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

Online Supplemental Material

Sylvia Hofmann PhD[1,*], Annegret Fischer PhD[1,*], Andreas Till PhD[1,*], Joachim Müller-Quernheim MD[2], Robert Häsler PhD[1], Andre Franke PhD[1], Karoline Gäde PhD[3], Heidi Schaarschmid MSc[1], GenPhenReSa Consortium[a], Philip Rosenstiel MD[1], Almut Nebel PhD[1], Manfred Schürmann MD[4], Michael Nothnagel PhD[5], Stefan Schreiber MD[1, 6]

[1] Institute for Clinical Molecular Biology; Christian-Albrechts-University; Kiel, Germany

[2] Department of Pneumology; University of Freiburg; Freiburg, Germany

[3] Department of Clinical Medicine; Research Center Borstel; Borstel, Germany

[4] Institute of Human Genetics; University of Lübeck; Lübeck, Germany

[5] Institute of Medical Informatics and Statistics; Christian-Albrechts-University; Kiel, Germany

[6] Department of General Internal Medicine; University Hospital Schleswig-Holstein; Christian-Albrechts-University; Kiel, Germany

[*] These authors contributed equally to the work.

[a] GenPhenReSa Consortium (see attachment with the detailed affiliations of the consortium members)

Table of contents **page**

.

Appendix S1: Affiliations of the members of the GenPhenReSa Consortium

Andreas Günther - Lung Center; University of Giessen and Lung Clinic Waldhof-Elgershausen; Greifenstein, 35753; Germany; e-mail: Andreas.Guenther@innere.med.uni-giessen.de

Anna Dubaniewicz - Department of Pneumonology; Medical University of Gdansk; Gdansk, 80211; Poland; e-mail: aduban@gumed.edu.pl

Stefan Pabst - Medical Clinic II, Department of Pneumology; University of Bonn; Bonn, 53105; Germany; e-mail: stefan.pabst@ukb.uni-bonn.de

Dragos Bumbacea - Marius Nasta Institute of Pulmonology; Bucharest, 050159; Romania; e-mail: d.bumbacea@gmail.com

Antje Prasse - Department of Pneumology, University Medical Center, Albrecht Ludwigs University; Freiburg, 70105; Germany; e-mail: antje.prasse@uniklinik-freiburg.de

Alexandar Grubanovic - Department of Pneumology, University Medical Center, Albrecht Ludwigs University; Freiburg, 70105; Germany; e-mail: aleksandar.grubanovic@uniklinik-freiburg.de

Paola Rottoli - Respiratory Diseases Section; Clinical Medicine and Immunology; Siena, 53100; Italy; e-mail: rottoli@unisi.it

Elena Bargagli - Respiratory Diseases Section; Clinical Medicine and Immunology; Siena, 53100; Italy; e-mail: bargagli2@unisi.it

Paul Bresser - Department of Pulmonology, F5-144; Academic Medical Center; Amsterdam 1100; The Netherlands; e-mail: p.bresser@amc.uva.nl

Venerino Poletti - Ospedale G.B. Morgagni; Dipartimento di Malattie dell'Apparato Respiratorio e del Torace U.O. Endoscopia Toracica; Forli, 47100; Italy; e-mail: v.poletti@ausl.fo.it

Maurizio Luisetti - Clinica Malattie Apparato Respiratorio - IRCCS Policlinico San Matteo-Pavia; University of Pavia; Pavia, 27100; Italy; e.mail: m.luisetti@smatteo.pv.it

Violeta Vucinic - Medical School Beograd; Beograd, 11000; Serbia; e-mail: violeta.vucinic@kcs.ac.yu

Jelica Videnovic-Ivanov - Medical School Beograd; Beograd, 11000; Serbia; e-mail: jelica.videnovic@kcs.ac.yu

Appendix S2: Patient recruitment and phenotyping

German sarcoidosis patients of panels A (GWAS), B (validation) and C (fine mapping) 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. All patients, but no control individuals, of panel A were included in the screening panel of our recent GWAS study [1], and the panels B and C were already used in previous studies [2]. Panel C comprises panel B and parts of panel A (1,837 sarcoidosis patients and 1,806 controls). Panel D (replication) comprised sarcoidosis patients from different European locations (n=19 Amsterdam, Netherlands; n=330 Belgrad, Serbia; n=38 Gdansk, Poland; n=15 Bukarest, Romania; n=8 Forli, n=28 Pavia, n=239 Sienna, all Italy; n=26 Berlin, n=20 Borstel, n=48 Gießen, n=164 Freiburg, all Germany) that were recruited within a European network Project. The diagnosis of the participating patients was based again on the International Consensus Statement (ICS) on Sarcoidosis [3], pulmonary function testing, routine serum parameters, and chest X-ray. In addition, a documented course of disease over at least 2 years prior to recruitment was required for all patients contacted within the GenPhenReSa project. The population-representative German controls (n=2,564) of panel D were recruited as reported in detail elsewhere [4]. Informed written consent was obtained from all study participants and all collection protocols were approved by the institutional review committees of the participating centres.

Appendix S3: SNP genotyping with the Affymetrix 100k Gene Chip Array

In the first, exploratory stage, we genotyped 384 sarcoidosis patients and 399 controls (panel A) using the Affymetrix GeneChip® Human Mapping 50K Xba and Hind Arrays (Affymetrix, Santa Clara, CA, USA). Genotypes of 116,204 SNPs were called with the GeneChip® DNA Analysis Software (GDAS v2.0, Affymetrix). Subjects whose percentage of missing genotypes was more than 5% (n=2) were removed from the data set using missingness statistics in PLINK v1.06 [5]. Pair-wise percentage IBS (identity-by-state) values were computed also by PLINK v1.06 for all individuals and all markers. “Outliers” were detected by comparing the distribution of the IBS values for each individual with the combined IBS distribution of the entire population. Two types of “outliers” were defined: 1) individuals less related to the entire population than expected were defined as those for whom >60% of the IBS values were smaller than the median minus three times the interquartile range (3×IQR) of the population distribution. In this case, the individual was removed from

the population (n=6); 2) individuals with a close cognate relative in the population were defined as those who had at least one observed IBS value above the median plus 3´IQR. In this case, the member of the cognate pair with the lower call rate was removed from the population (n=2).

This resulted in 381 sarcoidosis samples and 392 controls that were included in the analysis. Gender was checked by the proportion of heterozygous SNP genotypes on the X chromosome and could be confirmed for all samples. SNPs that had a sample call rate ≤ 90%, a genotyping success rate ≤ 95%, a minor allele frequency < 2% in controls, or deviated from Hardy-Weinberg equilibrium (HWE) in the controls ($P_{HWE} \leq 0.01$) were eliminated from subsequent analyses (n=19,116 SNPs/ 16.45%). Data quality protocols were enforced by using the implementation in PLINK v1.06. Experimental details concerning the genotyping of the 100k SNP set are provided in Matsuzaki *et al.* [6].

Appendix S4: Selection criteria of SNPs for validation and fine mapping

We applied the following criteria to select SNPs for the validation stage:

1) Ranking of markers by their allelic p value (allelic χ^2 test with one degree of freedom, multiplicative risk model and adjusted for population stratification with λ_{GC}=1.076; for details see [7]). The 250 top-ranking SNPs were subjected to the subsequent selection steps.
2) Of these, the 26 SNPs that were located in the HLA region (*6p21.1-6p21.3*) were not included in the validation stage because a strong disease association has already been established for a number of HLA loci and for the neighbouring *BTNL2* gene [8].
3) For the remaining 224 SNPs, the correct assignment of genotypes by the calling algorithm was assured by post-hoc visual inspection of the respective cluster plots, leading to the exclusion of another 140 SNPs.
4) To avoid false-positive results due to genotyping errors, SNPs that passed the aforementioned quality criteria subjected to the clumping approach as implemented in PLINK v1.06 with the following parameters: $p_{clumping\ SNP} < 0.05$, $r^2 > 0.5$, maximal physical distance < 250kb. Fifty-nine SNPs that were not accompanied by a clumping SNP, or where the cluster plots of the clumping SNP(s) indicated insufficient genotype quality were subsequently excluded. Cluster plots of the remaining 25 variants are shown in the supplementary Figure E1.

We applied the following procedure to select SNPs for the fine-mapping stage 1:

1) The target region for fine-mapping stage 1 was defined considering the location of genes and the LD structure according to HapMap data (HapMap CEU trios: CEPH Utah residents with ancestry from northern and western Europe, phase II, build 36; available from URL (1))
2) The SNPs were selected via pairwise tagging using Haploview [9] (see URL (2)) in the HapMap CEU trio, applying the following criteria: minor allele frequency > 1%; tagging threshold $r^2 > 0.7$; $P_{HWE} > 0.01$; For technical reasons 44 of those SNPs were selected aiming at the representation of a maximum number of HapMap SNP and with a focus on functional variants.

We applied the following procedure to select SNPs for the extended fine-mapping (stage 2):

1) Focussed on the region associated as defined by the fine-mapping stage 1.
2) Another 46 tagSNPs were selected via pairwise tagging [9] as described for stage 1 except for a more stringent tagging threshold ($r^2 > 0.8$).

Appendix S5: SNPlex and TaqMan genotyping

For validation and the subsequent fine mapping of the validated 6p12.1 region, ligation-based SNPlex™ genotyping (Applied Biosystems, Foster City, CA, USA) was performed as previously described [10]. Genotype assignments were visually inspected using the Genemapper 4.0 (Applied Biosystems) software. For fine mapping, 90 tagSNPs were selected via pairwise tagging [9] (see URL (2)) in the HapMap CEU trio (CEPH Utah residents with ancestry from northern and western Europe, phase II, build 36) data (available from URL (1)) and applying the following criteria: minor allele frequency > 1%; tagging threshold $r^2 > 0.8$; $P_{HWE} > 0.01$; less than two Mendelian errors. A TaqMan® SNP Genotyping Assay (Applied Biosystems) was additionally used to genotype marker rs10484410 in the validation panel for technical validation. An overview of the study design is given in Figure S2.

Appendix S6: Statistical analysis

Single-marker case-control analysis of allele and genotype frequency data was performed using χ^2tests with one and two degrees of freedom respectively, using PLINK v1.06 (see URL (3)). An allelic $P_{CCA} < 0.05$ was considered nominally significant. All genotyped SNPs were tested for HWE using the PLINK v1.06 implementation of an exact test [11] before being considered for further analyses ($P_{HWE} > 0.01$). Nominal P-values were adjusted for multiple testing using Bonferroni correction, accounting for all 25 markers in the validation stage.

Downstream single-marker association analysis, permutation testing, calculation of pair-wise LD, and SNP selection were performed using Haploview 4.0 [9] (see URL (2)). To evaluate the overall significance of the genome-wide association results and the potential impact of population stratification the quantile-quantile plot (fig. 1, main text) was created using R v2.8.2 [12] (see URL (4)). We tested the independence of association signals at the 6p12.1 locus using a logistic regression that was carried out also in R v2.10.1, using the glm function from the stats package. Backward model selection was applied using Akaike's Information Criterion (AIC) [13] and the step function in R from the same package. Nonparametric bootstrapping was implemented in Perl using own scripts. Graphical summaries of LD were generated with GOLD [14] (see URL (5)). Power calculations were performed using PS Power and Sample Size v2.130 [15] (see URL (6)). To adjust for potential population stratification within the study sample [16], we used the genome-wide genotype data described elsewhere [17] as a reference for the genetic background of individuals. Values for each of the first six principal components were averaged over all individuals from each of the 23 studied subpopulations, respectively. Individuals from our study were then assigned to the closest of the reference subpopulations (cases: Amsterdam: NL from [17]; Belgrad: YU; Gdansk: PO; Bukarest: RO; Forli, Pavia, Sienna: I1; Berlin, Borstel, Gießen: NG; Freiburg: SG; German controls: NG and SG). The averaged first six principal component values were included in a logistic regression model to adjust the phenotypic association analysis for the sampling location of each individual. Significance of the phenotypic association was assessed by a likelihood-ratio test.

We assigned each individual from our panel to the closest subpopulation of a European data set described elsewhere [17] and subsequently included the subpopulation-specific average values of the first six principal components (PC) of the genome-wide data of this reference panel to adjust for population stratification.

Appendix S7: Analysis of tissue-specific expression by RT-PCR

For investigation of tissue-specific expression patterns of the candidate genes in the fine mapped region, we used a commercially available tissue panel from Clontech (Palo Alto, CA, USA) and standard RT-PCR procedures. Briefly, the PCR program used in general was applied as follows: Denaturation for 5 min at 95°C; 33 cycles of 20 sec at 95°C, 20 sec at 53°C, 60 sec at 72°C; final extension for 5 min at 72°C. The following primers were used for amplification of the five candidate genes located in the validated region:

BAG2-F CAACGCTAAAGCCAACGAG,
BAG2-R CAATAATCCTTGTGGCATGC;
C6orf65 (*BEND6*)-F CCAGTTTCCTTAAAGCCTGA,
C6orf65 (*BEND6*)-R CTGGCAGAAATTCTCTGGTT;
KIAA1586-F GACCATCGAGACCTGTTCTTG,
KIAA1586-R CCTTTGCCTTAGACAATCTGCA;
RAB23-F CAACAAATAGCTGAGGATCCAG,
RAB23-R GAGCCATTAGGAGCAAAGTCT;
ZNF415-F CTGTGAATGCAATCAGCGAT,
ZNF415-R GTTGTAGAGGATGATGGCAAAG.
To confirm the use of equal amounts of RNA in each experiment, all samples were checked in parallel for GAPDH expression using the following primers: GAPDH-F ccagccgagccacatcgc, GAPDH-R atgagccccagccttctccat. All amplified DNA fragments were analyzed on 1% agarose gels and subsequently documented on a BioDoc Analyzer (Biometra, Göttingen, Germany).

Appendix S8: Bronchoalveolar lavage

Bronchoalveolar lavage (BAL) cell samples were taken from five patients with active sarcoidosis without steroid treatment at the time of BAL (1 male; 4 female; average age: 33.6 ± SEM 5.4 years). The diagnosis of non-acute sarcoidosis was established retrospectively in accordance with previously defined criteria [18] including non-caseating granuloma identified by transbronchial biopsies. All sarcoidosis patients showed clinical signs for an active disease. Five individuals, who underwent bronchoscopy due to chronic cough, and who had retrospective unaffected lung, no infection and were non-malignant at time of lavage served as controls. 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 (n=5) gave their informed consent to the study. BAL was performed as previously described [19]. Briefly, 200-300 ml of sterile saline (0.9% NaCl) was instilled in 25 ml aliquots into a lingula or middle lobe segment. Each aliquot was immediately aspirated. Mean recovery was 70.5% and the average cell number/100 ml of BAL cells was 26.3 ± 9.3x106. The cells were centrifuged at 500 x g and washed three times with phosphate buffered saline (PBS) at +4°C. Cell differentials were

determined by counting at least 200 cells on a cytocentrifuge preparation (Cytospin II, Shandon Instruments, Sewickley, PA, USA). Cells were stained with Haemacolor (Merck, Darmstadt, FRG). The cell suspensions contained >92% alveolar macrophages. The viability of the cells was 95%. The cell pellet was shock frozen in liquid N_2.

Appendix S9: mRNA isolation and real-time PCR

Total RNA was isolated from snap-frozen BAL-derived cells using a commercial kit (RNeasy, Qiagen, Hilden, Germany). For detection of mRNA levels of *C6orf65 (BEND6), BAG2, KIAA1586, ZNF415* and *RAB23*, cDNA was synthesized from 500 ng of total RNA using the Advantage RT-for-PCR kit (Clontech Laboratories, Palo Alto, CA) according to the manufacturer's protocol. SYBR Green PCR Master Mix (Applied Biosystems) and the 7900HT Fast Real Time PCR system was used for quantitative Real-time PCR using the same target-specific primers as described above. Transcript amounts were normalized to those of the housekeeping gene *beta-Actin* using following primers: ACTB-F GATGGTGGGCATGGGTCAG, ACTB-R CTTAATGTCACGCACGATTTCC. Relative expression levels of the target genes were checked for significant differences between sarcoidosis patients and healthy controls (n=5 each) using non-parametric Mann-Whitney U test.

Appendix S10: eQTL anaylsis

Genome wide mRNA expression data, originating from from 25 HapMap individuals (children of HapMap trios) on a HumanWG-6 v1.0 Expression BeadChip (Illumina, Inc.) was obtained from the GENEVAR consortium (URL (7)). Differentially expressed genes were determined by using the Mann-Whitney u-test, and p-values were subsequently corrected for multiple testing using the Benjamini-Hochberg correction [20]. Genes with a corrected p-value < 0.05 were considered significantly differentially expressed. To generate functional transcript maps, clusters of significantly regulated transcripts were created for each genotype comparison individually and displayed as heatmaps (clustering method: UPGMA (unweighted average), distance measure: correlation, software: Spotfire DSMA 9.1, TIBCO, Somerville USA). Gene Ontology analysis was performed as previously published [21] by comparing significantly upregulated and downregulated transcripts. Biological processes associated to the transcripts were retrieved from the Gene Ontology Consortium (URL (8)).

Table S1: Summary of GWAS (panel A) and validation (panel B) results for the top 25 SNPs that were followed up.

Original genotype counts and analysis results are shown for sarcoidosis GWAS (light grey header, n=381 cases [**SA**], n=392 controls [**U**]) and validation stage (dark grey header, n=1,582 cases, n=1,783 controls). Lead variants of the clumped regions are marked by an asterisk and ranked according to their p value adjusted for λ_{GC} in the GWAs analysis. Asymptotic **P values** were calculated by a Pearson χ^2 test with one degree of freedom. Column **P value$_{corr.}$** shows p values that were corrected for multiple testing using Bonferroni correction in the validation stage (adjusted significance threshold: 2.0×10^{-3}). Frequencies are listed for the rarer allele **A1** in unaffected controls (**F_U_A1**), and sarcoidosis patients (**F_SA_A1**). Positions are from NCBI build 36.1. Odds ratios (**OR**) including 95% confidence intervals (**95% CI**) are shown for A1. **U P$_{HWE}$**: Hardy-Weinberg p value for the control sample. Results for the validated SNP rs10484410 are highlighted in bold.

Rank	Probe Set ID	dbSNP ID	Chr.	Position	A1	A2	U	U P_{HWE}	F_U_A1	SA	F_SA_A1	P value adjusted for λ_{GC}	OR (95% CI)	U	U P_{HWE}	F_U_A1	SA	F_SA_A1	P value$_{corr.}$	OR (95% CI)
1	SNP_A-1669361	rs1434308	9	28,841,909	C	T	77/189/126	0.68	0.44	107/195/78	0.54	1.37E-04	1.50 (1.23-183)	397/925/461	0.14	0.48	366/738/430	0.48	1.00×10^{0}	0.99 (0.90-1.09)
2	SNP_A-1715785	rs717996	12	82,657,951	C	T	82/183/127	0.31	0.44	48/169/164	0.35	2.38E-04	0.67 (0.55-0.82)	289/851/639	0.80	0.40	261/757/564	0.40	1.00×10^{0}	1.01 (0.92-1.12)
3	**SNP_A-1685894**	**rs10484410**	**6**	**57,104,647**	**G**	**A**	**5/84/301**	**1.00**	**0.12**	**14/117/250**	**0.19**	**2.64E-04**	**1.72 (1.29-2.27)**	**28/390/1349**	**0.39**	**0.13**	**44/398/1139**	**0.15**	**2.93×10^{-2}**	**1.26 (1.10-1.44)**
4	SNP_A-1753679	rs7962346	12	61,952,763	C	T	8/97/286	1.00	0.14	21/123/237	0.22	3.85E-04	1.64 (1.26-2.13)	54/494/1176	0.89	0.17	46/418/1083	0.16	1.00×10^{0}	0.93 (0.82-1.06)
5	SNP_A-1726532	rs10495266	1	225,313,639	G	C	4/70/317	1.00	0.10	12/99/270	0.16	5.19E-04	1.74 (1.28-2.35)	30/393/1355	0.00	0.13	29/373/1180	0.14	1.00×10^{0}	1.08 (0.94-1.25)
6	SNP_A-1755680	rs951901	12	56,684,632	G	T	60/192/140	0.75	0.40	42/153/186	0.31	5.77E-04	0.68 (0.55-0.84)	222/836/718	0.26	0.36	193/696/689	0.34	1.00×10^{0}	0.93 (0.84-1.02)
7	SNP_A-1696822	rs2383784	9	28,934,277	C	T	15/124/253	1.00	0.20	4/91/286	0.13	6.58E-04	0.61 (0.46-0.80)	45/507/1231	0.80	0.17	45/442/1047	0.17	1.00×10^{0}	1.02 (0.90-1.16)
8	SNP_A-1717685	rs1873942	16	58,553,352	T	C	12/117/263	1.00	0.18	25/143/213	0.25	7.22E-04	1.55 (1.21-1.98)	92/623/1064	0.69	0.23	59/512/1009	0.20	1.55×10^{-1}	0.85 (0.75-0.95)
9	SNP_A-1704771	rs2078539	16	60,141,739	C	T	41/189/161	0.22	0.35	27/146/205	0.26	7.74E-04	0.68 (0.55-0.84)	165/731/881	0.07	0.30	132/687/761	0.30	1.00×10^{0}	1.01 (0.91-1.12)
10	SNP_A-1688612	rs10494467	1	164,817,060	A	G	5/58/329	0.19	0.09	1/31/349	0.04	8.68E-04	0.48 (0.31-0.73)	11/246/1522	0.79	0.08	6/208/1361	0.07	1.00×10^{0}	0.92 (0.77-1.11)
11	SNP_A-1663956	rs4856497	3	80,215,335	A	G	76/185/131	0.47	0.43	101/191/88	0.52	9.35E-04	1.42 (1.16-1.74)	424/892/435	0.89	0.50	356/784/437	0.47	1.00×10^{0}	0.91 (0.83-1.01)
12	SNP_A-1656135	rs1362941	5	59,037,578	C	T	27/174/185	0.11	0.30	14/135/222	0.22	1.19E-03	0.67 (0.53-0.85)	129/708/938	0.50	0.27	102/596/874	0.25	1.00×10^{0}	0.91 (0.82-1.02)
13	SNP_A-1669289	rs10489924	1	99,231,562	C	T	6/97/288	0.67	0.14	3/59/319	0.09	1.20E-03	0.58 (0.42-0.80)	19/318/1440	0.13	0.10	15/266/1300	0.09	1.00×10^{0}	0.93 (0.79-1.09)
14	SNP_A-1685746	rs866632	9	28,845,071	G	A	51/172/169	0.51	0.35	27/152/201	0.27	1.33E-03	0.69 (0.56-0.86)	139/813/827	0.37	0.31	173/698/701	0.33	6.45×10^{-1}	1.12 (1.01-1.25)
15	SNP_A-1725266	rs694788	5	113,954,835	C	T	71/179/142	0.30	0.41	93/190/98	0.49	1.38E-03	1.41 (1.15-1.72)	378/846/553	0.11	0.45	320/778/483	0.45	1.00×10^{0}	0.99 (0.90-1.09)
16	SNP_A-1729116	rs7132697	12	66,819,108	T	A	40/173/178	0.91	0.32	28/133/220	0.25	1.56E-03	0.69 (0.55-0.86)	175/707/894	NA	0.30	136/647/798	0.29	1.00×10^{0}	0.97 (0.87-1.07)
17	SNP_A-1667448	rs185793	15	45,483,437	T	A	62/183/147	0.67	0.39	37/164/180	0.31	1.67E-03	0.71 (0.57-0.87)	188/786/805	0.12	0.33	194/696/692	0.34	1.00×10^{0}	1.08 (0.97-1.19)
18	SNP_A-1756072	rs10483261	14	21,426,391	C	T	106/185/101	0.27	0.51	67/189/125	0.42	1.73E-03	0.72 (0.59-0.88)	423/872/484	0.15	0.48	328/793/455	0.46	1.00×10^{0}	0.91 (0.83-1.00)
19	SNP_A-1690293	rs10483437	14	32,905,552	A	T	39/185/168	0.31	0.34	23/152/206	0.26	1.74E-03	0.70 (0.56-0.87)	145/749/885	0.75	0.29	145/658/776	0.30	1.00×10^{0}	1.04 (0.94-1.16)
20	SNP_A-1692734	rs2351010	5	96,319,685	A	G	81/208/101	0.19	0.47	56/187/138	0.39	1.75E-03	0.72 (0.58-0.88)	324/861/592	0.73	0.42	310/784/487	0.44	1.00×10^{0}	1.08 (0.98-1.19)
21	SNP_A-1650214	rs10487432	7	125,458,829	T	C	57/165/170	0.12	0.36	27/159/195	0.28	1.90E-03	0.70 (0.57-0.87)	174/774/835	0.96	0.32	154/636/744	0.31	1.00×10^{0}	0.96 (0.87-1.07)
22	SNP_A-1667216	rs920956	13	92,901,302	G	C	53/201/138	0.17	0.39	35/169/177	0.31	2.00E-03	0.71 (0.58-0.88)	242/842/694	0.86	0.37	187/714/681	0.34	3.33×10^{-1}	0.88 (0.80-0.97)
23	SNP_A-1733718	rs7920803	10	28,481,267	T	C	29/176/185	0.18	0.30	17/140/224	0.23	2.11E-03	0.69 (0.55-0.87)	109/666/980	0.69	0.25	83/588/904	0.24	1.00×10^{0}	0.93 (0.84-1.05)
24	SNP_A-1675007	rs1439172	14	45,797,214	G	A	19/122/251	0.44	0.20	33/141/205	0.27	2.17E-03	1.47 (1.16-1.86)	97/612/1062	0.80	0.23	82/538/952	0.22	1.00×10^{0}	0.98 (0.87-1.10)
25	SNP_A-1643846	rs4941920	13	38,740,373	T	G	43/169/180	0.73	0.33	28/136/217	0.25	2.19E-03	0.70 (0.56-0.87)	146/751/878	0.77	0.29	133/664/773	0.30	1.00×10^{0}	1.01 (0.91-1.12)

Table S2: Results of the fine mapping stage (panel C).

Results for the 44 HapMap tagging SNPs located around the 6p12.1 lead variant rs10484410 (marked by grey shading) that were successfully genotyped in panel C (1,806 controls [**U**], 1,837 sarcoidosis patients [**SA**]). Inter-marker distances are given in kilobases [**kb**]. Pairwise linkage disequilibrium (LD) is quantified by r^2, and r^2 values > 0.80 are highlighted in bold. Significant p values ($p < 0.05$) of the analysis of the whole sample [**SA all**] and of the subphenotypes [**chronic/acute**] are highlighted by bold italic (SNPs with a call-rate ≤ 0.95 were not considered). CR: call-rate in the overall sample. For further description of column headers, see Table S1.

#	dbSNP ID	Position	Locus	A1	A2	Distance [kb]	r^2 with rs104844104	r^2 with neighbouring SNP	CR	U P_{HWE}	U	F_U_A1	SA all	F_SA all_A1	SA all P value	SA chronic (n=1,065) P value	SA acute (n=603) P value
1	rs10484787	56,412,250		A	G	176	0.00	0.00	0.99	0.90	19/325/1447	0.10	17/325/1466	0.10	7.47×10^{-1}	8.95×10^{-1}	3.59×10^{-1}
2	rs11968228	56,429,841		T	C	36	0.00	0.04	0.99	0.26	2/81/1703	0.02	1/74/1738	0.02	4.08×10^{-1}	3.36×10^{-1}	8.31×10^{-1}
3	rs7749370	56,433,476	*DST*, intron	G	C	23			≤0.95	0.68	172/648/579	0.35	316/755/698	0.39	***1.99 x10^{-3}***	2.40×10^{-2}	2.48×10^{-2}
4	rs11758339	56,435,808	*DST*, missense /Ala⇒Thr)	T	C	695	0.00	0.01	0.97	0.34	98/604/1054	0.23	99/645/1043	0.24	3.86×10^{-1}	6.86×10^{-1}	3.89×10^{-1}
5	rs4256439	56,505,350	*DST*, intron	A	G	199	0.00	0.01	0.99	0.08	5/120/1662	0.04	2/121/1693	0.03	6.40×10^{-1}	9.15×10^{-1}	***4.89x10^{-2}***
6	rs4715630	56,525,241	*DST*, missense (Met⇒Ile)	C	T	3	0.00	0.79	0.99	0.80	106/650/1030	0.24	103/671/1030	0.24	8.18×10^{-1}	9.28×10^{-1}	8.17×10^{-1}
7	rs4715631	56,525,504	*DST*, missense (Thr⇒Ala)	T	C	30	0.00	**0.99**	0.99	0.34	141/693/953	0.27	128/728/953	0.27	9.76×10^{-1}	8.70×10^{-1}	6.29×10^{-1}
8	rs11756977	56,528,497	*DST*, missense (Arg⇒His)	T	C	37	0.00	0.44	0.99	0.28	139/687/963	0.27	127/722/960	0.27	9.56×10^{-1}	9.22×10^{-1}	6.72×10^{-1}
9	rs9349828	56,532,191	*DST*, intron	C	T	2	0.00	0.01	0.99	0.78	165/765/857	0.31	166/801/831	0.32	4.04×10^{-1}	5.75×10^{-1}	6.48×10^{-1}
10	rs4715634	56,532,380	*DST*, intron	G	A	109	0.00	0.00	0.99	1.00	6/209/1571	0.06	11/209/1598	0.06	7.52×10^{-1}	9.20×10^{-1}	6.38×10^{-1}
11	rs973089	56,543,316	*DST*, intron	C	T	281	0.01	0.04	0.99	0.36	11/225/1543	0.07	9/218/1582	0.07	4.92×10^{-1}	7.68×10^{-1}	1.28×10^{-1}
12	rs4712138	56,571,369	*DST*, intron	C	T	71	0.00	**0.93**	0.99	0.36	255/817/717	0.37	283/835/687	0.39	1.21×10^{-1}	1.27×10^{-1}	6.37×10^{-1}
13	rs13194995	56,578,510	*DST*, missense (Thr⇒Ser)	A	T	261	0.00	0.18	0.99	0.53	229/803/752	0.35	253/819/732	0.37	2.03×10^{-1}	1.54×10^{-1}	8.93×10^{-1}
14	rs10456737	56,604,608	*DST*, intron	C	A	105	0.01	0.07	0.99	0.43	14/324/1450	0.10	13/341/1458	0.10	6.82×10^{-1}	3.14×10^{-1}	4.31×10^{-1}
15	rs1024195	56,615,094	*DST*, intron	T	C	4	0.00	0.02	0.98	0.72	249/846/692	0.38	251/849/688	0.38	9.18×10^{-1}	8.84×10^{-1}	4.99×10^{-1}
16	rs1024196	56,615,448	*DST*, missense (Leu⇒Ser)	G	A	487	0.01	0.00	0.99	0.11	4/107/1675	0.03	1/116/1685	0.03	9.10×10^{-1}	5.37×10^{-1}	8.28×10^{-2}
17	rs6459169	56,664,098	*DST*, intron	G	A	31	0.00	0.01	0.99	1.00	1/87/1698	0.02	0/67/1742	0.02	6.07×10^{-2}	***1.01x10^{-2}***	8.31×10^{-1}
18	rs7767128	56,667,246	*DST*, intron	C	G	69	0.02	0.03	0.99	0.72	77/602/1109	0.21	83/616/1107	0.22	6.34×10^{-1}	5.82×10^{-1}	5.98×10^{-1}
19	rs9918490	56,674,108	*DST*, intron	A	G	452	0.00	0.06	0.99	0.89	18/319/1447	0.10	12/331/1468	0.10	8.40×10^{-1}	8.96×10^{-1}	7.58×10^{-1}
20	rs3002005	56,719,295	*DST*, intron	C	T	47	0.01	0.15	0.98	0.76	233/816/737	0.36	245/809/741	0.36	7.94×10^{-1}	6.87×10^{-1}	5.90×10^{-1}
21	rs12204927	56,724,042	*DST*, intron	C	T	503	0.00	0.39	0.99	0.32	73/612/1101	0.21	77/598/1129	0.21	7.23×10^{-1}	7.04×10^{-1}	8.29×10^{-1}
22	rs9382665	56,774,326	*DST*, intron	G	A	598	0.00	0.03	0.99	0.91	23/369/1395	0.12	24/393/1397	0.12	4.30×10^{-1}	1.50×10^{-1}	9.82×10^{-1}
23	rs16888184	56,834,104	*DST*	G	A	306	0.10	0.05	0.99	0.25	48/534/1201	0.18	72/537/1204	0.19	2.39×10^{-1}	6.17×10^{-1}	1.06×10^{-1}
24	rs6910217	56,864,654	*DST*	G	A	219	0.04	0.01	0.99	0.76	65/541/1179	0.19	68/564/1172	0.19	4.66×10^{-1}	2.06×10^{-1}	9.89×10^{-1}
25	rs2817577	56,886,538	*DST*	C	T	207	0.00	0.09	0.99	0.77	4/148/1627	0.04	6/147/1655	0.04	9.96×10^{-1}	9.18×10^{-1}	6.37×10^{-1}
26	rs2599697	56,907,221	*DST*	T	C	177	0.03	0.03	0.99	0.96	208/802/780	0.34	229/762/817	0.34	8.25×10^{-1}	9.04×10^{-1}	8.63×10^{-1}
27	rs9396240	56,924,938	*DST*	T	C	94	0.00	0.03	0.99	0.13	10/190/1589	0.06	5/151/1660	0.04	***5.39x10^{-3}***	***2.46 x10^{-3}***	5.10×10^{-1}
28	rs9382671	56,934,289	*BEND6*, intron	A	T	26	0.07	0.13	0.99	0.21	194/752/837	0.32	186/770/856	0.32	7.05×10^{-1}	5.87×10^{-1}	9.73×10^{-1}
29	rs6925107	56,936,899	*BEND6*, intron	A	G	144	0.41	0.02	0.98	0.72	74/593/1112	0.21	115/586/1087	0.23	***4.69x10^{-2}***	1.28×10^{-1}	2.56×10^{-1}
30	rs11751461	56,951,253	*BEND6*, intron	T	C	412	0.01	0.01	0.99	1.00	5/178/1604	0.05	3/168/1644	0.05	3.53×10^{-1}	5.34×10^{-1}	1.11×10^{-1}
31	rs2167460	56,992,410	*BEND6*, intron	A	G	331	0.48	0.60	0.99	0.94	65/548/1174	0.19	94/573/1144	0.21	***3.43x10^{-2}***	9.27×10^{-2}	2.24×10^{-1}
32	rs6926980	57,025,497	*KIAA1586*, missense (Val⇒Met)	A	G	157			≤0.95	≤0.01	110/528/1086	0.22	114/567/1053	0.23	2.39×10^{-1}	6.03×10^{-1}	1.45×10^{-1}
33	rs10949010	57,041,161		C	G	325	0.21	0.21	0.99	0.92	332/872/579	0.43	387/848/566	0.45	9.27×10^{-2}	1.07×10^{-1}	2.90×10^{-1}
34	rs3800023	57,073,655	*ZNF451*	C	G	190	**0.99**	0.02	0.98	0.17	22/409/1339	0.13	48/460/1281	0.16	***1.04x10^{-3}***	***7.12x10^{-3}***	***2.85x10^{-2}***
35	rs11965193	57,092,657	*ZNF451*	T	C	120	0.02	0.02	0.99	0.38	31/377/1379	0.12	33/397/1380	0.13	4.70×10^{-1}	3.35×10^{-1}	7.17×10^{-1}
	rs10484410	57,104,647	*ZNF451*	G	A	573	-	0.00									
36	rs182662	57,161,967	*RAB23*, 3'UTR	T	C	8	0.00	0.01	0.99	0.26	0/114/1667	0.03	0/90/1719	0.02	7.73×10^{-2}	1.90×10^{-1}	2.57×10^{-1}

37	rs1411578	57,162,802	*RAB23*, 3'UTR	G	C	5	**0.89**	0.02	0.99	0.14	28/450/1312	0.14	63/493/1256	0.17	***6.64x10^{-4}***	***1.27x10^{-2}***	***8.48x10^{-3}***
38	rs1040461	57,163,313	*RAB23*, missense (Gly⇒Ser)	T	C	18	0.02	0.02	0.99	0.14	23/307/1454	0.10	8/276/1515	0.08	***7.83x10^{-3}***	***8.37x10^{-3}***	4.43 x10^{-1}
39	rs1547226	57,165,142	*RAB23*, intron	T	C	68	0.76	0.35	0.99	0.29	39/492/1256	0.16	77/532/1198	0.19	***8.30x10^{-4}***	***2.75x10^{-2}***	***8.60x10^{-3}***
40	rs12195875	57,171,905	*RAB23*, intron	G	A	15	0.46	0.30	0.97	0.65	41/475/1248	0.16	51/515/1183	0.18	***3.56x10^{-2}***	2.60 x10^{-1}	1.11 x10^{-1}
41	rs12211611	57,173,402	*RAB23*, intron	C	G	52			0.96	≤0.01	109/475/1160	0.20	109/513/1110	0.21	2.20 x10^{-1}	3.82 x10^{-1}	5.04 x10^{-1}
42	rs9396263	57,178,643	*RAB23*, intron	A	G	205	0.01	0.01	0.99	0.68	5/206/1586	0.06	2/172/1618	0.05	**3.82x10^{-2}**	8.29 x10^{-2}	2.77 x10^{-1}
43	rs704480	57,199,125		G	A	315	0.02	0.09	0.99	0.09	49/435/1306	0.15	38/458/1305	0.15	9.88 x10^{-1}	4.39 x10^{-1}	1.54 x10^{-1}
44	rs9349850	57,230,586		A	G	-	0.08		0.98	0.83	207/800/754	0.35	226/783/778	0.35	9.75 x10^{-1}	7.60 x10^{-1}	4.65 x10^{-1}

Table S3: Results of additional fine mapping.

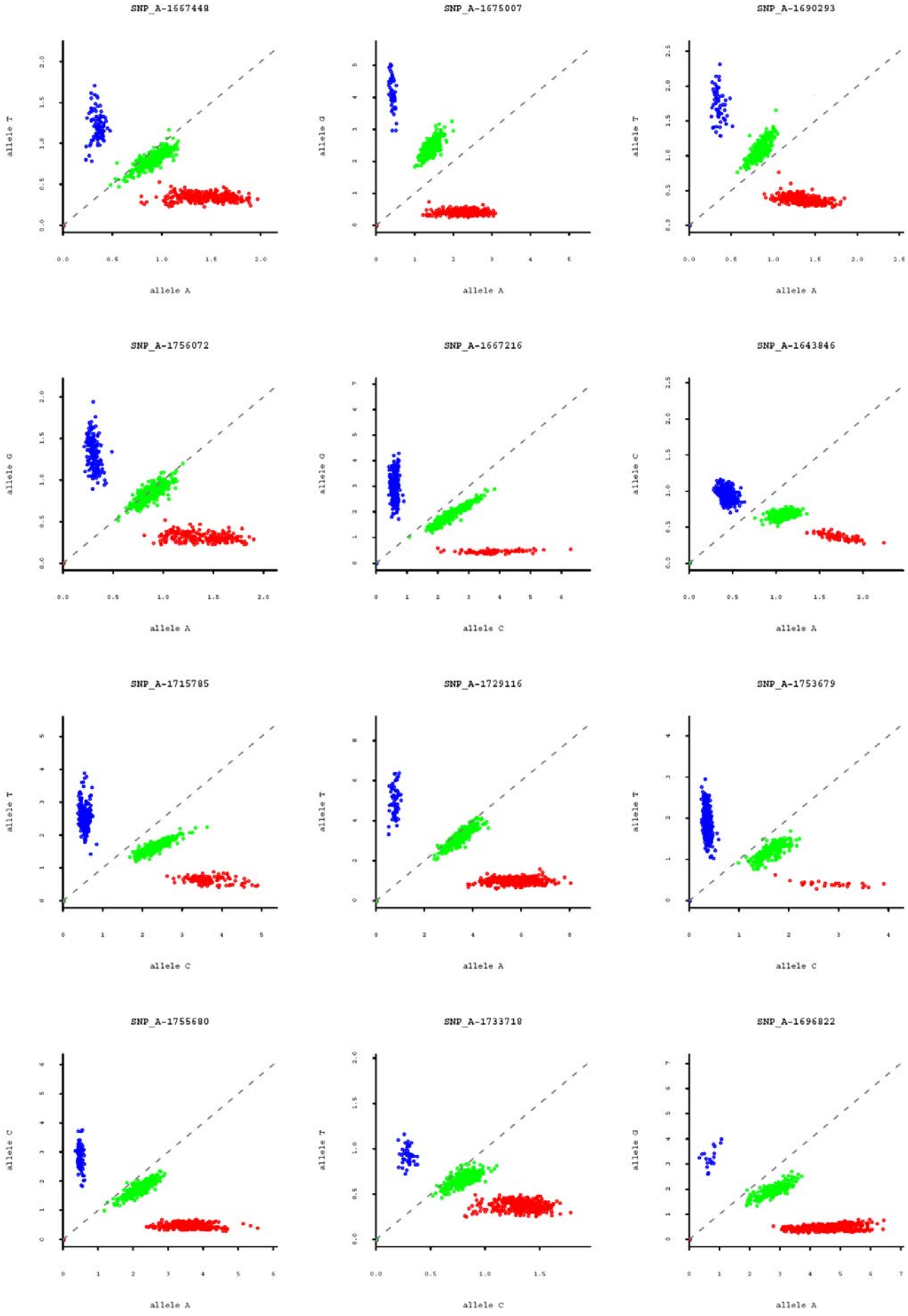
46 tagging SNPs were genotyped in panel C in addition to the 44 SNPs that are listed in Table S2 across the 6p12.1 region. For a description of column headers see Tables S1 and S2 and for a plot of association and LD of SNP that passed all quality criteria (incl. SNPs listed in Table S3) see Figure 1.

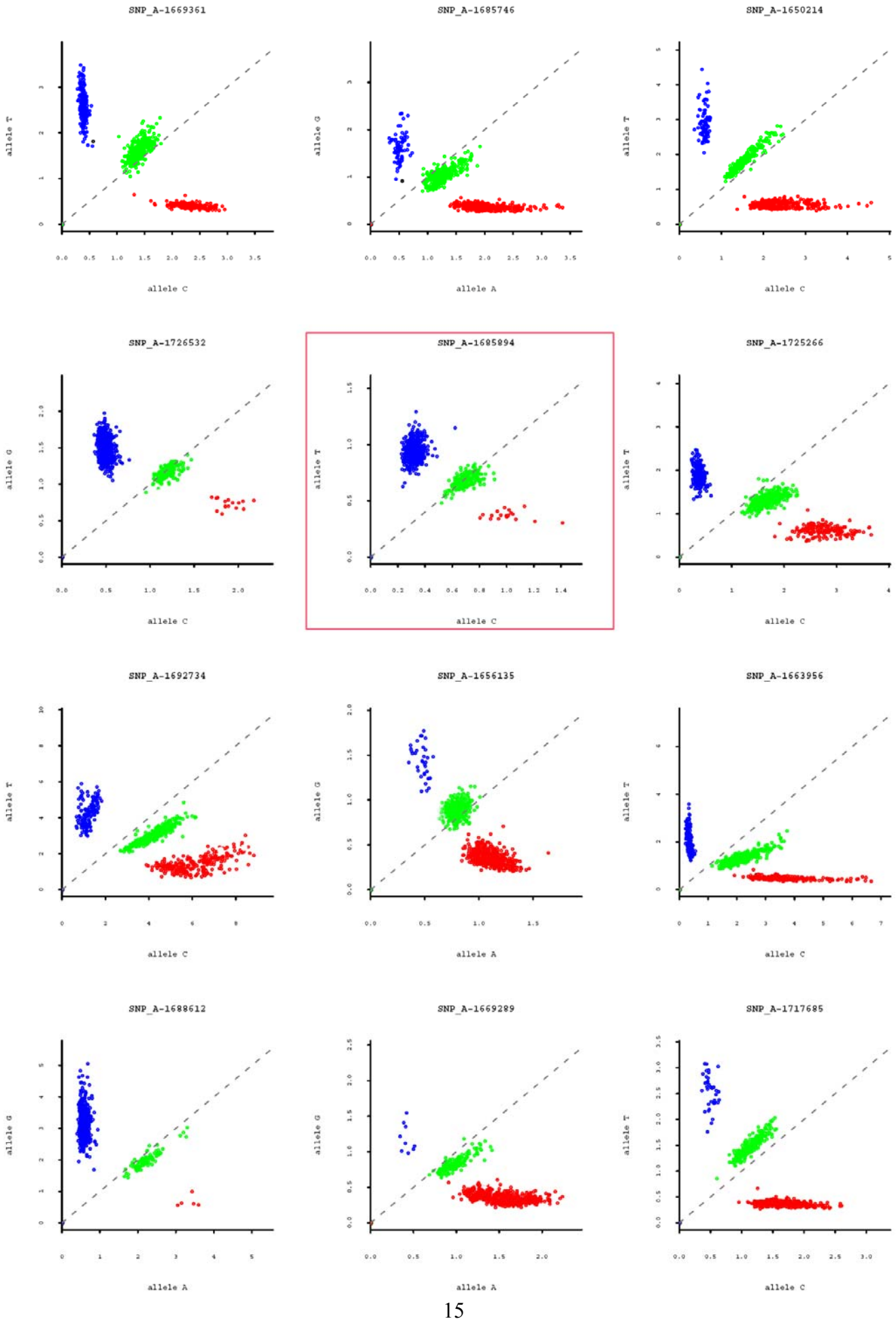
#	dbSNP ID	Position	Locus	A1	A2	Distance [kb]	CR	U P_{HWE}	U	F_U_A1	SA all	F_SAall_A1	SA all P value	SA chronic (n=1,065) P value	SA acute (n=603) P value
1	rs13203600	56,853,121	*DST*	A	G	344	≤0.95	0.24	12/194/519	0.15	65/513/1120	0.19	***1.50x10^{-3}***	***1.66 x10^{-2}***	***6.09x10^{-3}***
2	rs9464407	56,887,529	*DST*	G	A	99	0.99	0.30	94/670/1029	0.24	137/643/1028	0.25	1.75x10^{-1}	4.60 x10^{-1}	2.51x10^{-1}
3	rs4490669	56,897,424	*DST*	G	A	20	0.97	0.15	48/547/1201	0.18	88/537/1091	0.21	***2.65x10^{-3}***	***1.57x10^{-2}***	1.41x10^{-1}
4	rs17752435	56,899,454	*DST*	T	C	27	0.99	0.14	44/532/1219	0.17	83/554/1172	0.20	***4.73x10^{-3}***	***4.21x10^{-2}***	***8.35x10^{-2}***
5	rs13209289	56,902,113	*DST*	A	G	46	0.99	1.00	5/179/1606	0.05	3/163/1646	0.05	2.21x10^{-1}	4.68 x10^{-1}	6.63x10^{-2}
6	rs17758527	56,906,750	*DST*	A	T	4	0.99	0.12	10/188/1592	0.06	5/152/1659	0.04	***8.80x10^{-3}***	***6.73x10^{-3}***	5.11x10^{-1}
7	rs17685218	56,907,114	*DST*	C	G	16	0.99	0.55	92/651/1052	0.23	129/626/1046	0.25	2.21x10^{-1}	6.20 x10^{-1}	2.61x10^{-1}
8	rs13205191	56,908,751	*DST*	G	A	11	0.98	0.18	9/184/1596	0.06	5/152/1636	0.05	***2.81x10^{-2}***	***1.93x10^{-2}***	7.05x10^{-1}
9	rs1451707	56,909,836	*DST*	G	A	54	0.99	0.21	45/526/1219	0.17	82/546/1166	0.20	***5.61x10^{-3}***	5.34 x10^{-2}	8.33x10^{-2}
10	rs2122958	56,915,205	*DST*	G	C	33	0.99	0.67	77/607/1114	0.21	117/612/1088	0.23	***3.44x10^{-2}***	1.05 x10^{-1}	2.29x10^{-1}
11	rs1020548	56,918,498	*DST*	G	A	315	0.99	0.09	29/466/1298	0.15	67/499/1237	0.18	***7.97x10^{-4}***	***2.73x10^{-3}***	1.84x10^{-1}
12	rs9475762	56,950,011	*C6orf65*	A	G	75	0.96	0.17	9/182/1591	0.06	5/148/1573	0.05	***4.65x10^{-2}***	***2.75E-02***	7.27x10^{-1}
13	rs17685332	56,957,474	*BEND6*, intron	A	G	133	0.99	0.37	23/327/1440	0.10	22/359/1426	0.11	2.96 x10^{-1}	3.06 x10^{-1}	2.88x10^{-1}
14	rs6901944	56,970,764	*BEND6*, intron	G	A	292	0.99	0.11	30/468/1298	0.15	72/499/1234	0.18	***4.04x10^{-4}***	***1.17x10^{-3}***	1.62x10^{-1}
15	rs1044670	56,999,938	*BEND6*, 3'UTR	A	G	61	0.99	0.46	26/413/1355	0.13	56/476/1274	0.16	***7.93x10^{-5}***	***2.46x10^{-4}***	8.69x10^{-2}
16	rs1993118	57,005,993		C	T	72	≤0.95	0.07	26/307/1404	0.10	14/268/1417	0.09	2.03 x10^{-2}	***1.08x10^{-2}***	8.07x10^{-1}
17	rs3957366	57,013,233		A	G	56	0.98	0.42	28/426/1322	0.14	56/485/1262	0.17	***4.82x10^{-4}***	***3.46x10^{-3}***	***4.84x10^{-2}***
18	rs1157713	57,018,843	*KIAA1586*	G	A	99	≤0.95	0.11	44/236/439	0.23	109/597/987	0.24	2.92 x10^{-1}	4.51x10^{-1}	2.60x10^{-1}
19	rs12190575	57,028,763		G	A	336	0.98	0.20	22/405/1360	0.13	46/457/1286	0.15	***7.92x10^{-4}***	***7.20x10^{-3}***	***3.69x10^{-2}***
20	rs7753537	57,062,398	*ZNF451*	C	T	1	0.98	≤0.01	110/583/1090	0.23	112/632/1052	0.24	2.06 x10^{-1}	4.81x10^{-1}	1.57x10^{-1}
21	rs6459178	57,062,477	*ZNF451*	C	A	61	0.99	0.28	23/403/1353	0.13	50/465/1299	0.16	***3.74x10^{-4}***	***5.76x10^{-3}***	***1.43x10^{-2}***
22	rs4236146	57,068,571	*ZNF451, intron*	T	G	133	≤0.95	≤0.01	60/389/1298	0.15	73/389/1233	0.16	1.43 x10^{-1}	2.00x10^{-1}	1.66x10^{-1}
23	rs9464418	57,081,861	*ZNF451, intron*	C	T	201	≤0.95	0.06	12/128/667	0.09	6/167/897	0.08	2.40 x10^{-1}	2.14x10^{-1}	7.85x10^{-1}
24	rs9370564	57,101,983	*ZNF451*, intron	C	Tt	15	0.99	≤0.01	110/583/1094	0.23	115/640/1065	0.24	1.65 x10^{-1}	5.03x10^{-1}	1.00x10^{-1}
25	rs9396261	57,103,453	*ZNF451*, intron	T	A	10	0.98	0.15	20/285/1456	0.09	8/266/1517	0.08	***3.83x10^{-2}***	***2.82x10^{-2}***	7.18x10^{-1}
26	rs17619360	57,104,425	*ZNF451*, intron	A	C	2	0.99	0.28	23/406/1364	0.13	51/461/1294	0.16	***3.22x10^{-4}***	***2.70x10^{-3}***	***2.39x10^{-2}***
	rs10484410	57,104,647	*ZNF451*, intron	G	A	49	0.99	0.28	23/404/1357	0.13	52/460/1288	0.16	***2.41x10^{-4}***	***3.00x10^{-3}***	***1.39x10^{-2}***
27	rs9367713	57,109,534	*ZNF451*, intron	A	T	114	0.99	0.15	23/309/1463	0.10	9/280/1519	0.08	***1.37x10^{-2}***	***1.66x10^{-2}***	4.78x10^{-1}
28	rs3734738	57,120,889	*ZNF451*, synon	A	G	100	0.99	≤0.01	110/588/1090	0.23	115/630/1052	0.24	1.99 x10^{-1}	4.73x10^{-1}	1.63x10^{-1}
29	rs1556245	57,130,917	*ZNF451*, intron	A	G	111	0.99	0.23	22/307/1463	0.10	9/279/1515	0.08	***1.95x10^{-2}***	***2.13E-02***	5.49x10^{-1}
30	rs7756421	57,141,968	*ZNF451*, 3'UTR	G	A	170	0.98	0.28	22/398/1357	0.12	53/463/1283	0.16	***4.88x10^{-5}***	***1.38E-03***	***3.74x10^{-3}***

31	rs9296862	57,158,976		G	A	5	≤0.95	≤0.01	0/300/1465	0.09	0/254/1380	0.08	2.62×10^{-1}	6.70E-02	4.47×10^{-1}
32	rs6928715	57,159,524		C	T	24	0.99	≤0.01	113/592/1090	0.23	114/639/1051	0.24	2.32×10^{-1}	5.06×10^{-1}	1.88×10^{-1}
33	rs11398	57,161,906	*RAB23*, 3'UTR	G	A	150	0.99	0.17	28/447/1318	0.14	66/486/1251	0.17	***3.22×10^{-4}***	***4.56E-03***	***9.58×10^{-3}***
34	rs3800019	57,176,899	*RAB23*, intron	T	C	7	0.99	0.35	69/534/1188	0.19	86/578/1152	0.21	***4.88×10^{-2}***	1.48×10^{-1}	1.57×10^{-1}
35	rs3800018	57,177,565	*RAB23*, intron	G	A	53	0.99	0.34	24/404/1348	0.13	50/476/1281	0.16	***1.23E-04***	***2.19E-03***	***9.47×10^{-3}***
36	rs9357941	57,182,870	*RAB23*, intron	A	G	16	0.98	0.66	4/193/1561	0.06	3/166/1638	0.05	6.61×10^{-2}	1.10×10^{-1}	4.19×10^{-1}
37	rs16888392	57,184,511	*RAB23*, intron	A	G	132	≤0.95	0.86	9/154/556	0.12	31/261/776	0.15	***8.82×10^{-3}***	***1.45×10^{-2}***	2.02×10^{-1}
38	rs12173390	57,197,667		A	T	2	≤0.95	0.12	1/168/1526	0.05	0/139/1586	0.04	***4.73×10^{-2}***	***2.71×10^{-2}***	7.62×10^{-1}
49	rs9475792	57,197,901		G	A	314	0.99	1.00	0/73/1721	0.02	0/71/1741	0.02	8.09×10^{-1}	8.21×10^{-1}	9.62×10^{-1}
40	rs16888420	57,229,318		T	C	35	≤0.95	0.19	3/61/663	0.05	5/147/1559	0.05	9.45×10^{-1}	9.75×10^{-1}	5.98×10^{-1}
41	rs1013147	57,232,840		G	A	173	0.99	0.53	207/823/764	0.35	230/792/776	0.35	7.95×10^{-1}	9.54×10^{-1}	3.34×10^{-1}
42	rs9464431	57,250,125		A	G	185	0.99	0.34	26/345/1419	0.11	27/357/1419	0.11	6.49×10^{-1}	4.41×10^{-1}	6.32×10^{-1}
43	rs11758642	57,268,662		A	G	191	0.99	0.11	14/232/1548	0.07	13/243/1538	0.07	6.38×10^{-1}	4.65×10^{-1}	4.04×10^{-1}
44	rs9475837	57,287,806		C	T	23	0.99	1.00	29/399/1363	0.13	33/376/1395	0.12	5.38×10^{-1}	8.19×10^{-1}	1.89×10^{-1}
45	rs1888265	57,290,150	*PRIM2*	G	A	19	0.99	0.53	66/535/1197	0.19	62/533/1212	0.18	6.97×10^{-1}	9.31×10^{-1}	1.89×10^{-1}
46	rs1013893	57,292,072	*PRIM2*, intron	C	T	-	0.96	0.83	184/786/818	0.32	202/768/742	0.34	***8.49×10^{-2}***	2.96×10^{-1}	1.44×10^{-1}

Figure S1: Signal intensity/cluster plots

Scatter plots of normalized summary probe intensities for all 25 SNPs that were selected for replication genotyping, with each point representing one individual. Each point is colored according to the genotype assignment of the calling algorithm (blue or red=corresponding homozygote for one of the two alleles; green=heterozygote; black circle="null" or missing call). The replicated variant rs10484410 (SNP_A-1685894) is highlighted by a red box. The aim of examining a cluster plot is twofold: to determine whether the given SNP has been genotyped well — reflected as clear, distinct, clusters on the plot that would correspond to the three genotypes — and also to determine whether the calling algorithm has called the clusters correctly. If both of these are true, one can usually be confident that the genotype counts are accurate. If these are not true, any associations that we observe at such SNPs may well be caused by the resulting incorrect genotype counts.


SNP_A-1667448
allele T
allele A
SNP_A-1675007
allele G
allele A
SNP_A-1690293
allele T
allele A
SNP_A-1756072
allele G
allele A
SNP_A-1667216
allele G
allele C
SNP_A-1643846
allele C
allele A
SNP_A-1715785
allele T
allele C
SNP_A-1729116
allele T
allele A
SNP_A-1753679
allele T
allele C
SNP_A-1755680
allele C
allele A
SNP_A-1733718
allele T
allele C
SNP_A-1696822
allele G
allele A


SNP_A-1669361
allele T
allele C
SNP_A-1685746
allele G
allele A
SNP_A-1650214
allele T
allele C
SNP_A-1726532
allele G
allele C
SNP_A-1695094
allele T
allele C
SNP_A-1725266
allele T
allele C
SNP_A-1692734
allele T
allele C
SNP_A-1656135
allele G
allele A
SNP_A-1663956
allele T
allele C
SNP_A-1688612
allele G
allele A
SNP_A-1669289
allele G
allele A
SNP_A-1717685
allele T
allele C

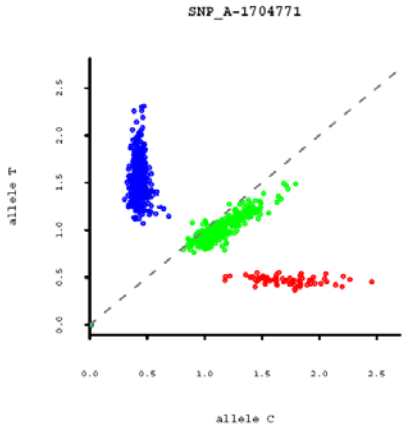

SNP_A-1704771
allele T
allele C
0.0
0.5
1.0
1.5
2.0
2.5

Figure S2: Workflow diagram summarizing the different stages of the experiment including specification of quality control parameters and selection criteria

Minor allele frequency in controls (MAF_{co}), p value from the test of deviation from Hardy-Weinberg equilibrium in controls (p_{HWE}), fine-mapping (FM).

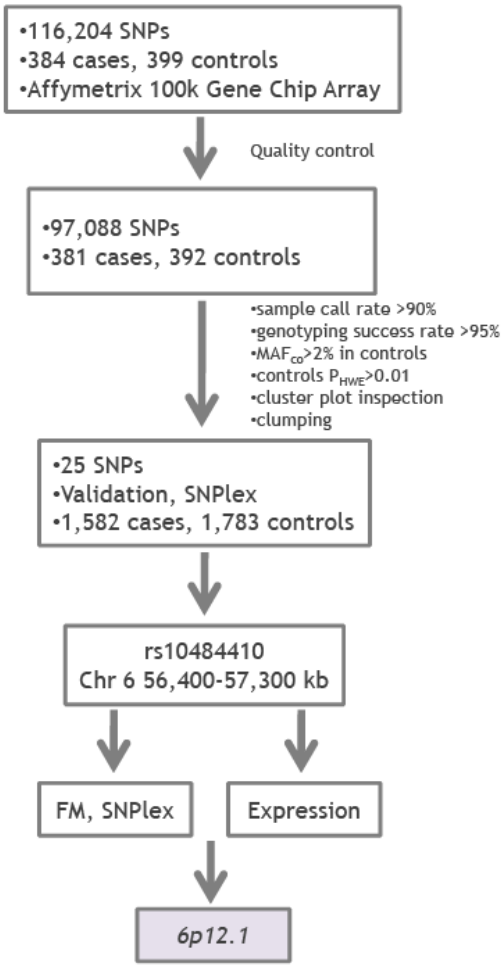


Figure S3: Power calculation

Power to detect a given allelic disease association (carriership of the rarer SNP allele) in the screening panel (381 cases/ 392 controls). The power is expressed as a function of the odds ratio. The calculations were performed for different allele frequencies at a significance level p=0.05 using PS Power and Sample Size v2.130[15].

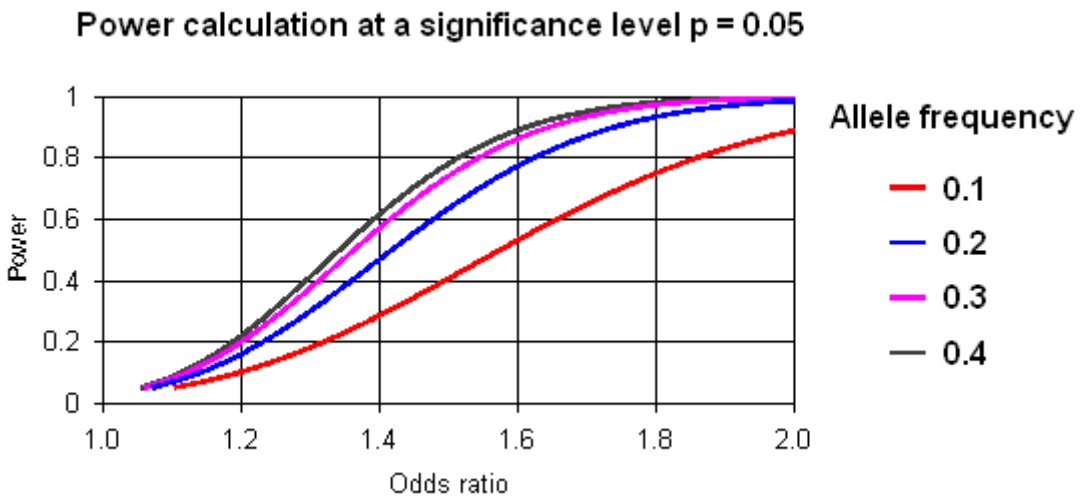


Figure S4: Genome-wide scan of sarcoidosis (panel A).

Genome-wide associations of sarcoidosis graphed by chromosome position and –log p value. The validated region at chromosome 6p12.1 is marked by an arrow.

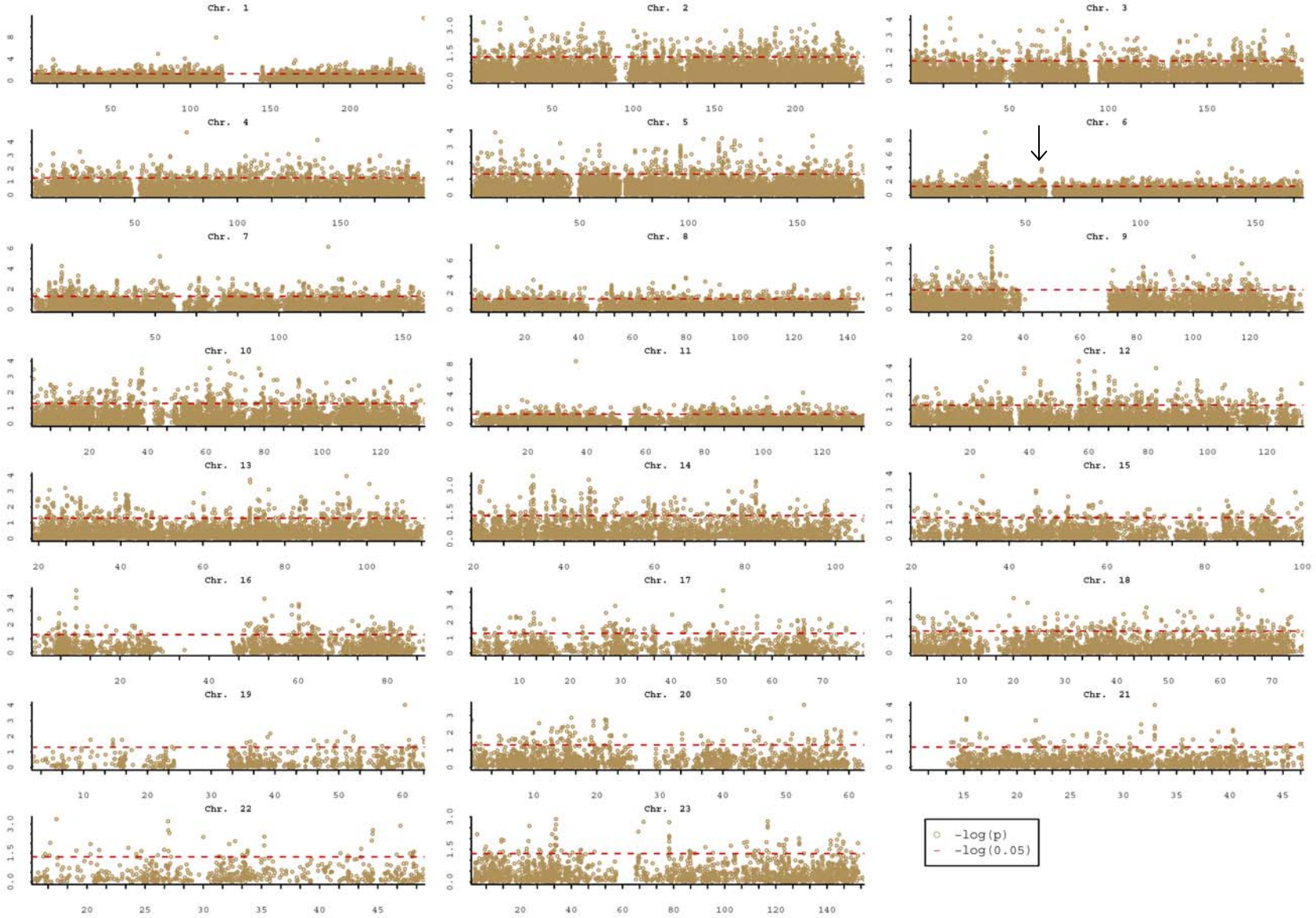


Figure S5A-7C: Results of the eQTL analysis for the three most significant associated SNPs: rs7756421 (Figures S5A-C), rs1044670 (Figures S6A-C) and the significantly associated non-synonymous SNP rs1040461 (Figure S7A-C). Results are presented as Hierarchical clusters (panel A), principle component analysis (panel B) and gene ontology analysis (panel C).

Hierarchical clustering (panel A) shows the transcript levels (arranged in rows) for each sample (arranged in columns) which are colored according to their expression intensity: red (high expression), green (low expression). To better visualize the expression differences within one transcript, colors were based on normalized expression intensity (z-score). The row dendrogram shows the similarities of the expression profile for each transcript while the column dendogram shows the similarities between the samples. Columns are labeled according to the genotype.

The principle component analysis (panel B) shows the two strongest components (PCA1 on the X-axis, PCA2 on the Y-axis), explaining the variation within the set of samples. Samples are colored according to their genotype, based on the significantly expressed genes.

The gene ontology analysis (panel C) lists biological processes (retrieved from www.geneontology.org) which were significantly associated to downregulated transcripts (green) and/or upregulated transcripts (red). The bar length displays the significance of enrichment or depletion of a specific process (-log(p)).

Figure S5A

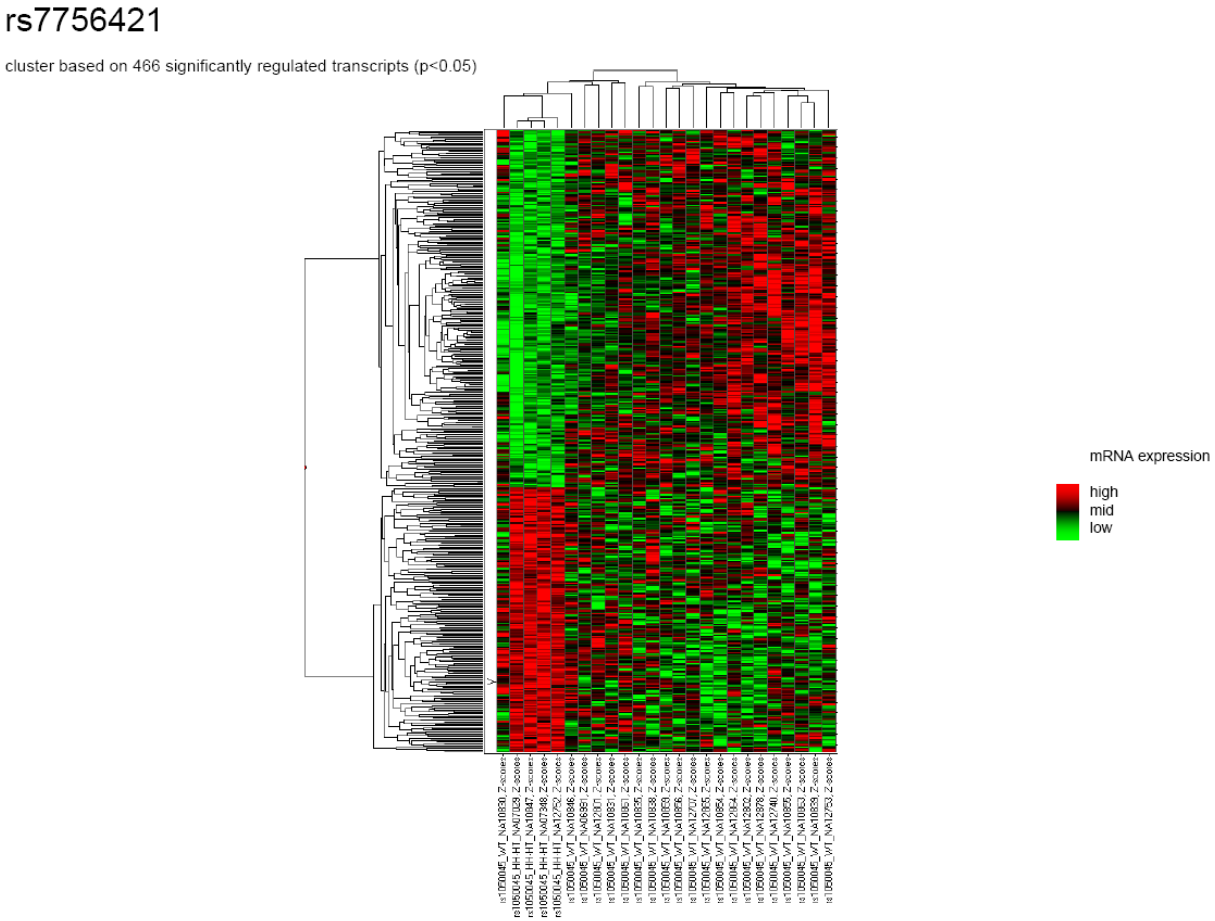


Figure S5B

rs7756421

PCA based on 466 significantly regulated transcripts (p<0.05)

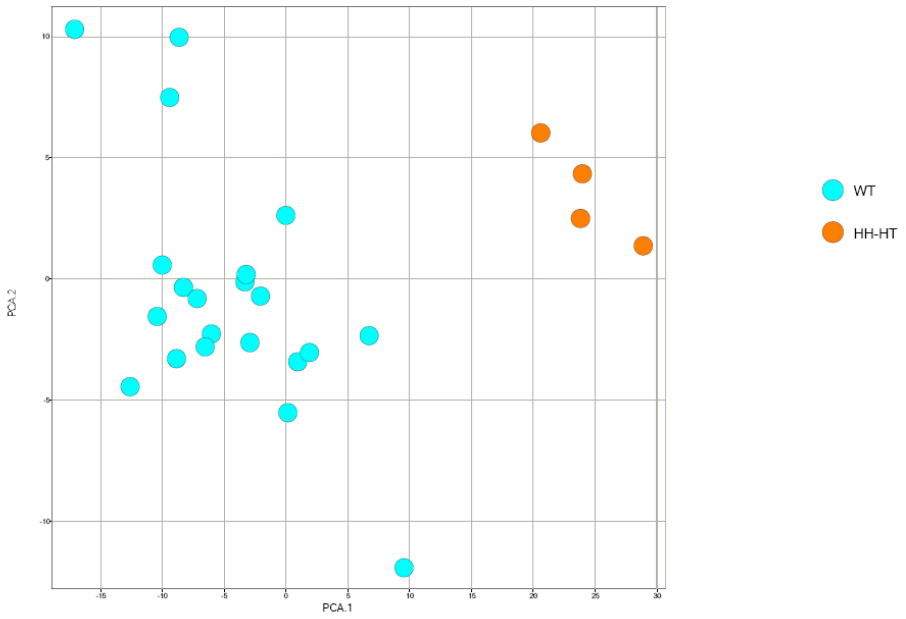


Figure S5C

rs7756421

Biological processes associated to 466 significantly regulated transcripts (p<0.05)

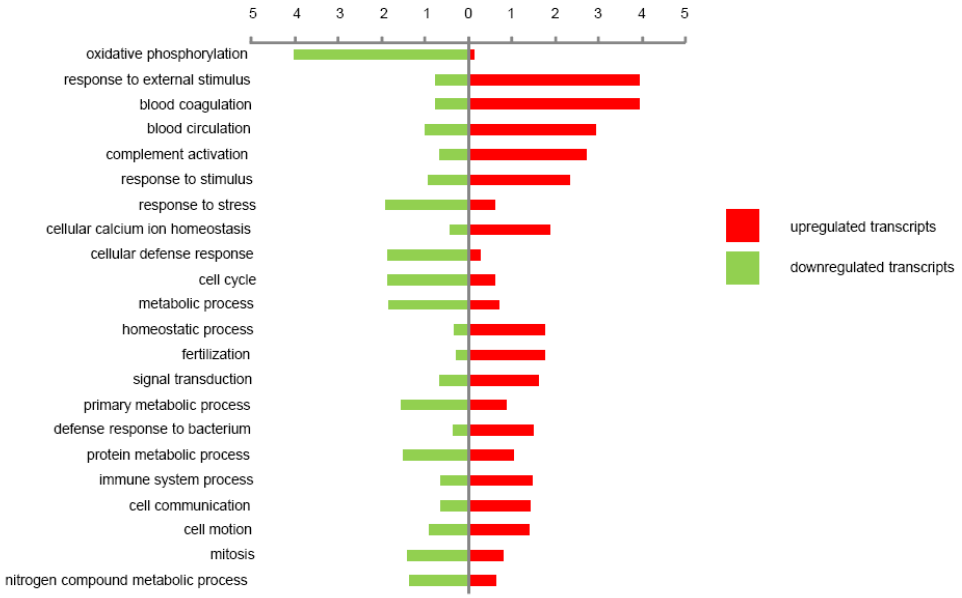


Figure S6A

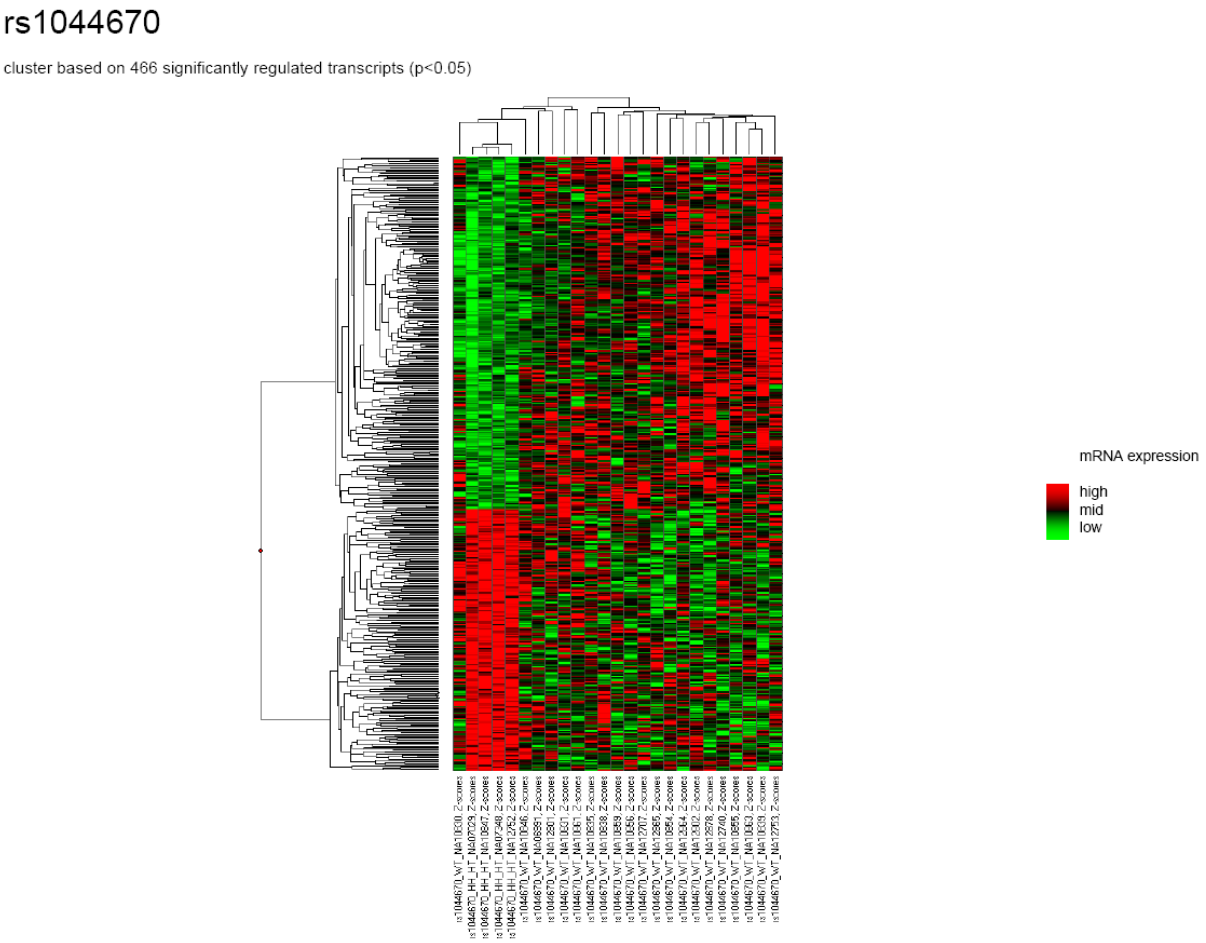


Figure S6B

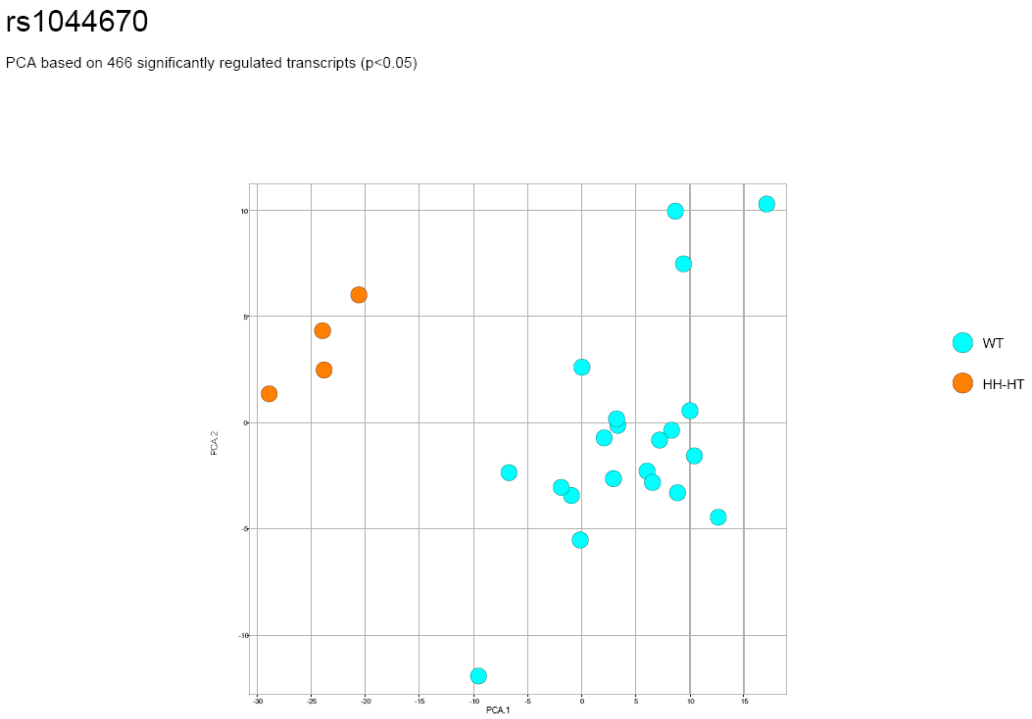


Figure S6C

rs1044670

Biological processes associated to 466 significantly regulated transcripts (p<0.05)

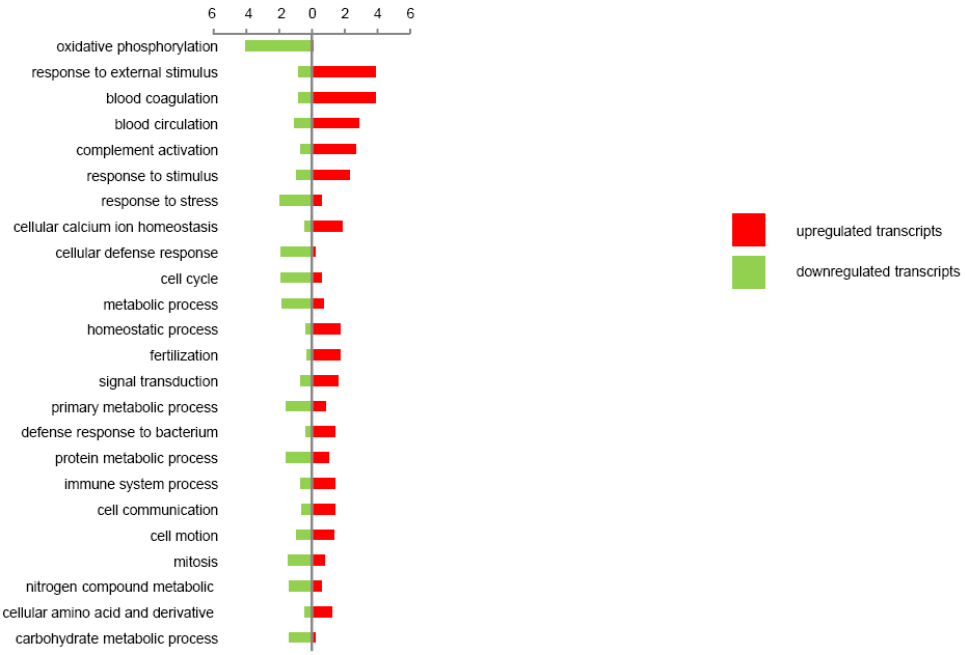


Figure S7A

rs1040461

cluster based on 466 significantly regulated transcripts (p<0.05)

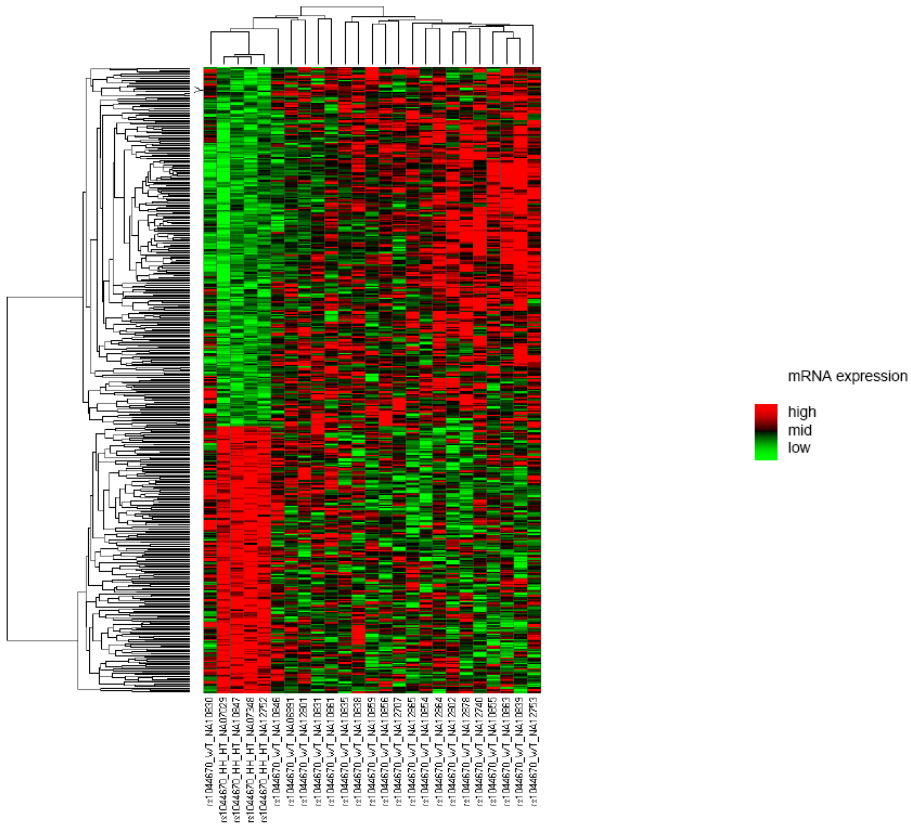


Figure S7B

rs1040461

PCA based on 466 significantly regulated transcripts (p<0.05)

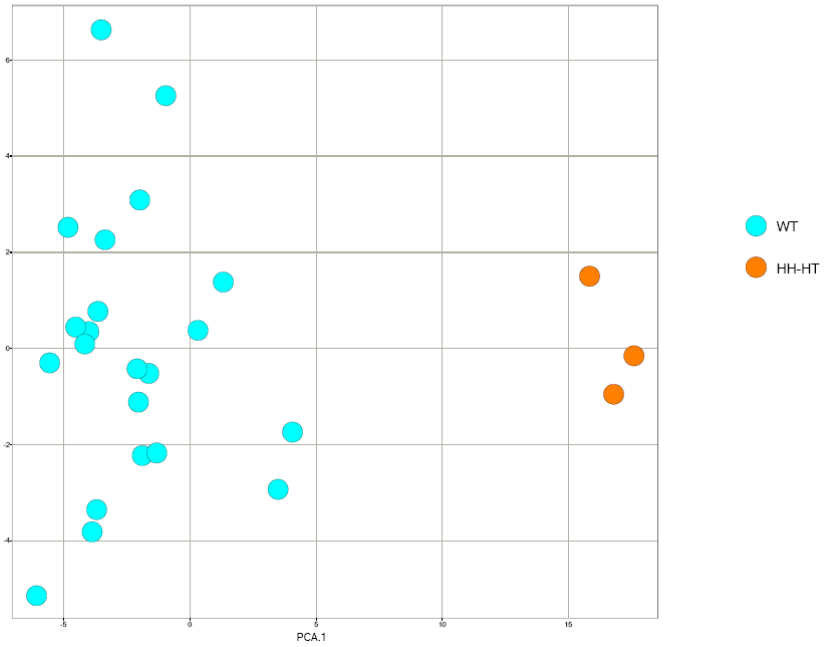
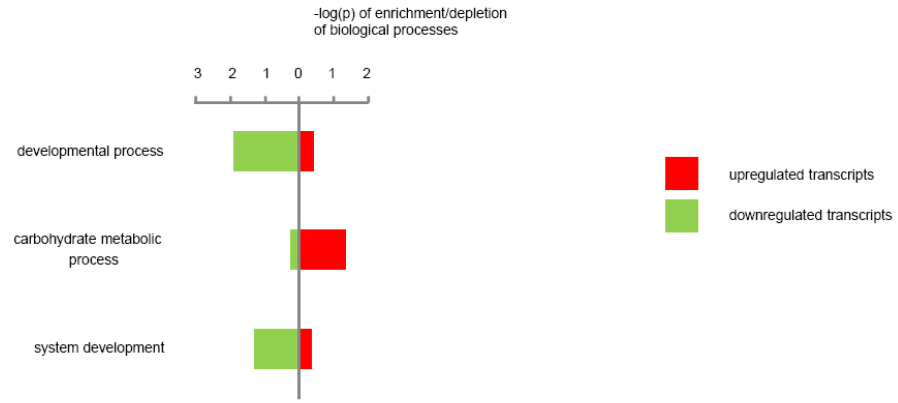


Figure S7C

rs1040461

Biological processes associated to 466 significantly regulated transcripts (p<0.05)



Web Resources

The URLs for data presented and software used herein are as follows:

(1) International HapMap project, http://www.hapmap.org/

(2) Haploview, http://www.broad.mit.edu/mpg/haploview/

(3) PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/

(4) R, http://www.R-project.org/

(5) GOLD, http://www.sph.umich.edu/csg/abecasis/gold/

(6) PS, http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize

(7) GENEVAR consortium, http://www.sanger.ac.uk/humgen/genevar/

(8) Gene Ontology, www.geneontology.org

Supplementary References

1. 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.

2. 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 IBD Risk Loci with Sarcoidosis and its Acute and Chronic Subphenotypes. *Eur Respir J* 2010.

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. *Am J Respir Crit Care Med* 1999: 160(2): 736-755.

4. Nebel A, Croucher PJP, Stiegeler R, Nikolaus S, Krawczak M, Schreiber S. No association between microsomal triglyceride transfer protein (MTP) haplotype and longevity in humans. *Proceedings of the National Academy of Sciences of the United States of America* 2005: 102(22): 7906-7909.

5. 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.

6. Matsuzaki H, Dong S, Loi H, Di X, Liu G, Hubbell E, Law J, Berntsen T, Chadha M, Hui H, Yang G, Kennedy GC, Webster TA, Cawley S, Walsh PS, Jones KW, Fodor SP, Mei R. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat Methods* 2004: 1(2): 109-111.

7. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999: 55(4): 997-1004.

8. 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.

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

10. 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.

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. R-DevelopmentCoreTeam. R: A language and environment for statistical computing. R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria, 2008.

13. 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.

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

15. Dupont WD, Plummer WD. PS power and sample size program available for free on the Internet. *Controlled Clin Trials* 1997: 18(274).

16. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006: 38(8): 904-909.

17. Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, Balascakova M, Bertranpetit J, Bindoff LA, Comas D, Holmlund G, Kouvatsi A, Macek M, Mollet I, Parson W, Palo J, Ploski R, Sajantila A, Tagliabracci A, Gether U, Werge T, Rivadeneira F, Hofman A, Uitterlinden AG, Gieger C, Wichmann HE, Ruther A, Schreiber S, Becker C, Nurnberg P, Nelson MR, Krawczak M, Kayser M. Correlation between genetic and geographic structure in Europe. *Curr Biol* 2008: 18(16): 1241-1248.

18. Consensus conference: activity of sarcoidosis. Third WASOG meeting, Los Angeles, USA, September 8-11, 1993. *European Respiratory Journal* 1994: 7(3): 624-627.

19. 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.

20. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 2001: 125(1-2): 279-284.

21. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. Systematic determination of genetic network architecture. *Nat Genet* 1999: 22(3): 281-285.