Genes and Immunity (2008), 1–7

www.nature.com/gene

ORIGINAL ARTICLE

Genome-wide SNP-based linkage analysis of tuberculosis in Thais

S Mahasirimongkol[1,2], H Yanai[2,3], N Nishida[4], C Ridruechai[5], I Matsushita[6], J Ohashi[4], S Summanapan[7], N Yamada[2], S Moolphate[2], C Chuchotaworn[8], A Chaiprasert[9], W Manosuthi[10], P Kantipong[11], S Kanitwittaya[12], T Sura[13], S Khusmith[5], K Tokunaga[4], P Sawanpanyalert[1] and N Keicho[6]

[1]Department of Medical Sciences, National Institute of Health, Ministry of Public Health, Nonthaburi, Thailand; [2]TB/HIV Research Project, Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Chiang Rai, Thailand; [3]Center for Collaborative Research and Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan; [4]Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo, Japan; [5]Department of Microbiology and Immunology, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand; [6]Department of Respiratory Diseases, Research Institute, International Medical Center of Japan, Shinjuku-ku, Tokyo, Japan; [7]Chiang Rai Provincial Health Office, Ministry of Public Health, Chiang Rai, Thailand; [8]Department of Medical Services, Chest Disease Institute, Ministry of Public Health, Nonthaburi, Thailand; [9]Department of Microbiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand; [10]Bamrasnaradura Infectious Diseases Institute, Ministry of Public Health, Nonthaburi, Thailand; [11]Chiang Rai Regional Hospital, Ministry of Public Health, Chiang Rai, Thailand; [12]Department of Medical Sciences, Chiang Rai Regional Medical Sciences Center, Ministry of Public Health, Chiang Rai, Thailand and [13]Department of Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

Tuberculosis, a potentially fatal infectious disease, affects millions of individuals annually worldwide. Human protective immunity that contains tuberculosis after infection has not been clearly defined. To gain insight into host genetic factors, nonparametric linkage analysis was performed using high-throughput microarray-based single nucleotide polymorphism (SNP) genotyping platform, a GeneChip array comprised 59 860 bi-allelic markers, in 93 Thai families with multiple siblings, 195 individuals affected with tuberculosis. Genotyping revealed a region on chromosome 5q showing suggestive evidence of linkage with tuberculosis (Z(lr) statistics = 3.01, logarithm of odds (LOD) score = 2.29, empirical P-value = 0.0005), and two candidate regions on chromosomes 17p and 20p by an ordered subset analysis using minimum age at onset of tuberculosis as the covariate (maximum LOD score = 2.57 and 3.33, permutation P-value = 0.0187 and 0.0183, respectively). These results imply a new evidence of genetic risk factors for tuberculosis in the Asian population. The significance of these ordered subset results supports a clinicopathological concept that immunological impairment in the disease differs between young and old tuberculosis patients. The linkage information from a specific ethnicity may provide unique candidate regions for the identification of the susceptibility genes and further help elucidate the immunopathogenesis of tuberculosis.
Genes and Immunity (2008) **0,** 000–000. doi:10.1038/gene.2008.81

Keywords: *linkage analysis; tuberculosis; Thais; 5q; 17p; 20p; age at onset*

Introduction

Tuberculosis remains a serious public health problem in the developing world, especially in view of recent outbreaks of virtually untreatable, extensive drug-resistant tuberculosis.[1] A majority of individuals in endemic areas are infected with the pathogen *Mycobacterium tuberculosis* when they reach adulthood. In 2005, around 5 million individuals were diagnosed as having tuberculosis according to a WHO surveillance report.[2]

Approximately 10% of those who are infected develop tuberculosis in their lifetime. Clinically and epidemiologically, three patterns of disease development after infection are often assumed, depending on the age at onset of tuberculosis and the prevalence of tuberculosis infection: (1) primary tuberculosis in adolescence; (2) reactivation of disease after infection and (3) exogenous re-infection in adulthood.[3] Primary tuberculosis is a disease that develops within the first few years after infection, often because of impaired host immunity, whereas reactivation of disease occurs later in life, after a long period of immune protection against development of the disease in infected individuals, together with the possibility of exogenous re-infection. Identification of a high-risk group corroborated by a specific mechanism for disease development would be one of the most desirable measures for controlling this disease in developing countries, in which use of chemoprophylaxis for

Correspondence: Dr N Keicho, Department of Respiratory Diseases, Research Institute, International Medical Center of Japan, 1-21-1, Toyama, Shinjuku-ku, Tokyo 162-8655, Japan.
E-mail: nkeicho-tky@umin.ac.jp
Received 30 June 2008; revised 11 August 2008; accepted 27 August 2008

all infected individuals is an unacceptable burden on national tuberculosis control programs.

Contribution of host genetic factors to the development of this infectious disease has been classically observed. Until Koch[4] discovered the tubercle bacilli as the causative agent of tuberculosis, familial aggregation resulted in the initial perception that tuberculosis was a hereditary disease in the 19th century. An early study showed that monozygotic twins have higher risks of tuberculosis than dizygotic twins, given tuberculosis affected one of the twins.[5] Individual genetic variations in innate immunity, adaptive immunity and intracellular bacterial killing ability have roles in different phenotypes of tuberculosis. Disruption of genes in the *IL12/23–IFNG* pathway results in disseminated mycobacterial infection in susceptible infants.[6] Various association studies implicated this as one of the key pathways for containing tuberculosis.[7–9] Mutation in genes with ubiquitination functions have also been implicated as risks in both tuberculosis and leprosy.[10,11] On the basis of databases of genetic association study, nearly 100 genes had been studied for their association with tuberculosis. From these candidate gene association studies, apart from genes in *IL12/23–IFNG* axis, only *SLC11A1* (formerly *NRAMP1*) had been consistently shown evidence of association with tuberculosis.[12]

Prior to candidate gene approach, reverse genetic study in humans is often useful. Susceptibility genes based on novel disease regulatory mechanisms can be identified by hypothesis-free studies, such as genome-wide linkage and genome-wide association studies. In African populations, linkage analysis for tuberculosis earlier demonstrated suggestive evidence for linkage in two regions on chromosome 15 and chromosome X. Subsequent fine mapping of the chromosome 15 regions revealed *UBE3A* as a candidate susceptibility gene for tuberculosis.[10,13] Suggestive evidence of linkage was also found on chromosomes 10q26, 11q12 and 20p12 in the Brazilian population.[14,15] The only significant linkage evidence (logarithm of odds (LOD) = 3.49) to date is a study of 96 Moroccan multiplex families that implicated an autosomal-dominant gene on chromosome 8q12–q13.[16]

Distribution of virulent *M. tuberculosis* strains such as Beijing strain differs among regions of the world,[17] which suggests its adaptation to different human populations.[18] In the context of the host–pathogen relationship, human genetic studies in Asia may also reveal novel susceptibility/resistance genes in this field. In this study, linkage analysis in the Thai population was performed to gain insight into host genetic epidemiology of tuberculosis in Asians.

Results

Description of studied families

The numbers of families with multiple siblings affected with tuberculosis are shown in Table 1. Because of limited genotyping resources, only affected siblings were genotyped, which provided better linkage evidence when compared with genotyping the unaffected individuals in these families.

A total of 199 individuals from 95 families were originally genotyped using the *Xba*I 50 K microarray system, which is part of the 100 K affymetrix genotyping system. In the relationship analysis of the GRR program, ambiguous relationships were clarified and corrections were made before linkage analysis. Two of the 95 families (four individuals) were excluded because the analysis revealed that the affected individuals were half-siblings on the basis of their genotypes.

Table 1 Number of families and sibling pairs in the linkage analysis

	Number of families
Two affected siblings	87
Three affected siblings	5
Four affected siblings	0
Five affected siblings	1
Total families	93
Total number of independent sibling pairs (number of affected siblings—~~one~~ per family)	101

A stringent genotype call strategy was used to reduce the number of genotyping errors; criteria for genotype calling were more stringent than the standard criteria proposed by the manufacturer. The average call rate (percentage of successful genotype calls among subjects) was 98.08%. By using more stringent call criteria, the final call rate was 99.48%.

Linkage analysis

Linkage analysis was carried out with MERLIN; analysis without linkage disequilibrium (LD) correction showed inflation of the linkage statistics and noisy linkage spike patterns throughout the genome. LD was taken into account for the analysis by using LD patterns described earlier and MERLIN, which implements a method accounting for LD by inferred haplotypes from single nucleotide polymorphisms (SNPs) that showed LD with each other and used the inferred haplotypes as multi-allelic markers for linkage analysis.

After LD correction, a maximum LOD score of ~~2.32~~ was observed on chromosome 5q23.2–31.3 at 138.3 cM, with the LOD-1 support interval between rs1515641 (SNP_A-1673674, 131.9086, 127.2 Mb) and rs252101 (SNP_A-1739565, 141.01 cM, 141.3 Mb) (Table 2 and Supporting information Figures 2a and b). Simulation tests showed this linkage to be genome-wide suggestive evidence of linkage (0.25 peaks per genome scan). Regions that showed LOD score >1 were located on chromosomes 4, 5, 6, 11 and 14 (Table 2). All of these regions could be good candidates for follow-up positional study.

Age at onset analysis

Ordered subset analysis was used to determine whether the evidence for linkage differs according to age at onset. The results at which LOD scores reached the uncorrected significance value (*P*-value <0.05) are shown (Table 3). Two regions on chromosome 17p13.3–13.1 and chromosome 20p13–12.3 showed significantly higher LOD scores (permutation *P*-value at 0.0187 and 0.0183, respectively) (Figures 1a and b).

Table 2 Chromosome regions with a non parametric LOD score with nominal significance ($P<0.01$)

Chromosome	*Peak SNP*	*Position (cM)*	*Physical position*	*LOD score*[a]	P-*value*
5q23.2–31.3	rs7706155	138.37	135 987 344	2.27	0.0006
	rs9327759	138.39	136 004 959		
	rs2188468	138.42	136 026 947		
11p15	rs1487214	29.85	19 588 715	1.62	0.003
14q32	rs8014257	109.60	98 321 756	1.48	0.005
6p12	rs10484980	78.13	54 559 585	1.33	0.007
4q26	rs2010003	121.27	118 807 065	1.3	0.007

Abbreviation: LOD, logarithm of odds.
[a]Calculated using the Kong and Cox linear model.

Table 3 Ordered subset analysis by minimum age at onset of tuberculosis in the families

Chromosome	*cM*	*Marker*	*No. of families in subset*	*Average age at onset (range)*	*Maximum LOD and (baseline LOD)*	P-*value*
17p13.3–13.1	13.5–13.7	rs2716912 rs10491086	32	19.47 (12–24) years	2.57 (0.14)	0.0187
20p13–12.3	10.6	rs750702	30	19.17 (12–23) years	3.33 (0.64)	0.0183

Abbreviation: LOD, logarithm of odds.

Discussion

In this study involving 93 Thai families (195 affected individuals), a region on chromosome 5q23.2–31.3 was shown to be a candidate region for tuberculosis susceptibility with suggestive evidence of linkage by genome-wide linkage analysis. From the ordered subset analysis, two regions on chromosome 17p13.3–13.1 and chromosome 20p13–12.3 were shown to have significant linkage with earlier onset of tuberculosis.

In the 5q candidate region, various genes, including a cluster of cytokine genes, *GM-CSF*, *IL3*, *IL4*, *IL5* and *IL13*, have been mapped. *IRF1*, a responsive element in the interferon-γ-mediated pathway, an antimicrobial peptide gene (*LEAP2*), and genes with ubiquitination activities, such as *APBB3* and *SPK1A*, have also been located. This chromosomal region has also been implicated in linkage analyses of the parasitic load of *Schistosoma mansoni*[19] and *Plasmodium falciparum*.[20] In addition, this region on chromosome 5 coincides with the *IBD5* region that showed significant evidence of linkage and association with Crohn's disease, an inflammatory bowel disease closely associated with *Mycobacterium paratuberculosis*.[21] It is interesting to note that the *IBD5* haplotype is distributed differently in Asian and European populations, and that the *IBD5* risk haplotype in Europeans is very rare in Asians.[22] Although candidate genes including *IL4*[23] and *CD14*[24,25] have been studied in this particular region, roles of these genes in tuberculosis remain inconclusive. Small sample size and ethnicity of the study populations may partially explain these equivocal association results. With this linkage evidence, polymorphisms within these genes should be the subjects for further validation in another Asian population.

From the functional viewpoint, this region has been reported to control the differentiation of Th2 lymphocytes, resulting in different levels of interleukin-4 (IL-4) and IL-5 from $CD4^+$ T-cell clones after exposure to parasite antigens, but not to affect the level of interferon-γ.[26] The predilection of this balance for Th2 may also play a role in susceptibility to tuberculosis. The predominant Th2 may partly influence tuberculosis susceptibility by the IL-4 and IL-13 capability of inhibiting autophagy by mycobacterium-infected macrophages.[27] IL-4 has been suggested to be a regulator of the immune response to tuberculosis antigen but not the disease *per se* in Brazilian families with multiple cases of tuberculosis.[23] This study suggests that the Th2 regulatory function of this region may play a more important role in tuberculosis susceptibility in the Thai population than in the Brazilian population. This genetic heterogeneity may reflect the host–pathogen interaction with specific mycobacterium strains: in Thailand, at least 20% of circulating strains are Beijing strains of Asian origin.[28] These virulent strains have been shown to preferentially stimulate Th2 response more than other non-virulent strains.[29]

From the ordered subset analysis of earlier onset of tuberculosis, the maximum linkage statistics (LOD = 3.33) on chromosome 20p were derived from 30 of the 93 families, and the average minimum age at onset of tuberculosis in these families was 19.17 years. The 20p region is of particular interest for tuberculosis susceptibility because it has been shown to have significant evidence of linkage with leprosy.[15,30] The susceptibility gene(s) in this region may play a role in susceptibility to primary tuberculosis in pediatric and adolescent cases. Several promising candidate genes in this region are a cluster of β-defensin genes that play important roles in innate immunity against respiratory infection.[31,32] The suggestive evidence of linkage for the region on chromosome 17p was derived from 32 families with the youngest age at onset, but with a lower maximum LOD score than for the 20p region. Although it is speculative at this stage, a possible candidate in the 17p13.3–13.1 region, *MYBBP1A* was recently identified as a co-repressor of NF-κB.[33] Although these two chromosome regions provide statistical evidence of possible linkage in subsets of these families, small number of families in the ordered subset analysis requires further confirmatory

Gene : NPGGENEgene200881

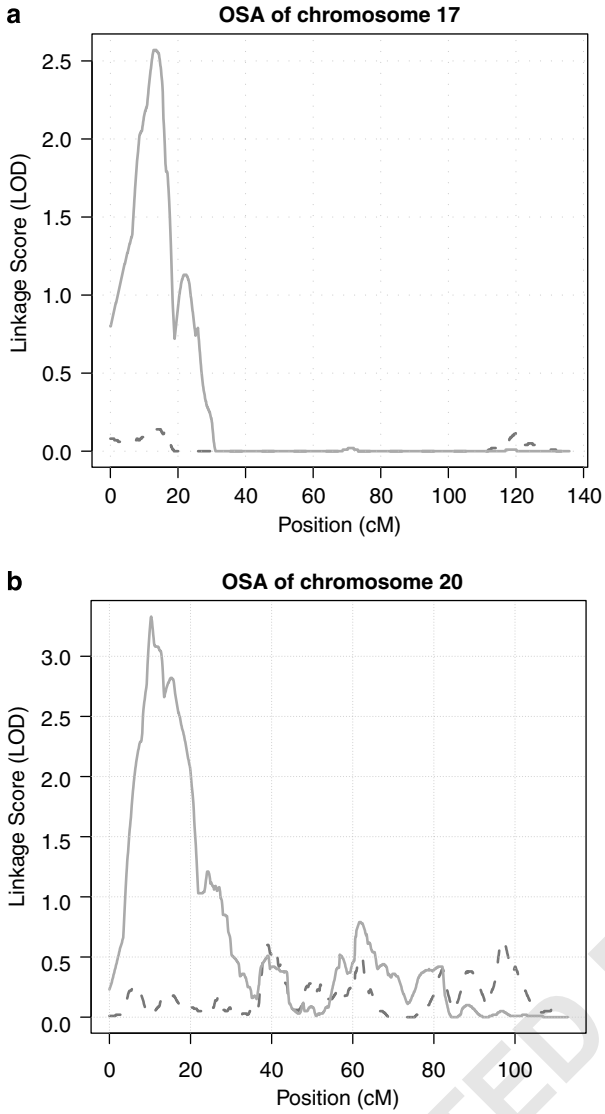

Figure 1 Ordered subset analysis (OSA) results on chromosome 17 and chromosome 20. The solid lines indicate the linkage statistics from the subset of families with younger age at onset. Thirty-two families and 30 families were grouped into the younger age at onset family for chromosome 17 (**a**) and chromosome 20 (**b**) respectively. The dotted lines indicate the statistics when all families were included in the analysis. (Image file = Fig1(a)_OSAChr17_black.pdf) (Image file = Fig1(b)_OSAChr20_black.pdf).

evidence either through additional linkage analysis or association study.

As for the regions on chromosomes 8, 15 and X reported in African populations, we could not find any evidence of linkage. Discordance of the results may be partially explained by the genetic heterogeneity in tuberculosis susceptibility in different human populations, and also by the differences in the distribution of *M. tuberculosis* strains.[18] We could improve the linkage evidence by genotyping more parental samples, but this may not be realistic given the small number of families in whom the parents are alive and available for enrollment. Confirmation of the linkage in another study with a larger sample size with extended genotyping of Asian families with tuberculosis and fine mapping around these candidate regions would be necessary to reinforce our findings in the future. A genome-wide association study provides an alternative tool for selective re-genotyping of the linkage regions and also is capable of revealing common low risk alleles even below the detection limit of linkage analysis. These hypothesis-free, reverse genetic approaches would elucidate the pathogenic mechanism and might enable effective treatment and prevention strategies for tuberculosis in the future.

Materials and methods

Ascertainment and collection of families

Families were ascertained mainly through a tuberculosis surveillance system in Chiang Rai province, the northernmost province of Thailand. Additional families were also recruited through the Central Chest Hospital in Bangkok, Thailand. In total, 95 families with at least two siblings affected with tuberculosis were ascertained. These families included 199 individuals affected with tuberculosis. A parent in a sibling pair family was genotyped because this parent was also affected with tuberculosis. Tuberculosis was diagnosed by clinical characteristics and microbiological confirmation by sputum culture or at least two out of three positive sputum smears. In a minority of cases, the diagnosis was obtained from earlier records in the tuberculosis surveillance system supplemented by an abnormal current chest X-ray. This study was reviewed and approved by The Ethical Review Committee for Research in Human Subjects (Ministry of Public Health, Thailand) and the Institutional Review Board of the International Medical Center of Japan. The patients were tested for HIV using the standard serological test, and HIV-infected patients were not included in this study. Venous blood samples were collected from the patients after individual informed consent was obtained by our field researchers.

Genotyping

Genomic DNA extraction from blood samples was performed using the QIAamp blood midi kit (Qiagen, Hilden, Germany) and quantification was performed using a nanodrop spectrometer to a concentration of 50 ng μl^{-1}. The *Xba*I microarray chip of the mapping 100 K Array (Affymetrix, Santa Clara, CA, USA) was used for genotyping, following the standard genotyping protocol for the GeneChip Mapping 100 K Array. In summary, 5 µl of 50 ng μl^{-1} (250 ng) of genomic DNA were digested with the restriction endonuclease (*Xba*I) and ligated to adaptors with the T4 ligase that recognizes the cohesive four base-pair overhangs. A generic primer that recognizes the adaptor sequence was used to amplify adaptor-ligated DNA fragments. Preferential amplification of fragments in the 250–2000 bp size range was performed by the optimized PCR conditions. The amplified DNA was purified by the DNA amplification clean-up kit (Takara Bio, Shiga, Japan) and then fragmented, labeled and hybridized to the *Xba*I chip of the mapping 100 K set.

Genotype calling algorithm and quality control

Genotype calls for each chip were given by the BRLMM algorithm using a confidence call threshold of 0.3. The

BRLMM algorithm is based on modification of the RLMM algorithm, which showed a call accuracy superior to that of the dynamic modeling algorithm.[34] The confidence call threshold was set at a lower threshold than the standard call threshold (0.5) to select only SNPs with higher call confidence and to reduce the effect of genotyping errors in linkage analysis. SNPs with missing rates of more than 5% were excluded from the later analysis. Thus, the average genotype call rate was 99.48% for 52 433 SNPs selected from 58 960 SNPs on the *Xba*I chip.

Quality control and Mendelian error checking

ALOHOROMA was used for quality control and preparation of input files and semi-automated linkage analysis using the Perl script.[35] The gender of each sample was checked with the number of heterozygosities at SNPs on chromosome X, and there was complete concordance between the reported sex and the genotyped sex. PedCheck was used for exclusion of markers that were in Mendelian inconsistence, and 339 genotypes data were deleted.[36]

Markers with significant deviation from Hardy–Weinberg's equilibrium (P-value <0.01) were filtered before the downstream analysis. The relatedness of these samples was analyzed using the GRR program, which analyzed the pair-wise relatedness for each pair of samples in the data set.[37] Mendelian error checking was also performed using MERLIN, with the error automatically wiped out before linkage analysis.[38] The map order and distances between SNP markers were based on the information provided by the manufacturer, with the flanking sequences of each of the SNPs aligned with the May 2004 release of the human genome sequence. LD patterns of these SNPs were determined using Haploview version 3.2. Because the tagging SNP selection algorithm had not been implemented for chromosome X in HAPLOVIEW V 3.2, we used the haplotype inference information based on a spine of strong LD ($D'>0.8$) for correction of LD in X chromosome analysis.[39] LD parameters were calculated by Haploview for every pair of markers situated within 1 Mb from each other. The LD information was used for the selection of tagSNPs to reduce the effect of LD on linkage statistics. TagSNPs were selected using a pairwise tagging strategy with $r^2>0.4$ as the threshold for the selection of independent markers. These tagSNPs in autosomal chromosomes and haplotype inferences in chromosome X were later used in linkage analysis by MERLIN. The final set of markers used for later analysis included 33 365 markers selected as the tagging set ($r^2<0.4$) from all available markers.

Linkage analysis

Nonparametric multipoint linkage analysis was carried out by MERLIN using all markers that passed the quality control criteria. LD was accounted for in the analysis by reanalyzing the data using haplotype block information derived from HAPLOVIEW (spine of strong LD, $D'>0.8$) on chromosome X and reanalyzing by using only tagSNPs ($r^2>0.4$ in the 1 MB region) in autosomal regions to reduce the effect of LD on linkage analysis. Population allele frequencies of these SNPs were inferred from unrelated individuals in the samples by MERLIN.

MERLIN recently provided a method to account for LD by determining the haplotype of SNPs in strong LD and using inferred haplotypes as multiallelic markers.[40] Effects of LD on the linkage analysis were investigated by comparison of nonparametric linkage statistics with and without haplotype block information and tagSNPs for linkage analysis (Supporting information Figure 2a and b). The significant levels of linkage statistics were assessed by 10 000 gene dropping simulation. The simulation was carried out with data from chromosome 22 to reduce the computation time. The numbers for each of the linkage statistics from the simulations were multiplied by 44 to compensate for the expected size of the whole genome compared with chromosome 22. The suggestive and significant evidences of linkage by this simulation were observed at LOD score levels of 1.79 and 3.12, respectively.

Ordered subset analysis with age at onset

Age at onset was taken into account for linkage analysis using the age at diagnosis of tuberculosis. When the age at diagnosis was not available, the patient's current age was imputed to replace the age at diagnosis. Only nine data points were imputed using the current age at ascertainment. Our ascertainment scheme only identified patients who developed tuberculosis within the past 10 years, and the current age at ascertainment should be fairly accurate for imputation. The age at onset was analyzed as a covariate using ordered subset analysis by the FLOSS program.[41] The minimum age at onset of tuberculosis was used as the covariate because it is more likely that tuberculosis occurred at a younger age caused by immunological impairment to tuberculosis rather than by repetitive external mycobacterium exposures or endogenous immune senescence when compared with tuberculosis in the elderly population. FLOSS accepts the MERLIN LOD score outputs per family for reanalyzing the linkage statistics by ordering the families by the supplied covariate. The linkage analysis was re-performed by adding one family at a time, and the highest linkage score from the ordered subset analysis was chosen as the plausible linkage when age at onset presented in the disease model. The significance of the ordered subset analysis was calculated by simulation through re-sampling of equal numbers of families that provided the maximum linkage statistics from all available families, and then the number of times the re-sampling statistics exceeded the highest observed statistics was counted. The significant levels of the ordered subset analysis can be considered chromosome-wide with a permutation P-value <0.025 and genome-wide significant with a permutation P-value <0.001. This analysis technique is robust against heterogeneity and should be useful in analysis of linkage in tuberculosis that showed two peaks of higher incidence in two different age groups, that is young adult and elderly populations.

Acknowledgements

This study was partly supported by International Cooperation Research Grant, the Ministry of Health, Labor and Welfare from 2002 to 2004 and by a Grant-in-Aid for Scientific Research on Priority Areas

'Comprehensive Genomics' from the Ministry of Education, Culture, Sports, Science and Technology of Japan. We thank Ms Masako Okochi (International Medical Center of Japan) for technical support. We extend their appreciation to all of the staff and collaborators of the TB/HIV Research Project, Thailand, a collaborative research project between the Research Institute of Tuberculosis (RIT), the Japan Anti-tuberculosis Association and the Thai Ministry of Public Health for collecting the clinical information and samples.

References

1 Raviglione MC, Smith IM. XDR tuberculosis—implications for global public health. *N Engl J Med* 2007; **356**: 656–659.
2 World Health Organization. Global tuberculosis control: surveillance, planning, financing: WHO report 2007. In: *World Health Organization*, 2007, p 26.
3 Sutherland I, Svandova E, Radhakrishna S. The development of clinical tuberculosis following infection with tubercle bacilli. 1. A theoretical model for the development of clinical tuberculosis following infection, linking from data on the risk of tuberculous infection and the incidence of clinical tuberculosis in the Netherlands. *Tubercle* 1982; **63**: 255–268.
4 Koch R. Die Aetiologie der Tuberculose [The aetiology of Tuberculosis]. *Berliner Klinische Wochenschrift* 1882; **19**: 221–230.
5 Comstock GW. Tuberculosis in twins: a re-analysis of the Prophit survey. *Am Rev Respir Dis* 1978; **117**: 621–624.
6 Casanova JL, Abel L. Genetic dissection of immunity to mycobacteria: the human model. *Annu Rev Immunol* 2002; **20**: 581–620.
7 Rossouw M, Nel HJ, Cooke GS, van Helden PD, Hoal EG. Association between tuberculosis and a polymorphic NFkappaB binding site in the interferon gamma gene. *Lancet* 2003; **361**: 1871–1872.
8 Kusuhara K, Yamamoto K, Okada K, Mizuno Y, Hara T. Association of IL12RB1 polymorphisms with susceptibility to and severity of tuberculosis in Japanese: a gene-based association analysis of 21 candidate genes. *Int J Immunogenet* 2007; **34**: 35–44.
9 Filipe-Santos O, Bustamante J, Chapgier A, Vogt G, de Beaucoudrey L, Feinberg J *et al.* Inborn errors of IL-12/23- and IFN-gamma-mediated immunity: molecular, cellular, and clinical features. *Semin Immunol* 2006; **18**: 347–361.
10 Cervino AC, Lakiss S, Sow O, Bellamy R, Beyers N, Hoal-van Helden E *et al.* Fine mapping of a putative tuberculosis-susceptibility locus on chromosome 15q11–13 in African families. *Hum Mol Genet* 2002; **11**: 1599–1603.
11 Mira MT, Alcais A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT *et al.* Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature* 2004; **427**: 636–640.
12 Li HT, Zhang TT, Zhou YQ, Huang QH, Huang J. SLC11A1 (formerly NRAMP1) gene polymorphisms and tuberculosis susceptibility: a meta-analysis. *Int J Tuberc Lung Dis* 2006; **10**: 3–12.
13 Bellamy R, Beyers N, McAdam KP, Ruwende C, Gie R, Samaai P *et al.* Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. *Proc Natl Acad Sci USA* 2000; **97**: 8005–8009.
14 Jamieson SE, Miller EN, Black GF, Peacock CS, Cordell HJ, Howson JM *et al.* Evidence for a cluster of genes on chromosome 17q11–q21 controlling susceptibility to tuberculosis and leprosy in Brazilians. *Genes Immun* 2004; **5**: 46–57.
15 Miller EN, Jamieson SE, Joberty C, Fakiola M, Hudson D, Peacock CS *et al.* Genome-wide scans for leprosy and tuberculosis susceptibility genes in Brazilians. *Genes Immun* 2004; **5**: 63–67.
16 Baghdadi JE, Orlova M, Alter A, Ranque B, Chentoufi M, Lazrak F *et al.* An autosomal dominant major gene confers predisposition to pulmonary tuberculosis in adults. *J Exp Med* 2006; **203**: 1679–1684.
17 Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect Dis* 2007; **7**: 328–337.
18 Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S *et al.* Variable host–pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 2006; **103**: 2869–2873.
19 Marquet S, Abel L, Hillaire D, Dessein H, Kalil J, Feingold J *et al.* Genetic localization of a locus controlling the intensity of infection by *Schistosoma mansoni* on chromosome 5q31–q33. *Nat Genet* 1996; **14**: 181–184.
20 Rihet P, Traore Y, Abel L, Aucan C, Traore-Leroux T, Fumoux F. Malaria in humans: *Plasmodium falciparum* blood infection levels are linked to chromosome 5q31–q33. *Am J Hum Genet* 1998; **63**: 498–505.
21 Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z *et al.* Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 2001; **29**: 223–228.
22 Negoro K, McGovern DP, Kinouchi Y, Takahashi S, Lench NJ, Shimosegawa T *et al.* Analysis of the IBD5 locus and potential gene–gene interactions in Crohn's disease. *Gut* 2003; **52**: 541–546.
23 Blackwell JM. Genetics of host resistance and susceptibility to intramacrophage pathogens: a study of multicase families of tuberculosis, leprosy and leishmaniasis in North-Eastern Brazil. *Int J Parasitol* 1998; **28**: 21–28.
24 Rosas-Taraco AG, Revol A, Salinas-Carmona MC, Rendon A, Caballero-Olin G, Arce-Mendoza AY. CD14 C(-159)T polymorphism is a risk factor for development of pulmonary tuberculosis. *J Infect Dis* 2007; **196**: 1698–1706.
25 Pacheco E, Fonseca C, Montes C, Zabaleta J, Garcia LF, Arias MA. CD14 gene promoter polymorphism in different clinical forms of tuberculosis. *FEMS Immunol Med Microbiol* 2004; **40**: 207–213.
26 Rodrigues Jr V, Piper K, Couissinier-Paris P, Bacelar O, Dessein H, Dessein AJ. Genetic control of schistosome infections by the SM1 locus of the 5q31–q33 region is linked to differentiation of type 2 helper T lymphocytes. *Infect Immun* 1999; **67**: 4689–4692.
27 Harris J, De Haro SA, Master SS, Keane J, Roberts EA, Delgado M *et al.* T helper 2 cytokines inhibit autophagic control of intracellular *Mycobacterium tuberculosis*. *Immunity* 2007; **27**: 505–517.
28 Rienthong D, Ajawatanawong P, Rienthong S, Smithtikarn S, Akarasewi P, Chaiprasert A *et al.* Restriction fragment length polymorphism study of nationwide samples of *Mycobacterium tuberculosis* in Thailand, 1997–1998. *Int J Tuberc Lung Dis* 2005; **9**: 576–581.
29 Manca C, Reed MB, Freeman S, Mathema B, Kreiswirth B, Barry III CE *et al.* Differential monocyte activation underlies strain-specific *Mycobacterium tuberculosis* pathogenesis. *Infect Immun* 2004; **72**: 5511–5514.
30 Tosh K, Meisner S, Siddiqui MR, Balakrishnan K, Ghei S, Golding M *et al.* A region of chromosome 20 is linked to leprosy susceptibility in a South Indian population. *J Infect Dis* 2002; **186**: 1190–1193.
31 Fu LM. The potential of human neutrophil peptides in tuberculosis therapy. *Int J Tuberc Lung Dis* 2003; **7**: 1027–1032.
32 Peschel A, Sahl HG. The co-evolution of host cationic antimicrobial peptides and microbial resistance. *Nat Rev Microbiol* 2006; **4**: 529–536.
33 Owen HR, Elser M, Cheung E, Gersbach M, Kraus WL, Hottiger MO. MYBBP1a is a novel repressor of NF-kappaB. *J Mol Biol* 2007; **366**: 725–736.

34 Rabbee N, Speed TP. A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* 2006; **22**: 7–12.
35 Ruschendorf F, Nurnberg P. ALOHOMORA: a tool for linkage analysis using 10K SNP array data. *Bioinformatics* 2005; **21**: 2123–2125.
36 O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63**: 259–266.
37 Abecasis GR, Cherny SS, Cookson WO, Cardon LR. GRR: graphical representation of relationship errors. *Bioinformatics* 2001; **17**: 742–743.
38 Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.
39 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
40 Abecasis GR, Wigginton JE. Handling marker–marker linkage disequilibrium: pedigree analysis with clustered markers. *Am J Hum Genet* 2005; **77**: 754–767.
41 Browning BL. FLOSS: flexible ordered subset analysis for linkage mapping of complex traits. *Bioinformatics* 2006; **22**: 512–513.

Supplementary Information accompanies the paper on Genes and Immunity website (http://www.nature.com/gene)