Multicohort analysis of genome-wide expression for diagnosis of tuberculosis

Purvesh Khatri
Assistant Professor
Institute for Immunity, Transplantation and Infection
Division of Biomedical Informatics Research
Department of Medicine
Stanford University

Email: pkhatri@stanford.edu
Twitter: @purveshkhatri
Traditional approach - reduce heterogeneity

- Single cohort
  - Clinical homogeneity
  - Minimize technical variance
  - Internal validation

- Does not capture heterogeneity of a disease
- Results are difficult to generalize
Embrace heterogeneity

- “Dirty” data - multiple datasets asking the same question
  - Clinical heterogeneity
  - Different treatments
  - Different technologies

- Generalizable results
- Unexpected results are more “believable”
- “Dirty data” - integration is challenging
Framework for leveraging heterogeneity

1. Data from public domain
   - Guideline: Re-normalize and log2 transform chosen datasets

2. Clinical question and systematic search
   - Guideline: At least 4-5 datasets, with close to even split between cases and controls

3. Split into discovery and validation cohorts
   - Guideline: Keep datasets with rich annotations as validation if possible
   - Guideline: Never separate datasets from the same research group between discovery and validation

4. Meta-analysis
   - Thresholds of at least:
     - FDR < 0.01-0.05
     - Effect Size > 1.3

5. Validate analysis in independent data

Sweeney et al. NAR 2016
Translational Medicine using Public Data

Diagnostic and prognostic markers

- Common rejection module across all solid organs
- Sepsis diagnosis 2-to-5 days prior
- Common host response to multiple viral infections
- Tuberculosis – satisfies WHO TPP
- Bacterial vs viral

Sweeney et al. Sci Trans Med 2015
Andres-Terre et al. Immunity 2015

Novel Drug Targets
Lung and Pancreatic Cancer

Mazur et al. Nature 2014
Chen et al. Cancer Res 2014

Drug repurposuring
Organ Transplant
Viral Infections

PTK7

Lofgren et al. (under submission)
Meeting Report

High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting

28–29 April 2014
Geneva, Switzerland

World Health Organization
Executive summary

- a **point-of-care non-sputum-based test** capable of detecting all forms of TB by identifying characteristic biomarkers or biosignatures (known as the biomarker test);

- a point-of-care **triage test**, which should be a **simple, low-cost test** that can be used by first-contact health-care providers to identify those who need further testing (the triage test);

- a point-of-care sputum-based test to replace smear microscopy for detecting pulmonary TB (the smear-replacement test);

- a rapid **drug-susceptibility test** that can be used at the microscopy-centre level of the health-care system to select first-line regimen-based therapy (the rapid DST test).
## Table: Summary of all datasets that matched inclusion criteria (whole blood, clinically active pulmonary tuberculosis)

<table>
<thead>
<tr>
<th>Year</th>
<th>Reference</th>
<th>Platform</th>
<th>Use</th>
<th>Country</th>
<th>Age</th>
<th>HIV status</th>
<th>Active tuberculosis</th>
<th>Latent tuberculosis disease</th>
<th>Disease</th>
<th>Other disease breakdown</th>
<th>Treatment</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019</td>
<td>Bery et al</td>
<td>GPL947</td>
<td>USA</td>
<td>South Africa</td>
<td>96</td>
<td>positive</td>
<td>24</td>
<td>33</td>
<td>59</td>
<td>Other disease breakdown: 28 ASAS, 87 SUL, 31 SUL, 59 other infections.</td>
<td>44 treatment samples not used.</td>
<td>405</td>
</tr>
<tr>
<td>2013</td>
<td>Merceño et al</td>
<td>GPL417</td>
<td>USA</td>
<td>South Africa</td>
<td>19</td>
<td>negative</td>
<td>6</td>
<td>86</td>
<td>94</td>
<td>Other disease breakdown: 90 patients with tuberculosis.</td>
<td>157 treatment measured at 12 and 26 weeks.</td>
<td>379</td>
</tr>
<tr>
<td>2013</td>
<td>Cliff et al</td>
<td>GSE4924</td>
<td>USA</td>
<td>South Africa</td>
<td>19</td>
<td>Negative</td>
<td>157</td>
<td>117</td>
<td>334</td>
<td>Other disease breakdown: 186 patients with tuberculosis.</td>
<td>157 treatment measured at 12 and 26 weeks.</td>
<td>534</td>
</tr>
<tr>
<td>2012</td>
<td>Merceño et al</td>
<td>GPL4133</td>
<td>USA</td>
<td>South Africa</td>
<td>19</td>
<td>Negative</td>
<td>157</td>
<td>117</td>
<td>334</td>
<td>Other disease breakdown: 186 patients with tuberculosis.</td>
<td>157 treatment measured at 12 and 26 weeks.</td>
<td>534</td>
</tr>
<tr>
<td>2012</td>
<td>Merceño et al</td>
<td>GPL4133</td>
<td>USA</td>
<td>South Africa</td>
<td>19</td>
<td>Negative</td>
<td>157</td>
<td>117</td>
<td>334</td>
<td>Other disease breakdown: 186 patients with tuberculosis.</td>
<td>157 treatment measured at 12 and 26 weeks.</td>
<td>534</td>
</tr>
<tr>
<td>2012</td>
<td>Merceño et al</td>
<td>GPL4133</td>
<td>USA</td>
<td>South Africa</td>
<td>19</td>
<td>Negative</td>
<td>157</td>
<td>117</td>
<td>334</td>
<td>Other disease breakdown: 186 patients with tuberculosis.</td>
<td>157 treatment measured at 12 and 26 weeks.</td>
<td>534</td>
</tr>
<tr>
<td>2012</td>
<td>Merceño et al</td>
<td>GPL4133</td>
<td>USA</td>
<td>South Africa</td>
<td>19</td>
<td>Negative</td>
<td>157</td>
<td>117</td>
<td>334</td>
<td>Other disease breakdown: 186 patients with tuberculosis.</td>
<td>157 treatment measured at 12 and 26 weeks.</td>
<td>534</td>
</tr>
<tr>
<td>2012</td>
<td>Merceño et al</td>
<td>GPL4133</td>
<td>USA</td>
<td>South Africa</td>
<td>19</td>
<td>Negative</td>
<td>157</td>
<td>117</td>
<td>334</td>
<td>Other disease breakdown: 186 patients with tuberculosis.</td>
<td>157 treatment measured at 12 and 26 weeks.</td>
<td>534</td>
</tr>
<tr>
<td>2012</td>
<td>Merceño et al</td>
<td>GPL4133</td>
<td>USA</td>
<td>South Africa</td>
<td>19</td>
<td>Negative</td>
<td>157</td>
<td>117</td>
<td>334</td>
<td>Other disease breakdown: 186 patients with tuberculosis.</td>
<td>157 treatment measured at 12 and 26 weeks.</td>
<td>534</td>
</tr>
<tr>
<td>2012</td>
<td>Merceño et al</td>
<td>GPL4133</td>
<td>USA</td>
<td>South Africa</td>
<td>19</td>
<td>Negative</td>
<td>157</td>
<td>117</td>
<td>334</td>
<td>Other disease breakdown: 186 patients with tuberculosis.</td>
<td>157 treatment measured at 12 and 26 weeks.</td>
<td>534</td>
</tr>
<tr>
<td>2012</td>
<td>Merceño et al</td>
<td>GPL4133</td>
<td>USA</td>
<td>South Africa</td>
<td>19</td>
<td>Negative</td>
<td>157</td>
<td>117</td>
<td>334</td>
<td>Other disease breakdown: 186 patients with tuberculosis.</td>
<td>157 treatment measured at 12 and 26 weeks.</td>
<td>534</td>
</tr>
<tr>
<td>2012</td>
<td>Merceño et al</td>
<td>GPL4133</td>
<td>USA</td>
<td>South Africa</td>
<td>19</td>
<td>Negative</td>
<td>157</td>
<td>117</td>
<td>334</td>
<td>Other disease breakdown: 186 patients with tuberculosis.</td>
<td>157 treatment measured at 12 and 26 weeks.</td>
<td>534</td>
</tr>
<tr>
<td>2012</td>
<td>Merceño et al</td>
<td>GPL4133</td>
<td>USA</td>
<td>South Africa</td>
<td>19</td>
<td>Negative</td>
<td>157</td>
<td>117</td>
<td>334</td>
<td>Other disease breakdown: 186 patients with tuberculosis.</td>
<td>157 treatment measured at 12 and 26 weeks.</td>
<td>534</td>
</tr>
<tr>
<td>2012</td>
<td>Merceño et al</td>
<td>GPL4133</td>
<td>USA</td>
<td>South Africa</td>
<td>19</td>
<td>Negative</td>
<td>157</td>
<td>117</td>
<td>334</td>
<td>Other disease breakdown: 186 patients with tuberculosis.</td>
<td>157 treatment measured at 12 and 26 weeks.</td>
<td>534</td>
</tr>
<tr>
<td>2012</td>
<td>Merceño et al</td>
<td>GPL4133</td>
<td>USA</td>
<td>South Africa</td>
<td>19</td>
<td>Negative</td>
<td>157</td>
<td>117</td>
<td>334</td>
<td>Other disease breakdown: 186 patients with tuberculosis.</td>
<td>157 treatment measured at 12 and 26 weeks.</td>
<td>534</td>
</tr>
<tr>
<td>2012</td>
<td>Merceño et al</td>
<td>GPL4133</td>
<td>USA</td>
<td>South Africa</td>
<td>19</td>
<td>Negative</td>
<td>157</td>
<td>117</td>
<td>334</td>
<td>Other disease breakdown: 186 patients with tuberculosis.</td>
<td>157 treatment measured at 12 and 26 weeks.</td>
<td>534</td>
</tr>
</tbody>
</table>

### Notes
- **DUSP3, GBP5, KLF2**
- 11 countries
- 14 cohorts
- 2,572 samples

### Genes

- **3 genes**
  - DUSP3
  - GBP5
  - KLF2
ATB Diagnosis vs healthy, LTB and other diseases

sensitivity = 86%; specificity = 86%; NPV = 99% @ 10% prevalence

Not confounded by HIV co-infection

May allow monitoring treatment response

Not confounded by BCG vaccination

A 3-gene signature distinguishes ATB in prospective cohorts

Zak et al. Lancet 2016
Adolescents
LTB vs ATB
RNAseq
3-gene signature distinguishes ATB in prospective cohorts

Adolescents
LTB vs ATB
RNAseq

Zak et al. *Tuberculosis* 2017
Adults
ATB vs controls
RNAseq
3-gene signature distinguishes ATB in prospective cohorts

Adolescents
LTB vs ATB
RNAseq

Zak et al. *Tuberculosis* 2017
Adults
ATB vs controls
RNAseq

Warsinske et al.
Active screen in adults
ATB vs controls
PCR
3-gene signature distinguishes ATB in prospective cohorts

Table 3  Maximized sensitivity values obtained from the ROC analysis of GBP5, DUSP3 and KLF2 combinations in WB cohort test.  

<table>
<thead>
<tr>
<th></th>
<th>GBP5</th>
<th>DUSP3</th>
<th>KLF2</th>
<th>GBP5,DUSP3</th>
<th>GBP5,KLF2</th>
<th>DUSP3,KLF2</th>
<th>GBP5,DUSP3,KLF2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATB vs HC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>0.85</td>
<td>0.73</td>
<td>0.62</td>
<td>0.84</td>
<td>0.86</td>
<td>0.77</td>
<td>0.85</td>
</tr>
<tr>
<td>95%CI</td>
<td>0.81-0.90</td>
<td>0.67-0.78</td>
<td>0.56-0.68</td>
<td>0.80-0.89</td>
<td>0.82-0.91</td>
<td>0.72-0.82</td>
<td>0.81-0.89</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>80.6%</td>
<td>61.8%</td>
<td>31.3%</td>
<td>77.8%</td>
<td>77.8%</td>
<td>66.0%</td>
<td>85.5%</td>
</tr>
<tr>
<td>Specificity</td>
<td>90.9%</td>
<td>78.0%</td>
<td>96.7%</td>
<td>89.5%</td>
<td>87.1%</td>
<td>82.3%</td>
<td>70.8%</td>
</tr>
</tbody>
</table>

OD vs ATB; PCR
3-gene signature predicts progression from LTB to ATB
3-gene signature detected for the spectrum of a *Mtb* infection

Predicts transition from LTB to ATB *6 months* prior (ACS cohort)

Identifies *treatment failure* at end-of-treatment (Catalysis cohort)

May also identify sub-clinically active TB

Active TB vs healthy, LTB, other Lung Dx (Brazil cohort, China cohort)

Healthy or Latent TB
Acknowledgements

Jason Andrews
Julio Croda
Tim Sweeney

SPADA

National Institute of Allergy and Infectious Diseases
Leading research to understand, treat, and prevent infectious, immunologic, and allergic diseases.

Human Immunology Project Consortium
Cooperative Centers on Human Immunology