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

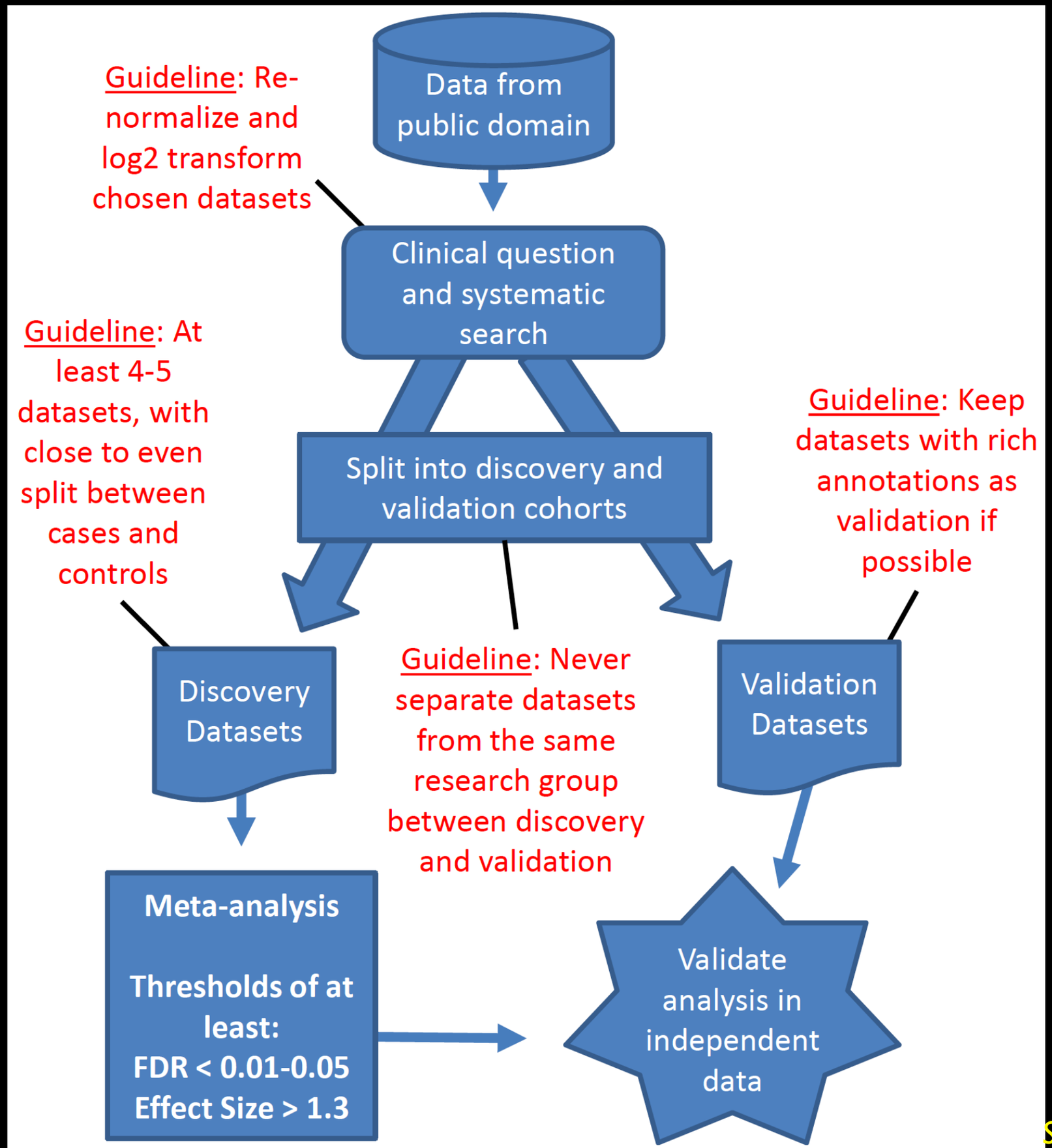
- Does not capture heterogeneity of a disease
- Results are difficult to generalize

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

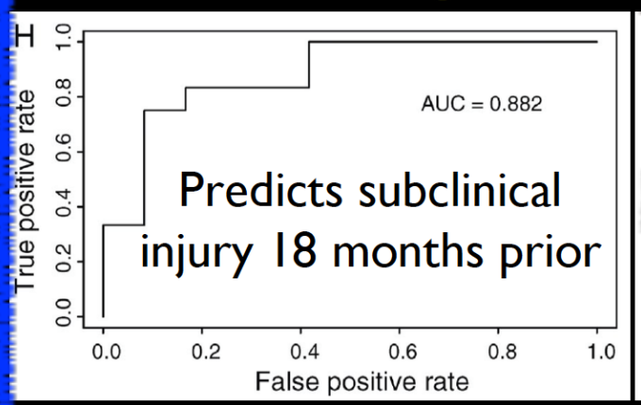


Sweeney et al. NAR 2016

Translational Medicine using Public Data

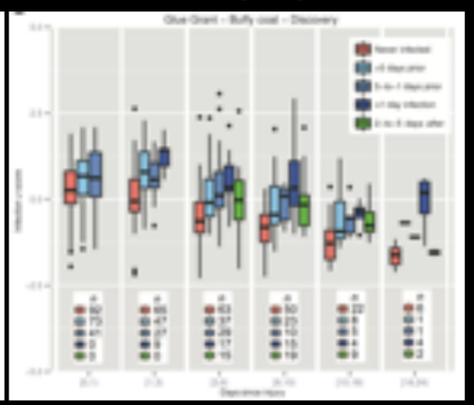
Diagnostic and prognostic markers

Common rejection module across all solid organs



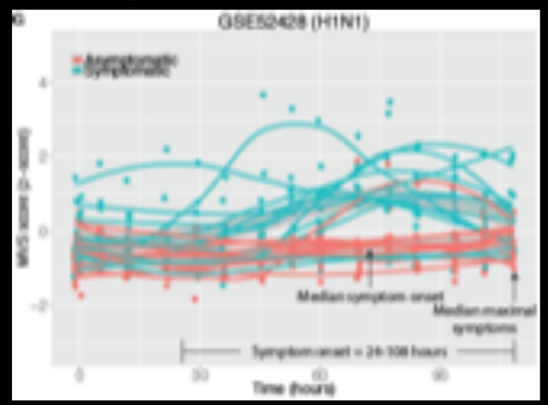
Khatri *et al.* *J Exp Med* 2013

Sepsis diagnosis 2-to-5 days prior



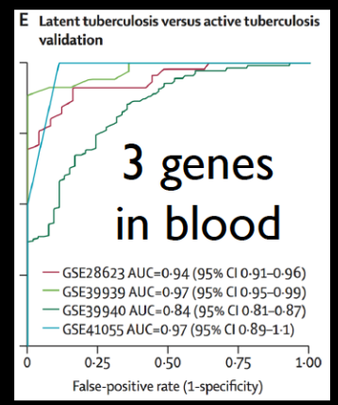
Sweeney *et al.* *Sci Trans Med* 2015

Common host response to multiple viral infections



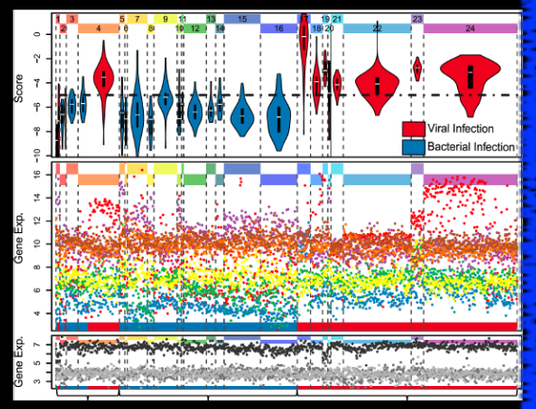
Andres-Terre *et al.* *Immunity* 2015

Tuberculosis – satisfies WHO TPP



Sweeney *et al.* *Lancet Res Med 2016*

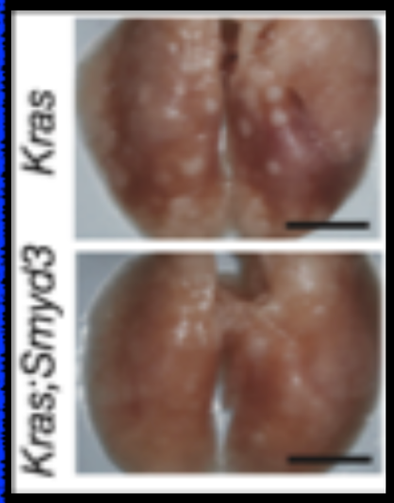
Bacterial vs viral

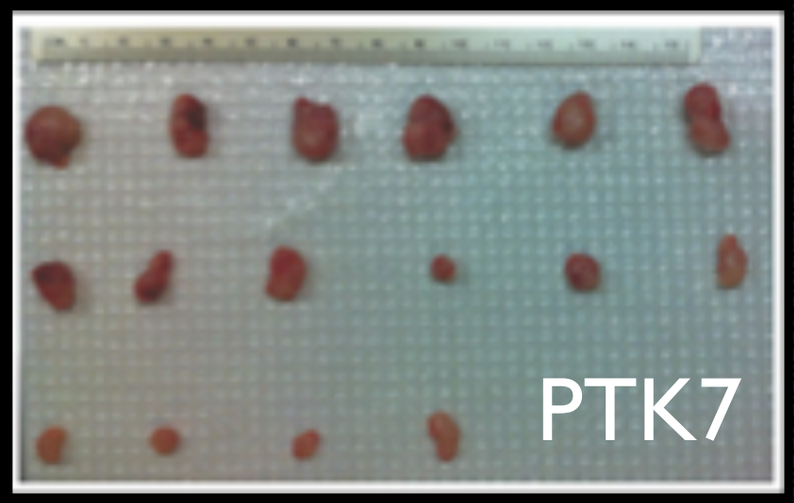


Sweeney *et al.* *Sci Trans Med 2016*

Novel Drug Targets

Lung and Pancreatic Cancer

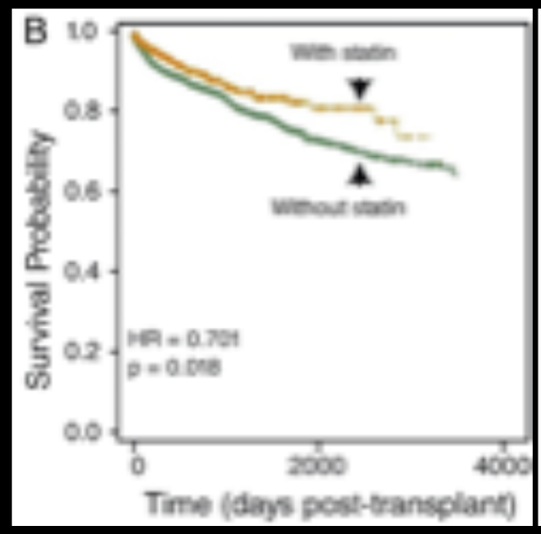




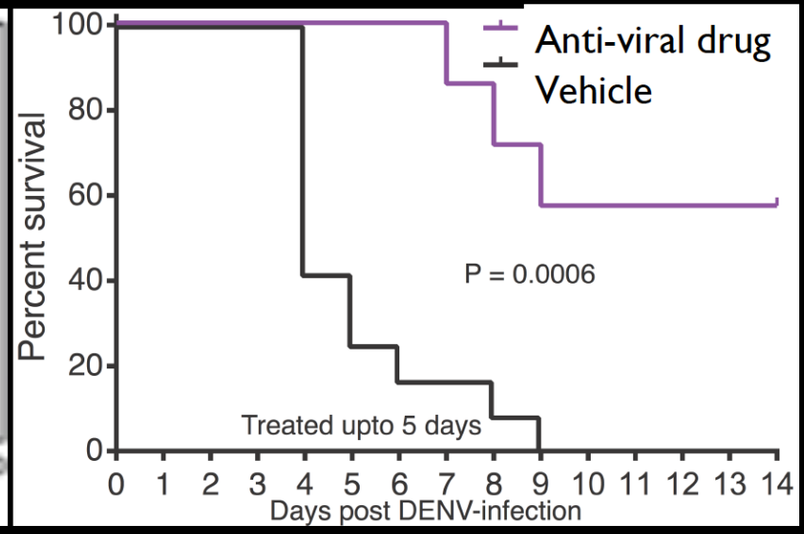
Mazur *et al. Nature* 2014 Chen *et al. Cancer Res* 2014

Drug repurposing

Organ Transplant



Viral Infections



Khatri *et al. J Exp Med* 2013 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).

	Year	Reference	Platform	Use	Country	Age	HIV status	Active tuberculosis culture or smear	Healthy controls	Latent tuberculosis	Other disease	Active tuberculosis	Treatment	Total	Miscellaneous
GSE19491	2010	Berry[8]	GPL6947	Discovery	South Africa, UK, USA	Adults	Negative	Positive	86	69	193	31	..	409	Other disease breakdown: 28 ASLE, 82 PSLE, 31 Still's, 52 *Streptococcus* and/or *Staphylococcus* infection; post-treatment samples not used.
GSE25534	2010	Maertzdorf[30]	GPL1708	Validation	South Africa	Adults	Negative	Positive	6	19	..	19	..	44	Two-colour array (on-chip comparisons between healthy controls, latent tuberculosis, and active tuberculosis)
GSE28623	2011	Maertzdorf[22]	GPL4133/ GPL6480	Validation	The Gambia	Adults	Negative	Positive	7	25	..	46	..	108	..
Cliff Combined Dataset	2013	Cliff[13]	GPL570	Validation	South Africa	Adults	Negative	Positive	..	..	..	36	117	153	Treatment measured at 1, 2, 4, and 26 weeks
GSE34608	2012	Maertzdorf[24]	GPL4133/ GPL6480	Validation	Germany	Adults	Negative	Positive	18	..	..	8	..	44	Other diseases all sarcoid
GSE37250	2014	Kaforou[7]	GPL10558	Discovery	Malawi, South Africa	Adults	Positive and negative	Positive	..	167	175	195	..	537	See reference for other disease distributions; 194 patients with other diseases reported but only 175 available with microarrays.
GSE39939	2014	Anderson[6]	GPL10558	Validation	Kenya	Child-ren	Positive and negative	Positive and negative	..	1..	6..	4.. negative, 3.. positive	..	157	Other diseases breakdown: 33 pneumonia, 5 sepsis, 7 malnutrition, 19 other
GSE39940		Anderson[6]		Validation	Malawi, South Africa	Child-ren	Positive and negative	Positive	..	54	169	111	..	334	Other diseases breakdown: 86 pneumonia, 8 CLD, 11 URI, 34 other infections, 12 malignancy, 18 other
GSE40553	2012	Bloom[9]	GPL10558	Validation	South Africa, UK	Adults	Negative	Positive	..	..	..	36	130	166	Treatment measured at 0·5, 2, 4, 6, and 12 months. Two cohorts followed. Latent tuberculosis not used; overlaps with GSE19491
GSE41055	2013	Verhagen[10]	GPL5175	Validation	Venezuela	Child-ren	Negative	Positive and negative	9	9	..	7 negative; 2 positive	..	27	..
GSE42834	2014	Bloom[9]	GPL10558	Discovery	UK, France	Adults	Negative	Positive	118	..	123	40	..	281	Other diseases breakdown: 83 sarcoidosis, 24 pneumonia, 16 cancer
GSE56153	2012	Ottenhoff[23]	GPL6883	Validation	Indonesia	Adults	Negative	Positive	18	..	..	18	35	71	Treatment measured at 8 and 28 weeks
GSE62147	2015	Tientcheu[29]	GPL6480	Validation	The Gambia	Adults	Negative	Positive	..	..	..	26	26	52	*M africanum* and *M tuberculosis*
GSE74092	2015	Maertzdor[12]	RT-PCR array GPL21040	Validation	India	Adults	Negative	Positive	76	..	..	113	..	189	*KLF2* not present in these data

ASLE=adult systemic lupus erythematosus. PSLE=paediatric systemic lupus erythematosus. CLD=chronic lung disease. URI=upper respiratory infection.

Table: **Summary table of all datasets that matched inclusion criteria (whole blood, clinically active pulmonary tuberculosis)**

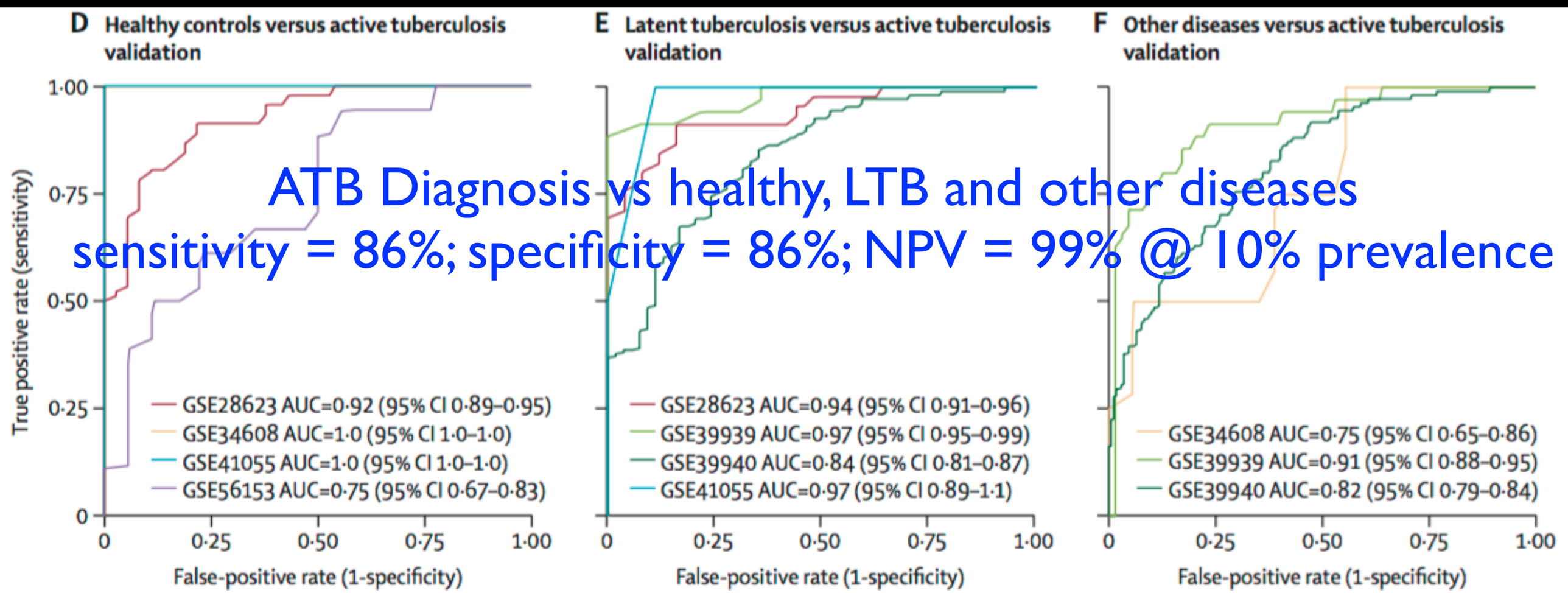
11 countries

14 cohorts

2,572 samples

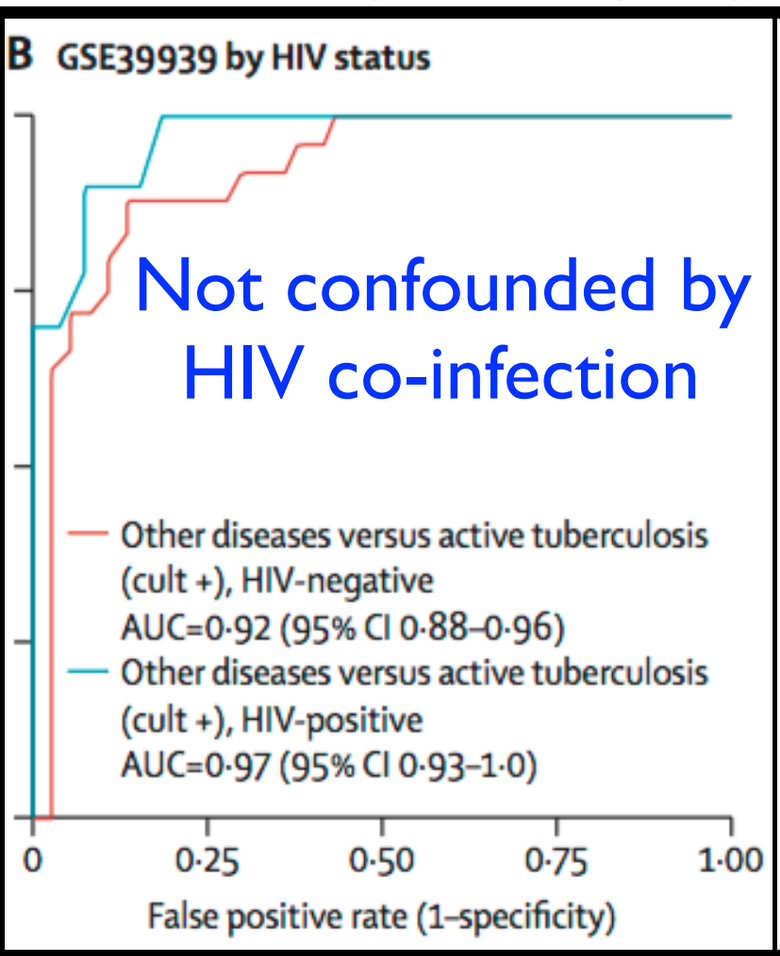
3 genes

(DUSP3, GBP5, KLF2)

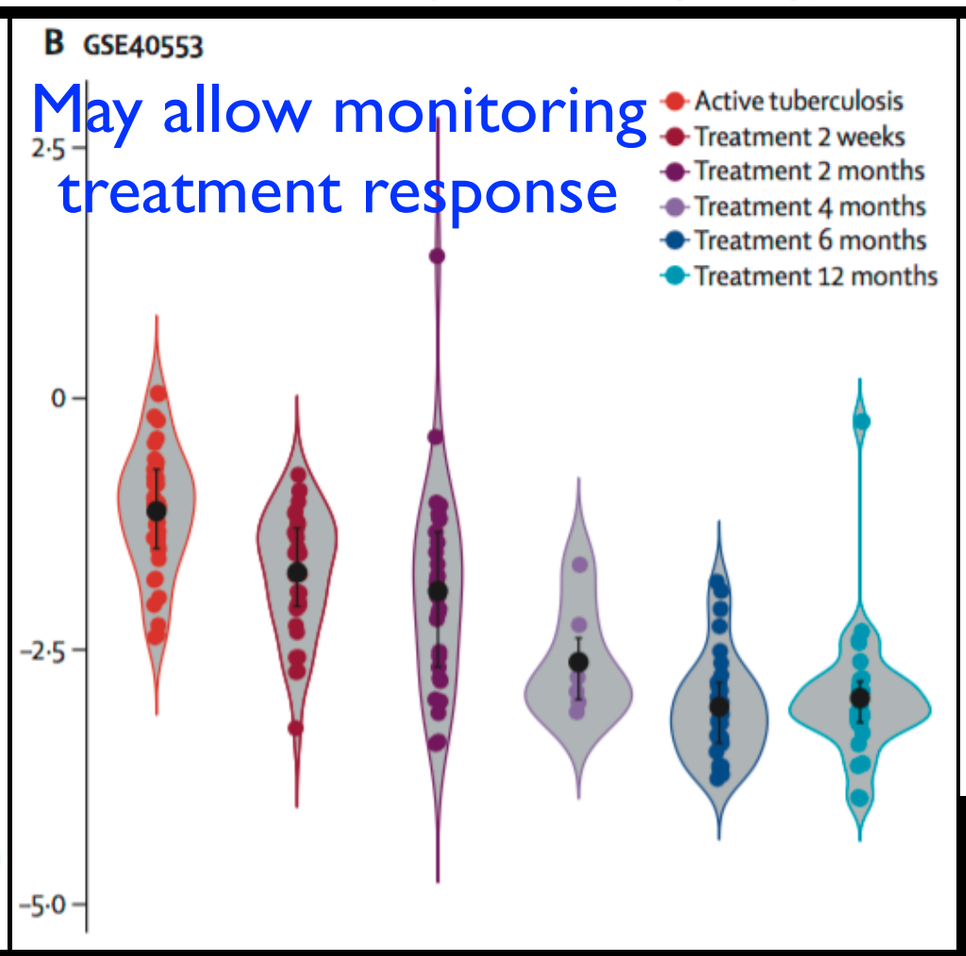


ATB Diagnosis vs healthy, LTB and other diseases

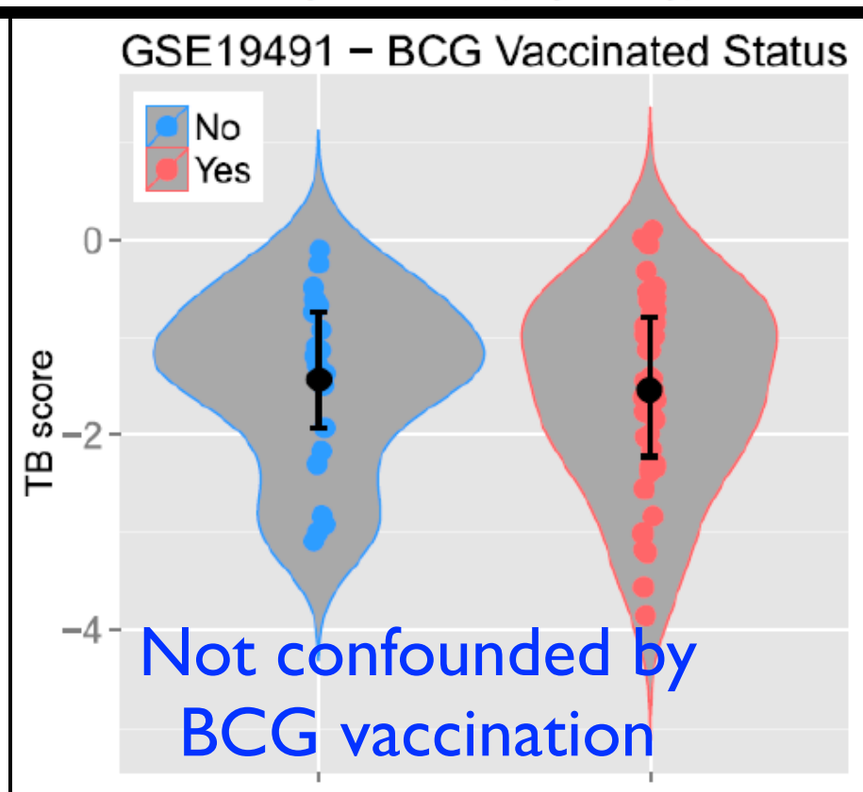
sensitivity = 86%; specificity = 86%; NPV = 99% @ 10% prevalence



Not confounded by HIV co-infection



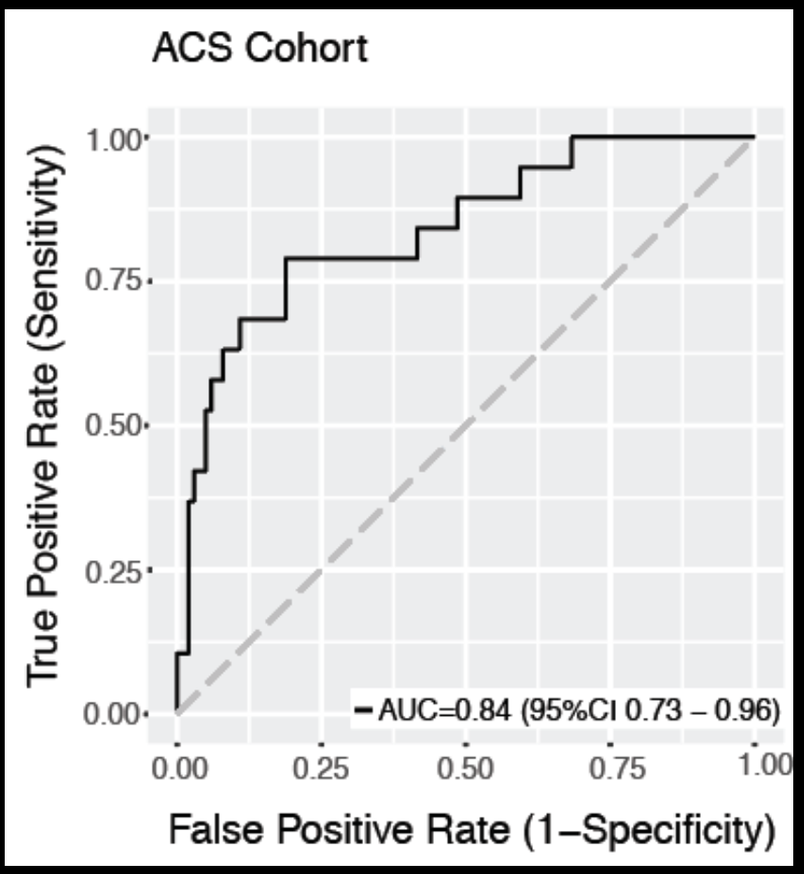
May allow monitoring treatment response



Not confounded by BCG vaccination

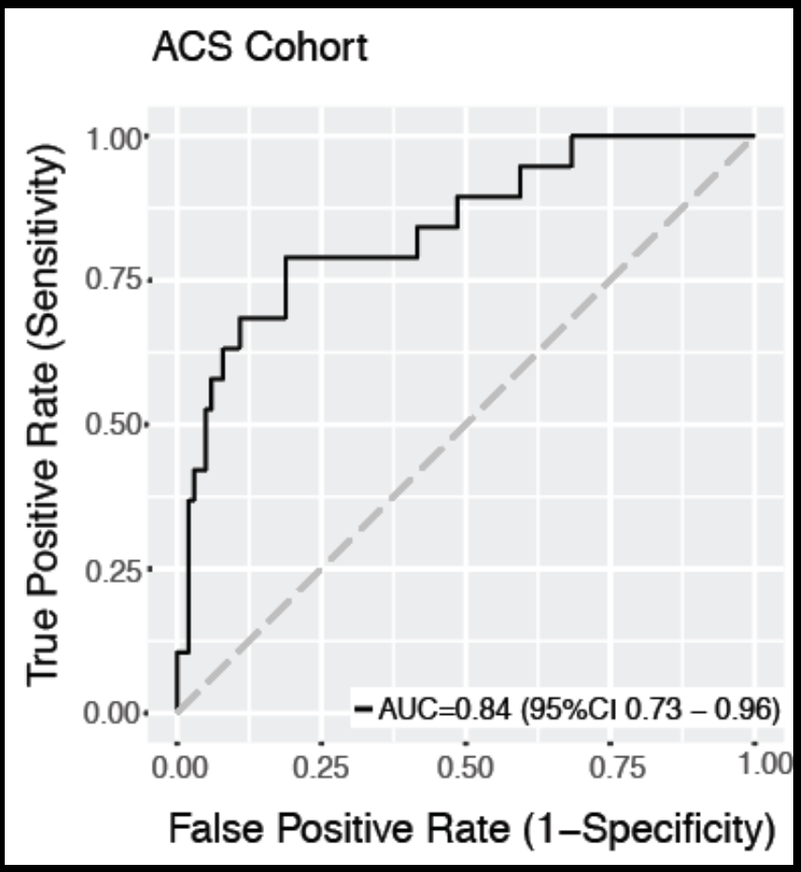
Sweeney *et al. Lancet Resp Med* 2016

3-gene signature distinguishes ATB in prospective cohorts

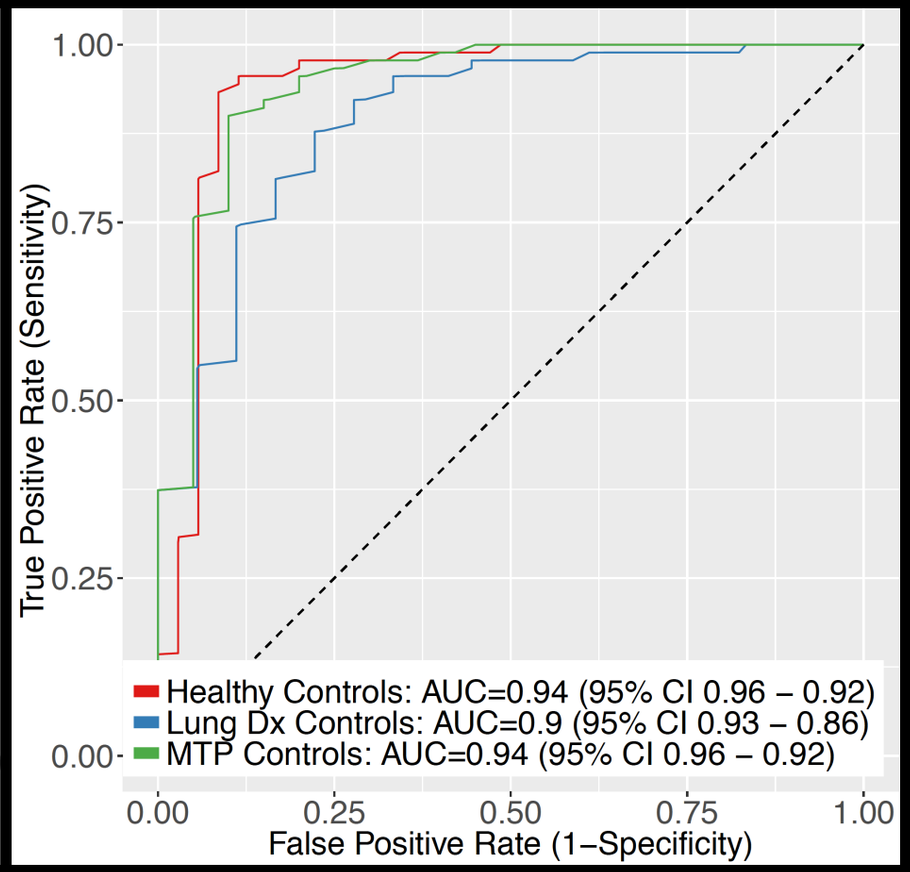


Zak *et al. Lancet* 2016
Adolescents
LTB vs ATB
RNAseq

3-gene signature distinguishes ATB in prospective cohorts

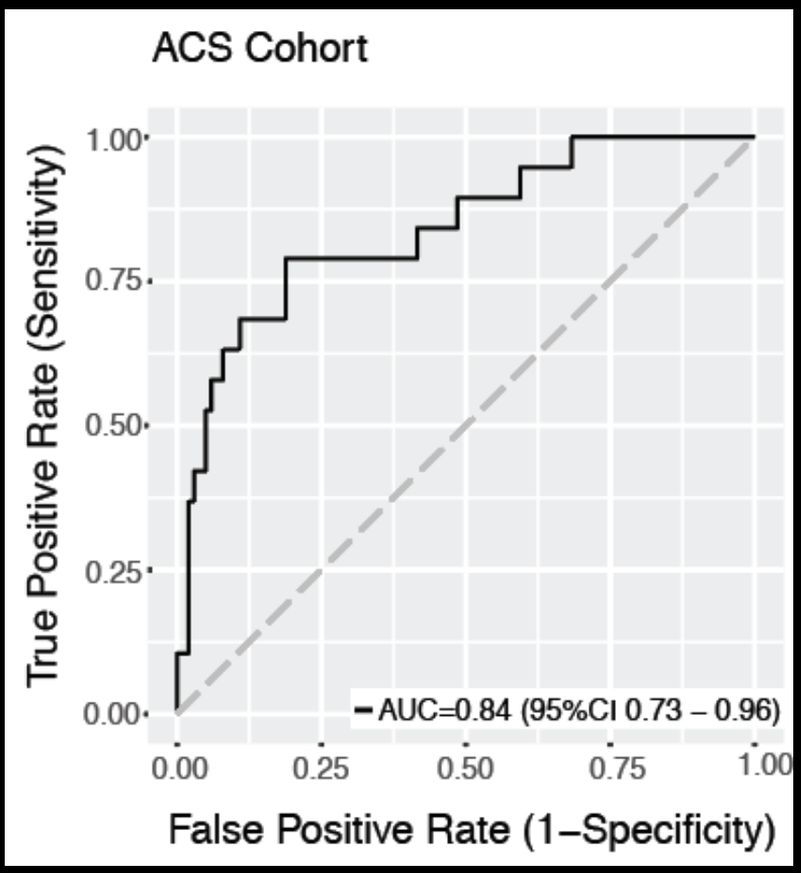


Zak *et al. Lancet* 2016
Adolescents
LTB vs ATB
RNAseq

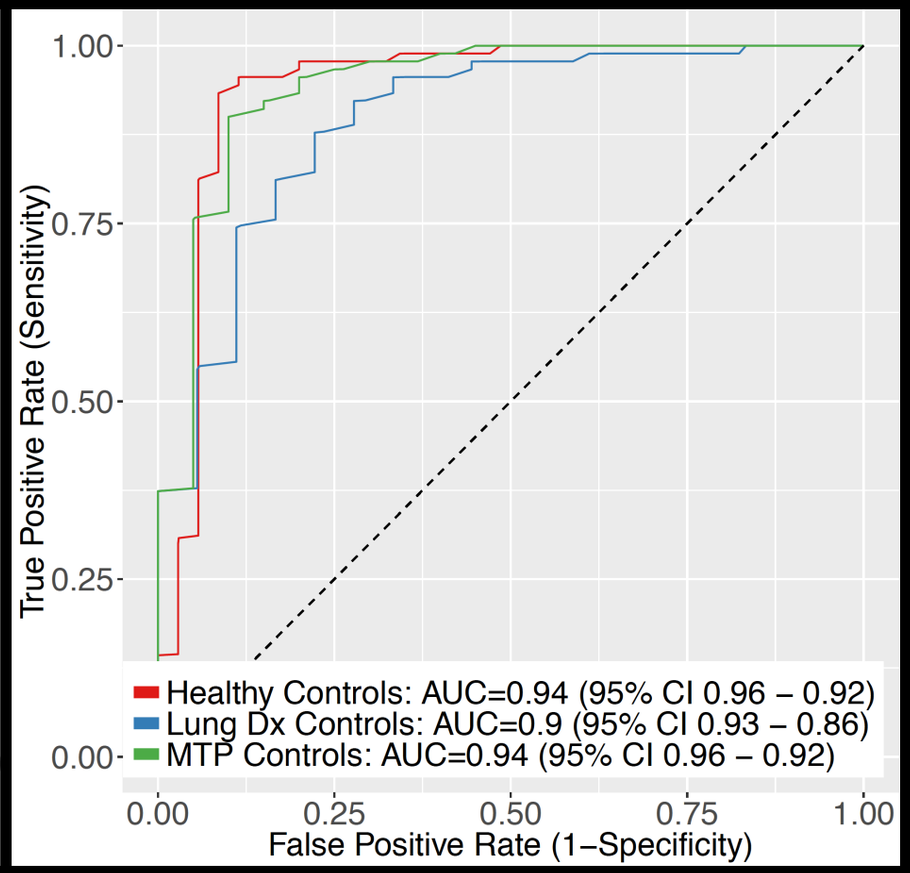


Zak *et al. Tuberculosis* 2017
Adults
ATB vs controls
RNAseq

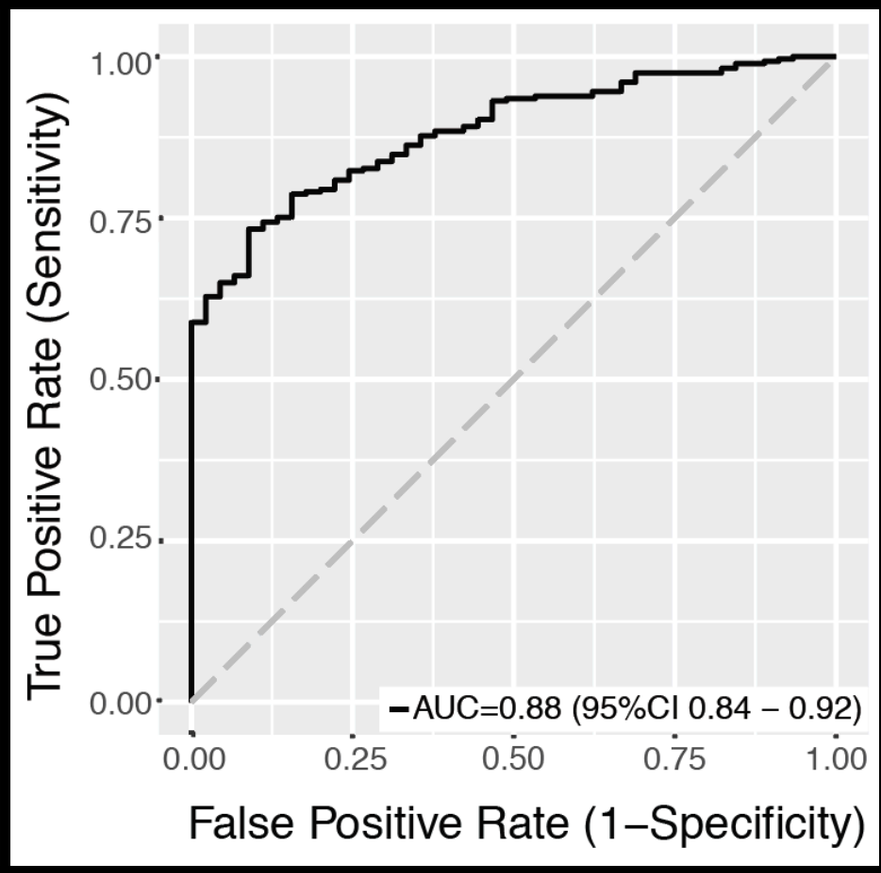
3-gene signature distinguishes ATB in prospective cohorts



Zak *et al. Lancet* 2016
Adolescents
LTB vs ATB
RNAseq

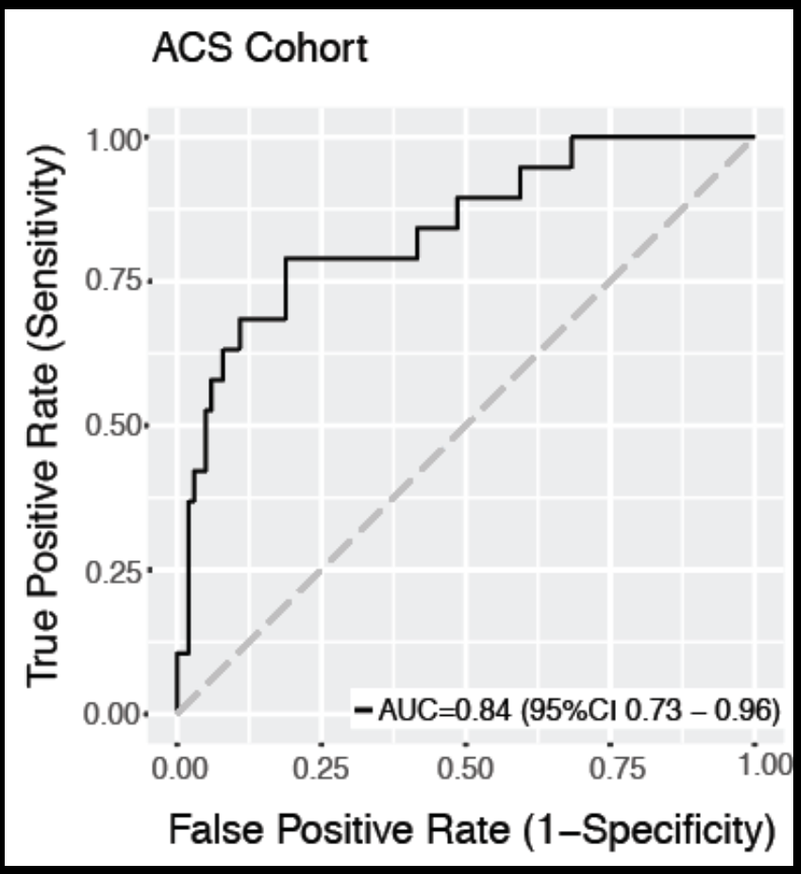


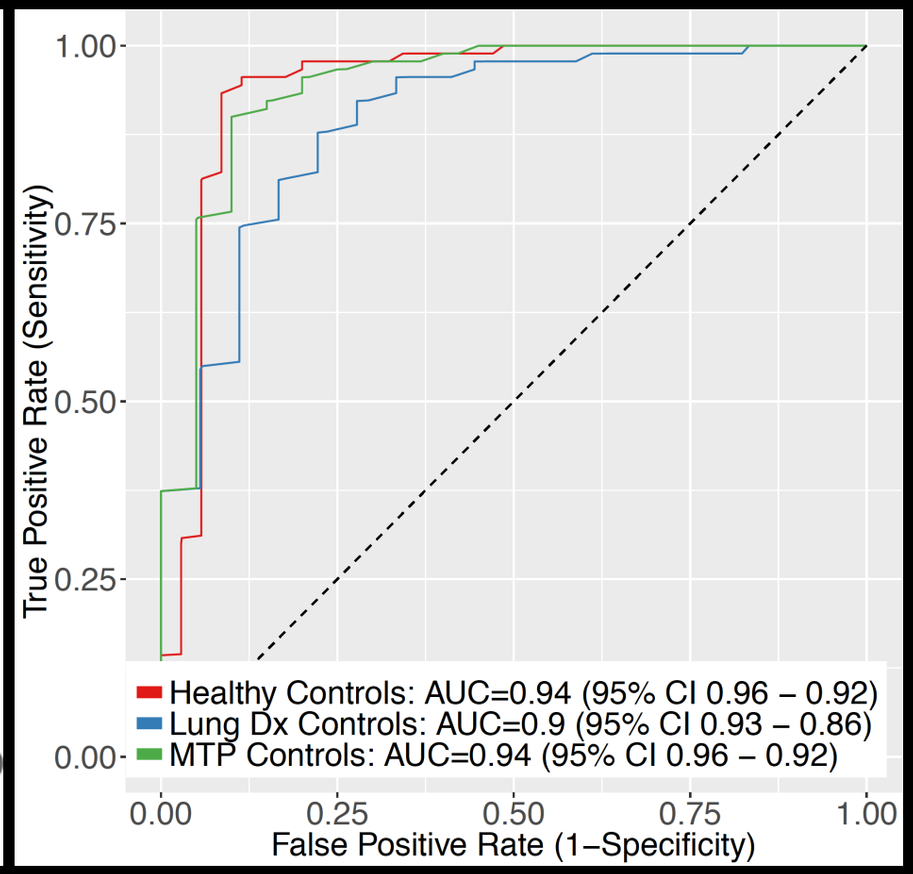
Zak *et al. Tuberculosis* 2017
Adults
ATB vs controls
RNAseq



Warsinske *et al.*
Active screen in adults
ATB vs controls
PCR

3-gene signature distinguishes ATB in prospective cohorts





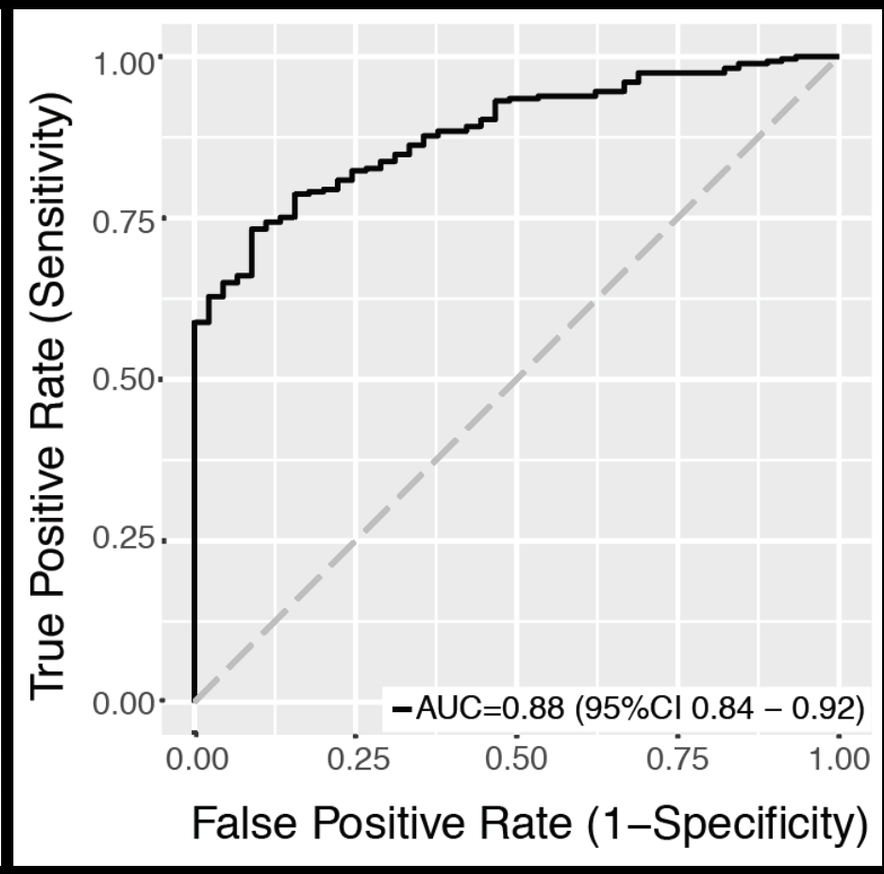
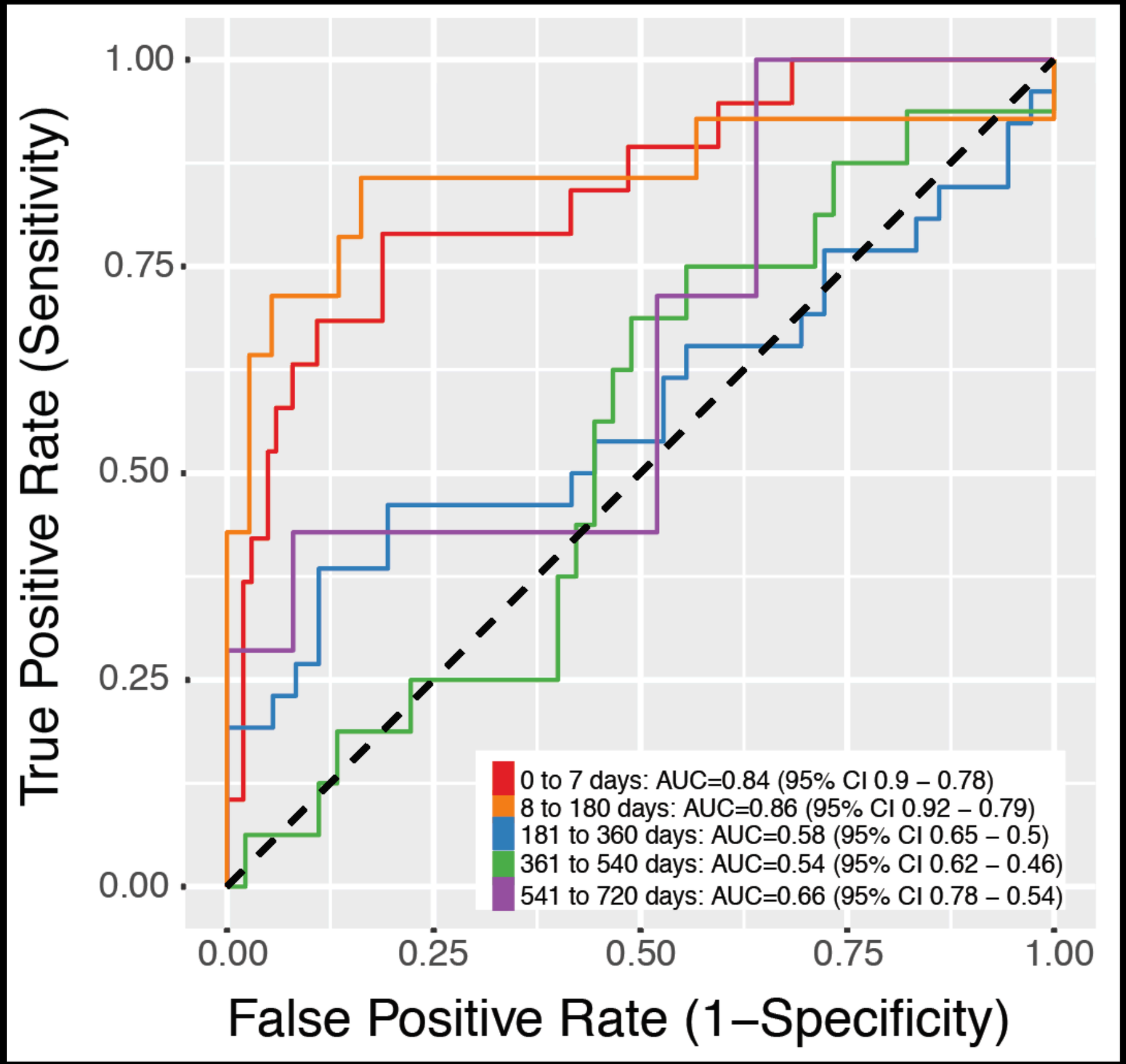


Table 3 Maximized sensitivity values obtained from the ROC analysis of *GBP5*, *DUSP3* and *KLF2* combinations in WB cohort test. Francisco *et al. J of Infection* 2017

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

OD vs ATB; PCR

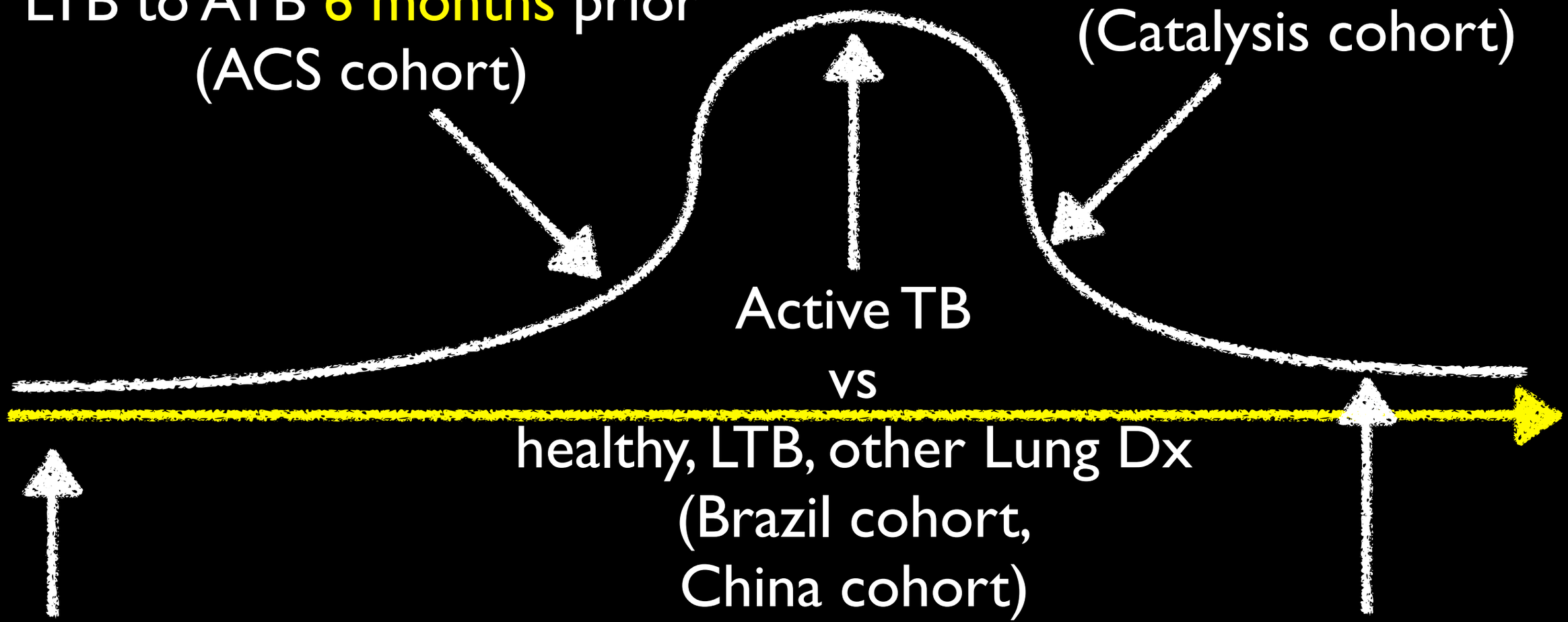
3-gene signature predicts progression from LTB to ATB



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)



Healthy or Latent TB

May also identify sub-clinically active TB

Acknowledgements

Jason Andrews Julio Croda Tim Sweeney

SPADA

NIH 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

CEDAR
CENTER FOR EXPANDED DATA ANNOTATION AND RETRIEVAL

BILL & MELINDA GATES foundation

novo nordisk®