Fine mapping of genetic polymorphisms of pulmonary tuberculosis within chromosome 18q11.2 in the Chinese population: a case-control study

Yaoyao Dai, Xia Zhang, Hongqiu Pan, Shaowen Tang, Hongbing Shen and Jianming Wang

Abstract

Background: Recently, one genome-wide association study identified a susceptibility locus of rs4331426 on chromosome 18q11.2 for tuberculosis in the African population. To validate the significance of this susceptibility locus in other areas, we conducted a case-control study in the Chinese population.

Methods: The present study consisted of 578 cases and 756 controls. The SNP rs4331426 and other six tag SNPs in the 100 Kbp up and down stream of rs4331426 on chromosome 18q11.2 were genotyped by using the Taqman-based allelic discrimination system.

Results: As compared with the findings from the African population, genetic variation of the SNP rs4331426 was rare among the Chinese. No significant differences were observed in genotypes or allele frequencies of the tag SNPs between cases and controls either before or after adjusting for age, sex, education, smoking, and drinking history. However, we observed strong linkage disequilibrium of SNPs. Constructed haplotypes within this block were linked the altered risks of tuberculosis. For example, in comparison with the common haplotype AA (rs8087945-rs12456774), haplotypes AG (rs8087945-rs12456774) and GA (rs8087945-rs12456774) were associated with a decreased risk of tuberculosis, with the adjusted odds ratio (95% confidence interval) of 0.34 (0.27-0.42) and 0.22 (0.16-0.29), respectively.

Conclusions: Susceptibility locus of rs4331426 discovered in the African population could not be validated in the Chinese population. None of genetic polymorphisms we genotyped were related to tuberculosis in the single-point analysis. However, haplotypes on chromosome 18q11.2 might contribute to an individual’s susceptibility. More work is necessary to identify the true causative variants of tuberculosis.
China, the world’s second largest country with TB epidemic, has a different genetic background, lifestyle, and disease prevalence from Africa [12]. It was estimated that 1.3 million people across the country developed active TB in 2009, of whom 600 000 had the highly infectious form. To validate the findings from the GWA study in Ghana and Gambia and to search for more susceptibility loci of TB, we performed a case-control study via the fine mapping analysis of the region of 18q11.2 in the Chinese population.

Methods
Study population
This case-control study was conducted in Jiangsu, a developed province located in the eastern part of China, with a total population of 77 million in 2009. We recruited 578 patients with pulmonary TB, including 368 (63.7%) new cases and 210 (36.3%) previously treated ones. For the definitions of new and previously treated cases, we referred to the WHO guidelines. In brief, a “new case” was termed as a newly registered episode of TB in a patient who, in response to direct questioning, denied having had any prior antituberculosis treatment (for up to one month); and in study sites where adequate documentation was available, there was no evidence of such history. A “previously treated case” was defined as a newly registered episode of TB in a patient who, in response to direct questioning admitted having been treated for TB for one month or more, or, in study sites where adequate documentation was available, there was evidence of such history. All patients were diagnosed with the evidence of sputum culture. Sputum samples were cultured on Lowenstein-Jensen (LJ) culture media. Identification of MTB was done by using the p-nitrobenzoic acid (PNB) and thiophene carboxylic acid hydrazine (TCH) resistance test. Growth in LJ medium containing PNB indicated that the bacilli did not belong to the MTB complex. We also recruited 756 controls from a pool of individuals who participated in the local community-based health examination program. Controls were frequency-matched to the cases by sex and age. These control subjects had no self-reported history of TB, diabetes and malignancy. All cases and controls had no prior HIV positive history. Each subject was individually interviewed in local health facilities by using a structured questionnaire and donated a blood sample for genotyping analysis.

SNPs selection and genotyping
The significant SNP rs4331426, which was identified in a GWA study in Ghana and Gambia, is located on the region of chromosome 18q11.2 [11]. Due to the low minor allele frequency (MAF) of this SNP (≤ 5%) among Han Chinese, we further searched for tag SNPs around it. Firstly, we downloaded all eligible SNPs in the 100 Kbp up and down stream of the SNP rs4331426 on chromosome 18q11.2 by using the Chinese Han population (CHB) database of HapMap http://www.hapmap.org/. All SNPs in this region were filtered by using the following criteria: (1) MAF≥0.05; (2) Hardy-Weinberg equilibrium test \( P \) value≥0.05. Then, tag SNPs were selected by using Haploview 4.2 software based on their ability to tag surrounding variants [13]. As a result, 7 SNPs (rs4330012, rs8087945, rs12456774, rs12457731, rs12958098, rs4800136 and rs4800417) were chosen for genotyping, which was performed by using TaqMan allelic discrimination technology on the ABI 7900 Real-Time PCR System (Applied Biosystems, Foster City, CA) [14]. The primers and probes for each SNP (Table 1) were designed by Nanjing Steed BioTechnologies Co., Ltd. Due to the technical limitation, we failed to design probes for detecting the SNP rs4330012 as another SNP rs9954441 neighboring to it. Preliminary experiments were carried out for each SNP and blank controls were set in each batch of samples. Both the laboratory personnel and the readers of genotyping were blinded to the status of cases and controls. The overall call rate of genotyping was > 95%.

Statistical analysis
Data were double entered with EpiData 3.1 (Denmark) and discrepancies were checked against the raw data. Continuous variables were described as mean ± SD and differences between groups were analyzed by using student-\( t \) test. Categorized variables were described as percentage and analyzed by using the Chi-square test. Unconditional logistic regression model was used to calculate odds ratio (OR) and 95% confidence interval (CI), as well as corresponding \( P \)-values. Hardy-Weinberg equilibrium was estimated using the \( \chi^2 \) goodness of fit test among controls. Haplotype blocks were selected with Haploview software by considering linkage disequilibrium (LD) blocks. The estimated frequency of polymorphic loci was calculated using PHASE 2.1 software. All analyses were performed using the SPSS software (SPSS Inc., USA). The \( P \)-value reported was two-sided and the values less than 0.05 were considered statistically significant.

Ethical consideration
This project has been approved by the Institutional Review Board of Nanjing Medical University. Written informed consents were obtained from all participants. Ethics has been respected throughout the whole study period.

Results
Overall, this study consisted of 578 cases (72.1% males, 27.9% females) and 756 controls (75.3% male, 24.7%
female). The age (mean ± SD) was 52.07 ± 18.01 years for cases and 52.85 ± 18.42 years for controls, respectively. As a result of frequency-matching, there were no significant differences in the distribution of gender and age between cases and controls. However, education level and the history of smoking and drinking were found to be different between the two groups. As shown in Table 2, the proportion of ever smoking was 52.5% in patients, which was significantly higher than that in controls (44.4%) ($\chi^2 = 8.543$, $P = 0.003$). In contrast, the proportion of alcohol drinking was 17.9% in TB patients, which was significantly lower than that in controls ($\chi^2 = 19.675$, $P < 0.001$).

As expected, genetic variants of rs4331426 were rare in the study population. The frequencies of AA, AG, and GG genotype were 93.70%, 6.14%, and 0.15%, respectively. No significant difference was observed in the distribution of either genotypes or alleles of this SNP between cases and controls. Six tag SNPs we genotyped were all in Hardy-Weinberg equilibrium [$\chi^2 = 0.330$, $P = 0.566$), rs12456774 ($\chi^2 = 0.438$, $P = 0.508$), rs12457731 ($\chi^2 = 0.236$, $P = 0.627$), rs12958098 ($\chi^2 = 0.757$, $P = 0.384$), rs4800136 ($\chi^2 = 0.047$, $P = 0.829$) and rs4800417 ($\chi^2 = 1.320$, $P = 0.251$)]. The genotype analysis showed that the minor allele frequencies of rs8087945, rs12456774, rs12457731, rs12958098, rs4800136 and rs4800417 were 22.41%, 29.96%, 5.10%, 47.14%, 3.91% and 19.34% in the cases and 22.00%, 29.93%, 4.37%, 48.99%, 4.03% and 18.66% in the controls, respectively. No significant difference was observed in genotypes or allele frequencies of the six tag SNPs between case and control groups either before or after adjusting for age, sex, education, smoking, and drinking history (Table 3). By considering both $D'$ and $r^2$, we analyzed two haplotypes as presented in the Table 4. For example, strong LD was observed between rs8087945 and rs12456774 ($D' = 1$, $r^2 = 0.12$). The common haplotype within this block was AA (rs8087945-rs12456774). Comparison of haplotype frequencies between case and control groups demonstrated that both the haplotype AG (rs8087945-rs12456774) and GA (rs8087945-rs12456774) were associated with the decreased risk of TB, with the adjusted OR(95% CI) of 0.34(0.27-0.42) and 0.22(0.16-0.29), respectively (Table 4). A strong linkage disequilibrium was also observed between rs4800417 and rs8087945 ($D' = 0.99$, $r^2 = 0.78$). As compared with the common haplotype CA (rs4800417-rs8087945), the haplotype TG (rs4800417-rs8087945) was associated with a decreased risk (aOR = 0.64, 95%CI: 0.50-0.81) whereas the haplotype CG (rs4800417-rs8087945) was related to an

Table 1 Information of primers and probes

| SNPs     | Primer sequence(5'–3')                  | Probe sequence                  |
|----------|----------------------------------------|---------------------------------|
| rs8087945| F-GGATGACAGAGTGACACTGACACACA           | C: FAM-AGAAGAAGAATCTGCCACTTT-MGB |
| A > G    | R-AGAAGACAGCAAGAAGTGCCAGGTG           | T: HEX-AGAAGAAGAATCTGCCACTTT-MGB |
| rs12456774| F-AAGTTGATACCCCAGCATAG                | C: FAM-CACACTTACCTCGGCTTTT-MGB  |
| A > G    | R-TGGAAGAAGTGCCATCTGAG                | T: HEX-CACACTTACCTCGGCTTTT-MGB  |
| rs12457731| F-ACCATCTAATGCCTGCTGAGAAGT            | G: FAM-CCTTATCCTAAATCTTACG-MGB  |
| A > G    | R-TAGATGCTTTCTCTCCAAGGCTCA            | A: HEX-CCTTATCCTAAATCTTACG-MGB |
| rs12958098| F-TCTGCAAAAGGCTCCAGAAGCTAT            | C: FAM-CAATTCCTGACAGT-MGB       |
| C > T    | R-GTAAACCAAGCCTACAGTTTCAATT          | G: HEX-CAATTCCTGACAGT-MGB       |
| rs4800136| F-CACCTGACCAACAAAGAAGGCGT             | C: FAM-CTCTTGGTCAATACCT-MGB     |
| A > C    | R-GACTGAAAGACAAAAGGCTCTC              | T: HEX-CTCTTGGTCAATACCT-MGB     |
| rs4800417| F-GTCCAGCTCTTGTGTTAGCT               | G: FAM-TCAACCTGACAGCG-MGB       |
| C > T    | R-TCCCACTTCTTACTTCTTTAGAA            | A: HEX-TCCCACTTCTTCTTTAGAA      |
| rs4331426| F-GGTTTTGATTAACAGCAGGCAAGAA           | G: FAM-TACCAATGCGACAG-MGB       |
| A > G    | R-ACACCTCTTGATAGTGAAAGAATTAAA         | A: HEX-TACCAATGCGACAG-MGB       |

Table 2 Basic characteristics of the cases and controls

| Variables     | case (N = 578) | control (N = 756) | P     |
|---------------|----------------|-------------------|-------|
| Age (mean ± SD) | 52.07 ± 18.01 | 52.85 ± 18.42     | 0.176 |
| Gender        |                |                   | 0.199 |
| Male          | 417 (72.1)     | 569 (75.3)        |       |
| Female        | 161 (27.9)     | 187 (24.7)        |       |
| Education     |                |                   | < 0.001|
| Illiterate or semi-illiterate | 164 (28.4) | 89 (11.8)        |       |
| Primary school | 133 (23)      | 215 (28.4)       |       |
| Junior high school | 201 (34.8) | 267 (35.3)       |       |
| Senior high school or above | 80 (13.8) | 185 (24.5)       |       |
| Smoking       |                |                   | 0.003 |
| Ever          | 290 (52.5)     | 336 (44.4)        |       |
| Never         | 270 (47.5)     | 420 (55.6)        |       |
| Drinking      |                |                   | < 0.001|
| Ever          | 101 (17.9)     | 215 (28.4)        |       |
| Never         | 463 (82.1)     | 541 (71.6)        |       |
increased risk of TB (OR = 3.19, 95%CI: 2.26-4.51) (Table 4).

Discussion
A puzzling feature of TB is that only a small proportion of infected persons will develop active diseases during their lifetimes [15], though nearly one third of global populations have been latently infected with the pathogen [2]. Host genetic factors can explain, at least in part, why some people resist infection more successfully than others [16,17]. Recently, a GWA study from Ghana and Gambia identified rs4331426 on the chromosome 18q11.2, as a susceptibility locus associating with the risk of TB [11]. Till now, this is the only GWA study relating to the susceptibility of TB, suggesting that a new non-MHC locus can be identified in an infectious

Table 3 Distribution of genotypes in cases and controls and their risks with pulmonary tuberculosis

| SNP              | Case (N) | Control (N) | cOR (95%CI) | P     | aOR (95%CI) a | P a | Model b OR (95%CI) | P |
|------------------|----------|-------------|-------------|--------|---------------|-----|--------------------|---|
| rs8087945 A > G | AA 338   | 459 61.20   | 1.04(0.82-1.31) | 0.742  | 1.04(0.81-1.33) | 0.763 | Dom 2.10(1.29) | 0.885 |
|                  | AG 193   | 252 33.60   | 1.01(0.61-1.67) | 0.970  | 0.89(0.52-1.51) | 0.660 | Add 0.99(0.82-1.21) | 0.946 |
|                  | GG 29    | 39 5.20     | 1.01(0.62-1.38) | 0.670  | 0.83(0.55-1.26) | 0.376 | Add 0.96(0.81-1.15) | 0.683 |
| rs12456774 A > G| AA 272   | 372 49.60   | 1.06(0.85-1.34) | 0.594  | 1.05(0.83-1.34) | 0.675 | Dom 1.01(0.80-1.27) | 0.942 |
|                  | AG 239   | 307 40.93   | 1.04(0.70-1.54) | 0.836  | 1.05(0.70-1.59) | 0.811 | Rec 0.81(0.54-1.21) | 0.299 |
|                  | GG 48    | 71 9.47     | 0.92(0.62-1.38) | 0.897  | 0.83(0.55-1.26) | 0.376 | Add 0.96(0.81-1.15) | 0.683 |
| rs12457731 A > G| AA 524   | 691 91.52   | 1.04(0.70-1.54) | 0.836  | 1.05(0.70-1.59) | 0.811 | Rec 0.81(0.54-1.21) | 0.299 |
|                  | AG 49    | 62 8.21     | 3.30(2.64-17.05) | 0.155  | 3.43(2.65-18.09) | 0.146 | Add 1.19(0.82-1.71) | 0.36 |
|                  | GG 5     | 2 0.36      | 1.01(0.80-1.29) | 0.885  | 1.02(0.80-1.29) | 0.885 | Dom 1.01(0.80-1.27) | 0.942 |

a: Adjusted for age, sex, education, smoking and drinking.
b: Dom, dominant model; Rec, recessive model; Add, additive mode.

Table 4 Haplotype frequencies and the risks of pulmonary tuberculosis

| Haplotypes        | Case (N) | Control (N) | cOR (95%CI) | P     | aOR (95% CI) a | P a | Model b OR (95%CI) | P |
|-------------------|----------|-------------|-------------|--------|---------------|-----|--------------------|---|
| rs8087945-rs12456774 | A/A 746   | 725(47.9)   | 1           | 1      | 1             | 1   | Dom 2.10(1.29) | 0.885 |
|                  | A/G 159   | 454(30.0)   | 0.34(0.28-0.42) | 0.0001 | 0.34(0.27-0.42) | 0.0001 | Add 0.99(0.82-1.21) | 0.946 |
|                  | G/G 75    | 333(22.0)   | 0.22(0.17-0.29) | 0.0001 | 0.22(0.16-0.29) | 0.0001 | Add 0.99(0.82-1.21) | 0.946 |
| rs4800417-rs8087945 | C/C 176   | 16(15.2)    | 0.994       | 0.994  | 0.994         | 0.994 | Add 0.99(0.82-1.21) | 0.946 |

a: Adjusted for age, sex, education, smoking and drinking.
disease caused by a highly polymorphic pathogen even in African populations [11]. The identified variant of SNP rs4331426 is common in the African, but is much rarer in other populations. No data of this SNP have been published yet in association with TB from other areas of the world. Considering the limitation of extensive genetic diversity and shorter LD ranges in African populations, we performed a study to validate this finding in China by searching for tag SNPs on the chromosome 18q11.2 in the 100 Kbp up and down stream of the SNP rs4331426. To our knowledge, since the publication of the GWA study by Thye et al [11], our work is the first one to explore the role of genetic polymorphisms in this region on the susceptibility to TB. Unfortunately, we observed no significant association between TB risk and selected SNPs individually. One possible explanation might be the heterogeneity of populations, which can be confirmed by the disparity of genotype frequency [18]. Another explanation was that we only detected tag SNPs on the chromosome 18q11.2 within 100 Kbp in the up and down stream of the SNP rs4331426, which could only represent a relatively narrow scope of the genetic loci. Even though none of polymorphisms we investigated were associated with TB in the single-point locus analysis, we found the haplotypes within this block might be associated with the altered risks of TB. For example, compared to individuals carrying the common haplotype A_r8087945A_r12456774, those with A_r8087945G_r12456774 or G_r8087945A_r12456774 had a significantly decreased risk. We should notice that in this study we only analyzed two haplotypes. Other haplotypes covering more SNPs might also contribute to the risk of TB.

Interestingly, chromosome 18q11.2 is a gene-desert region that is punctuated by evolutionarily conserved domains with regulatory potential [11]. Neither rs8087945 nor rs12456774 is located inside any gene or in the regulatory sequence. The nearest genes to these SNPs are GATA6, CTAGE1, RBBP8 and CABLES1, as well as a number of as yet unannotated open reading frames. Additional studies are required to ascertain their functional significance and any possible counterbalancing selective pressures. In addition, it must be noted that the association found in China could be population-specific; however, it could also be a false-positive result. For this reason, it is important that these findings should be replicated to confirm the association in other areas of China. Future work is needed to explore the nearest genes as well as a number of as yet unannotated open reading frames around this region.

Conclusions
Susceptibility locus of rs4331426 identified in the African population could not be validated in the Chinese population. Even though none of genetic polymorphisms we investigated was associated with TB in the single-point analysis, the haplotypes might contribute to the susceptibility to TB in the Chinese Han population. Additional studies are required to ascertain the causative variant, its functional significance and any possible counterbalancing selective pressures.

Additional material

Acknowledgements
This study is partly supported by National Natural Science Foundation of China (81072351), National S&T Major Project Foundation of China (2011ZX10004-902), Jiangsu social development project Foundation (BE2011841), and Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Author details
1Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, 210029 PR China. 2Department of Tuberculosis, Nanjing Chest Hospital, Nanjing, 210029 PR China. 3Department of Tuberculosis, Third Hospital of Zhenjiang City, Zhenjiang, 212005 PR China.

Authors' contributions
JW, ZX, HP, ST, and HS implemented the field study. JW and YD performed laboratory tests. YD and JW participated in the statistical analysis and drafted the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 8 March 2011 Accepted: 22 October 2011 Published: 22 October 2011

References
1. Maarten G, Wilkinson RJ: Tuberculosis, Lancet 2007, 370:2030-2043.
2. Young OB, Perkins MD, Duncan K, Barry CE: Confronting the scientific obstacles to global control of tuberculosis. J Clin Invest 2008, 118:1255-1265.
3. Bucher HC, Griffith LE, Guyatt GH, Sudre P, Naef M, Sendi P, Battegay M: Isoniazid prophylaxis for tuberculosis in HIV infection: a meta-analysis of randomized controlled trials. AIDS 1999, 13:501-507.
4. Girardi E, Raviglione MC, Antonucci G, Godfrey-Faussett P, Ippolito G: Impact of the HIV epidemic on the spread of other diseases: the case of tuberculosis. AIDS 2000, 14(Suppl 3):S47-S56.
5. Selwyn PA, Hartel D, Lewis VA, Schoenbaum EE, Vermund SH, Klein RS, Walker AT, Friedland GH: A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection. N Engl J Med 1989, 320:545-550.
6. Cornstock GW: Tuberculosis in twins: a re-analysis of the Prophilt survey. Am Rev Respir Dis 1978, 117:621-624.
7. Stead WW, Senner JW, Reddick WF, Lofgren JP: Racial differences in susceptibility to infection by Mycobacterium tuberculosis. N Engl J Med 1990, 322:422-427.
8. Wang J, Tang S, Shen H: Association of genetic polymorphisms in the IL12-IFNG pathway with susceptibility to and prognosis of pulmonary tuberculosis in a Chinese population. Eur J Clin Microbiol Infect Dis 2010, 29:1291-1295.
9. Wang TH, Wang HS: A genome-wide association study primer for clinicians. Taiwan J Obstet Gynecol 2009, 48:89-95.

10. de Bakker PI, Telenti A: Infectious diseases not immune to genome-wide association. Nat Genet 42:731-732.

11. Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, Siugó G, Siay-Joof F, Enimil A, Chinbuah MA, et al: Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat Genet 42:739-741.

12. Wang L, Liu J, Chin DP: Progress in tuberculosis control and the evolving public-health system in China. Lancet 2007, 369:691-696.

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

14. Teuber M, Wenz MH, Schreiber S, Franke A: GMFilter and SXTtestPlate: software tools for improving the SNPlex genotyping system. BMC Bioinformatics 2009, 10:81.

15. Bellamy R: Susceptibility to mycobacterial infections: the importance of host genetics. Genes Immun 2003, 4:4-11.

16. Yim JJ, Selvaraj P: Genetic susceptibility in tuberculosis. Respirology 2010, 15:241-256.

17. Pantelidis P: Tuberculosis: an ancient disease still confusing our genes. Respiration 2005, 72:347-348.

18. Ansari A, Talat N, Jamil B, Hasan Z, Razzaki T, Dawood G, Hussain R: Cytokine gene polymorphisms across tuberculosis clinical spectrum in Pakistani patients. PLoS One 2009, 4:e4776.

Pre-publication history
The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1471-2334/11/282/prepub

doi:10.1186/1471-2334-11-282
Cite this article as: Dai et al.: Fine mapping of genetic polymorphisms of pulmonary tuberculosis within chromosome 18q11.2 in the Chinese population: a case-control study. BMC Infectious Diseases 2011 11:282.