

ORIGINAL ARTICLE

# Diagnostic value of blood gene expression signatures in active tuberculosis in Thais: a pilot study

N Satproedprai<sup>1</sup>, N Wichukchinda<sup>1</sup>, S Suphankong<sup>1</sup>, W Inunchot<sup>1</sup>, T Kuntima<sup>2</sup>, S Kumpeerasart<sup>1</sup>, S Wattanapokayakit<sup>1</sup>, S Nedsuwan<sup>3</sup>, H Yanai<sup>4</sup>, K Higuchi<sup>5,6</sup>, N Harada<sup>5,6</sup> and S Mahasirimongkol<sup>1</sup>

Tuberculosis (TB) is a major global health problem. Routine laboratory tests or newly developed molecular detection are limited to the quality of sputum sample. Here we selected genes specific to TB by a minimum redundancy–maximum relevancy package using publicly available microarray data and determine level of selected genes in blood collected from a Thai TB cohort of 40 active TB patients, 38 healthy controls and 18 previous TB patients using quantitative real-time PCR. *FCGR1A*, *FCGR1B* variant 1, *FCGR1B* variant 2, *APOL1*, *GBP5*, *PSTPIP2*, *STAT1*, *KCNJ15*, *MAFB* and *KAZN* had significantly higher expression level in active TB individuals as compared with healthy controls and previous TB cases ( $P < 0.01$ ). A mathematical method was applied to calculate TB predictive score, which contains the level of expression of seven genes and this score can identify active TB cases with 82.5% sensitivity and 100% specificity as compared with conventional culture confirmation. In addition, TB predictive scores in active TB patients were reduced to normal after completion of standard short-course therapy, which was mostly in concordance with the disease outcome. These findings suggested that blood gene expression measurement and TB Sick Score could have potential value in terms of diagnosis of TB and anti-TB treatment monitoring.

*Genes and Immunity* (2015) 16, 253–260; doi:10.1038/gene.2015.4; published online 12 March 2015

## INTRODUCTION

Tuberculosis (TB) is a global health problem with high complexity in its transmission model, resulting in difficulty for controlling the epidemic. Tuberculosis disease development is caused by years of infection with acid-fast, slow-growing bacilli, *Mycobacterium tuberculosis* (MTB). It is estimated that one-third of the world population are infected with TB but only 10% among those eventually develops active disease.<sup>1</sup> Spread of the disease is fueled by globalization and expansion of previously underestimated reactivation risk for latently infected individuals such as in individuals with HIV or diabetes mellitus, aging population, war and poor nutritional status. In some clinical situations, such as smear-negative tuberculosis, disseminated TB and pediatric TB, when sputum samples are difficult to achieve, it makes accurate diagnosis impossible owing to the limited availability of bacteria for microbiological or even molecular diagnosis.<sup>2</sup>

Recently, blood transcription signatures have been used to discriminate between active tuberculosis, latent tuberculosis infection and healthy control.<sup>3</sup> The neutrophil-driven gene expression signatures were highly differentiated in the UK's active TB population and were validated in the second data set of South African active TB patients. These gene signatures were also confirmed by other concurrent studies carried out in another cohort from South Africa and Germany, highlighting the role of immune responses against TB infection dominated by *CD64* (*FCGR1A* and *FCGR1B*).<sup>4,5</sup> As *CD64* is also suggested as a marker for other systemic bacterial infection and most of the other TB responsive genes are correlated with *CD64* expression,<sup>6–9</sup> this

indicates that a nonspecific host response could be detected in blood during bacterial infection. Within systemic bacterial infections, only tuberculosis is presented with prolonged respiratory or systemic symptoms with few exceptions such as Melioidosis. In the clinical setting, where autoimmune diseases and cancer are the main differential diagnosis of tuberculosis, such kind of gene expression analysis could be very useful as a diagnostic tool to rule in tuberculosis and keep the clinician on high alert for possibility of chronic bacterial infection.

In the original study from UK populations, it seemed that the gene expression signatures is closely related to the disease severity and subsequently disappeared shortly after treatment of tuberculosis.<sup>3</sup> In another detailed study looking exclusively at the dynamic of TB-specific gene expression signatures along the course of anti-TB chemotherapy, it confirmed that the changes in gene expression level were seen as early as after 2 weeks of anti-TB chemotherapy.<sup>10</sup> Minimal set of gene expression should be tested in other populations to confirm this promising finding and accelerate the validation of these biomarkers.

The reported TB-specific gene expression signatures<sup>3</sup> also included *MAFB* and *KAZN*, which were also investigated in a genome-wide association study stating that these genes could have roles in host susceptibility/resistance to tuberculosis.<sup>11,12</sup> *MAFB* and *KAZN* were differentially expressed in active TB patients as compared with healthy individuals in the published data sets albeit lower ranking among the reported signatures.<sup>3</sup> Therefore, we hypothesized that *KAZN* and *MAFB* are associated with active

<sup>1</sup>Department of Medical Sciences, Medical Genetics Center, Medical Life Sciences Institute, Ministry of Public Health, Nonthaburi, Thailand; <sup>2</sup>TB/HIV Research Foundation, Chiangrai, Thailand; <sup>3</sup>Department of Medical Services, Chiangrai Hospital, Chiangrai, Thailand; <sup>4</sup>Department of Clinical Laboratory, Fukujuji Hospital, Tokyo, Japan; <sup>5</sup>Japan Anti Tuberculosis Association, Research Institute of Tuberculosis, Tokyo, Japan and <sup>6</sup>Research Institute of Immune Diagnosis, Tokyo, Japan. Correspondence: Dr S Mahasirimongkol, Department of Medical Sciences, Medical Genetics Center, Medical Life Sciences Institute, Ministry of Public Health, 88/7 Tiwanon Road, Taladkwan, Muang 11000, Nontaburi, Thailand.

E-mail: surakameth@hotmail.com

Received 7 November 2014; revised 17 January 2015; accepted 20 January 2015; published online 12 March 2015

disease activity and should be part of prognostic markers for active TB.

In this study, we analyzed the publicly available microarray data from the UK and South African study<sup>3</sup> using feature selection algorithm to define a minimal set of genes that can discriminate active TB disease from latent TB and healthy status. Then the minimal set of genes was validated using blood samples from a TB cohort in northern Thailand.

## RESULTS

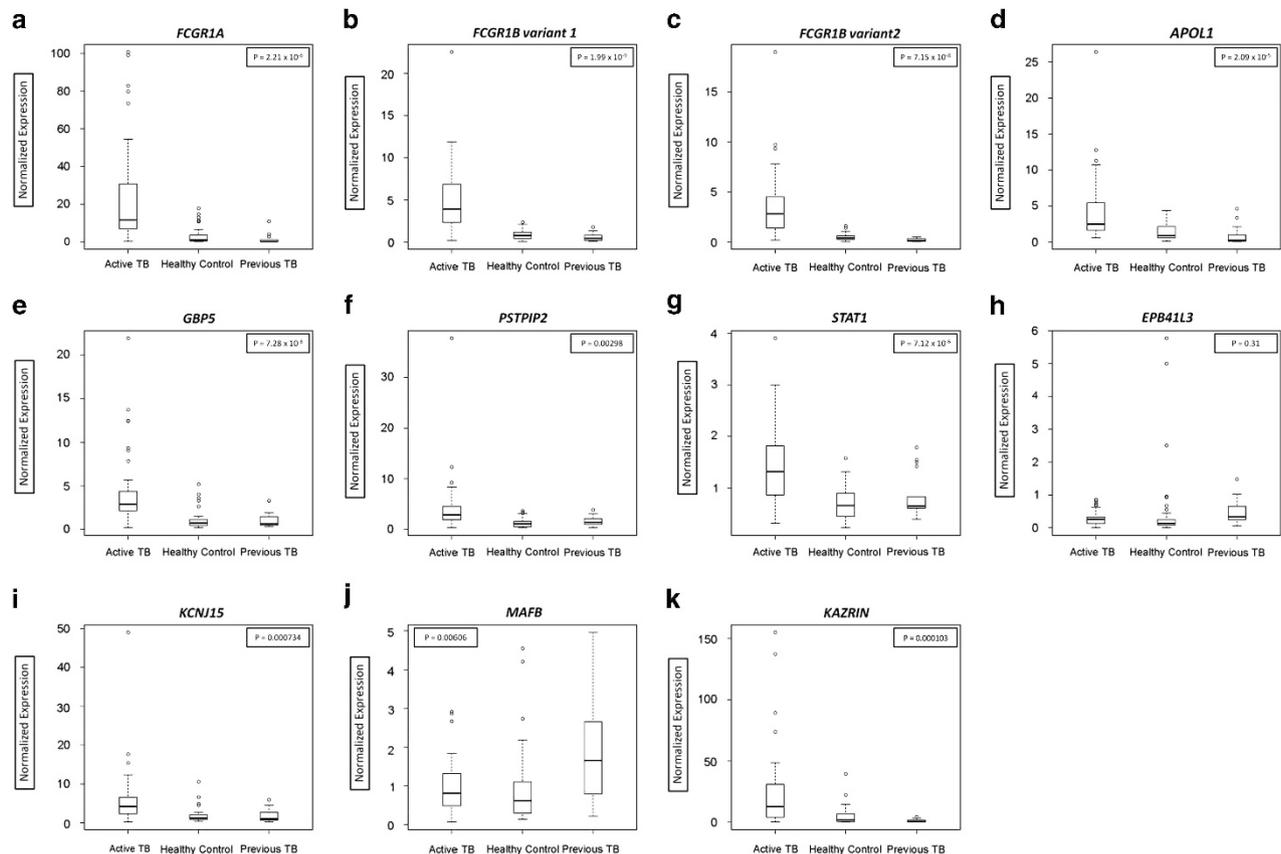
### Subject recruitment, demographic and clinical characteristics

All subjects were recruited and samples were collected from Chiangrai Prachanukroh Hospital in northern Thailand. The total number of subjects in each group was; 40 active TB patients (active TB), 18 previously treated TB patients (previous TB) and 38 healthy controls (healthy control). All TB patients were diagnosed of TB with abnormal chest radiography, positive sputum acid-fast bacilli staining (AFB), positive TB culture confirmation with no resistance to any of the four drugs in the standard TB regimen. All TB patients received standard short-course anti-TB chemotherapy containing rifampicin, isoniazid, pyrazinamide and ethambutol for 2 months following by rifampicin and isoniazid for another 4 months. All previous TB cases were recruited at least 1 year after they completed the anti-TB chemotherapy with cured result and no relapse. Quantiferon gold TB in-tube assay (QFT) were done for all subjects, the QFT-positive percentage in each group were 75%, 77.8% and 15.8% in active TB, previous TB and healthy controls,

respectively. Details of demographic and clinical characteristics were described in Supplementary Table 1 and 2.

Levels of selected genes were differentially expressed in active TB cases as compared to healthy controls and previous TB cases

Analysis using mRMR package yielded top genes that were differentially expressed in active TB individuals across different cohorts with high discriminatory power in previously published data sets<sup>3</sup> (Supplementary Table 4). First, we optimized the assay for measuring transcript levels from 13 target genes that were selected from mRMR package, which included the following genes; *SORT1*, *EPB41L3*, *KCNJ15*, *MAFB*, *PSTPIP2*, *GBP5*, *FCGR1A*, *FCGR1B* variant 1, *FCGR1B* variant 2, *KAZN*, *APOL1*, *STAT1*, *WARS* and two internal control genes; *GAPDH* and *HPRT1*. A real-time two-step TaqMan RT-PCR assay was optimized with single-color or dual-color, custom designed or commercially available probe and primer sets. Assay optimization was carried out in 10 active TB cases and 10 healthy controls. From 13 genes tested, 11 genes were successfully quantifiable. Those genes were *FCGR1A*, *FCGR1B* variant 1, *FCGR1B* variant 2, *APOL1*, *GBP5*, *PSTPIP2*, *STAT1*, *EPB41L3*, *KCNJ15*, *MAFB* and *KAZN*. Then we continued to measure gene expression level of those 11 genes in the total of 40 active TB cases, 38 healthy controls and 18 Previous TB cases (Figure 1). All target genes were differentially expressed in active TB as compared with other groups with statistical significance ( $P < 0.01$ ; Figures 1a–g and Figures 1i–k) except for *EPB41L3* (Figure 1h). The median of relative expression levels of each gene and the median expression fold changes between groups were



**Figure 1.** Boxplots of the level of gene expression in groups of active TB patients (active TB), healthy individuals (healthy controls) and previous TB patients who had completed anti-TB chemotherapy for  $> 1$  year without relapse (previous TB). Levels of gene expressions were normalized with internal control gene, *HPRT1*. The significance tests were carried out using ANOVA and *P*-value of each transcription was presented for each gene. (a) *FCGR1A*, (b) *FCGR1B* variant 1, (c) *FCGR1B* variant 2, (d) *APOL1*, (e) *GBP5*, (f) *PSTPIP2*, (g) *STAT1*, (h) *EPB41L3*, (i) *KCNJ15*, (j) *MAFB* and (k) *KAZN*.

listed in Supplementary Table 5 and 6. Then, the active TB group was further classified based on the number of days between the initiation of standard short-course chemotherapy and the blood sampling date. A significant deviation on *MAFB* and *KAZN* expression levels was observed between active TB cases whose blood samples were collected at day 0 to day 7 after the start of chemotherapy and active TB cases whose blood samples were taken at day 8 to day 15 after initiation of chemotherapy (Figure 2) whereas other genes had no significant deviation (data not shown).

Generation of active TB predictive score using minimum number of transcripts

The predictive logistic regression model 'TB Sick Score' was derived from the information of seven transcripts, which were determined by logistic regression analysis. The model information started to stabilize after the inclusion of four to seven transcripts in the calibration of the model construction and the addition of further transcripts did not provide more discrimination information as shown in Figure 3. List of seven genes and the calculation method for 'TB Sick Score' has been described in Materials and Methods. The constructed model gave us a score threshold at 1.0 to determine the probability of patient having active TB.

Active TB predictive scores generated from the minimum set of transcripts can differentiate active TB patients from healthy controls

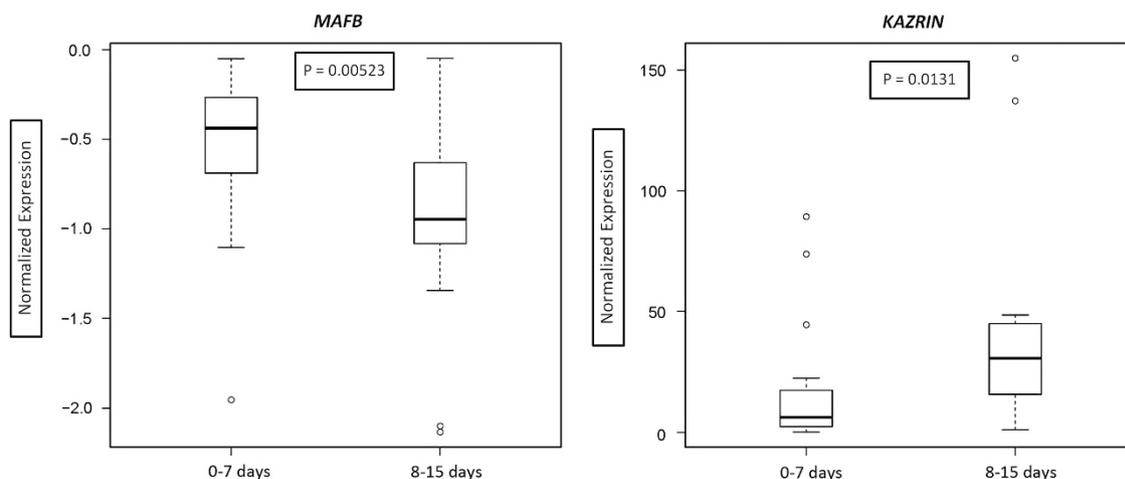
We calculated the active TB predictive score, which was named 'TB Sick Score' for every subject from all three groups. The TB Sick Score in the active TB group were varied with a median of 19.34 ( $0.074\text{--}1.13 \times 10^{13}$ ; Figure 4). Seven from 40 subjects in active TB group had TB Sick Score of  $< 1.0$ . All of them had positive acid-fast bacilli staining (AFB) and culture confirmation. In addition, we compared the clinical diagnosis such as chest X-ray and interferon gamma release assay (IGRA) status between active TB with TB Sick Score  $> 1.0$  and those with score  $< 1.0$ . However, we did not find any major different between these two groups. In contrast, none of the subjects in healthy control or previous TB had TB Sick Score over 1.0. The sensitivity and specificity of TB Sick Score were calculated using the cutoff of 1.0 to identify active TB and we found that the TB Sick Score has 82.5% sensitivity and 100% specificity as compared with clinical diagnosis (Table 1). The positive predictive value (PPV) and negative predictive value (NPV) of TB Sick Score were 100% and 88.89%, respectively

(Table 1). In addition, the receiver-operator curve analysis was applied and the AUC for TB Sick Score was 96.6 with 95% confidence interval of 92.5–100% (Figure 5).

Levels of TB Sick Score in most cases of active TB patients were reduced after completion of standard short-course chemotherapy. All active TB patients were followed up after completion of standard short-course chemotherapy and their treatment outcomes were classified according to WHO guidelines into five groups; cured ( $N=31$ ), treatment completed ( $N=4$ ), death ( $N=4$ ), treatment failure ( $N=4$ ) and treatment intercepted ( $N=3$ ). In addition, three patients were still on extended course of treatment after the standard short-course treatment was finished. Out of 40 active TB participants, we were able to collect another blood sample from 12 active TB cases to monitor the levels of each gene in the minimal set and TB Sick Score after completion of standard short-course chemotherapy. From 12 cases, 8 had reduced post-treatment TB Sick Score (0.178–1.821) whereas 4 cases had increased TB Sick Score as compared with at the time of diagnosis ( $430.64\text{--}1.75 \times 10^{17}$ ; Figure 6). All eight cases that had reduced TB Sick Score were classified as cured (100%). In contrast, four cases had post-treatment TB Sick Score that were significantly higher than the cutoff. When followed up, it was found that one active TB case failed to recover with the TB Sick Score after treatment completion significantly higher than the cutoff value (TB Sick Score =  $1.75 \times 10^{17}$ ),  $\sim 2 \times 10^{15}$  times higher than the TB Sick Score at time of diagnosis (TB Sick Score = 85.867). Another case was still on extended anti-TB chemotherapy with post-treatment TB Sick Score of 618.34, whereas the other two cases were classified as cured.

## DISCUSSION

TB is a disease that affects two-third of the world's population. No effective vaccine against TB is available and treatment of active TB disease is becoming increasingly challenging because of the emergence of multidrug resistant and extremely drug resistant MTB. Diagnostic methods that can identify active TB disease at the early stage and biomarkers that can be used for treatment monitoring are essential for effective administration of anti-TB chemotherapy to prevent drug resistance.<sup>13</sup> New methods were used to study responses to MTB infection and one in particular is the study of host gene expression profile using microarray that brings about the global view of host responses to MTB infection in high resolution.<sup>3,4,14–19</sup> In this study the publicly available microarray data that was published previously<sup>3</sup> was utilized for



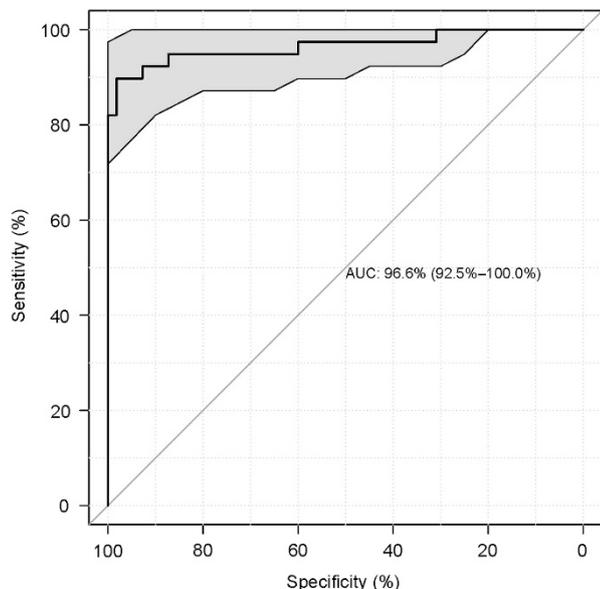
**Figure 2.** Significant deviation of gene expression level of (a) *MAFB* and (b) *KAZN* in the blood taken from active TB subjects who received anti-TB chemotherapy for 0–7 days and the blood taken from active TB subjects who received anti-TB chemotherapy for 8–15 days.



**Table 1.** Diagnostic characteristics of TB Sick Score for active TB classification as compared with diagnostic results (Sputum AFB and TB culture confirmation)

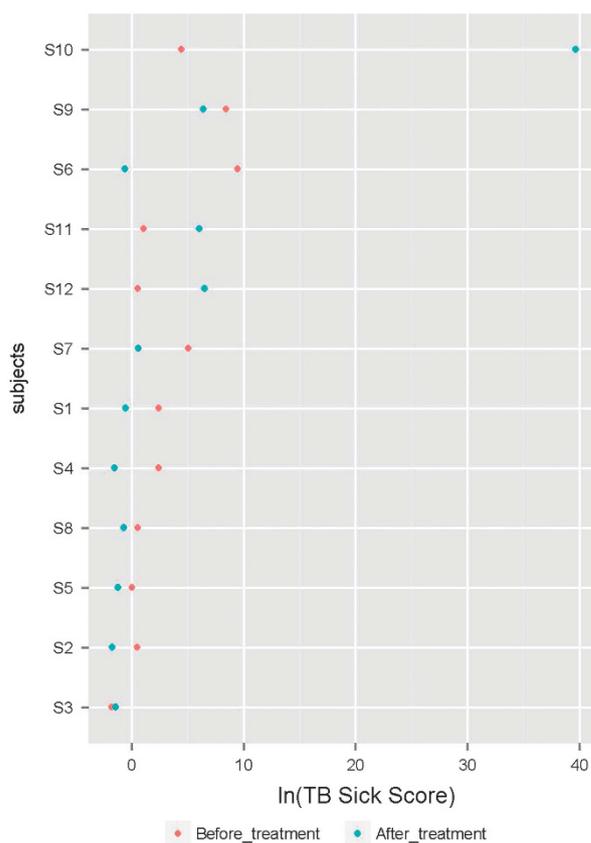
Diagnosis results	Active TB	Healthy control	Previous TB	Total
TB (TB Sick Score $\geq 1.0$ )	33	0	0	33
Not TB (TB Sick Score $< 1.0$ )	7	38	18	63
Total	40	38	18	96

Abbreviation: TB, tuberculosis. Sensitivity 82.5%. Specificity 100%. Positive predictive value (PPV) 100%. Negative predictive value (NPV) 88.89%.



**Figure 5.** Reverse transcription two-step quantitative real-time PCR of seven target genes; *FCGR1A*, *FCGR1B variant 1*, *FCGR1B variant 2*, *MAFB*, *APOL1*, *STAT1* and *KAZN* and a house keeping gene; *HPRT1* were done in 40 active TB cases, 38 healthy controls and 18 previous TB cases. All participants were HIV- Negative. TB Sick Score for all participants were analyzed using receiver operator curve analysis.

reduced after patients have started anti-TB chemotherapy for 2 weeks as compared with the level at diagnosis. Our results also demonstrated that in active TB patients who have received standard short-course chemotherapy for TB for  $< 7$  days tended to have significantly higher levels of *MAFB* and *KAZN* expression than a group of active TB patients who had received anti-TB treatment for  $> 7$  days. This means that, for a diagnostic purpose, gene expression measurement of *MAFB* and *KAZN* and consequently, the reported TB Sick Score should be used in naive patients who have not started a TB regimen or at most within 7 days after a TB regimen had started. This is because in active TB patients, who respond well with the anti-TB chemotherapy, the level of gene expression of *MAFB* and *KAZN* may have changed in such a way that they can affect the sensitivity of this method. In contrast, since the TB-specific gene expression signatures changes in response to anti-TB regimen, it means that these signatures can possibly be used for anti-TB treatment monitoring. From our study, 12 active TB cases were followed and the second blood sample were collected after the completion of anti-TB chemotherapy and queried for TB Sick Score. Most of them had TB Sick Score reduced to normal; however, some patients still had higher TB Sick Score after treatment completion. The treatment conclusion



**Figure 6.** TB Sick Score was calculated using the level of expression of seven target genes normalized by a housekeeping gene; *HPRT1*. TB Sick Score was applied to samples of 12 active TB cases, which were collected at different time points. First, samples were taken within 15 days after TB diagnosis and second, samples were taken after the patients completed standard short-course chemotherapy. Graphical representation shows comparison of TB Sick Score between at the time of diagnosis (red dot) and after the standard treatment finished (blue dot) of each patient.

revealed that some patients who had post-treatment TB Sick Score significantly over the cutoff limit was either having treatment failure or still on extended anti-TB treatment. Nevertheless, in case of patients with increased TB Sick Score after treatment completion, but treatment concluded as cured, it may owe to immune status or other undetected bacterial infections at the second sampling. Nonetheless, those cases were still followed up regularly for possible relapse. This gave us a glimpse that gene expression analysis could be used for monitoring effectiveness of anti-TB treatment for which no test is currently available for this purpose. Hence, further study with a larger sample number and in a different population should be done to increase accuracy of the gene expression analysis method for use as a TB diagnostic test and a longitudinal, prospective study could give a better insight whether gene expression analysis can possibly be used for relapse warning.

The mathematical method to calculate a TB Sick Score was applied in which the score of  $> 1.0$  was associated with active TB disease and none of the healthy controls or the previous TB patients with no relapse has the score of  $> 1.0$ . The TB Sick Score method can identify active TB cases with a high sensitivity and specificity as compared with culture confirmation. Many studies have attempted to apply mathematic calculation of predictive scores using only a small number of genes selected from microarray analysis to differentiate active TB with success.<sup>14,18</sup>

Similar to other studies, our TB Sick Score were strongly correlated with the outcome of the disease, in which patients whose TB treatment outcome were categorized as cured mostly had TB Sick Score at treatment completion < 1.0. This indicates the potential for the TB Sick Score to be used for treatment monitoring. However, because our cohort was focusing on TB patients who have positive AFB and culture with negative HIV status therefore this method should be further proved for its robustness in a real clinical setting with TB–HIV coinfection individuals and patients with diseases that have similar symptoms to TB, to evaluate the real diagnostic value of this test. However, evidence shows an encouraging possibility of arranging a small set of transcriptional biomarkers that can distinguish TB from other diseases, regardless of HIV coinfection or even difference in ethnicity.<sup>18</sup>

Some of the genes that were selected in this study such as *FCGR1A*, *FCGR1B* variant 1 and variant 2 were among the most differentially expressed genes between active TB and healthy control or latent TB across many cohorts in South Africa, Germany and UK.<sup>3,4,18–20</sup> *FCGR1A* and *FCGR1B* genes encode variants of the alpha subunit of CD64 or Fcγ receptor 1 that belongs to a family of surface glycoproteins. Fcγ receptor 1 binds to IgG and induces antibody-dependent cell cytotoxicity, endocytosis and phagocytosis.<sup>2,21,22</sup> It was found that CD64 is upregulated upon bacterial infection and is used as a marker for sepsis.<sup>7,9</sup> Fcγ receptor 1 is found on most cells in myeloid lineage, monocytes and macrophages and is induced by IFN-γ.<sup>21</sup> How Fcγ receptor 1 functions in anti-TB response is still unknown, but recently it was shown to be downregulated by microRNA-127 after an induced lung injury and the downregulation of Fcγ receptor 1 in alveolar macrophage seemed to alleviate lung inflammation.<sup>23</sup> In addition, it was shown that engagement of FcγRI stabilized the complex of leucotriene B4 receptor, BLT1 in the lipid raft and enhanced antimicrobial signals in rat alveolar macrophages.<sup>24</sup> This suggests that Fcγ receptor 1 could have a role in anti-TB response. However, more a detailed study should be conducted to see whether Fcγ receptor 1 could possibly be a target for TB treatment. Even though genes encoding the Fcγ receptor 1 may not be exclusively induced by TB infection but it can be useful in a setting where differential diagnosis of TB from other benign lung disease such as lung carcinoma is needed.

In addition, this study confirmed our previous finding that the level of *MAFB* expression was elevated in active TB individuals as compared with healthy individuals<sup>12,25</sup> with a significant *P*-value (*P*-value = 0.00606). *MAFB* is constitutively expressed in many cell types and it is an important determinant for monocyte–macrophage commitment in hematopoietic progenitor cells.<sup>26</sup> It was reported that *MAFB* is a transcriptional regulator and a type I interferon rheostat that maintains basal level expression of Type I interferon in the resting state.<sup>27</sup> When induced, *MAFB* acts as a negative regulator of Type I interferon upregulation by preventing activation of transcriptional factor IRF3. This leads to the possibility that *MAFB* may have a role in TB-induced regulation of Type I interferon but further functional study is needed to confirm this role of *MAFB*.

Interestingly, when the mRMR was performed to select genes from the data sets of Berry *et al.*, it was found that *APOL1* was among the top genes from the African cohort but not in a cohort from UK. *APOL1* encodes for a serum lipoprotein that binds to the high-density lipoprotein molecule.<sup>28</sup> *APOL1* protein also acts as an innate immune defense against blood parasites, *Trypanosoma brucei* that causes sleeping sickness in Africans, and is thought to be natural immune evolution against the parasites.<sup>29</sup> Certain genetic variations in the *APOL1* gene in people of African descent is associated with protective immunity against *T. brucei* but unfortunately, it is also associated with non-diabetes, chronic kidney disease.<sup>30</sup> Furthermore, *APOL1* is induced by many cytokines and known to be a downstream-expressed transcript in JAK-STAT signaling.<sup>31</sup> *APOL1* is upregulated by multiple proinflammatory cytokines such as IFN-γ, and TNF-α.<sup>32,33</sup> IFN-γ

induced upregulation of *APOL1* in a monocyte cell line was shown to restrict HIV-1 replication. Hence, there are evidences showing that *APOL1* has role in innate immune defense against parasites and HIV but its role in innate immunity against TB is awaiting further investigation.

Evidence from other studies and here have shown that the use of unbiased, global gene expression analysis can identify genes that are signatures of active TB disease. It is promising that these gene expression levels can be used for diagnosis of TB regardless of ethnicity and HIV coinfection status. The set of genes presented in this study was differentially expressed in active TB disease with possible biological function in TB immunity. A more detailed study is needed to further confirm roles of each gene and its contribution in immune response against TB. On the other hand, the mathematical method that we applied to calculate TB Sick Score demonstrated the possible usefulness of gene expression measurement as a method for diagnosis of TB. Nevertheless, the results that have been demonstrated here should be evaluated further in another independent cohort to ensure the diagnostic value of gene expression measurement for identification of active TB. Furthermore, the method for gene expression measurement can be further improved in such a way that it is simplified enough for use in a clinical lab in a limited resources setting.

## MATERIALS AND METHODS

### Ethics statement

This study protocol was reviewed and approved by the institutional review board of Chiangrai Prachanukroh hospital and the Institute for Development of Human Research Protection (IHRP). Informed consents were obtained from all participants in this study.

### Subjects and sample collection

All subjects in this study were enrolled from Chiangrai Prachanukroh Hospital in the northern province of Thailand. Forty patients (33 males, 7 females, median age of 48 (21–79) years) with active tuberculosis (active TB) were recruited. Active TB status was defined by acid-fast staining of sputum sample together with positive culture confirmation and abnormal chest X-ray. All active TB cases were HIV negative at the time of enrollment. Pre- and post-HIV test counseling was given to all subjects. Patients with active TB were given standard short-course chemotherapy. Blood samples were taken from Active TB patients within 15 days after the date when TB was diagnosed and TB regimens were given to these patients. Thirty-eight (24 males, 14 females, median age of 43.5 (25–64) years) healthy control were recruited from regular blood donors whose interferon gamma release assay (IGRAs) status had been determined. IGRAs method is described in Supplementary Information. In addition, 18 previous TB patients (named as “Previous TB”; 12 males, 4 females, median age of 52.5 (23–72) years) who had completed standard short-course chemotherapy and were, free of disease for at least 1 year, were also enrolled to the study to speculate the level of gene expression signatures after TB was treated.

### Gene expression level measurements

**Blood collection.** Three milliliters of venous blood were taken from subjects and injected into a Tempus blood RNA tube for RNA stabilization (Life technologies, Carlsbad, CA, USA). Briefly, whole blood were mixed with the RNA stabilizing solution in the Tempus tube and all cells were lysed to release RNA into RNA stabilizing solution by vigorously shaking the tube for 10–20 s. Blood RNA samples were frozen at –20 °C as soon as possible and then kept until further processes.

**RNA isolation and purification.** Briefly, a total of 9 ml of whole blood plus RNA stabilization buffer was vortexed with 3 ml of PBS in a 50 ml conical tube and centrifuged at 3000 *g*, 4 °C for 30 min then supernatant was carefully decanted leaving the RNA pellet at the bottom of the tube. RNA pellets were resuspended with RNase-free PBS and then subjected to purification using Purelink RNA mini kit (Life technologies, Carlsbad, CA, USA) according to the manufacturer's protocol. All purified RNA samples were quantified using a Nanodrop spectrophotometer (Thermo Scientific, Waltham, MA, USA). All RNA samples gave purity ratio of A260/280 between 1.9 and 2.1.

**Real time reverse transcription quantitative PCR.** RNA samples were reverse transcribed using oligo dT primer and Superscript III reverse transcriptase (Life technologies, Carlsbad, CA, USA). Expression of 13 target genes and two housekeeping genes were measured using TaqMan hydrolysis probe assay on an ABI 7900 real-time PCR machine (Life technologies, Carlsbad, CA, USA). Primers and probes for *SORT1*, *EPB41L3*, *KCNJ15*, *MAFB*, *PSTPIP2*, *GBP5*, *FCGR1A*, *FCGR1B* variant 1, *FCGR1B* variant 2, *KAZN*, *KAZN* isotype E and *GAPDH* were designed using Primer express software version 3.0 (Life technologies, Carlsbad, CA, USA). Primers and probes for *APOL1*, *STAT1* and *HPRT1* are commercially available on the Life technologies website. In-house designed primers and probes for each gene and their information are listed in Supplementary Table 3. We decided to use a two-step multiplex real-time RT-PCR to maximize sample utilization. All multiplex panels were validated and the amplification efficiency of each primer-probe set within a dual-color panel was similar to the efficiency of amplification of a single real-time PCR ( $R^2$  between 0.98 and 0.99). Pooled RNA samples were used as a standard sample to create a five-point standard curve with fivefold serial dilutions starting from 100 ng of RNA per reaction for quantification of each gene transcript. The assays were done in triplicate. All expression data were normalized against a housekeeping gene *HPRT1*.

### Statistical analysis

Transcripts were selected based on three selection strategies. Unbiased selection was made from reanalysis of Berry *et al.*<sup>3</sup> three data sets; GSE19444, GSE 19442 and GSE19439 from the super-series number GSE19491. The data set were retrieved from the Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo/>). The transcripts were selected based on their structure correlation and whether it complements the others. The minimal set of discriminative signatures were selected by applying the minimum redundancy feature selection from microarray data.<sup>34</sup> In each cohort, the transcripts were ranked based on their discriminative probability by feature selection algorithm implement in mRMR. The most significant transcript from each data set were selected, these transcripts were *APOL1*, *WARS* and *EPB41L3*. The overlapping transcripts across all data sets were selected based on lowest average rank from three cohorts and *STAT1*, *SORT1*, *KCNJ15*, *FCGR1A*, *FCGR1B* variant 1, *FCGR1B* variant 2, *PSTPIP2* and *GBP5* were included based on these criteria. Third, *MAFB* and *KAZN* were handpicked from Berry *et al.* data set owing to evidences of their genetic association in Thai tuberculosis populations.<sup>11</sup> In addition to 13 transcripts selected based on described strategies, two housekeeping genes were included in this evaluation for normalization purpose. ANOVA was applied for comparisons of gene expression level between each group, which was active TB, previous TB and healthy control,  $P$ -value < 0.05 were considered significant. On the basis of normalized expression levels of transcripts that were successfully measured, we used Glnet<sup>35</sup> to establish the optimum number of genes to develop the predictive score for active TB. The TB disease predictive score was made available based on logistic regression modeling by the Glnet package.<sup>35</sup>

### Calculation of TB Sick Score

The TB Sick Score were calculated using levels of seven genes; *FCGR1A*, *FCGR1B* variant 1, *FCGR1B* variant 2, *APOL1*, *STAT1*, *MAFB* and *KAZN*. The detail calculation method and equation is described in the Supplementary Information.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### ACKNOWLEDGEMENTS

We thank all participants in this study. In addition, the authors are really appreciated all assistance in subject recruitment and all administrative works from the TB-HIV research foundation. This work was financially supported by the Department of Medical Sciences, the Ministry of Public Health, Thailand, the National Research Council of Thailand and the Japan Society for the Promotion of Sciences. The method for gene expression measurement described in this work is in the process of patenting in the Kingdom of Thailand (No. 1401004750).

### REFERENCES

- O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MP. The immune response in tuberculosis. *Ann Rev Immunol* 2013; **31**: 475–527.
- Weyer K, Carai S, Nunn P. Viewpoint TB diagnostics: what does the world really need? *J Infect Dis* 2011; **204**: S1196–S1202.
- Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010; **466**: 973–977.
- Maertzdorf J, Reipsilber D, Parida SK, Stanley K, Roberts T, Black G *et al.* Human gene expression profiles of susceptibility and resistance in tuberculosis. *Genes Immun* 2011; **12**: 15–22.
- Jacobsen M, Reipsilber D, Gutschmidt A, Neher A, Feldmann K, Mollenkopf HJ *et al.* Candidate biomarkers for discrimination between infection and disease caused by Mycobacterium tuberculosis. *J Mol Med* 2007; **85**: 613–621.
- Streimish I, Bizzarro M, Northrup V, Wang C, Renna S, Koval N *et al.* Neutrophil CD64 as a diagnostic marker in neonatal sepsis. *Pediatr Infect Dis J* 2012; **31**: 777–781.
- Roussel M, Gros A, Sauvadet E, Gacouin A, Marque S, Chimot L *et al.* CD64, a reliable biomarker of bacterial infection in intensive care units? *Am J Respir Crit Care Med* 2012; **186**: 1058.
- Li S, Huang X, Chen Z, Zhong H, Peng Q, Deng Y *et al.* Neutrophil CD64 expression as a biomarker in the early diagnosis of bacterial infection: a meta-analysis. *Int J Infect Dis* 2012; **17**: e12–e23.
- Fjaertoft G, Hakansson L, Ewald U, Foucard T, Venge P. Neutrophils from term and preterm newborn infants express the high affinity Fcγ-receptor I (CD64) during bacterial infections. *Pediatr Res* 1999; **45**: 871–876.
- Bloom CI, Graham CM, Berry MP, Wilkinson KA, Oni T, Rozakeas F *et al.* Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. *PLoS One* 2012; **7**: e46191.
- Mahasirimongkol S, Yanai H, Mushiroda T, Promphittayarat W, Wattanapokayakit S, Phromjai J *et al.* Genome-wide association studies of tuberculosis in Asians identify distinct at-risk locus for young tuberculosis. *J Hum Genet* 2012; **57**: 363–367.
- Mahasirimongkol S, Yanai H, Nishida N, Ridruechai C, Matsushita I, Ohashi J *et al.* Genome-wide SNP-based linkage analysis of tuberculosis in Thais. *Genes Immun* 2009; **10**: 77–83.
- Parida SK, Kaufmann SH. The quest for biomarkers in tuberculosis. *Drug Discov Today* 2010; **15**: 148–157.
- Anderson ST, Kaforou M, Brent AJ, Wright VJ, Banwell CM, Chagaluka G *et al.* Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N Eng J Med* 2014; **370**: 1712–1723.
- Bloom CI, Graham CM, Berry MP, Rozakeas F, Redford PS, Wang Y *et al.* Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. *PLoS One* 2013; **8**: e70630.
- Cliff JM, Lee JS, Constantinou N, Cho JE, Clark TG, Ronacher K *et al.* Distinct phases of blood gene expression pattern through tuberculosis treatment reflect modulation of the humoral immune response. *J Infect Dis* 2013; **207**: 18–29.
- Jacobsen M, Mattow J, Reipsilber D, Kaufmann SH. Novel strategies to identify biomarkers in tuberculosis. *Biol Chem* 2008; **389**: 487–495.
- Kaforou M, Wright VJ, Oni T, French N, Anderson ST, Bangani N *et al.* Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med* 2013; **10**: e1001538.
- Maertzdorf J, Ota M, Reipsilber D, Mollenkopf HJ, Weiner J, Hill PC *et al.* Functional correlations of pathogenesis-driven gene expression signatures in tuberculosis. *PLoS One* 2011; **6**: e26938.
- Sutherland JS, Loxton AG, Haks MC, Kassa D, Ambrose L, Lee JS *et al.* Differential gene expression of activating Fcγ receptor classifies active tuberculosis regardless of human immunodeficiency virus status or ethnicity. *Clin Microbiol Infect* 2014; **20**: O230–O238.
- Siberil S, Duterte CA, Boix C, Bonnin E, Menez R, Stura E *et al.* Molecular aspects of human FcγR interactions with IgG: functional and therapeutic consequences. *Immunol Lett* 2006; **106**: 111–118.
- van der Poel CE, Spaapen RM, van de Winkel JG, Leusen JH. Functional characteristics of the high affinity IgG receptor, FcγRI. *J Immunol* 2011; **186**: 2699–2704.
- Xie T, Liang J, Liu N, Wang Q, Li Y, Noble PW *et al.* MicroRNA-127 inhibits lung inflammation by targeting IgG Fcγ receptor I. *J Immunol* 2012; **188**: 2437–2444.
- Serezani CH, Aronoff DM, Sitrin RG, Peters-Golden M. FcγRII ligation leads to a complex with BLT1 in lipid rafts that enhances rat lung macrophage antimicrobial functions. *Blood* 2009; **114**: 3316–3324.
- Ridruechai C, Mahasirimongkol S, Phromjai J, Yanai H, Nishida N, Matsushita I *et al.* Association analysis of susceptibility candidate region on chromosome 5q31 for tuberculosis. *Genes Immun* 2010; **11**: 416–422.

- 26 Sarrazin S, Mossadegh-Keller N, Fukao T, Aziz A, Mourcin F, Vanhille L *et al*. MafB restricts M-CSF-dependent myeloid commitment divisions of hematopoietic stem cells. *Cell* 2009; **138**: 300–313.
- 27 Kim H, Seed B. The transcription factor MafB antagonizes antiviral responses by blocking recruitment of coactivators to the transcription factor IRF3. *Nat Immunol* 2010; **11**: 743–750.
- 28 Smith EE, Malik HS. The apolipoprotein L family of programmed cell death and immunity genes rapidly evolved in primates at discrete sites of host-pathogen interactions. *Genome Res* 2009; **19**: 850–858.
- 29 Pays E, Vanhollebeke B. Human innate immunity against African trypanosomes. *Curr Opin Immunol* 2009; **21**: 493–498.
- 30 Wasser WG, Tzur S, Wolday D, Adu D, Baumstein D, Rosset S *et al*. Population genetics of chronic kidney disease: the evolving story of APOL1. *J Nephrol* 2012; **25**: 603–618.
- 31 Hartman SE, Bertone P, Nath AK, Royce TE, Gerstein M, Weissman S *et al*. Global changes in STAT target selection and transcription regulation upon interferon treatments. *Genes Dev* 2005; **19**: 2953–2968.
- 32 Monajemi H, Fontijn RD, Pannekoek H, Horrevoets AJ. The apolipoprotein L gene cluster has emerged recently in evolution and is expressed in human vascular tissue. *Genomics* 2002; **79**: 539–546.
- 33 Zhaorigetu S, Wan G, Kaini R, Jiang Z, ApoL1 Hu CA. a BH3-only lipid-binding protein, induces autophagic cell death. *Autophagy* 2008; **4**: 1079–1082.
- 34 Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. *J Bioinform Comput Biol* 2005; **3**: 185–205.
- 35 Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010; **33**: 1–22.

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