

論文の内容の要旨

論文題目 Identification of susceptibility loci to tuberculosis in Thais

(タイ人における結核感受性遺伝子座の同定)

氏名 スラカメ マハシリモンコン
Surakameth Mahasirimongkol

Tuberculosis (TB) is the leading cause of death in developing countries, despite the advanced insights into the mechanisms of disease development, from basic science studies with animal models of host immunity and extensive studies of the virulence factors in *Mycobacterium tuberculosis*. Among the host immune mechanisms postulated from studies of disease development in the animal models, confirmation of their roles in human host response to *Mycobacterium tuberculosis* infection can be obtained, by identifying functional polymorphisms influencing the variability of disease development after infection in humans. Studies of polymorphisms distributed differently in the human populations that affect disease development in TB would provide an unbiased estimate of the role of each immune mechanism in humans.

The candidate gene association studies have been applied in the search for genetic risks for leprosy, a related mycobacterial disease caused by *Mycobacterium leprae*. In addition, linkage analyses (Mira, Alcais et al. 2003; Mira, Alcais et al. 2004; Alcais, Alter et al. 2007) and a genome-wide association study (GWAS) in leprosy (Zhang, Huang et al. 2009) identified multiple host genetic factors with moderate to large effects located in biologically relevant candidate genes, such as *LTA*, *PARK2*, *HLA-DR-DQ*, *C13orf31*, *CCDC122*, and *RIPK2*. For several common infectious diseases, susceptibility loci, location of the genetic risk in the human genome, where recombination rarely occurred, with moderate to large effects (at risk odds ratios > 1.5), were successfully identified by GWAS. For example, associations have been identified between *HbS*, variant of hemoglobin gene causing Sickle Cell disease, and malaria (Jallow, Teo et al. 2009); and between *CFH* locus and *Neisseria meningitidis* (Davila, Wright et al. 2010). Recently, a GWAS of two African TB populations identified a gene desert in 18q11.2 as a novel candidate locus for TB, but unlike GWASs in other common infectious diseases, GWASs of TB was not successful in identifying any genetic factors with moderate to large effect sizes (Thye, Vannberg et al. 2010).

Clinical and epidemiological classifications of TB have been based on the timing of the expression of signs and symptoms after infection. Classically, three groups of TB based on this epidemiological classification were proposed: 1) primary disease after infection; 2) exogenous re-infection; and 3) endogenous re-activation

(Sutherland, Svandova et al. 1982). The primary infection and exogenous re-infection are clinical diseases that develop within 5 years after infection, while endogenous re-activation is a disease that develops later than 5 years after infection. While there are overlapping features of clinical diseases by this classification, a proportion of primary infection TB expresses distinct clinical features, and it is hypothesized that genetic risk factors might play a major role in these TB cases (Alcais, Fieschi et al. 2005). Epidemiological studies and mathematical modeling indicate that the majority of old TB cases in developed countries with a comparatively low TB incidence are mainly endogenous re-activation, while young patients in developing countries are mainly primary TB. These transmission models are typically used for prioritization of limited resources in developing countries to minimize TB incidence. The biological significance of these disease development models for TB has not been clearly defined, but age at onset of TB may be the best available classifier to identify the subset of TB patients based on the disease development model.

Interestingly, GWASs of TB *per se* in Africans and in my original GWAS from two Asian populations could not identify any obvious genetic risks to TB. The sample sizes of the present TB GWAS were similar to those of other infectious diseases, and it appears reasonable that one should be able to identify the common moderate- to high-risk polymorphisms that are present in other common infectious diseases. Failure in replication of top signals from the meta-analysis of GWASs of TB *per se* in Asian populations in the present study reinforce the results of GWASs in African TB (Thye, Vannberg et al. 2010); no common polymorphisms with moderate to high risks to TB could be identified despite the ubiquitous nature of TB in global populations. GWASs in Asian populations does not have the problem of sparse linkage disequilibrium like GWASs in African populations, and they have been successfully utilized to identify common variants with large effects in Thais (Nuinoon, Makarasara et al. 2009). Thus, the failure of the initial GWAS in Asian TB is not due to the problem of SNP coverage in the current microarray genotyping platform.

Taking into consideration the caveats that might affect the power of a GWAS in identifying genetic risks to common diseases (Manolio, Collins et al. 2009) and our experience of the increasing power of linkage analysis of TB by the subset ordered analysis based on the age at onset of TB sibpairs (Mahasirimongkol, Yanai et al. 2009), we hypothesized that a gene-environmental effect (G×E) interaction, such as BCG vaccination, exposure to different strains of *Mycobacterium tuberculosis* (Rienthong, Ajawatanawong et al. 2005), or misclassification bias, might cause heterogeneity in TB. All of these factors contribute to different age at onset contributed to genetic heterogeneity in TB. Thus, GWAS analysis of TB stratified by age at onset might lead to identification of moderate- to high-risk polymorphisms by reducing the genetic heterogeneity.

Having access to two GWAS data sets, the TB cases were split into two groups based on age at onset of TB: the young TB case/control data sets (age at onset <45 years: GWAS-T_{Young}=137/295, GWAS-J_{Young}=60/249) and the old TB case/control data sets (age at onset ≥45 years: GWAS-T_{Old}=300/295, GWAS-J_{Old}=123/685). This age cut-off is based on the bimodal distribution of age at onset of TB of patients within this study and the age distribution of TB patients in Japan and in Thailand. Meta-analysis of Young TB and Old TB provided non-overlapping lists of SNPs and suggested that the top association signals that contributed to Young TB and Old TB are distinct. The Significant findings were replicated in two independent replication samples (Young TB; Rep-T_{Young}=155/249, Rep-J_{Young}=41/462 and old TB; Rep-T_{Old}=212/187, Rep-J_{Old}=71/619). A total of 100 SNPs was genotyped in both Thai and Japanese replication samples; replication in these two case control data sets identified the HSPEP1-MAFB locus as a novel susceptibility locus on chromosome 20q12 for Young TB on GWASs. This locus had moderate effect sizes with an odds ratio (CI) of 1.73 (1.42-2.11). rs6071980 is located 450 kb proximal to *MAFB*, a transcription factor determining the fate of monocyte/macrophage differentiation and 300 kb distal to *HSPEP1* (Chaperonin 10), a heat shock 10-kDa protein suggested to be an auto-antigen in autoimmune hepatitis and type I diabetes. Interestingly, rs6028945 and rs6071980, two closely located SNPs with high correlation in Caucasians, have also been suggested as genetic markers for anti-TNF responsiveness by a GWAS (Liu, Batliwalla et al. 2008). *MAFB* is highly expressed in active TB compared to latent TB in an extensive study of whole blood gene expression signatures that also differentiated active TB cases from health individuals (Berry, Graham et al. 2010). These variants might influence the development of TB through an effect on the reactive response to *M. Tuberculosis* infection by expression of *MAFB* from non-lymphocytic population of white cells. Other loci with lower association evidence ($P_{4M-H} < 10^{-5}$) need additional replication evidence in other populations.

This study demonstrated that consistent replication in TB gene identification is achievable by stratified analysis based on age at onset of TB, despite the smaller number of cases available for analyses. The heterogeneity by age might reflect the complexity of the epidemiology of TB in Asia: introduction of BCG vaccination in the late 40s in Thailand (Ritz and Curtis 2009) and Japan (Yamamoto and Yamamoto 2007); viability of individuals who carried genetic risks of TB; host adaptation to various strains of *M. tuberculosis*; misclassification in the control population due to non-exposure to *M. tuberculosis*; temporal immunosuppression by immunosuppressive drugs (Ledingham, Wilkinson et al. 2005); immune senescence; and other conventional risk factors of TB. In different aged cohorts of TB, the intertwining of these factors contributed to disease development and caused genetic

heterogeneity. Stratification based on age at onset of TB as a classifier to homogenize other uncontrollable factors might be a simple and efficient method for identifying TB susceptibility genes. To increase the power of GWAS for difficult traits similar to TB in this thesis, improvement of the current analysis method is needed to maximize the discovery of a possible hypothesis from the GWAS study. A novel analysis method of GWAS was also developed and presented in the appendix of this thesis; this method shall be applicable to difficult phenotypes such as TB in the near future.