

35 **One Sentence Summary:**

36 Eight S protein-derived peptides, particularly S2-78 (aa 1148-1159), are of high
37 performance for diagnosis of COVID-19 as well as discrimination of other coronaviruses.

38 **Abstract**

39 Serological test plays an essential role in monitoring and combating COVID-19
40 pandemic. Recombinant spike protein (S protein), especially S1 protein is one of the
41 major reagents for serological tests. However, the high cost in production of S protein,
42 and the possible cross-reactivity with other human coronaviruses poses unneglectable
43 challenges. Taking advantage of a peptide microarray of full spike protein coverage, we
44 analyzed 2,434 sera from 858 COVID-19 patients, sera from 63 asymptomatic patients
45 and 610 controls collected from multiple clinical centers. Based on the results of the
46 peptide microarray, we identified several S protein derived 12-mer peptides that have
47 high diagnosis performance. Particularly, for monitoring IgG response, one peptide (aa
48 1148-1159 or S2-78) has a comparable sensitivity (95.5%, 95% CI 93.7-96.9%) and
49 specificity (96.7%, 95% CI 94.8-98.0%) to that of S1 protein for detection of both
50 COVID-19 patients and asymptomatic infections. Furthermore, the performance of S2-78
51 IgG for diagnosis was successfully validated by ELISA with an independent sample
52 cohort. By combining S2-78/ S1 with other peptides, a two-step strategy was proposed to
53 ensure both the sensitivity and specificity of S protein based serological assay. The
54 peptide/s identified in this study could be applied independently or in combination with
55 S1 protein for accurate, affordable, and accessible COVID-19 diagnosis.

56

57 COVID-19 is caused by SARS-CoV-2[1,2] and is a global pandemic. By August 7, 2020,
58 18,982,658 cases were diagnosed and 712,266 lives were claimed
59 (<https://coronavirus.jhu.edu/map.html>)[3]. To put the pandemic under control, one of the
60 essential options is to perform fast, reliable and affordable diagnosis. Although nucleic
61 acid test (NAT) is the reference standard for diagnosing COVID-19 with high sensitivity
62 and accuracy, however, false negative results were commonly observed[4,5]. The
63 immunological/ serological test, for example, monitoring the SARS-CoV-2 specific IgG
64 and IgM responses, provides important information to improve the accuracy of
65 diagnosis[4,5]. In addition, serological test is suitable for population screening in high
66 risk regions or among close-contact people, as well as surveillance of the pandemic
67 spreading or assess the infection rate of general population[6–8]. Moreover, antibody
68 response is reported to associated with disease severity or clinical outcomes[9,10].

69

70 S protein is the preferential antigen for serological assay. The key reagent of the S
71 protein based serological assay is the recombinant protein. However, the production of
72 the S protein is difficult and costly[11]. The inconsistency among different manufacturers
73 or even batches might contribute to the variability of commercial assays with the same
74 antigen[7,12]. Limited production capacity and high cost of recombinant protein
75 preparation is the bottleneck, particularly for remote regions or poor countries. In
76 addition, it should be concerned that the cross-reactivity of infections of other human
77 coronaviruses may cause false positive results, especially for those four common cold
78 causing coronaviruses, *i.e.*, HCoV-OC43, HKU1, NL63 and 229E, which are circulating
79 in population[4,11,13]. It was reported that S1, compared with full length S protein,
80 exhibit less cross-reactivity due to the less similarity of S1 subunit among the human
81 coronaviruses than that of S2[4]. To develop highly specific serological test, more efforts
82 are needed to identify sections of S protein that are highly immunogenic and less
83 homologous to other related coronaviruses[6,11].

84

85 Spike protein derived peptides that can elicit antibodies in COVID-19 patients has
86 been reported in several studies[14–16], including one of our previous work on epitope
87 mapping with a small sample cohort[17]. For instance, antibody against S2-78 (aa

88 1148-1159) and S2-22 (aa 812-823) have high positive rates in COVID-19 patients.
89 However, whether those peptides are suitable for diagnosis is still unknown. Herein, to
90 fully evaluate the diagnostic value of the S protein derived peptides, a total of four
91 cohorts of sera, consisted of 2,434 sera from 858 COVID-19 patients, sera from 63
92 asymptomatic patients and 610 controls were used. Eight peptides were verified to have
93 high potential for diagnosis, particularly, one peptide, S2-78 has a comparable diagnosis
94 performance as that of S1 protein for COVID-19 patients and asymptomatic infections.
95 By combining S2-78 IgG/ S1 IgG with other peptide/s, we purposed a two-step strategy
96 that can ensure both sensitive and specific diagnosis for COVID-19.

97

98 **Results**

99

100 *Four independent cohorts of samples were collected and designed*

101 To fully evaluate the diagnostic potential of the spike derived peptides, sera from four
102 cohorts of COVID-19 patients and controls from multiple medical centers of China were
103 collected (**Table 1**). 1) Cohort 1 consists of 55 sera from convalescent COVID-19
104 patients and 18 controls[17]; 2) Cohort 2 includes 2,360 sera from 784 in-hospital
105 COVID-19 patients and 542 sera from a variety of controls. To accurately evaluate the
106 peptides for diagnosis, for each patient, one serum sample that collected at least 21 days
107 after onset of symptoms was selected according to the suggestion of WHO (World Health
108 Organization) for antibody laboratory test[18]. As a result, in total, 729 sera were selected.
109 The control groups include two types of samples. The first type is sera collected from
110 hospitals, including sera from healthy people (n=92), upper respiratory infections (URI,
111 n=104), patients with autoimmune diseases (AID, n=120), lung cancer patients (n=41)
112 and patients with other diseases (n=112) that consist of cardiovascular or cerebrovascular
113 diseases (34.2%), diabetes (9%), non-lung cancers (7.2%) and others. The second type is
114 negative references provided by National Institutes for Food and Drug Control, including
115 N11-N25 of National Reference Panel for 2019-nCoV (SARS-CoV-2) IgG Antibody
116 Detection Kit (370096-202001) and other identified controls. 3) Cohort 3 includes 19
117 COVID-19 patients from another hospital and 50 healthy controls. 4) Cohort 4 consists of
118 asymptomatic patients defined as positive either on NAT (nucleic acid test) or antibody

119 test conducted by a commercial assay, and with a SARS-CoV-2 exposure history (see
120 **methods**).

121

122 *Peptide S2-78 and several other peptides exhibit high diagnostic values*

123 To identify which section of S protein has diagnostic value, it is necessary to survey the
124 entire protein on a systematic way. We took advantage of a previously constructed
125 peptide microarray which has full coverage of S protein[17], analyzed 55 convalescent
126 sera of COVID-19 patients along with 18 control sera (Cohort 1, **Table 1**). Overall,
127 significant bindings of both IgG and IgM were observed in patient group, while the
128 signals were low in the control group (**Fig. 1a**). Several peptides, *e. g.*, S2-78 and S1-93
129 exhibit strong IgG antibody bindings and high response frequencies in patients, we define
130 these peptides as “significant” peptides, these peptides may have diagnostic values (**Fig.**
131 **1b**). To test whether the peptide specific IgG antibody responses are concentration
132 dependent, 3x serially diluted peptides (0.9, 0.3 and 0.1 mg/mL) were printed and
133 immobilized on the microarray. As expected, the averaged signals of the patient group,
134 but not the control group, are proportional to the concentrations of the peptides (**Fig. 1c**).

135 To extensively evaluate the peptides for diagnostic application, a larger cohort (Cohort
136 2, **Table 1**) of samples were screened by a revised peptide microarray that contains only
137 one peptide concentration (0.3 mg/mL) for high-throughput analysis. To ensure the data
138 generated from different microarrays are comparable, we prepared a positive reference
139 sample by pooling 50 randomly selected patient sera. This reference sample was then
140 tested on all the microarrays for normalization (see **methods**). Consistent results were
141 achieved for most of the peptides. AUC (area under curve) values of IgG or IgM for each
142 peptide were calculated. Eight peptides, *i. e.*, S2-78, S1-97, S1-93, S1-101, S1-111, S2-97,
143 S1-105 and S2-22 are of high performance, the AUCs of IgG or IgM against these
144 peptides for both Cohort 1 and 2 are above 0.85. Specifically, for Cohort 2 the AUC
145 values with 95% CI (confidential intervals) for S2-78, S1-97, S1-93, S1-101, S1-111,
146 S2-97, S1-105 and S2-22 were 0.99 (0.986-0.995), 0.954 (0.942-0.965), 0.934
147 (0.92-0.948), 0.932 (0.917-0.947), 0.929 (0.915-0.943), 0.922 (0.907-0.938), 0.909
148 (0.893-0.926), and 0.866 (0.846-0.886), respectively (**Fig. 2a**). For IgM, only S2-78 has a
149 AUC >0.85, *i. e.*, 0.953 (0.941-0.964) (**Fig. 2a**). As expected, IgG and IgM for both N

150 protein and S1 protein exhibited high performance. We further examined the antibody
151 responses in different groups for each peptide (**Fig. 2b-e, Fig. S1**). Consistently, signals
152 for COVID-19 group are significantly higher than that of the negative samples in all
153 groups. It is noted that the signal intensity for S1 IgG is generally higher than that of any
154 single peptide in the group of COVID-19 patients, this may because there are multiple
155 antibody binding sites on S1 protein. However, slightly higher signal is also observed in
156 control group for protein antigens, demonstrating non-specific binding while it is largely
157 eliminated for synthetic peptides. It is suggested that more sensitive detection platforms
158 or higher antigen concentration might improve the performance of peptides for diagnosis.

159

160 *The diagnostic performance of S2-78 IgG is comparable to that of S1 IgG for COVID-19*
161 *patients*

162 We next focused on S2-78, the peptide of best performance for detection. Optimal
163 Youden index of ROC (receiver operating characteristic) curve was used to set the cutoff
164 value. The specificity, sensitivity and overall accuracy (95% CI) of S2-78 IgG for
165 detection of COVID-19 are 96.7 (94.8-98.0%), 95.5% (93.7-96.9%) and 96% (94.8-97%),
166 respectively, which are slightly lower than that of S1 IgG (**Table 2**). Since serological
167 test is essential for population screening, we calculated the PPV (positive predict value)
168 and NPV (negative predict value) of two assumed prevalence rates. One is 0.04 for
169 general population originated from the situations of Wuhan, China[19] and
170 Netherlands[20]. The other is 0.5 for a high risk population[7]. For prevalence rate of 0.5,
171 both PPV and NPV of S2-78 IgG is similar to that of S1 IgG, however, for prevalence
172 rate of 0.04, the PPV is only 54.7%, although the NPV is extremely high, suggesting the
173 antibody detection of S2-78 could effectively exclude negative ones but may generate
174 high false positive rate at low rate. However, for low prevalence rate, since the number of
175 real positive are very low, it might be acceptable to perform additional test by using other
176 antigens to improve the overall performance.

177

178 We next investigated the consistency between S1 IgG and S2-78 IgG (**Fig. 3a, b**). The
179 overall consistencies for COVID-19 group and Control group are 96.3% and 96.7%,
180 respectively. Interestingly, there are eight samples from COVID-19 patients are negative

181 for S1 IgG but positive for S2-78 IgG, suggesting it will be of diagnostic value to
182 combine S2-78 and S1. It is known that the immune response may correlate to some key
183 clinical parameters, such as gender, disease severity, age and the final outcome[5,9]. To
184 test whether the positive rate of S2-78 IgG is associated with these clinical parameters,
185 we analyzed the detection sensitivities among subgroups, *i. e.*, male vs. female, ≥ 60 vs.
186 < 60 for age, severe vs. non-severe cases and survivor vs. non-survivor with critical
187 diseases. Similar to that of S1 IgG (**Fig. S2a**), no significant difference was observed in
188 all these subgroups.

189

190 Virus specific antibodies would persistently increase and usually conduct
191 seroconversion within one or two weeks after symptom onset, and the portion of patients
192 with positive antibodies reach $\sim 100\%$ after two or three weeks[21,22]. To test whether
193 S2-78 IgG is suitable to detect the dynamic change of antibody response, we analyzed the
194 positive rates at different time points after the onset of symptoms. Expectedly, the
195 positive rate of S2-78 IgG continuously increased and reach the plateau about three
196 weeks after onset, which is similar as S1 and N IgG (**Fig. 3d**). Similar trends were
197 observed for S2-78 IgM (**Fig. 3e**) and IgG antibodies against other peptides (**Fig. S2b**).
198 These observations suggest that the antibodies against S2-78 and other peptides could be
199 applied for monitoring virus specific antibody dynamics.

200

201 *The diagnostic performance of S2-78 IgG is comparable to that of S1 IgG for*
202 *asymptomatic infections*

203 Monitoring of asymptomatic infections is essential to put SARS-CoV-2 infection under
204 control. It is thought that asymptomatic infection usually has a weak immune
205 response[23]. To test whether S2-78 IgG could be used for detection of asymptomatic
206 infection, we analyzed 63 asymptomatic patients (Cohort 4, **Table 1**) defined as positive
207 either in NAT test or antibody test conducted by a commercial assay (see **methods**) but
208 without obvious symptom. Four subgroups were divided according to positive or not for
209 NAT, IgG or IgM[24]. It was shown that for S1 IgG and S2-78 IgG, all samples ($n=4$) of
210 IgG^- group were negative, while of the 59 IgG^+ asymptomatic infections, 47 and 45 were
211 positive for S1 IgG and S2-78 IgG, respectively. (**Fig 4a, b**). The consistency between S1

212 IgG and S2-78 IgG was also high (93.7%). The contradictory results in IgG⁺ group
213 between our data and the commercial assay may due to the differences of antigens
214 involved. S and N recombinant proteins are used for the commercial assay, for which
215 slightly lower specificity is common[7]. Overall, these results demonstrate the diagnostic
216 and screening value of S2-78 IgG for asymptomatic infections.

217

218 *The diagnostic value of S2-78 IgG was validated by ELISA*

219 ELISA (enzyme linked immunosorbent assay) is common for commercial SARS-CoV-2
220 antibody assays[8,25]. To verify the efficacy and applicability of S2-78 IgG, we
221 established an ELISA assay. Firstly, to test the consistency of the peptide microarray and
222 ELISA, we randomly selected 31 sera from COVID-19 patients of Cohort 1 and tested by
223 ELISA. High consistency was achieved with a Pearson correlation of 0.926 (**Fig. 5a**),
224 demonstrating the validity of the microarray results. To further validate the performance
225 of S2-78 IgG, we screened another independent cohort of samples collected from a
226 different medical center (Cohort 3, **Table 1**). As expected, high performance of S2-78
227 IgG for specific detection of COVID-19 was achieved by ELISA (**Fig. 5b, c**).

228

229 *A two-step strategy by combining S2-78 and other peptides for specific detection of*
230 *SARS-CoV-2 infections*

231

232 Because of the highly homologous genomes among human coronaviruses, one
233 unneglectable risk for antibody based diagnosis of SARS-CoV-2 infection is the possible
234 cross-reactivity with other coronaviruses, especially for those common cold causing
235 coronaviruses, *i. e.*, HCoV-OC43, HKU1, NL63 and 229E. The protein sequences of S
236 protein among those viruses are of high similarity, selection of sections of S protein that
237 are unique for SARS-CoV-2 is important for diagnosis[11,13]. To investigate whether the
238 identified peptides are specific to SARS-CoV-2, we performed homology analysis among
239 SARS-CoV-2 and 6 other coronaviruses (**Extended data Fig 3**). High homologies were
240 observed for S2-78 and S2-22, suggesting they alone may not suitable for specific
241 detection of SARS-Cov-2 infection. In contrast, other peptides, *i. e.*, S1-93,
242 97,101,105,111 and S2-97, exhibit low similarities with other coronaviruses, particularly
243 the four coronaviruses that cause common cold in human, suggesting they could be

244 applied for specifically detection of SARS-Cov-2 infection. S1-97 with the best
245 performance among these peptides was selected for further investigation. Considering the
246 relatively low sensitivity (86.2%) of S1-97 IgG (**Fig. 2, Fig. S4**), we proposed a two-step
247 strategy by combining S1-97 IgG and S2-78 IgG for detection (**Fig. 6a**). To simplify this
248 strategy, there are three assumptions: 1) S2-78 IgG can detect the infection of
249 SARS-CoV-2 and related coronaviruses with the same sensitivity (95.5%) and generates
250 a false positive rate of 3.3% (1-specificity, 1- 96.7%) for healthy and patients of other
251 diseases; 2) S1-97 IgG can specially detect SARS-CoV-2 infection with 86.2% sensitivity
252 and no cross-reactivity with related coronaviruses, so it would generate a false positive
253 rate of 6.8% (1-specificity, 1-93.2%) for the infection of related coronaviruses, healthy
254 control and patients with other diseases; 3) The positive rate of S2-78 IgG and S1-97 IgG
255 in each group is independent to each other. According to the two-step strategy (**Fig. 6a**),
256 for any given sample, the first step is to detect S2-78 IgG; For the positive ones, the
257 second step is to detect S1-97 IgG. The samples positive for both S2-78 and S1-97 are
258 defined as final positive, while the samples negative for either one is defined as negative.
259 As a result, the sensitivity for SARS-CoV-2 infection detection would be 82.3%, the
260 specificity for related coronavirus infections would be 93.5% (*specificity 1*), and the
261 specificity for control group would be 99.8% (*specificity 2*). To further improve the
262 performance of the two-step strategy, a panel (Panel-A) of peptides was composed by
263 bivariate regression analysis based on the specificity of each peptide to SARS-CoV-2.
264 The linear function for Panel-A is $y = 0.014 * x_1 + 0.02 * x_2 - 0.003 * x_3 + 0.006 * x_4 - 2.593$,
265 where y represents signals of Panel-A and x_1, x_2, x_3, x_4 represents the signals of IgG
266 against S1-93, 97, 101 and 105, respectively. The sensitivity and specificity of Panel A,
267 based on the data generated from Cohort 2, is 88.3% and 96.7%, respectively (**Fig. 6b, c**).
268 Follow the two-way strategy by combining S2-78 IgG and Panel-A, the final sensitivity
269 for the detection of SARS-CoV-2 infection, specificity for related coronavirus infection
270 (*specificity 1*), and the specificity for control group (*specificity 2*) are 84.3%, 96.8% and
271 99.9%, respectively (**Fig. 6d**). When S1 protein and Panel-A are combined, the final
272 sensitivity for the detection of SARS-CoV-2 infection, specificity for related coronavirus
273 infection (*specificity 1*), and the specificity for control group (*specificity 2*) are 85.3%,
274 96.8% and 99.99%, respectively (**Fig. S4b**). These results demonstrate that combination

275 of human coronavirus conserved peptides or S1 protein and other SARS-CoV-2 specific
276 peptides (“significant” peptides) enable specific detection of SARS-CoV-2 infection with
277 a high specificity and acceptable sensitivity.

278

279 **Discussion**

280 In this study, we took advantage of a peptide microarray of full S protein coverage
281 [17,26], analyzed 2,434 sera from 858 COVID-19 patients, sera from 63 asymptomatic
282 patients and 610 controls collected from multiple medical centers. We identified eight
283 12-mer peptides (“significant” peptides) which exhibit high diagnostic values as antigens
284 to detect SARS-CoV-2 specific IgG or IgM. Among the “significant” peptides, S2-78
285 IgG has a comparable diagnosis performance to that of S1 protein for the detection of
286 COVID-19 and asymptomatic infections, suggesting S2-78 has the potential for
287 serological prevalence investigation. By combining S2-78 and other “significant”
288 peptides of low homologous among common human coronaviruses, we purposed a
289 two-step strategy for accurate and affordable detection of SARS-CoV-2 infection.

290

291 Serological test plays an important role in diagnosis and monitoring of COVID-19.
292 Recombinant proteins, particular S1, is one of the key reagents to build immunoassay for
293 detecting SARS-CoV-2 IgG, IgM or IgA. However, expressing S1 protein in the right
294 conformation is usually difficult, and in some cases, the antibodies that recognize
295 membrane spike protein are unable to bind recombinant S protein[27]. Moreover, the
296 inconsistency of the proteins from different manufacturers and even different batches
297 from the same manufacturer may result in high variation. High cost and insufficient
298 capacity to produce enough amount of high quality recombinant S1 protein limits the
299 accessibility of the immunoassay in poor or remote regions around the world.
300 Alternatively, peptide-based immunoassay provides a superior choice to S1 protein
301 assays. The reasons are as follows: 1) The peptide synthesis could be easily scaled up
302 when required. A large amount of peptide could be easily synthesized within a very short
303 period of time, if necessary, the peptide could be synthesized in a GMP facility; 2) The
304 consistency and purity of peptide synthesis is high, there is almost no batch-to-batch

305 variation; 3) Peptide is very stable, as reagent, it could be easily stored and transported; 4)
306 The cost of peptide is about 2-3 magnitude lower than that of S1 protein.

307

308 We identified S2-78 as a candidate of high diagnostic value. Applying large cohorts of
309 COVID-19 patients and a variety types of controls, we comprehensively verified S2-78
310 IgG as a good candidate for diagnosis of COVID-19, with comparable specificity and
311 sensitivity to that of S1 IgG. Expectedly, we found that the overall consistency of S2-78
312 IgG and S1 IgG is high, suggesting S2-78 has the potential to replace S1 protein. It is
313 notable that, the signal intensity level of S2-78 IgG is lower than that of S1 IgG, which is
314 reasonable since there are multiple sites on S1 protein that could be recognized by
315 antibodies in COVID-19 sera. However, because it is hard to ensure extremely high
316 purity when purifying recombinant proteins, the background of S1 IgG for control group
317 is also higher than that of peptides. These results imply the sensitivity of S2-78 IgG might
318 be further evaluated when raise the antigen concentration or adopt a more sensitive
319 platform for detection, such as electrochemical platform or single-molecule detection
320 technologies[28,29]. Indeed, we also performed S1 based ELISA assay by a commercial
321 kit with the same set of samples, the overall performance of S2-78 IgG based ELISA is
322 comparable to that of the commercial kits (data not shown).

323

324 Immunoassay is the major tool to assess the extent of virus circulation in population
325 and the likelihood of protection against re-infection by screening population to identify
326 infected individuals without clinical symptoms[7,12,19]. We verified that S2-78 IgG
327 could be applied to detect asymptomatic infections with a comparable sensitivity and
328 specificity with S1 IgG, though these individuals are thought to have weaker immune
329 responses[23]. We assessed PPV and NPV for two prevalence rates, *i. e.*, 0.04 and 0.5.
330 For 0.04, the PPV is 54.7% and NPV is 99.8%, indicating a high false positive rate, while
331 it can effectively and accurately exclude the negative ones. Due to the small number of
332 positive cases under this circumstance, one possible solution is to re-test the suspected
333 samples by additional assays. Moreover, to meet the requirement of population-wide
334 application, the performance of S2-78 might be improved by optimizing the parameters
335 or adoption of other platforms.

336

337 S protein shares high sequence similarities with other seasonal circulated human
338 coronaviruses. Theoretically, cross-reactivity may exist when S/S1 is applied as antigen
339 for immunological test, thus cause false positive. It is thus necessary to pinpoint specific
340 regions/ sites of SARS-CoV-2 S protein and eliminate the potential cross-reactivity.
341 Selection of peptides with high antibody responses and low sequence similarities with
342 other coronaviruses will improve the performance of diagnosis. Among the identified
343 peptides, except S2-78 and S2-22, other peptides are very distinct to the four circulating
344 human coronaviruses, implying that they could be served as the specific antigen to
345 eliminate the potential cross-activity. However, these peptides exhibit slightly lower
346 sensitivity and specificity. To take advantage of these “significant” peptides, we proposed
347 a two-step strategy that combine S2-78 with other significant peptide/s to discriminate
348 SARS-CoV-2 from related infections as well as non-infections. This study could be
349 further strengthened by testing sera collected from the infections of common human
350 coronaviruses, *i. e.*, HCoV-OC43, HKU1, NL63 and 229E.

351

352 In summary, we identified and verified eight peptides derived from S protein that
353 exhibit high diagnostic values. These peptides might be used in different circumstances
354 alone or in combination as candidates to build immunoassay/s for monitoring COVID-19.
355 In comparison to the current protein based immunoassays, the peptide based assays will
356 be highly affordable and accessible.

357

358 **Materials and Methods**

359 *Patients and samples*

360 The study was approved by the Ethical Committee of Tongji Hospital, Tongji Medical
361 College, Huazhong University of Science and Technology, Wuhan, China
362 (ITJ-C20200128), Institutional Ethics Review Committee of Foshan Fourth Hospital,
363 Foshan, China (202005) and the Ethical Committee of The Fifth Affiliated Hospital of
364 Sun Yat-sen University, Zhuhai, China (K14-2). Written informed consent was obtained
365 from all participants enrolled in this study. COVID-19 patients were hospitalized and

366 received treatment in multiple medical center during the period from 25 January 2020
367 and 28 April 2020. Sera of the control group from healthy donors, lung cancer patients,
368 patients with autoimmune diseases were collected from Ruijin Hospital, Shanghai, China
369 or Tongren Hospital, Shanghai, China. The negative reference samples were from
370 National Institutes for Food and Drug Control. Sera of 63 asymptomatic patients were
371 also from Tongji Hospital, Wuhan. The IgM and IgG antibodies against recombinant
372 nucleoprotein and spike protein of SARS-CoV-2 in sera of these patients were detected
373 by a commercial kit (YHLO Biotech, Shenzhen, China). According to the instruction of
374 the kit, the antibody level ≥ 10 AU/ mL is positive, and < 10 AU/mL is negative. All the
375 samples were stored at -80°C until use.

376

377 *Peptide synthesis and conjugation with BSA*

378 The N-terminal amidated peptides were synthesized by GL Biochem, Ltd. (Shanghai,
379 China). Each peptide was individually conjugated with BSA using Sulfo-SMCC (Thermo
380 Fisher Scientific, MA, USA) according to the manufacture's instruction. Briefly, BSA
381 was activated by Sulfo-SMCC in a molar ratio of 1: 30, followed by dialysis in PBS
382 buffer. The peptide with cysteine was added in a w/w ratio of 1:1 and incubated for 2 h,
383 followed by dialysis in PBS to remove free peptides. A few conjugates were randomly
384 selected for examination by SDS-PAGE. For the conjugates of biotin-BSA-peptide,
385 before conjugation, BSA was labelled with biotin by using NHS-LC-Biotin reagent
386 (Thermo Fisher Scientific, MA, USA) with a molar ratio of 1: 5, and then activated by
387 Sulfo-SMCC.

388

389 *Peptide microarray fabrication*

390 The peptide-BSA conjugates as well as S1 protein, RBD protein and N protein of
391 SARS-CoV-2, along with the negative (BSA) and positive controls (anti-Human IgG and
392 IgM antibody), were printed in triplicate on PATH substrate slide (Grace Bio-Labs,
393 Oregon, USA) to generate identical arrays in a 1 x 7 (for the peptide microarray with
394 three concentrations) or 2 x 7 subarray format (for the peptide microarray with one

395 concentration) using Super Marathon printer (Arrayjet, UK). The microarrays were stored
396 at -80°C until use.

397

398 *Microarray-based serum analysis*

399 A 7 or 14-chamber rubber gasket was mounted onto each slide to create individual
400 chambers for the 7 or 14 identical subarrays. The microarray was used for serum
401 profiling as described previously with minor modifications[30]. Briefly, the arrays stored
402 at -80°C were warmed to room temperature and then incubated in blocking buffer (3%
403 BSA in 1×PBS buffer with 0.1% Tween 20) for 3 h. A total of 400 µL for 7-subarray
404 format or 200 µL 14-subarray format of diluted sera or antibodies was incubated with
405 each subarray for 2 h. The sera were diluted at 1:200 for most samples and for
406 competition experiment, free peptides were added at a concentration of 0.25 mg/mL. For
407 the enriched antibodies, 0.1-0.5 µg antibodies were included in 200 µL incubation buffer.
408 The arrays were washed with 1×PBST and bound antibodies were detected by incubating
409 with Cy3-conjugated goat anti-human IgG and Alexa Fluor 647-conjugated donkey
410 anti-human IgM (Jackson ImmunoResearch, PA, USA), which were diluted for 1: 1,000
411 in 1×PBST. The incubation was carried out at room temperature for 1 h. The microarrays
412 were then washed with 1×PBST and dried by centrifugation at room temperature and
413 scanned by LuxScan 10K-A (CapitalBio Corporation, Beijing, China) with the
414 parameters set as 95% laser power/ PMT 550 and 95% laser power/ PMT 480 for IgM
415 and IgG, respectively. The fluorescent intensity was extracted by GenePix Pro 6.0
416 software (Molecular Devices, CA, USA).

417

418 *Data analysis of peptide microarray*

419 For each spot, signal intensity was defined as the foreground subtracted by the
420 background. The signal intensities of the triplicate spots for each peptide or protein were
421 averaged. For the samples of Cohort 2 and 3, normalization among microarray slides
422 were performed. For each slide, block #14 was incubated with the positive reference
423 sample which was generated by pooling of 50 randomly selected sera from COVID-19

424 patients. For each slide, the data generated from the positive reference sample were
425 subjected to build a linear regression function, and the data from other samples were
426 subjected to linear normalization according to the function. Graphpad 6.0 was used to
427 generate ROC plots and calculated AUC values.

428

429 *ELISA*

430 Briefly, 96-well microplates with high binding polystyrene surface (Corning, New York,
431 USA) were coated with 100 μ L BSA conjugated peptide (S2-78) at 100 μ g/mL and
432 incubated overnight at 4°C. The plates were washed once with PBST buffer (PBS buffer
433 with 0.1% Tween 20), and blocked with 3% BSA (bovine serum albumin) for 1 h at room
434 temperature, followed by one wash with PBST. The sera were diluted at 1:50 in PBST
435 buffer with 1% BSA, 1% FBS and 3% horse serum, 100 μ L of the preparation was loaded
436 to each well, the incubation was carried out at 37°C for 1.5 h in 100 μ L. After six washes
437 with PBST, the secondary antibody, *i.e.*, anti-human IgG-peroxidase (Sangon Biotech,
438 Shanghai, China) was diluted at 1:10000 and incubated at 37 ° C for 1 h, followed by
439 eight washes. Tetramethylbenzidine substrate (Sigma-Aldrich, Missouri, USA) was then
440 added and incubated for 20 min. Finally, 50 μ L sulfuric acid (2 M) was added to stop the
441 reaction. The optical density was read at 450 nm using a Behring EL311 ELISA
442 microplate reader (Dade Behring Marburg GmbH, Berlin, Germany). The assays were
443 repeated twice for each sample.

444 **Abbreviations:** AID: autoimmune diseases; AUC: area under curve; CI: confidence
445 interval; CoV: coronavirus; ELISA: enzyme linked immunosorbent assay; GMP: good
446 manufacturing practice; NAT: nucleic acid test; NPV: negative predictive value; PPV:
447 positive predictive value; ROC: receiver operating characteristic; SARS: severe acute
448 respiratory syndrome; URI: upper respiratory infections; WHO: world health
449 organization.

450 **Acknowledgments:** We thank Dr. Daniel M. Czajkowsky for critical reading and editing.

451 **Funding:** This work was partially supported by National Key Research and Development
452 Program of China Grant (No. 2016YFA0500600), Science and Technology Commission

453 of Shanghai Municipality (No. 19441911900), Interdisciplinary Program of Shanghai
454 Jiao Tong University (No. YG2020YQ10), National Natural Science Foundation of
455 China (No. 31900112, 21907065, 31970130 and 31670831).

456 **Author contributions:** S-C.T. developed the conceptual ideas and designed the study.
457 Z-Y.S., F.W., H-Y.H., Y-D.Z., X-S. L., Z-J.Y., H-M.S., J-X.W., L-Y.C., S-Q.L., P-F.P.,
458 and H.S. collected the sera samples. X-L.F., Y.L., D-Y.L., Q.L., Z-W.X., M-L. M., B.Z.,
459 H.C., C-Z.Y., J-B.X., X-N.W., Y-X.Z., H-N.Z., H-W.J., H.Q. Y-D.Z., X-S.L., Z-J.Y and
460 S-J.G. performed the experiments and analysis. S-C.T. and Y.L. wrote the manuscript
461 with suggestions from other authors.

462 **Competing interests:** Four relevant patents which have been applied in China
463 (application number: 202010413901.9, 202010415057.3, 202010415053.5 and
464 202010413911.2) are filed by Shanghai Jiao Tong University. S-C.T., Y.L., D-Y. L,
465 H-W.J., H-N.Z. and H.Q. are relevant to these patents. Other authors declare no conflicts
466 of interest.

467 **Data and material availability:** The peptide microarray data generated in this study has
468 been deposited to Protein Microarray Database (<http://www.proteinmicroarray.cn>) under
469 the accession number PMDE242 and PMDE244 Additional data related to this study may
470 be requested from the authors.

471

472 **Supplementary Materials**

473

474 Fig. S1. Evaluation of the significant peptides by Cohort 2 – other peptides.

475 Fig. S2. Performance of peptides and S1 for diagnosis in subgroups.

476 Fig. S3. Homology analysis of the significant peptides among SARS-CoV-2,

477 SARS-CoV, MERS-CoV and other four human coronaviruses.

478 Fig. S4. Discrimination of related coronaviruses by combining S1 and peptides of
479 Panel-A.

480 Table S1. Peptides of SARS-CoV-2 spike protein used in this study

481

482

483 References

484

- 485 1. Zhou P, Yang X Lou, Wang XG, et al. A pneumonia outbreak associated with a
486 new coronavirus of probable bat origin. *Nature* [Internet]. 2020; 579: 270–3.
487 Available at: <http://dx.doi.org/10.1038/s41586-020-2012-7>
- 488 2. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory
489 disease in China. *Nature*. 2020; 579: 265–9.
- 490 3. Dong E, Du H, Gardner L. An interactive web-based dashboard to track
491 COVID-19 in real time. *Lancet Infect Dis* [Internet]. 2020; 20: 533–4. Available at:
492 [http://dx.doi.org/10.1016/S1473-3099\(20\)30120-1](http://dx.doi.org/10.1016/S1473-3099(20)30120-1)
- 493 4. Lee CYP, Lin RTP, Renia L, Ng LFP. Serological Approaches for COVID-19:
494 Epidemiologic Perspective on Surveillance and Control. *Front Immunol*. 2020; 11:
495 879.
- 496 5. Long QX, Liu BZ, Deng HJ, et al. Antibody responses to SARS-CoV-2 in patients
497 with COVID-19. *Nat Med*. 2020; 26: 845–8.
- 498 6. Peeling RW, Wedderburn CJ, Garcia PJ, et al. Serology testing in the COVID-19
499 pandemic response. *Lancet Infect Dis* [Internet]. 2020; 3099: 30517–X. Available
500 at: <http://www.ncbi.nlm.nih.gov/pubmed/32687805>
- 501 7. GeurtsvanKessel CH, Okba NMA, Igloi Z, et al. An evaluation of COVID-19
502 serological assays informs future diagnostics and exposure assessment. *Nat*
503 *Commun*. 2020; 11: 3436.
- 504 8. Yan Y, Chang L, Wang L. Laboratory testing of SARS-CoV, MERS-CoV, and
505 SARS-CoV-2 (2019-nCoV): Current status, challenges, and countermeasures. *Rev*
506 *Med Virol*. 2020; 30: e2106.
- 507 9. Jiang H, Li Y, Zhang H, et al. Global profiling of SARS-CoV-2 specific IgG/ IgM
508 responses of convalescents using a proteome microarray. *Nat Commun*. 2020; 11:
509 3581.
- 510 10. Vabret N, Britton GJ, Gruber C, et al. Immunology of COVID-19: Current State of
511 the Science. *Immunity*. 2020; 52: 910–41.
- 512 11. Petherick A. Developing antibody tests for SARS-CoV-2. *Lancet*. 2020; 395:
513 1101–2.
- 514 12. Lisboa Bastos M, Tavaziva G, Abidi SK, et al. Diagnostic accuracy of serological
515 tests for covid-19: Systematic review and meta-analysis. *BMJ*. 2020; 370: m2516.
- 516 13. Hicks J, Klumpp-Thomas C, Kalish H, et al. Serologic cross-reactivity of
517 SARS-CoV-2 with endemic and seasonal Betacoronaviruses. *medRxiv* [Internet].
518 2020; DOI: 10.1101/2020.06.22.20137695. Available at:
519 [https://www.medrxiv.org/content/10.1101/2020.06.22.20137695v1%0Ahttps://ww
520 w.medrxiv.org/content/10.1101/2020.06.22.20137695v1.abstract](https://www.medrxiv.org/content/10.1101/2020.06.22.20137695v1%0Ahttps://www.medrxiv.org/content/10.1101/2020.06.22.20137695v1.abstract)
- 521 14. Wang H, Hou X, Wu X, et al. SARS-CoV-2 proteome microarray for mapping
522 COVID-19 antibody interactions at amino acid resolution. *bioRxiv* [Internet]. 2020;
523 DOI: 10.1101/2020.03.26.994756. Available at:
524 <http://biorxiv.org/content/early/2020/03/28/2020.03.26.994756>
- 525 15. Poh CM, Carissimo G, Wang B, et al. Two linear epitopes on the SARS-CoV-2
526 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat*
527 *Commun*. 2020; 11: 2806.

- 528 16. Zamecnik CR, Rajan J V, Yamauchi KA, et al. ReScan, a Multiplex Diagnostic
529 Pipeline, Pans Human Sera for SARS-CoV-2 Antigens. medRxiv. 2020;
530 DOI:10.1101/2020.05.11.20092528.
- 531 17. Li Y, Lai D, Zhang H, et al. Linear epitopes of SARS-CoV-2 spike protein elicit
532 neutralizing antibodies in COVID-19 patients. medRxiv [Internet]. 2020;
533 DOI:10.1101/2020.06.07.20125096. Available at:
534 <http://medrxiv.org/content/early/2020/06/09/2020.06.07.20125096.abstract>
- 535 18. WHO. WHO Laboratory testing for 2019 novel coronavirus (2019-nCoV) in
536 suspected human cases. 2020.
- 537 19. Xu X, Sun J, Nie S, et al. Seroprevalence of immunoglobulin M and G antibodies
538 against SARS-CoV-2 in China. Nat Med. 2020; DOI: 10.1038/s41591-020-0949-6.
- 539 20. Slot E, Hogema BM, Reusken CBEM, Reimerink JH, Molier M, Karregat JH.
540 Herd immunity is not a realistic exit strategy during a COVID-19 outbreak. Nat
541 Res. 2020; DOI: 10.21203/rs.3.rs-25862/v1.
- 542 21. Sikora M, von Bülow S, Blanc FEC, Gecht M, Covino R, Hummer G. Map of
543 SARS-CoV-2 spike epitopes not shielded by glycans. bioRxiv [Internet]. 2020;
544 DOI: 10.1101/2020.07.03.186825. Available at:
545 <http://biorxiv.org/content/early/2020/07/03/2020.07.03.186825.abstract>
- 546 22. Xiang F, Wang X, He X, et al. Antibody Detection and Dynamic Characteristics in
547 Patients with COVID-19. Clin Infect Dis. 2020; DOI: 10.1093/cid/ciaa461.
- 548 23. Long QX, Tang XJ, Shi QL, et al. Clinical and immunological assessment of
549 asymptomatic SARS-CoV-2 infections. Nat Med. 2020; DOI:
550 10.1038/s41591-020-0965-6.
- 551 24. Lei Q, Li Y, Hou H, et al. Antibody dynamics to SARS-CoV-2 in asymptomatic
552 COVID-19 infections. medRxiv. 2020; DOI: [org/10.1101/2020.07.09.20149633](https://doi.org/10.1101/2020.07.09.20149633).
- 553 25. Zhang X, Wu X, Wang D, et al. Proteome-wide analysis of
554 differentially-expressed SARS-CoV-2 antibodies in early COVID-19 infection.
555 medRxiv [Internet]. 2020; DOI:10.1101/2020.04.14.20064535. Available at:
556 [https://www.medrxiv.org/content/10.1101/2020.04.14.20064535v2?%253fcollecti
557 on=](https://www.medrxiv.org/content/10.1101/2020.04.14.20064535v2?%253fcollection=)
- 558 26. Li Y, Ma M, Lei Q, et al. Linear epitope landscape of SARS-CoV-2 Spike protein
559 constructed from 1,051 COVID-19 patients. medRxiv [Internet]. 2020;
560 DOI:10.1101/2020.07.13.20152587. Available at:
561 <http://medrxiv.org/content/early/2020/07/14/2020.07.13.20152587.abstract>
- 562 27. Wan J, Xing S, Ding L, et al. Human IgG cell neutralizing monoclonal antibodies
563 block SARS- CoV-2 infection. Cell Rep. 2020; 32: 107918.
- 564 28. Croop B, Zhang C, Lim Y, Gelfand RM, Han KY. Recent advancement of
565 light-based single-molecule approaches for studying biomolecules. Wiley
566 Interdiscip Rev Syst Biol Med. 2019; 11: e1445.
- 567 29. McEachern F, Harvey E, Merle G. Emerging Technologies for the Electrochemical
568 Detection of Bacteria. Biotechnol J. 2020; e2000140.
- 569 30. Li Y, Li C-Q, Guo S-J, et al. Longitudinal serum autoantibody repertoire profiling
570 identifies surgery-associated biomarkers in lung adenocarcinoma. EBioMedicine.
571 2020; 53: 102674.
- 572
- 573

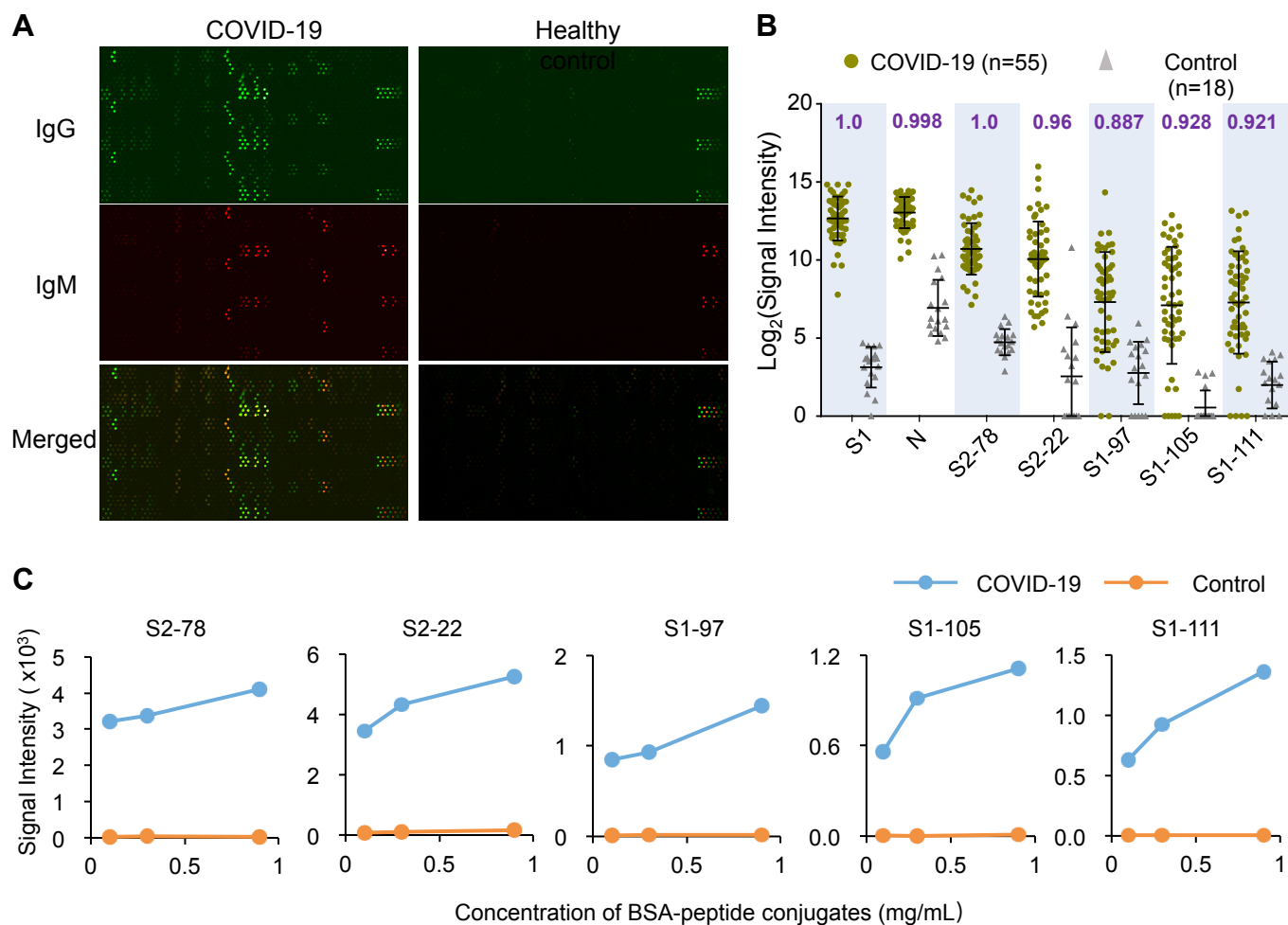


Figure 1. Peptides with strong IgG antibody responses were identified by Cohort 1. **A.** Representative images of peptide microarray profiled by sera from a COVID patient and a healthy control. IgG (green) and IgM (red) were detected simultaneously. **B.** Box plot of IgG antibody responses against S1 protein, N protein and some significant peptides for COVID-19 patient group (n=55) and control group (n=18) of Cohort 1. Each spot indicates one serum sample. Data are presented as box plots where the middle line is the mean value, and the upper and lower hinges are mean values \pm SD. AUC (area under curve) values are labeled for each peptide or protein on the top of the box plots. **C.** Averaged signal intensities of IgG antibody responses against the indicated peptides at different concentrations, *i.e.* 0.1, 0.3 and 0.9 mg/mL, both for COVID-19 patients (blue line) and control group (orange line).

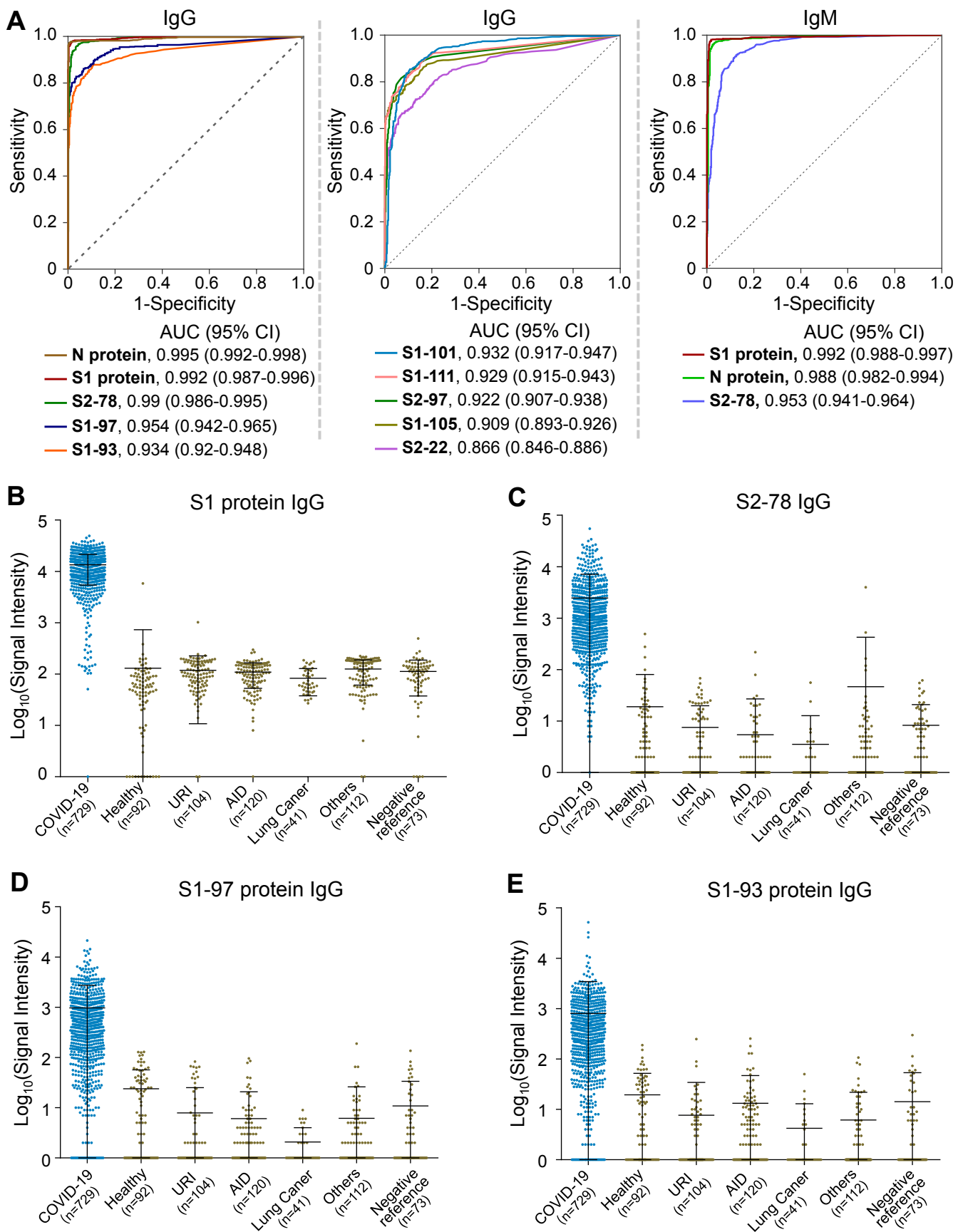
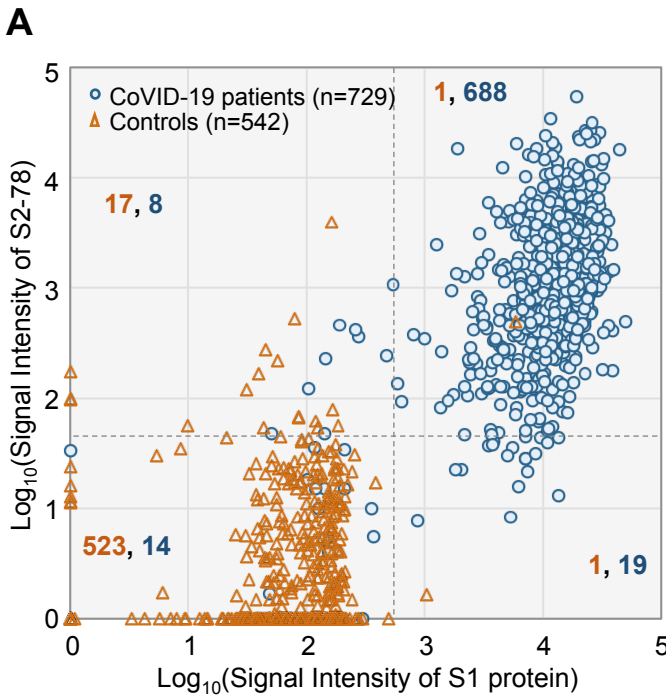


Figure 2. Evaluation significant peptides by Cohort 2. **A.** ROCs (receiver operating curves) of the peptides or proteins for discrimination of the COVID-19 group (n=729) and the control group (n=542). The AUC values with 95% CI (confidential intervals) are provided for all the peptides and proteins. **B-E.** signal distributions of anti-S1 IgG (B), anti-S2-78 IgG (C), anti-S1-97 IgG (D) and anti-S1-93 (E) in COVID-19 patients (blue) and control groups (yellow). Sample size is indicated for each group. URI: upper respiratory infection, AID: patients with autoimmune diseases.



B

		S2-78		
		Positive	Negative	Consistency
S1	COVID-19 Positive	688	19	96.3%
	COVID-19 Negative	8	14	(94.7-97.5%)
Control	Control Positive	1	1	96.7%
	Control Negative	17	523	(94.8-98%)

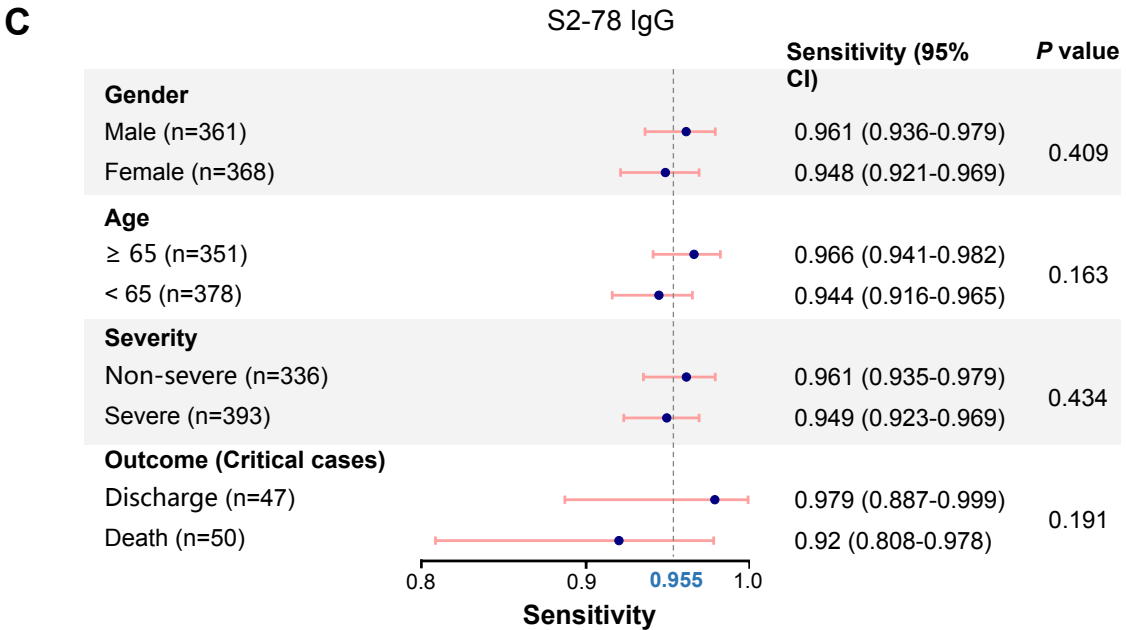
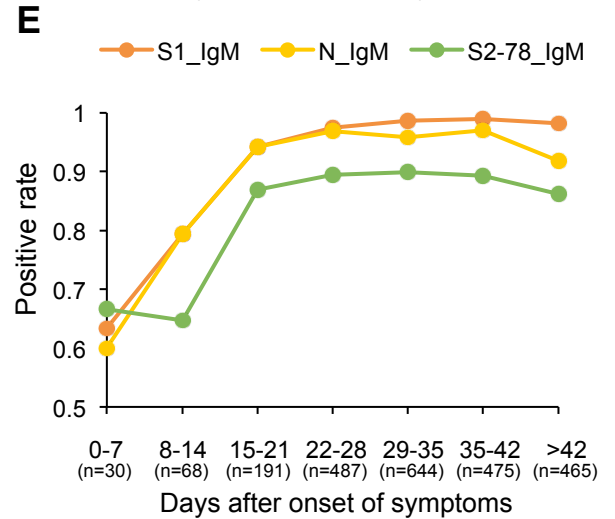
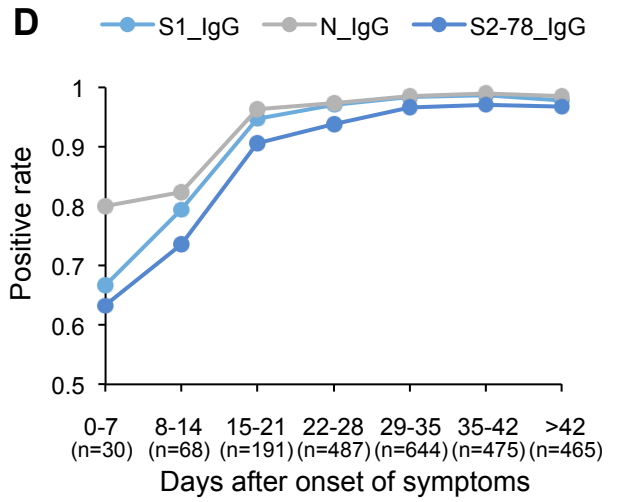


Figure 3. S2-78 IgG for diagnosis of COVID-19. **A.** Scatter plots of sera samples from COVID-19 (blue dots) and controls (orange triangle) for S1 IgG vs. S2-8 IgG. The grey lines indicate the cut-off values set as the optimal Youden index based on the ROC. The orange and blue numbers indicate the sample counts of control and patient on each quadrant, respectively. **B.** consistency between S1 protein and S2-78. The consistency values are provided with 95% CI. **C.** forest plot of sensitivities of S2-78 IgG among subgroups, i.e., age, gender, severity and outcome. The dots indicate the sensitivities while the error bars indicate the 95% CI. The exact values are also provided. P values were calculated with χ^2 test. **D-E.** Graph of positive rates of IgG or IgM against the S, N proteins or S2-78 versus days after symptom onset in 2,360 serum samples from 784 patients.

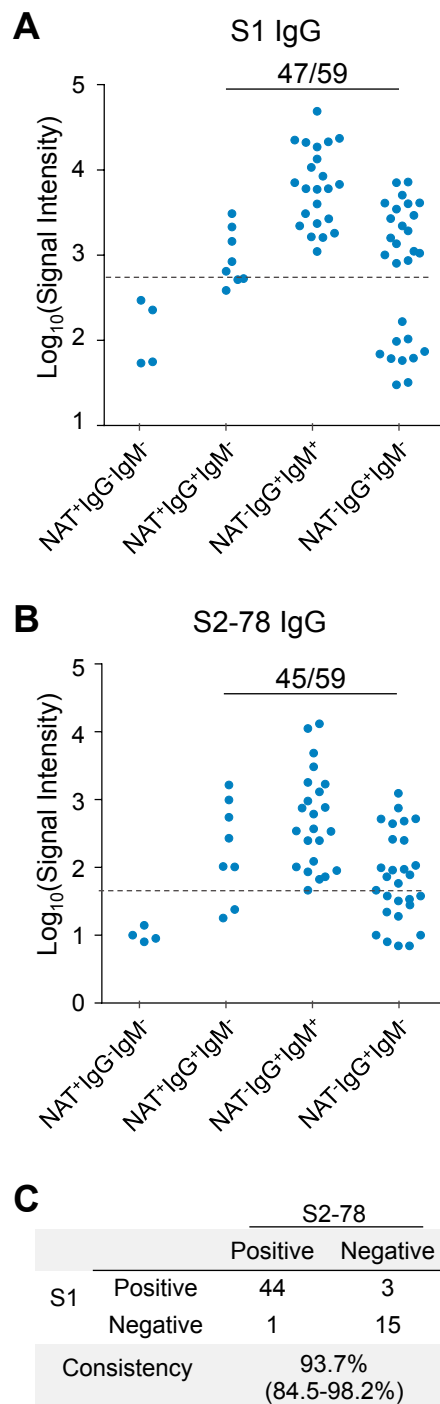


Figure 4. S2-78 IgG for diagnosis of asymptomatic patients. A-B. Levels of S1 IgG and S2-78 IgG in asymptomatic patient groups divided by positive or negative among three tests, i.e., NAT (nucleic acid test), SARS-CoV-2 virus specific IgG and IgM antibodies. The dashed lines indicate the cut-off values. The number of positive samples as well as the total number of samples of IgG⁺ (IgG positive) groups were presented for S1 IgG and S2-78 IgG. **C.** The consistency between S1 IgG and S2-78 IgG.

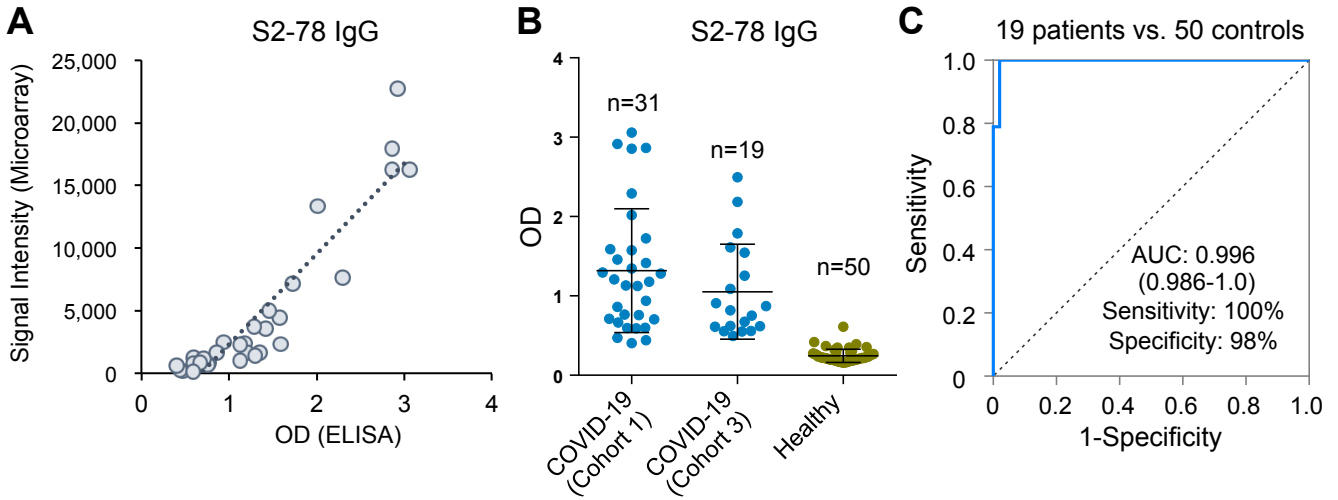


Figure 5. ELISA validation of S2-78 by Cohort 3. **A.** Correlation of the signals of S2-78 IgG response between peptide microarray and ELISA with Cohort 1 (n=31). **B.** Levels of S2-78 IgG in samples from COVID-19 patients either from Cohort 1 (n=31) or Cohort 3 (n=19) and healthy controls (n=50). **C.** ROC of S2-78 IgG for discrimination of 19 COVID-patients (Cohort 3) and 50 healthy controls.

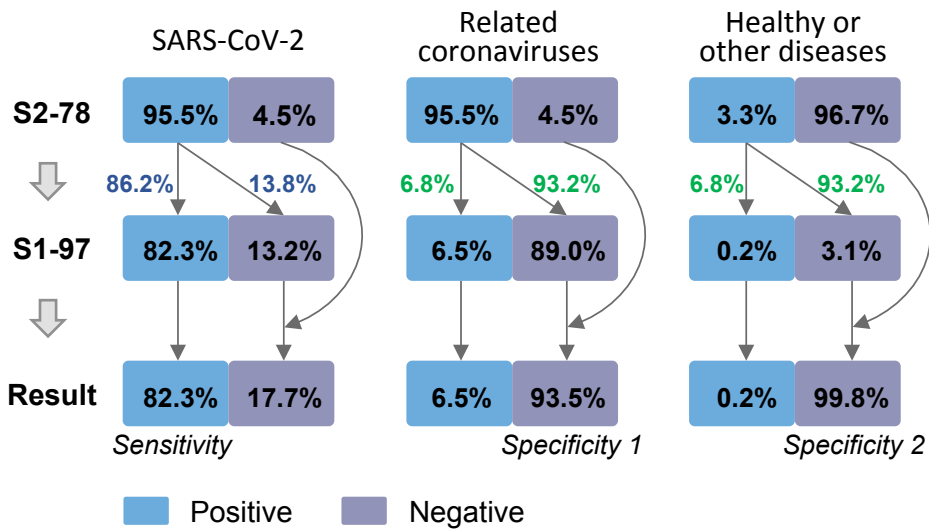
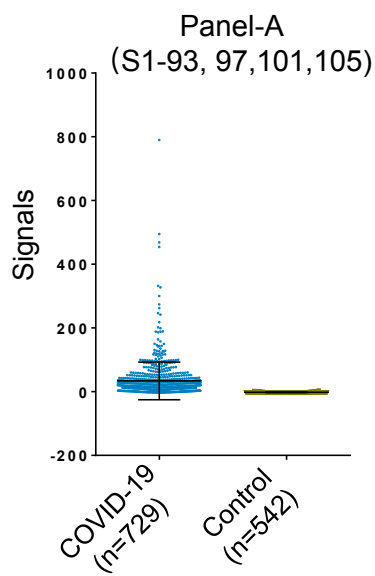
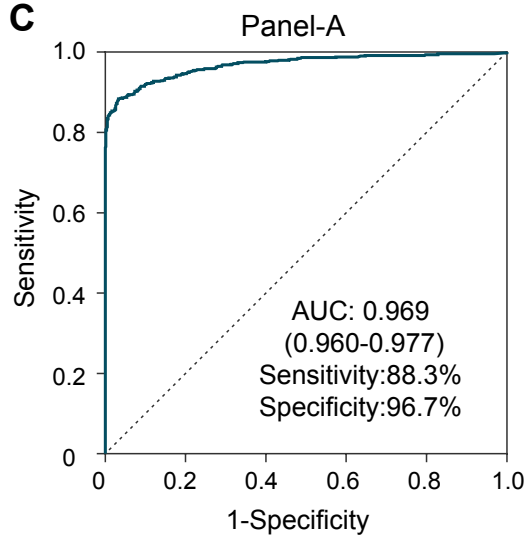
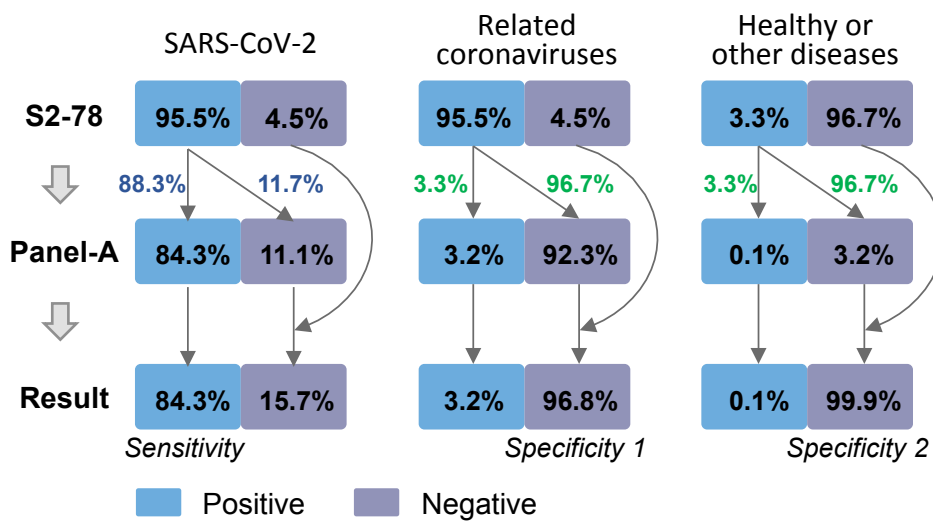
A**B****C****D**

Figure 6. Discrimination of related coronaviruses with combination of peptides.

A. Illustration of the strategy with combination of S2-78 IgG and S1-97 IgG based on the results of Cohort 2 (729 patients and 542 controls). For any given sample, the first step is to detect S2-78 IgG response. For the positive ones, the second step is to detect S1-97 IgG response. The blue numbers indicate the positive or negative rates for SARS-CoV-2 infections, while the green numbers indicated the assumed positive or negative rates for samples not belong the SARS-CoV-2 group. The last step summaries the overall sensitivities and specificities for different sample groups. **B.** Signals of Panel-A (S1-93, 97, 101 and 105 in COVID-19 patients and control groups. **C.** The corresponding ROC of Panel-A. **D.** The performance of the combination of S2-78 IgG and Panel-A.

Table 2. The overall performance of S1 and S2-78 for diagnosis

		S1		S2-78	
		Positive	Negative	Positive	Negative
COVID-19		707	22	696	33
Control		2	540	18	524
Specificity (95% CI)		99.6% (98.7-100%)		96.7% (94.8-98.0%)	
Sensitivity (95% CI)		97% (95.5-98.1%)		95.5% (93.7-96.9%)	
Accuracy (95% CI)		98.1% (97.2-98.8%)		96% (94.8-97%)	
Prevalence					
PPV	0.04	91.0%		54.7%	
	0.5	99.6%		97.5%	
NPV	0.04	99.9%		99.8%	
	0.5	97.1%		95.6%	

PPV: positive predictive value

NPV: negative predictive value

Table 1. Patients and controls involved in this study.

		Cohort 1		Cohort 2			Cohort 3		Cohort 4
Group		COVID-19	Control	COVID-19	Control-1	Control-2	COVID-19	Control	Asymptomatic patient ¹
Numbers		55	18	729	469	73	19	50	63
Stage		Convalescent	-	In hospital	-	-	Convalescent	-	-
Age		41.5±14.9	50.4±12.5	61.4±14.5	53.7±20.6	N/A	53.9±13.2	46.7±20.2	44.9±16.5
Gender	Male	27	8	361	224	N/A	10	24	25
	Female	28	10	368	245	N/A	9	26	38
Severity	Severe	0	-	393	-	-	0	-	-
	non-severe	55	-	336	-	-	19	-	-
Outcome	Discharge	55	-	679	-	-	19	-	-
	Death	0	-	50	-	-	0	-	-
Days after onset		27.5±7.7	-	31.1±8.8	-	-	42.6±9.5	-	-
Source		Foshan 4th Hospital, Guangdong	Ruijin Hospital, Shanghai	Tongji Hospital, Wuhan	Tongren Hospital, Shanghai Ruijin Hospital, Shanghai	National Institutes for Food and Drug Control, Beijing	The Fifth Affiliated Hospital Sun Yat-Sen University, Guangdong	Tongren Hospital, Shanghai	Tongji Hospital, Wuhan
Subtypes and numbers		-	Healthy: 8; LC ⁴ : 8	This is a subgroup of a cohort of 784 cases with a total of 2,360 longitudinal sera.	Healthy: 92; URI ² : 104; AID ³ : 120; LC ⁴ : 41; Other diseases: 112	Negative reference samples	-	-	NAT+IgG-IgM-: 4; NAT+IgG+IgM-: 8; NAT-IgG+IgM+: 23; NAT-IgG+IgM-: 28

¹Asymptomatic patients: positive for NAT (nucleic acid test) or antibody (IgG or IgM) performed in hospital.

²URI: Upper respiratory infection

³AID: Autoimmune diseases

⁴LC: Lung cancer

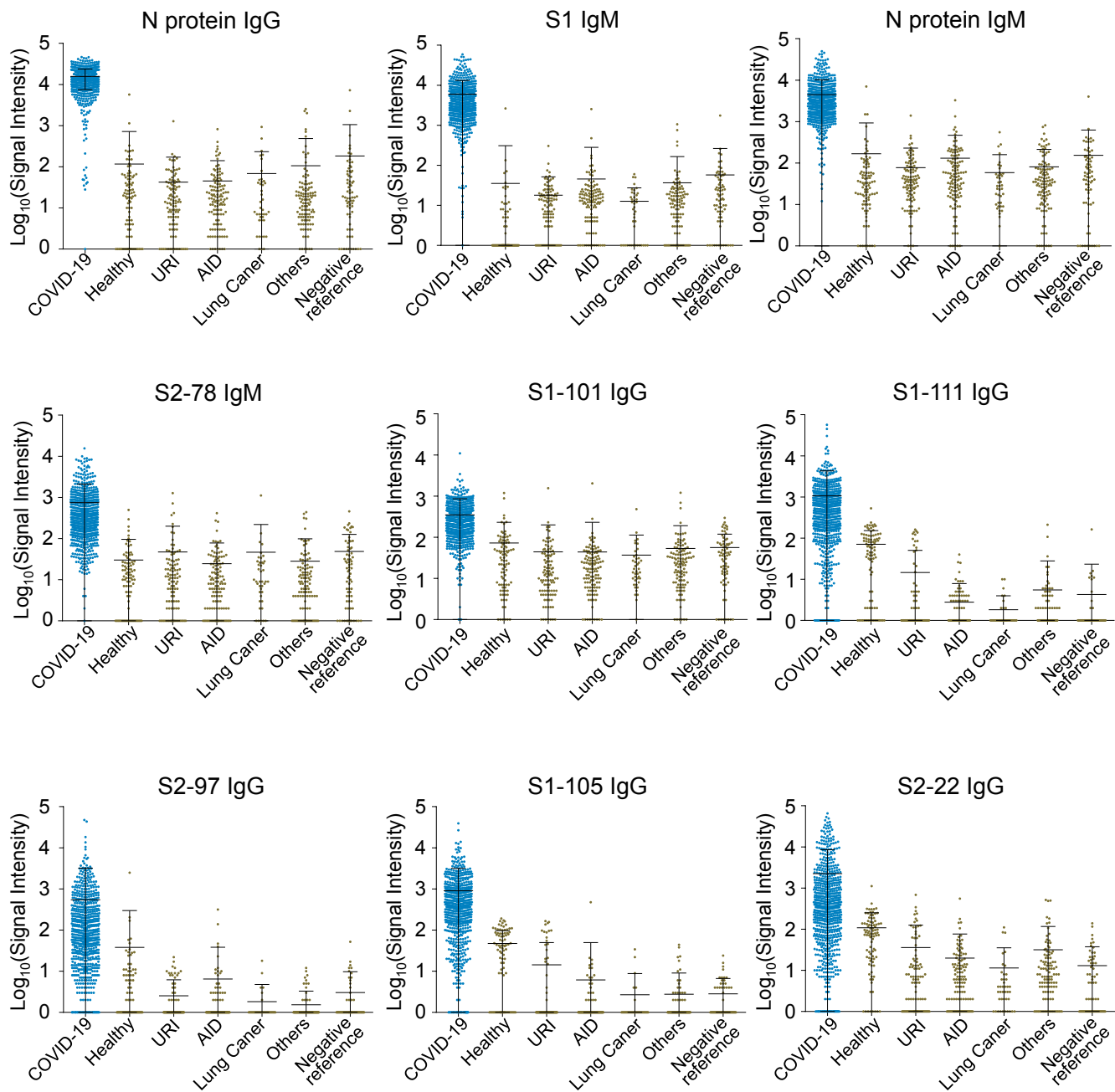


Figure S1. Evaluation of the significant peptides by Cohort 2 – other peptides. Signal levels of the antibodies against the indicated peptides in COVID-19 patients (n=729), Healthy controls (n=92), upper respiratory infections (URI, n=104), patients with autoimmune diseases (AID, n=120), lung cancer patients (n=41), patients with other diseases (n=112) and negative reference samples (n=73) of Cohort 2.

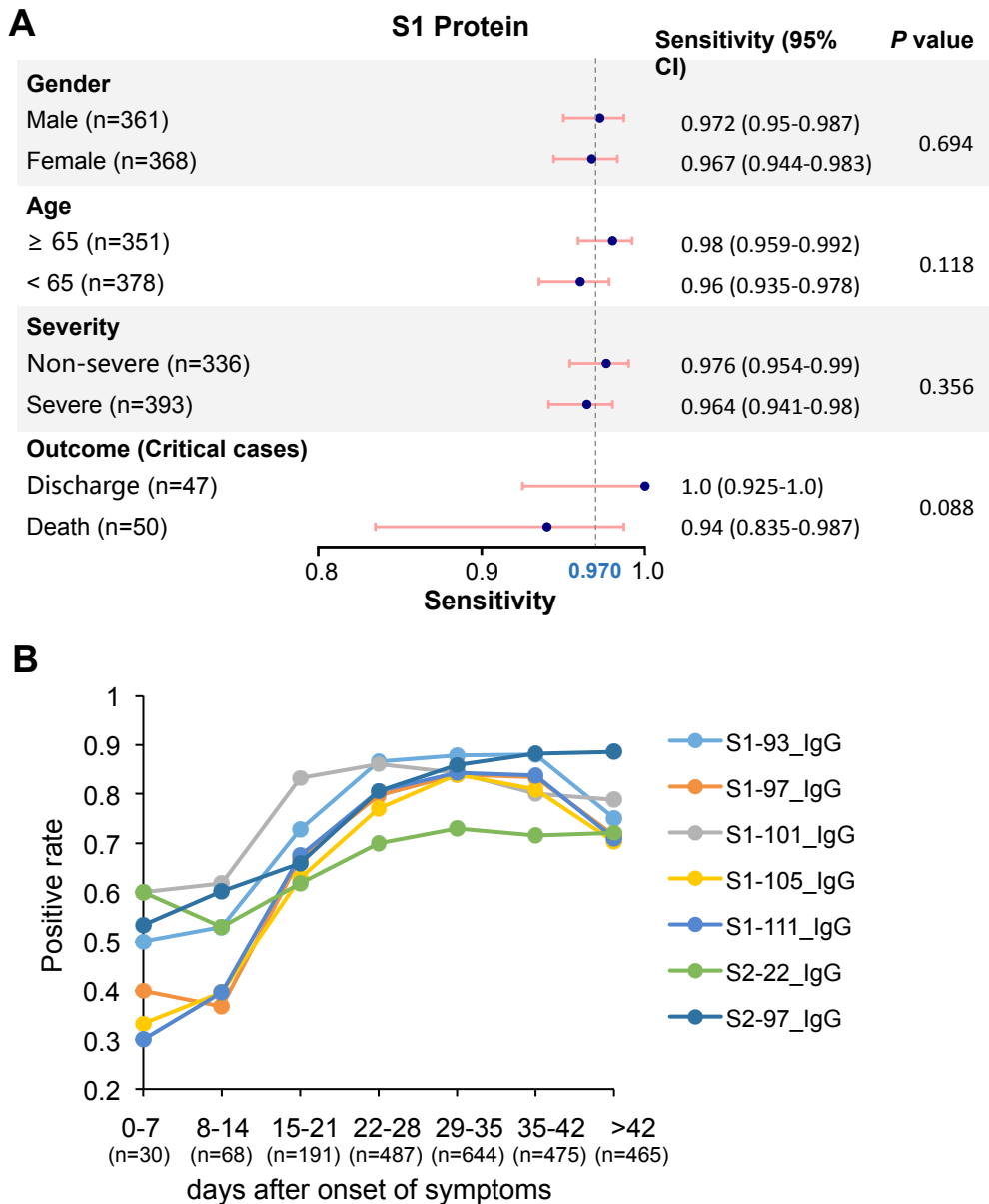


Figure S2. Performance of peptides and S1 for diagnosis in subgroups. A. forest plot of sensitivities of S1 IgG in different subgroups, i.e., age, gender, severity and outcome. The dots indicate the sensitivities while the error bars indicate the 95% CI. The exact values are also provided. P values were calculated with χ^2 test. **B.** Graph of positive rates of IgG antibodies against the indicated peptides versus days after symptom onset in 2,360 serum samples from 784 patients.

S1-93			S1-97				
SARS-CoV-2	553	TESNKKFL-PFQQ	564	SARS-CoV-2	577	RDPQTLEILDIT	588
SARS-CoV	539	TPSSKRFQ-PFQQ	550	SARS-CoV	563	RDPKTSEILDIS	574
MERS-CoV	618	QNCTAVGV-RQQR	629	MERS-CoV	643	SD--DGNYCLR	652
HCoV-0C43	636	VEVNATYYNSWQN	648	HCoV-0C43	661	RDYIINRTFMIR	672
HCoV-HKU1	636	KEVSAVYYNSWQN	648	HCoV-HKU1	661	KDFVTNKTYNIF	672
HCoV-NL63	642	RS SNQ SLAGG---	651	HCoV-NL63	665	KNVSTGNIFIVT	676
HCoV-229E	458	RI SND TFLNG---	467	HCoV-229E	481	KDVTNGTIYSIT	492

S1-101			S1-105				
SARS-CoV-2	601	GTNTSNQVAVLY	612	SARS-CoV-2	625	HADQLT--PTWRVY	636
SARS-CoV	587	GTNASSEVAVLY	598	SARS-CoV	611	HADQLT--PAWRIY	622
MERS-CoV	665	--KETK THATLF	674	MERS-CoV	686	SQYSRSTRSMLKRR	697
HCoV-0C43	685	--AN SSEPALLF	694	HCoV-0C43	706	L TRQLQ --PINYFD	717
HCoV-HKU1	685	--Q NASSLALLY	694	HCoV-HKU1	706	S---- LATQ -----	710
HCoV-NL63	689	Q-----	689	HCoV-NL63	701	SRYGLQ NLLQLP--	712
HCoV-229E	505	Q-----	505	HCoV-229E	517	TSYGF SNVVEMP--	528

S1-111			S2-78				
SARS-CoV-2	661	ECDIPIGAGICA	672	SARS-CoV-2	1148	FKEELDKYFKNH	1159
SARS-CoV	647	ECDIPIGAGICA	658	SARS-CoV	1130	FKEELDKYFKNH	1141
MERS-CoV	726	D CKLPLGQSLCA	737	MERS-CoV	1231	FQDELDEFKQNV	1242
HCoV-0C43	737	T CDLTVGSGYCV	748	HCoV-0C43	1233	FKEELDQWFKNQ	1244
HCoV-HKU1	685	SCALRMGSGFCV	694	HCoV-HKU1	1234	FSELSHWFKNQ	1245
HCoV-NL63	727	AVMTYS NFGICA	738	HCoV-NL63	1217	VNKT LQEF AQNL	1228
HCoV-229E	543	AVLTYSS FQVCA	554	HCoV-229E	1033	VNKT LQEL SYKL	1044

S2-22			S2-97				
SARS-CoV-2	812	PSKRSFIEDLLF	823	SARS-CoV-2	1262	EPVLKGVKLHYT	1273
SARS-CoV	794	PTKRSFIEDLLF	805	SARS-CoV	1244	EPVLKGVKLHYT	1255
MERS-CoV	884	RSARSA IEDLLF	895	MERS-CoV	1346	EPHKVHVH ----	1353
HCoV-0C43	900	ASSR SAIEDLLF	911	HCoV-0C43	1343	ELVIK -TSHDD-	1353
HCoV-HKU1	902	- SSRSFFEDLLF	912	HCoV-HKU1	1347	DFVIK -TSHDD-	1356
HCoV-NL63	867	IAGR SAIEDLLF	878	HCoV-NL63	1352	VHVQ-----	1356
HCoV-229E	683	VAGR SAIEDILF	694	HCoV-229E	1166	IHIQ-----	1170

Figure S3. Homology analysis of the significant peptides among SARS-CoV-2, SARS-CoV, MERS-CoV and other four human coronaviruses. The line for each position indicates the gap. The bold letters indicate the positions with the same amino acids as that of SARS-CoV-2. The homology analysis was performed by an online tool called Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).

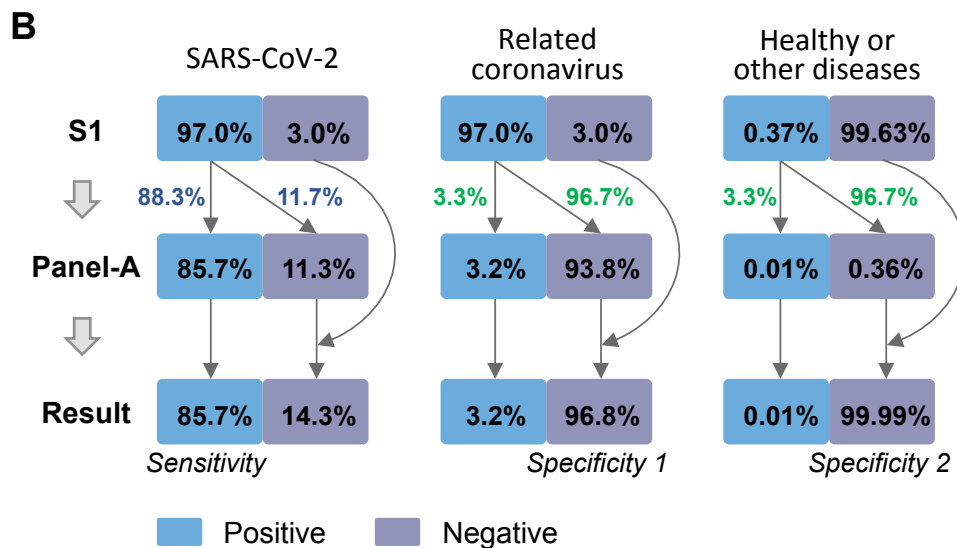
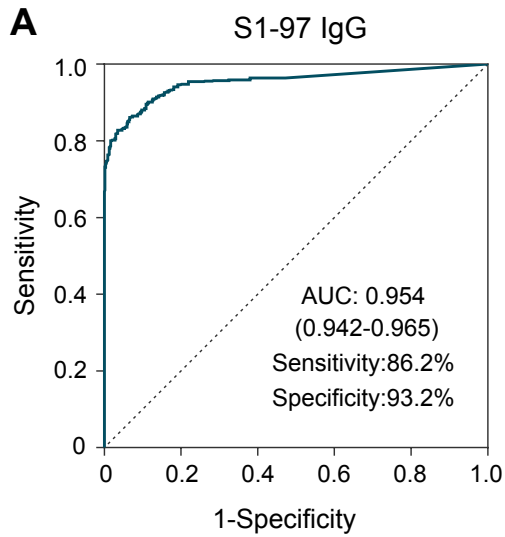


Figure S4. Discrimination of related coronaviruses by combining S1 and peptides of Panel-A. **A.** ROC of S1-97 IgG based on the results of Cohort 2. **B.** The combination of S1 IgG and Panel-A based on the results of Cohort 2 (729 patients and 542 controls).

Table S1. The peptides synthesized in this study

NO.	Peptide ID	Sart Position	Amino acid sequence	End Position	Note
1	S1-1	1	MFVFLVLLPLVS	12	N/A
2	S1-2	7	LLPLVSSQCVNL	18	
3	S1-3	13	SQCVNLTTRTQL	24	N/A
4	S1-4	19	TTRTQLPPAYTN	30	
5	S1-5	25	PPAYTNSFTRGV	36	
6	S1-6	31	SFTRGVVYYPDKV	42	
7	S1-7	37	YYPDKVFRSSVL	48	
8	S1-8	43	FRSSVLHSTQDL	54	
9	S1-9	49	HSTQDLFLPFFS	60	
10	S1-10	55	FLPFFSNVTWFH	66	N/A
11	S1-11	61	NVTWFHAIHVSG	72	
12	S1-12	67	AIHVSGTNGTKR	78	
13	S1-13	73	TNGTKRFDNPVL	84	
14	S1-14	79	FDNPVLPFNDGV	90	
15	S1-15	85	PFNDGVYFASTE	96	
16	S1-16	91	YFASTEKSNIR	102	
17	S1-17	97	KSNIRGWIFGT	108	
18	S1-18	103	GWIFGTTLDSKT	114	
19	S1-19	109	TLDSKTQSLIV	120	
20	S1-20	115	QSLIVNNATNV	126	
21	S1-21	121	NNATNVVIKVCE	132	
22	S1-22	127	VIKVICEFQFCND	138	
23	S1-23	133	FQFCNDPFLGVY	144	
24	S1-24	139	PFLGVVYHKNNK	150	
25	S1-25	145	YHKNNKSWMESE	156	
26	S1-26	151	SWMESEFRVYSS	162	
27	S1-27	157	FRVYSSANNCTF	168	
28	S1-28	163	ANNCTFEYVSQP	174	
29	S1-29	169	EYVSQPFMDLE	180	
30	S1-30	175	FLMDLEGKQGNF	186	
31	S1-31	181	GKQGNFKNLREF	192	
32	S1-32	187	KNLREFVFKNID	198	N/A
33	S1-33	193	VFKNIDGYFKIY	204	N/A
34	S1-34	199	GYFKIYSKHTPI	210	
35	S1-35	205	SKHTPINLVRDL	216	N/A
36	S1-36	211	NLVRDLPQGFSFA	222	
37	S1-37	217	PQGFSALEPLVD	228	N/A
38	S1-38	223	LEPLVDLPIGIN	234	
39	S1-39	229	LPIGINITRFQT	240	
40	S1-40	235	ITRFQTLALHR	246	N/A
41	S1-41	241	LLALHRSYLTPG	252	
42	S1-42	247	SYLTPGDSSSGW	258	
43	S1-43	253	DSSSGWTAGAAA	264	
44	S1-44	259	TAGAAAYVGYL	270	
45	S1-45	265	YVGYLQPRTFL	276	
46	S1-46	271	QPRTFLLYNEN	282	N/A
47	S1-47	277	LKYNENGTITDA	288	
48	S1-48	283	GTITDAVDCALD	294	
49	S1-49	289	VDCALDPLSETK	300	
50	S1-50	295	PLSETKCTLKSF	306	
51	S1-51	301	CTLKSFTVEKGI	312	
52	S1-52	307	TVEKGIYQTSNF	318	
53	S1-53	313	YQTSNFRVQPTE	324	
54	S1-54	319	RVQPTEIVRFP	330	
55	S1-55	325	SIVRFPNITNLC	336	
56	S1-56	331	NITNLCPFGEVF	342	
57	S1-57	337	PFGEVFNATRFA	348	
58	S1-58	343	NATRFAVYAWN	354	
59	S1-59	349	SVYAWNKRKRISN	360	
60	S1-60	355	RKRISNCVADYS	366	
61	S1-61	361	CVADYSVLYNSA	372	N/A
62	S1-62	367	VLYNSASFSTFK	378	
63	S1-63	373	SFSTFKCYGVSP	384	
64	S1-64	379	CYGVSPTKLNDL	390	

65	S1-65	385	TKLNDLCFTNVY	396	
66	S1-66	391	CFTNVYADSFVI	402	
67	S1-67	397	ADSFVIRGDEV	408	
68	S1-68	403	RGDEVQRQIAPGQ	414	
69	S1-69	409	QIAPGQTGKIAD	420	
70	S1-70	415	TGKIADYNYKLP	426	
71	S1-71	421	YNYKLPDDFTGC	432	
72	S1-72	427	DDFTGCVIAWNS	438	
73	S1-73	433	VIAWNSNNLDSK	444	
74	S1-74	439	NNLDSKVGGNYN	450	
75	S1-75	445	VGGNYNYLYRLF	456	N/A
76	S1-76	451	YLYRLFRRKSNLK	462	
77	S1-77	457	RKSNLKPFERDI	468	
78	S1-78	463	PFERDISTEIYQ	474	
79	S1-79	469	STEIYQAGSTPC	480	
80	S1-80	475	AGSTPCNGVEGF	486	
81	S1-81	481	NGVEGFNCYFPL	492	
82	S1-82	487	NCYFPLQSYGFQ	498	
83	S1-83	493	QSYGFQPTNGVG	504	
84	S1-84	499	PTNGVGYQPYRV	510	
85	S1-85	505	YQPYRVVLSFE	516	
86	S1-86	511	VVLSFELLHAPA	522	
87	S1-87	517	LLHAPATVCGPK	528	
88	S1-88	523	TVCGPKKSTNLV	534	
89	S1-89	529	KSTNLVKNKCVN	540	
90	S1-90	535	KNKCVNFNFNGL	546	
91	S1-91	541	FNFNGLTGTGVL	552	
92	S1-92	547	TGTGVLTESNKK	558	
93	S1-93	553	TESNKKFLPFQQ	564	
94	S1-94	559	FLPFQQFGRDIA	570	
95	S1-95	565	FGRDIADTTDAV	576	
96	S1-96	571	DTTDAVRDPQTL	582	
97	S1-97	577	RDPQTLLEILDIT	588	
98	S1-98	583	EILDITPCSFQGG	594	
99	S1-99	589	PCSFQGGVSVITP	600	
100	S1-100	595	VSVITPGTNTSN	606	
101	S1-101	601	GTNTSNQVAVLY	612	
102	S1-102	607	QVAVLYQDVNCT	618	
103	S1-103	613	QDVNCTEVPVAI	624	
104	S1-104	619	EVPVAIHADQLT	630	
105	S1-105	625	HADQLTPTWRVY	636	
106	S1-106	631	PTWRVYSTGSNV	642	
107	S1-107	637	STGSNVFQTRAG	648	N/A
108	S1-108	643	FQTRAGCLIGAE	654	
109	S1-109	649	CLIGAEHVNNNSY	660	
110	S1-110	655	HVNNSYECDIPI	666	
111	S1-111	661	ECDIPIGAGICA	672	
112	S1-112	667	GAGICASYQTQT	678	
113	S1-113	673	SYQTQTNSPRRA	684	
114	S1-114	679	NSPRRARGGGGS	685	
115	S2-1	686	SVASQSIIAYTM	697	N/A
116	S2-2	692	IIAYTMSLGAEN	703	N/A
117	S2-3	698	SLGAENSVAYSN	709	
118	S2-4	704	SVAYSNNSIAIP	715	
119	S2-5	710	NSIAIPTNFTIS	721	
120	S2-6	716	TNFTISVTTEIL	727	
121	S2-7	722	VTTEILPVSMTK	733	
122	S2-8	728	PVSMTKTSVDCT	739	
123	S2-9	734	TSVDCTMYICGD	745	
124	S2-10	740	MYICGDSTECNS	751	
125	S2-11	746	STECNSLLQYQ	757	
126	S2-12	752	LLLQYGSFCTQL	763	
127	S2-13	758	SFCTQLNRALTG	769	N/A
128	S2-14	764	NRALTGIAVEQD	775	
129	S2-15	770	IAVEQDKNTQEV	781	
130	S2-16	776	KNTQEVFAQVKQ	787	
131	S2-17	782	FAQVKQIYKTPP	793	

132	S2-18	788	IYKTPPIKDFGG	799	
133	S2-19	794	IKDFGGFNFSQI	805	
134	S2-20	800	FNFSQILPDPSK	811	
135	S2-21	806	LPDPSKPSKRSF	817	
136	S2-22	812	PSKRSFIEDLLF	823	
137	S2-23	818	IEDLLFNKVTLA	829	
138	S2-24	824	NKVTLADAGFIK	835	
139	S2-25	830	DAGFIKQYGDCL	841	
140	S2-26	836	QYGDCLGDIAAR	847	
141	S2-27	842	GDIAARDLICAQ	853	
142	S2-28	848	DLICAQKFNGLT	859	
143	S2-29	854	KFNGLTVLPPLL	865	
144	S2-30	860	VLPPLLTDEMIA	871	
145	S2-31	866	TDEMIAQYTSAL	877	
146	S2-32	872	QYTSALLAGTIT	883	
147	S2-33	878	LAGTITSGWTFG	889	
148	S2-34	884	SGWTFGAGAALQ	895	
149	S2-35	890	AGAALQIPFAMQ	901	
150	S2-36	896	IPFAMQMAYRFN	907	
151	S2-37	902	MAYRFNGIGVTQ	913	
152	S2-38	908	GIGVTQNVLYEN	919	
153	S2-39	914	NVLYENQKLIAN	925	
154	S2-40	920	QKLIANQFNSAI	931	
155	S2-41	926	QFNSAIGKIQDS	937	
156	S2-42	932	GKIQDSLSTAS	943	
157	S2-43	938	LSSTASALGKLQ	949	
158	S2-44	944	ALGKLQDVVNQN	955	
159	S2-45	950	DVVNQNAAQALNT	961	
160	S2-46	956	AQALNTLVKQLS	967	
161	S2-47	962	LVKQLSSNFGAI	973	
162	S2-48	968	SNFGAISSVLND	979	
163	S2-49	974	SSVLNDILSRLD	985	
164	S2-50	980	ILSRLDKVEAEV	991	
165	S2-51	986	KVEAEVQIDRLI	997	
166	S2-52	992	QIDRLITGRLQS	1003	
167	S2-53	998	TGRLQSLQTYVT	1009	
168	S2-54	1004	LQTYVTQQLIRA	1015	
169	S2-55	1010	QQLIRAAEIRAS	1021	
170	S2-56	1016	AEIRASANLAAT	1027	
171	S2-57	1022	ANLAATKMSECV	1033	
172	S2-58	1028	KMSECVLGQSKR	1039	
173	S2-59	1034	LGQSKRVDFCGK	1045	
174	S2-60	1040	VDFCGKGYHLMS	1051	
175	S2-61	1046	GYHLMSFPQSAP	1057	
176	S2-62	1052	FPQSAPHGVVFL	1063	
177	S2-63	1058	HGVVFLHVTYVP	1069	
178	S2-64	1064	HVTYVPAQEKNF	1075	
179	S2-65	1070	AQEKNFTTAPAI	1081	
180	S2-66	1076	TTAPAICHDGKA	1087	
181	S2-67	1082	CHDGKAHFPRREG	1093	
182	S2-68	1088	HFPREGVFSVNG	1099	
183	S2-69	1094	VFVSNGTHWFVT	1105	
184	S2-70	1100	THWFVTQRNFYE	1111	
185	S2-71	1106	QRNFYEPQIITT	1117	
186	S2-72	1112	PQIITTDNTFVS	1123	
187	S2-73	1118	DNTFVSGNCDVV	1129	
188	S2-74	1124	GNCDVVIGIVNN	1135	N/A
189	S2-75	1130	IGIVNNTVYDPL	1141	
190	S2-76	1136	TVYDPLQPELDS	1147	
191	S2-77	1142	QPELDSFKEELD	1153	
192	S2-78	1148	FKEELDKYFKNH	1159	
193	S2-79	1154	KYFKNHTSPDVD	1165	
194	S2-80	1160	TSPDVDLGDISG	1171	
195	S2-81	1166	LGDISGINASVV	1177	
196	S2-82	1172	INASVVNIQKEI	1183	
197	S2-83	1178	NIQKEIDRLNEV	1189	
198	S2-84	1184	DRLNEVAKNLNE	1195	

199	S2-85	1190	AKNLNESLIDLQ	1201	
200	S2-86	1196	SLIDLQELGKYE	1207	
201	S2-87	1202	ELGKYEQYIKWP	1213	
202	S2-88	1208	QYIKWPWYIWLG	1219	
203	S2-89	1214	WYIWLGFIAGLI	1225	N/A
204	S2-90	1220	FIAGLIAIVMVT	1231	N/A
205	S2-91	1226	AIVMVTIMLCCM	1237	N/A
206	S2-92	1232	IMLCCMTSCCSC	1243	N/A
207	S2-93	1238	TSCCCLKGCCS	1249	
208	S2-94	1244	LKGCCSCGSCCK	1255	N/A
209	S2-95	1250	CGSCCKFDEDDS	1261	
210	S2-96	1256	FDEDDSEPVKLG	1267	
211	S2-97	1262	EPVLKGVKLHYT	1273	
