

# 1 **A systems serology analysis of correlates of protection against cholera**

2  
3 Kirsten E. Wiens<sup>1,\*</sup>, Anita S. Iyer<sup>2,3,\*</sup>, Taufiqur R. Bhuiyan<sup>4</sup>, Lenette L. Lu<sup>5,6,7</sup>, Deniz Cizmeci<sup>8</sup>,  
4 Matthew J. Gorman<sup>8</sup>, Dansu Yuan<sup>8</sup>, Rachel L. Becker<sup>2</sup>, Edward T. Ryan<sup>2,3,10</sup>, Stephen B.  
5 Calderwood<sup>2,3</sup>, Regina C. LaRocque<sup>2,3</sup>, Fahima Chowdhury<sup>4</sup>, Ashraful I. Khan<sup>4</sup>, Myron M.  
6 Levine<sup>9</sup>, Wilbur H. Chen<sup>9</sup>, Richelle C. Charles<sup>2,3,10</sup>, Andrew S. Azman<sup>1,#</sup>, Firdausi Qadri<sup>4,#</sup>, Galit  
7 Alter<sup>3,8,#</sup>, Jason B. Harris<sup>2,11,#,†</sup>

8  
9 \* Contributed equally

10 # Contributed equally

## 11 12 **Author affiliations:**

13 <sup>1</sup>Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore,  
14 MD, USA

15 <sup>2</sup>Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA, USA

16 <sup>3</sup>Department of Medicine, Harvard Medical School, Boston, MD, USA

17 <sup>4</sup>Infectious Diseases Division, International Centre for Diarrheal Disease Research,  
18 Bangladesh, Dhaka, Bangladesh

19 <sup>5</sup>Division of Infectious Diseases and Geographic Medicine, Department of Internal Medicine,  
20 UTSW Medical Center, Dallas, TX, USA

21 <sup>6</sup>Department of Immunology, UTSW Medical Center, Dallas, TX, USA

22 <sup>7</sup>Parkland Health and Hospital System, Dallas, TX, USA

23 <sup>8</sup>Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA, USA

24 <sup>9</sup>Center for Vaccine Development and Global Health, University of Maryland School of Medicine,  
25 Baltimore, MD, USA

26 <sup>10</sup>Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston,  
27 MA, USA

28 <sup>11</sup>Department of Pediatrics, Harvard Medical School, Boston, MA, USA

29

30 † **For correspondence:** Jason B. Harris ([jbharris@mgh.harvard.edu](mailto:jbharris@mgh.harvard.edu))

31

32

### 33 **Abstract**

34 Vibriocidal antibodies are the best characterized correlate of protection against cholera and are  
35 used to gauge immunogenicity in vaccine trials. However, there is no vibriocidal titer threshold  
36 associated with absolute protection against infection with *Vibrio cholerae* or with symptomatic  
37 disease in infected individuals. While other circulating antibody responses have also been  
38 associated with a decreased risk of *V. cholerae* infection, there has been no comprehensive  
39 comparison of correlates of protection against cholera. To address this, we analyzed 58 serum  
40 antibody biomarkers as correlates of protection against both *V. cholerae* infection and against  
41 cholera diarrhea in infected individuals. The study was performed in two cohorts: (1) household  
42 contacts of patients with cholera in Bangladesh, and (2) North American volunteers who were  
43 vaccinated with a single dose of CVD 103-HgR live oral cholera vaccine and then challenged  
44 with virulent *V. cholerae* O1 El Tor Inaba. In household contacts, we identified 20 antibody  
45 markers that were correlated with protection against *V. cholerae* infection, though there was  
46 overlap between distributions. Conditional random forest models identified serum antibody-  
47 dependent complement deposition, targeting the *V. cholerae* O1 antigen, as the most predictive  
48 individual correlate of protection from infection, while vibriocidal antibody titers were less  
49 predictive. The model that most accurately predicted protection from infection included five  
50 biomarkers, with a cross-validated area under the curve (cvAUC) of 79% (95% CI 73-85).  
51 Similarly, in North American volunteers who were challenged with *V. cholerae* after vaccination,

52 a different five-biomarker model predicted protection from the development of cholera diarrhea  
53 with a cvAUC of 78% (95% CI 66-91). Thus, while several new biomarkers predict protection  
54 better than vibriocidal titers, it remains difficult to consistently predict whether an individual will  
55 be protected from future mucosal infection or symptoms using current serologic markers.

56

## 57 **Introduction**

58 Circulating antibodies that correlate with protection from infection or disease are valuable for  
59 understanding whether segments of the population are likely to be protected due to prior  
60 infection or vaccination. For cholera, the most well characterized and widely used correlate of  
61 protection is the serum vibriocidal antibody titer (Iyer and Harris, 2021). Vibriocidal titers remain  
62 elevated in the blood for up to a year or longer following infection (Azman et al., 2019), but are  
63 not thought to represent a mechanistic form of mucosal immunity (Iyer and Harris, 2021). Serum  
64 IgA antibodies specific to several *Vibrio cholerae* antigens are also correlated with protection  
65 from infection (Harris et al., 2008; Kaiser et al., 2021). While intestinal IgA antibodies could  
66 represent a direct form of protection, circulating IgA responses are short-lived following *V.*  
67 *cholerae* infection (Azman et al., 2019). For all of these known correlates, there is no titer that is  
68 associated with absolute protection.

69 In addition to serum vibriocidal and IgA responses, extensive household-based studies  
70 conducted in Dhaka, Bangladesh have shown that age, ABO blood group, and certain species  
71 in the intestinal microbiota are associated with infection and disease severity (Harris et al.,  
72 2008; Levade et al., 2021; Midani et al., 2018; Ritter et al., 2019; Weil et al., 2009). However,  
73 taken together, these variables do not entirely account for the variation in susceptibility or  
74 disease severity seen in this population. This residual variation is likely attributable to a  
75 combination of individual-level risk factors, exposure histories, and unmeasured or undetected  
76 immune responses. The degree to which unmeasured immune responses contribute to variation

77 in susceptibility is unknown, and no comprehensive comparison of the multiple known correlates  
78 of protection against cholera has been conducted.

79

80 Here, our objective was to perform a comprehensive analysis of antibody-mediated  
81 correlates of protection from both *V. cholerae* infection and cholera diarrhea. Specifically, we  
82 asked: (1) Can we identify novel serum correlates of susceptibility to infection and/or cholera  
83 diarrhea after exposure to *V. cholerae* following both prior infection and vaccination? and (2) Do  
84 pre-infection biomarkers allow for accurate identification of who will be infected after exposure  
85 and/or who will go on to develop symptoms?

86 To this end, we conducted a systems serology analysis (Arnold and Chung, 2018;  
87 Chung and Alter, 2017) of 58 pre-infection serum biomarkers in two cohorts of individuals that  
88 were followed prospectively to monitor *V. cholerae* infection and cholera diarrhea. The  
89 biomarkers evaluated included all previously identified antigens associated with protective  
90 immunity as well as functional markers of antibody-mediated antibacterial immunity. The first  
91 cohort included contacts in the household of patients who had cholera in Dhaka, Bangladesh,  
92 where cholera is endemic (Levade et al., 2021; Weil et al., 2009). The second cohort consisted  
93 of North American volunteers who were vaccinated with a live attenuated oral cholera vaccine  
94 and subsequently challenged with *V. cholerae* (Chen et al., 2016). We tested single- and multi-  
95 biomarker panels in prediction using several machine-learning approaches. Our work identifies  
96 several novel biomarkers, as well as both the utility and limitations of using serum biomarkers to  
97 predict susceptibility to mucosal infection.

98

## 99 **Results**

### 100 Household contacts of index cholera cases

101 We included samples and data from 261 contacts of *V. cholerae* O1-Ogawa-infected cholera  
102 cases in 180 households in Dhaka, Bangladesh. The characteristics of this cohort are shown in

103 Table 1. The cohort included 4 individuals who were under five years old, all of whom went on to  
 104 develop infection. 77 of the 261 contacts had a positive stool culture (i.e., became infected) after  
 105 exposure to a case in the household. Among the infected contacts, 26 were symptomatic and  
 106 51 were asymptomatic.

107

108 **Table 1. Characteristics of the household contacts of index cholera cases in Dhaka, Bangladesh**

		Total		Infected	
Category	Group	N	%	N	%
Sex	Female	134	51.3	40	29.9
	Male	127	48.7	37	29.1
Age group	0-4	4	1.5	4	100.0
	5-17	74	28.4	25	33.8
	18-49	164	62.8	45	27.4
	50-71	19	7.3	3	15.8
Self-reported history of prior cholera	No	251	96.2	76	30.3
	Yes	10	3.8	1	10.0
Vibriocidal titer $\geq 320$ upon enrollment	No	226	86.6	74	32.7
	Yes	35	13.4	3	8.6
Index cholera case received antibiotics	No	181	69.3	60	33.1
	Yes	80	30.7	17	21.2
Relationship to index cholera case	Child	48	18.4	18	37.5
	Sibling	47	18.0	15	31.9
	Spouse	61	23.4	15	24.6
	Parent	101	38.7	29	28.7
Diarrhea duration in index case	0-6 days	67	25.7	17	25.4
	7-20 days	92	35.2	23	25.0
	14-20 days	38	14.6	12	31.6
	21-73 days	64	24.5	25	39.1

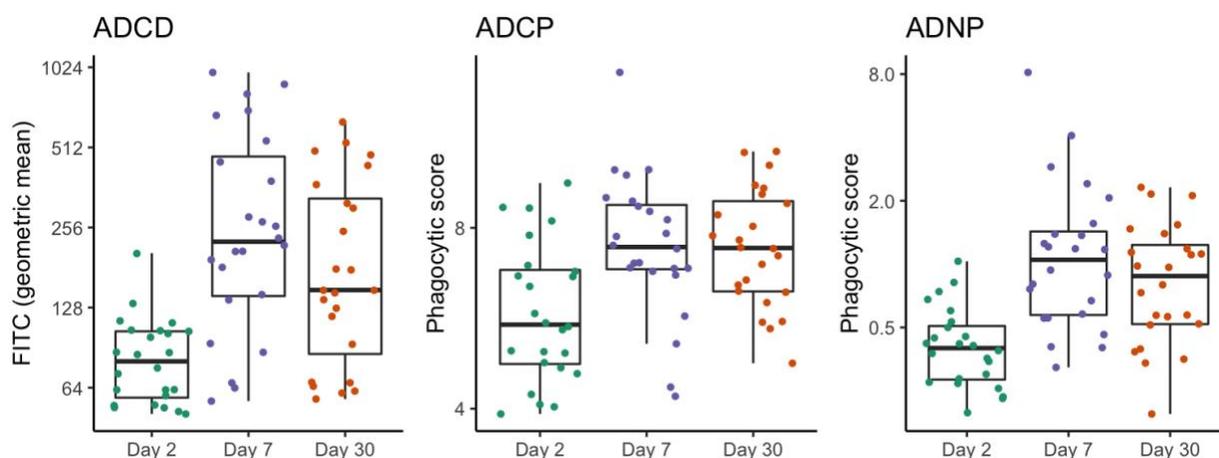
109

110 Correlates of protection in household contacts

111 We examined 58 serum antibody biomarkers in household contacts upon enrollment. These  
 112 included 54 antigen-isotype specific responses measured using a multiplex bead assay: IgA,  
 113 IgA1, IgA2, IgM, IgG, IgG1, IgG2, IgG3, and IgG4 specific for *V. cholerae* O1 Ogawa OSP:BSA,

114 O1 Inaba OSP:BSA, toxin coregulated pilus (TcpA), cholera holotoxin (CT-HT), cholera toxin B  
115 subunit (CtxB), and *V. cholerae* sialidase. These antigens represent dominant targets of the  
116 humoral immune response following cholera (Kauffman et al., 2016) (Azman et al., 2019). We  
117 also examined previous measurements of vibriocidal titers (Ritter et al., 2019) as well as three  
118 previously uninvestigated measures of functional antibody responses against cholera. These  
119 included measures of antibody-dependent complement deposition (ADCD), antibody-dependent  
120 cellular phagocytosis (ADCP), and antibody-dependent neutrophil phagocytosis (ADNP)  
121 targeting the *V. cholerae* O1 antigen. We verified that patients recovering from cholera mounted  
122 ADCD, ADCP and ADNP responses (Figure 1).

123



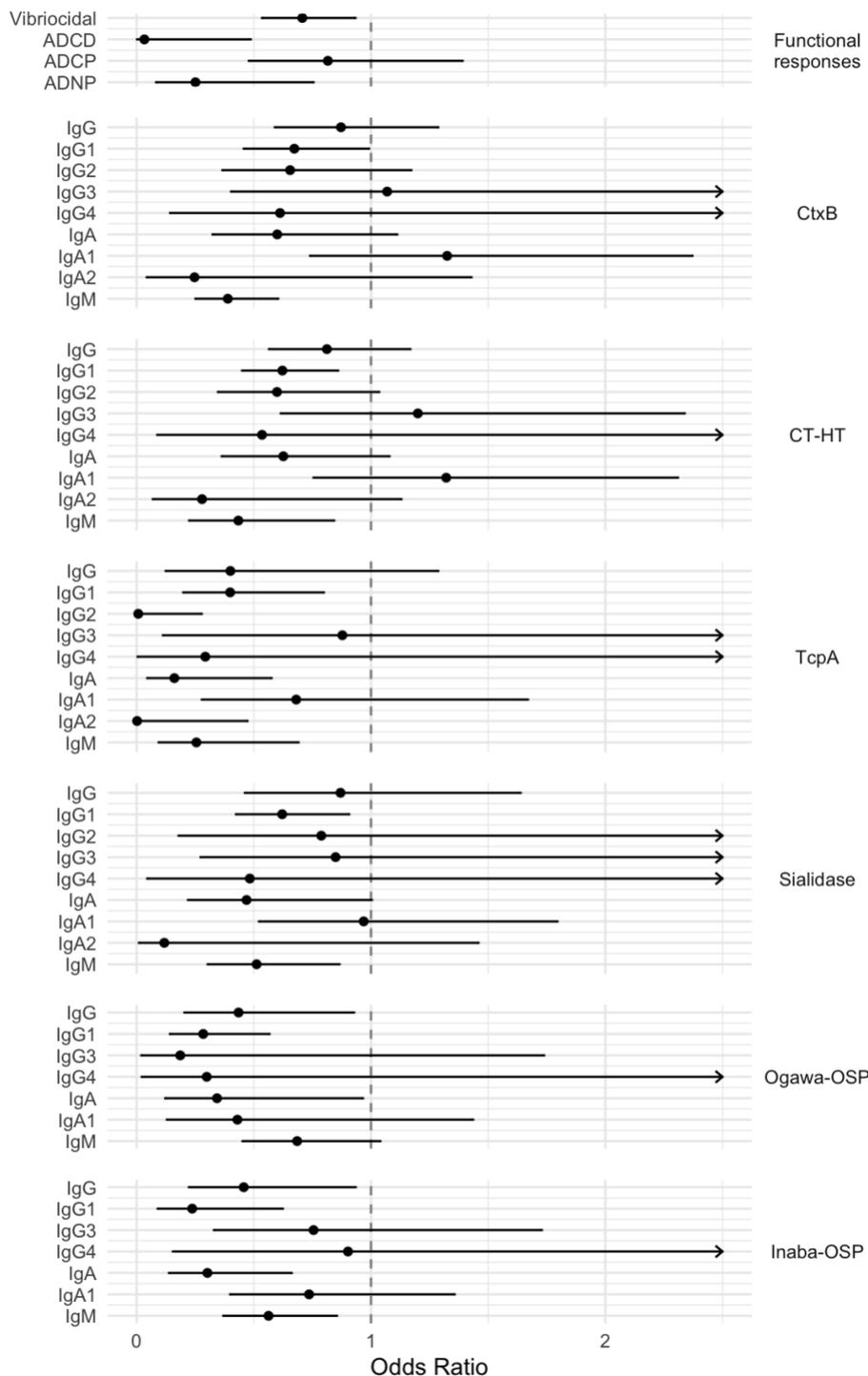
124

125 **Figure 1. Dynamics of functional antibody responses targeting the *V. cholerae* O1 antigen**  
126 **following *V. cholerae* infection.** Antibody-dependent complement deposition (ADCD), cellular  
127 phagocytosis (ADCP), and neutrophil phagocytosis (ADNP) in serum from 24 index cholera cases upon  
128 enrollment (i.e., Day 2 post presumed symptom onset) and days 7 and 30 post onset. Each point  
129 represents the geometric mean of complement deposition and phagocytic scores, respectively, across  
130 three replicates. Boxplots show 50% (median), 25%, and 75% quartiles; whiskers represent approximate  
131 95% confidence intervals for comparing medians (McGill et al., 1978).

132

133 In single biomarker analyses, after adjusting for age and household clustering, 20 biomarkers  
134 were elevated among household contacts that remained uninfected (Figure S1) and hence were  
135 correlates of protection from subsequent infection. Among individual biomarkers, the risk of  
136 becoming infected with *V. cholerae* O1 was particularly low in individuals with elevated ADCD,

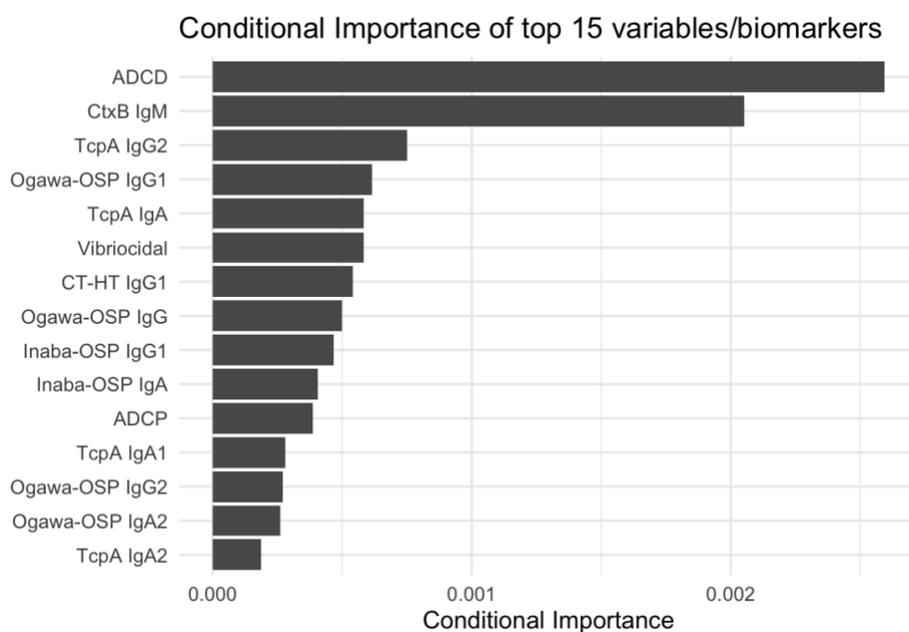
137 TcpA IgG2, and TcpA IgA2 titers (Figure 2). Correlates of protection from symptoms among  
138 infected individuals were distinct from correlates of protection from initial infection (Figure S2),  
139 though sample sizes were smaller (n=77) and there was substantial overlap in antibody  
140 distributions between outcomes (Figure S1).  
141



142

143 **Figure 2. Risk of infection in household contacts of an index cholera case.** Odds ratio of becoming  
 144 infected among household contacts of an index cholera case for every 2-fold increase in baseline (Day 2)  
 145 antibody titers, after adjusting for age and household clustering. Mean and 95% confidence interval are  
 146 shown for each biomarker analyzed independently. Arrows indicate where upper confidence interval  
 147 extends beyond the graph. Isotype is indicated on the left and antigen on the right for binding antibody  
 148 titers. Biomarkers for which logistic model fitted values were very close to 1 are excluded.

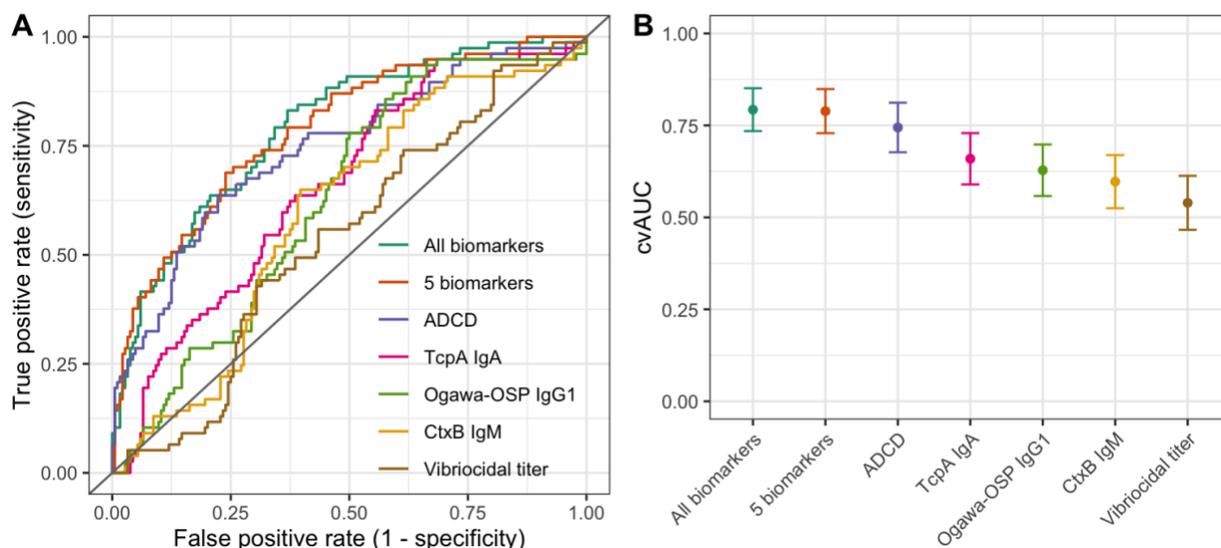
149 In an unsupervised multivariate analysis, we found limited separation of infection outcomes  
150 using all antibody responses (Figure S3), which were generally highly correlated (Figure S4).  
151 We then used conditional random forest models, designed to handle a large number of highly  
152 correlated variables (Strobl et al., 2007), to examine which biomarkers were most important for  
153 classifying individuals by outcome. ADCD was ranked most important for classifying infections,  
154 and there were several features with higher scores than vibriocidal titers (Figure 3).



155  
156 **Figure 3. Biomarkers important for classifying household contacts by infection outcome.** Top 15  
157 biomarkers important in classifying household contacts of index cholera cases as infected (i.e., becoming  
158 stool culture positive) vs. uninfected (i.e., remaining stool culture negative). Biomarkers are ranked by  
159 importance scores calculated using conditional random forest classification models.  
160

161 To test the validity of the classification models, we performed leave-one-out cross-validation for  
162 models constructed using different subsets of biomarkers and age. The model that most  
163 accurately predicted infection included five of the most informative biomarkers, with a cross-  
164 validated area under the curve (cvAUC) of 79% (95% CI 73-85) (Figure 4). This corresponded  
165 to a positive predictive value (PPV) of 0.85 and a negative predictive value (NPV) of 0.53. A  
166 model including ADCD and age performed comparably (cvAUC 74% (95% CI 68-81)) with PPV

167 0.83 and NPV 0.52. The vibriocidal model was less predictive (cvAUC 54% (95% CI 47-61))  
168 with PPV 0.74 and NPV 0.26 (Figure 4). These results were consistent when we performed  
169 classification using an ensemble or “super learner” of three different classification models  
170 (random forest, penalized logistic regression, and support vector machine) (Figure S5).



171  
172 **Figure 4. Predicting infection outcome among household contacts using different subsets of**  
173 **biomarkers. A)** Cross-validated receiver operator curves (cvROC) for classifying household contacts of  
174 index cholera cases that remain uninfected vs. become infected using random forest models with different  
175 subsets of biomarkers and age. “5 biomarkers” corresponds to five of the top biomarkers selected via  
176 conditional importance, including ADCD, CtxB IgM, TcpA IgG2, Ogawa-OSP IgG1, and Sialidase IgG1.  
177 True and false positive rates calculated using leave-one-out cross-validation. **B)** Cross-validated area  
178 under the curve (cvAUC) corresponding to the models in A. Influence-curve based 95% confidence  
179 intervals are shown for the cvAUC estimates.

#### 180 181 Correlates of protection in challenged vaccine recipients

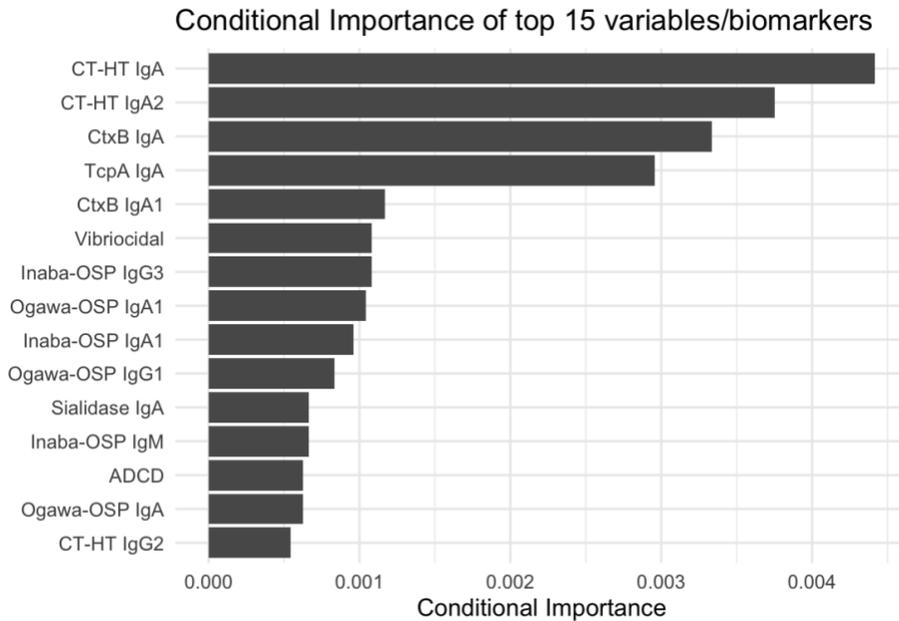
182 While household contacts of cholera cases have high risk of exposure in a real-world setting, a  
183 limitation of this study design for examining protection is that the degree and duration of  
184 exposure are unknown and varied among the study participants. To examine whether the  
185 primary correlates of protection for household contacts were consistent in vaccinees and in the  
186 setting of a controlled exposure, we replicated the analysis in a cohort of North American  
187 volunteers that received a live attenuated oral cholera vaccine (CVD 103-HgR) and were then  
188 challenged with approximately  $1 \times 10^5$  colony forming units of the wild type *V. cholerae* O1 El

189 Tor Inaba strain N16961 (Chen et al., 2016). This cohort included 34 individuals challenged 10  
 190 days post vaccination and 33 individuals challenged 90 days post vaccination (Table 2).  
 191 Participants were 18-45 years old, 16 (24%) were female, and 20 (30%) developed mild to  
 192 severe diarrhea following exposure (Table 2).

193  
 194 **Table 2. Characteristics of North American volunteers who were vaccinated and then challenged**  
 195 **with *V. cholerae***

Characteristic	Category	N	No qualifying diarrhea	Mild	Moderate	Severe
Challenge day post vaccination	10	34	29 (85.3%)	3 (8.8%)	1 (2.9%)	1 (2.9%)
	90	33	18 (54.5%)	11 (33.3%)	2 (6.1%)	2 (6.1%)
Age group	18-25	15	12 (80%)	3 (20%)	0 (0%)	0 (0%)
	26-35	34	24 (70.6%)	7 (20.6%)	1 (2.9%)	2 (5.9%)
	36-45	18	11 (61.1%)	4 (22.2%)	2 (11.1%)	1 (5.6%)
Sex	Female	16	11 (68.8%)	3 (18.8%)	0 (0%)	2 (12.5%)
	Male	51	36 (70.6%)	11 (21.6%)	3 (5.9%)	1 (2%)

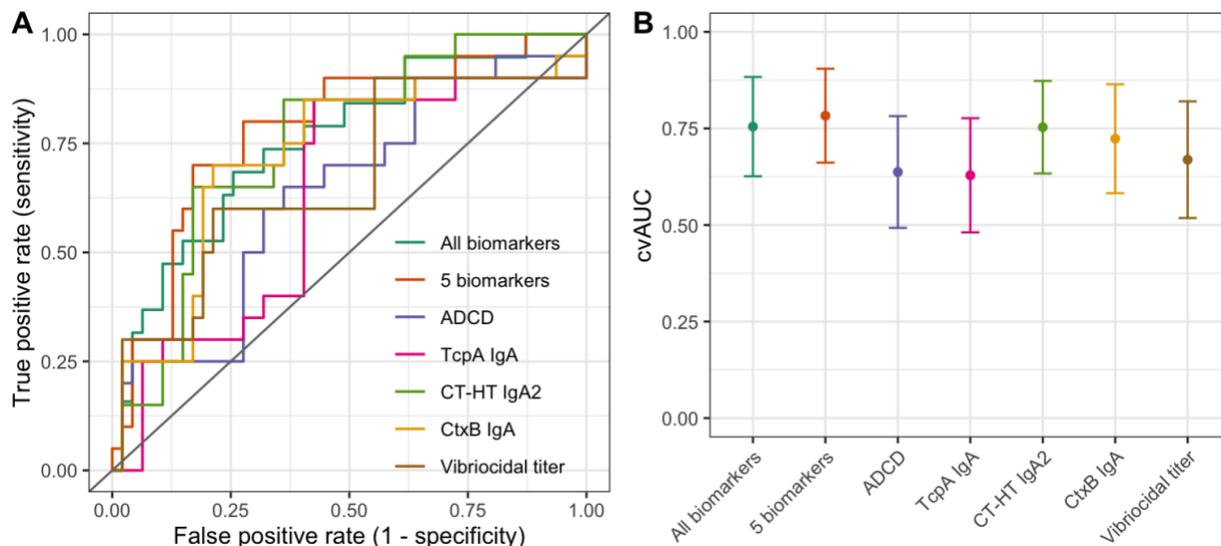
196  
 197  
 198 We measured serum antibody responses for the same set of biomarkers as above, apart from  
 199 ADCP (which was not measured due to limited sample volumes). As with the household  
 200 contacts, we found no separation of vaccinees by diarrhea outcome in an unsupervised analysis  
 201 of day-of-challenge serum biomarkers (Figure S6) and titers were highly correlated (Figure S7).  
 202 In contrast to the household contacts, CT-HT IgA and CtxB IgA responses were ranked most  
 203 conditionally important for classifying by diarrhea status (Figure 5). As with the household  
 204 contacts, several features ranked higher than vibriocidal titers (Figure 3, Figure 5). Differences  
 205 in biomarker importance between the cohorts persisted when we restricted the household  
 206 contacts to the same age groups as the vaccinees (Figure S8). Relative ranking of biomarkers  
 207 differed when we subset the vaccinees to those challenged on either day 10 or day 90 (Figure  
 208 S9), though sample sizes were small in the subsets (Table 2) and uncertainty was high (Figure  
 209 S10).



210

211 **Figure 5. Biomarkers important for classifying vaccinees by whether they developed diarrhea**  
212 **following *V. cholerae* challenge.** Top 15 biomarkers important in classifying vaccinees by developing  
213 either no qualifying diarrhea or mild to severe diarrhea following *V. cholerae* challenge. Biomarkers were  
214 selected using serum collected from participants on the day of challenge. Biomarkers are ranked by  
215 importance scores calculated using conditional random forest classification models.  
216

217 We found comparable cross-validated results for the vaccinees as the household contacts. The  
218 model that most accurately predicted developing cholera diarrhea included five of the top  
219 biomarkers and age with cvAUC 78% (95% CI, 66-91%) with a PPV of 0.83 and NPV of 0.65  
220 (Figure 6). Models that included only age and either CT-HT IgA2 or CtxB IgA performed  
221 comparably, followed by vibriocidal titers and then ADCD and TcpA IgA. The fold-increase in  
222 vibriocidal titers by day 10 following vaccination was less predictive of cholera diarrhea than  
223 day-of-challenge titers (Figure S11). These results were consistent when classification was  
224 performed using the ensemble model (Figure S12). Furthermore, the 5-biomarker fitted with the  
225 household contact data (Figure 4) performed nearly as well when predicting cholera diarrhea in  
226 the vaccinees with cvAUC 77% (95% CI, 64-90%) (Figure S13). In contrast, the 5-biomarker  
227 model fitted with the vaccinee data (Figure 6) performed poorly when predicting *V. cholerae*  
228 infection in the household contacts with cvAUC 60% (95% CI, 52-67%) (Figure S13).



229

230 **Figure 6. Predicting whether vaccinees develop diarrhea using different subsets of biomarkers. A)**

231 Cross-validated receiver operator curves (cvROC) for classifying vaccinees that develop diarrhea vs.

232 those that do not using random forest models with different subsets of biomarkers and age. “5

233 biomarkers” corresponds to five of the top biomarkers selected via conditional importance, including CT-

234 HT IgA, CtxB IgA, CT-HT IgA2, TcpA IgA, and Sialidase IgA2. True and false positive rates calculated

235 using leave-one-out cross-validation. **B)** Cross-validated area under the curve (cvAUC) corresponding to

236 the models in A. Influence-curve based 95% confidence intervals are shown for the cvAUC estimates.

237

## 238 Discussion

239 Vibriocidal antibody titers are considered the best accepted correlate of protection against *V.*

240 *cholerae* and are used as a proxy for predicting vaccine efficacy in Phase I/II clinical trials for

241 cholera vaccines (Iyer and Harris, 2021). Using a systems serology approach, we identified 19

242 additional biomarkers that were also associated with reduced risk of infection in household

243 contacts of patients with cholera. In models evaluating multiple immune measures, vibriocidal

244 antibodies were less important in classifying risk of infection than other markers. This was true

245 both in a cohort of household contacts and in a cohort of challenged cholera vaccinees. Our

246 finding, that numerous markers individually and in combination are more predictive of protection

247 from infection than vibriocidal titers, challenges the current standard that vibriocidal antibodies

248 are indeed the best available correlate of protection against cholera.

249           Though we identified many antibody correlates of protection against cholera, our ability  
250 to distinguish between immunologically protected and susceptible individuals using current  
251 serologic markers was not perfect. This may reflect non-immunologic factors that influence the  
252 outcome of infection, or the inherent challenges involved in using serum biomarkers to predict  
253 mucosal immunity. Some of the unexplained protection may be attributable to *V. cholerae* O1-  
254 specific memory B cell responses, which have been shown to be associated with protection  
255 even in the absence of elevated levels of circulating antibodies (Haney et al., 2018; Patel et al.,  
256 2012). In addition, secretory IgA and innate immune signaling at the mucosal surface play  
257 important roles in modulating susceptibility (Karlsson et al., 2013; Weil et al., 2019) but cannot  
258 be measured directly in serum. Systems serology approaches to distinguishing active vs. latent  
259 tuberculosis (Lu et al., 2016) and mild vs. severe SARS-CoV-2 (Bartsch et al., 2021) were more  
260 predictive than ours, with cross-validated accuracies of up to 91-97% for certain outcomes. In  
261 contrast, a systems serology analysis of protection from typhoid fever found predictive  
262 accuracies similar to ours (73-86%) (Jin et al., 2020). These discrepancies could reflect inherent  
263 differences in immunity in the lungs compared to the gut, or differences between intracellular  
264 and extracellular pathogens, and how those responses correlate with circulating antibodies.

265           Despite lack of a perfect predictive model, we were able to identify robust correlates of  
266 protection in cholera. Some of these correlates had been identified previously (Iyer and Harris,  
267 2021) while others, such as ADCD, are novel. Interestingly, ADCD was the single most  
268 predictive single marker of protection in household contacts and shares several similarities the  
269 vibriocidal titer marker. Both ADCD and vibriocidal titers are complement-dependent measures  
270 of functional antibody-mediated immunity. However, while vibriocidal titers reflect the  
271 downstream bactericidal activity of antibodies which activate complement via the classical  
272 pathway, ADCD is a bead-based, antigen-specific measure of C3 binding to *V. cholerae* O1-  
273 targeted antibodies. While there is ongoing debate whether circulating antibodies, transuded  
274 from the vascular compartment, could mediate complement dependent killing of *V. cholerae* and

275 other enteric pathogens on the surface of the intestinal mucosa (Iyer and Harris, 2021), it is  
276 notable that these predictive markers both implicate complement-dependent antibody  
277 responses. Considering this, and that exposure of *V. cholerae* to O-antigen binding IgG induces  
278 bacterial responses which protect against complement mediated killing (Baranova et al., 2018),  
279 raises the possibility that complement could play a role in protective immunity against *V.*  
280 *cholerae*.

281 In this study, we did not find a ‘universal signature’ or single set of biomarkers that was  
282 optimal for classifying outcomes across both cohorts (household contacts and challenged  
283 vaccinees) or outcomes (infection and illness). However, the model which best predicted  
284 susceptibility in household contacts also predicted susceptibility in vaccinees with similar  
285 accuracy. In contrast, a model trained to predict protection in challenged vaccinees in a 10 to  
286 90-day window did not accurately predict susceptibility among household contacts. This may  
287 reflect that a model trained in a single experimental setting is less likely to identify broadly  
288 applicable correlates of protection than models based on observed conditions in a cholera-  
289 endemic population. In other words, the household contact model may reflect correlates of  
290 protection which are applicable across a wider range of circumstances (e.g., different times  
291 between exposures, different strains of the bacteria, different populations) and to reflect multiple  
292 immune pathways which may lead to protection against cholera.

293 This study has limitations. We do not know if uninfected household contacts of index  
294 cholera cases were exposed to sufficient *V. cholerae* through contaminated water or food  
295 sources to be considered exposed yet protected. This may contribute to the variation seen in  
296 this cohort and may explain why predictive accuracy was slightly higher for vaccinees with  
297 known exposure status. In addition, we relied on previous measurements of vibriocidal titers in  
298 this cohort. In a subset of samples where we were able to repeat the vibriocidal assays, we  
299 found that the old and new responses were highly correlated (Figure S14) and gave comparable  
300 results in prediction (Figure S15). This suggests that changes over time in sample quality or

301 handling may not have affected the other biomarkers but does not eliminate the possibility.  
302 Lastly, the conditional permutation method we used to rank biomarkers in the random forest  
303 models has some inherent instability and may give preference to highly correlated variables  
304 (Debeer and Strobl, 2020). However, the consistency of the results from this approach with that  
305 of ensemble of three models (Figure S5, Figure S12) supports the robustness of our overall  
306 conclusions.

307 In summary, we have identified a set of serum biomarkers that can be used to monitor  
308 susceptibility to *V. cholerae* infection and symptoms, with varying sensitivity and specificity.  
309 These new markers could allow for flexibility in designing studies of population susceptibility to  
310 infection, provided that predictive accuracy metrics are used to adjust final estimates. In  
311 addition, future studies could examine whether antibody markers in alternative samples such as  
312 stool and/or saliva correlate with mucosal immunity and allow for more accurate prediction.  
313 Finally, these findings illustrate the need to design better observational studies that allow us to  
314 more accurately measure *V. cholerae* exposure, for example by including questionnaires on  
315 frequency and duration of interactions with index cases in addition to measures of  
316 infectiousness in household surveys (Coit et al., 2019). Such analyses would provide important  
317 insights into how immunity and other risk factors contribute to both individual outcomes and  
318 secondary transmission.

319

## 320 **Materials and Methods**

### 321 Study populations and sample collection

#### 322 *Case-ascertainment cohorts in Dhaka, Bangladesh*

323 We included samples and data from participants in Dhaka, Bangladesh from Dec. 2006 to Sept.  
324 2018 (Ritter et al., 2019). Symptom histories, blood, and fecal samples were collected  
325 prospectively from household contacts of hospital-ascertained cholera cases in Dhaka,  
326 Bangladesh. Households were defined as individuals who shared the same cooking pot for  $\geq 3$

327 days. Stool samples and rectal swabs were tested for presence of *V. cholerae*. Blood  
328 specimens were tested for serum antibody titers and ABO blood group. For this study, we  
329 included 261 household contacts who were stool negative upon enrollment, did not report  
330 diarrhea symptoms in the preceding week or at enrollment, and had sufficient sample material  
331 for laboratory analyses. We excluded infected contacts whose *V. cholerae* serotype did not  
332 match the index case and contacts with “possible infections” (i.e., stool negative with  $\geq 4$ -fold  
333 increase in vibriocidal titer and/or diarrhea symptoms).

334

#### 335 *Human challenge trial in North American volunteers*

336 We included samples and data from 67 North American volunteers with no history of cholera  
337 who were vaccinated with  $5 \times 10^8$  colony-forming units (CFU) of the live attenuated oral cholera  
338 vaccine CVD 103-HgR and then challenged with approximately  $5 \times 10^5$  *V. cholerae* O1 El Tor  
339 Inaba strain N16961 on either 10 ( $n = 34$ ) or 90 ( $n = 33$ ) days post vaccination (Chen et al.,  
340 2016). Challenged participants were monitored for diarrhea symptoms and each stool was  
341 graded; all loose stools (grades 3-5) were weighed, and volumes calculated.

342

#### 343 Outcomes and definitions

##### 344 *Case-ascertainment cohorts in Dhaka, Bangladesh*

345 The primary outcome in the household contacts cohort was *V. cholera* infection. Infection was  
346 defined as developing a positive stool culture result on day 7 or 30 after enrollment of the  
347 household’s index cholera case. A secondary outcome among infected household contacts was  
348 developing symptoms. Symptoms included watery diarrhea ( $\geq 3$  loose or watery stools in a 24-hr  
349 period), abdominal pain, vomiting, and/or fever.

350

##### 351 *Human challenge trial in North American volunteers*

352 The primary outcome in the vaccine challenge cohort was developing symptomatic cholera  
353 diarrhea. Diarrhea was defined as  $\geq 2$  loose stools of  $\geq 200$  mL each or a single loose stool of  
354  $\geq 300$  mL over a 48-hour period. Moderate diarrhea was passage of  $\geq 3$  L and severe diarrhea  
355 was passage of  $\geq 5$  L loose stool (Chen et al., 2016).

356

## 357 Laboratory methods

### 358 *Antigens*

359 *V. cholerae* Inaba and Ogawa OSP:BSA conjugates, TcpA, and *V. cholerae* sialidase were  
360 derived as previously described (Charles et al., 2017). *V. cholerae* holotoxin (List Biological,  
361 catalog 100B) and CtxB (Sigma, catalog C9903) were purchased from commercial vendors as  
362 indicated.

363

### 364 *Luminex assay for antigen-specific antibody responses*

365 We measured antigen-specific Ig class (IgG, IgA, IgM) and subclass (IgG1, IgG2, IgG, IgG4,  
366 IgA1, IgA2) responses against antigens using a customized Luminex assay. Each antigen was  
367 conjugated to unique, phycoerythrin (PE)-labeled, MagPlex-C Microsphere (Luminex Corp,  
368 Austin, TX) bead region at a concentration of 5  $\mu\text{g}$  antigen per  $1 \times 10^6$  microspheres using the  
369 principles of carbodiimide coupling and as per manufacturer's instructions (Luminex, xMAP®  
370 Antibody Coupling kit, catalog 40-50016).

371

372 Plasma samples were heat-inactivated by incubation at 56°C for 30 minutes and diluted in an  
373 assay buffer (0.1% BSA in 1X DPBS) to a final dilution of 1:100. Five  $\mu\text{l}$  of diluted plasma and  
374 45  $\mu\text{l}$  of diluted antigen-conjugated beads were added to a black 384-well polystyrene plate  
375 (Greiner Bio-One™, Monroe, NC) at a final concentration of 15 beads per  $\mu\text{l}$  (total: 675 beads  
376 per well). Plates were incubated overnight at 4°C with shaking (800 rpm) and then sonicated

377 and washed thrice with 0.1%BSA in 1X PBS, 0.05% Tween20 before addition of PE-conjugated  
378 anti-human IgG, IgM and IgA (Southern Biotech; 40 µl per well) for a final dilution of 1:154. After  
379 incubation at 1 hour with shaking, plates were sonicated and washed thrice with 0.1% BSA in  
380 1X PBS, 0.05% Tween20. The beads were resuspended in 40µl of sheath fluid (Fisher  
381 Scientific, Waltham, MA) and analyzed using Bioplex/Flexmap machines (BioRad). Samples  
382 were tested in duplicate, and data were analyzed as geometric mean fluorescence intensities.

383

#### 384 *Antigen biotinylation*

385 Ogawa OSP:BSA antigen (2 mg/ml stock) and Inaba OSP:BSA (2.5mg/ml) was biotinylated at  
386 20mM excess as per manufacturer's instructions (Thermofisher Scientific, catalog: A39256)  
387 followed by removal of unbound biotin using Zeba Spin column 7K MWCO (Thermofisher,  
388 catalog: 89882).

389

#### 390 *Antibody-dependent neutrophil phagocytosis (ADNP assay)*

391 Biotinylated Ogawa OSP:BSA antigen (2 mg/ml stock) and Inaba OSP:BSA (2.5mg/ml) was  
392 conjugated with 1.0µm yellow/green fluorescent neutravidin labeled microspheres  
393 (Thermofisher, catalog F8776) at 1:1 ratio and incubated at 37°C for 2 hours. The antigen-  
394 conjugated beads were pelleted, washed twice, and resuspended in 0.1% BSA in 1X DPBS and  
395 used within 4 days. Ten µl of heat inactivated (56°C for 30 mins) plasma diluted 1:4 in 0.1%  
396 BSA in 1X DPBS was added per well to a 96-well plate (Costar catalog:3799) followed by the  
397 addition of 10 µl/well of antigen-bead conjugate. Plates were incubated at 37°C 5% CO<sub>2</sub> for 2  
398 hours.

399

400 Blood collected from healthy donors was lysed using ammonium-chloride-potassium lysis buffer  
401 (150 mM NH<sub>4</sub>Cl, 10 mM KHCO<sub>3</sub>, 0.1 mM Na<sub>2</sub>EDTA, pH 7.4) at 1:10 ratio and incubated for 5  
402 mins at room temperature. Cells were pelleted at 500xg for 5 mins, washed with cold 1X DPBS

403 and resuspended in RPMI with 10% Fetal Calf Serum and 1X Penicillin/Streptomycin.  
404 Granulocytes were resuspended at a concentration of  $2.5 \times 10^5$  cells/ml. At the end of the 2-  
405 hour incubation, beads were pelleted, washed twice with 0.1% BSA in 1X DPBS and  
406 supernatants flicked. 200ul of resuspended cells (50,000 total cells) were added per well to the  
407 96-well plate and plates were incubated at 37°C for 4 hours. Cells were pelleted (5 min, 4°C,  
408 500 ×g), stained with CD66b-PacBlue (Biolegend), washed, and fixed with 4%  
409 paraformaldehyde. Samples were acquired within 48 hours on BD LSR Fortessa.

410  
411 Neutrophil bead internalization was quantified using FlowJo (FlowJo, LLC) software by gating  
412 for granulocytes (FSC/SSC)/neutrophils (CD66b+)/bead+neutrophils (FITC+). Phagocytic score  
413 was calculated by multiplying the percentage of neutrophils that internalized the beads with  
414 geometric mean fluorescence intensity of bead+ neutrophils+ cells divided by 10,000. Samples  
415 were run in singlicate and results are an average of 3 independent experiments run using 3  
416 different blood donors.

417  
418 *Antibody-dependent complement deposition (ADCD assay)*

419 12 µl of biotinylated Ogawa OSP:BSA or Inaba OSP:BSA antigen was conjugated with 1.0 µm  
420 red fluorescent, neutravidin labeled microspheres (Thermofisher, catalog F8775) per 96-well  
421 plate at 37°C for 2 hours. The antigen:conjugated beads were pelleted and washed twice with  
422 0.1%BSA in 1X DPBS followed by resuspension in 1200ul of 0.1% BSA in 1X DPBS. They were  
423 used within 4 days.

424  
425 Ten µl of heat inactivated (56°C for 30 mins) plasma diluted 1:4 in 5% BSA in 1X DPBS was  
426 added to a 96-well plate (Costar catalog:3799) followed by the addition of 10 µl of antigen-bead  
427 conjugate and incubation at 37°C, 5% CO<sub>2</sub>. Four µl of reconstituted guinea pig complement  
428 (Cedarlane; catalog CL4051) diluted 1:50 in Gelatin Veronal Buffer was added per well to the

429 96-well plate and incubated at 37°C for 20 mins. At the end of incubation, beads were pelleted  
430 at 2000×g for 10 mins and then washed twice with cold 15mM EDTA in 1X DPBS.  
431  
432 Fifty µl of FITC labeled goat anti-guinea pig C3 diluted 1:200 in 1X PBS was added to the plate  
433 and incubated in the dark at room temperature for 15 mins. Plates were washed twice, and the  
434 beads were resuspended in 100µl PBS. Geometric MFI of FITC gated on red, neutravidin beads  
435 were calculated using FlowJo software. Samples were tested in singlicate and results are an  
436 average of 3 independent experiments.

437

#### 438 *Antibody-dependent cellular phagocytosis*

439 Ogawa antigen bead conjugates were prepared as described for the ADNP assay. THP-1 cells  
440 were maintained in RPMI 1640 media (ATCC) containing 2 mM L-Glutamine (Corning), 10%  
441 Fetal Bovine Serum (Sigma), 10 mM HEPES (Corning), 55 µM beta-mercaptoethanol (Gibco),  
442 and 1X Penicillin/Streptomycin (Corning). A volume of 10 µL diluted human monoclonal  
443 antibody or plasma sample was added to each well, and immune complexes were formed over  
444 a 2-hour incubation at 37°C. After washing to remove non-specific unbound antibody, THP-1  
445 cells were added (200 µL/well) at a concentration of  $1.25 \times 10^5$  cells/mL ( $2.5 \times 10^4$  cells/well)  
446 and incubated with the immune complexed beads for 16 hours at 37°C. Cells were then fixed  
447 with 4% PFA and acquired on a BD LSRFortessa flow cytometer. The phagocytic score was  
448 calculated by multiplying the percentage of bead-positive cells by the geometric mean  
449 fluorescence intensities of the bead-positive cells divided by 10,000.

450

#### 451 Analytical methods

##### 452 *Single-marker correlates of infection*

453 All analyses were performed in the R Statistical Software v4.0.3 (R Core Team, 2020). We  
454 examined risk of infection (i.e., developing a positive stool culture result) among all household

455 contacts after exposure to an index cholera case, and risk of developing symptoms among  
456 infected household contacts. Odds ratios of developing infection or symptomatic cholera for  
457 each 2-fold increase in baseline biomarker titers were calculated using generalized estimating  
458 equations using the gee package (Carey et al., 2019), adjusting for household clustering and  
459 age. Biomarkers were centered and scaled prior to analysis. Biomarkers in logistic models with  
460 fitted values very close to 1 were excluded.

461

#### 462 *Unsupervised analysis*

463 We examined whether participants were separated by individual outcomes (*V. cholerae*  
464 infection and cholera diarrhea) using generalized principal component analysis (GLM-PCA),  
465 which allows for dimension reduction in non-normally distributed data. Analyses were run using  
466 the glmpca package with a negative binomial likelihood (Townes et al., 2019).

467

#### 468 *Supervised analyses*

469 We used conditional random forest models to examine which baseline biomarkers were most  
470 important for classifying individuals that went on to develop infection vs. those that remained  
471 uninfected and, of those infected, which went on to develop symptoms vs. which remained  
472 asymptomatic in the household contacts. We used the same approach to classify cholera  
473 diarrhea vs. no qualifying diarrhea in the vaccine challenge cohort. Models were run with 1000  
474 trees and class weights to account for imbalances in the data, and otherwise default settings  
475 from the party package were used (Strobl et al., 2008, 2007). Conditional permutation  
476 importance was used to rank variables via the perimp package (Debeer and Strobl, 2020).

477

478 We used leave-one-out cross-validation to examine the predictive validity of the conditional  
479 random forest models for the outcomes described above using different subsets of biomarkers  
480 and age. Predictions were visualized using receiver operator curves (cvROC) built using the

481 pROC package (Robin et al., 2011). The area under those curves (cvAUC) and corresponding  
482 influence curve-based confidence intervals were calculated using the cvAUC package (LeDell et  
483 al., 2014). In addition, we compared cvROC and cvAUC for the conditional random forest  
484 models to that of an ensemble of three different machine learning models: random forest,  
485 penalized logistic regression, and support vector machine. Analyses were run using class  
486 weights to account for imbalances in the dataset and otherwise default settings in the  
487 SuperLearner package (Polley et al., 2021).

488

#### 489 **Data availability**

490 All data generated in this study are available in the manuscript and Supplementary file 1. All  
491 input datasets and analytical code are available at: [https://github.com/HopkinsIDD/cholera-](https://github.com/HopkinsIDD/cholera-systems-serology)  
492 [systems-serology](https://github.com/HopkinsIDD/cholera-systems-serology) and <https://doi.org/10.5281/zenodo.6626079>.

493

#### 494 **Acknowledgements**

495 This research was supported by the icddr,b and extramural grants from the National Institutes of  
496 Health, including the National Institute of Allergy and Infectious Diseases, (R01 AI137164  
497 [J.B.H., R.C.C., G.A., F.Q.], R01 AI106878 [E.T.R., F.Q.], R01 AI130378 [T.R.B.], R01  
498 AI135115 [A.S.A., K.E.W.]). icddr,b is thankful to the donors for their support to its research  
499 efforts. icddr,b is also grateful to the governments of Bangladesh, Canada, Sweden, and the  
500 United Kingdom for providing core/unrestricted support.

501

#### 502 **Competing interests**

503 The authors declare that they have no competing interests.

504

#### 505 **Ethics**

506 This study was approved by icddr,b Ethical Review Committee and Massachusetts General  
507 Hospital's Institutional Review Boards.

508

## 509 **References**

510 Arnold KB, Chung AW. 2018. Prospects from systems serology research. *Immunology*

511 **153**:279–289. doi:10.1111/imm.12861

512 Azman AS, Lessler J, Luquero FJ, Bhuiyan TR, Khan AI, Chowdhury F, Kabir A, Gurwith M,

513 Weil AA, Harris JB, Calderwood SB, Ryan ET, Qadri F, Leung DT. 2019. Estimating

514 cholera incidence with cross-sectional serology. *Science Translational Medicine* **11**.

515 doi:10.1126/scitranslmed.aau6242

516 Baranova DE, Levinson KJ, Mantis NJ. 2018. *Vibrio cholerae* O1 secretes an extracellular

517 matrix in response to antibody-mediated agglutination. *PLoS One* **13**:e0190026.

518 doi:10.1371/journal.pone.0190026

519 Bartsch YC, Wang C, Zohar T, Fischinger S, Atyeo C, Burke JS, Kang J, Edlow AG, Fasano A,

520 Baden LR, Nilles EJ, Woolley AE, Karlson EW, Hopke AR, Irimia D, Fischer ES, Ryan

521 ET, Charles RC, Julg BD, Lauffenburger DA, Yonker LM, Alter G. 2021. Humoral

522 signatures of protective and pathological SARS-CoV-2 infection in children. *Nat Med*

523 **27**:454–462. doi:10.1038/s41591-021-01263-3

524 Carey VJ, Lumley T, Ripley B, Moler C. 2019. gee: generalized estimation equation solver. R

525 package version 4.13-20.

526 Charles RC, Nakajima R, Liang L, Jasinskas A, Berger A, Leung DT, Kelly M, Xu P, Kováč P,

527 Giffen SR, Harbison JD, Chowdhury F, Khan AI, Calderwood SB, Bhuiyan TR, Harris JB,

528 Felgner PL, Qadri F, Ryan ET. 2017. Plasma and mucosal immunoglobulin M,

529 immunoglobulin A, and immunoglobulin G responses to the *Vibrio cholerae* O1 protein

- 530 immunome in adults with cholera in Bangladesh. *The Journal of Infectious Diseases*  
531 **216**:125–134. doi:10.1093/infdis/jix253
- 532 Chen WH, Cohen MB, Kirkpatrick BD, Brady RC, Galloway D, Gurwith M, Hall RH, Kessler RA,  
533 Lock M, Haney D, Lyon CE, Pasetti MF, Simon JK, Szabo F, Tennant S, Levine MM.  
534 2016. Single-dose live oral cholera vaccine CVD 103-HgR protects against human  
535 experimental infection With *Vibrio cholerae* O1 El Tor. *Clinical Infectious Diseases*  
536 **62**:1329–1335. doi:10.1093/cid/ciw145
- 537 Chung AW, Alter G. 2017. Systems serology: profiling vaccine induced humoral immunity  
538 against HIV. *Retrovirology* **14**:57. doi:10.1186/s12977-017-0380-3
- 539 Coit J, Mendoza M, Pinedo C, Marin H, Chiang SS, Lecca L, Franke M. 2019. Performance of a  
540 household tuberculosis exposure survey among children in a Latin American setting.  
541 *The International Journal of Tuberculosis and Lung Disease* **23**:1223–1227.  
542 doi:10.5588/ijtld.18.0841
- 543 Debeer D, Strobl C. 2020. Conditional permutation importance revisited. *BMC Bioinformatics*  
544 **21**:307. doi:10.1186/s12859-020-03622-2
- 545 Haney DJ, Lock MD, Gurwith M, Simon JK, Ishioka G, Cohen MB, Kirkpatrick BD, Lyon CE,  
546 Chen WH, Sztein MB, Levine MM, Harris JB. 2018. Lipopolysaccharide-specific memory  
547 B cell responses to an attenuated live cholera vaccine are associated with protection  
548 against *Vibrio cholerae* infection. *Vaccine* **36**:2768–2773.  
549 doi:10.1016/j.vaccine.2018.04.011
- 550 Harris JB, LaRocque RC, Chowdhury F, Khan AI, Logvinenko T, Faruque ASG, Ryan ET, Qadri  
551 F, Calderwood SB. 2008. Susceptibility to *Vibrio cholerae* infection in a cohort of  
552 household contacts of patients with cholera in Bangladesh. *PLOS Neglected Tropical*  
553 *Diseases* **2**:e221. doi:10.1371/journal.pntd.0000221
- 554 Iyer AS, Harris JB. 2021. Correlates of protection for cholera. *The Journal of Infectious*  
555 *Diseases* **224**:S732–S737. doi:10.1093/infdis/jiab497

- 556 Jin C, Hill J, Gunn BM, Yu W-H, Dahora LC, Jones E, Johnson M, Gibani MM, Spreng RL, Alam  
557 SM, Nebykova A, Juel HB, Dennison SM, Seaton KE, Fallon JK, Tomaras GD, Alter G,  
558 Pollard AJ. 2020. Vi-specific serological correlates of protection for typhoid fever. *J Exp*  
559 *Med* **218**:e20201116. doi:10.1084/jem.20201116
- 560 Kaisar MH, Bhuiyan MS, Akter A, Saleem D, Iyer AS, Dash P, Hakim A, Chowdhury F, Khan AI,  
561 Calderwood SB, Harris JB, Ryan ET, Qadri F, Charles RC, Bhuiyan TR. 2021. *Vibrio*  
562 *cholerae* sialidase-specific immune responses are associated with protection against  
563 cholera. *mSphere* **6**:e01232-20. doi:10.1128/mSphere.01232-20
- 564 Karlsson EK, Harris JB, Tabrizi S, Rahman A, Shlyakhter I, Patterson N, O'Dushlaine C,  
565 Schaffner SF, Gupta S, Chowdhury F, Sheikh A, Shin OS, Ellis C, Becker CE, Stuart  
566 LM, Calderwood SB, Ryan ET, Qadri F, Sabeti PC, LaRocque RC. 2013. Natural  
567 selection in a Bangladeshi population from the cholera-endemic Ganges River Delta.  
568 *Science Translational Medicine* **5**:192ra86-192ra86. doi:10.1126/scitranslmed.3006338
- 569 Kauffman RC, Bhuiyan TR, Nakajima R, Mayo-Smith LM, Rashu R, Hoq MR, Chowdhury F,  
570 Khan AI, Rahman A, Bhaumik SK, Harris L, O'Neal JT, Trost JF, Alam NH, Jasinskas A,  
571 Dotsey E, Kelly M, Charles RC, Xu P, Kováč P, Calderwood SB, Ryan ET, Felgner PL,  
572 Qadri F, Wrammert J, Harris JB. 2016. Single-cell analysis of the plasmablast response  
573 to *Vibrio cholerae* demonstrates expansion of cross-reactive memory B cells. *mBio*  
574 **7**:e02021-16. doi:10.1128/mBio.02021-16
- 575 LeDell E, Petersen M, Laan M van der. 2014. cvAUC: cross-validated area under the ROC  
576 curve confidence intervals.
- 577 Levade I, Saber MM, Midani FS, Chowdhury F, Khan AI, Begum YA, Ryan ET, David LA,  
578 Calderwood SB, Harris JB, LaRocque RC, Qadri F, Shapiro BJ, Weil AA. 2021.  
579 Predicting *Vibrio cholerae* infection and disease severity using metagenomics in a  
580 prospective cohort study. *J Infect Dis* **223**:342–351. doi:10.1093/infdis/jiaa358

581 Lu LL, Chung AW, Rosebrock T, Ghebremichael M, Yu WH, Grace PS, Schoen MK, Tafesse F,  
582 Martin C, Leung V, Mahan AE, Sips M, Kumar M, Tedesco J, Robinson H, Tkachenko E,  
583 Draghi M, Freedberg KJ, Streeck H, Suscovich TJ, Lauffenburger D, Restrepo BI, Day  
584 C, Fortune SM, Alter G. 2016. A functional role for antibodies in tuberculosis. *Cell*  
585 **167**:433-443.e14. doi:10.1016/j.cell.2016.08.072

586 McGill R, Tukey JW, Larsen WA. 1978. Variations of box plots. *The American Statistician*  
587 **32**:12–16. doi:10.2307/2683468

588 Midani FS, Weil AA, Chowdhury F, Begum YA, Khan AI, Debela MD, Durand HK, Reese AT,  
589 Nimmagadda SN, Silverman JD, Ellis CN, Ryan ET, Calderwood SB, Harris JB, Qadri F,  
590 David LA, LaRocque RC. 2018. Human gut microbiota predicts susceptibility to *Vibrio*  
591 *cholerae* infection. *J Infect Dis* **218**:645–653. doi:10.1093/infdis/jiy192

592 Patel SM, Rahman MA, Mohasin M, Riyadh MA, Leung DT, Alam MM, Chowdhury F, Khan AI,  
593 Weil AA, Aktar A, Nazim M, LaRocque RC, Ryan ET, Calderwood SB, Qadri F, Harris  
594 JB. 2012. Memory B cell responses to *Vibrio cholerae* O1 lipopolysaccharide are  
595 associated with protection against infection from household contacts of patients with  
596 cholera in Bangladesh. *Clinical and Vaccine Immunology* **19**:842–848.  
597 doi:10.1128/CVI.00037-12

598 Polley E, LeDell E, Kennedy C, Lendle S, Laan M van der. 2021. SuperLearner: super learner  
599 prediction.

600 R Core Team. 2020. R: a language and environment for statistical computing. R Foundation for  
601 Statistical Computing, Vienna, Austria.

602 Ritter AS, Chowdhury F, Franke MF, Becker RL, Bhuiyan TR, Khan AI, Saha NC, Ryan ET,  
603 Calderwood SB, LaRocque RC, Harris JB, Qadri F, Weil AA. 2019. Vibriocidal titer and  
604 protection from cholera in children. *Open Forum Infectious Diseases* **6**.  
605 doi:10.1093/ofid/ofz057

- 606 Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. 2011. pROC: an  
607 open-source package for R and S+ to analyze and compare ROC curves. *BMC*  
608 *Bioinformatics* **12**:77. doi:10.1186/1471-2105-12-77
- 609 Strobl C, Boulesteix A-L, Kneib T, Augustin T, Zeileis A. 2008. Conditional variable importance  
610 for random forests. *BMC Bioinformatics* **9**:307. doi:10.1186/1471-2105-9-307
- 611 Strobl C, Boulesteix A-L, Zeileis A, Hothorn T. 2007. Bias in random forest variable importance  
612 measures: Illustrations, sources and a solution. *BMC Bioinformatics* **8**:25.  
613 doi:10.1186/1471-2105-8-25
- 614 Townes FW, Hicks SC, Aryee MJ, Irizarry RA. 2019. Feature selection and dimension reduction  
615 for single-cell RNA-Seq based on a multinomial model. *Genome Biology* **20**:295.  
616 doi:10.1186/s13059-019-1861-6
- 617 Weil AA, Becker RL, Harris JB. 2019. Vibrio cholerae at the intersection of immunity and the  
618 microbiome. *mSphere* **4**. doi:10.1128/mSphere.00597-19
- 619 Weil AA, Khan AI, Chowdhury F, LaRocque RC, Faruque ASG, Ryan ET, Calderwood SB,  
620 Qadri F, Harris JB. 2009. Clinical outcomes in household contacts of patients with  
621 cholera in Bangladesh. *Clin Infect Dis* **49**:1473–1479. doi:10.1086/644779  
622