Identification of Antibody Targets for Tuberculosis Serology using High-Density Nucleic Acid Programmable Protein Arrays

Lusheng Song[1], Garrick Wallstrom[1], Xiaobo Yu[2], Marika Hopper[1], Jennifer Van Duine[1], Jason Steel[1], Jin Park[1], Peter Wiktor[1,3], Peter Kahn[3], Al Brunner[3], Douglas Wilson[4], Elizabeth R. Jenny-Avital[5], Ji Qiu[1], Joshua Labaer[1], D. Mitchell Magee[1*] and Jacqueline M. Achkar[5,6*]

[1]The Virginia G Piper Center for Personalized Diagnostics, The Biodesign Institute, Arizona State University, Tempe, AZ, 85287, United States

[2] State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences-Beijing (PHOENIX Center), Beijing Institute of Radiation Medicine, Beijing, 102206, China

[3]Engineering Arts LLC, Tempe, AZ 85287, United States

[4]Department of Internal Medicine, Edendale Hospital, University of KwaZulu-Natal, Pietermaritzburg, South Africa

Departments of Medicine[5] and Microbiology and Immunology[6], Albert Einstein College of Medicine, Bronx, New York, 10461, United States

*Co-Corresponding Authors

D. Mitchell Magee, Ph.D.

Center for Personalized Diagnostics

The Biodesign Institute at Arizona State University

1001 S. McAllister BLVD

Tempe, AZ 85287

Telephone: 480-727-0857

FAX: 480-965-3051

Email: mitch.magee@asu.edu

Jacqueline M. Achkar, MD, MS, FIDSA

Division of Infectious Diseases, Departments of Medicine and Microbiology and Immunology

Block Bld., Rm. 115

Albert Einstein College of Medicine

1300 Morris Park Avenue

Bronx, NY 10461

Telephone: 718-430-8763

Fax: 718-430-8767

Email: jacqueline.achkar@einstein.yu.edu

Running Title: Discovery of Antibody Targets for Tuberculosis Serology

Supplementary Figure 1. Demonstration of the M-HD-NAPPA for *Mtb* biomarker discovery.

(A). Array image showing protein display as assessed by anti-GST binding of both single protein and mixture of 3 proteins/spot. (B). Dotplot of anti-GST expression intensity for both single protein and mixture of 3 proteins/spot. Greater than 96% of the proteins express at the cutoff level of 1 X 10^6 fluorescent intensity units. Blue bars represent median reactivity. (C). Ball and stick heat map of single and multiplex protein responses to pool serum. This ball-map figure is based on the normalized response value of a multiplex protein spots for the central circle. Each of the component three proteins orbits the central circle and their color also represents the normalized response as individual protein spot. The darker the red color indicates higher responses, with signal to background ratio (S/B) values ranging from 2.0 to >5.0. One of our key criteria is a minimal S/B value of 2.0 to establish a level of minimal reactivity and 16 individual proteins met this cutoff value. We determined that 4 proteins represented low-level reactivity (S/B 2.0-3.0), 5 proteins revealed mid-level reactivity (S/B 3.0-5.0), and 7 individual proteins revealed high-level responses (S/B > 5.0).

Supplementary Figure 2. Quality control evaluation of *Mtb* M-HD-NAPPA.

(A). Array image of protein display as assessed by anti-GST binding to *Mtb* 3 protein/spot M-HD-NAPPA array. (B). Dot plot of the protein display of full *Mtb* M-HD-NAPPA array across subarray and print batch. Blue bars represent median reactivity. (C). Pearson correlation of the anti-GST expression intensity intra- and inter-slides. (D). Pearson correlation of Pool serum binding to two slides of the M-HD-NAPPA array.

Supplementary Figure 3. Quality control evaluation of *Mtb* HD-NAPPA used in deconvolution and validation studies.

(A). Array image of protein display as assessed by anti-GST binding to single protein/spot *Mtb* HD-NAPPA array. Spots that exhibit overall fluorescence intensity higher than 1x10^6 a.u.

exceed the cutoff of the successful protein expression (B). Dot plot of the protein display of full *Mtb* HD-NAPPA array across subarray and print batch. The success protein display rate on the multiplex HD-NAPPA were as high as 98.50% ± 0.34% across 4 subarrays across two slides. Blue bars represent median reactivity. (C). The intra-slide and inter-slide Pearson correlation reveal great reproducibility with intra-slide subarray correlation (R=0.963) and the inter-slide correlation (R=0.936) (D). Reproducibility testing using binding of our positive control TB+ pool revealed excellent inter-slide Pearson correlations (0.970).

Supplementary Figure 4. *Mtb* protein RAPID ELISA quality controls.

(A) Protein expression of the 8 selected antigens and negative control protein (Rv1553). The target proteins were captured onto ELISA plates and the samples probed with monoclonal anti-GST and detected via chemiluminescence to assess protein capture. All 8 targets and 1 negative control proteins were detected between 2.0 x10^6 a.u. and 1.2 x10^7 a.u., which is more than 100 fold greater than our cutoff of 2.0 x 10^4 a.u. (blank intensity plus two time standard deviation) and indicates good expression of the target protein. Blue bars represent the standard deviation

(B) Serum response to negative control protein (Rv1553) across runs with an average value of median at 4.10 ± 1.39 x10^4 a.u with significant correlations (r>0.8;). A cutoff was then set as 4.0 x10^4 a.u post subtracting normalization of the data, which was the median response of Rv1553 across all samples. Blue bars represent the median response.

Supplementary Figure 5. ROC analysis according to subgroups for candidate *Mtb* proteins tested via RAPID ELISA. Single hit ROC analysis (AUC>0.7, p<0.05) across four subgroups:

US/HIV-: Rv0831c (A.) and Rv2031c (B.). US/HIV+: Rv0054 (C.) and Rv3405c (D.).

SA/HIV-: Rv2031c (E.), Rv0054 (F.), Rv0831c (G.) and Rv3405c (H.). SA/HIV+: Rv3405c (I.).

Supplementary Figure 6. Individual *Mtb* protein expression quality control on NAPPA.

(A) Array images of Individual *Mtb* NAPPA arrays (4,045 *Mtb* proteins) printed onto two glass slides and the expression with IVTT were assessed by anti-GST binding to individual protein (on the right) on the NAPPA array. (B) The individual *Mtb* protein expression quantification of two of both TB array 01 and TB array 02. The TB array 01 has a display rate of 91.3% to 93.7% while the TB array 02 has a display rate of 99.8% at the cutoff of $5*10^6$ a.u. (Master mix spots (MM) plus 3 standard deviation).

Supplementary Data

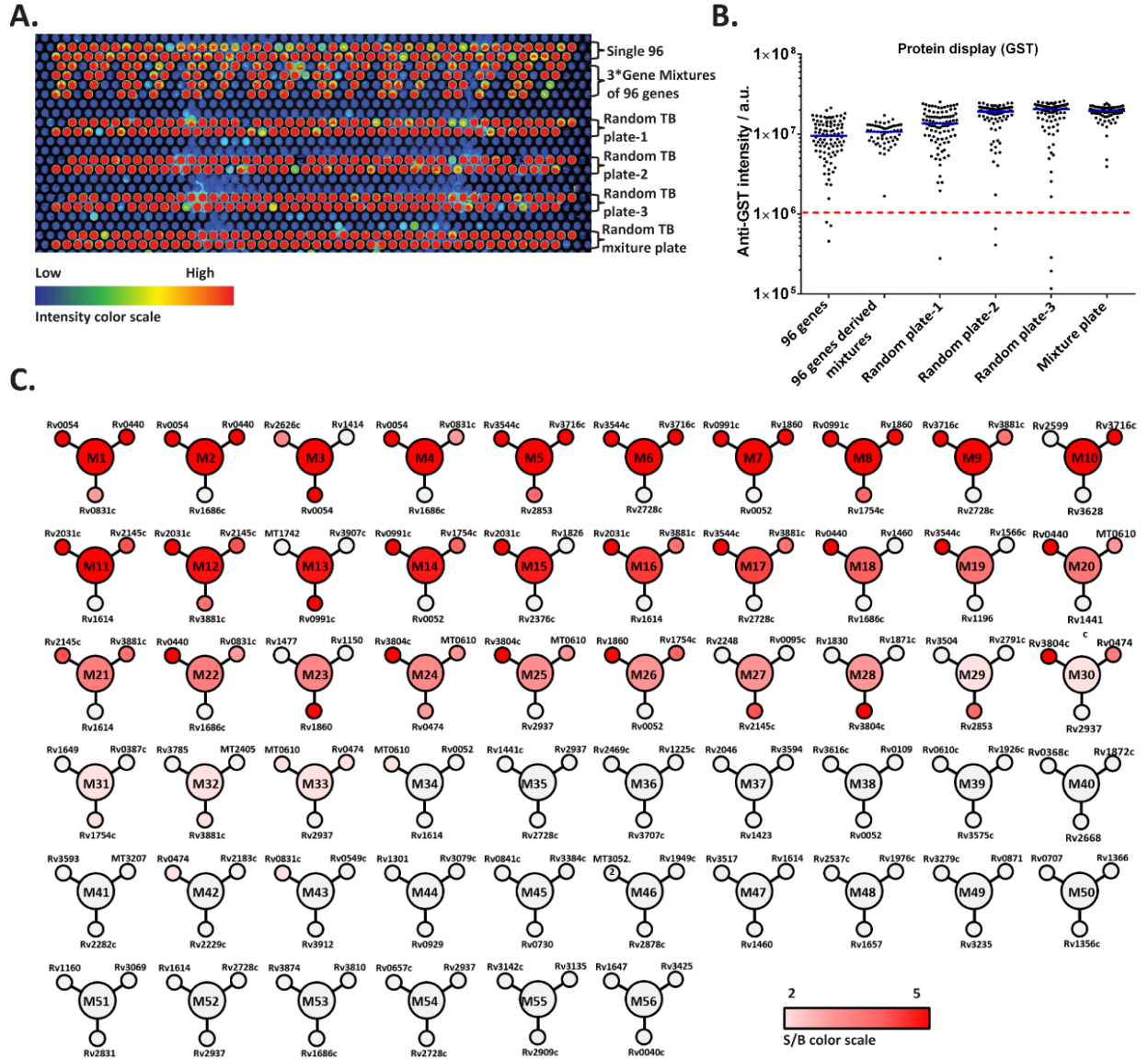Supplementary Table 1. Top IgG and IgA hits from HD-NAPPA validation analysis.
Top IgG and IgA response antigens across 4 subgroups which showed more than 20%
sensitivity at cutoff 1.4, odd's value>1.5 and AUC value>0.55. Top IgG and IgA hits for US/HIV-
(A), US/HIV+ (B), SA/HIV- (C) and SA/HIV+ (D).

A. Top IgG hits for US/HIV-						
Protein	Cutoff-1.4		Odd's ratio		AUC	
	Sensitivity	Specificity	Value	P value	Value	P value
Rv0054	34.8%	90.9%	3.83	0.18	0.81	0.00
Rv3544c	34.8%	90.9%	3.83	0.18	0.68	0.05
Rv1731	34.8%	81.8%	1.91	0.29	0.60	0.16
Rv3260c	30.4%	100.0%	6.09	0.25	0.64	0.10
Rv0178	30.4%	100.0%	6.09	0.17	0.60	0.16
Rv2770c	30.4%	81.8%	1.67	0.53	0.58	0.23
Rv2306c	26.1%	100.0%	5.22	0.14	0.82	0.00
Rv2031c	26.1%	100.0%	5.22	0.23	0.73	0.02
Rv2668	26.1%	100.0%	5.22	0.30	0.72	0.02
Rv3897c	26.1%	90.9%	2.87	0.25	0.66	0.06
Rv0983	26.1%	100.0%	5.22	0.30	0.65	0.08
Rv0474	26.1%	100.0%	5.22	0.23	0.62	0.14
Rv3822	26.1%	100.0%	5.22	0.25	0.60	0.16
MT3033	26.1%	100.0%	5.22	0.23	0.59	0.19
Rv2831	21.7%	100.0%	4.35	0.23	0.83	0.00
Rv3912	21.7%	100.0%	4.35	0.39	0.70	0.03
Rv1860	21.7%	100.0%	4.35	0.40	0.65	0.08
Rv1904	21.7%	100.0%	4.35	0.23	0.64	0.10
Rv0040c	21.7%	100.0%	4.35	0.40	0.64	0.10
Rv2922c	21.7%	90.9%	2.39	0.30	0.63	0.11
Top IgA hits for US/HIV-						
Rv0052	47.8%	81.8%	2.63	0.25	0.72	0.02
Rv2922c	21.7%	100.0%	4.35	0.17	0.64	0.09
Rv0509	21.7%	90.9%	2.39	0.23	0.55	0.31

B. Top IgG hits for US/HIV+					
Protein	Cutoff-1.4		Odd's ratio		AUC	
	Sensitivity	Specificity	Value	P value	Value	P value
Rv2853	88.9%	63.6%	2.44	0.10	0.66	0.12
MT3033	55.6%	90.9%	6.11	0.12	0.79	0.02
Rv3810	44.4%	100.0%	8.89	0.08	0.77	0.02
Rv0109	44.4%	100.0%	8.89	0.25	0.75	0.03
Rv0040c	44.4%	72.7%	1.63	0.47	0.68	0.09
Rv3260c	44.4%	90.9%	4.89	0.47	0.62	0.19
Rv1748	33.3%	100.0%	6.67	0.13	0.83	0.01
Rv3835	33.3%	90.9%	3.67	0.25	0.67	0.10
Rv0247c	33.3%	100.0%	6.67	0.13	0.63	0.17
Top IgA hits for US/HIV+						
Rv1411c	44.4%	81.8%	2.44	0.25	0.70	0.07
Rv1754c	33.3%	100.0%	6.67	0.48	0.60	0.24
Rv0983	33.3%	100.0%	6.67	0.23	0.56	0.34

C. Top IgG hits for SA/HIV-					
Protein	Cutoff-1.4		Odd's ratio		AUC	
	Sensitivity	Specificity	Value	P value	Value	P value
Rv2668	66.7%	75.0%	2.67	0.34	0.68	0.11
Rv0638	50.0%	83.3%	3.00	0.19	0.85	0.01
Rv0831c	50.0%	91.7%	6.00	0.09	0.81	0.02
Rv3405c	50.0%	66.7%	1.50	0.24	0.64	0.17
Rv1651c	50.0%	91.7%	6.00	0.09	0.60	0.26
Top IgA hits for SA/HIV-						
Rv1566c	50.0%	83.3%	3.00	0.45	0.74	0.06

D. Top IgG hits for SA/HIV+					
Protein	Cutoff-1.4		Odd's ratio		AUC	
	Sensitivity	Specificity	Value	P value	Value	P value
Rv3405c	62.5%	58.3%	1.50	0.49	0.56	0.25
Rv3822	27.5%	83.3%	1.65	0.52	0.58	0.20
Rv2770c	27.5%	91.7%	3.30	0.20	0.57	0.23
Rv0652	20.0%	100.0%	4.00	0.20	0.68	0.03
Top IgA hits for SA/HIV+						

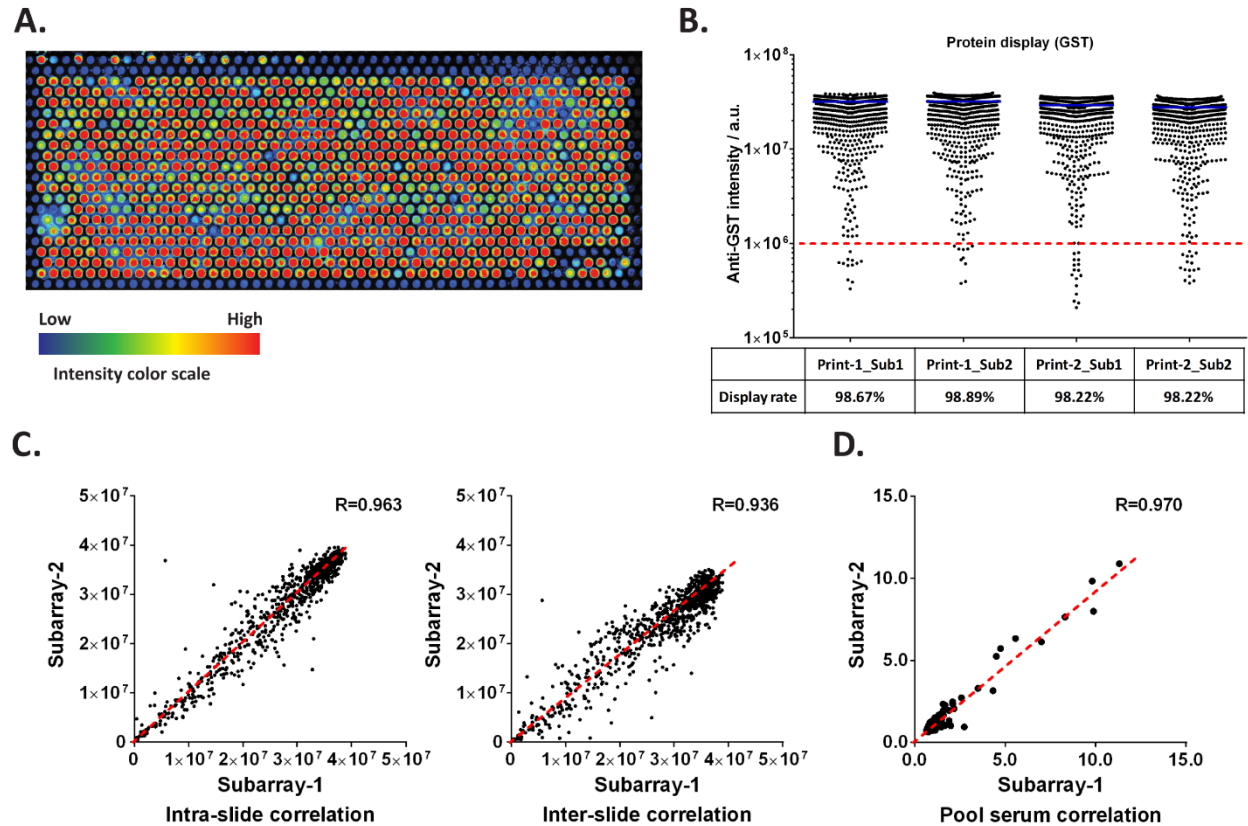Supplementary Figure 1. Demonstration of the M-HD-NAPPA for *Mtb* biomarker discovery.

(A). Array image showing protein display as assessed by anti-GST binding of both single protein and mixture of 3 proteins/spot. (B). Dotplot of anti-GST expression intensity for both single protein and mixture of 3 proteins/spot. Greater than 96% of the proteins express at the cutoff level of 1 X 10^6 fluorescent intensity units. Blue bars represent median reactivity. (C). Ball and stick heat map of single and multiplex protein responses to pool serum. This ball-map figure is based on the normalized response value of a multiplex protein spots for the central circle. Each of the component three proteins orbits the central circle and their color also represents the

normalized response as individual protein spot. The darker the red color indicates higher responses, with signal to background ratio (S/B) values ranging from 2.0 to >5.0. One of our key criteria is a minimal S/B value of 2.0 to establish a level of minimal reactivity and 16 individual proteins met this cutoff value. We determined that 4 proteins represented low-level reactivity (S/B 2.0-3.0), 5 proteins revealed mid-level reactivity (S/B 3.0-5.0), and 7 individual proteins revealed high-level responses (S/B > 5.0).
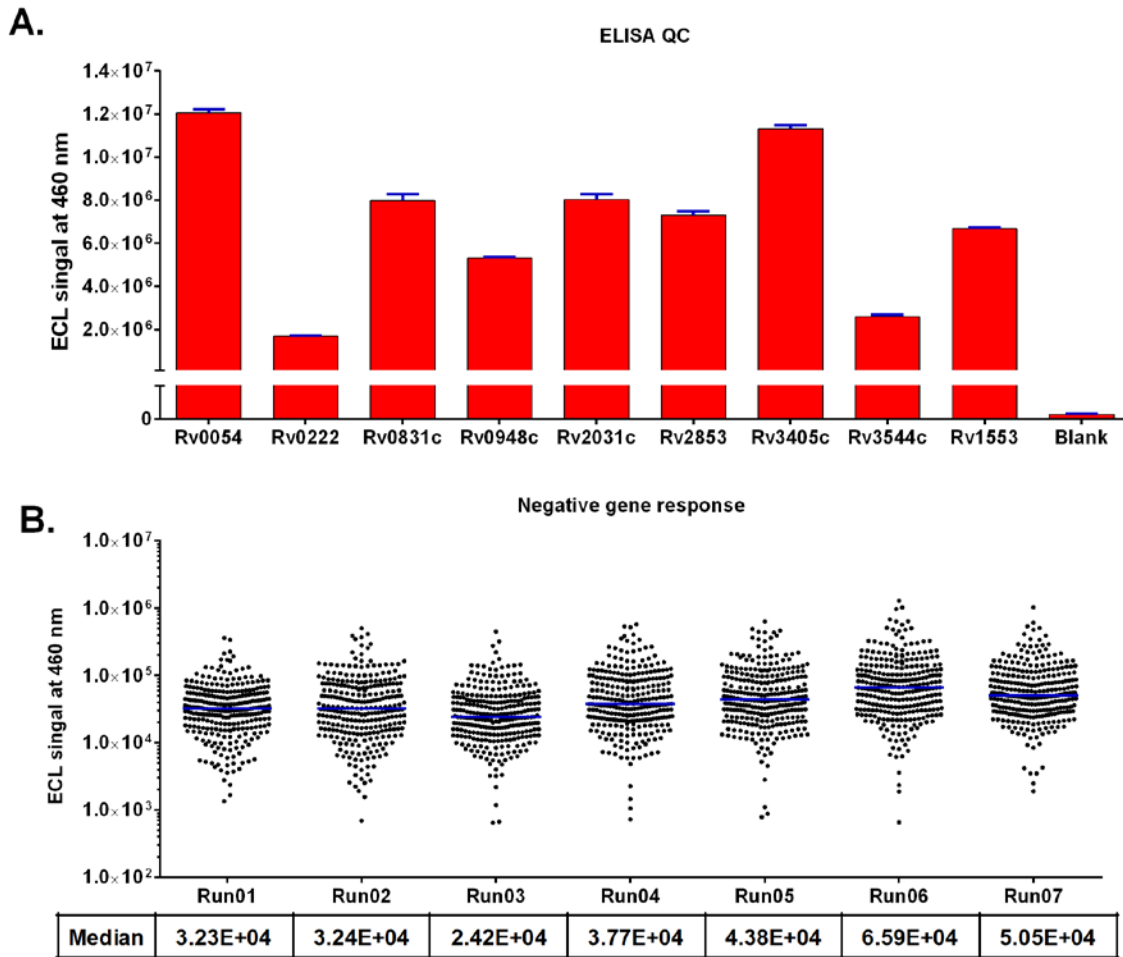
A.

3*genes mixtures

96 *Mtb* genes
virus genes

Low　　High

Intensity color scale

B.

Protein display (GST)

Print	Print-1				Print-2			
Subarray	Sub-1	Sub-2	Sub-3	Sub-4	Sub-1	Sub-2	Sub-3	Sub-4
Display rate	99.42%	99.22%	99.29%	99.48%	99.61%	99.16%	99.29%	99.16%

C.

R=0.903

Intra-slide correlation

R=0.829

Inter-slide correlation

D.

R=0.977

Pool serum correlation

Supplementary Figure 2. Quality control evaluation of *Mtb* M-HD-NAPPA.

(A). Array image of protein display as assessed by anti-GST binding to *Mtb* 3 protein/spot M-HD-NAPPA array. (B). Dot plot of the protein display of full *Mtb* M-HD-NAPPA array across subarray and print batch. Blue bars represent median reactivity. (C). Pearson correlation of the anti-GST expression intensity intra- and inter-slides. (D). Pearson correlation of Pool serum binding to two slides of the M-HD-NAPPA array.

Supplementary Figure 3. Quality control evaluation of *Mtb* HD-NAPPA used in deconvolution and validation studies.

(A). Array image of protein display as assessed by anti-GST binding to single protein/spot *Mtb* HD-NAPPA array. Spots that exhibit overall fluorescence intensity higher than 1×10^6 a.u. exceed the cutoff of the successful protein expression (B). Dot plot of the protein display of full *Mtb* HD-NAPPA array across subarray and print batch. The success protein display rate on the multiplex HD-NAPPA were as high as 98.50% ± 0.34% across 4 subarrays across two slides. Blue bars represent median reactivity. (C). The intra-slide and inter-slide Pearson correlation reveal great reproducibility with intra-slide subarray correlation (R=0.963) and the inter-slide correlation (R=0.936) (D). Reproducibility testing using binding of our positive control TB+ pool revealed excellent inter-slide Pearson correlations (0.970).
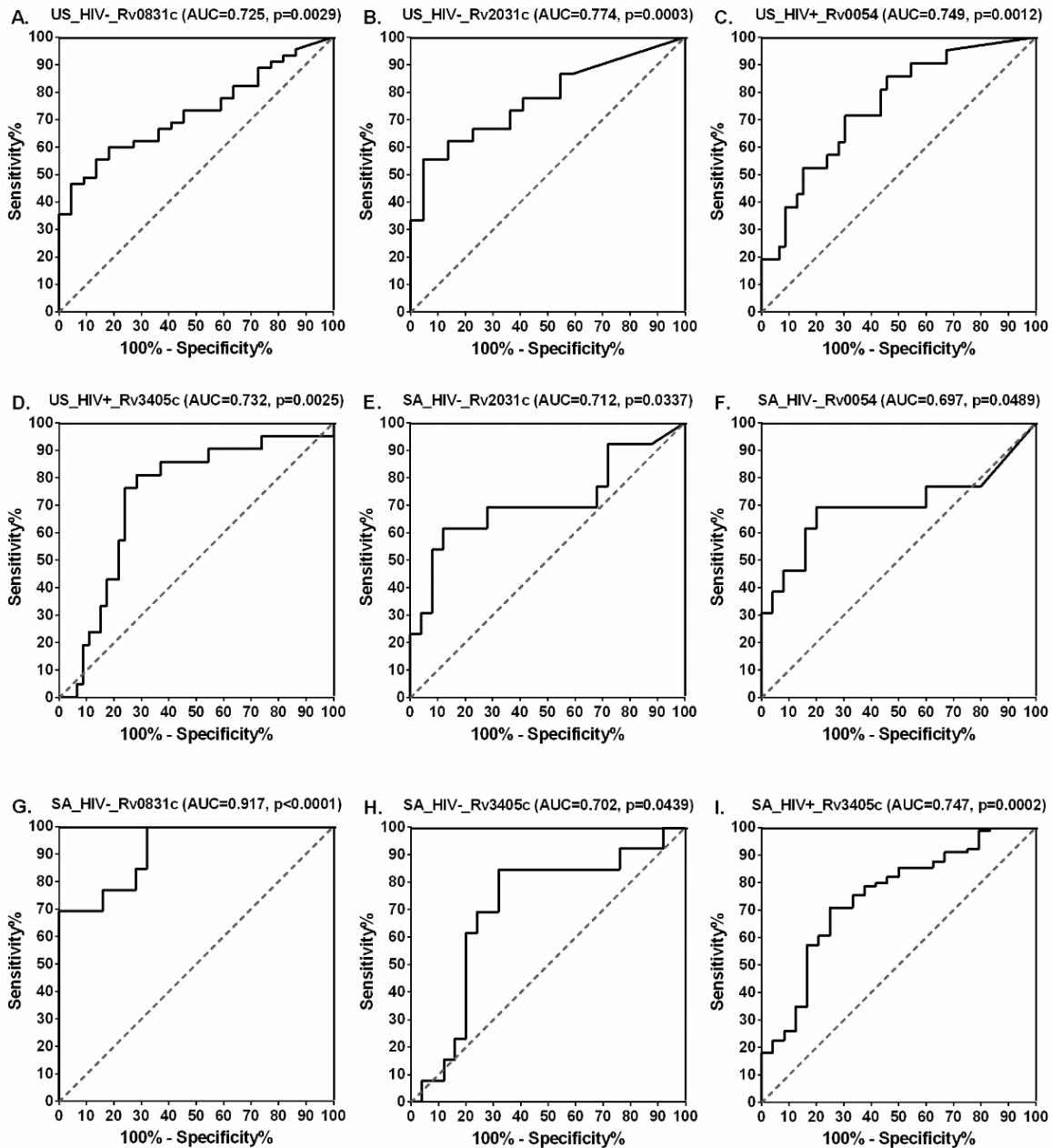
Supplementary Figure 4. *Mtb* protein RAPID ELISA quality controls.

(A) Protein expression of the 8 selected antigens and negative control protein (Rv1553). The target proteins were captured onto ELISA plates and the samples probed with monoclonal anti-GST and detected via chemiluminescence to assess protein capture. All 8 targets and 1 negative control proteins were detected between 2.0×10^6 a.u. and 1.2×10^7 a.u., which is more than 100 fold greater than our cutoff of 2.0×10^4 a.u. (blank intensity plus two time standard deviation) and indicates good expression of the target protein. Blue bars represent the standard deviation
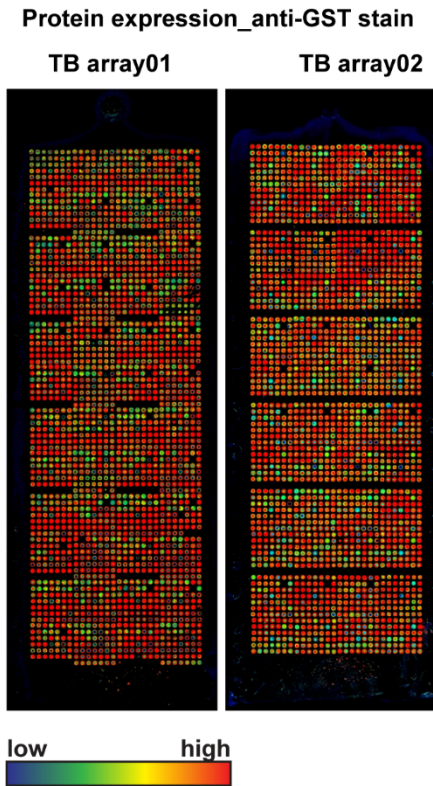
(B) Serum response to negative control protein (Rv1553) across runs with an average value of median at $4.10 \pm 1.39 \times 10^4$ a.u with significant correlations (r>0.8;). A cutoff was then set as

4.0 x10^4 a.u post subtracting normalization of the data, which was the median response of

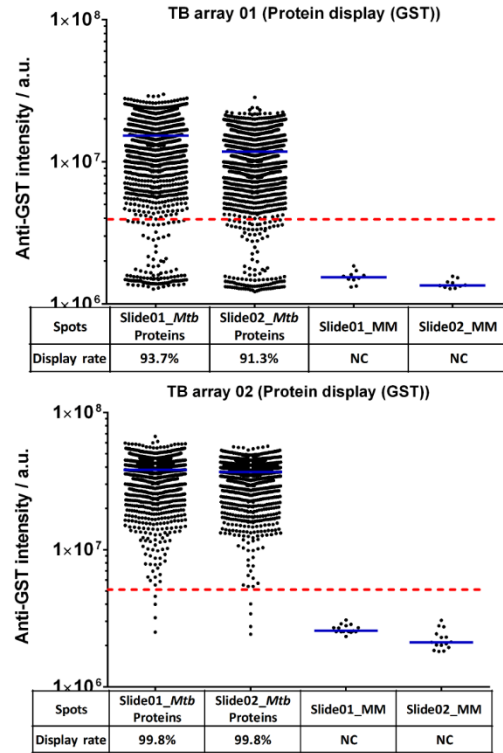Rv1553 across all samples. Blue bars represent the median response.

Supplementary Figure 5. ROC analysis according to subgroups for candidate *Mtb* proteins tested via RAPID ELISA. Single hit ROC analysis (AUC>0.7, p<0.05) across four subgroups: US/HIV-: Rv0831c (A.) and Rv2031c (B.). US/HIV+: Rv0054 (C.) and Rv3405c (D.). SA/HIV-: Rv2031c (E.), Rv0054 (F.), Rv0831c (G.) and Rv3405c (H.). SA/HIV+: Rv3405c (I.).

Supplementary Figure 6. Individual *Mtb* protein expression quality control on NAPPA.

(A) Array images of Individual *Mtb* NAPPA arrays (4,045 *Mtb* proteins) printed onto two glass slides and the expression with IVTT were assessed by anti-GST binding to individual protein (on the right) on the NAPPA array. (B) The individual *Mtb* protein expression quantification of two of both TB array 01 and TB array 02. The TB array 01 has a display rate of 91.3% to 93.7% while the TB array 02 has a display rate of 99.8% at the cutoff of $5*10^6$ a.u. (Master mix spots (MM) plus 3 standard deviation).