**Supplementary Figure 1-5**  
**Supplementary Figure 1. RNA electrophoresis, amplification curve of qRT-PCR and standard curve of control cDNA  
Supplementary Figure 2. Hierarchical clustering and volcano plot for differentially expressed lncRNA profiles in the Screening Cohort  
Supplementary Figure 3. LncRNA expressions between clinically diagnosed PTB patients and non-TB disease controls in the Selection and Validation Cohorts Supplementary Figure 4. qPCR analysis of *ENST00000497872* expression in BCG-infected THP-1 cells**

**Supplementary Figure 5. Ten-fold cross-validation ROC of "EHR+lncRNA" model developed using the data from the Selection Cohort.**

** Supplementary Figure 1.** **RNA electrophoresis, amplification curve of qRT-PCR and standard curve of control cDNA.**

(A), RNA electrophoresis. Marker: DNA marker (100-2000bp).

(B), Amplification curve of qRT-PCR. The left set of amplification curves indicates the endogenous control *GAPDH*, and the right set of amplification curves indicates the lncRNA gene.

(C), Standard curve of control cDNA.Cq values (i.e., crossing points) are plotted versus the log10 [cDNA quantity]. Some information for the standard curve: Y = -3.252 X + 36.524, R2 = 0.9973, reaction efficiency = 2.030, error = 0.0154.



**Supplementary Figure 2. Hierarchical clustering and volcano plot for differentially expressed lncRNA profiles in the Screening Cohort.**

(A), Heatmap of lncRNA profile generated by hierarchical clustering. Pink bar, clinically diagnosed PTB patients with negative MTB evidence, herein referred to as NPTB for short; cyan bar, healthy subjects (HS, for short). Downregulated lncRNAs are depicted in green and upregulated in orange.

(B), Volcano plot. The signal values were distributed in the corresponding area after the data normalization. The green dots and orange dots are significantly down- and upregulated lncRNA distributions, respectively.



**Supplementary Figure 3. LncRNA expressions between clinically diagnosed PTB patients and non-TB disease controls in the Selection and Validation Cohorts.** Clinically diagnosed PTB patients with negative MTB evidence, herein referred to as NPTB for short. *ENST00000497872* and *n333737* show statistical significance (p-value < 0.0001 after age-adjustment for both the Selection and Validation Cohorts) in group comparison, and n335265 indicates a p-value of 0.080 and 0.110 in the Selection Cohort and Validation Cohort, respectively. These error bar plots demonstrate the median and interquartile range. Mann-Whitney U test was used to compare the differences between groups.

../../流式%20和参考文献/1%20Supplemental%20Figure%208%20lncRNA%20AL.pdf

**Supplementary Figure 4. qPCR analysis of *ENST00000497872* expression in BCG-infected THP-1 cells.** Following infection with BCG from 6 h to 72 h, the relative expression level of *ENST00000497872* (named lnc AL, for short) was significantly elevated in LV-lnc AL cells and reduced in shRNA-lnc AL cells. Negative control group, BCG treated THP-1 cells with stably expressing LV-control or shRNA-control; shRNA-lnc AL group, BCG treated THP-1 cells with stably expressing shRNA-lnc AL; LV-lnc AL group, BCG treated THP-1 cells with stably overexpressing LV-lnc AL.

The expression results of lnc AL in BCG treated THP-1 cells with stably expressing LV-control and shRNA-control are similar, and the average values of these two kinds of negative controls are regarded as the values in the negative control; Blank control (BCG treated THP-1 cells) was not shown in the figure. Results shown are the average of triplicate experiments.

****

**Supplementary Figure 5. Ten-fold cross-validation ROC of "EHR+lncRNA" model developed using the data from the Selection Cohort.**

The "EHR+lncRNA" models were trained using data from the Selection Cohort. The AUC of "EHR+lncRNA" models was 0.89 (0.84-0.93). Sensitivity and specificity of the "EHR+lncRNA" model for prediction of clinically diagnosed PTB was 0.89 (0.82-0.93) and 0.80 (0.73-0.85) at a cutoff of 0.37. The blue line represents the ROC for the entire training dataset. The grey lines refer to the ROC developed for each of the 10-fold cross-validations.