Combining antibody markers for serosurveillance of SARS-CoV-2 to estimate seroprevalence and time-since-infection

Md S. Bhuiyan[1], Ben J. Brintz[2], Alana L. Whitcombe[3,4], Alena J. Markmann[5], Luther A. Bartelt[5], Nicole J. Moreland[3,4] (iD), Andrew S. Azman[6,7] (iD) and Daniel T. Leung[1,8] (iD)

[1]Division of Infectious Disease, University of Utah School of Medicine, Salt Lake City, UT, USA; [2]Division of Epidemiology, University of Utah School of Medicine, Salt Lake City, UT, USA; [3]Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand; [4]Maurice Wilkins Center, University of Auckland, Auckland, New Zealand; [5]Department of Medicine, Division of Infectious Diseases, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA; [6]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; [7]Faculty of Medicine, Institute of Global Health, University of Geneva, Geneva, Switzerland and [8]Division of Microbiology & Immunology, University of Utah School of Medicine, Salt Lake City, UT, USA

Abstract

Serosurveillance is an important epidemiologic tool for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), used to estimate infection rates and the degree of population immunity. There is no general agreement on which antibody biomarker(s) should be used, especially with the rollout of vaccines globally. Here, we used random forest models to demonstrate that a single spike or receptor-binding domain (RBD) antibody was adequate for classifying prior infection, while a combination of two antibody biomarkers performed better than any single marker for estimating time-since-infection. Nucleocapsid antibodies performed worse than spike or RBD antibodies for classification, but can be useful for estimating time-since-infection, and in distinguishing infection-induced from vaccine-induced responses. Our analysis has the potential to inform the design of serosurveys for SARS-CoV-2, including decisions regarding a number of antibody biomarkers measured.

Introduction

Increasingly, cross-sectional serosurveillance is being used to estimate the fraction of the population previously infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Representative seroepidemiological studies reveal the immune landscape of the population, and compared to the use of data based on viral detection, they can provide more accurate insights into the infection fatality rate, the amplitude of transmission in different populations, and highlight disparities in infection rates without typical health-seeking behaviour biases [1]. Further, such population-level surveys, when coupled with statistical and/or mechanistic models, could be used to estimate the probability and timing of future waves of disease, measure the impact of interventions such as physical distancing or vaccination, and in later stages, confirm the absence of transmission [2]. However, current knowledge of the kinetics of antibody responses to SARS-CoV-2 infection is insufficient to fully realise the array of use-cases for data from population-level seroepidemiological studies. For those designing serosurveys, the choice of antibody assays can be daunting given the number of available tests that target different antigens and isotypes. The aim of this study was to provide new evidence to highlight the best types of antibody biomarkers for estimation of seroprevalence and time-since-SARS-CoV-2 infection, and whether a combination of antibody biomarkers could improve such estimations.

Methods

Data sources

We identified studies in the literature or on preprint servers that measured multiple (>3) antibody responses at varying time points greater than a median of 50 days after PCR-confirmed SARS-CoV-2 infection [3–7]. We used no inclusion/exclusion criteria based on case symptoms nor C_t values. We chose this cut off to optimise the capture of data during the period of antibody decay post-infection, considering the potential range of post-infection time points in population-based serosurveys. Data that were not publicly available were obtained on request from study authors. Antibody responses examined included IgG, IgM and IgA responses

against spike (S), receptor-binding domain (RBD) and nucleocapsid (N) antigens as determined by ELISA or multiplex bead assays. For each serologic measurement, we extracted the time between the date of serologic sample collection and either date of symptom onset or the date of PCR confirmation (preferred if available), which was termed 'time since infection.' For subjects with antibody response measurements at more than one time point, only the last time point was used. Given the lack of a gold standard for specific antibody responses, we do not explicitly account for test performance in antibody detection. We assume that all immunoassays have 100% specificity for detecting recent infection and their decay over time since infections are reflected through decreased sensitivity.

Outcomes and predictor variables

We explored how individual, and combinations of, antibody measurements could identify those who were infected with SARS-CoV-2 and, if infected, their time since the last infection. Using antibody biomarkers measured at different time points post-infection and those collected before the SARS-CoV-2 pandemic, we evaluated the importance and performance of IgG, IgM and IgA antibody isotypes against the nucleocapsid (N), the spike surface protein (S) and RBD antigens in (1) identifying previously infected individuals and (2) their time since infection. We only used binding antibody biomarkers and excluded neutralising antibody results due to the complexity of the assay and variability in methodology.

Model development

We used random forest models to both determine the order of importance of biomarkers and to make our final predictions (1000 random trees, 3 biomarkers per split). Due to the highly correlated nature of the biomarkers, we measured importance using the conditional permutation importance algorithm [8], which measures the importance of each biomarker conditioned on other associated biomarkers in the model. We used this order of importance to train and test various sized sets of antibody response predictors. We fit the models for each study separately, using the cforest and party packages, and measured importance using the permimp package, in RBD [8, 9].

Model comparison

In order to assess the predictive performance of each model, we used repeated 5-fold cross-validation (CV) with 100 iterations, where each iteration contained screening (variable importance) and model-fitting steps. Within the cross-validation, we considered predictor sets of sizes 1, 2 or 3 variables. We developed reduced compact models with a maximum of three predictor antibody responses to make our models more applicable to public health practice. We also investigated the performance of the full model in all datasets with available biomarker variables to understand the upper limit. Using this process, we developed two independent models: (1) a model to identify PCR-confirmed infections using biomarkers and (2) a model to estimate time since infection among those previously infected. For the first model, we used PCR confirmed cases and pre-pandemic controls to train a random forest model and assessed performance using the cross-validated area under the curve (cvAUC) [10]. Within each iteration of cross-validation, we compared model ROC

curves using a permutation test with the function roc.test. We then summarised the *P*-values across iterations to compare model performance [11]. For the second model, we used only confirmed cases to train a random forest model for estimating time-since infection and assessed performance using cross-validated mean absolute error (MAE), the mean of the absolute differences of the predictions from the true time since infection.

Results

We extracted and analysed data on PCR-confirmed SARS-CoV-2 infections from five separate studies, with a total of 834 subjects (Table 1). Of the five studies, four of them used enzyme immunoassays, while three of them used Luminex bead array methods. Except for one study (Peluso *et al.*), all assays were laboratory-developed and not commercial assays. Among a total of 834 subjects, the median time from infection to serologic sampling was 76 (IQR 51–98) days, the median age was 49 (IQR 33–60) years, and the proportion of males was 46.4%. Among the 5 studies, the proportion of patients with severe disease or those who were hospitalised ranged from 7% to 24%.

A single antibody biomarker sufficiently identifies the previous infection

We first explored the classification performance of single antibody and isotype thresholds in identifying infection using four of the five datasets with pre-pandemic control data available. We show that across all studies, a single RBD or spike IgG biomarker performs similarly to the combination of the best two biomarkers in identifying prior infection (Spike/RBD IgG *vs*. Best Two biomarkers, mean *P*-value: 0.19–0.62; the proportion of iterations with *P*-value <0.05: 0 to 0.48; Table 1). The addition of a third biomarker did not increase discriminatory performance (Best Two *vs*. Best Three biomarkers, mean *P*-value 0.16–0.30; the proportion of iterations with *P*-value <0.05: 0.11–0.48) in any of the studies examined and further addition of biomarkers resulted in no additional performance benefit (*P* > 0.05 for all studies).

Two antibody biomarkers are better than one for prediction of time-since-infection

Next, we explored the performance of single *vs*. multiple biomarker thresholds in predicting time-since-infection. In all five datasets, combining two antibody biomarkers performed better than the best single IgG for estimation of time-since-infection (Table 2). We found that in the four datasets where multiple antibody isotypes are measured, the best two antibody biomarkers included a combination of an IgG and an IgM (or IgA in the one dataset where IgM was not measured, Fig. 1). Addition of the third marker results in a marginal (within s.D.) increase in prediction performance in three of the five datasets (Table 2).

Nucleocapsid antibody biomarkers are suboptimal for classification of the previous infection, but adequate for estimating time-since-infection

Given that all vaccines approved for use in the USA/EU at the time of writing induce only spike or RBD antibody responses, we examined the performance of nucleocapsid-only combinations of antibodies. For identification of the previous infection, nucleocapsid IgG performed statistically significantly worse than RBD/

Table 1. Summary of the characteristics of datasets used in this analysis, with cross-validated AUC (95% CI) from classifying the previous infection on four of the published datasets

	Dan et al. [3]	Peluso et al. [6]	Whitcombe et al. [7]	Markmann et al. [5]	Isho et al. [4]
No. of patients	101	122	112	156	343
Age, Median (IQR)	38 (29.5–46.5)	48 (19–85)	48 (31–57)	46.5 (32–60)	36 (21–44)
Sex (male, %)	45	54.9	43.7	44.0	52.1
% patients with severe disease (hospitalised)	7	24.2	13.7	17.7	n/a
Median time since infection (IQR)	62 (47.5–87)	115 (95–124)	109 (94.7–190.5)	57 (49–79)	72 (51–86)
Antibody isotypes measured	IgA, IgG	IgG	IgA, IgG, IgM	IgA, IgG, IgM	IgA, IgG, IgM
Immunoassay platform	ELISA	ELISA/Luminex	Luminex	ELISA/Luminex	ELISA
Cross validated AUC (95% CI)					
RBD IgG	n/a	97.8 (97.2–98.5)	99.7 (99.4–99.9)	98.7 (97.9–99.5)	96.9 (96.6–97.3)
Spike IgG	n/a	98.8 (98.3–99.2)	99.7 (99.5–99.9)	n/a	99.1 (98.8–99.3)
Nucleocapsid IgG	n/a	97.9 (97.4–98.4)	93.9 (93.2–94.5)	98.9 (98.2–99.5)	94.4 (93.9–94.9)
Nucleocapsid IgG vs. best of spike/RBD IgG. Mean (s.d.) *P*-value, proportion of iterations with *P*-value <0.05	n/a	0.27 (0.27), 0.24	0.38 (0.28), 0.11	0.36 (0.30), 0.15	0.40 (0.28), 0.1
Best two biomarkers	n/a	98.8 (98.4–99.2)	99.9 (99.9–99.9)	99.2 (98.6–99.8)	99.4 (99.2–99.5)
Spike/RBD IgG vs. best two biomarkers. Mean (s.d.) *P*-value, proportion of iterations with *P*-value <0.05		0.62 (0.31), 0	0.28 (0.28), 0.22	0.23 (0.29), 0.28	0.19 (0.26), 0.48
Best Three biomarkers	n/a	99.0 (98.7–99.4)	99.9 (99.9–99.9)	n/a	99.2 (99.1–99.4)
Best two vs. best three biomarkers. Mean (s.d.) *P*-value, proportion of iterations with *P*-value <0.05	n/a	0.30 (0.31), 0.11	0.35 (0.29), 0.14	n/a	0.16 (0.25), 0.48
Best two nucleocapsid biomarkers	n/a	n/a	95.0 (94.5–95.5)	n/a	95.9 (95.5–96.2)
Full/saturated model	n/a	99.0 (98.8–99.3)	99.9 (99.9–99.9)	n/a	99.4 (99.3–99.5)

The rows with row-name starting with 'Best' include a screening step in which the biomarkers are ordered by importance for classification (ever-infected) using the random forest conditional permutation algorithm and only the top biomarkers from that iteration are used when training the model.

Table 2. Mean (standard deviation) of MAE from predicting time since infection from repeated cross-validation on five published datasets

	Dan *et al.* [3]	Peluso *et al.* [6]	Whitcombe *et al.* [7]	Markmann *et al.* [5]	Isho *et al.* [4]
Antibody isotypes measured	IgA, IgG	IgG	IgA, IgG, IgM	IgA, IgG, IgM	IgA, IgG, IgM
Best RBD IgG	18.7 (1.7)	29.2 (3.7)	59.8 (5.4)	21.4 (5.7)	17.3 (1.1)
Best Spike IgG	19.6 (1.9)	26.6 (3.6)	61.1 (6.9)	n/a	17.5 (1.3)
Nucleocapsid IgG	18.8 (2.0)	23.9 (4.0)	55.9 (7.4)	21.2 (6.3)	18.3 (1.4)
Best Two biomarkers	17.1 (1.9)	22.6 (4.4)	53.1 (5.1)	20.9 (5.3)	15.7 (1.4)
Best Three biomarkers	17.1 (1.9)	22.4 (4.0)	52.5 (5.3)	n/a	15.3 (1.4)
Best two Nucleocapsid biomarkers	n/a	24.7 (3.9)	51.3 (7.1)	n/a	17.5 (1.1)
Full/saturated model	17.8 (1.9)	22.2 (3.8)	51.5 (6.3)	20.6 (5.9)	15.1 (1.1)

The rows with row-name starting with 'Best' include a screening step in which the biomarkers are ordered by importance for time-since-infection using the random forest conditional permutation algorithm and only the top biomarkers from that iteration are used when training the model (low MAE indicates better performance).

spike IgG in two of four studies examined. In the two studies where data were available, the combination of the two top nucleocapsid markers (IgG plus either IgM or IgA) improved discriminatory performance (Table 1). On the other hand, for predicting time-since-infection (Fig. 1), a combination of the two top nucleocapsid markers performed similar to, or better than, RBD or spike IgG alone (Table 2).

Discussion

The current COVID-19 pandemic is a major public health concern worldwide, and assessment of infection burden in populations is crucial towards efforts to mitigate its spread and inform policy and decision-making. Population-level serosurveillance has emerged to be a useful method to provide accurate estimates of disease burden, as when done under a representative sampling framework, is not subject to biases related to health-seeking behaviour or testing availability. However, there are limited studies to inform the choice and numbers of antibody biomarkers for SARS-CoV-2 serosurveillance. Here, we leverage antibody decay and differing time-varying sensitivity of various assays to build models using serologic data from five studies of individuals with confirmed SARS-CoV-2 infection, to examine which biomarker(s) are best for identifying prior infection and prediction of time-since-infection. Our results show that while Spike/RBD IgG alone are adequate for discrimination/classification of those who have been infected, combinations of antibody markers may be best for estimation of time-since-infection.

An important consideration in the design of serosurveys is the selection of the biomarker(s), with a goal of minimising cost while capturing enough information about infection, transmission or immunity. Population-level serosurveys are able to not only provide estimates for seroprevalence (proportion with circulating anti-SARS-CoV-2 antibodies), they also have the potential to provide data towards estimating the time-since-infection, which could help with accurate incidence estimation and tracking transmission changes on a population level. Our analysis, using MAE as a performance measure, suggests that a combination of antibodies are the best predictors of time-since-infection. For the majority of studies examined, we found using that using three or more biomarkers only performed slightly better than the use of only two biomarkers. In addition, we show that clinical-demographic factors such as age (and less so severity) were potentially important predictors that should be considered in model-building. Further studies are needed to assess the combined performance of both stages of this model by

recreating the epidemic curve through estimation and comparing it to a known epidemic curve.

As COVID-19 vaccines are increasingly made available worldwide, a distinction of vaccine-induced immune responses from that elicited by natural infection is important in the design of seroepidemiologic studies. Identifying infections in vaccinated populations will help estimate the rate of spread. Unfortunately, the most widely used antibody markers for SARS-CoV-2 serosurveillance are the IgG to S or RBD, which is also the target of all currently approved vaccines in the US/EU. Thus, future serosurveillance efforts aimed at assessing infection attack rates may increasingly depend on the nucleocapsid antibody. Our analysis suggests that while nucleocapsid specific IgG alone is inferior to spike and RBD for classification of infection, combinations of N antibodies may improve performance. Notably, for estimation of time-since-infection, the best two N antibodies performed similarly or better than any single S or RBD antibody. Thus, monitoring nucleocapsid specific antibodies may be of utility in distinguishing infection-related antibodies from vaccine-induced antibody response.

A number of research questions and goals remain for SARS-CoV-2 seroepidemiology. First, our conclusions regarding the use of biomarkers for SARS-CoV-2 serosurveillance are based on internal cross-validation of models built using datasets featuring antibody responses for up to 200 days from the time of symptoms onset or diagnosis. Detailed characterisation of the kinetics of serologic responses through longitudinal cohort studies of infected persons of varying severity will enable the development of more tailored and precise statistical models of recent infection. Second, in addition to commercial platforms, standardisation of serosurvey reagents, such as through publicly available monoclonal antibody standards, and/or reference positive sera, will enable a broader application and validation of seroepidemiological analytical models. Third, the development of point-of-care antibody testing will enable serosurveillance to be better performed in more austere environments. The use of dried blood spots from finger pricks in low resources settings could reduce cost while obviating the need for cold-chain storage [7]. Fourth, high-throughput multiplex platforms such as Luminex technology [12] could enable the measurement of numerous SARS-CoV-2 serological markers alongside markers against other infectious pathogens of interest.

There were a number of limitations in this analysis. The lack of longitudinal immune responses and lack of detailed time-since-infection data may have led to larger error predicting time-since-infection. Our analysis was limited to studies of adults
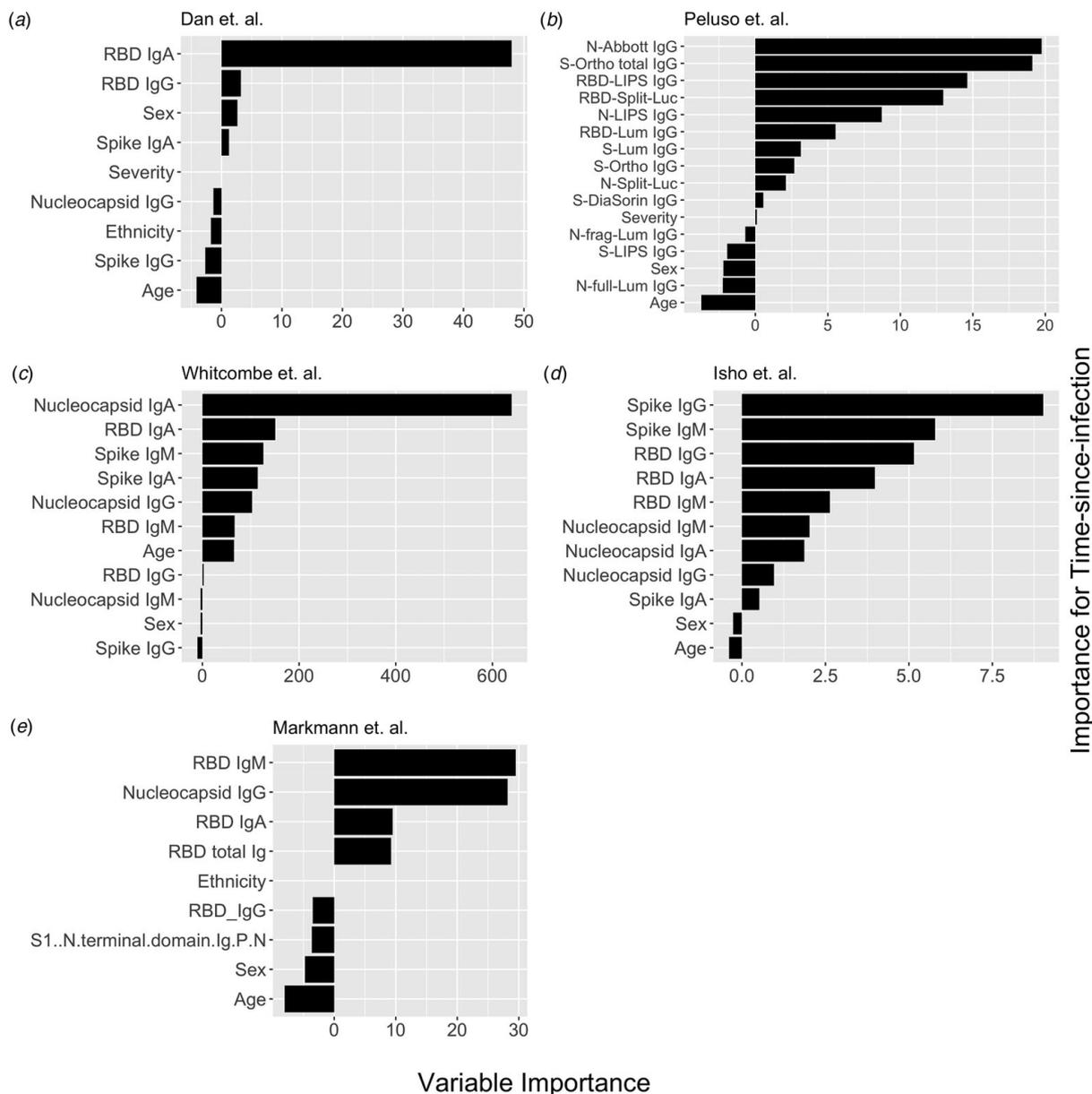
Fig. 1. Conditional permutation variable importance from random forest regression measured by mean decrease in accuracy. Negative importance indicates that the variables inclusion has decreased mean accuracy, probably due to overfitting or random error. Each column represents the order of importance of biomarkers in five datasets. In Peluso *et al.* dataset, S_Ortho_Ig and S_Ortho_IgG indicate total Ig and S IgG by Ortho Clinical Diagnostics VITROS kits; N_abbott indicate Abbot ARCHITECT (IgG); S_DiaSorin is Spike IgG by DiaSorin LIASON(IgG); Neu_Monogram is Monogram PhenoSense (neutralising antibodies); RBD_LIPS, S_LIPS, N_LIPS is IgG by Luciferase Immunoprecipition System (LIPS); RBD_Split_Luc, N_Split_Lum, S_Lum, N.full_Lum, N.frag_Lum indicate IgG to respective antigens by Luminex assay.

in high-income countries, and thus our results cannot be generalised to low- and middle-income countries, or to paediatric populations, and underscore the need for a better understanding of the kinetics of SARS-CoV-2 antibody responses across diverse populations. Despite this, our findings contribute towards informing the choice of antibody responses for seroepidemiological investigations of SARS-CoV-2.

References

1. **Bryant JE** *et al.* (2020) Serology for SARS-CoV-2: apprehensions, opportunities, and the path forward. *Science Immunology* **5**, eabc6347.

2. **Patel R *et al.*** (2020) Report from the American society for microbiology COVID-19 international summit, 23 March 2020: value of diagnostic testing for SARS-CoV-2/COVID-19. *mBio* **11**, e00722-20.

3. **Dan JM *et al.*** (2021) Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science (New York, N.Y.)* **371**, eabf4063.

4. **Isho B *et al.*** (2020) Persistence of serum and saliva antibody responses to SARS-CoV-2 spike antigens in COVID-19 patients. *Science Immunology* **5**, eabe5511.

5. **Markmann AJ *et al.*** (2021) Sex disparities and neutralizing-antibody durability to SARS-CoV-2 infection in convalescent individuals. *mSphere* **6**, e0027521.

6. **Peluso MJ *et al.*** (2021) SARS-CoV-2 antibody magnitude and detectability are driven by disease severity, timing, and assay. *Science Advances* **7**, eabh3409.

7. **Whitcombe AL *et al.*** (2021) Comprehensive analysis of SARS-CoV-2 antibody dynamics in New Zealand. *Clinical Translational Immunology* **10**, e1261.

8. **Strobl C *et al.*** (2008) Conditional variable importance for random forests. *BMC Bioinformatics* **9**, 307.

9. **Strobl C *et al.*** (2007) Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics* **8**, 25.

10. **LeDell EP, Maya L and van der Laan MJ** (2012) Computationally Efficient Confidence Intervals for Cross-validated Area Under the ROC Curve Estimates. *UC Berkeley Division of Biostatistics Working Paper Series*; Working Paper 304.

11. **Venkatraman ES and Begg CB** (1996) A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment. *Biometrika* **83**, 835–848.

12. **Rosado J *et al.*** (2021) Multiplex assays for the identification of serological signatures of SARS-CoV-2 infection: an antibody-based diagnostic and machine learning study. *Lancet Microbe* **2**, e60–e69.