27th ECCMID Vienna, Austria 22 – 25 April 2017
The congress of ESCMID

Session: OS174 Let's diagnose parasites

Category: 7a. Diagnostic parasitology

25 April 2017, 09:24 - 09:34
OS0858

A robust host-based gene expression diagnostic for malaria versus other infectious diseases

Aditya Rao[1], Timothy Sweeney[2], Purvesh Khatri*[3]

[1]*Stanford University*

[2]*Stanford University; Institute for Immunity, Transplantation and Infections*

[3]*Stanford University; Medicine*

Background: The global incidence of malarial infection is approximately 200 million, with annual mortality of at least 438,000. While early diagnosis of malaria can lead to rapid treatment and cure, the similarity of symptoms to other infectious diseases and the long incubation time causes delayed and incorrect diagnoses. The gold-standard diagnostic, blood smear, is time-consuming and not widely available. Rapid molecular tests for malarial parasites suffer from inaccuracy and relatively poor sensitivity. We have previously shown that diagnostics based on host gene expression can be an effective way to diagnose infectious diseases such as tuberculosis, sepsis, and bacterial infections. Here we hypothesized that transcriptomic profiles of patients infected with malaria would yield a robust diagnostic for malaria versus other infectious diseases.

Material/methods: We performed a systematic search for gene expression datasets profiling malaria or other common tropical infectious diseases (dengue, typhoid, and other parasitic infections) using either PBMCs or whole blood. We used our previously-described COCONUT co-normalization platform to pool gene expression data from all 10 cohorts into a single group. This group was randomly split into 70% / 30% training/validation groups. We then used machine learning to derive a set of genes optimized for diagnostic power in the training group, which we then tested in the held-out validation group.

Results: Our systematic search identified 10 datasets composed of 846 patient samples (177 acute malaria, 214 controls, 217 acute dengue, 119 typhoid, 103 brucellosis, 16 leishmaniasis). After COCONUT normalization, the pooled data had a total of 6,212 host genes in common. We trained multiple penalized regression models and found that a core group of 7 host genes were present in most regression models. We then ran logistic regression based on those 7 genes comparing malaria to other infectious diseases. This model had an AUROC of 0.951 in the training data. When we applied the model to the held-out test data, the 7-gene model had an AUROC of 0.962. Four of the genes included were from well-known pro-inflammatory pathways, while three are not traditionally associated with inflammatory responses.

Conclusions: We here show that a small set of host genes can be used to form a robust classifier for malaria vs. other infectious diseases that are within the differential diagnosis for malarial symptoms. As has been shown with the GeneXpert system (among others), gene-expression based diagnostics can be made to be rapid and affordable enough to become useful in low- and middle-income countries. Our 7-gene classifier will need prospective validation prior to clinical translation.