Article type: Original Article

Common Polymorphisms in the *TLR4* Gene are Associated with Susceptibility to Pulmonary Tuberculosis in Sudanese

H. Y. Zaki,[*†] K. H. Leung,[†] W. C. Yiu,[†] N. Gasmelseed,[*] N.E.M. Elwali,[*] S. P. Yip[†]

[*]Molecular Biology Department, National Cancer Institute, University of Gezira, Wad Medani, Sudan; [†]Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong SAR, China.

Correspondence to: Shea Ping Yip, Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China. Tel: (+852) 3400 8571. Fax: (+852) 2362 4365. e-mail: shea.ping.yip@inet.polyu.edu.hk

Running head: *TLR4* gene and pulmonary tuberculosis

Word count: 2,547 words

Keywords: Toll-like receptor 4, case-control study, single nucleotide polymorphism, genetic predisposition

ABSTRACT

SETTING: Host genetic risk factors influence the susceptibility to tuberculosis (TB). There is ample evidence supporting the involvement of toll-like receptor 4 (TLR4) in mycobacterial infection.

OBJECTIVE: To study the relationship between the *TLR4* gene and TB susceptibility in Sudanese.

DESIGN: A case-control study was conducted by recruiting 207 patients with pulmonary TB and 395 healthy controls. Ten tag single nucleotide polymorphisms (SNPs) of the *TLR4* gene were genotyped using restriction digestion or hybridization assays, and analyzed.

RESULTS: The genotypes were in Hardy-Weinberg equilibrium. With Mantel-Haenszel tests controlled for gender, four SNPs showed significant differences between cases and controls even after correction of multiple comparisons by Bonferroni procedure. The Mantel-Haenszel estimates of allelic odds ratios for the high-risk alleles were 1.67 for rs1927911 (*P*=0.0001), 1.85 for rs5030725 (*P*=0.0008), 2.14 for rs7869402 (*P*=1.87e-07), and 2.31 for rs1927906 (*P*=1.23e-10). Haplotype analysis showed that rs1927911 and rs5030725 were in one haplotype block, and rs7869402 and rs1927906 were in another haplotype block. Conditional haplotype analysis suggests the presence of one causal variant downstream of a recombination hotspot at the 3' region of the *TLR4* gene.

CONCLUSION: This is the first study to show that common TLR4 polymorphisms are associated with TB susceptibility in a Sudanese population.

(Abstract: 203 words)

INTRODUCTION

Tuberculosis (TB) is a major cause of illness and death worldwide. In 2008, World Health Organization (WHO) estimated 8.9-9.9 million new cases and 1.1-1.7 million deaths from TB globally.[1] In Africa, over four million people suffer from active TB resulting in ~650,000 deaths annually.[2] Sudan ranks among countries with the highest TB prevalence with 154,933 tuberculous patients and 27,450 deaths in 2007.[1] The re-emergence of TB can be attributed to multidrug-resistant strains of *Mycobacterium tuberculosis* (*Mtb*), and epidemics of human immunodeficiency virus (HIV) infection.[3] One third of the world's population is infected with *Mtb*, but only 5-10% of the infected individuals will develop clinical signs of tuberculosis.[4] Genetic and non-genetic factors of both the bacterium and the host influence the host immune response to *Mtb*.[5] Many host susceptibility genes have been identified in the last decade.[6]

Toll-like receptors (TLRs) can recognize specific conserved microbial components.[7] Upon binding of ligands, TLRs trigger signal transduction that results in the expression of pro-inflammatory cytokines and the functional maturation of antigen presenting cells of the innate immune system, and also prepares for the adaptive immunity. Several TLRs, particularly TLR2, TLR4 and TLR9, are known to be important in mycobacterial infections.[8]

Many *in vivo* studies testify the protective role of TLR4 in the host defence against mycobacterial infections[9-11] although there are also conflicting results.[12-14] Employing mice as the host, these studies differed in terms of the type of mutant mice used, the mycobacterial species used, inoculation dose, route of infection and methods of assessing the infection. Housing conditions of animals and the handling protocols have also been suggested to affect the immunoregulatory responses of the animals.[12-14] While the genuine causes of discrepant results remain to be determined, there are still many other *in vitro* studies supporting the role of TLR4 in mycobacterial infections. Many ligands of mycobacterial origin have been found to mediate TLR4-dependent activation,[15] e.g., different acylated forms of lipomannans purified from mycobacteria.[16] Moreover, TLR4 is a known sensor for autophagy[17] – an important mechanism in the immunity again *Mtb*.[18] The TLR4-mediated

signaling is also crucial in maintaining the balance between apoptosis vs. necrosis of *Mtb*-infected macrophages.[19] Overall, there exists ample evidence supporting the functional role of TLR4 in mycobacterial infections.

We therefore hypothesized that common polymorphisms in the human *TLR4* gene might predispose humans to TB. We undertook a case-control study to investigate the relationship between common *TLR4* single nucleotide polymorphisms (SNPs) and the susceptibility to pulmonary TB in an African population. The *TLR4* gene (GeneID = 7099; 9q33.1) spans a genomic region of ~13.3 kb, has three exons and encodes a protein of 839 amino acids (NCBI; http://www.ncbi.nlm.nih.gov/). TLR4 is expressed in many tissues, and particularly at high levels in monocytes/macrophages.[7]

MATERIALS AND METHODS

Subjects

The study was approved by IRB of both universities. Cases were Sudanese patients with pulmonary TB and were recruited from the Chest Department at Wad Medani Teaching Hospital in the Gezira State of Sudan. Diagnosis was based on clinical presentation and sputum smear positive of acid-fast bacilli by the Ziehl-Neelsen method on *all three* consecutive sputum specimens; our case definition was within those adopted by WHO.[20] Controls were healthy Sudanese from the same region (blood donors or staff members of the University of Geriza) with normal chest radiograph and no past/family history of TB; use of community controls is common in such studies.[21,22] Controls were not tested by tuberculin skin test because its interpretation was complicated by Bacillus Calmette-Guérin (BCG) vaccination and BCG coverage rate was ~70% in Sudan.[23] In fact, most of the cases and the controls were BCG scar positive. Patients and controls were Nilo-Saharan speaking individuals residing in Wad Medani, had similar socioeconomic status and the same major ethnic category (Nilotic). All subjects gave written informed consent. Subjects with diabetes, HIV infection or taking immunosuppressive drugs were excluded. All subjects were screened for HIV1/2 with Enzygnost HIV

Integral II kit (Dade Behring). Whole blood samples were collected from the subjects. Genomic DNA was extracted by a salting-out method[24] or the FlexiGene DNA Kit (Qiagen).

SNP selection and genotyping

Tag SNPs from a 19-kb genomic region encompassing the *TLR4* locus and 3 kb upstream and downstream of the gene (Figure 1), based on the HapMap data for African subjects (release 23a/phaseII Mar08; http://hapmap.ncbi.nlm.nih.gov/index.html.en). Note that flanking sequences were included to capture SNPs in potential regulatory regions. The online Tagger software[25] was used for tag SNP selection with the pairwise tagging algorithm, $r^2 \geq 0.8$ and minor allele frequency (MAF) ≥ 0.2. The choice of this MAF served to strike a balance for the following factors: the number of SNPs to be genotyped, the available sample size and a reasonable power of the study (at least 80%). The SNPs were genotyped using restriction fragment length polymorphism analysis or hybridization assay (details provided as an online appendix).

Statistical analysis

Hardy-Weinberg equilibrium (HWE) was tested for genotypes of cases and controls separately by exact tests as executed in Haploview[26] (version 4.2; http://www.broad.mit.edu/mpg/haploview/). Genetic association analysis was performed with the GenAssoc program[27] (http://www-gene.cimr.cam.ac.uk/clayton/software/) executed with the statistical package STATA (version 8.2; StataCorp). Allele and genotype frequencies were compared between cases and controls by Mantel-Haenszel χ^2 tests controlled for gender. Since the genetic models involved were not known, only the additive model was also tested by the trend test, which is more conservative but does not require the assumption of HWE among the genotypes.[28] Mantel-Haenszel estimates of gender-adjusted odds ratios (ORs) and the 95% confidence intervals were calculated as appropriate. Bonferroni procedure was applied for correction of multiple comparisons. In order to have an overall significance level (α') of 0.05, the significance level (α) for each individual comparison was obtained from $\alpha = 1 - (1 - 0.05)^{1/30}$, and found to be 1.708×10^{-3}. Haplotype blocks were constructed using Haploview as defined

by Gabriel's definition.[26] Haplotype frequencies were estimated by an accelerated expectation-maximization algorithm and then compared between cases and controls. Multiple comparisons of haplotype frequencies were corrected by generating empirical *P* values via 100,000 permutations.

RESULTS

Characteristics of subjects

The main characteristics of the study groups are shown in Table 1.

Single-marker analysis

Ten tag SNPs were selected from the HapMap database for Africans (Figure 1). For easy reference, these 10 SNPs were also designated as S01 to S10 in the 5' to 3' direction (Table 2). The genotypes were all in HWE except for S08 (rs1927906; $P = 0.0454$) in the control group. This is not unexpected at a significance level of 0.05 because we expect one significant result due to random chance in 20 tests for HWE as in our case. As such, S08 was still included for analysis.

There were significantly fewer males in the case group than in the control group (70.5% vs 89.9%; $P < 0.0001$). To avoid potential confounding due to gender, comparison between cases and controls was performed by stratification into males and females and the results were combined by Mantel-Haenszel procedure. Four SNPs showed significant differences between cases and controls in allele and genotype frequencies and under an additive genetic model even at Bonferroni-adjusted significance level of 0.0017 (Table 2): S03 (rs1927911), S06 (rs5030725; borderline for genotype frequency comparison, $P = 0.0027$), S07 (7869402) and S08 (rs1927906). Among these four SNPs, the Mantel-Haenszel ORs was lowest for S03 and highest for S08. The high-risk allele was the less common allele for three SNPs (G for S06, T for S07 and C for S08), but was the more frequent allele for S03 (A). Note that ORs are expressed as greater than 1.00 in Table 2 to show the greater risks in the corresponding alleles or genotypes.

Two SNPs (S04 and S09) showed significant differences between cases and controls in at least two scenarios at the nominal *P* value of 0.05, but did not survive the correction of multiple testing (Table 2). The remaining 4 SNPs (S01, S02, S05 and S10) did not differ significantly between cases and controls in all scenarios even at the nominal cut-off *P* value of 0.05.

Haplotype blocks and haplotype analysis

Two haplotype blocks were constructed for the ten tag SNPs by Haploview as defined by Gabriel's definition.[25] Block 1 consisted of the six SNPs (S01 to S06) at the 5' half of the *TLR4* gene (Figure 1). Block 2 contained three SNPs (S07 to S09) at the 3' untranslated region and 3' flanking region of the gene. Haplotype frequencies were compared between cases and controls for the two blocks separately (Table 3). For block 1, five haplotypes differed significantly in frequencies between cases and controls at the nominal cut-off *P* value of 0.05, but only two (TTGCAT and TTACAG, see Table 3 for the meaning of underlined bases) survived the correction of multiple comparisons by permutation. The haplotype TTGCAT was present at a lower frequency in cases than in controls (0.318 vs 0.427) and hence protective in nature. The haplotype TTACAG was present at a higher frequency in cases than in controls (0.176 vs 0.105) and hence high-risk in nature. For block 2, four haplotypes were significantly different between cases and controls, but only two (CTT and TCT) were still significant after correction of multiple testing (Table 3). By the same token, the haplotype CTT was protective while the haplotype TCT was high-risk.

As defined by Gabriel's definition,[26] block 2 did not include S10 (Figure 1). In fact, forced inclusion of S10 in block 2 did not change the frequencies and the chi-squared statistics of the two commonest haplotypes that were associated with TB susceptibility (data not shown). This indicates that S10 did not add extra information to the associated haplotypes of S07-S08-S09.

DISCUSSION

This study selected tag SNPs based on the HapMap data for Yoruban from Nigeria, West Africa. These tag SNPs fall within two haplotype blocks (data not shown), which are basically similar to those of the Sudanese population in Eastern Africa under study (Figure 1B). Two SNPs (S03 and S06) in haplotype block 1 (5' end of the gene) were associated with pulmonary TB, and two SNPs (S07 and S08) in haplotype block 2 (3' end of the end) also demonstrated association with the disease phenotype (Tables 2 and 3, and Figure 1).

Our finding raises the question of whether all four associated SNPs were in strong linkage disequilibrium (LD) with another *single* causal SNP somewhere in the gene or whether the positive signals in these two haplotype blocks implies the presence of *two different* causal variants in the gene. Pertinent to this argument is the presence of the intragenic recombination hotspot that spans a genomic region of ~4 kb, encompasses exon 3 of the gene (Figure 1A), and separates the two haplotype blocks (Figure 1B). The recombination rate is ~6 cM/Mb in this region and is higher than the average 1 cM/Mb, and this recombination hotspot is *not* unique to the African populations.

We explore these two possibilities via sex-adjusted conditional haplotype-based association analysis using the Plink package.[29] We perform haplotype analysis across *nine* SNPs from S01 (rs10116253) to S09 (rs7044464) with or without controlling for associated SNPs or haplotype blocks (Table 3 and Figure 1B). Expectedly, significant haplotype association is obtained (P=1.74e-06) if *not* controlling for any SNPs or haploptypes. The haplotype association remains significant even if controlling for S03 and S06 (P=0.0003) or haplotype block 1 (S01 to S06, P=0.0013). However, haplotype association disappears if controlling for S07 and S08 (P=0.072) or haplotype block 2 (S07 to S09, P=0.071). This indicates that S07, S08 or haplotype block 2 is tagging an independent association signal probably due to a *single* causal SNP. On the other hand, the association signal produced by S03, S06 or haplotype block 1 is not independent and is probably due to the signal from haplotype block 2. This probably also explains the much higher P values for association with S03, S06 and haplotype block 1 (Tables 2 and 3). In summary, our findings suggest that there is *one causal*

TLR4 variant associated with susceptibility to pulmonary TB, and this causal variant is downstream of the recombination hotspot.

A recent study did not find any association of *TLR4* SNPs with pulmonary TB in an Indonesian population.[30] This study investigated 13 SNPs: the MAF was <9% for six SNPs, and >10% for the remaining seven SNPs in Indonesians. For the six less common SNPs, the study might not be powerful enough to detect association, if any. Of the seven common SNPs, four were not genotyped in our current study: two had an MAF of 4% or less *in Africans*, and two had an MAF of ~15% and an $r^2 \leq 0.41$ with S03 *in Africans*. The remaining three common SNPs were genotyped in the present study: S01, S03 and S04. Since S03 (rs1927911) was found to be associated with TB in our African study, there seems to be discrepancy here. Ethnic difference in genetic susceptibility genes could be one of the reasons. Alternatively, this may indicate that the genuine causal SNP is in strong LD with S03 and S06 and is on a chromosome background defined by the associated haplotypes in our study. That this signal was not detected by S03 in the Indonesian study might be due to a much weaker LD between S03 and the causal SNP in this population as a result of a different population history. Further investigation is needed to confirm this hypothesis. Intriguingly, there is a functional SNP located 100 bp upstream of S03: rs41426344 was found to alter gene expression in luciferase-based reporter gene assays.[31,32] We genotyped this SNP for our African samples (n=602). It turned out that this SNP was monomorphic in our African samples.

We did not investigate the role of the less common non-synonymous polymorphisms, particularly rs4986790 (Asp299Gly) and rs4986791 (Thr399Ile), of the *TLR4* gene in TB susceptibility. Ten studies have addressed this issue in twelve population groups.[21,22,30,33-39] Except an Asian Indian study of pulmonary TB and a Spanish study of TB in HIV-infected patients,[36,38] all the remaining eight studies did not support a role of the less common or rare non-synonymous *TLR4* polymorphisms in TB susceptibility. It is worth noting that exon 3 is the largest exon of the gene and harbors most of coding SNPs identified,[35] and that exon 3 is within the recombination hotspot (Figure 1). Therefore, irrespective of the allele frequencies, coding SNPs in this region would not be tagged

efficiently by other SNPs outside the hotspot. In particular, none of the associated SNPs are in strong LD with the known functional SNPs in this region. For instance, the r^2 values are <0.04 between rs4986790 (Asp299Gly) and any one of the associated SNPs reported here. In addition, rs4986790 (Asp299Gly) and rs4986791 (Thr399Ile) are in very weak LD ($r^2 < 0.07$) with the associated SNP S08 (rs1927906) in African populations as recorded in the HapMap database.

In summary, we systematically investigated ten tag SNPs of the *TLR4* gene and its immediate flanking regions, and identified four SNPs to be associated with susceptibility to TB in Sudanese. Our results also suggest the presence of one causal variant in 3' flanking region of the gene and downstream of the recombination hotspot encompassing exon 3 of the gene. This is the first study reporting the association of *TLR4* common polymorphisms with TB susceptibility in a Sudanese population. Replication in independent sample sets from different populations is needed to confirm our initial positive findings.

ACKNOWLEDGEMENTS

We thank all the subjects, technicians and clinicians who participated in the study. We are grateful to colleagues at the Department of Health Technology and Informatics, The Hong Kong Polytechnic University, and the National Cancer Institute, University of Gezira, for providing help whenever necessary. The present study would not have been possible without the financial support from The Hong Kong Polytechnic University, and the Ministry of Higher Education and Scientific Research in Sudan.

REFERENCES

1 World Health Organization. Global tuberculosis control: epidemiology, strategy, financing. WHO report 2009. Geneva, Switzerland: WHO Press, 2009.

2 Kaufmann SH, Parida SK. Tuberculosis in Africa: learning from pathogenesis for biomarker identification. Cell Host Microbe 2008; 4: 219-228.

3 Wells CD, Cegielski JP, Nelson LJ, et al. HIV infection and multidrug-resistant tuberculosis: the perfect storm. J Infect Dis 2007; 196(Suppl 1): S86-S107.

4 Harries AD, Dye C. Tuberculosis. Ann Trop Med Parasitol 2006; 100: 415-431.

5 Schurr E. Is susceptibility to tuberculosis acquired or inherited? J Intern Med 2007; 261: 106-111.

6 Lykouras D, Sampsonas F, Kaparianos A, et al. Human genes in TB infection: their role in immune response. Monaldi Arch Chest Dis 2008; 69: 24-31.

7 Takeda K, Kaisho T, Akira S. Toll-like receptors. Annu Rev Immunol 2003; 21: 335-376.

8 Jo EK, Yang CS, Choi CH, Harding CV. Intracellular signalling cascades regulating innate immune responses to Mycobacteria: branching out from Toll-like receptors. Cell Microbiol 2007; 9: 1087-1098.

9 Abel B, Thieblemont N, Quesniaux VJ, et al. Toll-like receptor 4 expression is required to control chronic Mycobacterium tuberculosis infection in mice. J Immunol 2002; 169: 3155-3162.

10 Fremond CM, Nicolle DM, Torres DS, Quesniaux VF. Control of Mycobacterium bovis BCG infection with increased inflammation in TLR4-deficient mice. Microbes Infect 2003; 5: 1070-1081.

11 Branger J, Leemans JC, Florquin S, et al. Toll-like receptor 4 plays a protective role in pulmonary tuberculosis in mice. Int Immunol 2004; 16: 509-516.

12 Shim TS, Turner OC, Orme IM. Toll-like receptor 4 plays no role in susceptibility of mice to Mycobacterium tuberculosis infection. Tuberculosis (Edinb) 2004; 83: 367-371.

13 Shi S, Blumenthal A, Hickey CM, et al. Expression of many immunologically important genes in Mycobacterium tuberculosis-infected macrophages is independent of both TLR2 and TLR4 but dependent on IFN-alphabeta receptor and STAT1. J Immunol 2005; 175: 3318-3328.

14 Holscher C, Reiling N, Schaible UE, et al. Containment of aerogenic Mycobacterium tuberculosis infection in mice does not require MyD88 adaptor function for TLR2, -4 and -9. Eur J Immunol 2008; 38: 680-694.

15 Jo EK. Mycobacterial interaction with innate receptors: TLRs, C-type lectins, and NLRs. Curr Opin Infect Dis. 2008;21:279-86.

16 Doz E, Rose S, Nigou J, et al. (2007) Acylation determines the toll-like receptor (TLR)-dependent positive versus TLR2-, mannose receptor-, and SIGNR1-independent negative regulation of pro-inflammatory cytokines by mycobacterial lipomannan. J Biol Chem 282: 26014-26025.

17 Xu Y, Jagannath C, Liu XD, et al. Toll-like receptor 4 is a sensor for autophagy associated with innate immunity. Immunity 2007; 27: 135-144.

18 Deretic V, Delgado M, Vergne I, et al. Autophagy in immunity against mycobacterium tuberculosis: a model system to dissect immunological roles of autophagy. Curr Top Microbiol Immunol 2009; 335: 169-188.

19 Sanchez D, Rojas M, Hernandez I, et al. Role of TLR2- and TLR4-mediated signaling in Mycobacterium tuberculosis-induced macrophage death. Cell Immunol 2010; 260: 128-136.

20 World Health Organization, International Union Against Tuberculosis and Lung Disease, Royal Netherlands Tuberculosis Association. Revised international definitions in tuberculosis control. Int. J. Tuberc Lung Dis 2000; 5, 213-215.

21 Newport MJ, Allen A, Awomoyi AA, et al. The toll-like receptor 4 Asp299Gly variant: no influence on LPS responsiveness or susceptibility to pulmonary tuberculosis in The Gambia. Tuberculosis (Edinb) 2004; 84: 347-352.

22 Xue Y, Zhao ZQ, Wang HJ, et al. Toll-like receptors 2 and 4 gene polymorphisms in a southeastern Chinese population with tuberculosis. Int J Immunogenet 2010; 37: 135-138.

23 Al-Jahdali H, Memish ZA, Menzies D. The utility and interpretation of tuberculin skin tests in the Middle East. Am J Infect Control 2005; 33: 151-156.

24 Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988; 16: 1215.

25 de Bakker PI, Yelensky R, Pe'er I, et al. Efficiency and power in genetic association studies. Nat Genet 2005; 37: 1217-1223.

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

27 Cordell HJ, Clayton DG. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. Am J Hum Genet 2002; 70: 124-141.

28 Balding DJ. A tutorial on statistical methods for population association studies. Nat Rev Genet 2006; 7: 781-791.

29 Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559-575.

30 Davila S, Hibberd ML, Hari Dass R, et al. Genetic association and expression studies indicate a role of toll-like receptor 8 in pulmonary tuberculosis. PLoS Genet 2008; 4: e1000218.

31 Song C, Chen LZ, Zhang RH, Yu XJ, Zeng YX. Functional variant in the 3'-untranslated region of Toll-like receptor 4 is associated with nasopharyngeal carcinoma risk. Cancer Biol Ther 2006; 5: 1285-1291.

32 Duan ZX, Zhu PF, Dong H, et al. Functional significance of the TLR4/11367 polymorphism identified in Chinese Han population. Shock 2007; 28: 160-164.

33 Ferwerda B, Kibiki GS, Netea MG, Dolmans WM, van der Ven AJ. The toll-like receptor 4 Asp299Gly variant and tuberculosis susceptibility in HIV-infected patients in Tanzania. AIDS 2007; 21: 1375-1377.

34 Fitness J, Floyd S, Warndorff DK, et al. Large-scale candidate gene study of tuberculosis susceptibility in the Karonga district of northern Malawi. Am J Trop Med Hyg 2004; 71: 341-349.

35 Ma X, Liu Y, Gowen BB, et al. (2007) Full-exon resequencing reveals toll-like receptor variants contribute to human susceptibility to tuberculosis disease. PLoS One 2007; 2: e1318.

36 Najmi N, Kaur G, Sharma SK, Mehra NK. Human Toll-like receptor 4 polymorphisms TLR4 Asp299Gly and Thr399Ile influence susceptibility and severity of pulmonary tuberculosis in the Asian Indian population. Tissue Antigens 2010; 76:102-109.

37 Olesen R, Wejse C, Velez DR, et al. DC-SIGN (CD209), pentraxin 3 and vitamin D receptor gene variants associate with pulmonary tuberculosis risk in West Africans. Genes Immun 2007; 8:456-467.

38 Pulido I, Leal M, Genebat M, et al. The TLR4 ASP299GLY polymorphism is a risk factor for active tuberculosis in Caucasian HIV-infected patients. Curr HIV Res 2010; 8: 253-258.

39 Selvaraj P, Harishankar M, Singh B, Jawahar MS, Banurekha VV. Toll-like receptor and TIRAP gene polymorphisms in pulmonary tuberculosis patients of South India. Tuberculosis (Edinb) 2010; 90: 306-310.

FIGURE LEGEND

Figure 1. (A) The *TLR4* gene: recombination rate, gene structure and tag SNPs. The recombination rate (cM/Mb) across a 19kb region encompassing the gene is shown above the gene, and the locations of the tag SNPs investigated in this study are shown below the gene. Exon 3 is the largest exon and its coding part is located within a recombination hotspot of about 4 kb in size. The 3' boundary of the recombination hotspot falls within the 3' untranslated region of the gene. Note that SNPs rs5030717 (S05) and rs5030725 (S06) are at the 5' boundary of the recombination hotspot while rs7869402 (S07) is at the 3' boundary of the hotspot. (Source: http://hapmap.ncbi.nlm.nih.gov/index.html.en) (B) Linkage disequilibrium (LD) measures (D' and r^2) of tag SNPs and haplotype blocks of the *TLR4* gene. The LD measures of the combined group of cases and controls are calculated by Haploview, with shades of gray indicating the magnitude of the measures and black equal to 100%, which is omitted in the diagram to avoid cluttering the display. Haplotype blocks are constructed using Gabriel's definition (one of the standard features in Haploview). As calculated by Haploview, the multiallelic LD measure (D') between these two haplotype blocks is only 0.57.

Table 1 Main characteristics of study groups

	Cases *	Controls *
Definition	Patients with pulmonary tuberculosis: Clinical presentation, Smear positive for acid-fast bacilli on all three consecutive sputum specimens	Healthy subjects: Normal chest radiograph, No past/family history of tuberculosis
Total no.	207	395
Males (%)	146 (70.5%) †	355 (89.9%) †
Mean age ± SD (yr)	30.0 ± 11.6	28.5 ± 6.1

* All cases and controls were Nilo-Saharan speaking Sudanese residing in Wad Medani (Sudan), had similar socioeconomic status and shared the same major ethnic category (Nilotic). Note that most cases and controls were Bacillus Calmette-Guérin scare positive. Controls were not tested for tuberculin skin test. Subjects with diabetes, HIV infection or taking immunosuppressive drugs were excluded.

† The proportion of males differed significantly between cases and controls (P <0.0001).

Table 2 *TLR4* SNPs: comparison of genetic data between subjects with pulmonary tuberculosis (cases) and healthy subjects (controls)*

SNP†	Gt	Ca/Co (207/395) ‡	MH-OR (95% CI) ‡	*P* ‡	Common MH-OR (95% CI) §	*P* §	Allele	Ca/Co (414/790) ‖	MH-OR (95% CI) ‖	*P* ‖
rs10116253	TT	88/168	1.00 (reference)	0.9500	0.98 (0.76-1.26)	0.8498	T	270/511	1.00 (reference)	0.9000
(S01)	TC	94/175	1.01 (0.69-1.47)				C	144/279	0.98 (0.75-1.26)	
	CC	25/52	0.92 (0.53-1.62)							
rs10759932	TT	107/242	1.00 (reference)	0.1131	1.29 (0.98-1.70)	0.0691	T	297/610	1.00 (reference)	0.0711
(S02)	TC	83/126	1.47 (1.02-2.13)				C	117/180	1.31 (0.99-1.73)	
	CC	17/27	1.36 (0.70-2.63)							
rs1927911	GG	28/90	1.00 (reference)	**0.0008**	1.56 (1.23-1.87)	**0.0002**	G	134/356	1.00 (reference)	**0.0001**
(S03)	AG	78/176	1.35 (0.81-2.27)				A	280/434	1.67 (1.29-2.15)	
	AA	101/129	2.38 (1.42-3.99)							
rs1927907	CC	102/244	1.00 (reference)	0.0232	1.45 (1.10-1.90)	0.0075	C	288/613	1.00 (reference)	0.0069
(S04)	CT	84/125	1.57 (1.08-2.28)				T	126/177	1.47 (1.21-1.95)	
	TT	21/26	1.84 (0.98-3.45)							
rs5030717	AA	119/262	1.00 (reference)	0.0666	1.25 (0.92-1.69)	0.1505	A	318/639	1.00 (reference)	0.1718
(S05)	AG	80/115	1.53 (1.05-2.22)				G	96/151	1.25 (0.92-1.68)	
	GG	8/18	0.90 (0.38-2.12)							
rs5030725	TT	140/316	1.00 (reference)	0.0027	1.89 (1.31-2.72)	**0.0006**	T	341/707	1.00 (reference)	**0.0008**
(S06)	TG	61/75	1.89 (1.26-2.83)				G	73/83	1.85 (1.30-2.63)	
	GG	6/4	3.42 (0.85-13.85)							
rs7869402	CC	99/270	1.00 (reference)	**2.25e-06**	2.07 (1.56-2.75)	**3.81e-07**	C	280/651	1.00 (reference)	**1.87e-07**
(S07)	CT	82/111	1.95 (1.34-2.84)				T	134/139	2.14 (1.61-2.84)	
	TT	26/14	4.56 (2.22-9.36)							
rs1927906	TT	61/212	1.00 (reference)	**9.33e-09**	2.11 (1.66-2.68)	**1.19e-09**	T	214/567	1.00 (reference)	**1.23e-10**
(S08)	TC	92/143	2.24 (1.51-2.24)				C	200/223	2.31 (1.79 -2.98)	
	CC	54/40	4.41 (2.63-7.40)							

rs7044464 (S09)	TT	155/258	1.00 (reference)	0.0615	0.72 (0.52-0.99)	0.0411	T	355/640	1.00 (reference)	0.0469
	TA	45/124	0.61 (0.41-0.92)				A	59/150	0.70 (0.50-0.98)	
	AA	7/13	0.82 (0.32-2.11)							
rs7856729 (S10)	GG	141/236	1.00 (reference)	0.1258	0.77 (0.57-1.04)	0.0896	G	339/615	1.00 (reference)	0.1061
	GT	57/143	0.67 (0.46-0.98)				T	75/75	0.76 (0.56-1.04)	
	TT	9/16	0.87 (0.37-2.05)							

Abbreviations: Ca, cases; CI, confidence intervals; Co, controls; Gt, genotype; MH-OR, Mantel-Haenszel odds ratio; SNP, single-nucleotide polymorphism.

* The columns Ca/Co (third from the left and third from the right) show the distributions of genotypes and alleles, respectively, in cases and controls without stratification by gender. However, there are significantly fewer males in cases than in controls (70.5% vs 89.9%, $P < 0.0001$). To control for potential confounding due to gender, comparison of genotypes and alleles between cases and controls is performed using Mantel-Haenszel χ^2 test with stratification based on gender. There are a total of 30 comparisons in the table, and *P* values in **boldface** indicate that the differences between the cases and the controls are significant at $P = 0.0017$ (the significant level for each individual test adjusted for 30 comparisons by Bonferroni procedure at an overall significance level of 0.05).

† The SNPs are listed down the column in order from the 5' end to the 3' end of the *TLR4* gene and are also designated as S01 to S10 in this order for the sake of easy reference in the text.

‡ Indicates the number of subjects with a given genotype for Ca/Co. There are a total of 207 Ca and 395 Co. The ORs are calculated with the homozygote listed first for each SNP as the reference. The genotype counts of Ca and Co are compared by the generalized Mantel-Haenszel χ^2 test controlled for gender with 2 degrees of freedom.

§ The common OR is calculated for each copy increase of the risk allele (the allele listed second in a heterozygous genotype or in the Allele column). This additive genetic model is tested by the Mantel-Haenszel χ^2 test for trend and controlled for gender with 1 degree of freedom.

‖ Indicates the number of chromosomes with a given allele for Ca/Co. The total number of alleles or chromosomes is 414 (2 times the number of subjects) for Ca and 790 for Co. The OR is calculated with the allele listed first for each SNP as the reference. The allele counts of Ca and Co are compared by the Mantel-Haenszel χ^2 test controlled for gender with 1 degree of freedom.

Table 3 Association tests of *TLR4* haplotypes

Haplotype Block*	Haplotype*	Estimated freq in cases (n=414) †	Estimated freq in controls (n=790) †	χ^2 test *P* value	Empirical *P* value‡
Block 1	TTGCAT	0.318	0.427	**0.0002**	**0.0024**
	CCATGT	0.219	0.187	0.1887	0.8426
	TTACAG	0.176	0.105	**0.0004**	**0.0043**
	TTACAT	0.133	0.111	0.2592	0.9697
	CTACAT	0.06	0.106	0.0084	0.0819
	CCATAT	0.061	0.032	0.0180	0.1862
	CTGCAT	0.006	0.021	0.0445	0.3796
Block 2	CTT	0.374	0.528	**4.09e-07**	**0.00001**
	TCT	0.324	0.175	**4.96e-09**	**<0.00001**
	CTA	0.143	0.189	0.0425	0.3604
	CCT	0.159	0.108	0.0104	0.0991

* Haplotype block 1 consists of 6 SNPs with haplotypes shown in the order S01, S02, S03, S04, S05 and S06. Haplotype block 2 consists of 3 SNPs with haplotypes shown in the order of S07, S087, and S09. See Figure 1 for details of haplotype blocks. Underlined bases refer to the alleles of the SNPs showing association in single-marker analysis: S03 and S06 in block 1, and S07 and S08 in block 2. The reference allele is underlined with single-line, and the risk allele with double-line.

† Haplotype frequencies in cases and controls are estimated by Haploview using an accelerated expectation-maximization algorithm. There are 207 cases and 395 controls, and hence 414 and 790 chromosomes in these two groups, respectively.

‡ Empirical *P* values are generated by Haploview using 100,000 permutations. *P* values in **boldface** indicate significant differences of the haplotype frequencies between the cases and the controls at $P = 0.05$ after correction of multiple comparisons by permutation. The original χ^2 test *P* values are also indicated in **boldface** accordingly.

Online appendix

Common Polymorphisms in the *TLR4* Gene are Associated with Susceptibility to Pulmonary Tuberculosis in Sudanese

H. Y. Zaki, K. H. Leung, W. C. Yiu, N. Gasmelseed, N.E.M. Elwali, S. P. Yip

SNP genotyping

Of the ten tag SNPs selected, 8 were genotyped using the method of restriction fragment length polymorphism (RFLP) under conditions recommended by the manufacturer (MBI Fermentas) (Table S1). Polymerase chain reaction (PCR) was carried out in a 10-μL reaction mixture containing 20 ng of genomic DNA, 3.5 mmol/L $MgCl_2$, 0.3 or 0.5 μmol/L each of the forward and reverse primers (Table S1), 0.2 mmol/L of each dNTP, 1× PCR Buffer, and 0.2 U of HotStarTaq *Plus* DNA Polymerase (Qiagen). Amplification was performed in 96-well plates with a GeneAmp PCR System 9700 thermocycler (Applied Biosystems), including an initial denaturation step of 95°C for 5 min, 38 cycles of 95°C for 30 s, annealing temperature for 30 or 60 s, and 72°C for 40 or 60 s (Table S1), plus a final extension step of 72°C for 5 min. Digested PCR products were separated by electrophoresis in horizontal polyacrylamide gels of appropriate concentration and prepared using an in-house tailor-made casting cassette. Other than the polymorphic restriction site of the SNP to be genotyped, all PCR fragments also contained an invariant internal control restriction site that was either a natural site within the fragment or an artificial restriction site introduced through a mismatched primer. We found this internal control restriction site very crucial to accurate genotyping of a large number of samples because it avoids wrong genotype calls due to inactive restriction enzyme or incomplete digestion.

Two remaining SNPs were genotyped using hybridization probe assays based on induced fluorescence resonance energy transfer (FRET).[1] Asymmetric PCR was performed as described above with the following exceptions: 0.1 μmol/L forward primer and 0.5 μmol/L reverse primer in a reaction volume of 15 μL. After amplification, a 15-μL melting mixture was prepared as follows: 10 μL of PCR product, 3 μL of 10× SYBR Green I solution and 2 μL of 10 μmol/L Cy5-labeled probe solution.

Melting analysis was performed in LightCycler 480 (Roche) with the following settings: denaturation at 94°C for 2 min, cooling at a rate of 20°C/s to 40°C, holding at 40°C for 2 min, continuous melting curve acquisition during a 0.2°C/s to 90°C, and final cooling to 40°C. The wavelength was 483 nm for excitation, and 670 nm for emission. Genotype was called based on the temperature(s) of the melting peak(s) in the derivative melting curve. Finally, irrespective of the genotyping methods, representative samples from each genotype were sequenced with a cycle sequencing kit (Applied Biosystems) to confirm the genotypes.

REFERENCES

1 Lo CL, Yip SP, Cheng PK, To TS, Lim WW, Leung PH. One-step rapid reverse transcription-PCR assay for detecting and typing dengue viruses with GC tail and induced fluorescence resonance energy transfer techniques for melting temperature and color multiplexing. Clin Chem 2007; 53: 594-599.

Table S1 *TLR4* single nucleotide polymorphisms: genotyping method and PCR conditions

SNP*	Primer / Probe sequences (5'>3')†	Genotyping method‡	PCR conditions§				
			Size (bp)	Primer conc. (μM)	Annealing temp. (°C) / time (sec)	Extension time (sec)	Cycle no.
rs10116253 (S01)	F: $(T)_{23}$ GCC CTG GTA AGT GCG GTC R: GTA GCA AGT GCA ATG TAA GTT TCT GTT A	RFLP by *Eco*47I	508	0.3	59 / 60	60	38
rs10759932 (S02)	F: TGC TAT GAT TAA AAG TGA TTA CCA R: TTC TCC ACA GCA TCT GTA C P: CAG ACC AGA AAG TAA TAA TAA GCG TTG-Cy5	Induced FRET	132	0.1 (F) / 0.5 (R)	54 / 30	40	55
rs1927911 (S03)	F: TGA GAG CAT TCA GAA ATT AGA TGG GA R: CTT ATG TAA TAA ACC TGC ATG CTC TG	RFLP by *Eco*103I	385	0.3	61 / 30	40	38
rs1927907 (S04)	F: CTG TAC GAA AAG GCA AAG GAT GT R: TTT TTT TTT TTT CAA TCA AGA AGT AGT TGT TC	RFLP by *Hpy*8I	215	0.3	58 / 30	40	38
rs5030717 (S05)	F: AAC CTT CTT TCT CCA CAA CTC CAT AT R: TGC TAT AAG AGT TCA CAT CCA GAG AA	RFLP by *Tas*I	258	0.3	60 / 30	40	38
rs5030725 (S06)	F: CCA AGA AAA TCA AAG TAG GTT T R: GTA GCA TGG CTT TCA TTT C P: GGA ATC ATG ACA AAT AGC TTC-Cy5	Induced FRET	297	0.1 (F) / 0.5 (R)	53 / 30	40	55
rs7869402 (S07)	F: TCA CTC GAT GTC ATT CCA AAG TTA TT R: $(T)_{20}$ GTA GTT CTC CTG GCC GTG	RFLP by *Bse*LI	381	0.5	59 / 60	60	38
rs1927906 (S08)	F: $(T)_{20}$ CTG AGA AGG AAG ATC AAC ATC CAA R: $(T)_{20}$ CTC ACC TGC CAT TTC TAG GAT TT	RFLP by *Bse*JI	161	0.3	59 / 30	40	38
rs7044464 (S09)	F: $(T)_{20}$ TGC TTT GAG GCT CAG GTC TGA AT R: $(T)_4$ TTT GTA AAT TGG AAA TAC TTT CAT ATC TTT CGA AT	RFLP by *Eco*RI	228	0.5	55 / 60	60	38
rs7856729 (S10)	F: TCA CAT CAA TGA ATT AGA AAG ATA CGA AA R: TTT GTT CAC TGT TAA AAA ATC TTT GGT CA	RFLP by *Nmu*cI	356	0.5	58 / 30	40	38

* The single nucleotide polymorphisms (SNPs) are listed down the column in order from the 5' end to the 3' end of the *TLR4* gene and are also designated as S01 to S10 in this order for the sake of easy reference in the text.

† Primer or probe names are indicated before the colon. Names starting with F indicate forward primers; names with R, reverse primers; and names with P, probes. Underlined bases indicate mismatches with genomic sequences. Mismatched bases are introduced near the 3' end of the primers to create restriction sites for the SNP to be genotyped or for internal control sites for restriction digestion. Other mismatches are introduced to avoid hairpin structure or primer-dimer formation. Poly-T tails indicated as $(T)_n$ are added to the 5' end of several primers so that restricted fragments of similar sizes can be distinguished from each other more easily in the electrophoretograms. Probes for induced FRET are labeled with the fluorescence dye Cy5 at the 3' end.

‡ RFLP stands for restriction fragment length polymorphism, and FRET for fluorescence resonance energy transfer. For RFLP, the digestion was carried out for ~16 hours at 37°C, except for *Tas*I and *Bse*JI (65°C), and *Bse*LI (55°C).

§ See main text for details.