⁻³	1.23 [1.05-1.43]	0.28	74/409/514	3.39×10 ⁻¹	1.07 [0.94-1.21]
rs1678536	56,265,457	G	0.30	122/606/696	0.26	121/672/953	1.18×10 ⁻³	0.83 [0.75-0.93]	0.26	38/223/308	2.54×10 ⁻²	0.84 [0.72-0.98]	0.26	71/390/556	5.01×10 ⁻³	0.83 [0.73-0.95]
rs1545783	56,268,521	T	0.25	82/560/779	0.28	124/738/876	1.01×10 ⁻²	1.16 [1.04-1.30]	0.31	47/254/268	1.08×10 ⁻³	1.29 [1.11-1.50]	0.27	66/415/527	2.00×10 ⁻¹	1.09 [0.96-1.24]
rs1284185	56,275,310	A	0.03	1/86/1338	0.03	0/114/1639	7.11×10 ⁻¹	1.06 [0.80-1.40]	0.03	0/36/536	9.20×10 ⁻¹	1.02 [0.69-1.51]	0.04	0/72/947	3.86×10 ⁻¹	1.15 [0.84-1.58]
rs812315	56,279,757	C	0.22	63/494/865	0.24	98/630/1016	7.62×10 ⁻²	1.11 [0.99-1.25]	0.26	38/221/310	3.77×10 ⁻³	1.27 [1.08-1.48]	0.22	51/346/617	8.22×10 ⁻¹	1.02 [0.89-1.17]
rs2306390	56,288,866	T	0.28	97/589/737	0.24	94/653/1002	1.63×10 ⁻³	0.83 [0.74-0.93]	0.23	32/201/339	4.95×10 ⁻³	0.80 [0.68-0.93]	0.24	53/386/576	1.07×10 ⁻²	0.84 [0.74-0.96]
rs2277323	56,295,639	A	0.25	81/542/798	0.24	93/633/1002	3.22×10 ⁻¹	0.94 [0.84-1.06]	0.25	42/198/320	7.81×10 ⁻¹	1.02 [0.87-1.20]	0.23	42/375/592	1.06×10 ⁻¹	0.90 [0.78-1.02]
rs1552842	56,298,430	A	0.41	248/648/502	0.44	326/845/530	1.44×10 ⁻²	1.14 [1.03-1.26]	0.45	117/266/172	1.96×10 ⁻²	1.18 [1.03-1.36]	0.44	181/508/303	4.59×10 ⁻²	1.13 [1.00-1.27]
rs2277324	56,299,442	A	0.38	210/665/547	0.42	307/869/569	4.63×10 ⁻⁴	1.20 [1.08-1.33]	0.43	113/265/193	5.07×10 ⁻³	1.22 [1.06-1.40]	0.43	167/528/318	2.24×10 ⁻³	1.20 [1.07-1.35]
rs923828	56,301,761	T	0.40	232/676/509	0.44	323/865/529	2.60×10 ⁻³	1.17 [1.06-1.29]	0.45	115/269/174	1.07×10 ⁻²	1.20 [1.04-1.38]	0.44	178/517/303	1.60×10 ⁻²	1.15 [1.03-1.30]
rs1871417	56,305,246	A	0.33	129/676/618	0.31	161/751/827	9.46×10 ⁻²	0.91 [0.82-1.02]	0.28	50/220/296	5.55×10 ⁻³	0.81 [0.69-0.94]	0.32	92/461/458	5.11×10 ⁻¹	0.96 [0.85-1.08]
rs813516	56,307,092	T	0.02	1/59/1362	0.02	0/56/1695	1.08×10 ⁻¹	0.74 [0.51-1.07]	0.02	0/21/550	5.40×10 ⁻¹	0.86 [0.52-1.41]	0.02	0/33/986	1.89×10 ⁻¹	0.75 [0.49-1.15]
rs12320537	56,307,358	C	0.23	67/510/834	0.21	77/588/1078	1.43×10 ⁻¹	0.91 [0.81-1.03]	0.23	32/192/345	8.34×10 ⁻¹	0.98 [0.83-1.16]	0.20	36/342/635	4.87×10 ⁻²	0.87 [0.76-1.00]

rs715930	56,310,248	T	0.23	67/512/846	0.21	76/589/1084	1.55×10 ⁻¹	0.92 [0.81-1.03]	0.23	32/193/345	9.42×10 ⁻¹	0.99 [0.84-1.17]	0.20	35/342/641	4.33×10 ⁻²	0.87 [0.75-1.00]
rs1689595	56,343,030	G	0.36	164/672/571	0.33	190/782/766	8.02×10 ⁻²	0.91 [0.82-1.01]	0.31	62/223/280	4.02×10 ⁻³	0.80 [0.69-0.93]	0.35	109/483/420	5.29×10 ⁻¹	0.96 [0.85-1.09]
rs10876994	56,351,004	C	0.27	86/595/742	0.24	94/663/996	1.49×10 ⁻²	0.87 [0.78-0.97]	0.23	33/200/339	1.64×10 ⁻²	0.82 [0.70-0.96]	0.24	51/397/571	5.43×10 ⁻²	0.88 [0.77-1.00]
rs12831104	56,357,782	G	0.13	30/316/1077	0.12	19/374/1355	8.67×10 ⁻²	0.88 [0.76-1.02]	0.11	9/111/450	1.06×10 ⁻¹	0.84 [0.68-1.04]	0.12	9/223/785	1.61×10 ⁻¹	0.88 [0.74-1.05]
rs11172300	56,362,782	T	0.40	234/665/526	0.44	334/858/554	1.54×10 ⁻³	1.18 [1.06-1.30]	0.46	133/259/180	4.09×10 ⁻⁴	1.28 [1.12-1.47]	0.43	174/519/319	3.35×10 ⁻²	1.13 [1.01-1.27]
rs1689585	56,366,407	G	0.36	161/676/565	0.33	176/773/750	4.02×10 ⁻²	0.90 [0.81-1.00]	0.30	54/220/278	4.98×10 ⁻⁴	0.77 [0.66-0.89]	0.34	102/476/409	4.25×10 ⁻¹	0.95 [0.84-1.07]
rs1628552	56,368,409	C	0.36	165/683/575	0.33	191/786/775	5.92×10 ⁻²	0.90 [0.82-1.00]	0.30	61/223/288	1.10×10 ⁻³	0.78 [0.67-0.91]	0.35	110/486/422	5.19×10 ⁻¹	0.96 [0.85-1.08]
rs4760168	56,374,004	T	0.35	160/679/581	0.33	188/775/776	8.23×10 ⁻²	0.91 [0.82-1.01]	0.30	60/223/286	2.52×10 ⁻³	0.80 [0.69-0.92]	0.34	107/477/425	5.12×10 ⁻¹	0.96 [0.85-1.08]
rs7979246	56,378,821	A	0.34	156/663/598	0.33	182/746/778	1.18×10 ⁻¹	0.92 [0.83-1.02]	0.29	57/205/285	1.80×10 ⁻³	0.79 [0.68-0.91]	0.34	105/468/426	7.48×10 ⁻¹	0.98 [0.87-1.11]
rs10783844	56,384,692	A	0.36	169/680/576	0.34	197/779/772	7.10×10 ⁻²	0.91 [0.82-1.01]	0.30	63/221/286	1.57×10 ⁻³	0.79 [0.68-0.91]	0.35	113/482/422	5.23×10 ⁻¹	0.96 [0.85-1.08]
rs1050045	56,401,538	C	0.40	231/662/510	0.45	346/863/522	1.10×10 ⁻⁴	1.22 [1.10-1.35]	0.47	132/262/171	2.08×10 ⁻⁴	1.30 [1.13-1.50]	0.44	186/522/300	3.20×10 ⁻³	1.19 [1.06-1.34]
rs701008	56,403,912	G	0.39	217/678/528	0.35	205/808/727	8.36×10 ⁻⁴	0.84 [0.76-0.93]	0.36	69/266/234	3.74×10 ⁻²	0.86 [0.74-0.99]	0.35	114/473/424	1.85×10 ⁻³	0.83 [0.74-0.93]
rs12307841	56,411,705	C	0.13	29/311/1075	0.12	19/371/1358	1.07×10 ⁻¹	0.88 [0.76-1.03]	0.11	8/113/449	1.40×10 ⁻¹	0.85 [0.69-1.05]	0.12	10/217/789	1.55×10 ⁻¹	0.88 [0.74-1.05]
rs2270777	56,431,423	A	0.41	236/686/502	0.45	354/864/531	6.15×10 ⁻⁴	1.19 [1.08-1.32]	0.46	129/267/175	2.31×10 ⁻³	1.24 [1.08-1.42]	0.44	194/513/309	1.12×10 ⁻²	1.16 [1.03-1.30]
rs1048691	56,439,215	T	0.23	72/496/852	0.22	86/592/1073	4.93×10 ⁻¹	0.96 [0.85-1.08]	0.19	24/173/373	2.99×10 ⁻²	0.83 [0.70-0.98]	0.23	56/351/612	8.70×10 ⁻¹	1.01 [0.88-1.16]
rs4646537	56,443,548	C	0.04	1/114/1310	0.04	6/113/1634	2.95×10 ⁻¹	0.87 [0.67-1.13]	0.03	3/33/536	3.30×10 ⁻¹	0.83 [0.58-1.21]	0.04	3/69/947	4.91×10 ⁻¹	0.90 [0.67-1.21]
rs8176345	56,444,825	A	0.04	1/98/1324	0.03	1/113/1634	6.24×10 ⁻¹	0.93 [0.71-1.23]	0.03	0/39/532	8.81×10 ⁻¹	0.97 [0.67-1.42]	0.03	0/62/953	3.80×10 ⁻¹	0.87 [0.63-1.20]
rs703842	56,449,006	C	0.34	177/627/621	0.31	153/769/823	2.20×10 ⁻³	0.85 [0.76-0.94]	0.31	51/254/266	5.11×10 ⁻²	0.86 [0.75-1.00]	0.30	87/441/484	3.27×10 ⁻³	0.83 [0.74-0.94]
rs11172335	56,461,468	T	0.34	177/622/625	0.31	151/770/831	1.83×10 ⁻³	0.85 [0.76-0.94]	0.31	51/253/268	5.14×10 ⁻²	0.86 [0.75-1.00]	0.30	85/443/490	2.34×10 ⁻³	0.83 [0.73-0.93]
rs17565804	56,462,687	T	0.05	2/131/1292	0.04	7/136/1605	3.93×10 ⁻¹	0.90 [0.71-1.14]	0.04	4/42/524	6.37×10 ⁻¹	0.92 [0.66-1.29]	0.04	3/78/936	3.15×10 ⁻¹	0.87 [0.66-1.15]
rs10877020	56,474,281	A	0.42	251/696/470	0.46	369/870/511	3.46×10 ⁻³	1.16 [1.05-1.28]	0.47	129/283/160	4.18×10 ⁻³	1.22 [1.07-1.40]	0.46	212/502/302	2.38×10 ⁻²	1.14 [1.02-1.28]
rs2277326	56,488,686	T	0.37	196/670/557	0.41	295/846/596	1.15×10 ⁻³	1.18 [1.07-1.31]	0.43	105/277/189	1.97×10 ⁻³	1.25 [1.08-1.43]	0.41	164/491/352	1.98×10 ⁻²	1.15 [1.02-1.29]
rs12582311	56,489,531	G	0.23	76/512/832	0.23	93/616/1033	7.35×10 ⁻¹	0.98 [0.87-1.10]	0.22	27/193/350	2.49×10 ⁻¹	0.91 [0.77-1.07]	0.24	61/356/594	8.23×10 ⁻¹	1.02 [0.89-1.16]
rs10783853	56,520,529	A	0.30	126/586/701	0.27	122/688/925	1.42×10 ⁻²	0.87 [0.78-0.97]	0.28	39/231/291	1.92×10 ⁻¹	0.90 [0.77-1.05]	0.27	72/397/545	2.43×10 ⁻²	0.86 [0.76-0.98]
rs4760340	56,561,443	T	0.50	360/699/364	0.47	386/869/497	1.64×10 ⁻²	0.89 [0.80-0.98]	0.46	122/284/166	3.59×10 ⁻²	0.86 [0.75-0.99]	0.48	230/508/280	1.16×10 ⁻¹	0.91 [0.81-1.02]
rs17119981	56,573,469	A	0.05	1/145/1278	0.05	7/145/1571	3.15×10 ⁻¹	0.89 [0.71-1.12]	0.05	3/47/511	5.73×10 ⁻¹	0.91 [0.66-1.26]	0.04	3/84/917	2.82×10 ⁻¹	0.86 [0.66-1.13]
rs12309291	56,579,931	G	0.05	3/150/1270	0.06	9/199/1535	2.11×10 ⁻¹	1.15 [0.93-1.42]	0.05	2/54/510	6.55×10 ⁻¹	0.93 [0.68-1.27]	0.07	6/130/879	2.92×10 ⁻²	1.30 [1.03-1.64]
rs7977734	56,587,329	C	0.05	5/142/1278	0.06	8/181/1551	5.70×10 ⁻¹	1.07 [0.86-1.32]	0.05	2/55/512	8.53×10 ⁻¹	0.97 [0.71-1.32]	0.06	6/113/892	2.05×10 ⁻¹	1.17 [0.92-1.49]
rs12816216	56,633,587	A	0.04	3/113/1300	0.05	7/167/1569	6.57×10 ⁻²	1.25 [0.99-1.58]	0.05	0/53/516	5.21×10 ⁻¹	1.11 [0.80-1.55]	0.05	6/98/909	4.60×10 ⁻²	1.31 [1.00-1.71]
rs4362183	56,649,511	C	0.30	122/619/681	0.28	150/695/905	9.54×10 ⁻²	0.91 [0.82-1.02]	0.28	51/215/303	1.23×10 ⁻¹	0.89 [0.76-1.03]	0.28	82/413/524	1.29×10 ⁻¹	0.91 [0.80-1.03]

Table E7: Haplotype association analysis around the lead SNP rs1050045. Haplotypes of the lead SNP rs1050045 and six neighboring SNPs in the *OS9* gene region in the fine-mapping panel D (1,708 sarcoidosis cases, 1,429 controls). Alleles are given for the SNPs in the following order: rs11172300, rs1689585, rs1628552, rs4760168, rs7979246, rs10783844 and rs1050045. **F Co**: haplotype frequency in controls. **P**: P value of a score test assuming an additive effect of the respective haplotype. **SA**: controls vs. all sarcoidosis patients. **Acute SA**: controls vs. patients with the acute subphenotype (n = 570). **Chronic SA**: controls vs. chronic subphenotype (n = 1016). **Global P**: P-value for a difference in the joint frequency distribution of all haplotypes between cases and controls. **Ns**: not significant.

Haplotype	F Co	P SA	P acute SA	P chronic SA
C A C T A A C	0.001	ns	-	ns
C A T G A A T	0.001	ns	-	ns
C A T G G A C	0.001	2.54x10 ⁻⁰²	-	ns
C A T G G C C	0.006	2.59x10 ⁻⁰²	ns	2.30x10 ⁻⁰²
C A T G G C T	0.229	3.86x10 ⁻⁰²	ns	1.90x10 ⁻⁰²
C A C G A A T	0.004	ns	ns	ns
C A C T A A C	0.001	ns	-	ns
C A C T A A T	0.325	ns	4.96 x10 ⁻⁰³	ns
C A C T G A T	0.010	ns	ns	ns
C A C T G C T	0.001	ns	-	ns
T A T G G C C	0.416	2.44x10 ⁻⁰³	5.80x10 ⁻⁴	1.50x10 ⁻⁰²
T A T G G C T	0.002	ns	ns	ns
global P		9.13x10 ⁻⁰³	4.68x10 ⁻⁰³	ns

Table E8: Prediction of SNP function

The potential functional effect of the seven strongly associated SNPs in the *OS9* gene region (highlighted in yellow) and 26 additional SNPs that are in strong LD ($r^2 > 0.9$) with those were investigated using the NIEHS SNPinfo web server [31]. Positions refer to human genome build 18, and p values from the fine-mapping stage are given for patients with sarcoidosis (P_{SA}) and the acute (P_{acute}) and chronic subphenotype ($P_{chronic}$) compared to controls. Only functional categories with a reliable predictive value are shown in the table:

Effect on transcription by location of the variant in a transcription factor binding site (TFBS), effect on splicing due to location of the SNP in an exonic splicing enhancer or exonic splicing silencer (Splicing), effect miRNA binding according to the miRanda database (miRNA), the regulatory potential according to ESPERR Regulatory Potential (RegPot) and conservation according to the Vertebrate Multiz Alignment and Conservation Score in 17 species (Cons). The two latter parameters (RegPot and Cons) are given only for SNP outside the coding region. For references of the individual prediction tools and a detailed description please see <http://snpinfo.niehs.nih.gov/guide.htm#snpfunc>.

dbSNP ID	Position	P _{SA}	P _{acute}	P _{chronic}	TFBS	Splicing	miRNA	RegPot	Cons	Nearby Gene	Allele
rs7314152	56,303,968	-	-	-	--	--	--	0.13	0.09	SLC26A10	T
rs774887	56,321,102	-	-	-	--	--	--			B4GALNT1	A
rs1082502	56,321,973	-	-	-	--	--	--			B4GALNT1	A
rs1629032	56,324,953	-	-	-	--	--	--			B4GALNT1	T
rs1678514	56,332,984	-	-	-	--	--	--			B4GALNT1	C
rs1678516	56,333,725	-	-	-	--	--	--			B4GALNT1	G
rs1678520	56,335,586	-	-	-	--	--	--			B4GALNT1	C
rs1689595	56,343,030	8.02 x10 ⁻⁰²	4.02 x10 ⁻⁰³	5.29 x10 ⁻⁰¹	--	--	--			B4GALNT1	C
rs1689592	56,344,040	-	-	-	--	--	--			B4GALNT1	C
rs10876993	56,348,934	-	-	-	--	--	--			B4GALNT1	C
rs1678540	56,351,269	-	-	-	Y	--	--			B4GALNT1	A
rs2640629	56,351,715	-	-	-	Y	--	--			B4GALNT1	C
rs2640631	56,353,799	-	-	-	Y	--	--			B4GALNT1	A
rs774890	56,354,655	-	-	-	Y	--	--			B4GALNT1	T
rs1689587	56,361,665	-	-	-	--	--	--			OS9	A
rs11172300	56,362,782	1.54 x10 ⁻⁰³	4.09 x10 ⁻⁰⁴	3.35 x10 ⁻⁰²	--	--	--		0.07	OS9	C
rs2640637	56,363,853	-	-	-	--	--	--			OS9	A
rs1689585	56,366,407	4.02 x10 ⁻⁰²	4.98 x10 ⁻⁰⁴	4.25 x10 ⁻⁰¹	--	--	--		0.04	OS9	T
rs1628552	56,368,409	5.92 x10 ⁻⁰²	1.10 x10 ⁻⁰³	5.19 x10 ⁻⁰¹	--	--	--			OS9	C
rs10877001	56,369,344	-	-	-	Y	--	--			OS9	T
rs1678504	56,371,765	-	-	-	Y	--	--			OS9	G
rs7311632	56,372,607	-	-	-	Y	--	--		0.01	OS9	A
rs4760168	56,374,004	8.23 x10 ⁻⁰²	2.52 x10 ⁻⁰³	5.12 x10 ⁻⁰¹	Y	--	--	0.35		OS9	T
rs3825078	56,375,680	-	-	-	--	--	--			OS9	A
rs7979246	56,378,821	1.18 x10 ⁻⁰¹	1.80 x10 ⁻⁰³	7.48 x10 ⁻⁰¹	--	--	--	0.01		OS9	A
rs7952989	56,382,256	-	-	-	--	--	--			OS9	T
rs10783844	56,384,692	7.10 x10 ⁻⁰²	1.57 x10 ⁻⁰³	5.23 x10 ⁻⁰¹	--	--	--			OS9	A
rs7295422	56,386,003	-	-	-	--	--	--			OS9	G
rs7976852	56,386,818	-	-	-	--	--	--			OS9	G
rs10877008	56,389,859	-	-	-	--	--	--			OS9	A
rs10082911	56,390,371	-	-	-	--	--	--		0.01	OS9	C
rs799265	56,398,456	-	-	-	--	Y	--	0.59	1.00	OS9	A
rs1050045	56,401,538	1.10 x10 ⁻⁰⁴	2.08 x10 ⁻⁰⁴	3.20 x10 ⁻⁰³	Y	Y	Y	0.07	0.01	OS9	T

3. References

1. Fischer A, Valentonyte R, Nebel A, Nothnagel M, Müller-Quernheim J, Schürmann M, Schreiber S. Female-specific association of C-C chemokine receptor 5 gene polymorphisms with Lofgren's syndrome. *J Mol Med* 2008; 86(5): 553-561.
2. Hofmann S, Franke A, Fischer A, Jacobs G, Nothnagel M, Gaede KI, Schürmann M, Müller-Quernheim J, Krawczak M, Rosenstiel P, Schreiber S. Genome-wide association study identifies ANXA11 as a new susceptibility gene for sarcoidosis. *Nat Genet* 2008; 40: 1103 - 1106.
3. Prasse A, Katic C, Germann M, Buchwald A, Zissel G, Muller-Quernheim J. Phenotyping sarcoidosis from a pulmonary perspective. *Am J Respir Crit Care Med* 2008; 177(3): 330-336.
4. Krawczak M, Nikolaus S, von Eberstein H, Croucher PJP, El Mokhtari NE, Schreiber S. PopGen: Population-Based Recruitment of Patients and Controls for the Analysis of Complex Genotype-Phenotype Relationships. *Community Genet* 2006; 9(1): 55-61.
5. Valentonyte R, Hampe J, Huse K, Rosenstiel P, Albrecht M, Stenzel A, Nagy M, Gaede KI, Franke A, Haesler R, Koch A, Lengauer T, Seegert D, Reiling N, Ehlers S, Schwinger E, Platzer M, Krawczak M, Müller-Quernheim J, Schürmann M, Schreiber S. Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nat Genet* 2005; 37(4): 357-364.
6. 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* 1999; 160(2): 736-755.
7. Klareskog L, Stolt P, Lundberg K, Kallberg H, Bengtsson C, Grunewald J, Ronnelid J, Harris HE, Ulfgren AK, Rantapaa-Dahlqvist S, Eklund A, Padyukov L, Alfredsson L. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum* 2006; 54(1): 38-46.
8. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007; 449(7164): 851-861.
9. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* 2010; <http://www.R-project.org>.
10. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MR, Bender D, Maller J, Sklar P, de Bakker PW, Daly M, Sham P. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* 2007; 81(3): 559-575.
11. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet* 2005; 76(5): 887-893.
12. Franke A, Balschun T, Sina C, Ellinghaus D, Hasler R, Mayr G, Albrecht M, Wittig M, Buchert E, Nikolaus S, Gieger C, Wichmann HE, Sventoraityte J, Kupcinskis L, Onnie CM, Gazouli M, Anagnou NP, Strachan D, McArdle WL, Mathew CG, Rutgeerts P, Vermeire S, Vatn MH, Krawczak M, Rosenstiel P, Karlsen TH, Schreiber S. Genome-wide association study for ulcerative colitis identifies risk loci at 7q22 and 22q13 (IL17REL). *Nat Genet* 2010; 42(4): 292-294.
13. Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, Zhao JH, Ramasamy A, Zhai G, Vitart V, Huffman JE, Igl W, Albrecht E, Deloukas P, Henderson J, Granell R, McArdle WL, Rudnicka AR, Barroso I, Loos RJ, Wareham NJ, Mustelin L, Rantanen T, Surakka I, Imboden M, Wichmann HE, Grkovic I, Jankovic S, Zgaga L, Hartikainen AL, Peltonen L, Gyllenstein U, Johansson A, Zaboli G, Campbell H, Wild SH, Wilson JF, Glaser S, Homuth G, Volzke H, Mangino M, Soranzo N, Spector TD, Polasek O, Rudan I, Wright AF, Heliovaara M, Ripatti S, Pouta A, Naluai AT, Olin AC, Toren K, Cooper MN, James AL,

- Palmer LJ, Hingorani AD, Wannamethee SG, Whincup PH, Smith GD, Ebrahim S, McKeever TM, Pavord ID, MacLeod AK, Morris AD, Porteous DJ, Cooper C, Dennison E, Shaheen S, Karrasch S, Schnabel E, Schulz H, Grallert H, Bouatia-Naji N, Delplanque J, Froguel P, Blakey JD, Britton JR, Morris RW, Holloway JW, Lawlor DA, Hui J, Nyberg F, Jarvelin MR, Jackson C, Kahonen M, Kaprio J, Probst-Hensch NM, Koch B, Hayward C, Evans DM, Elliott P, Strachan DP, Hall IP, Tobin MD. Genome-wide association study identifies five loci associated with lung function. *Nat Genet* 2010: 42(1): 36-44.
14. Neale BM, Fagerness J, Reynolds R, Sobrin L, Parker M, Raychaudhuri S, Tan PL, Oh EC, Merriam JE, Souied E, Bernstein PS, Li B, Frederick JM, Zhang K, Brantley MA, Jr., Lee AY, Zack DJ, Campochiaro B, Campochiaro P, Ripke S, Smith RT, Barile GR, Katsanis N, Allikmets R, Daly MJ, Seddon JM. Genome-wide association study of advanced age-related macular degeneration identifies a role of the hepatic lipase gene (LIPC). *Proc Natl Acad Sci U S A* 2010: 107(16):7395-400.
15. Ellinghaus E, Stanulla M, Richter G, Ellinghaus D, Te Kronnie G, Cario G, Cazzaniga G, Horstmann M, Panzer Grumayer R, Cave H, Trka J, Cinek O, Teigler-Schlegel A, Elsharawy A, Hasler R, Nebel A, Meissner B, Bartram T, Lescai F, Franceschi C, Giordan M, Nurnberg P, Heinzow B, Zimmermann M, Schreiber S, Schrappe M, Franke A. Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia. *Leukemia* 2012: 26(5): 902-909.
16. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, Albrecht M, Mayr G, De La Vega FM, Briggs J, Gunther S, Prescott NJ, Onnie CM, Hasler R, Sipos B, Folsch UR, Lengauer T, Platzer M, Mathew CG, Krawczak M, Schreiber S. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* 2007: 39(2): 207-211.
17. De la Vega FM, Lazaruk KD, Rhodes MD, Wenz MH. Assessment of two flexible and compatible SNP genotyping platforms: TaqMan SNP Genotyping Assays and the SNPlex Genotyping System. *Mutat Res* 2005: 573(1-2): 111-135.
18. Teuber M, Koch WA, Manaster C, Wächter S, Hampe J, Schreiber S. Improving quality control and workflow management in high-throughput single-nucleotide polymorphism genotyping environments. *JALA* 2005: 10(1): 43-47.
19. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Csaaki BNPaF, ed. International Symposium on Information Theory 2nd ed. Akademiai Kiado, Budapest, 1973; pp. 267-281.
20. Abecasis GR, Cookson WOC. GOLD--Graphical Overview of Linkage Disequilibrium. *Bioinformatics* 2000: 16(2): 182-183.
21. Schaid D, Rowland C, Tines D, Jacobson R, Poland G. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002: 70(2): 425-434.
22. Dupont WD, Plummer WD. PS power and sample size program available for free on the Internet. *Controlled Clin Trials* 1997: 18(274).
23. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005: 21(2): 263-265.
24. Hunninghake GW, Crystal RG. Pulmonary sarcoidosis: a disorder mediated by excess helper T-lymphocyte activity at sites of disease activity. *N Engl J Med* 1981: 305(8): 429-434.
25. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods and Protocols: Methods in Molecular Biology* 2000: 132:365-386.
26. Hofmann S, Fischer A, Till A, Muller-Quernheim J, Hasler R, Franke A, Gade KI, Schaarschmidt H, Rosenstiel P, Nebel A, Schurmann M, Nothnagel M, Schreiber S. A genome-wide association study reveals evidence of association with sarcoidosis at 6p12.1. *Eur Respir J* 2011: 38(5):1127-35.

27. Franke A, Fischer A, Nothnagel M, Becker C, Grabe N, Till A, Lu T, Muller-Quernheim J, Wittig M, Hermann A, Balschun T, Hofmann S, Niemiec R, Schulz S, Hampe J, Nikolaus S, Nurnberg P, Krawczak M, Schurmann M, Rosenstiel P, Nebel A, Schreiber S. Genome-wide association analysis in sarcoidosis and Crohn's disease unravels a common susceptibility locus on 10p12.2. *Gastroenterology* 2008; 135(4): 1207-1215.
28. Valentonyte R, Hampe J, Huse K, Rosenstiel P, Albrecht M, Stenzel A, Nagy M, Gaede KI, Franke A, Haesler R, Koch A, Lengauer T, Seegert D, Reiling N, Ehlers S, Schwinger E, Platzer M, Krawczak M, Muller-Quernheim J, Schurmann M, Schreiber S. Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nat Genet* 2005; 37(4): 357-364.
29. Hofmann S, Fischer A, Till A, Muller-Quernheim J, Hasler R, Franke A, Gade KI, Schaarschmidt H, Rosenstiel P, Nebel A, Schurmann M, Nothnagel M, Schreiber S. A genome-wide association study reveals evidence of association with sarcoidosis at 6p12.1. *Eur Respir J* 2011; 38(5): 1127-1135.
30. Fischer A, Nothnagel M, Franke A, Jacobs G, Saadati HR, Gaede KI, Rosenstiel P, Schurmann M, Muller-Quernheim J, Schreiber S, Hofmann S. Association of inflammatory bowel disease risk loci with sarcoidosis, and its acute and chronic subphenotypes. *Eur Respir J* 2011; 37(3): 610-616.
31. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res* 2009; 37(Web Server issue): W600-605.