Robust SARS-COV-2 serological population screens via multi-antigen rules-based approach

One Sentence Summary: **Clinical and Population-level performance of single and multiplex SARS-CoV-2 serological assays.**

Christos Fotis[1*], Nikolaos Meimetis[1*], Nikos Tsolakos[2*], Marianna Politou[3*], Karolina Akinosoglou[4*], Vicky Pliaka[2], Angeliki Minia[2], Evangelos Terpos[3], Ioannis P. Trougakos[5], Andreas Mentis[6], Markos Marangos[4], George Panayiotakopoulos[7,8], Meletios A. Dimopoulos[3], Charalampos Gogos[4], Alexandros Spyridonidis[9#$], Leonidas G. Alexopoulos[1,2#$]

1. Biomedical Systems Laboratory, National Technical University of Athens, Athens, Greece
2. ProtATonce Ltd, Demokritos Science Park, Athens, Greece
3. Department of Clinical Therapeutics, Alexandra General Hospital, National and Kapodistrian University of Athens, Athens, Greece
4. Department of Internal Medicine, Division of Infectious Diseases, University Hospital of Patras, Patras, Greece
5. Department of Cell Biology and Biophysics, Faculty of Biology, National and Kapodistrian University of Athens, Athens, Greece
6. Medicinal Microbiology Laboratory, Hellenic Pasteur Institute, Athens, Greece
7. Pharmacology Laboratory, University of Patras, Patras, Greece
8. National Public Health Organization, Athens, Greece
9. Department of Internal Medicine, BMT Unit and CBMDP Donor Center, University of Patras, Patras, Greece

*co-first authors

#co-last authors

$ Correspondence: E-mails: leo@mail.ntua.gr (LGA) and spyridonidis@upatras.gr (AS)

ABSTRACT

More than 300 SARS-COV-2 serological tests have recently been developed using either the nucleocapsid phosphoprotein (N), the spike glycoprotein subunit (S1), and more recently the receptor binding domain (RBD). Most of the assays report very good clinical performance characteristics in well-controlled clinical settings. However, there is a growing belief that good performance characteristics that are obtained during clinical performance trials might not be sufficient to deliver good diagnostic results in population-wide screens that are usually characterized with low seroprevalence. In this paper, we developed a serological assay against N, S1 and RBD using a bead-based multiplex platform and a rules-based computational approach to assess the performance of single and multi-antigen readouts in well-defined clinical samples and in a population-wide serosurvey from blood donors. Even though assays based on single antigen readouts performed similarly well in the clinical samples, there was a striking difference between the antigens on the population-wide screen. Asymptomatic individuals with low antibody titers and sub-optimal assay specificity might contribute to the large discrepancies in population studies with low seroprevalence. A multi-antigen assay requiring partial agreement between RBD, N and S1 readouts exhibited enhanced specificity, less dependency on assay cut-off values and an overall more robust performance in both sample settings. Our data suggest that assays based on multiple antigen readouts combined with a rules-based computational consensus can provide a more robust platform for routine antibody screening.

INTRODUCTION

There is an urgent need for reliable and highly accurate SARS-CoV-2 serological tests that can be used for the diagnosis of recent or prior infection and to screen for possible immunity in population-wide seroprevalence studies (*1, 2*). More than 300 new SARS-CoV-2 serological tests are currently in development (updated at https://www.finddx.org/covid-19/pipeline). Most of these assays report good sensitivity and specificity with samples usually obtained from PCR-positive hospitalized patients (thereafter referred to as clinical cases) and matched with negative blood samples usually obtained before the COVID-19 era (*3*). However, there is a growing awareness that assays with seemingly good clinical performance characteristics might not lead to reliable diagnostic outcomes at low seroprevalence population screening where asymptomatic carriers with low antibody titers are overrepresented (*2, 4*).

Serological tests detect antibodies against SARS-CoV-2 antigens and differ in terms of (i) the targeted antigen, i.e. the nucleocapsid phosphoprotein (N), the S1 subunit of spike glycoprotein (S) or the receptor binding domain of S (RBD), (ii) the class of immunoglobulin detected (e.g. IgG, IgM, IgA, or total), (iii) the detection principle (e.g. enzyme-linked immunosorbent assay, fluorescence, colloidal gold, lateral flow, etc.) and (iv) the assay readout (i.e. quantitative, semi-quantitative or qualitative) (*3*). Assessing antibody presence with a single-readout assay is typically performed by selecting a cut-off value above which the antibody is considered present. Typically, serological data are assumed to follow a normal or half-normal distribution and a cut-off value that is at least 3 standard deviations (SD) above the negative mean distribution is considered as a valid threshold point for assessing positivity (*5*). Setting higher cut-offs could increase the specificity of the test, however at cost of its sensitivity. In many cases, overlapping distributions originating from cross-reactive negatives (*6*) and low-titer positive samples makes the cut-off zone more uncertain. Multi-antigen readouts and employment of "AND" and "OR" logic gates between the different antigens could be used to enhance the diagnostic performance, accuracy, and robustness of the test, especially in borderline

cases. Indeed, a small number of multi-antigen SARS-CoV-2 serological assays have been developed reporting a better overall assay performance in comparison to single-antigen assays (*7-10*). To our knowledge, how this enhanced performance influence results in low-prevalence settings has not been evaluated yet.

Here, we report the development of a bead-based multiplex serological assay for the simultaneous detection of antibody responses against N, S1 and RBD SARS-CoV-2 antigens. The developed multiplex platform combined with a rules-designed approach was used to assess the performance of single- or multi-antigen(s) readouts in well-defined clinical cases, as well as in a population-level study. The rules-based multiplex SARS-CoV-2 serological assay exhibited enhanced specificity, less dependency on cut-off values and an overall more robust performance in both sample settings.

RESULTS

Antibody responses to SARS-CoV-2 and endemic human coronaviruses in clinical samples

We developed a multiplex serological assay utilizing the Luminex® xMAP™ technology to simultaneously detect antibody responses to three SARS-CoV-2 antigens (N, S1 and RBD) and one antigen from each one of the four endemic human coronaviruses HCoV-OC43 (S1+S2), HCoV-HKU1 (S1), HCoV-229E (S1) and HCoV-NL63 (S1). This setup was applied for the detection of IgG, IgM and IgA isotypes as well as for total antibodies (IgG/IgM/IgA). The analytical performance of the assay together with the parameters optimized during assay development can be found in Supplementary Material.

A total of 155 serum samples from 77 PCR-confirmed COVID-19 cases and 78 pre-epidemic individuals were screened for the existence of reactive antibodies against antigens of the five different coronaviruses (Table 1). SARS-CoV-2 infection was deemed asymptomatic and mild needing no hospitalization in 8% and 60% of the cases, respectively, whereas 32% of the participants required hospitalization. Antibody

detection was performed at a median of 46 days (range 13-87) post SARS-CoV-2 infection. Figure 1 shows the strong reactivity of sera to each SARS-CoV-2 antigen for the different antibody isotypes. Antibody levels against the SARS-CoV-2 antigens N, S1 and RBD were significantly higher in the SARS-CoV-2 infected population as compared to samples from non-infected individuals ($p<0.001$, Supplementary Table S3); this finding was true for all isotypes. The analyzed blood samples from SARS-CoV-2 tested positive donors also showed significantly high antibody titers against the HCoV-OC43 S1/S2 antigens in all isotypes and against the HCoV-HKU1 S1 antigen in IgG detection, possibly as a result of antibodies cross-reactivity against S proteins that are conserved between SARS-CoV-2, HCoV-HKU1 and HCoV-OC43 coronaviruses.

We further assessed the validity of our results for the different SARS-CoV-2 antigens by looking at the correlation of the normalized MFI values between antigens in both positive and negative cases (Table 2). The SARS-CoV-2 S1 and RBD antigens-related data correlated strongly across all Ig isotypes in COVID-19 positive samples with Pearson's correlation coefficient (r) values > 0.96; a high correlation was also observed between N and S1 or RBD readouts in IgG and total antibody assays (r values between 0.72-0.82) in COVID-19 positive samples. Readouts between the N and S1 or RBD antigens correlated poorly in IgA and IgM detection. Results from negative samples showed no correlation across all antigen comparisons and Ig isotypes with the notable exception of the S1 and RBD antigens in IgM isotype ($r=0.78$).

Diagnostic Performance of SARS-CoV-2 serological assays using single and multiple antigen readouts in clinical samples.

The sensitivity and specificity were calculated for the three single antigen readouts and for 11 multi-antigen rules. Calculations were first performed using a cut-off value of mean plus 3 standard deviation (SD) cut-off for each antigen by assuming normal distribution (Table 3). When assessed individually, antigens were equally specific producing one or two false positives out of the 78 negative samples tested (97.4-98.7% specificity). RBD

was the most sensitive among the three antigens in detecting total, IgG and IgA antibodies (98.7%, 97.4% and 93.5%, respectively), while S1 was the most sensitive in the detection of the IgM isotype (83.1%). Assays detecting total or IgG antibodies were more sensitive as compared to those detecting IgA or IgM regardless of the antigen (94.8-98.7% compared to 22.1-93.5%). In the IgA detection assay, the S1 and N antigens produced many false negative predictions (48/78 and 37/78, respectively), while in the IgM assay a very high number of false negative results were observed only for the N antigen (60/78). Similar results were obtained when different cut-off values based on ROC analysis were used (Table S4, Figure S3 in Supplementary Material).

Rules-based approaches requiring at least two of the antigens to be above cut-off (i.e. rules using an AND gate between two antigens) for reporting a positive result improved assay specificity in all isotypes *vs*. individual antigens (Table 3). These AND-based rules also showed comparable sensitivity to individual antigens in total, IgG and IgM antibody detection. Rules did not improve assay performance parameters for IgA detection where RBD alone was the best performing antigen. Rules requiring at least one/any of the antigens (OR-based rules) to be above the cut-off for scoring a positive result did not improve assay performance when compared to the best performing individual antigen for the respective antibody isotype. Overall, rules that utilized all three antigens and set a positive result when at least two of them were above the cut-off (Antigen A AND [B OR C] reported as A&B|C) showed consistent performance and were further investigated.

To assess the effect of the different rules on assay performance as compared to their individual counterparts, we investigated their performance profiles across an extended range of cut-off values. Thus, sensitivity, specificity and accuracy were calculated for gradually increasing threshold values, based on the negative sample distribution of each antigen and antibody isotype. Rules using all antigens and requiring two or all three to be higher than the threshold were included in the analysis (Figure 2). Across all isotypes, all rules exhibited a more robust profile and were less affected by changes in the cut-off

thresholds vs. individual antigens. Rules provided a clear benefit in assay specificity with specificities of 100% achieved at much lower cut-offs as compared to individual antigens. Additionally, for total and IgG antibody detection, assay sensitivity was retained at high levels across a wide range of cut-offs resulting in an overall more robust and accurate assay. The RBD&N|S1 rule was shown to outperform all other rules for detection of total and IgG antibodies. In terms of the IgA and IgM assays, although all rules resulted in improved specificity profiles, their sensitivity was mostly driven by the sensitivity of the antigen that formed the basis of the rule and was not improved by the inclusion of the results from the other antigens.

Assay validation against commercial antibody tests

The performance of the assays (single and multi-antigen based) was validated using commercially available antibody tests developed by Abbott and Euroimmun AG against the N and S1 antigens, respectively. As shown in Figure 3A, both N and S1 antigen readouts of the multiplex assay were highly correlated with their commercial counterparts (Pearson's r=0.98 for N and 0.9 for S1). Their diagnostic agreement (positive/negative call) was 100% when compared to Euroimmun S1 assay (60 out of 60 predictions) and Abbott N assay (31 out of 31 predictions). A strong agreement was also observed between the commercial assays and the RBD&N|S1 multi-antigen rule (Figure 3B). Specifically, the rule agreed in 59 out of the 60 samples tested with the Euroimmun S1 assay and in 29 out of the 31 samples tested with the Abbott N assay. The three samples in which the assays disagreed were called negative by the commercial assays but showed positive readouts in the other two antigens measured in our multiplex assay and were thus called positive by the multi-antigen rule.

Population level screening

An important application of SARS-CoV-2 serological assays is the identification of seroconverted individuals at the population level. Since asymptomatic infection may induce low SARS-CoV-2 related antibodies titers which may also decline over time, a

major requirement for this type of analysis is a highly sensitive assay that does not compromise specificity. Therefore, we used the total (IgG/IgA/IgM) SARS-CoV-2 antibody multiplex assay to assess seroprevalence in 1,225 asymptomatic blood donors with no known history for SARS-CoV-2 exposure. Seroprevalence was calculated based on the single antigen readouts and the multi-antigen RBD&N|S1 rule and was found to be strongly influenced both by the antigen analyzed and the cut-off value used to determine the diagnostic outcome (Figure 4). SARS-CoV-2 seroprevalence from single antigen readouts ranged between 0.8% (N, mean plus 5 SD cut-off) and 7.5% (S1, mean plus 3 SD cut-off), indicating a wide range of potentially indeterminate cases. When using the RBD&N|S1 rule, seroprevalence ranged between 0.6% (mean plus 5 SD cut-off for each antigen) and 1.2% (mean plus 3 SD cut-off), in line with the robust performance of the multi-antigen assay in clinical samples. We examined the overlap of positive individuals being diagnosed by the single antigens or the RBD&N|S1 rule using the stringent cut-off of mean plus 5 SD (Figure 5). We observed a strikingly low agreement between antigens, in that the different antigens resulted in vastly different subsets of positive individuals. Specifically, N shared 2 positive samples with RBD (5.9% agreement) and 1 with S1 (1.8% agreement) while S1 and RBD shared 6 positive samples (9.1% agreement). The RBD&N|S1 rule had a total of 7 positive calls, 5 of which were samples with S1 and RBD positive readouts, 1 sample with RBD and N positive readouts and 1 sample with all three antigens above the cut-off. We re-analyzed all 1,225 samples with an independent commercially available test which detects IgG N-specific antibodies (Abbott). From the 6 samples that were counted positive (estimated seroprevalence 0.5%) only 2 also scored positive with the multiplex RBD&N|S1 readout. An analysis of the seroprevalence, agreement rates and overlap of positive samples between the other rules of the multiplex assay (N&RBD|S1, S1&RBD|N) and the results of commercial Abbott test (N-specific IgG) are shown in Supplemental Figures S4,S5,S6.

DISCUSSION

Serological tests play an instrumental role during the COVID-19 pandemic for quantifying seroconversion, seroprevalence, vaccination status and for the diagnosis of recent or prior SARS-COV2 infections. More than 300 assays have been developed using either N, S1 or RBD antigens and many of them have already reached the market in unforeseen development speeds (*3*). However, it is questionable whether single-antigen assays with strong performance characteristics on clinical samples show sufficient diagnostic performance at population screening studies (*4*). Here, we developed and compared single and multi-antigen-based assays in the two different settings, the clinical level which determines clinical assay performance and the population level which presents a wider range of responses and likely low seroprevalence.

The bead-based multiplex platform developed here enabled the simultaneous detection of diverse antibody responses against N, S1 and RBD within a single diagnostic run. In line with other clinical performance reports, we found a high variability of IgM and IgA responses to different antigens in 77 previously infected SARS-CoV-2 individuals (*11*). In contrast, nearly all individuals produced IgG specific antibodies against all three antigens, with the RBD being the most sensitive SARS-CoV-2 epitope. Notably, antibodies against N, S1 and RBD antigens were also detected in pre-COVID19 negative samples potentially due to cross-reactivity to other common coronaviruses, thus making no single readout of our assay as 100% specific when a mean plus 3 SD cut-off was used(*6*). Other studies have also evaluated N, S1, and RBD antigens in single or multiplex formats and reported similar or even higher sensitivities and specificities for total or IgG against RBD vs N and/or S1 in samples tested beyond two to three weeks post infection (*7, 12-14*). The validation of our single assays against other commercial single-antigen tests (Abbott - N and Euroimmun - S1) show 100% agreement with our N and S1 multiplex readouts (Figure 3A). However, diagnostic outcomes are slightly different amongst single and multi-antigen assays and previously infected SARS-CoV-2 individuals may be missed by single antigen-based tests (Figure 3B). In support, the absence of antibodies against certain SARS-CoV-2 antigens is

also apparent in studies that have compared assays targeting different antigens (*7, 14-16*); yet, the biological basis of these observations is currently not well understood.

The computational framework implemented in this study aimed to provide an intuitive approach to combine multiple antigen readouts and possibly further enhance the specificity, accuracy and robustness of multi-antigen approaches. With the rules-based method, SARS-CoV-2 seroconversion is given using "AND" and "OR" logic between the single antigen readouts. The use of the "OR" rule between antigens (i.e. positive S1 OR positive RBD OR positive N) increases the sensitivity at the expense of specificity whereas the use of the "AND" rule (i.e. positive S1 AND positive RBD AND positive N) increases the specificity at the expense of sensitivity. Such a multi-antigen and rules-based approach has been shown to enhance the diagnostic performance of serological assays for common SARS-CoV (*17*) and for SARS-CoV-2 (*10*). We found that an optimally selected combination of AND plus OR rules could increase specificity and the overall accuracy with little or no cost on sensitivity for IgG and total antibody detection in the clinical samples. All potential combinations were tested, and the optimal combination was built by the best performing readout (RBD antigen) whose specificity was further enhanced by requiring consensus (AND) with either S1, N, or both readouts (Table 3). Most importantly, the rules-based approach improved assay robustness (i.e. how small deviations of the cut off value affect the diagnostic outcome). The "RBD AND [N OR S1]" (reported as RBD&N|S1) rule resulted in an assay with consistent accuracy across a wider range of cut-off values compared to single-antigen assays suggesting that this multi-antigen combinatorial assay can achieve optimal performance at lower cut-off. It is worth noting out that multi-antigen rules, while improved IgG and total antibody detection, did not improve performance of IgA and IgM assays.

The use and application of our multiplex strategy in over a thousand samples from randomly selected blood donors aimed at providing a dataset of experimental settings that mimic routine low-prevalence screens. To our knowledge, this is the first time that a

multi-antigen serological assay was applied at population-level analysis. In contrast to clinical samples where almost all positive donors showed high antibody titers for all three antigens, a substantial number of screened donors revealed a variability of responses to the N, S1 and RBD antigens and just one of out of the 1,225 donors showed high antibody titers for all three antigens. With such high dissensus on single-antigen responses, it is obvious that diagnosis of seroconversion in the community becomes challenging. Even the highly concordant S1 and RBD antigens exhibited only 9% agreement in their positive calls in the population screen (6 out of 66 positive calls). One explanation for this antigen disagreement is that SARS-CoV-2 infected persons in population-wide screens include mostly asymptomatic individuals with likely lower antibody titers and unknown time since infection, during which antibodies to specific epitopes may have already waned below detection limit (*18-20*). Another important reason for the large disagreement between antigens is the inadequate single-antigen specificity at low seroprevalence settings that can lead to a very large number of false predictions (*4*). For example, at 1% seroprevalence and 100% sensitivity, the maximum 97.4% specificity we observed from the single antigen readouts in the clinical performance study would result in ~31 false positive predictions in our serosurvey whereas just 12 samples are expected to be truly positive (*21*). In line with that false positive predictions, the estimated 7.5 seroprevalence that was observed in figure 4 at mean plus 3 SD with S1 may be mostly false positive predictions. The striking differences in the seroconversion calls between the different SARS-CoV-2 antigens were reflected not only in the wide range of estimated seroprevalence rates (0.6%-7.5%) but more importantly in the vastly different subsets of potentially SARS-CoV-2 positive individuals. Consistently to our findings, a side-by-side comparison of three fully automated SARS-CoV-2 antibody assays (Abbott against N, Roche against N, and DiaSorin against S1/S2 antigens) showed good agreement in 65 samples from COVID-19 patients, but had profound discrepancies of positive predictability at 1% seroprevalence (*22*). Likewise, in an epidemiological study in Iceland including 18,609 individuals tested for total anti-N and anti-S1-RBD antibodies using independent commercial kits, 158 were found positive for either N (n=56) or S1-RBD

(n=102) but only 39 of them had antibodies for both epitopes (*23*). Similar discrepancies between N- and S1-specific serological assays were observed in a big epidemiological study in Spain (*24*).

Our results have profound implications for the seroprevalence rates presented by serosurveys based on single antigen assays, as well as for the diagnostic value of such screens. To achieve the most accurate seroprevalence rates at low seroprevalence setting, an ideal 100% specificity assay is required. One approach to increase specificity would employ raising the cut-off value for the assay used, an approach that IVD manufacturers prefer to adopt to be on the safe side of a true positive prediction. However, raising the cut-off value can have a profound underestimation of the seroprevalence rates using single-antigen approaches (as shown in Figure 4). A better approach to enhance serosurveillance accuracy would be confirmatory testing with an independent assay that uses a different antigenic target for a positive outcome (*25*). In this paper, we showed that the higher specificity of the multi-antigen assay can be achieved by consensus rather by increased cut-off values. On this front, our rules-method achieved an almost consistent seroprevalence rate for a large range of cut-off values between 3 to 5 standard deviations above the mean (Figure 4). Consequently, we believe that the RBD&N|S1 rule, which revealed the best performance characteristics in the clinical study, combined with a mean plus 5 SD threshold provided a more realistic estimation of the seroprevalence figures in the community screen and more accurate identification of seroconverted individuals. Notably, in our study we used cut-off thresholds between 3 to 5 SD's whereas using the same distribution assumptions the manufacturer's recommended 1.4 cut-off value of the IgG N-specific Abbot test was calculated to correspond to more than 10 SD above the negative mean. Such high cut-off value can undermine sensitivity and thus, not surprisingly, when using the rules-based multiplex assay (total IgG/IgA/IgM for RBD&N|S1) as reference in the community screen the positive predictive value of the IgG N-specific Abbot test was only 29% (2 out of 7 detected), though both assays showed excellent agreement in detecting seroconversion

in the clinical performance study. An important limitation on such comparisons between serological assays in population-wide surveys is the fact that there is no gold standard method to identify whether positively called individuals were truly asymptomatic SARS-CoV-2 positive cases.

In conclusion, our study has demonstrated that serological assays based on single antigens, while good at diagnosing infected individuals in a clinical setting, may not be ideal in low seroprevalence, population-wide COVID19 screens where low antibody responses from mostly asymptomatic individuals are expected. A multi-antigen approach combined with a rules-based computational framework for diagnostic decisions can provide a better alternative in such contexts through its enhanced specificity and reduced dependency on cut-off thresholds. We believe that such multi-antigen approach should be performed in a single multiplex assay, thus diminishing possible differences attributed to operational issues of independent assay formats (*11*). An added advantage of multiplexing is the reduced usage of resources and time. The embrace by the scientific community of multiplex assays or multiple single antigen-based assays for serological analysis can eventually lead to more accurate and reliable results regarding SARS-CoV-2 spread in the general population.

MATERIALS AND METHODS

Serum Samples

All samples were acquired under approved clinical protocols and informed consent (see Ethics section). The list of serum samples used throughout the study is presented in Table 1. A total of 155 clinical serum samples were analyzed, of which78 were negative as were banked sera collected prior to the COVID-19 pandemic (2018-2019) and 77 were positive samples from PCR-confirmed SARS-CoV-2 individuals collected at least two weeks after SARS-CoV-2 infection. Nineteen additional serum samples from PCR-confirmed SARS-CoV-2 individuals were used for assay validation against other commercial SARS-CoV-2

serological assays (see Supplementary Material). For the general population screening, we obtained 1,225 serum samples between the 23rd and 25th week of 2020, e.g. 5 weeks after lockdown in Greece that took place between the 13th and 18th week 2020; blood donors were from 13 different geographical regions in Greece. Eligibility for donation included an extended detailed questionnaire for previous possible signs of infection during 2020. All samples were acquired under approved clinical protocols and informed consent (see Ethics section).

Multiplex Immunoassay Development

A magnetic bead-based immunoassay was developed using the xMAP Luminex technology against SARS-CoV-2 antigens N, S1 and RBD. One antigen from each one of the four endemic coronaviruses was also included in the assay. Specifically, the S1 subunit of HCoV-HKU1, HCoV-229E and HCoV-NL63 and the S1+S2 subunits from HCoV-OC43 were used. The SARS-CoV-2 N and S1 antigens were purchased from the Native Antigen Company (Kidlington, UK). All other antigens were from Sino Biological Europe GmbH (Eschborn, Germany). Each antigen was covalently coupled to a distinct magnetic bead region (Luminex Corp, Austin, Texas) by carbodiimide coupling at a ratio of 15 µg per 5 million beads (*17*). Coupling efficiency was confirmed by incubation of 5,000 beads from each coupled region with a phycoerythrin-conjugated anti-6x HisTag antibody (Abcam, Cambridge, UK) at a concentration of 32 µg/mL for 15 min at room temperature. Coupled beads were mixed to a final concentration of 50 beads/µL and stored in PBS supplemented with 1% bovine serum albumin, 0.02% Tween-20 and 0.05% sodium azide at 4°C until use. For analysis of serum samples, 25 µL of the bead mix (corresponding to 1,250 beads per antigen) were added to each well of a 96-well plate, washed twice with 100 µL Assay Buffer (PBS supplemented with 1% BSA and 0.05% sodium azide) and incubated with 50 µL of serum diluted in LowCross-Buffer® (CANDOR Bioscience GmbH, Wangen, Germany) for 2 hrs at room temperature in a plate shaker (900 rpm). A serum dilution of 1/400 was used for testing all immunoglobulin types except for IgA that was assayed at a 1:100 serum dilution was used. Unbound material was removed by two

washes with 100 µL assay buffer and beads were incubated with 20 µL of biotinylated anti-human immunoglobulin antibodies (Jackson ImmunoResearch Europe Ltd, Ely, UK) for 1 hr at room temperature in a plate shaker (900 rpm). Antibodies were diluted in assay buffer at 1:1,600 for IgG/IgM/IgA, 1:800 for IgG, 1:3,200 for IgA and 1:800 for IgM. Beads were washed twice with 100 µL assay buffer and incubated with streptavidin R-phycoerythrin (Jackson ImmunoResearch Europe Ltd, Ely, UK) diluted 1:100 in assay buffer for 15 min at room temperature in a plate shaker (900 rpm). Beads were washed again as before, reconstituted in 130 µL assay buffer, and measured in a FLEXMAP 3D instrument (Luminex Corp, Austin, Texas). Instrument settings included standard PMT, 100 µL sample volume, a bead count of 50 beads per antigen and doublet discrimination gate set at 3,000-20,000.

Diagnostic Assay Performance Analysis

The Median Fluorescent Intensity (MFI) values of each SARS-CoV-2 antigen were first divided by the average MFI of the negative control samples (made from a pool of negative sera) for the same antigen. These "normalized MFI" values were used in all subsequent calculations. Cut-off values for determining the diagnostic outcome (positive/negative) regarding the presence (or not) of SARS-CoV-2 specific antibodies were calculated for each SARS-CoV-2 antigen based on its on its distribution at the negative samples. The diagnostic performance of the assay was assessed for cut-off values ranging from mean *plus* one standard deviation (SD) up to mean plus five SD. Performance was evaluated in terms of sensitivity, specificity and accuracy, while the corresponding 95% confidence intervals (CI) were calculated using the Wilson approximation (*26*). Furthermore, for each antibody isotype and antigen, the Receiver Operating Characteristic (ROC) curves and the corresponding area under the curve (AUC) was calculated (see Supplementary Material).

Rules-based Method

To assess assay performance based on the multi-antigen readout, a rules-based method was developed. First, the single antigen "normalized MFI" values were transformed to

positive/negative predictions by comparing them with the appropriate cut-off value. Then, logic circuits that utilize the “AND” (represented by the symbol “&”) and “OR” (represented by the symbol “|”) logic gates were implemented. These logic circuits take as input the single-antigen predictions and output the final rules-based prediction that corresponds to a positive/negative call for the presence of SARS-CoV-2 specific antibodies. All possible simple circuits that could be formed using the “AND” and “OR” logic gates to combine the predictions of the RBD, S1 and N predictions were examined.

Assay Validation

Assay validation was performed in two separate subsets of matched clinical samples against two widely used, commercially available SARS-CoV-2 antibody tests developed by Euroimmun (Euroimmun Medizinische Labordiagnostika AG, Lubeck, Germany) and Abbott (Abbott Diagnostics, Illinois, USA) which detect IgG antibodies against S1 and N, respectively (Supplementary Table S5). Single antigen readouts for S1 and N from our IgG multiplex assay were plotted against the results of the commercial tests and the Pearson’s correlation coefficient values calculated using GraphPad Prism (version 8.4.2). For determining the diagnostic outcome (positive/negative calls) we used the mean plus three SD cut-off values for our assays and the manufacturer’s recommended cut-offs for the commercial assays (1.1 for Euroimmun and 1.4 for Abbott). Commercially available RBD-based IVD assays provided in lateral flow formats were incompatible with our sample collection procedure and could not be used for validation of RBD readouts.

Population-level Analysis

For the analysis of the 1,225 samples from blood donors, the total (IgG/IgA/IgM) assay was used, and the diagnostic outcome was assessed across cut-off values ranging from mean plus 3 SD to mean plus 5 SD from the negative sample distribution. For comparisons between single antigen readouts, the diagnostic outcomes from the mean plus 5 SD cut off values were used. Frozen back-up samples (n=1,225) were sent to the Immunology Laboratory of the National Public Health Organization, Athens, Greece and analyzed for

the presence of IgG antibodies against the SARS-CoV-2 N antigen using the ABBOTT SARS-CoV-2 IgG assay with the ARCHITECT i2000SR analyzer (Abbott, Illinois, United States). The manufacturer's recommended cut-offs (1.4) was used to determine positivity.

SUPPLEMENTARY MATERIALS

Supplementary methods and results: 1) Optimization of Assay Parameters. 2) Analytical Assay Performance. 3) Correlation of SARS-CoV-2 antibody responses to antibody levels from other infectious agents. 4) Comparison of positive and negative distributions across corona virus antigen. 5) ROC analysis. 6) Assay Validation. 7) Seroprevalence and agreement rates of single and multi-antigen rules in population screening.

Figure S1. Serum Dilution Optimization.

Figure S2. Secondary Antibody Dilution Optimization.

Table S1. Intra- and Inter-assay variability.

Table S2. Correlation of SARS-CoV-2 antibody responses to antibody levels from other infectious agents.

Table S3. Antibody responses to coronavirus antigens in SARS-CoV-2 positive and negative cases.

Table S4. Diagnostic Accuracy and AUC of SARS-CoV-2 individual antigens

Figure S3. ROC curves for the 3 antigens and for the different antibody isotypes.

Table S5. Commercial SARS-CoV-2 serological assays and samples used for assay validation.

Figure S4. Percentage of asymptomatic individuals (n=1,255) positive for total (IgG/IgA/IgM) SARS-CoV-2 antibodies using single antigen readouts and multi-antigen rules.

Figure S5. Overlap of positive samples across antigens and combinations.

Figure S6. Overlap of positive samples between methods.

REFERENCES

1. M. Lipsitch, D. L. Swerdlow, L. Finelli, Defining the Epidemiology of Covid-19 - Studies Needed. *The New England journal of medicine* **382**, 1194-1196file:///C:/Users/User/Downloads/nejmp2002125.pdf (2020).
2. M. C. Weinstein, K. A. Freedberg, E. P. Hyle, A. D. Paltiel, Waiting for Certainty on Covid-19 Antibody Tests — At What Cost? *New England Journal of Medicine* **383**, e37 (2020).
3. J. J. Deeks *et al.*, Antibody tests for identification of current and past infection with SARS-CoV-2. *The Cochrane database of systematic reviews* **6**, Cd013652 (2020).
4. J. E. Bryant *et al.*, Serology for SARS-CoV-2: Apprehensions, opportunities, and the path forward. *Science immunology* **5**, (2020).
5. S. GREENLAND, Basic Methods for Sensitivity Analysis of Biases. *International Journal of Epidemiology* **25**, 1107-1116 (1996).
6. H. Lv *et al.*, Cross-reactive Antibody Response between SARS-CoV-2 and SARS-CoV Infections. *Cell reports* **31**, 107725 (2020).
7. P. R. Randad *et al.*, COVID-19 serology at population scale: SARS-CoV-2-specific antibody responses in saliva. *medRxiv : the preprint server for health sciences*, (2020).
8. J. Rosado *et al.*, Serological signatures of SARS-CoV-2 infection: Implications for antibody-based diagnostics. *medRxiv : the preprint server for health sciences*, 2020.2005.2007.20093963 (2020).
9. G. den Hartog *et al.*, SARS-CoV-2-specific antibody detection for sero-epidemiology: a multiplex analysis approach accounting for accurate seroprevalence. *The Journal of infectious diseases*, (2020).
10. C. Dobaño *et al.*, Highly sensitive and specific multiplex antibody assays to quantify immunoglobulins M, A and G against SARS-CoV-2 antigens. *bioRxiv : the preprint server for biology*, 2020.2006.2011.147363 (2020).
11. J. D. Whitman *et al.*, Evaluation of SARS-CoV-2 serology assays reveals a range of test performance. *Nat Biotechnol*, (2020).

12. W. Liu *et al.*, Evaluation of Nucleocapsid and Spike Protein-Based Enzyme-Linked Immunosorbent Assays for Detecting Antibodies against SARS-CoV-2. *Journal of clinical microbiology* **58**, (2020).

13. R. Lassaunière *et al.*, Evaluation of nine commercial SARS-CoV-2 immunoassays. *medRxiv : the preprint server for health sciences*, 2020.2004.2009.20056325 (2020).

14. N. M. A. Okba *et al.*, Severe Acute Respiratory Syndrome Coronavirus 2-Specific Antibody Responses in Coronavirus Disease Patients. *Emerging infectious diseases* **26**, 1478-1488 (2020).

15. G. Zhang, S. Nie, Z. Zhang, Z. Zhang, Longitudinal Change of Severe Acute Respiratory Syndrome Coronavirus 2 Antibodies in Patients with Coronavirus Disease 2019. *The Journal of infectious diseases* **222**, 183-188 (2020).

16. L. Grzelak *et al.*, A comparison of four serological assays for detecting anti–SARS-CoV-2 antibodies in human serum samples from different populations. *Science Translational Medicine* **12**, eabc3103 (2020).

17. S. U. Trivedi *et al.*, Development and Evaluation of a Multiplexed Immunoassay for Simultaneous Detection of Serum IgG Antibodies to Six Human Coronaviruses. *Scientific reports* **9**, 1390 (2019).

18. F. Amanat *et al.*, A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nature medicine* **26**, 1033-1036 (2020).

19. Q. X. Long *et al.*, Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nature medicine*, (2020).

20. F. J. Ibarrondo *et al.*, Rapid Decay of Anti-SARS-CoV-2 Antibodies in Persons with Mild Covid-19. *The New England journal of medicine*, (2020).

21. P. Diamandis, I. Prassas, E. P. Diamandis, Antibody tests for COVID-19: drawing attention to the importance of analytical specificity. *Clinical chemistry and laboratory medicine* **58**, 1144-1145 (2020).

22. T. Perkmann *et al.*, Side by side comparison of three fully automated SARS-CoV-2 antibody assays with a focus on specificity. *medRxiv : the preprint server for health sciences*, 2020.2006.2004.20117911 (2020).
23. D. F. Gudbjartsson *et al.*, Humoral Immune Response to SARS-CoV-2 in Iceland. *New England Journal of Medicine*, (2020).
24. M. Pollán *et al.*, Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet (London, England)*, (2020).
25. National Institutes of Health. Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Available at https://www.covid19treatmentguidelines.nih.gov/.
26. E. B. Wilson, Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association* **22**, 209-212 (1927).

ACKNOWLEDGMENTS: We thank the patients and individuals who donated their blood. We greatly acknowledge all healthcare workers who were involved in this study. **FUNDING:** Does not apply. **AUTHOR CONTRIBUTIONS:** Conceptualization: LGA, AS; Cohort management and sample collection: MP, KA, ET, AM, MM, GP, MAD, CG, AS; Methodology and sample measurements: CF, NM, NT, VP, AM, AM; Data Analysis: CF, NM, NT, IPT, AS, LG; Writing: CF, NM, NT, IPT, AS, LG; All authors have read and agreed to the final version of the manuscript. **CONFLICT OF INTEREST STATEMENT:** LGA, NT, AM VP, are members of ProtATonce Ltd. All other authors declare no competing interests related to this article. **ETHICAL STANDARDS:** Sampling from SARS-CoV-2 positive individuals was done with informed consent and under approved institutional and ethic review board approved clinical protocols conducted in full compliance with the principles of Good Clinical Practice and the Declaration of Helsinki (University Hospital of Patras EC 164/27.04.2020 and IRB 216/08.05.2020, Alexandra General Hospital NCT04408209 trial). Stored negative samples were acquired in accordance with local ethical approvals. The population-level study was approved by the Research Ethics Committee of the University of Patras (Ref. Number 6099) and all participants gave written informed consent.

FIGURES

Figure 1.

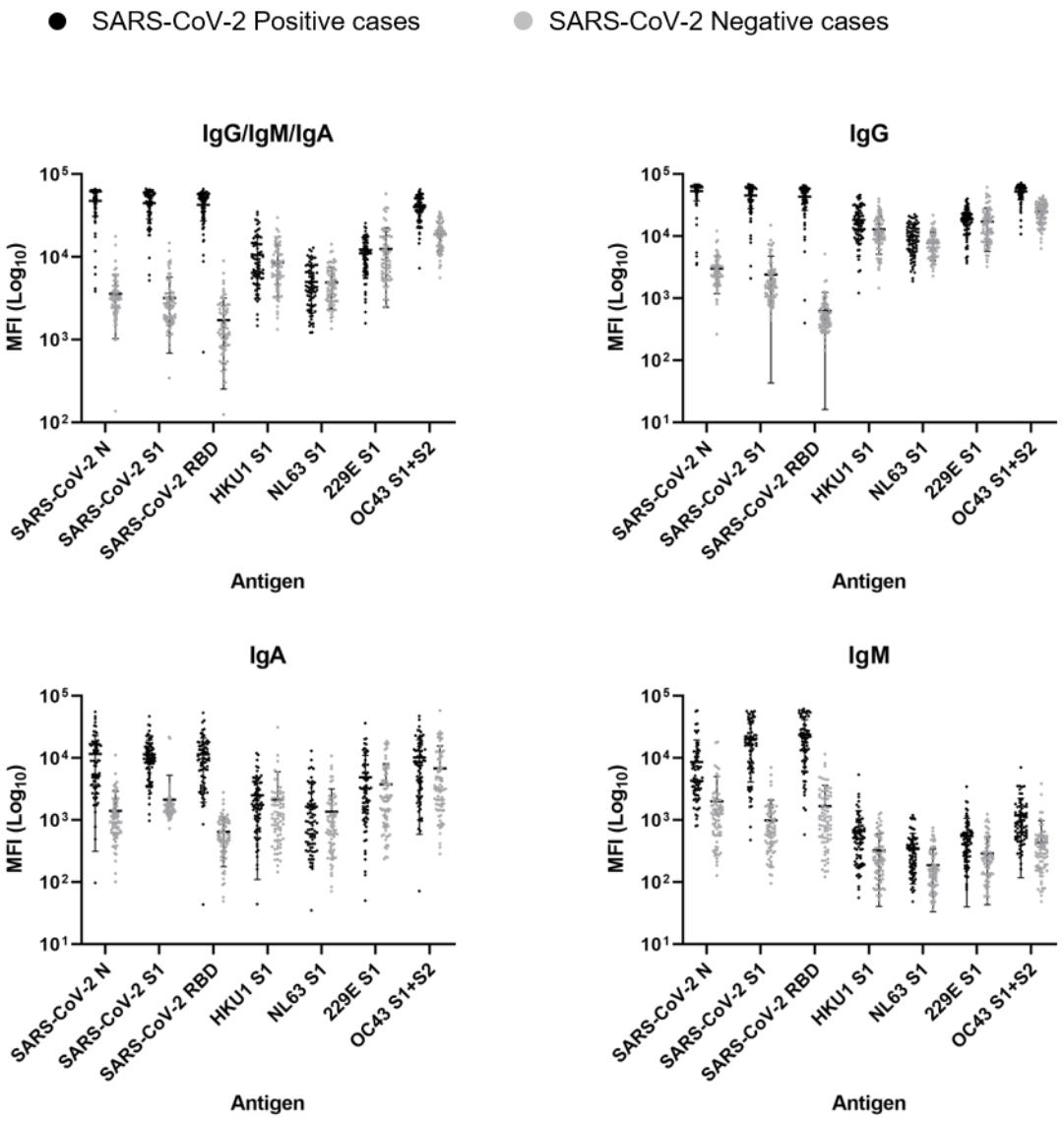

Figure 1. Antibody responses against SARS-CoV-2 N, S1 and RBD antigens and 4 endemic coronaviruses antigens in serum samples from SARS-CoV-2 PCR-tested positive and negative (banked blood samples from 2018-2019) individuals. The Median Fluorescence Intensities (MFI, log10 scale) measured in positive and negative samples were plotted for each antigen. Graphs indicated the different antibody isotypes tested. Mean plus standard deviation (SD) error bars are presented.

Figure 2.

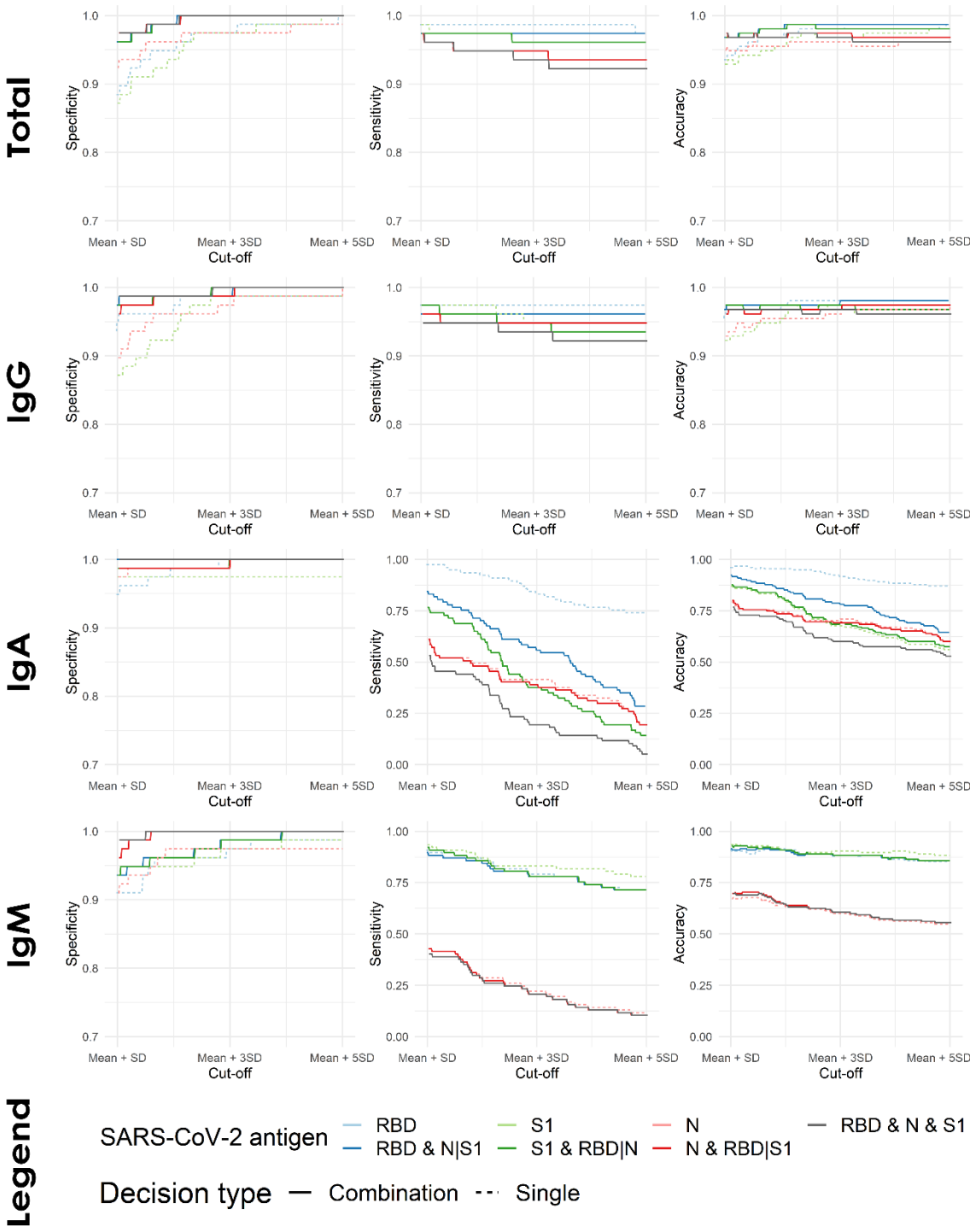

Figure 2. Assay performance parameters for different cut-off thresholds. For each antibody isotype, sensitivity, specificity, and accuracy values were calculated for increasing cut-off thresholds based on the distribution of each antigen in negative samples.

Figure 3.

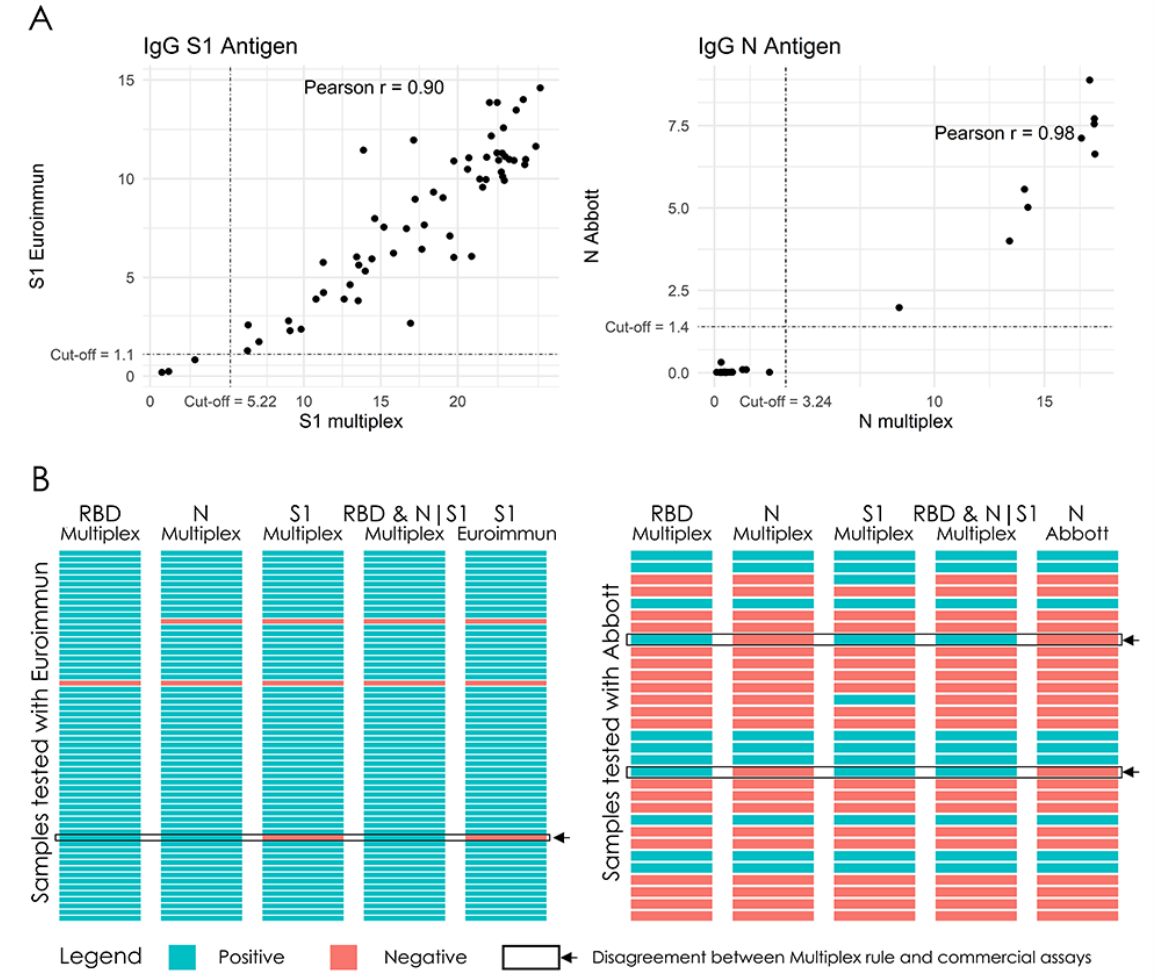

Figure 3. Validation of the multiplex assay against commercial serological assays. (A) Scatter plot of the S1 and N single antigen readouts of the multiplex IgG assay (termed S1 multiplex and N multiplex) against results from the Euroimmun S1 (n=60) and Abbott N (n=31) commercial assays, respectively. Dotted lines correspond to assay cut-offs for positivity. (B) Heatmap of diagnostic outcomes depicting the agreement between the multiplex and the commercial assays.

Figure 4.

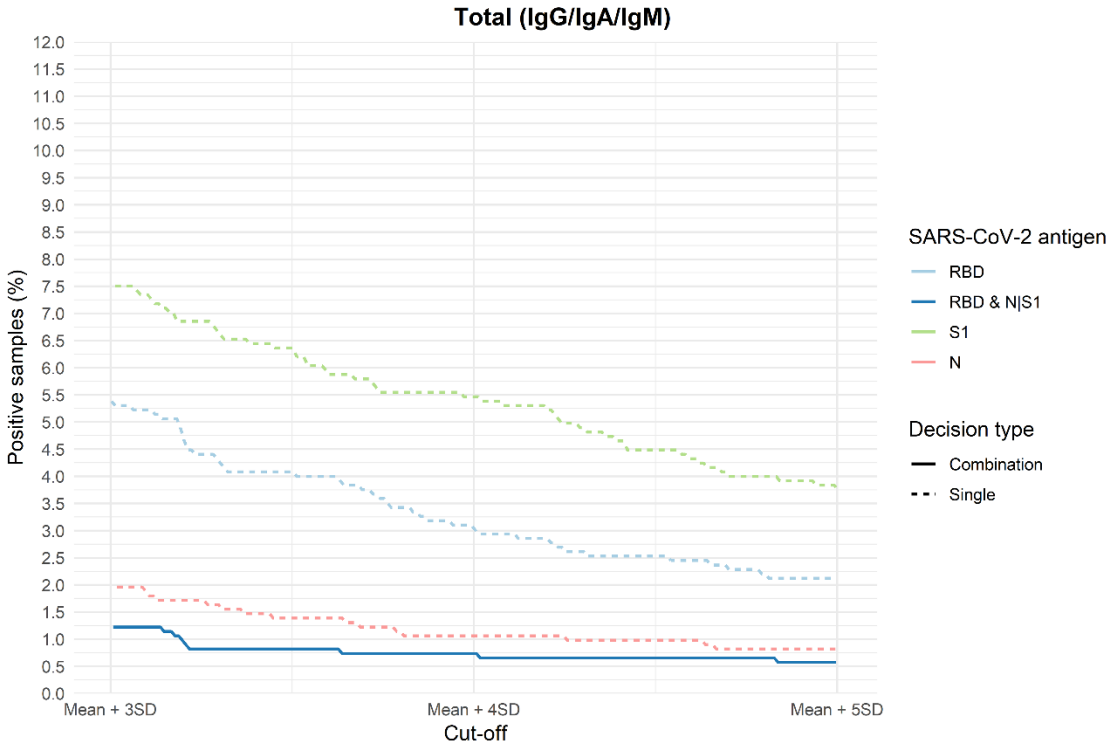

Figure 4. Percentage of asymptomatic individuals (n=1,255) found positive for total (IgG/IgA/IgM) antibodies using single-antigen readouts or the RBD&N|S1 multi-antigen rule. Seroprevalence is calculated using cut-off values that are based on the distribution of negative samples from the clinical performance analysis (n=78).

Figure 5.

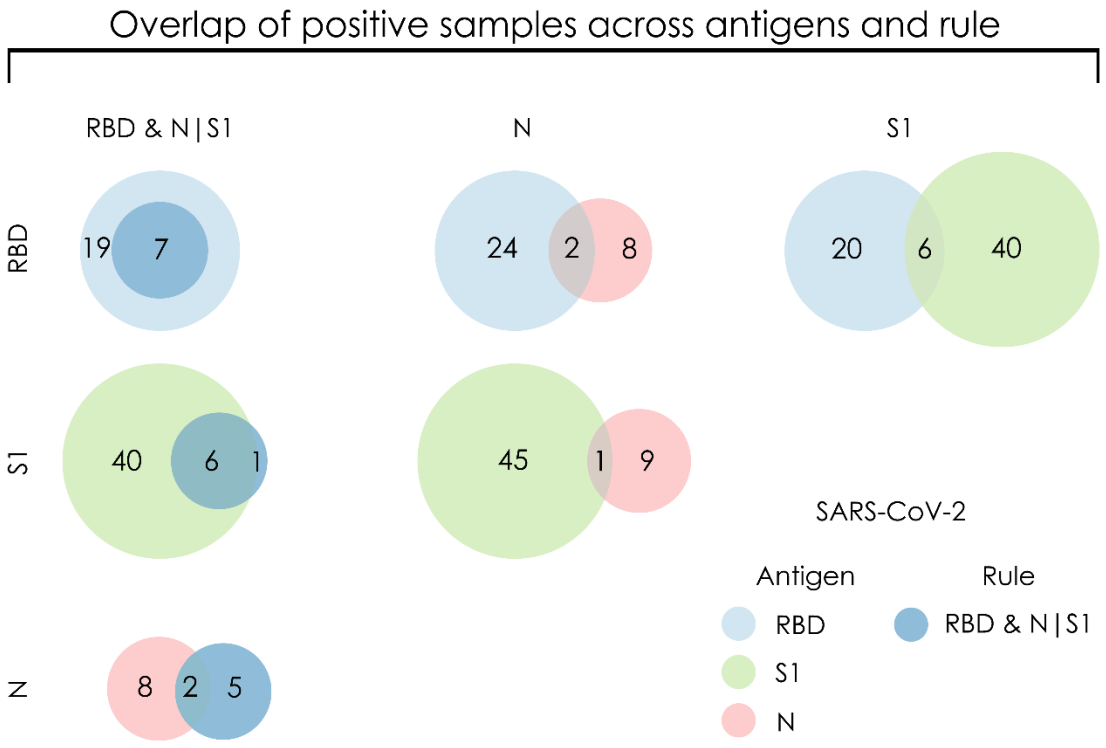

Figure 5. Consensus between positive samples for total (IgG/IgA/IgM) antibodies using N, RBD, S1 and the RBD&N|S1 multi-antigen rule and a mean plus 5 SD cut-off in the population screen.

TABLES

Table 1. List of samples used throughout this study.

Cohort Type	Sample Type	Sample Source	No of Samples	Demographics		
				Median Age (range)	Gender	Days post infection (range)
Clinical*	Positive Serum (SARS-CoV-2 PCR-confirmed cases)	Alexandra General Hospital, Athens	60	49 (20-74)	65% M 35% F	46 (13-87)
		University Hospital of Patras	17			
	Negative Serum (banked samples from 2018-2019)	University Hospital of Patras	78	57 (17-71)	51% M 49% F	N/A
Population screen	Serum from blood donors	Multi-center/ University of Patras	1225	42 (17-65)	74% M 26% F	N/A

Table 2. Correlation of normalized MFI values between SARS-CoV-2 antigens for the different antibody isotypes in positive and negative cases.

Isotype	Comparison	SARS-CoV-2 Positive Cases	SARS-CoV-2 Negative Cases
Total (IgG/IgM/IgA)	N *vs* S1	0.82	0.29
	N *vs* RBD	0.81	0.32
	S1 *vs* RBD	0.98	0.23
IgG	N *vs* S1	0.75	0.22
	N *vs* RBD	0.72	0.38
	S1 *vs* RBD	0.98	0.34
IgA	N *vs* S1	0.27	0.56
	N *vs* RBD	0.24	0.21
	S1 *vs* RBD	0.96	-0.03
IgM	N *vs* S1	0.39	0.37
	N *vs* RBD	0.35	0.27
	S1 *vs* RBD	0.96	0.78

The Pearson's correlation coefficient (r) is presented for each comparison. Color formatting was applied to all r values with green representing higher values and red representing lower or negative values.

Table 3. Diagnostic Sensitivity and Specificity of assays from single antigen readouts and multi-antigen rules.

		Total (IgG/IgM/IgA)		IgG		IgA		IgM	
Rule	**Description**	**Sensitivity** [95% CI]	**Specificity** [95% CI]	**Sensitivity** [95% CI]	**Specificity** [95% CI]	**Sensitivity** [95% CI]	**Specificity** [95% CI]	**Sensitivity** [95% CI]	**Specificity** [95% CI]
N	Positive for N	0.948 [0.874,0.980]	0.974 [0.911,0.993]	0.948 [0.874, 0.980]	0.974 [0.911, 0.993]	0.519 [0.410, 0.627]	0.987 [0.927, 0.998]	0.221 [0.143, 0.325]	0.974 [0.911, 0.993]
S1	Positive for S1	0.961 [0.892,0.987]	0.974 [0.911,0.993]	0.948 [0.874, 0.980]	0.987 [0.927, 0.998]	0.377 [0.277, 0.488]	0.974 [0.911, 0.993]	0.831 [0.732, 0.899]	0.974 [0.911, 0.993]
RBD	Positive for RBD	**0.987** [0.930,0.998]	0.974 [0.911,0.993]	**0.974** [0.910, 0.993]	0.987 [0.927, 0.998]	0.935 [0.857, 0.972]	0.974 [0.911, 0.993]	0.792 [0.689, 0.868]	0.974 [0.911, 0.993]
N & S1	Positive for N and S1	0.935 [0.857,0.972]	**1.000** [0.953,1.000]	0.935 [0.857, 0.972]	**1.000** [0.953, 1.000]	0.247 [0.164, 0.354]	0.987 [0.927, 0.998]	0.208 [0.132, 0.311]	**1.000** [0.953, 1.000]
RBD & S1	Positive for RBD and S1	0.961 [0.892,0.987]	**1.000** [0.953,1.000]	0.948 [0.874, 0.980]	**1.000** [0.953, 1.000]	0.377 [0.277, 0.488]	**1.000** [0.953, 1.000]	0.779 [0.675, 0.857]	0.987 [0.927, 0.998]
RBD & N	Positive for N and RBD	0.948 [0.874,0.980]	**1.000** [0.953,1.000]	0.948 [0.874, 0.980]	0.987 [0.927, 0.998]	0.506 [0.397, 0.615]	**1.000** [0.953, 1.000]	0.208 [0.132, 0.311]	**1.000** [0.953, 1.000]
RBD&N| S1	Positive for RBD and N or S1	0.974 [0.910,0.993]	**1.000** [0.953,1.000]	0.961 [0.892, 0.987]	0.987 [0.927, 0.998]	0.636 [0.525, 0.735]	**1.000** [0.953, 1.000]	0.779 [0.675, 0.857]	0.987 [0.927, 0.998]
N & RBD|S1	Positive for N and RBD or S1	0.948 [0.874,0.980]	**1.000** [0.953,1.000]	0.948 [0.874, 0.980]	0.987 [0.927, 0.998]	0.506 [0.397, 0.615]	0.987 [0.927, 0.998]	0.208 [0.132, 0.311]	**1.000** [0.953, 1.000]
S1 & RBD|N	Positive for S1 and RBD or N	0.961 [0.892,0.987]	**1.000** [0.953,1.000]	0.948 [0.874, 0.980]	**1.000** [0.953, 1.000]	0.377 [0.277, 0.488]	0.987 [0.927, 0.998]	0.779 [0.675, 0.857]	0.987 [0.927, 0.998]
RBD | N&S1	Positive for RBD or N and S1	**0.987** [0.930,0.998]	0.974 [0.911,0.993]	**0.974** [0.910, 0.993]	0.987 [0.927, 0.998]	0.935 [0.857, 0.972]	0.962 [0.893, 0.987]	0.792 [0.689, 0.868]	0.974 [0.911, 0.993]
S1 | RBD&N	Positive for S1 or RBD and N	0.974 [0.910,0.993]	0.974 [0.911,0.993]	0.961 [0.892, 0.987]	0.974 [0.911, 0.993]	0.636 [0.525, 0.735]	0.974 [0.911, 0.993]	0.831 [0.732, 0.899]	0.974 [0.911, 0.993]
N | RBD&S1	Positive for N or RBD and S1	0.974 [0.910,0.993]	0.974 [0.911,0.993]	0.961 [0.892, 0.987]	0.974 [0.911, 0.993]	0.649 [0.538, 0.747]	0.987 [0.927, 0.998]	0.792 [0.689, 0.868]	0.962 [0.893, 0.987]
RBD & S1 & N	Positive for all three	0.935 [0.857,0.972]	**1.000** [0.953,1.000]	0.935 [0.857, 0.972]	**1.000** [0.953, 1.000]	0.247 [0.164, 0.354]	**1.000** [0.953, 1.000]	0.208 [0.132, 0.311]	**1.000** [0.953, 1.000]
RBD | N | S1	Positive for any	**0.987** [0.930,0.998]	0.923 [0.842,0.964]	**0.974** [0.910, 0.993]	0.962 [0.893, 0.987]	**0.948** [0.874, 0.980]	0.948 [0.875, 0.980]	**0.857** [0.762, 0.918]	0.936 [0.856, 0.972]