

Primary category: Basic Research

Title: Identification of a six-cytokine biosignature discriminating active tuberculosis from latent infection.

Authors and Affiliations:

**Vaishnavi Kaipilyawar**<sup>1\*</sup>, Jason Zhao<sup>\*2,3</sup>, Noyal M. Joseph<sup>4</sup>, Sonali Sarkar<sup>4</sup>, C. Robert Horsburgh<sup>5,6</sup>, Natasha S. Hochberg<sup>5,6,7</sup>, Padmini Salgame<sup>1</sup>, W. Evan Johnson<sup>2,3</sup>, and Jerrold J Ellner<sup>1</sup>

\*Co-first authors

<sup>1</sup> Department of Medicine, Center for Emerging Pathogens, Rutgers-New Jersey Medical School, Newark, NJ, USA

<sup>2</sup> Bioinformatics Program, Boston University, Boston, MA, USA

<sup>3</sup> Division of Computational Biomedicine, Boston University School of Medicine, Boston, MA, USA

<sup>4</sup> Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India

<sup>5</sup> Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA

<sup>6</sup> Section of Infectious Diseases, Boston University School of Medicine, Boston, MA, USA

<sup>7</sup> Boston Medical Center, Boston, MA, USA

Background: The high cost and logistical difficulty of current sputum-based diagnostic tests impedes Tuberculosis (TB) diagnosis in high-burden, resource-poor settings. The development of a cost-effective and rapid triage test would significantly improve the efficiency of TB detection and treatment, further facilitating the reduction in the global TB burden. Our goal was to develop a blood-based biosignature that segregated tuberculosis from latent infection (LTBI) to be subsequently deployed as a triage test.

Methods: Multiplexed, quantitative detection of 23 cytokines/chemokines was performed on plasma samples from a South Indian dataset of 155 TB cases and 208 individuals with LTBI, using Meso Scale Discovery's custom V-plex assays. A random stratified sampling was first applied to split the dataset into a training set (80%) and a validation set (20%). From the training set, an ensemble feature selection pipeline (lasso logistic regression, random forest, and xgboost models and leave-one-out cross-validation) was used to generate a parsimonious cytokine biomarker with robust and unbiased power in differentiating TB from LTBI. An ensemble learning classifier from the training set was then used to evaluate the biomarker's quantitative performance in the validation set.

Results: A biomarker comprising six cytokines (Eotaxin, IFN $\gamma$ , IL-15, IL-1 $\beta$ , IL-6, and IP-10) 'CYTO6' was derived from the training set. When tested in the validation set, CYTO6 achieved 0.99 AUC, 100% sensitivity, 92.68% specificity, 60.29% PPV and 100% NPV.

Conclusions: CYTO6 demonstrates strong performance in distinguishing TB from LTBI and thus shows promise as a rapid triage test. Future studies will evaluate the performance of CYTO6 as a community-based triage test.