nature research

https://doi.org/10.1038/s41551-020-00611-x

Supplementary information

Ultrasensitive high-resolution profiling of early seroconversion in patients with COVID-19

In the format provided by the authors and unedited

Table of contents

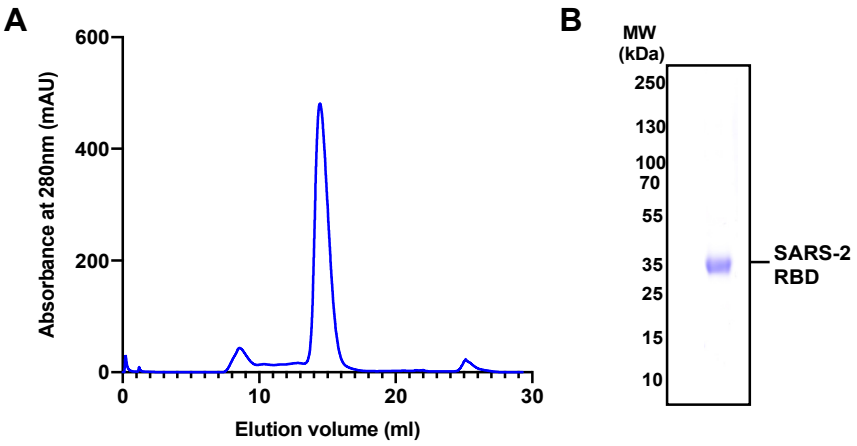
Figure S1: RBD sequence and purification characterization

RBD represents amino acid residues 319-529 from full-length SARS-CoV-2 spike glycoprotein (GenBank Accession no. MN975262.1)

>SARS CoV2 RBD pVRC
GCTAGCCGCGTACAGCCCACTGAAAGCATTGTTCGCTTTCCCAACATAACAAACTTGTGCCCCTTTGGA
GAGGTCTTCAATGCTACGAGATTTGCCTCCGTCTACGCTTGGAATCGGAAACGGATTAGCAACTGCGTG
GCTGACTATTCAGTGCTGTACAACTCTGCTTCTTTCTCAACATTCAAGTGTTACGGGGTCTCTCCAACTA
AACTGAATGATCTTTGTTTTACGAACGTATACGCAGATTCTTTCGTAATACGAGGCGATGAGGTGAGGCA
GATCGCCCCCGGGCAAACTGGTAAAATAGCTGACTACAATTACAAGTTGCCGGACGACTTCACTGGGT
GTGTTATTGCCTGGAATTCTAATAACTTGGATTCCAAAGTAGGAGGCAATTACAACTATCTTTACAGGCT
TTTCCGGAAATCTAACTTGAAACCGTTTGAAAGGGACATCTCAACGGAGATCTACCAGGCAGGGTCCAC
GCCTTGCAATGGGGTAGAGGGTTTTAATTGTTATTTTCCCTTGCAGTCATATGGTTTCCAACCGACAAAT
GGTGTGGGTTACCAGCCGTACCGGGTCGTGGTATTGTCCTTCGAGTTGCTCCATGCGCCAGCGACTGT
ATGTGGGCCCAAAAAGGGCGCCGGCTCCAGCCTTGAGGTCCTCTTCCAAGGACCCGGTTCAGGCTCCA
GCCACCATCACCATCATCATCACCACGGGGGTTCTGGCTCCAGCATGGATGAGAAAACCACCGGCTGG
AGAGGCGGTCACGTCGTAGAGGGGTTGGCTGGCGAACTCGAACAACTTCGAGCGAGATTGGAACACC
ATCCCCAGGGTCAGCGGGAGCCTTGATGAGCGGCCGCT
NheI
KasI
NotI

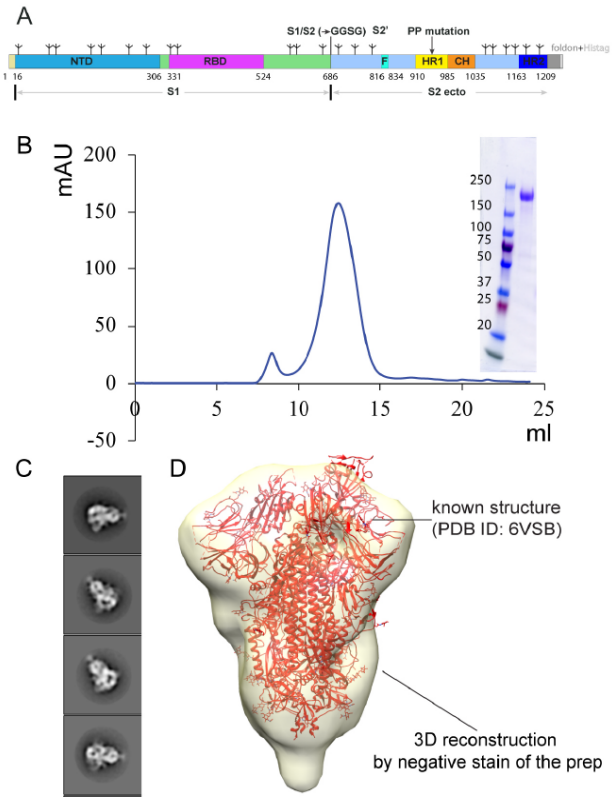
>translation SARS CoV2 RBD pVRC
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLC
FTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKP
FERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKGAGSSL
EVLFQGPGSGSSHHHHHHHHGGSGSSMDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP

HRV 3C Cleavage Site
8xHis Tag
Streptavidin-Binding Peptide (SBP) Tag



Supplementary Figure 1. Purification and biochemical characterization of SARS-CoV-2 (SARS-2) RBD. A) Representative size exclusion chromatography profile of SARS-2 RBD eluting at approximately 14.5 ml. B) SDS-PAGE analysis (reducing conditions) of the purified SARS-2 RBD.

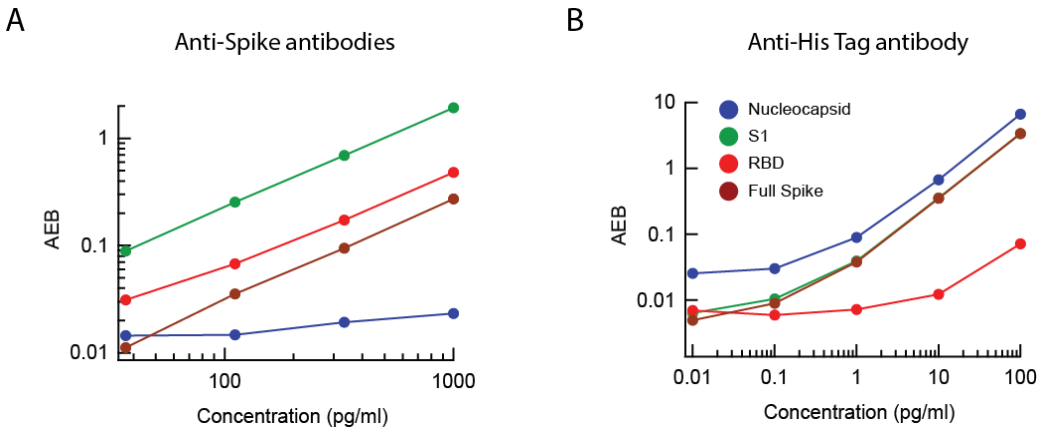
Figure S2: Spike protein sequence and purification validation



Supplementary Figure 2. Production of a stabilized S ectodomain trimer in the prefusion conformation. (A) Schematic representation of the construct design. Various segments include: NTD, N-terminal domain; RBD, receptor-binding domain; S1/S2, S1/S2 cleavage site; S2', S2' cleavage site; F, fusion peptide; HR1, heptad repeat 1; PP mutation, two proline residues introduced to stabilize the prefusion conformation; CH, central helix region; HR2, heptad repeat 2; foldon trimerization tag and a His tag; and tree-like symbols for glycans. (B) The purified S ectodomain trimer was resolved by gel-filtration chromatography. Inset, fractions of the peak were analyzed by Coomassie stained SDS-PAGE. (C) Negative stain EM of the S trimer showing representative 2D class averages, box size 220 Å. (D) 3D reconstruction showing the defined shape of S trimer, superimposed with the atomic model of S trimer (PDB: 6VSB).

Figure S3: Bead coupling validation

In order to confirm that coupling was successful, we tested the beads in two ways. First, we tested the three proteins which had His tags. To that end we loaded all four beads and used a biotinylated anti-his tag antibody for detection in decreasing quantities. As expected, the three His tagged proteins gave high signal while the RBD protein, which had its His tag removed, gave low signal. Next, to ensure the RBD protein had also been conjugated properly to the bead, we introduced a recombinant human anti-RBD domain antibody (CR3022) and used an anti-human IgG biotinylated detector to complete the immunocomplex. Since the Spike protein and S1 also contain the RBD domain we expected that the antibody would bind to all three, but that it would not bind to the nucleocapsid protein.



Supplementary Tables 1-3: Detector antibody screening

In order to identify the highest affinity anti-IgM and anti-IgA antibodies, we screened candidates using a sandwich Simoa immunoassay for detection of each isotype. In this screening process commercially available human-plasma-derived immunoglobulins (IgM Sigma-Aldrich I8260 and IgA Sigma-Aldrich I4036) were used as our protein standards in a sandwich Simoa assay. We then cross-tested each anti-IgM or anti-IgA antibody as a capture antibody against all other candidate IgM or IgA antibodies as detectors in order to establish high binding affinity on the Simoa platform. This was done separately for IgM and IgA. Tight binders were selected for further analysis. Below, are tables indicating signal to noise ratios (SNR) for each antibody pair.

Supplementary Table 1: IgA detector cross-testing

IgA detector	411502	MAB4787	AB214003	AB224184	A80109B	AB128739	AB128731
411502		10.59	41.94	48.74	115.36	14.79	1.25
MAB4787	0.16		121.78	4.81	25.7	64.51	4.13
AB214003	61.49	13.28		28.72	58.25	3.88	1.08

Supplementary Table 2: IgM detector cross-testing

IgM detector	MII401	MII402	MAB9435	MAB94351	314502	A80-100B
MII401		3.62	1.59	1.29	41.60	NaN
MII402	33.01		1.14	1.30	23.48	NaN
MAB9435	1.04	1.00		1.10	0.19	0.89
MAB94351	0.94	1.04	0.32		NaN	0.35
314502	15.75	2.72	0.80	0.91		5.49

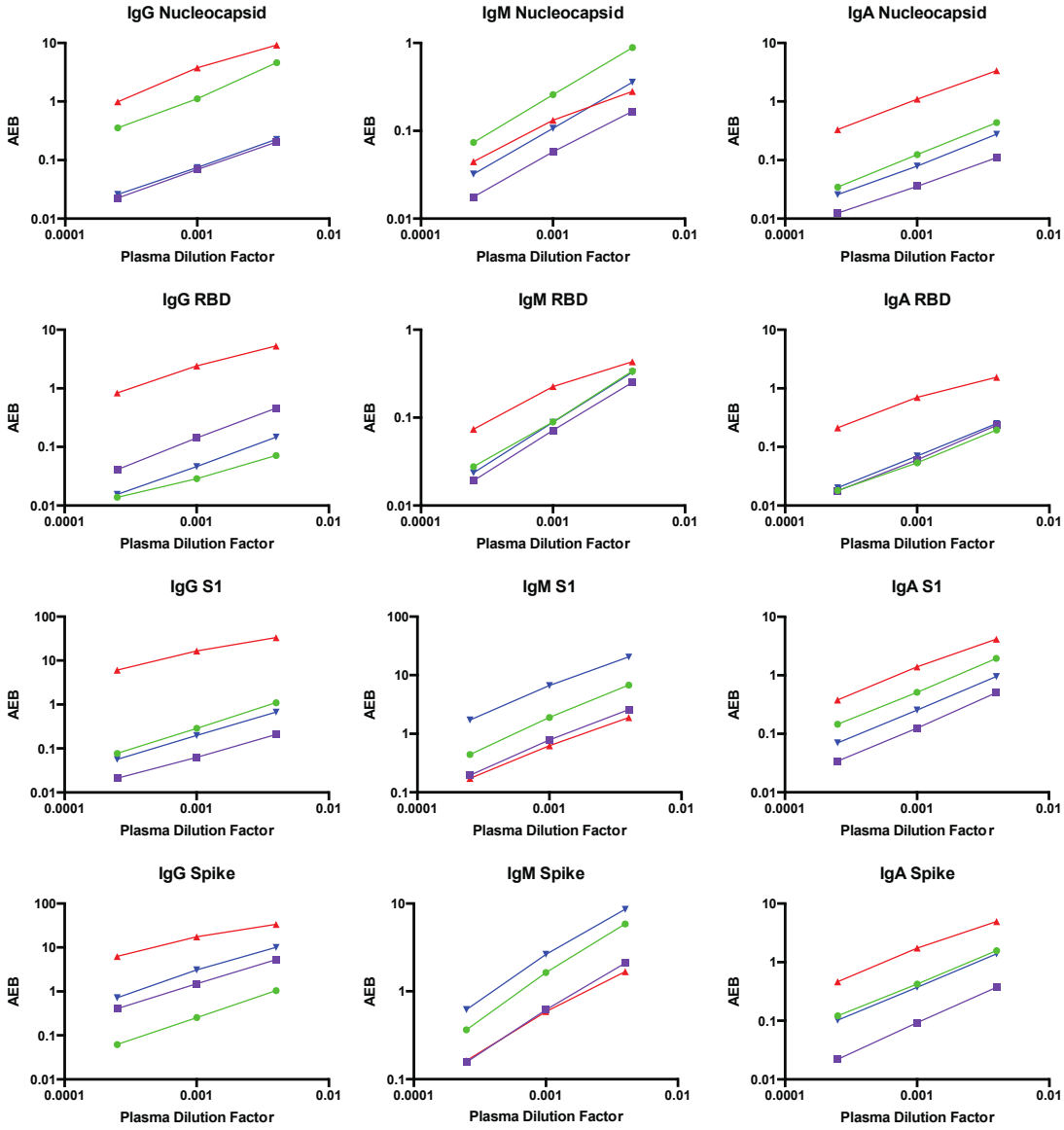
Since the antibodies to be screened for IgG detection were mouse, rabbit and goat IgGs, we could not screen them using the above method because of species cross-reactivity. Instead, we screened anti-IgG detector antibodies using a mixture of three commercially available recombinant human IgGs to S1 (Creative Biolabs CR3022, CBFYR-0119, CBFYR-0120) and the four viral epitope coated beads (spike, S1, RBD and nucleocapsid). As expected, the beads that show a high SNR are those that have an RBD domain (RBD, S1, Spike) but not the nucleocapsid. A80-148B was chosen for further analysis due to its high SNR.

Supplementary Table 3:

IgG Detector	A80-148B	410701	409302
Nucleocapsid bead SNR	**2.68**	1.34	1.44
S1 bead SNR	**216.29**	143.28	10.11
RBD bead SNR	**66.31**	15.77	24.42
Spike bead SNR	**651.27**	4.69	34.34

Supplementary Figure 4: Dilution linearity for all immunoglobulin assays

Four COVID 19+ patient samples were diluted three times (250x, 1000x & 4000x). Below we demonstrate that in all 12 assays the AEB values (signal) diluted linearly for all patients.



Supplementary Table 4: Spike and recovery for IgG assay

In order to establish that the assays were accurate and quantitative, we created a calibration curve using three recombinant human IgG antibodies to S1 that we serially diluted (Creative Biolabs CR3022, CBFYR-0119, CBFYR-0120). We then spiked these antibodies into pre-pandemic pooled human plasma (BIOIVT) in order to show good recovery at relevant dilution factors. We did not have recombinant IgM or IgA human antibodies to SARS-CoV-2, so this validation could only be done for IgG.

Spiked concentration (pg/ml)	Dilution factor	AEB	Measured concentration (pg/ml)	% Recovery
250	250	0.128359	230.5	92.2
250	500	0.138653	251.1	100.4
250	1000	0.142077	258.1	103.2
1000	250	0.425218	892.7	89.3
1000	500	0.423456	888.1	88.8
1000	1000	0.408272	849.8	85.0

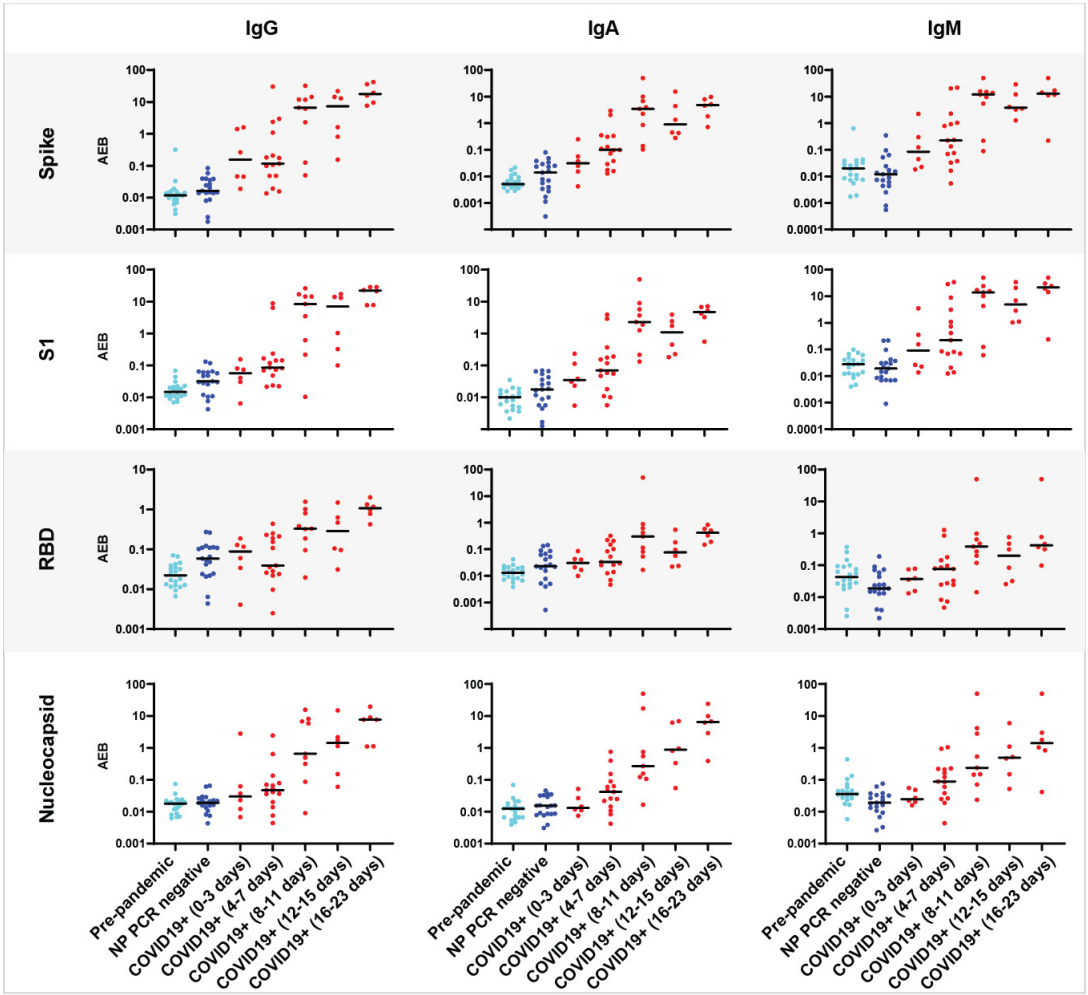
Supplementary Table 5: Clinical characteristics of the discovery cohort

Patient characteristics	NP PCR Positive (n= 21)	NP PCR Negative (n=19)	Pre-pandemic control (n=20)
Age (25^{th}, 75^{th} quartile)	58 (45, 72)	61 (41 ,75)	55 (33, 59)
Male (%)	20 (90.9)	11 (57.9)	8 (40)
Symptoms at presentation (%)			
Fever	16 (72.7)	4 (21)	NA
Cough	17 (77.2)	5 (26.3)	NA
SOB	15 (68.2)	5 (26.3)	NA
Sore throat	0 (0)	1 (5.3)	NA
Myalgia	4 (18.2)	2 (10.5)	NA
Risk Factors			
Diabetes	4 (18.2)	NA	NA
Immunosuppression	2 (9.1)	NA	NA
Lung disease	2 (9.1)	NA	NA
Heart disease	7 (31.8)	NA	NA
Clinical Outcome (%)			
Required Ventilation	12 (54.5)	NA	NA
Discharged from hospital	11 (0.5)	NA	NA
Still admitted	9 (45)	NA	NA
Death	2 (10)	NA	NA
CoV related death	2 (10)	NA	NA

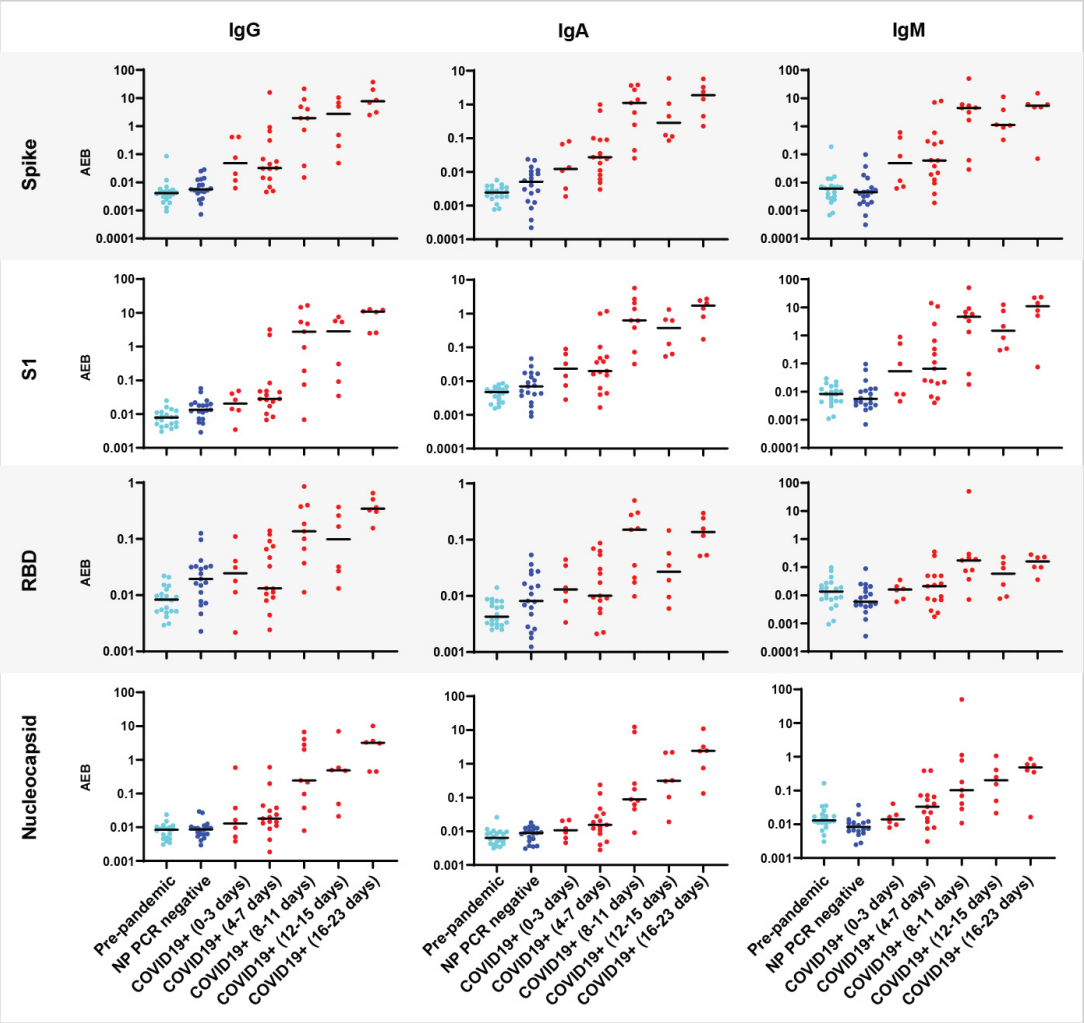
Supplementary Figure 5: Discovery cohort results

In order to establish whether our assay detects seroconversion in SARS-CoV-2 infected patients, we initially tested 81 plasma samples from Massachusetts General Hospital (MGH) in Boston, MA, USA. This sample cohort consisted of three groups: (1) patient plasma collected before the COVID-19 outbreak, (2) plasma collected during the pandemic from patients presenting with symptoms of SARS-CoV-2 at the Emergency Department (ED) with a negative NP RNA test, and (3) serial plasma samples collected during the pandemic from patients who had tested positive for SARS-CoV-2 infection and were hospitalized for COVID-19. Clinical characteristics of each patient group are shown in Supplementary Table 5. Samples were measured at two different dilution factors and similar results were observed. Plasma samples measured at a 4000x dilution factor show that all four viral antigens elicit an IgG, IgM, and IgA immune response in COVID-19 positive patients. All twelve combinations (i.e. each of the four viral targets against each of the three immunoglobulins) show an overall increase in immunoglobulin levels over the course of the disease. Only one patient had a decrease in immunoglobulin levels; this person did not survive their infection with SARS-CoV-2. In addition, a separate patient who displayed low immunoglobulins levels at the late time point was on immunosuppressive medication.

A. Simoa serological assay results using 1000x dilution



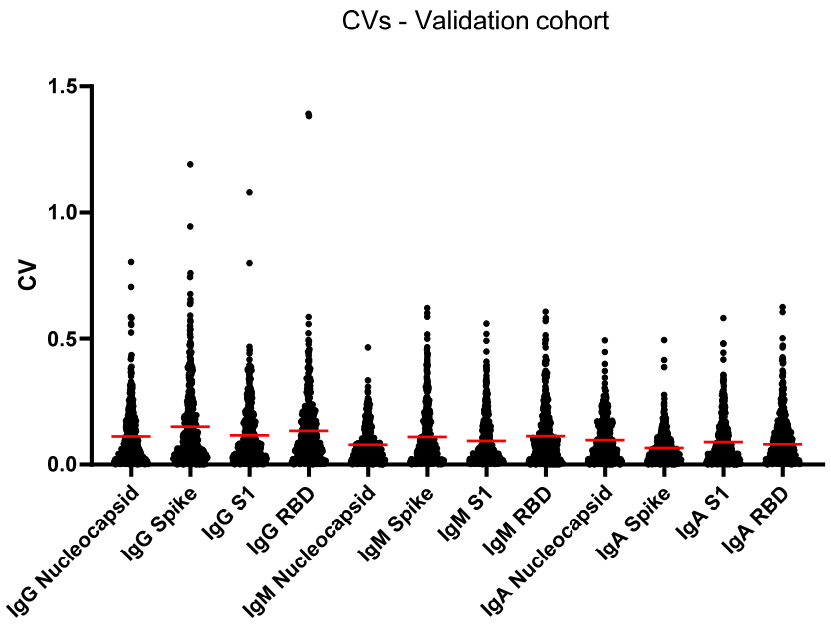
B. Simoa serological assay results using 4000x dilution



Supplementary Figure 5. Simoa serological assay results for IgG, IgM, and IgA against the four viral targets: Spike, S1 subunit, Receptor Binding Domain, and Nucleocapsid for pre-pandemic samples (light blue, n=20), NP PCR negative samples (dark blue, n=19) and COVID-19 positive samples (red, n=21, two timepoints each). The COVID-19 positive samples were divided into five groups according to time since symptom onset. Black lines indicate the mean AEB (average enzyme per bead) value of each population.

Supplementary Figure 6: CV values – validation cohort

Coefficients of Variance (CV) of the duplicate measurements for each of the samples in the validation cohort. Each dot represents the CV of duplicate measurements for each of the 472 datapoints show in Figure 2. The red line indicates the median of the CVs. As can be seen, all CVs were in an acceptable range indicating the accuracy of the measurements.



Supplementary Table 6: U-test results – validation cohort

Mann-Whitney U test results for each of the four subgroups (0-3 days, 4-7 days, 8-14 days, and >14 days after positive NP RT-PCR) against the pre-pandemic control. All U-tests are two-tailed but do not correct for multiple comparisons.

	IgG Nuc	IgG Spike	IgG S1	IgG RBD	IgM Nuc	IgM Spike	IgM S1	IgM RBD	IgA Nuc	IgA Spike	IgA S1	IgA RBD
0-3 days	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	p=0.0101	P<0.0001	P<0.0001	P<0.0001	P<0.0001
4-7 days	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
8-14 days	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
>15 days	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001

Supplementary Table 7: Five-fold cross validation results for early stage cases

	Fold 1		Fold 2		Fold 3		Fold 4		Fold 5	
	Antibody	**VI**	**Antibody**	**VI**	**Antibody**	**VI**	**Antibody**	**VI**	**Antibody**	**VI**
Training Set	IgA S1	100	IgA S1	100	IgA S1	100	IgA S1	100	IgG Spike	100
	IgA Nuc	96	IgG S1	89	IgG Nuc	71	IgG Spike	97	IgA RBD	95
	IgG Nuc	91	IgG Nuc	83	IgA Nuc	68	IgG Nuc	77	IgA S1	85
	IgG Spike	62	IgA Nuc	81	IgG RBD	61	IgA Nuc	56	IgG Nuc	83
	IgG S1	45	IgM RBD	0	IgG S1	52	IgA RBD	54	IgA Nuc	72
	IgG RBD	3	IgG Spike	0	IgA RBD	48	IgG RBD	0	IgG RBD	69
	IgM Spike	1	IgM S1	0	IgG Spike	33	IgG S1	0	IgM RBD	2
	IgM S1	0	IgA RBD	0	IgA Spike	0	IgA Spike	0	IgM Nuc	0
	IgA Spike	0							IgA Spike	0

AUC, All Test Sets Combined: 98% (95% CI 96%, 100%)

Each of the 290 participants in the cross validation was randomly assigned to one of five groups that included 58 participants. In each of five folds, one group of 58 was set aside as a fold-specific test set and the other four groups were combined to create a training set of 232 participants. For each fold, the model was identified through a backwards-selection process starting with predictors of all 12 antibodies measured. Variable importance rounded to the nearest integer. Variables that dropped out entirely had importance of 0.

AUC=Area Under Receiver Operating Characteristic Curve, Nuc=Nucleocapsid, RBD=RNA Binding Domain, VI=Variable Importance

Supplementary Table 8: Five-fold cross validation results for late stage cases

	Fold 1		Fold 2		Fold 3		Fold 4		Fold 5	
	Antibody	**VI**	**Antibody**	**VI**	**Antibody**	**VI**	**Antibody**	**VI**	**Antibody**	**VI**
Training Set	IgM S1	100	IgA S1	100	IgA S1	100	IgA S1	100	IgA S1	100
	IgA Spike	41	IgA Spike	0			IgG S1	29		
	IgA S1	29					IgM S1	17		
	IgG RBD	5					IgA RBD	13		
	IgM Spike	-8					IgG Nuc	2		
	IgG Nuc	-10					IgG Spike	0		
							IgM RBD	0		

AUC, All Test Sets Combined: 100% (95% CI 100%, 100%)

Each of the 249 participants in the cross validation was randomly assigned to one of five groups that included 50 participants (one group had only 49 participants). In each of five folds, one group of 50 (or 49 in one fold) was set aside as a fold-specific test set and the other four groups were combined to create a training set of 199 participants (or 200 in one fold). For each fold, the model was identified through a backwards-selection process starting with predictors of all 12 antibodies measured. Variable importance rounded to the nearest integer. Variables that dropped out entirely had importance of 0.

AUC=Area Under Receiver Operating Characteristic Curve, Nuc=Nucleocapsid, RBD=RNA Binding Domain, VI=Variable Importance

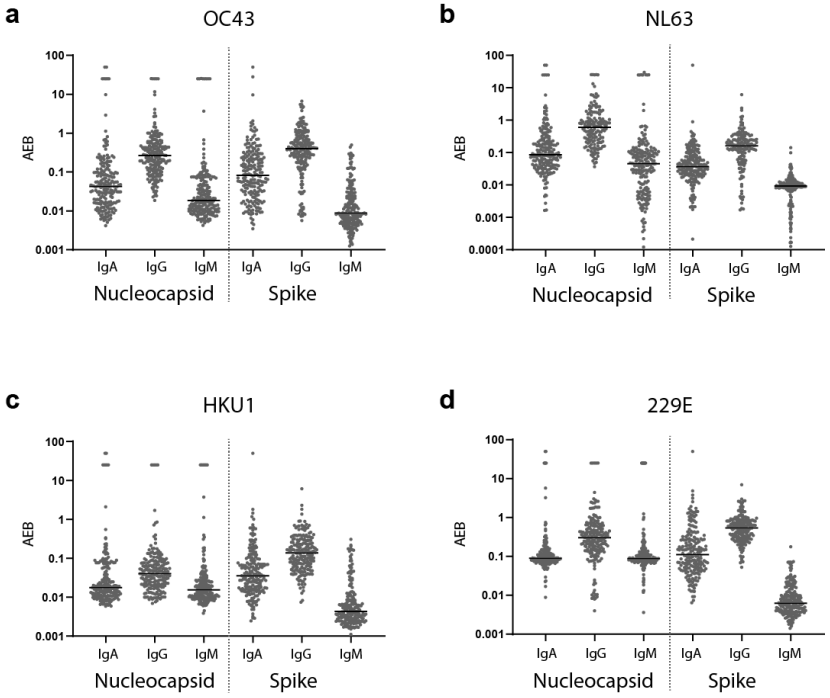
Supplementary Table 9: Classification accuracy

Case Group	Model	Positive Cases	Total Cases	Sensitivity (%)	Negative Controls	Total Controls	Specificity (%)	Overall Accuracy (%)
All	Early-Stage	124	141	88	195	199	98	94
All	Late-Stage	122	141	87	192	199	96	92
All	Full	127	141	90	194	199	97	94
Early-Stage	Early-Stage	74	91	81	195	199	98	93
Early-Stage	Full	77	91	85	194	199	97	93
Late-Stage	Late-Stage	50	50	100	199	199	100	100
Late-Stage	Full	50	50	100	199	199	100	100

Reference group was always All Controls. Predicted case status was case if model-specific predicted probability was at least 50%, otherwise predicted to be a control if predicted probability was less than 50%.

Supplementary Figure 7: Cross reactivity of the control samples with common human coronaviruses

In order to assess reactivity of pre-pandemic samples with other coronaviruses, which cause the common cold, we developed assays for Nucleocapsid and S1/RBD of the four common human coronaviruses: 229e (alpha coronavirus), NL63 (alpha coronavirus), OC43 (beta coronavirus), HKU1 (beta coronavirus). All antigens were purchased commercially from Sino Biological with the exception of the OC43 RBD, which was produced by the Aaron Schmidt Lab: 229E N Nucleocapsid (40640-V07E), NL63 Nucleocapsid (40641-V07E), OC43 Nucleocapsid (40643-V07E), HKU1 Nucleocapsid (40642-V07E), 229E S1 (40601-V08H), NL63 S1 (40600-V08H), HKU S1 (40602-V08H). We hypothesized that many of the individuals in our pre-pandemic cohort had likely experienced an infection with one or more of these coronaviruses, and we could therefore assess how antibodies to these assays correlated with activity in our SARS-CoV-2 serological assay. To accomplish this, we coupled two sets of multiplex beads, one for the N and one for the S1/RBD antigen. Each of the four multiplex beads (488, 647, 700 and 750) was coupled to the antigen from one of four common human coronaviruses: 229e, NL63, OC43 and HKU1. Using the same assay format as was used for the SARS-CoV-2 serological assay, we tested these two multiplexed assays using detection antibodies to IgG, IgM and IgA yielding 24 values per sample. We tested the 200 pre-pandemic samples for reactivity with these other coronaviruses. Several pre-pandemic control samples had high levels of antibodies for these common coronaviruses, likely indicating prior infection with these species.



Codes caption:

1. Cross validation and models training.

This SAS code is for the analyses of all cases and all controls in the training set. Predicted probabilities were estimates of case status (case vs. control) as predicted by different sets of circulating antibody measurements. Two cross validations were run to identify and validate important markers: 1) among early-stage cases vs. all controls and 2) among late-stage cases vs. all controls. The predicted probabilities were used to estimate areas under receiver operating characteristic curves, calibration plots, and classification accuracies (sensitivity and specificity). For comparisons of early-stage cases vs. all controls, the code also identifies the maximum sensitivity at which a specificity of 100% is maintained for both the full 12-marker model and the 4-marker early-stage model.

2. Classification of the blinded validation set according to the three models

This R code trains the three models (early stage model, late stage model and full marker panel model) with all cases and controls in the training set, and uses those models to classify the blinded validation cohort. The output includes all the ROC curves, which are used to determine the classification threshold, and the probabilities for all the unknown cases, which are then used for the classification according to a chosen threshold.