



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

Journal:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

Peer Review Only

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 - Tuberculosis, 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 tuberculosis 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 ($\chi^2 = 12.849$, $P = 0.0003$, OR = 0.703, 95%CI = 0.579 - 0.852), T allele in rs1049550 ($\chi^2 = 24.420$, $P = 0.0007$, OR = 0.617, 95%CI= 0.503 - 0.758) and the rs2819941 C allele frequency ($\chi^2 = 11.803$, $P = 0.001$, 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

1
2
3 haplotypes (block 4) ($P = 0.027$) were found in the patients with sarcoidosis, but did
4
5 not pass the threshold value ($P = 0.025$).
6
7

8 **Conclusions:** These findings point to a role for ANXA11 gene polymorphisms in
9
10 sarcoidosis in the Han Chinese, and may be informative for future genetic or
11
12 biological studies on sarcoidosis.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 Löfgren'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 sarcoidosis^{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 division⁵⁻⁷. 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

1
2
3
4 been fully elucidated. However, it is generally assumed that the induction of apoptosis
5
6 and / or by the withdrawal of inflammatory cytokines participates in the
7
8 disappearance of granulomas^{10 11}. In patients with sarcoidosis, their peripheral blood
9
10 monocytes were characterized by apoptosis⁹.
11
12

13
14 In previous studies, case - control / family ‘hypothesis - driven’ studies and low
15
16 density linkage scans were used to identify genetic factors conferring susceptibility to
17
18 sarcoidosis¹². Notably, the first genome - wide association study (GWAS) in
19
20 sarcoidosis conducted in a German population has recently revealed an association:
21
22 the leading rs2789679 SNP located in the 3’ - untranslated region (3’UTR) and a
23
24 common nonsynonymous SNP (rs1049550, C > T, Arg230Cys) were associated with
25
26 increased risk of sarcoidosis⁴. This association has recently been supported by
27
28 another report from the same population¹³. Thus, more studies should be performed
29
30 to demonstrate the following item: whether these SNPs modulate the risk of
31
32 sarcoidosis by themselves or whether they correlate with other causative SNPs and are
33
34 repeated in other populations. The published sarcoidosis - association studies for
35
36 ANXA11 summarized in Table 1. We hypothesize that common variants in the
37
38 ANXA11 gene may significantly contribute to the predisposition to develop
39
40 sarcoidosis.
41
42
43
44
45
46
47
48

49 In this study, we investigated fifteen loci in a Chinese population from He’nan
50
51 province (China) to verify the putative association between ANXA11 polymorphisms
52
53 and sarcoidosis.
54
55

56 **SUBJECTS AND METHODS**

57
58
59
60

Subjects

Four - hundred and twelve patients with sarcoidosis (mean \pm 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 \pm 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,

1
2
3
4 rs2573356, rs2819945, rs11202059 and rs2819941 in intron 14. Marker selection was
5
6 done according to previous studies ^{4 12-14}, and preliminary analysis was performed
7
8 using the HapMap data (Fig 1). We examined tagSNPs in the Haploview software
9
10 v4.2, using the Chinese Han in Beijing (CHB) population and a minor allele
11
12 frequency cut-off (MAF) \geq 5% (the HapMap Data Release 27). The linkage
13
14 disequilibrium (LD) pattern of this gene was determined in the Chinese population
15
16 using the preliminary data from the HapMap. These SNPs were further analyzed in an
17
18 association study.
19
20
21
22

23 24 **Genotyping**

25
26 Peripheral blood was collected from a vein into a sterile tube coated with
27
28 ethylenediamine tetraacetic acid (EDTA). Plasma samples were stored at -20°C.
29
30 Genomic DNA was extracted from the frozen peripheral blood samples using a
31
32 QIAmp Blood Mini Kit (Qiagen Inc., Valencia, CA, USA) according to the
33
34 manufacturer's protocols. The selected SNPs were genotyped in cases and controls by
35
36 using the MALDI - TOF in the MassARRAY system (Sequenom Inc., San Diego, CA,
37
38 USA). Probes and primers were designed using the Assay Design Software
39
40 (Sequenom, San Diego, CA, USA). The completed genotyping reactions were spotted
41
42 onto a 384 well spectroCHIP (Sequenom) using the MassARRAY Nanodispenser
43
44 (Sequenom) and determined by the matrix - assisted laser desorption ionization time -
45
46 of - flight mass spectrometer. Genotype calling was performed in realtime with the
47
48 MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer
49
50 software version 3.4 (Sequenom).
51
52
53
54
55
56
57
58
59
60

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 4.0¹⁵.

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,

1
2
3
4 and haplotypes in the patients with sarcoidosis and the healthy controls are showed in
5
6 Tables 2 and Table 3.
7

8
9 Comparison of genotype and allele frequency distribution revealed significant
10
11 differences between the patients with sarcoidosis and healthy controls for 3 SNPs:
12
13 rs2789679, rs1049550 and rs2819941. The frequency of the T allele in rs2789679 (χ^2
14
15 = 12.849, $P = 0.0003$, OR = 0.703, 95% CI = 0.579 - 0.852), T allele in rs1049550 (χ^2
16
17 = 24.420, $P = 0.0007$, OR = 0.617, 95% CI = 0.503 - 0.758) and the rs2819941 C
18
19 allele frequency ($\chi^2 = 11.803$, $P = 0.001$, OR = 0.713, 95% CI = 0.588 - 0.865) in the
20
21 patients with sarcoidosis was significantly lower than that in the controls, and these
22
23 differences remained statistically significant after Bonferroni corrections.
24
25
26
27

28
29 Due to the positive association between 2 SNPs (rs1049550 and rs2819941), and
30
31 sarcoidosis, we performed an association analysis to determine whether the haplotype
32
33 was associated with the risk of sarcoidosis. Compared with the healthy controls, the T
34
35 - C haplotype occurred significantly less frequently ($P = 0.001$) and the C - C
36
37 haplotypes occurred more frequently ($P = 0.0001$) in block 2 in the patients with
38
39 sarcoidosis. The significantly more G - G - C haplotypes (block 4) ($P = 0.027$) were
40
41 found in the patients with sarcoidosis, but did not pass the threshold value ($P =$
42
43
44
45
46
47 0.025).
48

49 DISCUSSION

50
51 A key step in linkage and association studies is to identify common risk variants
52
53 in different populations. To determine if common risk variants exist in distinct
54
55 populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene,
56
57
58
59
60

1
2
3
4 were genotyped in samples from the patients with sarcoidosis and healthy controls in
5
6 a Chinese Han population. Recently, new sarcoidosis loci have been identified by
7
8 genome - wide association studies ^{4 13}. With the fast development of genome-wide
9
10 association studies, increasing number of susceptibility loci for sarcoidosis have been
11
12 reported in various populations ^{4 13-14}. However, these observations should be
13
14 confirmed in other genetically independent populations. In this study, we conducted
15
16 the first large genetic association study of the ANXA11 gene in a Chinese Han
17
18 population. The evidence of markers associated with sarcoidosis was presented, and
19
20 these markers were mapped to different locations in the ANXA11 gene (81897864 -
21
22 81951001). The association signals in the region were identified, and some
23
24 significantly associated haplotypes also appeared this region.
25
26
27
28
29
30

31 Hoffman et al. reported an association between the T allele of the non -
32
33 synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a
34
35 German population ⁴. Similar results were also obtained in other two European
36
37 populations ^{13 14}. As a part of the sarcoidosis GWAS done in Americans ¹⁶, we further
38
39 confirmed this association in a Chinese Han population. In this study, the frequency of
40
41 ANXA11 rs1049550 T allele was significantly lower in the patients with sarcoidosis
42
43 than that in the healthy controls. The “protective” effect of ANXA11 T allele
44
45 increased with the number of its copies in the genotype, which is consistent with the
46
47 result obtained in the patients with sarcoidosis in a German population ⁴, Furthermore,
48
49 the T allele carriers among the patients were protected from infiltration of lung
50
51 parenchyma (radiographic stages II-IV) ¹⁴. We have also demonstrated that rs1049550
52
53
54
55
56
57
58
59
60

1
2
3
4 is significant cross-ethnically at the gene level after adjustment for the single SNP
5
6 association tests performed. The mechanism by which rs1049550 SNP affects the
7
8 susceptibility to sarcoidosis may be that the SNP may affect the function of the
9
10 ANXA11 protein rather than affect its expression. In humans, the ANXA11 protein
11
12 consists of an N - terminal proline - tyrosine - glycine (PYG) - rich region, followed
13
14 by four annexin core domains. The SNP rs1049550 leads to an amino - acid exchange
15
16 (basic arginine to a polar cysteine) at the evolutionarily conserved position 230
17
18 (R230C) in the first annexin domain, which is responsible for Ca²⁺ - dependent
19
20 trafficking of the protein in the cell ¹⁷.

21
22 In the GWAS conducted by Hofmann et al ⁴, the strongest association signal was
23
24 observed for SNP rs2789679. This marker is located in the 3' - UTR of the ANXA11
25
26 gene. The T allele had a frequency of 35% among the patients with sarcoidosis and 44%
27
28 in controls, with 13% of the patients with sarcoidosis and 19% of controls being TT
29
30 homozygous. Fine - mapping in the Black Women's Health Study (BWHS) revealed
31
32 rs2819941 to be the top SNP, associated with a nonsignificant ($P = 0.06$) reduction in
33
34 sarcoidosis risk (17% reduction per copy of the C - allele). In this case - control
35
36 association study, the frequency of the T allele in rs2789679 and the rs2819941 C
37
38 allele frequency in the patients with sarcoidosis was significantly lower than that in
39
40 the controls. Several lines of evidence suggest that the observed associations are
41
42 unlikely to be an artifact. First, both the single - SNP and the haplotype - based
43
44 association analyses support the current finding. Second, population stratification is
45
46 an impossible reason, because all of our samples were from the same geographical
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 region. Finally, consistent results were obtained from two genetically independent
5
6 populations (Han Chinese and Europeans). Collectively, our results confirmed the
7
8 strong association between variations in the ANXA11 gene and sarcoidosis, and
9
10 suggest that ANXA11 represents a strong genetic risk factor for sarcoidosis.
11
12

13
14 We further investigated the interaction among polymorphisms and observed
15
16 strong LD. The haplotype analysis revealed that significantly more C - C (block 2),
17
18 and G - G - C (block 4) haplotypes and significantly fewer T - C haplotypes (block 2)
19
20 were found in the patients with sarcoidosis. There were significant point - wise
21
22 associations of these variants with sarcoidosis (rs2789679, rs1049550 and rs2819941).
23
24 These results indicated that the patients with C - C (block 2), and G - G - C (block 4)
25
26 haplotypes of the ANXA11 gene were more prone to sarcoidosis. Significantly higher
27
28 frequencies of T - C haplotypes were detected in the healthy controls than in the
29
30 patients with sarcoidosis, suggesting that they may show a protective effect against
31
32 sarcoidosis. To some extent, this finding further supports a role of ANXA11
33
34 polymorphisms in sarcoidosis, while ethnic group difference may exist.
35
36
37
38
39
40

41
42 In conclusion, our study suggests a potential role of the ANXA11 SNPs and their
43
44 related haplotypes in the genetic susceptibility to sarcoidosis. Further studies are
45
46 needed to investigate how these SNPs affect the function of ANXA11.
47
48

49 **Contributors**

50
51 LG Zhang was involved in conception and design of the study, acquisition of
52
53 patient data, genotyping and drafted the paper. ZJ Xiao and GC Shi were involved in
54
55 design of the study, genotyping and interpretation of results. J Huang and CX Zhang
56
57
58
59
60

1
2
3 were involved in design of the study, acquisition of patient data and interpretation of
4
5
6 result. All authors revised the draft paper.
7

8 9 **Funding**

10
11 The project was supported by the Research and Development Foundation of
12
13 Science and Technology of Henan Province (2012k15).
14

15 16 **Competing interests**

17
18 None.
19

20 21 **Patient consent**

22
23 Obtained.
24

25 26 **Ethics approval**

27
28 Ethics approval was provided by the First Hospital Affiliated to the Xinxiang
29
30 Medical College.
31

32 33 **Provenance and peer review**

34
35 Not commissioned; externally peer reviewed.
36
37

38 39 **Data sharing statement**

40
41 No additional data are available.
42
43
44
45

46 47 **Figure Legend**

48
49 **Fig. 1** The structure of the human ANXA11 gene and ten SNPs located on the gene.
50

51
52 **Fig. 2** The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the
53
54 pair - wise calculation of r^2 (left) or D' (right). Black squares indicate $r^2 = 1$ (i.e.
55
56 perfect LD between a pair of SNPs). Empty squares indicate $D' = 1$ (i.e. complete LD
57
58
59
60

1
2
3
4 between a pair of SNPs).

5
6 **Reference**

- 7
8
9 1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med*
10
11 2007;357:2153-2165.
12
13
14 2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-1118.
15
16
17 3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis.
18
19 *Semin Respir Crit Care Med* 2010;31:380-389.
20
21
22 4. Hofmann S, Franke A, Fischer A, et al. Genome-wide association study identifies
23
24 ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet*
25
26 2008;40:1103-1106.
27
28
29 5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev*
30
31 2002;82:331-371.
32
33
34 6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using
35
36 recombinant human annexin XI-specific autoantibodies cloned by phage
37
38 display. *J Biol Chem* 2003;278:33120-33126.
39
40
41 7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and
42
43 completion of the terminal phase of cytokinesis. *J Cell Biol*
44
45 2004;165:813-822.
46
47
48
49 8. Petzmann S, Maercker C, Markert E, et al. Enhanced proliferation and decreased
50
51 apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis, vasculitis,*
52
53 *and diffuse lung diseases : official journal of WASOG / World Association of*
54
55 *Sarcoidosis and Other Granulomatous Disorders* 2006;23:190-200.
56
57
58
59
60

- 1
2
3
4 9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, Dubaniewicz A,
5
6 Singh M, Mysliwski A. Comparative analysis of mycobacterial heat shock
7
8 proteins-induced apoptosis of peripheral blood mononuclear cells in
9
10 sarcoidosis and tuberculosis. *J Clin Immunol* 2006;26:243-250.
11
12
13
14 10. Agostini C, Trentin L, Facco M, et al. Role of IL-15, IL-2, and their receptors in
15
16 the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol*
17
18 1996;157:910-918.
19
20
21 11. Jullien D, Sieling PA, Uyemura K, Mar ND, Rea TH, Modlin RL. IL-15, an
22
23 immunomodulator of T cell responses in intracellular infection. *J Immunol*
24
25 1997;158:800-806.
26
27
28
29 12. Muller-Quernheim J, Schurmann M, Hofmann S, et al. Genetics of sarcoidosis.
30
31 *Clin Chest Med* 2008;29:391-414.
32
33
34 13. Li Y, Pabst S, Kubisch C, Grohe C, Wollnik B. First independent replication study
35
36 confirms the strong genetic association of ANXA11 with sarcoidosis. *Thorax*
37
38 2010;65:939-940.
39
40
41 14. Mrazek F, Stahelova A, Kriegova E, et al. Functional variant ANXA11 R230C:
42
43 true marker of protection and candidate disease modifier in sarcoidosis. *Genes*
44
45 *Immun* 2011;12:490-494.
46
47
48
49 15. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD
50
51 and haplotype maps. *Bioinformatics* 2005;21:263-265.
52
53
54 16. Levin AM, Iannuzzi MC, Montgomery CG, et al. Association of ANXA11 genetic
55
56 variation with sarcoidosis in African Americans and European Americans.
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

For peer review only

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

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)			
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)			
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)			
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)			
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)			
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)			
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)			

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

^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 ($p > 0.05$).

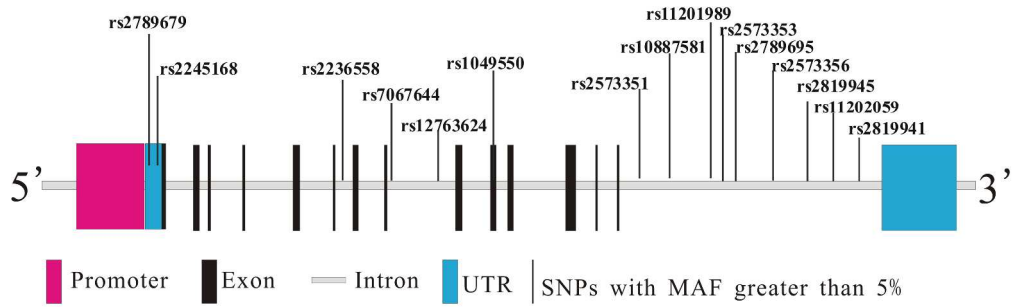
For peer review only

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	<i>P</i>	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).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



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



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).



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

Journal:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3
4 Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a Han
5
6 Chinese population - a case-control study
7

8
9 LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang

10
11 Internal Medicine - Tuberculosist, First Hospital Affiliated to the Xinxiang Medical
12
13 College, Xinxiang, Henan, PR China
14

15
16 *Corresponding author: Zhijian Xiao, Internal Medicine - Tuberculosisthe, First
17
18 Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China
19

20
21 No.88 Jiankang Road, Weihui City, Henan, 463000, PR China
22

23
24 E-mail: zhangligongys@163.com
25

26
27 Tel: +86 0373 - 4402229 (office)
28

29
30 Fax: +86 0373 -4402251
31

32
33 **Key words:** Sarcoidosis; Annexin A11; Single-nucleotide polymorphism
34

35
36 Word count: 2909
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 tuberculosis 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

1
2
3
4 blocks ($D' > 0.9$). In block 2 (rs1049550-rs2573351), the T - C haplotype occurred
5
6 significantly less frequently ($P = 0.001$), whereas the C - C haplotypes occurred more
7
8 frequently ($P = 0.0001$).
9

10
11 **Conclusions:** These findings point to a role for ANXA11 gene polymorphisms in
12
13 sarcoidosis in a Han Chinese population, and may be informative for future genetic or
14
15 biological studies on sarcoidosis.
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 Löfgren'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 sarcoidosis^{1 3}.

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

1
2
3
4 chronic diseases^{4,5}. One member of the Annexin gene family, ANXA11, located on
5
6 chromosome 10q22.3, is involved in calcium signaling, apoptosis, vesicle trafficking,
7
8 cell growth and terminal phase of cell division⁵⁻⁷. In this context, the development
9
10 and maintenance of the granulomatous inflammation in sarcoidosis have been
11
12 repeatedly associated with the impaired apoptosis of activated inflammatory cells^{8,9}.
13
14 In most patients with sarcoidosis, the early stages of this disease are characterized by
15
16 mononuclear cell alveolitis dominated by activated CD4⁺ T cells and macrophages.
17
18 Between these cells, the uncoordinated interplay results in the formation of typical
19
20 non - caseating granulomas. The mechanism by which the granulomas resolve has not
21
22 been fully elucidated. However, it is generally assumed that the induction of apoptosis
23
24 and / or by the withdrawal of inflammatory cytokines participates in the
25
26 disappearance of granulomas^{10,11}. In the patients with sarcoidosis, their peripheral
27
28 blood monocytes were characterized by apoptosis⁹.

29
30
31 In previous studies, case - control / family 'hypothesis - driven' studies and low
32
33 density linkage scans were used to identify genetic factors conferring the genetic
34
35 susceptibility to sarcoidosis¹². Notably, the first genome - wide association study
36
37 (GWAS) in sarcoidosis conducted in a German population has recently revealed an
38
39 association: the leading rs2789679 SNP (single - nucleotide polymorphism) located in
40
41 the 3' - untranslated region (3' - UTR) and a common nonsynonymous SNP
42
43 (rs1049550, C > T, Arg230Cys) were associated with increased risk of sarcoidosis^{4,13}.
44
45 This association has recently been supported by another report from the same
46
47 population¹⁴. Thus, more studies should be performed to demonstrate the following
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 item: whether these SNPs modulate the risk of sarcoidosis by themselves or whether
5
6 they correlate with other causative SNPs and are repeated in other populations. The
7
8 published sarcoidosis - association studies for ANXA11 are summarized in Table 1.
9
10 We hypothesize that common variants in the ANXA11 gene may significantly
11
12 contribute to the predisposition to develop sarcoidosis.
13
14

15
16 In this study, we investigated fifteen loci in a Chinese population from He'nan
17
18 province (China) to verify the putative association between ANXA11 polymorphisms
19
20 and sarcoidosis.
21
22

23 **SUBJECTS AND METHODS**

24 **Subjects**

25
26 Four hundred and twelve patients with sarcoidosis (mean \pm SD age: 53.6 ± 4.6
27
28 years; 155 men and 257 women) were recruited from our hospital between May 2005
29
30 and March 2013. Sarcoidosis was diagnosed by the evidence of non - caseating
31
32 epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities.
33
34 Chronic sarcoidosis was defined as a disease over at least 2 years or at least two
35
36 episodes in a lifetime. A cute sarcoidosis was defined as one episode of acute
37
38 sarcoidosis which had totally resolved at the date of the examination. The control
39
40 group consisted of 418 unrelated healthy subjects (mean \pm SD age: 54.2 ± 5.3 years;
41
42 160 men and 258 women) who underwent health examinations in the Medical
43
44 Examination Center of the First Affiliated Hospital of the Xinxiang Medical College
45
46 (Xinxiang, China) between Oct 2009 and Sept 2013. None of the individuals in the
47
48 control group had a history of lung diseases or showed any symptoms of the lung or
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 other diseases by chest radiography or laboratory blood tests. Participants were
5
6 excluded if they: were taking other prescribed medications that could affect the
7
8 central nervous system; had a history of seizures, hematological diseases, or severe
9
10 liver or kidney impairment; or were pregnant. No familial relationship was known
11
12 between the study participants. The study was performed according to the Guidelines
13
14 of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang, China).
15
16 Written informed consent was obtained from each participant in this study.
17
18
19

20 21 **SNP selection**

22
23 Tagging SNPs were selected from the catalogs of the International HapMap
24
25 Project. The rs2789679 and rs2245168 are located in 5'-UTR, rs2236558 in intron 6,
26
27 rs7067644 and rs12763624 in intron 8, rs1049550 in exon 10, rs2573351 in intron 13,
28
29 rs10887581, rs11201989, rs2573353, rs2789695, rs2573356, rs2819945, rs11202059
30
31 and rs2819941 in intron 14 (Figure 1). Marker selection was done according to
32
33 previous studies^{4 12 14 15}, and preliminary analysis was performed using the HapMap
34
35 data and the following criteria. First, we examined tagSNPs in the Haploview (v4.2),
36
37 using the CHB population and a minor allele frequency cut - off (MAF) $\geq 5\%$
38
39 (HapMap Data Release 27). We found that there were a total of 29 potential tagSNPs
40
41 in all. As a first screen of the most common SNPs in the sarcoidosis sample from an
42
43 Eastern Chinese Han population, a MAF $\geq 20\%$ with pair - wise tagging and $r^2 \geq 0.8$
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
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

1
2
3
4 'confidence intervals' method^{17 18}. These SNPs were further analyzed in an
5
6 association study.
7

8 9 **Genotyping**

10
11 Peripheral blood was collected from a vein into a sterile tube coated with
12
13 ethylenediamine tetraacetic acid (EDTA). Plasma samples were stored at -20°C.
14
15 Genomic DNA was extracted from the frozen peripheral blood samples using a
16
17 QIAmp Blood Mini Kit (Qiagen Inc., Valencia, CA, USA) according to the
18
19 manufacturer's protocols. The selected SNPs were genotyped in cases and controls by
20
21 using the MALDI - TOF in the MassARRAY system (Sequenom Inc., San Diego, CA,
22
23 USA). Probes and primers were designed using the Assay Design Software
24
25 (Sequenom, San Diego, CA, USA). The completed genotyping reactions were spotted
26
27 onto a 384 well spectroCHIP (Sequenom) using the MassARRAY Nanodispenser
28
29 (Sequenom) and determined by the matrix - assisted laser desorption ionization time -
30
31 of - flight mass spectrometer. Genotype calling was performed in realtime with the
32
33 MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer
34
35 software version 3.4 (Sequenom).
36
37
38
39
40
41
42
43

44 **Statistical analysis**

45
46 All data were analyzed using the SPSS 17.0 software (SPSS Inc., Chicago, IL,
47
48 USA). Hardy-Weinberg equilibrium were evaluated by Chi - square tests. Differences
49
50 between the cases and controls in the frequency of the alleles, genotypes and
51
52 haplotypes were evaluated by the Fisher's exact test or the Pearson Chi - square test.
53
54
55 Unconditional logistic regression was used to calculate the odds ratio (OR) and 95%
56
57
58
59
60

1
2
3
4 confidence interval (CI) in independent association between each locus and the
5
6 presence of sarcoidosis. Generalized linear regression was used to evaluate the
7
8 interaction effects between gene and gender or age. Gender and age of subjects were
9
10 treated as covariants in binary logistic regression. *P* - value was calculated based on
11
12 codominant, dominant for the rare allele, heterosis and recessive for the rare allele
13
14 models of inheritance. The Bonferroni correction was used to adjust the test level
15
16 when multiple comparisons were conducted, and the *P* value was divided by the total
17
18 number of loci. Haplotype blocks were defined according to the criteria developed by
19
20 Gabriel et al.¹⁸. Pair - wise LD statistics (D' and r^2) and haplotype frequency were
21
22 calculated, and haplotype blocks were constructed using the Haploview 4.0¹⁷. To
23
24 ensure that the LD blocks most closely reflect the population level LD patterns,
25
26 definition of the blocks were based on the control samples alone. The haplotype
27
28 frequencies were estimated using GENECOUNTING, which computes maximum -
29
30 likelihood estimates of haplotype frequencies from unknown phase data by utilizing
31
32 an expectation - maximization algorithm¹⁹⁻²². The significance of any haplotypic
33
34 association was evaluated using a likelihood ratio test, followed by permutation
35
36 testing that compared estimated haplotype frequencies in cases and controls^{19 21}.
37
38
39
40
41
42
43
44
45

46 RESULTS

47
48 The genotype distribution of the fifteen polymorphisms was consistent with the
49
50 Hardy - Weinberg equilibrium ($P > 0.05$). The analysis of strong LD in the patients
51
52 with sarcoidosis and healthy controls revealed that 2 SNPs (rs2245168, rs2236558), 2
53
54 SNPs (rs1049550, rs2573351), 2 SNPs (rs2789695, rs2573356) and 3 SNPs
55
56
57
58
59
60

1
2
3
4 (rs2819945, rs11202059, rs2819941) were located in haplotype block 1, block 2,
5
6 block 3 and block 4 ($D' > 0.9$, Fig. 2). The genotype distribution, allelic frequencies,
7
8 and haplotypes in the patients with sarcoidosis and healthy controls are showed in
9
10 tables 2, table 4 and table 5.

11
12
13
14 Comparison of genotype and allele frequency distribution revealed significant
15
16 differences between the patients with sarcoidosis and healthy controls for 3 SNPs:
17
18 rs2789679, rs1049550 and rs2819941. The frequency of the T allele in rs2789679 (P
19
20 = 0.0003, OR = 0.70, 95%CI = 0.58 - 0.85), T allele in rs1049550 (P = 0.0007, OR =
21
22 0.62, 95%CI = 0.50 - 0.76) and the rs2819941 C allele frequency (P = 0.001, OR =
23
24 0.71, 95%CI = 0.59 - 0.87) in the patients with sarcoidosis was significantly lower
25
26 than that in the controls, and these differences remained statistically significant after
27
28 Bonferroni corrections. There were no interactions among rs2789679, rs1049550 and
29
30 rs2819941 ($P > 0.05$).

31
32
33
34
35
36 We performed an association analysis to determine whether the haplotype was
37
38 associated with the risk of sarcoidosis (Global $P < 0.05$ in block 2 and block 3).
39
40 Compared with the healthy controls, the T - C haplotype occurred significantly less
41
42 frequently (P = 0.001) and the C - C haplotypes occurred more frequently (P = 0.0001)
43
44 in block 2 (rs1049550 - rs2573351) in the patients with sarcoidosis.

45
46
47
48
49 To assess particular disease phenotypes, the patients with sarcoidosis were
50
51 divided into the subgroups according to their chest radiographic (CXR) stage. CXR
52
53 stage I (isolated bilateral hilar lymphadenopathy): N = 196; CXR stages II - IV
54
55 (infiltration of parenchyma): 167; CXR stage 0 (absence on pulmonary disease): N =
56
57
58
59
60

1
2
3
4 28; the information on CXR stage was not available for 21 patients. Comparison of
5
6 genotype frequency distribution revealed significant differences between stage I and
7
8 stages II - IV for rs1049550. The significantly more CC genotype ($P = 0.012$) were
9
10 found in the patients with stage I sarcoidosis (table 3).
11
12

13 **DISCUSSION**

14
15 A key step in linkage and association studies is to identify common risk variants
16
17 in different populations. To determine if common risk variants exist in distinct
18
19 populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene,
20
21 were genotyped in samples from the patients with sarcoidosis and healthy controls in
22
23 a Chinese Han population. Recently, new sarcoidosis loci have been identified by
24
25 genome - wide association studies ^{4 14}. With the fast development of genome - wide
26
27 association studies, increasing number of susceptibility loci for sarcoidosis have been
28
29 reported in different populations ^{4 14 15}. However, these observations should be
30
31 confirmed in other genetically independent populations. In this study, we conducted
32
33 the first large genetic association study of the ANXA11 gene in a Chinese Han
34
35 population. The evidence of markers associated with sarcoidosis was presented, and
36
37 these markers were mapped to different locations in the ANXA11 gene (81897864 -
38
39 81951001). The association signals in the region were identified, and some
40
41 significantly associated haplotypes also appeared this region.
42
43
44
45
46
47
48
49
50

51 Hoffman et al. reported an association between the T allele of the non -
52
53 synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a
54
55 German population ⁴. Similar results were also obtained in other two European
56
57
58
59
60

1
2
3 populations^{14 15}. As a part of the sarcoidosis GWAS done in Americans²³, we further
4
5 confirmed this association in a Chinese Han population. In this study, the frequency of
6
7 ANXA11 rs1049550T allele was significantly lower in the patients with sarcoidosis
8
9 than that in the healthy controls. Comparison of stages II - IV, the significantly more
10
11 CC genotype was found in the patients with stages I sarcoidosis. The “protective”
12
13 effect of ANXA11 T allele increased with the number of its copies in the genotype,
14
15 which is consistent with the result obtained in the patients with sarcoidosis in a
16
17 German population⁴. Furthermore, the T allele carriers among the patients were
18
19 protected from infiltration of lung parenchyma (radiographic stages II-IV)¹⁵. We have
20
21 also demonstrated that rs1049550 is significant cross - ethnically at the gene level
22
23 after adjustment for the single SNP association tests performed. The mechanism by
24
25 which rs1049550 affects the susceptibility to sarcoidosis may be that the SNP may
26
27 affect the function of the ANXA11 protein rather than affect its expression. In humans,
28
29 the ANXA11 protein consists of an N - terminal proline - tyrosine - glycine (PYG) -
30
31 rich region, followed by four annexin core domains. The rs1049550 leads to an amino
32
33 - acid exchange (basic arginine to a polar cysteine) at the evolutionarily conserved
34
35 position 230 (R230C) in the first annexin domain, which is responsible for Ca²⁺ -
36
37 dependent trafficking of the protein in the cell²⁴.
38
39
40
41
42
43
44
45
46
47
48

49 In the GWAS conducted by Hofmann et al⁴, the strongest association signal was
50
51 observed for rs2789679. This marker is located in the 3' - UTR of the ANXA11 gene.
52
53 The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in
54
55 the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls
56
57
58
59
60

1
2
3
4 being TT homozygous. Fine - mapping in the Black Women's Health Study (BWHS)
5
6 revealed rs2819941 to be the top SNP, associated with a nonsignificant ($P = 0.06$)
7
8 reduction in sarcoidosis risk (17% reduction per copy of the C-allele). In this case -
9
10 control association study, the frequency of the T allele in rs2789679 and the
11
12 rs2819941 C allele frequency in the patients with sarcoidosis was significantly lower
13
14 than that in the healthy controls. Several lines of evidence suggest that the observed
15
16 associations are unlikely to be an artifact. First, both the single - SNP and the
17
18 haplotype - based association analyses support the current finding. Second, population
19
20 stratification is an impossible reason, because all of our samples were from the same
21
22 geographical region. Finally, consistent results were obtained from two genetically
23
24 independent populations (Chinese Han and Europeans). Collectively, our results
25
26 confirmed the strong association between variations in the ANXA11 gene and
27
28 sarcoidosis, and suggest that ANXA11 represents a strong genetic risk factor for
29
30 sarcoidosis.
31
32
33
34
35
36
37
38

39 We further investigated the interaction among polymorphisms and observed
40
41 strong LD. The haplotype analysis revealed that significantly more C - C (block 2,
42
43 rs1049550 - rs2573351) and significantly fewer T - C haplotypes (block 2) were
44
45 found in the patients with sarcoidosis. These results indicated that the patients with C
46
47 - C (block 2) haplotypes of the ANXA11 gene were more prone to sarcoidosis.
48
49 Significantly higher frequencies of T - C haplotypes were detected in the healthy
50
51 controls than in the patients with sarcoidosis, suggesting that they may show a
52
53 protective effect against sarcoidosis. Our sample size can detect SNP and haplotype
54
55
56
57
58
59
60

1
2
3
4 associations with 90% and 85% power, respectively, at a false positive rate of 5%,
5
6 disease prevalence of 1%, disease allele / haplotype frequency of 0.05 / 0.03, and a
7
8 presumed odds ratio (OR) of 1.5. To some extent, this finding further supports a role
9
10 of ANXA11 polymorphisms in sarcoidosis, while ethnic group difference may exist.
11
12

13
14 In conclusion, our study suggests a potential role of the ANXA11 SNPs and their
15
16 related haplotypes in the genetic susceptibility to sarcoidosis. Further studies are
17
18 needed to investigate how these SNPs affect the function of ANXA11.
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

44 **Funding**

45
46 The project was supported by the Research and Development Foundation of
47
48 Science and Technology of Henan Province (2012k15).
49
50

51 **Contributors**

52
53 LG Zhang was involved in conception and design of the study, acquisition of
54
55 patient data, genotyping and drafted the paper. ZJ Xiao and GC Shi were involved in
56
57
58
59
60

1
2
3 design of the study, genotyping and interpretation of results. J Huang and CX Zhang
4
5 were involved in design of the study, acquisition of patient data and interpretation of
6
7 result. All authors revised the draft paper.
8
9

10 11 **Competing interests**

12
13 None.

14 15 **Patient consent**

16
17 Obtained.

18 19 **Ethics approval**

20
21 Ethics approval was provided by the First Hospital Affiliated to the Xinxiang
22
23 Medical College.
24
25

26 27 **Provenance and peer review**

28
29 Not commissioned; externally peer reviewed.

30 31 **Data sharing statement**

32
33 No additional data are available.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

- 1
2
3
4 analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral
5
6 blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol*
7
8 2006;26:243-50.
9
10
11 10. Agostini C, Trentin L, Facco M, *et al.* Role of IL-15, IL-2, and their receptors in
12
13 the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol*
14
15 1996;157:910-8
16
17
18 11. Jullien D, Sieling PA, Uyemura K, *et al.* IL-15, an immunomodulator of T cell
19
20 responses in intracellular infection. *J Immunol* 1997;158:800-6
21
22
23
24 12. Muller-Quernheim J, Schurmann M, Hofmann S, *et al.* Genetics of sarcoidosis.
25
26 *Clin Chest Med* 2008;29:391-414.
27
28
29 13. Morais A, Lima B, Peixoto M, *et al.* Annexin A11 gene polymorphism (R230C
30
31 variant) and sarcoidosis in a Portuguese population. *Tissue antigens*
32
33 2013;82:186-91.
34
35
36 14. Li Y, Pabst S, Kubisch C, *et al.* First independent replication study confirms the
37
38 strong genetic association of ANXA11 with sarcoidosis. *Thorax*
39
40 2010;65:939-40.
41
42
43 15. Mrazek F, Stahelova A, Kriegova E, *et al.* Functional variant ANXA11 R230C:
44
45 true marker of protection and candidate disease modifier in sarcoidosis. *Genes*
46
47 *Immun* 2011;12:490-4.
48
49
50
51 16. de Bakker PI, Yelensky R, Pe'er I, *et al.* Efficiency and power in genetic
52
53 association studies. *Nat Genet* 2005;37:1217-23.
54
55
56 17. Barrett JC, Fry B, Maller J, *et al.* Haploview: analysis and visualization of LD and
57
58
59
60

- 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	rs27896 79, rs70915 65, rs10495 50
Mrazek F, 2011	Caucasian	Czech	Case-control	245	254	1	rs10495 50
Cozier YC, 2012	Caucasian	America	Case-control	486	943	10p12-10q22	rs27896 79, rs28199 41
Li Y, 2013	Caucasian	Germany	Case-control	325	364	1	rs10495 50

Levin AM, 2013	Caucasian	American	Case-control	1689	1252	25	rs1049550
Morais A, 2013	Caucasian	Portuguese	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)		P-value ^a	OR, 95% CI	Sarcoidosis (n=412)		P-value ^a	OR, 95% CI	
	N, %	N, %			N, %	N, %			
rs278			0.00		rs108		0.2		
9679			07		8758		90		
					1				
AA	79, 18.9	132, 32.0	0.00	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	0.01	1.42, 1.08-1.87	TC	190, 45.5	208, 50.5	0.148	0.82, 0.62-1.08

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

1										
2										
3		49.3	44.2	7	0.94-1.68		45.0	44.7	20	0.77-1.
4										33
5										1.08,
6		60,	54,	0.59	1.11,		52,	48,	0.7	0.71-1.
7	GG	14.4	13.1	7	0.75-1.65	CC	12.4	11.7	3	64
8										0.96,
9										0.78-1.
10	Per G	326,	290,	0.10	0.85,	Per C	292,	280,	0.6	17
11	allele	38.0	35.2	9	0.70-1.04	allele	34.9	34.0	9	
12										
13										
14	rs127			0.11		rs257			0.8	
15	63624			8		3356			92	
16										1.05,
17	TT	126,	150,	0.00	0.68,	CC	118,	112,	0.7	0.78-1.
18		75.6	36.4	8	0.51-0.91		28.2	27.2	42	43
19										0.94,
20	CT	216,	186,	0.05	1.30,	CT	190,	194,	0.6	0.71-1.
21		51.7	45.2	9	0.99-1.71		45.5	47.1	39	23
22										
23										1.03,
24	CC	76,	76,	0.41	1.16,	TT	110,	106,	0.8	0.76-1.
25		18.2	18.5	1	0.81-1.67		26.3	25.7	43	41
26										0.99,
27	Per C	368,	338,	0.21	0.88,	Per T	410,	406,	0.9	0.82-1.
28	allele	44.0	41.0	6	0.73-1.08	allele	49.0	49.3	26	20
29										
30										
31										
32										
33	rs104			0.00		rs281			0.1	
34	9550			02		9945			01	
35										0.85,
36	CC	154,	208,	0.00	0.57,	GG	119,	130,	0.2	0.63-1.
37		36.8	50.5	07	0.43-0.75		28.5	31.6	94	15
38										1.35,
39	CT	190,	170,	0.22	1.19,	GA	221,	188,	0.0	1.03-1.
40		45.5	41.3	1	0.90-1.56		52.9	45.6	30	78
41										0.78,
42	TT	74,	34,	0.00	2.39,	AA	78,	94,	0.1	0.55-1.
43		17.7	8.3	08	1.55-3.68		18.7	22.8	38	09
44										1.02,
45	Per T	338,	238,	0.00	0.60,	Per A	377,	376,	0.8	0.84-1.
46	allele	40.4	28.8	07	0.49-0.73	allele	45.1	45.6	27	24
47										
48										
49										
50										
51										
52	rs257			0.43		rs112			0.0	
53	3351			7		0205			64	
54						9				
55										
56	TT	120,	104,	0.26	1.19,	GG	106,	128,	0.0	0.75,
57		28.7	25.2	3	0.88-1.62		25.4	31.1	67	0.56-1.
58										
59										
60										

									02
									1.38,
	CT	202,	216,	0.23	0.85,	GA	226,	190,	0.0
		48.3	52.4	8	0.65-1.12		54.1	46.1	21
									1.05-1.81
									0.88,
	CC	96,	92,	0.82	1.04,	AA	86,	94,	0.4
		23.0	22.3	6	0.75-1.44		20.6	22.8	29
									0.63-1.22
	Per C	394,	400,	0.56	1.06,	Per A	398,	378,	0.4
	allele	47.1	48.5	4	0.87-1.28	allele	47.6	45.9	79
									0.93,
									0.77-1.13
	rs281			0.00					
	9941			02					
	CC	114,	94,	0.14	1.27,				
		27.3	22.8	1	0.93-1.74				
	CT	226,	190,	0.02	1.38,				
		54.1	46.1	1	1.05-1.81				
	TT	78,	128,	0.00	0.51,				
		18.7	31.1	04	0.37-0.70				
	Per T	382,	446,	0.00	0.71,				
	allele	45.7	54.1	1	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)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

For peer review only

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 ($P < 0.05$).

1
2
3
4 1 **Table 4** ANXA11 haplotype in block 2 frequencies and the results of their associations with risk
5
6 2 of sarcoidosis
7
8
9 3

Haplotype			Genecounting (frequency %)			
ID	rs1049550	rs2573351	Cases	Controls	<i>P</i> -value ^a	Global <i>P</i> ^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	

4 ^a Based on 10,000 permutations.

5 ^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	<i>P</i> -value ^a	Global <i>P</i> ^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.

1
2
3
4 1 **Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a**
5
6 2 **Han Chinese population**
7

8
9 3 LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang

10
11 4 Internal Medicine - Tuberculosist, First Hospital Affiliated to the Xinxiang Medical
12
13 5 College, Xinxiang, Henan, PR China

14
15
16 6 *Corresponding author: Zhijian Xiao, Internal Medicine - Tuberculosisthe, First
17
18 7 Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China

19
20
21 8 No.88 Jiankang Road, Weihui City, Henan, 463000, PR China

22
23
24 9 E-mail: zhangligongys@163.com

25
26 10 Tel: +86 0373 - 4402229 (office)

27
28
29 11 Fax: +86 0373 -4402251

30
31 12 **Key words:** Sarcoidosis; Annexin A11; Single-nucleotide polymorphism

32
33
34 13 Word count: 2909

35
36
37 14

1
2
3
4 **Abstract**

5
6 **Objectives:** To further identify the single - nucleotide polymorphisms (SNPs) that
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
3 contribute to the genetic susceptibility to sarcoidosis, we examined the potential
4 association between sarcoidosis and fifteen SNPs of the ANXA11 gene.

5 **Design:** A case - control study.

6 **Setting:** A tuberculosis unit in a university hospital in China.

7 **Participants:** Participants included 412 patients with sarcoidosis and 418 healthy
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
8 controls.

9 **Methods:** The selected SNPs were genotyped in cases and controls by using the
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
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).

13 **Results:** The results showed that statistically significant differences were observed in
14
15
16
17
18
19
20
21
22
15 the allelic or genotypic frequencies of the rs2789679, rs1049550 and rs2819941 in the
16
17
18
19
20
21
22
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 ($\chi^2 = 12.849$, $P = 0.0003$, OR = 0.703,
95%CI = 0.579 - 0.852), T allele in rs1049550 ($\chi^2 = 24.420$, $P = 0.0007$, OR = 0.617,
95%CI = 0.503 - 0.758) and the rs2819941 C allele frequency ($\chi^2 = 11.803$, $P = 0.001$,
OR = 0.713, 95%CI = 0.588 - 0.865) in the patients with sarcoidosis was significantly

1
2
3
4 1 ~~lower than that in the controls.~~ Comparison of genotype frequency distribution
5
6 2 revealed significant differences between chest radiographic (CXR) stage I and stages
7
8
9 3 II - IV for rs1049550. The significantly more CC genotype ($P = 0.012$) were found in
10
11 4 the patients with stages I sarcoidosis. Furthermore, strong linkage disequilibrium (LD)
12
13
14 5 was observed in ~~two~~four blocks ($D' > 0.9$). In block 2 (rs1049550-rs2573351), the T -
15
16 6 C haplotype occurred significantly less frequently ($P = 0.001$), whereas the C - C
17
18
19 7 haplotypes occurred more frequently ($P = 0.0001$).

20
21 8 **Conclusions:** These findings point to a role for ANXA11 gene polymorphisms in
22
23
24 9 sarcoidosis in a Han Chinese population, and may be informative for future genetic or
25
26
27 10 biological studies on sarcoidosis.

28
29 11
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 **1 Article summary**

5
6 **2 Article focus**

7
8
9 3 1. The first genome - wide association study (GWAS) in sarcoidosis conducted in a
10 German population has revealed that the rs1049550 (C > T, Arg230Cys) located in
11 exon 10 were associated with sarcoidosis.
12

13
14
15
16 6 2. More studies should be performed to demonstrate the following item: whether these
17 SNPs modulate the risk of sarcoidosis by themselves or whether they correlate with
18 other causative SNPs and are repeated in other populations.
19

20
21
22 9 3. We hypothesize that common variants in the ANXA11 gene may significantly
23 contribute to the predisposition to develop sarcoidosis in a Chinese Han population.
24
25
26
27

28
29 **11 Key messages**

30
31 12 1. ANXA11 rs2789679 (3'UTR) was found to be associated with sarcoidosis.

32
33 13 2. ANXA11 rs1049550 (exon 10) was significantly associated with sarcoidosis.

34
35 14 3. Another significant association was observed for rs2819941 (intron 14).
36
37
38

39 **15 Strengths and limitations of this study**

40
41 16 1. Functional SNPs in the promoter region, 5'- and 3'-UTR, exons of ANXA11 gene
42 were systematically screened and homogeneity of the study subjects, representing the
43 Han Chinese, is the main strength of the current study.
44
45
46
47

48
49 19 2. Functional characterization is the potential limitation of the study that could have
50 further helped in proving the positive association observed for rs2789679 and
51 rs1049550. The lack of correlation of serum ANXA11 levels are needed to
52
53
54
55
56
57 investigate.
58
59
60

~~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 Löfgren'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 sarcoidosis¹³.

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 division⁵⁻⁷. 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.

1
2
3
4 1 Between these cells, the uncoordinated interplay results in the formation of typical
5
6 2 non - caseating granulomas. The mechanism by which the granulomas resolve has not
7
8
9 3 been fully elucidated. However, it is generally assumed that the induction of apoptosis
10
11 4 and / or by the withdrawal of inflammatory cytokines participates in the
12
13 5 disappearance of granulomas ^{10 11}. In the patients with sarcoidosis, their peripheral
14
15 6 blood monocytes were characterized by apoptosis ⁹.

17
18
19 7 In previous studies, case - control / family ‘hypothesis - driven’ studies and low
20
21 8 density linkage scans were used to identify genetic factors conferring the genetic
22
23 9 susceptibility to sarcoidosis ¹². Notably, the first genome - wide association study
24
25 10 (GWAS) in sarcoidosis conducted in a German population has recently revealed an
26
27 11 association: the leading rs2789679 SNP (single - nucleotide polymorphism) located in
28
29 12 the 3’ - untranslated region (3’ - UTR) and a common nonsynonymous SNP
30
31 13 (rs1049550, C > T, Arg230Cys) were associated with increased risk of sarcoidosis ^{4 13}.
32
33 14 This association has recently been supported by another report from the same
34
35 15 population ¹⁴. Thus, more studies should be performed to demonstrate the following
36
37 16 item: whether these SNPs modulate the risk of sarcoidosis by themselves or whether
38
39 17 they correlate with other causative SNPs and are repeated in other populations. The
40
41 18 published sarcoidosis - association studies for ANXA11 are summarized in Table 1.
42
43 19 We hypothesize that common variants in the ANXA11 gene may significantly
44
45 20 contribute to the predisposition to develop sarcoidosis.

46
47
48
49 21 In this study, we investigated fifteen loci in a Chinese population from He’nan
50
51 22 province (China) to verify the putative association between ANXA11 polymorphisms
52
53
54
55
56
57
58
59
60

1 and sarcoidosis.

2 SUBJECTS AND METHODS

3 Subjects

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

1 ~~2009 and Sept 2013. No familial relationship was known between the study~~
2 ~~participants. All participants were from a non-genetically related Chinese Han~~
3 ~~population in He'nan Province (China).~~ The study was performed according to the
4 Guidelines of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang,
5 China). Written informed consent was obtained from each participant in this study.

6 **SNP selection**

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

~~Marker selection was done according to previous studies⁴⁻¹²⁻¹⁴⁻¹⁵, 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) \geq 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

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

1 rs2789679, rs1049550 and rs2819941. The frequency of the T allele in rs2789679 (P
2 = 0.0003, OR = 0.70, 95%CI = 0.58 - 0.85), T allele in rs1049550 (P = 0.0007, OR =
3 0.62, 95%CI = 0.50 - 0.76) and the rs2819941 C allele frequency (P = 0.001, OR =
4 0.71, 95%CI = 0.59 - 0.87) in the patients with sarcoidosis was significantly lower
5 than that in the controls, and these differences remained statistically significant after
6 Bonferroni corrections. There were no interactions among rs2789679, rs1049550 and
7 rs2819941 ($P > 0.05$).

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

13 To assess particular disease phenotypes, the patients with sarcoidosis were
14 divided into the subgroups according to their chest radiographic (CXR) stage. CXR
15 stage I (isolated bilateral hilar lymphadenopathy): N = 196; CXR stages II - IV
16 (infiltration of parenchyma): 167; CXR stage 0 (absence on pulmonary disease): N =
17 28; the information on CXR stage was not available for 21 patients. Comparison of
18 genotype frequency distribution revealed significant differences between stage I and
19 stages II - IV for rs1049550. The significantly more CC genotype ($P = 0.012$) were
20 found in the patients with stage I sarcoidosis (table 3).

21 DISCUSSION

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

1
2
3
4 1 in different populations. To determine if common risk variants exist in distinct
5
6 2 populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene,
7
8
9 3 were genotyped in samples from the patients with sarcoidosis and healthy controls in
10
11 4 a Chinese Han population. Recently, new sarcoidosis loci have been identified by
12
13 5 genome - wide association studies ^{4 14}. With the fast development of genome - wide
14
15 6 association studies, increasing number of susceptibility loci for sarcoidosis have been
16
17 7 reported in different populations ^{4 14 15}. However, these observations should be
18
19 8 confirmed in other genetically independent populations. In this study, we conducted
20
21 9 the first large genetic association study of the ANXA11 gene in a Chinese Han
22
23 10 population. The evidence of markers associated with sarcoidosis was presented, and
24
25 11 these markers were mapped to different locations in the ANXA11 gene (81897864 -
26
27 12 81951001). The association signals in the region were identified, and some
28
29 13 significantly associated haplotypes also appeared this region.
30
31
32
33
34
35

36
37 14 Hoffman et al. reported an association between the T allele of the non -
38
39 15 synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a
40
41 16 German population ⁴. Similar results were also obtained in other two European
42
43 17 populations ^{14 15}. As a part of the sarcoidosis GWAS done in Americans ²³, we further
44
45 18 confirmed this association in a Chinese Han population. In this study, the frequency of
46
47 19 ANXA11 rs1049550T allele was significantly lower in the patients with sarcoidosis
48
49 20 than that in the healthy controls. Comparison of stages II - IV, the significantly more
50
51 21 CC genotypewas found in the patients with stages I sarcoidosis. The “protective”
52
53 22 effect of ANXA11 T allele increased with the number of its copies in the genotype.
54
55
56
57
58
59
60

1 which is consistent with the result obtained in the patients with sarcoidosis in a
2 German population⁴, Furthermore, the T allele carriers among the patients were
3 protected from infiltration of lung parenchyma (radiographic stages II-IV)¹⁵. We have
4 also demonstrated that rs1049550 is significant cross - ethnically at the gene level
5 after adjustment for the single SNP association tests performed. The mechanism by
6 which rs1049550 affects the susceptibility to sarcoidosis may be that the SNP may
7 affect the function of the ANXA11 protein rather than affect its expression. In humans,
8 the ANXA11 protein consists of an N - terminalproline - tyrosine - glycine (PYG) -
9 rich region, followed by four annexin core domains. The rs1049550 leads to an amino
10 - acid exchange (basic arginine to a polar cysteine) at the evolutionarily conserved
11 position 230 (R230C) in the first annexin domain, which is responsible for Ca²⁺ -
12 dependent trafficking of the protein in the cell²⁴.

13 In the GWAS conducted by Hofmann et al⁴, the strongest association signal was
14 observed for rs2789679. This marker is located in the 3' - UTR of the ANXA11 gene.
15 The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in
16 the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls
17 being TT homozygous. Fine - mapping in the Black Women's Health Study (BWHS)
18 revealed rs2819941 to be the top SNP, associated with a nonsignificant ($P = 0.06$)
19 reduction in sarcoidosis risk (17% reduction per copy of the C-allele). In this case -
20 control association study, the frequency of the T allele in rs2789679 and the
21 rs2819941 C allele frequency in the patients with sarcoidosis was significantly lower
22 than that in the healthy controls. Several lines of evidence suggest that the observed

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

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

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

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

2 **Contributors**

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

8 **Funding**

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

11 **Competing interests**

12 None.

13 **Patient consent**

14 Obtained.

15 **Ethics approval**

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

18 **Provenance and peer review**

19 Not commissioned; externally peer reviewed.

20 **Data sharing statement**

21 No additional data are available.

22 **Figure Legend**

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

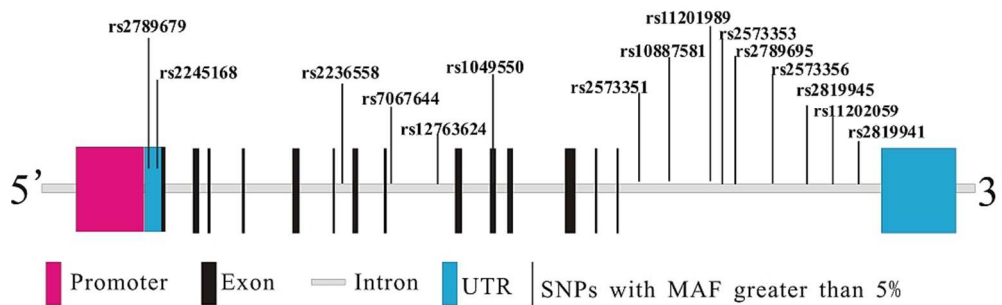
2 **Fig. 2** The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the
3 pair-wise calculation of r^2 (left) or D' (right). Black squares indicate $r^2 = 1$ (i.e. perfect
4 LD between a pair of SNPs). Empty squares indicate $D' = 1$ (i.e. complete LD
5 between a pair of SNPs).

6 Reference

- 7 1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007;
8 357:2153-65.
- 9 2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-8
- 10 3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis.
11 *Semin Respir Crit Care Med* 2010;31:380-9.
- 12 4. Hofmann S, Franke A, Fischer A, *et al.* Genome-wide association study identifies
13 ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet*
14 2008;40:1103-6.
- 15 5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev*
16 2002;82:331-71.
- 17 6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using
18 recombinant human annexin XI-specific autoantibodies cloned by phage
19 display. *J Biol Chem* 2003;278:33120-6.
- 20 7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and
21 completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-22.
- 22 8. Petzmann S, Maercker C, Markert E, *et al.* Enhanced proliferation and decreased

- 1
2
3
4 1 apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis Vasc Diffuse*
5
6 2 *Lung Dis* 2006;23:190-200
7
8
9 3 9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, *et al.* Comparative
10
11 4 analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral
12
13 5 blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol*
14
15 6 2006;26:243-50.
16
17
18 7 10. Agostini C, Trentin L, Facco M, *et al.* Role of IL-15, IL-2, and their receptors in
19
20 8 the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol*
21
22 9 1996;157:910-8
23
24
25 10 11. Jullien D, Sieling PA, Uyemura K, *et al.* IL-15, an immunomodulator of T cell
26
27 11 responses in intracellular infection. *J Immunol* 1997;158:800-6
28
29
30 12 12. Muller-Quernheim J, Schurmann M, Hofmann S, *et al.* Genetics of sarcoidosis.
31
32 13 *Clin Chest Med* 2008;29:391-414.
33
34
35 14 13. Morais A, Lima B, Peixoto M, *et al.* Annexin A11 gene polymorphism (R230C
36
37 15 variant) and sarcoidosis in a Portuguese population. *Tissue antigens*
38
39 16 2013;82:186-91.
40
41
42 17 14. Li Y, Pabst S, Kubisch C, *et al.* First independent replication study confirms the
43
44 18 strong genetic association of ANXA11 with sarcoidosis. *Thorax*
45
46 19 2010;65:939-40.
47
48
49 20 15. Mrazek F, Stahelova A, Kriegova E, *et al.* Functional variant ANXA11 R230C:
50
51 21 true marker of protection and candidate disease modifier in sarcoidosis. *Genes*
52
53 22 *Immun* 2011;12:490-4.
54
55
56
57
58
59
60

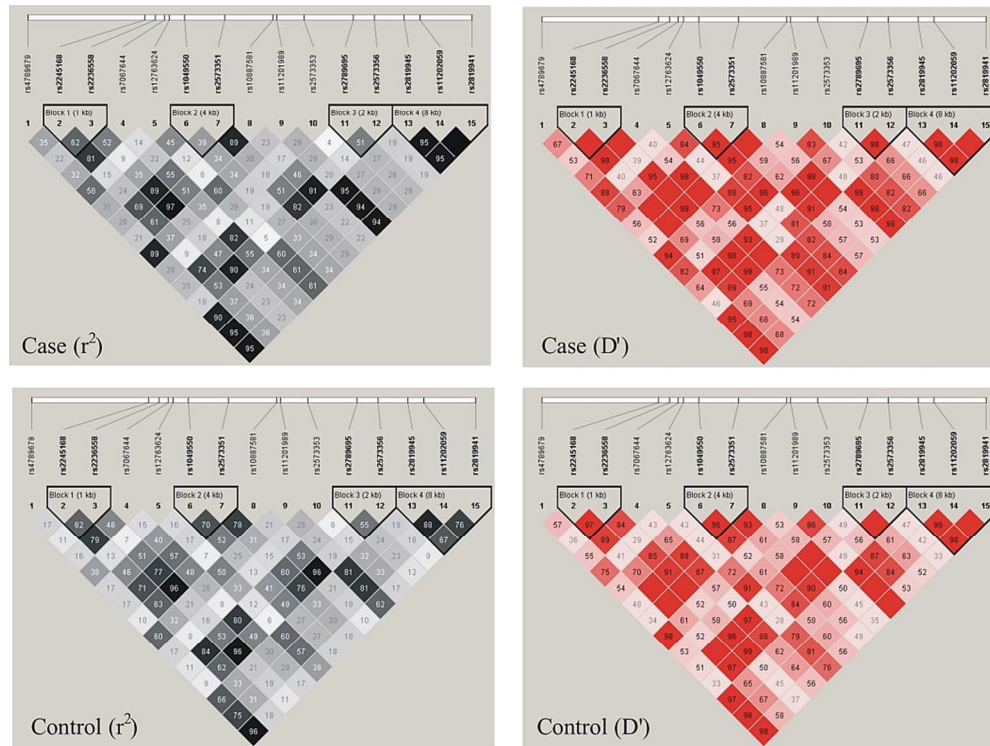
- 1
2
3
4 16. de Bakker PI, Yelensky R, Pe'er I, *et al.* Efficiency and power in genetic
5
6 association studies. *Nat Genet* 2005;37:1217-23.
7
8
9 17. Barrett JC, Fry B, Maller J, *et al.* Haploview: analysis and visualization of LD and
10
11 haplotype maps. *Bioinformatics* 2005;21:263-5.
12
13
14 18. Gabriel SB, Schaffner SF, Nguyen H, *et al.* The structure of haplotype blocks in
15
16 the human genome. *Science* 2002;296:2225-9.
17
18
19 19. Curtis D, Knight J, Sham PC. Program report: GENECOUNTING support
20
21 programs. *Ann Hum Genet* 2006;70:277-9.
22
23
24 20. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic
25
26 associations. *Hum Hered* 2000;50:133-9.
27
28
29 21. Zhao JH, Lissarrague S, Essioux L, *et al.* GENECOUNTING: haplotype analysis
30
31 with missing genotypes. *Bioinformatics* 2002;18:1694-5
32
33
34 22. Zhao JH, Sham PC. Generic number systems and haplotype analysis. *Comput*
35
36 *Methods Programs Biomed* 2003;70:1-9
37
38
39 23. Levin AM, Iannuzzi MC, Montgomery CG, *et al.* Association of ANXA11 genetic
40
41 variation with sarcoidosis in African Americans and European Americans.
42
43 *Genes Immun* 2013;14:13-8.
44
45
46 24. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11
47
48 with the nuclear envelope. *J Biol Chem* 2003;278:20210-6.
49
50
51
52
53
54
55
56
57
58
59
60



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

peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



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).
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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

only

1
2
3
4 1 **Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a**

5
6 2 **Han Chinese population - a case-control study**

7
8
9 3 LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang

10
11 4 The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang

12
13
14 5 Medical College, Xinxiang, Henan, PR China

15
16 6 *Corresponding author: Zhijian Xiao, The Third Department of Tuberculosis, the First

17
18
19 7 Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China.

20
21 8 No.88 Jiankang Road, Weihui City, Henan, 463000, PR China

22
23
24 9 E-mail: zhangligongys@163.com

25
26 10 Tel: +86 0373 - 4402229 (office)

27
28
29 11 Fax: +86 0373 -4402251

30
31 12 **Key words:** Sarcoidosis; Annexin A11; Single-nucleotide polymorphism

32
33
34 13 Word count: 3787

35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60 14

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 Abstract

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

5 **Design:** A case - control study.

6 **Setting:** A tuberculosis unit in a hospital of the university in China.

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

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

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

1
2
3
4 1 rs1049550, CC genotype was significantly more in the patients with chest
5
6 2 radiographic (CXR) stage I sarcoidosis than the patients with CXR stage II-IV
7
8
9 3 sarcoidosis ($P = 0.012$).

10
11 4 **Conclusions:** These findings point to a role for the polymorphisms of ANXA11 in
12
13 5 sarcoidosis in a Chinese Han population, and may be informative for future genetic
14
15 6 or biological studies on sarcoidosis.
16
17
18
19 7

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 **Article summary**

2 **Article focus**

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

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

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

11 **Key messages**

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

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

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

15 **Strengths and limitations of this study**

16 1. A systematical screening of the functional SNPs in the promoter region, 5'- and
17 3'-UTR, exons of the ANXA11 gene, and the homogeneity of the study subjects
18 representing the Chinese Han are the main strengths of this study. Chinese Han

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

22

1 INTRODUCTION

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

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

1
2
3
4 1 However, it is generally assumed that the induction of apoptosis and/or by the
5
6 2 withdrawal of inflammatory cytokines participates in the disappearance of
7
8 3 granulomas^{10 11}. In the patients with sarcoidosis, their peripheral blood monocytes
9
10 4 were characterized by apoptosis⁹.

11
12
13
14 5 In previous studies, case - control/family ‘hypothesis - driven’ studies and low
15
16 6 density linkage scans were used to identify genetic factors conferring the genetic
17
18 7 susceptibility to sarcoidosis¹². Notably, the first genome - wide association study
19
20 8 (GWAS) in sarcoidosis conducted in a German population has recently revealed an
21
22 9 association: the leading rs2789679 SNP (single - nucleotide polymorphism) located in
23
24 10 the 3’-untranslated region (3’-UTR) and a common nonsynonymous SNP (rs1049550,
25
26 11 C > T, Arg230Cys) were associated with the increased risk of sarcoidosis^{4 13}. This
27
28 12 association has recently been supported by another report from the same population¹⁴.
29
30 13 Thus, more studies should be performed to demonstrate the following items: whether
31
32 14 these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate
33
34 15 with other causative SNPs and are repeated in other populations. The published
35
36 16 studies about the association of sarcoidosis and ANXA11 are summarized in
37
38 17 Supplemental table 1. We hypothesize that common variants in the ANXA11 gene
39
40 18 may significantly contribute to the predisposition to develop sarcoidosis.
41
42
43
44
45
46
47
48

49 19 In this study, we investigated fifteen loci in a Chinese population from He’nan
50
51 20 province (China) to verify the putative association between ANXA11 polymorphisms
52
53 21 and sarcoidosis.
54
55

56 22 **SUBJECTS AND METHODS**

57
58
59
60

1 **Subjects**

2 Four hundred and twelve patients with sarcoidosis (mean \pm SD age: 53.6 ± 4.6
3 years; 155 men and 257 women) were recruited from our hospital between May 2005
4 and March 2013. Sarcoidosis was diagnosed by the evidence of non-caseating
5 epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities.
6 Chronic sarcoidosis was defined as a disease over at least 2 years or at least two
7 episodes in a lifetime. A cute sarcoidosis was defined as one episode of acute
8 sarcoidosis which had totally resolved at the date of the examination. The control
9 group consisted of 418 unrelated healthy subjects (mean \pm SD age: 54.2 ± 5.3 years;
10 160 men and 258 women) who underwent health examinations in the Medical
11 Examination Center of the First Affiliated Hospital of the Xinxiang Medical College
12 (Xinxiang, China) between Oct 2009 and Sept 2013. None of the individuals in the
13 control group had a history of lung diseases or showed any symptoms of the lung or
14 other diseases by chest radiography or laboratory blood tests. Participants were
15 excluded if they: were taking other prescribed medications that could affect the
16 central nervous system; had a history of seizures, hematological diseases, or severe
17 liver or kidney impairment; or were pregnant. No familial relationship was known
18 between the study participants. The study was performed according to the Guidelines
19 of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang, China).
20 Written informed consent was obtained from each participant in this study.

21 **SNP selection**

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

12 Genotyping

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

1
2
3
4 1 of - flight mass spectrometer. Genotype calling was performed in realtime with the
5
6 2 MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer
7
8
9 3 software version 3.4 (Sequenom).

10 4 **Statistical analysis**

11
12
13
14 5 All data were analyzed using the SPSS 17.0 software (SPSS Inc., Chicago, IL,
15
16 6 USA). Hardy-Weinberg equilibrium was evaluated by Chi-square tests. Differences
17
18 7 between the cases and controls in the frequency of the alleles, genotypes and
19
20 8 haplotypes were evaluated by the Fisher's exact test or the Pearson Chi-square test.
21
22 9 Unconditional logistic regression was used to calculate the odds ratio (OR) and 95%
23
24 10 confidence interval (CI) in independent association between each locus and the
25
26 11 presence of sarcoidosis. Logistic regression was used to evaluate the interaction
27
28 12 effects between gene and gender or age. Gender and age of subjects were treated as
29
30 13 covariants in binary logistic regression. *P* values were calculated based on
31
32 14 codominant, dominant for the rare allele, heterosis and recessive for the rare allele
33
34 15 models of inheritance. The Bonferroni correction was used to adjust the test level
35
36 16 when multiple comparisons were conducted, and the *P* value was divided by the total
37
38 17 number of loci. Haplotype blocks were defined according to the criteria developed by
39
40 18 Gabriel et al.¹⁵. Pair-wise LD statistics (D' and r^2) and haplotype frequency were
41
42 19 calculated, and haplotype blocks were constructed using the Haploview 4.0¹⁶. To
43
44 20 ensure that the LD blocks most closely reflect the population level LD patterns,
45
46 21 definition of the blocks were based on the control samples alone. The haplotype
47
48 22 frequencies were estimated using GENECOUNTING, which computes maximum -
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 likelihood estimates of haplotype frequencies from unknown phase data by utilizing
5
6 an expectation - maximization algorithm¹⁷⁻²⁰. The significance of any haplotypic
7
8 association was evaluated using a likelihood ratio test, followed by permutation
9
10 testing that compared estimated haplotype frequencies in cases and controls^{17 19}.
11
12

13 RESULTS

14
15
16 The genotype distribution of the fifteen polymorphisms was consistent with the
17
18 Hardy-Weinberg equilibrium ($P > 0.05$). The analysis of strong LD in the patients
19
20 with sarcoidosis and healthy controls revealed that 2 SNPs (rs2245168, rs2236558), 2
21
22 SNPs (rs1049550, rs2573351), 2 SNPs (rs2789695, rs2573356) and 3 SNPs
23
24 (rs2819945, rs11202059, rs2819941) were located in haplotype block 1, block 2,
25
26 block 3 and block 4 ($D' > 0.9$, Fig. 1). The genotype distribution, allelic frequencies,
27
28 and haplotypes in the patients with sarcoidosis and healthy controls are showed in
29
30 Tables 1-3 and Supplemental table 2
31
32
33
34
35

36
37 Comparison of genotype and allele frequency distribution revealed significant
38
39 differences between the patients with sarcoidosis and healthy controls for 3 SNPs:
40
41 rs2789679, rs1049550 and rs2819941. The frequency of the T allele in rs2789679 (P
42
43 = 0.0003, OR = 0.70, 95%CI = 0.58 - 0.85) and rs1049550 (P = 0.0007, OR = 0.62,
44
45 95%CI = 0.50 - 0.76), and C allele in rs2819941 (P = 0.001, OR = 0.71, 95%CI =
46
47 0.59 - 0.87) in the patients with sarcoidosis was significantly lower than that in the
48
49 controls, and these differences remained statistically significant after Bonferroni
50
51 corrections. We presented a multi-SNP model of sarcoidosis risk. We screen out the
52
53 most significant model to predict the risk by forward stepwise strategy in logistic
54
55
56
57
58
59
60

1 regression, and found that a model includes rs2789679 and rs1049550 is the best one
2
3
4
5
6 (Supplemental table 3), which suggests that there is a interaction between them.
7

8
9 We performed an association analysis to determine whether the haplotype was
10
11 associated with the risk of sarcoidosis (Global $P < 0.05$ in block 2 and block 3).
12
13 Compared with the controls, the T - C haplotype occurred significantly less frequently
14
15 ($P = 0.001$) but the C - C haplotypes occurred more frequently ($P = 0.0001$) in block 2
16
17 (rs1049550 - rs2573351) in the patients with sarcoidosis.
18
19

20
21 To assess particular disease phenotypes, the patients with sarcoidosis were
22
23 divided into the subgroups according to their chest radiographic (CXR) stage. CXR
24
25 stage I (isolated bilateral hilar lymphadenopathy): N = 196; CXR stages II - IV
26
27 (infiltration of parenchyma): 167; CXR stage 0 (absence on pulmonary disease): N =
28
29 28; the information on CXR stage was not available for 21 patients. Genotype
30
31 frequency distribution revealed that, in rs1049550, CC genotype was significantly
32
33 more in the patients with chest radiographic (CXR) stage I sarcoidosis than the
34
35 patients with CXR stage II-IV sarcoidosis ($P = 0.012$). (Table 4).
36
37
38
39

40 41 **DISCUSSION**

42
43 A key step in linkage and association studies is to identify common risk variants
44
45 in different populations. To determine if common risk variants exist in distinct
46
47 populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene,
48
49 were genotyped in samples from the patients with sarcoidosis and healthy controls in
50
51 a Chinese Han population. Recently, new sarcoidosis loci have been identified by
52
53 genome-wide association studies^{4 14}. With the fast development of genome - wide
54
55
56
57
58
59
60

1
2
3
4 1 association studies, an increasing number of susceptibility loci for sarcoidosis have
5
6 2 been found in different populations ^{4 14 21}. However, these observations should be
7
8
9 3 confirmed in other genetically independent populations. In this study, we conducted
10
11 4 the first large genetic association study of the ANXA11 gene in a Chinese Han
12
13
14 5 population. The evidence of markers associated with sarcoidosis was presented, and
15
16 6 these markers were mapped to different locations in the ANXA11 gene (81897864 -
17
18
19 7 81951001). The association signals in the region were identified, and some
20
21 8 significantly associated haplotypes also appeared this region.
22

23
24 9 Hoffman et al. reported an association between the T allele of the non -
25
26 10 synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a
27
28
29 11 German population ⁴. Similar results were also obtained in other two European
30
31 12 populations ^{14 21}. As a part of the sarcoidosis GWAS done in Americans ²², we further
32
33
34 13 confirmed this association in a Chinese Han population. In this study, the frequency of
35
36 14 ANXA11 rs1049550T allele was significantly lower in the patients with sarcoidosis
37
38
39 15 than the healthy controls. There is interaction between rs2789679 and rs1049550.
40
41 16 Genotype frequency distribution revealed that, in rs1049550, CC genotype was
42
43
44 17 significantly more in the patients with stage I sarcoidosis than the patients with stage
45
46 18 II-IV sarcoidosis. The “protective” effect of ANXA11 T allele increased with the
47
48
49 19 number of its copies in the genotype, which is consistent with the result obtained in
50
51 20 the patients with sarcoidosis in a German population ⁴, Furthermore, the T allele
52
53
54 21 carriers among the patients were protected from infiltration of lung parenchyma
55
56 22 (radiographic stages II-IV) ²¹. We have also demonstrated that rs1049550 is
57
58
59
60

1
2
3
4 1 significant cross - ethnically at the gene level after adjustment for the single SNP
5
6 2 association tests performed. The mechanism by which rs1049550 affects the
7
8 3 susceptibility to sarcoidosis is that the SNP may affect the function of the ANXA11
9
10 4 protein rather than affect its expression. In humans, the ANXA11 protein consists of
11
12 5 an N - terminalproline - tyrosine - glycine (PYG) - rich region, followed by four
13
14 6 annexin core domains. The rs1049550 leads to an amino - acid exchange (basic
15
16 7 arginine to a polar cysteine) at the evolutionarily conserved position 230 (R230C) in
17
18 8 the first annexin domain, which is responsible for Ca^{2+} - dependent trafficking of the
19
20 9 protein in the cell ²³.

21
22
23
24
25
26 10 In the GWAS conducted by Hofmann et al ⁴, the strongest association signal was
27
28 11 observed for rs2789679. This marker is located in the 3'-UTR of the ANXA11 gene.
29
30 12 The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in
31
32 13 the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls
33
34 14 being TT homozygous. Fine-mapping in the Black Women's Health Study (BWHS)
35
36 15 revealed rs2819941 to be the top SNP, associated with a non-significant ($P = 0.06$)
37
38 16 reduction in the risk of sarcoidosis (17% reduction per copy of the C-allele). In this
39
40 17 case-control association study, the frequency of the T allele in rs2789679 and the
41
42 18 rs2819941 C allele frequency in the patients with sarcoidosis was significantly lower
43
44 19 than that in the healthy controls. Several lines of evidence suggest that the observed
45
46 20 association is unlikely to be an artifact. First, both the single-SNP and the
47
48 21 haplotype-based association analyses support the current finding. Second, our samples
49
50 22 were from the same geographical region, excluding the . Finally, consistent results
51
52
53
54
55
56
57
58
59
60

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

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

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

20 **Contributors**

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

For peer review only

1
2
3
4 **Funding**

5
6 The project was supported by the Research and Development Foundation of
7
8
9 Science and Technology of Henan Province (2012k15).

10
11 **Contributorship Statement**

12
13
14 LG Zhang was involved in conception and design of the study, acquisition of
15
16 patient data, genotyping and drafted the paper. ZJ Xiao and GC Shi were involved in
17
18 design of the study, genotyping and interpretation of results. J Huang and CX Zhang
19
20 were involved in design of the study, acquisition of patient data and interpretation of
21
22 result. All authors revised the draft paper.
23
24

25
26 **Competing interests**

27
28
29 None.

30
31 **Data sharing statement**

32
33
34 No additional data are available.
35

36
37 **Patient consent**

38
39
40 Obtained.

41
42 **Ethics approval**

43
44
45 Ethics approval was provided by the First Hospital Affiliated to the Xinxiang
46
47 Medical College.
48

49
50 **Provenance and peer review**

51
52
53 Not commissioned; externally peer reviewed.
54
55

56
57
58
59
60

1
2
3
4 **Reference**
5

- 6 1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007;
7
8 357:2153-65.
9
10 2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-8
11
12 3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis.
13
14 *Semin Respir Crit Care Med* 2010;31:380-9.
15
16 4. Hofmann S, Franke A, Fischer A, *et al.* Genome-wide association study identifies
17
18 ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet*
19
20 2008;40:1103-6.
21
22 5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev*
23
24 2002;82:331-71.
25
26 6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using
27
28 recombinant human annexin XI-specific autoantibodies cloned by phage
29
30 display. *J Biol Chem* 2003;278:33120-6.
31
32 7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and
33
34 completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-22.
35
36 8. Petzmann S, Maercker C, Markert E, *et al.* Enhanced proliferation and decreased
37
38 apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis Vasc Diffuse*
39
40 *Lung Dis* 2006;23:190-200
41
42 9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, *et al.* Comparative
43
44 analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral
45
46 blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol*
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4 1 2006;26:243-50.
5
6 2 10. Agostini C, Trentin L, Facco M, *et al.* Role of IL-15, IL-2, and their receptors in
7
8 3 the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol*
9
10 4 1996;157:910-8
11
12 5 11. Jullien D, Sieling PA, Uyemura K, *et al.* IL-15, an immunomodulator of T cell
13
14 6 responses in intracellular infection. *J Immunol* 1997;158:800-6
15
16 7 12. Muller-Quernheim J, Schurmann M, Hofmann S, *et al.* Genetics of sarcoidosis.
17
18 8 *Clin Chest Med* 2008;29:391-414.
19
20 9 13. Morais A, Lima B, Peixoto M, *et al.* Annexin A11 gene polymorphism (R230C
21
22 10 variant) and sarcoidosis in a Portuguese population. *Tissue antigens*
23
24 11 2013;82:186-91.
25
26 12 14. Li Y, Pabst S, Kubisch C, *et al.* First independent replication study confirms the
27
28 13 strong genetic association of ANXA11 with sarcoidosis. *Thorax*
29
30 14 2010;65:939-40.
31
32 15 15. Gabriel SB, Schaffner SF, Nguyen H, *et al.* The structure of haplotype blocks in
33
34 16 the human genome. *Science* 2002;296:2225-9.
35
36 17 16. Barrett JC, Fry B, Maller J, *et al.* Haploview: analysis and visualization of LD and
37
38 18 haplotype maps. *Bioinformatics* 2005;21:263-5.
39
40 19 17. Curtis D, Knight J, Sham PC. Program report: GENECOUNTING support
41
42 20 programs. *Ann Hum Genet* 2006;70:277-9.
43
44 21 18. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic
45
46 22 associations. *Hum Hered* 2000;50:133-9.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4 19. Zhao JH, Lissarrague S, Essioux L, et al. GENECOUNTING: haplotype analysis
5
6 with missing genotypes. *Bioinformatics* 2002;18:1694-5
7
8
9 20. Zhao JH, Sham PC. Generic number systems and haplotype analysis. *Comput*
10
11 *Methods Programs Biomed* 2003;70:1-9
12
13
14 21. Mrazek F, Stahelova A, Kriegova E, et al. Functional variant ANXA11 R230C:
15
16 true marker of protection and candidate disease modifier in sarcoidosis. *Genes*
17
18 *Immun* 2011;12:490-4.
19
20
21 22. Levin AM, Iannuzzi MC, Montgomery CG, et al. Association of ANXA11 genetic
22
23 variation with sarcoidosis in African Americans and European Americans.
24
25 *Genes Immun* 2013;14:13-8.
26
27
28
29 24. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11
30
31 with the nuclear envelope. *J Biol Chem* 2003;278:20210-6.
32
33
34
35
36
37
38

15 **Figure Legend**

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

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

22

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)

1

2

3

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	<i>P</i> -value ^a	Global <i>P</i> ^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	

4

^a Based on 10,000 permutations.

5

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

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

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	<i>P</i> -value ^a	Global <i>P</i> ^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 1
5
6 2
7
8
9 3
10
11 4
12
13
14 5
15
16 6
17
18
19 7
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1
2
3
4 **Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a**
5
6 **Han Chinese population - a case-control study**
7

8 LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang
9

10
11 ~~The Third Department of Tuberculosis~~~~Internal Medicine, the First Hospital Affiliated~~
12 ~~to the Xinxiang Medical College—Tuberculosis, First Hospital Affiliated to the~~
13 ~~Xinxiang Medical College, Xinxiang, Henan, PR China~~
14
15
16
17

18
19 *Corresponding author: Zhijian Xiao, The Third Department of Tuberculosis, the First
20 Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR
21 China.~~Internal Medicine—Tuberculosis~~~~the, First Hospital Affiliated to the Xinxiang~~
22 ~~Medical College, Xinxiang, Henan, PR China~~
23
24
25
26
27

28
29 No.88 Jiankang Road, Weihui City, Henan, 463000, PR China
30

31 E-mail: zhangligongys@163.com
32

33
34 Tel: +86 0373 - 4402229 (office)
35

36 Fax: +86 0373 -4402251
37

38
39 **Key words:** Sarcoidosis; Annexin A11; Single-nucleotide polymorphism
40

41 Word count: 2909
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 tuberculosis 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

1
2
3
4 frequently ($P = 0.0001$) in the patients with sarcoidosis than the controls. Furthermore,
5
6 e-Comparison of genotype frequency distribution revealed that, in rs1049550, CC
7
8 genotype was significantly more in the patients with chest radiographic (CXR) stage I
9
10 sarcoidosis than the patients with CXR stage II-IV sarcoidosis ($P = 0.012$). significant
11
12 differences between chest radiographic (CXR) stage I and stages II–IV for rs1049550.
13
14 The significantly more CC genotype ($P = 0.012$) were found in the patients with
15
16 stages I sarcoidosis.
17
18 –Conclusions: These findings point to a role for the polymorphisms of ANXA11
19
20 gene polymorphisms in sarcoidosis in a Han-ChineseChinese Han population, and
21
22 may be informative for future genetic or biological studies on sarcoidosis.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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: w~~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~~3'-UTR) ~~was found to be associated~~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 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

1
2
3
4 [rs1049550 is a potential limitation of this study. Furthermore, the association of the](#)
5 [serum level of ANXA11 with sarcoidosis still needs to be investigated.](#)~~Functional~~
6 ~~characterization is the potential limitation of the study that could have further helped~~
7 ~~in proving the positive association observed for rs2789679 and rs1049550. The lack~~
8 ~~of correlation of serum ANXA11 levels are needed to investigate.~~

16 INTRODUCTION

17
18
19 Sarcoidosis is a systemic autoimmune disease characterized by destructive, non-
20 ~~caseating~~ epithelioid granulomatous lesions, with accumulated phagocytes and
21 ~~caseating~~ epithelioid granulomatous lesions, with accumulated phagocytes and
22 activated CD4⁺ T helper type 1 lymphocytes^{1 2}. The typical manifestations of
23 sarcoidosis include bilateral hilar lymphadenopathy, pulmonary infiltration and ocular
24 and skin lesions. The clinical course and prognosis of sarcoidosis are variable. The
25 acute forms of sarcoidosis such as Löfgren's syndrome (LS) can resolve
26 spontaneously in most cases. However, chronic pulmonary sarcoidosis may progress
27 to lung fibrosis ~~which and~~ eventually ~~result in causes~~ respiratory failure. Recent
28 studies have demonstrated that sarcoidosis can occur in a pattern of family clustering,
29 and its racial incidence is also different, suggesting that some genetic factors may
30 contribute to the risk and severity of sarcoidosis^{1 3}.

31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46 The Annexin gene family is involved in the etiology of several autoimmune and
47 chronic diseases^{4 5}. One member of the Annexin gene family, ANXA11, located on
48 chromosome 10q22.3, is involved in calcium signaling, apoptosis, vesicle trafficking,
49 cell growth and terminal phase of cell division⁵⁻⁷. In this context, the development
50 and maintenance of the granulomatous inflammation in sarcoidosis have been
51
52
53
54
55
56
57
58
59
60

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 apoptosis⁹.

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 sarcoidosis¹². 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 population¹⁴. 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

1
2
3
4 | [studies for ANXA11](#) are summarized in Supplemental table 21. We hypothesize that
5
6 common variants in the ANXA11 gene may significantly contribute to the
7
8 predisposition to develop sarcoidosis.
9

10
11 In this study, we investigated fifteen loci in a Chinese population from He'nan
12
13 province (China) to verify the putative association between ANXA11 polymorphisms
14
15 and sarcoidosis.
16
17

18 19 **SUBJECTS AND METHODS**

20 21 **Subjects**

22
23 Four hundred and twelve patients with sarcoidosis (mean \pm SD age: 53.6 ± 4.6
24
25 years; 155 men and 257 women) were recruited from our hospital between May 2005
26
27 and March 2013. Sarcoidosis was diagnosed by the evidence of non—caseating
28
29 epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities.
30
31
32 Chronic sarcoidosis was defined as a disease over at least 2 years or at least two
33
34 episodes in a lifetime. A acute sarcoidosis was defined as one episode of acute
35
36 sarcoidosis which had totally resolved at the date of the examination. The control
37
38 group consisted of 418 unrelated healthy subjects (mean \pm SD age: 54.2 ± 5.3 years;
39
40 160 men and 258 women) who underwent health examinations in the Medical
41
42 Examination Center of the First Affiliated Hospital of the Xinxiang Medical College
43
44 (Xinxiang, China) between Oct 2009 and Sept 2013. None of the individuals in the
45
46 control group had a history of lung diseases or showed any symptoms of the lung or
47
48 other diseases by chest radiography or laboratory blood tests. Participants were
49
50 excluded if they: were taking other prescribed medications that could affect the
51
52
53
54
55
56
57
58
59
60

1
2
3
4 central nervous system; had a history of seizures, hematological diseases, or severe
5
6 liver or kidney impairment; or were pregnant. No familial relationship was known
7
8 between the study participants. The study was performed according to the Guidelines
9
10 of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang, China).
11
12 Written informed consent was obtained from each participant in this study.
13
14

15 16 **SNP selection**

17
18 Tagging SNPs were selected from the catalogs of the International HapMap
19
20 Project. The rs2789679 and rs2245168 are located in 5'-UTR, rs2236558 in intron 6,
21
22 rs7067644 and rs12763624 in intron 8, rs1049550 in exon 10, rs2573351 in intron 13,
23
24 rs10887581, rs11201989, rs2573353, rs2789695, rs2573356, rs2819945, rs11202059
25
26 and rs2819941 in intron 14 (Supplemental figure 1). Marker selection was based on
27
28 the following criteria. (a) We used the CHB data from the HapMap (release 27) to
29
30 select tagSNPs for ANXA11. (b) We restricted our search for tagSNPs from base pair
31
32 80155124 to 80205677. (c) Further, we limited the SNPs to tag to those with a minor
33
34 allele frequency ≥ 0.2 in the CHB. (d) Based on these restrictions, there were a total
35
36 of 29 SNPs. (e) Using HAPLOVIEW with an r^2 threshold ≥ 0.8 , 15 tagSNP were
37
38 selected and used for subsequent analyses.

39
40
41 ~~Marker selection was done according to previous studies⁴⁻¹²⁻¹⁴⁻¹⁵, and preliminary~~
42
43 ~~analysis was performed using the HapMap data and the following criteria. First, we~~
44
45 ~~examined tagSNPs in the Haploview (v4.2), using the CHB population and a minor~~
46
47 ~~allele frequency cut-off (MAF) $\geq 5\%$ (HapMap Data Release 27). We found that~~
48
49 ~~there were a total of 29 potential tagSNPs in all. As a first screen of the most common~~
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 SNPs in the sarcoidosis sample from an Eastern Chinese Han population, a MAF $\geq 20\%$
5
6 with pair wise tagging and $r^2 \geq 0.8$ ¹⁶ was used as the cut off when choosing
7
8 tagSNPs. Second, the LD pattern of this gene was determined in the Chinese
9
10 population using the preliminary data from the HapMap. Five LD blocks across the
11
12 ANXA11 were defined using Haploview's 'confidence intervals' method^{17,18}. These
13
14 SNPs were further analyzed in an association study.
15
16
17
18

19 Genotyping

20
21 Peripheral blood was collected from a vein into a sterile tube coated with
22
23 ethylenediamine tetraacetic acid (EDTA). Plasma samples were stored at -20°C.
24
25 Genomic DNA was extracted from the frozen peripheral blood samples using a
26
27 QIAmp Blood Mini Kit (Qiagen Inc., Valencia, CA, USA) according to the
28
29 manufacturer's protocols. The selected SNPs were genotyped in cases and controls by
30
31 using the MALDI-TOF in the MassARRAY system (Sequenom Inc., San Diego,
32
33 CA, USA). Probes and primers were designed using the Assay Design Software
34
35 (Sequenom, San Diego, CA, USA). The completed genotyping reactions were spotted
36
37 onto a 384 well spectroCHIP (Sequenom) using the MassARRAY Nanodispenser
38
39 (Sequenom) and determined by the matrix - assisted laser desorption ionization time -
40
41 of - flight mass spectrometer. Genotype calling was performed in realtime with the
42
43 MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer
44
45 software version 3.4 (Sequenom).
46
47
48
49
50
51
52

53 Statistical analysis

54
55
56 All data were analyzed using the SPSS 17.0 software (SPSS Inc., Chicago, IL,
57
58
59
60

1
2
3
4 USA). Hardy-Weinberg equilibrium ~~were~~was evaluated by Chi—square tests.
5
6 Differences between the cases and controls in the frequency of the alleles, genotypes
7
8 and haplotypes were evaluated by the Fisher’s exact test or the Pearson Chi—square
9
10 test. Unconditional logistic regression was used to calculate the odds ratio (OR) and
11
12 95% confidence interval (CI) in independent association between each locus and the
13
14 presence of sarcoidosis. ~~Logistic regression~~Generalized linear regression was used to
15
16 evaluate the interaction effects between gene and gender or age. Gender and age of
17
18 subjects were treated as covariants in binary logistic regression. ~~P—~~ values ~~was~~were
19
20 calculated based on codominant, dominant for the rare allele, heterosis and recessive
21
22 for the rare allele models of inheritance. The Bonferroni correction was used to adjust
23
24 the test level when multiple comparisons were conducted, and the *P* value was
25
26 divided by the total number of loci. Haplotype blocks were defined according to the
27
28 criteria developed by Gabriel et al. ¹⁵. Pair—wise LD statistics (D' and r^2) and
29
30 haplotype frequency were calculated, and haplotype blocks were constructed using
31
32 the Haploview 4.0 ¹⁶. To ensure that the LD blocks most closely reflect the population
33
34 level LD patterns, definition of the blocks were based on the control samples alone.
35
36 The haplotype frequencies were estimated using GENECOUNTING, which computes
37
38 maximum - likelihood estimates of haplotype frequencies from unknown phase data
39
40 by utilizing an expectation - maximization algorithm ¹⁷⁻²⁰. The significance of any
41
42 haplotypic association was evaluated using a likelihood ratio test, followed by
43
44 permutation testing that compared estimated haplotype frequencies in cases and
45
46 controls ^{17 19}.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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. 21). The genotype distribution, allelic frequencies, and haplotypes in the patients with sarcoidosis and healthy controls are showed in [Tables 21, 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 in 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.](#)

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

1
2
3
4 We performed an association analysis to determine whether the haplotype was
5
6 associated with the risk of sarcoidosis (Global $P < 0.05$ in block 2 and block 3).
7
8 Compared with the ~~healthy~~ controls, the T - C haplotype occurred significantly less
9
10 frequently ($P = 0.001$) ~~and but~~ the C - C haplotypes occurred more frequently ($P =$
11
12 0.0001) in block 2 (-rs1049550 - rs2573351) in the patients with sarcoidosis.—
13
14

15
16 To assess particular disease phenotypes, the patients with sarcoidosis were
17
18 divided into the subgroups according to their chest radiographic (CXR) stage. CXR
19
20 stage I (isolated bilateral hilar lymphadenopathy): N = 196; CXR stages II - IV
21
22 (infiltration of parenchyma): 167; CXR stage 0 (absence on pulmonary disease): N =
23
24 28; the information on CXR stage was not available for 21 patients. Genotype
25
26 frequency distribution revealed that, in rs1049550, CC genotype was significantly
27
28 more in the patients with chest radiographic (CXR) stage I sarcoidosis than the
29
30 patients with CXR stage II-IV sarcoidosis ($P = 0.012$). ~~Comparison of genotype~~
31
32 ~~frequency distribution revealed significant differences between stage I and stages II-~~
33
34 ~~IV for rs1049550. The significantly more CC genotype ($P = 0.012$) were found in the~~
35
36 ~~patients with stage I sarcoidosis (table Table 34).~~
37
38
39
40
41
42
43

44 DISCUSSION

45
46 A key step in linkage and association studies is to identify common risk variants
47
48 in different populations. To determine if common risk variants exist in distinct
49
50 populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene,
51
52 were genotyped in samples from the patients with sarcoidosis and healthy controls in
53
54 a Chinese Han population. Recently, new sarcoidosis loci have been identified by
55
56
57
58
59
60

1
2
3 genome-wide association studies^{4,14}. With the fast development of genome-wide
4
5
6 association studies, an increasing number of susceptibility loci for sarcoidosis have
7
8
9 been ~~reported~~ found in different populations^{4,14,21}. However, these observations
10
11 should be confirmed in other genetically independent populations. In this study, we
12
13 conducted the first large genetic association study of the ANXA11 gene in a Chinese
14
15 Han population. The evidence of markers associated with sarcoidosis was presented,
16
17 and these markers were mapped to different locations in the ANXA11 gene
18
19 (81897864 - 81951001). The association signals in the region were identified, and
20
21 some significantly associated haplotypes also appeared this region.
22
23
24

25
26 Hoffman et al. reported an association between the T allele of the non -
27
28 synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a
29
30 German population⁴. Similar results were also obtained in other two European
31
32 populations^{14,21}. As a part of the sarcoidosis GWAS done in Americans²², we further
33
34 confirmed this association in a Chinese Han population. In this study, the frequency of
35
36 ANXA11 rs1049550T allele was significantly lower in the patients with sarcoidosis
37
38 than the healthy controls. There is interaction between rs2789679 and rs1049550.
39
40 Genotype frequency distribution revealed that, in rs1049550, CC genotype was
41
42 significantly more in the patients with stage I sarcoidosis than the patients with stage
43
44 II-IV sarcoidosis ~~Comparison of stages II-IV, the significantly more CC genotypewas~~
45
46 ~~found in the patients with stages I sarcoidosis.~~ The “protective” effect of ANXA11 T
47
48 allele increased with the number of its copies in the genotype, which is consistent
49
50 with the result obtained in the patients with sarcoidosis in a German population⁴,
51
52
53
54
55
56
57
58
59
60

1
2
3
4 Furthermore, the T allele carriers among the patients were protected from infiltration
5
6 of lung parenchyma (radiographic stages II-IV) ²¹. We have also demonstrated that
7
8 rs1049550 is significant cross - ethnically at the gene level after adjustment for the
9
10 single SNP association tests performed. The mechanism by which rs1049550 affects
11
12 the susceptibility to sarcoidosis ~~may be~~ that the SNP may affect the function of the
13
14 ANXA11 protein rather than affect its expression. In humans, the ANXA11 protein
15
16 consists of an N - terminal proline - tyrosine - glycine (PYG) - rich region, followed
17
18 by four annexin core domains. The rs1049550 leads to an amino - acid exchange
19
20 (basic arginine to a polar cysteine) at the evolutionarily conserved position 230
21
22 (R230C) in the first annexin domain, which is responsible for Ca²⁺ - dependent
23
24 trafficking of the protein in the cell ²³.

25
26
27
28
29
30
31 In the GWAS conducted by Hofmann et al ⁴, the strongest association signal was
32
33 observed for rs2789679. This marker is located in the 3' UTR of the ANXA11 gene.
34
35 The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in
36
37 the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls
38
39 being TT homozygous. Fine-mapping in the Black Women's Health Study (BWHS)
40
41 revealed rs2819941 to be the top SNP, associated with a non-significant ($P = 0.06$)
42
43 reduction in the risk of sarcoidosis ~~risk~~ (17% reduction per copy of the C-allele). In
44
45 this case-control association study, the frequency of the T allele in rs2789679 and
46
47 the rs2819941 C allele frequency in the patients with sarcoidosis was significantly
48
49 lower than that in the healthy controls. Several lines of evidence suggest that the
50
51 observed associations ~~are~~ unlikely to be an artifact. First, both the single SNP and
52
53
54
55
56
57
58
59
60

1
2
3
4 the haplotype—based association analyses support the current finding. Second, our
5 samples were from the same geographical region, excluding the population
6 stratification is an impossible reason, because all of our samples were from the same
7 geographical region. Finally, consistent results were obtained from two genetically
8
9 independent populations (Chinese Han and Europeans). Collectively, our results
10
11 confirmed the strong association between ~~variations in the~~ ANXA11 polymorphisms
12 gene and sarcoidosis, ~~and~~ suggesting that ANXA11 represents a strong genetic risk
13
14 factor for sarcoidosis.
15
16
17
18
19
20
21
22
23

24 We further investigated the interaction among polymorphisms and observed
25 strong LD. The haplotype analysis revealed that significantly more C - C (block 2,
26 rs1049550 - rs2573351) and significantly fewer T - C haplotypes (block 2) were
27
28 found in the patients with sarcoidosis. These results indicated that the patients with C
29
30 - C (block 2) haplotypes of the ANXA11 gene were more prone to sarcoidosis.
31
32 Significantly higher frequencies of T - C haplotypes were detected in the healthy
33
34 controls than in the patients with sarcoidosis, suggesting that they may show ~~a~~
35
36 protective effects against sarcoidosis. Our sample size can detect SNP and haplotype
37
38 associations with 90% and 85% power, respectively, at a false positive rate of 5%,
39
40 disease prevalence of 1%, disease allele / haplotype frequency of 0.05 / 0.03, and a
41
42 presumed odds ratio (OR) of 1.5. To some extent, this finding further supports a role
43
44 of ANXA11 polymorphisms in sarcoidosis, while ethnic group difference may exist.
45
46
47
48
49
50
51
52

53 In conclusion, our study suggests a potential role of the ANXA11 SNPs and their
54
55 related haplotypes in the genetic susceptibility to sarcoidosis. Further studies are
56
57
58
59
60

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.~~

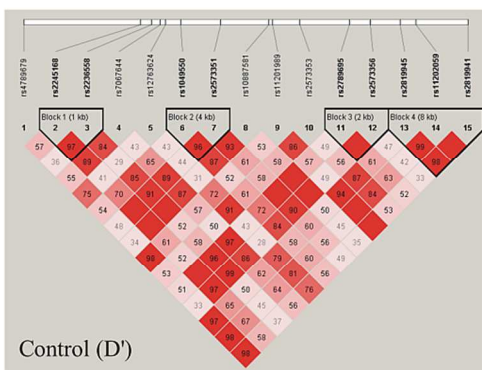
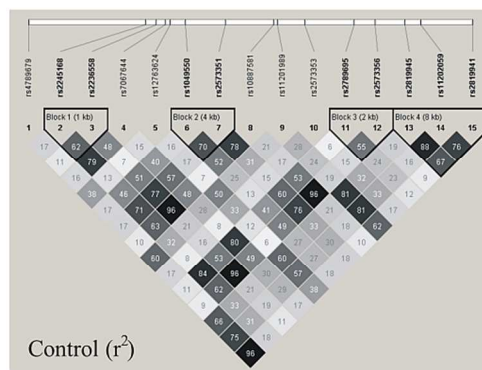
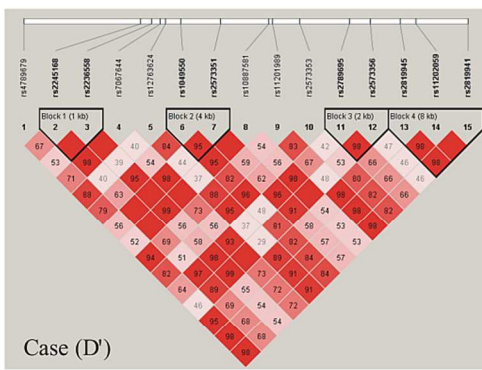
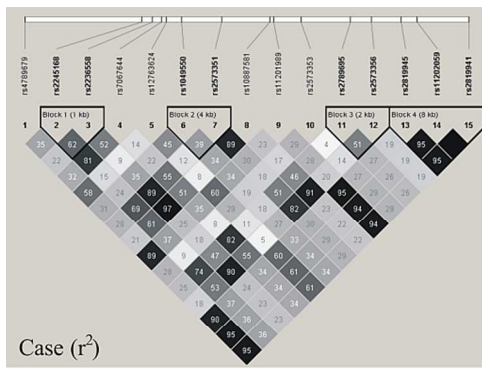
Reference

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.

- 1
2
3
4 7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and
5
6 completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-22.
7
8
9 8. Petzmann S, Maercker C, Markert E, *et al.* Enhanced proliferation and decreased
10
11 apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis Vasc Diffuse*
12
13 *Lung Dis* 2006;23:190-200
14
15
16 9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, *et al.* Comparative
17
18 analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral
19
20 blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol*
21
22 2006;26:243-50.
23
24
25
26 10. Agostini C, Trentin L, Facco M, *et al.* Role of IL-15, IL-2, and their receptors in
27
28 the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol*
29
30 1996;157:910-8
31
32
33
34 11. Jullien D, Sieling PA, Uyemura K, *et al.* IL-15, an immunomodulator of T cell
35
36 responses in intracellular infection. *J Immunol* 1997;158:800-6
37
38
39 12. Muller-Quernheim J, Schurmann M, Hofmann S, *et al.* Genetics of sarcoidosis.
40
41 *Clin Chest Med* 2008;29:391-414.
42
43
44 13. Morais A, Lima B, Peixoto M, *et al.* Annexin A11 gene polymorphism (R230C
45
46 variant) and sarcoidosis in a Portuguese population. *Tissue antigens*
47
48 2013;82:186-91.
49
50
51 14. Li Y, Pabst S, Kubisch C, *et al.* First independent replication study confirms the
52
53 strong genetic association of ANXA11 with sarcoidosis. *Thorax*
54
55 2010;65:939-40.
56
57
58
59
60

- 1
2
3
4 15. Gabriel SB, Schaffner SF, Nguyen H, *et al.* The structure of haplotype blocks in
5
6 the human genome. *Science* 2002;296:2225-9.
7
8
9 16. Barrett JC, Fry B, Maller J, *et al.* Haploview: analysis and visualization of LD and
10
11 haplotype maps. *Bioinformatics* 2005;21:263-5.
12
13
14 17. Curtis D, Knight J, Sham PC. Program report: GENECOUNTING support
15
16 programs. *Ann Hum Genet* 2006;70:277-9.
17
18
19 18. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic
20
21 associations. *Hum Hered* 2000;50:133-9.
22
23
24 19. Zhao JH, Lissarrague S, Essioux L, *et al.* GENECOUNTING: haplotype analysis
25
26 with missing genotypes. *Bioinformatics* 2002;18:1694-5
27
28
29 20. Zhao JH, Sham PC. Generic number systems and haplotype analysis. *Comput*
30
31 *Methods Programs Biomed* 2003;70:1-9
32
33
34 21. Mrazek F, Stahelova A, Kriegova E, *et al.* Functional variant ANXA11 R230C:
35
36 true marker of protection and candidate disease modifier in sarcoidosis. *Genes*
37
38 *Immun* 2011;12:490-4.
39
40
41 22. Levin AM, Iannuzzi MC, Montgomery CG, *et al.* Association of ANXA11 genetic
42
43 variation with sarcoidosis in African Americans and European Americans.
44
45 *Genes Immun* 2013;14:13-8.
46
47
48
49 24. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11
50
51 with the nuclear envelope. *J Biol Chem* 2003;278:20210-6.
52
53
54
55
56
57
58
59
60

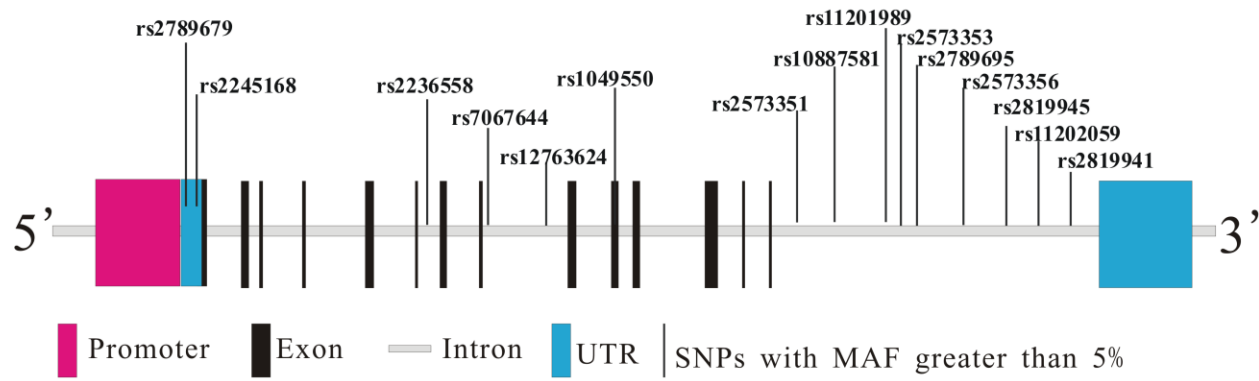
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



90x66mm (300 x 300 DPI)

Review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



For peer review only

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	

1							
2							
3							
4							
5							
6							
7	TT	120	28.7	104	25.2	0.263	1.19, 0.88-1.62
8	CT	202	48.3	216	52.4	0.238	0.85, 0.65-1.12
9	CC	96	23.0	92	22.3	0.826	1.04, 0.75-1.44
10	Per C allele	394	47.1	400	48.5	0.564	1.06, 0.87-1.28
11	rs10887581					0.290	
12	TT	114	27.3	96	23.3	0.190	1.23, 0.90-1.69
13	TC	190	45.5	208	50.5	0.148	0.82, 0.62-1.07
14	CC	114	27.3	108	26.2	0.729	1.06, 0.78-1.44
15	Per C allele	418	50.0	424	51.5	0.583	1.06, 0.87-1.29
16	rs11201989					0.144	
17	GG	130	31.1	128	31.1	0.995	1.00, 0.75-1.34
18	GC	218	52.2	194	47.1	0.141	1.23, 0.93-1.62
19	CC	70	16.8	90	21.9	0.060	0.71, 0.50-1.01
20	Per C allele	358	42.8	374	45.4	0.29	1.11, 0.91-1.35
21	rs2573353					0.918	
22	CC	178	42.6	176	42.7	0.957	0.99, 0.75-1.31
23	CA	196	46.9	192	46.6	0.921	1.01, 0.77-1.33
24	AA	44	10.5	44	10.7	0.941	0.98, 0.63-1.53
25	Per A allele	284	34.0	280	34.0	0.997	1.00, 0.82-1.23
26	rs2789695					0.918	
27	TT	178	42.6	180	34.0	0.740	0.95, 0.73-1.26
28	CT	188	45.0	184	44.7	0.920	1.01, 0.77-1.33
29	CC	52	12.4	48	11.7	0.73	1.08, 0.71-1.64
30	Per C allele	292	34.9	280	34.0	0.69	0.96, 0.78-1.17
31	rs2573356					0.892	
32	CC	118	28.2	112	27.2	0.742	1.05, 0.78-1.43
33							
34							
35							
36							
37							
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							
48							
49							

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

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	<i>P</i> 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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

only

1
2
3
4 1 **Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a**

5
6 2 **Han Chinese population - a case-control study**

7
8
9 3 LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang

10
11 4 The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang

12
13
14 5 Medical College, Xinxiang, Henan, PR China

15
16 6 *Corresponding author: Zhijian Xiao, The Third Department of Tuberculosis, the First

17
18
19 7 Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China.

20
21 8 No.88 Jiankang Road, Weihui City, Henan, 463000, PR China

22
23
24 9 E-mail: zhangligongys@163.com

25
26 10 Tel: +86 0373 - 4402229 (office)

27
28
29 11 Fax: +86 0373 -4402251

30
31 12 **Key words:** Sarcoidosis; Annexin A11; Single-nucleotide polymorphism

32
33
34 13 Word count: 3918

35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60 14

1
2
3
4 **Abstract**

5
6 **Objectives:** To further identify the single-nucleotide polymorphisms (SNPs) that
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
3 contribute to the genetic susceptibility to sarcoidosis, we examined the potential
4 association between sarcoidosis and fifteen SNPs of the ANXA11 gene.

5 **Design:** A case - control study.

6 **Setting:** A tuberculosis unit in a hospital of the university in China.

7 **Participants:** Participants included 412 patients with sarcoidosis and 418 healthy
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
controls.

9 **Methods:** The selected SNPs were genotyped using the MALDI-TOF in the
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
MassARRAY system. Probes and primers were designed using the Assay Design
Software (Sequenom, San Diego, CA, USA).

12 **Results:** The statistical significant differences were found in the allelic or genotypic
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
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

1
2
3
4 1 than the controls. Furthermore, genotype frequency distribution revealed that, in
5
6 2 rs1049550, CC genotype was significantly more in the patients with chest
7
8 3 radiographic (CXR) stage I sarcoidosis than the patients with CXR stage II-IV
9
10 4 sarcoidosis ($P = 0.012$).

11
12
13 5 **Conclusions:** These findings point to a role for the polymorphisms of ANXA11 in
14
15 6 sarcoidosis in a Chinese Han population, and may be informative for future genetic
16
17 7 studies on sarcoidosis.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 **1 Article summary**

5
6 **2 Article focus**

7
8
9 3 1. The first genome-wide association study (GWAS) in sarcoidosis conducted in a
10
11 4 German population has revealed that the rs1049550 (C > T, Arg230Cys) located in
12
13 5 exon 10 were associated with sarcoidosis.

14
15
16 6 2. Whether these SNPs modulate the risk of sarcoidosis by themselves or whether
17
18 7 they correlate with other causative SNPs and are repeated in other populations remain
19
20 8 unclear.

21
22 9 3. We hypothesize that common variants in the ANXA11 gene may significantly
23
24 10 contribute to the predisposition to develop sarcoidosis in a Chinese Han population.

25
26
27
28
29 **11 Key messages**

30
31 12 1. ANXA11 rs2789679 (3'-UTR) was significantly associated with sarcoidosis.

32
33 13 2. ANXA11 rs1049550 (exon 10) was significantly associated with sarcoidosis.

34
35 14 3. Another significant association was observed for rs2819941 (intron 14).

36
37
38
39 **15 Strengths and limitations of this study**

40
41 16 1. A systematical screening of the functional SNPs in the promoter region, 5'- and
42
43 17 3'-UTR, exons of the ANXA11 gene, and the homogeneity of the study subjects
44
45 18 representing the Chinese Han are the main strengths of this study. Chinese Han

46
47
48 19 2. The lack of data proving the positive association observed for rs2789679 and
49
50 20 rs1049550 is a potential limitation of this study. Furthermore, the association of the
51
52 21 serum level of ANXA11 with sarcoidosis still needs to be investigated.

53
54
55
56
57
58
59
60 22

1 INTRODUCTION

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

13 The Annexin gene family is involved in the etiology of several autoimmune and
14 chronic diseases^{4 5}. One member of the Annexin gene family, ANXA11, located on
15 chromosome 10q22.3, is involved in calcium signaling, apoptosis, vesicle trafficking,
16 cell growth and terminal phase of cell division⁵⁻⁷. In this context, the development
17 and maintenance of the granulomatous inflammation in sarcoidosis have been
18 repeatedly associated with the impaired apoptosis of activated inflammatory cells^{8 9}.
19 Most of the early-stage sarcoidosis is characterized by mononuclear cell alveolitis
20 dominated by activated CD4⁺ T cells and macrophages. Between these cells, the
21 uncoordinated interplay results in the formation of typical non-caseating granulomas.
22 The mechanism by which the granulomas resolve has not been fully elucidated.

1
2
3
4 1 However, it is generally assumed that the induction of apoptosis and/or by the
5
6 2 withdrawal of inflammatory cytokines participates in the disappearance of
7
8 3 granulomas^{10 11}. In the patients with sarcoidosis, their peripheral blood monocytes
9
10 4 were characterized by apoptosis⁹.

11
12
13
14 5 In previous studies, case - control/family ‘hypothesis - driven’ studies and low
15
16 6 density linkage scans were used to identify genetic factors conferring the genetic
17
18 7 susceptibility to sarcoidosis¹². Notably, the first genome - wide association study
19
20 8 (GWAS) in sarcoidosis conducted in a German population has recently revealed an
21
22 9 association: the leading rs2789679 SNP (single - nucleotide polymorphism) located in
23
24 10 the 3’-untranslated region (3’-UTR) and a common nonsynonymous SNP (rs1049550,
25
26 11 C > T, Arg230Cys) were associated with the increased risk of sarcoidosis^{4 13}. This
27
28 12 association has recently been supported by another report from the same population¹⁴.
29
30 13 Thus, more studies should be performed to demonstrate the following items: whether
31
32 14 these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate
33
34 15 with other causative SNPs and are repeated in other populations. The published
35
36 16 studies about the association of sarcoidosis and ANXA11 are summarized in
37
38 17 Supplemental table 1. We hypothesize that common variants in the ANXA11 gene
39
40 18 may significantly contribute to the predisposition to develop sarcoidosis.
41
42
43
44
45
46
47
48

49 19 In this study, we investigated fifteen loci in a Chinese population from He’nan
50
51 20 province (China) to verify the putative association between ANXA11 polymorphisms
52
53 21 and sarcoidosis.
54
55

56 22 **SUBJECTS AND METHODS**

1 **Subjects**

2 Four hundred and twelve patients with sarcoidosis (mean \pm SD age: 53.6 ± 4.6
3 years; 155 men and 257 women) were recruited from our hospital between May 2005
4 and March 2013. Sarcoidosis was diagnosed by the evidence of non-caseating
5 epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities.
6 Chronic sarcoidosis was defined as a disease over at least 2 years or at least two
7 episodes in a lifetime. A cute sarcoidosis was defined as one episode of acute
8 sarcoidosis which had totally resolved at the date of the examination. The control
9 group consisted of 418 unrelated healthy subjects (mean \pm SD age: 54.2 ± 5.3 years;
10 160 men and 258 women) who underwent health examinations in the Medical
11 Examination Center of the First Affiliated Hospital of the Xinxiang Medical College
12 (Xinxiang, China) between Oct 2009 and Sept 2013. None of the individuals in the
13 control group had a history of lung diseases or showed any symptoms of the lung or
14 other diseases by chest radiography or laboratory blood tests. Participants were
15 excluded if they: were taking other prescribed medications that could affect the
16 central nervous system; had a history of seizures, hematological diseases, or severe
17 liver or kidney impairment; or were pregnant. No familial relationship was known
18 between the study participants. The study was performed according to the Guidelines
19 of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang, China).
20 Written informed consent was obtained from each participant in this study.

21 **SNP selection**

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

12 Genotyping

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

1 of - flight mass spectrometer. Genotype calling was performed in realtime with the
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 of - flight mass spectrometer. Genotype calling was performed in realtime with the
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

4 **Statistical analysis**

5 All data were analyzed using the SPSS 17.0 software (SPSS Inc., Chicago, IL,
6 USA). Hardy-Weinberg equilibrium was evaluated by Chi-square tests. Differences
7 between the cases and controls in the frequency of the alleles, genotypes and
8 haplotypes were evaluated by the Fisher's exact test or the Pearson Chi-square test.
9 Unconditional logistic regression was used to calculate the odds ratio (OR) and 95%
10 confidence interval (CI) in independent association between each locus and the
11 presence of sarcoidosis. Gender and age of subjects were treated as covariants in
12 binary logistic regression. *P* values were calculated based on codominant, dominant
13 for the rare allele, heterosis and recessive for the rare allele models of inheritance.
14 Models of multiple logistic regression were used to test the independence of
15 individual allelic effect. In detail, the most significant SNP was chosen to be the
16 conditional SNP (covariate in the regression model) when testing other significant
17 SNPs; Meanwhile, a multi-SNP model including all significant SNPs was also
18 performed. The Bonferroni correction was used to adjust the test level when multiple
19 comparisons were conducted, and the *P* value was divided by the total number of loci.
20 Haplotype blocks were defined according to the criteria developed by Gabriel et al.¹⁵.
21 Pair-wise LD statistics (D' and r^2) and haplotype frequency were calculated, and
22 haplotype blocks were constructed using the Haploview 4.0¹⁶. To ensure that the LD

1 blocks most closely reflect the population level LD patterns, definition of the blocks
2 were based on the control samples alone. The haplotype frequencies were estimated
3 using GENECOUNTING, which computes maximum - likelihood estimates of
4 haplotype frequencies from unknown phase data by utilizing an expectation -
5 maximization algorithm¹⁷⁻²⁰. The significance of any haplotypic association was
6 evaluated using a likelihood ratio test, followed by permutation testing that compared
7 estimated haplotype frequencies in cases and controls^{17 19}.

8 RESULTS

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

17 Comparison of genotype and allele frequency distribution revealed significant
18 differences between the patients with sarcoidosis and healthy controls for 3 SNPs:
19 rs2789679, rs1049550 and rs2819941. The rs2789679 A allele ($P = 0.00004$, OR =
20 1.42, 95%CI = 1.17-1.73) and rs2819941 T allele ($P = 0.0006$, OR = 1.41, 95%CI =
21 1.16-1.71) were significantly more frequent in the patients with sarcoidosis compared
22 to controls. The frequency of the rs1049550 T allele ($P = 0.000002$, OR = 0.61,

1
2
3
4 1 95%CI= 0.49-0.74) in the patients with sarcoidosis was significantly lower than that
5
6 2 in the controls. These differences remained statistically significant after Bonferroni
7
8 3 corrections.
9

10
11 4 The multi-SNP model showed only rs1049550 present a significant effect on the
12
13 5 disease phenotype ($p < 0.001$). No independent effect was found for rs2789679 or
14
15 6 rs2819941 when adjusting for the effect of rs1049550. Both individual SNP and multi -
16
17 7 SNP analysis supported rs1049550: T allele was an important protective factor for
18
19 8 affecting sarcoidosis (odds ratio around 0.6). In another word, carriers of rs1049550:
20
21 9 C allele have higher susceptibility to sarcoidosis in our Chinese Han population
22
23 10 (Supplemental table 3).
24
25
26
27

28
29 11 We performed an association analysis to determine whether the haplotype was
30
31 12 associated with the risk of sarcoidosis (Global $P < 0.05$ in block 2 and block 3).
32
33 13 Compared with the controls, the T - C haplotype occurred significantly less frequently
34
35 14 ($P = 0.001$) but the C - C haplotypes occurred more frequently ($P = 0.0001$) in block 2
36
37 15 (rs1049550 - rs2573351) in the patients with sarcoidosis.
38
39
40

41
42 16 To assess particular disease phenotypes, the patients with sarcoidosis were
43
44 17 divided into the subgroups according to their chest radiographic (CXR) stage. CXR
45
46 18 stage I (isolated bilateral hilar lymphadenopathy): $N = 196$; CXR stages II - IV
47
48 19 (infiltration of parenchyma): 167; CXR stage 0 (absence on pulmonary disease): $N =$
49
50 20 28; the information on CXR stage was not available for 21 patients. Genotype
51
52 21 frequency distribution revealed that, in rs1049550, CC genotype was significantly
53
54 22 more in the patients with chest radiographic (CXR) stage I sarcoidosis than the
55
56
57
58
59
60

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

2 **DISCUSSION**

3 A key step in linkage and association studies is to identify common risk variants
4 in different populations. To determine if common risk variants exist in distinct
5 populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene,
6 were genotyped in samples from the patients with sarcoidosis and healthy controls in
7 a Chinese Han population. Recently, new sarcoidosis loci have been identified by
8 genome-wide association studies^{4 14}. With the fast development of genome - wide
9 association studies, an increasing number of susceptibility loci for sarcoidosis have
10 been found in different populations^{4 14 21}. However, these observations should be
11 confirmed in other genetically independent populations. In this study, we conducted
12 the first large genetic association study of the ANXA11 gene in a Chinese Han
13 population. The evidence of markers associated with sarcoidosis was presented, and
14 these markers were mapped to different locations in the ANXA11 gene (81897864 -
15 81951001). The association signals in the region were identified, and some
16 significantly associated haplotypes also appeared this region.

17 Hoffman et al. reported an association between the T allele of the non -
18 synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a
19 German population⁴. Similar results were also obtained in other two European
20 populations^{14 21}. As a part of the sarcoidosis GWAS done in Americans²², we further
21 confirmed this association in a Chinese Han population. In this study, the frequency of
22 ANXA11 rs1049550 T allele was significantly lower in the patients with sarcoidosis

1
2
3
4 1 than the healthy controls. Genotype frequency distribution revealed that, in rs1049550,
5
6 2 CC genotype was significantly more in the patients with stage I sarcoidosis than the
7
8 3 patients with stage II-IV sarcoidosis. The “protective” effect of ANXA11 T allele
9
10 4 increased with the number of its copies in the genotype, which is consistent with the
11
12 5 result obtained in the patients with sarcoidosis in a German population ⁴, Furthermore,
13
14 6 the T allele carriers among the patients were protected from infiltration of lung
15
16 7 parenchyma (radiographic stages II-IV) ²¹. We have also demonstrated that rs1049550
17
18 8 is significant cross - ethnically at the gene level after adjustment for the single SNP
19
20 9 association tests performed. The mechanism by which rs1049550 affects the
21
22 10 susceptibility to sarcoidosis is that the SNP may affect the function of the ANXA11
23
24 11 protein rather than affect its expression. In humans, the ANXA11 protein consists of
25
26 12 an N - terminalproline - tyrosine - glycine (PYG) - rich region, followed by four
27
28 13 annexin core domains. The rs1049550 leads to an amino - acid exchange (basic
29
30 14 arginine to a polar cysteine) at the evolutionarily conserved position 230 (R230C) in
31
32 15 the first annexin domain, which is responsible for Ca²⁺ - dependent trafficking of the
33
34 16 protein in the cell ²³.

35
36
37
38
39
40
41
42
43 17 In the GWAS conducted by Hofmann et al ⁴, the strongest association signal was
44
45 18 observed for rs2789679. This marker is located in the 3'-UTR of the ANXA11 gene.
46
47 19 The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in
48
49 20 the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls
50
51 21 being TT homozygous. Fine-mapping in the Black Women's Health Study (BWHS)
52
53 22 revealed rs2819941 to be the top SNP, associated with a non-significant ($P = 0.06$)
54
55
56
57
58
59
60

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

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

1
2
3
4 1 protective effects against sarcoidosis. Our sample size can detect SNP and haplotype
5
6 2 associations with 90% and 85% power, respectively, at a false positive rate of 5%,
7
8
9 3 disease prevalence of 1%, disease allele / haplotype frequency of 0.05 / 0.03, and a
10
11 4 presumed odds ratio (OR) of 1.5. To some extent, this finding further supports a role
12
13
14 5 of ANXA11 polymorphisms in sarcoidosis, while ethnic group difference may exist.

16 In conclusion, our study suggests a potential role of the ANXA11 SNPs and their
17
18
19 7 related haplotypes in the genetic susceptibility to sarcoidosis. Further studies are
20
21 8 needed to investigate how these SNPs affect the function of ANXA11.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 **Funding**

5
6 The project was supported by the Research and Development Foundation of
7
8
9 Science and Technology of Henan Province (2012k15).

10
11 **Contributors**

12
13 LG Zhang was involved in conception and design of the study, acquisition of
14
15 patient data, genotyping and drafted the paper. ZJ Xiao and GC Shi were involved in
16
17 design of the study, genotyping and interpretation of results. J Huang and CX Zhang
18
19 were involved in the design of the study, acquisition of patient data and interpretation
20
21 of result. All authors revised the draft paper.
22
23
24
25

26 **Competing interests**

27
28
29 None.

30
31 **Data sharing statement**

32
33 No additional data are available.
34
35

36 **Patient consent**

37
38
39 Obtained.
40

41 **Ethics approval**

42
43
44 Ethics approval was provided by the First Hospital Affiliated to the Xinxiang
45
46 Medical College.
47
48

49 **Provenance and peer review**

50
51 Not commissioned; externally peer reviewed.
52
53
54
55
56
57
58
59
60

1
2
3
4 **Figure Legends**
5

6 **Fig. 1** The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the
7
8 pair-wise calculation of r^2 (left) or D' (right). Black squares indicate $r^2 = 1$ (i.e. perfect
9
10 LD between a pair of SNPs). Empty squares indicate $D' = 1$ (i.e. complete LD
11
12 between a pair of SNPs).
13
14
15

16 **Supplemental Fig. 1** The structure of the human ANXA11 gene and 15 SNPs located
17
18 on the gene.
19
20

21 8

22 9

23 10

24 11

25 12

26 13

27 14

28 15

29 16

30 17

31 18

32 19

33 20

34 21

35 22

1
2
3
4 **Reference**
5

- 6 1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007;
7
8 357:2153-65.
9
10 2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-8
11
12 3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis.
13
14 *Semin Respir Crit Care Med* 2010;31:380-9.
15
16 4. Hofmann S, Franke A, Fischer A, *et al.* Genome-wide association study identifies
17
18 ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet*
19
20 2008;40:1103-6.
21
22 5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev*
23
24 2002;82:331-71.
25
26 6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using
27
28 recombinant human annexin XI-specific autoantibodies cloned by phage
29
30 display. *J Biol Chem* 2003;278:33120-6.
31
32 7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and
33
34 completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-22.
35
36 8. Petzmann S, Maercker C, Markert E, *et al.* Enhanced proliferation and decreased
37
38 apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis Vasc Diffuse*
39
40 *Lung Dis* 2006;23:190-200
41
42 9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, *et al.* Comparative
43
44 analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral
45
46 blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol*
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4 1 2006;26:243-50.
5
6 2 10. Agostini C, Trentin L, Facco M, *et al.* Role of IL-15, IL-2, and their receptors in
7
8 3 the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol*
9
10 4 1996;157:910-8
11
12 5 11. Jullien D, Sieling PA, Uyemura K, *et al.* IL-15, an immunomodulator of T cell
13
14 6 responses in intracellular infection. *J Immunol* 1997;158:800-6
15
16 7 12. Muller-Quernheim J, Schurmann M, *et al.* Genetics of sarcoidosis. *Clin Chest*
17
18 8 *Med* 2008;29:391-414.
19
20 9 13. Morais A, Lima B, Peixoto M, *et al.* Annexin A11 gene polymorphism (R230C
21
22 10 variant) and sarcoidosis in a Portuguese population. *Tissue antigens*
23
24 11 2013;82:186-91.
25
26 12 14. Li Y, Pabst S, Kubisch C, *et al.* First independent replication study confirms the
27
28 13 strong genetic association of ANXA11 with sarcoidosis. *Thorax*
29
30 14 2010;65:939-40.
31
32 15 15. Gabriel SB, Schaffner SF, Nguyen H, *et al.* The structure of haplotype blocks in
33
34 16 the human genome. *Science* 2002;296:2225-9.
35
36 17 16. Barrett JC, Fry B, Maller J, *et al.* Haploview: analysis and visualization of LD and
37
38 18 haplotype maps. *Bioinformatics* 2005;21:263-5.
39
40 19 17. Curtis D, Knight J, Sham PC. Program report: GENECOUNTING support
41
42 20 programs. *Ann Hum Genet* 2006;70:277-9.
43
44 21 18. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic
45
46 22 associations. *Hum Hered* 2000;50:133-9.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4 19. Zhao JH, Lissarrague S, Essioux L, et al. GENECOUNTING: haplotype analysis
5
6 with missing genotypes. *Bioinformatics* 2002;18:1694-5
7
8
9 20. Zhao JH, Sham PC. Generic number systems and haplotype analysis. *Comput*
10
11 *Methods Programs Biomed* 2003;70:1-9
12
13
14 21. Mrazek F, Stahelova A, Kriegova E, et al. Functional variant ANXA11 R230C:
15
16 true marker of protection and candidate disease modifier in sarcoidosis. *Genes*
17
18 *Immun* 2011;12:490-4.
19
20
21 22. Levin AM, Iannuzzi MC, Montgomery CG, et al. Association of ANXA11 genetic
22
23 variation with sarcoidosis in African Americans and European Americans.
24
25 *Genes Immun* 2013;14:13-8.
26
27
28
29 24. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11
30
31 with the nuclear envelope. *J Biol Chem* 2003;278:20210-6.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 **Table 1** Genotypic and allelic frequencies of ANXA11 polymorphisms in the controls
 2 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

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

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

1

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

Haplotype			Genecounting (frequency %)			
ID	rs1049550	rs2573351	Cases	Controls	<i>P</i> -value ^a	Global <i>P</i> ^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	

4 ^a Based on 10,000 permutations.

5 ^b Based on comparison of frequency distribution of all haplotypes for the combination of SNPs.

6

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

Haplotype			Genecounting (frequency %)			
ID	rs2789695	rs2573356	Cases	Controls	<i>P</i> -value ^a	Global <i>P</i> ^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	

9 ^a Based on 10,000 permutations.

10 ^b Based on comparison of frequency distribution of all haplotypes for the combination of SNPs.

11

12

13

14

15

1 **Table 4** Chest radiographic (CXR) stages of sarcoidosis patients by genotype of
 2 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		

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

4

1
2
3
4 1 **Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a**

5
6 2 **Han Chinese population - a case-control study**

7
8
9 3 LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang

10
11 4 The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang

12
13
14 5 Medical College, Xinxiang, Henan, PR China

15
16 6 *Corresponding author: Zhijian Xiao, The Third Department of Tuberculosis, the First

17
18
19 7 Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China.

20
21 8 No.88 Jiankang Road, Weihui City, Henan, 463000, PR China

22
23
24 9 E-mail: zhangligongys@163.com

25
26 10 Tel: +86 0373 - 4402229 (office)

27
28
29 11 Fax: +86 0373 -4402251

30
31 12 **Key words:** Sarcoidosis; Annexin A11; Single-nucleotide polymorphism

32
33
34 13 Word count: 37873918

35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60 14

1
2
3
4 **Abstract**

5
6 **Objectives:** To further identify the single-nucleotide polymorphisms (SNPs) that
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
3 contribute to the genetic susceptibility to sarcoidosis, we examined the potential
4 association between sarcoidosis and fifteen SNPs of the ANXA11 gene.

5 **Design:** A case - control study.

6 **Setting:** A tuberculosis unit in a hospital of the university in China.

7 **Participants:** Participants included 412 patients with sarcoidosis and 418 healthy
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
8 controls.

9 **Methods:** The selected SNPs were genotyped using the MALDI-TOF in the
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
MassARRAY system. Probes and primers were designed using the Assay Design
Software (Sequenom, San Diego, CA, USA).

12 **Results:** The statistical significant differences were found in the allelic or genotypic
13
14
15
16
17
18
19
20
21
22
13 frequencies of the rs2789679, rs1049550 and rs2819941 in the ANXA11 gene
14 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

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

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

11

1
2
3
4 **1 Article summary**

5
6 **2 Article focus**

7
8
9 3 1. The first genome-wide association study (GWAS) in sarcoidosis conducted in a
10
11 4 German population has revealed that the rs1049550 (C > T, Arg230Cys) located in
12
13 5 exon 10 were associated with sarcoidosis.

14
15
16 6 2. Whether these SNPs modulate the risk of sarcoidosis by themselves or whether
17
18 7 they correlate with other causative SNPs and are repeated in other populations remain
19
20 8 unclear.

21
22 9 3. We hypothesize that common variants in the ANXA11 gene may significantly
23
24 10 contribute to the predisposition to develop sarcoidosis in a Chinese Han population.

25
26
27
28
29 **11 Key messages**

30
31 12 1. ANXA11 rs2789679 (3'-UTR) was significantly associated with sarcoidosis.

32
33 13 2. ANXA11 rs1049550 (exon 10) was significantly associated with sarcoidosis.

34
35 14 3. Another significant association was observed for rs2819941 (intron 14).

36
37
38
39 **15 Strengths and limitations of this study**

40
41 16 1. A systematical screening of the functional SNPs in the promoter region, 5'- and
42
43 17 3'-UTR, exons of the ANXA11 gene, and the homogeneity of the study subjects
44
45 18 representing the Chinese Han are the main strengths of this study. Chinese Han

46
47
48
49 19 2. The lack of data proving the positive association observed for rs2789679 and
50
51 20 rs1049550 is a potential limitation of this study. Furthermore, the association of the
52
53 21 serum level of ANXA11 with sarcoidosis still needs to be investigated.

54
55
56 **22 INTRODUCTION**

1 Sarcoidosis is a systemic autoimmune disease characterized by destructive,
2 non-caseating epithelioid granulomatous lesions, with accumulated phagocytes and
3 activated CD4⁺ T helper type 1 lymphocytes^{1 2}. The typical manifestations of
4 sarcoidosis include bilateral hilar lymphadenopathy, pulmonary infiltration and ocular
5 and skin lesions. The clinical course and prognosis of sarcoidosis are variable. The
6 acute forms of sarcoidosis such as Löfgren's syndrome (LS) can resolve
7 spontaneously in most cases. However, chronic pulmonary sarcoidosis may progress
8 to lung fibrosis which eventually causes respiratory failure. Recent studies have
9 demonstrated that sarcoidosis can occur in a pattern of family clustering, and its racial
10 incidence is also different, suggesting that some genetic factors may contribute to the
11 risk and severity of sarcoidosis¹³.

12 The Annexin gene family is involved in the etiology of several autoimmune and
13 chronic diseases^{4 5}. One member of the Annexin gene family, ANXA11, located on
14 chromosome 10q22.3, is involved in calcium signaling, apoptosis, vesicle trafficking,
15 cell growth and terminal phase of cell division⁵⁻⁷. In this context, the development
16 and maintenance of the granulomatous inflammation in sarcoidosis have been
17 repeatedly associated with the impaired apoptosis of activated inflammatory cells^{8 9}.
18 Most of the early-stage sarcoidosis is characterized by mononuclear cell alveolitis
19 dominated by activated CD4⁺ T cells and macrophages. Between these cells, the
20 uncoordinated interplay results in the formation of typical non-caseating granulomas.
21 The mechanism by which the granulomas resolve has not been fully elucidated.
22 However, it is generally assumed that the induction of apoptosis and/or by the

1
2
3
4 1 withdrawal of inflammatory cytokines participates in the disappearance of
5
6 2 granulomas^{10 11}. In the patients with sarcoidosis, their peripheral blood monocytes
7
8
9 3 were characterized by apoptosis⁹.

10
11 4 In previous studies, case - control/family ‘hypothesis - driven’ studies and low
12
13 5 density linkage scans were used to identify genetic factors conferring the genetic
14
15 6 susceptibility to sarcoidosis¹². Notably, the first genome - wide association study
16
17 7 (GWAS) in sarcoidosis conducted in a German population has recently revealed an
18
19 8 association: the leading rs2789679 SNP (single - nucleotide polymorphism) located in
20
21 9 the 3’-untranslated region (3’-UTR) and a common nonsynonymous SNP (rs1049550,
22
23 10 C > T, Arg230Cys) were associated with the increased risk of sarcoidosis^{4 13}. This
24
25 11 association has recently been supported by another report from the same population¹⁴.
26
27 12 Thus, more studies should be performed to demonstrate the following items: whether
28
29 13 these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate
30
31 14 with other causative SNPs and are repeated in other populations. The published
32
33 15 studies about the association of sarcoidosis and ANXA11 are summarized in
34
35 16 Supplemental table 1. We hypothesize that common variants in the ANXA11 gene
36
37 17 may significantly contribute to the predisposition to develop sarcoidosis.
38
39
40
41
42
43
44
45

46 18 In this study, we investigated fifteen loci in a Chinese population from He’nan
47
48 19 province (China) to verify the putative association between ANXA11 polymorphisms
49
50 20 and sarcoidosis.
51

52 21 **SUBJECTS AND METHODS**

53 22 **Subjects**

1
2
3
4 1 Four hundred and twelve patients with sarcoidosis (mean \pm SD age: 53.6 ± 4.6
5
6 2 years; 155 men and 257 women) were recruited from our hospital between May 2005
7
8
9 3 and March 2013. Sarcoidosis was diagnosed by the evidence of non-caseating
10
11 4 epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities.
12
13
14 5 Chronic sarcoidosis was defined as a disease over at least 2 years or at least two
15
16 6 episodes in a lifetime. A acute sarcoidosis was defined as one episode of acute
17
18
19 7 sarcoidosis which had totally resolved at the date of the examination. The control
20
21 8 group consisted of 418 unrelated healthy subjects (mean \pm SD age: 54.2 ± 5.3 years;
22
23
24 9 160 men and 258 women) who underwent health examinations in the Medical
25
26
27 10 Examination Center of the First Affiliated Hospital of the Xinxiang Medical College
28
29 11 (Xinxiang, China) between Oct 2009 and Sept 2013. None of the individuals in the
30
31 12 control group had a history of lung diseases or showed any symptoms of the lung or
32
33
34 13 other diseases by chest radiography or laboratory blood tests. Participants were
35
36
37 14 excluded if they: were taking other prescribed medications that could affect the
38
39 15 central nervous system; had a history of seizures, hematological diseases, or severe
40
41 16 liver or kidney impairment; or were pregnant. No familial relationship was known
42
43
44 17 between the study participants. The study was performed according to the Guidelines
45
46 18 of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang, China).
47
48
49 19 Written informed consent was obtained from each participant in this study.

50 51 **SNP selection**

52
53
54 21 Tagging SNPs were selected from the catalogs of the International HapMap
55
56 22 Project. The rs2789679 and rs2245168 are located in 5'-UTR, rs2236558 in intron 6,
57
58
59
60

1
2
3
4 1 rs7067644 and rs12763624 in intron 8, rs1049550 in exon 10, rs2573351 in intron 13,
5
6 2 rs10887581, rs11201989, rs2573353, rs2789695, rs2573356, rs2819945, rs11202059
7
8
9 3 and rs2819941 in intron 14 (Supplemental figure 1). Marker selection was based on
10
11 4 the following criteria. (a) We used the CHB data from the HapMap (release 27) to
12
13 5 select tagSNPs for ANXA11. (b) We restricted our search for tagSNPs from base pair
14
15 6 80155124 to 80205677. (c) Further, we limited the SNPs to tag to those with a minor
16
17 7 allele frequency ≥ 0.2 in the CHB. (d) Based on these restrictions, there were a total
18
19 8 of 29 SNPs. (e) Using HAPLOVIEW with an r^2 threshold ≥ 0.8 , 15 tagSNP were
20
21 9 selected and used for subsequent analyses.
22
23
24
25

26 **Genotyping**

27
28
29 11 Peripheral blood was collected from a vein into a sterile tube coated with
30
31 12 ethylenediamine tetraacetic acid (EDTA). Plasma samples were stored at -20°C .
32
33 13 Genomic DNA was extracted from the frozen peripheral blood samples using a
34
35 14 QIAmp Blood Mini Kit (Qiagen Inc., Valencia, CA, USA) according to the
36
37 15 manufacturer's protocols. The selected SNPs were genotyped in cases and controls by
38
39 16 using the MALDI-TOF in the MassARRAY system (Sequenom Inc., San Diego, CA,
40
41 17 USA). Probes and primers were designed using the Assay Design Software
42
43 18 (Sequenom, San Diego, CA, USA). The completed genotyping reactions were spotted
44
45 19 onto a 384 well spectroCHIP (Sequenom) using the MassARRAY Nanodispenser
46
47 20 (Sequenom) and determined by the matrix - assisted laser desorption ionization time -
48
49 21 of - flight mass spectrometer. Genotype calling was performed in realtime with the
50
51 22 MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer
52
53
54
55
56
57
58
59
60

1 software version 3.4 (Sequenom).

2 **Statistical analysis**

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

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

7 RESULTS

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

16 Comparison of genotype and allele frequency distribution revealed significant
17 differences between the patients with sarcoidosis and healthy controls for 3 SNPs:
18 rs2789679, rs1049550 and rs2819941. The rs2789679 A allele ($P = 0.00004$, OR =
19 1.42, 95%CI = 1.17-1.73) and rs2819941 T allele ($P = 0.0006$, OR = 1.41, 95%CI =
20 1.16-1.71) were significantly more frequent in the patients with sarcoidosis compared
21 to controls. The frequency of the rs1049550 T allele ($P = 0.000002$, OR = 0.61,
22 95%CI= 0.49-0.74) in the patients with sarcoidosis was significantly lower than that

1 ~~in the controls. The frequency of the T allele in rs2789679 ($P = 0.0003$, OR = 0.70,~~
2 ~~95%CI = 0.58–0.85) and rs1049550 ($P = 0.0007$, OR = 0.62, 95%CI = 0.50–0.76),~~
3 ~~and C allele in rs2819941 ($P = 0.001$, OR = 0.71, 95%CI = 0.59–0.87) in the patients~~
4 ~~with sarcoidosis was significantly lower than that in the controls, and these~~
5 differences remained statistically significant after Bonferroni corrections.

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

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

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

22 To assess particular disease phenotypes, the patients with sarcoidosis were

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

8 DISCUSSION

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

1 Hoffman et al. reported an association between the T allele of the non -
2 synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a
3 German population ⁴. Similar results were also obtained in other two European
4 populations ^{14 21}. As a part of the sarcoidosis GWAS done in Americans ²², we further
5 confirmed this association in a Chinese Han population. In this study, the frequency of
6 ANXA11 rs1049550 T allele was significantly lower in the patients with sarcoidosis
7 than the healthy controls. ~~There is interaction between rs2789679 and rs1049550.~~
8 Genotype frequency distribution revealed that, in rs1049550, CC genotype was
9 significantly more in the patients with stage I sarcoidosis than the patients with stage
10 II-IV sarcoidosis. The “protective” effect of ANXA11 T allele increased with the
11 number of its copies in the genotype, which is consistent with the result obtained in
12 the patients with sarcoidosis in a German population ⁴. Furthermore, the T allele
13 carriers among the patients were protected from infiltration of lung parenchyma
14 (radiographic stages II-IV) ²¹. We have also demonstrated that rs1049550 is
15 significant cross - ethnically at the gene level after adjustment for the single SNP
16 association tests performed. The mechanism by which rs1049550 affects the
17 susceptibility to sarcoidosis is that the SNP may affect the function of the ANXA11
18 protein rather than affect its expression. In humans, the ANXA11 protein consists of
19 an N - terminalproline - tyrosine - glycine (PYG) - rich region, followed by four
20 annexin core domains. The rs1049550 leads to an amino - acid exchange (basic
21 arginine to a polar cysteine) at the evolutionarily conserved position 230 (R230C) in
22 the first annexin domain, which is responsible for Ca²⁺ - dependent trafficking of the

1 protein in the cell ²³.

2 In the GWAS conducted by Hofmann et al ⁴, the strongest association signal was
3 observed for rs2789679. This marker is located in the 3'-UTR of the ANXA11 gene.

4 The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in
5 the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls
6 being TT homozygous. Fine-mapping in the Black Women's Health Study (BWHS)
7 revealed rs2819941 to be the top SNP, associated with a non-significant ($P = 0.06$)

8 reduction in the risk of sarcoidosis (17% reduction per copy of the C-allele). In this
9 case-control association study, the frequency of the ~~T~~A allele in rs2789679 and the
10 rs2819941 ~~C~~T allele frequency in the patients with sarcoidosis was significantly
11 ~~lower~~higher than that in the healthy controls. Several lines of evidence suggest that

12 the observed association is unlikely to be an artifact. First, both the single-SNP and
13 the haplotype-based association analyses support the current finding. Second, our
14 samples were from the same geographical region, excluding the. Finally, consistent
15 results were obtained from two genetically independent populations (Chinese Han and

16 Europeans). Collectively, our results confirmed the strong association between
17 ANXA11 polymorphisms and sarcoidosis, suggesting that ANXA11 represents a

18 strong genetic risk factor for sarcoidosis. Since causal SNP in ANXA11 was not
19 known, all the SNPs significantly associated with sarcoidosis were actually indirect
20 association through the real causal one. Though rs2789679 and rs2819941 were not in
21 LD with rs1049550, it may not tag the causal SNP as well as rs1049550. Hence, their
22 individual effects were eliminated when controlling for the effect of rs1049550.

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

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

16 **Contributors**

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

22 **Funding**

1
2
3
4 1 The project was supported by the Research and Development Foundation of
5
6 2 Science and Technology of Henan Province (2012k15).
7

8
9 **3 Competing interests**

10
11 4 None.

12
13 **5 Patient consent**

14
15
16 6 Obtained.

17
18 **7 Ethics approval**

19
20
21 8 Ethics approval was provided by the First Hospital Affiliated to the Xinxiang
22
23 9 Medical College.

24
25 **10 Provenance and peer review**

26
27
28 11 Not commissioned; externally peer reviewed.

29
30 **12 Data sharing statement**

31
32
33 13 No additional data are available.

34
35 **14 Figure Legend**

36
37
38 **15 Fig. 1** The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the
39
40 pair-wise calculation of r^2 (left) or D' (right). Black squares indicate $r^2 = 1$ (i.e. perfect
41
42 LD between a pair of SNPs). Empty squares indicate $D' = 1$ (i.e. complete LD
43
44 between a pair of SNPs).
45
46
47

48
49 **19 Supplemental Fig. 1** The structure of the human ANXA11 gene and 15 SNPs located
50
51 on the gene.
52

53
54 **21 Reference**

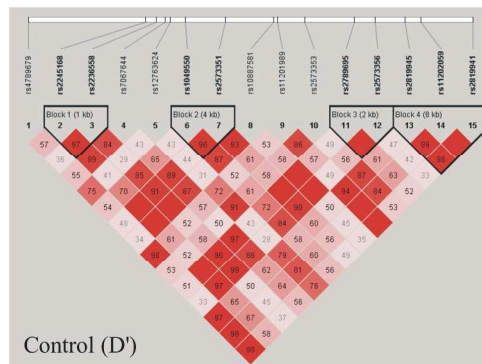
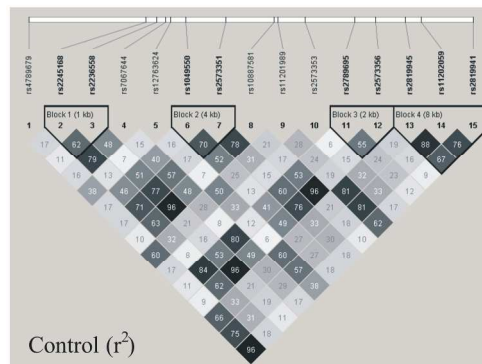
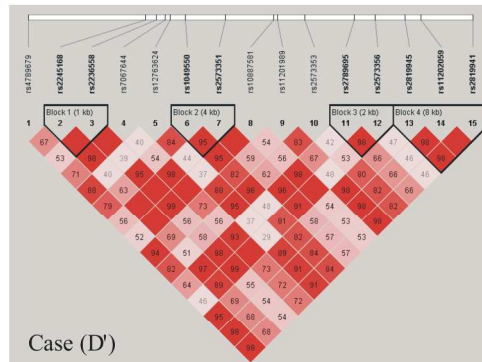
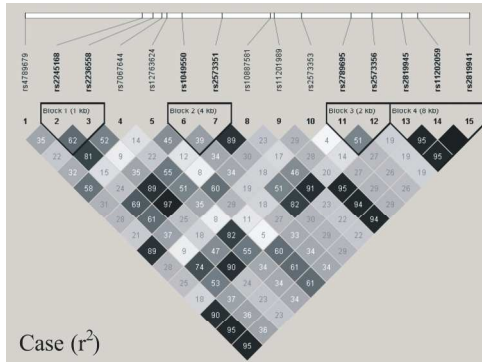
55
56 22 1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007;
57
58
59
60

- 1
2
3
4 1 357:2153-65.
5
6 2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-8
7
8
9 3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis.
10
11 *Semin Respir Crit Care Med* 2010;31:380-9.
12
13 4. Hofmann S, Franke A, Fischer A, *et al.* Genome-wide association study identifies
14 ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet*
15 2008;40:1103-6.
16
17
18 5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev*
19 2002;82:331-71.
20
21 6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using
22 recombinant human annexin XI-specific autoantibodies cloned by phage
23 display. *J Biol Chem* 2003;278:33120-6.
24
25
26 7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and
27 completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-22.
28
29 8. Petzmann S, Maercker C, Markert E, *et al.* Enhanced proliferation and decreased
30 apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis Vasc Diffuse*
31 *Lung Dis* 2006;23:190-200
32
33 9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, *et al.* Comparative
34 analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral
35 blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol*
36 2006;26:243-50.
37
38
39 10. Agostini C, Trentin L, Facco M, *et al.* Role of IL-15, IL-2, and their receptors in
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4 1 the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol*
5
6 2 1996;157:910-8
7
8
9 3 11. Jullien D, Sieling PA, Uyemura K, *et al.* IL-15, an immunomodulator of T cell
10
11 4 responses in intracellular infection. *J Immunol* 1997;158:800-6
12
13 5 12. Muller-Quernheim J, Schurmann M, Hofmann S, *et al.* Genetics of sarcoidosis.
14
15 6 *Clin Chest Med* 2008;29:391-414.
16
17 7 13. Morais A, Lima B, Peixoto M, *et al.* Annexin A11 gene polymorphism (R230C
18
19 8 variant) and sarcoidosis in a Portuguese population. *Tissue antigens*
20
21 9 2013;82:186-91.
22
23 10 14. Li Y, Pabst S, Kubisch C, *et al.* First independent replication study confirms the
24
25 11 strong genetic association of ANXA11 with sarcoidosis. *Thorax*
26
27 12 2010;65:939-40.
28
29 13 15. Gabriel SB, Schaffner SF, Nguyen H, *et al.* The structure of haplotype blocks in
30
31 14 the human genome. *Science* 2002;296:2225-9.
32
33 15 16. Barrett JC, Fry B, Maller J, *et al.* Haploview: analysis and visualization of LD and
34
35 16 haplotype maps. *Bioinformatics* 2005;21:263-5.
36
37 17 17. Curtis D, Knight J, Sham PC. Program report: GENECOUNTING support
38
39 18 programs. *Ann Hum Genet* 2006;70:277-9.
40
41 19 18. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic
42
43 20 associations. *Hum Hered* 2000;50:133-9.
44
45 21 19. Zhao JH, Lissarrague S, Essioux L, *et al.* GENECOUNTING: haplotype analysis
46
47 22 with missing genotypes. *Bioinformatics* 2002;18:1694-5
48
49
50
51
52
53
54
55
56
57
58
59
60

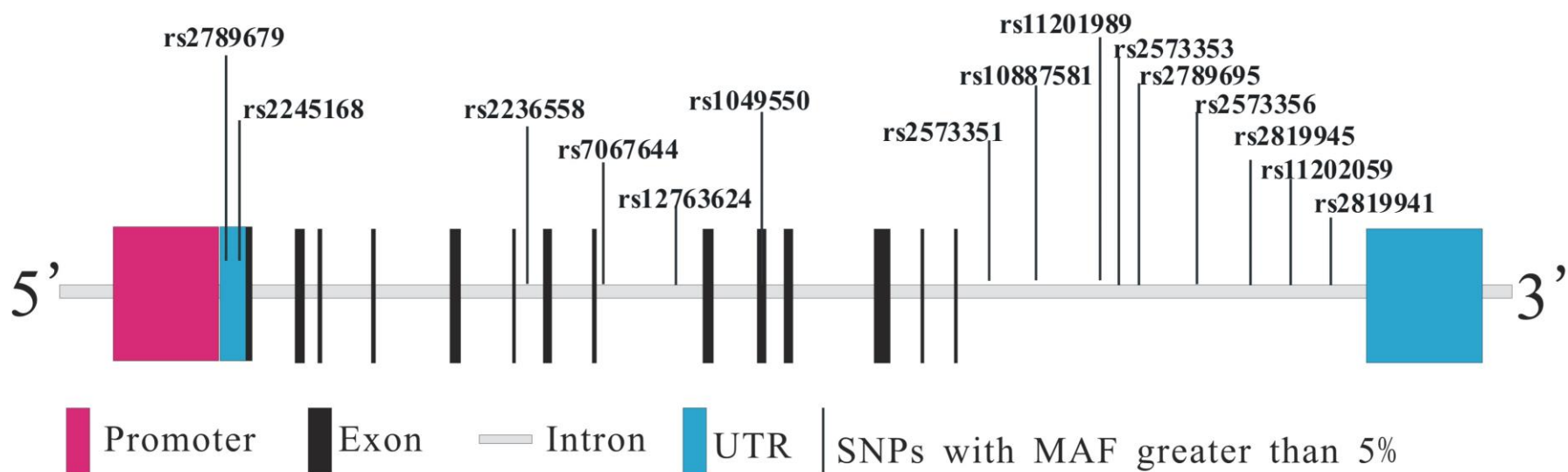
- 1
2
3
4 20. Zhao JH, Sham PC. Generic number systems and haplotype analysis. *Comput*
5
6 *Methods Programs Biomed* 2003;70:1-9
7
8
9 21. Mrazek F, Stahelova A, Kriegova E, *et al.* Functional variant ANXA11 R230C:
10
11 true marker of protection and candidate disease modifier in sarcoidosis. *Genes*
12
13 *Immun* 2011;12:490-4.
14
15
16 22. Levin AM, Iannuzzi MC, Montgomery CG, *et al.* Association of ANXA11 genetic
17
18 variation with sarcoidosis in African Americans and European Americans.
19
20
21 *Genes Immun* 2013;14:13-8.
22
23
24 24. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11
25
26 with the nuclear envelope. *J Biol Chem* 2003;278:20210-6.
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



273x203mm (300 x 300 DPI)

Review only



review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

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	

1							
2							
3							
4							
5							
6							
7	TT	120	28.7	104	25.2	0.263	1.19, 0.88-1.62
8	CT	202	48.3	216	52.4	0.238	0.85, 0.65-1.12
9	CC	96	23.0	92	22.3	0.826	1.04, 0.75-1.44
10	Per C allele	394	47.1	400	48.5	0.564	1.06, 0.87-1.28
11	rs10887581					0.290	
12	TT	114	27.3	96	23.3	0.190	1.23, 0.90-1.69
13	TC	190	45.5	208	50.5	0.148	0.82, 0.62-1.07
14	CC	114	27.3	108	26.2	0.729	1.06, 0.78-1.44
15	Per C allele	418	50.0	424	51.5	0.583	1.06, 0.87-1.29
16	rs11201989					0.144	
17	GG	130	31.1	128	31.1	0.995	1.00, 0.75-1.34
18	GC	218	52.2	194	47.1	0.141	1.23, 0.93-1.62
19	CC	70	16.8	90	21.9	0.060	0.71, 0.50-1.01
20	Per C allele	358	42.8	374	45.4	0.29	1.11, 0.91-1.35
21	rs2573353					0.918	
22	CC	178	42.6	176	42.7	0.957	0.99, 0.75-1.31
23	CA	196	46.9	192	46.6	0.921	1.01, 0.77-1.33
24	AA	44	10.5	44	10.7	0.941	0.98, 0.63-1.53
25	Per A allele	284	34.0	280	34.0	0.997	1.00, 0.82-1.23
26	rs2789695					0.918	
27	TT	178	42.6	180	34.0	0.740	0.95, 0.73-1.26
28	CT	188	45.0	184	44.7	0.920	1.01, 0.77-1.33
29	CC	52	12.4	48	11.7	0.73	1.08, 0.71-1.64
30	Per C allele	292	34.9	280	34.0	0.69	0.96, 0.78-1.17
31	rs2573356					0.892	
32	CC	118	28.2	112	27.2	0.742	1.05, 0.78-1.43
33							
34							
35							
36							
37							
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							
48							
49							

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: rs2789679+rs1049550+rs2819941	rs1049550: T	0.0002	0.65	0.52-0.82

For peer review only

BMJ Open

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

Journal:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

only

1
2
3
4 1 **Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a**
5
6 2 **Han Chinese population - a case-control study**

7
8
9 3 LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang

10
11 4 The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang
12
13 5 Medical College, Xinxiang, Henan, PR China

14
15
16 6 *Corresponding author: Zhijian Xiao, The Third Department of Tuberculosis, the First
17
18 7 Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China.

19
20
21 8 No.88 Jiankang Road, Weihui City, Henan, 463000, PR China

22
23
24 9 E-mail: zhangligongys@163.com

25
26 10 Tel: +86 0373 - 4402229 (office)

27
28
29 11 Fax: +86 0373 -4402251

30
31 12 **Key words:** Sarcoidosis; Annexin A11; Single-nucleotide polymorphism

32
33
34 13 Word count: 3918

35
36
37 14

1 **Abstract**

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

5 **Design:** A case - control study.

6 **Setting:** A tuberculosis unit in a hospital of the university in China.

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

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

11 **Results:** The statistical significant differences were found in the allelic or genotypic
12 frequencies of the rs2789679, rs1049550 and rs2819941 in the ANXA11 gene
13 between the patients with sarcoidosis and the controls. The rs2789679 A allele ($P =$
14 0.00004 , OR = 1.42, 95%CI = 1.17-1.73) and rs2819941 T allele ($P = 0.0006$, OR =
15 1.41, 95%CI = 1.16-1.71) were significantly more frequent in the patients with
16 sarcoidosis compared to controls. The frequency of the rs1049550 T allele ($P =$
17 0.000002 , OR = 0.61, 95%CI= 0.49 - 0.74) in the patients with sarcoidosis was
18 significantly lower than that in the controls. The multi-SNP model reveals that
19 rs1049550 is the only independent SNP association effect after accounting for the
20 other two marginally associated SNPs. In block 2 (rs1049550-rs2573351), the T - C
21 haplotype occurred significantly less frequently ($P = 0.001$), whereas the C - C
22 haplotypes occurred more frequently ($P = 0.0001$) in the patients with sarcoidosis

1
2
3
4 1 than the controls. Furthermore, genotype frequency distribution revealed that, in
5
6 2 rs1049550, CC genotype was significantly more in the patients with chest
7
8 3 radiographic (CXR) stage I sarcoidosis than the patients with CXR stage II-IV
9
10 4 sarcoidosis ($P = 0.012$).

11
12
13 5 **Conclusions:** These findings point to a role for the polymorphisms of ANXA11 in
14
15 6 sarcoidosis in a Chinese Han population, and may be informative for future genetic
16
17 7 studies on sarcoidosis.
18
19
20

21 8
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 **1 Article summary**

5
6 **2 Article focus**

7
8
9 3 1. The first genome-wide association study (GWAS) in sarcoidosis conducted in a
10
11 4 German population has revealed that the rs1049550 (C > T, Arg230Cys) located in
12
13 5 exon 10 were associated with sarcoidosis.

14
15
16 6 2. Whether these SNPs modulate the risk of sarcoidosis by themselves or whether
17
18 7 they correlate with other causative SNPs and are repeated in other populations remain
19
20 8 unclear.

21
22 9 3. We hypothesize that common variants in the ANXA11 gene may significantly
23
24 10 contribute to the predisposition to develop sarcoidosis in a Chinese Han population.

25
26
27
28
29 **11 Key messages**

30
31 12 1. ANXA11 rs2789679 (3'-UTR) was significantly associated with sarcoidosis.

32
33 13 2. ANXA11 rs1049550 (exon 10) was significantly associated with sarcoidosis.

34
35 14 3. Another significant association was observed for rs2819941 (intron 14).

36
37
38
39 **15 Strengths and limitations of this study**

40
41 16 1. A systematical screening of the functional SNPs in the promoter region, 5'- and
42
43 17 3'-UTR, exons of the ANXA11 gene, and the homogeneity of the study subjects
44
45 18 representing the Chinese Han are the main strengths of this study.

46
47
48 19 2. The lack of data proving the positive association observed for rs2789679 and
49
50 20 rs1049550 is a potential limitation of this study. Furthermore, the association of the
51
52 21 serum level of ANXA11 with sarcoidosis still needs to be investigated.

53
54
55
56
57
58
59
60 22

1 INTRODUCTION

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

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

1
2
3
4 1 However, it is generally assumed that the induction of apoptosis and/or by the
5
6 2 withdrawal of inflammatory cytokines participates in the disappearance of
7
8 3 granulomas^{10 11}. In the patients with sarcoidosis, their peripheral blood monocytes
9
10 4 were characterized by apoptosis⁹.

11
12
13
14 5 In previous studies, case - control/family ‘hypothesis - driven’ studies and low
15
16 6 density linkage scans were used to identify genetic factors conferring the genetic
17
18 7 susceptibility to sarcoidosis¹². Notably, the first genome - wide association study
19
20 8 (GWAS) in sarcoidosis conducted in a German population has recently revealed an
21
22 9 association: the leading rs2789679 SNP (single - nucleotide polymorphism) located in
23
24 10 the 3’-untranslated region (3’-UTR) and a common nonsynonymous SNP (rs1049550,
25
26 11 C > T, Arg230Cys) were associated with the increased risk of sarcoidosis^{4 13}. This
27
28 12 association has recently been supported by another report from the same population¹⁴.
29
30 13 Thus, more studies should be performed to demonstrate the following items: whether
31
32 14 these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate
33
34 15 with other causative SNPs and are repeated in other populations. The published
35
36 16 studies about the association of sarcoidosis and ANXA11 are summarized in
37
38 17 Supplemental table 1. We hypothesize that common variants in the ANXA11 gene
39
40 18 may significantly contribute to the predisposition to develop sarcoidosis.
41
42
43
44
45
46
47
48

49 19 In this study, we investigated fifteen loci in a Chinese population from He’nan
50
51 20 province (China) to verify the putative association between ANXA11 polymorphisms
52
53 21 and sarcoidosis.
54
55

56 22 **SUBJECTS AND METHODS**

57
58
59
60

1 **Subjects**

2 Four hundred and twelve patients with sarcoidosis (mean \pm SD age: 53.6 ± 4.6
3 years; 155 men and 257 women) were recruited from our hospital between May 2005
4 and March 2013. Sarcoidosis was diagnosed by the evidence of non-caseating
5 epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities.
6 Chronic sarcoidosis was defined as a disease over at least 2 years or at least two
7 episodes in a lifetime. A cute sarcoidosis was defined as one episode of acute
8 sarcoidosis which had totally resolved at the date of the examination. The control
9 group consisted of 418 unrelated healthy subjects (mean \pm SD age: 54.2 ± 5.3 years;
10 160 men and 258 women) who underwent health examinations in the Medical
11 Examination Center of the First Affiliated Hospital of the Xinxiang Medical College
12 (Xinxiang, China) between Oct 2009 and Sept 2013. None of the individuals in the
13 control group had a history of lung diseases or showed any symptoms of the lung or
14 other diseases by chest radiography or laboratory blood tests. Participants were
15 excluded if they: were taking other prescribed medications that could affect the
16 central nervous system; had a history of seizures, hematological diseases, or severe
17 liver or kidney impairment; or were pregnant. No familial relationship was known
18 between the study participants. The study was performed according to the Guidelines
19 of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang, China).
20 Written informed consent was obtained from each participant in this study.

21 **SNP selection**

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

12 Genotyping

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

1 of - flight mass spectrometer. Genotype calling was performed in realtime with the
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 of - flight mass spectrometer. Genotype calling was performed in realtime with the
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

4 **Statistical analysis**

5 All data were analyzed using the SPSS 17.0 software (SPSS Inc., Chicago, IL,
6 USA). Hardy-Weinberg equilibrium was evaluated by Chi-square tests. Differences
7 between the cases and controls in the frequency of the alleles, genotypes and
8 haplotypes were evaluated by the Fisher's exact test or the Pearson Chi-square test.
9 Unconditional logistic regression was used to calculate the odds ratio (OR) and 95%
10 confidence interval (CI) in independent association between each locus and the
11 presence of sarcoidosis. Gender and age of subjects were treated as covariants in
12 binary logistic regression. *P* values were calculated based on codominant, dominant
13 for the rare allele, heterosis and recessive for the rare allele models of inheritance.
14 Models of multiple logistic regression were used to test the independence of
15 individual allelic effect. In detail, the most significant SNP was chosen to be the
16 conditional SNP (covariate in the regression model) when testing other significant
17 SNPs; Meanwhile, a multi-SNP model including all significant SNPs was also
18 performed. The Bonferroni correction was used to adjust the test level when multiple
19 comparisons were conducted, and the *P* value was divided by the total number of loci.
20 Haplotype blocks were defined according to the criteria developed by Gabriel et al.¹⁵.
21 Pair-wise LD statistics (D' and r^2) and haplotype frequency were calculated, and
22 haplotype blocks were constructed using the Haploview 4.0¹⁶. To ensure that the LD

1 blocks most closely reflect the population level LD patterns, definition of the blocks
2 were based on the control samples alone. The haplotype frequencies were estimated
3 using GENECOUNTING, which computes maximum - likelihood estimates of
4 haplotype frequencies from unknown phase data by utilizing an expectation -
5 maximization algorithm¹⁷⁻²⁰. The significance of any haplotypic association was
6 evaluated using a likelihood ratio test, followed by permutation testing that compared
7 estimated haplotype frequencies in cases and controls^{17 19}.

8 RESULTS

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

17 Comparison of genotype and allele frequency distribution revealed significant
18 differences between the patients with sarcoidosis and healthy controls for 3 SNPs:
19 rs2789679, rs1049550 and rs2819941. The frequency of the rs1049550 T allele ($P =$
20 0.000002, OR = 0.61, 95%CI= 0.49-0.74) in the patients with sarcoidosis was
21 significantly lower than that in the controls. These differences remained statistically
22 significant after Bonferroni corrections. The rs1049550 was the most significant of the

1
2
3
4 1 three. A multi-SNP model revealed rs1049550 to be the only significant independent
5
6 2 SNP association with sarcoidosis risk.
7

8
9 3 The multi-SNP model showed only rs1049550 present a significant effect on the
10
11 4 disease phenotype ($p < 0.001$). No independent effect was found for rs2789679 or
12
13 5 rs2819941 when adjusting for the effect of rs1049550. Both individual SNP and multi -
14
15 6 SNP analysis supported rs1049550: T allele was an important protective factor for
16
17 7 affecting sarcoidosis (odds ratio around 0.6). In another word, carriers of rs1049550:
18
19 8 C allele have higher susceptibility to sarcoidosis in our Chinese Han population
20
21 9 (Supplemental table 3).
22
23
24

25
26 10 We performed an association analysis to determine whether the haplotype was
27
28 11 associated with the risk of sarcoidosis (Global $P < 0.05$ in block 2 and block 3).
29
30 12 Compared with the controls, the T - C haplotype occurred significantly less frequently
31
32 13 ($P = 0.001$) but the C - C haplotypes occurred more frequently ($P = 0.0001$) in block 2
33
34 14 (rs1049550 - rs2573351) in the patients with sarcoidosis.
35
36
37
38

39 15 To assess particular disease phenotypes, the patients with sarcoidosis were
40
41 16 divided into the subgroups according to their chest radiographic (CXR) stage. CXR
42
43 17 stage I (isolated bilateral hilar lymphadenopathy): $N = 196$; CXR stages II - IV
44
45 18 (infiltration of parenchyma): 167; CXR stage 0 (absence on pulmonary disease): $N =$
46
47 19 28; the information on CXR stage was not available for 21 patients. Genotype
48
49 20 frequency distribution revealed that, in rs1049550, CC genotype was significantly
50
51 21 more in the patients with chest radiographic (CXR) stage I sarcoidosis than the
52
53 22 patients with CXR stage II-IV sarcoidosis ($P = 0.012$). (Table 4).
54
55
56
57
58
59
60

1 DISCUSSION

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

16 Hoffman et al. reported an association between the T allele of the non -
17 synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a
18 German population⁴. Similar results were also obtained in other two European
19 populations^{14 21}. As a part of the sarcoidosis GWAS done in Americans²², we further
20 confirmed this association in a Chinese Han population. In this study, the frequency of
21 ANXA11 rs1049550 T allele was significantly lower in the patients with sarcoidosis
22 than the healthy controls. Genotype frequency distribution revealed that, in rs1049550,

1
2
3
4 1 CC genotype was significantly more in the patients with stage I sarcoidosis than the
5
6 2 patients with stage II-IV sarcoidosis. The “protective” effect of ANXA11 T allele
7
8
9 3 increased with the number of its copies in the genotype, which is consistent with the
10
11 4 result obtained in the patients with sarcoidosis in a German population ⁴, Furthermore,
12
13
14 5 the T allele carriers among the patients were protected from infiltration of lung
15
16 6 parenchyma (radiographic stages II-IV) ²¹. We have also demonstrated that rs1049550
17
18
19 7 is significant cross - ethnically at the gene level after adjustment for the single SNP
20
21 8 association tests performed. The mechanism by which rs1049550 affects the
22
23
24 9 susceptibility to sarcoidosis is that the SNP may affect the function of the ANXA11
25
26 10 protein rather than affect its expression. In humans, the ANXA11 protein consists of
27
28
29 11 an N - terminalproline - tyrosine - glycine (PYG) - rich region, followed by four
30
31 12 annexin core domains. The rs1049550 leads to an amino - acid exchange (basic
32
33
34 13 arginine to a polar cysteine) at the evolutionarily conserved position 230 (R230C) in
35
36 14 the first annexin domain, which is responsible for Ca²⁺ - dependent trafficking of the
37
38
39 15 protein in the cell ²³.

40
41 16 In the GWAS conducted by Hofmann et al ⁴, the strongest association signal was
42
43
44 17 observed for rs2789679. This marker is located in the 3'-UTR of the ANXA11 gene.
45
46 18 The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in
47
48
49 19 the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls
50
51 20 being TT homozygous. Fine-mapping in the Black Women's Health Study (BWHS)
52
53
54 21 revealed rs2819941 to be the top SNP, associated with a non-significant ($P = 0.06$)
55
56 22 reduction in the risk of sarcoidosis (17% reduction per copy of the C-allele). In this
57
58
59
60

1
2
3
4 1 case-control association study, the frequency of the A allele in rs2789679 and the
5
6 2 rs2819941 T allele frequency in the patients with sarcoidosis was significantly higher
7
8
9 3 than that in the healthy controls. Several lines of evidence suggest that the observed
10
11 4 association is unlikely to be an artifact. First, both the single-SNP and the
12
13 5 haplotype-based association analyses support the current finding. Second, our samples
14
15 6 were from the same geographical region, excluding the. Finally, consistent results
16
17 7 were obtained from two genetically independent populations (Chinese Han and
18
19 8 Europeans). Collectively, our results confirmed the strong association between
20
21 9 ANXA11 polymorphisms and sarcoidosis, suggesting that ANXA11 represents a
22
23 10 strong genetic risk factor for sarcoidosis. Since causal SNP in ANXA11 was not
24
25 11 known, all the SNPs significantly associated with sarcoidosis were actually indirect
26
27 12 association through the real causal one. Though rs2789679 and rs2819941 were not in
28
29 13 LD with rs1049550, it may not tag the causal SNP as well as rs1049550. Hence, their
30
31 14 individual effects were eliminated when controlling for the effect of rs1049550.

32
33
34
35
36
37
38
39 15 We further investigated the interaction among polymorphisms and observed
40
41 16 strong LD. The haplotype analysis revealed that significantly more C - C (block 2,
42
43 17 rs1049550 - rs2573351) and significantly fewer T - C haplotypes (block 2) were
44
45 18 found in the patients with sarcoidosis. These results indicated that the patients with C
46
47 19 - C (block 2) haplotypes of the ANXA11 gene were more prone to sarcoidosis.
48
49 20 Significantly higher frequencies of T - C haplotypes were detected in the healthy
50
51 21 controls than in the patients with sarcoidosis, suggesting that they may show
52
53 22 protective effects against sarcoidosis. Our sample size can detect SNP and haplotype
54
55
56
57
58
59
60

1
2
3
4 1 associations with 90% and 85% power, respectively, at a false positive rate of 5%,
5
6 2 disease prevalence of 1%, disease allele / haplotype frequency of 0.05 / 0.03, and a
7
8
9 3 presumed odds ratio (OR) of 1.5. To some extent, this finding further supports a role
10
11 4 of ANXA11 polymorphisms in sarcoidosis, while ethnic group difference may exist.

12
13
14 5 In conclusion, our study suggests a potential role of the ANXA11 SNPs and their
15
16 6 related haplotypes in the genetic susceptibility to sarcoidosis. Further studies are
17
18 7 needed to investigate how these SNPs affect the function of ANXA11. A broader
19
20 8 examination of the genetic variation in ANXA11 in the Han Chinese may reveal other
21
22 9 variants associated with disease risk.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 **Contributors**

5
6 LG Zhang was involved in conception and design of the study, acquisition of
7
8 patient data, genotyping and drafted the paper. ZJ Xiao and GC Shi were involved in
9
10 design of the study, genotyping and interpretation of results. J Huang and CX Zhang
11
12 were involved in the design of the study, acquisition of patient data and interpretation
13
14 of result. All authors revised the draft paper.
15
16

17
18
19 **Funding**

20
21 The project was supported by the Research and Development Foundation of
22
23 Science and Technology of Henan Province (2012k15).
24
25

26 **Competing interests**

27
28
29 None.

30
31 **Patient consent**

32
33
34 Obtained.

35
36 **Ethics approval**

37
38
39 Ethics approval was provided by the First Hospital Affiliated to the Xinxiang
40
41 Medical College.
42
43

44 **Provenance and peer review**

45
46
47 Not commissioned; externally peer reviewed.
48

49 **Data sharing statement**

50
51
52 No additional data are available.
53
54
55
56
57
58
59
60

1
2
3
4 **Reference**

- 5
6 1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007;
7
8 357:2153-65.
9
10 2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-8
11
12 3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis.
13
14 *Semin Respir Crit Care Med* 2010;31:380-9.
15
16 4. Hofmann S, Franke A, Fischer A, *et al.* Genome-wide association study identifies
17
18 ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet*
19
20 2008;40:1103-6.
21
22 5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev*
23
24 2002;82:331-71.
25
26 6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using
27
28 recombinant human annexin XI-specific autoantibodies cloned by phage
29
30 display. *J Biol Chem* 2003;278:33120-6.
31
32 7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and
33
34 completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-22.
35
36 8. Petzmann S, Maercker C, Markert E, *et al.* Enhanced proliferation and decreased
37
38 apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis Vasc Diffuse*
39
40 *Lung Dis* 2006;23:190-200
41
42 9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, *et al.* Comparative
43
44 analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral
45
46 blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol*
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4 1 2006;26:243-50.
5
6 2 10. Agostini C, Trentin L, Facco M, *et al.* Role of IL-15, IL-2, and their receptors in
7
8 3 the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol*
9
10 4 1996;157:910-8
11
12 5 11. Jullien D, Sieling PA, Uyemura K, *et al.* IL-15, an immunomodulator of T cell
13
14 6 responses in intracellular infection. *J Immunol* 1997;158:800-6
15
16 7 12. Muller-Quernheim J, Schurmann M, Hofmann S, *et al.* Genetics of sarcoidosis.
17
18 8 *Clin Chest Med* 2008;29:391-414.
19
20 9 13. Morais A, Lima B, Peixoto M, *et al.* Annexin A11 gene polymorphism (R230C
21
22 10 variant) and sarcoidosis in a Portuguese population. *Tissue antigens*
23
24 11 2013;82:186-91.
25
26 12 14. Li Y, Pabst S, Kubisch C, *et al.* First independent replication study confirms the
27
28 13 strong genetic association of ANXA11 with sarcoidosis. *Thorax*
29
30 14 2010;65:939-40.
31
32 15 15. Gabriel SB, Schaffner SF, Nguyen H, *et al.* The structure of haplotype blocks in
33
34 16 the human genome. *Science* 2002;296:2225-9.
35
36 17 16. Barrett JC, Fry B, Maller J, *et al.* Haploview: analysis and visualization of LD and
37
38 18 haplotype maps. *Bioinformatics* 2005;21:263-5.
39
40 19 17. Curtis D, Knight J, Sham PC. Program report: GENECOUNTING support
41
42 20 programs. *Ann Hum Genet* 2006;70:277-9.
43
44 21 18. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic
45
46 22 associations. *Hum Hered* 2000;50:133-9.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4 19. Zhao JH, Lissarrague S, Essioux L, et al. GENECOUNTING: haplotype analysis
5
6 with missing genotypes. *Bioinformatics* 2002;18:1694-5
7
8
9 20. Zhao JH, Sham PC. Generic number systems and haplotype analysis. *Comput*
10
11 *Methods Programs Biomed* 2003;70:1-9
12
13
14 21. Mrazek F, Stahelova A, Kriegova E, et al. Functional variant ANXA11 R230C:
15
16 true marker of protection and candidate disease modifier in sarcoidosis. *Genes*
17
18 *Immun* 2011;12:490-4.
19
20
21 22. Levin AM, Iannuzzi MC, Montgomery CG, et al. Association of ANXA11 genetic
22
23 variation with sarcoidosis in African Americans and European Americans.
24
25 *Genes Immun* 2013;14:13-8.
26
27
28
29 24. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11
30
31 with the nuclear envelope. *J Biol Chem* 2003;278:20210-6.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 **Figure Legend**

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

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

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

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)

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

3 ^a Based on 10,000 permutations.

4 ^b Based on comparison of frequency distribution of all haplotypes for the combination of SNPs.

5
 6 **Table 3** ANXA11 haplotype in block 3 frequencies and the results of their associations with risk
 7 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	

8 ^a Based on 10,000 permutations.

9 ^b Based on comparison of frequency distribution of all haplotypes for the combination of SNPs.

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

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 ($P < 0.05$).

4

5

6

1
2
3
4 1 **Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a**

5
6 2 **Han Chinese population - a case-control study**

7
8
9 3 LG Zhang, ZJ Xiao*, GC Shi, J Huang, CX Zhang

10
11 4 The Third Department of Tuberculosis, the First Hospital Affiliated to the Xinxiang

12
13
14 5 Medical College, Xinxiang, Henan, PR China

15
16 6 *Corresponding author: Zhijian Xiao, The Third Department of Tuberculosis, the First

17
18
19 7 Hospital Affiliated to the Xinxiang Medical College, Xinxiang, Henan, PR China.

20
21 8 No.88 Jiankang Road, Weihui City, Henan, 463000, PR China

22
23
24 9 E-mail: zhangligongys@163.com

25
26 10 Tel: +86 0373 - 4402229 (office)

27
28
29 11 Fax: +86 0373 -4402251

30
31 12 **Key words:** Sarcoidosis; Annexin A11; Single-nucleotide polymorphism

32
33
34 13 Word count: 3918

35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60 14

1
2
3
4 **Abstract**

5
6 **Objectives:** To further identify the single-nucleotide polymorphisms (SNPs) that
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
3 contribute to the genetic susceptibility to sarcoidosis, we examined the potential
4 association between sarcoidosis and fifteen SNPs of the ANXA11 gene.

5 **Design:** A case - control study.

6 **Setting:** A tuberculosis unit in a hospital of the university in China.

7 **Participants:** Participants included 412 patients with sarcoidosis and 418 healthy
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
8 controls.

9 **Methods:** The selected SNPs were genotyped using the MALDI-TOF in the
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
MassARRAY system. ~~Probes and primers were designed using the Assay Design
Software (Sequenom, San Diego, CA, USA).~~

12 **Results:** The statistical significant differences were found in the allelic or genotypic
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
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

1
2
3
4 1 haplotype occurred significantly less frequently ($P = 0.001$), whereas the C - C
5
6 2 haplotypes occurred more frequently ($P = 0.0001$) in the patients with sarcoidosis
7
8
9 3 than the controls. Furthermore, genotype frequency distribution revealed that, in
10
11 4 rs1049550, CC genotype was significantly more in the patients with chest
12
13
14 5 radiographic (CXR) stage I sarcoidosis than the patients with CXR stage II-IV
15
16 6 sarcoidosis ($P = 0.012$).

17
18
19 7 **Conclusions:** These findings point to a role for the polymorphisms of ANXA11 in
20
21 8 sarcoidosis in a Chinese Han population, and may be informative for future genetic
22
23
24 9 studies on sarcoidosis.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 **1 Article summary**

5
6 **2 Article focus**

7
8
9 3 1. The first genome-wide association study (GWAS) in sarcoidosis conducted in a
10
11 4 German population has revealed that the rs1049550 (C > T, Arg230Cys) located in
12
13 5 exon 10 were associated with sarcoidosis.

14
15
16 6 2. Whether these SNPs modulate the risk of sarcoidosis by themselves or whether
17
18 7 they correlate with other causative SNPs and are repeated in other populations remain
19
20 8 unclear.

21
22 9 3. We hypothesize that common variants in the ANXA11 gene may significantly
23
24 10 contribute to the predisposition to develop sarcoidosis in a Chinese Han population.

25
26
27
28
29 **11 Key messages**

30
31 12 1. ANXA11 rs2789679 (3'-UTR) was significantly associated with sarcoidosis.

32
33 13 2. ANXA11 rs1049550 (exon 10) was significantly associated with sarcoidosis.

34
35 14 3. Another significant association was observed for rs2819941 (intron 14).

36
37
38
39 **15 Strengths and limitations of this study**

40
41 16 1. A systematical screening of the functional SNPs in the promoter region, 5'- and
42
43 17 3'-UTR, exons of the ANXA11 gene, and the homogeneity of the study subjects
44
45 18 representing the Chinese Han are the main strengths of this study.

46
47
48 19 2. The lack of data proving the positive association observed for rs2789679 and
49
50 20 rs1049550 is a potential limitation of this study. Furthermore, the association of the
51
52 21 serum level of ANXA11 with sarcoidosis still needs to be investigated.

53
54
55
56 **22 INTRODUCTION**

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

12 The Annexin gene family is involved in the etiology of several autoimmune and
13 chronic diseases^{4 5}. One member of the Annexin gene family, ANXA11, located on
14 chromosome 10q22.3, is involved in calcium signaling, apoptosis, vesicle trafficking,
15 cell growth and terminal phase of cell division⁵⁻⁷. In this context, the development
16 and maintenance of the granulomatous inflammation in sarcoidosis have been
17 repeatedly associated with the impaired apoptosis of activated inflammatory cells^{8 9}.
18 Most of the early-stage sarcoidosis is characterized by mononuclear cell alveolitis
19 dominated by activated CD4⁺ T cells and macrophages. Between these cells, the
20 uncoordinated interplay results in the formation of typical non-caseating granulomas.
21 The mechanism by which the granulomas resolve has not been fully elucidated.
22 However, it is generally assumed that the induction of apoptosis and/or by the

1 withdrawal of inflammatory cytokines participates in the disappearance of
2 granulomas^{10 11}. In the patients with sarcoidosis, their peripheral blood monocytes
3 were characterized by apoptosis⁹.

4 In previous studies, case - control/family ‘hypothesis - driven’ studies and low
5 density linkage scans were used to identify genetic factors conferring the genetic
6 susceptibility to sarcoidosis¹². Notably, the first genome - wide association study
7 (GWAS) in sarcoidosis conducted in a German population has recently revealed an
8 association: the leading rs2789679 SNP (single - nucleotide polymorphism) located in
9 the 3’-untranslated region (3’-UTR) and a common nonsynonymous SNP (rs1049550,
10 C > T, Arg230Cys) were associated with the increased risk of sarcoidosis^{4 13}. This
11 association has recently been supported by another report from the same population¹⁴.
12 Thus, more studies should be performed to demonstrate the following items: whether
13 these SNPs modulate the risk of sarcoidosis by themselves or whether they correlate
14 with other causative SNPs and are repeated in other populations. The published
15 studies about the association of sarcoidosis and ANXA11 are summarized in
16 Supplemental table 1. We hypothesize that common variants in the ANXA11 gene
17 may significantly contribute to the predisposition to develop sarcoidosis.

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

21 **SUBJECTS AND METHODS**

22 **Subjects**

1
2
3
4 1 Four hundred and twelve patients with sarcoidosis (mean \pm SD age: 53.6 ± 4.6
5
6 2 years; 155 men and 257 women) were recruited from our hospital between May 2005
7
8
9 3 and March 2013. Sarcoidosis was diagnosed by the evidence of non-caseating
10
11 4 epitheloid cell granuloma in biopsy specimens and chest radiographic abnormalities.
12
13
14 5 Chronic sarcoidosis was defined as a disease over at least 2 years or at least two
15
16 6 episodes in a lifetime. A acute sarcoidosis was defined as one episode of acute
17
18
19 7 sarcoidosis which had totally resolved at the date of the examination. The control
20
21 8 group consisted of 418 unrelated healthy subjects (mean \pm SD age: 54.2 ± 5.3 years;
22
23
24 9 160 men and 258 women) who underwent health examinations in the Medical
25
26 10 Examination Center of the First Affiliated Hospital of the Xinxiang Medical College
27
28
29 11 (Xinxiang, China) between Oct 2009 and Sept 2013. None of the individuals in the
30
31 12 control group had a history of lung diseases or showed any symptoms of the lung or
32
33
34 13 other diseases by chest radiography or laboratory blood tests. Participants were
35
36 14 excluded if they: were taking other prescribed medications that could affect the
37
38
39 15 central nervous system; had a history of seizures, hematological diseases, or severe
40
41 16 liver or kidney impairment; or were pregnant. No familial relationship was known
42
43
44 17 between the study participants. The study was performed according to the Guidelines
45
46 18 of the Medical Ethical Committee of Xinxiang Medical College (Xinxiang, China).
47
48
49 19 Written informed consent was obtained from each participant in this study.

50 51 **SNP selection**

52
53
54 21 Tagging SNPs were selected from the catalogs of the International HapMap
55
56 22 Project. The rs2789679 and rs2245168 are located in 5'-UTR, rs2236558 in intron 6,
57
58
59
60

1
2
3
4 1 rs7067644 and rs12763624 in intron 8, rs1049550 in exon 10, rs2573351 in intron 13,
5
6 2 rs10887581, rs11201989, rs2573353, rs2789695, rs2573356, rs2819945, rs11202059
7
8
9 3 and rs2819941 in intron 14 (Supplemental figure 1). Marker selection was based on
10
11 4 the following criteria. (a) We used the CHB data from the HapMap (release 27) to
12
13 5 select tagSNPs for ANXA11. (b) We restricted our search for tagSNPs from base pair
14
15 6 80155124 to 80205677. (c) Further, we limited the SNPs to tag to those with a minor
16
17 7 allele frequency ≥ 0.2 in the CHB. (d) Based on these restrictions, there were a total
18
19 8 of 29 SNPs. (e) Using HAPLOVIEW with an r^2 threshold ≥ 0.8 , 15 tagSNP were
20
21 9 selected and used for subsequent analyses.
22
23
24
25

26 **Genotyping**

27
28
29 11 Peripheral blood was collected from a vein into a sterile tube coated with
30
31 12 ethylenediamine tetraacetic acid (EDTA). Plasma samples were stored at -20°C .
32
33 13 Genomic DNA was extracted from the frozen peripheral blood samples using a
34
35 14 QIAmp Blood Mini Kit (Qiagen Inc., Valencia, CA, USA) according to the
36
37 15 manufacturer's protocols. The selected SNPs were genotyped in cases and controls by
38
39 16 using the MALDI-TOF in the MassARRAY system (Sequenom Inc., San Diego, CA,
40
41 17 USA). Probes and primers were designed using the Assay Design Software
42
43 18 (Sequenom, San Diego, CA, USA). The completed genotyping reactions were spotted
44
45 19 onto a 384 well spectroCHIP (Sequenom) using the MassARRAY Nanodispenser
46
47 20 (Sequenom) and determined by the matrix - assisted laser desorption ionization time -
48
49 21 of - flight mass spectrometer. Genotype calling was performed in realtime with the
50
51 22 MassARRAY RT software version 3.0.0.4 and analyzed using the MassARRAY Typer
52
53
54
55
56
57
58
59
60

1 software version 3.4 (Sequenom).

2 **Statistical analysis**

3 All data were analyzed using the SPSS 17.0 software (SPSS Inc., Chicago, IL,
4 USA). Hardy-Weinberg equilibrium was evaluated by Chi-square tests. Differences
5 between the cases and controls in the frequency of the alleles, genotypes and
6 haplotypes were evaluated by the Fisher's exact test or the Pearson Chi-square test.
7 Unconditional logistic regression was used to calculate the odds ratio (OR) and 95%
8 confidence interval (CI) in independent association between each locus and the
9 presence of sarcoidosis. Gender and age of subjects were treated as covariants in
10 binary logistic regression. *P* values were calculated based on codominant, dominant
11 for the rare allele, heterosis and recessive for the rare allele models of inheritance.
12 Models of multiple logistic regression were used to test the independence of
13 individual allelic effect. In detail, the most significant SNP was chosen to be the
14 conditional SNP (covariate in the regression model) when testing other significant
15 SNPs; Meanwhile, a multi-SNP model including all significant SNPs was also
16 performed. The Bonferroni correction was used to adjust the test level when multiple
17 comparisons were conducted, and the *P* value was divided by the total number of loci.
18 Haplotype blocks were defined according to the criteria developed by Gabriel et al.¹⁵.
19 Pair-wise LD statistics (D' and r^2) and haplotype frequency were calculated, and
20 haplotype blocks were constructed using the Haploview 4.0¹⁶. To ensure that the LD
21 blocks most closely reflect the population level LD patterns, definition of the blocks
22 were based on the control samples alone. The haplotype frequencies were estimated

1 using GENECOUNTING, which computes maximum - likelihood estimates of
2 haplotype frequencies from unknown phase data by utilizing an expectation -
3 maximization algorithm¹⁷⁻²⁰. The significance of any haplotypic association was
4 evaluated using a likelihood ratio test, followed by permutation testing that compared
5 estimated haplotype frequencies in cases and controls^{17 19}.

6 RESULTS

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

15 Comparison of genotype and allele frequency distribution revealed significant
16 differences between the patients with sarcoidosis and healthy controls for 3 SNPs:
17 rs2789679, rs1049550 and rs2819941. ~~The rs2789679 A allele ($P = 0.00004$, OR =~~
18 ~~1.42, 95%CI = 1.17-1.73) and rs2819941 T allele ($P = 0.0006$, OR = 1.41, 95%CI =~~
19 ~~1.16-1.71) were significantly more frequent in the patients with sarcoidosis compared~~
20 ~~to controls.~~ The frequency of the rs1049550 T allele ($P = 0.000002$, OR = 0.61,
21 95%CI= 0.49-0.74) in the patients with sarcoidosis was significantly lower than that
22 in the controls. These differences remained statistically significant after Bonferroni

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

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

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

16 To assess particular disease phenotypes, the patients with sarcoidosis were
17 divided into the subgroups according to their chest radiographic (CXR) stage. CXR
18 stage I (isolated bilateral hilar lymphadenopathy): $N = 196$; CXR stages II - IV
19 (infiltration of parenchyma): 167; CXR stage 0 (absence on pulmonary disease): $N =$
20 28; the information on CXR stage was not available for 21 patients. Genotype
21 frequency distribution revealed that, in rs1049550, CC genotype was significantly
22 more in the patients with chest radiographic (CXR) stage I sarcoidosis than the

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

2 **DISCUSSION**

3 A key step in linkage and association studies is to identify common risk variants
4 in different populations. To determine if common risk variants exist in distinct
5 populations, fifteen SNPs which span approximately 0.53 Mb of the ANXA11 gene,
6 were genotyped in samples from the patients with sarcoidosis and healthy controls in
7 a Chinese Han population. Recently, new sarcoidosis loci have been identified by
8 genome-wide association studies^{4 14}. With the fast development of genome - wide
9 association studies, an increasing number of susceptibility loci for sarcoidosis have
10 been found in different populations^{4 14 21}. However, these observations should be
11 confirmed in other genetically independent populations. In this study, we conducted
12 the first large genetic association study of the ANXA11 gene in a Chinese Han
13 population. The evidence of markers associated with sarcoidosis was presented, and
14 these markers were mapped to different locations in the ANXA11 gene (81897864 -
15 81951001). The association signals in the region were identified, and some
16 significantly associated haplotypes also appeared this region.

17 Hoffman et al. reported an association between the T allele of the non -
18 synonymous SNP rs1049550 and significantly decreased risk of sarcoidosis in a
19 German population⁴. Similar results were also obtained in other two European
20 populations^{14 21}. As a part of the sarcoidosis GWAS done in Americans²², we further
21 confirmed this association in a Chinese Han population. In this study, the frequency of
22 ANXA11 rs1049550 T allele was significantly lower in the patients with sarcoidosis

1
2
3
4 1 than the healthy controls. Genotype frequency distribution revealed that, in rs1049550,
5
6 2 CC genotype was significantly more in the patients with stage I sarcoidosis than the
7
8 3 patients with stage II-IV sarcoidosis. The “protective” effect of ANXA11 T allele
9
10 4 increased with the number of its copies in the genotype, which is consistent with the
11
12 5 result obtained in the patients with sarcoidosis in a German population ⁴, Furthermore,
13
14 6 the T allele carriers among the patients were protected from infiltration of lung
15
16 7 parenchyma (radiographic stages II-IV) ²¹. We have also demonstrated that rs1049550
17
18 8 is significant cross - ethnically at the gene level after adjustment for the single SNP
19
20 9 association tests performed. The mechanism by which rs1049550 affects the
21
22 10 susceptibility to sarcoidosis is that the SNP may affect the function of the ANXA11
23
24 11 protein rather than affect its expression. In humans, the ANXA11 protein consists of
25
26 12 an N - terminalproline - tyrosine - glycine (PYG) - rich region, followed by four
27
28 13 annexin core domains. The rs1049550 leads to an amino - acid exchange (basic
29
30 14 arginine to a polar cysteine) at the evolutionarily conserved position 230 (R230C) in
31
32 15 the first annexin domain, which is responsible for Ca²⁺ - dependent trafficking of the
33
34 16 protein in the cell ²³.

35
36
37
38
39
40
41
42
43 17 In the GWAS conducted by Hofmann et al ⁴, the strongest association signal was
44
45 18 observed for rs2789679. This marker is located in the 3'-UTR of the ANXA11 gene.
46
47 19 The T allele had a frequency of 35% among the patients with sarcoidosis and 44% in
48
49 20 the healthy controls, with 13% of the patients with sarcoidosis and 19% of controls
50
51 21 being TT homozygous. Fine-mapping in the Black Women's Health Study (BWHS)
52
53 22 revealed rs2819941 to be the top SNP, associated with a non-significant ($P = 0.06$)
54
55
56
57
58
59
60

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

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

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

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

11

12 **Contributors**

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

18 **Funding**

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

21 **Competing interests**

22 None.

1
2
3
4 **1 Patient consent**

5
6 2 Obtained.

7
8
9 **3 Ethics approval**

10
11 4 Ethics approval was provided by the First Hospital Affiliated to the Xinxiang
12
13
14 5 Medical College.

15
16 **6 Provenance and peer review**

17
18
19 7 Not commissioned; externally peer reviewed.

20
21 **8 Data sharing statement**

22
23
24 9 No additional data are available.

25
26 **10 Figure Legend**

27
28
29 **Fig. 1** The LD plot of the fifteen SNPs in the ANXA11 gene. Values in squares are the
30
31 pair-wise calculation of r^2 (left) or D' (right). Black squares indicate $r^2 = 1$ (i.e. perfect
32
33 LD between a pair of SNPs). Empty squares indicate $D' = 1$ (i.e. complete LD
34
35 between a pair of SNPs).
36
37

38
39 **Supplemental Fig. 1** The structure of the human ANXA11 gene and 15 SNPs located
40
41 on the gene.
42

43
44 **17 Reference**

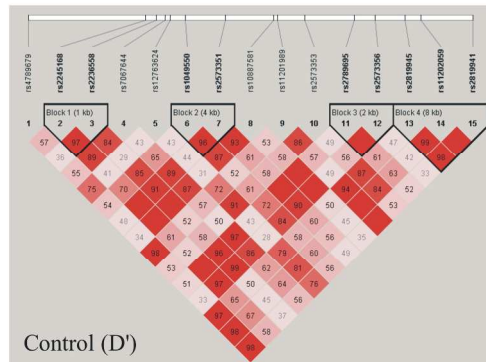
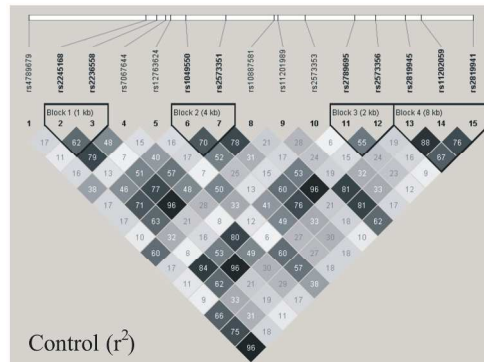
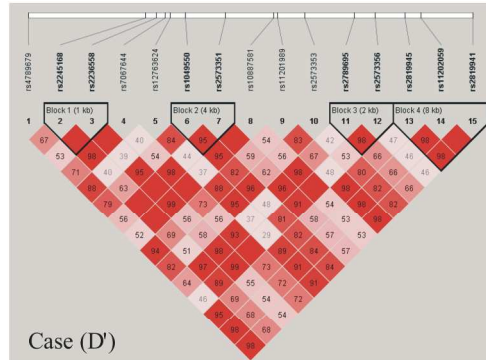
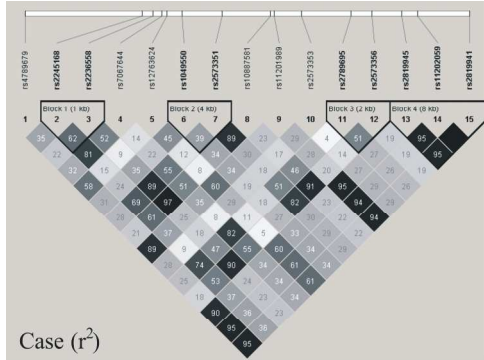
- 45
46 18 1. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med* 2007;
47
48 357:2153-65.
49
50 20 2. Baughman RP, Lower EE, du Bois RM. Sarcoidosis. *Lancet* 2003;361:1111-8
51
52
53 21 3. Grunewald J. Review: role of genetics in susceptibility and outcome of sarcoidosis.
54
55
56 22 *Semin Respir Crit Care Med* 2010;31:380-9.
57
58
59
60

- 1 4. Hofmann S, Franke A, Fischer A, *et al.* Genome-wide association study identifies
2 ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet*
3 2008;40:1103-6.
- 4 5. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev*
5 2002;82:331-71.
- 6 6. Farnaes L, Ditzel HJ. Dissecting the cellular functions of annexin XI using
7 recombinant human annexin XI-specific autoantibodies cloned by phage
8 display. *J Biol Chem* 2003;278:33120-6.
- 9 7. Tomas A, Futter C, Moss SE. Annexin 11 is required for midbody formation and
10 completion of the terminal phase of cytokinesis. *J Cell Biol* 2004;165:813-22.
- 11 8. Petzmann S, Maercker C, Markert E, *et al.* Enhanced proliferation and decreased
12 apoptosis in lung lavage cells of sarcoidosis patients. *Sarcoidosis Vasc Diffuse*
13 *Lung Dis* 2006;23:190-200
- 14 9. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, *et al.* Comparative
15 analysis of mycobacterial heat shock proteins-induced apoptosis of peripheral
16 blood mononuclear cells in sarcoidosis and tuberculosis. *J Clin Immunol*
17 2006;26:243-50.
- 18 10. Agostini C, Trentin L, Facco M, *et al.* Role of IL-15, IL-2, and their receptors in
19 the development of T cell alveolitis in pulmonary sarcoidosis. *J Immunol*
20 1996;157:910-8
- 21 11. Jullien D, Sieling PA, Uyemura K, *et al.* IL-15, an immunomodulator of T cell
22 responses in intracellular infection. *J Immunol* 1997;158:800-6

- 1
2
3
4 12. Muller-Quernheim J, Schurmann M, Hofmann S, *et al.* Genetics of sarcoidosis.
5
6 2 *Clin Chest Med* 2008;29:391-414.
7
8
9 3 13. Morais A, Lima B, Peixoto M, *et al.* Annexin A11 gene polymorphism (R230C
10
11 4 variant) and sarcoidosis in a Portuguese population. *Tissue antigens*
12
13 5 2013;82:186-91.
14
15
16 6 14. Li Y, Pabst S, Kubisch C, *et al.* First independent replication study confirms the
17
18 7 strong genetic association of ANXA11 with sarcoidosis. *Thorax*
19
20 8 2010;65:939-40.
21
22
23 9 15. Gabriel SB, Schaffner SF, Nguyen H, *et al.* The structure of haplotype blocks in
24
25 10 the human genome. *Science* 2002;296:2225-9.
26
27
28 11 16. Barrett JC, Fry B, Maller J, *et al.* Haploview: analysis and visualization of LD and
29
30 12 haplotype maps. *Bioinformatics* 2005;21:263-5.
31
32
33 13 17. Curtis D, Knight J, Sham PC. Program report: GENECOUNTING support
34
35 14 programs. *Ann Hum Genet* 2006;70:277-9.
36
37
38 15 18. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic
39
40 16 associations. *Hum Hered* 2000;50:133-9.
41
42
43 17 19. Zhao JH, Lissarrague S, Essioux L, *et al.* GENECOUNTING: haplotype analysis
44
45 18 with missing genotypes. *Bioinformatics* 2002;18:1694-5
46
47
48 19 20. Zhao JH, Sham PC. Generic number systems and haplotype analysis. *Comput*
49
50 20 *Methods Programs Biomed* 2003;70:1-9
51
52
53 21 21. Mrazek F, Stahelova A, Kriegova E, *et al.* Functional variant ANXA11 R230C:
54
55 22 true marker of protection and candidate disease modifier in sarcoidosis. *Genes*
56
57
58
59
60

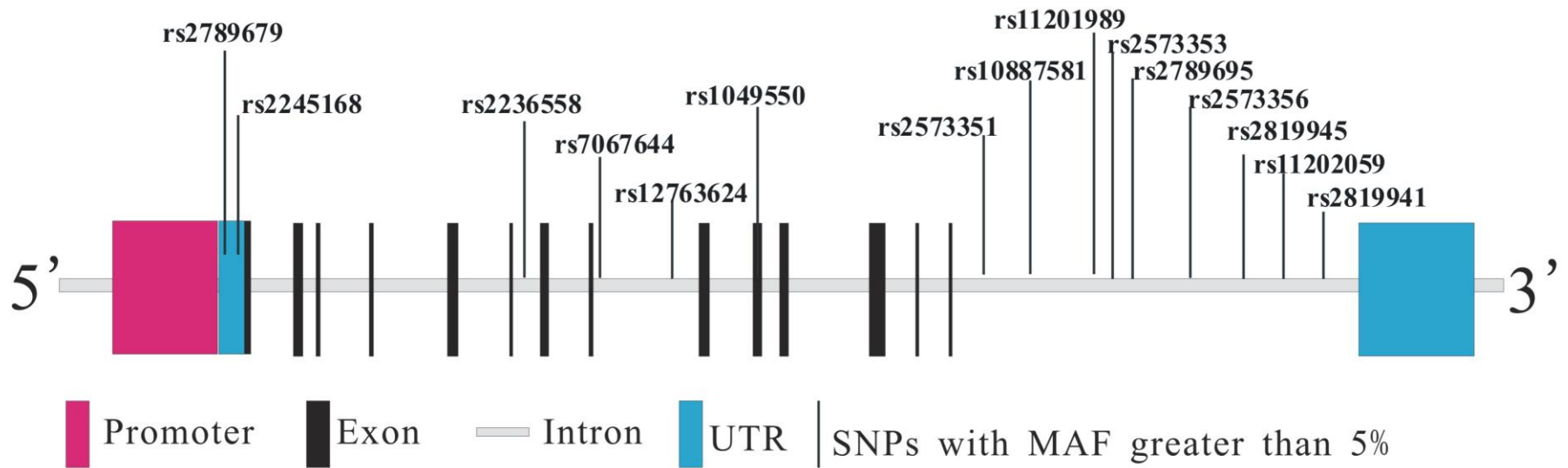
- 1
2
3
4 1 *Immun* 2011;12:490-4.
5
6 22. Levin AM, Iannuzzi MC, Montgomery CG, *et al.* Association of ANXA11 genetic
7
8 variation with sarcoidosis in African Americans and European Americans.
9
10 *Genes Immun* 2013;14:13-8.
11
12 24. Tomas A, Moss SE. Calcium- and cell cycle-dependent association of annexin 11
13
14 with the nuclear envelope. *J Biol Chem* 2003;278:20210-6.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



273x203mm (300 x 300 DPI)

Review only



review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

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

review only

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

Variable	Controls (n=418)		Sarcoidosis (n=412)		<i>P</i> -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	

1							
2							
3							
4							
5							
6							
7	TT	120	28.7	104	25.2	0.263	1.19, 0.88-1.62
8	CT	202	48.3	216	52.4	0.238	0.85, 0.65-1.12
9	CC	96	23.0	92	22.3	0.826	1.04, 0.75-1.44
10	Per C allele	394	47.1	400	48.5	0.564	1.06, 0.87-1.28
11	rs10887581					0.290	
12	TT	114	27.3	96	23.3	0.190	1.23, 0.90-1.69
13	TC	190	45.5	208	50.5	0.148	0.82, 0.62-1.07
14	CC	114	27.3	108	26.2	0.729	1.06, 0.78-1.44
15	Per C allele	418	50.0	424	51.5	0.583	1.06, 0.87-1.29
16	rs11201989					0.144	
17	GG	130	31.1	128	31.1	0.995	1.00, 0.75-1.34
18	GC	218	52.2	194	47.1	0.141	1.23, 0.93-1.62
19	CC	70	16.8	90	21.9	0.060	0.71, 0.50-1.01
20	Per C allele	358	42.8	374	45.4	0.29	1.11, 0.91-1.35
21	rs2573353					0.918	
22	CC	178	42.6	176	42.7	0.957	0.99, 0.75-1.31
23	CA	196	46.9	192	46.6	0.921	1.01, 0.77-1.33
24	AA	44	10.5	44	10.7	0.941	0.98, 0.63-1.53
25	Per A allele	284	34.0	280	34.0	0.997	1.00, 0.82-1.23
26	rs2789695					0.918	
27	TT	178	42.6	180	34.0	0.740	0.95, 0.73-1.26
28	CT	188	45.0	184	44.7	0.920	1.01, 0.77-1.33
29	CC	52	12.4	48	11.7	0.73	1.08, 0.71-1.64
30	Per C allele	292	34.9	280	34.0	0.69	0.96, 0.78-1.17
31	rs2573356					0.892	
32	CC	118	28.2	112	27.2	0.742	1.05, 0.78-1.43
33							
34							
35							
36							
37							
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							
48							
49							

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: rs2789679+rs1049550+rs2819941	rs1049550: T	0.0002	0.65	0.52-0.82

For peer review only



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

1			
2		assessing outcomes) and how	
3			
4		11b If relevant, description of the similarity of interventions	8
5	Statistical methods	12a Statistical methods used to compare groups for primary and secondary outcomes	8-9
6		12b Methods for additional analyses, such as subgroup analyses and adjusted analyses	8
7			
8	Results		
9	Participant flow (a	13a For each group, the numbers of participants who were randomly assigned, received intended treatment, and	10-11
10	diagram is strongly	were analysed for the primary outcome	
11	recommended)	13b For each group, losses and exclusions after randomisation, together with reasons	10-11
12	Recruitment	14a Dates defining the periods of recruitment and follow-up	10
13		14b Why the trial ended or was stopped	10
14			
15	Baseline data	15 A table showing baseline demographic and clinical characteristics for each group	10
16	Numbers analysed	16 For each group, number of participants (denominator) included in each analysis and whether the analysis was	11
17		by original assigned groups	
18			
19	Outcomes and	17a For each primary and secondary outcome, results for each group, and the estimated effect size and its	11
20	estimation	precision (such as 95% confidence interval)	
21		17b For binary outcomes, presentation of both absolute and relative effect sizes is recommended	11
22	Ancillary analyses	18 Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing	11
23		pre-specified from exploratory	
24			
25	Harms	19 All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	
26			
27	Discussion		
28	Limitations	20 Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	11-15
29	Generalisability	21 Generalisability (external validity, applicability) of the trial findings	12-14
30	Interpretation	22 Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	12
31			
32	Other information		
33	Registration	23 Registration number and name of trial registry	13
34	Protocol	24 Where the full trial protocol can be accessed, if available	14
35	Funding	25 Sources of funding and other support (such as supply of drugs), role of funders	15
36			

37

38 *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

39 recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials.

40 Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

41

42