Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in Han Chinese

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-004466
Article Type:	Research
Date Submitted by the Author:	14-Nov-2013
Complete List of Authors:	Zhang, LG Xiao, ZJ Shi, GC J, Huang Zhang, CX
Primary Subject Heading:	Infectious diseases
Secondary Subject Heading:	Genetics and genomics
Keywords:	GENETICS, IMMUNOLOGY, INFECTIOUS DISEASES

Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in Han Chinese

LG Zhang, ZJ Xiao *, GC Shi, J Huang, CX Zhang

Internal Medicine - Tuberculosist, First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China

* Corresponding author: Zhijian Xiao, Internal Medicine - Tuberculosisthe, First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China

No.88 Jiankang Road, Weihui City, Henan, 463000, PR China

E-mail: zhangligongys@163.com

Tel: +86 0373 - 4402229 (office)

Fax: +86 0373 - 4402251

Key words: Sarcoidosis; Annexin A11; Single-nucleotide polymorphism

Word count: 2324

Abstract

Objectives: To further identify SNPs that contribute to the genetic susceptibility to sarcoidosis, we examined the potential association between sarcoidosis and fifteen SNPs of the ANXA11 gene.

Design: Case - control study.

Setting: A university hospital tuberculosist units in China.

Participants: Participants included 412 patients with sarcoidosis and 418 healthy controls.

Methods: The selected SNPs were genotyped in cases and controls by using the MALDI - TOF in the MassARRAY system (Sequenom Inc., San Diego, CA, USA). Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA).

Results: The results showed that statistically significant differences were observed in the allelic or genotypic frequencies of the rs2789679, rs1049550 and rs2819941 in the ANXA11 gene. The frequency of the T allele in rs2789679 (χ^2 = 12.849, P = 0.0003, OR = 0.703, 95%CI = 0.579 - 0.852), T allele in rs1049550 (χ^2 = 24.420, P = 0.0007, OR = 0.617, 95%CI= 0.503 - 0.758) and the rs2819941 C allele frequency (χ^2 = 11.803, P = 0.001, OR = 0.713, 95%CI = 0.588 - 0.865) in the patients with sarcoidosis was significantly lower than that in the controls. Furthermore, strong linkage disequilibrium (LD) was observed in two blocks (D' > 0.9). In block 2, the T - C haplotype occurred significantly less frequently (P = 0.001), whereas the C - C haplotypes occurred more frequently (P = 0.0001). Significantly more G - G - C

haplotypes (block 4) (P = 0.027) were found in the patients with sarcoidosis, but did not pass the threshold value (P = 0.025).

Conclusions: These findings point to a role for ANXA11 gene polymorphisms in sarcoidosis in the Han Chinese, and may be informative for future genetic or biological studies on sarcoidosis.

Article summary

Article focus

1. The first genome - wide association study (GWAS) in sarcoidosis conducted in a German population has revealed an association: the rs1049550 (C > T, Arg230Cys) located in exon 10 was associated with sarcoidosis.

2. More studies should be performed to demonstrate the following item: whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations.

3. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis in Han Chinese.

Key messages

1. ANXA11 rs2789679 (3'UTR) was found to be associated with sarcoidosis.

2. ANXA11 rs1049550 (exon 10) was significantly associated with sarcoidosis.

3. Another significant association was observed for rs2819941 (intron 14).

Strengths and limitations of this study

1. Functional SNPs in the promoter region, 5'- and 3'-UTR, exons of ANXA11 gene were systematically screened and homogeneity of the study subjects, representing the Han Chinese, is the main strength of the current study.

2. Functional characterisation is the potential limitation of the study that could have further helped in proving the positive association observed for rs2789679 and rs1049550. The lack of correlation of serum ANXA11 levels and sample size are other limitations.

INTRODUCTION

Sarcoidosis is a systemic autoimmune disease characterized by destructive, noncaseating epithelioid granulomatous lesions, with accumulated phagocytes and activated $CD4^+$ T helper type 1 lymphocytes [1 2]. The typical manifestations of sarcoidosis include bilateral hilar lymphadenopathy, pulmonary infiltration and ocular and skin lesions. The clinical course and prognosis of sarcoidosis is variable. The acute forms of sarcoidosis such as L□fgren's syndrome (LS) can resolve spontaneously in most cases. However, chronic pulmonary sarcoidosis may progress to lung fibrosis and eventually result in respiratory failure. Recent studies demonstrated that sarcoidosis occurs in a pattern of family clustering, and its racial incidence is also different, suggesting that some genetic factors may contribute to the risk and severity of sacroidosis [1 3].

Annexin gene family is involved in the etiology of several autoimmune and chronic diseases [4 5]. One member of Annexin gene family, ANXA11, located on chromosome 10q22.3, is involved in calcium signaling, apoptosis, vesicle trafficking, cell growth and terminal phase of cell division [5-7]. In this context, the development and maintenance of granulomatous inflammation in sarcoidosis have been repeatedly associated with the impaired apoptosis of activated inflammatory cells [8 9]. In most patients with sarcoidosis, the early stages of this disease are characterized by mononuclear cell alveolitis dominated by activated $CD4^+$ T cells and macrophages. Between these cells, the uncoordinated interplay results in the formation of typical noncaseating granulomas. The mechanism by which the granulomas resolve has not

been fully elucidated. However, it is generally assumed that the induction of apoptosis and / or by the withdrawal of inflammatory cytokines participates in the disappearance of granulomas [10 11]. In patients with sarcoidosis, their peripheral blood monocytes were characterized by apoptosis [9].

In previous studies, case - control / family 'hypothesis - driven' studies and low density linkage scans were used to identify genetic factors conferring susceptibility to sarcoidosis [12]. Notably, the first genome - wide association study (GWAS) in sarcoidosis conducted in a German population has recently revealed an association: the leading rs2789679 SNP located in the 3' - untranslated region (3'UTR) and a common nonsynonymous SNP (rs1049550, C > T, Arg230Cys) were associated with increased risk of sarcoidosis [4]. This association has recently been supported by another report from the same population [13]. Thus, more studies should be performed to demonstrate the following item: whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations. The published sarcoidosis - association studies for ANXA11 summarized in Table 1. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis.

In this study, we investigated fifteen loci in a Chinese population from He'nan province (China) to verify the putative association between ANXA11 polymorphisms and sarcoidosis.

SUBJECTS AND METHODS

Subjects

Four - hundred and twelve patients with sarcoidosis (mean ± SD age; 53.6 ± 4.6 years) were recruited from our hospital between May 2005 and March 2013. Sarcoidosis was diagnosed by the evidence of non - caseating epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities. Chronic sarcoidosis was defined as a disease over at least 2 years or at least two episodes in a lifetime. Acute sarcoidosis was defined as one episode of acute sarcoidosis which had totally resolved at the date of the examination. The control group consisted of 418 unrelated healthy subjects (mean ± SD age; 54.2 ± 5.3 years) who underwent health examinations in the Medical Examination Center of the First Affiliated Hospital of the Xin'xiang Medical College (Xinxiang, China). None of the individuals in the control group had a history of lung diseases or showed any symptoms of the lung or other diseases by chest radiography or laboratory blood tests. All participants were from a non - genetically related Chinese Han population in He'nan Province (China). The study was performed according to the Guidelines of the Medical Ethical Committee of Xin'xiang Medical College (Xinxiang, China). Written informed consent was obtained from each participant in this study.

SNP selection

Tagging SNPs were selected from the catalogs of the International HapMap Project and the 1000 Genomes Project. The rs2789679 and rs2245168 are located in 5' - UTR, rs2236558 in intron 6, rs7067644 and rs12763624 in intron 8, rs1049550 in exon10, rs2573351 in intron 13, rs10887581, rs11201989, rs2573353, rs2789695,

rs2573356, rs2819945, rs11202059 and rs2819941 in intron 14. Marker selection was done according to previous studies [4 12-14], and preliminary analysis was performed using the HapMap data (Fig 1). We examined tagSNPs in the Haploview software v4.2, using the Chinese Han in Beijing (CHB) population and a minor allele frequency cut-off (MAF) ≥ 5% (the HapMap Data Release 27). The linkage disequilibrium (LD) pattern of this gene was determined in the Chinese population using the preliminary data from the HapMap. These SNPs were further analyzed in an association study.

Genotyping

Peripheral blood was collected from a vein into a sterile tube coated with ethylenediamine tetraacetic acid (EDTA). Plasma samples were stored at -20°C. Genomic DNA was extracted from the frozen peripheral blood samples using a QIAmp Blood Mini Kit (Qiagen Inc., Valencia, CA, USA) according to the manufacturer's protocols. The selected SNPs were genotyped in cases and controls by using the MALDI - TOF in the MassARRAY system (Sequenom Inc., San Diego, CA, USA). Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA). The completed genotyping reactions were spotted onto a 384 well spectroCHIP (Sequenom) using the MassARRAY Nanodispenser (Sequenom) and determined by the matrix - assisted laser desorption ionization time - of - flight mass spectrometer. Genotype calling was performed in realtime with the MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer software version 3.4 (Sequenom).

Statistical analysis

All statistical analyses were performed using the SPSS 17.0 software (SPSS Inc., Chicago, IL, USA). Allele and genotype frequencies for each individual polymorphism and Hardy - Weinberg equilibrium were evaluated by Chi - square tests. Differences between the cases and controls in the frequency of the alleles, genotypes and haplotypes were evaluated by the Fisher's exact test or the Pearson Chi - square test. Unconditional logistic regression was used to calculate the odds ratio (OR) and 95% confidence interval (CI) in independent association between each locus and the presence of sarcoidosis. The Bonferroni correction was used to adjust the test level when multiple comparisons were conducted, and the *P* value was divided by the total number of loci or haplotypes. Haplotype blocks were defined according to the criteria developed by Gabriel et al. [14], as implemented in the Haploview 4.0, to examine if some SNPs significant in the single - marker association analysis also exist in the haplotype blocks. Pair - wise LD statistics (D' and r^2) and haplotype frequency were calculated, and haplotype blocks were constructed using the Haploview 4.0 [15].

RESULTS

The genotype distribution of the fifteen polymorphisms was consistent with the Hardy - Weinberg equilibrium ($P > 0.05$). The analysis of strong LD in the patients with sarcoidosis and the healthy controls revealed that 2 SNPs (rs2245168, rs2236558), 2 SNPs (rs1049550, rs2573351), 2 SNPs (rs2789695, rs2573356) and 3 SNPs (rs2819945, rs11202059, rs2819941) were located in haplotype block 1, block 2, block 3 and block 4 (D' > 0.9, Fig. 2). The genotype distribution, allelic frequencies,

and haplotypes in the patients with sarcoidosis and the healthy controls are showed in Tables 2 and Table 3.

Comparison of genotype and allele frequency distribution revealed significant differences between the patients with sarcoidosis and healthy controls for 3 SNPs: rs2789679, rs1049550 and rs2819941. The frequency of the T allele in rs2789679 (χ^2 = 12.849, P = 0.0003, OR = 0.703, 95% CI = 0.579 - 0.852), T allele inrs1049550 (χ^2 = 24.420, P = 0.0007, OR = 0.617, 95% CI = 0.503 - 0.758) and the rs2819941 C allele frequency (χ^2 = 11.803, P = 0.001, OR = 0.713, 95% CI = 0.588 - 0.865) in the patients with sarcoidosis was significantly lower than that in the controls, and these differences remained statistically significant after Bonferroni corrections.

Due to the positive association between 2 SNPs (rs1049550 and rs2819941), and sarcoidosis, we performed an association analysis to determine whether the haplotype was associated with the risk of sarcoidosis. Compared with the healthy controls, the T - C haplotype occurred significantly less frequently (P = 0.001) and the C - C haplotypes occurred more frequently (P = 0.0001) in block 2 in the patients with sarcoidosis. The significantly more G - G -C haplotypes (block 4) (P = 0.027) were found in the patients with sarcoidosis, but did not pass the threshold value (P = 0.025).

DISCUSSION

A key step in linkage and association studies is to identify common risk variants in different populations. To determine if common risk variants exist in distinct populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene,

were genotyped in samples from the patients with sarcoidosis and healthy controls in a Chinese Han population. Recently, new sarcoidosis loci have been identified by genome - wide association studies [4 13]. With the fast development of genome-wide association studies, increasing number of susceptibility loci for sarcoidosis have been reported in various populations [4 13-14]. However, these observations should be confirmed in other genetically independent populations. In this study, we conducted the first large genetic association study of the ANXA11 gene in a Chinese Han population. The evidence of markers associated with sarcoidosis was presented, and these markers were mapped to different locations in the ANXA11 gene (81897864 - 81951001). The association signals in the region were identified, and some significantly associated haplotypes also appeared this region.

Hoffman et al. reported an association between the T allele of the non - synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a German population [4]. Similar results were also obtained in other two European populations [13 14]. As a part of the sarcoidosis GWAS done in Americans [16], we further confirmed this associationin a Chinese Han population. In this study, the frequency of ANXA11 rs1049550 T allele was significantly lower in the patients with sarcoidosis than that in the healthy controls. The “protective” effect of ANXA11 T allele increased with the number of its copies in the genotype, which is consistent with the result obtained in the patients with sarcoidosis in a German population [4], Furthermore, the T allele carriers among the patients were protected from infiltration of lung parenchyma (radiographic stages II-IV) [14]. We have also demonstrated that rs1049550

is significant cross-ethnically at the gene level after adjustment for the single SNP association tests performed. The mechanism by which rs1049550 SNP affects the susceptibility to sarcoidosis may be that the SNP may affect the function of the ANXA11 protein rather than affect its expression. In humans, the ANXA11 protein consists of an N - terminalproline - tyrosine - glycine (PYG) - rich region, followed by four annexincore domains. The SNP rs1049550 leads to an amino - acid exchange (basic arginine to a polar cysteine) at the evolutionarily conserved position 230 (R230C) in the first annexin domain, which is responsible for Ca^{2+} - dependent trafficking of the protein inthe cell [17].

In the GWAS conducted by Hofmann et al 4, the strongest association signal was observed for SNP rs2789679. This marker is located in the 3' - UTR of the ANXA11 gene. The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in controls, with 13% of the patients with sarcoidosis and 19% of controls being TT homozygous. Fine - mapping in the Black Women's Health Study (BWHS) revealed rs2819941 to be the top SNP, associated with a nonsignificant ($P = 0.06$) reduction in sarcoidosis risk (17% reduction per copy of the C - allele). In this case - control association study, the frequency of the T allele in rs2789679 and the rs2819941 C allele frequency in the patients with sarcoidosis was significantly lower than that in the controls. Several lines of evidence suggest that the observed associations are unlikely to be anartifact. First, both the single - SNP and the haplotype - based association analyses support the current finding. Second, population stratification is an impossible reason, because all of our samples were from the same geographical

region. Finally, consistent results were obtained from two genetically independent populations (Han Chinese and Europeans). Collectively, our results confirmed the strong association between variations in the ANXA11 gene and sarcoidosis, and suggest that ANXA11 represents a strong genetic risk factor for sarcoidosis.

We further investigated the interaction among polymorphisms and observed strong LD. The haplotype analysis revealed that significantly more C - C (block 2), and G - G - C (block 4) haplotypes and significantly fewer T - C haplotypes (block 2) were found in the patients with sarcoidosis. There were significant point - wise associations of these variants with sarcoidosis (rs2789679, rs1049550 and rs2819941). These results indicated that the patients with C - C (block 2), and G - G - C (block 4) haplotypes of the ANXA11 gene were more prone to sarcoidosis. Significantly higher frequencies of T - C haplotypes were detected in the healthy controls than in the patients with sarcoidosis, suggesting that they may show a protective effect against sarcoidosis. To some extent, this finding further supports a role of ANXA11 polymorphisms in sarcoidosis, while ethnic group difference may exist.

In conclusion, our study suggests a potential role of the ANXA11 SNPs and their related haplotypes in the genetic susceptibility to sarcoidosis. Further studies are needed to investigate how these SNPs affect the function of ANXA11.

Contributors

LG Zhang was involved in conception and design of the study, acquisition of patient data, genotyping and drafted the paper. ZJ Xiao and GC Shi were involved in design of the study, genotyping and interpretation of results. J Huang and CX Zhang

were involved in design of the study, acquisition of patient data and interpretation of result. All authors revised the draft paper.

Funding

The project was supported by the Research and Development Foundation of Science and Technology of Henan Province (2012k15).

Competing interests

None.

Patient consent

Obtained.

Ethics approval

Ethics approval was provided by the First Hospital Affiliated to the Xinxiang Medical College.

Provenance and peer review

Not commissioned; externally peer reviewed.

Data sharing statement

No additional data are available.

Figure Legend

Fig. 1 The structure of the human ANXA11 gene and ten SNPs located on the gene.

Fig. 2 The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the pair - wise calculation of r^2 (left) or D' (right). Black squares indicate $r^2 = 1$ (i.e. perfect LD between a pair of SNPs). Empty squares indicate D' = 1 (i.e. complete LD

between a pair of SNPs).

Reference

1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007;357:2153-2165.

2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-1118.

3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis. *Semin Respir Crit Care Med* 2010;31:380-389.

4. Hofmann S, Franke A, Fischer A, et al. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet* 2008;40:1103-1106.

5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev* 2002;82:331-371.

6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using recombinant human annexin XI-specific autoantibodies cloned by phage display. *J Biol Chem* 2003;278:33120-33126.

7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-822.

8. Petzmann S, Maercker C, Markert E, et al. Enhanced proliferation and decreased apoptosis in lung lavage cells of sarcoidosis patients. Sarcoidosis, vasculitis, and diffuse lung diseases : official journal of WASOG / World Association of Sarcoidosis and Other Granulomatous Disorders 2006;23:190-200.

9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, Dubaniewicz A, Singh M, Mysliwski A. Comparative analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol* 2006;26:243-250.

10. Agostini C, Trentin L, Facco M, et al. Role of IL-15, IL-2, and their receptors in the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol* 1996;157:910-918.

11. Jullien D, Sieling PA, Uyemura K, Mar ND, Rea TH, Modlin RL. IL-15, an immunomodulator of T cell responses in intracellular infection. *J Immunol* 1997;158:800-806.

12. Muller-Quernheim J, Schurmann M, Hofmann S, et al. Genetics of sarcoidosis. *Clin Chest Med* 2008;29:391-414.

13. Li Y, Pabst S, Kubisch C, Grohe C, Wollnik B. First independent replication study confirms the strong genetic association of ANXA11 with sarcoidosis. *Thorax* 2010;65:939-940.

14. Mrazek F, Stahelova A, Kriegova E, et al. Functional variant ANXA11 R230C: true marker of protection and candidate disease modifier in sarcoidosis. *Genes Immun* 2011;12:490-494.

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

16. Levin AM, Iannuzzi MC, Montgomery CG, et al. Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans.

Genes Immun 2013;14:13-18.

17. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11 with the nuclear envelope. *J Biol Chem* 2003;278:20210-20216.

Table 1 Summary of all published sarcoidosis-association studies for ANXA11

Study	Population		Type of Study	Sample Size (n)		Number of SNPs typed	Positive SNPs
	Ethnic group	Country		Sarcoidosis	Control		
Hofmann S, 2008	Caucasian	Germany	Case-control	499	490	GWAS	rs2789679, rs7091565, rs1049550
Mrazek F, 2011	Caucasian	Czech	Case-control	245	254	1	rs1049550
Cozier YC, 2012	Caucasian	America	Case-control	486	943	10p12-10q22	rs2789679, rs2819941
Li Y, 2013	Caucasian	Germany	Case-control	325	364	1	rs1049550
Levin AM, 2013	Caucasian	America	Case-control	1689	1252	25	rs1049550

Table 2 Genotype and allele frequencies of the ANXA11 gene polymorphisms in Controls (n=418) and sarcoidosis (n=412) and their associated risk

Variable	Location	MAF	Group	Genotype (n, %)			Allele (n, %)		P^a	P^b	OR, 95%CI[c]
				AA	AT	TT	A	T			
rs2789679	3'UTR	0.388	Controls	79 (18.900)	225 (53.828)	114 (27.273)	383 (45.813)	453 (54.187)	0.0007	0.0003	1.423, 1.173-1.726
			Cases	132 (32.039)	186 (45.146)	94 (22.816)	450 (54.612)	374 (45.388)			
				CC	CT	TT	C	T			
rs2245168	3'UTR	0.388	Controls	156 (37.321)	200 (47.847)	62 (14.833)	512 (61.244)	324 (38.756)	0.920	0.698	1.040, 0.854-1.266
			Cases	150 (36.408)	197 (47.816)	65 (15.777)	497 (60.316)	327 (39.684)			
				TT	TG	GG	T	G			
rs2236558	intron 6	0.488	Controls	110 (26.316)	190 (45.455)	118 (28.230)	410 (49.043)	426 (50.957)	0.172	0.917	1.010, 0.833-1.025
			Cases	94 (22.816)	214 (51.942)	104 (25.243)	402 (48.786)	422 (51.214)			
				GG	GA	AA	G	A			
rs7067644	intron 8	0.390	Controls	60 (14.354)	206 (49.282)	152 (36.364)	326 (38.995)	510 (61.005)	0.173	0.109	1.177, 0.964-1.437
			Cases	54 (13.107)	182 (44.175)	176 (42.718)	290 (35.194)	534 (64.806)			
				CC	CT	TT	C	C			
rs12763624	intron 8	0.440	Controls	76 (18.182)	216 (51.675)	126 (30.144)	368 (44.019)	468 (55.981)	0.118	0.216	1.131, 0.931-1.374
			Cases	76 (18.447)	186 (45.146)	150 (36.408)	338 (69.547)	486 (58.981)			
				CC	CT	TT	C	T			
rs1049550	exon 10	0.404	Controls	154 (36.842)	190 (45.455)	74 (17.703)	498 (59.569)	338 (40.431)	0.0002	0.0007	0.598, 0.488-0.734
			Cases	208 (50.485)	170 (41.262)	34 (8.252)	586 (71.117)	238 (28.883)			
				CC	CT	TT	C	T			
rs2573351	intron 13	0.471	Controls	96 (22.967)	202 (48.325)	120 (28.708)	394 (47.129)	442 (52.871)	0.437	0.564	0.945, 0.779-1.146

			Cases	92 (22.330)	216 (52.427)	104 (25.243)	400 (48.544)	424 (51.456)			
				TT	TC	CC	T	C			
rs10887581	intron 14	0.500	Controls	114 (27.273)	190 (45.455)	114 (27.273)	418 (50.000)	418 (50.000)	0.290	0.553	1.060, 0.874-1.285
			Cases	96 (23.301)	208 (50.485)	108 (26.214)	400 (48.544)	424 (51.456)			
				GG	GC	CC	G	C			
rs11201989	intron 14	0.425	Controls	130 (31.100)	218 (52.153)	70 (16.746)	478 (57.177)	358 (42.823)	0.144	0.293	1.110, 0.914-1.347
			Cases	128 (31.068)	194 (47.087)	90 (21.845)	450 (54.612)	374 (45.388)			
				CC	CA	AA	C	A			
rs2573353	intron 14	0.340	Controls	178 (42.584)	196 (46.890)	44 (10.526)	552 (66.029)	284 (33.971)	0.995	0.863	1.000, 0.817-1.226
			Cases	176 (42.718)	192 (46.602)	44 (10.680)	544 (66.019)	280 (33.981)			
				CC	CT	TT	C	T			
rs2789695	intron 14	0.349	Controls	52 (12.440)	188 (44.976)	178 (42.584)	292 (34.928)	544 (65.072)	0.918	0.685	1.043, 0.852-1.277
			Cases	48 (11.650)	184 (44.660)	180 (43.689)	280 (33.981)	544 (66.019)			
				CC	CT	TT	C	T			
rs2573356	intron 14	0.490	Controls	118 (28.230)	190 (45.455)	110 (26.316)	426 (50.957)	410 (49.043)	0.892	0.926	1.009, 0.833-1.223
			Cases	112 (27.184)	194 (47.087)	106 (25.728)	418 (50.728)	406 (49.272)			
				GG	GA	AA	G	A			
rs2819945	intron 14	0.452	Controls	119 (28.469)	221 (52.871)	78 (18.660)	459 (54.904)	377 (45.096)	0.101	0.827	1.022, 0.842-1.240
			Cases	130 (31.553)	188 (45.631)	94 (22.816)	448 (54.369)	376 (45.631)			
				GG	GA	AA	G	A			
rs11202059	intron 14	0.476	Controls	106 (25.359)	226 (54.067)	86 (20.574)	438 (52.392)	398 (47.608)	0.064	0.479	0.933, 0.769-1.131
			Cases	128 (31.068)	190 (46.117)	94 (22.816)	446 (54.126)	378 (45.874)			
				CC	CT	TT	C	T			
rs2819941	intron 14	0.457	Controls	114 (27.273)	226 (54.067)	78 (18.660)	454 (54.306)	382 (45.694)	0.0002	0.001	1.402, 1.156-1.701
			Cases	94 (22.816)	190 (46.117)	128 (31.068)	378 (45.874)	446 (54.126)			

[a] p values for genotype frequency distribution. [b] p values for allele frequency distribution. [c] OR and 95%CI values for allele frequency distribution.

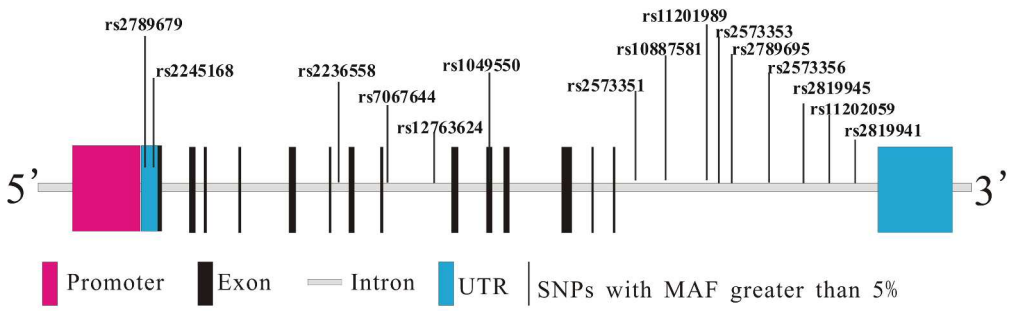
[*] *P* value is adjusted by Bonferroni correction and statistically significant results ($P < 0.003$).

All SNPs were found to be in Hardy-Weinberg equilibrium ($p > 0.05$).

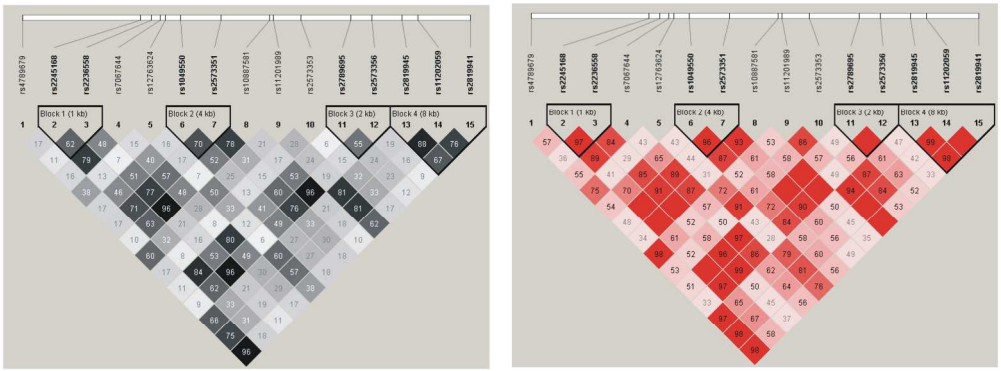
Table 3 ANXA11 haplotype in block 1-4 frequencies and the results of their associations with risk of sarcoidosis

Block	Haplotype	Cases(n, %)	Controls(n, %)	Statistics			
				χ^2	*P*	OR	95%CI
1	C - T	201 (48.786)	212 (50.718)	0.310	0,578	0.926	0.705～1.215
	T - G	163 (39.563)	160 (38.278)	0.144	0.104	1.056	0.199～1.395
	C - G	47 (11.408)	48 (11.483)	0.001	0.973	0.993	0.647～1.522
2	C - T	210 (50.971)	218 (52.153)	0.116	0.733	0.954	0.726～1.252
	T - C	117 (27.725)	166 (39.713)	11.822	0.001	0.602	0.450～0.805
	C - C	82 (19.903)	31 (7.416)	27.507	0.0001	3.102	2.001～4.810
3	C - T	208 (50.485)	213 (50.957)	0.018	0.892	0.981	0.747～1.288
	T - C	139 (33.738)	146 (34.928)	0.130	0.718	0.949	0.712～1.263
	C - C	63 (15,291)	59 (14.115)	0.229	0.632	1.098	0.748～1.613
4	G - G -C	219 (53.155)	190 (45.455)	4.923	0.027	1.362	1.036～1.789
	G - A - C	185 (44.903)	188 (44.876)	0.0001	0.997	0.758	0.712～1.311

* *P* value is adjusted by Bonferroni correction and statistically significant results ($P < 0.025$ in block 1-3; $P < 0.0167$ in block 4).



The structure of the human ANXA11 gene and ten SNPs located on the gene.



r^2 D'

The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the pair - wise calculation of r2 (left) or D’ (right). Black squares indicate r2 = 1 (i.e. perfect LD between a pair of SNPs). Empty squares indicate D’ = 1 (i.e. complete LD between a pair of SNPs).

BMJ open

Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a Han Chinese population - a case-control study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-004466.R1
Article Type:	Research
Date Submitted by the Author:	11-Feb-2014
Complete List of Authors:	Zhang, LG Xiao, ZJ Shi, GC J, Huang Zhang, CX
Primary Subject Heading:	Infectious diseases
Secondary Subject Heading:	Genetics and genomics
Keywords:	GENETICS, IMMUNOLOGY, INFECTIOUS DISEASES

Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a Han Chinese population - a case-control study

LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang

Internal Medicine - Tuberculosist, First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China

*Corresponding author: Zhijian Xiao, Internal Medicine - Tuberculosisthe, First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China

No.88 Jiankang Road, Weihui City, Henan, 463000, PR China

E-mail: zhangligongys@163.com

Tel: +86 0373 - 4402229 (office)

Fax: +86 0373 -4402251

Key words: Sarcoidosis; Annexin A11; Single-nucleotide polymorphism

Word count: 2909

Abstract

Objectives: To further identify the single - nucleotide polymorphisms (SNPs) that contribute to the genetic susceptibility to sarcoidosis, we examined the potential association between sarcoidosis and fifteen SNPs of the ANXA11 gene.

Design: A case - control study.

Setting: A tuberculosist unit in a university hospital in China.

Participants: Participants included 412 patients with sarcoidosis and 418 healthy controls.

Methods: The selected SNPs were genotyped in cases and controls by using the MALDI - TOF in the MassARRAY system (Sequenom Inc., San Diego, CA, USA). Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA).

Results: The results showed that statistically significant differences were observed in the allelic or genotypic frequencies of the rs2789679, rs1049550 and rs2819941 in the ANXA11 gene. The frequency of the T allele in rs2789679 (P = 0.0003, OR = 0.70, 95%CI = 0.58 - 0.85), T allele in rs1049550 (P = 0.0007, OR = 0.62, 95%CI= 0.50 - 0.76) and the rs2819941 C allele frequency (P = 0.001, OR = 0.71, 95%CI = 0.59 - 0.87) in the patients with sarcoidosis was significantly lower than that in the controls. Comparison of genotype frequency distribution revealed significant differences between chest radiographic (CXR) stage I and stages II - IV for rs1049550. The significantly more CC genotype (P = 0.012) were found in the patients with stages I sarcoidosis. Furthermore, strong linkage disequilibrium (LD) was observed in four

blocks (D' > 0.9). In block 2 (rs1049550-rs2573351), the T - C haplotype occurred significantly less frequently (P = 0.001), whereas the C - C haplotypes occurred more frequently (P = 0.0001).

Conclusions: These findings point to a role for ANXA11 gene polymorphisms in sarcoidosis in a Han Chinese population, and may be informative for future genetic or biological studies on sarcoidosis.

Article summary

Article focus

1. The first genome - wide association study (GWAS) in sarcoidosis conducted in a German population has revealed that the rs1049550 (C > T, Arg230Cys) located in exon 10 were associated with sarcoidosis.

2. More studies should be performed to demonstrate the following item: whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations.

3. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis in a Chinese Han population.

Key messages

1. ANXA11 rs2789679 (3'UTR) was found to be associated with sarcoidosis.

2. ANXA11 rs1049550 (exon 10) was significantly associated with sarcoidosis.

3. Another significant association was observed for rs2819941 (intron 14).

Strengths and limitations of this study

1. Functional SNPs in the promoter region, 5'- and 3'-UTR, exons of ANXA11 gene were systematically screened and homogeneity of the study subjects, representing the Han Chinese, is the main strength of the current study.

2. Functional characterization is the potential limitation of the study that could have further helped in proving the positive association observed for rs2789679 and rs1049550. The lack of correlation of serum ANXA11 levels are needed to investigate.

INTRODUCTION

Sarcoidosis is a systemic autoimmune disease characterized by destructive, non - caseating epithelioid granulomatous lesions, with accumulated phagocytes and activated $CD4^+$ T helper type 1 lymphocytes [1 2]. The typical manifestations of sarcoidosis include bilateral hilar lymphadenopathy, pulmonary infiltration and ocular and skin lesions. The clinical course and prognosis of sarcoidosis are variable. The acute forms of sarcoidosis such as L□fgren's syndrome (LS) can resolve spontaneously in most cases. However, chronic pulmonary sarcoidosis may progress to lung fibrosis and eventually result in respiratory failure. Recent studies have demonstrated that sarcoidosis can occur in a pattern of family clustering, and its racial incidence is also different, suggesting that some genetic factors may contribute to the risk and severity of sacroidosis [1 3].

The Annexin gene family is involved in the etiology of several autoimmune and

chronic diseases [4 5.] One member of the Annexin gene family, ANXA11, located on chromosome 10q22.3, is involved in calcium signaling, apoptosis, vesicle trafficking, cell growth and terminal phase of cell division [5-7]. In this context, the development and maintenance of the granulomatous inflammation in sarcoidosis have been repeatedly associated with the impaired apoptosis of activated inflammatory cells [8 9]. In most patients with sarcoidosis, the early stages of this disease are characterized by mononuclear cell alveolitis dominated by activated $CD4^+$ T cells and macrophages. Between these cells, the uncoordinated interplay results in the formation of typical non - caseating granulomas. The mechanism by which the granulomas resolve has not been fully elucidated. However, it is generally assumed that the induction of apoptosis and / or by the withdrawal of inflammatory cytokines participates in the disappearance of granulomas [10 11]. In the patients with sarcoidosis, their peripheral blood monocytes were characterized by apoptosis [9].

In previous studies, case - control / family 'hypothesis - driven' studies and low density linkage scans were used to identify genetic factors conferring the genetic susceptibility to sarcoidosis [12]. Notably, the first genome - wide association study (GWAS) in sarcoidosis conducted in a German population has recently revealed an association: the leading rs2789679 SNP (single - nucleotide polymorphism) located in the 3' - untranslated region (3' - UTR) and a common nonsynonymous SNP (rs1049550, C > T, Arg230Cys) were associated with increased risk of sarcoidosis [4 13]. This association has recently been supported by another report from the same population [14]. Thus, more studies should be performed to demonstrate the following

item: whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations. The published sarcoidosis - association studies for ANXA11 are summarized in Table 1. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis.

In this study, we investigated fifteen loci in a Chinese population from He'nan province (China) to verify the putative association between ANXA11 polymorphisms and sarcoidosis.

SUBJECTS AND METHODS

Subjects

Four hundred and twelve patients with sarcoidosis (mean ± SD age: 53.6 ± 4.6 years; 155 men and 257 women) were recruited from our hospital between May 2005 and March 2013. Sarcoidosis was diagnosed by the evidence of non - caseating epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities. Chronic sarcoidosis was defined as a disease over at least 2 years or at least two episodes in a lifetime. A cute sarcoidosis was defined as one episode of acute sarcoidosis which had totally resolved at the date of the examination. The control group consisted of 418 unrelated healthy subjects (mean ± SD age: 54.2 ± 5.3 years; 160 men and 258 women) who underwent health examinations in the Medical Examination Center of the First Affiliated Hospital of the Xinxiang Medical College (Xinxiang, China) between Oct 2009 and Sept 2013. None of the individuals in the control group had a history of lung diseases or showed any symptoms of the lung or

other diseases by chest radiography or laboratory blood tests. Participants were excluded if they: were taking other prescribed medications that could affect the central nervous system; had a history of seizures, hematological diseases, or severe liver or kidney impairment; or were pregnant. No familial relationship was known between the study participants. The study was performed according to the Guidelines of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang, China). Written informed consent was obtained from each participant in this study.

SNP selection

Tagging SNPs were selected from the catalogs of the International HapMap Project. The rs2789679 and rs2245168 are located in 5'-UTR, rs2236558 in intron 6, rs7067644 and rs12763624 in intron 8, rs1049550 in exon 10, rs2573351 in intron 13, rs10887581, rs11201989, rs2573353, rs2789695, rs2573356, rs2819945, rs11202059 and rs2819941 in intron 14 (Figure 1). Marker selection was done according to previous studies [4 12 14 15], and preliminary analysis was performed using the HapMap data and the following criteria. First, we examined tagSNPs in the Haploview (v4.2), using the CHB population and a minor allele frequency cut - off (MAF) $\geq$ 5% (HapMap Data Release 27). We found that there were a total of 29 potential tagSNPs in all. As a first screen of the most common SNPs in the sarcoidosis sample from an Eastern Chinese Han population, a MAF $\geq$ 20% with pair - wise tagging and $r^2 \geq 0.8$ [16] was used as the cut - off when choosing tagSNPs. Second, the LD pattern of this gene was determined in the Chinese population using the preliminary data from the HapMap. Five LD blocks across the ANXA11 were defined using Haploview's

'confidence intervals' method [17 18]. These SNPs were further analyzed in an association study.

Genotyping

Peripheral blood was collected from a vein into a sterile tube coated with ethylenediamine tetraacetic acid (EDTA). Plasma samples were stored at -20°C. Genomic DNA was extracted from the frozen peripheral blood samples using a QIAmp Blood Mini Kit (Qiagen Inc., Valencia, CA, USA) according to the manufacturer's protocols. The selected SNPs were genotyped in cases and controls by using the MALDI - TOF in the MassARRAY system (Sequenom Inc., San Diego, CA, USA). Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA). The completed genotyping reactions were spotted onto a 384 well spectroCHIP (Sequenom) using the MassARRAY Nanodispenser (Sequenom) and determined by the matrix - assisted laser desorption ionization time - of - flight mass spectrometer. Genotype calling was performed in realtime with the MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer software version 3.4 (Sequenom).

Statistical analysis

All data were analyzed using the SPSS 17.0 software (SPSS Inc., Chicago, IL, USA). Hardy-Weinberg equilibrium were evaluated by Chi - square tests. Differences between the cases and controls in the frequency of the alleles, genotypes and haplotypes were evaluated by the Fisher's exact test or the Pearson Chi - square test. Unconditional logistic regression was used to calculate the odds ratio (OR) and 95%

confidence interval (CI) in independent association between each locus and the presence of sarcoidosis. Generalized linear regression was used to evaluate the interaction effects between gene and gender or age. Gender and age of subjects were treated as covariants in binary logistic regression. *P* - value was calculated based on codominant, dominant for the rare allele, heterosis and recessive for the rare allele models of inheritance. The Bonferroni correction was used to adjust the test level when multiple comparisons were conducted, and the *P* value was divided by the total number of loci. Haplotype blocks were defined according to the criteria developed by Gabriel et al. [18]. Pair - wise LD statistics (D' and r^2) and haplotype frequency were calculated, and haplotype blocks were constructed using the Haploview 4.0 [17]. To ensure that the LD blocks most closely reflect the population level LD patterns, definition of the blocks were based on the control samples alone. The haplotype frequencies were estimated using GENECOUNTING, which computes maximum - likelihood estimates of haplotype frequencies from unknown phase data by utilizing an expectation - maximization algorithm [19-22]. The significance of any haplotypic association was evaluated using a likelihood ratio test, followed by permutation testing that compared estimated haplotype frequencies in cases and controls [19 21].

RESULTS

The genotype distribution of the fifteen polymorphisms was consistent with the Hardy - Weinberg equilibrium ($P > 0.05$). The analysis of strong LD in the patients with sarcoidosis and healthy controls revealed that 2 SNPs (rs2245168, rs2236558), 2 SNPs (rs1049550, rs2573351), 2 SNPs (rs2789695, rs2573356) and 3 SNPs

(rs2819945, rs11202059, rs2819941) were located in haplotype block 1, block 2, block 3 and block 4 (D' > 0.9, Fig. 2). The genotype distribution, allelic frequencies, and haplotypes in the patients with sarcoidosis and healthy controls are showed in tables 2, table 4 and table 5.

Comparison of genotype and allele frequency distribution revealed significant differences between the patients with sarcoidosis and healthy controls for 3 SNPs: rs2789679, rs1049550 and rs2819941. The frequency of the T allele in rs2789679 (P = 0.0003, OR = 0.70, 95%CI = 0.58 - 0.85), T allele in rs1049550 (P = 0.0007, OR = 0.62, 95%CI = 0.50 - 0.76) and the rs2819941 C allele frequency (P = 0.001, OR = 0.71, 95%CI = 0.59 - 0.87) in the patients with sarcoidosis was significantly lower than that in the controls, and these differences remained statistically significant after Bonferroni corrections. There were no interactions among rs2789679, rs1049550 and rs2819941 (P > 0.05).

We performed an association analysis to determine whether the haplotype was associated with the risk of sarcoidosis (Global P < 0.05 in block 2 and block 3). Compared with the healthy controls, the T - C haplotype occurred significantly less frequently (P = 0.001) and the C - C haplotypes occurred more frequently (P = 0.0001) in block 2 (rs1049550 - rs2573351) in the patients with sarcoidosis.

To assess particular disease phenotypes, the patients with sarcoidosis were divided into the subgroups according to their chest radiographic (CXR) stage. CXR stage I (isolated bilateral hilar lymphadenopathy): N = 196; CXR stages II - IV (infiltration of parenchyma): 167; CXR stage 0 (absence on pulmonary disease): N =

28; the information on CXR stage was not available for 21 patients. Comparison of genotype frequency distribution revealed significant differences between stage I and stages II - IV for rs1049550. The significantly more CC genotype (P = 0.012) were found in the patients with stage I sarcoidosis (table 3).

DISCUSSION

A key step in linkage and association studies is to identify common risk variants in different populations. To determine if common risk variants exist in distinct populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene, were genotyped in samples from the patients with sarcoidosis and healthy controls in a Chinese Han population. Recently, new sarcoidosis loci have been identified by genome - wide association studies [4 14]. With the fast development of genome - wide association studies, increasing number of susceptibility loci for sarcoidosis have been reported in different populations [4 14 15]. However, these observations should be confirmed in other genetically independent populations. In this study, we conducted the first large genetic association study of the ANXA11 gene in a Chinese Han population. The evidence of markers associated with sarcoidosis was presented, and these markers were mapped to different locations in the ANXA11 gene (81897864 - 81951001). The association signals in the region were identified, and some significantly associated haplotypes also appeared this region.

Hoffman et al. reported an association between the T allele of the non - synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a German population [4]. Similar results were also obtained in other two European

populations [14 15]. As a part of the sarcoidosis GWAS done in Americans [23], we further confirmed this associationin a Chinese Han population. In this study, the frequency of ANXA11 rs1049550T allele was significantly lower in the patients with sarcoidosis than that in the healthy controls. Comparison of stages II - IV, the significantly more CC genotypewas found in the patients with stages I sarcoidosis. The "protective" effect of ANXA11 T allele increased with the number of its copies in the genotype, which is consistent with the result obtained in the patients with sarcoidosis in a German population [4], Furthermore, the T allele carriers among the patients were protected from infiltration of lung parenchyma (radiographic stages II-IV) [15]. We have also demonstrated that rs1049550 is significant cross - ethnically at the gene level after adjustment for the single SNP association tests performed. The mechanism by which rs1049550 affects the susceptibility to sarcoidosis may be that the SNP may affect the function of the ANXA11 protein rather than affect its expression. In humans, the ANXA11 protein consists of an N - terminalproline - tyrosine - glycine (PYG) - rich region, followed by four annexin core domains. The rs1049550 leads to an amino - acid exchange (basic arginine to a polar cysteine) at the evolutionarily conserved position 230 (R230C) in the first annexin domain, which is responsible for Ca^{2+} - dependent trafficking of the protein in the cell [24].

In the GWAS conducted by Hofmann et al [4], the strongest association signal was observed for rs2789679. This marker is located in the 3' - UTR of the ANXA11 gene. The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls

being TT homozygous. Fine - mapping in the Black Women's Health Study (BWHS) revealed rs2819941 to be the top SNP, associated with a nonsignificant (P = 0.06) reduction in sarcoidosis risk (17% reduction per copy of the C-allele). In this case - control association study, the frequency of the T allele in rs2789679 and the rs2819941 C allele frequency in the patients with sarcoidosis was significantly lower than that in the healthy controls. Several lines of evidence suggest that the observed associations are unlikely to be anartifact. First, both the single - SNP and the haplotype - based association analyses support the current finding. Second, population stratification is an impossible reason, because all of our samples were from the same geographical region. Finally, consistent results were obtained from two genetically independent populations (Chinese Han and Europeans). Collectively, our results confirmed the strong association between variations in the ANXA11 gene and sarcoidosis, and suggest that ANXA11 represents a strong genetic risk factor for sarcoidosis.

We further investigated the interaction among polymorphisms and observed strong LD. The haplotype analysis revealed that significantly more C - C (block 2, rs1049550 - rs2573351) and significantly fewer T - C haplotypes (block 2) were found in the patients with sarcoidosis. These results indicated that the patients with C - C (block 2) haplotypes of the ANXA11 gene were more prone to sarcoidosis. Significantly higher frequencies of T - C haplotypes were detected in the healthy controls than in the patients with sarcoidosis, suggesting that they may show a protective effect against sarcoidosis. Our sample size can detect SNP and haplotype

associations with 90% and 85% power, respectively, at a false positive rate of 5%, disease prevalence of 1‰, disease allele / haplotype frequency of 0.05 / 0.03, and a presumed odds ratio (OR) of 1.5. To some extent, this finding further supports a role of ANXA11 polymorphisms in sarcoidosis, while ethnic group difference may exist.

In conclusion, our study suggests a potential role of the ANXA11 SNPs and their related haplotypes in the genetic susceptibility to sarcoidosis. Further studies are needed to investigate how these SNPs affect the function of ANXA11.

Funding

The project was supported by the Research and Development Foundation of Science and Technology of Henan Province (2012k15).

Contributors

LG Zhang was involved in conception and design of the study, acquisition of patient data, genotyping and drafted the paper. ZJ Xiao and GC Shi were involved in

design of the study, genotyping and interpretation of results. J Huang and CX Zhang were involved in design of the study, acquisition of patient data and interpretation of result. All authors revised the draft paper.

Competing interests

None.

Patient consent

Obtained.

Ethics approval

Ethics approval was provided by theFirst Hospital Affiliated to the Xinxiang Medical College.

Provenance and peer review

Not commissioned; externally peer reviewed.

Data sharing statement

No additional data are available.

Reference

1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007; 357:2153-65.

2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-8

3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis. *Semin Respir Crit Care Med* 2010;31:380-9.

4. Hofmann S, Franke A, Fischer A, *et al*. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet* 2008;40:1103-6.

5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev* 2002;82:331-71.

6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using recombinant human annexin XI-specific autoantibodies cloned by phage display. *J Biol Chem* 2003;278:33120-6.

7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-22.

8. Petzmann S, Maercker C, Markert E, *et al*. Enhanced proliferation and decreased apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis Vasc Diffuse Lung Dis* 2006;23:190-200

9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, *et al*. Comparative

analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol* 2006;26:243-50.

10. Agostini C, Trentin L, Facco M, *et al*. Role of IL-15, IL-2, and their receptors in the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol* 1996;157:910-8
11. Jullien D, Sieling PA, Uyemura K, *et al*. IL-15, an immunomodulator of T cell responses in intracellular infection. *J Immunol* 1997;158:800-6
12. Muller-Quernheim J, Schurmann M, Hofmann S, *et al*. Genetics of sarcoidosis. *Clin Chest Med* 2008;29:391-414.
13. Morais A, Lima B, Peixoto M, *et al*. Annexin A11 gene polymorphism (R230C variant) and sarcoidosis in a Portuguese population. *Tissue antigens* 2013;82:186-91.
14. Li Y, Pabst S, Kubisch C, *et al*. First independent replication study confirms the strong genetic association of ANXA11 with sarcoidosis. *Thorax* 2010;65:939-40.
15. Mrazek F, Stahelova A, Kriegova E, *et al*. Functional variant ANXA11 R230C: true marker of protection and candidate disease modifier in sarcoidosis. *Genes Immun* 2011;12:490-4.
16. de Bakker PI, Yelensky R, Pe'er I, *et al*. Efficiency and power in genetic association studies. *Nat Genet* 2005;37:1217-23.
17. Barrett JC, Fry B, Maller J, *et al*. Haploview: analysis and visualization of LD and

haplotype maps. *Bioinformatics* 2005;21:263-5.

18. Gabriel SB, Schaffner SF, Nguyen H, *et al*. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-9.

19. Curtis D, Knight J, Sham PC. Program report: GENECOUNTING support programs. *Ann Hum Genet* 2006;70:277-9.

20. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic associations. *Hum Hered* 2000;50:133-9.

21. Zhao JH, Lissarrague S, Essioux L, et al. GENECOUNTING: haplotype analysis with missing genotypes. *Bioinformatics* 2002;18:1694-5

22. Zhao JH, Sham PC. Generic number systems and haplotype analysis. *Comput Methods Programs Biomed* 2003;70:1-9

23. Levin AM, Iannuzzi MC, Montgomery CG, *et al*. Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes Immun* 2013;14:13-8.

24. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11 with the nuclear envelope. *J Biol Chem* 2003;278:20210-6.

Figure Legend

Fig. 1 The structure of the human ANXA11 gene and 15 SNPs located on the gene.

Fig. 2 The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the pair-wise calculation of r^2 (left) or D' (right). Black squares indicate $r^2 = 1$ (i.e. perfect

LD between a pair of SNPs). Empty squares indicate D' = 1 (i.e. complete LD between a pair of SNPs).

Table 1 Summary of all published sarcoidosis-association studies for ANXA11

Study	Population		Type of Study	Sample Size (n)		Number of SNPs typed	Positive SNPs
	Ethnic group	Country		Sarcoidosis	Control		
Hofmann S, 2008	Caucasian	Germany	Case-control	499	490	GWAS	rs2789679, rs7091565, rs1049550
Mrazek F, 2011	Caucasian	Czech	Case-control	245	254	1	rs1049550
Cozier YC, 2012	Caucasian	America	Case-control	486	943	10p12-10q22	rs2789679, rs2819941
Li Y, 2013	Caucasian	Germany	Case-control	325	364	1	rs1049550

Levin AM, 2013	Caucasian	America	Case-control	1689	1252	25	rs1049550
Morais A, 2013	Caucasian	Portugal	Case-control	208	197	1	rs1049550

Table 2 Genotypic and allelic frequencies of ANXA11 polymorphisms in the controls and patients with sarcoidosis

Variable	Controls (n=418)	sarcoidosis (n=412)	*P*-value[a]	OR, 95% CI	Variable	Controls (n=418)	sarcoidosis (n=412)	*P*-value[a]	OR, 95% CI
	N, %	N, %				N, %	N, %		
rs2789679			0.0007		rs10887581			0.290	
AA	79, 18.9	132, 32.0	0.002	0.49, 0.36-0.68	TT	114, 27.3	96, 23.3	0.190	1.23, 0.90-1.69
AT	225, 53.8	186, 45.2	0.012	1.42, 1.08-1.87	TC	190, 45.5	208, 50.5	0.148	0.82, 0.62-1.

									07
TT	114, 27.3	94, 22.8	0.14 0	1.27, 0.93-1.74	CC	114, 27.3	108, 26.2	0.7 29	1.06, 0.78-1. 44
Per T allele	383, 45.8	450, 54.6	0.00 03	0.70, 0.58-0.85	Per C allele	418, 50.0	424, 51.5	0.5 83	1.06,0.8 7-1.29
rs224 5168			0.92 0		rs112 0198 9			0.1 44	
CC	156, 37.3	150, 36.4	0.79 1	1.04, 0.74-1.38	GG	130, 31.1	128, 31.1	0.9 95	1.00, 0.75-1. 34
CT	200, 47.9	197, 47.8	0.95 5	0.99, 0.76-1.30	GC	218, 52.2	194, 47.1	0.1 41	1.23, 0.93-1. 62
TT	62, 14.8	65, 15.8	0.78 1	0.95, 0.65-1.39	CC	70, 16.8	90, 21.9	0.0 60	0.71, 0.50-1. 01
Per T allele	324, 38.8	327, 39.7	0.69 8	1.04, 0.85-1.27	Per C allele	358, 42.8	374, 45.4	0.2 9	1.11, 0.91-1. 35
rs223 6558			0.17 2		rs257 3353			0.9 18	
GG	118, 28.2	104, 25.2	0.72	1.06, 0.77-1.45	CC	178, 42.6	176, 42.7	0.9 57	0.99, 0.75-1. 31
TG	190, 45.5	214, 51.9	0.06 0	0.77, 0.59-1.01	CA	196, 46.9	192, 46.6	0.9 21	1.01, 0.77-1. 33
TT	110, 26.3	94, 22.8	0.07 3	1.33, 0.97-1.82	AA	44, 10.5	44, 10.7	0.9 41	0.98, 0.63-1. 53
Per T allele	410, 49.0	402, 48.8	0.91 7	0.99, 0.82-1.20	Per A allele	284, 34.0	280, 34.0	0.9 97	1.00, 0.82-1. 23
rs706 7644			0.17 3		rs278 9695			0.9 18	
AA	152, 36.4	176, 42.7	0.05 9	0.76, 0.58-1.01	TT	178, 42.6	180, 34.0	0.7 40	0.95, 0.73-1. 26
GA	206,	182,	0.13	1.23,	CT	188,	184,	0.9	1.01,

	49.3	44.2	7	0.94-1.68		45.0	44.7	20	0.77-1.33
GG	60, 14.4	54, 13.1	0.597	1.11, 0.75-1.65	CC	52, 12.4	48, 11.7	0.73	1.08, 0.71-1.64
Per G allele	326, 38.0	290, 35.2	0.109	0.85, 0.70-1.04	Per C allele	292, 34.9	280, 34.0	0.69	0.96, 0.78-1.17
rs12763624			0.118		rs2573356			0.892	
TT	126, 75.6	150, 36.4	0.008	0.68, 0.51-0.91	CC	118, 28.2	112, 27.2	0.742	1.05, 0.78-1.43
CT	216, 51.7	186, 45.2	0.059	1.30, 0.99-1.71	CT	190, 45.5	194, 47.1	0.639	0.94, 0.71-1.23
CC	76, 18.2	76, 18.5	0.411	1.16, 0.81-1.67	TT	110, 26.3	106, 25.7	0.843	1.03, 0.76-1.41
Per C allele	368, 44.0	338, 41.0	0.216	0.88, 0.73-1.08	Per T allele	410, 49.0	406, 49.3	0.926	0.99, 0.82-1.20
rs1049550			0.0002		rs2819945			0.101	
CC	154, 36.8	208, 50.5	0.0007	0.57, 0.43-0.75	GG	119, 28.5	130, 31.6	0.294	0.85, 0.63-1.15
CT	190, 45.5	170, 41.3	0.221	1.19, 0.90-1.56	GA	221, 52.9	188, 45.6	0.030	1.35, 1.03-1.78
TT	74, 17.7	34, 8.3	0.0008	2.39, 1.55-3.68	AA	78, 18.7	94, 22.8	0.138	0.78, 0.55-1.09
Per T allele	338, 40.4	238, 28.8	0.0007	0.60, 0.49-0.73	Per A allele	377, 45.1	376, 45.6	0.827	1.02, 0.84-1.24
rs2573351			0.437		rs1120205 9			0.064	
TT	120, 28.7	104, 25.2	0.263	1.19, 0.88-1.62	GG	106, 25.4	128, 31.1	0.067	0.75, 0.56-1.

CT	202, 48.3	216, 52.4	0.238	0.85, 0.65-1.12	GA	226, 54.1	190, 46.1	0.021	1.38, 1.05-1.81
CC	96, 23.0	92, 22.3	0.826	1.04, 0.75-1.44	AA	86, 20.6	94, 22.8	0.429	0.88, 0.63-1.22
Per C allele	394, 47.1	400, 48.5	0.564	1.06, 0.87-1.28	Per A allele	398, 47.6	378, 45.9	0.479	0.93, 0.77-1.13
rs2819941			0.0002						
CC	114, 27.3	94, 22.8	0.141	1.27, 0.93-1.74					
CT	226, 54.1	190, 46.1	0.021	1.38, 1.05-1.81					
TT	78, 18.7	128, 31.1	0.0004	0.51, 0.37-0.70					
Per T allele	382, 45.7	446, 54.1	0.001	0.71, 0.59-0.87					

[a] *P*-value was calculated by 2×3 and 2×2 chi-squared tests based on codominant, dominant for the rare allele, heterosis and recessive for the rare allele models of inheritance.

Alpha value is adjusted by Bonferroni correction and statistically significant results (*P*<0.003)

Table 3 Chest radiographic (CXR) stages of sarcoidosis patients by genotype of rs2789679, rs1049550 and rs2819941

Stages	rs2789679 (n, %)			rs1049550* (n, %)			rs2819941 (n, %)		
	AA	AT	TT	CC	CT	TT	**CC**	**CT**	**TT**
stage I	59 (30.1)	92 (47.0)	45 (23.0)	101 (51.5)	82 (41.8)	13 (6.6)	47 (25.0)	88 (44.9)	61 (31.1)
stages II - IV	46 (27.6)	75 (44.9)	46 (27.6)	72 (43.1)	68 (40.7)	27 (16.2)	41 (24.6)	73 (43.7)	53 (31.7)
χ^2, P		1.041, 0.594			8.807, 0.012			0.052, 0.975	

* P values for genotype frequency distribution ($P < 0.05$).

Table 4 ANXA11 haplotype in block 2 frequencies and the results of their associations with risk of sarcoidosis

Haplotype			Genecounting (frequency %)			
ID	**rs1049550**	**rs2573351**	**Cases**	**Controls**	***P*-value[a]**	**Global *P*[b]**
HAP1	T	C	27.7	39.7	0.001	0.003
HAP2	C	C	19.9	7.4	0.0001	
HAP3	C	T	50.5	51.0	0.892	

[a] Based on 10,000 permutations.

[b] Based on comparison of frequency distribution of all haplotypes for the combination of SNPs.

Table 5 ANXA11 haplotype in block 3 frequencies and the results of their associations with risk of sarcoidosis

Haplotype			Genecounting (frequency %)			
ID	**rs2789695**	**rs2573356**	**Cases**	**Controls**	***P*-value[a]**	**Global *P*[b]**
HAP1	C	T	33.7	34.9	0.718	0.045
HAP2	T	T	15.3	14.1	0.632	
HAP2	T	T	15.6	14.1	0.565	

[a] Based on 10,000 permutations.

[b] Based on comparison of frequency distribution of all haplotypes for the combination of SNPs.

Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a Han Chinese population

LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang

Internal Medicine - Tuberculosist, First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China

*Corresponding author: Zhijian Xiao, Internal Medicine - Tuberculosisthe, First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China

No.88 Jiankang Road, Weihui City, Henan, 463000, PR China

E-mail: zhangligongys@163.com

Tel: +86 0373 - 4402229 (office)

Fax: +86 0373 -4402251

Key words: Sarcoidosis; Annexin A11; Single-nucleotide polymorphism

Word count: 2909

Abstract

Objectives: To further identify the single - nucleotide polymorphisms (SNPs) that contribute to the genetic susceptibility to sarcoidosis, we examined the potential association between sarcoidosis and fifteen SNPs of the ANXA11 gene.

Design: A case - control study.

Setting: A tuberculosist unit in a university hospital in China.

Participants: Participants included 412 patients with sarcoidosis and 418 healthy controls.

Methods: The selected SNPs were genotyped in cases and controls by using the MALDI - TOF in the MassARRAY system (Sequenom Inc., San Diego, CA, USA). Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA).

Results: The results showed that statistically significant differences were observed in the allelic or genotypic frequencies of the rs2789679, rs1049550 and rs2819941 in the ANXA11 gene. The frequency of the T allele in rs2789679 (P = 0.0003, OR = 0.70, 95%CI = 0.58 - 0.85), T allele in rs1049550 (P = 0.0007, OR = 0.62, 95%CI= 0.50 - 0.76) and the rs2819941 C allele frequency (P = 0.001, OR = 0.71, 95%CI = 0.59 - 0.87) in the patients with sarcoidosis was significantly lower than that in the controls. ~~The frequency of the T allele in rs2789679 (χ^2 = 12.849, P = 0.0003, OR = 0.703, 95%CI = 0.579 - 0.852), T allele in rs1049550 (χ^2 = 24.420, P = 0.0007, OR = 0.617, 95%CI= 0.503 - 0.758) and the rs2819941 C allele frequency (χ^2 = 11.803, P = 0.001, OR = 0.713, 95%CI = 0.588 - 0.865) in the patients with sarcoidosis was significantly~~

~~lower than that in the controls.~~ Comparison of genotype frequency distribution revealed significant differences between chest radiographic (CXR) stage I and stages II - IV for rs1049550. The significantly more CC genotype (P = 0.012) were found in the patients with stages I sarcoidosis. Furthermore, strong linkage disequilibrium (LD) was observed in ~~two~~ four blocks (D' > 0.9). In block 2 (rs1049550-rs2573351), the T - C haplotype occurred significantly less frequently (P = 0.001), whereas the C - C haplotypes occurred more frequently (P = 0.0001).

Conclusions: These findings point to a role for ANXA11 gene polymorphisms in sarcoidosis in a Han Chinese population, and may be informative for future genetic or biological studies on sarcoidosis.

Article summary

Article focus

1. The first genome - wide association study (GWAS) in sarcoidosis conducted in a German population has revealed that the rs1049550 (C > T, Arg230Cys) located in exon 10 were associated with sarcoidosis.

2. More studies should be performed to demonstrate the following item: whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations.

3. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis in a Chinese Han population.

Key messages

1. ANXA11 rs2789679 (3'UTR) was found to be associated with sarcoidosis.

2. ANXA11 rs1049550 (exon 10) was significantly associated with sarcoidosis.

3. Another significant association was observed for rs2819941 (intron 14).

Strengths and limitations of this study

1. Functional SNPs in the promoter region, 5'- and 3'-UTR, exons of ANXA11 gene were systematically screened and homogeneity of the study subjects, representing the Han Chinese, is the main strength of the current study.

2. Functional characterization is the potential limitation of the study that could have further helped in proving the positive association observed for rs2789679 and rs1049550. The lack of correlation of serum ANXA11 levels are needed to investigate.

~~The lack of correlation of serum ANXA11 levels and sample size are other limitations.~~

INTRODUCTION

Sarcoidosis is a systemic autoimmune disease characterized by destructive, non-caseating epithelioid granulomatous lesions, with accumulated phagocytes and activated $CD4^+$ T helper type 1 lymphocytes [1 2]. The typical manifestations of sarcoidosis include bilateral hilar lymphadenopathy, pulmonary infiltration and ocular and skin lesions. The clinical course and prognosis of sarcoidosis are variable. The acute forms of sarcoidosis such as L□fgren's syndrome (LS) can resolve spontaneously in most cases. However, chronic pulmonary sarcoidosis may progress to lung fibrosis and eventually result in respiratory failure. Recent studies have demonstrated that sarcoidosis can occur in a pattern of family clustering, and its racial incidence is also different, suggesting that some genetic factors may contribute to the risk and severity of sacroidosis [1 3].

The Annexin gene family is involved in the etiology of several autoimmune and chronic diseases [4 5.] One member of the Annexin gene family, ANXA11, located on chromosome 10q22.3, is involved in calcium signaling, apoptosis, vesicle trafficking, cell growth and terminal phase of cell division [5-7]. In this context, the development and maintenance of the granulomatous inflammation in sarcoidosis have been repeatedly associated with the impaired apoptosis of activated inflammatory cells [8 9]. In most patients with sarcoidosis, the early stages of this disease are characterized by mononuclear cell alveolitis dominated by activated $CD4^+$ T cells and macrophages.

Between these cells, the uncoordinated interplay results in the formation of typical non - caseating granulomas. The mechanism by which the granulomas resolve has not been fully elucidated. However, it is generally assumed that the induction of apoptosis and / or by the withdrawal of inflammatory cytokines participates in the disappearance of granulomas [10 11]. In the patients with sarcoidosis, their peripheral blood monocytes were characterized by apoptosis [9].

In previous studies, case - control / family 'hypothesis - driven' studies and low density linkage scans were used to identify genetic factors conferring the genetic susceptibility to sarcoidosis [12]. Notably, the first genome - wide association study (GWAS) in sarcoidosis conducted in a German population has recently revealed an association: the leading rs2789679 SNP (single - nucleotide polymorphism) located in the 3' - untranslated region (3' - UTR) and a common nonsynonymous SNP (rs1049550, C > T, Arg230Cys) were associated with increased risk of sarcoidosis [4 13]. This association has recently been supported by another report from the same population [14]. Thus, more studies should be performed to demonstrate the following item: whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations. The published sarcoidosis - association studies for ANXA11 are summarized in Table 1. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis.

In this study, we investigated fifteen loci in a Chinese population from He'nan province (China) to verify the putative association between ANXA11 polymorphisms

and sarcoidosis.

SUBJECTS AND METHODS

Subjects

Four hundred and twelve patients with sarcoidosis (mean ± SD age: 53.6 ± 4.6 years; 155 men and 257 women) ~~(mean ± SD age: 53.6 ± 4.6 years)~~ were recruited from our hospital between May 2005 and March 2013. Sarcoidosis was diagnosed by the evidence of non - caseating epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities. Chronic sarcoidosis was defined as a disease over at least 2 years or at least two episodes in a lifetime. A cute sarcoidosis was defined as one episode of acute sarcoidosis which had totally resolved at the date of the examination. The control group consisted of 418 unrelated healthy subjects (mean ± SD age: 54.2 ± 5.3 years; 160 men and 258 women) who underwent health examinations in the Medical Examination Center of the First Affiliated Hospital of the Xinxiang Medical College (Xinxiang, China) between Oct 2009 and Sept 2013. None of the individuals in the control group had a history of lung diseases or showed any symptoms of the lung or other diseases by chest radiography or laboratory blood tests. Participants were excluded if they: were taking other prescribed medications that could affect the central nervous system; had a history of seizures, hematological diseases, or severe liver or kidney impairment; or were pregnant. ~~The control group consisted of 418 unrelated healthy subjects (mean ± SD age: 54.2 ± 5.3 year) who underwent health examinations in the Medical Examination Center of the First Affiliated Hospital of the Xinxiang Medical College (Xinxiang, China) between Oct~~

~~2009 and Sept 2013.~~ No familial relationship was known between the study participants.~~All participants were from a non - genetically related Chinese Han population in He'nan Province (China).~~ The study was performed according to the Guidelines of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang, China). Written informed consent was obtained from each participant in this study.

SNP selection

Tagging SNPs were selected from the catalogs of the International HapMap Project. The rs2789679 and rs2245168 are located in 5'-UTR, rs2236558 in intron 6, rs7067644 and rs12763624 in intron 8, rs1049550 in exon 10, rs2573351 in intron 13, rs10887581, rs11201989, rs2573353, rs2789695, rs2573356, rs2819945, rs11202059 and rs2819941 in intron 14 (Figure 1). Marker selection was done according to previous studies [4 12 14 15], and preliminary analysis was performed using the HapMap data and the following criteria. First, we examined tagSNPs in the Haploview (v4.2), using the CHB population and a minor allele frequency cut - off (MAF) ≥ 5% (HapMap Data Release 27). We found that there were a total of 29 potential tagSNPs in all. As a first screen of the most common SNPs in the sarcoidosis sample from an Eastern Chinese Han population, a MAF ≥ 20% with pair - wise tagging and $r^2 \geq 0.8$ [16] was used as the cut - off when choosing tagSNPs. Second, the LD pattern of this gene was determined in the Chinese population using the preliminary data from the HapMap. Five LD blocks across the ANXA11 were defined using Haploview's 'confidence intervals' method [17 18]. These SNPs were further analyzed in an association study.

~~Marker selection was done according to previous studies [4 12 14 15], and preliminary analysis was performed using the HapMap data and the following criteria. We examined tagSNPs in the Haploview software v4.2, using the Chinese Han in Beijing (CHB) population and a minor allele frequency cut-off (MAF) ≥ 5% (the HapMap Data Release 27). The linkage disequilibrium (LD) pattern of this gene was determined in the Chinese population using the preliminary data from the HapMap. These SNPs were further analyzed in an association study.~~

Genotyping

Peripheral blood was collected from a vein into a sterile tube coated with ethylenediamine tetraacetic acid (EDTA). Plasma samples were stored at -20°C. Genomic DNA was extracted from the frozen peripheral blood samples using a QIAmp Blood Mini Kit (Qiagen Inc., Valencia, CA, USA) according to the manufacturer's protocols. The selected SNPs were genotyped in cases and controls by using the MALDI - TOF in the MassARRAY system (Sequenom Inc., San Diego, CA, USA). Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA). The completed genotyping reactions were spotted onto a 384 well spectroCHIP (Sequenom) using the MassARRAY Nanodispenser (Sequenom) and determined by the matrix - assisted laser desorption ionization time - of - flight mass spectrometer. Genotype calling was performed in realtime with the MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer software version 3.4 (Sequenom).

Statistical analysis

All data were analyzed using the SPSS 17.0 software (SPSS Inc., Chicago, IL, USA). Hardy-Weinberg equilibrium were evaluated by Chi - square tests. Differences between the cases and controls in the frequency of the alleles, genotypes and haplotypes were evaluated by the Fisher's exact test or the Pearson Chi - square test. Unconditional logistic regression was used to calculate the odds ratio (OR) and 95% confidence interval (CI) in independent association between each locus and the presence of sarcoidosis. Generalized linear regression was used to evaluate the interaction effects between gene and gender or age. Gender and age of subjects were treated as covariants in binary logistic regression. *P* - value was calculated based on codominant, dominant for the rare allele, heterosis and recessive for the rare allele models of inheritance. The Bonferroni correction was used to adjust the test level when multiple comparisons were conducted, and the *P* value was divided by the total number of loci. Haplotype blocks were defined according to the criteria developed by Gabriel et al. [18]. Pair - wise LD statistics (D' and r^2) and haplotype frequency were calculated, and haplotype blocks were constructed using the Haploview 4.0 [17]. To ensure that the LD blocks most closely reflect the population level LD patterns, definition of the blocks were based on the control samples alone. The haplotype frequencies were estimated using GENECOUNTING, which computes maximum - likelihood estimates of haplotype frequencies from unknown phase data by utilizing an expectation - maximization algorithm [19-22]. The significance of any haplotypic association was evaluated using a likelihood ratio test, followed by permutation testing that compared estimated haplotype frequencies in cases and controls [19 21].

All data were analyzed using the SPSS 17.0 software (SPSS Inc., Chicago, IL, USA). Hardy-Weinberg equilibrium were evaluated by Chi - square tests. Differences between the cases and controls in the frequency of the alleles, genotypes and haplotypes were evaluated by the Fisher's exact test or the Pearson Chi - square test. Unconditional logistic regression was used to calculate the odds ratio (OR) and 95% confidence interval (CI) in independent association between each locus and the presence of sarcoidosis. The Bonferroni correction was used to adjust the test level when multiple comparisons were conducted, and the *P* value was divided by the total number of loci. Haplotype blocks were defined according to the criteria developed by Gabriel et al. [18]. Pair - wise LD statistics (D' and r^2) and haplotype frequency were calculated, and haplotype blocks were constructed using the Haploview 4.0 [17].

RESULTS

The genotype distribution of the fifteen polymorphisms was consistent with the Hardy - Weinberg equilibrium ($P > 0.05$). The analysis of strong LD in the patients with sarcoidosis and healthy controls revealed that 2 SNPs (rs2245168, rs2236558), 2 SNPs (rs1049550, rs2573351), 2 SNPs (rs2789695, rs2573356) and 3 SNPs (rs2819945, rs11202059, rs2819941) were located in haplotype block 1, block 2, block 3 and block 4 (D' > 0.9, Fig. 2). The genotype distribution, allelic frequencies, and haplotypes in the patients with sarcoidosis and healthy controls are showed in tables 2, table 4 and table 5.

Comparison of genotype and allele frequency distribution revealed significant differences between the patients with sarcoidosis and healthy controls for 3 SNPs:

rs2789679, rs1049550 and rs2819941. The frequency of the T allele in rs2789679 (P = 0.0003, OR = 0.70, 95%CI = 0.58 - 0.85), T allele in rs1049550 (P = 0.0007, OR = 0.62, 95%CI = 0.50 - 0.76) and the rs2819941 C allele frequency (P = 0.001, OR = 0.71, 95%CI = 0.59 - 0.87) in the patients with sarcoidosis was significantly lower than that in the controls, and these differences remained statistically significant after Bonferroni corrections. There were no interactions among rs2789679, rs1049550 and rs2819941 ($P > 0.05$).

We performed an association analysis to determine whether the haplotype was associated with the risk of sarcoidosis (Global $P < 0.05$ in block 2 and block 3). Compared with the healthy controls, the T - C haplotype occurred significantly less frequently (P = 0.001) and the C - C haplotypes occurred more frequently (P = 0.0001) in block 2 (rs1049550 - rs2573351) in the patients with sarcoidosis.

To assess particular disease phenotypes, the patients with sarcoidosis were divided into the subgroups according to their chest radiographic (CXR) stage. CXR stage I (isolated bilateral hilar lymphadenopathy): N = 196; CXR stages II - IV (infiltration of parenchyma): 167; CXR stage 0 (absence on pulmonary disease): N = 28; the information on CXR stage was not available for 21 patients. Comparison of genotype frequency distribution revealed significant differences between stage I and stages II - IV for rs1049550. The significantly more CC genotype (P = 0.012) were found in the patients with stage I sarcoidosis (table 3).

DISCUSSION

A key step in linkage and association studies is to identify common risk variants

in different populations. To determine if common risk variants exist in distinct populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene, were genotyped in samples from the patients with sarcoidosis and healthy controls in a Chinese Han population. Recently, new sarcoidosis loci have been identified by genome - wide association studies [4 14]. With the fast development of genome - wide association studies, increasing number of susceptibility loci for sarcoidosis have been reported in different populations [4 14 15]. However, these observations should be confirmed in other genetically independent populations. In this study, we conducted the first large genetic association study of the ANXA11 gene in a Chinese Han population. The evidence of markers associated with sarcoidosis was presented, and these markers were mapped to different locations in the ANXA11 gene (81897864 - 81951001). The association signals in the region were identified, and some significantly associated haplotypes also appeared this region.

Hoffman et al. reported an association between the T allele of the non - synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a German population [4]. Similar results were also obtained in other two European populations [14 15]. As a part of the sarcoidosis GWAS done in Americans [23], we further confirmed this associationin a Chinese Han population. In this study, the frequency of ANXA11 rs1049550T allele was significantly lower in the patients with sarcoidosis than that in the healthy controls. Comparison of stages II - IV, the significantly more CC genotypewas found in the patients with stages I sarcoidosis. The “protective” effect of ANXA11 T allele increased with the number of its copies in the genotype,

which is consistent with the result obtained in the patients with sarcoidosis in a German population [4], Furthermore, the T allele carriers among the patients were protected from infiltration of lung parenchyma (radiographic stages II-IV) [15]. We have also demonstrated that rs1049550 is significant cross - ethnically at the gene level after adjustment for the single SNP association tests performed. The mechanism by which rs1049550 affects the susceptibility to sarcoidosis may be that the SNP may affect the function of the ANXA11 protein rather than affect its expression. In humans, the ANXA11 protein consists of an N - terminalproline - tyrosine - glycine (PYG) - rich region, followed by four annexin core domains. The rs1049550 leads to an amino - acid exchange (basic arginine to a polar cysteine) at the evolutionarily conserved position 230 (R230C) in the first annexin domain, which is responsible for Ca^{2+} - dependent trafficking of the protein in the cell [24].

In the GWAS conducted by Hofmann et al [4], the strongest association signal was observed for rs2789679. This marker is located in the 3' - UTR of the ANXA11 gene. The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls being TT homozygous. Fine - mapping in the Black Women's Health Study (BWHS) revealed rs2819941 to be the top SNP, associated with a nonsignificant (P = 0.06) reduction in sarcoidosis risk (17% reduction per copy of the C-allele). In this case - control association study, the frequency of the T allele in rs2789679 and the rs2819941 C allele frequency in the patients with sarcoidosis was significantly lower than that in the healthy controls. Several lines of evidence suggest that the observed

associations are unlikely to be anartifact. First, both the single - SNP and the haplotype - based association analyses support the current finding. Second, population stratification is an impossible reason, because all of our samples were from the same geographical region. Finally, consistent results were obtained from two genetically independent populations (Chinese Han and Europeans). Collectively, our results confirmed the strong association between variations in the ANXA11 gene and sarcoidosis, and suggest that ANXA11 represents a strong genetic risk factor for sarcoidosis.

We further investigated the interaction among polymorphisms and observed strong LD. The haplotype analysis revealed that significantly more C - C (block 2, rs1049550 - rs2573351) and significantly fewer T - C haplotypes (block 2) were found in the patients with sarcoidosis. These results indicated that the patients with C - C (block 2) haplotypes of the ANXA11 gene were more prone to sarcoidosis. Significantly higher frequencies of T - C haplotypes were detected in the healthy controls than in the patients with sarcoidosis, suggesting that they may show a protective effect against sarcoidosis. Our sample size can detect SNP and haplotype associations with 90% and 85% power, respectively, at a false positive rate of 5%, disease prevalence of 1‰, disease allele / haplotype frequency of 0.05 / 0.03, and a presumed odds ratio (OR) of 1.5. To some extent, this finding further supports a role of ANXA11 polymorphisms in sarcoidosis, while ethnic group difference may exist.

In conclusion, our study suggests a potential role of the ANXA11 SNPs and their related haplotypes in the genetic susceptibility to sarcoidosis. Further studies are

needed to investigate how these SNPs affect the function of ANXA11.

Contributors

LG Zhang was involved in conception and design of the study, acquisition of patient data, genotyping and drafted the paper. ZJ Xiao and GC Shi were involved in design of the study, genotyping and interpretation of results. J Huang and CX Zhang were involved in design of the study, acquisition of patient data and interpretation of result. All authors revised the draft paper.

Funding

The project was supported by the Research and Development Foundation of Science and Technology of Henan Province (2012k15).

Competing interests

None.

Patient consent

Obtained.

Ethics approval

Ethics approval was provided by theFirst Hospital Affiliated to the Xinxiang Medical College.

Provenance and peer review

Not commissioned; externally peer reviewed.

Data sharing statement

No additional data are available.

Figure Legend

Fig. 1 The structure of the human ANXA11 gene and 15 SNPs located on the gene.

Fig. 2 The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the pair-wise calculation of r^2 (left) or D' (right). Black squares indicate $r^2 = 1$ (i.e. perfect LD between a pair of SNPs). Empty squares indicate D' = 1 (i.e. complete LD between a pair of SNPs).

Reference

1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007; 357:2153-65.

2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-8

3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis. *Semin Respir Crit Care Med* 2010;31:380-9.

4. Hofmann S, Franke A, Fischer A, *et al*. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet* 2008;40:1103-6.

5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev* 2002;82:331-71.

6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using recombinant human annexin XI-specific autoantibodies cloned by phage display. *J Biol Chem* 2003;278:33120-6.

7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-22.

8. Petzmann S, Maercker C, Markert E, *et al*. Enhanced proliferation and decreased

apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis Vasc Diffuse Lung Dis* 2006;23:190-200

9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, *et al*. Comparative analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol* 2006;26:243-50.

10. Agostini C, Trentin L, Facco M, *et al*. Role of IL-15, IL-2, and their receptors in the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol* 1996;157:910-8

11. Jullien D, Sieling PA, Uyemura K, *et al*. IL-15, an immunomodulator of T cell responses in intracellular infection. *J Immunol* 1997;158:800-6

12. Muller-Quernheim J, Schurmann M, Hofmann S, *et al*. Genetics of sarcoidosis. *Clin Chest Med* 2008;29:391-414.

13. Morais A, Lima B, Peixoto M, *et al*. Annexin A11 gene polymorphism (R230C variant) and sarcoidosis in a Portuguese population. *Tissue antigens* 2013;82:186-91.

14. Li Y, Pabst S, Kubisch C, *et al*. First independent replication study confirms the strong genetic association of ANXA11 with sarcoidosis. *Thorax* 2010;65:939-40.

15. Mrazek F, Stahelova A, Kriegova E, *et al*. Functional variant ANXA11 R230C: true marker of protection and candidate disease modifier in sarcoidosis. *Genes Immun* 2011;12:490-4.

16. de Bakker PI, Yelensky R, Pe'er I, *et al*. Efficiency and power in genetic association studies. *Nat Genet* 2005;37:1217-23.

17. Barrett JC, Fry B, Maller J, *et al*. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-5.

18. Gabriel SB, Schaffner SF, Nguyen H, *et al*. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-9.

19. Curtis D, Knight J, Sham PC. Program report: GENECOUNTING support programs. *Ann Hum Genet* 2006;70:277-9.

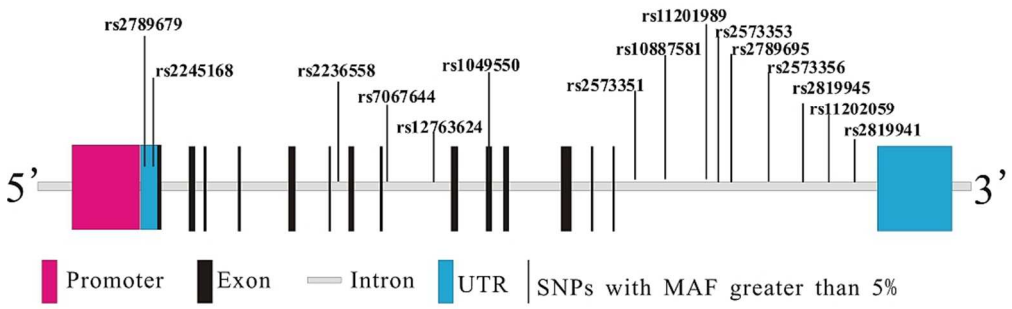
20. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic associations. *Hum Hered* 2000;50:133-9.

21. Zhao JH, Lissarrague S, Essioux L, et al. GENECOUNTING: haplotype analysis with missing genotypes. *Bioinformatics* 2002;18:1694-5

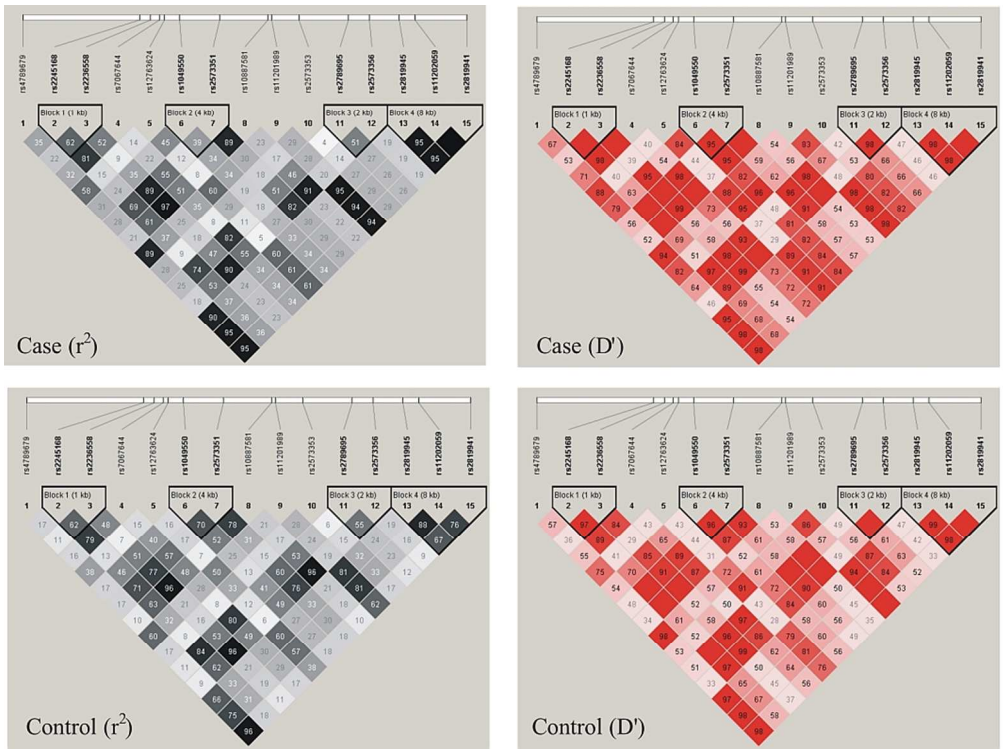
22. Zhao JH, Sham PC. Generic number systems and haplotype analysis. *Comput Methods Programs Biomed* 2003;70:1-9

23. Levin AM, Iannuzzi MC, Montgomery CG, *et al*. Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes Immun* 2013;14:13-8.

24. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11 with the nuclear envelope. *J Biol Chem* 2003;278:20210-6.



The structure of the human ANXA11 gene and 15 SNPs located on the gene.
90x27mm (300 x 300 DPI)



The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the pair-wise calculation of r2 (left) or D' (right). Black squares indicate r2 = 1 (i.e. perfect LD between a pair of SNPs). Empty squares indicate D' = 1 (i.e. complete LD between a pair of SNPs).
90x66mm (300 x 300 DPI)

BMJ Open

Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a Han Chinese population - a case-control study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-004466.R2
Article Type:	Research
Date Submitted by the Author:	27-Mar-2014
Complete List of Authors:	Zhang, LG; the First Hospital Affiliated to the Xinxiang Medical College, The Third Department of Tuberculosis Xiao, ZJ; the First Hospital Affiliated to the Xinxiang Medical College, The Third Department of Tuberculosis Shi, GC; the First Hospital Affiliated to the Xinxiang Medical College, The Third Department of Tuberculosis J, Huang; the First Hospital Affiliated to the Xinxiang Medical College, The Third Department of Tuberculosis Zhang, CX; the First Hospital Affiliated to the Xinxiang Medical College, The Third Department of Tuberculosis
Primary Subject Heading:	Infectious diseases
Secondary Subject Heading:	Genetics and genomics
Keywords:	GENETICS, IMMUNOLOGY, INFECTIOUS DISEASES

Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a Han Chinese population - a case-control study

LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang

The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China

*Corresponding author: Zhijian Xiao, The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China. No.88 Jiankang Road, Weihui City, Henan, 463000, PR China

E-mail: zhangligongys@163.com

Tel: +86 0373 - 4402229 (office)

Fax: +86 0373 -4402251

Key words: Sarcoidosis; Annexin A11; Single-nucleotide polymorphism

Word count: 3787

Abstract

Objectives: To further identify the single-nucleotide polymorphisms (SNPs) that contribute to the genetic susceptibility to sarcoidosis, we examined the potential association between sarcoidosis and fifteen SNPs of the ANXA11 gene.

Design: A case - control study.

Setting: A tuberculosist unit in a hospital of the university in China.

Participants: Participants included 412 patients with sarcoidosis and 418 healthy controls.

Methods: The selected SNPs were genotyped using the MALDI-TOF in the MassARRAY system. Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA).

Results: The results showed that statistical significant differences were found in the allelic or genotypic frequencies of the rs2789679, rs1049550 and rs2819941 in the ANXA11 gene between the patients with sarcoidosis and the controls. The frequency of the T allele in rs2789679 (P = 0.0003, OR = 0.70, 95%CI = 0.58 - 0.85) and rs1049550 (P = 0.0007, OR = 0.62, 95%CI= 0.50 - 0.76), and C allele in rs2819941 (P = 0.001, OR = 0.71, 95%CI = 0.59 - 0.87) in the patients with sarcoidosis was significantly lower than that in the controls. The strong linkage disequilibrium (LD) was observed in four blocks (D' > 0.9). In block 2 (rs1049550-rs2573351), the T - C haplotype occurred significantly less frequently (P = 0.001), whereas the C - C haplotypes occurred more frequently (P = 0.0001) in the patients with sarcoidosis than the controls. Furthermore, genotype frequency distribution revealed that, in

rs1049550, CC genotype was significantly more in the patients with chest radiographic (CXR) stage I sarcoidosis than the patients with CXR stage II-IV sarcoidosis ($P = 0.012$).

Conclusions: These findings point to a role for the polymorphisms of ANXA11 in sarcoidosis in a Chinese Han population, and may be informative for future genetic or biological studies on sarcoidosis.

Article summary

Article focus

1. The first genome-wide association study (GWAS) in sarcoidosis conducted in a German population has revealed that the rs1049550 (C > T, Arg230Cys) located in exon 10 were associated with sarcoidosis.

2. Whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations remain unclear.

3. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis in a Chinese Han population.

Key messages

1. ANXA11 rs2789679 (3'-UTR) was significantly associated with sarcoidosis.

2. ANXA11 rs1049550 (exon 10) was significantly associated with sarcoidosis.

3. Another significant association was observed for rs2819941 (intron 14).

Strengths and limitations of this study

1. A systematical screening of the functional SNPs in the promoter region, 5'- and 3'-UTR, exons of the ANXA11 gene, and the homogeneity of the study subjects representing the Chinese Han are the main strengths of this study.Chinese Han

2. The lack of data proving the positive association observed for rs2789679 and rs1049550 is a potential limitation of this study. Furthermore, the association of the serum level of ANXA11 with sarcoidosis still needs to be investigated.

INTRODUCTION

Sarcoidosis is a systemic autoimmune disease characterized by destructive, non-caseating epithelioid granulomatous lesions, with accumulated phagocytes and activated $CD4^+$ T helper type 1 lymphocytes [1 2]. The typical manifestations of sarcoidosis include bilateral hilar lymphadenopathy, pulmonary infiltration and ocular and skin lesions. The clinical course and prognosis of sarcoidosis are variable. The acute forms of sarcoidosis such as L□fgren's syndrome (LS) can resolve spontaneously in most cases. However, chronic pulmonary sarcoidosis may progress to lung fibrosis which eventually causes respiratory failure. Recent studies have demonstrated that sarcoidosis can occur in a pattern of family clustering, and its racial incidence is also different, suggesting that some genetic factors may contribute to the risk and severity of sacroidosis [1 3].

The Annexin gene family is involved in the etiology of several autoimmune and chronic diseases [4 5.] One member of the Annexin gene family, ANXA11, located on chromosome 10q22.3, is involved in calcium signaling, apoptosis, vesicle trafficking, cell growth and terminal phase of cell division [5-7]. In this context, the development and maintenance of the granulomatous inflammation in sarcoidosis have been repeatedly associated with the impaired apoptosis of activated inflammatory cells [8 9]. Most of the early-stage sarcoidosis is characterized by mononuclear cell alveolitis dominated by activated $CD4^+$ T cells and macrophages. Between these cells, the uncoordinated interplay results in the formation of typical non-caseating granulomas. The mechanism by which the granulomas resolve has not been fully elucidated.

However, it is generally assumed that the induction of apoptosis and/or by the withdrawal of inflammatory cytokines participates in the disappearance of granulomas [10 11]. In the patients with sarcoidosis, their peripheral blood monocytes were characterized by apoptosis [9].

In previous studies, case - control/family 'hypothesis - driven' studies and low density linkage scans were used to identify genetic factors conferring the genetic susceptibility to sarcoidosis [12]. Notably, the first genome - wide association study (GWAS) in sarcoidosis conducted in a German population has recently revealed an association: the leading rs2789679 SNP (single - nucleotide polymorphism) located in the 3'-untranslated region (3'-UTR) and a common nonsynonymous SNP (rs1049550, C > T, Arg230Cys) were associated with the increased risk of sarcoidosis [4 13]. This association has recently been supported by another report from the same population [14]. Thus, more studies should be performed to demonstrate the following items: whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations. The published studies about the association of sarcoidosis and ANXA11 are summarized in Supplemental table 1. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis.

In this study, we investigated fifteen loci in a Chinese population from He'nan province (China) to verify the putative association between ANXA11 polymorphisms and sarcoidosis.

SUBJECTS AND METHODS

Subjects

Four hundred and twelve patients with sarcoidosis (mean ± SD age: 53.6 ± 4.6 years; 155 men and 257 women) were recruited from our hospital between May 2005 and March 2013. Sarcoidosis was diagnosed by the evidence of non-caseating epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities. Chronic sarcoidosis was defined as a disease over at least 2 years or at least two episodes in a lifetime. A cute sarcoidosis was defined as one episode of acute sarcoidosis which had totally resolved at the date of the examination. The control group consisted of 418 unrelated healthy subjects (mean ± SD age: 54.2 ± 5.3 years; 160 men and 258 women) who underwent health examinations in the Medical Examination Center of the First Affiliated Hospital of the Xinxiang Medical College (Xinxiang, China) between Oct 2009 and Sept 2013. None of the individuals in the control group had a history of lung diseases or showed any symptoms of the lung or other diseases by chest radiography or laboratory blood tests. Participants were excluded if they: were taking other prescribed medications that could affect the central nervous system; had a history of seizures, hematological diseases, or severe liver or kidney impairment; or were pregnant. No familial relationship was known between the study participants. The study was performed according to the Guidelines of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang, China). Written informed consent was obtained from each participant in this study.

SNP selection

Tagging SNPs were selected from the catalogs of the International HapMap Project. The rs2789679 and rs2245168 are located in 5'-UTR, rs2236558 in intron 6, rs7067644 and rs12763624 in intron 8, rs1049550 in exon 10, rs2573351 in intron 13, rs10887581, rs11201989, rs2573353, rs2789695, rs2573356, rs2819945, rs11202059 and rs2819941 in intron 14 (Supplemental figure 1). Marker selection was based on the following criteria. (a) We used the CHB data from the HapMap (release 27) to select tagSNPs for ANXA11. (b) We restricted our search for tagSNPs from base pair 80155124 to 80205677. (c) Further, we limited the SNPs to tag to those with a minor allele frequency ≥ 0.2 in the CHB. (d) Based on these restrictions, there were a total of 29 SNPs. (e) Using HAPLOVIEW with an r^2 threshold ≥ 0.8, 15 tagSNP were selected and used for subsequent analyses.

Genotyping

Peripheral blood was collected from a vein into a sterile tube coated with ethylenediamine tetraacetic acid (EDTA). Plasma samples were stored at -20°C. Genomic DNA was extracted from the frozen peripheral blood samples using a QIAmp Blood Mini Kit (Qiagen Inc., Valencia, CA, USA) according to the manufacturer's protocols. The selected SNPs were genotyped in cases and controls by using the MALDI-TOF in the MassARRAY system (Sequenom Inc., San Diego, CA, USA). Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA). The completed genotyping reactions were spotted onto a 384 well spectroCHIP (Sequenom) using the MassARRAY Nanodispenser (Sequenom) and determined by the matrix - assisted laser desorption ionization time -

of - flight mass spectrometer. Genotype calling was performed in realtime with the MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer software version 3.4 (Sequenom).

Statistical analysis

All data were analyzed using the SPSS 17.0 software (SPSS Inc., Chicago, IL, USA). Hardy-Weinberg equilibrium was evaluated by Chi-square tests. Differences between the cases and controls in the frequency of the alleles, genotypes and haplotypes were evaluated by the Fisher's exact test or the Pearson Chi-square test. Unconditional logistic regression was used to calculate the odds ratio (OR) and 95% confidence interval (CI) in independent association between each locus and the presence of sarcoidosis. Logistic regression was used to evaluate the interaction effects between gene and gender or age. Gender and age of subjects were treated as covariants in binary logistic regression. *P* values were calculated based on codominant, dominant for the rare allele, heterosis and recessive for the rare allele models of inheritance. The Bonferroni correction was used to adjust the test level when multiple comparisons were conducted, and the *P* value was divided by the total number of loci. Haplotype blocks were defined according to the criteria developed by Gabriel et al. [15]. Pair-wise LD statistics (D' and r^2) and haplotype frequency were calculated, and haplotype blocks were constructed using the Haploview 4.0 [16]. To ensure that the LD blocks most closely reflect the population level LD patterns, definition of the blocks were based on the control samples alone. The haplotype frequencies were estimated using GENECOUNTING, which computes maximum -

likelihood estimates of haplotype frequencies from unknown phase data by utilizing an expectation - maximization algorithm [17-20]. The significance of any haplotypic association was evaluated using a likelihood ratio test, followed by permutation testing that compared estimated haplotype frequencies in cases and controls [17 19].

RESULTS

The genotype distribution of the fifteen polymorphisms was consistent with the Hardy-Weinberg equilibrium ($P > 0.05$). The analysis of strong LD in the patients with sarcoidosis and healthy controls revealed that 2 SNPs (rs2245168, rs2236558), 2 SNPs (rs1049550, rs2573351), 2 SNPs (rs2789695, rs2573356) and 3 SNPs (rs2819945, rs11202059, rs2819941) were located in haplotype block 1, block 2, block 3 and block 4 (D' > 0.9, Fig. 1). The genotype distribution, allelic frequencies, and haplotypes in the patients with sarcoidosis and healthy controls are showed in Tables 1-3 and Supplemental table 2

Comparison of genotype and allele frequency distribution revealed significant differences between the patients with sarcoidosis and healthy controls for 3 SNPs: rs2789679, rs1049550 and rs2819941. The frequency of the T allele in rs2789679 ($P = 0.0003$, OR = 0.70, 95%CI = 0.58 - 0.85) and rs1049550 ($P = 0.0007$, OR = 0.62, 95%CI = 0.50 - 0.76), and C allele in rs2819941 ($P = 0.001$, OR = 0.71, 95%CI = 0.59 - 0.87) in the patients with sarcoidosis was significantly lower than that in the controls, and these differences remained statistically significant after Bonferroni corrections. We presented a multi-SNP model of sarcoidosis risk. We screen out the most significant model to predict the risk by forward stepwise strategy in logistic

regression, and found that a model includes rs2789679 and rs1049550 is the best one (Supplemental table 3), which suggests that there is a interaction between them.

We performed an association analysis to determine whether the haplotype was associated with the risk of sarcoidosis (Global $P < 0.05$ in block 2 and block 3). Compared with the controls, the T - C haplotype occurred significantly less frequently ($P = 0.001$) but the C - C haplotypes occurred more frequently ($P = 0.0001$) in block 2 (rs1049550 - rs2573351) in the patients with sarcoidosis.

To assess particular disease phenotypes, the patients with sarcoidosis were divided into the subgroups according to their chest radiographic (CXR) stage. CXR stage I (isolated bilateral hilar lymphadenopathy): N = 196; CXR stages II - IV (infiltration of parenchyma): 167; CXR stage 0 (absence on pulmonary disease): N = 28; the information on CXR stage was not available for 21 patients. Genotype frequency distribution revealed that, in rs1049550, CC genotype was significantly more in the patients with chest radiographic (CXR) stage I sarcoidosis than the patients with CXR stage II-IV sarcoidosis ($P = 0.012$). (Table 4).

DISCUSSION

A key step in linkage and association studies is to identify common risk variants in different populations. To determine if common risk variants exist in distinct populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene, were genotyped in samples from the patients with sarcoidosis and healthy controls in a Chinese Han population. Recently, new sarcoidosis loci have been identified by genome-wide association studies [4 14]. With the fast development of genome - wide

association studies, an increasing number of susceptibility loci for sarcoidosis have been found in different populations [4 14 21]. However, these observations should be confirmed in other genetically independent populations. In this study, we conducted the first large genetic association study of the ANXA11 gene in a Chinese Han population. The evidence of markers associated with sarcoidosis was presented, and these markers were mapped to different locations in the ANXA11 gene (81897864 - 81951001). The association signals in the region were identified, and some significantly associated haplotypes also appeared this region.

Hoffman et al. reported an association between the T allele of the non - synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a German population [4]. Similar results were also obtained in other two European populations [14 21]. As a part of the sarcoidosis GWAS done in Americans [22], we further confirmed this association in a Chinese Han population. In this study, the frequency of ANXA11 rs1049550T allele was significantly lower in the patients with sarcoidosis than the healthy controls. There is interaction between rs2789679 and rs1049550. Genotype frequency distribution revealed that, in rs1049550, CC genotype was significantly more in the patients with stage I sarcoidosis than the patients with stage II-IV sarcoidosis. The “protective” effect of ANXA11 T allele increased with the number of its copies in the genotype, which is consistent with the result obtained in the patients with sarcoidosis in a German population [4], Furthermore, the T allele carriers among the patients were protected from infiltration of lung parenchyma (radiographic stages II-IV) [21]. We have also demonstrated that rs1049550 is

significant cross - ethnically at the gene level after adjustment for the single SNP association tests performed. The mechanism by which rs1049550 affects the susceptibility to sarcoidosis is that the SNP may affect the function of the ANXA11 protein rather than affect its expression. In humans, the ANXA11 protein consists of an N - terminalproline - tyrosine - glycine (PYG) - rich region, followed by four annexin core domains. The rs1049550 leads to an amino - acid exchange (basic arginine to a polar cysteine) at the evolutionarily conserved position 230 (R230C) in the first annexin domain, which is responsible for Ca^{2+} - dependent trafficking of the protein in the cell [23].

In the GWAS conducted by Hofmann et al [4], the strongest association signal was observed for rs2789679. This marker is located in the 3'-UTR of the ANXA11 gene. The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls being TT homozygous. Fine-mapping in the Black Women's Health Study (BWHS) revealed rs2819941 to be the top SNP, associated with a non-significant (P = 0.06) reduction in the risk of sarcoidosis (17% reduction per copy of the C-allele). In this case-control association study, the frequency of the T allele in rs2789679 and the rs2819941 C allele frequency in the patients with sarcoidosis was significantly lower than that in the healthy controls. Several lines of evidence suggest that the observed association is unlikely to be an artifact. First, both the single-SNP and the haplotype-based association analyses support the current finding. Second, our samples were from the same geographical region, excluding the . Finally, consistent results

were obtained from two genetically independent populations (Chinese Han and Europeans). Collectively, our results confirmed the strong association between ANXA11 polymorphisms and sarcoidosis, suggesting that ANXA11 represents a strong genetic risk factor for sarcoidosis.

We further investigated the interaction among polymorphisms and observed strong LD. The haplotype analysis revealed that significantly more C - C (block 2, rs1049550 - rs2573351) and significantly fewer T - C haplotypes (block 2) were found in the patients with sarcoidosis. These results indicated that the patients with C - C (block 2) haplotypes of the ANXA11 gene were more prone to sarcoidosis. Significantly higher frequencies of T - C haplotypes were detected in the healthy controls than in the patients with sarcoidosis, suggesting that they may show protective effects against sarcoidosis. Our sample size can detect SNP and haplotype associations with 90% and 85% power, respectively, at a false positive rate of 5%, disease prevalence of 1‰, disease allele / haplotype frequency of 0.05 / 0.03, and a presumed odds ratio (OR) of 1.5. To some extent, this finding further supports a role of ANXA11 polymorphisms in sarcoidosis, while ethnic group difference may exist.

In conclusion, our study suggests a potential role of the ANXA11 SNPs and their related haplotypes in the genetic susceptibility to sarcoidosis. Further studies are needed to investigate how these SNPs affect the function of ANXA11.

Contributors

LG Zhang was involved in conception and design of the study, acquisition of patient data, genotyping and drafted the paper. ZJ Xiao and GC Shi were involved in

design of the study, genotyping and interpretation of results. J Huang and CX Zhang were involved in the design of the study, acquisition of patient data and interpretation of result. All authors revised the draft paper.

Funding

The project was supported by the Research and Development Foundation of Science and Technology of Henan Province (2012k15).

Contributorship Statement

LG Zhang was involved in conception and design of the study, acquisition of patient data, genotyping and drafted the paper. ZJ Xiao and GC Shi were involved in design of the study, genotyping and interpretation of results. J Huang and CX Zhang were involved in design of the study, acquisition of patient data and interpretation of result. All authors revised the draft paper**.**

Competing interests

None.

Data sharing statement

No additional data are available.

Patient consent

Obtained.

Ethics approval

Ethics approval was provided by the First Hospital Affiliated to the Xinxiang Medical College.

Provenance and peer review

Not commissioned; externally peer reviewed.

Reference

1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007; 357:2153-65.
2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-8
3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis. *Semin Respir Crit Care Med* 2010;31:380-9.
4. Hofmann S, Franke A, Fischer A, *et al*. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet* 2008;40:1103-6.
5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev* 2002;82:331-71.
6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using recombinant human annexin XI-specific autoantibodies cloned by phage display. *J Biol Chem* 2003;278:33120-6.
7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-22.
8. Petzmann S, Maercker C, Markert E, *et al*. Enhanced proliferation and decreased apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis Vasc Diffuse Lung Dis* 2006;23:190-200
9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, *et al*. Comparative analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol*

2006;26:243-50.

10. Agostini C, Trentin L, Facco M, *et al*. Role of IL-15, IL-2, and their receptors in the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol* 1996;157:910-8

11. Jullien D, Sieling PA, Uyemura K, *et al*. IL-15, an immunomodulator of T cell responses in intracellular infection. *J Immunol* 1997;158:800-6

12. Muller-Quernheim J, Schurmann M, Hofmann S, *et al*. Genetics of sarcoidosis. *Clin Chest Med* 2008;29:391-414.

13. Morais A, Lima B, Peixoto M, *et al*. Annexin A11 gene polymorphism (R230C variant) and sarcoidosis in a Portuguese population. *Tissue antigens* 2013;82:186-91.

14. Li Y, Pabst S, Kubisch C, *et al*. First independent replication study confirms the strong genetic association of ANXA11 with sarcoidosis. *Thorax* 2010;65:939-40.

15. Gabriel SB, Schaffner SF, Nguyen H, *et al*. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-9.

16. Barrett JC, Fry B, Maller J, *et al*. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-5.

17. Curtis D, Knight J, Sham PC. Program report: GENECOUNTING support programs. *Ann Hum Genet* 2006;70:277-9.

18. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic associations. *Hum Hered* 2000;50:133-9.

19. Zhao JH, Lissarrague S, Essioux L, et al. GENECOUNTING: haplotype analysis with missing genotypes. *Bioinformatics* 2002;18:1694-5

20. Zhao JH, Sham PC. Generic number systems and haplotype analysis. *Comput Methods Programs Biomed* 2003;70:1-9

21. Mrazek F, Stahelova A, Kriegova E, *et al*. Functional variant ANXA11 R230C: true marker of protection and candidate disease modifier in sarcoidosis. *Genes Immun* 2011;12:490-4.

22. Levin AM, Iannuzzi MC, Montgomery CG, *et al*. Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes Immun* 2013;14:13-8.

24. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11 with the nuclear envelope. *J Biol Chem* 2003;278:20210-6.

Figure Legend

Fig. 1 The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the pair-wise calculation of r^2 (left) or D' (right). Black squares indicate $r^2 = 1$ (i.e. perfect LD between a pair of SNPs). Empty squares indicate D' = 1 (i.e. complete LD between a pair of SNPs).

Supplemental Fig. 1 The structure of the human ANXA11 gene and 15 SNPs located on the gene.

Table 1 Genotypic and allelic frequencies of ANXA11 polymorphisms in the controls and patients with sarcoidosis

Variable	**Controls** (n=418)		**Sarcoidosis** (n=412)		*P*-value[a]	OR, 95% CI
	No.	%	No.	%		
rs2789679					0.0007	
AA	79	18.9	132	32.0	0.0002	0.49, 0.36-0.68
AT	225	53.8	186	45.2	0.012	1.42, 1.08-1.87
TT	114	27.3	94	22.8	0.140	1.27, 0.93-1.74
Per T allele	383	45.8	450	54.6	0.0003	0.70, 0.58-0.85
rs1049550					0.0002	
CC	154	36.8	208	50.5	0.0007	0.57, 0.43-0.75
CT	190	45.5	170	41.3	0.221	1.19, 0.90-1.56
TT	74	17.7	34	8.3	0.0008	2.39, 1.55-3.68
Per T allele	338	40.4	238	28.8	0.0007	0.60, 0.49-0.73
rs2819941					0.0002	
CC	114	27.3	94	22.8	0.141	1.27, 0.93-1.74
CT	226	54.1	190	46.1	0.021	1.38, 1.05-1.81
TT	78	18.7	128	31.1	0.0004	0.51, 0.37-0.70
Per T allele	382	45.7	446	54.1	0.001	0.71, 0.59-0.87

[a] *P*-value was calculated by 2×3 and 2×2 chi-squared tests based on codominant, dominant for the rare allele, heterosis and recessive for the rare allele models of inheritance.

Alpha value is adjusted by Bonferroni correction and statistically significant results (P<0.003)

Table 2 ANXA11 haplotype in block 2 frequencies and the results of their associations with risk of sarcoidosis

Haplotype			Genecounting (frequency %)			
ID	**rs1049550**	**rs2573351**	**Cases**	**Controls**	***P*-value[a]**	**Global *P*[b]**
HAP1	T	C	27.7	39.7	0.001	0.003
HAP2	C	C	19.9	7.4	0.0001	
HAP3	C	T	50.5	51.0	0.892	

[a] Based on 10,000 permutations.

[b] Based on comparison of frequency distribution of all haplotypes for the combination of SNPs.

Table 3 ANXA11 haplotype in block 3 frequencies and the results of their associations with risk of sarcoidosis

Haplotype			Genecounting (frequency %)			
ID	**rs2789695**	**rs2573356**	**Cases**	**Controls**	***P*-value[a]**	**Global *P*[b]**
HAP1	C	T	33.7	34.9	0.718	0.045
HAP2	T	T	15.3	14.1	0.632	
HAP2	T	T	15.6	14.1	0.565	

[a] Based on 10,000 permutations.

[b] Based on comparison of frequency distribution of all haplotypes for the combination of SNPs.

Table 4 Chest radiographic (CXR) stages of sarcoidosis patients by genotype of rs2789679, rs1049550 and rs2819941

Stages	rs2789679 (n, %)			rs1049550* (n, %)			rs2819941 (n, %)		
	AA	AT	TT	CC	CT	TT	CC	CT	TT
stage I	59 (30.1)	92 (47.0)	45 (23.0)	101 (51.5)	82 (41.8)	13 (6.6)	47 (25.0)	88 (44.9)	61 (31.1)
stages II - IV	46 (27.6)	75 (44.9)	46 (27.6)	72 (43.1)	68 (40.7)	27 (16.2)	41 (24.6)	73 (43.7)	53 (31.7)
χ^2, *P*	1.041, 0.594			8.807, 0.012			0.052, 0.975		

* *P* values for genotype frequency distribution ($P < 0.05$).

1

2

3

4

5

6

7

Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a Han Chinese population - a case-control study

LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang

The Third Department of Tuberculosis~~Internal Medicine~~, the First Hospital Affiliated to the Xinxiang Medical College~~, Tuberculosist, First Hospital Affiliated to the Xinxiang Medical College~~, Xinxiang, Henan, PR China

*Corresponding author: Zhijian Xiao, The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China.~~Internal Medicine, Tuberculosisthe, First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China~~

No.88 Jiankang Road, Weihui City, Henan, 463000, PR China

E-mail: zhangligongys@163.com

Tel: +86 0373 - 4402229 (office)

Fax: +86 0373 -4402251

Key words: Sarcoidosis; Annexin A11; Single-nucleotide polymorphism

Word count: 2909

Abstract

Objectives: To further identify the single-nucleotide polymorphisms (SNPs) that contribute to the genetic susceptibility to sarcoidosis, we examined the potential association between sarcoidosis and fifteen SNPs of the ANXA11 gene.

Design: A case - control study.

Setting: A tuberculosist unit in a ~~university~~ hospital of the university in China.

Participants: Participants included 412 patients with sarcoidosis and 418 healthy controls.

Methods: The selected SNPs were genotyped ~~in cases and controls by~~ using the MALDI-TOF in the MassARRAY system (Sequenom Inc., San Diego, CA, USA). Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA).

Results: The results showed that statistical~~statistically~~ significant differences were ~~observed~~ found in the allelic or genotypic frequencies of the rs2789679, rs1049550 and rs2819941 in the ANXA11 gene between the patients with sarcoidosis and the controls. The frequency of the T allele in rs2789679 (P = 0.0003, OR = 0.70, 95%CI = 0.58 - 0.85)~~, T allele in~~ and rs1049550 (P = 0.0007, OR = 0.62, 95%CI= 0.50 - 0.76), and C allele ~~the~~ in rs2819941 ~~C allele frequency~~ (P = 0.001, OR = 0.71, 95%CI = 0.59 - 0.87) in the patients with sarcoidosis was significantly lower than that in the controls. ~~Furthermore,~~ The strong linkage disequilibrium (LD) was observed in four blocks (D' > 0.9). In block 2 (rs1049550-rs2573351), the T - C haplotype occurred significantly less frequently (P = 0.001), whereas the C - C haplotypes occurred more

frequently (P = 0.0001) in the patients with sarcoidosis than the controls. Furthermore, ~~c Comparison of~~ genotype frequency distribution ~~–~~revealed that, in rs1049550, CC genotype was significantly more in the patients with chest radiographic (CXR) stage I sarcoidosis than the patients with CXR stage II-IV sarcoidosis (P = 0.012). ~~significant differences between chest radiographic (CXR) stage I and stages II - IV for rs1049550. The significantly more CC genotype (P = 0.012) were found in the patients with stages I sarcoidosis.~~

~~–~~**Conclusions**: These findings point to a role for the polymorphisms of ANXA11 ~~gene polymorphisms~~ in sarcoidosis in a ~~Han Chinese~~Chinese Han population, and may be informative for future genetic or biological studies on sarcoidosis.

Article summary

Article focus

1. The first genome-wide association study (GWAS) in sarcoidosis conducted in a German population has revealed that the rs1049550 (C > T, Arg230Cys) located in exon 10 were associated with sarcoidosis.

2. More studies should be performed to demonstrate the following item: wWhether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations remain unclear.

3. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis in a Chinese Han population.

Key messages

1. ANXA11 rs2789679 (3'UTR3'-UTR) was found to be associatedwas significantly associated –with sarcoidosis.

2. ANXA11 rs1049550 (exon 10) was significantly associated with sarcoidosis.

3. Another significant association was observed for rs2819941 (intron 14).

Strengths and limitations of this study

1. A systematical screening of the Ffunctional SNPs in the promoter region, 5'- and 3'-UTR, exons of the ANXA11 gene, and the homogeneity of the study subjects representing the Chinese Han are the main strengths of this study.were systematically screened and homogeneity of the study subjects, representing the Han ChineseChinese Han , is the main strength of the current study.

2. The lack of data proving the positive association observed for rs2789679 and

rs1049550 is a potential limitation of this study. Furthermore, the association of the serum level of ANXA11 with sarcoidosis still needs to be investigated. ~~Functional characterization is the potential limitation of the study that could have further helped in proving the positive association observed for rs2789679 and rs1049550. The lack of correlation of serum ANXA11 levels are needed to investigate.~~

INTRODUCTION

Sarcoidosis is a systemic autoimmune disease characterized by destructive, non-caseating epithelioid granulomatous lesions, with accumulated phagocytes and activated $CD4^+$ T helper type 1 lymphocytes [1 2]. The typical manifestations of sarcoidosis include bilateral hilar lymphadenopathy, pulmonary infiltration and ocular and skin lesions. The clinical course and prognosis of sarcoidosis are variable. The acute forms of sarcoidosis such as Löfgren's syndrome (LS) can resolve spontaneously in most cases. However, chronic pulmonary sarcoidosis may progress to lung fibrosis which ~~and~~ eventually ~~result in~~causes respiratory failure. Recent studies have demonstrated that sarcoidosis can occur in a pattern of family clustering, and its racial incidence is also different, suggesting that some genetic factors may contribute to the risk and severity of sacroidosis [1 3].

The Annexin gene family is involved in the etiology of several autoimmune and chronic diseases [4 5]. One member of the Annexin gene family, ANXA11, located on chromosome 10q22.3, is involved in calcium signaling, apoptosis, vesicle trafficking, cell growth and terminal phase of cell division [5-7]. In this context, the development and maintenance of the granulomatous inflammation in sarcoidosis have been

repeatedly associated with the impaired apoptosis of activated inflammatory cells [8 9]. ~~In most patients with sarcoidosis~~Most of the early-stage sarcoidosis~~, the early stages of this disease are~~ is characterized by mononuclear cell alveolitis dominated by activated $CD4^+$ T cells and macrophages. Between these cells, the uncoordinated interplay results in the formation of typical non~~-~~caseating granulomas. The mechanism by which the granulomas resolve has not been fully elucidated. However, it is generally assumed that the induction of apoptosis and~~/~~/or by the withdrawal of inflammatory cytokines participates in the disappearance of granulomas [10 11]. In the patients with sarcoidosis, their peripheral blood monocytes were characterized by apoptosis [9].

In previous studies, case - control~~/~~/family 'hypothesis - driven' studies and low density linkage scans were used to identify genetic factors conferring the genetic susceptibility to sarcoidosis [12]. Notably, the first genome - wide association study (GWAS) in sarcoidosis conducted in a German population has recently revealed an association: the leading rs2789679 SNP (single - nucleotide polymorphism) located in the 3'~~-~~untranslated region (3'~~-~~UTR) and a common nonsynonymous SNP (rs1049550, C > T, Arg230Cys) were associated with the increased risk of sarcoidosis [4 13]. This association has recently been supported by another report from the same population [14]. Thus, more studies should be performed to demonstrate the following items: whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations. The published studies about the association of sarcoidosis and ANXA11~~ association~~

studies for ANXA11 are summarized in Supplemental table 21. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis.

In this study, we investigated fifteen loci in a Chinese population from He'nan province (China) to verify the putative association between ANXA11 polymorphisms and sarcoidosis.

SUBJECTS AND METHODS

Subjects

Four hundred and twelve patients with sarcoidosis (mean ± SD age: 53.6 ± 4.6 years; 155 men and 257 women) were recruited from our hospital between May 2005 and March 2013. Sarcoidosis was diagnosed by the evidence of non-caseating epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities. Chronic sarcoidosis was defined as a disease over at least 2 years or at least two episodes in a lifetime. A cute sarcoidosis was defined as one episode of acute sarcoidosis which had totally resolved at the date of the examination. The control group consisted of 418 unrelated healthy subjects (mean ± SD age: 54.2 ± 5.3 years; 160 men and 258 women) who underwent health examinations in the Medical Examination Center of the First Affiliated Hospital of the Xinxiang Medical College (Xinxiang, China) between Oct 2009 and Sept 2013. None of the individuals in the control group had a history of lung diseases or showed any symptoms of the lung or other diseases by chest radiography or laboratory blood tests. Participants were excluded if they: were taking other prescribed medications that could affect the

central nervous system; had a history of seizures, hematological diseases, or severe liver or kidney impairment; or were pregnant. No familial relationship was known between the study participants. The study was performed according to the Guidelines of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang, China). Written informed consent was obtained from each participant in this study.

SNP selection

——Tagging SNPs were selected from the catalogs of the International HapMap Project. The rs2789679 and rs2245168 are located in 5'-UTR, rs2236558 in intron 6, rs7067644 and rs12763624 in intron 8, rs1049550 in exon 10, rs2573351 in intron 13, rs10887581, rs11201989, rs2573353, rs2789695, rs2573356, rs2819945, rs11202059 and rs2819941 in intron 14 (Supplemental figure 1). Marker selection was based on the following criteria. (a) We used the CHB data from the HapMap (release 27) to select tagSNPs for ANXA11. (b) We restricted our search for tagSNPs from base pair 80155124 to 80205677. (c) Further, we limited the SNPs to tag to those with a minor allele frequency ≥ 0.2 in the CHB. (d) Based on these restrictions, there were a total of 29 SNPs. (e) Using HAPLOVIEW with an r^2 threshold ≥ 0.8, 15 tagSNP were selected and used for subsequent analyses.

~~Marker selection was done according to previous studies [4 12 14 15], and preliminary analysis was performed using the HapMap data and the following criteria. First, we examined tagSNPs in the Haploview (v4.2), using the CHB population and a minor allele frequency cut - off (MAF) ≥ 5% (HapMap Data Release 27). We found that there were a total of 29 potential tagSNPs in all. As a first screen of the most common~~

~~SNPs in the sarcoidosis sample from an Eastern Chinese Han population, a MAF ≥ 20% with pair - wise tagging and r^2 ≥ 0.8 [16] was used as the cut - off when choosing tagSNPs. Second, the LD pattern of this gene was determined in the Chinese population using the preliminary data from the HapMap. Five LD blocks across the ANXA11 were defined using Haploview's 'confidence intervals' method [17,18]. These SNPs were further analyzed in an association study.~~

Genotyping

Peripheral blood was collected from a vein into a sterile tube coated with ethylenediamine tetraacetic acid (EDTA). Plasma samples were stored at -20°C. Genomic DNA was extracted from the frozen peripheral blood samples using a QIAmp Blood Mini Kit (Qiagen Inc., Valencia, CA, USA) according to the manufacturer's protocols. The selected SNPs were genotyped in cases and controls by using the MALDI-TOF in the MassARRAY system (Sequenom Inc., San Diego, CA, USA). Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA). The completed genotyping reactions were spotted onto a 384 well spectroCHIP (Sequenom) using the MassARRAY Nanodispenser (Sequenom) and determined by the matrix - assisted laser desorption ionization time - of - flight mass spectrometer. Genotype calling was performed in realtime with the MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer software version 3.4 (Sequenom).

Statistical analysis

All data were analyzed using the SPSS 17.0 software (SPSS Inc., Chicago, IL,

USA). Hardy-Weinberg equilibrium ~~were~~ was evaluated by Chi-square tests. Differences between the cases and controls in the frequency of the alleles, genotypes and haplotypes were evaluated by the Fisher's exact test or the Pearson Chi-square test. Unconditional logistic regression was used to calculate the odds ratio (OR) and 95% confidence interval (CI) in independent association between each locus and the presence of sarcoidosis. Logistic regression~~Generalized linear regression~~ was used to evaluate the interaction effects between gene and gender or age. Gender and age of subjects were treated as covariants in binary logistic regression. *P*- values ~~was~~ were calculated based on codominant, dominant for the rare allele, heterosis and recessive for the rare allele models of inheritance. The Bonferroni correction was used to adjust the test level when multiple comparisons were conducted, and the *P* value was divided by the total number of loci. Haplotype blocks were defined according to the criteria developed by Gabriel et al. [15]. Pair-wise LD statistics (D' and r^2) and haplotype frequency were calculated, and haplotype blocks were constructed using the Haploview 4.0 [16]. To ensure that the LD blocks most closely reflect the population level LD patterns, definition of the blocks were based on the control samples alone. The haplotype frequencies were estimated using GENECOUNTING, which computes maximum - likelihood estimates of haplotype frequencies from unknown phase data by utilizing an expectation - maximization algorithm [17-20]. The significance of any haplotypic association was evaluated using a likelihood ratio test, followed by permutation testing that compared estimated haplotype frequencies in cases and controls [17 19].

RESULTS

The genotype distribution of the fifteen polymorphisms was consistent with the Hardy—-Weinberg equilibrium ($P > 0.05$). The analysis of strong LD in the patients with sarcoidosis and healthy controls revealed that 2 SNPs (rs2245168, rs2236558), 2 SNPs (rs1049550, rs2573351), 2 SNPs (rs2789695, rs2573356) and 3 SNPs (rs2819945, rs11202059, rs2819941) were located in haplotype block 1, block 2, block 3 and block 4 (D' > 0.9, Fig. ~~2~~1). The genotype distribution, allelic frequencies, and haplotypes in the patients with sarcoidosis and healthy controls are showed in ~~tables~~ Tables ~~2~~1, -3 and Supplemental table 2~~table 4 and table 5~~.

Comparison of genotype and allele frequency distribution revealed significant differences between the patients with sarcoidosis and healthy controls for 3 SNPs: rs2789679, rs1049550 and rs2819941. The frequency of the T allele in rs2789679 ($P = 0.0003$, OR = 0.70, 95%CI = 0.58 - 0.85)~~, T allele~~ and~~in~~ rs1049550 ($P = 0.0007$, OR = 0.62, 95%CI = 0.50 - 0.76), and C allele in rs2819941~~the rs2819941 C allele frequency~~ ($P = 0.001$, OR = 0.71, 95%CI = 0.59 - 0.87) in the patients with sarcoidosis was significantly lower than that in the controls, and these differences remained statistically significant after Bonferroni corrections. We presented a multi-SNP model of sarcoidosis risk. We screen out the most significant model to predict the risk by forward stepwise strategy in logistic regression, and found that a model includes rs2789679 and rs1049550 is the best one (Supplemental table 3) , which suggests that there is a interaction between them.

~~There were no interactions among rs2789679, rs1049550 and rs2819941 ($P > 0.05$).~~

We performed an association analysis to determine whether the haplotype was associated with the risk of sarcoidosis (Global $P < 0.05$ in block 2 and block 3). Compared with the ~~healthy~~ controls, the T - C haplotype occurred significantly less frequently ($P = 0.001$) ~~and~~ but the C - C haplotypes occurred more frequently ($P = 0.0001$) in block 2 (~~-~~rs1049550 - rs2573351) in the patients with sarcoidosis.~~—~~

To assess particular disease phenotypes, the patients with sarcoidosis were divided into the subgroups according to their chest radiographic (CXR) stage. CXR stage I (isolated bilateral hilar lymphadenopathy): N = 196; CXR stages II - IV (infiltration of parenchyma): 167; CXR stage 0 (absence on pulmonary disease): N = 28; the information on CXR stage was not available for 21 patients. Genotype frequency distribution revealed that, in rs1049550, CC genotype was significantly more in the patients with chest radiographic (CXR) stage I sarcoidosis than the patients with CXR stage II-IV sarcoidosis ($P = 0.012$).~~Comparison of genotype frequency distribution revealed significant differences between stage I and stages II - IV for rs1049550. The significantly more CC genotype ($P = 0.012$) were found in the patients with stage I sarcoidosis~~ (~~table~~ Table ~~3~~4).

DISCUSSION

A key step in linkage and association studies is to identify common risk variants in different populations. To determine if common risk variants exist in distinct populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene, were genotyped in samples from the patients with sarcoidosis and healthy controls in a Chinese Han population. Recently, new sarcoidosis loci have been identified by

genome-wide association studies [4 14]. With the fast development of genome - wide association studies, an increasing number of susceptibility loci for sarcoidosis have been ~~reported~~ found in different populations [4 14 21]. However, these observations should be confirmed in other genetically independent populations. In this study, we conducted the first large genetic association study of the ANXA11 gene in a Chinese Han population. The evidence of markers associated with sarcoidosis was presented, and these markers were mapped to different locations in the ANXA11 gene (81897864 - 81951001). The association signals in the region were identified, and some significantly associated haplotypes also appeared this region.

Hoffman et al. reported an association between the T allele of the non - synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a German population [4]. Similar results were also obtained in other two European populations [14 21]. As a part of the sarcoidosis GWAS done in Americans [22], we further confirmed this association in a Chinese Han population. In this study, the frequency of ANXA11 rs1049550T allele was significantly lower in the patients with sarcoidosis than the healthy controls. There is interaction between rs2789679 and rs1049550. Genotype frequency distribution revealed that, in rs1049550, CC genotype was significantly more in the patients with stage I sarcoidosis than the patients with stage II-IV sarcoidosis~~Comparison of stages II - IV, the significantly more CC genotypewas found in the patients with stages I sarcoidosis~~. The "protective" effect of ANXA11 T allele increased with the number of its copies in the genotype, which is consistent with the result obtained in the patients with sarcoidosis in a German population [4],

Furthermore, the T allele carriers among the patients were protected from infiltration of lung parenchyma (radiographic stages II-IV) [21]. We have also demonstrated that rs1049550 is significant cross - ethnically at the gene level after adjustment for the single SNP association tests performed. The mechanism by which rs1049550 affects the susceptibility to sarcoidosis ~~may be~~is that the SNP may affect the function of the ANXA11 protein rather than affect its expression. In humans, the ANXA11 protein consists of an N - terminalproline - tyrosine - glycine (PYG) - rich region, followed by four annexin core domains. The rs1049550 leads to an amino - acid exchange (basic arginine to a polar cysteine) at the evolutionarily conserved position 230 (R230C) in the first annexin domain, which is responsible for Ca^{2+} - dependent trafficking of the protein in the cell [23].

In the GWAS conducted by Hofmann et al [4], the strongest association signal was observed for rs2789679. This marker is located in the 3'-UTR of the ANXA11 gene. The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls being TT homozygous. Fine-mapping in the Black Women's Health Study (BWHS) revealed rs2819941 to be the top SNP, associated with a non-significant (P = 0.06) reduction in the risk of sarcoidosis ~~risk~~ (17% reduction per copy of the C-allele). In this case-control association study, the frequency of the T allele in rs2789679 and the rs2819941 C allele frequency in the patients with sarcoidosis was significantly lower than that in the healthy controls. Several lines of evidence suggest that the observed associations ~~are~~is unlikely to be an artifact. First, both the single-SNP and

the haplotype-based association analyses support the current finding. Second, our samples were from the same geographical region, excluding the ~~population stratification is an impossible reason, because all of our samples were from the same geographical region~~. Finally, consistent results were obtained from two genetically independent populations (Chinese Han and Europeans). Collectively, our results confirmed the strong association between ~~variations in the~~ ANXA11 polymorphisms ~~gene~~ and sarcoidosis, ~~and~~ suggesting that ANXA11 represents a strong genetic risk factor for sarcoidosis.

We further investigated the interaction among polymorphisms and observed strong LD. The haplotype analysis revealed that significantly more C - C (block 2, rs1049550 - rs2573351) and significantly fewer T - C haplotypes (block 2) were found in the patients with sarcoidosis. These results indicated that the patients with C - C (block 2) haplotypes of the ANXA11 gene were more prone to sarcoidosis. Significantly higher frequencies of T - C haplotypes were detected in the healthy controls than in the patients with sarcoidosis, suggesting that they may show ~~a~~ protective effects against sarcoidosis. Our sample size can detect SNP and haplotype associations with 90% and 85% power, respectively, at a false positive rate of 5%, disease prevalence of 1‰, disease allele / haplotype frequency of 0.05 / 0.03, and a presumed odds ratio (OR) of 1.5. To some extent, this finding further supports a role of ANXA11 polymorphisms in sarcoidosis, while ethnic group difference may exist.

In conclusion, our study suggests a potential role of the ANXA11 SNPs and their related haplotypes in the genetic susceptibility to sarcoidosis. Further studies are

needed to investigate how these SNPs affect the function of ANXA11.

Contributors

LG Zhang was involved in conception and design of the study, acquisition of patient data, genotyping and drafted the paper. ZJ Xiao and GC Shi were involved in design of the study, genotyping and interpretation of results. J Huang and CX Zhang were involved in the design of the study, acquisition of patient data and interpretation of result. All authors revised the draft paper.

Funding

The project was supported by the Research and Development Foundation of Science and Technology of Henan Province (2012k15).

Competing interests

None.

Patient consent

Obtained.

Ethics approval

Ethics approval was provided by the First Hospital Affiliated to the Xinxiang Medical College.

Provenance and peer review

Not commissioned; externally peer reviewed.

Data sharing statement

No additional data are available.

Figure Legend

~~**Fig. 1** The structure of the human ANXA11 gene and 15 SNPs located on the gene.~~

Fig. ~~2~~ 1 The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the pair-wise calculation of r^2 (left) or D' (right). Black squares indicate $r^2 = 1$ (i.e. perfect LD between a pair of SNPs). Empty squares indicate D' = 1 (i.e. complete LD between a pair of SNPs).

Supplemental Fig. 1~~**Fig. 1**~~ The structure of the human ANXA11 gene and 15 SNPs located on the gene.

Reference

1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007; 357:2153-65.
2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-8
3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis. *Semin Respir Crit Care Med* 2010;31:380-9.
4. Hofmann S, Franke A, Fischer A, *et al*. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet* 2008;40:1103-6.
5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev* 2002;82:331-71.
6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using recombinant human annexin XI-specific autoantibodies cloned by phage display. *J Biol Chem* 2003;278:33120-6.

7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-22.

8. Petzmann S, Maercker C, Markert E, *et al*. Enhanced proliferation and decreased apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis Vasc Diffuse Lung Dis* 2006;23:190-200

9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, *et al*. Comparative analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol* 2006;26:243-50.

10. Agostini C, Trentin L, Facco M, *et al*. Role of IL-15, IL-2, and their receptors in the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol* 1996;157:910-8

11. Jullien D, Sieling PA, Uyemura K, *et al*. IL-15, an immunomodulator of T cell responses in intracellular infection. *J Immunol* 1997;158:800-6

12. Muller-Quernheim J, Schurmann M, Hofmann S, *et al*. Genetics of sarcoidosis. *Clin Chest Med* 2008;29:391-414.

13. Morais A, Lima B, Peixoto M, *et al*. Annexin A11 gene polymorphism (R230C variant) and sarcoidosis in a Portuguese population. *Tissue antigens* 2013;82:186-91.

14. Li Y, Pabst S, Kubisch C, *et al*. First independent replication study confirms the strong genetic association of ANXA11 with sarcoidosis. *Thorax* 2010;65:939-40.

15. Gabriel SB, Schaffner SF, Nguyen H, *et al*. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-9.

16. Barrett JC, Fry B, Maller J, *et al*. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-5.

17. Curtis D, Knight J, Sham PC. Program report: GENECOUNTING support programs. *Ann Hum Genet* 2006;70:277-9.

18. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic associations. *Hum Hered* 2000;50:133-9.

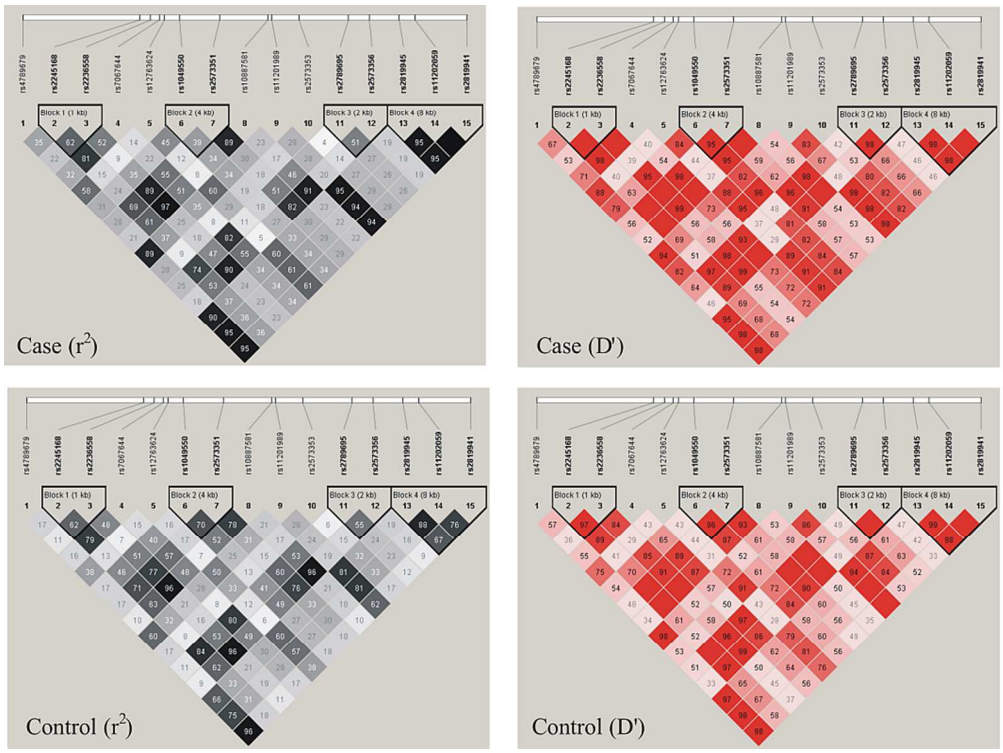
19. Zhao JH, Lissarrague S, Essioux L, et al. GENECOUNTING: haplotype analysis with missing genotypes. *Bioinformatics* 2002;18:1694-5

20. Zhao JH, Sham PC. Generic number systems and haplotype analysis. *Comput Methods Programs Biomed* 2003;70:1-9

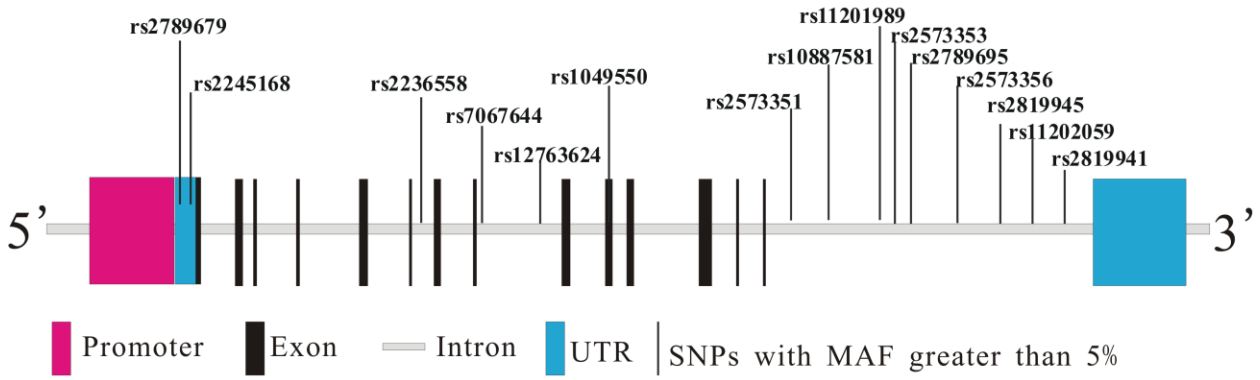
21. Mrazek F, Stahelova A, Kriegova E, *et al*. Functional variant ANXA11 R230C: true marker of protection and candidate disease modifier in sarcoidosis. *Genes Immun* 2011;12:490-4.

22. Levin AM, Iannuzzi MC, Montgomery CG, *et al*. Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes Immun* 2013;14:13-8.

24. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11 with the nuclear envelope. *J Biol Chem* 2003;278:20210-6.



90x66mm (300 x 300 DPI)



Supplemental table 1 Summary of all published sarcoidosis-association studies for ANXA11

Study	Population		Type of Study	Sample Size (n)		Number of SNPs typed	Positive SNPs
	Ethnic group	Country		Sarcoidosis	Control		
Hofmann S, 2008	Caucasian	Germany	Case-control	499	490	GWAS	rs2789679, rs7091565, rs1049550
Mrazek F, 2011	Caucasian	Czech	Case-control	245	254	1	rs1049550
Cozier YC, 2012	Caucasian	America	Case-control	486	943	10p12-10q22	rs2789679, rs2819941
Li Y, 2013	Caucasian	Germany	Case-control	325	364	1	rs1049550
Levin AM, 2013	Caucasian	America	Case-control	1689	1252	25	rs1049550
Morais A, 2013	Caucasian	Portugal	Case-control	208	197	1	rs1049550

Supplemental Table 2 Genotypic and allelic frequencies of ANXA11 polymorphisms in the controls and patients with sarcoidosis

Variable	**Controls** (n=418)		**Sarcoidosis** (n=412)		*P*-value[a]	OR, 95% CI
	No.	%	No.	%		
rs2245168					0.920	
CC	156	37.3	150	36.4	0.791	1.04, 0.74-1.38
CT	200	47.9	197	47.8	0.955	0.99, 0.76-1.30
TT	62	14.8	65	15.8	0.781	0.95, 0.65-1.39
Per T allele	324	38.8	327	39.7	0.698	1.04, 0.85-1.27
rs2236558					0.172	
GG	118	28.2	104	25.2	0.72	1.06, 0.77-1.45
TG	190	45.5	214	51.9	0.060	0.77, 0.59-1.01
TT	110	26.3	94	22.8	0.073	1.33, 0.97-1.82
Per T allele	410	49.0	402	48.8	0.917	0.99, 0.82-1.20
rs7067644					0.173	
AA	152	36.4	176	42.7	0.059	0.76, 0.58-1.01
GA	206	49.3	182	44.2	0.137	1.23, 0.94-1.68
GG	60	14.4	54	13.1	0.597	1.11, 0.75-1.65
Per G allele	326	38.0	290	35.2	0.109	0.85, 0.70-1.04
rs12763624					0.118	
TT	126	75.6	150	36.4	0.008	0.68, 0.51-0.91
CT	216	51.7	186	45.2	0.059	1.30, 0.99-1.71
CC	76	18.2	76	18.5	0.411	1.16, 0.81-1.67
Per C allele	368	44.0	338	41.0	0.216	0.88, 0.73-1.08
rs2573351					0.437	

TT	120	28.7	104	25.2	0.263	1.19, 0.88-1.62
CT	202	48.3	216	52.4	0.238	0.85, 0.65-1.12
CC	96	23.0	92	22.3	0.826	1.04, 0.75-1.44
Per C allele	394	47.1	400	48.5	0.564	1.06, 0.87-1.28
rs10887581					0.290	
TT	114	27.3	96	23.3	0.190	1.23, 0.90-1.69
TC	190	45.5	208	50.5	0.148	0.82, 0.62-1.07
CC	114	27.3	108	26.2	0.729	1.06, 0.78-1.44
Per C allele	418	50.0	424	51.5	0.583	1.06,0.87-1.29
rs11201989					0.144	
GG	130	31.1	128	31.1	0.995	1.00, 0.75-1.34
GC	218	52.2	194	47.1	0.141	1.23, 0.93-1.62
CC	70	16.8	90	21.9	0.060	0.71, 0.50-1.01
Per C allele	358	42.8	374	45.4	0.29	1.11, 0.91-1.35
rs2573353					0.918	
CC	178	42.6	176	42.7	0.957	0.99, 0.75-1.31
CA	196	46.9	192	46.6	0.921	1.01, 0.77-1.33
AA	44	10.5	44	10.7	0.941	0.98, 0.63-1.53
Per A allele	284	34.0	280	34.0	0.997	1.00, 0.82-1.23
rs2789695					0.918	
TT	178	42.6	180	34.0	0.740	0.95, 0.73-1.26
CT	188	45.0	184	44.7	0.920	1.01, 0.77-1.33
CC	52	12.4	48	11.7	0.73	1.08, 0.71-1.64
Per C allele	292	34.9	280	34.0	0.69	0.96, 0.78-1.17
rs2573356					0.892	
CC	118	28.2	112	27.2	0.742	1.05, 0.78-1.43

CT	190	45.5	194	47.1	0.639	0.94, 0.71-1.23
TT	110	26.3	106	25.7	0.843	1.03, 0.76-1.41
Per T allele	410	49.0	406	49.3	0.926	0.99, 0.82-1.20
rs2819945					0.101	
GG	119	28.5	130	31.6	0.294	0.85, 0.63-1.15
GA	221	52.9	188	45.6	0.030	1.35, 1.03-1.78
AA	78	18.7	94	22.8	0.138	0.78, 0.55-1.09
Per A allele	377	45.1	376	45.6	0.827	1.02, 0.84-1.24
rs11202059					0.064	
GG	106	25.4	128	31.1	0.067	0.75, 0.56-1.02
GA	226	54.1	190	46.1	0.021	1.38, 1.05-1.81
AA	86	20.6	94	22.8	0.429	0.88, 0.63-1.22
Per A allele	398	47.6	378	45.9	0.479	0.93, 0.77-1.13

[a] *P*-value was calculated by 2×3 and 2×2 chi-squared tests based on codominant, dominant for the rare allele, heterosis and recessive for the rare allele models of inheritance.

Alpha value is adjusted by Bonferroni correction and statistically significant results (P<0.003)

Supplemental table 3 Comparison of logistic regression models with and without Genotypes for rs2789679 and rs1049550

Model	Log Likelihood	Df	*P* value
Constant + rs2789679	575.3	1	< 0.001
Constants+ rs2789679+rs1049550	565.8	2	0.001

BMJ Open

Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a Han Chinese population - a case-control study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-004466.R3
Article Type:	Research
Date Submitted by the Author:	13-May-2014
Complete List of Authors:	Zhang, LG; the First Hospital Affiliated to the Xinxiang Medical College, The Third Department of Tuberculosis Xiao, ZJ; the First Hospital Affiliated to the Xinxiang Medical College, The Third Department of Tuberculosis Shi, GC; the First Hospital Affiliated to the Xinxiang Medical College, The Third Department of Tuberculosis J, Huang; the First Hospital Affiliated to the Xinxiang Medical College, The Third Department of Tuberculosis Zhang, CX; the First Hospital Affiliated to the Xinxiang Medical College, The Third Department of Tuberculosis
Primary Subject Heading:	Infectious diseases
Secondary Subject Heading:	Genetics and genomics
Keywords:	GENETICS, IMMUNOLOGY, INFECTIOUS DISEASES

Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a Han Chinese population - a case-control study

LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang

The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China

*Corresponding author: Zhijian Xiao, The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China. No.88 Jiankang Road, Weihui City, Henan, 463000, PR China

E-mail: zhangligongys@163.com

Tel: +86 0373 - 4402229 (office)

Fax: +86 0373 -4402251

Key words: Sarcoidosis; Annexin A11; Single-nucleotide polymorphism

Word count: 3918

Abstract

Objectives: To further identify the single-nucleotide polymorphisms (SNPs) that contribute to the genetic susceptibility to sarcoidosis, we examined the potential association between sarcoidosis and fifteen SNPs of the ANXA11 gene.

Design: A case - control study.

Setting: A tuberculosist unit in a hospital of the university in China.

Participants: Participants included 412 patients with sarcoidosis and 418 healthy controls.

Methods: The selected SNPs were genotyped using the MALDI-TOF in the MassARRAY system. Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA).

Results: The statistical significant differences were found in the allelic or genotypic frequencies of the rs2789679, rs1049550 and rs2819941 in the ANXA11 gene between the patients with sarcoidosis and the controls. The rs2789679 A allele (P = 0.00004, OR = 1.42, 95%CI = 1.17-1.73) and rs2819941 T allele (P = 0.0006, OR = 1.41, 95%CI = 1.16-1.71) were significantly more frequent in the patients with sarcoidosis compared to controls. The frequency of the rs1049550 T allele (P = 0.000002, OR = 0.61, 95%CI= 0.49 - 0.74) in the patients with sarcoidosis was significantly lower than that in the controls. The strong linkage disequilibrium (LD) was observed in four blocks (D' > 0.9). In block 2 (rs1049550-rs2573351), the T - C haplotype occurred significantly less frequently (P = 0.001), whereas the C - C haplotypes occurred more frequently (P = 0.0001) in the patients with sarcoidosis

than the controls. Furthermore, genotype frequency distribution revealed that, in rs1049550, CC genotype was significantly more in the patients with chest radiographic (CXR) stage I sarcoidosis than the patients with CXR stage II-IV sarcoidosis ($P = 0.012$).

Conclusions: These findings point to a role for the polymorphisms of ANXA11 in sarcoidosis in a Chinese Han population, and may be informative for future genetic studies on sarcoidosis.

Article summary

Article focus

1. The first genome-wide association study (GWAS) in sarcoidosis conducted in a German population has revealed that the rs1049550 (C > T, Arg230Cys) located in exon 10 were associated with sarcoidosis.

2. Whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations remain unclear.

3. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis in a Chinese Han population.

Key messages

1. ANXA11 rs2789679 (3'-UTR) was significantly associated with sarcoidosis.

2. ANXA11 rs1049550 (exon 10) was significantly associated with sarcoidosis.

3. Another significant association was observed for rs2819941 (intron 14).

Strengths and limitations of this study

1. A systematical screening of the functional SNPs in the promoter region, 5'- and 3'-UTR, exons of the ANXA11 gene, and the homogeneity of the study subjects representing the Chinese Han are the main strengths of this study.Chinese Han

2. The lack of data proving the positive association observed for rs2789679 and rs1049550 is a potential limitation of this study. Furthermore, the association of the serum level of ANXA11 with sarcoidosis still needs to be investigated.

INTRODUCTION

Sarcoidosis is a systemic autoimmune disease characterized by destructive, non-caseating epithelioid granulomatous lesions, with accumulated phagocytes and activated $CD4^+$ T helper type 1 lymphocytes [1 2]. The typical manifestations of sarcoidosis include bilateral hilar lymphadenopathy, pulmonary infiltration and ocular and skin lesions. The clinical course and prognosis of sarcoidosis are variable. The acute forms of sarcoidosis such as L□fgren's syndrome (LS) can resolve spontaneously in most cases. However, chronic pulmonary sarcoidosis may progress to lung fibrosis which eventually causes respiratory failure. Recent studies have demonstrated that sarcoidosis can occur in a pattern of family clustering, and its racial incidence is also different, suggesting that some genetic factors may contribute to the risk and severity of sacroidosis [1 3].

The Annexin gene family is involved in the etiology of several autoimmune and chronic diseases [4 5.] One member of the Annexin gene family, ANXA11, located on chromosome 10q22.3, is involved in calcium signaling, apoptosis, vesicle trafficking, cell growth and terminal phase of cell division [5-7]. In this context, the development and maintenance of the granulomatous inflammation in sarcoidosis have been repeatedly associated with the impaired apoptosis of activated inflammatory cells [8 9]. Most of the early-stage sarcoidosis is characterized by mononuclear cell alveolitis dominated by activated $CD4^+$ T cells and macrophages. Between these cells, the uncoordinated interplay results in the formation of typical non-caseating granulomas. The mechanism by which the granulomas resolve has not been fully elucidated.

However, it is generally assumed that the induction of apoptosis and/or by the withdrawal of inflammatory cytokines participates in the disappearance of granulomas [10 11]. In the patients with sarcoidosis, their peripheral blood monocytes were characterized by apoptosis [9].

In previous studies, case - control/family 'hypothesis - driven' studies and low density linkage scans were used to identify genetic factors conferring the genetic susceptibility to sarcoidosis [12]. Notably, the first genome - wide association study (GWAS) in sarcoidosis conducted in a German population has recently revealed an association: the leading rs2789679 SNP (single - nucleotide polymorphism) located in the 3'-untranslated region (3'-UTR) and a common nonsynonymous SNP (rs1049550, C > T, Arg230Cys) were associated with the increased risk of sarcoidosis [4 13]. This association has recently been supported by another report from the same population [14]. Thus, more studies should be performed to demonstrate the following items: whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations. The published studies about the association of sarcoidosis and ANXA11 are summarized in Supplemental table 1. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis.

In this study, we investigated fifteen loci in a Chinese population from He'nan province (China) to verify the putative association between ANXA11 polymorphisms and sarcoidosis.

SUBJECTS AND METHODS

Subjects

Four hundred and twelve patients with sarcoidosis (mean ± SD age: 53.6 ± 4.6 years; 155 men and 257 women) were recruited from our hospital between May 2005 and March 2013. Sarcoidosis was diagnosed by the evidence of non-caseating epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities. Chronic sarcoidosis was defined as a disease over at least 2 years or at least two episodes in a lifetime. A cute sarcoidosis was defined as one episode of acute sarcoidosis which had totally resolved at the date of the examination. The control group consisted of 418 unrelated healthy subjects (mean ± SD age: 54.2 ± 5.3 years; 160 men and 258 women) who underwent health examinations in the Medical Examination Center of the First Affiliated Hospital of the Xinxiang Medical College (Xinxiang, China) between Oct 2009 and Sept 2013. None of the individuals in the control group had a history of lung diseases or showed any symptoms of the lung or other diseases by chest radiography or laboratory blood tests. Participants were excluded if they: were taking other prescribed medications that could affect the central nervous system; had a history of seizures, hematological diseases, or severe liver or kidney impairment; or were pregnant. No familial relationship was known between the study participants. The study was performed according to the Guidelines of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang, China). Written informed consent was obtained from each participant in this study.

SNP selection

Tagging SNPs were selected from the catalogs of the International HapMap Project. The rs2789679 and rs2245168 are located in 5'-UTR, rs2236558 in intron 6, rs7067644 and rs12763624 in intron 8, rs1049550 in exon 10, rs2573351 in intron 13, rs10887581, rs11201989, rs2573353, rs2789695, rs2573356, rs2819945, rs11202059 and rs2819941 in intron 14 (Supplemental figure 1). Marker selection was based on the following criteria. (a) We used the CHB data from the HapMap (release 27) to select tagSNPs for ANXA11. (b) We restricted our search for tagSNPs from base pair 80155124 to 80205677. (c) Further, we limited the SNPs to tag to those with a minor allele frequency ≥ 0.2 in the CHB. (d) Based on these restrictions, there were a total of 29 SNPs. (e) Using HAPLOVIEW with an r^2 threshold ≥ 0.8, 15 tagSNP were selected and used for subsequent analyses.

Genotyping

Peripheral blood was collected from a vein into a sterile tube coated with ethylenediamine tetraacetic acid (EDTA). Plasma samples were stored at -20°C. Genomic DNA was extracted from the frozen peripheral blood samples using a QIAmp Blood Mini Kit (Qiagen Inc., Valencia, CA, USA) according to the manufacturer's protocols. The selected SNPs were genotyped in cases and controls by using the MALDI-TOF in the MassARRAY system (Sequenom Inc., San Diego, CA, USA). Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA). The completed genotyping reactions were spotted onto a 384 well spectroCHIP (Sequenom) using the MassARRAY Nanodispenser (Sequenom) and determined by the matrix - assisted laser desorption ionization time -

of - flight mass spectrometer. Genotype calling was performed in realtime with the MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer software version 3.4 (Sequenom).

Statistical analysis

All data were analyzed using the SPSS 17.0 software (SPSS Inc., Chicago, IL, USA). Hardy-Weinberg equilibrium was evaluated by Chi-square tests. Differences between the cases and controls in the frequency of the alleles, genotypes and haplotypes were evaluated by the Fisher's exact test or the Pearson Chi-square test. Unconditional logistic regression was used to calculate the odds ratio (OR) and 95% confidence interval (CI) in independent association between each locus and the presence of sarcoidosis. Gender and age of subjects were treated as covariants in binary logistic regression. *P* values were calculated based on codominant, dominant for the rare allele, heterosis and recessive for the rare allele models of inheritance. Models of multiple logistic regression were used to test the independence of individual allelic effect. In detail, the most significant SNP was chosen to be the conditional SNP (covariate in the regression model) when testing other significant SNPs; Meanwhile, a multi-SNP model including all significant SNPs was also performed. The Bonferroni correction was used to adjust the test level when multiple comparisons were conducted, and the *P* value was divided by the total number of loci. Haplotype blocks were defined according to the criteria developed by Gabriel et al. [15]. Pair-wise LD statistics (D' and r^2) and haplotype frequency were calculated, and haplotype blocks were constructed using the Haploview 4.0 [16]. To ensure that the LD

blocks most closely reflect the population level LD patterns, definition of the blocks were based on the control samples alone. The haplotype frequencies were estimated using GENECOUNTING, which computes maximum - likelihood estimates of haplotype frequencies from unknown phase data by utilizing an expectation - maximization algorithm [17-20]. The significance of any haplotypic association was evaluated using a likelihood ratio test, followed by permutation testing that compared estimated haplotype frequencies in cases and controls [17 19].

RESULTS

The genotype distribution of the fifteen polymorphisms was consistent with the Hardy-Weinberg equilibrium ($P > 0.05$). The analysis of strong LD in the patients with sarcoidosis and healthy controls revealed that 2 SNPs (rs2245168, rs2236558), 2 SNPs (rs1049550, rs2573351), 2 SNPs (rs2789695, rs2573356) and 3 SNPs (rs2819945, rs11202059, rs2819941) were located in haplotype block 1, block 2, block 3 and block 4 (D' > 0.9, Fig. 1). The genotype distribution, allelic frequencies, and haplotypes in the patients with sarcoidosis and healthy controls are showed in Tables 1-3 and Supplemental table 2

Comparison of genotype and allele frequency distribution revealed significant differences between the patients with sarcoidosis and healthy controls for 3 SNPs: rs2789679, rs1049550 and rs2819941. The rs2789679 A allele ($P = 0.00004$, OR = 1.42, 95%CI = 1.17-1.73) and rs2819941 T allele ($P = 0.0006$, OR = 1.41, 95%CI = 1.16-1.71) were significantly more frequent in the patients with sarcoidosis compared to controls. The frequency of the rs1049550 T allele ($P = 0.000002$, OR = 0.61,

95%CI= 0.49-0.74) in the patients with sarcoidosis was significantly lower than that in the controls. These differences remained statistically significant after Bonferroni corrections.

The multi-SNP model showed only rs1049550 present a significant effect on the disease phenotype ($p < 0.001$). No independent effect was found for rs2789679 or rs2819941 when adjusting for the effect of rs1049550. Both individual SNP and multi - SNP analysis supported rs1049550: T allele was an important protective factor for affecting sarcoidosis (odds ratio around 0.6). In another word, carriers of rs1049550: C allele have higher susceptibility to sarcoidosis in our Chinese Han population (Supplemental table 3).

We performed an association analysis to determine whether the haplotype was associated with the risk of sarcoidosis (Global $P < 0.05$ in block 2 and block 3). Compared with the controls, the T - C haplotype occurred significantly less frequently ($P = 0.001$) but the C - C haplotypes occurred more frequently ($P = 0.0001$) in block 2 (rs1049550 - rs2573351) in the patients with sarcoidosis.

To assess particular disease phenotypes, the patients with sarcoidosis were divided into the subgroups according to their chest radiographic (CXR) stage. CXR stage I (isolated bilateral hilar lymphadenopathy): N = 196; CXR stages II - IV (infiltration of parenchyma): 167; CXR stage 0 (absence on pulmonary disease): N = 28; the information on CXR stage was not available for 21 patients. Genotype frequency distribution revealed that, in rs1049550, CC genotype was significantly more in the patients with chest radiographic (CXR) stage I sarcoidosis than the

patients with CXR stage II-IV sarcoidosis ($P = 0.012$). (Table 4).

DISCUSSION

A key step in linkage and association studies is to identify common risk variants in different populations. To determine if common risk variants exist in distinct populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene, were genotyped in samples from the patients with sarcoidosis and healthy controls in a Chinese Han population. Recently, new sarcoidosis loci have been identified by genome-wide association studies [4 14]. With the fast development of genome - wide association studies, an increasing number of susceptibility loci for sarcoidosis have been found in different populations [4 14 21]. However, these observations should be confirmed in other genetically independent populations. In this study, we conducted the first large genetic association study of the ANXA11 gene in a Chinese Han population. The evidence of markers associated with sarcoidosis was presented, and these markers were mapped to different locations in the ANXA11 gene (81897864 - 81951001). The association signals in the region were identified, and some significantly associated haplotypes also appeared this region.

Hoffman et al. reported an association between the T allele of the non - synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a German population [4]. Similar results were also obtained in other two European populations [14 21]. As a part of the sarcoidosis GWAS done in Americans [22], we further confirmed this association in a Chinese Han population. In this study, the frequency of ANXA11 rs1049550 T allele was significantly lower in the patients with sarcoidosis

than the healthy controls. Genotype frequency distribution revealed that, in rs1049550, CC genotype was significantly more in the patients with stage I sarcoidosis than the patients with stage II-IV sarcoidosis. The "protective" effect of ANXA11 T allele increased with the number of its copies in the genotype, which is consistent with the result obtained in the patients with sarcoidosis in a German population [4], Furthermore, the T allele carriers among the patients were protected from infiltration of lung parenchyma (radiographic stages II-IV) [21]. We have also demonstrated that rs1049550 is significant cross - ethnically at the gene level after adjustment for the single SNP association tests performed. The mechanism by which rs1049550 affects the susceptibility to sarcoidosis is that the SNP may affect the function of the ANXA11 protein rather than affect its expression. In humans, the ANXA11 protein consists of an N - terminalproline - tyrosine - glycine (PYG) - rich region, followed by four annexin core domains. The rs1049550 leads to an amino - acid exchange (basic arginine to a polar cysteine) at the evolutionarily conserved position 230 (R230C) in the first annexin domain, which is responsible for Ca^{2+} - dependent trafficking of the protein in the cell [23].

In the GWAS conducted by Hofmann et al [4], the strongest association signal was observed for rs2789679. This marker is located in the 3'-UTR of the ANXA11 gene. The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls being TT homozygous. Fine-mapping in the Black Women's Health Study (BWHS) revealed rs2819941 to be the top SNP, associated with a non-significant (P = 0.06)

reduction in the risk of sarcoidosis (17% reduction per copy of the C-allele). In this case-control association study, the frequency of the A allele in rs2789679 and the rs2819941 T allele frequency in the patients with sarcoidosis was significantly higher than that in the healthy controls. Several lines of evidence suggest that the observed association is unlikely to be an artifact. First, both the single-SNP and the haplotype-based association analyses support the current finding. Second, our samples were from the same geographical region, excluding the. Finally, consistent results were obtained from two genetically independent populations (Chinese Han and Europeans). Collectively, our results confirmed the strong association between ANXA11 polymorphisms and sarcoidosis, suggesting that ANXA11 represents a strong genetic risk factor for sarcoidosis. Since causal SNP in ANXA11 was not known, all the SNPs significantly associated with sarcoidosis were actually indirect association through the real causal one. Though rs2789679 and rs2819941 were not in LD with rs1049550, it may not tag the causal SNP as well as rs1049550. Hence, their individual effects were eliminated when controlling for the effect of rs1049550.

We further investigated the interaction among polymorphisms and observed strong LD. The haplotype analysis revealed that significantly more C - C (block 2, rs1049550 - rs2573351) and significantly fewer T - C haplotypes (block 2) were found in the patients with sarcoidosis. These results indicated that the patients with C - C (block 2) haplotypes of the ANXA11 gene were more prone to sarcoidosis. Significantly higher frequencies of T - C haplotypes were detected in the healthy controls than in the patients with sarcoidosis, suggesting that they may show

protective effects against sarcoidosis. Our sample size can detect SNP and haplotype associations with 90% and 85% power, respectively, at a false positive rate of 5%, disease prevalence of 1‰, disease allele / haplotype frequency of 0.05 / 0.03, and a presumed odds ratio (OR) of 1.5. To some extent, this finding further supports a role of ANXA11 polymorphisms in sarcoidosis, while ethnic group difference may exist.

In conclusion, our study suggests a potential role of the ANXA11 SNPs and their related haplotypes in the genetic susceptibility to sarcoidosis. Further studies are needed to investigate how these SNPs affect the function of ANXA11.

Funding

The project was supported by the Research and Development Foundation of Science and Technology of Henan Province (2012k15).

Contributors

LG Zhang was involved in conception and design of the study, acquisition of patient data, genotyping and drafted the paper. ZJ Xiao and GC Shi were involved in design of the study, genotyping and interpretation of results. J Huang and CX Zhang were involved in the design of the study, acquisition of patient data and interpretation of result. All authors revised the draft paper.

Competing interests

None.

Data sharing statement

No additional data are available.

Patient consent

Obtained.

Ethics approval

Ethics approval was provided by the First Hospital Affiliated to the Xinxiang Medical College.

Provenance and peer review

Not commissioned; externally peer reviewed.

Figure Legends

Fig. 1 The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the pair-wise calculation of r^2 (left) or D' (right). Black squares indicate $r^2 = 1$ (i.e. perfect LD between a pair of SNPs). Empty squares indicate D' = 1 (i.e. complete LD between a pair of SNPs).

Supplemental Fig. 1 The structure of the human ANXA11 gene and 15 SNPs located on the gene.

Reference

1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007; 357:2153-65.

2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-8

3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis. *Semin Respir Crit Care Med* 2010;31:380-9.

4. Hofmann S, Franke A, Fischer A, *et al*. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet* 2008;40:1103-6.

5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev* 2002;82:331-71.

6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using recombinant human annexin XI-specific autoantibodies cloned by phage display. *J Biol Chem* 2003;278:33120-6.

7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-22.

8. Petzmann S, Maercker C, Markert E, *et al*. Enhanced proliferation and decreased apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis Vasc Diffuse Lung Dis* 2006;23:190-200

9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, *et al*. Comparative analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol*

2006;26:243-50.

10. Agostini C, Trentin L, Facco M, *et al*. Role of IL-15, IL-2, and their receptors in the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol* 1996;157:910-8

11. Jullien D, Sieling PA, Uyemura K, *et al*. IL-15, an immunomodulator of T cell responses in intracellular infection. *J Immunol* 1997;158:800-6

12. Muller-Quernheim J, Schurmann M, *et al*. Genetics of sarcoidosis. *Clin Chest Med* 2008;29:391-414.

13. Morais A, Lima B, Peixoto M, *et al*. Annexin A11 gene polymorphism (R230C variant) and sarcoidosis in a Portuguese population. *Tissue antigens* 2013;82:186-91.

14. Li Y, Pabst S, Kubisch C, *et al*. First independent replication study confirms the strong genetic association of ANXA11 with sarcoidosis. *Thorax* 2010;65:939-40.

15. Gabriel SB, Schaffner SF, Nguyen H, *et al*. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-9.

16. Barrett JC, Fry B, Maller J, *et al*. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-5.

17. Curtis D, Knight J, Sham PC. Program report: GENECOUNTING support programs. *Ann Hum Genet* 2006;70:277-9.

18. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic associations. *Hum Hered* 2000;50:133-9.

19. Zhao JH, Lissarrague S, Essioux L, et al. GENECOUNTING: haplotype analysis with missing genotypes. *Bioinformatics* 2002;18:1694-5

20. Zhao JH, Sham PC. Generic number systems and haplotype analysis. *Comput Methods Programs Biomed* 2003;70:1-9

21. Mrazek F, Stahelova A, Kriegova E, *et al*. Functional variant ANXA11 R230C: true marker of protection and candidate disease modifier in sarcoidosis. *Genes Immun* 2011;12:490-4.

22. Levin AM, Iannuzzi MC, Montgomery CG, *et al*. Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes Immun* 2013;14:13-8.

24. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11 with the nuclear envelope. *J Biol Chem* 2003;278:20210-6.

Table 1 Genotypic and allelic frequencies of ANXA11 polymorphisms in the controls and patients with sarcoidosis

Variable	**Controls** (n=418)		**Sarcoidosis** (n=412)		*P*-value[a]	OR, 95% CI
	No.	%	No.	%		
rs2789679					0.0007	
TT	114	27.3	94	22.8	0.140	1.27, 0.93-1.74
AT	225	53.8	186	45.2	0.012	1.42, 1.08-1.87
AA	79	18.9	132	32.0	0.0002	0.49, 0.36-0.68
Per A allele	383	45.8	450	54.6	0.00004	1.42, 1.17-1.73
rs1049550					0.0002	
CC	154	36.8	208	50.5	0.0007	0.57, 0.43-0.75
CT	190	45.5	170	41.3	0.221	1.19, 0.90-1.56
TT	74	17.7	34	8.3	0.0008	2.39, 1.55-3.68
Per T allele	338	40.4	238	28.8	0.000002	0.61, 0.49-0.74
rs2819941					0.0002	
CC	114	27.3	94	22.8	0.141	1.27, 0.93-1.74
CT	226	54.1	190	46.1	0.021	1.38, 1.05-1.81
TT	78	18.7	128	31.1	0.0004	0.51, 0.37-0.70
Per T allele	382	45.7	446	54.1	0.0006	1.41, 1.16-1.71

[a] *P*-value was calculated by 2×3 and 2×2 chi-squared tests based on codominant, dominant for the rare allele, heterosis and recessive for the rare allele models of inheritance.

Alpha value is adjusted by Bonferroni correction and statistically significant results (P<0.003)

Table 2 ANXA11 haplotype in block 2 frequencies and the results of their associations with risk of sarcoidosis

Haplotype			Genecounting (frequency %)			
ID	**rs1049550**	**rs2573351**	**Cases**	**Controls**	***P*-value[a]**	**Global *P*[b]**
HAP1	T	C	27.7	39.7	0.001	0.003
HAP2	C	C	19.9	7.4	0.0001	
HAP3	C	T	50.5	51.0	0.892	

[a] Based on 10,000 permutations.

[b] Based on comparison of frequency distribution of all haplotypes for the combination of SNPs.

Table 3 ANXA11 haplotype in block 3 frequencies and the results of their associations with risk of sarcoidosis

Haplotype			Genecounting (frequency %)			
ID	**rs2789695**	**rs2573356**	**Cases**	**Controls**	***P*-value[a]**	**Global *P*[b]**
HAP1	C	T	33.7	34.9	0.718	0.045
HAP2	T	T	15.3	14.1	0.632	
HAP2	T	T	15.6	14.1	0.565	

[a] Based on 10,000 permutations.

[b] Based on comparison of frequency distribution of all haplotypes for the combination of SNPs.

Table 4 Chest radiographic (CXR) stages of sarcoidosis patients by genotype of rs2789679, rs1049550 and rs2819941

Stages	rs2789679 (n, %)			rs1049550* (n, %)			rs2819941 (n, %)		
	AA	AT	TT	CC	CT	TT	**CC**	**CT**	**TT**
stage I	59 (30.1)	92 (47.0)	45 (23.0)	101 (51.5)	82 (41.8)	13 (6.6)	47 (25.0)	88 (44.9)	61 (31.1)
stages II - IV	46 (27.6)	75 (44.9)	46 (27.6)	72 (43.1)	68 (40.7)	27 (16.2)	41 (24.6)	73 (43.7)	53 (31.7)
χ^2, P	1.041, 0.594			8.807, 0.012			0.052, 0.975		

* P values for genotype frequency distribution ($P < 0.05$).

Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a Han Chinese population - a case-control study

LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang

The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China

*Corresponding author: Zhijian Xiao, The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China. No.88 Jiankang Road, Weihui City, Henan, 463000, PR China

E-mail: zhangligongys@163.com

Tel: +86 0373 - 4402229 (office)

Fax: +86 0373 -4402251

Key words: Sarcoidosis; Annexin A11; Single-nucleotide polymorphism

Word count: ~~3787~~3918

Abstract

Objectives: To further identify the single-nucleotide polymorphisms (SNPs) that contribute to the genetic susceptibility to sarcoidosis, we examined the potential association between sarcoidosis and fifteen SNPs of the ANXA11 gene.

Design: A case - control study.

Setting: A tuberculosist unit in a hospital of the university in China.

Participants: Participants included 412 patients with sarcoidosis and 418 healthy controls.

Methods: The selected SNPs were genotyped using the MALDI-TOF in the MassARRAY system. Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA).

Results: The statistical significant differences were found in the allelic or genotypic frequencies of the rs2789679, rs1049550 and rs2819941 in the ANXA11 gene between the patients with sarcoidosis and the controls. The rs2789679 A allele (P = 0.00004, OR = 1.42, 95%CI = 1.17-1.73) and rs2819941 T allele (P = 0.0006, OR = 1.41, 95%CI = 1.16-1.71) were significantly more frequent in the patients with sarcoidosis compared to controls. The frequency of the rs1049550 T allele (P = 0.000002, OR = 0.61, 95%CI= 0.49 - 0.74) in the patients with sarcoidosis was significantly lower than that in the controls. ~~The frequency of the T allele in rs2789679 (P = 0.0003, OR = 0.70, 95%CI = 0.58 - 0.85) and rs1049550 (P = 0.0007, OR = 0.62, 95%CI= 0.50 - 0.76), and C allele in rs2819941 (P = 0.001, OR = 0.71, 95%CI = 0.59 - 0.87) in the patients with sarcoidosis was significantly lower than that~~

~~in the controls.~~ The strong linkage disequilibrium (LD) was observed in four blocks (D' > 0.9). In block 2 (rs1049550-rs2573351), the T - C haplotype occurred significantly less frequently (P = 0.001), whereas the C - C haplotypes occurred more frequently (P = 0.0001) in the patients with sarcoidosis than the controls. Furthermore, genotype frequency distribution revealed that, in rs1049550, CC genotype was significantly more in the patients with chest radiographic (CXR) stage I sarcoidosis than the patients with CXR stage II-IV sarcoidosis (P = 0.012).

Conclusions: These findings point to a role for the polymorphisms of ANXA11 in sarcoidosis in a Chinese Han population, and may be informative for future genetic studies on sarcoidosis.

Article summary

Article focus

1. The first genome-wide association study (GWAS) in sarcoidosis conducted in a German population has revealed that the rs1049550 (C > T, Arg230Cys) located in exon 10 were associated with sarcoidosis.

2. Whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations remain unclear.

3. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis in a Chinese Han population.

Key messages

1. ANXA11 rs2789679 (3'-UTR) was significantly associated with sarcoidosis.

2. ANXA11 rs1049550 (exon 10) was significantly associated with sarcoidosis.

3. Another significant association was observed for rs2819941 (intron 14).

Strengths and limitations of this study

1. A systematical screening of the functional SNPs in the promoter region, 5'- and 3'-UTR, exons of the ANXA11 gene, and the homogeneity of the study subjects representing the Chinese Han are the main strengths of this study.Chinese Han

2. The lack of data proving the positive association observed for rs2789679 and rs1049550 is a potential limitation of this study. Furthermore, the association of the serum level of ANXA11 with sarcoidosis still needs to be investigated.

INTRODUCTION

Sarcoidosis is a systemic autoimmune disease characterized by destructive, non-caseating epithelioid granulomatous lesions, with accumulated phagocytes and activated $CD4^+$ T helper type 1 lymphocytes [1 2]. The typical manifestations of sarcoidosis include bilateral hilar lymphadenopathy, pulmonary infiltration and ocular and skin lesions. The clinical course and prognosis of sarcoidosis are variable. The acute forms of sarcoidosis such as L□fgren's syndrome (LS) can resolve spontaneously in most cases. However, chronic pulmonary sarcoidosis may progress to lung fibrosis which eventually causes respiratory failure. Recent studies have demonstrated that sarcoidosis can occur in a pattern of family clustering, and its racial incidence is also different, suggesting that some genetic factors may contribute to the risk and severity of sacroidosis [1 3].

The Annexin gene family is involved in the etiology of several autoimmune and chronic diseases [4 5.] One member of the Annexin gene family, ANXA11, located on chromosome 10q22.3, is involved in calcium signaling, apoptosis, vesicle trafficking, cell growth and terminal phase of cell division [5-7]. In this context, the development and maintenance of the granulomatous inflammation in sarcoidosis have been repeatedly associated with the impaired apoptosis of activated inflammatory cells [8 9]. Most of the early-stage sarcoidosis is characterized by mononuclear cell alveolitis dominated by activated $CD4^+$ T cells and macrophages. Between these cells, the uncoordinated interplay results in the formation of typical non-caseating granulomas. The mechanism by which the granulomas resolve has not been fully elucidated. However, it is generally assumed that the induction of apoptosis and/or by the

withdrawal of inflammatory cytokines participates in the disappearance of granulomas [10 11]. In the patients with sarcoidosis, their peripheral blood monocytes were characterized by apoptosis [9].

In previous studies, case - control/family 'hypothesis - driven' studies and low density linkage scans were used to identify genetic factors conferring the genetic susceptibility to sarcoidosis [12]. Notably, the first genome - wide association study (GWAS) in sarcoidosis conducted in a German population has recently revealed an association: the leading rs2789679 SNP (single - nucleotide polymorphism) located in the 3'-untranslated region (3'-UTR) and a common nonsynonymous SNP (rs1049550, C > T, Arg230Cys) were associated with the increased risk of sarcoidosis [4 13]. This association has recently been supported by another report from the same population [14]. Thus, more studies should be performed to demonstrate the following items: whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations. The published studies about the association of sarcoidosis and ANXA11 are summarized in Supplemental table 1. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis.

In this study, we investigated fifteen loci in a Chinese population from He'nan province (China) to verify the putative association between ANXA11 polymorphisms and sarcoidosis.

SUBJECTS AND METHODS

Subjects

Four hundred and twelve patients with sarcoidosis (mean ± SD age: 53.6 ± 4.6 years; 155 men and 257 women) were recruited from our hospital between May 2005 and March 2013. Sarcoidosis was diagnosed by the evidence of non-caseating epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities. Chronic sarcoidosis was defined as a disease over at least 2 years or at least two episodes in a lifetime. A cute sarcoidosis was defined as one episode of acute sarcoidosis which had totally resolved at the date of the examination. The control group consisted of 418 unrelated healthy subjects (mean ± SD age: 54.2 ± 5.3 years; 160 men and 258 women) who underwent health examinations in the Medical Examination Center of the First Affiliated Hospital of the Xinxiang Medical College (Xinxiang, China) between Oct 2009 and Sept 2013. None of the individuals in the control group had a history of lung diseases or showed any symptoms of the lung or other diseases by chest radiography or laboratory blood tests. Participants were excluded if they: were taking other prescribed medications that could affect the central nervous system; had a history of seizures, hematological diseases, or severe liver or kidney impairment; or were pregnant. No familial relationship was known between the study participants. The study was performed according to the Guidelines of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang, China). Written informed consent was obtained from each participant in this study.

SNP selection

Tagging SNPs were selected from the catalogs of the International HapMap Project. The rs2789679 and rs2245168 are located in 5'-UTR, rs2236558 in intron 6,

rs7067644 and rs12763624 in intron 8, rs1049550 in exon 10, rs2573351 in intron 13, rs10887581, rs11201989, rs2573353, rs2789695, rs2573356, rs2819945, rs11202059 and rs2819941 in intron 14 (Supplemental figure 1). Marker selection was based on the following criteria. (a) We used the CHB data from the HapMap (release 27) to select tagSNPs for ANXA11. (b) We restricted our search for tagSNPs from base pair 80155124 to 80205677. (c) Further, we limited the SNPs to tag to those with a minor allele frequency ≥ 0.2 in the CHB. (d) Based on these restrictions, there were a total of 29 SNPs. (e) Using HAPLOVIEW with an r^2 threshold ≥ 0.8, 15 tagSNP were selected and used for subsequent analyses.

Genotyping

Peripheral blood was collected from a vein into a sterile tube coated with ethylenediamine tetraacetic acid (EDTA). Plasma samples were stored at -20°C. Genomic DNA was extracted from the frozen peripheral blood samples using a QIAmp Blood Mini Kit (Qiagen Inc., Valencia, CA, USA) according to the manufacturer's protocols. The selected SNPs were genotyped in cases and controls by using the MALDI-TOF in the MassARRAY system (Sequenom Inc., San Diego, CA, USA). Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA). The completed genotyping reactions were spotted onto a 384 well spectroCHIP (Sequenom) using the MassARRAY Nanodispenser (Sequenom) and determined by the matrix - assisted laser desorption ionization time - of - flight mass spectrometer. Genotype calling was performed in realtime with the MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer

software version 3.4 (Sequenom).

Statistical analysis

All data were analyzed using the SPSS 17.0 software (SPSS Inc., Chicago, IL, USA). Hardy-Weinberg equilibrium was evaluated by Chi-square tests. Differences between the cases and controls in the frequency of the alleles, genotypes and haplotypes were evaluated by the Fisher's exact test or the Pearson Chi-square test. Unconditional logistic regression was used to calculate the odds ratio (OR) and 95% confidence interval (CI) in independent association between each locus and the presence of sarcoidosis. ~~Logistic regression was used to evaluate the interaction effects between gene and gender or age.~~ Gender and age of subjects were treated as covariants in binary logistic regression. *P* values were calculated based on codominant, dominant for the rare allele, heterosis and recessive for the rare allele models of inheritance. Models of multiple logistic regression were used to test the independence of individual allelic effect. In detail, the most significant SNP was chosen to be the conditional SNP (covariate in the regression model) when testing other significant SNPs; Meanwhile, a multi-SNP model including all significant SNPs was also performed. The Bonferroni correction was used to adjust the test level when multiple comparisons were conducted, and the *P* value was divided by the total number of loci. Haplotype blocks were defined according to the criteria developed by Gabriel et al. [15]. Pair-wise LD statistics (D' and r^2) and haplotype frequency were calculated, and haplotype blocks were constructed using the Haploview 4.0 [16]. To ensure that the LD blocks most closely reflect the population level LD patterns,

definition of the blocks were based on the control samples alone. The haplotype frequencies were estimated using GENECOUNTING, which computes maximum - likelihood estimates of haplotype frequencies from unknown phase data by utilizing an expectation - maximization algorithm [17-20]. The significance of any haplotypic association was evaluated using a likelihood ratio test, followed by permutation testing that compared estimated haplotype frequencies in cases and controls [17 19].

RESULTS

The genotype distribution of the fifteen polymorphisms was consistent with the Hardy-Weinberg equilibrium ($P > 0.05$). The analysis of strong LD in the patients with sarcoidosis and healthy controls revealed that 2 SNPs (rs2245168, rs2236558), 2 SNPs (rs1049550, rs2573351), 2 SNPs (rs2789695, rs2573356) and 3 SNPs (rs2819945, rs11202059, rs2819941) were located in haplotype block 1, block 2, block 3 and block 4 (D' > 0.9, Fig. 1). The genotype distribution, allelic frequencies, and haplotypes in the patients with sarcoidosis and healthy controls are showed in Tables 1-3 and Supplemental table 2

Comparison of genotype and allele frequency distribution revealed significant differences between the patients with sarcoidosis and healthy controls for 3 SNPs: rs2789679, rs1049550 and rs2819941. The rs2789679 A allele ($P = 0.00004$, OR = 1.42, 95%CI = 1.17-1.73) and rs2819941 T allele ($P = 0.0006$, OR = 1.41, 95%CI = 1.16-1.71) were significantly more frequent in the patients with sarcoidosis compared to controls. The frequency of the rs1049550 T allele ($P = 0.000002$, OR = 0.61, 95%CI= 0.49-0.74) in the patients with sarcoidosis was significantly lower than that

in the controls. ~~The frequency of the T allele in rs2789679 (*P* = 0.0003, OR = 0.70, 95%CI = 0.58 - 0.85) and rs1049550 (*P* = 0.0007, OR = 0.62, 95%CI = 0.50 - 0.76), and C allele in rs2819941 (*P* = 0.001, OR = 0.71, 95%CI = 0.59 - 0.87) in the patients with sarcoidosis was significantly lower than that in the controls, and~~ These differences remained statistically significant after Bonferroni corrections.

The multi-SNP model showed only rs1049550 present a significant effect on the disease phenotype ($p < 0.001$). No independent effect was found for rs2789679 or rs2819941 when adjusting for the effect of rs1049550. Both individual SNP and multi - SNP analysis supported rs1049550: T allele was an important protective factor for affecting sarcoidosis (odds ratio around 0.6). In another word, carriers of rs1049550: C allele have higher susceptibility to sarcoidosis in our Chinese Han population (Supplemental table 3).

~~We presented a multi-SNP model of sarcoidosis risk. We screen out the most significant model to predict the risk by forward stepwise strategy in logistic regression, and found that a model includes rs2789679 and rs1049550 is the best one (Supplemental table 3), which suggests that there is a interaction between them.~~

We performed an association analysis to determine whether the haplotype was associated with the risk of sarcoidosis (Global $P < 0.05$ in block 2 and block 3). Compared with the controls, the T - C haplotype occurred significantly less frequently ($P = 0.001$) but the C - C haplotypes occurred more frequently ($P = 0.0001$) in block 2 (rs1049550 - rs2573351) in the patients with sarcoidosis.

To assess particular disease phenotypes, the patients with sarcoidosis were

divided into the subgroups according to their chest radiographic (CXR) stage. CXR stage I (isolated bilateral hilar lymphadenopathy): N = 196; CXR stages II - IV (infiltration of parenchyma): 167; CXR stage 0 (absence on pulmonary disease): N = 28; the information on CXR stage was not available for 21 patients. Genotype frequency distribution revealed that, in rs1049550, CC genotype was significantly more in the patients with chest radiographic (CXR) stage I sarcoidosis than the patients with CXR stage II-IV sarcoidosis ($P = 0.012$). (Table 4).

DISCUSSION

A key step in linkage and association studies is to identify common risk variants in different populations. To determine if common risk variants exist in distinct populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene, were genotyped in samples from the patients with sarcoidosis and healthy controls in a Chinese Han population. Recently, new sarcoidosis loci have been identified by genome-wide association studies [4 14]. With the fast development of genome - wide association studies, an increasing number of susceptibility loci for sarcoidosis have been found in different populations [4 14 21]. However, these observations should be confirmed in other genetically independent populations. In this study, we conducted the first large genetic association study of the ANXA11 gene in a Chinese Han population. The evidence of markers associated with sarcoidosis was presented, and these markers were mapped to different locations in the ANXA11 gene (81897864 - 81951001). The association signals in the region were identified, and some significantly associated haplotypes also appeared this region.

Hoffman et al. reported an association between the T allele of the non - synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a German population [4]. Similar results were also obtained in other two European populations [14 21]. As a part of the sarcoidosis GWAS done in Americans [22], we further confirmed this association in a Chinese Han population. In this study, the frequency of ANXA11 rs1049550 T allele was significantly lower in the patients with sarcoidosis than the healthy controls. ~~There is interaction between rs2789679 and rs1049550.~~ Genotype frequency distribution revealed that, in rs1049550, CC genotype was significantly more in the patients with stage I sarcoidosis than the patients with stage II-IV sarcoidosis. The "protective" effect of ANXA11 T allele increased with the number of its copies in the genotype, which is consistent with the result obtained in the patients with sarcoidosis in a German population [4], Furthermore, the T allele carriers among the patients were protected from infiltration of lung parenchyma (radiographic stages II-IV) [21]. We have also demonstrated that rs1049550 is significant cross - ethnically at the gene level after adjustment for the single SNP association tests performed. The mechanism by which rs1049550 affects the susceptibility to sarcoidosis is that the SNP may affect the function of the ANXA11 protein rather than affect its expression. In humans, the ANXA11 protein consists of an N - terminalproline - tyrosine - glycine (PYG) - rich region, followed by four annexin core domains. The rs1049550 leads to an amino - acid exchange (basic arginine to a polar cysteine) at the evolutionarily conserved position 230 (R230C) in the first annexin domain, which is responsible for Ca^{2+} - dependent trafficking of the

protein in the cell [23].

In the GWAS conducted by Hofmann et al [4], the strongest association signal was observed for rs2789679. This marker is located in the 3'-UTR of the ANXA11 gene. The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls being TT homozygous. Fine-mapping in the Black Women's Health Study (BWHS) revealed rs2819941 to be the top SNP, associated with a non-significant (P = 0.06) reduction in the risk of sarcoidosis (17% reduction per copy of the C-allele). In this case-control association study, the frequency of the ~~T~~ A allele in rs2789679 and the rs2819941 ~~C~~ T allele frequency in the patients with sarcoidosis was significantly ~~lower~~ higher than that in the healthy controls. Several lines of evidence suggest that the observed association is unlikely to be an artifact. First, both the single-SNP and the haplotype-based association analyses support the current finding. Second, our samples were from the same geographical region, excluding the. Finally, consistent results were obtained from two genetically independent populations (Chinese Han and Europeans). Collectively, our results confirmed the strong association between ANXA11 polymorphisms and sarcoidosis, suggesting that ANXA11 represents a strong genetic risk factor for sarcoidosis. Since causal SNP in ANXA11 was not known, all the SNPs significantly associated with sarcoidosis were actually indirect association through the real causal one. Though rs2789679 and rs2819941 were not in LD with rs1049550, it may not tag the causal SNP as well as rs1049550. Hence, their individual effects were eliminated when controlling for the effect of rs1049550.

We further investigated the interaction among polymorphisms and observed strong LD. The haplotype analysis revealed that significantly more C - C (block 2, rs1049550 - rs2573351) and significantly fewer T - C haplotypes (block 2) were found in the patients with sarcoidosis. These results indicated that the patients with C - C (block 2) haplotypes of the ANXA11 gene were more prone to sarcoidosis. Significantly higher frequencies of T - C haplotypes were detected in the healthy controls than in the patients with sarcoidosis, suggesting that they may show protective effects against sarcoidosis. Our sample size can detect SNP and haplotype associations with 90% and 85% power, respectively, at a false positive rate of 5%, disease prevalence of 1‰, disease allele / haplotype frequency of 0.05 / 0.03, and a presumed odds ratio (OR) of 1.5. To some extent, this finding further supports a role of ANXA11 polymorphisms in sarcoidosis, while ethnic group difference may exist.

In conclusion, our study suggests a potential role of the ANXA11 SNPs and their related haplotypes in the genetic susceptibility to sarcoidosis. Further studies are needed to investigate how these SNPs affect the function of ANXA11.

Contributors

LG Zhang was involved in conception and design of the study, acquisition of patient data, genotyping and drafted the paper. ZJ Xiao and GC Shi were involved in design of the study, genotyping and interpretation of results. J Huang and CX Zhang were involved in the design of the study, acquisition of patient data and interpretation of result. All authors revised the draft paper.

Funding

The project was supported by the Research and Development Foundation of Science and Technology of Henan Province (2012k15).

Competing interests

None.

Patient consent

Obtained.

Ethics approval

Ethics approval was provided by the First Hospital Affiliated to the Xinxiang Medical College.

Provenance and peer review

Not commissioned; externally peer reviewed.

Data sharing statement

No additional data are available.

Figure Legend

Fig. 1 The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the pair-wise calculation of r^2 (left) or D' (right). Black squares indicate $r^2 = 1$ (i.e. perfect LD between a pair of SNPs). Empty squares indicate D' = 1 (i.e. complete LD between a pair of SNPs).

Supplemental Fig. 1 The structure of the human ANXA11 gene and 15 SNPs located on the gene.

Reference

1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007;

357:2153-65.

2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-8

3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis. *Semin Respir Crit Care Med* 2010;31:380-9.

4. Hofmann S, Franke A, Fischer A, *et al*. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet* 2008;40:1103-6.

5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev* 2002;82:331-71.

6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using recombinant human annexin XI-specific autoantibodies cloned by phage display. *J Biol Chem* 2003;278:33120-6.

7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-22.

8. Petzmann S, Maercker C, Markert E, *et al*. Enhanced proliferation and decreased apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis Vasc Diffuse Lung Dis* 2006;23:190-200

9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, *et al*. Comparative analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol* 2006;26:243-50.

10. Agostini C, Trentin L, Facco M, *et al*. Role of IL-15, IL-2, and their receptors in

the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol* 1996;157:910-8

11. Jullien D, Sieling PA, Uyemura K, *et al*. IL-15, an immunomodulator of T cell responses in intracellular infection. *J Immunol* 1997;158:800-6

12. Muller-Quernheim J, Schurmann M, Hofmann S, *et al*. Genetics of sarcoidosis. *Clin Chest Med* 2008;29:391-414.

13. Morais A, Lima B, Peixoto M, *et al*. Annexin A11 gene polymorphism (R230C variant) and sarcoidosis in a Portuguese population. *Tissue antigens* 2013;82:186-91.

14. Li Y, Pabst S, Kubisch C, *et al*. First independent replication study confirms the strong genetic association of ANXA11 with sarcoidosis. *Thorax* 2010;65:939-40.

15. Gabriel SB, Schaffner SF, Nguyen H, *et al*. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-9.

16. Barrett JC, Fry B, Maller J, *et al*. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-5.

17. Curtis D, Knight J, Sham PC. Program report: GENECOUNTING support programs. *Ann Hum Genet* 2006;70:277-9.

18. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic associations. *Hum Hered* 2000;50:133-9.

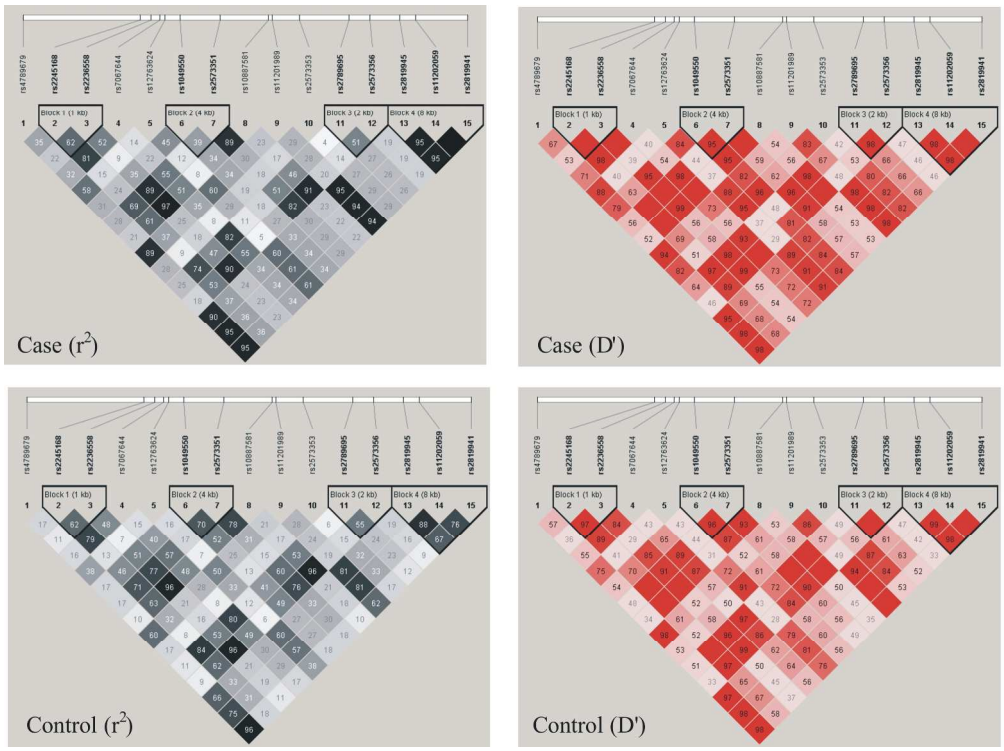
19. Zhao JH, Lissarrague S, Essioux L, et al. GENECOUNTING: haplotype analysis with missing genotypes. *Bioinformatics* 2002;18:1694-5

20. Zhao JH, Sham PC. Generic number systems and haplotype analysis. *Comput Methods Programs Biomed* 2003;70:1-9

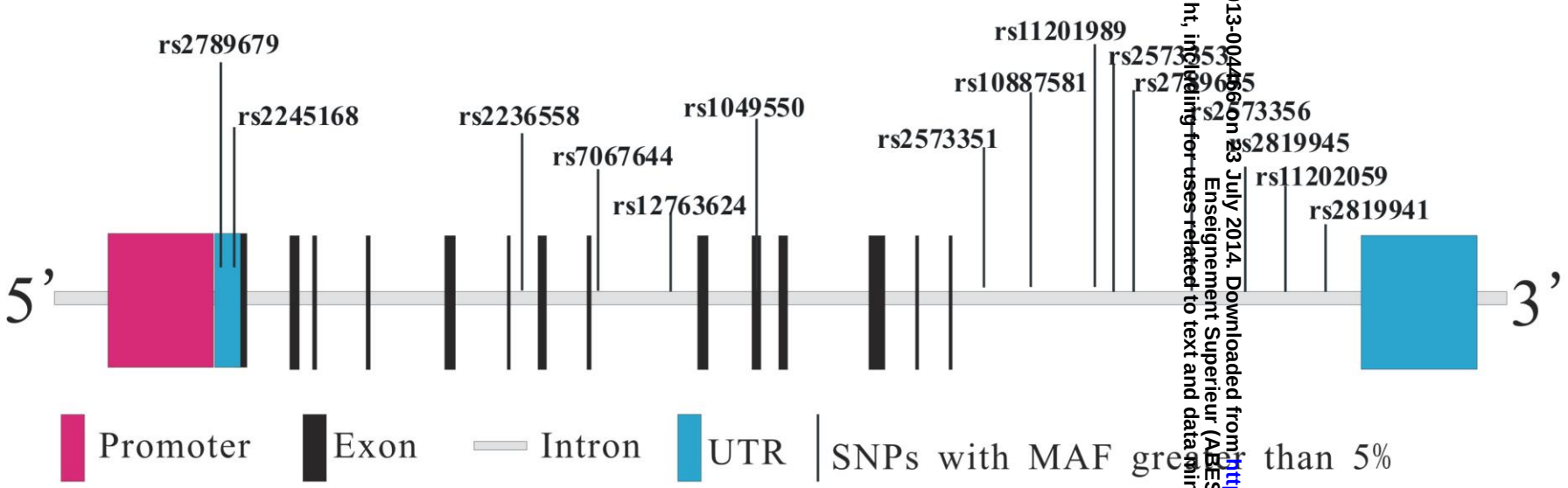
21. Mrazek F, Stahelova A, Kriegova E, *et al*. Functional variant ANXA11 R230C: true marker of protection and candidate disease modifier in sarcoidosis. *Genes Immun* 2011;12:490-4.

22. Levin AM, Iannuzzi MC, Montgomery CG, *et al*. Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes Immun* 2013;14:13-8.

24. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11 with the nuclear envelope. *J Biol Chem* 2003;278:20210-6.



273x203mm (300 x 300 DPI)



Supplemental table 1 Summary of all published sarcoidosis-association studies for ANXA11

Study	Population		Type of Study	Sample Size (n)		Number of SNPs typed	Positive SNPs
	Ethnic group	Country		Sarcoidosis	Control		
Hofmann S, 2008	Caucasian	Germany	Case-control	499	490	GWAS	rs2789679, rs7091565, rs1049550
Mrazek F, 2011	Caucasian	Czech	Case-control	245	254	1	rs1049550
Cozier YC, 2012	Caucasian	America	Case-control	486	943	10p12-10q22	rs2789679, rs2819941
Li Y, 2013	Caucasian	Germany	Case-control	325	364	1	rs1049550
Levin AM, 2013	Caucasian	America	Case-control	1689	1252	25	rs1049550
Morais A, 2013	Caucasian	Portugal	Case-control	208	197	1	rs1049550

Supplemental Table 2 Genotypic and allelic frequencies of ANXA11 polymorphisms in the controls and patients with sarcoidosis

Variable	**Controls** (n=418)		**Sarcoidosis** (n=412)		***P*-value**[a]	**OR, 95% CI**
	No.	%	No.	%		
rs2245168					0.920	
CC	156	37.3	150	36.4	0.791	1.04, 0.74-1.38
CT	200	47.9	197	47.8	0.955	0.99, 0.76-1.30
TT	62	14.8	65	15.8	0.781	0.95, 0.65-1.39
Per T allele	324	38.8	327	39.7	0.698	1.04, 0.85-1.27
rs2236558					0.172	
GG	118	28.2	104	25.2	0.72	1.06, 0.77-1.45
TG	190	45.5	214	51.9	0.060	0.77, 0.59-1.01
TT	110	26.3	94	22.8	0.073	1.33, 0.97-1.82
Per T allele	410	49.0	402	48.8	0.917	0.99, 0.82-1.20
rs7067644					0.173	
AA	152	36.4	176	42.7	0.059	0.76, 0.58-1.01
GA	206	49.3	182	44.2	0.137	1.23, 0.94-1.68
GG	60	14.4	54	13.1	0.597	1.11, 0.75-1.65
Per G allele	326	38.0	290	35.2	0.109	0.85, 0.70-1.04
rs12763624					0.118	
TT	126	75.6	150	36.4	0.008	0.68, 0.51-0.91
CT	216	51.7	186	45.2	0.059	1.30, 0.99-1.71
CC	76	18.2	76	18.5	0.411	1.16, 0.81-1.67
Per C allele	368	44.0	338	41.0	0.216	0.88, 0.73-1.08
rs2573351					0.437	

TT	120	28.7	104	25.2	0.263	1.19, 0.88-1.62
CT	202	48.3	216	52.4	0.238	0.85, 0.65-1.12
CC	96	23.0	92	22.3	0.826	1.04, 0.75-1.44
Per C allele	394	47.1	400	48.5	0.564	1.06, 0.87-1.28
rs10887581					0.290	
TT	114	27.3	96	23.3	0.190	1.23, 0.90-1.69
TC	190	45.5	208	50.5	0.148	0.82, 0.62-1.07
CC	114	27.3	108	26.2	0.729	1.06, 0.78-1.44
Per C allele	418	50.0	424	51.5	0.583	1.06,0.87-1.29
rs11201989					0.144	
GG	130	31.1	128	31.1	0.995	1.00, 0.75-1.34
GC	218	52.2	194	47.1	0.141	1.23, 0.93-1.62
CC	70	16.8	90	21.9	0.060	0.71, 0.50-1.01
Per C allele	358	42.8	374	45.4	0.29	1.11, 0.91-1.35
rs2573353					0.918	
CC	178	42.6	176	42.7	0.957	0.99, 0.75-1.31
CA	196	46.9	192	46.6	0.921	1.01, 0.77-1.33
AA	44	10.5	44	10.7	0.941	0.98, 0.63-1.53
Per A allele	284	34.0	280	34.0	0.997	1.00, 0.82-1.23
rs2789695					0.918	
TT	178	42.6	180	34.0	0.740	0.95, 0.73-1.26
CT	188	45.0	184	44.7	0.920	1.01, 0.77-1.33
CC	52	12.4	48	11.7	0.73	1.08, 0.71-1.64
Per C allele	292	34.9	280	34.0	0.69	0.96, 0.78-1.17
rs2573356					0.892	
CC	118	28.2	112	27.2	0.742	1.05, 0.78-1.43

CT	190	45.5	194	47.1	0.639	0.94, 0.71-1.23
TT	110	26.3	106	25.7	0.843	1.03, 0.76-1.41
Per T allele	410	49.0	406	49.3	0.926	0.99, 0.82-1.20
rs2819945					0.101	
GG	119	28.5	130	31.6	0.294	0.85, 0.63-1.15
GA	221	52.9	188	45.6	0.030	1.35, 1.03-1.78
AA	78	18.7	94	22.8	0.138	0.78, 0.55-1.09
Per A allele	377	45.1	376	45.6	0.827	1.02, 0.84-1.24
rs11202059					0.064	
GG	106	25.4	128	31.1	0.067	0.75, 0.56-1.02
GA	226	54.1	190	46.1	0.021	1.38, 1.05-1.81
AA	86	20.6	94	22.8	0.429	0.88, 0.63-1.22
Per A allele	398	47.6	378	45.9	0.479	0.93, 0.77-1.13

[a] *P*-value was calculated by 2×3 and 2×2 chi-squared tests based on codominant, dominant for the rare allele, heterosis and recessive for the rare allele models of inheritance.

Alpha value is adjusted by Bonferroni correction and statistically significant results (P<0.003)

Supplementary table 3 Multi-SNP analysis of rs2789679, rs1049550 and rs2819941

Models	testing SNP	*P*-value	OR	95%CI
model1: rs2789679+ rs1049550	rs2789679: A	0.11	1.19	0.96-1.48
model2: rs2819941+rs1049550	rs2819941: T	0.169	1.17	0.94-1.45
model3: rs2789670+rs1049550+rs2819941	rs1049550: T	0.0002	0.65	0.52-0.82

BMJ Open

Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a Han Chinese population - a case-control study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-004466.R4
Article Type:	Research
Date Submitted by the Author:	19-Jun-2014
Complete List of Authors:	Zhang, LG; the First Hospital Affiliated to the Xinxiang Medical College, The Third Department of Tuberculosis Xiao, ZJ; the First Hospital Affiliated to the Xinxiang Medical College, The Third Department of Tuberculosis Shi, GC; the First Hospital Affiliated to the Xinxiang Medical College, The Third Department of Tuberculosis J, Huang; the First Hospital Affiliated to the Xinxiang Medical College, The Third Department of Tuberculosis Zhang, CX; the First Hospital Affiliated to the Xinxiang Medical College, The Third Department of Tuberculosis
Primary Subject Heading:	Infectious diseases
Secondary Subject Heading:	Genetics and genomics
Keywords:	GENETICS, IMMUNOLOGY, INFECTIOUS DISEASES

Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a Han Chinese population - a case-control study

LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang

The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China

*Corresponding author: Zhijian Xiao, The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China. No.88 Jiankang Road, Weihui City, Henan, 463000, PR China

E-mail: zhangligongys@163.com

Tel: +86 0373 - 4402229 (office)

Fax: +86 0373 -4402251

Key words: Sarcoidosis; Annexin A11; Single-nucleotide polymorphism

Word count: 3918

Abstract

Objectives: To further identify the single-nucleotide polymorphisms (SNPs) that contribute to the genetic susceptibility to sarcoidosis, we examined the potential association between sarcoidosis and fifteen SNPs of the ANXA11 gene.

Design: A case - control study.

Setting: A tuberculosist unit in a hospital of the university in China.

Participants: Participants included 412 patients with sarcoidosis and 418 healthy controls.

Methods: The selected SNPs were genotyped using the MALDI-TOF in the MassARRAY system.

Results: The statistical significant differences were found in the allelic or genotypic frequencies of the rs2789679, rs1049550 and rs2819941 in the ANXA11 gene between the patients with sarcoidosis and the controls. The rs2789679 A allele (P = 0.00004, OR = 1.42, 95%CI = 1.17-1.73) and rs2819941 T allele (P = 0.0006, OR = 1.41, 95%CI = 1.16-1.71) were significantly more frequent in the patients with sarcoidosis compared to controls. The frequency of the rs1049550 T allele (P = 0.000002, OR = 0.61, 95%CI= 0.49 - 0.74) in the patients with sarcoidosis was significantly lower than that in the controls. The multi-SNP model reveals that rs1049550 is the only independent SNP association effect after accounting for the other two marginally associated SNPs. In block 2 (rs1049550-rs2573351), the T - C haplotype occurred significantly less frequently (P = 0.001), whereas the C - C haplotypes occurred more frequently (P = 0.0001) in the patients with sarcoidosis

than the controls. Furthermore, genotype frequency distribution revealed that, in rs1049550, CC genotype was significantly more in the patients with chest radiographic (CXR) stage I sarcoidosis than the patients with CXR stage II-IV sarcoidosis ($P = 0.012$).

Conclusions: These findings point to a role for the polymorphisms of ANXA11 in sarcoidosis in a Chinese Han population, and may be informative for future genetic studies on sarcoidosis.

Article summary

Article focus

1. The first genome-wide association study (GWAS) in sarcoidosis conducted in a German population has revealed that the rs1049550 (C > T, Arg230Cys) located in exon 10 were associated with sarcoidosis.

2. Whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations remain unclear.

3. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis in a Chinese Han population.

Key messages

1. ANXA11 rs2789679 (3'-UTR) was significantly associated with sarcoidosis.

2. ANXA11 rs1049550 (exon 10) was significantly associated with sarcoidosis.

3. Another significant association was observed for rs2819941 (intron 14).

Strengths and limitations of this study

1. A systematical screening of the functional SNPs in the promoter region, 5'- and 3'-UTR, exons of the ANXA11 gene, and the homogeneity of the study subjects representing the Chinese Han are the main strengths of this study.

2. The lack of data proving the positive association observed for rs2789679 and rs1049550 is a potential limitation of this study. Furthermore, the association of the serum level of ANXA11 with sarcoidosis still needs to be investigated.

INTRODUCTION

Sarcoidosis is a systemic autoimmune disease characterized by destructive, non-caseating epithelioid granulomatous lesions, with accumulated phagocytes and activated $CD4^+$ T helper type 1 lymphocytes [1 2]. The typical manifestations of sarcoidosis include bilateral hilar lymphadenopathy, pulmonary infiltration and ocular and skin lesions. The clinical course and prognosis of sarcoidosis are variable. The acute forms of sarcoidosis such as Lòfgren's syndrome (LS) can resolve spontaneously in most cases. However, chronic pulmonary sarcoidosis may progress to lung fibrosis which eventually causes respiratory failure. Recent studies have demonstrated that sarcoidosis can occur in a pattern of family clustering, and its racial incidence is also different, suggesting that some genetic factors may contribute to the risk and severity of sacroidosis [1 3].

The Annexin gene family is involved in the etiology of several autoimmune and chronic diseases [4 5.] One member of the Annexin gene family, ANXA11, located on chromosome 10q22.3, is involved in calcium signaling, apoptosis, vesicle trafficking, cell growth and terminal phase of cell division [5-7]. In this context, the development and maintenance of the granulomatous inflammation in sarcoidosis have been repeatedly associated with the impaired apoptosis of activated inflammatory cells [8 9]. Most of the early-stage sarcoidosis is characterized by mononuclear cell alveolitis dominated by activated $CD4^+$ T cells and macrophages. Between these cells, the uncoordinated interplay results in the formation of typical non-caseating granulomas. The mechanism by which the granulomas resolve has not been fully elucidated.

However, it is generally assumed that the induction of apoptosis and/or by the withdrawal of inflammatory cytokines participates in the disappearance of granulomas [10 11]. In the patients with sarcoidosis, their peripheral blood monocytes were characterized by apoptosis [9].

In previous studies, case - control/family 'hypothesis - driven' studies and low density linkage scans were used to identify genetic factors conferring the genetic susceptibility to sarcoidosis [12]. Notably, the first genome - wide association study (GWAS) in sarcoidosis conducted in a German population has recently revealed an association: the leading rs2789679 SNP (single - nucleotide polymorphism) located in the 3'-untranslated region (3'-UTR) and a common nonsynonymous SNP (rs1049550, C > T, Arg230Cys) were associated with the increased risk of sarcoidosis [4 13]. This association has recently been supported by another report from the same population [14]. Thus, more studies should be performed to demonstrate the following items: whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations. The published studies about the association of sarcoidosis and ANXA11 are summarized in Supplemental table 1. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis.

In this study, we investigated fifteen loci in a Chinese population from He'nan province (China) to verify the putative association between ANXA11 polymorphisms and sarcoidosis.

SUBJECTS AND METHODS

Subjects

Four hundred and twelve patients with sarcoidosis (mean ± SD age: 53.6 ± 4.6 years; 155 men and 257 women) were recruited from our hospital between May 2005 and March 2013. Sarcoidosis was diagnosed by the evidence of non-caseating epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities. Chronic sarcoidosis was defined as a disease over at least 2 years or at least two episodes in a lifetime. A cute sarcoidosis was defined as one episode of acute sarcoidosis which had totally resolved at the date of the examination. The control group consisted of 418 unrelated healthy subjects (mean ± SD age: 54.2 ± 5.3 years; 160 men and 258 women) who underwent health examinations in the Medical Examination Center of the First Affiliated Hospital of the Xinxiang Medical College (Xinxiang, China) between Oct 2009 and Sept 2013. None of the individuals in the control group had a history of lung diseases or showed any symptoms of the lung or other diseases by chest radiography or laboratory blood tests. Participants were excluded if they: were taking other prescribed medications that could affect the central nervous system; had a history of seizures, hematological diseases, or severe liver or kidney impairment; or were pregnant. No familial relationship was known between the study participants. The study was performed according to the Guidelines of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang, China). Written informed consent was obtained from each participant in this study.

SNP selection

Tagging SNPs were selected from the catalogs of the International HapMap Project. The rs2789679 and rs2245168 are located in 5'-UTR, rs2236558 in intron 6, rs7067644 and rs12763624 in intron 8, rs1049550 in exon 10, rs2573351 in intron 13, rs10887581, rs11201989, rs2573353, rs2789695, rs2573356, rs2819945, rs11202059 and rs2819941 in intron 14 (Supplemental figure 1). Marker selection was based on the following criteria. (a) We used the CHB data from the HapMap (release 27) to select tagSNPs for ANXA11. (b) We restricted our search for tagSNPs from base pair 80155124 to 80205677. (c) Further, we limited the SNPs to tag to those with a minor allele frequency ≥ 0.2 in the CHB. (d) Based on these restrictions, there were a total of 29 SNPs. (e) Using HAPLOVIEW with an r^2 threshold ≥ 0.8, 15 tagSNP were selected and used for subsequent analyses.

Genotyping

Peripheral blood was collected from a vein into a sterile tube coated with ethylenediamine tetraacetic acid (EDTA). Plasma samples were stored at -20°C. Genomic DNA was extracted from the frozen peripheral blood samples using a QIAmp Blood Mini Kit (Qiagen Inc., Valencia, CA, USA) according to the manufacturer's protocols. The selected SNPs were genotyped in cases and controls by using the MALDI-TOF in the MassARRAY system (Sequenom Inc., San Diego, CA, USA). Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA). The completed genotyping reactions were spotted onto a 384 well spectroCHIP (Sequenom) using the MassARRAY Nanodispenser (Sequenom) and determined by the matrix - assisted laser desorption ionization time -

of - flight mass spectrometer. Genotype calling was performed in realtime with the MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer software version 3.4 (Sequenom).

Statistical analysis

All data were analyzed using the SPSS 17.0 software (SPSS Inc., Chicago, IL, USA). Hardy-Weinberg equilibrium was evaluated by Chi-square tests. Differences between the cases and controls in the frequency of the alleles, genotypes and haplotypes were evaluated by the Fisher's exact test or the Pearson Chi-square test. Unconditional logistic regression was used to calculate the odds ratio (OR) and 95% confidence interval (CI) in independent association between each locus and the presence of sarcoidosis. Gender and age of subjects were treated as covariants in binary logistic regression. *P* values were calculated based on codominant, dominant for the rare allele, heterosis and recessive for the rare allele models of inheritance. Models of multiple logistic regression were used to test the independence of individual allelic effect. In detail, the most significant SNP was chosen to be the conditional SNP (covariate in the regression model) when testing other significant SNPs; Meanwhile, a multi-SNP model including all significant SNPs was also performed. The Bonferroni correction was used to adjust the test level when multiple comparisons were conducted, and the *P* value was divided by the total number of loci. Haplotype blocks were defined according to the criteria developed by Gabriel et al. [15]. Pair-wise LD statistics (D' and r^2) and haplotype frequency were calculated, and haplotype blocks were constructed using the Haploview 4.0 [16]. To ensure that the LD

blocks most closely reflect the population level LD patterns, definition of the blocks were based on the control samples alone. The haplotype frequencies were estimated using GENECOUNTING, which computes maximum - likelihood estimates of haplotype frequencies from unknown phase data by utilizing an expectation - maximization algorithm [17-20]. The significance of any haplotypic association was evaluated using a likelihood ratio test, followed by permutation testing that compared estimated haplotype frequencies in cases and controls [17 19].

RESULTS

The genotype distribution of the fifteen polymorphisms was consistent with the Hardy-Weinberg equilibrium ($P > 0.05$). The analysis of strong LD in the patients with sarcoidosis and healthy controls revealed that 2 SNPs (rs2245168, rs2236558), 2 SNPs (rs1049550, rs2573351), 2 SNPs (rs2789695, rs2573356) and 3 SNPs (rs2819945, rs11202059, rs2819941) were located in haplotype block 1, block 2, block 3 and block 4 (D' > 0.9, Fig. 1). The genotype distribution, allelic frequencies, and haplotypes in the patients with sarcoidosis and healthy controls are showed in Tables 1-3 and Supplemental table 2

Comparison of genotype and allele frequency distribution revealed significant differences between the patients with sarcoidosis and healthy controls for 3 SNPs: rs2789679, rs1049550 and rs2819941. The frequency of the rs1049550 T allele (P = 0.000002, OR = 0.61, 95%CI= 0.49-0.74) in the patients with sarcoidosis was significantly lower than that in the controls. These differences remained statistically significant after Bonferroni corrections. The rs1049550 was the most significant of the

three. A multi-SNP model revealed rs1049550 to be the only significant independent SNP association with sarcoidosis risk.

The multi-SNP model showed only rs1049550 present a significant effect on the disease phenotype ($p < 0.001$). No independent effect was found for rs2789679 or rs2819941 when adjusting for the effect of rs1049550. Both individual SNP and multi - SNP analysis supported rs1049550: T allele was an important protective factor for affecting sarcoidosis (odds ratio around 0.6). In another word, carriers of rs1049550: C allele have higher susceptibility to sarcoidosis in our Chinese Han population (Supplemental table 3).

We performed an association analysis to determine whether the haplotype was associated with the risk of sarcoidosis (Global $P < 0.05$ in block 2 and block 3). Compared with the controls, the T - C haplotype occurred significantly less frequently ($P = 0.001$) but the C - C haplotypes occurred more frequently ($P = 0.0001$) in block 2 (rs1049550 - rs2573351) in the patients with sarcoidosis.

To assess particular disease phenotypes, the patients with sarcoidosis were divided into the subgroups according to their chest radiographic (CXR) stage. CXR stage I (isolated bilateral hilar lymphadenopathy): N = 196; CXR stages II - IV (infiltration of parenchyma): 167; CXR stage 0 (absence on pulmonary disease): N = 28; the information on CXR stage was not available for 21 patients. Genotype frequency distribution revealed that, in rs1049550, CC genotype was significantly more in the patients with chest radiographic (CXR) stage I sarcoidosis than the patients with CXR stage II-IV sarcoidosis ($P = 0.012$). (Table 4).

DISCUSSION

A key step in linkage and association studies is to identify common risk variants in different populations. To determine if common risk variants exist in distinct populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene, were genotyped in samples from the patients with sarcoidosis and healthy controls in a Chinese Han population. Recently, new sarcoidosis loci have been identified by genome-wide association studies [4 14]. With the fast development of genome - wide association studies, an increasing number of susceptibility loci for sarcoidosis have been found in different populations [4 14 21]. However, these observations should be confirmed in other genetically independent populations. In this study, we conducted the first large genetic association study of the ANXA11 gene in a Chinese Han population. The evidence of markers associated with sarcoidosis was presented, and these markers were mapped to different locations in the ANXA11 gene (81897864 - 81951001). The association signals in the region were identified, and some significantly associated haplotypes also appeared this region.

Hoffman et al. reported an association between the T allele of the non - synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a German population [4]. Similar results were also obtained in other two European populations [14 21]. As a part of the sarcoidosis GWAS done in Americans [22], we further confirmed this association in a Chinese Han population. In this study, the frequency of ANXA11 rs1049550 T allele was significantly lower in the patients with sarcoidosis than the healthy controls. Genotype frequency distribution revealed that, in rs1049550,

CC genotype was significantly more in the patients with stage I sarcoidosis than the patients with stage II-IV sarcoidosis. The "protective" effect of ANXA11 T allele increased with the number of its copies in the genotype, which is consistent with the result obtained in the patients with sarcoidosis in a German population [4], Furthermore, the T allele carriers among the patients were protected from infiltration of lung parenchyma (radiographic stages II-IV) [21]. We have also demonstrated that rs1049550 is significant cross - ethnically at the gene level after adjustment for the single SNP association tests performed. The mechanism by which rs1049550 affects the susceptibility to sarcoidosis is that the SNP may affect the function of the ANXA11 protein rather than affect its expression. In humans, the ANXA11 protein consists of an N - terminalproline - tyrosine - glycine (PYG) - rich region, followed by four annexin core domains. The rs1049550 leads to an amino - acid exchange (basic arginine to a polar cysteine) at the evolutionarily conserved position 230 (R230C) in the first annexin domain, which is responsible for Ca^{2+} - dependent trafficking of the protein in the cell [23].

In the GWAS conducted by Hofmann et al [4], the strongest association signal was observed for rs2789679. This marker is located in the 3'-UTR of the ANXA11 gene. The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls being TT homozygous. Fine-mapping in the Black Women's Health Study (BWHS) revealed rs2819941 to be the top SNP, associated with a non-significant (P = 0.06) reduction in the risk of sarcoidosis (17% reduction per copy of the C-allele). In this

case-control association study, the frequency of the A allele in rs2789679 and the rs2819941 T allele frequency in the patients with sarcoidosis was significantly higher than that in the healthy controls. Several lines of evidence suggest that the observed association is unlikely to be an artifact. First, both the single-SNP and the haplotype-based association analyses support the current finding. Second, our samples were from the same geographical region, excluding the. Finally, consistent results were obtained from two genetically independent populations (Chinese Han and Europeans). Collectively, our results confirmed the strong association between ANXA11 polymorphisms and sarcoidosis, suggesting that ANXA11 represents a strong genetic risk factor for sarcoidosis. Since causal SNP in ANXA11 was not known, all the SNPs significantly associated with sarcoidosis were actually indirect association through the real causal one. Though rs2789679 and rs2819941 were not in LD with rs1049550, it may not tag the causal SNP as well as rs1049550. Hence, their individual effects were eliminated when controlling for the effect of rs1049550.

We further investigated the interaction among polymorphisms and observed strong LD. The haplotype analysis revealed that significantly more C - C (block 2, rs1049550 - rs2573351) and significantly fewer T - C haplotypes (block 2) were found in the patients with sarcoidosis. These results indicated that the patients with C - C (block 2) haplotypes of the ANXA11 gene were more prone to sarcoidosis. Significantly higher frequencies of T - C haplotypes were detected in the healthy controls than in the patients with sarcoidosis, suggesting that they may show protective effects against sarcoidosis. Our sample size can detect SNP and haplotype

associations with 90% and 85% power, respectively, at a false positive rate of 5%, disease prevalence of 1‰, disease allele / haplotype frequency of 0.05 / 0.03, and a presumed odds ratio (OR) of 1.5. To some extent, this finding further supports a role of ANXA11 polymorphisms in sarcoidosis, while ethnic group difference may exist.

In conclusion, our study suggests a potential role of the ANXA11 SNPs and their related haplotypes in the genetic susceptibility to sarcoidosis. Further studies are needed to investigate how these SNPs affect the function of ANXA11. A broader examination of the genetic variation in ANXA11 in the Han Chinese may reveal other variants associated with disease risk.

Contributors

LG Zhang was involved in conception and design of the study, acquisition of patient data, genotyping and drafted the paper. ZJ Xiao and GC Shi were involved in design of the study, genotyping and interpretation of results. J Huang and CX Zhang were involved in the design of the study, acquisition of patient data and interpretation of result. All authors revised the draft paper.

Funding

The project was supported by the Research and Development Foundation of Science and Technology of Henan Province (2012k15).

Competing interests

None.

Patient consent

Obtained.

Ethics approval

Ethics approval was provided by the First Hospital Affiliated to the Xinxiang Medical College.

Provenance and peer review

Not commissioned; externally peer reviewed.

Data sharing statement

No additional data are available.

Reference

1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007; 357:2153-65.
2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-8
3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis. *Semin Respir Crit Care Med* 2010;31:380-9.
4. Hofmann S, Franke A, Fischer A, *et al*. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet* 2008;40:1103-6.
5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev* 2002;82:331-71.
6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using recombinant human annexin XI-specific autoantibodies cloned by phage display. *J Biol Chem* 2003;278:33120-6.
7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-22.
8. Petzmann S, Maercker C, Markert E, *et al*. Enhanced proliferation and decreased apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis Vasc Diffuse Lung Dis* 2006;23:190-200
9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, *et al*. Comparative analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol*

2006;26:243-50.

10. Agostini C, Trentin L, Facco M, *et al*. Role of IL-15, IL-2, and their receptors in the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol* 1996;157:910-8

11. Jullien D, Sieling PA, Uyemura K, *et al*. IL-15, an immunomodulator of T cell responses in intracellular infection. *J Immunol* 1997;158:800-6

12. Muller-Quernheim J, Schurmann M, Hofmann S, *et al*. Genetics of sarcoidosis. *Clin Chest Med* 2008;29:391-414.

13. Morais A, Lima B, Peixoto M, *et al*. Annexin A11 gene polymorphism (R230C variant) and sarcoidosis in a Portuguese population. *Tissue antigens* 2013;82:186-91.

14. Li Y, Pabst S, Kubisch C, *et al*. First independent replication study confirms the strong genetic association of ANXA11 with sarcoidosis. *Thorax* 2010;65:939-40.

15. Gabriel SB, Schaffner SF, Nguyen H, *et al*. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-9.

16. Barrett JC, Fry B, Maller J, *et al*. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-5.

17. Curtis D, Knight J, Sham PC. Program report: GENECOUNTING support programs. *Ann Hum Genet* 2006;70:277-9.

18. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic associations. *Hum Hered* 2000;50:133-9.

19. Zhao JH, Lissarrague S, Essioux L, et al. GENECOUNTING: haplotype analysis with missing genotypes. *Bioinformatics* 2002;18:1694-5

20. Zhao JH, Sham PC. Generic number systems and haplotype analysis. *Comput Methods Programs Biomed* 2003;70:1-9

21. Mrazek F, Stahelova A, Kriegova E, *et al*. Functional variant ANXA11 R230C: true marker of protection and candidate disease modifier in sarcoidosis. *Genes Immun* 2011;12:490-4.

22. Levin AM, Iannuzzi MC, Montgomery CG, *et al*. Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes Immun* 2013;14:13-8.

24. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11 with the nuclear envelope. *J Biol Chem* 2003;278:20210-6.

Figure Legend

Fig. 1 The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the pair-wise calculation of r^2 (left) or D' (right). Black squares indicate $r^2 = 1$ (i.e. perfect LD between a pair of SNPs). Empty squares indicate D' = 1 (i.e. complete LD between a pair of SNPs).

Supplemental Fig. 1 The structure of the human ANXA11 gene and 15 SNPs located on the gene.

Table 1 Genotypic and allelic frequencies of ANXA11 polymorphisms in the controls and patients with sarcoidosis

Variable	**Controls** (n=418)		**Sarcoidosis** (n=412)		*P*-value[a]	OR, 95% CI
	No.	%	No.	%		
rs2789679					0.0007	
TT	114	27.3	94	22.8	0.140	1.27, 0.93-1.74
AT	225	53.8	186	45.2	0.012	1.42, 1.08-1.87
AA	79	18.9	132	32.0	0.0002	0.49, 0.36-0.68
Per A allele	383	45.8	450	54.6	0.00004	1.42, 1.17-1.73
rs1049550					0.0002	
CC	154	36.8	208	50.5	0.0007	0.57, 0.43-0.75
CT	190	45.5	170	41.3	0.221	1.19, 0.90-1.56
TT	74	17.7	34	8.3	0.0008	2.39, 1.55-3.68
Per T allele	338	40.4	238	28.8	0.000002	0.61, 0.49-0.74
rs2819941					0.0002	
CC	114	27.3	94	22.8	0.141	1.27, 0.93-1.74
CT	226	54.1	190	46.1	0.021	1.38, 1.05-1.81
TT	78	18.7	128	31.1	0.0004	0.51, 0.37-0.70
Per T allele	382	45.7	446	54.1	0.0006	1.41, 1.16-1.71

[a] *P*-value was calculated by 2×3 and 2×2 chi-squared tests based on codominant, dominant for the rare allele, heterosis and recessive for the rare allele models of inheritance.

Alpha value is adjusted by Bonferroni correction and statistically significant results (P<0.003)

Table 2 ANXA11 haplotype in block 2 frequencies and the results of their associations with risk of sarcoidosis

Haplotype			Genecounting (frequency %)			
ID	**rs1049550**	**rs2573351**	**Cases**	**Controls**	***P*-value[a]**	**Global *P*[b]**
HAP1	T	C	27.7	39.7	0.001	0.003
HAP2	C	C	19.9	7.4	0.0001	
HAP3	C	T	50.5	51.0	0.892	

[a] Based on 10,000 permutations.

[b] Based on comparison of frequency distribution of all haplotypes for the combination of SNPs.

Table 3 ANXA11 haplotype in block 3 frequencies and the results of their associations with risk of sarcoidosis

Haplotype			Genecounting (frequency %)			
ID	**rs2789695**	**rs2573356**	**Cases**	**Controls**	***P*-value[a]**	**Global *P*[b]**
HAP1	C	T	33.7	34.9	0.718	0.045
HAP2	T	T	15.3	14.1	0.632	
HAP2	T	T	15.6	14.1	0.565	

[a] Based on 10,000 permutations.

[b] Based on comparison of frequency distribution of all haplotypes for the combination of SNPs.

Table 4 Chest radiographic (CXR) stages of sarcoidosis patients by genotype of rs2789679, rs1049550 and rs2819941

Stages	rs2789679 (n, %)			rs1049550* (n, %)			rs2819941 (n, %)		
	AA	AT	TT	CC	CT	TT	**CC**	**CT**	**TT**
stage I	59 (30.1)	92 (47.0)	45 (23.0)	101 (51.5)	82 (41.8)	13 (6.6)	47 (25.0)	88 (44.9)	61 (31.1)
stages II - IV	46 (27.6)	75 (44.9)	46 (27.6)	72 (43.1)	68 (40.7)	27 (16.2)	41 (24.6)	73 (43.7)	53 (31.7)
χ^2, *P*	1.041, 0.594			8.807, 0.012			0.052, 0.975		

* *P* values for genotype frequency distribution ($P < 0.05$).

Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a Han Chinese population - a case-control study

LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang

The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China

*Corresponding author: Zhijian Xiao, The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China. No.88 Jiankang Road, Weihui City, Henan, 463000, PR China

E-mail: zhangligongys@163.com

Tel: +86 0373 - 4402229 (office)

Fax: +86 0373 -4402251

Key words: Sarcoidosis; Annexin A11; Single-nucleotide polymorphism

Word count: 3918

Abstract

Objectives: To further identify the single-nucleotide polymorphisms (SNPs) that contribute to the genetic susceptibility to sarcoidosis, we examined the potential association between sarcoidosis and fifteen SNPs of the ANXA11 gene.

Design: A case - control study.

Setting: A tuberculosist unit in a hospital of the university in China.

Participants: Participants included 412 patients with sarcoidosis and 418 healthy controls.

Methods: The selected SNPs were genotyped using the MALDI-TOF in the MassARRAY system. ~~Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA).~~

Results: The statistical significant differences were found in the allelic or genotypic frequencies of the rs2789679, rs1049550 and rs2819941 in the ANXA11 gene between the patients with sarcoidosis and the controls. The rs2789679 A allele (P = 0.00004, OR = 1.42, 95%CI = 1.17-1.73) and rs2819941 T allele (P = 0.0006, OR = 1.41, 95%CI = 1.16-1.71) were significantly more frequent in the patients with sarcoidosis compared to controls. The frequency of the rs1049550 T allele (P = 0.000002, OR = 0.61, 95%CI= 0.49 - 0.74) in the patients with sarcoidosis was significantly lower than that in the controls. The multi-SNP model reveals that rs1049550 is the only independent SNP association effect after accounting for the other two marginally associated SNPs. ~~The strong linkage disequilibrium (LD) was observed in four blocks (D' > 0.9).~~ In block 2 (rs1049550-rs2573351), the T - C

haplotype occurred significantly less frequently (P = 0.001), whereas the C - C haplotypes occurred more frequently (P = 0.0001) in the patients with sarcoidosis than the controls. Furthermore, genotype frequency distribution revealed that, in rs1049550, CC genotype was significantly more in the patients with chest radiographic (CXR) stage I sarcoidosis than the patients with CXR stage II-IV sarcoidosis (P = 0.012).

Conclusions: These findings point to a role for the polymorphisms of ANXA11 in sarcoidosis in a Chinese Han population, and may be informative for future genetic studies on sarcoidosis.

Article summary

Article focus

1. The first genome-wide association study (GWAS) in sarcoidosis conducted in a German population has revealed that the rs1049550 (C > T, Arg230Cys) located in exon 10 were associated with sarcoidosis.

2. Whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations remain unclear.

3. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis in a Chinese Han population.

Key messages

1. ANXA11 rs2789679 (3'-UTR) was significantly associated with sarcoidosis.

2. ANXA11 rs1049550 (exon 10) was significantly associated with sarcoidosis.

3. Another significant association was observed for rs2819941 (intron 14).

Strengths and limitations of this study

1. A systematical screening of the functional SNPs in the promoter region, 5'- and 3'-UTR, exons of the ANXA11 gene, and the homogeneity of the study subjects representing the Chinese Han are the main strengths of this study.

2. The lack of data proving the positive association observed for rs2789679 and rs1049550 is a potential limitation of this study. Furthermore, the association of the serum level of ANXA11 with sarcoidosis still needs to be investigated.

INTRODUCTION

Sarcoidosis is a systemic autoimmune disease characterized by destructive, non-caseating epithelioid granulomatous lesions, with accumulated phagocytes and activated $CD4^+$ T helper type 1 lymphocytes [1 2]. The typical manifestations of sarcoidosis include bilateral hilar lymphadenopathy, pulmonary infiltration and ocular and skin lesions. The clinical course and prognosis of sarcoidosis are variable. The acute forms of sarcoidosis such as Lòfgren's syndrome (LS) can resolve spontaneously in most cases. However, chronic pulmonary sarcoidosis may progress to lung fibrosis which eventually causes respiratory failure. Recent studies have demonstrated that sarcoidosis can occur in a pattern of family clustering, and its racial incidence is also different, suggesting that some genetic factors may contribute to the risk and severity of sacroidosis [1 3].

The Annexin gene family is involved in the etiology of several autoimmune and chronic diseases [4 5.] One member of the Annexin gene family, ANXA11, located on chromosome 10q22.3, is involved in calcium signaling, apoptosis, vesicle trafficking, cell growth and terminal phase of cell division [5-7]. In this context, the development and maintenance of the granulomatous inflammation in sarcoidosis have been repeatedly associated with the impaired apoptosis of activated inflammatory cells [8 9]. Most of the early-stage sarcoidosis is characterized by mononuclear cell alveolitis dominated by activated $CD4^+$ T cells and macrophages. Between these cells, the uncoordinated interplay results in the formation of typical non-caseating granulomas. The mechanism by which the granulomas resolve has not been fully elucidated. However, it is generally assumed that the induction of apoptosis and/or by the

withdrawal of inflammatory cytokines participates in the disappearance of granulomas [10 11]. In the patients with sarcoidosis, their peripheral blood monocytes were characterized by apoptosis [9].

In previous studies, case - control/family 'hypothesis - driven' studies and low density linkage scans were used to identify genetic factors conferring the genetic susceptibility to sarcoidosis [12]. Notably, the first genome - wide association study (GWAS) in sarcoidosis conducted in a German population has recently revealed an association: the leading rs2789679 SNP (single - nucleotide polymorphism) located in the 3'-untranslated region (3'-UTR) and a common nonsynonymous SNP (rs1049550, C > T, Arg230Cys) were associated with the increased risk of sarcoidosis [4 13]. This association has recently been supported by another report from the same population [14]. Thus, more studies should be performed to demonstrate the following items: whether these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with other causative SNPs and are repeated in other populations. The published studies about the association of sarcoidosis and ANXA11 are summarized in Supplemental table 1. We hypothesize that common variants in the ANXA11 gene may significantly contribute to the predisposition to develop sarcoidosis.

In this study, we investigated fifteen loci in a Chinese population from He'nan province (China) to verify the putative association between ANXA11 polymorphisms and sarcoidosis.

SUBJECTS AND METHODS

Subjects

Four hundred and twelve patients with sarcoidosis (mean ± SD age: 53.6 ± 4.6 years; 155 men and 257 women) were recruited from our hospital between May 2005 and March 2013. Sarcoidosis was diagnosed by the evidence of non-caseating epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities. Chronic sarcoidosis was defined as a disease over at least 2 years or at least two episodes in a lifetime. A cute sarcoidosis was defined as one episode of acute sarcoidosis which had totally resolved at the date of the examination. The control group consisted of 418 unrelated healthy subjects (mean ± SD age: 54.2 ± 5.3 years; 160 men and 258 women) who underwent health examinations in the Medical Examination Center of the First Affiliated Hospital of the Xinxiang Medical College (Xinxiang, China) between Oct 2009 and Sept 2013. None of the individuals in the control group had a history of lung diseases or showed any symptoms of the lung or other diseases by chest radiography or laboratory blood tests. Participants were excluded if they: were taking other prescribed medications that could affect the central nervous system; had a history of seizures, hematological diseases, or severe liver or kidney impairment; or were pregnant. No familial relationship was known between the study participants. The study was performed according to the Guidelines of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang, China). Written informed consent was obtained from each participant in this study.

SNP selection

Tagging SNPs were selected from the catalogs of the International HapMap Project. The rs2789679 and rs2245168 are located in 5'-UTR, rs2236558 in intron 6,

rs7067644 and rs12763624 in intron 8, rs1049550 in exon 10, rs2573351 in intron 13, rs10887581, rs11201989, rs2573353, rs2789695, rs2573356, rs2819945, rs11202059 and rs2819941 in intron 14 (Supplemental figure 1). Marker selection was based on the following criteria. (a) We used the CHB data from the HapMap (release 27) to select tagSNPs for ANXA11. (b) We restricted our search for tagSNPs from base pair 80155124 to 80205677. (c) Further, we limited the SNPs to tag to those with a minor allele frequency ≥ 0.2 in the CHB. (d) Based on these restrictions, there were a total of 29 SNPs. (e) Using HAPLOVIEW with an r^2 threshold ≥ 0.8, 15 tagSNP were selected and used for subsequent analyses.

Genotyping

Peripheral blood was collected from a vein into a sterile tube coated with ethylenediamine tetraacetic acid (EDTA). Plasma samples were stored at -20°C. Genomic DNA was extracted from the frozen peripheral blood samples using a QIAmp Blood Mini Kit (Qiagen Inc., Valencia, CA, USA) according to the manufacturer's protocols. The selected SNPs were genotyped in cases and controls by using the MALDI-TOF in the MassARRAY system (Sequenom Inc., San Diego, CA, USA). Probes and primers were designed using the Assay Design Software (Sequenom, San Diego, CA, USA). The completed genotyping reactions were spotted onto a 384 well spectroCHIP (Sequenom) using the MassARRAY Nanodispenser (Sequenom) and determined by the matrix - assisted laser desorption ionization time - of - flight mass spectrometer. Genotype calling was performed in realtime with the MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer

software version 3.4 (Sequenom).

Statistical analysis

All data were analyzed using the SPSS 17.0 software (SPSS Inc., Chicago, IL, USA). Hardy-Weinberg equilibrium was evaluated by Chi-square tests. Differences between the cases and controls in the frequency of the alleles, genotypes and haplotypes were evaluated by the Fisher's exact test or the Pearson Chi-square test. Unconditional logistic regression was used to calculate the odds ratio (OR) and 95% confidence interval (CI) in independent association between each locus and the presence of sarcoidosis. Gender and age of subjects were treated as covariants in binary logistic regression. *P* values were calculated based on codominant, dominant for the rare allele, heterosis and recessive for the rare allele models of inheritance. Models of multiple logistic regression were used to test the independence of individual allelic effect. In detail, the most significant SNP was chosen to be the conditional SNP (covariate in the regression model) when testing other significant SNPs; Meanwhile, a multi-SNP model including all significant SNPs was also performed. The Bonferroni correction was used to adjust the test level when multiple comparisons were conducted, and the *P* value was divided by the total number of loci. Haplotype blocks were defined according to the criteria developed by Gabriel et al. [15]. Pair-wise LD statistics (D' and r^2) and haplotype frequency were calculated, and haplotype blocks were constructed using the Haploview 4.0 [16]. To ensure that the LD blocks most closely reflect the population level LD patterns, definition of the blocks were based on the control samples alone. The haplotype frequencies were estimated

using GENECOUNTING, which computes maximum - likelihood estimates of haplotype frequencies from unknown phase data by utilizing an expectation - maximization algorithm [17-20]. The significance of any haplotypic association was evaluated using a likelihood ratio test, followed by permutation testing that compared estimated haplotype frequencies in cases and controls [17 19].

RESULTS

The genotype distribution of the fifteen polymorphisms was consistent with the Hardy-Weinberg equilibrium ($P > 0.05$). The analysis of strong LD in the patients with sarcoidosis and healthy controls revealed that 2 SNPs (rs2245168, rs2236558), 2 SNPs (rs1049550, rs2573351), 2 SNPs (rs2789695, rs2573356) and 3 SNPs (rs2819945, rs11202059, rs2819941) were located in haplotype block 1, block 2, block 3 and block 4 (D' > 0.9, Fig. 1). The genotype distribution, allelic frequencies, and haplotypes in the patients with sarcoidosis and healthy controls are showed in Tables 1-3 and Supplemental table 2

Comparison of genotype and allele frequency distribution revealed significant differences between the patients with sarcoidosis and healthy controls for 3 SNPs: rs2789679, rs1049550 and rs2819941. ~~The rs2789679 A allele ($P = 0.00004$, OR = 1.42, 95%CI = 1.17-1.73) and rs2819941 T allele ($P = 0.0006$, OR = 1.41, 95%CI = 1.16-1.71) were significantly more frequent in the patients with sarcoidosis compared to controls.~~ The frequency of the rs1049550 T allele ($P = 0.000002$, OR = 0.61, 95%CI= 0.49-0.74) in the patients with sarcoidosis was significantly lower than that in the controls. These differences remained statistically significant after Bonferroni

corrections. The rs1049550 was the most significant of the three. A multi-SNP model revealed rs1049550 to be the only significant independent SNP association with sarcoidosis risk.

The multi-SNP model showed only rs1049550 present a significant effect on the disease phenotype ($p < 0.001$). No independent effect was found for rs2789679 or rs2819941 when adjusting for the effect of rs1049550. Both individual SNP and multi - SNP analysis supported rs1049550: T allele was an important protective factor for affecting sarcoidosis (odds ratio around 0.6). In another word, carriers of rs1049550: C allele have higher susceptibility to sarcoidosis in our Chinese Han population (Supplemental table 3).

We performed an association analysis to determine whether the haplotype was associated with the risk of sarcoidosis (Global $P < 0.05$ in block 2 and block 3). Compared with the controls, the T - C haplotype occurred significantly less frequently ($P = 0.001$) but the C - C haplotypes occurred more frequently ($P = 0.0001$) in block 2 (rs1049550 - rs2573351) in the patients with sarcoidosis.

To assess particular disease phenotypes, the patients with sarcoidosis were divided into the subgroups according to their chest radiographic (CXR) stage. CXR stage I (isolated bilateral hilar lymphadenopathy): N = 196; CXR stages II - IV (infiltration of parenchyma): 167; CXR stage 0 (absence on pulmonary disease): N = 28; the information on CXR stage was not available for 21 patients. Genotype frequency distribution revealed that, in rs1049550, CC genotype was significantly more in the patients with chest radiographic (CXR) stage I sarcoidosis than the

patients with CXR stage II-IV sarcoidosis ($P = 0.012$). (Table 4).

DISCUSSION

A key step in linkage and association studies is to identify common risk variants in different populations. To determine if common risk variants exist in distinct populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene, were genotyped in samples from the patients with sarcoidosis and healthy controls in a Chinese Han population. Recently, new sarcoidosis loci have been identified by genome-wide association studies [4 14]. With the fast development of genome - wide association studies, an increasing number of susceptibility loci for sarcoidosis have been found in different populations [4 14 21]. However, these observations should be confirmed in other genetically independent populations. In this study, we conducted the first large genetic association study of the ANXA11 gene in a Chinese Han population. The evidence of markers associated with sarcoidosis was presented, and these markers were mapped to different locations in the ANXA11 gene (81897864 - 81951001). The association signals in the region were identified, and some significantly associated haplotypes also appeared this region.

Hoffman et al. reported an association between the T allele of the non - synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a German population [4]. Similar results were also obtained in other two European populations [14 21]. As a part of the sarcoidosis GWAS done in Americans [22], we further confirmed this association in a Chinese Han population. In this study, the frequency of ANXA11 rs1049550 T allele was significantly lower in the patients with sarcoidosis

than the healthy controls. Genotype frequency distribution revealed that, in rs1049550, CC genotype was significantly more in the patients with stage I sarcoidosis than the patients with stage II-IV sarcoidosis. The "protective" effect of ANXA11 T allele increased with the number of its copies in the genotype, which is consistent with the result obtained in the patients with sarcoidosis in a German population [4], Furthermore, the T allele carriers among the patients were protected from infiltration of lung parenchyma (radiographic stages II-IV) [21]. We have also demonstrated that rs1049550 is significant cross - ethnically at the gene level after adjustment for the single SNP association tests performed. The mechanism by which rs1049550 affects the susceptibility to sarcoidosis is that the SNP may affect the function of the ANXA11 protein rather than affect its expression. In humans, the ANXA11 protein consists of an N - terminalproline - tyrosine - glycine (PYG) - rich region, followed by four annexin core domains. The rs1049550 leads to an amino - acid exchange (basic arginine to a polar cysteine) at the evolutionarily conserved position 230 (R230C) in the first annexin domain, which is responsible for Ca^{2+} - dependent trafficking of the protein in the cell [23].

In the GWAS conducted by Hofmann et al [4], the strongest association signal was observed for rs2789679. This marker is located in the 3'-UTR of the ANXA11 gene. The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls being TT homozygous. Fine-mapping in the Black Women's Health Study (BWHS) revealed rs2819941 to be the top SNP, associated with a non-significant (P = 0.06)

reduction in the risk of sarcoidosis (17% reduction per copy of the C-allele). In this case-control association study, the frequency of the A allele in rs2789679 and the rs2819941 T allele frequency in the patients with sarcoidosis was significantly higher than that in the healthy controls. Several lines of evidence suggest that the observed association is unlikely to be an artifact. First, both the single-SNP and the haplotype-based association analyses support the current finding. Second, our samples were from the same geographical region, excluding the. Finally, consistent results were obtained from two genetically independent populations (Chinese Han and Europeans). Collectively, our results confirmed the strong association between ANXA11 polymorphisms and sarcoidosis, suggesting that ANXA11 represents a strong genetic risk factor for sarcoidosis. Since causal SNP in ANXA11 was not known, all the SNPs significantly associated with sarcoidosis were actually indirect association through the real causal one. Though rs2789679 and rs2819941 were not in LD with rs1049550, it may not tag the causal SNP as well as rs1049550. Hence, their individual effects were eliminated when controlling for the effect of rs1049550.

We further investigated the interaction among polymorphisms and observed strong LD. The haplotype analysis revealed that significantly more C - C (block 2, rs1049550 - rs2573351) and significantly fewer T - C haplotypes (block 2) were found in the patients with sarcoidosis. These results indicated that the patients with C - C (block 2) haplotypes of the ANXA11 gene were more prone to sarcoidosis. Significantly higher frequencies of T - C haplotypes were detected in the healthy controls than in the patients with sarcoidosis, suggesting that they may show

protective effects against sarcoidosis. Our sample size can detect SNP and haplotype associations with 90% and 85% power, respectively, at a false positive rate of 5%, disease prevalence of 1‰, disease allele / haplotype frequency of 0.05 / 0.03, and a presumed odds ratio (OR) of 1.5. To some extent, this finding further supports a role of ANXA11 polymorphisms in sarcoidosis, while ethnic group difference may exist.

In conclusion, our study suggests a potential role of the ANXA11 SNPs and their related haplotypes in the genetic susceptibility to sarcoidosis. Further studies are needed to investigate how these SNPs affect the function of ANXA11. A broader examination of the genetic variation in ANXA11 in the Han Chinese may reveal other variants associated with disease risk.

Contributors

LG Zhang was involved in conception and design of the study, acquisition of patient data, genotyping and drafted the paper. ZJ Xiao and GC Shi were involved in design of the study, genotyping and interpretation of results. J Huang and CX Zhang were involved in the design of the study, acquisition of patient data and interpretation of result. All authors revised the draft paper.

Funding

The project was supported by the Research and Development Foundation of Science and Technology of Henan Province (2012k15).

Competing interests

None.

Patient consent

Obtained.

Ethics approval

Ethics approval was provided by the First Hospital Affiliated to the Xinxiang Medical College.

Provenance and peer review

Not commissioned; externally peer reviewed.

Data sharing statement

No additional data are available.

Figure Legend

Fig. 1 The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the pair-wise calculation of r^2 (left) or D' (right). Black squares indicate $r^2 = 1$ (i.e. perfect LD between a pair of SNPs). Empty squares indicate D' = 1 (i.e. complete LD between a pair of SNPs).

Supplemental Fig. 1 The structure of the human ANXA11 gene and 15 SNPs located on the gene.

Reference

1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007; 357:2153-65.

2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-8

3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis. *Semin Respir Crit Care Med* 2010;31:380-9.

4. Hofmann S, Franke A, Fischer A, *et al*. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet* 2008;40:1103-6.

5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev* 2002;82:331-71.

6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using recombinant human annexin XI-specific autoantibodies cloned by phage display. *J Biol Chem* 2003;278:33120-6.

7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-22.

8. Petzmann S, Maercker C, Markert E, *et al*. Enhanced proliferation and decreased apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis Vasc Diffuse Lung Dis* 2006;23:190-200

9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, *et al*. Comparative analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol* 2006;26:243-50.

10. Agostini C, Trentin L, Facco M, *et al*. Role of IL-15, IL-2, and their receptors in the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol* 1996;157:910-8

11. Jullien D, Sieling PA, Uyemura K, *et al*. IL-15, an immunomodulator of T cell responses in intracellular infection. *J Immunol* 1997;158:800-6

12. Muller-Quernheim J, Schurmann M, Hofmann S, *et al*. Genetics of sarcoidosis. *Clin Chest Med* 2008;29:391-414.

13. Morais A, Lima B, Peixoto M, *et al*. Annexin A11 gene polymorphism (R230C variant) and sarcoidosis in a Portuguese population. *Tissue antigens* 2013;82:186-91.

14. Li Y, Pabst S, Kubisch C, *et al*. First independent replication study confirms the strong genetic association of ANXA11 with sarcoidosis. *Thorax* 2010;65:939-40.

15. Gabriel SB, Schaffner SF, Nguyen H, *et al*. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-9.

16. Barrett JC, Fry B, Maller J, *et al*. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-5.

17. Curtis D, Knight J, Sham PC. Program report: GENECOUNTING support programs. *Ann Hum Genet* 2006;70:277-9.

18. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic associations. *Hum Hered* 2000;50:133-9.

19. Zhao JH, Lissarrague S, Essioux L, et al. GENECOUNTING: haplotype analysis with missing genotypes. *Bioinformatics* 2002;18:1694-5

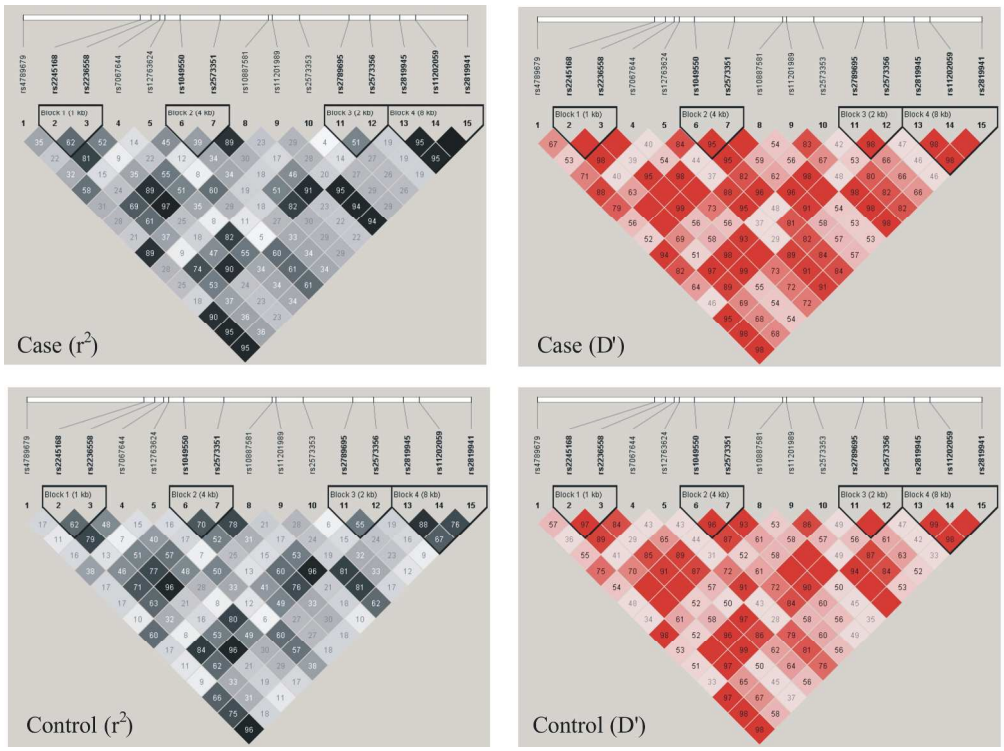
20. Zhao JH, Sham PC. Generic number systems and haplotype analysis. *Comput Methods Programs Biomed* 2003;70:1-9

21. Mrazek F, Stahelova A, Kriegova E, *et al*. Functional variant ANXA11 R230C: true marker of protection and candidate disease modifier in sarcoidosis. *Genes*

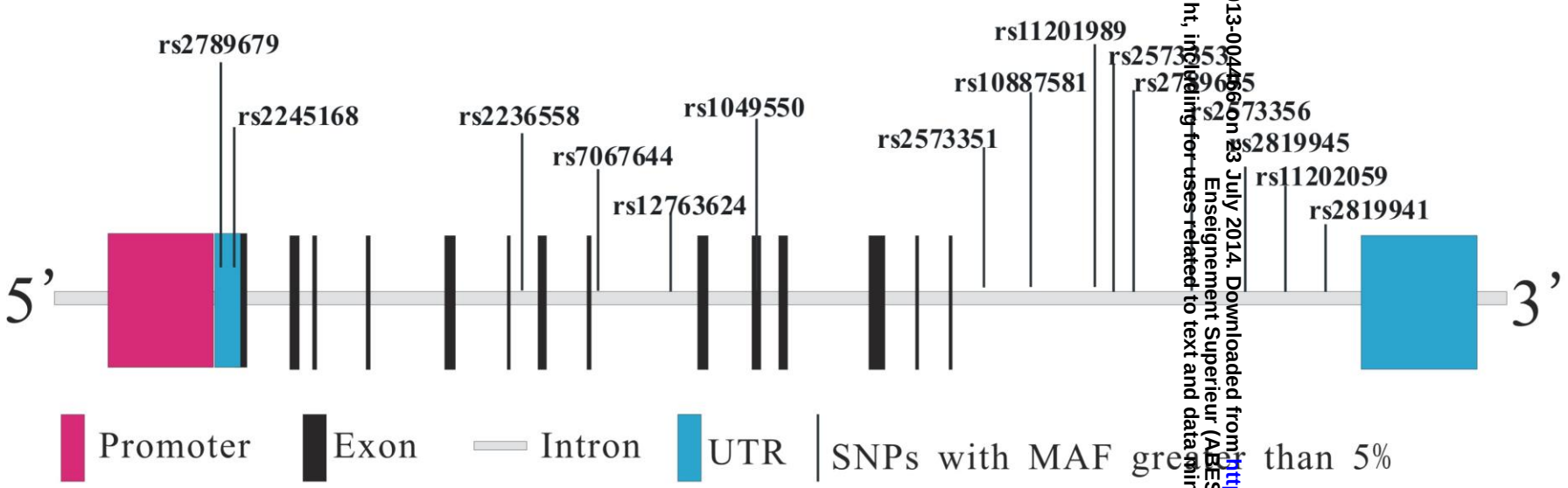
Immun 2011;12:490-4.

22. Levin AM, Iannuzzi MC, Montgomery CG, *et al*. Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes Immun* 2013;14:13-8.

24. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11 with the nuclear envelope. *J Biol Chem* 2003;278:20210-6.



273x203mm (300 x 300 DPI)



Supplemental table 1 Summary of all published sarcoidosis-association studies for ANXA11

Study	Population		Type of Study	Sample Size (n)		Number of SNPs typed	Positive SNPs
	Ethnic group	Country		Sarcoidosis	Control		
Hofmann S, 2008	Caucasian	Germany	Case-control	499	490	GWAS	rs2789679, rs7091565, rs1049550
Mrazek F, 2011	Caucasian	Czech	Case-control	245	254	1	rs1049550
Cozier YC, 2012	Caucasian	America	Case-control	486	943	10p12-10q22	rs2789679, rs2819941
Li Y, 2013	Caucasian	Germany	Case-control	325	364	1	rs1049550
Levin AM, 2013	Caucasian	America	Case-control	1689	1252	25	rs1049550
Morais A, 2013	Caucasian	Portugal	Case-control	208	197	1	rs1049550

Supplemental Table 2 Genotypic and allelic frequencies of ANXA11 polymorphisms in the controls and patients with sarcoidosis

Variable	**Controls** (n=418)		**Sarcoidosis** (n=412)		*P*-value[a]	OR, 95% CI
	No.	%	No.	%		
rs2245168					0.920	
CC	156	37.3	150	36.4	0.791	1.04, 0.74-1.38
CT	200	47.9	197	47.8	0.955	0.99, 0.76-1.30
TT	62	14.8	65	15.8	0.781	0.95, 0.65-1.39
Per T allele	324	38.8	327	39.7	0.698	1.04, 0.85-1.27
rs2236558					0.172	
GG	118	28.2	104	25.2	0.72	1.06, 0.77-1.45
TG	190	45.5	214	51.9	0.060	0.77, 0.59-1.01
TT	110	26.3	94	22.8	0.073	1.33, 0.97-1.82
Per T allele	410	49.0	402	48.8	0.917	0.99, 0.82-1.20
rs7067644					0.173	
AA	152	36.4	176	42.7	0.059	0.76, 0.58-1.01
GA	206	49.3	182	44.2	0.137	1.23, 0.94-1.68
GG	60	14.4	54	13.1	0.597	1.11, 0.75-1.65
Per G allele	326	38.0	290	35.2	0.109	0.85, 0.70-1.04
rs12763624					0.118	
TT	126	75.6	150	36.4	0.008	0.68, 0.51-0.91
CT	216	51.7	186	45.2	0.059	1.30, 0.99-1.71
CC	76	18.2	76	18.5	0.411	1.16, 0.81-1.67
Per C allele	368	44.0	338	41.0	0.216	0.88, 0.73-1.08
rs2573351					0.437	

TT	120	28.7	104	25.2	0.263	1.19, 0.88-1.62
CT	202	48.3	216	52.4	0.238	0.85, 0.65-1.12
CC	96	23.0	92	22.3	0.826	1.04, 0.75-1.44
Per C allele	394	47.1	400	48.5	0.564	1.06, 0.87-1.28
rs10887581					0.290	
TT	114	27.3	96	23.3	0.190	1.23, 0.90-1.69
TC	190	45.5	208	50.5	0.148	0.82, 0.62-1.07
CC	114	27.3	108	26.2	0.729	1.06, 0.78-1.44
Per C allele	418	50.0	424	51.5	0.583	1.06,0.87-1.29
rs11201989					0.144	
GG	130	31.1	128	31.1	0.995	1.00, 0.75-1.34
GC	218	52.2	194	47.1	0.141	1.23, 0.93-1.62
CC	70	16.8	90	21.9	0.060	0.71, 0.50-1.01
Per C allele	358	42.8	374	45.4	0.29	1.11, 0.91-1.35
rs2573353					0.918	
CC	178	42.6	176	42.7	0.957	0.99, 0.75-1.31
CA	196	46.9	192	46.6	0.921	1.01, 0.77-1.33
AA	44	10.5	44	10.7	0.941	0.98, 0.63-1.53
Per A allele	284	34.0	280	34.0	0.997	1.00, 0.82-1.23
rs2789695					0.918	
TT	178	42.6	180	34.0	0.740	0.95, 0.73-1.26
CT	188	45.0	184	44.7	0.920	1.01, 0.77-1.33
CC	52	12.4	48	11.7	0.73	1.08, 0.71-1.64
Per C allele	292	34.9	280	34.0	0.69	0.96, 0.78-1.17
rs2573356					0.892	
CC	118	28.2	112	27.2	0.742	1.05, 0.78-1.43

CT	190	45.5	194	47.1	0.639	0.94, 0.71-1.23
TT	110	26.3	106	25.7	0.843	1.03, 0.76-1.41
Per T allele	410	49.0	406	49.3	0.926	0.99, 0.82-1.20
rs2819945					0.101	
GG	119	28.5	130	31.6	0.294	0.85, 0.63-1.15
GA	221	52.9	188	45.6	0.030	1.35, 1.03-1.78
AA	78	18.7	94	22.8	0.138	0.78, 0.55-1.09
Per A allele	377	45.1	376	45.6	0.827	1.02, 0.84-1.24
rs11202059					0.064	
GG	106	25.4	128	31.1	0.067	0.75, 0.56-1.02
GA	226	54.1	190	46.1	0.021	1.38, 1.05-1.81
AA	86	20.6	94	22.8	0.429	0.88, 0.63-1.22
Per A allele	398	47.6	378	45.9	0.479	0.93, 0.77-1.13

[a] *P*-value was calculated by 2×3 and 2×2 chi-squared tests based on codominant, dominant for the rare allele, heterosis and recessive for the rare allele models of inheritance.

Alpha value is adjusted by Bonferroni correction and statistically significant results (P<0.003)

Supplementary table 3 Multi-SNP analysis of rs2789679, rs1049550 and rs2819941

Models	testing SNP	*P*-value	OR	95%CI
model1: rs2789679+ rs1049550	rs2789679: A	0.11	1.19	0.96-1.48
model2: rs2819941+rs1049550	rs2819941: T	0.169	1.17	0.94-1.45
model3: rs2789670+rs1049550+rs2819941	rs1049550: T	0.0002	0.65	0.52-0.82

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	2-3
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	5-6
	2b	Specific objectives or hypotheses	6
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	7
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	7
Participants	4a	Eligibility criteria for participants	7
	4b	Settings and locations where the data were collected	7
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	8
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	7
	6b	Any changes to trial outcomes after the trial commenced, with reasons	7
Sample size	7a	How sample size was determined	7
	7b	When applicable, explanation of any interim analyses and stopping guidelines	7
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	8
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	8
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	7
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	7
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those	7

		assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	8
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	8-9
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	8
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	10-11
	13b	For each group, losses and exclusions after randomisation, together with reasons	10-11
Recruitment	14a	Dates defining the periods of recruitment and follow-up	10
	14b	Why the trial ended or was stopped	10
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	10
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	11
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	11
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	11
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	11
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	11-15
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	12-14
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	12
Other information			
Registration	23	Registration number and name of trial registry	13
Protocol	24	Where the full trial protocol can be accessed, if available	14
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	15

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.