



# Evaluation of SARS-CoV-2 serology assays reveals a range of test performance

Jeffrey D. Whitman<sup>1,38</sup>, Joseph Hiatt<sup>2,3,4,5,6,7,38</sup>, Cody T. Mowery<sup>2,3,5,6,7,38</sup>, Brian R. Shy<sup>1,38</sup>, Ruby Yu<sup>5,7,38</sup>, Tori N. Yamamoto<sup>5,7</sup>, Ujjwal Rathore<sup>4,5,6,7</sup>, Gregory M. Goldgof<sup>1</sup>, Caroline Whitty<sup>1,5,7</sup>, Jonathan M. Woo<sup>5,6,7</sup>, Antonia E. Gallman<sup>2,5,8</sup>, Tyler E. Miller<sup>9</sup>, Andrew G. Levine<sup>1</sup>, David N. Nguyen<sup>5,6,10</sup>, Sagar P. Bapat<sup>1,5,7</sup>, Joanna Balcerek<sup>1</sup>, Sophia A. Bylsma<sup>11</sup>, Ana M. Lyons<sup>12</sup>, Stacy Li<sup>12</sup>, Allison Wai-yi Wong<sup>2</sup>, Eva Mae Gillis-Buck<sup>13</sup>, Zachary B. Steinhart<sup>5,7</sup>, Youjin Lee<sup>5</sup>, Ryan Apathy<sup>5,6,7</sup>, Mitchell J. Lipke<sup>5,7</sup>, Jennifer Anne Smith<sup>7</sup>, Tina Zheng<sup>2,3,14,15</sup>, Ian C. Boothby<sup>2,16</sup>, Erin Isaza<sup>2,17</sup>, Jackie Chan<sup>5</sup>, Dante D. Acenas II<sup>5</sup>, Jinwoo Lee<sup>2,18</sup>, Trisha A. Macrae<sup>2,18</sup>, Than S. Kyaw<sup>2,5</sup>, David Wu<sup>2,3</sup>, Dianna L. Ng<sup>15,19</sup>, Wei Gu<sup>1</sup>, Vanessa A. York<sup>20</sup>, Haig Alexander Eskandarian<sup>20</sup>, Perri C. Callaway<sup>20,21</sup>, Lakshmi Warriar<sup>20</sup>, Mary E. Moreno<sup>20</sup>, Justine Levan<sup>20</sup>, Leonel Torres<sup>20</sup>, Lila A. Farrington<sup>20</sup>, Rita P. Loudermilk<sup>22</sup>, Kanishka Koshal<sup>22</sup>, Kelsey C. Zorn<sup>23</sup>, Wilfredo F. Garcia-Beltran<sup>9</sup>, Diane Yang<sup>9</sup>, Michael G. Astudillo<sup>9</sup>, Bradley E. Bernstein<sup>9</sup>, Jeffrey A. Gelfand<sup>24</sup>, Edward T. Ryan<sup>24</sup>, Richelle C. Charles<sup>24</sup>, A. John Iafrate<sup>9</sup>, Jochen K. Lennerz<sup>9</sup>, Steve Miller<sup>1</sup>, Charles Y. Chiu<sup>1,10,25</sup>, Susan L. Stramer<sup>26</sup>, Michael R. Wilson<sup>3,22</sup>, Aashish Manglik<sup>27,28</sup>, Chun Jimmie Ye<sup>29,30,31,32,33,34</sup>, Nevan J. Krogan<sup>4,35,36</sup>, Mark S. Anderson<sup>7</sup>, Jason G. Cyster<sup>5,8</sup>, Joel D. Ernst<sup>20</sup>, Alan H. B. Wu<sup>1</sup>, Kara L. Lynch<sup>1</sup>, Caryn Bern<sup>34</sup>✉, Patrick D. Hsu<sup>6,11,37</sup>✉ and Alexander Marson<sup>4,5,6,7,10,15,29,30,31,32</sup>✉

**Appropriate use and interpretation of serological tests for assessments of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) exposure, infection and potential immunity require accurate data on assay performance. We conducted a head-to-head evaluation of ten point-of-care-style lateral flow assays (LFAs) and two laboratory-based enzyme-linked immunosorbent assays to detect anti-SARS-CoV-2 IgM and IgG antibodies in 5-d time intervals from symptom onset and studied the specificity of each assay in pre-coronavirus disease 2019 specimens. The percent of seropositive individuals increased with time, peaking in the latest time interval tested (>20 d after symptom onset). Test specificity ranged from 84.3% to 100.0% and was predominantly affected by variability in IgM results. LFA specificity could be increased by considering weak bands as negative, but this decreased detection of antibodies (sensitivity) in a subset of SARS-CoV-2 real-time PCR-positive cases. Our results underline the importance of seropositivity threshold determination and reader training for reliable LFA deployment. Although there was no standout serological assay, four tests achieved more than 80% positivity at later time points tested and more than 95% specificity.**

To date, hundreds of thousands of deaths have been attributed to coronavirus disease 2019 (COVID-19)<sup>1</sup>. Millions of infections by SARS-CoV-2, the virus responsible for COVID-19, have been reported, although its full extent has yet to be determined owing to limited testing<sup>2</sup>. Government interventions to slow viral spread have disrupted daily life and economic activity for billions of people. Strategies to ease restraints on human mobility and interaction without provoking a major resurgence of transmission and mortality will depend on accurate estimates of population levels of infection and immunity<sup>3</sup>. Current testing for the virus largely depends on labor-intensive molecular techniques<sup>4</sup>. Individuals with positive molecular tests represent only a small fraction of all infections, given limited deployment and the brief time window when real-time (RT)-PCR testing has the highest sensitivity<sup>5-7</sup>.

The proportion of undocumented cases in the original epidemic focus was estimated to be as high as 86%<sup>8</sup>, and asymptomatic infections are suspected to play a substantial role in transmission<sup>9-14</sup>.

Widely available, reliable antibody detection assays would enable more accurate estimates of SARS-CoV-2 prevalence and incidence. On February 4, 2020, the Secretary of the US Department of Health and Human Services issued an emergency use authorization (EUA) for the diagnosis of SARS-CoV-2<sup>15</sup>, allowing nucleic acid detection and immunoassay tests to be offered based on manufacturer-reported data without formal US Food and Drug Administration (FDA) clearance<sup>16</sup>. In response, dozens of companies began to market laboratory-based immunoassays and point-of-care (POC) tests. Rigorous, comparative performance data are crucial to inform clinical care and public health responses.

A full list of affiliations appears at the end of the paper.

**Table 1 | Demographics and clinical characteristics of patients who tested positive for SARS-CoV-2 by RT-PCR<sup>a</sup>**

Variable	All patients (n = 79)	0–5 d (n = 28)	6–10 d (n = 36)	11–15 d (n = 34)	16–20 d (n = 19)	>20 d (n = 11)
Age, mean (s.d.)	52.9 (15)	48.2 (15.0)	53.3 (15.1)	58.1 ± 15.1	56.6 (13.2)	55.5 (14.8)
Male sex (%)	54 (68)	15 (54)	24 (67)	21 (62)	12 (63)	8 (73)
<b>Racial or ethnic group</b>						
Hispanic/Latinx (%)	54 (68)	18 (64)	29 (81)	23 (68)	12 (63)	7 (64)
Asian (%)	7 (9)	3 (11)	2 (6)	4 (12)	3 (16)	0 (0)
White (%)	7 (9)	3 (11)	1 (3)	2 (6)	2 (11)	0 (0)
Black (%)	6 (8)	2 (7)	3 (8)	4 (12)	1 (5)	2 (18)
Other/not reported (%)	5 (6)	2 (7)	1 (3)	1 (3)	1 (5)	2 (18)
<b>Presenting symptoms</b>						
Cough (%)	72 (91)	24 (86)	33 (92)	31 (91)	17 (89)	9 (82)
Fever (%)	68 (86)	23 (82)	30 (83)	29 (85)	17 (89)	9 (82)
Myalgia (%)	29 (37)	8 (29)	12 (33)	13 (38)	8 (42)	3 (27)
Chest pain (%)	20 (25)	5 (18)	8 (22)	7 (21)	5 (26)	4 (36)
Headache (%)	20 (25)	4 (14)	11 (31)	9 (26)	6 (32)	4 (36)
Chills (%)	19 (24)	5 (18)	9 (25)	7 (21)	7 (37)	2 (18)
Sore throat (%)	19 (24)	4 (14)	11 (31)	8 (24)	5 (26)	3 (27)
Malaise (%)	17 (22)	4 (14)	7 (19)	9 (26)	4 (21)	1 (9)
Diarrhea (%)	13 (16)	4 (14)	7 (19)	6 (18)	4 (21)	1 (9)
Anorexia (%)	8 (10)	2 (7)	1 (3)	2 (6)	4 (21)	1 (9)
Nausea and/or vomiting (%)	8 (10)	2 (7)	2 (6)	2 (6)	2 (11)	1 (9)
Anosmia and/or dysgeusia (%)	4 (5)	1 (4)	1 (3)	2 (6)	0 (0)	1 (9)
<b>Chronic medical conditions</b>						
Hypertension (%)	36 (46)	11 (39)	17 (47)	21 (62)	11 (58)	6 (55)
Type 2 diabetes mellitus (%)	33 (42)	11 (39)	17 (47)	19 (56)	8 (42)	6 (55)
Obesity (%)	19 (24)	7 (25)	9 (25)	11 (32)	6 (32)	6 (55)
Chronic kidney disease (%)	10 (13)	4 (14)	3 (8)	6 (18)	4 (21)	3 (27)
Hypothyroid (%)	6 (8)	3 (11)	3 (8)	3 (9)	0 (0)	0 (0)
Solid organ transplant (%)	6 (8)	2 (7)	0 (0)	2 (6)	2 (11)	2 (18)
Coronary artery disease (%)	5 (6)	1 (4)	1 (3)	2 (6)	2 (11)	3 (27)
Asthma (%)	4 (5)	1 (4)	1 (3)	3 (9)	2 (11)	0 (0)
Congestive heart failure (%)	3 (4)	2 (7)	2 (6)	2 (6)	1 (5)	0 (0)
Liver disease (%)	3 (4)	0 (0)	1 (3)	2 (6)	1 (5)	1 (9)
Malignancy (%)	3 (4)	1 (4)	2 (6)	1 (3)	2 (11)	0 (0)
Emphysema (%)	2 (3)	0 (0)	1 (3)	1 (3)	1 (5)	1 (9)
Prior stroke (%)	2 (3)	1 (4)	0 (0)	0 (0)	0 (0)	1 (9)
HIV (%)	1 (1)	1 (4)	0 (0)	0 (0)	0 (0)	0 (0)
Other immune-compromised condition <sup>b</sup> (%)	5 (6)	1 (4)	1 (3)	3 (9)	2 (11)	1 (9)
<b>Highest level of care</b>						
Ambulatory <sup>c</sup> (%)	14 (18)	9 (32)	2 (6)	3 (9)	0 (0)	0 (0)
Admitted (%)	36 (46)	11 (39)	19 (53)	12 (35)	5 (26)	4 (36)
ICU (%)	29 (37)	8 (29)	15 (42)	19 (56)	14 (74)	7 (64)

<sup>a</sup>Baseline demographic characteristics, presenting symptoms, chronic medical conditions, initial disposition and highest-level outcome for all participants whose samples were included in each time interval for serological testing. Only one sample per patient was included in each time interval, and some individuals are represented by multiple samples in different time intervals. In total, we tested 128 samples taken from 79 patients who tested positive for SARS-CoV-2 by RT-PCR. <sup>b</sup>Other immune-compromised condition includes rheumatology patients (rheumatoid arthritis, psoriasis, Crohn's disease, ankylosing spondylitis and reactive arthritis), all of whom were taking immune modulating/suppressing therapies. <sup>c</sup>Ambulatory care includes outpatients as well as patients seen in the emergency department and not admitted.

We conducted a head-to-head comparison of serology tests available to our group in early April, comprising ten immunochromatographic LFAs and two enzyme-linked immunosorbent assays (ELISAs) (for details, see Supplementary Table 1). Specimens were

obtained from patients with SARS-CoV-2 that was confirmed by RT-PCR, contemporaneous patients with other respiratory pathogen testing and/or without SARS-CoV-2 by RT-PCR and blood donor specimens collected before 2019. We included analyses of

**Table 2 | Summary statistics for immunochromatographic LFAs and ELISAs<sup>a</sup>**

Percentage of positive specimens from patients who tested positive for SARS-CoV-2 by RT-PCR												
Assay	IgM				IgG				IgM or IgG			
	Total n	Positive	%	95% CI	Total n	Positive	%	95% CI	Total n	Positive	%	95% CI
<b>Immunochromatographic LFAs</b>												
Biomedomics												
1-5 d	27	7	<b>25.9</b>	11.1-46.3	27	6	<b>22.2</b>	8.6-42.3	27	8	<b>29.6</b>	13.8-50.2
6-10 d	36	22	<b>61.1</b>	43.5-76.9	36	19	<b>52.8</b>	35.5-69.6	36	23	<b>63.9</b>	46.2-79.2
11-15 d	33	25	<b>75.8</b>	57.7-88.9	33	23	<b>69.7</b>	51.3-84.4	33	26	<b>78.8</b>	61.1-91.0
16-20 d	19	16	<b>84.2</b>	60.4-96.6	19	14	<b>73.7</b>	48.8-90.9	19	17	<b>89.5</b>	66.9-98.7
>20 d	11	9	<b>81.8</b>	48.2-97.7	11	9	<b>81.8</b>	48.2-97.7	11	9	<b>81.8</b>	48.2-97.7
Bioperfectus												
1-5 d	28	11	<b>39.3</b>	21.5-59.4	28	7	<b>25.0</b>	10.7-44.9	28	11	<b>39.3</b>	21.5-59.4
6-10 d	35	26	<b>74.3</b>	56.7-87.5	35	23	<b>65.7</b>	47.8-80.9	35	27	<b>77.1</b>	59.9-89.6
11-15 d	34	28	<b>82.4</b>	65.5-93.2	34	27	<b>79.4</b>	62.1-91.3	34	30	<b>88.2</b>	72.5-96.7
16-20 d	19	16	<b>84.2</b>	60.4-96.6	19	14	<b>73.7</b>	48.8-90.9	19	17	<b>89.5</b>	66.9-98.7
>20 d	10	10	<b>100.0</b>	69.2-100.0	10	9	<b>90.0</b>	55.5-99.7	10	10	<b>100.0</b>	69.2-100.0
DecomBio												
1-5 d	26	8	<b>30.8</b>	14.3-51.8	26	7	<b>26.9</b>	11.6-47.8	26	8	<b>30.8</b>	14.3-51.8
6-10 d	36	24	<b>66.7</b>	49.0-81.4	36	24	<b>66.7</b>	49.0-81.4	36	24	<b>66.7</b>	49.0-81.4
11-15 d	33	29	<b>87.9</b>	71.8-96.6	33	29	<b>87.9</b>	71.8-96.6	33	29	<b>87.9</b>	71.8-96.6
16-20 d	18	14	<b>77.8</b>	52.4-93.6	18	14	<b>77.8</b>	52.4-93.6	18	14	<b>77.8</b>	52.4-93.6
>20 d	11	10	<b>90.9</b>	58.7-99.8	11	10	<b>90.9</b>	58.7-99.8	11	10	<b>90.9</b>	58.7-99.8
DeepBlue												
1-5 d	28	12	<b>42.9</b>	24.5-62.8	28	6	<b>21.4</b>	8.3-41.0	28	12	<b>42.9</b>	24.5-62.8
6-10 d	36	28	<b>77.8</b>	60.8-89.9	36	18	<b>50.0</b>	32.9-67.1	36	28	<b>77.8</b>	60.8-89.9
11-15 d	34	28	<b>82.4</b>	65.5-93.2	34	21	<b>61.8</b>	43.6-77.8	34	28	<b>82.4</b>	65.5-93.2
16-20 d	19	16	<b>84.2</b>	60.4-96.6	19	15	<b>78.9</b>	54.4-93.9	19	17	<b>89.5</b>	66.9-98.7
>20 d	11	10	<b>90.9</b>	58.7-99.8	11	9	<b>81.8</b>	48.2-97.7	11	10	<b>90.9</b>	58.7-99.8
Innovita												
1-5 d	27	4	<b>14.8</b>	4.2-33.7	27	7	<b>25.9</b>	11.1-46.3	27	7	<b>25.9</b>	11.1-46.3
6-10 d	36	12	<b>33.3</b>	18.6-51.0	36	17	<b>47.2</b>	30.4-64.5	36	20	<b>55.6</b>	38.1-72.1
11-15 d	31	12	<b>38.7</b>	21.8-57.8	32	25	<b>78.1</b>	60.0-90.7	32	25	<b>78.1</b>	60.0-90.7
16-20 d	13	4	<b>30.8</b>	9.1-61.4	13	9	<b>69.2</b>	38.6-90.9	13	9	<b>69.2</b>	38.6-90.9
>20 d	6	1	<b>16.7</b>	0.4-64.1	6	4	<b>66.7</b>	22.3-95.7	6	5	<b>83.3</b>	35.9-99.6
Premier												
1-5 d	28	10	<b>35.7</b>	18.6-55.9	28	6	<b>21.4</b>	8.3-41.0	28	10	<b>35.7</b>	18.6-55.9
6-10 d	35	25	<b>71.4</b>	53.7-85.4	35	18	<b>51.4</b>	34.0-68.6	35	25	<b>71.4</b>	53.7-85.4
11-15 d	34	28	<b>82.4</b>	65.5-93.2	34	22	<b>64.7</b>	46.5-80.3	34	29	<b>85.3</b>	68.9-95.0
16-20 d	19	16	<b>84.2</b>	60.4-96.6	19	14	<b>73.7</b>	48.8-90.9	19	17	<b>89.5</b>	66.9-98.7
>20 d	11	10	<b>90.9</b>	58.7-99.8	11	9	<b>81.8</b>	48.2-97.7	11	10	<b>90.9</b>	58.7-99.8
Sure												
1-5 d	28	3	<b>10.7</b>	2.3-28.2	28	5	<b>17.9</b>	6.1-36.9	28	5	<b>17.9</b>	6.1-36.9
6-10 d	35	15	<b>42.9</b>	26.3-60.6	35	19	<b>54.3</b>	36.6-71.2	35	19	<b>54.3</b>	36.6-71.2
11-15 d	34	22	<b>64.7</b>	46.5-80.3	34	25	<b>73.5</b>	55.6-87.1	34	25	<b>73.5</b>	55.6-87.1
16-20 d	19	14	<b>73.7</b>	48.8-90.9	19	14	<b>73.7</b>	48.8-90.9	19	15	<b>78.9</b>	54.4-93.9
>20 d	11	8	<b>72.7</b>	39.0-94.0	11	10	<b>90.9</b>	58.7-99.8	11	10	<b>90.9</b>	58.7-99.8
UCP												
1-5 d	28	7	<b>25.0</b>	10.7-44.9	28	7	<b>25.0</b>	10.7-44.9	28	7	<b>25.0</b>	10.7-44.9
6-10 d	36	21	<b>58.3</b>	40.8-74.5	36	18	<b>50.0</b>	32.9-67.1	36	21	<b>58.3</b>	40.8-74.5

Continued

**Table 2 | Summary statistics for immunochromatographic LFAs and ELISAs<sup>a</sup> (continued)**

Percentage of positive specimens from patients who tested positive for SARS-CoV-2 by RT-PCR												
Assay	IgM				IgG				IgM or IgG			
	Total n	Positive	%	95% CI	Total n	Positive	%	95% CI	Total n	Positive	%	95% CI
11-15 d	34	26	<b>76.5</b>	58.8-89.3	34	25	<b>73.5</b>	55.6-87.1	34	27	<b>79.4</b>	62.1-91.3
16-20 d	19	15	<b>78.9</b>	54.4-93.9	19	14	<b>73.7</b>	48.8-90.9	19	15	<b>78.9</b>	54.4-93.9
>20 d	11	10	<b>90.9</b>	58.7-99.8	11	9	<b>81.8</b>	48.2-97.7	11	10	<b>90.9</b>	58.7-99.8
<b>VivaChek</b>												
1-5 d	25	7	<b>28.0</b>	12.1-49.4	25	7	<b>28.0</b>	12.1-49.4	25	7	<b>28.0</b>	12.1-49.4
6-10 d	35	22	<b>62.9</b>	44.9-78.5	35	22	<b>62.9</b>	44.9-78.5	35	22	<b>62.9</b>	44.9-78.5
11-15 d	30	26	<b>86.7</b>	69.3-96.2	30	25	<b>83.3</b>	65.3-94.4	30	26	<b>86.7</b>	69.3-96.2
16-20 d	19	15	<b>78.9</b>	54.4-93.9	19	14	<b>73.7</b>	48.8-90.9	19	15	<b>78.9</b>	54.4-93.9
>20 d	10	9	<b>90.0</b>	55.5-99.7	10	9	<b>90.0</b>	55.5-99.7	10	9	<b>90.0</b>	55.5-99.7
<b>Wondfo</b>												
1-5 d									26	10	<b>38.5</b>	20.2-59.4
6-10 d									36	24	<b>66.7</b>	49.0-81.4
11-15 d									32	27	<b>84.4</b>	67.2-94.7
16-20 d									19	17	<b>89.5</b>	66.9-98.7
>20 d									11	9	<b>81.8</b>	48.2-97.7
<b>ELISAs</b>												
<b>Epitope</b>												
1-5 d	28	5	<b>17.9</b>	6.1-36.9	28	11	<b>39.3</b>	21.5-59.4	28	11	<b>39.3</b>	39.3-21.5
6-10 d	36	19	<b>52.8</b>	35.5-69.6	36	28	<b>77.8</b>	60.8-89.9	36	29	<b>80.6</b>	80.6-64.0
11-15 d	34	27	<b>79.4</b>	62.1-91.3	34	31	<b>91.2</b>	76.3-98.1	34	31	<b>91.2</b>	91.2-76.3
16-20 d	19	14	<b>73.7</b>	48.8-90.9	19	16	<b>84.2</b>	60.4-96.6	19	17	<b>89.5</b>	89.5-66.9
>20 d	11	9	<b>81.8</b>	48.2-97.7	11	10	<b>90.9</b>	58.7-99.8	11	10	<b>90.9</b>	90.9-58.7
<b>In-house<sup>a</sup></b>												
1-5 d									28	10	<b>35.7</b>	18.6-55.9
6-10 d									36	26	<b>72.2</b>	54.8-85.8
11-15 d									34	32	<b>94.1</b>	80.3-99.3
16-20 d									19	17	<b>89.5</b>	66.9-98.7
>20 d									11	9	<b>81.8</b>	48.2-97.7
<b>Specificity in 108 blood donor plasma specimens collected before July 2018</b>												
Assay	IgM				IgG				IgM or IgG			
	Total n	Positive	Specificity (%)	95% CI	Total n	Positive	Specificity (%)	95% CI	Total n	Positive	Specificity (%)	95% CI
<b>Immunochromatographic LFAs</b>												
Biomedomics	107	13	<b>87.9</b>	80.1-93.4	107	4	<b>96.3</b>	90.7-99.0	107	14	<b>86.9</b>	79.0-92.7
Bioperfectus	104	3	<b>97.1</b>	91.8-99.4	104	2	<b>98.1</b>	93.2-99.8	104	5	<b>95.2</b>	89.1-98.4
DecomBio	107	10	<b>90.7</b>	83.5-95.4	107	9	<b>91.6</b>	84.6-96.1	107	11	<b>89.7</b>	82.3-94.8
DeepBlue	108	17	<b>84.3</b>	76.0-90.6	108	1	<b>99.1</b>	94.9-100.0	108	17	<b>84.3</b>	76.0-90.6
Innovita	108	4	<b>96.3</b>	90.8-99.0	108	0	<b>100.0</b>	96.6-100.0	108	4	<b>96.3</b>	90.8-99.0
Premier	108	2	<b>98.1</b>	93.5-99.8	108	1	<b>99.1</b>	94.9-100.0	108	3	<b>97.2</b>	92.1-99.4
Sure	108	0	<b>100.0</b>	96.6-100.0	108	0	<b>100.0</b>	96.6-100.0	108	0	<b>100.0</b>	96.6-100.0
UCP	107	2	<b>98.1</b>	93.4-99.8	107	2	<b>98.1</b>	93.4-99.8	107	2	<b>98.1</b>	93.4-99.8
VivaChek	99	5	<b>94.9</b>	88.6-98.3	99	4	<b>96.0</b>	90.0-98.9	99	5	<b>94.9</b>	88.6-98.3
Wondfo									106	1	<b>99.1</b>	94.9-100.0
<b>ELISAs</b>												
Epitope	108	3	<b>97.2</b>	92.1-99.4	108	10	<b>90.7</b>	83.6-95.5	108	11	<b>89.8</b>	82.5-94.8
In-house <sup>b</sup>									108	1	<b>99.1</b>	94.9-100.0

Continued

**Table 2 | Summary statistics for immunochromatographic LFAs and ELISAs<sup>a</sup> (continued)**

Percentage of positive specimens from individuals who tested positive for non-SARS-CoV-2 viral infections and/or tested negative for SARS-CoV-2 by RT-PCR

Assay	IgM				IgG				IgM or IgG			
	Total n	Positive	%	95% CI	Total n	Positive	%	95% CI	Total n	Positive	%	95% CI
<b>Immunochromatographic LFA</b>												
Biomedomics	51	8	15.7	7.0–28.6	51	4	7.8	2.2–18.9	51	11	21.6	11.3–35.3
Bioperfectus	45	5	11.1	3.7–24.1	45	6	13.3	5.1–26.8	45	8	17.8	8.0–32.1
DecomBio	51	5	9.8	3.3–21.4	51	2	3.9	0.5–13.5	51	6	11.8	4.4–23.9
DeepBlue	51	14	27.5	15.9–41.7	51	7	13.7	5.7–26.3	51	14	27.5	15.9–41.7
Innovita	28	2	7.1	0.9–23.5	28	2	7.1	0.9–23.5	28	3	10.7	2.3–28.2
Premier	51	0	0.0	0.0–7.0	51	1	2.0	0.0–10.4	51	2	3.9	0.5–13.5
Sure	51	0	0.0	0.0–7.0	51	0	0.0	0.0–7.0	51	0	0.0	0.0–7.0
UCP	51	3	5.9	1.2–16.2	51	2	3.9	0.5–13.5	51	3	5.9	1.2–16.2
VivaChek	48	4	8.3	2.3–20.0	48	1	2.1	0.1–11.1	48	4	8.3	2.3–20.0
WondFo									41	0	0.0	0.0–8.6
<b>ELISAs</b>												
Epitope	51	2	3.9	0.5–13.5	51	7	13.7	5.7–26.3	51	8	15.7	7.0–28.6
In-house <sup>b</sup>									51	7	13.7	5.7–26.3

<sup>a</sup>Samples are binned by time after patient-reported symptom onset for SARS-CoV-2 RT-PCR-positive cases. Percent of seropositivity assessed by each assay in SARS-CoV-2 RT-PCR-positive samples is reported with 95% CIs. The column 'IgM or IgG' refers to positivity of either isotype. Specificity is determined relative to pre-COVID-19 negative control serum samples. Percent of seropositivity assessed by each assay is reported with 95% CIs for samples from individuals who tested positive for non-SARS-CoV-2 viral infections and/or tested negative for SARS-CoV-2 by RT-PCR. <sup>b</sup>In-house ELISA secondary antibody is directed against human IgG F(ab)<sub>2</sub>. Per the manufacturer, some cross-reactivity with IgM and IgA isotypes cannot be excluded.

goal was to provide well-controlled performance data to help guide the use of serology in the response to COVID-19.

## Results

**Study population.** This study included 128 plasma or serum specimens from 79 individuals who tested positive for SARS-CoV-2 and who were diagnosed in the University of California, San Francisco (UCSF) hospital system and Zuckerberg San Francisco General (ZSFG) Hospital. Patients ranged from 22 to over 90 years of age (Table 1). Most patients were Hispanic/Latinx (68%), reflecting the ZSFG patient population and demographics of the epidemic in San Francisco<sup>17,18</sup>. Most presented with cough (91%) and fever (86%). Chronic medical conditions, such as hypertension, type 2 diabetes mellitus, obesity and chronic kidney disease, were frequent. Of the 79 individuals, 18% were not admitted, 46% were inpatients without intensive care unit (ICU) care and 37% required ICU care. There were no reported deaths at the time of chart review.

**Test performance.** Because we lacked a gold standard against which to benchmark the 12 tests in our study, we assessed the positive percent agreement (positivity) compared with the RT-PCR assay. The percentage of specimens testing positive rose with increasing time from symptom onset (Table 2 and Fig. 1a), reaching the highest levels in the 16–20-d and >20-d time intervals. The highest detection rate was achieved by combining IgM and IgG results (Fig. 1b). However, 95% confidence intervals (CIs) for later time intervals showed substantial overlap with those for earlier intervals (Fig. 1b). Four assays (Bioperfectus, Premier, Wondfo and in-house ELISA) achieved more than 80% positivity in the later two time intervals (16–20 d and >20 d) while maintaining more than 95% specificity. Some tests were not performed on a subset of specimens owing to exhausted sample material, which might have affected reported percent positivity; the sample size tested is reflected in 95% CIs. IgM detection was less consistent than IgG for nearly all assays. Kappa agreement statistic ranged from 0.95 to 0.99 for IgG and from 0.81 to 1.00 for IgM for standardized intensity score (Supplementary Table 2 and

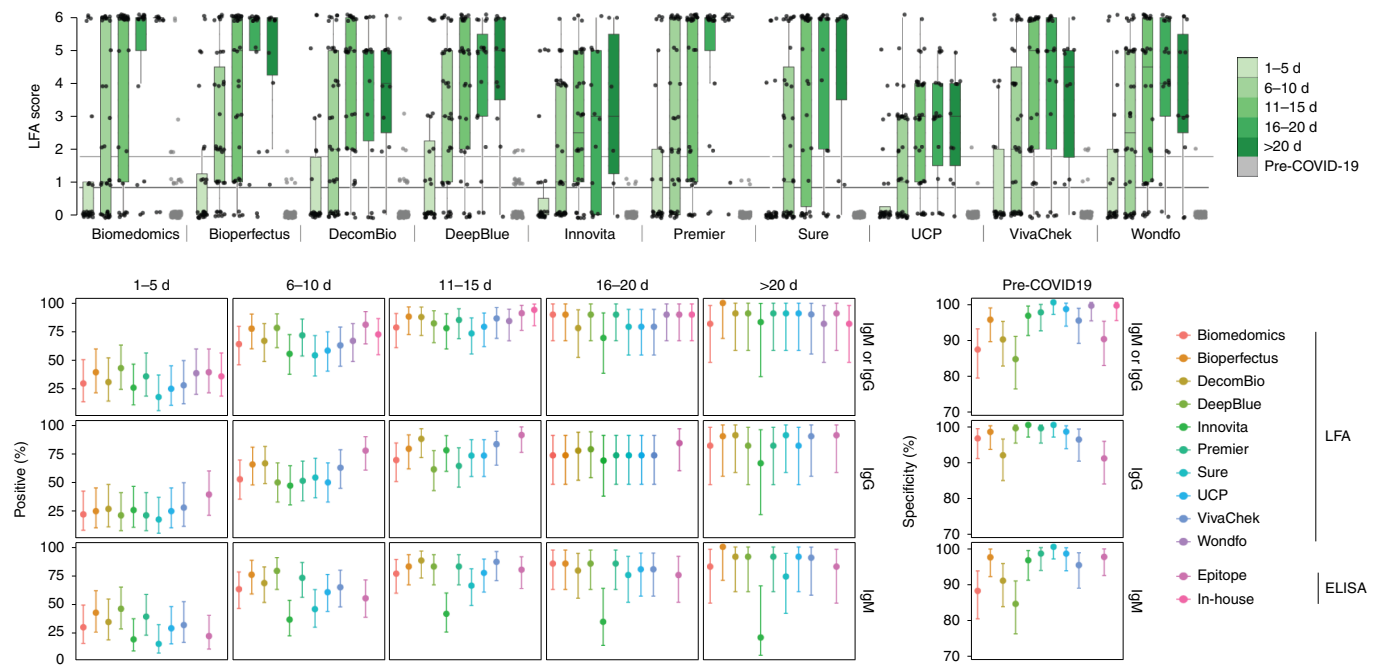
and reader training are available in the 'Immunochromatographic LFAs' section within Methods. Although mean band intensities varied among different assays, the approximate rate of sample positivity was generally consistent (Fig. 2). For ELISA tests, a normalized value of sample optical density at 450 nm (OD450) divided by calculated cutoff (signal-to-cutoff (S/CO)) was used to capture quantitative data about antibody levels for each specimen. S/CO values provide a quantitative value comparable between plates. Our ability to perform end-point dilutions was limited by specimen and assay availability.

We observed a trend toward higher percent positivity by LFA for patients admitted to the ICU compared to those having milder disease, but the specimen numbers per time interval were low, limiting statistical power (Supplementary Fig. 3).

Test specificity in 108 pre-COVID-19 blood donor plasma samples ranged from 84.3% to 100.0%, with 39 samples demonstrating false-positive results by at least one LFA (Table 2 and Fig. 2b). Of the false-positive results, 61.5% (24/39) had a weak intensity score of 1. Intensity scores of 2–3 were seen in 30.8% (12/39), and scores of 4–6 were seen in 7.7% (3/39).

We evaluated the tradeoff between percent positivity in samples from RT-PCR-positive individuals and specificity as a function of LFA reader score. RT-PCR measures the presence of viral nucleotides. Individuals with RT-PCR-proven SARS-CoV-2 infection are expected to seroconvert and develop anti-SARS-CoV-2 antibodies, although frequency and kinetics of seroconversion can vary<sup>5,6,19–22</sup>. We, therefore, assessed percent positivity at various time intervals after onset of symptoms. Changing the positive LFA threshold from 1 to 2 decreased the mean overall percent positivity across tests from 67.2% (range, 57.9–75.4%) to 57.8% (range, 44.7–65.6%) and increased the average specificity from 94.2% (range, 84.3–100.0%) to 98.1% (range, 94.4–100.0%) (Fig. 3).

An independent study at Massachusetts General Hospital (MGH) compared three LFAs, of which BioMedomics was also assessed in the current study (Supplementary Table 3). Although study design and methods differed between sites, precluding direct



**Fig. 1 | Performance data for immunochromatographic LFAs.** **a**, The second reader's score (0–6 based on band intensity) is reported for each assay, binned by time after patient-reported symptom onset. Biologically independent samples for each test are as follows:  $n=126$ , Biomedomics;  $n=126$ , Bioperfectus;  $n=124$ , Decombio;  $n=128$ , DeepBlue;  $n=114$ , Innovita;  $n=127$ , Premier;  $n=127$ , Sure;  $n=128$ , UCP;  $n=119$ , VivaChek;  $n=124$ , Wondfo. The second reader's score for pre-COVID-19 samples is also displayed ( $n=107$ , Biomedomics;  $n=104$ , Bioperfectus;  $n=107$ , Decombio;  $n=108$ , DeepBlue;  $n=108$ , Innovita;  $n=108$ , Premier;  $n=108$ , Sure;  $n=107$ , UCP;  $n=99$ , VivaChek;  $n=106$ , Wondfo). For tests with separate IgG and IgM bands, the higher score is reported. Joint IgM/IgG signal is represented by a single band in Wondfo. The lower, dark gray line refers to the positivity threshold (score greater than or equal to 1) used in this study. The upper, light gray line refers to an alternative positivity threshold (score greater than or equal to 2) discussed in the text and Fig. 3. Box spans 25th to 75th percentiles with median indicated by the black bar; whiskers show maximum and minimum value within 1.5 $\times$  the interquartile range from the box. **b**, Percent of SARS-CoV-2 RT-PCR-positive samples testing positive by each LFA and ELISA are plotted relative to time after patient-reported symptom onset ( $n=126$ , Biomedomics;  $n=126$ , Bioperfectus;  $n=124$ , Decombio;  $n=128$ , DeepBlue;  $n=114$ , Innovita;  $n=127$ , Premier;  $n=127$ , Sure;  $n=128$ , UCP;  $n=119$ , VivaChek;  $n=124$ , Wondfo;  $n=128$ , Epitope;  $n=128$ , in-house). The 'IgM or IgG' category refers to positivity of either isotype. **c**, Specificity is plotted for each test using pre-COVID-19 negative control samples ( $n=107$ , Biomedomics;  $n=104$ , Bioperfectus;  $n=107$ , Decombio;  $n=108$ , DeepBlue;  $n=108$ , Innovita;  $n=108$ , Premier;  $n=108$ , Sure;  $n=107$ , UCP;  $n=99$ , VivaChek;  $n=106$ , Wondfo;  $n=108$ , Epitope;  $n=108$ , in-house). For **b** and **c**, all nodes refer to the calculated percent positivity or specificity, respectively. Error bars signify 95% CIs.

comparison of results (see 'Study design' in Methods), test validation efforts at another site provided additional useful data. Overall, both studies showed a trend for increased detection of SARS-CoV-2-specific antibodies with increased time from symptom onset. However, the MGH study displayed increased specificity with lower percent positivity at early time points after symptom onset. MGH positivity thresholds were set higher to prioritize test specificity (Fig. 3b,c).

A set of specimens collected during the COVID-19 outbreak that had negative SARS-CoV-2 RT-PCR testing and/or alternative respiratory pathogen testing demonstrated higher numbers of positive results compared to the pre-COVID-19 sample set (Fig. 2c). Five specimens had positive results by more than three tests, all with respiratory symptoms and concurrent negative or un-performed SARS-CoV-2 RT-PCR testing (Fig. 2c, arrows). One patient was positive on eight different tests, including the in-house ELISA. In this limited panel, no consistent pattern of cross-reactivity was identified with non-SARS-CoV-2 respiratory viruses, including two strains of seasonal coronavirus (one coronavirus OC43 and three coronavirus HKU1).

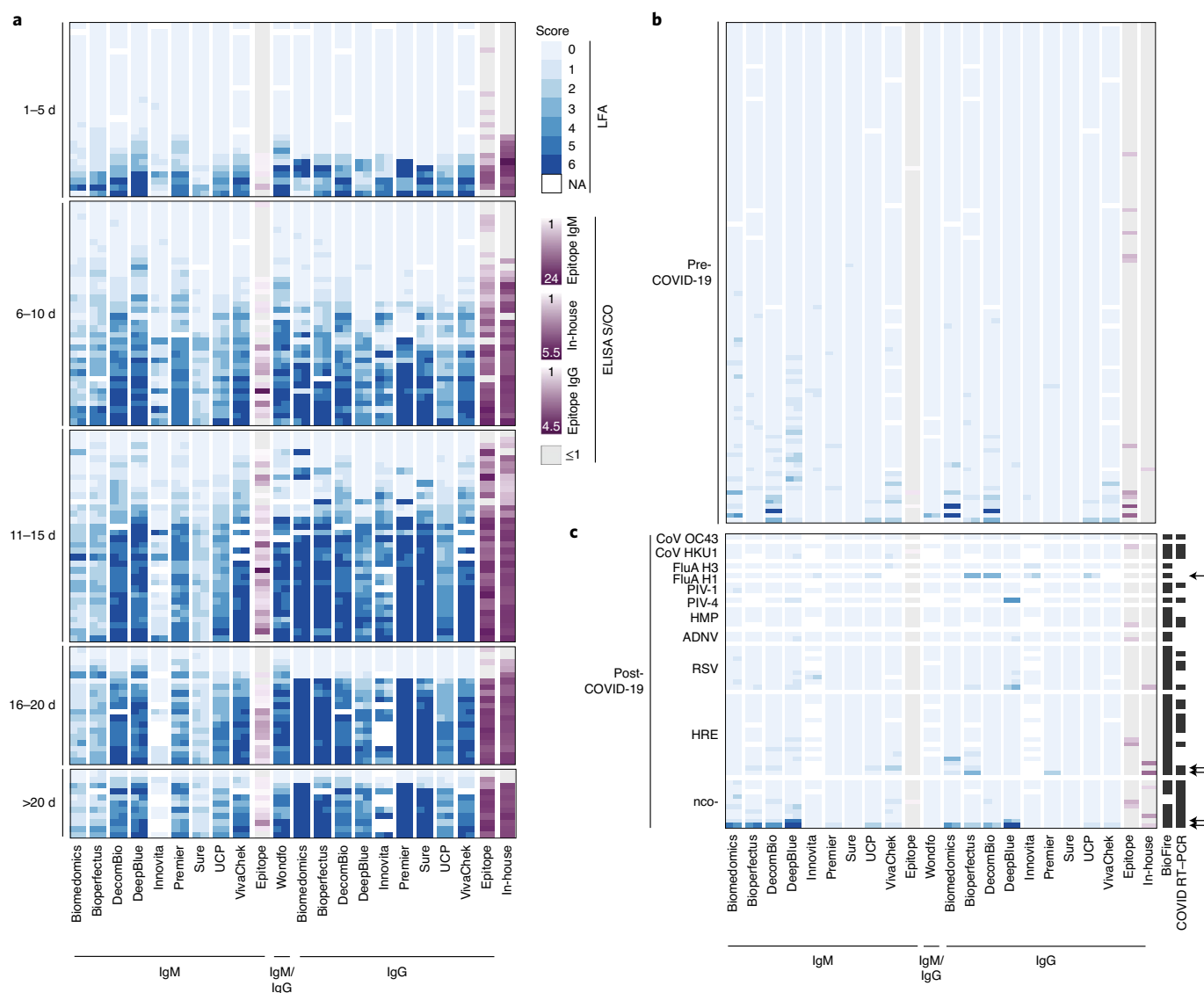
Agreement among results of LFAs with those of IgG and IgM Epitope ELISAs ranged from 75.7% to 85.6%, whereas agreement with the in-house ELISA ranged from 83.5% to 94.8% (Fig. 4a). LFA band intensity scores showed a direct correlation with ELISA S/CO values (Fig. 4b).

## Discussion

This study describes test performance for 12 COVID-19 serology assays on a panel of 128 samples from 79 individuals with RT-PCR-confirmed SARS-CoV-2 infection and 108 pre-COVID-19 specimens. In April 2020, when we performed this analysis, there was no assay with sufficient performance data for use as a proven reference standard; only three serological assays had an FDA EUA<sup>23</sup>; and anti-SARS-CoV-2 IgM and IgG kinetics were poorly understood. We, therefore, chose a specimen set covering the first several weeks after illness onset in patients with SARS-CoV-2 proven by RT-PCR to avoid the potential bias of assuming superiority of one assay over the others. To date, no single assay or combination of assays has been accepted as a gold standard comparator for antibody testing. Additionally, we surveyed 51 specimens from individuals who were tested for other respiratory viral pathogens and/or had negative molecular testing for SARS-CoV-2 to evaluate potential cross-reactivity or infections detected only by serology. Our data are also available on a dedicated website (<https://covidtestingproject.org>). We hope these data will inform the use of serology by the medical and public health communities and provide feedback to test developers about areas of success and necessary improvement.

We focused on comparisons of percent positivity by time interval, rather than reporting the 'sensitivity' of each assay, both because of the lack of a gold standard to test against and our expectation that percent positivity would rise with increasing time after symptom



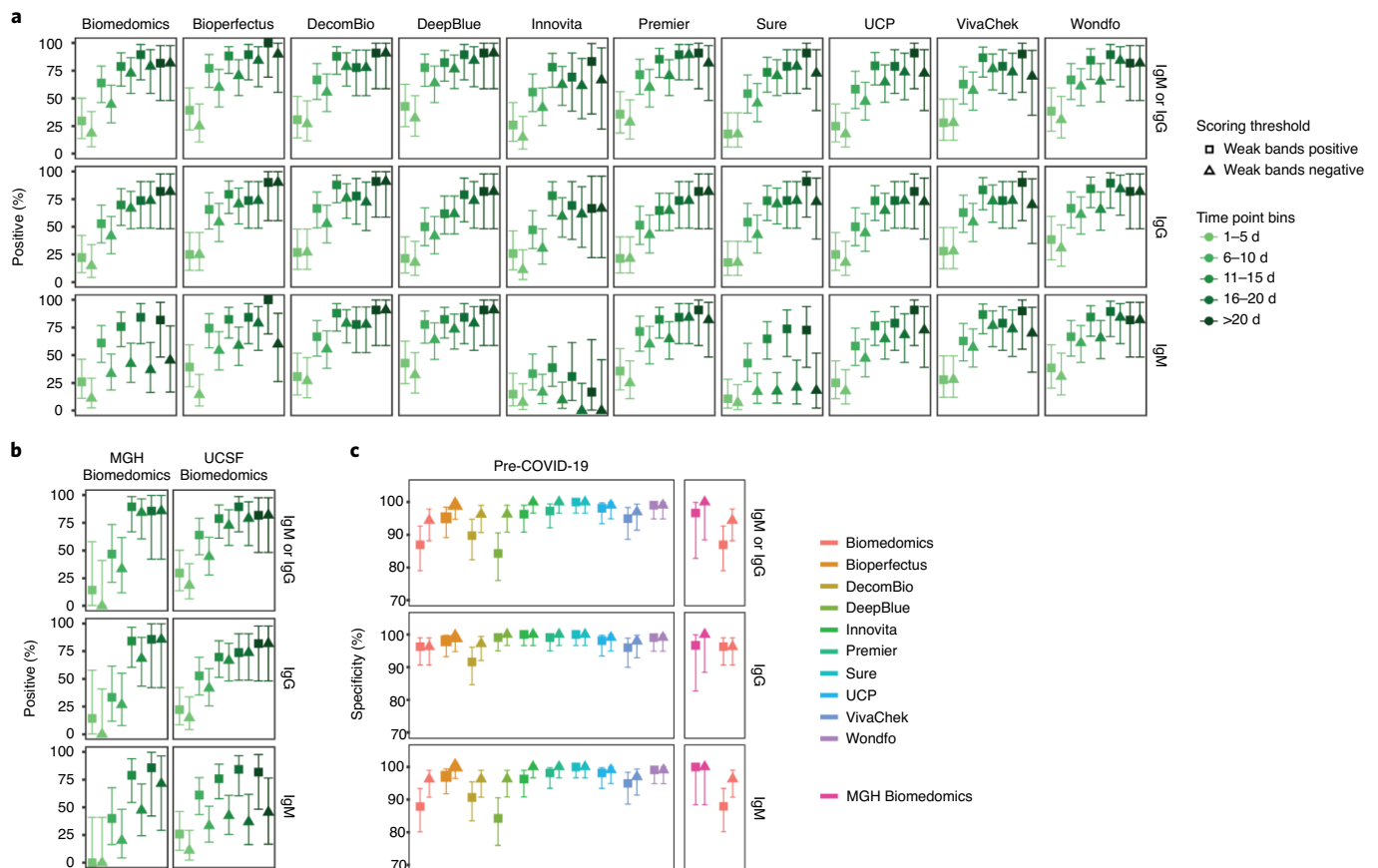


**Fig. 2 | LFA and ELISA values by serological assay. a**, LFA scores for each of two readers (blue) and mean ELISA S/CO (purple) for each specimen are grouped by binned time after patient-reported symptom onset and plotted by assay. White cells indicate samples not run with the corresponding assay. For ELISAs, gray indicates S/CO less than or equal to 1. The same legend applies to **b** and **c**. The F(ab')<sub>2</sub> specific secondary antibody used in our in-house ELISA preferentially binds the IgG light chain but per the manufacturer has some reactivity for other isotypes (IgM and IgA). **b**, LFA score and ELISA S/CO values are plotted for pre-COVID-19 historical control serum samples to determine assay specificity. **c**, LFA score and ELISA S/CO values are plotted for serum samples obtained from 51 individuals after the emergence of COVID-19 (post-COVID-19), some of which received BioFire FilmArray (BioFire Diagnostics) and/or SARS-CoV-2 RT-PCR testing (all negative) as indicated (black cells) in the appropriate columns. Arrows highlight specimens from five individuals with moderate to strong band intensity further discussed in the text. Specimens are grouped by positive testing for coronavirus HKU1 (CoV HKU1), coronavirus OC43 (CoV OC43), influenza A virus A/H3 (FluA H3), influenza A virus A/H1 2009 (FluA H1), parainfluenza type 1 virus (PIV-1), parainfluenza type 4 virus (PIV-4), human metapneumovirus (HMP), adenovirus (ADNV), respiratory syncytial virus (RSV), human rhinovirus/enterovirus (HRE) or negative testing for SARS-CoV-2 and other viruses (nco-).

onset<sup>5,6,19–22,24,25</sup>. Percent positivity above 80% was not reached until at least 2 weeks into clinical illness; diagnosis early in the course of illness remains dependent on viral detection methods. Our data are consistent with growing evidence that IgM and IgG tend to rise around the same time in COVID-19<sup>5,19</sup>. The assays showed a trend to higher positive rates within time intervals for more severe disease, but this finding should be interpreted with caution, owing to the limited data from ambulatory cases. Most samples more than 20 d after symptom onset had detectable anti-SARS-CoV-2 antibodies, suggesting good to excellent sensitivity for all evaluated tests in hospitalized patients three or more weeks into their disease

course. Additional studies assessing frozen versus fresh specimens and matrix effects between serum versus plasma will be useful in understanding potential limitations of our current test performance evaluations. Looking forward, well-powered studies testing ambulatory or asymptomatic individuals, including LFA performance with fresh capillary blood, will be essential to guide appropriate use of serology.

Our data demonstrate specificity of more than 95% for most tests evaluated and more than 99% for two LFAs (Wondfo and Sure Biotech) and the in-house ELISA (adapted from Amanat et al., 2020)<sup>26</sup>. We observed moderate to strong positive bands in several



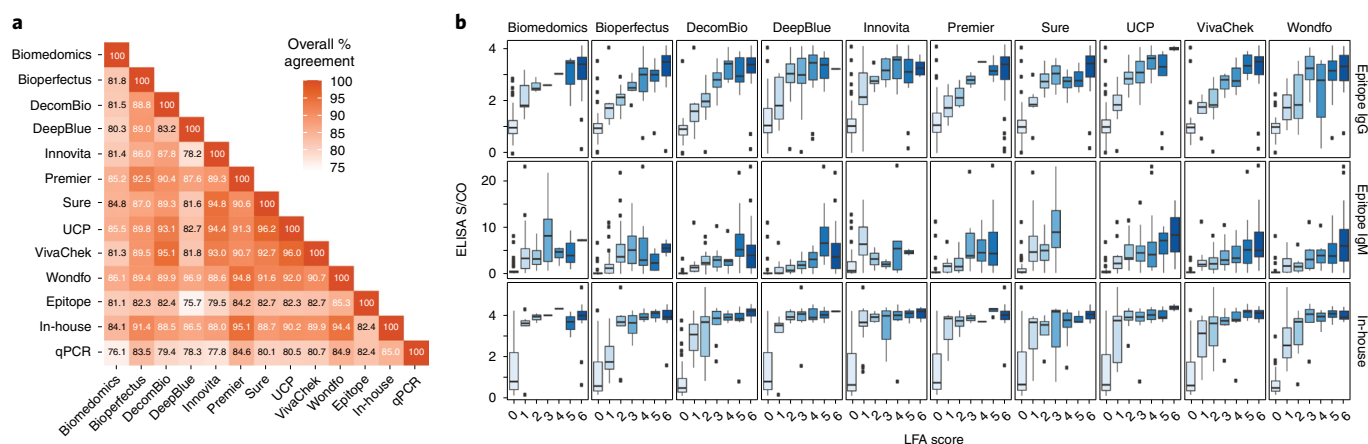
**Fig. 3 | Comparison of the effect of different positivity thresholds on percent positivity and specificity. a**, The percent positivity of each assay tested on serum from patients who tested positive for SARS-CoV-2 by RT-PCR is plotted by time after patient-reported symptom onset. Biologically independent samples for each test are as follows:  $n=126$ , Biomedomics;  $n=126$ , Bioperfectus;  $n=124$ , Decombio;  $n=128$ , DeepBlue;  $n=114$ , Innovita;  $n=127$ , Premier;  $n=127$ , Sure;  $n=128$ , UCP;  $n=119$ , VivaChek;  $n=124$ , Wondfo. Squares indicate percent positivity using reader score  $>0$  ('Weak bands positive') as the positivity threshold. Triangles indicate percent positivity using reader score  $>1$  ('Weak bands negative') as the positivity threshold. 'IgM or IgG' signifies detection of either isotype. Wondfo reports a single band for IgM and IgG together, and the results are plotted here as both 'IgM' and 'IgG' for horizontal comparison across assays. **b**, Comparison of percent positivity at each timepoint for BioMedomics assay at either the MGH (left,  $n=48$ ) or UCSF (right,  $n=126$ ) study site using low (square) or high (triangle) positivity thresholds. Note that a weak score at MGH is not directly equivalent to a 1 at UCSF owing to differences in reader training. **c**, The specificity of all assays on historical pre-COVID-19 serum using low (square) or high (triangle) positivity thresholds. UCSF BioMedomics data are plotted again in the right subpanel column for direct comparison to MGH BioMedomics data. All nodes refer to the calculated percent positivity or specificity (as designated), and all error bars indicate 95% CIs.

pre-COVID-19 blood donor specimens, some of them positive by multiple assays, suggesting the possibility of non-specific binding of plasma proteins, non-specific antibodies (potentially including auto-antibodies) or cross-reactivity with antibodies against other viruses. Three of the pre-COVID-19 specimens (2.8%) were scored positive by more than three assays. Intriguingly, the fraction of positive tests was higher in a set of recent specimens obtained during the COVID-19 outbreak from individuals undergoing respiratory infection workup, many with negative SARS-CoV-2 RT-PCR. Five of these (9.8%) had positive results by more than three assays, without relation to a specific viral pathogen, suggesting non-specific reactivity and/or missed COVID-19 diagnosis. Recent reports demonstrate that RT-PCR from nasopharyngeal swabs might yield false-negative results in over 20% of cases<sup>5,27</sup>, and co-infection with other respiratory pathogens might be significantly higher than previously anticipated<sup>28</sup>. One specimen was positive by 8 of 12 assays, including the in-house ELISA. The patient was over 90 years old and presented with altered mental status, fever and ground glass opacities on chest radiological imaging. SARS-CoV-2 RT-PCR was negative, and ancillary laboratory testing suggested a urinary

tract infection. This case could represent COVID-19 not detected by RT-PCR, reinforcing the importance of caution in interpreting negative molecular results as ruling out the infection. Appropriate algorithms for serology testing, including confirmatory or reflexive testing, have yet to be determined. These algorithms will be affected by test performance characteristics and prevalence of disease, as well as pre-test probability of infection.

Importantly, we still do not know the extent to which positive results by serology reflect a protective immune response, nor how long such protection might last<sup>29</sup>. Neutralization assays measure the ability of blood-derived samples to prevent viral (most commonly pseudovirus) infection of cultured cells in vitro<sup>30,31</sup>. Although these assays provide information on the functional capabilities of an individual's antibodies, their correlation with total IgG antibodies to serological test antigens (primarily spike and nucleocapsid proteins) is not well established. Additionally, most antibody neutralization assays are research laboratory based with limited test performance data and inter-lab standardization measures. Antibody neutralization assays should be harmonized across laboratories to establish the extent to which conventional serology assays correlate with





**Fig. 4 | Agreement of serological assays for SARS-CoV-2. a**, Percent agreement is plotted across all assay combinations, and values signify the binomial regression of the two assays across all tests. Samples were labeled ‘positive’ if any antibody isotype was detected for each assay. **b**, IgM or IgG LFA scores for each assay are compared to S/CO from three different ELISAs for all SARS-CoV-2 RT-PCR-positive samples. Biologically independent samples for each test are as follows:  $n = 126$ , Biomedomics;  $n = 126$ , Bioperfectus;  $n = 124$ , Decombio;  $n = 128$ , DeepBlue;  $n = 114$ , Innovita;  $n = 127$ , Premier;  $n = 127$ , Sure;  $n = 128$ , UCP;  $n = 119$ , VivaChek;  $n = 124$ , Wondfo. Joint IgM/IgG signal is represented by a single band in Wondfo, so data were plotted as IgM or IgG depending on ELISA comparison. The  $F(ab')_2$ -specific secondary antibody used in our in-house ELISA preferentially binds the IgG light chain but per the manufacturer contains some reactivity for other isotypes (IgM and IgA); it is compared in **b** to IgG band intensity. For **b**, the box spans the 25th to 75th percentiles with median indicated by the black bar; whiskers show maximum and minimum value within  $1.5\times$  the interquartile range from the box.

neutralization. Further studies are needed to assess the relationships among positive serological testing, *in vitro* viral neutralization results and clinical protection from future SARS-CoV-2 infection and transmission. Epidemiological data and results from convalescent plasma treatment trials should help guide clinical and public health policies for use of serological testing.

High specificity testing is crucial in low-prevalence settings. One approach to increase specificity would employ confirmatory testing with an independent assay (perhaps recognizing a distinct epitope or antigen). Our comparison of UCSF and MGH data suggests that reclassifying faint bands as ‘negative’ or ‘inconclusive’ can change test performance characteristics by increasing specificity, albeit at the expense of sensitivity. However, the subjectivity of calling faint bands by individual readers might be difficult to standardize without specific control materials, operator training and/or objective methods of analyzing LFAs. In the clinical setting, these parameters and protocols should be independently assessed and validated by clinical laboratories for operation under the Clinical Laboratory Improvement Amendments<sup>32</sup>.

Objective methods to standardize LFA reading, such as digital image analysis, are potentially attractive. Image analysis tools can be benchtop or mobile (for example, smartphone applications). However, introduction of a separate device for reading LFAs will require specific validation. Variables, including lighting, camera quality, image compression and quantification algorithms, must all be assessed rigorously to ensure accuracy and precision.

A consensus has emerged that serological testing provides an essential tool in the pandemic response, but inadequate data on test performance characteristics in some early surveys and important gaps in immunological knowledge have impeded agreement on appropriate implementation strategies<sup>33,34</sup>. Our study highlights the need for rigorous assay validation using standardized sample sets with: 1) known positives from individuals with a range of clinical presentations at multiple time points after onset of symptoms; 2) pre-COVID-19 outbreak samples for specificity; and 3) samples from individuals with other viral and inflammatory illnesses as cross-reactivity controls. Coordinated efforts to ensure widespread availability of validated sample sets would facilitate data-driven decisions on the use of serology. The updated guidance released by

the FDA in early May 2020<sup>35</sup> and the initiative recently launched by the FDA and the US National Cancer Institute/National Institutes of Health<sup>36</sup> to systematize data generation for EUAs are substantive steps toward this goal and will help build the essential evidence base to guide serological testing during the COVID-19 pandemic.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-020-0659-0>.

Received: 9 June 2020; Accepted: 29 July 2020;  
Published online: 27 August 2020

### References

- Johns Hopkins University. *COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)*, <https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6> (2020).
- Infectious Diseases Society of America. *IDSA Statement on COVID-19 Testing*, <https://www.idsociety.org/globalassets/idsa/public-health/covid-19-idsa-testing-intro.pdf> (2020).
- Weitz, J. S. et al. Modeling shield immunity to reduce COVID-19 epidemic spread. *Nat. Med.* <https://doi.org/10.1038/s41591-020-0895-3> (2020).
- World Health Organization. *Laboratory Testing for Coronavirus Disease 2019 (COVID-19) in Suspected Human Cases: Interim Guidance, 2 March 2020* (World Health Organization, 2020).
- Zhao, J. et al. Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. *Clin. Infect. Dis.*, <https://doi.org/10.1093/cid/ciaa344> (2020).
- Wölfel, R. et al. Virological assessment of hospitalized patients with COVID-2019. *Nature* <https://doi.org/10.1038/s41586-020-2196-x> (2020).
- He, X. et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* <https://doi.org/10.1038/s41591-020-0869-5> (2020).
- Li, R. et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* <https://doi.org/10.1126/science.abb3221> (2020).
- Bai, Y. et al. Presumed asymptomatic carrier transmission of COVID-19. *JAMA* <https://doi.org/10.1001/jama.2020.2565> (2020).

10. Du, Z. et al. Serial interval of COVID-19 among publicly reported confirmed cases. *Emerg. Infect. Dis.* <https://doi.org/10.3201/eid2606.200357> (2020).
11. Kimball, A. et al. Asymptomatic and presymptomatic SARS-CoV-2 infections in residents of a long-term care skilled nursing facility - King County, Washington, March 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 377–381 (2020).
12. Mizumoto, K., Kagaya, K., Zarebski, A. & Chowell, G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Euro. Surveill.* <https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180> (2020).
13. Qiu, H. et al. Clinical and epidemiological features of 36 children with coronavirus disease 2019 (COVID-19) in Zhejiang, China: an observational cohort study. *Lancet Infect. Dis.* [https://doi.org/10.1016/S1473-3099\(20\)30198-5](https://doi.org/10.1016/S1473-3099(20)30198-5) (2020).
14. Tong, Z. D. et al. Potential presymptomatic transmission of SARS-CoV-2, Zhejiang Province, China, 2020. *Emerg. Infect. Dis.* **26**, 1052–1054 (2020).
15. Secretary of Health and Human Services. Determination of a public health emergency and declaration that circumstances exist justifying authorizations pursuant to section 564(b) of the Federal Food, Drug, and Cosmetic Act, 21 U.S.C. § 360bbb-3. <https://www.fda.gov/media/135010/download> (2020).
16. Food and Drug Administration. Policy for diagnostic tests for coronavirus disease-2019 during the public health emergency. <https://www.fda.gov/media/135659/download> (2020).
17. San Francisco Department of Public Health. *San Francisco Department of Public Health Annual Report 2017-2018*, <https://www.sfdph.org/dph/files/reports/PolicyProcOfc/DPHAnnualReportFY17-18.pdf> (2018).
18. DataSF. *San Francisco COVID-19 Data Tracker*, <https://data.sfgov.org/stories/s/San-Francisco-COVID-19-Data-Tracker/fjki-2fab/> (2020).
19. Long, Q. X. et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat. Med.* <https://doi.org/10.1038/s41591-020-0897-1> (2020).
20. Tan, W. et al. Viral kinetics and antibody responses in patients with COVID-19. Preprint at *medRxiv* <https://doi.org/10.1101/2020.03.24.20042382> (2020).
21. Ng, D. et al. SARS-CoV-2 seroprevalence and neutralizing activity in donor and patient blood from the San Francisco Bay Area. Preprint at *medRxiv* <https://doi.org/10.1101/2020.05.19.20107482> (2020).
22. Lynch, K. L. et al. Magnitude and kinetics of anti-SARS-CoV-2 antibody responses and their relationship to disease severity. Preprint at *medRxiv* <https://doi.org/10.1101/2020.06.03.20121525> (2020).
23. U.S. Food and Drug Administration. *Emergency Use Authorization*, <https://www.fda.gov/emergency-preparedness-and-response/mcm-legal-regulatory-and-policy-framework/emergency-use-authorization> (2020).
24. Lassaunière, R. et al. Evaluation of nine commercial SARS-CoV-2 immunoassays. Preprint at *medRxiv* <https://doi.org/10.1101/2020.04.09.20056325> (2020).
25. To, K. K. et al. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect. Dis.* [https://doi.org/10.1016/S1473-3099\(20\)30196-1](https://doi.org/10.1016/S1473-3099(20)30196-1) (2020).
26. Amanat, F. et al. A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nat. Med.* <https://doi.org/10.1038/s41591-020-0913-5> (2020).
27. Wang, W. et al. Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA* <https://doi.org/10.1001/jama.2020.3786> (2020).
28. Kim, D., Quinn, J., Pinsky, B., Shah, N. H. & Brown, I. Rates of co-infection between SARS-CoV-2 and other respiratory pathogens. *JAMA* <https://doi.org/10.1001/jama.2020.6266> (2020).
29. Jiang, S., Hillyer, C. & Du, L. Neutralizing antibodies against SARS-CoV-2 and other human coronaviruses. *Trends Immunol.* <https://doi.org/10.1016/j.it.2020.03.007> (2020).
30. Nie, J. et al. Establishment and validation of a pseudovirus neutralization assay for SARS-CoV-2. *Emerg. Microbes Infect.* **9**, 680–686 (2020).
31. Crawford, K. H. D. et al. Protocol and reagents for pseudotyping lentiviral particles with SARS-CoV-2 spike protein for neutralization assays. *Viruses* <https://doi.org/10.3390/v12050513> (2020).
32. Food and Drug Administration. *Clinical Laboratory Improvement Amendments (CLIA)*, <https://www.fda.gov/medical-devices/ivd-regulatory-assistance/clinical-laboratory-improvement-amendments-clia> (2020).
33. Altmann, D. M., Douek, D. C. & Boyton, R. J. What policy makers need to know about COVID-19 protective immunity. *Lancet* **395**, 1527–1529 (2020).
34. Weinstein, M. C., Freedberg, K. A., Hyle, E. P. & Paltiel, A. D. Waiting for certainty on COVID-19 antibody tests — at what cost? *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMp2017739> (2020).
35. U.S. Food and Drug Administration. *Policy for Coronavirus Disease-2019 Tests During the Public Health Emergency (Revised)*, <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/policy-coronavirus-disease-2019-tests-during-public-health-emergency-revised> (2020).
36. National Cancer Institute. *NCI Part of Federal Effort to Evaluate Antibody Tests for Novel Coronavirus*, <https://www.cancer.gov/news-events/cancer-currents-blog/2020/covid-19-nci-antibody-testing-review> (2020).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

<sup>1</sup>Department of Laboratory Medicine, University of California, San Francisco, San Francisco, CA, USA. <sup>2</sup>Medical Scientist Training Program, University of California, San Francisco, San Francisco, CA, USA. <sup>3</sup>Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, CA, USA. <sup>4</sup>J. David Gladstone Institutes, San Francisco, CA, USA. <sup>5</sup>Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, CA, USA. <sup>6</sup>Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA, USA. <sup>7</sup>Diabetes Center, University of California, San Francisco, San Francisco, CA, USA. <sup>8</sup>Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, USA. <sup>9</sup>Department of Pathology, Massachusetts General Hospital/Harvard Medical School, Boston, MA, USA. <sup>10</sup>Department of Medicine, Division of Infectious Diseases, University of California, San Francisco, San Francisco, CA, USA. <sup>11</sup>Department of Bioengineering, University of California, Berkeley, Berkeley, CA, USA. <sup>12</sup>Department of Integrative Biology, University of California, Berkeley, Berkeley, CA, USA. <sup>13</sup>Department of Surgery, University of California, San Francisco, San Francisco, CA, USA. <sup>14</sup>Department of Neurology, University of California, San Francisco, San Francisco, CA, USA. <sup>15</sup>Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, USA. <sup>16</sup>Department of Dermatology, University of California, San Francisco, San Francisco, CA, USA. <sup>17</sup>Program in Quantitative Biology, University of California, San Francisco, San Francisco, CA, USA. <sup>18</sup>School of Medicine, University of California, San Francisco, San Francisco, CA, USA. <sup>19</sup>Department of Pathology, University of California, San Francisco, San Francisco, CA, USA. <sup>20</sup>Division of Experimental Medicine, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA. <sup>21</sup>Infectious Diseases and Immunity Graduate Group, University of California, Berkeley, Berkeley, CA, USA. <sup>22</sup>Weill Institute for Neurosciences, Department of Neurology, University of California, San Francisco, San Francisco, CA, USA. <sup>23</sup>Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA, USA. <sup>24</sup>Division of Infectious Diseases, Massachusetts General Hospital/Harvard Medical School, Boston, MA, USA. <sup>25</sup>UCSF-Abbott Viral Diagnostics and Discovery Center, San Francisco, CA, USA. <sup>26</sup>Scientific Affairs, American Red Cross, Gaithersburg, MD, USA. <sup>27</sup>Department of Anesthesia and Perioperative Care, University of California, San Francisco, San Francisco, CA, USA. <sup>28</sup>Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA, USA. <sup>29</sup>Parker Institute for Cancer Immunotherapy, San Francisco, CA, USA. <sup>30</sup>Chan Zuckerberg Biohub, San Francisco, CA, USA. <sup>31</sup>Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA. <sup>32</sup>Division of Rheumatology, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA. <sup>33</sup>Institute of Computational Health Sciences, University of California, San Francisco, San Francisco, CA, USA. <sup>34</sup>Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA. <sup>35</sup>Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA, USA. <sup>36</sup>Quantitative Biosciences Institute, University of California, San Francisco, San Francisco, CA, USA. <sup>37</sup>Center for Computational Biology, University of California, Berkeley, Berkeley, CA, USA. <sup>38</sup>These authors contributed equally: Jeffrey D. Whitman, Joseph Hiatt, Cody T. Mowery, Brian R. Shy, Ruby Yu. <sup>✉</sup>e-mail: [caryn.bern2@ucsf.edu](mailto:caryn.bern2@ucsf.edu); [pdhsu@berkeley.edu](mailto:pdhsu@berkeley.edu); [alexander.marson@ucsf.edu](mailto:alexander.marson@ucsf.edu)

## Methods

**Ethical approvals.** This study was approved by institutional review boards at the UCSF/ZSFG and MGH.

**Study design.** The study population included individuals with symptomatic infection and positive SARS-CoV-2 RT-PCR testing of nasopharyngeal or oropharyngeal swabs who had remnant serum and plasma specimens in clinical laboratories serving the UCSF and ZSFG medical center networks. All samples were obtained from venous blood draws, with serum being collected in either uncoated or serum separator tubes and plasma from lithium heparin tubes depending on other ancillary testing orders. All samples were drawn in an outpatient or hospital setting, professionally couriered to the clinical laboratory and acquisitioned for routine testing within the clinical laboratory within the same day. Samples were stored at 4 °C and aliquoted for freezing at -20 °C within 1 week of the initial blood draw. Serum and plasma were used interchangeably. All but one assay (Epitope ELISA) noted that either specimen type could be used. We included multiple specimens per individual but no more than one sample per time interval (1–5, 6–10, 11–15, 16–20 and >20 d after symptom onset). If an individual had more than one specimen for a given time interval, only the later specimen was included. For specificity, we included 108 pre-COVID-19 plasma specimens from eligible blood donors collected before July 2018<sup>37</sup>. We assessed detection of SARS-CoV-2 antibodies in 51 specimens from 2020: 49 with test results for detection of other respiratory viruses (BioFire FilmArray, BioFire Diagnostics) and 31 with negative results by SARS-CoV-2 RT-PCR. For these specimens, the median days from symptom onset was four with a range of 0–107 d, the latter end of the range owing to unresolving respiratory viral infection in the setting of HIV infection.

We based minimum sample size calculations on expected binomial exact 95% confidence limits. A total of 287 samples were included in the final analysis, including 128 from 79 individuals who tested positive for SARS-CoV-2 by RT-PCR. Some specimens were exhausted during the analysis and were not included in all tests. Data obtained from serial specimens that did not conform to our study design were excluded.

Clinical data were extracted from electronic health records and entered in a Health Insurance Portability and Accountability Act-secure REDCap database hosted by UCSF. Data included demographic information, major comorbidities, patient-reported symptom onset date, symptoms and indicators of severity.

Independent data from testing efforts at MGH, with slight deviations in methods, are included as Supplementary Data (Supplementary Fig. 3). Briefly, 48 heat-inactivated serum/plasma samples from 44 individuals who tested positive for SARS-CoV-2 by RT-PCR were included. For specificity, the MGH study included 60 heat-inactivated pre-COVID-19 samples from 30 asymptomatic adults and 30 individuals admitted with febrile and/or respiratory illness with a confirmed pathogen.

**Sample preparation.** Samples from UCSF and ZSFG were assigned a random well position in one of four 96-well plates. Samples were thawed at 37 °C, and up to 200 µl was transferred to the assigned well without heat inactivation. Samples were then sub-aliquoted (12.5 µl) to replica plates for testing. Replica plates were stored at -20 °C until needed and then thawed for 10 min at room temperature and briefly centrifuged before testing. All sample handling followed UCSF biosafety committee-approved practices.

For the MGH study, samples were heat inactivated at 56 °C for 60 min, aliquoted and stored at 4 °C and -20 °C. Samples stored at 4 °C were used within 7 d. Frozen aliquots were stored until needed with only a single freeze-thaw cycle for any sample. All samples were brought to room temperature and briefly centrifuged before adding the recommended volume to the LFA cartridge.

**Immunochromatographic LFAs.** Ten LFAs were evaluated (Supplementary Table 1). At the time of testing, cartridges were labeled by randomized sample location (plate and well). The appropriate sample volume was transferred from the plate to the indicated sample port, followed by provided diluent, following manufacturer instructions. The lateral flow cartridges were incubated for the recommended time at room temperature before readings. Each cartridge was assigned an integer score (0 for negative, 1–6 for positive) for test line intensity by two independent readers blinded to specimen status and to each other's scores (Supplementary Fig. 1). Readers were trained to score intensity from images representative of each value from a previous LFA test performance evaluation<sup>37</sup>. Test line scoring was performed for research purposes to capture semi-quantitative data about the LFA readout and reproducibility of subjective interpretation, considering that these are the major analytical factors that affect test performance. These tests are prescribed to be interpreted qualitatively, and test performance characteristics in this report are derived from qualitative scoring of any interpreted band color. For some cartridges (DeepBlue, UCP and Bioperfectus), the positive control indicator failed to appear after addition of diluent in a significant fraction of tests. For these tests, two further drops of diluent were added to successfully recover control indicators in all affected tests. These results were included in analyses. During testing, two plates were transposed 180°, and assays were run in the opposite order from the wells documented on cartridges. These data were

corrected, and accuracy was confirmed by empty well position and verification of a subset of results.

**ELISAs.** Epitope Diagnostics assays were carried out according to manufacturer instructions with minor deviations, including the mixed use of plasma and serum specimens (instead of serum only), use of frozen specimens (versus same day), blanking all specimens and controls instead of using raw OD450 values and performing samples in singlicate for three of four 96-well plates (instead of duplicate). Plate 4 was run in duplicate owing to availability of samples and assay wells. For IgM detection, 100 µl of control samples or 10 µl of patient serum and 100 µl of sample diluent were added to indicated wells. Plates were incubated for 30 min at 37 °C and manually washed five times in provided Wash Buffer. Each well received 100 µl of horseradish peroxidase (HRP)-labeled COVID-19 antigen, was incubated for 30 min at 37 °C and was manually washed five times in provided Wash Buffer. Each well then received 100 µl of colorimetric substrate, was incubated for 20 min and then received 100 µl of Stop Solution. The OD450 was measured using a Synergy H1 Microplate Reader (BioTek Instruments) within 10 min of adding Stop Solution. Positive cutoff for IgM detection was calculated as described in the Epitope Diagnostics protocol: IgM positive cutoff =  $1.1 \times ((\text{average of negative control readings}) + 0.10)$ . Values less than or equal to the positive cutoff were interpreted as negative. For IgG detection, 1 µl of serum was diluted 1:100 in Sample Diluent and loaded into designated wells. Plates were incubated for 30 min at room temperature and manually washed five times in provided Wash Buffer. Each well received 100 µl of provided HRP-labeled COVID-19 Tracer Antibody; plates were incubated for 30 min at room temperature and manually washed five times in provided Wash Buffer. Then, each well received 100 µl of Substrate, was incubated for 20 min and then received 100 µl of Stop Solution. The absorbance at OD450 was measured using a Synergy H1 Microplate Reader (BioTek Instruments) within 10 min of adding Stop Solution. Positive cutoffs for IgG detection were calculated as described in the Epitope Diagnostics protocol: IgG positive cutoff =  $1.1 \times ((\text{average of negative control readings}) + 0.18)$ . Values less than or equal to the positive cutoff were interpreted as negative.

An in-house receptor binding domain (RBD)-based ELISA was performed with minor deviations from a published protocol (Amanat et al.<sup>26</sup>, Kramer Lab, Mount Sinai School of Medicine). SARS-CoV-2 RBD protein was produced using the published construct (NR-52306, BEI Resources) by Aashish Manglik (UCSF). Next, 96-well plates (3855, Thermo Fisher Scientific) were coated with 2 µg ml<sup>-1</sup> RBD protein and stored at 4 °C for up to 5 d before use. Specimen aliquots (12 µl) were diluted 1:5 in 1× phosphate-buffered saline (PBS) (10010-023, Gibco), mixed and heat inactivated at 56 °C for 1 h. RBD-treated plates were washed three times with PBS-Tween (PBST, BP337-500, Fisher Bioreagents) using a 405 TS Microplate Washer (BioTek Instruments) and blocked with PBST-Milk (3% wt/vol, AB10109-01000, AmericanBio) for 1 h at 20 °C. Samples were further diluted 1:10 (1:50 final) in PBST-Milk (1% wt/vol), and 100 µl was transferred to the blocked ELISA plates in duplicate plates. Samples were incubated for 2 h at 20 °C and washed three times with PBST. The peroxidase AffiniPure Goat Anti-human IgG (Fab')<sub>20</sub>-specific secondary antibody (109-035-097, lot 146576, Jackson ImmunoResearch) used in this study binds the IgG light chain and has some reactivity for other isotypes (IgM and IgA). This secondary antibody was diluted 1:750 in PBST-Milk (1% wt/vol), 50 µl was added to each sample well and samples were incubated for 1 h at 20 °C. Plates were subsequently washed three times with PBST. We dispensed 100 µl of 1× SigmaFast OPD Solution (P9187, Sigma-Aldrich) to each sample well and incubated plates for 10 min at room temperature. We added 50 µl of 3M HCl (A144-212, Fisher Chemical) to stop the reaction and immediately read the optical density at 490 nm (OD490) using a Synergy H1 Microplate Reader (BioTek Instruments). OD490 values were corrected for each plate by subtracting the mean value of each plate's blank wells. To determine a cutoff for positive values, we calculated the mean value of negative wells for each plate, plus three standard deviations.

**Data analysis.** For LFA testing, the second reader's scores were used for performance calculations, and the first reader's scores were used to calculate inter-reader agreement statistics. Percent seropositivity among RT-PCR-confirmed cases was calculated by time interval from symptom onset. Specificity was based on results in pre-COVID-2019 samples. Binomial exact 95% CIs were calculated for all estimates. Analyses were conducted in R (3.6.3) and SAS (9.4).

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The source data for figures are provided as a supplementary file or can be viewed at <https://covidtestingproject.org>.

## References

- Whitman, J. D. et al. Chagas disease serological test performance in United States blood donor specimens. *J. Clin. Microbiol.* <https://doi.org/10.1128/JCM.01217-19> (2019).

## Acknowledgements

We thank all members of the Marson lab and the Hsu lab, P. Kim, S. Boyd, J. DeRisi, S. Quake, B. Greenhouse, C. Tato, J. Doudna, F. Urnov, D. Friedberg, D. Neeleman, J. Hering, C. Cheng, N. Khosla, M. Krisiloff, L. Groom, C. Xu, D. Fontenot, J. Karkania, G. Worthington, B. Burkholder, C. Craik, XPrize Pandemic Alliance, W. Bokhari, Z. Joaquin, S. Sarlati, S. Nesbit, W. Poe, S. Broder, Verily, C. Kim, A. Kijac, M. Solit and the Coronavirus Standards Working Group, D. Havlir, J. Engel, P. Farley, J. MacGregor, K. Hou, B. Sanders, S. Yang and S. Parker. We thank Y. E. Castro-Sesquen for sharing her semi-quantitative LFA scale, which was adapted for use in our current study. This work was supported by gifts from Anthem Blue Cross Blue Shield, the Chan Zuckerberg Biohub and anonymous philanthropy. We thank the following sources for donation of test kits: the manufacturers of Bioperfectus, DecomBio, Sure Biotech, UCP Biosciences; D. Friedberg, J. Hering and H. Schein. The Wilson lab has received support from the Rachleff Family Foundation. The Hsu lab has received support from S. Altman, V. and N. Khosla, D. and S. Deb, the Curci Foundation and Emergent Ventures. P.D.H. holds a Deb Faculty Fellowship from the University of California, Berkeley College of Engineering and is the recipient of the Rainwater Foundation Prize for Innovative Early-Career Scientist. The Marson lab has received gifts from J. Aronov, G. Hoskin, K. Jordan, B. Bakar, the Caulfield family and funds from the Innovative Genomics Institute, the Northern California JDRF Center of Excellence and the Parker Institute for Cancer Immunotherapy. We thank the National Institutes of Health for its support (to J.D.W., R38HL143581; to A.E.G., F30AI150061; to D.N.N., L40 AI140341; to S.P.B., NHLBI R38HL143581, to G.M.G., NHLBI R38HL143581; to T.A.M., 1F30HD093116; to D.W., 1F31NS106868-01; to J.G.C., R01 AI40098; to E.T.R. and R.C.C., CDC U01CK000490; MSTP students are supported by T32GM007618). R.Y. was supported by an AP Giannini Postdoctoral Fellowship. J.A.S. was supported by the Larry L. Hillblom Foundation (2019-D-006-FEL). A.M. holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund, is an investigator at the Chan Zuckerberg Biohub and is a recipient of the Cancer Research Institute Lloyd J. Old STAR grant.

## Author contributions

As co-first authors, J.D.W., J.H., C.T.M., B.R.S. and R.Y. were leaders of study design, data management and statistical analysis, test procurement, laboratory analysis, clinical correlation and manuscript drafting. A.M. and P.D.H. conceived this project and, with C.B., co-supervised this work and manuscript drafting. T.N.Y., U.R., C.W., J.M.W.,

A.E.G., S.A.B., A.M.L., S.L., A.W.W., E.M.G.B., Z.B.S., Y.L., R.A., M.J.L., J.A.S., T.Z., I.C.B., E.I., J.C., D.D.A., J.L., T.A.M., T.S.K., D.W., D.L.N., W.G., V.A.Y., H.A.E., P.C.C., L.W., M.E.M., J.L., L.T., L.F., R.L. K.K. and K.C.Z. were involved in test kit procurement, laboratory analysis and laboratory administration. G.M.G., A.G.L., D.N.N., S.P.B., J.B. and S.B. contributed to chart review and clinical correlation. S.M., C.Y.C., S.L.S., M.R.W., A.M., C.J.Y., N.J.K., M.S.A., J.G.C., J.D.E., A.H.B.W. and K.L.L. contributed to study design and supervision of UCSF co-authors. T.E.M., W.F.G.B., D.Y., M.G.A., B.E.B., J.A.G., E.T.R., R.C.C., A.J.I. and J.K.L. contributed to study design, laboratory analysis and data analysis at Massachusetts General Hospital. All authors contributed to manuscript review and editing.

## Competing interests

This work was supported by gifts from Anthem Blue Cross Blue Shield, the Chan Zuckerberg Biohub and anonymous philanthropy. C.Y.C. is the director of the UCSF-Abbott Viral Diagnostics and Discovery Center, receives research support funding from Abbott Laboratories and is on the scientific advisory board of Mammoth Biosciences. C.J.Y. is co-founder of DropPrint Genomics and serves as an advisor to them. M.S.A. holds stock in Medtronic and Merck. P.D.H. is a co-founder of Spotlight Therapeutics and serves on the board of directors and scientific advisory board and is an advisor to Serotiny. P.D.H. holds stock in Spotlight Therapeutics and Editas Medicine. A.M. is a co-founder of Spotlight Therapeutics and Arsenal Biosciences and serves on their board of directors and scientific advisory board. A.M. has served as an advisor to Juno Therapeutics, was a member of the scientific advisory board at PACT Pharma and was an advisor to Trizell. A.M. owns stock in Arsenal Biosciences, Spotlight Therapeutics and PACT Pharma. R.Y. owns stock in AbbVie, Bluebird Bio, Bristol-Myers Squibb, Cara Therapeutics, Editas Medicine, Esperion and Gilead Sciences. Unrelated to this current work, the Marson lab has received sponsored research support from Juno Therapeutics, Epinomics, Sanofi and GlaxoSmithKline and a gift from Gilead.

## Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41587-020-0659-0>.

**Correspondence and requests for materials** should be addressed to C.B., P.D.H. or A.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.

For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a                                 | Confirmed                           |  |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of all covariates tested   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection: No software was used for data collection.

Data analysis: Analyses were conducted in R (3.6.3), RStudio(1.1.419), Tidyverse (1.3.0), irr package (v.0.84.1) and SAS (9.4).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data Availability: The source data for figures are provided as a supplementary file or can be viewed at <https://covidtestingproject.org/>.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Minimum sample size was determined using 95% confidence intervals assuming a realistic range of specimen numbers and hypothetical sensitivity/specificity. Maximum specimen numbers included were determined by assay/reagent availability.
Data exclusions	Data that did not fit our study design was excluded after the fact. This includes all data and statistics derived from specimens from individuals who were mis-assigned to a data analysis group (including time interval from days from symptom onset, or RT-PCR status), duplicate patient specimens not originally identified prior to obtaining results, or data from confirmatory spot testing described in the manuscript.
Replication	Due the COVID-19 pandemic, assay/reagent availability precluded most instances of assessing reproducibility/precision. Positive and negative quality control standards run in duplicate on ELISA plates were available to assess reproducibility/precision inter and intra-run (4 plates).
Randomization	Specimens were randomized to location on 96-well plates.
Blinding	Specimen status was blinded to the researchers performing the tests. Randomization of specimens among 96-well plates was only available to the data analysis team. Individual lateral flow assay reads were performed separately (within the appropriate time frame).

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	Secondary antibodies were provided with/in each commercial assay listed in supplementary table 1 without specific details. For the in-house ELISA AffiniPure Goat Anti-human IgG (F(ab') <sub>2</sub> specific) secondary antibody (109-035-097, Lot 146576, Jackson ImmunoResearch) was utilized.
Validation	AffiniPure Goat Anti-human IgG (F(ab') <sub>2</sub> specific) secondary antibody was tested with a range of dilutions against a confirmed SARS-CoV-2 positive plasma sample and a pre-SARS-CoV-2 blond bank negative control sample. The ELISA area under the curve (AUC) was calculated for the positive and negative control samples at each dilution, and the 1:750 dilution (final antibody concentration on the plate of 1.067 micrograms/mL) was selected for use in the assay according to the highest positive/negative AUC ratio.  Additional citations of this antibody use are available at: <a href="https://www.citeab.com/antibodies/2036596-109-035-097-peroxidase-affinipure-goat-anti-human-ig?utm_campaign=Widget+All+Citations&amp;utm_medium=Widget&amp;utm_source=Jackson+ImmunoResearch&amp;utm_term=Jackson+ImmunoResearch">https://www.citeab.com/antibodies/2036596-109-035-097-peroxidase-affinipure-goat-anti-human-ig?utm_campaign=Widget+All+Citations&amp;utm_medium=Widget&amp;utm_source=Jackson+ImmunoResearch&amp;utm_term=Jackson+ImmunoResearch</a> .

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Study population characteristics are available in the methods section and Table 1 of the manuscript.
----------------------------	--



Recruitment

Sample inclusion was based on criteria related to SARS-CoV-2 RT-PCR status performed for routine clinical purposes.

Ethics oversight

University of California, San Francisco, Zuckerberg San Francisco General Hospital, Massachusetts General Hospital.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

---

**Supplementary information**

---

**Evaluation of SARS-CoV-2 serology assays reveals a range of test performance**

---

In the format provided by the authors and unedited

## **Supplementary Information**

**Title:** Test performance evaluation of SARS-CoV-2 serological assays

### **Table of Contents**

Supplementary Table 1: Immunoassay Kit and Manufacturer Information	2
Supplementary Table 2: Reader Agreement on Immunochromatographic Lateral Flow Assays	4
Supplementary Table 3: MGH Study Assay Performance	5
Supplementary Figure 1: Representative Images of LFA scoring	6
Supplementary Figure 2: Comparison of Reader 1 and Reader 2 LFA scores	7
Supplementary Figure 3: LFA Scores by Serological Assay and Patient Disposition	8

Assay	Supplier	Product	Antigen*	Format**	Lot(s)	Product Number	Distributor	Kit Acquisition for Study	Performance Notes
LFAs	BioMedomics Inc, Morrisville, NC, USA	COVID-19 IgM-IgG Rapid Test	RBD	1	20200 32103	51-002-20	Henry Schein, Melville, NY, USA	Provided by Distributor Free of Charge	Some control band inconsistency
	Bioperfectus Technologies Co Ltd, Jiangsu, China	PerfectPOC Novel Corona Virus (SARS-CoV-2) IgM/IgG Rapid Test Kit	NP, SP	1	20200 313, 20210 312	SC30201 W	-----	Provided by Supplier Free of Charge	Extra diluent necessary
	Decombio Biotechnology Co Ltd, Beijing, China	Novel Coronavirus (SARS-CoV-2) IgM/IgG Combo Rapid Test-Cassette	-----	1	-----	-----	-----	Provided by Supplier Free of Charge	Some control band inconsistency
	DeepBlue Medical Technology Co Ltd, Anhui, China	COVID-19 (SARS-CoV-2) IgG/IgM Antibody Test Kit (Colloidal Gold)	-----	1	20200 305	-----	-----	Donated by John Hering, who purchased from supplier	Extra diluent necessary, Some control band inconsistency
	Innovita Biological Technology Co Ltd, Qian'an, China	Novel Coronavirus (2019-nCoV) Ab Test (Colloidal Gold)	NP, SP	2	20200 304	-----	20/20 GeneSystems, Rockville, MD, USA	Purchased from Distributor	Some band smearing
	Premier Biotech, Minneapolis, MN, USA	COVID-19 IgG/IgM Rapid Test Cassette	-----	1	COV2 00300 71	INGM- MC42S	-----	Purchased from Supplier	Some band smearing
	Sure Biotech, New York, NY, USA; Wan Chai, Hong Kong	SARS-CoV-2 IgM/IgG Antibody Rapid Test	NP, SP	1	COV1 25200 3B	VC01210 3	-----	Provided by Supplier Free of Charge	-----
	UCP Biosciences, San Jose, CA, USA	Coronavirus IgG/IgM Antibody (COVID-19) Test Cassette	-----	1	SMP2 02003 12, SMP2 02003 13	U-CoV- 102	-----	Provided by Supplier Free of Charge	Extra diluent necessary
	VivaChek Biotech Co, Hangzhou, China	VivaDiag™ SARS-CoV-2 IgM/IgG Rapid Test (COVID-19 IgM/IgG Rapid Test)	-----	1	E2003 002	VID35- 08-011	Everest Links Pte Ltd, Singapore	Purchased from Distributor	Some band smearing
	Wondfo Biotech Co Ltd, Guangzhou, China	SARS-CoV-2 Antibody Test (Lateral Flow Method)	-----	3	W195 00318	W195	-----	Donated by David Friedberg, who purchased from supplier	Some band smearing

Assay	Supplier	Product	Antigen*	Format**	Lot(s)	Product Number	Distributor	Kit Acquisition for Study	Performance Notes
<b>MGH LFAs</b>	<b>SD Biosensor</b> , Suwon-si, Gyeonggi-do, Republic of Korea	STANDARD Q COVID-19 IgM/IgG Duo	NP	2	QCO1 02000 6	Q- NCOV-01D	Henry Schein, Melville, NY, USA	Provided by Distributor Free of Charge	-----
	<b>Biolidics</b> Limited, Mapex, Singapore	2019-nCoV IgG/IgM antibody detection kit	NP, RBD	1	V2020 0330	CBB-F015016 -V	-----	Purchased from Supplier	-----
	<b>Biomedomics</b> Inc, Morrisville, NC, USA	COVID-19 IgM and IgG Rapid Test	RBD	1	20200 22702 20200 32103	51-002-20	Henry Schein, Melville, NY, USA	Lot 1 provided by Distributor Free of Charge; Lot 2 purchased from Supplier	-----
<b>ELISAs</b>	<b>Epitope</b> Diagnostics, San Diego, CA, USA	KT-1033 EDI™ Novel Coronavirus COVID-19 <b>IgM</b> ELISA Kit	NP	--	P630C	KT-1032	-----	Purchased from Supplier	-----
	<b>Epitope</b> Diagnostics, San Diego, CA, USA	KT-1032 EDI™ Novel Coronavirus COVID-19 <b>IgG</b> ELISA Kit	NP	--	P637U	KT-1033	-----	Purchased from Supplier	-----
	<b>In-House</b> ELISA	Peroxidase AffiniPure Goat Anti-human IgG (F(ab') <sub>2</sub> specific) secondary antibody (Jackson ImmunoResearch)	RBD	--	14657 6	109-035-097	Adapted from Krammer Lab, Icahn School of Medicine at Mt. Sinai, New York, NY, USA	Lab-developed test	-----

**\*Antigen:**

NP = Nucleocapsid protein

SP = Spike protein

RBD = Receptor binding domain, Spike protein

**\*\*LFA Test Cartridge Format:**

1: Single lane, separate IgM and IgG bands

2: Separate IgM and IgG lanes

3: Single lane, single band for both IgM and

IgG

**Supplementary Table 1. Immunoassay Kit and Manufacturer Information. Bold** signifies labels used in text and figures.

Supplier	IgG			IgM		
	n	Positive Kappa Correlation	Weighted Kappa Correlation	n	Positive Kappa Correlation	Weighted Kappa correlation
BioMedomics	284	0.9649	0.9580	284	0.8238	0.8253
Bioperfectus	275	0.9586	0.9488	275	0.9130	0.8631
DecomBio	282	0.9761	0.9529	282	0.9845	0.9660
DeepBlue	287	0.9547	0.8971	287	0.9213	0.9377
Innovita	250	0.9588	0.8490	249	0.8084	0.8029
Premier	286	0.9718	0.9880	286	0.9680	0.9340
Sure	286	0.9908	0.9665	286	0.9300	0.7967
UCP	286	0.9565	0.9574	286	1.0000	0.9484
VivaChek	266	0.9912	0.9669	266	0.9332	0.9439
Wondfo*	271	0.9916	0.9542	-	-	-

\*Wondfo kits detect combined IgG and IgM.

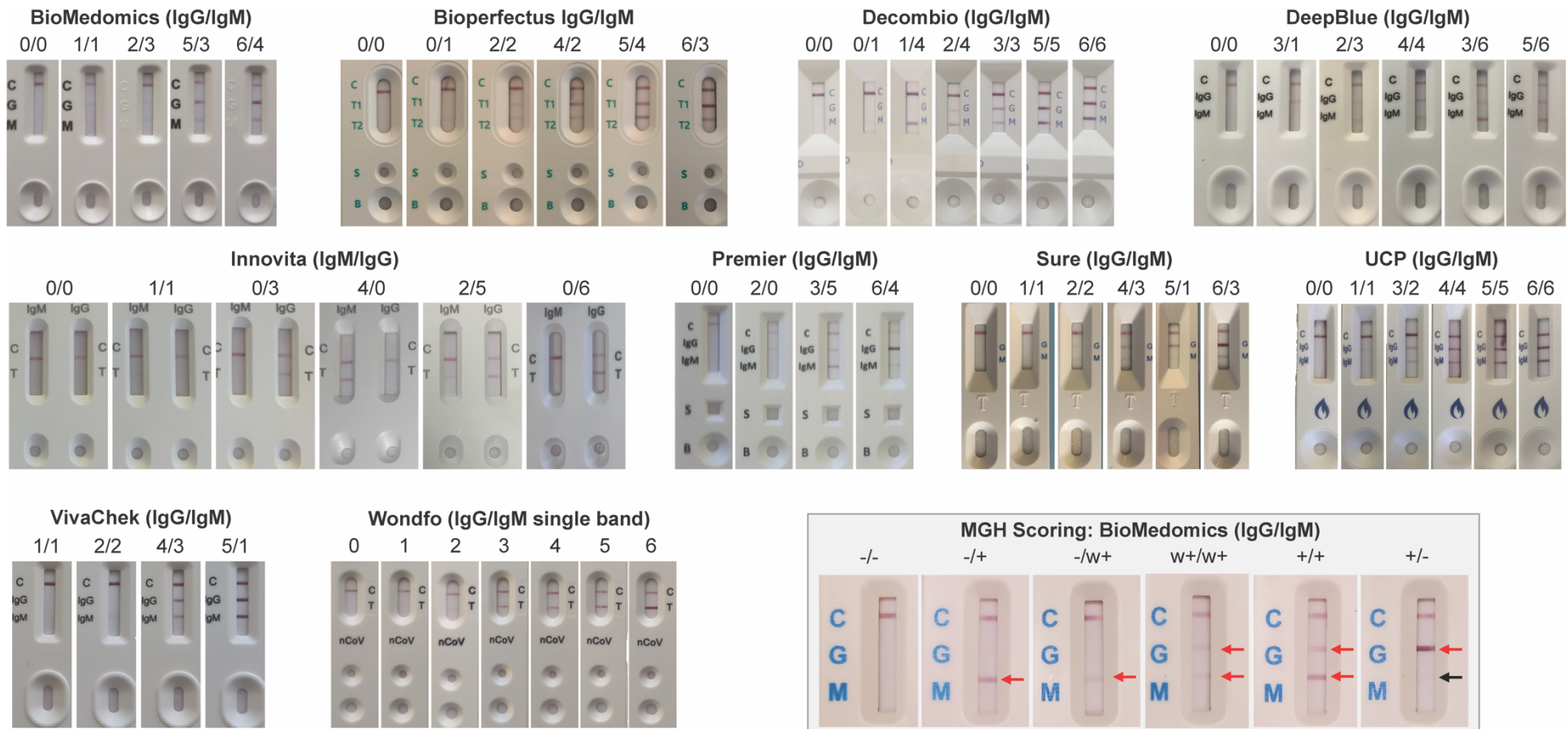
**Supplementary Table 2. Reader Agreement on Immunochromatographic Lateral Flow Assays (LFAs).** Cohen’s Kappa correlations were calculated for scores of the IgG band (left) and IgM band (right) of each LFA. The LFA manufactured by Wondfo has a single band for IgG and IgM detection and is displayed here as IgG for convenience. Positive Kappa Correlation: unweighted inter-reader agreement on positive (LFA score > 0) vs. negative (LFA score = 0) reads. Weighted Kappa Correlation: inter-reader agreement on LFA score (0-6), weighted by the square of the difference in reads. All correlations were calculated with the *irr* package version 0.84.1 in R version 3.6.1 using RStudio.



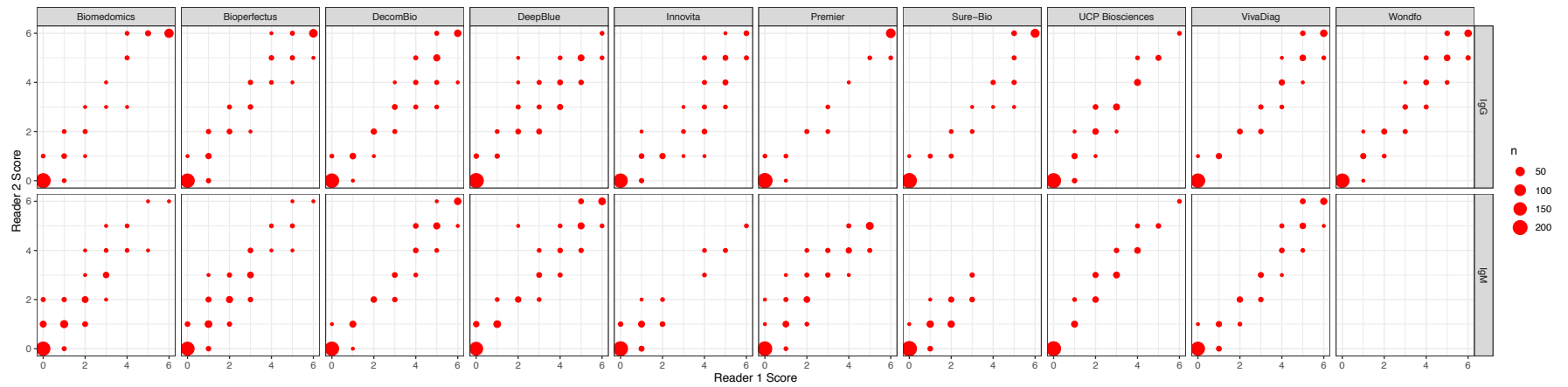
**MGH Serology Test Performance Evaluation**

Assay	IgM					IgG					IgM or IgG				
	Total N	positive	%	95% CI Lower	95% CI Upper	Total N	positive	%	95% CI Lower	95% CI Upper	Total N	positive	%	95% CI Lower	95% CI Upper
<b>LFA</b>															
<b>SD Biosensor</b>															
1-5 days	7	0	0.00	0.00	40.96	7	1	14.29	0.36	57.87	7	1	14.29	0.36	57.87
6-10 days	15	6	40.00	16.34	67.71	15	5	33.33	11.82	61.62	15	7	46.67	21.27	73.41
11-15 days	19	15	78.95	54.43	93.95	19	16	84.21	60.42	96.62	19	17	89.47	66.86	98.70
>16 days	7	6	85.71	42.13	99.64	7	6	85.71	42.13	99.64	7	6	85.71	42.13	99.64
Pre-COVID-19	60	0				60	1				60	1			
<b>Biolidics</b>															
1-5 days	7	0	0.00	0.00	40.96	7	0	0.00	0.00	40.96	7	0	0.00	0.00	40.96
6-10 days	15	2	13.33	1.66	40.46	15	8	53.33	26.59	78.73	15	8	53.33	26.59	78.73
11-15 days	19	9	47.37	24.45	71.14	19	16	84.21	60.42	96.62	19	16	84.21	60.42	96.62
>16 days	7	4	57.14	18.41	90.10	7	6	85.71	42.13	99.64	7	6	85.71	42.13	99.64
Pre-COVID-19	60	0				60	0				60	0			
<b>BioMedomics</b>															
1-5 days	7	1	14.29	0.36	57.87	7	0	0.00	0.00	40.96	7	1	14.29	0.36	57.87
6-10 days	15	6	40.00	16.34	67.71	15	6	40.00	16.34	67.71	15	7	46.67	21.27	73.41
11-15 days	19	14	73.68	48.80	90.85	19	14	73.68	48.80	90.85	19	15	78.95	54.43	93.95
>16 days	7	6	85.71	42.13	99.64	7	6	85.71	42.13	99.64	7	6	85.71	42.13	99.64
Pre-COVID-19	60	0				60	0				60	0			

**Supplementary Table 3. Assay performance on validation cohort performed at MGH using positivity thresholds based on concordance studies to an MGH-group in-house ELISA.** Comparison of MGH and UCSF percent positivity at different positivity thresholds is performed in Supplementary Figure 4. Note, the one negative patient included in the >16-day timepoint was immunocompromised.

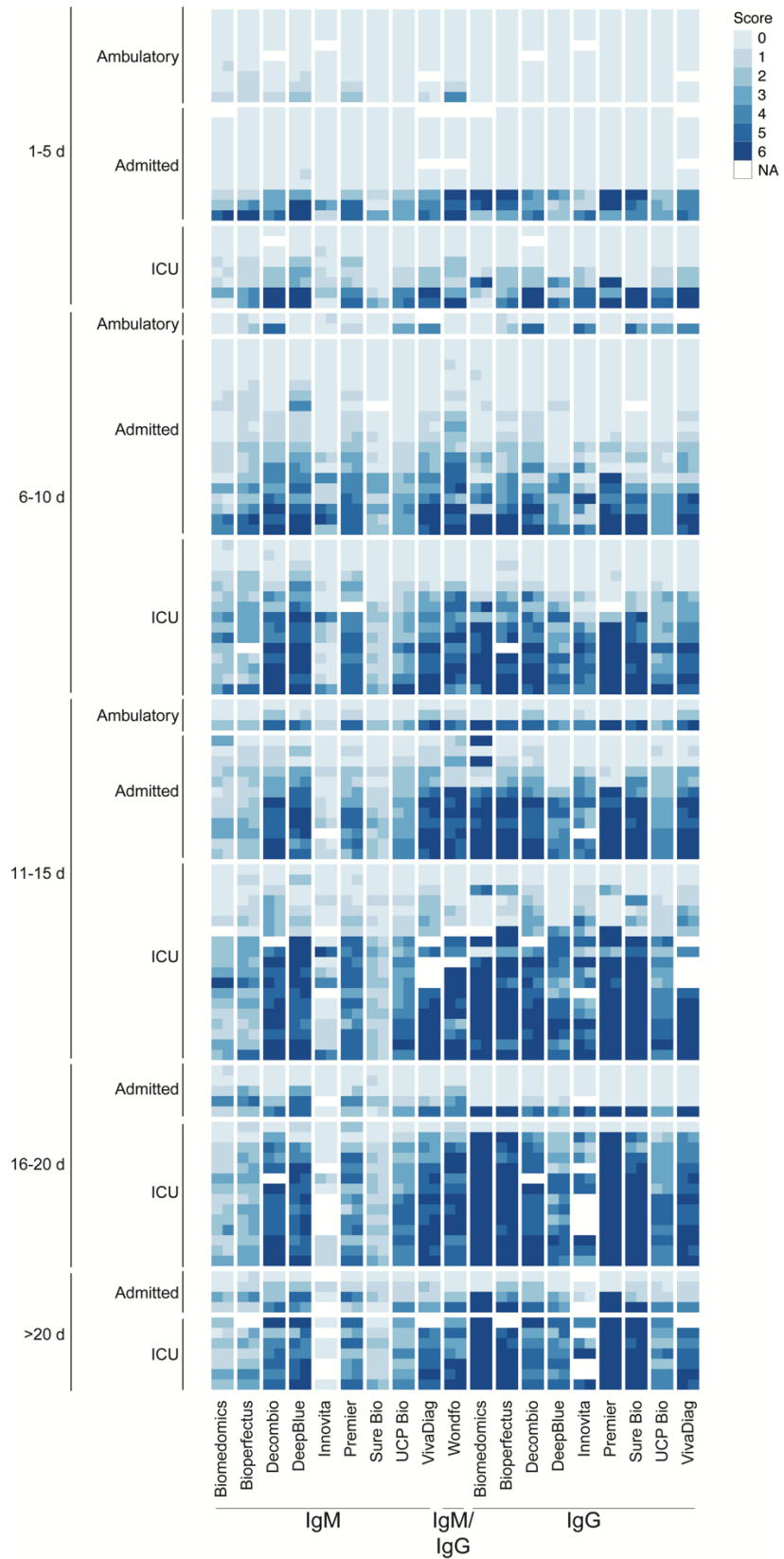


**Supplementary Figure 1: Representative images of LFA scoring.** The intensity score is noted at the top of each image for IgG and IgM separately (unless stated otherwise).



**Supplementary Figure 2: Comparison of Reader 1 and Reader 2 LFA scores.**

The size of each point signifies the number of tests with the indicated reader 1-to-reader 2 score combination. The LFA manufactured by Wondfo has a single band for IgG and IgM detection and is displayed here as IgG for convenience.



**Supplementary Figure 3: LFA scores by serological assay according to highest-level clinical care received by the patient.**