ERJ Express. Published on August 30, 2012 as doi: 10.1183/09031936.00033812

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

Sylvia Hofmann[1*], Annegret Fischer[1*], Michael Nothnagel[2], Gunnar Jacobs[1,3], Benjamin Schmid[1], Michael Wittig[1], Andre Franke[1], Karoline I. Gaede[4], Manfred Schürmann[5], Martin Petrek[6], Frantisek Mrazek[6], Stefan Pabst[7], Christian Grohé[8], Johan Grunewald[9], Marcus Ronninger[10], Anders Eklund[9], Philip Rosenstiel[1], Kerstin Höhne[11], Gernot Zissel[11], Joachim Müller-Quernheim[11], Stefan Schreiber[1,3,12]

1 Institute of Clinical Molecular Biology, Christian-Albrechts University, Kiel, Germany
2 Institute of Medical Informatics and Statistics, Christian-Albrechts University; Kiel, Germany
3 Popgen Biobank, University Hospital Schleswig-Holstein; Kiel, Germany
4 Department of Pneumology, Research Center Borstel, Borstel, Germany
5 Institute of Human Genetics, University of Lübeck, Lübeck, Germany
6 Laboratory of Immunogenomics and Immunoproteomics, Faculty of Medicine and Dentistry, Palacky University, Olomouc, Czech Republic
7 Medical Clinic II, Department of Pneumology, University of Bonn, Germany
8 Department of Respiratory Medicine, Evangelische Lungenklinik Berlin-Buch, Germany
9 Respiratory Medicine Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden
10 Department of Medicine, Rheumatology Unit, Karolinska Institute, Stockholm, Sweden
11 Department of Pneumology, University of Freiburg, Freiburg, Germany
12 Department of General Internal Medicine, University Hospital Schleswig-Holstein, Kiel, Germany

*Both authors contributed equally to this work.

Corresponding author: Dr. Sylvia Hofmann, Institute of Clinical Molecular Biology, Christian-Albrechts University, Arnold-Heller Str.3, Haus 5, 24105 Kiel, Germany; Tel.: +49-431-597-1306, Fax.: +49-431-597-2196; E-mail: s.hofmann@ikmb.uni-kiel.de

Abstract

Sarcoidosis is a systemic inflammatory disease of unknown etiology, influenced by genetic and environmental factors. However, the loci so far identified for sarcoidosis explain only a part of its assumed heritability.

To identify further susceptibility loci, we performed a genome-wide association analysis using the Affymetrix 6.0 Human GeneChip followed by validation and replication stages.

After quality control, 637 cases, 1,233 controls and 677,619 SNPs were available for an initial screening. 99 SNPs were selected for validation in an independent study panel (1,664 patients, 2,932 controls).

SNP rs1050045 was significantly associated with sarcoidosis (corrected $p = 0.0215$) in the validation panel and yielded a p value of 9.22×10^{-8} (OR = 1.24) in the meta-analysis of the screening and validation stage. A meta-analysis of three populations from Germany, Czech Republic and Sweden confirmed this finding ($p = 0.024$; OR = 1.14). Fine-mapping and mRNA expression studies pointed to *osteosarcoma amplified 9 (OS9*) as the most likely candidate for the underlying risk factor.

The OS9 protein plays an important role in the ER-associated protein degradation and acts during Toll-like receptor induced activation of myeloid cells. Expression analyses of *OS9* mRNA provide evidence for a functional mechanism underlying the detected association signal.

Introduction

Sarcoidosis (MIM 181000) is a systemic inflammatory disease of unknown etiology that is characterized by non-caseating epitheloid cell granulomas. Although any organ system can be affected, granulomas are most frequently found in the lung and the lymph nodes. The pathogenesis is characterized by high activity of macrophages and CD4+ helper T cells after exposure to a yet elusive antigen under the regulatory influence of cytokines produced by local mononuclear phagocytes, T cells, dendritic cells and fibroblasts [1, 2]. According to the course of the disease, patients can be classified as being affected by acute or chronic sarcoidosis [3]. In brief, acute sarcoidosis is characterized by sudden complaints and recovery within two years. It includes Löfgren´s syndrome, which is characterized by erythema nodosum, bilateral hilar lymphadenopathy and polyarthritis. On the contrary, chronic sarcoidosis patients exhibit subtly intensifying early symptoms, followed by enduring disease activity for two years or longer.

Sarcoidosis is a rare disease with a prevalence rate of about 40/100,000 inhabitants in Germany, and mainly affects young adults (20-40 years) and preferably women [4]. It is thought to be triggered by a complex combination of environmental and genetic factors with an estimated heritability of 66% [5]. The genetic underpinning of the disease is supported by the identification of a number of risk genes like *BTNL2* [6-8], *ANXA11* [9, 10], *TNF-α* [11] and several *HLA*-loci (for review see [12]). Several candidates await further support like *Rab23* [13] and the chemokine receptors *CCR2* and *CCR5* and *IL23R* [14]. With the present study, we therefore aimed at the identification of further susceptibility loci for sarcoidosis using the Affymetrix SNP array 6.0, which comprises nearly one million SNPs and thus yields a significantly higher and also partially different coverage than the previously published genome-wide association studies (GWAS) [9, 13]. Since we expected rather small effect sizes and therefore no results of genome-wide significance in the initial genome-wide

screening, we included an independent validation step and a replication step following a multi-stage design (supplement Figure E1).

Materials & Methods

Patient and control subjects

Sarcoidosis patients were classified as having chronic or acute sarcoidosis as previously described [9, 15, 16], according to the course and the presentation of the disease based on all available information (questionnaires completed by patients and physicians, hospital records and interview information). Briefly, subtly intensifying early symptoms followed by enduring disease activity for 2 years or longer defined the chronic sarcoidosis sample (further on referred to as "chronic"). Patients of the acute sarcoidosis sample ("acute") suffered from sudden complaints and recovered within 2 years. Only patients who could be unequivocally categorized to acute or chronic were recruited to those subphenotypes. Thus, for the categorization into acute and chronic a disease course of at least 2 years was analyzed. All patients showed evidence of disease involvement in the thorax.

Before quality control, the screening panel A comprised 640 patients, including 191 acute and 401 chronic patients, and 1,256 control subjects. Panel A almost completely overlapped with the panel (also termed panel A) that was used in a previous sarcoidosis association screen using the Affymetrix 5.0 GeneChip [9]. Moreover, parts of the panel A had already been used in the former association analyses [6, 17]. For validation 1,664 sarcoidosis patients, including 563 individuals with acute and 947 individuals with chronic sarcoidosis and 2,932 healthy individuals were available before quality control, together forming panel B. Replication panel C-I comprised 303 German sarcoidosis patients and 281 controls and had no overlap with any other panel. Information on the subphenotype status was available only for a limited number

of these patients (acute: n = 40; chronic: n = 61). Replication panel C-II consisted of 267 sarcoidosis cases and 330 controls from the Czech Republic and substantially overlapped with a sample described elsewhere [10]. No subphenotype information was available for this sample. The Swedish samples (panel C-III) comprised 1066 cases recruited at the outpatient clinic at the Pulmonary Division at the Karolinska University Hospital, Solna, Sweden, of which 333 patients were diagnosed with Löfgren syndrome, an acute form of sarcoidosis. For the remaining cases Löfgren syndrome was either excluded or no subphenotype information was available. The 940 Swedish controls were contributed by the Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA) study [18]. Fine-mapping (panel D) was carried out in 1,829 German sarcoidosis patients, comprising all patients from panel B and parts of panel A, including 597 acutely and 1,055 chronically affected patients, and in 1,465 German controls from panel B. Among the patients with the acute course of sarcoidosis a total of 123 individuals showed the classical symptoms of Löfgren´s syndrome. Recruitment and diagnosis of patients of panels A, B, C-I and D were accomplished as described above. All study participants of panels A, B, C-I and D were of German origin. For details on diagnosis and recruitment, see the online data supplement methods section. The 45 sarcoidosis patients and 45 control individuals (panel E) used for sequencing of the *OS9* gene region were selected from fine-mapping panel D in order to enrich carriers of the rs1050045 risk allele, since these individuals have a higher chance to carry the causative variant(s), which are assumed to be in LD with rs1050045.

Genotyping and quality control

Genotyping of panels A and B was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 (Santa Clara, CA, USA) and SNPlex™ technology (Applied Biosystems, Foster City, CA, USA), respectively. Additional genotyping of panels C-I, C-II, C-III, D and

E was performed using Taqman® technology (Applied Biosystems, Foster City, CA, USA). Conservative and established quality filters were following common practice [19]; for details of quality control and genotyping of each panel see supplementary material. Briefly, all individuals had to have <10% missing genotypes. Samples that showed evidence for cryptic relatedness (identical by state value > 0.8; see also Figure E2 in the supplementary material) to other samples were removed from the data set. For each panel (screening, validation and replication), SNPs were checked for missing genotypes (threshold for exclusion: < 95% in either patients or control subjects), minor allele frequency (< 2% in patients or control subjects), and deviation from Hardy-Weinberg equilibrium (HWE) in the control sample ($p \leq 0.01$), which led to the exclusion of 257,349 markers (27.5%) from the GWAS dataset. Any SNPs selected for validation underwent visual inspection of its cluster plot (see Figure E3 for the cluster plot of lead SNP rs1050045).

SNP selection and statistical analysis

Those markers that ranked top with their p value in the genome-wide association analysis of panel A and for which at least one additional correlated SNP ($r^2 > 0.5$ with $p < 10^{-3}$ were selected for validation. Markers from the *HLA* region (*6p21.1-6p21.3*) and from the *ANXA11* gene region (chr10, position 81,850-82,000 kb) were not included in the validation stage because a strong disease association of those loci had already been established based on the same study population [6, 9]. Statistical analysis of genotype data was carried out using PLINK v.1.06 [20] unless stated otherwise. In the entire experiment, single-marker allele-based association analysis was performed using a χ^2 test ($df = 1$). Visualization of linkage disequlibrium (LD) was carried out with GOLD [21]. Logistic regression, backward model selection using AIC [22] and haplotype analysis using the haplo.stats package [23] were conducted in R v2.10.1 and v2.15.0 [24]. HapMap tagging SNPs were selected for fine-

mapping using Haploview [25]. The population attributable risk (PAR) for the GWAS lead SNP was calculated using the following formula: PAR = [f* (rr-1)]/[f * (rr-1) + 1], where f denotes the allele frequency in the risk population and rr equals the allelic relative risk, as estimated by the corresponding odds ratio [26]. For a detailed description see the online data supplement methods section. Correction for population stratification in panel A and B was conducted using an estimated genomic inflation factor of 1.149. Meta-analysis of panel A and B was conducted using the *inverse normal* method [27], while meta-analysis of replication panels C-I, -II and –III was carried out using the *fixed-effect* model implemented in PLINK v.1.07.

Interaction analysis was performed using the *epistasis* option in PLINK v.1.07. A significant result in this test indicates a deviation of the combined effects of the associated SNPs from the multiplicative model.

Additionally, the association analysis of markers was adjusted for the effects of previously reported markers using logistic regression models (see supplementary material Table E1). All markers were considered under a genotypic risk model. Statistical backward model selection was performed using the step function in R and was based on the default AIC criterion. Only samples that had no missing genotypes at any of the considered markers were included in this analysis (637 cases, 1,233 controls). Significance was assessed by a likelihood-ratio test.

Sanger sequencing

The exonic, exon-flanking and regulatory regions of *OS9* were sequenced using standard Sanger sequencing technology on ABI PRISM 3700 DNA Analyzer (Applied Biosystems). Primers were designed using Primer3 [28]. Primer sequences are given in the online data supplement Table E2.

Analysis of tissue-specific expression by PCR

For investigation of tissue-specific expression patterns of the candidate genes we used a commercially available tissue and immune cell panel from Clontech (Palo Alto, CA, USA) and a semi-quantitative PCR. Expressions were normalized on *GAPDH* expression. For the respective primer sequences see online data supplement Table E3.

BAL preparation, mRNA isolation and qRT-PCR

BAL cell samples of BAL panel I were matched by their portion of alveolar macrophages (see online supplement material for details). Total RNA was isolated from snap-frozen BAL cells using a commercial kit (RNeasy, Qiagen, Hilden, Germany) and cDNA was synthesized using the Advantage RT-for-PCR kit (Clontech Laboratories, Palo Alto, CA, USA) according to the manufacturer's protocol. Sequences of target-specific primers for qRT-PCR are given in the online data supplement Table E3. Transcript amounts were normalized to *GAPDH* mRNA levels. Relative expression levels of the target genes were tested for significant differences between sarcoidosis patients and unaffected individuals (n = 4 each) using a non-parametric Mann-Whitney U test as implemented in the Graphpad statistical software (Graphpad, Inc.; La Jolla, CA).

A second BAL series of BAL-samples was obtained from 46 patients with active sarcoidosis and 8 controls (BAL panel II). BAL cell smears were dried and stained using May-Grünwald-Gimsa staining. Cell differentials were determined by counting at least 200 cells. For analysis of HLA-DR expression on lymphocytes cells were fixed on poly-L-lysine coated slides, incubated with monoclonal antibodies directed against HLA-DR at concentrations suggested by the supplier and developed with a peroxidase-antiperoxidase technique. The sequences of the primers used for qRT-PCR are given in Table E4. The primers do not distinguish between

the three known isoforms of *OS9* mRNA. Cycle numbers of *GAPDH* and *OS9* were equal (35 cycles). A threshold cycle value (Ct) was calculated and used to calculate the relative expression (rE) level of mRNA for each sample by using the following formula: $rE = 2^{(CtGapDH-Ct-Target)}$ x10,000. The relative expression is given as a dimension-free ratio. Statistical analysis of this cohort was performed using StatView (SAS Institute, Cary, NC) using a Mann-Whitney U test. DNA was extracted from blood, and genotyping was performed as described above. For a description of the BAL samples and further details see the online data supplement methods section.

Immunohistochemistry

Sections of lung tissue fixed with Hepes-glutamic acid buffer-mediated organic solvent protection effect (HOPE) from anonymized normal controls (n = 4) and active sarcoidosis (n = 8) were stained with the rabbit polyclonal OS9 antibody (Novus biotechnicals, NB100-519B) using standard protocols at a primary antibody dilution of 1:100 [29]. Omitting the primary antibody and using irrelevant primary antibodies served as negative and positive control, respectively. Photomicrographs were taken on a Zeiss Axio Imager Z1 (Zeiss, Oberkochen, Germany).

Results

GWAS analysis

After applying conservative and established quality filters to the data set, 1,870 samples (panel A: 637 cases, 1,233 controls) and 677,619 SNPs were included in the initial genome-wide screening. The assessed population heterogeneity was moderate in panel A, with a genomic inflation factor of $\lambda_{GC} = 1.15$ based on a median χ^2-distribution [30], where $\lambda_{GC} = 1.0$

corresponds to no inflation. The QQ-plot and association signals for known sarcoidosis risk loci are given in the online data supplementary results and Fig. E4.

In addition to the SNPs carried forward to the validation stage (see below), multiple SNPs in the *BTNL2* gene (rs2076533, rs2076530, rs9268480, rs3806156) and in several *HLA* loci (rs7194, rs7195, rs3177928, HLA-DRA; rs9277550, rs1431403, rs2856816 HLA-DPB1; rs2071475, rs2071473, HLA-DOB) on chromosome 6p21.3 (supplementary Fig. E5 online), a region that is characterized by patterns of high LD, were found in the GWAS to be significantly associated with sarcoidosis. The respective SNPs were strongly associated with nominal P values between 1.01×10^{-15} (rs2076533) and 9.17×10^{-5} (rs2856816).

Validation of lead variants

Ninety-nine SNPs that passed the pre-defined selection criteria were genotyped in an independent validation sample (panel B). After quality control, 2,770 German controls and 1,572 German sarcoidosis patients, including 894 chronic and 530 acute patients, comprising 99 individuals with Löfgren's syndrome, were included in the analysis. Twenty-one markers showed a nominally significant association with sarcoidosis in the validation stage (Table 1). One variant, rs1050045, located at 12q13.3-q14.1, was associated with sarcoidosis with an uncorrected p value of 7.38×10^{-5}. Since screening panel A and validation panel B originate from the same German population, we assumed an inflation of the test statistic due to population stratification of $\lambda = 1.15$ for the combined panel, as estimated for panel A. The result for rs1050045 remained significant after correction for this effect and after Bonferroni correction for multiple testing [corrected p value (p_{corr}). $= 99 \times 2.18\times10^{-4} = 0.0215$; OR = 1.20 with a 95% confidence interval (95% CI) of 1.10-1.31]. The SNP conferred a population attributable risk (PAR) of 8 % to 12 % (based on the frequencies obtained in the validation and screening panel, respectively). In a meta-analysis of panels A and B it was associated

with sarcoidosis with a p_{corr} of 9.22×10^{-8} (OR = 1.24) [31]. No significant SNP-SNP interactions of this SNP with known susceptibility variants for sarcoidosis in the *ANXA11*, *BTNL2, Rab23* and the *IL23R* loci were observed in panel A (data not shown).

In order to determine whether rs1050045 exerts a statistically independent influence on sarcoidosis from previously reported susceptibility variants in *BTNL2*, *ANXA11*, *Rab23*, *IL23R* and *HLA*, we performed a backward model selection in a logistic regression model based on AIC. Additional to marker rs1050045, another 21 markers entered the model before the selection, namely rs6664119 (*IL23R*), rs644045 (chr6 p21.33), rs9268402, rs9391858, rs2076533, rs2076530 (all *BTNL2*), rs3177928, rs7194, rs7195 (all *HLA-DRA*), rs502771 (*HLA-DRB1/5*), rs4530903 (*HLA-DRB1/DQA1* region), rs9275371, rs9275418, rs2856717, rs9275522, rs9275523 (all *HLA-DQ* region), rs9277550, rs3117242, rs3128923 (all *HLA-DPB*), rs3957366 (*BEND6*) and rs1953600 (*ANXA11*). All markers were considered under a genotypic risk model The final model contained, additional to rs1050045, markers rs6664119, rs644045, rs2076533, rs3177928, rs502771, rs9275371, rs9275418, rs2856717, rs3128923, rs3957366, rs1953600 and. After adjustment for the other 11 markers, rs1050045 was still significantly associated with sarcoidosis ($p=6.2\times10^{-4}$).

Subphenotype-specific analysis revealed a stronger association of rs1050045 with acute sarcoidosis after correction for population stratification [corrected $p_{acute} = 6.75\times10^{-4}$; OR = 1.30; 95% CI (1.14-1.49)] compared to the chronic subphenotype [corrected $p_{chronic} = 0.021$; OR = 1.16; 95% CI (1.04-1.30)]. Patients with Löfgren´s syndrome showed an even stronger effect [corrected $p_{Löfgren} = 0.044$; OR = 1.41; 95% CI (1.05-1.88)]. Genotypes were verified using TaqMan® SNP genotyping as an independent technology (98.74% genotype concordance). Risk allele C had a frequency of 42% in controls, 46% in cases, 48% in patients with acute sarcoidosis and even 50% in cases with Löfgren´s syndrom.

Detailed results including genotyping counts for all SNPs under study in the validation stage are presented in the online data supplement Table E5.

Replication in independent samples from different European populations

In order to replicate the detected association of lead SNP rs1050045 with sarcoidosis, we performed a meta-analysis of this SNP in independent case-control samples from Germany, the Czech Republic and Sweden (panels C-I, C-II and C-III, respectively). The variant showed significant association with sarcoidosis in a meta-analysis of these three sample sets ($p = 0.023$; OR = 1.14). Risk allele frequencies (and corresponding odds ratios) varied between the populations, ranging from 41.9% in controls vs. 43.1% in cases [OR = 1.05 (0.83-1.32)] in panel C-I through 37.7% vs. 40.1% [OR = 1.11; 95% CI (0.97-1.26)] in panel C-III up to 42.7% vs. 48.6% [OR = 1.27; 95% CI (1.00-1.61)] in panel C-II. A subphenotype analysis of panel C-I was not promising due to its small sample size (power < 15%), and no subphenotype information at all was available for panel C-II. In panel C-III though, we found a significant association with Löfgren syndrome, an acute subform of sarcoidosis [$p = 0.015$; OR = 1.26; 95% CI (1.05-1.52)].

Fine-mapping around rs1050045 (chromosome 12q13.3-q14.1)

In addition to the lead SNP rs1050045, we selected 57 tagging SNPs from HapMap CEU for fine-mapping of ~500 kb of the 12q13.3-q14.1 region around rs1050045 in the fine-mapping panel D. After quality control, genotypes of 53 SNPs from 1,753 cases, including 570 acute and 1,016 chronically affected patients, and 1,429 control individuals were available for the analysis. Twenty-two markers yielded a nominal p value < 0.05 in the association analysis. Again, the strongest association signal was observed with lead SNP rs1050045

[$p = 1.10\times10^{-4}$; OR = 1.22; 95% CI (1.10-1.35)]. Complete analysis results are shown in the online data supplement Table E6. Figure 1 gives an overview of the association signals, the genes, recombination rates and linkage disequilibrium (LD) structure at the 12q13.3-q14.1 locus, showing that the strength of the association signal decreases gradually with increasing genetic and physical distance from the lead SNP rs1050045, which is located in the 3′-UTR of the *osteosarcoma amplified 9 (OS9*) gene. In panel D, this marker was in strong linkage disequilibrium with rs11172300 ($r^2 = 0.96$ in controls), which is located 12 kb upstream of *OS9* and as well significantly associated with sarcoidosis [$p = 1.5\times10^{-3}$; OR = 1.18; 95% CI (1.06-1.30)].

Five SNPs, namely rs1689585, rs1628552, rs4760168, rs7979246 and rs10783844, located between rs1050045 and rs11172300 in the *OS9* gene region, were in strong LD with each other ($r^2 = 0.91$-0.99), but neither with the lead SNP rs1050045 nor with rs11172300. One of these five SNPs, namely rs1689585, showed nominally significant association with the general sarcoidosis phenotype [$p = 4.0\times10^{-2}$; OR = 0.90; 95% CI (0.81-1.00)] and remarkably strong association in the acute subsample [$p_{acute} = 5.0\times10^{-4}$; OR = 0.77; 95% CI (0.66-0.89)]. he remaining four SNPs were significantly associated with acute sarcoidosis only (see online data supplement Figure E6 and Table E6). Haplotype analysis of these seven SNPs (rs1050045, rs11172300, rs1689585, rs1628552, rs4760168, rs7979246 and rs10783844) revealed a significant difference in the haplotype frequency distribution between sarcoidosis patients and controls ($p = 9.1\times10^{-3}$) and in patients with the acute subphenotype compared to controls ($p = 4.7\times10^{-3}$). See online data supplement Table E7 for complete results. However, backward model selection for a logistic regression model using Akaike's Information Criterion (AIC) gave inconclusive results whether rs1050045 represents the only source for phenotypic association in the region (data not shown). Prediction of the functional consequences using the NIEHS SNPinfo web server [32] revealed possible influence of these

seven associated SNPs, and SNPs that are in strong LD ($r^2 > 0.9$) with them, on transcription factor binding sites as well as miRNA binding sites. See online data supplement Table E8 for complete results.

Sequencing of OS9 coding regions

In order to verify existing variants and to identify novel mutations at the associated locus, we sequenced the *OS9* regulatory and exonic regions of 45 sarcoidosis patients and 45 control individuals (panel E) using Sanger sequencing technology. We detected seven known and two novel SNPs, including a non-synonymous SNP in exon 2, named *OS9*-SNP1 (Table 2 and online data supplement Figure E7). Sequencing of exon 10 failed due to technical reasons. Two of the nine detected SNPs (OS9-SNP1 and rs74368191) were confirmed by genotyping and investigated in the fine-mapping panel D, both using Taqman technology. OS9-SNP1 turned out to be an extremely rare variant, with one control individual being the only heterozygous carrier. rs74368191, which was detected only in cases in the sequencing panel E, was slightly more frequent in the cases (MAF = 0.014) than in controls (MAF = 0.011) in panel D, but showed a weaker effect than the lead SNP rs1050045 (OR = 1.22; $p > 0.05$).

Expression analysis of candidate genes in chromosome 12q13.3-q14.1 region

Based on the results of the fine-mapping experiment, we selected eight genes located near to the lead SNP as candidates for the putative susceptibility gene driving the association signal (see Figure 1). We hypothesized that the causative variant(s) might influence susceptibility to sarcoidosis by changing the expression levels of one or several of these genes, namely *OS9*, *ArfGAP with GTPase domain, ankyrin repeat and PH domain 2* (*AGAP2*), *tetraspanin 31*

(*TSPAN31*), *cyclin-dependent kinase 4* (*CDK4*), *membrane-associated ring finger (C3HC4) 9* (*MARCH9*), *cytochrome P450, family 27, subfamily B, polypeptide 1* (*CYP27B1*), *methyltransferase like 1* (*METTL1*) and *family with sequence similarity 119, member B* (*FAM119B*). To prove this hypothesis, we first assessed the expression of the eight transcripts in healthy tissue and immune cell types which are relevant to the pathogenesis of sarcoidosis by a semi-quantitative PCR. All transcripts showed moderate to high expression in healthy lung tissue. Several transcripts, namely *OS9*, *AGAP2, TSPAN31* and *MARCH9*, were down-regulated in activated mononuclear cells, CD4+ and CD8+ T cells compared to their resting counterparts (online supplement Figure E8).

We further hypothesized that the causative variant(s) might confer the increased risk to sarcoidosis by changing the expression levels of one or several of these genes. We therefore analyzed the expression levels of the candidate genes in bronchoalveaolar lavage (BAL) cells from sarcoidosis patients and unaffected persons (BAL panel I, $n = 4$ per group) by performing quantitative real time (qRT)-PCR on cDNA. Expression of all eight candidate genes was detectable in BAL cells, while expression of *OS9*, *TSPAN31* and *FAM119B* was significantly increased in sarcoidosis BAL samples compared to controls (each with nominal $p = 0.029$, Figure 2).

Allele-specific expression of OS9 and immunohistochemistry

From the results of the fine-mapping experiment and the expression studies, *OS9* seemed to be the most promising candidate gene in the associated region. We therefore investigated *OS9* expression in the independent BAL panel II. Here, expression of *OS9* mRNA was detected in 8/8 healthy controls but only in 38/47 sarcoidosis patients using qRT-PCR. A more detailed analysis showed that in patients expressing *OS9*, termed *OS9*-positive patients from now on,

OS9 expression was significantly higher compared to controls (3774±2794 (n = 38) versus 1587±857 (n = 8), online supplement Figure E9, $p = 7.0 \times 10^{-3}$). Since the BAL of sarcoidosis patients is characterized by a low proportion of alveolar macrophages [33], it is interesting that BAL cell composition differed between *OS9*-positive patients and patients not expressing OS9 (termed OS9-negative patients) and controls with regards to the percentage of alveolar macrophages (online supplement Figure E10). The two groups also differed significantly with respect to the percentage of HLA-DR^+ T cells with the highest percentage in *OS9*-neagtiv cases (online supplement Figure E11). Stratification of the BAL samples from sarcoidosis patients according to the genotype at the lead SNP rs1050045 revealed a significant negative correlation of *OS9* mRNA expression level with the CC genotype (AA: 2879, AC: 3078, CC: 824, CC vs. AA p = 0.019; CC vs. AC p = 0.016; Figure 3).

Next, we verified the *in situ* localization of OS9 protein expression by immunohistochemistry in lung biopsies of a third cohort (n = 4 normal controls; n = 8 active sarcoidosis patients). Ubiquitous OS9 protein expression could be observed with a marked perinuclear and granular cytoplasmic staining pattern concurrent with the reported endoplasmatic reticulum (ER) localization of OS9. Strong immunoreactivity could be observed in alveolar macrophages and lymphocytic cells in both diseased samples and normal controls. The granuloma structures in the sarcoidosis biopsies also stained positive for OS9 (Figure 4).

Discussion

We performed a genome-wide association analysis of 677,619 SNPs in 637 German sarcoidosis cases and 1,233 controls and identified a new sarcoidosis susceptibility locus at chromosome 12q13.3-q14.1, which was validated in an independent German case-control

population. The association of lead SNP rs1050045 was replicated by a meta-analysis of three independent cohorts from Germany, the Czech Republic and Sweden (panel C-I, C-II and C-III). Subphenotype-specific analysis revealed a stronger association of this SNP with the acute subform of sarcoidosis than with the overall sarcoidosis phenotype.

The locus was overlooked in the previous genome-wide studies due to a lack of coverage of the region and different analysis strategies although using almost the same primary study population [9, 13].

The region 12q13.3-q14.1 has been reported before to be associated with rheumatoid arthritis [34-36], with SNP rs1678542 showing a similar effect as for sarcoidosis (OR = 0.88-0.94), as well as with type I diabetes [37-39], multiple sclerosis [40] and celiac disease [41]. It therefore represents the first genetic risk locus shared by these clinically distinct diseases. However, since the associated region harbors a number of potential risk genes and no fine-mapping was conducted for any of the mentioned diseases except sarcoidosis, no conclusion can be drawn on whether the different diseases share a single risk factor or are affected by different factors in this region.

For sarcoidosis, fine-mapping of the region and expression studies suggest *OS9* as the most likely candidate for the underlying risk gene. Sequencing of the exonic and exon-flanking regions of *OS9* revealed two novel SNPs, but no obvious candidate for the causative variant(s). Besides rs1050045, which yielded the strongest association of all investigated SNPs in the fine-mapping, only one additional SNP in the 3´-UTR of OS9 was found. Since 16 SNPs are annotated in this region according to dbSNP build 132, 14 of which were not detected, this result dramatically reduces the number of potentially regulatory 3´-UTR SNPs and emphasizes the potential functional relevance of rs1050045.

In silico functional analysis of the most strongly associated SNPs in the OS9 region suggested an influence on gene expression via altering transcription factor as well as miRNA binding and on splicing events. SNP rs799265 which is in perfect LD with rs1050045 ($r^2 = 1$ according to HapMap rel 27) showed high regulatory potential as well as maximal conservation over 17 vertebrate species making this SNP an interesting candidate for further functional analysis. Analysis of *OS9* mRNA from BAL panel I, which contained only samples with high proportion of macrophages, found elevated *OS9* expression levels of sarcoidosis patients compared to controls. This finding was confirmed in a subgroup of patients from BAL panel II, which was characterized by a high percentage of macrophages. In addition, we found that absence of *OS9* mRNA co-occurs with a reduced ratio of $CD4^+/CD8^+$ T cells. Although not significant, this finding may indicate that OS9 is involved in the active state of sarcoidosis, which is characterized by a high $CD4^+/CD8^+$ ratio [33], while a lack of *OS9* mRNA may be associated with the chronic and less active form of sarcoidosis. Preliminary investigation of a potential allele-specific expression of *OS9* in BAL (panel II) showed a significant reduction of *OS9* expression in homozygote carriers of the rs1050045 risk allele. Possible mechanistic explanations for this differential expression include a shorter half-life of the *OS9* mRNA or a different binding behavior of the microRNA (miRNA)-145. This regulatory miRNA binds to the 3´-UTR of *OS9*. It is expressed in human trachea and lung tissue according to the UCSC genome browser [42] and has been shown to be involved in airway inflammation in mouse [43]. Results from a recent publication showed a 2.8-fold higher expression of this miRNA in full blood of sarcoidosis patients compared to healthy controls ($p = 3.9 \times 10^{-3}$) [44]. Further functional experiments are now warranted to elucidate the functional consequences of reasonable candidates for the causative variants, e.g. rs1050045 or rs799265.

The putative risk gene *OS9* consists of 15 exons and encodes four different splice variants [45, 46], all of which contain a mannose 6-phosphate receptor homology domain [47]. The OS9 protein plays an important role in the ER-associated degradation of misfolded or unassembled proteins [48, 49]. This may be of importance, since only peptides derived from proteins, which are common in the human body, were found in the groove of HLA-DR of alveolar macrophages from sarcoidosis patients in a recent study [50]. Moreover, it was demonstrated that these molecules induce an antigenic T cell response [51]. OS9 may therefore act as an autoimmune component in the immunopathogenesis of sarcoidosis. In addition, it has recently been shown that OS9 interacts with the cytoplasmatic tail of the dendritic cell-specific transmembrane (DC-STAMP) protein during toll-like receptor-induced maturation of DCs suggesting a role for OS9 in myeloid differentiation and cell fusion [52]. DCs have been discussed as important mediators of sarcoidosis immunology [53] and have been widely investigated with regards to various aspects of sarcoidosis [53-56].

Based on our results, we cannot, however, exclude that functional change in genes at 12q13.3-q14.1 other than *OS9* influence susceptibility to sarcoidosis. From a clinical point of view, *CYP27B1* represents a further plausible candidate for sarcoidosis susceptibility as discussed similary for multiple sclerosis [40]. This gene encodes a member of the cytochrome P450 superfamily. The enzyme activates vitamin D3, which has a well-established immunoregulatory function in general [57, 58] and also plays a role in lung immunity and in sarcoidosis [59, 60]. Moreover, this gene has very recently been reported to be differentially expressed in sarcoidosis patients with a progressive-fibrotic compared to a self-limiting course of the disease [61]. Since the entire 12q13.3-q14.1 locus showed a subphenotype-specific genetic association pattern in our study, it is conceivable that specific genetic variants in the associated region may influence the regulation of CYP27B1 expression restricted to

certain stages or subtypes of the disease. To date no obvious functional implication of the remaining gene products, namely, AGAP2, TSPAN31, CDK4, MARCH9, METTL1 and FAM119B, with sarcoidosis pathogenesis can be drawn from the literature.

In summary, this is the first report of an association of chromosomal region 12q13.3-q14.1 with sarcoidosis. Fine-mapping of the region and preliminary expression studies suggest *OS9* as the most likely candidate for the underlying susceptibility gene and may support the notion of an autoimmune reaction in the immunopathogenesis of sarcoidosis. Based on the data presented here, more detailed studies on the reported genetic association and OS9 function in the context of sarcoidosis are now needed, in order to define the causative variant(s) and the molecular mechanisms underlying our observations. Moreover, similar studies need to be done in patients with a different ethnic background to determine whether ancestry is linked with the new sarcoidosis susceptibility locus.

Sources of Support

This work was supported by grants of the Federal Ministry for Education and Research in Germany (BMBF) through the National Genome Research Network (NGFN), by the Cluster of Excellence “Inflammation at Interfaces” and GenPhenReSa MU692/8-1, both German Research Foundation (DFG), by the Network for Diffus Parenchymal Lung Disease (GOLD.net), in part by Palacky University IGA PU LF 2010_08, by the Swedish Heart-Lung Foundation, the Swedish Medical Research Council, and through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and the Karolinska Institutet.

Acknowledgements

The authors wish to thank all patients, families and physicians for their cooperation. The support of the German Sarcoidosis Patients Organization (Deutsche Sarkoidose-Vereinigung e.V.), the PopGen biobank and of the contributing pulmonologists is gratefully acknowledged. We gratefully acknowledge technical assistance from the staff of the Institute of Clinical Molecular Biology (Kiel, Germany). Dr. Leonid Padyukov contributed with selection of controls for the Swedish cohort and the authors are grateful to members of the Swedish EIRA study for collection of control samples.

Conflict of Interest

There is no conflict of interest for any of the contributing authors.

References

1. Zissel G, Prasse A, Muller-Quernheim J. Sarcoidosis--immunopathogenetic concepts. *Semin Respir Crit Care Med* 2007: 28(1): 3-14.

2. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007: 357(21): 2153-2165.

3. 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(2): 736-755.

4. Muller-Quernheim J. Sarcoidosis: immunopathogenetic concepts and their clinical application. *Eur Respir J* 1998: 12(3): 716-738.

5. Sverrild A, Backer V, Kyvik KO, Kaprio J, Milman N, Svendsen CB, Thomsen SF. Heredity in sarcoidosis: a registry-based twin study. *Thorax* 2008: 63(10): 894-896.

6. 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.

7. Li Y, Wollnik B, Pabst S, Lennarz M, Rohmann E, Gillissen A, Vetter H, Grohe C. BTNL2 gene variant and sarcoidosis. *Thorax* 2006: 61(3): 273-274.

8. Rybicki BA, Walewski JL, Maliarik MJ, Kian H, Iannuzzi MC, Group AR. The BTNL2 Gene and Sarcoidosis Susceptibility in African Americans and Whites. *Am J Hum Genet* 2005: 77(3): 491-499.

9. Hofmann S, Franke A, Fischer A, Jacobs G, Nothnagel M, Gaede KI, Schurmann M, Muller-Quernheim J, Krawczak M, Rosenstiel P, Schreiber S. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet* 2008: 40(9): 1103-1106.

10. Mrazek F, Stahelova A, Kriegova E, Fillerova R, Zurkova M, Kolek V, Petrek M. Functional variant ANXA11 R230C: true marker of protection and candidate disease modifier in sarcoidosis. *Genes Immun* 2011: 12(6): 490-494.

11. Grunewald J, Idali F, Kockum I, Seddighzadeh M, Nisell M, Eklund A, Padyukov L. Major histocompatibility complex class II transactivator gene polymorphism: associations with Lofgren's syndrome. *Tissue Antigens*: 76(2): 96-101.

12. Muller-Quernheim J, Schurmann M, Hofmann S, Gaede KI, Fischer A, Prasse A, Zissel G, Schreiber S. Genetics of sarcoidosis. *Clin Chest Med* 2008: 29(3): 391-414, viii.

13. 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.

14. 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*: 37(3): 610-616.

15. Fischer A, Valentonyte R, Nebel A, Nothnagel M, Muller-Quernheim J, Schurmann 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.

16. 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.

17. 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.

18. 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.

19. 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, Kupcinskas 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.

20. 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.

21. Abecasis GR, Cookson WOC. GOLD--Graphical Overview of Linkage Disequilibrium. *Bioinformatics* 2000: 16(2): 182-183.

22. Akaike H. Information theory and an extension of the maximum likelihood principle. *In:* Csaaki BNPaF, ed. International Symposium on Information Theory 2nd ed. Acadeemiai Kiadi, Budapest, 1973; pp. 267–281.

23. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002: 70(2): 425-434.

24. Team RDC. A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* 2010: http://www.R-project.org.

25. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005: 21(2): 263-265.

26. Pearce N. Analytical implications of epidemiological concepts of interaction. *Int J Epidemiol* 1989: 18(4): 976-980.

27. Stouffer S, Suchman E, DeVinney L, Star S, Williams J. Studies in Social Psychology in World War II: The American Soldier. Vol. 1, Adjustment During Army Life. Princeton, NJ: Princeton University Press, 1949.

28. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 2000: 132: 365-386.

29. 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.

30. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007: 447(7145): 661-678.

31. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol* 2008: 32(3): 227-234.

32. Xu Z, Taylor J. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Research* 2009: 37(Web Server issue): W600-605.

33. Zissel G, Prasse A, Muller-Quernheim J. Immunologic response of sarcoidosis. *Semin Respir Crit Care Med*: 31(4): 390-403.

34. Barton A, Thomson W, Ke X, Eyre S, Hinks A, Bowes J, Gibbons L, Plant D, Wilson AG, Marinou I, Morgan A, Emery P, Steer S, Hocking L, Reid DM, Wordsworth P, Harrison P, Worthington J. Re-evaluation of putative rheumatoid arthritis susceptibility genes in the post-genome wide association study era and hypothesis of a key pathway underlying susceptibility. *Hum Mol Genet* 2008: 17(15): 2274-2279.

35. Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, Burtt NP, Gianniny L, Korman BD, Padyukov L, Kurreeman FA, Chang M, Catanese JJ, Ding B, Wong S, van der Helm-van Mil AH, Neale BM, Coblyn J, Cui J, Tak PP, Wolbink GJ, Crusius JB, van der Horst-Bruinsma IE, Criswell LA, Amos CI, Seldin MF, Kastner DL, Ardlie KG, Alfredsson L, Costenbader KH, Altshuler D, Huizinga TW, Shadick NA, Weinblatt ME, de Vries N, Worthington J, Seielstad M, Toes RE, Karlson EW, Begovich AB, Klareskog L, Gregersen PK, Daly MJ, Plenge RM. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nat Genet* 2008: 40(10): 1216-1223.

36. Plant D, Flynn E, Mbarek H, Dieude P, Cornelis F, Arlestig L, Dahlqvist SR, Goulielmos G, Boumpas DT, Sidiropoulos P, Johansen JS, Ornbjerg LM, Hetland ML, Klareskog L, Filer A, Buckley CD, Raza K, Witte T, Schmidt RE, Worthington J. Investigation of potential non-HLA rheumatoid arthritis susceptibility loci in a European cohort increases the evidence for nine markers. *Ann Rheum Dis*: 69(8): 1548-1553.

37. Fung EY, Smyth DJ, Howson JM, Cooper JD, Walker NM, Stevens H, Wicker LS, Todd JA. Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. *Genes Immun* 2009: 10(2): 188-191.

38. Cooper JD, Walker NM, Healy BC, Smyth DJ, Downes K, Todd JA. Analysis of 55 autoimmune disease and type II diabetes loci: further confirmation of chromosomes 4q27, 12q13.2 and 12q24.13 as type I diabetes loci, and support for a new locus, 12q13.3-q14.1. *Genes Immun* 2009: 10 Suppl 1: S95-120.

39. Bailey R, Cooper JD, Zeitels L, Smyth DJ, Yang JH, Walker NM, Hypponen E, Dunger DB, Ramos-Lopez E, Badenhoop K, Nejentsev S, Todd JA. Association of the

vitamin D metabolism gene CYP27B1 with type 1 diabetes. *Diabetes* 2007: 56(10): 2616-2621.

40. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat Genet* 2009: 41(7): 824-828.

41. Zhernakova A, Stahl EA, Trynka G, Raychaudhuri S, Festen EA, Franke L, Westra HJ, Fehrmann RS, Kurreeman FA, Thomson B, Gupta N, Romanos J, McManus R, Ryan AW, Turner G, Brouwer E, Posthumus MD, Remmers EF, Tucci F, Toes R, Grandone E, Mazzilli MC, Rybak A, Cukrowska B, Coenen MJ, Radstake TR, van Riel PL, Li Y, de Bakker PI, Gregersen PK, Worthington J, Siminovitch KA, Klareskog L, Huizinga TW, Wijmenga C, Plenge RM. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet*: 7(2): e1002004.

42. Liang Y, Ridzon D, Wong L, Chen C. Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics* 2007: 8: 166.

43. Collison A, Mattes J, Plank M, Foster PS. Inhibition of house dust mite-induced allergic airways disease by antagonism of microRNA-145 is comparable to glucocorticoid treatment. *J Allergy Clin Immunol*: 128(1): 160-167 e164.

44. Keller A, Leidinger P, Bauer A, Elsharawy A, Haas J, Backes C, Wendschlag A, Giese N, Tjaden C, Ott K, Werner J, Hackert T, Ruprecht K, Huwer H, Huebers J, Jacobs G, Rosenstiel P, Dommisch H, Schaefer A, Muller-Quernheim J, Wullich B, Keck B, Graf N, Reichrath J, Vogel B, Nebel A, Jager SU, Staehler P, Amarantos I, Boisguerin V, Staehler C, Beier M, Scheffler M, Buchler MW, Wischhusen J, Haeusler SF, Dietl J, Hofmann S, Lenhof HP, Schreiber S, Katus HA, Rottbauer W, Meder B, Hoheisel JD, Franke A, Meese E. Toward the blood-borne miRNome of human diseases. *Nat Methods*: 8(10): 841-843.

45. Kimura Y, Nakazawa M, Tsuchiya N, Asakawa S, Shimizu N, Yamada M. Genomic organization of the OS-9 gene amplified in human sarcomas. *J Biochem* 1997: 122(6): 1190-1195.

46. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR, Siepel A, Smith KE, Sugnet CW, Sultan-Qurraie A, Thomas DJ, Trumbower H, Weber RJ, Weirauch M, Zweig AS, Haussler D, Kent WJ. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* 2006: 34(Database issue): D590-598.

47. Munro S. The MRH domain suggests a shared ancestry for the mannose 6-phosphate receptors and other N-glycan-recognising proteins. *Curr Biol* 2001: 11(13): R499-501.

48. Christianson JC, Shaler TA, Tyler RE, Kopito RR. OS-9 and GRP94 deliver mutant alpha1-antitrypsin to the Hrd1-SEL1L ubiquitin ligase complex for ERAD. *Nat Cell Biol* 2008: 10(3): 272-282.

49. Alcock F, Swanton E. Mammalian OS-9 is upregulated in response to endoplasmic reticulum stress and facilitates ubiquitination of misfolded glycoproteins. *J Mol Biol* 2009: 385(4): 1032-1042.

50. Wahlstrom J, Dengjel J, Persson B, Duyar H, Rammensee HG, Stevanovic S, Eklund A, Weissert R, Grunewald J. Identification of HLA-DR-bound peptides presented by human bronchoalveolar lavage cells in sarcoidosis. *J Clin Invest* 2007: 117(11): 3576-3582.

51. Wahlstrom J, Dengjel J, Winqvist O, Targoff I, Persson B, Duyar H, Rammensee HG, Eklund A, Weissert R, Grunewald J. Autoimmune T cell responses to antigenic peptides presented by bronchoalveolar lavage cell HLA-DR molecules in sarcoidosis. *Clin Immunol* 2009: 133(3): 353-363.

52. Jansen BJ, Eleveld-Trancikova D, Sanecka A, van Hout-Kuijer M, Hendriks IA, Looman MG, Leusen JH, Adema GJ. OS9 interacts with DC-STAMP and modulates its intracellular localization in response to TLR ligation. *Mol Immunol* 2009: 46(4): 505-515.

53. Zaba LC, Smith GP, Sanchez M, Prystowsky SD. Dendritic cells in the pathogenesis of sarcoidosis. *Am J Respir Cell Mol Biol*: 42(1): 32-39.

54. Kulakova N, Urban B, McMichael AJ, Ho LP. Functional analysis of dendritic cell-T cell interaction in sarcoidosis. *Clin Exp Immunol*: 159(1): 82-86.

55. Mathew S, Bauer KL, Fischoeder A, Bhardwaj N, Oliver SJ. The anergic state in sarcoidosis is associated with diminished dendritic cell function. *J Immunol* 2008: 181(1): 746-755.

56. Bordignon M, Rottoli P, Agostini C, Alaibac M. Adaptive immune responses in primary cutaneous sarcoidosis. *Clin Dev Immunol*: 2011: 235142.

57. Liu PT, Stenger S, Li H, Wenzel L, Tan BH, Krutzik SR, Ochoa MT, Schauber J, Wu K, Meinken C, Kamen DL, Wagner M, Bals R, Steinmeyer A, Zugel U, Gallo RL, Eisenberg D, Hewison M, Hollis BW, Adams JS, Bloom BR, Modlin RL. Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response. *Science* 2006: 311(5768): 1770-1773.

58. Zasloff M. Fighting infections with vitamin D. *Nat Med* 2006: 12(4): 388-390.

59. Hansdottir S, Monick MM, Hinde SL, Lovan N, Look DC, Hunninghake GW. Respiratory epithelial cells convert inactive vitamin D to its active form: potential effects on host defense. *J Immunol* 2008: 181(10): 7090-7099.

60. Kavathia D, Buckley JD, Rao D, Rybicki B, Burke R. Elevated 1, 25-dihydroxyvitamin D levels are associated with protracted treatment in sarcoidosis. *Respir Med*: 104(4): 564-570.

61. Lockstone HE, Sanderson S, Kulakova N, Baban D, Leonard A, Kok WL, McGowan S, McMichael AJ, Ho LP. Gene set analysis of lung samples provides insight into pathogenesis of progressive, fibrotic pulmonary sarcoidosis. *Am J Respir Crit Care Med*: 181(12): 1367-1375.

Figure Legends

Figure 1: Regional plot of the validated sarcoidosis association at 12q13.3-q14.1

(a) Nominal-$\log_{10}$ p values obtained from an allelic χ^2 test with one degree of freedom in the fine-mapping in panel D. Fifty-three tagging SNPs were genotyped across the 500-kb region surrounding the GWAS lead SNP rs1050045 (highlighted by orange vertical line).The black dashed line corresponds to a threshold of 0.05. (**b**) Plots of the recombination intensity (cM/Mb) and the cumulative genetic distance (cM) from the lead SNP according to HapMap (CEU trios, Phase I + II). **(c)** Position and intron-exon structure of transcripts according to the NCBI Reference Sequence (RefSeq) collection. **(d)** rPairwise linkage disequilibrium values (r^2) in control individuals. Positions are given as NCBI build 36 coordinates.

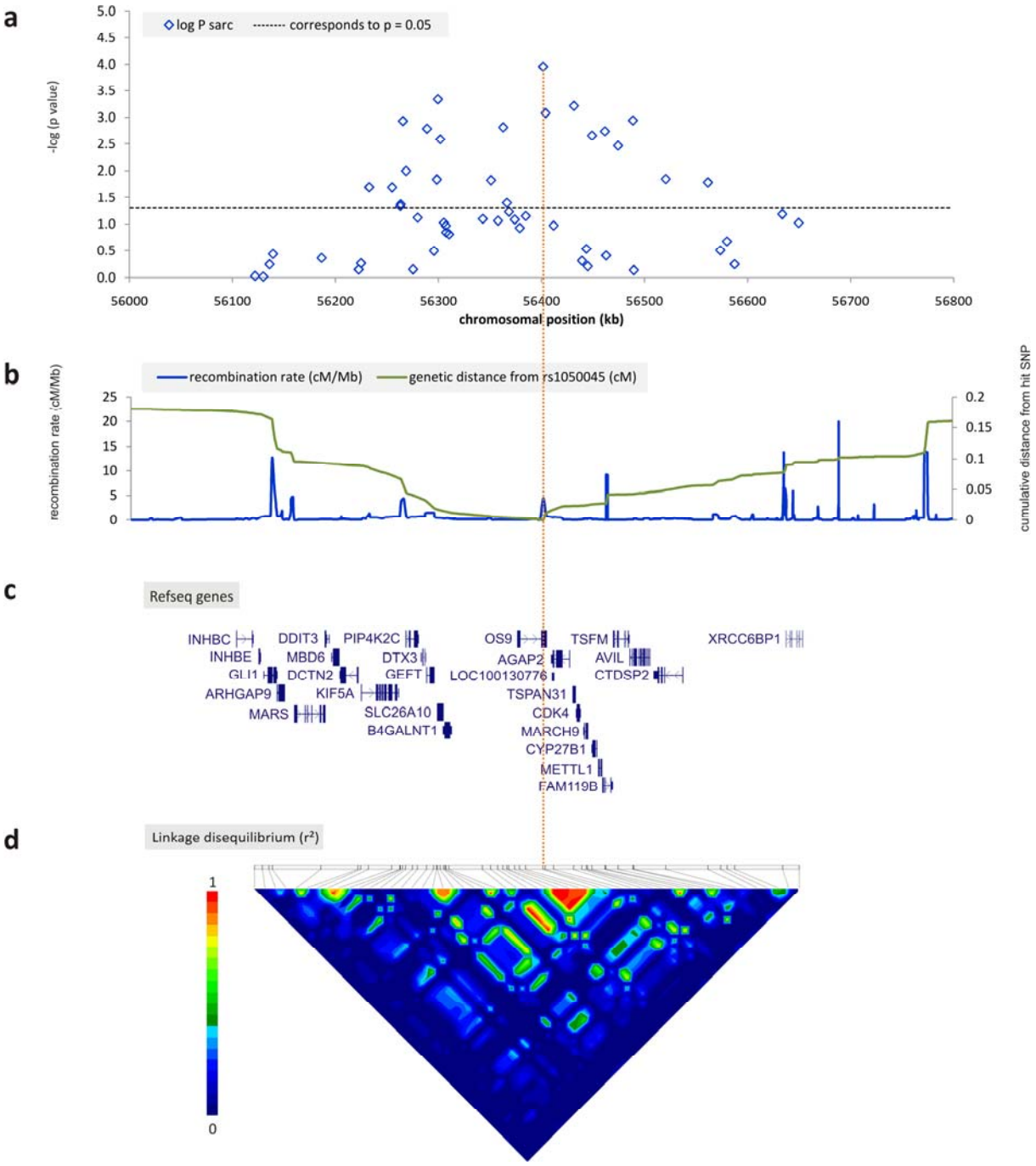


Figure 2: Expression of the eight candidate genes in BALs from sarcoidosis patients (n = 4) compared to unaffected individuals (n = 4, BAL panel I).

Transcription levels of *OS9*, *AGAP2*, *TSPAN31*, *CDK4*, *MARCH9*, *CYP27B1*, *METTL1* and *FAM119B* were normalized on *GAPDH* mRNA levels. The asterisk denotes significance at a five percent level based on a non-parametric Mann-Whitney U test.

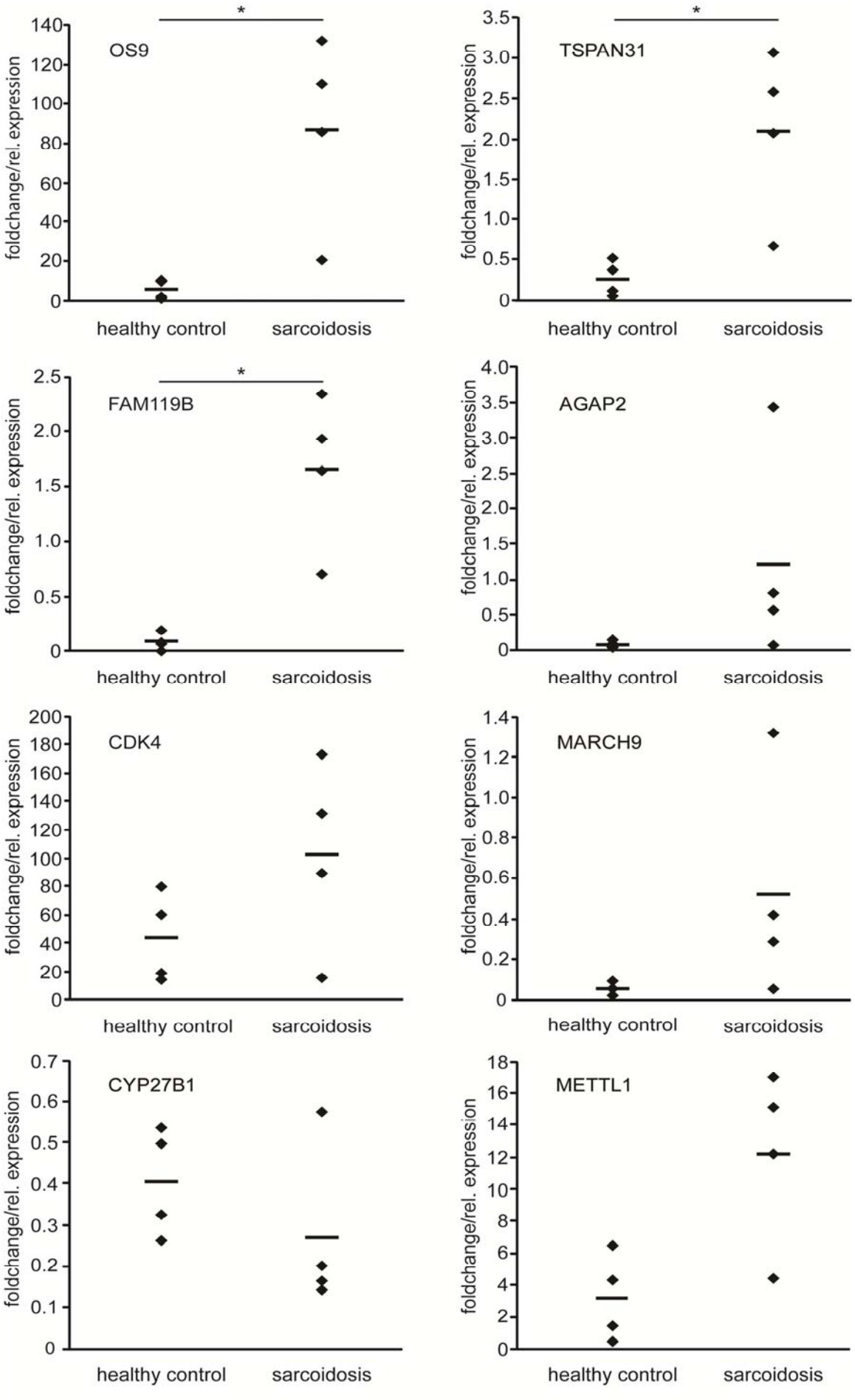

OS9
foldchange/rel. expression
140
120
100
80
60
40
20
0
*
healthy control
sarcoidosis
TSPAN31
foldchange/rel. expression
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
*
healthy control
sarcoidosis
FAM119B
foldchange/rel. expression
2.5
2.0
1.5
1.0
0.5
0
*
healthy control
sarcoidosis
AGAP2
foldchange/rel. expression
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
healthy control
sarcoidosis
CDK4
foldchange/rel. expression
200
180
160
140
120
100
80
60
40
20
0
healthy control
sarcoidosis
MARCH9
foldchange/rel. expression
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
healthy control
sarcoidosis
CYP27B1
foldchange/rel. expression
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
healthy control
sarcoidosis
METTL1
foldchange/rel. expression
18
16
14
12
10
8
6
4
2
0
healthy control
sarcoidosis

Figure 3: Allele-specific expression of OS9

Transcript levels of OS9 were measured using qRT-PCR in the patients of BAL panel II (n = 46). A change of mRNA expression pattern was significantly associated with the homozygote risk genotype of rs1050045 “CC” (p=0.016).

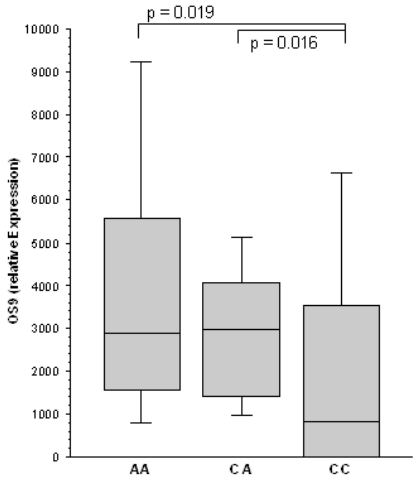
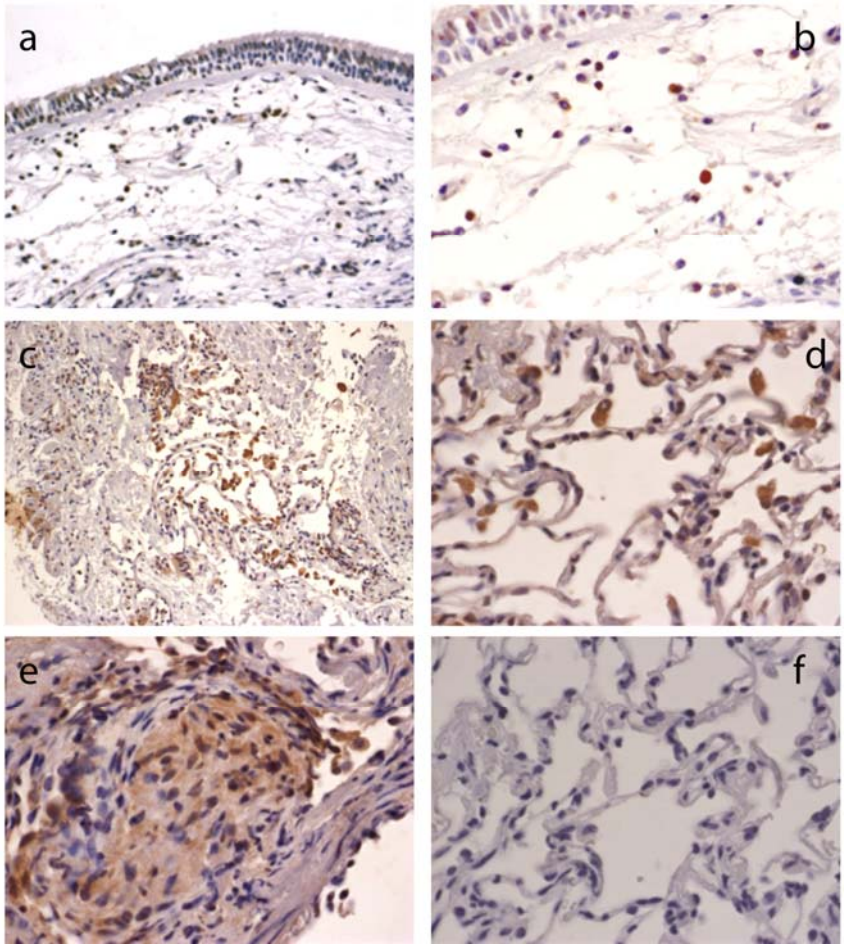


Figure 4: OS9 immunohistochemistry staining

Staining for OS9 in lung biopsy material (a, b), respiratory epithelium and submucosa of a healthy subject. Note strong OS9 staining (brown) of monocytic and lymphocytic cells. (c, d) Specimen from a sarcoidosis patient demonstrate positive staining of alveolar macrophages . (e) Diffuse granular staining of a granulomatous structure in a Sarcoidosis specimen. (f) Negative control of adjacent section from (d). Pictures show representative results from n = 4

healthy subjects and n = 8 Sarcoidosis patients. Nuclei were counterstained with Hematoxylin.



Tables

Table 1: Results for SNPs with nominally significant p values in the validation panel

Ninety-nine SNPs were tested for association with sarcoidosis (SA) and the acute and chronic subphenotypes in validation panel B. Data are shown for the 21 SNPs that were nominally significant in this analysis ($p < 0.05$), ranked by their P value. Nucleotide positions refer to NCBI build 36. **A1**: rare allele in controls. **AF**: allele frequencies. P: P value obtained from an allele-based χ^2 test with one degree of freedom. Significant P values after Bonferroni correction for multiple testing are highlighted in bold ($p < 0.05$). **OR [95 % CI]**: allelic odds ratios and 95% confidence intervals for allele A1. **Chr:** chromosome (Chr). **Co**: controls. **SA**: sarcoidosis patients. **chronic/acute**: patients with the respective subphenotype. **us/ds/i**: location of SNP upstream, downstream or intronic of gene locus, respectively. **ns**: not significant (ns). * denotes the p value of the lead SNP rs1050045 with the corresponding OR [95 % CI] = 1.30 [1.14-1.49]

				Screening panel A					Validation panel B							
						SA				SA			acute		chronic	
Chr	**Position (bp)**	**dbSNP ID**	**Gene**	**A1**	**AF Co**	**AF**	**P**	**OR [95% CI]**	**AF Co**	**AF val**	**P**	**OR [95% CI]**	**AF**	**P**	**AF**	**P**
12	**56,401,538**	**rs1050045**	***OS9, 3´UTR***	**C**	**0.40**	**0.48**	**1.62**$\times10^{-5}$	**1.35 [1.18-1.55]**	**0.42**	**0.46**	**7.38**$\times10^{-5}$	**1.20 [1.10-1.31]**	**0.48**	**9.36**$\times10^{-5}$*****	0.45	7.80×10^{-3}
9	17,696,884	rs16935914	*SH3GL2, i*	G	0.05	0.08	5.89×10^{-4}	1.59 [1.22-2.07]	0.07	0.09	6.05×10^{-4}	1.34 [1.13-1.58]	0.08	ns	0.09	8.03×10^{-4}
3	69,788,306	rs6549245	*MITF, us*	C	0.37	0.31	1.78×10^{-4}	0.76 [0.66-0.88]	0.36	0.33	1.44×10^{-3}	0.86 [0.78-0.94]	0.34	ns	0.32	1.89×10^{-3}
8	96,169,351	rs6471513	*AX747981, us*	C	0.49	0.43	3.08×10^{-4}	0.78 [0.68-0.89]	0.49	0.46	1.63×10^{-3}	0.87 [0.79-0.95]	0.43	9.92×10^{-4}	0.46	2.32×10^{-2}
2	113,633,383	rs12475781	*PSD4, us*	G	0.38	0.32	5.60×10^{-4}	0.78 [0.67-0.90]	0.37	0.34	2.06×10^{-3}	0.86 [0.79-0.95]	0.34	ns	0.34	3.52×10^{-2}
5	55,707,498	rs7736704	*ANKDR55, us*	C	0.26	0.32	4.36×10^{-5}	1.37 [1.18-1.59]	0.27	0.30	4.59×10^{-3}	1.15 [1.04-1.27]	0.29	ns	0.30	2.26×10^{-2}
18	17,050,176	rs9962826	*ROCK1, us*	G	0.01	0.04	2.40×10^{-5}	2.54 [1.62-3.97]	0.02	0.03	4.63×10^{-3}	1.50 [1.13-1.99]	0.03	2.61×10^{-2}	0.03	2.27×10^{-2}
2	203,264,771	rs6748088	*FAM117B, i*	C	0.30	0.36	5.92×10^{-4}	1.29 [1.11-1.48]	0.32	0.35	7.58×10^{-3}	1.14 [1.03-1.25]	0.33	ns	0.36	1.93×10^{-3}
14	58,969,773	rs2774052	*GPR135, i*	T	0.41	0.47	3.34×10^{-4}	1.28 [1.12-1.47]	0.43	0.45	8.01×10^{-3}	1.13 [1.03-1.23]	0.45	ns	0.46	1.32×10^{-2}
4	86,409,257	rs11735414	*ARHGAP24, us*	G	0.18	0.23	4.45×10^{-4}	1.35 [1.14-1.59]	0.19	0.21	8.38×10^{-3}	1.16 [1.04-1.29]	0.20	ns	0.21	2.08×10^{-2}
7	28,504,694	rs217498	*CREB5, i*	C	0.32	0.38	4.21×10^{-4}	1.29 [1.12-1.49]	0.33	0.36	8.42×10^{-3}	1.13 [1.03-1.24]	0.37	7.66×10^{-3}	0.35	ns
11	105,832,978	rs12798744	*GUCY1A2, ds*	A	0.30	0.36	3.34×10^{-4}	1.30 [1.13-1.50]	0.30	0.33	1.03×10^{-2}	1.13 [1.03-1.25]	0.33	ns	0.34	1.41×10^{-2}
2	45,348,647	rs17033293	*UNQ6957, us*	C	0.14	0.09	9.12×10^{-5}	0.64 [0.52-0.80]	0.12	0.10	1.10×10^{-2}	0.83 [0.72-0.96]	0.11	ns	0.10	1.74×10^{-2}
14	64,352,830	rs2285003	*SPTB, i*	C	0.07	0.11	3.39×10^{-5}	1.65 [1.30-2.09]	0.08	0.09	1.35×10^{-2}	1.22 [1.04-1.42]	0.10	7.62×10^{-3}	0.09	ns
18	54,697,058	rs17694691	*ZNF532, i*	C	0.27	0.34	3.01×10^{-5}	1.36 [1.18-1.58]	0.28	0.30	2.84×10^{-2}	1.11 [1.01-1.23]	0.30	ns	0.30	ns
16	50,359,507	rs10163352	*intergenic*	G	0.34	0.28	5.87×10^{-4}	0.77 [0.67-0.89]	0.32	0.30	2.97×10^{-2}	0.90 [0.82-0.99]	0.29	ns	0.30	ns
2	217,520,622	rs1921998	*TNP1, us*	C	0.47	0.41	6.06×10^{-4}	0.79 [0.69-0.90]	0.46	0.43	3.63×10^{-2}	0.91 [0.83-0.99]	0.45	ns	0.43	4.20×10^{-2}
12	99,167,581	rs3887427	*DEPDC4, i*	T	0.06	0.09	3.93×10^{-5}	1.70 [1.32-2.20]	0.07	0.08	3.69×10^{-2}	1.19 [1.01-1.40]	0.08	ns	0.08	ns
9	119,649,893	rs995988	*TLR4, ds*	C	0.49	0.42	1.04×10^{-4}	0.76 [0.67-0.88]	0.49	0.46	3.75×10^{-2}	0.91 [0.83-0.99]	0.45	1.57×10^{-2}	0.48	ns
11	55,537,527	rs1552151	*OR5F1, us*	T	0.14	0.10	2.41×10^{-4}	0.67 [0.54-0.83]	0.14	0.12	4.52×10^{-2}	0.87 [0.77-1.00]	0.12	ns	0.12	2.53×10^{-2}
11	50,052,671	rs2201637	*OR4C12, us*	T	0.13	0.09	1.73×10^{-4}	0.65 [0.52-0.82]	0.12	0.11	4.58×10^{-2}	0.87 [0.76-1.00]	0.12	ns	0.10	7.30×10^{-3}

Table 2: Detected variants in the *OS9* exonic, exon-flanking and regulatory regions

Samples were enriched for homozygote and heterozygote carriers of the rs1050045 risk allele C and sequenced using Sanger sequencing technology. The location is given relative to the *OS9* gene region and the numbers of homozygotes for the minor allele, heterozygotes and homozygotes for the major allele are denoted by "A1A1/A1A2/A2A2".

SNP	Position (bp, hg19)	Location	Allele A1>A2	Context sequence	A1A1/A1A2/A2A2 cases	A1A1/A1A2/A2A2 controls
rs4760319	58,087,486	upstream	T>C		41/4/0	40/4/1
rs4760168	58,087,737	upstream	G>T		26/13/6	30/13/2
OS9-SNP1	58,088,665	non-synon (Pro>Thr)	C>A	AAGGGAGGAGGAAACACCTGCTTACCAAG GGCCTGGGATC[C/A]CTGAGTTGTTGAGCCC AATGAGAGATGCTCCCTGCTTGCTGA	45/0/0	41/1/0
rs114532828	58,090,051	intron	T>C		45/0/0	44/1/0
rs73125477	58,109,707	coding-synon	T>G		45/0/0	44/1/0
rs799265	58,112,189	coding-synon	G>A		22/11/8	24/12/9
OS9-SNP2	58,114,081	intron	C>T	CTGCAGGTGGGCCCTGGAGGGCGGCTGGAC CCAGTGCTGT[C/T]GGAAGGGCAAGCTGCC GGAAGTGGAGGGGCTGGGACCAGT	45/0/0	44/1/0
rs1050045	58,115,271	3′-UTR	T>C		11/12/24	10/9/24
rs74368191	58,115,286	3′-UTR	G>A		43/2/0	45/0/0

Abbreviations

95% CI	95% confidence interval
AIC	Akaike's Information Criterion
BAL	Bronchoalveaolar lavage
DC	Dendritic cell
ER	Endoplasmatic reticulum
GWAS	Genome-wide association study
HWE	Hardy-Weinberg equilibrium
LD	Linkage disequilibrium
MAF	Minor allele frequency
OR	Odds ratio
PAR	Population attributable risk
qRT-PCR	quantitative real-time PCR