

1 Probabilistic approaches for classifying highly variable anti-SARS-CoV-2 2 antibody responses

3
4 Xaquín C Dopico^{1#}, Leo Hanke^{1°}, Daniel J. Sheward^{1°}, Sandra Muschiol^{2*}, Soo Aleman^{3*},
5 Nastasiya F. Grinberg⁴, Monika Adori¹, Murray Christian¹, Laura Perez Vidakovics¹, Changil
6 Kim¹, Sharesta Khoenkhoen¹, Pradeepa Pushparaj¹, Ainhua Moliner Morro¹, Marco
7 Mandolesi¹, Marcus Ahl³, Mattias Forsell⁵, Jonathan Coquet¹, Martin Corcoran¹, Joanna
8 Rorbach^{6,7}, Joakim Dillner⁸, Gordana Bogdanovic², Gerald M. McInerney¹, Tobias
9 Allander^{1,2}, Ben Murrell¹, Chris Wallace^{4,9}, Jan Albert^{1,2}, Gunilla B. Karlsson Hedestam^{1#}

11 Affiliations:

12 ¹Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm 171 77, Sweden

13 ²Department of Clinical Microbiology, Karolinska University Hospital, Stockholm 171 76, Sweden

14 ³Department of Infectious Diseases, Karolinska Universitetssjukhuset, Huddinge 141 52, Sweden

15 ⁴Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), Jeffrey Cheah Biomedical
16 Centre, Cambridge Biomedical Campus, University of Cambridge, Cambridge, CB2 0AW

17 ⁵Department of Clinical Microbiology, Umeå Universitet, Umeå 901 85, Sweden

18 ⁶Department of Molecular Biochemistry & Biophysics, Karolinska Institutet, Stockholm 171 77, Sweden

19 ⁷Max Planck Institute-Biology of Ageing, Karolinska Institutet Laboratory, Stockholm 171 77, Sweden

20 ⁸Department of Laboratory Medicine, Division of Pathology, Karolinska Institutet, Huddinge 141 52, Sweden

21 ⁹MRC Biostatistics Unit, University of Cambridge CB2 0SR, United Kingdom

22
23 °*Equal contribution

24 #Correspondence

25

26

27

28 **Abstract**

29

30 Antibody responses vary widely between individuals¹, complicating the correct classification
31 of low-titer measurements using conventional assay cut-offs. We found all participants in a
32 clinically diverse cohort of SARS-CoV-2 PCR+ individuals ($n=105$) – and $n=33$ PCR+
33 hospital staff – to have detectable IgG specific for pre-fusion-stabilized spike (S)
34 glycoprotein trimers, while 98% of persons had IgG specific for the receptor-binding domain
35 (RBD). However, anti-viral IgG levels differed by several orders of magnitude between
36 individuals and were associated with disease severity, with critically ill patients displaying
37 the highest anti-viral antibody titers and strongest *in vitro* neutralizing responses. Parallel
38 analysis of random healthy blood donors and pregnant women ($n=1,000$) of unknown
39 serostatus, further demonstrated highly variable IgG titers amongst seroconverters, although
40 these were generally lower than in hospitalized patients and included several measurements
41 that scored between the classical 3 and 6SD assay cut-offs. Since the correct classification of
42 seropositivity is critical for individual- and population-level metrics, we compared different
43 probabilistic algorithms for their ability to assign likelihood of past infection. To do this, we
44 used tandem anti-S and -RBD IgG responses from our PCR+ individuals ($n=138$) and a large
45 cohort of historical negative controls ($n=595$) as training data, and generated an equal-
46 weighted learner from the output of support vector machines and linear discriminant analysis.
47 Applied to test samples, this approach provided a more quantitative way to interpret anti-viral
48 titers over a large continuum, scrutinizing measurements overlapping the negative control
49 background more closely and offering a probability-based diagnosis with potential clinical
50 utility. Especially as most SARS-CoV-2 infections result in asymptomatic or mild disease,
51 these platform-independent approaches improve individual and epidemiological estimates of
52 seropositivity, critical for effective management of the pandemic and monitoring the response
53 to vaccination.

54

55

56 Keywords: SARS-CoV-2; antibody testing; serology; probabilistic analyses; viral immunity

57

58 Introduction

59

60 The characterization of nascent SARS-CoV-2-specific antibody responses is critical to our
61 understanding of the infection at individual and population levels. Although several groups
62 have carried out elegant work in this regard²⁻⁵, consensus on several key issues remains
63 outstanding, such as: whether all infected persons develop an antibody response to the virus;
64 what the duration of these responses is following peak levels; and what titers provide
65 protective immunity against re-infection⁶⁻⁸.

66

67 Antibody responses to the SARS-CoV-2 spike glycoprotein (S) are particularly relevant, as
68 S-directed antibody specificities mediate virus neutralizing activity and new S variants (such
69 as B.1.1.7) have emerged. Indeed, the vast majority of COVID-19 vaccines are based on S
70 surface antigens, as the goal is to induce neutralizing antibodies that block viral entry into
71 ACE2-positive target cells^{9,10}. Furthermore, because serological studies play such a central
72 role in immunosurveillance, there is a pressing need for robust assays and quantitative
73 statistical tools to examine antibody titers of varying levels after vaccination and natural
74 infection in different target groups.

75

76 To meet these needs, we developed highly sensitive and specific IgM, IgG and IgA ELISA
77 assays based on mammalian cell-expressed pre-fusion-stabilized soluble trimers of the
78 SARS-CoV-2 spike (S) glycoprotein and the receptor-binding domain (RBD), and used them
79 in tandem to survey serum samples from large cohort of individuals PCR+ for SARS-CoV-2.
80 To validate our assays, we repeatedly analyzed a large set of serum samples from historical
81 blood donors as negative controls ($n=595$) - critical for determining the assay background.

82

83 As we show, and as has been reported by others^{4,7,11}, the magnitude of response varied
84 greatly between seropositive individuals and was associated with disease severity. Those with
85 most pronounced symptoms had the highest anti-viral antibody titers, while those with
86 asymptomatic or mild disease (including otherwise healthy blood donors and pregnant
87 women) exhibited a range of antibody levels, with many measurements in close
88 approximation to the negative control background, complicating their correct classification.

89 To improve upon the dichotomization of a continuous variable – which is common to many
90 clinical tests but results in a loss of information^{12,13} – we used tandem anti-S and RBD IgG
91 data from confirmed infections and negative controls to train different probabilistic

92 algorithms to assign likelihood of past infection. Compared to strictly thresholding the assay
93 at 3 or 6 standard deviations (SD) from the mean of negative control measurements, these
94 more quantitative approaches modelled the probability a sample was positive, improving the
95 identification of low titer values and paving the way for a greater utility to antibody test
96 results.
97

98 **Results**

99

100 Study samples are detailed in Fig. 1A and Table 1.

101

102 Antibody test development

103

104 We developed ELISA protocols to profile IgM, IgG and IgA specific for a pre-fusion-
105 stabilized spike (S) glycoprotein trimer¹⁴, the RBD, and the nucleocapsid (N). Trimer
106 conformation was confirmed in each batch by cryo-EM¹⁵ and a representative subset of study
107 samples was used for assay development (Fig. S1A). In contrast to other studies reporting
108 significant cross-reactivity to S in the UK population¹⁶, we did not observe reproducible IgG
109 reactivity to S or RBD across all 595 historical controls in the study, although two individuals
110 who were PCR-positive for endemic coronaviruses (ECV+) in the last six months displayed
111 reproducible IgM reactivity to both SARS-CoV-2 N and S, and two 2019 blood donors (from
112 $n=72$ tested) had low anti-S IgM reactivity (Fig. S1B). Thus, further investigation is required
113 to establish the contribution of potential cross-reactive memory SARS-CoV-2 responses¹⁷.

114

115 Responses to S and the RBD were highly correlated and our assay revealed a greater than
116 1,000-fold difference in anti-viral IgG titers between Ab-positive individuals when
117 examining serially diluted sera (Fig. S1C and D). In SARS-CoV-2 PCR+ individuals, anti-
118 viral IgG titers were comparable for S ($EC_{50}=3,064$; 95% CI [1,197 - 3,626]) and N
119 ($EC_{50}=2,945$; 95% CI [543 - 3,936]) and lower for RBD [$EC_{50}=1,751$; 95% CI 966 - 1,595].
120 Notably, a subset (*ca.* 10%) of the SARS-CoV-2-confirmed individuals did not have
121 detectable IgG responses against the SARS-CoV-2 nucleocapsid protein (N) (Fig. S1C), as
122 previously reported¹⁸. Therefore, we did not explore responses to N further. These results
123 highlight that the choice of antigen is critical for seropositivity estimates.

124

125 Elevated anti-viral Ab titers and neutralizing responses are associated with increased disease 126 severity

127

128 When screening samples from SARS-CoV-2 PCR+ individuals from whom clinical
129 information was available ($n=105$), we detected potent IgG responses against S in 100% of
130 participants, and against RBD in 97% of persons (Fig. 1B), supporting that natural infection
131 engenders a robust B cell response in the majority of cases, as reported⁸. IgM and IgA

132 responses were generally weaker and more variable and also spread over a large range (Fig.
133 1B).

134

135 To examine this further, PCR+ individuals were grouped according to their clinical status:
136 non-hospitalized (Cat. 1), hospitalized (Cat. 2) or admitted to the intensive care unit (Cat. 3).
137 To validate our clinical classification, we measured serum IL-6 levels in a random subset of
138 PCR+ individuals ($n=64$). IL-6 feeds Ab production¹⁹⁻²², and as has been reported²³, was
139 increased in samples from individuals with severe disease (Fig. 1C). Furthermore,
140 multivariate analyses (accounting for the effects of age, sex and days from symptom
141 onset/PCR test) revealed increased anti-viral IgM, IgG and IgA to be associated with disease
142 severity, as has been reported⁷ (Fig. 1C and S1D-E, Table S1). Severe disease was most
143 strongly associated with virus-specific IgA, suggestive of mucosal pathology. We did not
144 observe an association between ICU or IL-6 status and IgM levels, supporting that levels of
145 the cytokine and IgA mark a more severe clinical course of COVID-19 (Fig. S1D). Anti-RBD
146 IgA responses were slightly lower in non-hospitalized and hospitalized females compared to
147 males, and trended similarly for S (Fig. S1D and Table S1), consistent with females
148 developing less severe disease⁴.

149

150 Across all PCR+ individuals (sampled up to two months from PCR test), anti-viral IgG levels
151 were maintained, while IgM and IgA decreased, in agreement with their circulating $t_{1/2}$ and
152 viral clearance (Fig. S1D and Table S1). In longitudinal patient samples (sequential sampling
153 of PCR+ individuals in the study) where we observed seroconversion, IgM, IgG and IgA
154 peaked with similar kinetics when all three isotypes developed, although IgA was not always
155 generated in non-hospitalized or hospitalized individuals (Fig 1E), supporting a more diverse
156 antibody response in severe disease.

157

158 To extend these observations, we characterized the *in vitro* virus neutralizing antibody
159 response in PCR+ patients. Using an established pseudotype virus neutralization assay²⁴, we
160 detected neutralizing antibodies in the serum of all SARS-CoV-2 PCR+ individuals screened
161 ($n=48$) (Fig. 1F). Neutralizing responses were not seen in samples before seroconversion or
162 negative controls (Fig. 1E and F). A large range of neutralizing ID₅₀ titers was apparent, with
163 binding and neutralization being highly correlated (Fig. S1D). In agreement with the binding
164 data, the strongest neutralizing responses were observed in samples from patients in intensive
165 care (g.mean ID₅₀=5,058; 95% CI [2,422 - 10,564]) (Fig 1E).

166

167 In healthy blood donors and pregnant women ($n=1,000$ collected between weeks 17-21 2020
168 – the same time as the patient cohort), who did not have signs or symptoms of COVID-19 for
169 two weeks prior to sampling, and had not been hospitalized for COVID-19, IgG titers varied
170 greatly but were generally lower than hospitalized COVID-19 patients, and were comparable
171 to titers in PCR+ hospital staff ($n=33$) who also had never been hospitalized following
172 infection (Fig. 1G).

173

174 Probabilistic analyses of positivity

175

176 As SARS-CoV-2 results in asymptomatic or mild disease in the majority of cases, and
177 antibody titers decline following peak responses and viral clearance, the correct classification
178 of low titer values is critical to individual and population-level estimates of antibody-
179 positivity for COVID-19. Indeed, several healthy donor test samples screen in this study had
180 optical densities between the 3 and 6 SD cut-offs for both or a single antigen (Fig. 1G),
181 highlighting the problem of assigning case to *low responder* values.

182

183 To further our understanding of the assay boundary, we repeatedly analyzed a large number
184 of historical (SARS-CoV-2-negative) controls (blood donors from the spring of 2019, $n=595$)
185 alongside test samples throughout the study. We considered the spread of negative values
186 critical, since the use of a small and unrepresentative set of controls can lead to an incorrectly
187 set threshold, which can considerably skew the seropositivity estimate. This is illustrated by
188 the random sub-sampling of non-overlapping groups of negative controls, resulting in a 40%
189 difference in the positivity estimate (Fig. 2A). Worryingly, many clinically approved tests use
190 a ratio between a known positive and negative serum calibrator to classify seropositivity²⁵,
191 although we show here that these are highly variable within the population.

192

193 Therefore, to exploit individual titers generated against multiple antigens, we used anti-S and
194 RBD data from PCR+ individuals and negative controls to train probabilistic algorithms to
195 assign likelihood of past infection. To this end, we compared different probabilistic
196 algorithms – logistic regression (LOG), linear discriminant analysis (LDA), linear support
197 vector machines (SVM) and quadratic SVM (SVM2) – suited to ELISA data (Fig. 2B,
198 Materials and Methods). Using ten-fold cross validation and training models on both proteins
199 simultaneously (S and RBD), we found all methods worked well, with sensitivity >98% and

200 specificity >99.6% (Fig. 2D). On these metrics, LDA gave the highest specificity. Logistic
201 regression had similarly high specificity on some folds of the training data, but with higher
202 sensitivity. However, we deliberately considered balanced and unbalanced folds (where
203 case:control ratios varied between folds) and found LOG to show the least consistency across
204 strategies, which reflects that the proportion of cases in a sample directly informs a logistic
205 model's estimated parameters. SVM methods had lower specificity than LDA in the training
206 data, but higher sensitivity.

207

208 The standard methods, calling positives by a fixed number of SD above the mean of negative
209 controls, displayed two extreme behaviors: 3-SD had the highest sensitivity (100%) while 6-
210 SD had the highest specificity, and the lowest sensitivity (Fig. S2A), emphasizing that the
211 number of SD above the mean is a key parameter, but one which is not learnt in any formal
212 data-driven manner. Both SVM and LDA offer linear classification boundaries, but we can
213 see that the probability transition from negative to positive cases is much sharper for LDA
214 (Fig. 2B) – potentially resulting in false negatives when applied to the test data, but giving the
215 model high specificity in the training data under cross-validation. SVM exhibits a softer
216 probability transition around its classification boundary, offering a much more nuanced
217 approach to the points lying in the mid-range of the two proteins. SVM2 creates a nonlinear
218 boundary, but the cross validation suggested that this didn't improve performance relative to
219 linear SVM.

220

221 Given these results, we chose to create ensemble learners, which were unweighted averages
222 of SVM (linear) or SVM2 (quadratic) and LDA (*ENS* and *ENS2*, respectively), as well as a
223 LOG-LDA learner, to balance the benefits of each approach. The ensemble learners seemed
224 to combine the benefits of their parent methods (Fig. S2A). Test data points in the lower right
225 region of each plot are the hardest to classify due to the relative scarcity of observations in
226 this region in the training dataset and *ENS* (SVM-LDA) showed the greatest uncertainty in
227 these regions, appropriately. Given these results, we chose to use *ENS* (SVM-LDA), with an
228 average sensitivity >99.1% and specificity >99.8%, to analyze test data. When applied to the
229 serology data, the output of *ENS* is the probability of each sample being antibody-positive.

230

231 In healthy donor test data, the *ENS* learner estimated 7.8% (95% CI [4.8-12.5]) positivity in
232 samples collected in week 21 of 2020 (Fig. S2B, Table S2). This is in contrast to the SD
233 thresholding, which identified 12% and 10% positivity for S and RBD, respectively, at 3 SD,

234 and 8% and 7.5, respectively, at 6 SD (Table S2). Therefore, apart from providing more
235 accurate population-level estimates – critical to seroprevalence studies, where we have
236 applied these and related tools in a large cohort²⁶ – these methods have the potential to
237 provide more nuanced information about titers to an individual after an antibody test. For
238 example, test samples with a 30-60% chance of being antibody positive (Fig. S2B) can be
239 targeted for further investigation or help inform vaccine boosting, as antibody titers decline
240 over time from peak responses. Moreover, such tools are applicable to other clinical metrics
241 where a continuous scale is dichotomized and code for implementation is freely available via
242 our online repositories.
243

244 **Discussion**

245

246 Benefitting from a robust antibody test developed alongside a diagnostic clinical laboratory
247 responsible for monitoring sero-reactivity during the pandemic, we profiled SARS-CoV-2
248 antibody responses in three cohorts of clinical interest. COVID-19 patients receiving
249 intensive care showed the highest anti-viral Ab titers, developing augmented serum IgA and
250 IL-6 with worsening disease and more advanced respiratory and/or gastrointestinal pathology.
251 These results support the use of cytokine and isotype-level measures for patient
252 management²³.

253

254 Importantly, our neutralization data illustrated that nearly all SARS-CoV-2 PCR+ individuals
255 developed neutralizing antibodies capable of preventing S-mediated cell entry, albeit at
256 different titers. These data support that SARS-CoV-2 infection generates a functional B cell
257 response in the majority of people⁸ and serve as a useful comparator to titers engendered by
258 vaccination. Indeed, the first generation of mRNA vaccines have been reported to generate
259 neutralizing titers comparable to samples from individuals with mild infection in our
260 study^{9,10}.

261

262 Outside of the severe disease setting, it is critical to accurately determine who and how many
263 people have seroconverted for clinical and epidemiological reasons. However, this is
264 complicated by low titer values, which in some cases – and increasingly with time since
265 exposure or vaccination^{5,27} – can overlap outlier values among negative control samples.
266 Test samples with true low anti-viral titers fall into this range, highlighting the need to better
267 understand the assay boundary. To improve upon strictly thresholding the assay, we
268 developed probabilistic approaches for ELISA data that characterized the uncertainty in
269 individual measures. These approaches provide more statistically sound measurements at the
270 level of cohorts and the potential to communicate more nuanced information to individual
271 patients – although the communication of probability needs to be approached with care to
272 ensure what is described matches what an individual interprets. Furthermore, such
273 approaches will aid the analysis of data from assay platforms measuring the responses to
274 multiple antigens; longitudinal studies of the duration of immunity after SARS-CoV-2 spike-
275 based vaccines and natural infection; and facilitate comparison of responses in different
276 cohorts.

277 **Materials and methods**

278

279 Human samples and ethical declaration

280 Samples from PCR+ individuals and admitted COVID-19 patients ($n=105$) were collected by
281 the attending clinicians and processed through the Departments of Medicine and Clinical
282 Microbiology at the Karolinska University Hospital. Samples were used in accordance with
283 approval by the Swedish Ethical Review Authority (registration no. 2020-02811). All
284 personal identifiers were pseudo-anonymized, and all clinical feature data were blinded to the
285 researchers carrying out experiments until data generation was complete. PCR testing for
286 SARS-CoV-2 RNA was by nasopharyngeal swab or upper respiratory tract sampling at
287 Karolinska University Hospital. As viral RNA levels were determined using different qPCR
288 platforms (with the same reported sensitivity and specificity) between participants, we did not
289 analyze these alongside other features. PCR+ individuals ($n=105$) were questioned about the
290 date of symptom onset at their initial consultation and followed-up for serology during their
291 care, up to 2 months post-diagnosis. Serum from SARS-CoV-2 PCR+ individuals was
292 collected 6-61 days post-test, with the median time from symptom onset to PCR being 5
293 days. In addition, longitudinal samples from 10 of these patients were collected to monitor
294 seroconversion and isotype persistence.

295

296 Hospital workers at Karolinska University Hospital were invited to test for the presence of
297 SARS-CoV-2 RNA in throat swabs in April 2020 and virus-specific IgG in serum in July
298 2020. We screened 33 PCR+ individuals to provide additional training data for ML
299 approaches. All participants provided written informed consent. The study was approved by
300 the National Ethical Review Agency of Sweden (2020-01620) and the work was performed
301 accordingly.

302

303 Anonymized samples from blood donors ($n=100/\text{week}$) and pregnant women ($n=100/\text{week}$)
304 were randomly selected from their respective pools by the department of Clinical
305 Microbiology, Karolinska University Hospital. No metadata, such as age or sex information
306 were available for these samples in this study. Pregnant women were sampled as part of
307 routine for infectious diseases screening during the first trimester of pregnancy. Blood donors
308 ($n=595$) collected through the same channels a year previously were randomly selected for
309 use as negative controls. Serum samples from individuals testing PCR+ for endemic
310 coronaviruses, 229E, HKU1, NL63, OC43 ($n=20$, ECV+) in the prior 2-6 months, were used

311 as additional negative controls. The use of study samples was approved by the Swedish
312 Ethical Review Authority (registration no. 2020-01807). Stockholm County death and
313 Swedish mortality data was sourced from the ECDC and the Swedish Public Health Agency,
314 respectively. Study samples are defined in Table 1.

315

316 Serum sample processing

317 Blood samples were collected by the attending clinical team and serum isolated by the
318 department of Clinical Microbiology. Samples were anonymized, barcoded and stored at -
319 20°C until use. Serum samples were not heat-inactivated for ELISA protocols but were heat-
320 inactivated at 56°C for 60 min for neutralization experiments.

321

322 SARS-CoV-2 antigen generation

323 The plasmid for expression of the SARS-CoV-2 prefusion-stabilized spike ectodomain with a
324 C-terminal T4 fibrin trimerization motif was obtained from¹⁴. The plasmid was used to
325 transiently transfect FreeStyle 293F cells using FreeStyle MAX reagent (Thermo Fisher
326 Scientific). The ectodomain was purified from filtered supernatant on Streptactin XT resin
327 (IBA Lifesciences), followed by size-exclusion chromatography on a Superdex 200 in 5 mM
328 Tris pH 8, 200 mM NaCl.

329

330 The RBD domain (RVQ – QFG) of SARS-CoV-2 was cloned upstream of a Sortase A
331 recognition site (LPETG) and a 6xHIS tag, and expressed in 293F cells as described above.
332 RBD-HIS was purified from filtered supernatant on His-Pur Ni-NTA resin (Thermo Fisher
333 Scientific), followed by size-exclusion chromatography on a Superdex 200. The nucleocapsid
334 was purchased from Sino Biological.

335

336 Anti-SARS-CoV-2 ELISA

337 96-well ELISA plates (Nunc MaxiSorp) were coated with SARS-CoV-2 S trimers, RBD or
338 nucleocapsid (100 µl of 1 ng/µl) in PBS overnight at 4°C. Plates were washed six times with
339 PBS-Tween-20 (0.05%) and blocked using PBS-5% no-fat milk. Human serum samples were
340 thawed at room temperature, diluted (1:100 unless otherwise indicated), and incubated in
341 blocking buffer for 1h (with vortexing) before plating. Serum samples were incubated
342 overnight at 4°C before washing, as before. Secondary HRP-conjugated anti-human
343 antibodies were diluted in blocking buffer and incubated with samples for 1 hour at room
344 temperature. Plates were washed a final time before development with TMB Stabilized

345 Chromogen (Invitrogen). The reaction was stopped using 1M sulphuric acid and optical
346 density (OD) values were measured at 450 nm using an Asys Expert 96 ELISA reader
347 (Biochrom Ltd.). Secondary antibodies (all from Southern Biotech) and dilutions used: goat
348 anti-human IgG (2014-05) at 1:10,000; goat anti-human IgM (2020-05) at 1:1000; goat anti-
349 human IgA (2050-05) at 1:6,000. All assays of the same antigen and isotype were developed
350 for their fixed time and samples were randomized and run together on the same day when
351 comparing binding between PCR+ individuals. Negative control samples were run alongside
352 test samples in all assays and raw data were log transformed for statistical analyses.

353

354 *In vitro* virus neutralisation assay

355 Pseudotyped viruses were generated by the co-transfection of HEK293T cells with plasmids
356 encoding the SARS-CoV-2 spike protein harboring an 18 amino acid truncation of the
357 cytoplasmic tail¹⁴; a plasmid encoding firefly luciferase; a lentiviral packaging plasmid
358 (Addgene 8455) using Lipofectamine 3000 (Invitrogen). Media was changed 12-16 hours
359 post-transfection and pseudotyped viruses harvested at 48- and 72-hours, filtered through a
360 0.45 µm filter and stored at -80°C until use. Pseudotyped neutralisation assays were adapted
361 from protocols validated to characterize the neutralization of HIV, but with the use of
362 HEK293T-ACE2 cells. Briefly, pseudotyped viruses sufficient to generate ~100,000 RLUs
363 were incubated with serial dilutions of heat-inactivated serum for 60 min at 37°C.
364 Approximately 15,000 HEK293T-ACE2 cells were then added to each well and the plates
365 incubated at 37°C for 48 hours. Luminescence was measured using Bright-Glo (Promega)
366 according to the manufacturer's instructions on a GM-2000 luminometer (Promega) with an
367 integration time of 0.3s. The limit of detection was at a 1:45 serum dilution.

368

369 IL-6 cytometric bead array

370 Serum IL-6 levels were measured in a subset of PCR+ serum samples ($n=64$) using an
371 enhanced sensitivity cytometric bead array against human IL-6 from BD Biosciences (Cat #
372 561512). Protocols were carried out according to the manufacturer's recommendations and
373 data acquired using a BD Celesta flow cytometer.

374

375 Statistical analysis of SARS-CoV-2 PCR+ data

376 All univariate comparisons were performed using non-parametric analyses (Kruskal-Wallis,
377 stratified Mann-Whitney, hypergeometric exact tests and Spearman rank correlation), as
378 indicated, while multivariate comparisons were performed using linear regression of log

379 transformed measures and Wald tests. For multivariate tests, all biochemical measures (IL-6,
380 PSV ID50 neut., IgG, IgA, IgM) were log transformed to improve the symmetry of the
381 distribution. As “days since first symptom” and ”days since PCR+ test” are highly correlated,
382 we cannot include both in any single analysis. Instead, we show results for one, then the other
383 (Supp. Table 1).

384

385 Probabilistic algorithms for classifying antibody positivity

386 Prior to analysis, each sample OD was standardized by dividing by the mean OD of ”no
387 sample controls” on that plate or other plates run on the same day. This resulted in more
388 similar distributions for 2019 blood donor samples with 2020 blood donors and pregnant
389 volunteers, as well as smaller coefficients of variation amongst PCR+ COVID patients for
390 both SPIKE and RBD.

391

392 Our probabilistic learning approach consisted of evaluating different algorithms suited to
393 ELISA data, which we compared through ten-fold cross validation (CV): logistic regression
394 (LOG), linear discriminant analysis (LDA), support vector machines (SVM) with a linear
395 kernel, and quadratic SVM (SVM2). Logistic regression and linear discriminant analysis both
396 model log odds of a sample being case as a linear equation with a resulting linear decision
397 boundary. The difference between the two methods is in how the coefficients for the linear
398 models are estimated from the data. When applied to new data, the output of logistic
399 regression and LDA is the probability of each new sample being a case. Support vector
400 machines is an altogether different approach. We opted for a linear kernel, once again
401 resulting in a linear boundary. SVM constructs a boundary that maximally separates the
402 classes (i.e. the margin between the closest member of any class and the boundary is as wide
403 as possible), hence points lying far away from their respective class boundaries do not play an
404 important role in shaping it. SVM thus puts more weight on points closest to the class
405 boundary, which in our case is far from being clear. Linear SVM has one tuning parameter C ,
406 a cost, with larger values resulting in narrower margins. We tuned C on a vector of values
407 (0.001, 0.01, 0.5, 1, 2, 5, 10) via an internal 5-fold CV with 5 repeats (with the winning
408 parameter used for the final model for the main CV iteration). We also note that the natural
409 output of SVM are class labels rather than class probabilities, so the latter are obtained via the
410 method of Platt²⁸.

411

412 We considered three strategies for cross-validation: i) random: individuals were sampled into
413 folds at random, ii) stratified: individuals were sampled into folds at random, subject to
414 ensuring the balance of cases:controls remained fixed and iii) unbalanced: individuals were
415 sampled into folds such that each fold was deliberately skewed to under or over-represent
416 cases compared to the total sample. We sought a method with performance that was
417 consistently good across all cross-validation sampling schemes, because the true proportion
418 of cases in the test data is unknown, and we want a method that is not overly sensitive to the
419 proportion of cases in the training data. We chose to assess performance using sensitivity and
420 specificity, as well as consistency.

421

422 Given the good performance of all learners (described in the results), we considered the
423 prediction surface associated with each SVM, LDA, SVM-LDA ensemble, and the standard
424 3-SD, 6-SD hard decision boundaries. Note that while methods trained on both proteins can
425 draw decision contours at any angle, SD methods are limited to vertical or horizontal lines.
426 We can see that success, or failure, of the SD cut-offs depends on how many positive and
427 negative cases overlap for a given measure (S or RBD) in the training sample. In the training
428 data the two classes are nearly linearly separable when each protein is considered on its own,
429 which explains good performance of 3-SD and 6-SD thresholds. However, the test data
430 contain many more points in the mid-range of S-RBD, which makes hard cut-offs a
431 problematic choice for classifying test samples.

432

433 We trained the learners on all 733 training samples and used these to predict the probability
434 of anti-SARS-CoV-2 antibodies in blood donors and pregnant volunteers sampled in 2020.
435 We inferred the proportion of the sampled population with positive antibody status each week
436 using multiple imputation. We repeatedly (1,000 times) imputed antibody status for each
437 individual randomly according to the ensemble prediction, and then analyzed each of the
438 1,000 datasets in parallel, combining inference using Rubin's rules, derived for the Wilson
439 binomial proportion confidence interval²⁹.

440

441 **Data and code availability statement**

442

443 Data generated as part of the study, along with custom code for statistical analyses, is openly
444 available via our GitHub repository: <https://github.com/chrlswallace/elisa-paper>.

445

446 **Author contributions**

447

448 GKH and XCD designed the study and wrote the manuscript with input from co-authors. JA,
449 TA, JD, SM, GB, MA and SA provided the study serum samples and clinical information.
450 LH, LPV, AMM, DJS, KCI, BM and GM generated SARS-CoV-2 antigens and pseudotyped
451 viruses. MF and XCD developed the ELISA protocols and XCD generated the data. DJS and
452 BM performed the neutralization assay. CW and NFG executed machine learning approaches
453 and statistical analyses, with input from MCh and BM. MA, SK, PP, MM, JC, MCo and JR
454 carried out wet lab experiments and assisted with data analysis.

455

456 **Acknowledgments**

457

458 We would like to thank the study participants and attending clinical teams. Secondly, we
459 extend our thanks to Björn Reinius, Marc Panas, Julian Stark, Remy M. Muts and Darío Solis
460 Sayago for their input and discussion. Funding for this work was provided by a Distinguished
461 Professor grant from the Swedish Research Council (agreement 2017-00968) and NIH
462 (agreement 400 SUM1A44462-02). CW and NFG are funded by the Wellcome Trust
463 (WT107881) and MRC (MC_UP_1302/5). For the purpose of Open Access, the author has
464 applied a CC BY public copyright licence to any Author Accepted Manuscript version arising
465 from this submission.

466

467 **Conflict of interest**

468

469 The study authors declare no competing interests related to the work.

470

471 **References**

472

- 473 1. Jonsson, S. *et al.* Identification of sequence variants influencing immunoglobulin
474 levels. *Nat. Genet.* (2017) doi:10.1038/ng.3897.
- 475 2. Robbiani, D. F. *et al.* Convergent Antibody Responses to SARS-CoV-2 Infection in
476 Convalescent Individuals. *Nature* (2020) doi:10.1101/2020.05.13.092619.
- 477 3. Dan, J. M. *et al.* Immunological memory to SARS-CoV-2 assessed for greater than
478 eight months after infection. *bioRxiv* (2020).
- 479 4. Shrock, E. *et al.* Viral epitope profiling of COVID-19 patients reveals cross-reactivity

- 480 and correlates of severity. *Science* (80-.). (2020) doi:10.1126/science.abd4250.
- 481 5. Long, Q. X. *et al.* Clinical and immunological assessment of asymptomatic SARS-
482 CoV-2 infections. *Nat. Med.* (2020) doi:10.1038/s41591-020-0965-6.
- 483 6. Sekine, T. *et al.* Robust T Cell Immunity in Convalescent Individuals with
484 Asymptomatic or Mild COVID-19. *Cell* (2020) doi:10.1016/j.cell.2020.08.017.
- 485 7. Cervia, C. *et al.* Systemic and mucosal antibody responses specific to SARS-CoV-2
486 during mild versus severe COVID-19. *J. Allergy Clin. Immunol.* (2020)
487 doi:10.1016/j.jaci.2020.10.040.
- 488 8. Gudbjartsson, D. F. *et al.* Spread of SARS-CoV-2 in the Icelandic Population. *N. Engl.*
489 *J. Med.* (2020) doi:10.1056/nejmoa2006100.
- 490 9. Mulligan, M. J. *et al.* Phase I/II study of COVID-19 RNA vaccine BNT162b1 in
491 adults. *Nature* (2020) doi:10.1038/s41586-020-2639-4.
- 492 10. Jackson, L. A. *et al.* An mRNA Vaccine against SARS-CoV-2 — Preliminary Report.
493 *N. Engl. J. Med.* (2020) doi:10.1056/nejmoa2022483.
- 494 11. Röltgen, K. *et al.* Defining the features and duration of antibody responses to SARS-
495 CoV-2 infection associated with disease severity and outcome. *2Science Immunol.* **5**,
496 DOI: 10.1126/sciimmunol.abe0240 (2020).
- 497 12. Altman, D. G. & Royston, P. The cost of dichotomising continuous variables. *British*
498 *Medical Journal* (2006) doi:10.1136/bmj.332.7549.1080.
- 499 13. MacCallum, R. C., Zhang, S., Preacher, K. J. & Rucker, D. D. On the practice of
500 dichotomization of quantitative variables. *Psychol. Methods* (2002) doi:10.1037//1082-
501 989x.7.1.19.
- 502 14. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion
503 conformation. *Science* (80-.). (2020) doi:10.1126/science.aax0902.
- 504 15. Hanke, L. *et al.* An alpaca nanobody neutralizes SARS-CoV-2 by blocking receptor
505 interaction. *Nat. Commun.* (2020) doi:10.1038/s41467-020-18174-5.
- 506 16. Ng, K. W. *et al.* Preexisting and de novo humoral immunity to SARS-CoV-2 in
507 humans. *Science* (80-.). (2020) doi:10.1126/science.abe1107.
- 508 17. Herzenberg, L. A. & Herzenberg, L. A. Toward a layered immune system. *Cell* (1989)
509 doi:10.1016/0092-8674(89)90748-4.
- 510 18. Long, Q. X. *et al.* Antibody responses to SARS-CoV-2 in patients with COVID-19.
511 *Nat. Med.* (2020) doi:10.1038/s41591-020-0897-1.
- 512 19. Eto, D. *et al.* IL-21 and IL-6 are critical for different aspects of B cell immunity and
513 redundantly induce optimal follicular helper CD4 T cell (Tfh) differentiation. *PLoS*

- 514 *One* **6**, e17739 (2011).
- 515 20. Dienz, O. *et al.* The induction of antibody production by IL-6 is indirectly mediated by
516 IL-21 produced by CD4 + T cells. *J. Exp. Med.* (2009) doi:10.1084/jem.20081571.
- 517 21. Maeda, K., Mehta, H., Drevets, D. A. & Coggeshall, K. M. IL-6 increases B-cell IgG
518 production in a feed-forward proinflammatory mechanism to skew hematopoiesis and
519 elevate myeloid production. *Blood* (2010) doi:10.1182/blood-2009-07-230631.
- 520 22. Beagley, K. W. *et al.* Interleukins and IgA synthesis. Human and murine interleukin 6
521 induce high rate IgA secretion in IgA-committed B cells. *J. Exp. Med.* (1989)
522 doi:10.1084/jem.169.6.2133.
- 523 23. Del Valle, D. M. *et al.* An inflammatory cytokine signature predicts COVID-19
524 severity and survival. *Nat. Med.* (2020) doi:10.1038/s41591-020-1051-9.
- 525 24. Bartosch, B. *et al.* In vitro assay for neutralizing antibody to hepatitis C virus:
526 Evidence for broadly conserved neutralization epitopes. *Proc. Natl. Acad. Sci. U. S. A.*
527 (2003) doi:10.1073/pnas.2335981100.
- 528 25. Beavis, K. G. *et al.* Evaluation of the EUROIMMUN Anti-SARS-CoV-2 ELISA
529 Assay for detection of IgA and IgG antibodies. *J. Clin. Virol.* (2020)
530 doi:10.1016/j.jcv.2020.104468.
- 531 26. Dopico, X. *et al.* Seropositivity in blood donors and pregnant women during 9-months
532 of SARS-CoV-2 transmission in Stockholm, Sweden. *MedRxiv* <https://doi.org/10.1101/2020.07.09.20148429>, (2020).
- 533 27. Seow, J. *et al.* Longitudinal evaluation and decline of antibody responses in SARS-
534 CoV-2 infection. *Nat. Microbiol.* (2020) doi:10.1101/2020.07.09.20148429.
- 535 28. Platt, J. & others. Probabilistic outputs for support vector machines and comparisons to
536 regularized likelihood methods. *Adv. large margin Classif.* (1999).
- 537 29. Lott, A. & Reiter, J. P. Wilson Confidence Intervals for Binomial Proportions With
538 Multiple Imputation for Missing Data. *Am. Stat.* (2020)
539 doi:10.1080/00031305.2018.1473796.
- 540

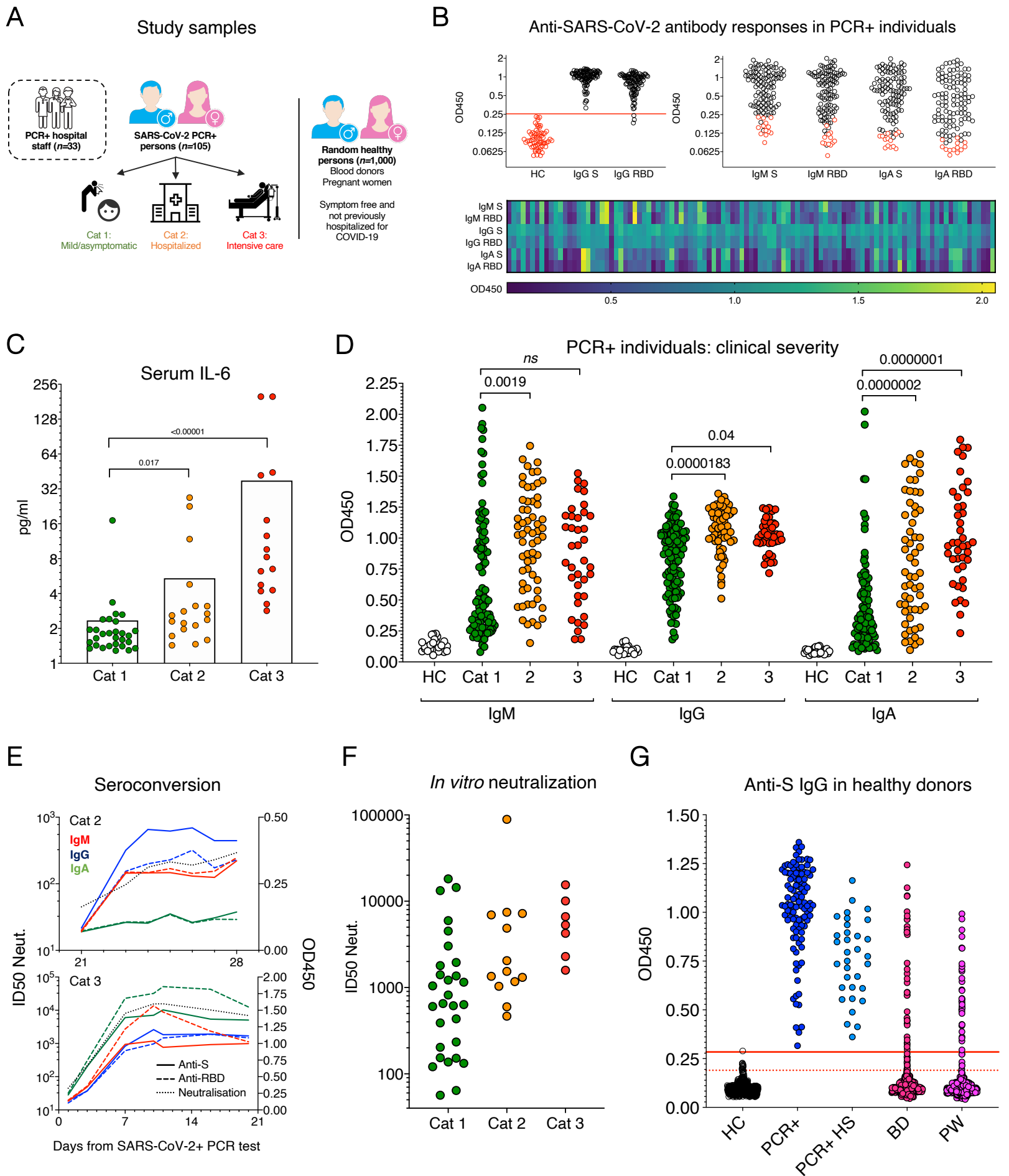
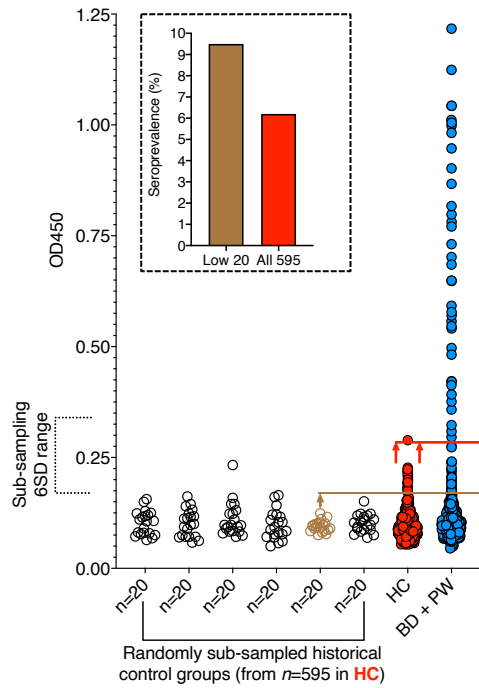


Figure 1: Anti-SARS-CoV-2 Ab phenotypes in COVID-19 patients, PCR+ individuals, blood donors and pregnant women.

(A) Study samples. **(B)** Raw optical density (450nm) for anti-S and -RBD IgG (first graph) IgM and IgA responses (second graph) in SARS-CoV-2 PCR+ individuals ($n=105$). A 6SD cut-off based on $n=595$ control (HC) values is shown for the IgG assay and a small number of HC samples are shown (red) for the IgM and IgA assays. Individual responses for the three isotypes are shown by a heatmap. **(C)** Circulating IL-6 levels in serum are associated with disease severity. **(D)** Anti-viral antibody levels are associated with disease severity, most pronounced for anti-viral IgA. Anti-S and RBD responses are graphed together. P values for S are shown. **(E)** Two discordant longitudinal profiles of seroconversion and neutralisation capacity are shown in hospitalized COVID-19 patients. **(F)** *In vitro* pseudotyped virus neutralization ID_{50} titers are associated with disease severity, with the highest titers observed in Cat 3 (ICU) patients. $n=48$ SARS-CoV-2 PCR+ individuals were analyzed in duplicate. **(G)** Comparison of anti-S IgG levels between PCR+ individuals ($n=105$), PCR+ hospital staff (PCR+ HS, $n=33$), blood donors (BD, $n=500$) and pregnant women (PW, $n=500$). 3 and 6 SD cut-offs are shown by red lines.

A



B

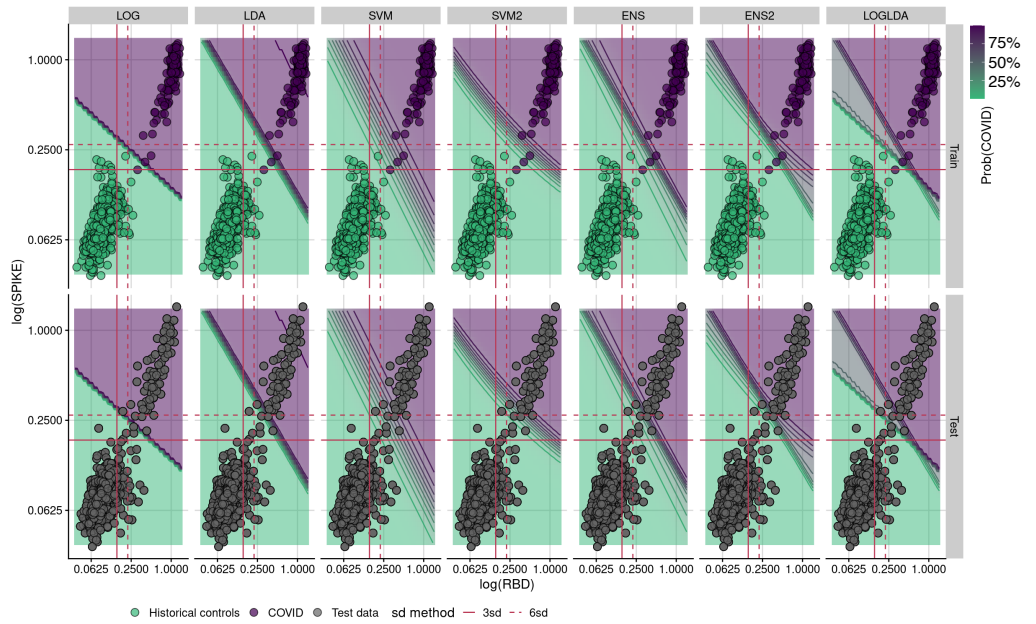


Figure 2: Probability-based seropositivity estimates in blood donors and pregnant women

(A) Random sub-sampling of non-overlapping negative controls illustrates how the range of negative control (C) values can influence a conventional test cut-off, here 6 SD from the mean of the respective C groups. In the test data, depending on the control values used to set the test threshold for positivity, SP estimates varied 40%. Blood donor and pregnant women sample values are used as an example. Anti-S IgG values are shown. **(B)** Comparison of probabilistic algorithms suited to ELISA measurements. Logistic regression (LOG), linear discriminant analysis (LDA), support vector machines (SVM) and quadratic SVM (SVM2). Learners were trained using anti-S and RBD IgG data from 595 negative control values and 138 SARS-CoV-2 PCR+ individual samples. Ensemble (ENS) learners were generated from the output of SVM-LDA, SVM2-LDA and LOG-LDA.

Table 1 – Study samples

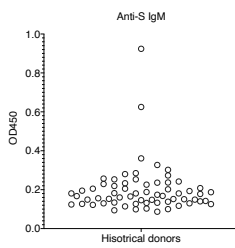
SARS-CoV-2 PCR+ individuals[§]	<i>n</i> =105
Females	44 (41.9%)
Males	61 (58.1%)
Median age (years)	53.0 (49-61)
Females	51.5 (48-56.2)
Males	55.0 (49-63)
Non-hospitalized (<i>n</i> =)	53
Females, males	28, 25
Hospitalized patients (<i>n</i> =)	31
Females, males	12, 17
Intensive care (ICU) patients (<i>n</i> =)	21
Females, males	3, 17
SARS-CoV2+ PCR (<i>n</i> =)	105
Sample collection dates	March-May 2020
SARS-CoV-2 PCR+ KI hospital staff *	<i>n</i> =33
Sample collection dates	July 2020
Blood donors *	<i>n</i> =500
Sample collection dates	Weeks 17-21 (March-May) 2020
Pregnant women *	<i>n</i> =500
Sample collection dates	Weeks 17-21 (March-May) 2020
Historical blood donors *	<i>n</i> =595
Sample collection dates	March-June 2019
ECV+ donors *	<i>n</i> =20
Sample collection dates	July-December 2019
[§] Under the care of Karolinska University Hospital	
*No additional metadata available for any samples	

A

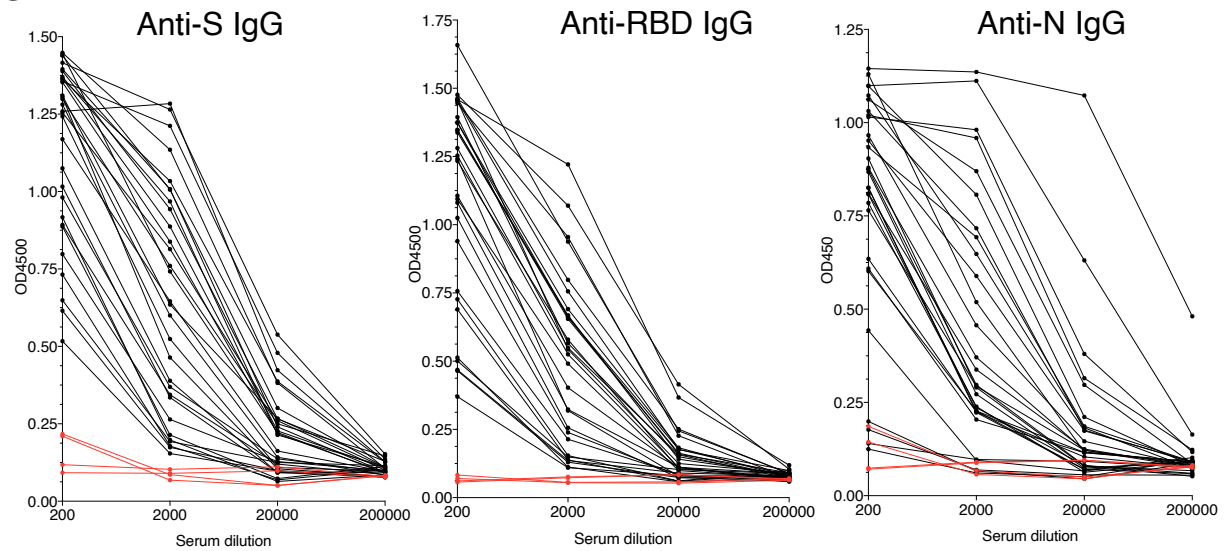
Samples used for assay development:

- 100 historical controls
- 38 PCR+ individuals
- 100 blood donor samples
- 20 individuals PCR+ for endemic coronaviruses in the past six months

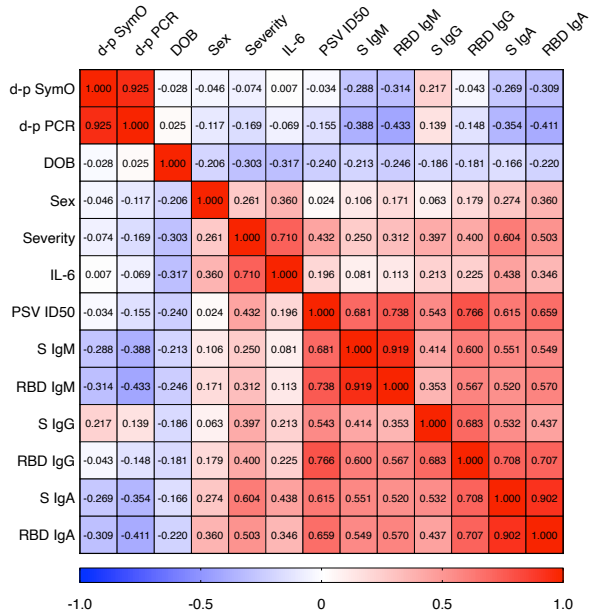
B



C



D



E

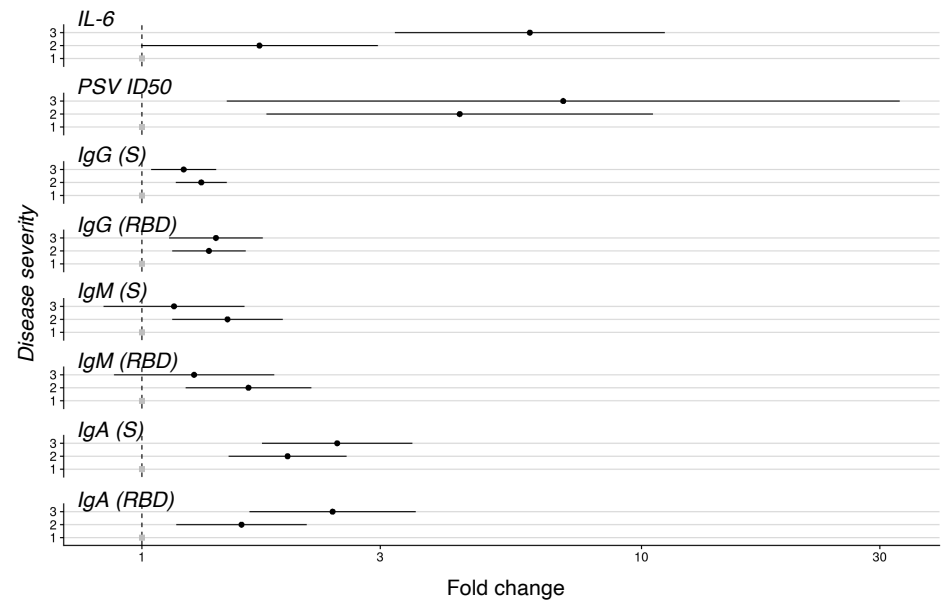
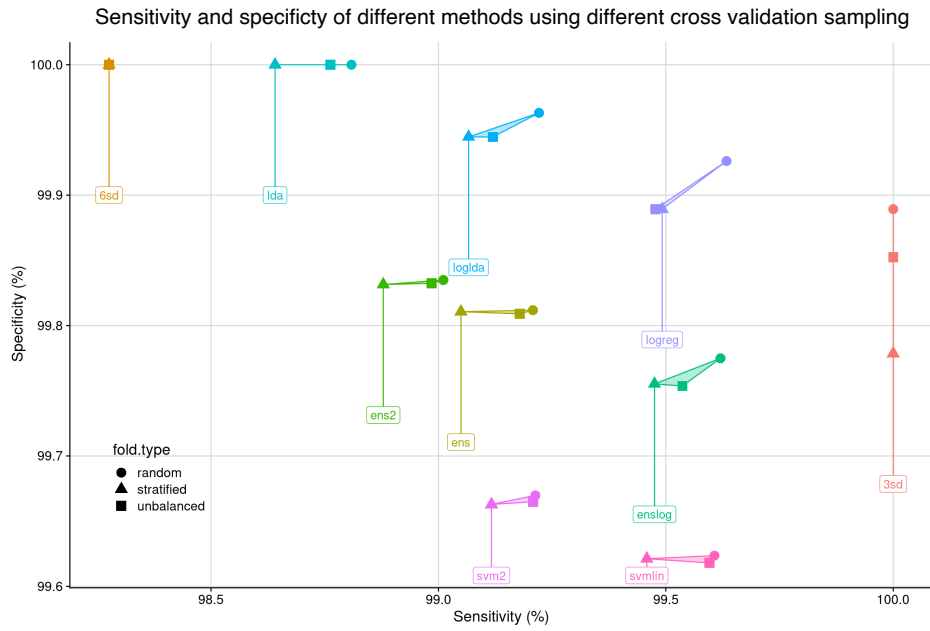


Figure S1: Antibody phenotypes in PCR+ individuals.

(A) Study samples used for assay development. (B) Anti-S IgM reactivity observed in a random subset of historical controls. Binding was confirmed in these samples in an independent experiment. No reproducible IgG reactivity to S trimers of the RBD was observed across all historical controls in the study. (C) Serial dilution of $n=30$ random PCR+ individuals. ECV+ ($n=4$) controls are shown in red. (D) Spearman's rank correlation of PCR+ dataset features and antibody levels. DOB - *date of birth*; d-p SymO - *days post-symptom onset*; d-p PCR - *days post SARS-CoV-2+ PCR*; PSV ID50 - *neutralizing titer*. (E) Adjusted fold-change compared to Category 1 PCR+ individuals. The effects of age (DOB), sex, days from PCR test were considered.

A



B

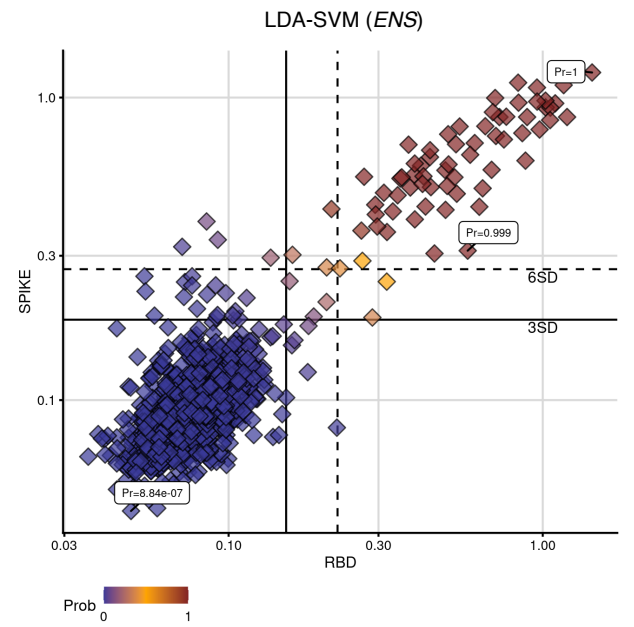


Figure S2: Performance of different probabilistic approaches.

(A) Comparisons of specificity and sensitivity for the different probabilistic methods (and 3 and 6 SD thresholding) using different cross-validation strategies. (B) *ENS* probabilities when applied to healthy donor test data, providing a highly sensitive, specific and consistent multi-dimensional solution to the problem of low responders, and assigning each data point a probability of being positive.