SUPPLEMENTAL INFORMATION

METHODS

Real-Time Quantitative TaqMan RT-PCR. Peripheral blood mononuclear cells (PBMC) were plated into 24-well round bottom plates and cultured in media with and without 1.25 µg/mL recombinant Protective Antigen (rPA). The cells were incubated at 37°C/5% CO2 for 24 h (Group 1-5) and 64 h (Group 6-12).. The cell suspension was centrifuged and the cell pellet was resuspended in TRIzol Reagent (Invitrogen, Carlsbad, CA) and stored frozen until processing according to the manufacturer's instructions with the addition of PhaseLock Gel (5 Prime 3 Prime, Inc., Boulder, CO). The RNA pellet was dissolved in RNA Storage Solution (Ambion, Austin, TX). RNA was quantitated spectrophotometrically based on an absorbance at 260 nm of one equal to an RNA concentration of 40 µg/mL. Total RNA (0.65 µg) was reverse transcribed into cDNA using SuperScript III^{TM} First-Strand Synthesis System for RT-PCR (Life Technologies, Gaithersburg, MD) according to the manufacturer's instructions.

Cytokine mRNA levels were measured by real-time quantitative RT-PCR using a PE Applied Biosystems Prism 7700 sequence detection instrument. NHP primer and probe sets for IFN-γ, IL-2, IL-4, IL-6, IL-1β, and TNF-α were designed using the Assay-by-Design service of Applied Biosystems (Applied Biosystems, Foster City, CA). Gene accession numbers are in Table S1 and NHP primer probe sequences are in Table S2. Assays were performed in duplicate and averaged. No-template controls and reverse transcriptase minus controls were negative for amplification.

Threshold cycle (C*t*), which correlates inversely with the target mRNA levels, was measured as the cycle number at which the reporter fluorescent emission increased above a threshold level. The comparative C*t* method was used to determine relative quantitation. C*t* values for cytokine amplification were normalized by subtracting the C*t* values for 18S rRNA using the equation: $Ct_{(cytokine)} - Ct_{(18S\ rRNA)} = \Delta Ct$. The cytokine stimulated ΔC*t* was subtracted from the unstimulated ΔC*t* to calculate the fold change in cytokine expression: $\Delta Ct_{(stimulated)} - \Delta Ct_{(unstimulated)} = \Delta\Delta Ct$. Fold increases in cytokine expression were calculated by the following equation according to ABI User Bulletin #2: $2^{-\Delta\Delta Ct}$ = fold change in expression.

Cytokine Secretion Analyses. Secreted cytokine levels in unstimulated and rPA stimulated PBMC were assayed in duplicate using commercially available ELISA kits (Table S3) according to the manufacturer's instructions. The threshold levels of detection were 15.6 pg/mL IFN-γ, 31.2 pg/mL IL-2, 7.8 pg/mL IL-4, 3.12 pg/mL IL-6, 3.9 pg/mL IL-1β, and 7.8 pg/mL TNF-α. Cytokine levels below the limit of detection were set to one-half the minimum detectable level for the assay. Cytokine levels above the limit of detection were repeated at a higher dilution if sufficient sample was available. If not, values were then set at the maximum limit of detection for each assay, 1000 pg/mL IFN-γ, 2000 pg/mL IL-2, 500 pg/mL IL-4, 300 pg/mL IL-6, 250 pg/mL IL-1β, and 500 pg/mL TNF-α. The stimulated to unstimulated ratio for each cytokine was calculated.

Lethal Toxin Neutralization Activity (TNA) Assay: TNA assays were done according to Li et al. (1) using human reference standard AVR801 (2). Reportable values were the reciprocal serum sample dilution effecting 50% neutralization of anthrax lethal toxin (ED50). Endpoints were calculated using SAS® version 9.0 (SAS Institute Inc. Cary, NC USA). The LOD and LLOQ were ED50 of 11 and 36 respectively (1). ED50 values <LOD were replaced with ½ the LOD for the statistical analyses.

Anti-PA IgG ELISA: Immulon® 2 HB microtiter plates (Thermo Labsystems, Franklin, MA) were coated with rPA (2 μg/mL) in phosphate buffered saline (PBS) pH 7.4 (Life Technologies, Gaithersburg, MD). Plates were washed 3x with PBS containing, 0.1% Tween 20. Test sera were added to wells pre-loaded with 100 μl of PBS containing 5% skim milk (wt/vol) and 0.5% Tween-20 (vol/vol), pH 7.4, mixed on the plate and serially transferred to make an 8-point dilution series with a 100 μl/well. After washing, bound anti-PA IgG was detected with horseradish peroxidase-conjugated goat anti-monkey IgG (Research Diagnostics, Inc, Flanders, NJ) and color developed with ABTS substrate (Kirkegaard and Perry Laboratories, Gaithersburg, MD). Data were analyzed using a four-parameter logistic-log curve-fitting model with ELISA for Windows software (Version 2.15). Reportable values of anti-PA IgG for rhesus macaques were in μg/mL using a calibration factor of 171.9 μg/mL for reference serum

AVR731. The lower limits of detection (LOD) and quantification (LLOQ) were 0.4 and 2.3 g/mL anti-PA IgG respectively (3). Concentration values <LOD were replaced with ½ the LOD for the statistical analyses.

Anti-PA IgG Avidity: Serum samples with ≥5 μg/mL total anti-PA IgG were evaluated for avidity, an indirect assessment of polyclonal antibody affinity, immune response maturation and a surrogate for memory B cell persistence (4). The avidity indices (AI) were determined by anti-PA IgG elution from immobilized rPA with ammonium thiocyanate (NH4SCN; 0.078 - 5M) (Sigma). A 4-PL dissociation curve was generated for percent maximum detected signal versus NH4SCN concentration and the avidity index (AI) reported as the concentration of NH4SCN required to elute 50% of bound anti-PA IgG.

Detection of IFN-γ and IL-4 Secreting Cells: PBMC were prepared as described previously (4). IFN-γ and IL-4 producing cells were enumerated by ELISpot assay following *in vitro* re-stimulation with 1 μg/mL rPA (24 h for IFN-γ assays and 36 h for IL-4 assays). Staphylococcal enterotoxin B at 2 μg/well (Toxin Technology, Sarasota, FL) was used as a positive control. Un-stimulated cultures served as negative controls. The frequency of IFN-γ+ or IL-4+ T cells specific for rPA was calculated by subtracting the average number of spot forming units (SFU) in unstimulated negative control triplicate wells from the average number of SFU in rPA stimulated triplicate wells and expressed as rPA-specific IFN-γ or IL-4 SFU/10^6 PBMC.

Lymphocyte Stimulation Indices: PBMC were plated in quadruplicate into 96-well round bottom microtiter plates containing 200 μl of either media alone or media containing 1.25 μg/mL rPA. The positive control was phytohemagglutinin (10 μg/mL). Cells were incubated for 96 h at 37°C, 5% CO_2. Cultures were then pulsed with 20 μL of a 50 μCi/mL ^{3}[H]-thymidine solution and incubated for 18 h at 37°C, 5% CO_2. Cells were harvested onto filter discs (Fisher, Pittsburgh, PA) and counted on a Packard scintillation counter (Packard, Meriden, CT). Stimulation indices (SI) were calculated as the quotient of [mean counts per minute of stimulated cells ÷ mean counts per minute of unstimulated cells].

Anti-PA IgG Specific B Cells: Antigen specific B cells were enumerated by ELISpot assay as described in detail elsewhere (5 - 7) and modified for the proliferation and detection of rhesus macaque IgG secreting cells. Macaque PBMC were plated in a 24-well plates at 5 x 10^5 cells/well in R-10 medium supplemented with a mix of polyclonal mitogens: 1/10,000 Pokeweed Mitogen extract, 6 µg/ml CpG ODN-2006, and 1/10,000 Staphylococcus Aureus, Cowan strain (SAC) (Sigma). Cells were cultured for 6 days at 37ºC, 6-8% CO_2. For ELISpot detection, 96-well filter plates (Millipore, MAHA N4510) were coated overnight with rPA at 1 µg/ml. KLH (2.5 µg/ml) was used as an antigen control. Total and rPA specific IgG-secreting cells were detected using 10 µg/ml goat anti-monkey Ig (Accurate Chem. Co). Data were represented as the frequency (percentage) of rPA-specific anti-PA secreting cells versus the total IgG^+ secreting cells in PBMC. The lower limit of detection (LOD) was 0.002 antigen-specific IgG^+ secreting cells per 10^6 PBMC.

Primary data set construction, variable masking, transformation and standardization. Data were from control and vaccinated animals that completed the study (Table 1). Except for vaccine dose and the interval between first vaccination and aerosol challenge ('duration') the primary data set was constructed with each variable corresponding to an assay with measurements approximately every four weeks. Values of TNA (1) and anti-PA IgG (3) that were lower than their lower limit of detection (TNA LOD = 11; anti-PA IgG LOD = 0.4 µg/mL) were replaced with half of their LOD values. Data from assays without an established LOD were transformed by scaling followed by addition of 1. The scaling was performed by multiplying each value within a variable with the same number so that the lowest non-zero value within the variable became 3. The addition of 1 prevents zeros from being lost during log transformation (8). For the non-zero data points that have values below 3, log transformation with addition of 1 significantly change the positions of these data points compared to log transformation without addition of 1. After making the smallest non-zero value of each variable be 3, addition of 1 followed by log transformation still preserves the positions of these low values compared to log transformation without addition of 1. For assessment of the relative contributions of humoral and cellular immune responses, ratio variables were generated by dividing Th2 response related variables by Th1 response related variables. The ratio variables were the ratio of IL-4 mRNA to IFN-γ mRNA (r_il4IFNm), the ratio of secreted IL-4 protein to secreted IFN-γ protein (R_IL4IFNe), and the ratio of the frequency of IL-4-secreting cells to that of IFN-γ-secreting cells (R_IL4IFNeli). All the assay variables were then log10 transformed and standardized with a mean of 0 and a standard deviation of 1.

Data set re-construction. The data set was re-constructed from the primary data set by converting the measurement at each time point into an individual variable. The measurements at different study time points were then treated as independent variables (e.g. anti-PA IgG at month 6 is one variable IgG_6, and anti-PA IgG at month 7 is a separate variable, IgG_7). Except for the last available time point prior to *B. anthracis* spore aerosol challenge, all time points after month 12 were excluded due to the fact that further time points were unavailable for animals challenged at month 12. The month 7 time point, which is 1 month after the priming series, was designated ‘Peak’, and the last available sample time point prior to challenge was designated ‘Last’ for all NHP. The final assay variables (n=80) used in the analysis are listed in Table 2.

Missing value imputation. To impute missing values, Proc MI (SAS® version 9.3, SAS Institute Inc. Cary, NC USA) with the expectation–maximization (EM) algorithm was used to generate 20 imputed data sets. Due to the presence of multicollinearity among some variables, Proc MI was performed separately at different study time points. At each study time point, mRNA variables, cytokine-ELISA variables and ratio variables were imputed separately. Vaccine dose was not included for imputations, because the collinearity of dose with other variables varies across different time points. Variables that made the EM algorithm not converge were excluded. These variables were anti-PA IgG1, IgG2, IgG3 and IgG4, and anti-PA IgG-specific B cells. In addition, variables at some time points with identical observed values across all the animals were excluded. The time points used for the variables were included in Table 2. In total, 80 immunological variables together with vaccine dose and duration (n = 82) were included in the 20 imputed data sets for selecting variables (Table 2). Table 3 summarizes the three imputations performed at each time point. All variables were standardized with a mean of 0 and a variance of 1 prior to evaluating for COP.

Variable selections by LASSO and elastic net penalized logistic regressions. Multiple methods were implemented in various software packages for the purpose of identifying correlations. Each method or package has various strengths and weaknesses. In order to have the highest confidence that the best correlates are identified, we selected software packages that employ two statistical approaches and differ in their optimization algorithms and penalty parameter tuning. Optimal or parsimonious LASSO and elastic net variable selections were performed in three R packages Glmnet (9), Elasticnet (10), Pensim (11), and the C++ software package BBR (Bayesian Binary Regression) (12). Penalized logistic regression is to maximizing the penalized log likelihood $l(\beta)_{\text{penalized}} = l(\beta) - \lambda_1(|\hat{\beta}_1| + \ldots + |\hat{\beta}_p|) - \lambda_2(\hat{\beta}_1^2 + \ldots + \hat{\beta}_p^2)$, where $l(\beta)$ is the log likelihood, λ_1 is a LASSO penalty parameter; λ_2 is a Ridge penalty parameter; and $\hat{\beta}_1, \ldots, \hat{\beta}_p$ are the parameter estimates for variables $X_{1,} \ldots, X_p$ respectively. λ_2 is equal to 0 in LASSO penalized logistic regression; λ_1 is equal to 0 in Ridge penalized logistic regression; and neither is 0 in elastic net penalized logistic regression. LASSO may undergo too stringent shrinkage and thus ignore important predictors, while elastic net has the grouping effect, selecting important predictors even if they are correlated. Elastic net may however select too many predictors, resulting in overfitting in the prediction model. Among the four packages, one dimensional or two dimensional penalty parameter tuning was done by repeated (60 times) 10-fold cross-validation, where the data were randomly and evenly split into 10 subsets and cross-validation was performed 10 times with each subset being used as the validation data set once for testing the model and the remaining 9 subsets as the training data set for the model. The 10 sets of results generated were then summarized to produce a single estimation of the prediction error. When feasible based on the software package features, both LASSO and elastic net approaches were used and two sets of variables were selected; an optimal set and a parsimonious set. The optimal set of variables was selected when the cross validation error was the minimum or the cross-validated likelihood was the maximum, thus minimizing prediction error. The parsimonious sets of variables were selected by applying the “1-standard error rule” (13), choosing the variables when the cross validation error reached the sum of the minimum cross validation error and one standard error, thus minimizing overfitting. BBR only performs LASSO selection by applying the

Laplace prior to the parameter space, while only the optimal variable set can be obtained from Pensim because its cross-validation is based on maximum likelihood and the “1-standard error rule” can not be applied.

With each of the twenty imputed data sets, variable selection was accomplished using each permutation of LASSO and elastic net with software package and optimal or parsimonious set. In order to summarize all the selected variables from twenty imputed data sets into a single variable set, rank, frequency and score were generated. Within each set of selected variables from each imputed data set, the variables were ordered from high to low according to their regression coefficients and were then assigned numbers in descending order from 82 with a difference of 1 between neighboring variables. Rank was obtained by adding these assigned numbers across the imputed data sets where the variable was selected among the twenty imputed data sets. Frequency indicates the number of times each variable was selected out of the 20 imputed data sets. Score is the product of rank and frequency. Variables with a frequency of ≥10 were chosen for further analyses (Tables S5-S10) (14, 15).

Evaluation of survival prediction models with selected sets of variables. Collinearity or multicollinearity, arising from the correlations among variables in the model, can generate large standard errors in the coefficient estimates in the model. When a sample set has a different collinearity or multicollinearity pattern from that used for building the model the cross-sample predictions are not reliable (16). Collinearity or multicollinearity diagnoses were performed by Proc REG in SAS® version 9.3, with cutoff values of 0.4 for tolerance, 2.5 for variance inflation factor (VIF) and 10 for condition number (17, 18).

For variable sets that were diagnosed as having multicollinearity, PCLR were performed by doing PCA using Prcomp in R followed by logistic regression using glm.fit in R with models ln[$\hat{\pi}$/(1- $\hat{\pi}$)] = $\hat{\beta}_0$ + $\hat{\beta}_1 Z_1$ + ... + $\hat{\beta}_p Z_p$, where $\hat{\pi}$ was the estimated probability of survival given scores Z_1, …, Z_p, the centered values multiplied by the eigenvectors generated from PCA for principal components 1…p respectively, $\hat{\beta}_0$ was the estimated intercept of the PCLR model, and $\hat{\beta}_1$, …, $\hat{\beta}_p$were the parameter estimates for scores Z_1, …, Z_p respectively. For variable sets that did not have collinearity or multicollinearity, logistic regressions were performed by glm.fit in R. The models were ln[$\hat{\pi}$/(1- $\hat{\pi}$)] = $\hat{\beta}_0$ + $\hat{\beta}_1 X_1$ + … + $\hat{\beta}_p X_p$, where $\hat{\pi}$ was the estimated probability of survival given variables X_1, …, X_p, $\hat{\beta}_0$was the estimated intercept of the logistic regression model, and $\hat{\beta}_1$, …, $\hat{\beta}_p$ were the parameter estimates for variables X_1, …, X_p respectively.

AUC was generated for each regression model. AUC was the probability for ranking a randomly chosen survivor NHP higher than a randomly chosen non-survivor NHP. The higher the AUC is, the higher the discriminative accuracy of the model. An AUC greater than 0.90 indicated high accuracy; AUC of 0.80–0.90 indicated good accuracy; 0.70–0.80 moderate accuracy and 0.50–0.70 indicated low accuracy approaching random probability (19, 20). To compare AUCs between models, paired permutation tests in R were performed (21, 22), with a Bonferroni-corrected significance level of 0.0025 for multiple comparisons.

SUPPLEMENTAL REFERENCES

1. **Li H, Soroka SD, Taylor TH Jr, Stamey KL, Stinson KW, Freeman AE, Abramson DR, Desai R, Cronin LX, Oxford JW, Caba J, Pleatman C, Pathak S, Schmidt DS, Semenova VA, Martin SK, Wilkins PP, Quinn CP.** 2008. Standardized, mathematical model-based and validated in vitro analysis of anthrax lethal toxin neutralization. J. Immunol. Methods. **333:**89-106.

2. **Semenova, V.A., E. Steward-Clark, K.L. Stamey, T.H. Taylor Jr., D.S. Schmidt, S.K. Martin, N. Marano N, and C.P. Quinn.** 2004. Mass value assignment of total and subclass immunoglobulin G in a human standard anthrax reference serum. Clin. Diagn. Lab. Immunol. **11**:919-923.

3. **Longworth, E., R. Borrow, D. Goldblatt, P. Balmer, M. Dawson, N. Andrews, E. Miller, and K. Cartwright.** 2002. Avidity maturation following vaccination with a meningococcal recombinant hexavalent PorA OMV vaccine in UK infants. Vaccine. **20**:2592-2596.

4. **Pahar, B., J. Li, T. Rourke, C.J. Miller, and M.B. McChesney.** 2003. Detection of antigen-specific T cell interferon gamma expression by ELISPOT and cytokine flow cytometry assays in rhesus macaques. J. Immunol. Methods. **282**:103-15.

5. **Crotty, S., R.D. Aubert, J. Glidewell, and R. Ahmed.** 2004. Tracking human antigen-specific memory B cells: a sensitive and generalized ELISPOT system. J. Immunol. Methods. **286**:111-122.

6. **Crotty S, P. Felgner, H. Davies, J. Glidewell, L. Villarreal, and R. Ahmed.** 2003. Cutting edge: long-term B cell memory in humans after smallpox vaccination. J. Immunol. **171**:4969–73.

7. **Quinn CP, Sabourin CL, Niemuth NA, Li H, Semenova VA, Rudge TL, Mayfield HJ, Schiffer J, Mittler RS, Ibegbu CC, Wrammert J, Ahmed R, Brys AM, Hunt RE, Levesque D, Estep JE, Barnewall RE, Robinson DM, Plikaytis BD, Marano N.** 2012. A Three-Dose Intramuscular Injection Schedule of Anthrax Vaccine Adsorbed Generates Sustained Humoral and Cellular Immune Responses to Protective Antigen and Provides Long-Term Protection against Inhalation Anthrax in Rhesus Macaques. Clin. Vaccine Immunol. **19:**1730-1745.

8. **Osborne, J.** 2002. Notes on the use of data transformations. Practical Assessment, Res. Eval. **8**. http://pareonline.net/getvn.asp?v=8&n=6

9. **Friedman J, Hastie T, Tibshirani R.** 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. J. Stat. Softw. **33:**1-22.

10. **Zou H, Hastie T.** 2005. Regularization and variable selection via the elastic net. J. Roy. Stat. Soc. B. **67:**301-320.

11. **Waldron L, Pintilie M, Tsao MS, Shepherd FA, Huttenhower C, Jurisica I.** 2011. Optimized application of penalized regression methods to diverse genomic data. Bioinformatics **27:**3399-3406.

12. **Genkin A, Lewis DD, Madigan D.** 2007. Large-scale Bayesian logistic regression for text categorization. Technometrics **49:**291-304.

13. **Hastie T, Tibshirani R, Friedman J.** 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed, Springer-Verlag, New York, NY.

14. **Heymans MW, van Buuren S, Knol DL, van Mechelen W, de Vet HCW.** 2007. Variable selection under multiple imputation using the bootstrap in a prognostic study. BMC Med. Res. Methodol. **7**:33.

15. **Austin PC, Tu JV.** 2004. Bootstrap Methods for Developing Predictive Models. The American Statistician, **58**:131-137.

16. **Chatterjee S, Hadi AS, Price B** 2000. Regression Analysis by Example, 3rd Edition, A Wiley-Interscience Publication, John Wiley and Sons.

17. **Allison, PD.** 1999. Multiple Regression: A Primer. Pine Forge Press, Thousand Oaks, CA.

18. **Belsley DA, Kuh K, Welsch RE.** 1980. Regression diagnostics: Identifying influential data and sources of collinearity, John Wiley & Sons, New York, NY.

19. **Swets, JA.** 1988. Measuring the accuracy of diagnostic systems. Science **240:**1285–1293.

20. **Wigton, RS, Connor, JL, Centor, RM.** 1986. Transportability of a decision rule for the diagnosis of streptococcal pharyngitis. Arch. Intern. Med. **146:**81-83.

21. **Venkatraman ES.** 2000. A permutation test to compare receiver operating characteristic curves. Biometrics **56:**1134-1138.

22. **Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M.** 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics **12:**77.

TABLE LEGENDS

TABLE S1 Gene Accession Numbers

Accession numbers for six cytokine genes that were determined by real-time quantitative TaqMan RT-PCR.

TABLE S2 NHP Assay by Design Primer/Probe Sets

Forward and reverse primers and probes were designed for real-time quantitative TaqMan RT-PCR of six cytokine genes.

TABLE S3 NHP Assay by Design Primer/Probe Sets

Six ELISA kits were used for detecting six cytokine proteins.

TABLE S4 Missing rates (%) for variables

NA not available.

TABLE S5 Imputations at each time point

Three sets of imputations were performed at each time point. In each set of imputation, variables were included if there was no multicollinearity present. Some variables were used in more than one set (e.g. TNA), and these variables were retained for analysis from only 1 imputed set.

TABLE S6 Introduction of variable selection software packages

Four software packages were used for variable selections, with different languages, optimization algorithms and criteria for tuning penalty parameters.

TABLE S7 Summary of optimal variable selections

Optimal variable selections were performed with 7 selection methods. * p value < 0.05 from Wald Chi-Square test of the parameter estimate; √ Variables that were selected by all the optimal selection methods.

TABLE S8 Variables selected by BBR

Variables were selected by LASSO with C++ package BBR. Parsimonious: Parsimonious variable set. Optimal: Optimal variable set. The selected variables were ordered from high to low by ordering the regression coefficients from high to low, and were then assigned numbers in descending order from 82 with a difference of 1 between neighboring variables. Rank was obtained by adding these assigned numbers across the imputed data sets where the variable was selected. Frequency indicates the number of times each variable was selected out of the 20 imputed data sets. Score is the product between rank and frequency.

TABLE S9 Parsimonious sets of variables selected by LASSO and elastic net with Elasticnet package

Parsimonious variables were selected by LASSO or elastic net with the Elasticnet package. The selected variables were ordered from high to low by ordering the regression coefficients from high to low, and were then assigned numbers in descending order from 82 with a difference of 1 between neighboring variables. Rank was obtained by adding these assigned numbers across the imputed data sets where the variable was selected. Frequency indicates the number of times each variable was selected out of the 20 imputed data sets. Score is the product between rank and frequency.

TABLE S10 Optimal sets of variables selected by LASSO and elastic net with Elasticnet package

Optimal variables were selected by LASSO or elastic net with the Elasticnet package. The selected variables were ordered from high to low by ordering the regression coefficients from high to low, and were then assigned numbers in descending order from 82 with a difference of 1 between neighboring variables. Rank was obtained by adding these assigned numbers across the imputed data sets where the variable was selected. Frequency indicates the number of times each variable was selected out of the 20 imputed data sets. Score is the product between rank and frequency.

TABLE S11 Optimal sets of variables selected by LASSO and elastic net with Pensim package

Optimal variables were selected by LASSO or elastic net with the Pensim package. The selected variables were ordered from high to low by ordering the regression coefficients from high to low, and were then assigned numbers in descending order from 82 with a difference of 1 between neighboring variables. Rank was obtained by adding these assigned numbers across the imputed data sets where the variable was selected. Frequency indicates the number of times each variable was selected out of the 20 imputed data sets. Score is the product between rank and frequency.

TABLE S12 Parsimonious sets of variables selected by LASSO and elastic net with Glmnet package

Parsimonious variables were selected by LASSO or elastic net with the Glmnet package. The selected variables were ordered from high to low by ordering the regression coefficients from high to low, and were then assigned numbers in descending order from 82 with a difference of 1 between neighboring variables. Rank was obtained by adding these assigned numbers across the imputed data sets where the variable was selected. Frequency indicates the number of times each variable was selected out of the 20 imputed data sets. Score is the product between rank and frequency.

TABLE S13 Optimal sets of variables selected by LASSO and elastic net with Glmnet package

Optimal variables were selected by LASSO or elastic net with the Glmnet package. The selected variables were ordered from high to low by ordering the regression coefficients from high to low, and were then assigned numbers in descending order from 82 with a difference of 1 between neighboring variables. Rank was obtained by adding these assigned numbers across the imputed data sets where the variable was selected. Frequency indicates the number of times each variable was selected out of the 20 imputed data sets. Score is the product between rank and frequency.

TABLE S14 Comparing performance of regression models with parsimonious variable sets

The AUCs of logistic regression and PCLR models were compared with that of the logistic regression model with variables ‘Last’ anti-PA IgG and SI at month 2 by paired permutation tests with the twenty imputed data sets, with a Bonferroni-corrected significance level of 0.0025 for multiple comparisons. * $p < 0.0025$.

TABLE S1 Gene Accession Numbers

Gene	Accession Number
IFN-γ	L26024
IL-2	U19847
IL-4	L26027
IL-6	L26028
IL-1β	U19845
TNF-α	U19850

Accession numbers for six cytokine genes that were determined by real-time quantitative TaqMan RT-PCR.

TABLE S2 NHP Assay by Design Primer/Probe Sets

Forward Primer Name	Forward Primer Sequence
IFN-gamma-366F	AAACGGGATGACTTTGAAAAGCT
IL-2-207F	ACCAGGATGCTCACATTTAAGTTTT
IL-4-360F	AACGGCTCGACAGGAACCT
IL-6-210F	CATCCTCGACGGCATCTCA
IL-1-47F	GAGCTCGCCAGTGAAATGATG
TNF-232F	CCCAAGGACCCCTCTCTAATCAG
Reverse Primer Name	**Reverse Primer Sequence**
IFN-gamma-366R	GCTTTGCGTTGGACATTTGAG
IL-2-207R	CCAGAGGTTTGAGTTCTTCTTCTAGAC
IL-4-360R	CTCTGGTTGGCTTCCTTCACA
IL-6-210R	TGCTTTCACACATGTTACTCCTGTT
IL-1-47R	CATCGACGTCAAAGAACAAGTCATC
TNF-232R	GGGCTACAGGCTTGTCACTT
Probe Name	**Probe Sequence**
IFN-gamma-366M2	CAGTTACCGAATAATTG
IL-2-207M2	CTGTGGCCTTCTTG
IL-4-360M2	AGGAGTTCAAGCCC
IL-6-210M1	CCTGAGAAAGGAGACATG
IL-1-47M2	ACTACAGCGGCAACGAG
TNF-232M2	CAGGCAGTCAGATCAT

Forward and reverse primers and probes were designed for real-time quantitative TaqMan RT-PCR of six cytokine genes.

TABLE S3 ELISA Kits

Cytokine	Company	Kit	Catalog Number
IFN-γ	R&D	Quantikine human IFN-γ	DIF50
IL-2	R&D	Quantikine human IL-2	D2050
IL-4	BD PharMingen	Opt EIA human IL-4	550614
IL-6	R&D	Quantikine human IL-6	D6050
IL-1β	R&D	Quantikine human IL-1β	DLB50
TNF-α	BD PharMingen	Opt EIA human TNF-α	550610

Six ELISA kits were used for detecting six cytokine proteins.

TABLE S4 Missing rates (%) for variables

Assay type	Variable Name	Target	Time Points (Months)				
			1	2	6	7	'Last'
ELISA	IgG	Anti-PA IgG protein	0.73	1.46	0	0	0
	IL1Be	IL-1β protein	27.74	10.22	64.23	13.14	NA
	IL2e	IL-2 protein	27.01	2.19	57.66	9.49	NA
	IL4e	IL-4 protein	27.74	2.19	60.58	9.49	NA
	IL6e	IL-6 protein	27.01	41.61	58.39	9.49	NA
	IFNe	IFN-γ protein	28.47	2.19	60.58	9.49	NA
	TNFe	TNF-α protein	28.47	3.65	58.39	9.49	NA
	R_IL4IFNe	Ratio of IL-4 protein to IFN-γ protein	29.20	2.19	63.50	9.49	NA
RT-PCR	IL1Bm	IL-1β mRNA	1.46	3.65	10.95	1.46	NA
	IL2m	IL-2 mRNA	1.46	3.65	10.95	1.46	NA
	IL4m	IL-4 mRNA	1.46	3.65	10.95	1.46	NA
	IL6m	IL-6 mRNA	1.46	3.65	10.95	1.46	NA
	IFNm	IFN-γ mRNA	1.46	3.65	10.95	1.46	NA
	TNFm	TNF-α mRNA	1.46	3.65	10.95	1.46	NA
	R_IL4IFNm	Ratio of IL-4 mRNA to IFN-γ mRNA	1.46	3.65	10.95	1.46	NA
Toxin neutralization assay	TNA	Toxin Neutralization Activity ED50	0.73	1.46	0	0	0
Lymphocyte stimulation	SI	Lymphocyte Stimulation Index	2.92	2.19	51.82	0.73	33.58

assay							
Avidity assay	AI	Anti-PA IgG avidity	61.31	23.36	77.37	18.24	NA
ELISpot	INFeli	Frequency of IFN-γ-secreting cells	54.01	53.28	13.14	21.17	9.49
	IL4eli	Frequency of IL-4-secreting cells	59.12	51.09	8.76	17.52	35.77
	R_IL4IFNeli	Ratio of frequency of IL-4-secreting cells to frequency of IFN-γ-secreting cells	70.07	57.66	13.14	29.93	35.77

NA not available.

TABLE S5 Imputations at each time point

	Imputation 1	Imputation 2	Imputation 3
Variables used for imputation	Survival	Survival	survival
	control	control	control
	IgG	IgG	IgG
	TNA	TNA	TNA
	IFNm	IFNeli	R_IL4IFNeli
	IL-1Bm	IL1Be	R_IL4IFNm
	IL2m	IL2e	R_IL4IFNe
	IL4m	IL4e	SI
	IL6m	IL4eli	AI
	TNFm	IL6e	
	SI	TNFe	
	AI	SI	
		AI	
Variables kept after imputation	IgG	IFNeli	R_IL4IFNeli
	TNA	IL1Be	R_IL4IFNm
	IFNm	IL2e	R_IL4IFNe
	IL1Bm	IL4e	
	IL2m	IL4eli	
	IL4m	IL6e	
	IL6m	TNFe	
	TNFm		
	SI		
	AI		

Three sets of imputations were performed at each time point. In each set of imputation, variables were included if there was no multicollinearity present. Some variables were used in more than one set (e.g. TNA), and these variables were retained for analysis from only 1 imputed set.

TABLE S6 Introduction of variable selection software packages

	Variable Selection Software Packages			
	BBR	**Elasticnet**	**Glmnet**	**Pensim**
Language	C++	R	R	R
Optimization algorithm	Imposes Laplace priors, cyclic coordinate descent	Least angle regression (LARS)	Cyclic coordinate descent	Combination of gradient ascent and Newton-Raphson
Criterion for tuning penalty parameters	Maximum cross-validated log-likelihood	Minimum cross-validated mean squared prediction error	Minimum deviance	Maximum cross-validated log-likelihood
Penalty parameter tunning in elastic net	NA	Successive one-dimensional tuning	Successive one-dimensional tuning	Two-dimensional tuning

Four software packages were used for variable selections, with different languages, optimization algorithms, criteria for tuning penalty parameters and penalty parameter tuning in elastic net. NA, not applicable due to the absence of elastic net variable selection in BBR.

Table S7 Summary of optimal variable selections

	Time	Simple logistic regression			BBR	Elasticnet		Pensim		Glmnet	
	Point	Intercept	Parameter	AUC			elastic		elastic		elastic
Variable	(month)	(p value)	(P value)	(95% CI)	LASSO	LASSO	net	LASSO	net	LASSO	net
IgG	6	0.8952	1.7332	0.7724	x		x				x
		(<0.0001)	(<0.0001)*	(0.6956-0.8491)							
	7	-0.6569	1.0009	0.7956		x	x	x	x		x
		(0.0582)	(<0.0001)*	(0.7208-0.8703)							
	last	0.7105	2.1628	0.8214	√	√	√	√	√	√	√
		(0.0015)	(<0.0001)*	(0.7514-0.8914)							
TNA	6	-1.346	1.8551	0.7416	x		x	x	x	x	x
		(0.0060)	(<0.0001)*	(0.6689-0.8143)							
	7	-1.9889	1.0808	0.7918		x	x	x	x	x	x
		(0.0006)	(<0.0001)*	(0.7158-0.8678)							
SI	2	-4.9416	1.4900	0.7860	√	√	√	√	√	√	√
		(<0.0001)	(<0.0001)*	(0.7086-0.8633)							
	6	-3.9370	1.4369	0.7095	√	√	√	√	√	√	√
		(0.0376)	(0.0111)*	(0.5809-0.8381)							
IL4eli	1	0.7150	0.3953	0.5961				x	x		x
		(0.1233)	(0.2996)	(0.4467-0.7455)							
	7	0.4176	0.2921	0.5804				x	x	x	x
		(0.2114)	(0.1621)	(0.4643-0.6965)							
	last	0.6154	0.0560	0.4842	√	√	√	√	√	√	√
		(0.2330)	(0.8384)	(0.3407-0.6277)							
IFNeli	6	-0.3719	0.7881	0.7073		x	x	x	x		x
		(0.3090)	(0.0004)*	(0.6077-0.8068)							
R_IL4IFNeli	6	0.6981	-0.1751	0.5414				x	x	x	x
		(0.0007)	(0.3884)	(0.4256-0.6573)							
	7	0.6808	-0.1508	0.5428		x	x	x	x	x	x
		(0.0020)	(0.3684)	(0.4195-0.6661)							
	last	0.7042	-0.0990	0.5640				x	x		x
		(0.0019)	(0.6792)	(0.4374-0.6906)							
IL1Be	2	-2.2035	1.3679	0.5107				x	x		
		(0.5649)	(0.4383)	(0.4636-0.5578)							

	6	-2.4384	1.3598	0.5395				x	x		
		(0.4219)	(0.3364)	(0.4105-0.6684)							
IL1Bm	7	-0.7914	0.5420	0.5930				x	x	x	x
		(0.4133)	(0.1135)	(0.4937-0.6923)							
IL4e	1	-2.0654	2.0972	0.5149	√	√	√	√	√	√	√
		(0.5856)	(0.4594)	(0.4790-0.5508)							
IL6e	2	-2.4945	1.5251	0.6359					x		
		(0.2042)	(0.0701)	(0.5203-0.7515)							
TNFe	1	3.2168	-1.4247	0.6627	√	√	√	√	√	√	√
		(0.0787)	(0.1622)	(0.5485-0.7769)							
	6	-21.8363	13.1804	0.5208	√	√	√	√	√	√	√
		(0.9858)	(0.9856)	(0.4800-0.5617)							
R_IL4IFNm	6	0.0685	-2.1127	0.5818				x	x	x	x
		(0.8324)	(0.0296)*	(0.4750-0.6885)							
	7	0.6608	-0.7249	0.5829		x	x	x	x	x	x
		(0.0004)	(0.0508)	(0.4844-0.6815)							
Number of Variables					9	12	14	21	22	15	20
Variable Set Identifier					**Opt_LASSO_BBR**	**Opt_LASSO_Elasticnet**	**Opt_Elastic_Elasticnet**	**Opt_LASSO_Pensim**	**Opt_Elastic_Pensim**	**Opt_LASSO_Glmnet**	**Opt_Elastic_Glmnet**

Optimal variable selections were performed with 7 selection methods. * p value < 0.05 from Wald Chi-Square test of the parameter estimate; √ Variables that were selected by all the optimal selection methods.

TABLE S8 Variables selected by BBR

Parsimonious				Optimal			
Variable	Rank	Frequency	Score	Variable	Rank	Frequency	Score
IgG_Last	1626	20	32520	IgG_Last	1625	20	32500
SI_2	1447	18	26046	SI_2	1447	18	26046
SI_6	1062	14	14868	SI_6	1060	14	14840
				TNFe_6	777	10	7770
				IL4e_1	776	10	7760
				TNA_6	764	10	7640
				IL4eLi_7	563	10	5630
				TNFe_1	532	10	5320
				IL4eLi_Last	521	10	5210

Variables were selected by LASSO with C++ package BBR. Parsimonious: Parsimonious variable set. Optimal: Optimal variable set. The selected variables were ordered from high to low by ordering the regression coefficients from high to low, and were then assigned numbers in descending order from 82 with a difference of 1 between neighboring variables. Rank was obtained by adding these assigned numbers across the imputed data sets where the variable was selected. Frequency indicates the number of times each variable was selected out of the 20 imputed data sets. Score is the product between rank and frequency.

TABLE S9 Parsimonious sets of variables selected by LASSO and elastic net with Elasticnet package

LASSO				elastic net			
Variable	Rank	Frequency	Score	Variable	Rank	Frequency	Score
IgG_Last	1640	20	32800	IgG_Last	1640	20	32800
SI_2	1117	14	15638	SI_2	1500	19	28500
SI_6	955	12	11460	IgG_7	1436	18	25848
				SI_6	875	11	9625
				TNA_7	789	10	7890

Parsimonious variables were selected by LASSO or elastic net with the Elasticnet package. The selected variables were ordered from high to low by ordering the regression coefficients from high to low, and were then assigned numbers in descending order from 82 with a difference of 1 between neighboring variables. Rank was obtained by adding these assigned numbers across the imputed data sets where the variable was selected. Frequency indicates the number of times each variable was selected out of the 20 imputed data sets. Score is the product between rank and frequency.

TABLE S10 Optimal sets of variables selected by LASSO and elastic net with Elasticnet package

LASSO				elastic net			
Variable	Rank	Frequency	Score	Variable	Rank	Frequency	Score
IgG_Last	1631	20	32620	IgG_Last	1629	20	32580
SI_2	1596	20	31920	SI_2	1587	20	31740
TNFe_1	1132	17	19244	IgG_7	1490	19	28310
IgG_7	1251	16	20016	TNA_7	1485	19	28215
SI_6	1150	15	17250	IgG_6	1134	15	17010
IL4eli_Last	890	14	12460	SI_6	1062	14	14868
TNA_7	1012	13	13156	IL4e_1	1001	13	13013
IL4e_1	1007	13	13091	TNA_6	963	13	12519
R_IL4IFNeli_7	872	13	11336	R_IL4IFNm_7	812	13	10556
R_IL4IFNm_7	867	13	11271	TNFe_1	769	13	9997
TNFe_6	876	11	9636	IL4eli_Last	758	13	9854
IFNeli_6	762	10	7620	TNFe_6	949	12	11388
				R_IL4IFNeli_7	614	10	6140
				IL4eli_7	595	10	5950

Optimal variables were selected by LASSO or elastic net with the Elasticnet package. The selected variables were ordered from high to low by ordering the regression coefficients from high to low, and were then assigned numbers in descending order from 82 with a difference of 1 between neighboring variables. Rank was obtained by adding these assigned numbers across the imputed data sets where the variable was selected. Frequency indicates the number of times each variable was selected out of the 20 imputed data sets. Score is the product between rank and frequency.

TABLE S11 Optimal sets of variables selected by LASSO and elastic net with Pensim package

LASSO				elastic net			
Variable	Rank	Frequency	Score	Variable	Rank	Frequency	Score
IgG_Last	1625	20	32500	IgG_Last	1626	20	32520
SI_2	1598	20	31960	SI_2	1600	20	32000
IL4eLi_Last	1200	20	24000	IL4e_1	1537	20	30740
IL4e_1	1469	19	27911	TNA_6	1521	20	30420
TNA_6	1449	19	27531	IL4eLi_Last	1119	20	22380
R_IL4IFNm_7	1199	19	22781	TNA_7	1443	19	27417
TNFe_1	1148	19	21812	R_IL4IFNm_7	1113	19	21147
SI_6	1354	18	24372	SI_6	1347	18	24246
IL4eLi_7	1141	18	20538	IL1Bm_7	1055	18	18990
IL1Bm_7	1071	17	18207	IL4eLi_7	1048	18	18864
R_IL4IFNeLi_7	1010	16	16160	TNFe_1	991	18	17838
TNFe_6	1182	15	17730	IgG_7	1250	17	21250
TNA_7	1166	15	17490	R_IL4IFNeLi_7	1008	17	17136
IL4eLi_1	873	13	11349	TNFe_6	1183	15	17745
R_IL4IFNm_6	858	13	11154	R_IL4IFNm_6	878	14	12292
R_IL4IFNeLi_6	838	13	10894	R_IL4IFNeLi_6	780	13	10140
IL1Be_2	773	12	9276	IL4eLi_1	785	12	9420
R_IL4IFNeLi_Last	793	11	8723	IL1Be_6	719	12	8628
IL1Be_6	700	11	7700	R_IL4IFNeLi_Last	752	11	8272
IFNeLi_6	744	10	7440	IL1Be_2	627	11	6897
IgG_7	729	10	7290	IL6e_2	733	10	7330
				IFNeLi_6	724	10	7240

Optimal variables were selected by LASSO or elastic net with the Pensim package. The selected variables were ordered from high to low by ordering the regression coefficients from high to low, and were then assigned numbers in descending order from 82 with a difference of 1 between neighboring variables. Rank was obtained by adding these assigned numbers across the imputed data sets where the variable was selected. Frequency indicates the number of times each variable was selected out of the 20 imputed data sets. Score is the product between rank and frequency.

TABLE S12 Parsimonious sets of variables selected by LASSO and elastic net with Glmnet package

LASSO				elastic net			
Variable	Rank	Frequency	Score	Variable	Rank	Frequency	Score
IgG_Last	1628	20	32560	IgG_Last	1639	20	32780
SI_2	1358	17	23086	SI_2	1515	19	28785
SI_6	799	10	7990	IgG_6	1471	19	27949
				IgG_7	1406	18	25308
				TNA_7	1395	18	25110
				R_IL4IFNm_7	1164	17	19788
				SI_6	1247	16	19952
				TNA_6	1126	15	16890
				IFNeli_6	902	12	10824
				R_IL4IFNeli_7	808	12	9696
				TNFe_1	794	12	9528
				IL4e_1	737	10	7370

Parsimonious variables were selected by LASSO or elastic net with the Glmnet package. The selected variables were ordered from high to low by ordering the regression coefficients from high to low, and were then assigned numbers in descending order from 82 with a difference of 1 between neighboring variables. Rank was obtained by adding these assigned numbers across the imputed data sets where the variable was selected. Frequency indicates the number of times each variable was selected out of the 20 imputed data sets. Score is the product between rank and frequency.

TABLE S13 Optimal sets of variables selected by LASSO and elastic net with Glmnet package

LASSO				elastic net			
Variable	Rank	Frequency	Score	Variable	Rank	Frequency	Score
IgG_Last	1624	20	32480	IgG_Last	1602	20	32040
SI_2	1596	20	31920	SI_2	1592	20	31840
R_IL4IFNm_7	1366	20	27320	IL4E_1	1544	20	30880
IL4e_1	1405	18	25290	TNA_7	1538	20	30760
TNA_6	1371	18	24678	TNA_6	1497	20	29940
SI_6	1299	17	22083	R_IL4IFNm_7	1160	20	23200
TNA_7	1297	17	22049	IL4eLi_Last	1159	20	23180
TNFe_1	1125	17	19125	SI_6	1352	18	24336
IL4eLi_Last	1109	17	18853	IL1Bm_7	1086	18	19548
R_IL4IFNeLi_7	951	14	13314	R_IL4IFNeLi_7	1067	18	19206
TNFe_6	939	12	11268	TNFe_1	1016	18	18288
IL1Bm_7	745	11	8195	IgG_7	1272	17	21624
IL4eLi_7	722	11	7942	IL4eLi_7	993	17	16881
R_IL4IFNm_6	692	10	6920	TNFe_6	1245	16	19920
R_IL4IFNeLi_6	686	10	6860	IgG_6	1163	16	18608
				R_IL4IFNm_6	865	14	12110
				R_IL4IFNeLi_6	801	13	10413
				IFNeLi_6	854	12	10248
				IL4eLi_1	717	11	7887
				R_IL4IFNeLi_Last	647	10	6470

Optimal variables were selected by LASSO or elastic net with the Glmnet package. The selected variables were ordered from high to low by ordering the regression coefficients from high to low, and were then assigned numbers in descending order from 82 with a difference of 1 between neighboring variables. Rank was obtained by adding these assigned numbers across the imputed data sets where the variable was selected. Frequency indicates the number of times each variable was selected out of the 20 imputed data sets. Score is the product between rank and frequency.

TABLE S14 Comparing performance of regression models

Model	'Last' anti-PA IgG, SI_2			Par_LASSO			Par_Elastic_Elasticnet			Par_Elastic_Glmnet		
Data Set	AUC	CI	p value	AUC	CI	p value	AUC	CI	p value	AUC	CI	p value
1	0.8419	0.7762-0.9075	Reference Model	0.8529	0.7883-0.9175	0.5185	0.8529	0.7890-0.9167	0.5175	0.8981	0.8426-0.9536	0.0050
2	0.8375	0.7705-0.9045	Reference Model	0.8419	0.7762-0.9076	0.9630	0.8412	0.7753-0.9070	0.8925	0.9186	0.8724-0.9648	0.0040
3	0.8382	0.7715-0.9050	Reference Model	0.8424	0.7775-0.9073	0.8830	0.8429	0.7782-0.9075	0.8755	0.8988	0.8475-0.9502	0.0155
4	0.8434	0.7783-0.9084	Reference Model	0.8458	0.7806-0.9110	0.8670	0.8456	0.7802-0.9109	0.5585	0.8913	0.8379-0.9446	0.0620
5	0.8448	0.7800-0.9097	Reference Model	0.8460	0.7813-0.9108	0.5500	0.8441	0.7789-0.9093	0.7590	0.8908	0.8346-0.9469	0.0335
6	0.8394	0.7730-0.9059	Reference Model	0.8473	0.7825-0.9120	0.9715	0.8460	0.7810-0.9111	0.9650	0.9179	0.8671-0.9687	0.0030
7	0.8404	0.7742-0.9067	Reference Model	0.8438	0.7783-0.9094	0.7965	0.8453	0.7800-0.9106	0.7945	0.9062	0.8544-0.9580	0.0080
8	0.8409	0.7748-0.9070	Reference Model	0.8495	0.7861-0.9128	0.7690	0.8514	0.7890-0.9139	0.8150	0.9128	0.8637-0.9618	0.0055
9	0.8380	0.7711-0.9048	Reference Model	0.8514	0.7874-0.9155	0.5675	0.8522	0.7884-0.9159	0.5170	0.8876	0.8300-0.9451	0.0120
10	0.8416	0.7759-0.9074	Reference Model	0.8436	0.7781-0.9091	0.6165	0.8451	0.7798-0.9103	0.6730	0.9137	0.8686-0.9589	0.0095
11	0.8424	0.7768-0.9080	Reference Model	0.8414	0.7756-0.9072	0.2700	0.8394	0.7730-0.9058	0.8885	0.8768	0.8188-0.9349	0.0755
12	0.8436	0.7785-0.9087	Reference Model	0.8443	0.7793-0.9094	0.9220	0.8456	0.7808-0.9103	0.8520	0.8893	0.8336-0.9450	0.0375
13	0.8382	0.7715-0.9050	Reference Model	0.8680	0.8052-0.9309	0.0345	0.8688	0.8063-0.9312	0.0410	0.9052	0.8505-0.9599	0.0020*
14	0.8392	0.7727-0.9057	Reference Model	0.8497	0.7862-0.9132	0.9115	0.8495	0.7859-0.9130	0.9540	0.8991	0.8456-0.9525	0.0180
15	0.8419	0.7761-0.9077	Reference Model	0.8524	0.7899-0.9149	0.5615	0.8517	0.7890-0.9143	0.6730	0.9152	0.8688-0.9616	0.0100
16	0.8385	0.7717-0.9052	Reference Model	0.8509	0.7878-0.9140	0.7515	0.8539	0.7916-0.9161	0.7895	0.8976	0.8417-0.9535	0.0075
17	0.8419	0.7762-0.9075	Reference Model	0.8460	0.7815-0.9106	0.3655	0.8456	0.7809-0.9102	0.6020	0.9079	0.8600-0.9557	0.0110
18	0.8436	0.7786-0.9086	Reference Model	0.8529	0.7897-0.9161	0.4330	0.8500	0.7858-0.9141	0.5885	0.9069	0.8590-0.9548	0.0185
19	0.8421	0.7765-0.9078	Reference Model	0.8592	0.7985-0.9200	0.6695	0.8631	0.8030-0.9233	0.6640	0.9064	0.8541-0.9587	0.0085
20	0.8399	0.7736-0.9063	Reference Model	0.8539	0.7908-0.9170	0.6600	0.8541	0.7912-0.9171	0.6165	0.9035	0.8514-0.9556	0.0085
Mean	0.8409		Reference Model	0.8492		0.6541	0.8494		0.7018	0.9022		0.0178
Median	0.8413		Reference Model	0.8484		0.6648	0.8478		0.7160	0.9044		0.0098
Min	0.8375		Reference Model	0.8414		0.0345	0.8394		0.0410	0.8768		0.0020*
Max	0.8448		Reference Model	0.8680		0.9715	0.8688		0.9650	0.9186		0.0755

The AUCs of logistic regression and PCLR models were compared with that of the logistic regression model with variables ‘Last’ anti-PA IgG and SI at month 2 by paired permutation tests with the twenty imputed data sets, with a Bonferroni-corrected significance level of 0.0025 for multiple comparisons. * p < 0.0025.