




OPEN

## Infection induced SARS-CoV-2 seroprevalence and heterogeneity of antibody responses in a general population cohort study in Catalonia Spain

Marianna Karachaliou<sup>1,10</sup>, Gemma Moncunill<sup>1,2,10</sup>, Ana Espinosa<sup>1</sup>, Gemma Castaño-Vinyals<sup>1</sup>, Alfons Jiménez<sup>1,3</sup>, Marta Vidal<sup>1</sup>, Rebeca Santano<sup>1</sup>, Diana Barrios<sup>1</sup>, Laura Puyol<sup>1</sup>, Anna Carreras<sup>4</sup>, Leonie Mayer<sup>1</sup>, Rocío Rubio<sup>1</sup>, Beatriz Cortés<sup>4</sup>, Vanessa Pleguezuelos<sup>5</sup>, Cristina O'Callaghan-Gordo<sup>1,3,6,7</sup>, Serena Fossati<sup>1</sup>, Ioar Rivas<sup>1,6,7</sup>, Delphine Casabonne<sup>3,8</sup>, Martine Vrijheid<sup>1,3,6</sup>, Luis Izquierdo<sup>1</sup>, Ruth Aguilar<sup>1</sup>, Xavier Basagaña<sup>1</sup>, Judith Garcia-Aymerich<sup>1,3,6</sup>, Rafael de Cid<sup>4</sup>, Carlota Dobaño<sup>1,2,11</sup> & Manolis Kogevinas<sup>1,3,6,9,11</sup>

Sparse data exist on the complex natural immunity to SARS-CoV-2 at the population level. We applied a well-validated multiplex serology test in 5000 participants of a general population study in Catalonia in blood samples collected from end June to mid November 2020. Based on responses to fifteen isotype-antigen combinations, we detected a seroprevalence of 18.1% in adults ( $n = 4740$ ), and modeled extrapolation to the general population of Catalonia indicated a 15.3% seroprevalence. Antibodies persisted up to 9 months after infection. Immune profiling of infected individuals revealed that with increasing severity of infection (asymptomatic, 1–3 symptoms,  $\geq 4$  symptoms, admitted to hospital/ICU), seroresponses were more robust and rich with a shift towards IgG over IgA and anti-spike over anti-nucleocapsid responses. Among seropositive participants, lower antibody levels were observed for those  $\geq 60$  years vs  $< 60$  years old and smokers vs non-smokers. Overweight/obese participants vs normal weight had higher antibody levels. Adolescents (13–15 years old) ( $n = 260$ ) showed a seroprevalence of 11.5%, were less likely to be tested seropositive compared to their parents and had dominant anti-spike rather than anti-nucleocapsid IgG responses. Our study provides an unbiased estimate of SARS-CoV-2 seroprevalence in Catalonia and new evidence on the durability and heterogeneity of post-infection immunity.

### Abbreviations

SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
COVID-19	Coronavirus disease 2019
IgM	Immunoglobulin M
IgG	Immunoglobulin G
IgA	Immunoglobulin A
COVICAT	COVID-19 Cohorts in CATalonia

<sup>1</sup>Barcelona Institute for Global Health (ISGlobal), 08036 Barcelona, Spain. <sup>2</sup>Centro de Investigação em Saúde de Manhiça (CISM), Maputo, Mozambique. <sup>3</sup>Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), 08036 Madrid, Spain. <sup>4</sup>Genomes for Life-GCAT Lab, Institute for Health Science Research Germans Trias i Pujol (IGTP), Badalona, Spain. <sup>5</sup>Banc de Sang i Teixits (BST), Barcelona, Spain. <sup>6</sup>Universitat Pompeu Fabra (UPF), Barcelona, Spain. <sup>7</sup>Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain. <sup>8</sup>Unit of Molecular and Genetic Epidemiology in Infections and Cancer (UNIC-Molecular), Cancer Epidemiology Research Programme, IDIBELL, Catalan Institute of Oncology, Hospitalet De Llobregat, Spain. <sup>9</sup>Hospital del Mar Medical Research Institute (IMIM), 08003 Barcelona, Spain. <sup>10</sup>These authors contributed equally: Marianna Karachaliou and Gemma Moncunill. <sup>11</sup>These authors jointly supervised this work: Carlota Dobaño and Manolis Kogevinas. email: marianna.karachaliou@isglobal.org

S	Spike full protein
S2	S2 fragment
RBD	Receptor binding domain
NFL	Nucleocapsid full protein
NCT	Nucleocapsid C-terminal region
ICU	Intensive Care Unit
BMI	Body mass index
INMA	Infancia y Medio Ambiente
MFI	Median fluorescence intensity

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) surveillance based on diagnostic testing, incomplete screening of all possible infections and imperfect test sensitivity may lead to a domino-like effect resulting in significant underestimation of the number of coronavirus disease 2019 (COVID-19) cases<sup>1</sup>. The high proportion of asymptomatic cases distorts even more the picture of the pandemic<sup>2</sup>.

Seroepidemiological studies have emerged across the world in order to provide us with a better estimate of the proportion of the population previously infected (vaccine-induced immunity is distinguishable)<sup>3</sup>. Nonetheless, many of these studies target specific populations (e.g. health care workers, previous hospitalized COVID-19 patients), use not well validated laboratory methods<sup>4</sup> and have been mainly geared toward studying IgG responses to only one antigen<sup>5</sup>. Multiplex serology may improve the diagnostic power of such studies given the considerable heterogeneity in antibody responses between individuals. In particular, the virus has several antigenic epitopes that are the target of antibodies but not everyone responds to the same antigens<sup>6</sup>. Additionally, detection of certain isotype responses depends on the time since infection<sup>7–9</sup>. Within days of symptom onset, specific immunoglobulins M (IgM) are detected and after a lag period strong immunoglobulins G (IgG) responses typically occur. Immunoglobulin A (IgA) responses are detected almost concurrently to IgM or earlier. With time, attenuation of antibody levels is expected due to decay of immune responses and transition of immunoglobulin production from short to long-lived plasma cell; thus cut-offs for seropositivity should take into account levels of waning immunity<sup>10,11</sup>. Moreover, the magnitude and type of antibody response correlates with disease severity. For example, most studies show that seroresponses are higher in more severe cases<sup>12,13</sup>. Recent data also show that multiplex serology is better correlated with levels of protective immunity<sup>14</sup>.

Limited data exist on the trajectories of antibody responses to SARS-CoV-2 over time and the factors that determine their heterogeneity. Notably, most studies consider individuals hospitalized or at least requiring some outpatient treatment<sup>9,15–20</sup>. Describing the characteristics of an effective immune response, as such encountered by asymptomatics or those with mild infections, is valuable. Early data show that some antigen and/or isotype responses dominate among milder infections<sup>8,19,20</sup>. Children are also facing effectively the infection, and studies comparing immune responses between SARS-CoV-2 infected children and adults have already provided some insights<sup>21,22</sup>. Considering members of the same family may resolve further questions related to time of infection, genetics, and other shared environmental exposures.

Taking advantage of multiplex serology to SARS-CoV-2, we describe the presence and heterogeneity of antibody responses in a population of 13–93 years old participants of existing cohort studies in Catalonia up to mid-November 2020. Catalonia in northeast Spain, has been among the hardest-hit populations in Europe from COVID-19.

## Results

**SARS-CoV-2 seroprevalence.** Among the 10,837 adult participants of the COVID-19 Cohorts in Catalonia (COVICAT) study, the 4740 (44%) who donated a blood sample for serological testing were more likely to have reported symptoms, not having been tested before, be of higher education and less likely to work in their usual workplace during confinement and be smokers before confinement compared to those who participated only with questionnaire data (Supplementary Resource 1). A blood sample was available for all adolescents.

Table 1 presents the seroprevalence of SARS-CoV-2 based on the serostatus of fifteen isotype-antigen combinations [three isotypes: IgM, IgA and IgG; five viral target antigens: spike full protein (S), S2 fragment (S2), receptor binding domain (RBD), nucleocapsid full protein (NFL) and nucleocapsid C-terminal region (NCT)]. Details on the contribution of each isotype-antigen combination in the overall serostatus are available in Supplementary Resource 2. The overall SARS-CoV-2 seroprevalence among adults was 18.1% (IgM 3.7%, IgA 14.6%, IgG 9.0%) while 11.6% had an undetermined status. The highest prevalences were observed for RBD IgG (8.0%), S IgG (7.4%), RBD IgA (7.1%) and S IgA (6.9%). Adult participants of the second sampling period ( $n = 1089$ , 23%) (median date: 30 October 2020, range 8 September–17 November 2020) were more likely to be seropositive compared to those of the first sampling period (median date: 19 July 2020, range 23 June–31 July 2020) (23.8% versus 16.4%) (Supplementary Resource 3). To extrapolate the study results to Catalonia's adult population, we used raked weights to balance the study sample characteristics (age, sex, educational level, health region, smoking) to those of the total population (more details in methods). We estimated a seroprevalence of 15.3% in adults in Catalonia. Among the 260 adolescents (13–15 years old), 11.5% were seropositive and 7.3% had an undetermined status (Table 1).

**Antibody levels in time since infection.** We examined the association of time since infection with antibody levels using cross-sectional data from all seropositive adults with an estimated time since infection ranging from 23 to 273 days (Fig. 1). For each isotype-antigen combination, levels are plotted irrespective of the serostatus to the specific combination to not affect the heterogeneity of responses expected with time. We observed that RBD, S, and S2 IgM levels decreased significantly over time, with NFL and NCT IgMs being less affected; very

	Adults (n = 4740)						Adolescents (n = 260)					
	Positive		Negative		Undetermined		Positive		Negative		Undetermined	
	n	%	n	%	n	%	n	%	n	%	n	%
Overall	858	18.1	3331	70.3	551	11.6	30	11.5	211	81.2	19	7.3
<b>By isotype</b>												
IgM	175	3.7	4392	92.7	173	3.6	3	1.2	250	96.2	7	2.7
IgA	694	14.6	3616	76.3	430	9.1	19	7.3	233	89.6	8	3.1
IgG	426	9.0	4134	87.2	180	3.8	24	9.2	225	86.5	11	4.2
<b>By isotype-antigen combination</b>												
IgM NFL	6	0.1	4694	99.0	40	0.8	0	0	260	100	0	0.0
IgM NCt	18	0.4	4674	98.6	48	1.0	0	0	258	99.2	2	0.8
IgM RBD	88	1.9	4552	96.0	100	2.1	1	0.4	253	97.3	6	2.3
IgM S	49	1.0	4601	97.1	90	1.9	0	0	259	99.6	1	0.4
IgM S2	20	0.4	4637	97.8	83	1.8	0	0	258	99.2	2	0.8
IgA NFL	140	3.0	4310	90.9	290	6.1	1	0.4	256	98.5	3	1.2
IgA NCt	129	2.7	4479	94.5	132	2.8	0	0	257	98.8	3	1.2
IgA RBD	338	7.1	4210	88.8	192	4.1	17	6.5	240	92.3	3	1.2
IgA S	325	6.9	4259	89.9	156	3.3	13	5.0	245	94.2	2	0.8
IgA S2	209	4.4	4201	88.6	330	7.0	4	1.5	255	98.1	1	0.4
IgG NFL	59	1.2	4473	94.4	208	4.4	2	0.8	246	94.6	12	4.6
IgG NCt	98	2.1	4481	94.5	161	3.4	5	1.9	245	94.2	10	3.8
IgG RBD	380	8.0	4308	90.9	52	1.1	21	8.1	237	91.2	2	0.8
IgG S	351	7.4	4307	90.9	82	1.7	21	8.1	233	89.6	6	2.3
IgG S2	104	2.2	4353	91.8	283	6.0	6	2.3	233	89.6	21	8.1

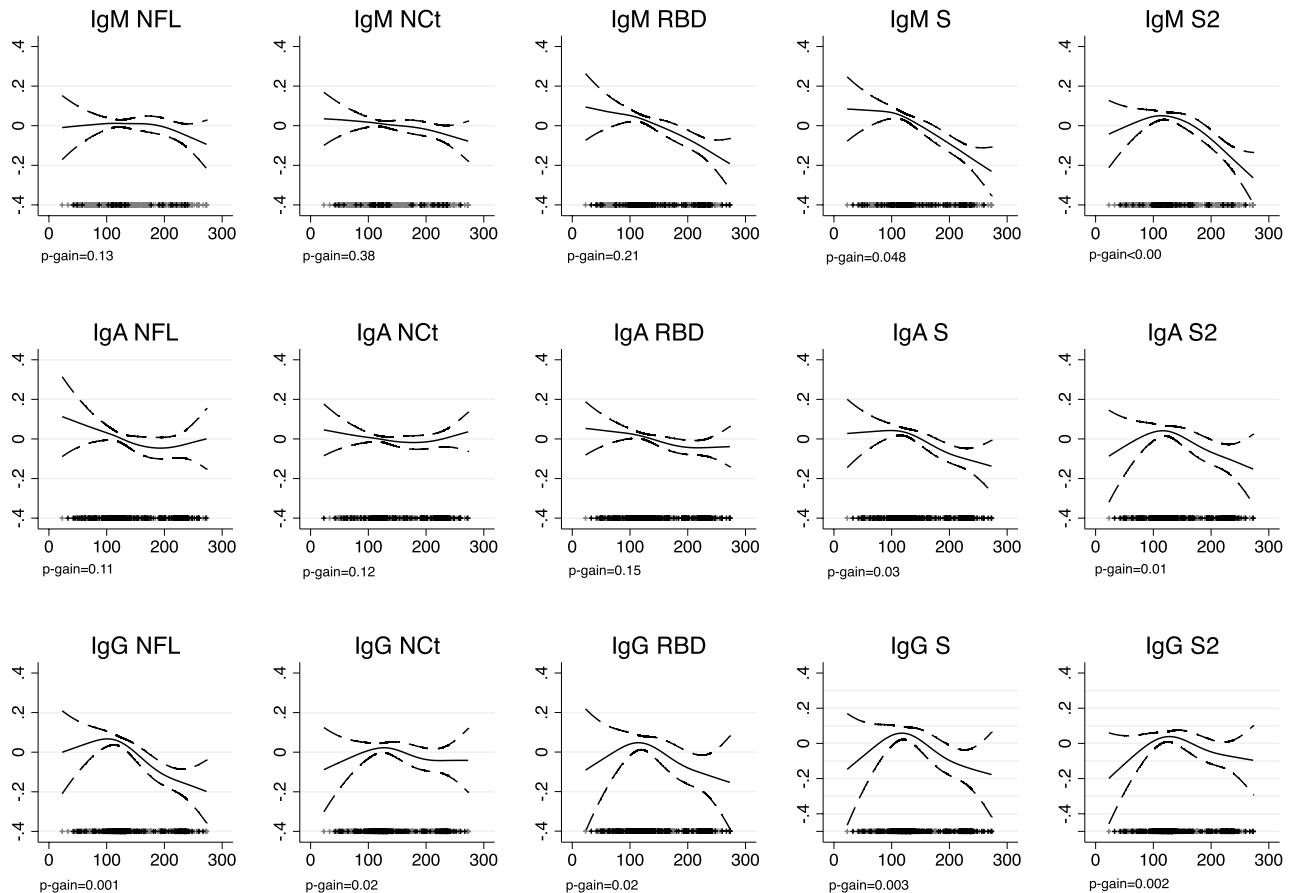
**Table 1.** SARS-CoV-2 seroprevalence (overall, by isotype, by isotype-antigen combination) in adult and adolescent participants of the COVICAT study in Catalonia (sampling period: end June to mid November 2020).

few participants were actually seropositive to the specific isotype-antigen combinations. IgA responses to NFL, NCt and RBD sustained in time but those of IgA to S and S2 declined after around 120 days to lower levels. IgG responses seemed to be stable or peaking at around 100 days after infection and then started to decline rapidly for NFL IgG, modestly for RBD and S IgG, while NCt and S2 IgG levels were minimally affected. In a subgroup of 99 participants who were previously tested positive (self-reported result), 92% had a positive multiplex serology at a median of 102 days after first diagnosis (range 13–233 days) (Supplementary Resource 4).

**SARS-CoV-2 serology by COVID-19 symptoms.** We present the distribution of symptoms among adults and adolescents by SARS-CoV-2 serostatus in Table 2. Among adults, all symptoms were more prevalent ( $p$ -value < 0.05) in SARS-CoV-2 seropositive versus seronegative participants with most remarkable differences seen for loss of odor/taste and fever. Among seropositives, 38.4% were asymptomatic and seropositives versus seronegatives were more likely to report  $\geq 4$  symptoms (28.5% vs 9.6%). The distribution of symptoms among seropositive adolescents was slightly different compared to adults, with a statistically significant higher proportion reporting chest pain (25% vs. 8% in adolescents and adults respectively), and a lower proportion reporting fever (16.7% vs. 30.6%) and respiratory symptoms (cough and dyspnea 25% vs. 39.1%). Having four or more symptoms, fatigue, chest pain, runny nose, loss of odor/taste and fever were statistically significant more prevalent in seropositive versus seronegative adolescents.

Demographic and clinical characteristics of adult participants according to SARS-CoV-2 serostatus and severity of infection are presented in Supplementary Resource 5. Participants reporting contact with a COVID-19 case and non-smokers were more likely to be seropositive. The proportion of participants reporting contact with a COVID-19 case, being previously tested, reporting any chronic disease or being overweight/obese increased with the severity of infection.

**Antibody responses by the severity of infection.** To determine whether the severity of infection is associated with quantitatively and qualitatively different antibody responses we performed four analyses. Firstly, we compared antibody levels of the 15 isotype-antigen combinations between asymptomatic individuals ( $n = 322$ ), those reporting 1–3 symptoms ( $n = 276$ ), those reporting  $\geq 4$  symptoms ( $n = 216$ ) and those admitted to hospital/ICU ( $n = 24$ ) (Fig. 2a). We observed lower levels among asymptomatics and higher levels among those admitted to hospital/ICU (apart from NFL and NCt IgMs). Gradient differences were most evident among IgG to NFL and NCt and all RBD, S and S2 responses. Secondly, we found that the breadth of positive immune responses (aggregate number of seropositive isotype-antigen combinations), increased with severity of infection (Fig. 2b). Thirdly, we explored for differences in isotype responses. We observed a higher increase in levels



**Figure 1.** Generalized additive models for associations (95% CIs) of days since infection with antibody responses to the fifteen isotype-antigen combinations in seropositive participants of the COVICAT study. Plus symbols (+) represent overall seropositive participants and with grey color are participants seronegative to the specific isotype-antigen combination. For each isotype-antigen combination, levels are plotted irrespective of the serostatus to the specific combination.

of IgG than IgA responses with increasing severity (Fig. 2a). In Fig. 2c, using IgA/IgG ratios, we observed that IgA levels were closer to IgG levels among asymptomatics and more likely to exceed them compared to those with more severe infection. Also, we found that among asymptomatics, a higher proportion (46%) had more positive IgA than IgG responses compared to the 13% displaying more IgG than IgA responses (Fig. 2d). We observed reverse findings among those hospitalized. Finally, we compared responses related to spike protein versus nucleocapsid antigens. Experiencing a more severe infection was associated with a shift towards spike over nucleocapsid antibody responses (Fig. 2e). The overall trend is reflected in the last graph of Fig. 2e depicting differences in the number of features that had greater spike than nucleocapsid related responses (ratios over one). Because time since infection may impact the associations mentioned above, we repeated all analyses in two strata of seropositive individuals those sampled before and after 120 days since infection. Results were materially unchanged (data not shown).

**Antibody responses by age, sex, and lifestyle characteristics.** We examined for differences in antibody levels and breadth of positive immune responses among seropositive adults with respect to age, sex, smoking, and body mass index (BMI) status before confinement (Table 3, Supplementary Resource 7). Participants 60 years of age or older had lower responses to almost all isotype-antigen combinations and a lower breadth of positive responses but had higher levels of NFL IgA. Females had statistically significant higher NFL, NCt and S2 IgM responses but lower NFL, NCt, and S2 IgA responses. Overweight or obese people had higher levels to almost all IgA and IgG responses and a higher breadth of positive responses. On the other hand, smokers displayed lower levels of almost all antibodies and a lower breadth of positive responses. We additionally adjusted for the severity of infection, considering it as a mediator of the associations. After adjustment, age  $\geq 60$  years old was associated with lower levels of IgM to S2 but higher levels of IgA to NFL. Associations with sex remained, with women showing an overall lower number of positive responses. Associations with overweight/obesity were largely diluted but remained positive. Smoking was consistently associated with lower levels and a lower breadth of positive responses. We repeated all the analyses excluding those seronegative for each isotype-antigen combination and results for smoking and BMI status were similar (Supplementary Resource 7).

**Table 2.** Symptoms prevalence (%) by SARS-CoV-2 serostatus in adult and adolescent participants of the COVICAT study in Catalonia.

Symptoms	Adults		Symptoms	Adolescents	
	seropositive	seronegative or undetermined		seropositive	seronegative or undetermined
Headache	35.1	23.4	Fatigue	37.5	17.3
Muscle/joint pain	35.0	17.7	Headache	33.3	26.2
Fever	30.6	9.4	Muscle/joint pain	29.2	14.9
Fatigue	30.5	15.0	Chest pain	25.0	6.0
Cough	27.0	13.9	Runny nose	25.0	10.7
Loss of odor/taste	22.2	1.6	Loss odor/taste	25.0	1.2
Runny nose	17.5	14.6	Nausea/vomiting	20.8	6.5
Diarrhea	17.5	10.2	Diarrhea	16.7	7.7
Dyspnea	12.1	5.7	Fever	16.7	5.4
Rash	8.6	4.2	Cough	16.7	10.7
Chest pain	8.0	5.1	Dyspnea	8.3	5.4
Nausea	5.8	3.0	Rash	0	4.2
<b>Number of symptoms</b>			<b>Number of symptoms</b>		
0 symptoms	38.4	54.6	0 symptoms	33.3	54.8
1-3 symptoms	33.1	35.8	1-3 symptoms	29.2	35.7
≥4 symptoms	28.5	9.6	≥4 symptoms	37.5	9.5

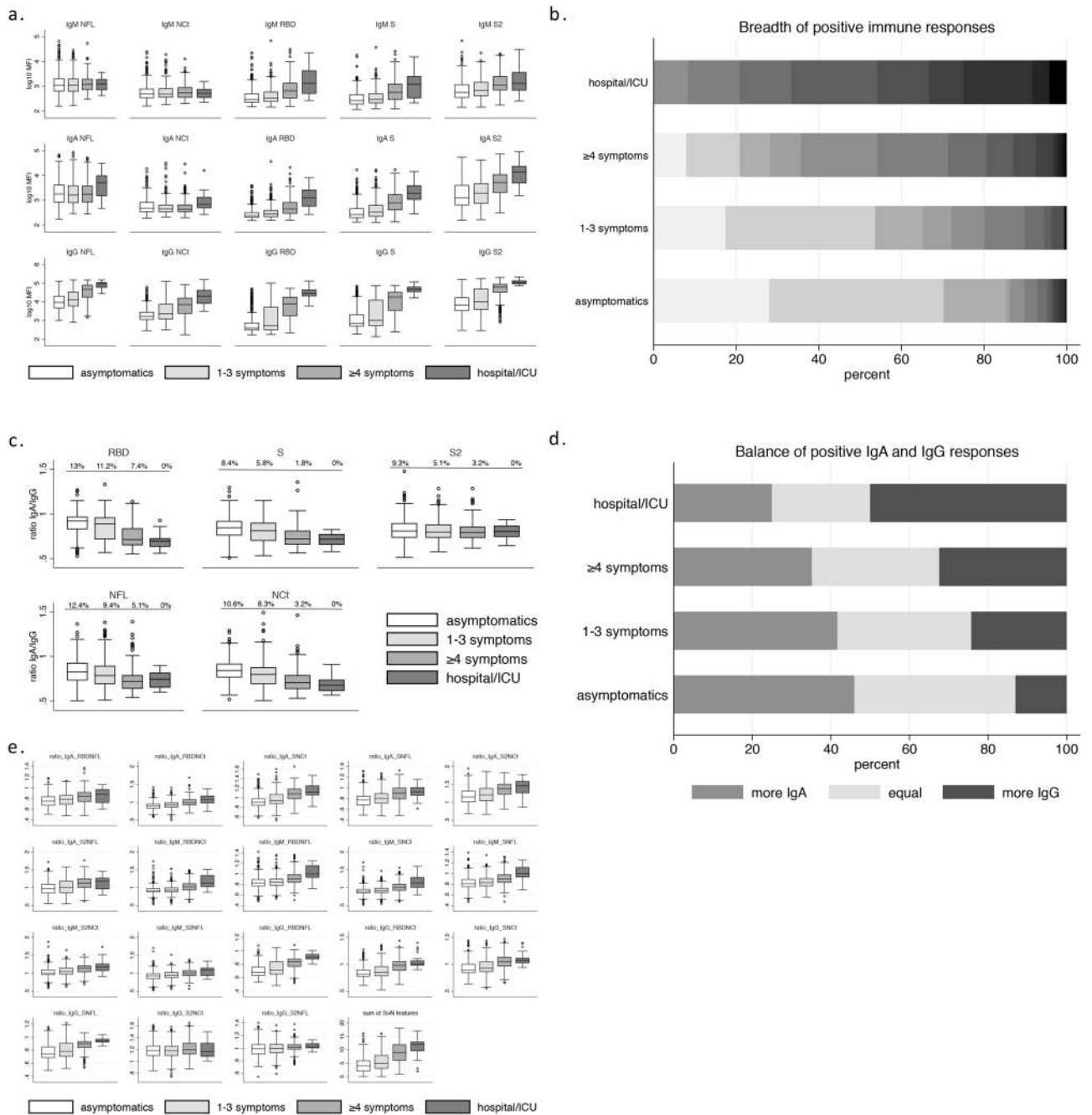
**Table 2.** Symptoms prevalence (%) by SARS-CoV-2 serostatus in adult and adolescent participants of the COVICAT study in Catalonia. Symptoms are sorted in adults and adolescents in decreasing frequency as observed among seropositives. Darker red = higher prevalence.

**Antibody responses among adolescents.** Serological data among 260 parent–child pairs showed a much lower risk for seropositivity among adolescents ( $n=30$ ) than their parents ( $n=50$ ) [RR: 0.6, 95% CI 0.39–0.91]. Sample collection took place at the same day for parents and their children. Among seropositive, adolescents had higher responses to S, S2 and RBD IgG, whereas parents had higher responses to NFL and NCT IgA (Supplementary Resource 8). The dominant IgG responses related to spike protein (S, S2, RBD) observed among adolescents were further confirmed when we compared the ratios of spike versus nucleocapsid responses (Fig. 3a) and the number of positive IgG compared to IgA or IgM responses (Fig. 3b). When we restricted our analysis to the 16 parent–child pairs that were both tested seropositive, we observed similar results although most were no longer statistically significant (data not shown).

## Discussion

The COVICAT study is one of the largest studies examining the complex natural immunity to SARS-CoV-2 at a population level. Based on multiplex serology testing of 5000 participants, we detected a SARS-CoV-2 seroprevalence of 18.1% in adults and much lower, of 11.5%, in adolescents. Additionally, 11.6% of adults and 7.3% of adolescents showed marginal seroresponses (undetermined). Severity of infection, determined the magnitude, breadth and specificity (towards certain antigens and/or isotypes) of immune responses long time after the acute phase of infection. We also identified diverse associations between individuals' characteristics, including age, sex, smoking, and BMI status, with antibody responses.



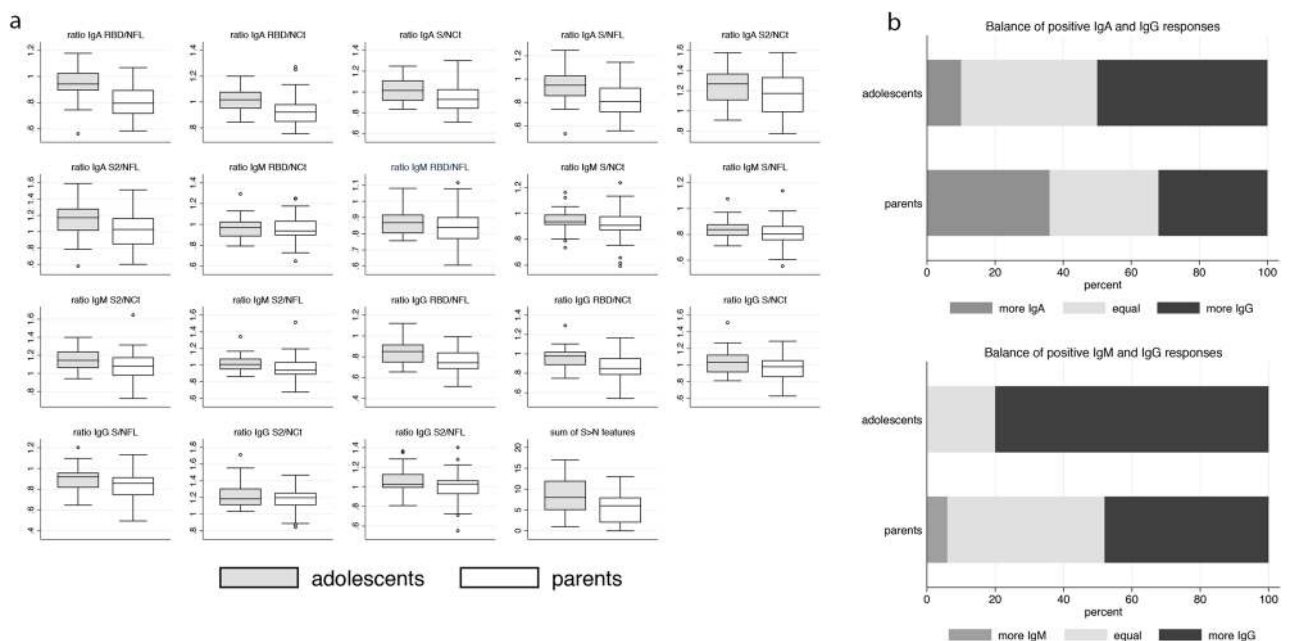


**Figure 2.** Differences in antibody responses according to the severity of infection among SARS-CoV-2 seropositive participants. Comparison of (a) antibody levels (log<sub>10</sub>-transformed values) of the fifteen isotype-antigen combinations, (b) breadth of positive immune responses ranging from 0 responses (light grey) to 13 responses (almost black), (c) ratio IgA/IgG antibody levels (log<sub>10</sub>-transformed values) for each of the antigens (the proportion of the participants showing ratio > 1 is listed in the top of each graph), (d) proportion of people showing more IgA, more IgG and equal number of positive IgA and IgG responses and (e) ratios of anti-spike (S, S2, RBD) over anti-nucleocapsid (NFL, NCI) responses for every combination and isotype. S for spike protein and N for nucleocapsid. Supplementary Resource 6 presents corresponding p-values.

Up to mid-November, there were 238,596 confirmed COVID-19 cases in Catalonia in people > 20 years old which corresponds to 3.9% of the population<sup>23</sup>. This proportion is much lower than our proportion of infected individuals based on serology (18.1%). We expected this difference, as surveillance systems are restricted by the emergency and load of testing clinically evident infections and high-risk individuals. Meanwhile, updated data from the fourth phase of a nationwide seroprevalence study in Spain (ENE-COVID) reported a seroprevalence of 9.9% in Spain and 11.6% in Catalonia until the end of November<sup>24</sup>. They used two tests, a point-of-care rapid test determining IgG against RBD and an immunoassay detecting IgG against nucleocapsid (not yet available

Characteristic	IgM NFL	IgM NCt	IgM RBD	IgM S	IgM S2	IgA NFL	IgA NCt	IgA RBD	IgA S	IgA S2	IgG NFL	IgG NCt	IgG RBD	IgG S	IgG S2	breadth
	≥60 years old	-0.02	-0.03	<b>-0.1</b>	<b>-0.08</b>	<b>-0.14</b>	<b>0.1</b>	0	<b>-0.07</b>	<b>-0.09</b>	<b>-0.15</b>	-0.04	<b>-0.11</b>	<b>-0.18</b>	<b>-0.17</b>	
Female sex	<b>0.15</b>	<b>0.12</b>	0.02	0.02	<b>0.07</b>	<b>-0.07</b>	<b>-0.06</b>	-0.03	-0.04	<b>-0.11</b>	0	-0.05	0.01	-0.01	-0.06	-0.18
Obese/overweight	-0.03	0.01	0.04	0.04	0	0.05	0.03	<b>0.11</b>	<b>0.08</b>	<b>0.11</b>	0.07	0.07	<b>0.15</b>	<b>0.14</b>	<b>0.11</b>	<b>0.71</b>
Smoker	0.03	0.02	-0.05	<b>-0.09</b>	<b>-0.09</b>	-0.07	-0.04	<b>-0.1</b>	<b>-0.16</b>	<b>-0.16</b>	<b>-0.22</b>	<b>-0.21</b>	<b>-0.35</b>	<b>-0.38</b>	<b>-0.3</b>	<b>-1.13</b>
<b>after adjustment for severity of infection</b>																
≥60 years old	-0.03	-0.03	-0.05	-0.03	<b>-0.1</b>	<b>0.11</b>	0	-0.02	-0.02	-0.06	0.05	-0.01	-0.02	-0.01	-0.03	0.07
Female sex	<b>0.16</b>	<b>0.12</b>	-0.01	-0.01	0.05	<b>-0.08</b>	<b>-0.06</b>	<b>-0.05</b>	<b>-0.07</b>	<b>-0.15</b>	-0.04	<b>-0.09</b>	-0.06	-0.08	<b>-0.12</b>	<b>-0.41</b>
Obese/overweight	-0.03	0.01	0.02	0.02	-0.02	0.05	0.03	<b>0.09</b>	0.05	0.07	0.03	0.04	0.09	0.08	0.06	<b>0.51</b>
Smoker	0.03	0.01	-0.01	-0.06	-0.06	-0.06	-0.04	-0.06	<b>-0.12</b>	-0.1	<b>-0.15</b>	<b>-0.14</b>	<b>-0.24</b>	<b>-0.27</b>	<b>-0.22</b>	<b>-0.78</b>

**Table 3.** Adjusted associations ( $\beta$ ) between each characteristic and antibody levels for each of the fifteen isotype-antigen combinations and the breadth of positive immune responses for SARS-CoV-2 seropositive participants, the COVICAT study. Associations ( $\beta$  coefficients) of each listed characteristic with  $\log_{10}$  transformed MFI values of each isotype-antigen combination adjusted for all the listed characteristics and days since infection. Bold indicates statistically significant associations (Supplementary Resource 7 presents corresponding p-values and 95% CI). The red color indicates a positive association while the blue color a negative association (associations with breadth are not comparable with those of antibody levels and thus not colored).



**Figure 3.** Differences in antibody responses between SARS-CoV-2 seropositive adolescents and parents. Comparison of antibody responses, (a) ratio of anti-spike (S, S2, RBD) over anti-nucleocapsid (NFL, NCt) responses for every combination and isotype and (b) proportion of people showing more IgA or IgM, more IgG and equal number of positive IgA or IgM and IgG responses.

for the fourth phase). Although the seroprevalence in ENE-COVID for Catalonia is lower than in our study it is not so different when compared with the seroprevalence for RBD and/or NCt IgG being 8.1% in our study. Thus, the difference in seroprevalence between the two studies could be partially attributable to the less extensive serological testing in ENE-COVID compared to our multiplex approach. Another argument for this scenario is that seroprevalence in adolescents, whose responses were primarily IgG anti-RBD, was not so different between the two studies (8.6% for 10–14 years of age in ENECOVID and 11.5% for 13–15 years old in our study). An increasing number of seroprevalence surveys now utilize multi-antigen and multi-isotype antibody responses because seroresponses might be skewed to different antigens and isotypes depending on clinical and individuals' characteristics<sup>16,25,26</sup>. Also, the importance of IgA isotype in diagnostic accuracy of SARS-CoV-2 serological tests is emerging<sup>7,17,27</sup>.

We found no differences in seroprevalence between females and males or with age among adults. Interestingly, seropositive participants older than 60 years of age had higher NFL IgA levels and women had lower antibody levels and number of seropositive responses (apart from NFL and NCT IgM). Age and sex-specific antibody responses against SARS-CoV-2 have been documented but results are mixed<sup>16,28,29</sup>. Similar to other studies, seroprevalence among young adolescents was lower than among adults<sup>30</sup>. Within one family, adolescents were at lower risk for seropositivity compared to their parents. We cannot make direct conclusions about children's role in transmitting SARS-CoV-2 within the household, but evidence argues for a reduced, marginal or conditional contribution<sup>30,31</sup>. It remains unclear why children are less susceptible to infection but mechanisms related to the number of ACE-2 receptors<sup>32</sup>, the naivety of innate immunity<sup>33</sup> and preexisting human coronaviruses-elicited immunity<sup>34</sup> are proposed.

A striking observation was that over 90% of previously tested positive participants had detectable antibodies up to 7 months after their first diagnosis, but most of them were either hospitalized or had experienced  $\geq 4$  symptoms. In the overall seropositive population of our study including infections of varying severity, we observed sustained levels for IgA and IgG responses at least 4 months after infection. More stable responses up to 9 months after infection were evident for NCT, RBD IgA and NCT, S2 IgG. We did not have repeated samples but a number of other studies did and showed limited loss of IgG antibodies and some loss of IgA antibodies over time<sup>9,16,18,35–37</sup>. More importantly, two recent studies showed that seropositive participants had a significantly decreased risk of re-infection up to 6 months after first infection<sup>38,39</sup>. It remains to be determined what levels of antibodies to what specific antigen epitopes protect people from recurrent infections.

Our findings are in agreement with previous reports showing that asymptomatics account for a significant proportion of the infected population (approximately 40%) and that a range of symptoms occurs with COVID-19 infection with the most specific being the loss of odor/taste and fever<sup>40,41</sup>. Importantly, the clinical spectrum of SARS-CoV-2 infection reflects a spectrum of antibody responses. With increasing severity of infection, we observed that hosts mounted more robust and rich responses. A recent study showed that the immune response of severely infected subjects was spread to subdominant viral antigens as well<sup>42</sup>. A novel finding in our study was that asymptomatics were more likely to have greater IgA than IgG responses compared to those experiencing more severe disease although the magnitude of IgA responses remained in lower levels in asymptomatics compared to those admitted to hospital/ICU. IgA could contribute to virus neutralization early in the infection to a greater extent compared with IgG<sup>8</sup>. Similar to us, most studies have described higher levels of antibodies among those with more severe disease, and some have suggested that a robust IgA response, in particular, may have a pathological role in SARS-CoV-2 infection<sup>9,16,17,22</sup>. Collectively with our data, it seems that IgA at low levels may be able to control the infection, but it could be associated with detrimental effects when boosted to higher levels along with other responses. In our study, we cannot disentangle to which extent the severity of infection drives these immune responses or whether these responses play a role in the pathogenesis of SARS-CoV-2.

Contradictory to two previous smaller studies, we found a shift towards spike over nucleocapsid responses with increasing severity of infection<sup>19,20</sup>. This discrepancy might be related to the fact that each study examined different immune features and we collected samples long after infection. Different rates of decay of anti-spike versus anti-nucleocapsid antibodies might have affected our results<sup>43,44</sup>. Although unclear, it is possible that the severity of infection determines the antibody production in the longterm (e.g. from long-lived plasma cells) in an antigen-specific manner<sup>45</sup>. For example, a recent study demonstrated that subjects with more severe disease mounted a larger memory B cell formation against the spike, but not the nucleocapsid<sup>42</sup>. Serological data from the group of adolescents point to very specific responses mainly of IgG against spike protein, consistent with previous studies<sup>21,22</sup>. We should note that children until late adolescence have a lower capacity of generating IgA<sup>46</sup>. These data suggest that adolescents use IgG alone to control the infection and that lack of anti-nucleocapsid responses might indicate a less widespread infection than adults. Such differences between adult and childhood immune responses should be delineated given the less harmful effects of SARS-CoV-2 infection in children.

Consistent with other studies, we detected higher levels of antibodies among overweight/obese participants. This association largely diluted when we adjusted for severity of infection, suggesting that higher levels were a consequence of the more severe disease experienced by obese people (adjustment by severity of infection was not considered in previous studies)<sup>16,47</sup>. In our population, the highest proportion of overweight/obese people was among those experiencing more severe infection. World Obesity Forum recently reported that COVID-19 mortality increased along with the countries' prevalence of obesity, even after adjusting for age and wealth<sup>48</sup>. These data suggest that despite displaying robust serological responses, overweight/obese infected people are more likely to develop severe infection than non-overweight/obese people. Reduced levels of SARS-CoV-2 antibodies and breadth of immune responses were detected among seropositive smokers compared to non-smokers irrespective of the severity of infection. Two studies report similar results concerning levels of antibodies<sup>16,49</sup>. These data suggest that smokers present a weakened immune response to SARS-CoV-2. Thus, the higher COVID-19 morbidity among smokers might be due to impaired immunity as reflected in lower antibody levels<sup>50</sup>. Simultaneously, the paradox of low prevalence of smoking among SARS-CoV-2 infected people might be due to low/non-detectable levels of antibodies. Of course, this would not affect results from PCR tests but it has been suggested, although not investigated, that smoking might decrease nasopharyngeal viral load resulting more often in false-negative results<sup>51</sup>. We need more studies in this perspective.

We acknowledge that this is not a random population-based study as it recruited participants from pre-existing cohorts. On the other hand, existing cohorts allow us to quickly contact a population, achieve a high participation rate and access pre-pandemic information. The study primarily consisted of people 40–70 years old ( $\approx 90\%$ ), owing to the age distribution in original cohorts but we combined several cohorts in order to include younger and older people. Volunteer bias is always of concern, as we observed a lower participation rate for those who had received a test; this would have led to underestimation of true seroprevalence. We assumed that all persons infected with SARS-CoV-2 have detectable antibodies at the time of sampling. The 8% of previously



tested positive participants with a negative serology in our study might indicate the proportion of the population who lost immunity, were antibody non-responders or had false-positive results in the first test. Moreover, we could not verify whether the symptoms reported were attributable to a SARS-CoV-2 infection given the limited access of the population to diagnostic tests at the initial months of the pandemic.

## Conclusion

Collectively, the data presented here argue for a higher number of exposed individuals to SARS-CoV-2 in Catalonia, than what has been described, but still the majority of the population remains unexposed. Although further analysis with repeated samples will allow us to describe the progression of antibody levels in time, we observed that even 4–9 months after infection, responses against SARS-CoV-2 were evident. Individuals presented strikingly heterogeneous immune responses depending on the severity of infection. Factors such as obesity and smoking that are related to significant COVID-19 morbidity and mortality probably determine antibody responses.

## Methods

**Study design and setting.** The COVICAT study includes participants from different pre-existing ongoing population-based cohorts in Catalonia and was developed following the COVID-19 pandemic. Eligible participants were from three adult population-based cohorts (GCAT, Genomes for life<sup>52</sup>; MCC-Spain, population controls from Catalunya<sup>53</sup>; and ECRHS, European Community Respiratory Health Survey<sup>54</sup>), two mother–child cohorts [INMA-Sabadell, (Infancia y Medio Ambiente)<sup>55</sup> children born in 2005–2007 and their mothers; BiSC, Barcelona Life Study Cohort pregnant women recruited immediately before and during the pandemic] and two small general population cohorts in special populations (Urban Training, older persons; and LeRAGs, rural population). The eligible adult population consists of 19,424 people and we were able to contact 18,737 (96%) using email and telephone messages or calls. Of them, 10,837 (58%) participated in the study. Just after the strict first confinement period, we invited them to complete an online questionnaire about COVID-19 compatible symptoms, diagnostics as well as about occupational and financial aspects, sociodemographic and lifestyle characteristics, mental health and chronic diseases. We also did telephone interviews for those unfamiliar with the use of online approaches. Participants were then asked to donate a blood sample in different facilities. We offered the option for older people and those living in remote areas of collecting a blood fingerprick sample at their residence, as well as a second opportunity for donating a blood sample during September–November 2020 for all participants not able to provide a sample earlier. Participants of the second sampling period completed an additional questionnaire related to COVID-19 symptoms and testing in order to update the relative information from the main questionnaire. We also had available samples and questionnaire information on 260 adolescent participants of the INMA-Sabadell, cohort. Details on the sources and methods of assessment of variables are presented in the Supplementary Resource 9.

All participants gave written informed consent before participation in the study. For individuals younger than 18 years, parents or a legal representative provided consent. The study was approved by the Parc de Salut Mar Drug Research Ethical Committee (IBR number: 2020/9307/I). All methods were performed in accordance with the relevant guidelines and regulations.

**Serology.** Blood samples were processed within 24 h of collection and were analyzed at the ISGlobal Immunology laboratory in Barcelona. The levels [median fluorescence intensity (MFI)] of IgG, IgM and IgA were assessed by high-throughput multiplex quantitative suspension array technology, including, as SARS-CoV-2 antigens, the S (aa 1–1213 expressed in Expi293 and His tag-purified) and the S2 fragment (purchased from SinoBiologicals), the RBD (donated by the Krammer lab, Mount Sinai, NY), the NFL and the specific NCT (expressed in *E. coli* and His tag-purified). Assay performance was previously established as 100% specificity and 95.78% sensitivity for seropositivity 14 days after symptoms onset<sup>56</sup>. Antigen-coupled microspheres were added to a 384-well  $\mu$ Clear<sup>®</sup> flat bottom plate (Greiner Bio-One, Frickenhausen, Germany) in multiplex (2000 microspheres per analyte per well) in a volume of 90  $\mu$ L of Luminex Buffer (1% BSA, 0.05% Tween 20, 0.05% sodium azide in PBS) using 384 channels Integra Viaflo semi-automatic device (96/384, 384 channel pipette). Hyperimmune pools were used as positive controls prepared at twofold, 8 serial dilutions from 1:12.5. Pre-pandemic samples were used as negative controls to estimate the cut-off of seropositivity. Ten microliter of each dilution of the positive control, negative controls and test samples (prediluted 1:50 in 96 round-bottom well plates), were added to a 384-well plate using Assist Plus Integra device with 12 channels Voyager pipette (final test sample dilution of 1:500). To quantify IgM, test samples and controls were pre-treated with anti-Human IgG (Gullorsorb) at 1:10 dilution, to avoid IgG interferences. Technical blanks consisting of Luminex Buffer and microspheres without samples were added in 4 wells to control for non-specific signals. Plates were incubated for 1 h at room temperature in agitation (Titramax 1000) at 900 rpm and protected from light. Then, the plates were washed three times with 200  $\mu$ L/well of PBS-T (0.05% Tween 20 in PBS), using BioTek 405 TS (384-well format). Twenty five microliter of goat anti-human IgG-phycoerythrin (PE) (GTIG-001, Moss Bio) diluted 1:400, goat anti-human IgA-PE (GTIA-001, Moss Bio) 1:200, or goat anti-human IgM-PE (GTIM-001, Moss Bio) 1:200 in Luminex buffer were added to each well and incubated for 30 min. Plates were washed and microspheres resuspended with 80  $\mu$ L of Luminex Buffer, covered with an adhesive film and sonicated 20 s on sonicator bath platform, before acquisition on the Flexmap 3D reader. At least 50 microspheres per analyte per well were acquired, and MFI was reported for each analyte. Assay positivity cut-offs specific for each isotype and analyte were calculated as 10 to the mean plus 3 standard deviations of  $\log_{10}$ -transformed MFI of 128 pre-pandemic controls. Results were defined as undetermined when the MFI levels for a given isotype-analyte were between the positivity threshold and an upper limit at 10 to the mean plus 4.5 standard deviations of the  $\log_{10}$ -transformed MFIs of pre-pandemic samples, and no other isotype-antigen combination was above the positivity cut-off.

**Statistics.** Descriptive analyses of the study population characteristics were conducted. We used raked weights to extrapolate seroprevalence to the total population of Catalonia aged more than 20 years. Briefly, raking calculates weights so that the weighted sample has the same marginal distribution than the reference population in terms of the variables used to calculate the weights<sup>57</sup>. In particular, we used the joint sex-age (in 10-year groups) distribution, educational level, health region and smoking. Population data were obtained from the National Statistics Institute and from the Catalan Health Survey<sup>58</sup>. The COVICAT study includes lower numbers of younger ages, which leads to overdispersed weights and gaps in the distribution of the weights. For this reason, we restricted the extremes weights by trimming the distribution at 99% of the weights. Generalized additive models were used to explore the shape of the relationship between days since infection and antibody levels to each of the fifteen isotype-antigen combination. Antibody levels were  $\log_{10}$ -transformed to normalise their distribution. Differences in antibody levels and ratios by severity of infection were examined using oneway Anova and pairwise comparisons were performed using Tukey post hoc-test. Differences in immune responses between adolescents and parents were examined with t-tests. Multivariable regression models were applied to examine among seropositive individuals the association between age, sex, smoking and BMI status before confinement (all in the same model adjusted also for days since infection) and  $\log_{10}$ -transformed antibody levels and the breadth of positive responses. All analyses were conducted using Stata version 16 (StataCorp LP, College Station, Texas).

Received: 24 May 2021; Accepted: 12 October 2021

Published online: 03 November 2021

## References

- Alwan, N. A. Surveillance is underestimating the burden of the COVID-19 pandemic. *The Lancet* **396**, e24 (2020).
- Yanes-Lane, M. *et al.* Proportion of asymptomatic infection among COVID-19 positive persons and their transmission potential: A systematic review and meta-analysis. *PLoS ONE* **15**, e0241536 (2020).
- Jones, J. M. *et al.* Estimated US infection- and vaccine-induced SARS-CoV-2 seroprevalence based on blood donations, July 2020–May 2021. *JAMA* **326**, 1400 (2021).
- Chen, X. *et al.* Serological evidence of human infection with SARS-CoV-2: A systematic review and meta-analysis. *Lancet Glob. Health* **9**, e598–e609 (2021).
- Lai, C.-C., Wang, J.-H. & Hsueh, P.-R. Population-based seroprevalence surveys of anti-SARS-CoV-2 antibody: An up-to-date review. *Int. J. Infect. Dis.* **101**, 314–322 (2020).
- Dai, L. & Gao, G. F. Viral targets for vaccines against COVID-19. *Nat. Rev. Immunol.* **21**, 73–82 (2021).
- Vogl, T., Leviatan, S. & Segal, E. SARS-CoV-2 antibody testing for estimating COVID-19 prevalence in the population. *Cell Rep. Med.* **2**, 100191 (2021).
- Sterlin, D. *et al.* IgA dominates the early neutralizing antibody response to SARS-CoV-2. *Sci. Transl. Med.* **13**, 577 (2021).
- Yu, H. *et al.* Distinct features of SARS-CoV-2-specific IgA response in COVID-19 patients. *Eur. Respir. J.* **56**, 2001526 (2020).
- Rosado, J. *et al.* Multiplex assays for the identification of serological signatures of SARS-CoV-2 infection: An antibody-based diagnostic and machine learning study. *Lancet Microbe* **2**, e60 (2020).
- Alter, G. & Seder, R. The power of antibody-based surveillance. *N. Engl. J. Med.* **383**, 1782–1784 (2020).
- van Tol, S. *et al.* Accurate serology for SARS-CoV-2 and common human coronaviruses using a multiplex approach. *Emerg. Microbes Infect.* **9**, 1965–1973 (2020).
- Long, Q.-X. *et al.* Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* **26**, 1200–1204 (2020).
- Cook, N. *et al.* Multiplex assessment of SARS-CoV-2 antibodies improves assay sensitivity and correlation with neutralizing antibodies. *Clin. Biochem.* **97**, 54 (2021).
- Post, N. *et al.* Antibody response to SARS-CoV-2 infection in humans: A systematic review. *PLoS ONE* **15**, e0244126 (2020).
- Gudbjartsson, D. F. *et al.* Humoral immune response to SARS-CoV-2 in Iceland. *N. Engl. J. Med.* **383**, 1724–1734 (2020).
- Grossberg, A. N. *et al.* A multiplex chemiluminescent immunoassay for serological profiling of COVID-19-positive symptomatic and asymptomatic patients. *Nat. Commun.* **12**, 740 (2021).
- Dan, J. M. *et al.* Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* **371**, 4063 (2021).
- Röltgen, K. *et al.* Defining the features and duration of antibody responses to SARS-CoV-2 infection associated with disease severity and outcome. *Sci. Immunol.* **5**, 0240 (2020).
- Atyeo, C. *et al.* Distinct early serological signatures track with SARS-CoV-2 survival. *Immunity* **53**, 524–532.e4 (2020).
- Weisberg, S. P. *et al.* Distinct antibody responses to SARS-CoV-2 in children and adults across the COVID-19 clinical spectrum. *Nat. Immunol.* **22**, 25–31 (2021).
- Bartsch, Y. C. *et al.* Humoral signatures of protective and pathological SARS-CoV-2 infection in children. *Nat. Med.* **27**, 454 (2021).
- <https://aguas.gencat.cat/ca/actualitat/ultimes-dades-coronavirus>. Accessed 1 June 2021.
- Ministerio de Sanidad, Consumo y Bienestar Social. Estudio ENE-COVID: Cuarta ronda estudio nacional de sero-epidemiología de la infección por SARS-CoV-2 en España. <https://www.msbs.gob.es>. Accessed 15 Dec 2020.
- García-Basteiro, A. L. *et al.* Seroprevalence of antibodies against SARS-CoV-2 among health care workers in a large Spanish reference hospital. *Nat. Commun.* **11**, 3500 (2020).
- Knabl, L. *et al.* High SARS-CoV-2 seroprevalence in children and adults in the austrian ski resort Ischgl. *Epidemiology*. <https://doi.org/10.1101/2020.08.20.20178533> (2020).
- Burgess, S., Ponsford, M. J. & Gill, D. Are we underestimating seroprevalence of SARS-CoV-2? *BMJ*. <https://doi.org/10.1136/bmj.m3364> (2020).
- Conti, P. & Younes, A. Coronavirus COV-19/SARS-CoV-2 affects women less than men: Clinical response to viral infection. *J. Biol. Regul. Homeost. Agents* **34**, 339–343 (2020).
- Dobaño, C. *et al.* Immunogenicity and crossreactivity of antibodies to the nucleocapsid protein of SARS-CoV-2: Utility and limitations in seroprevalence and immunity studies. *Transl. Res. J. Lab. Clin. Med.* **232**, 60 (2021).
- Viner, R. M. *et al.* Susceptibility to SARS-CoV-2 infection among children and adolescents compared with adults: A systematic review and meta-analysis. *JAMA Pediatr.* **175**, 143 (2021).
- Li, F. *et al.* Household transmission of SARS-CoV-2 and risk factors for susceptibility and infectivity in Wuhan: A retrospective observational study. *Lancet Infect. Dis.* **21**, 617 (2021).

32. Bunyavanich, S., Do, A. & Vicencio, A. Nasal gene expression of angiotensin-converting enzyme 2 in children and adults. *JAMA* **323**, 2427 (2020).
33. Carsetti, R. *et al.* The immune system of children: The key to understanding SARS-CoV-2 susceptibility? *Lancet Child. Adolesc. Health* **4**, 414–416 (2020).
34. Ng, K. W. *et al.* Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. *Science* **370**, 1339–1343 (2020).
35. Gaebler, C. *et al.* Evolution of antibody immunity to SARS-CoV-2. *Nature* **591**, 639 (2021).
36. Moncunill, G. *et al.* SARS-CoV-2 seroprevalence and antibody kinetics among health care workers in a Spanish hospital after 3 months of follow-up. *J. Infect. Dis.* **223**, 62–71 (2021).
37. Ortega, N. *et al.* Seven-month kinetics of SARS-CoV-2 antibodies and role of pre-existing antibodies to human coronaviruses. *Nat. Commun.* **12**, 4740 (2021).
38. Harvey, R. A. *et al.* Association of SARS-CoV-2 seropositive antibody test with risk of future infection. *JAMA Intern. Med.* **181**, 672 (2021).
39. Lumley, S. F. *et al.* Antibody status and incidence of SARS-CoV-2 infection in health care workers. *N. Engl. J. Med.* **384**, 533–540 (2021).
40. Oran, D. P. & Topol, E. J. The proportion of SARS-CoV-2 infections that are asymptomatic: A systematic review. *Ann. Intern. Med.* **175**, 655 (2021).
41. Struyf, T. *et al.* Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. *Cochrane Database Syst. Rev.* <https://doi.org/10.1002/14651858.CD013665.pub2> (2021).
42. Guthmiller, J. J. *et al.* SARS-CoV-2 Infection severity is linked to superior humoral immunity against the spike. *MBio* **12**, e02940 (2021).
43. Fenwick, C. *et al.* Changes in SARS-CoV-2 spike versus nucleoprotein antibody responses impact the estimates of infections in population-based seroprevalence studies. *J. Virol.* **95**, e01828 (2020).
44. Manisty, C. *et al.* Time series analysis and mechanistic modelling of heterogeneity and sero-reversion in antibody responses to mild SARS-CoV-2 infection. *EBioMedicine* **65**, 103259 (2021).
45. Akkaya, M., Kwak, K. & Pierce, S. K. B cell memory: Building two walls of protection against pathogens. *Nat. Rev. Immunol.* **20**, 229–238 (2020).
46. Buckley, R. H., Dees, S. C. & O'Fallon, W. M. Serum immunoglobulins. I. Levels in normal children and in uncomplicated childhood allergy. *Pediatrics* **41**, 600–611 (1968).
47. Racine-Brzostek, S. E. *et al.* Postconvalescent SARS-CoV-2 IgG and neutralizing antibodies are elevated in individuals with poor metabolic health. *J. Clin. Endocrinol. Metab.* **106**, e2025 (2021).
48. World Obesity Foundation. *COVID-19 and Obesity: The 2021 Atlas* (World Obesity Foundation, 2021).
49. Schaffner, A. *et al.* Characterization of a pan-immunoglobulin assay quantifying antibodies directed against the receptor binding domain of the SARS-CoV-2 S1-subunit of the spike protein: A population-based study. *J. Clin. Med.* **9**, 3989 (2020).
50. Alqahtani, J. S. *et al.* Prevalence, severity and mortality associated with COPD and smoking in patients with COVID-19: A rapid systematic review and meta-analysis. *PLoS ONE* **15**, e0233147 (2020).
51. de Lusignan, S. *et al.* Risk factors for SARS-CoV-2 among patients in the Oxford Royal College of General Practitioners Research and Surveillance Centre primary care network: A cross-sectional study. *Lancet Infect. Dis.* **20**, 1034–1042 (2020).
52. Obón-Santacana, M. *et al.* GCAT|Genomes for life: A prospective cohort study of the genomes of Catalonia. *BMJ Open* **8**, e018324 (2018).
53. Castaño-Vinyals, G. *et al.* Population-based multicase-control study in common tumors in Spain (MCC-Spain): Rationale and study design. *Gac. Sanit.* **29**, 308–315 (2015).
54. Burney, P. G., Luczynska, C., Chinn, S. & Jarvis, D. The European community respiratory health survey. *Eur. Respir. J.* **7**, 954–960 (1994).
55. Guxens, M. *et al.* Cohort profile: The INMA—Infancia y Medio Ambiente—(environment and childhood) project. *Int. J. Epidemiol.* **41**, 930–940 (2012).
56. Dobaño, C. *et al.* Highly sensitive and specific multiplex antibody assays to quantify immunoglobulins M, A, and G against SARS-CoV-2 antigens. *J. Clin. Microbiol.* **59**, e01731 (2020).
57. Dal Grande, E., Chittleborough, C. R., Camprostrini, S., Tucker, G. & Taylor, A. W. Health estimates using survey raked-weighting techniques in an Australian population health surveillance system. *Am. J. Epidemiol.* **182**, 544–556 (2015).
58. Garcia-Codina, O. *et al.* Determinants of health literacy in the general population: Results of the Catalan health survey. *BMC Public Health* **19**, 1122 (2019).

## Acknowledgements

We thank the volunteers who participate in the cohort studies, all the workers in different facilities of the Blood and Tissue Bank for sample recruitment, Jordi Chi for antigen expression and purification and Victor Moreno for data collection in MCC study. This study makes use of data generated by the GCAT-Genomes for Life. Cohort study of the Genomes of Catalonia, Fundacio IGTP. IGTP is part of the CERCA Program/Generalitat de Catalunya. GCAT is supported by Acció de Dinamització del ISCIII-MINECO and the Ministry of Health of the Generalitat of Catalunya (ADE 10/00026); the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) (2017-SGR 529). Full list of the investigators who contributed to the generation of the GCAT data is available from [www.genomesforlife.com](http://www.genomesforlife.com). LeRAGs is supported by Instituto de Salud Carlos III (PI17/01555). Beatriz Cortés is supported by ISCIII national Grant PI18/01512. Ioar Rivas is supported from the postdoctoral fellowships programme Beatriu de Pinós (2018 BP 00114), funded by the Secretary of Universities and Research (Government of Catalonia) and by the Horizon 2020 programme of research and innovation of the European Union under the Marie Skłodowska-Curie Grant Agreement No 801370.

## Author contributions

Conceptualization: M.Kar., G.M., C.D. and M.Kog; Methodology: M.Kar., G.M., C.D., D.B., L.P., R.A., L.M., R.R., L.I., L.M., R.R., L.I. and V.P.; Formal analysis: M.Kar., A.E. and X.B.; Investigation: A.J., M.V., R.S., D.B., L.P., R.A., L.M., R.R., L.I., V.P., S.F. and I.R.; Data curation: A.J., M.V., R.S., D.B., L.P., R.A., A.C., B.C., A.E. and G.C.V.; Writing-original draft preparation: M.Kar., G.M., C.D. and M.Kog.; Writing-review and editing: M.Kar., G.M., A.E., G.C.V., A.J., M.V., R.S., D.B., L.P., A.C., L.M., R.R., B.C., V.P., C.O.G., S.F., I.R., D.C., M.V., L.I., R.A., X.B., J.G.A., R.d.C., C.D. and M.Kog.; Visualization: M.Kar. and A.E.; Funding acquisition: G.M., C.D. and M.Kog.; Project administration: G.C.V.; Supervision: G.M., C.D. and M.Kog.

## Funding

This work was funded by Incentius a l'Avaluació de Centres CERCA (in\_CERCA); EIT HEALTH BP2020-20873-Certify.Health.; Fundació Privada Daniel Bravo Andreu; PID2019-110810RB-I00 Grant (Spanish Ministry of Science & Innovation). ISGlobal acknowledges support from the Spanish Ministry of Science and Innovation through the “Centro de Excelencia Severo Ochoa 2019–2023” Program (CEX2018-000806-S). ISGlobal and IGTP receive support from the Generalitat de Catalunya through the CERCA Program. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-00807-4>.

**Correspondence** and requests for materials should be addressed to M.K.

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

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021