

Features of acute COVID-19 associated with post-acute sequelae of SARS-CoV-2 phenotypes: results from the IMPACC study

Received: 17 March 2023

Accepted: 29 November 2023

Published online: 03 January 2024

 Check for updates

A list of authors and their affiliations appears at the end of the paper

Post-acute sequelae of SARS-CoV-2 (PASC) is a significant public health concern. We describe Patient Reported Outcomes (PROs) on 590 participants prospectively assessed from hospital admission for COVID-19 through one year after discharge. Modeling identified 4 PRO clusters based on reported deficits (minimal, physical, mental/cognitive, and multidomain), supporting heterogenous clinical presentations in PASC, with sub-phenotypes associated with female sex and distinctive comorbidities. During the acute phase of disease, a higher respiratory SARS-CoV-2 viral burden and lower Receptor Binding Domain and Spike antibody titers were associated with both the physical predominant and the multidomain deficit clusters. A lower frequency of circulating B lymphocytes by mass cytometry (CyTOF) was observed in the multidomain deficit cluster. Circulating fibroblast growth factor 21 (FGF21) was significantly elevated in the mental/cognitive predominant and the multidomain clusters. Future efforts to link PASC to acute anti-viral host responses may help to better target treatment and prevention of PASC.

Post-acute sequelae of SARS-CoV-2 (PASC), also known as long coronavirus disease 2019 (long COVID-19), represents a growing concern in public healthcare. While consensus for the clinical definition of PASC is evolving, the National Institute for Health and Care Excellence (NICE) defines PASC as signs and symptoms that develop during or following an infection consistent with COVID-19, continue for more than four weeks and are not explained by an alternative diagnosis^{1,2}. Several prospective cohorts of PASC have been described, each differing in case definition, size, and composition of the study population, symptoms evaluated, as well as follow-up frequency and duration. The prevalence of PASC is especially high amongst those COVID-19 patients who needed hospitalization: according to some reports³, up to half of hospitalized patients reported at least one physical, cognitive, or mental impairment, months after COVID-19 diagnosis. PASC is, thus, having a substantial impact on quality of life, healthcare costs, and economic productivity⁴.

The immunobiology of PASC is currently under intensive investigation with some leading hypotheses⁵ invoking the persistence of viral components driving immune stimulation, reactivation of viral

infections such as EBV, dysbiosis of microbiome or virome, unrepaired tissue damage, and autoimmunity^{6,7}.

Here, we describe patient-reported recovery data prospectively assessed from the acute infection through one year after hospital discharge for COVID-19. We identify predisposing factors and immune profiles from the acute phase of the disease that are associated with impaired clinical and functional recovery during the year following hospital discharge.

Results

Demographics and descriptive statistics

In all, 1164 participants were enrolled between May 5th, 2020 and March 19th, 2021 and followed up to 28 days while hospitalized. Of the 702 participants who survived hospitalization and were alive and on study at 3 months post discharge, 590 (84%) completed at least one quarterly set of surveys post discharge (survey respondent cohort) (Fig. 1A) with 29% (170) completing all 4 quarterly surveys and most completing 2 or more surveys (494; 84%)⁷. The participants who responded appear to be demographically representative

✉ e-mail: nroupha@emory.edu

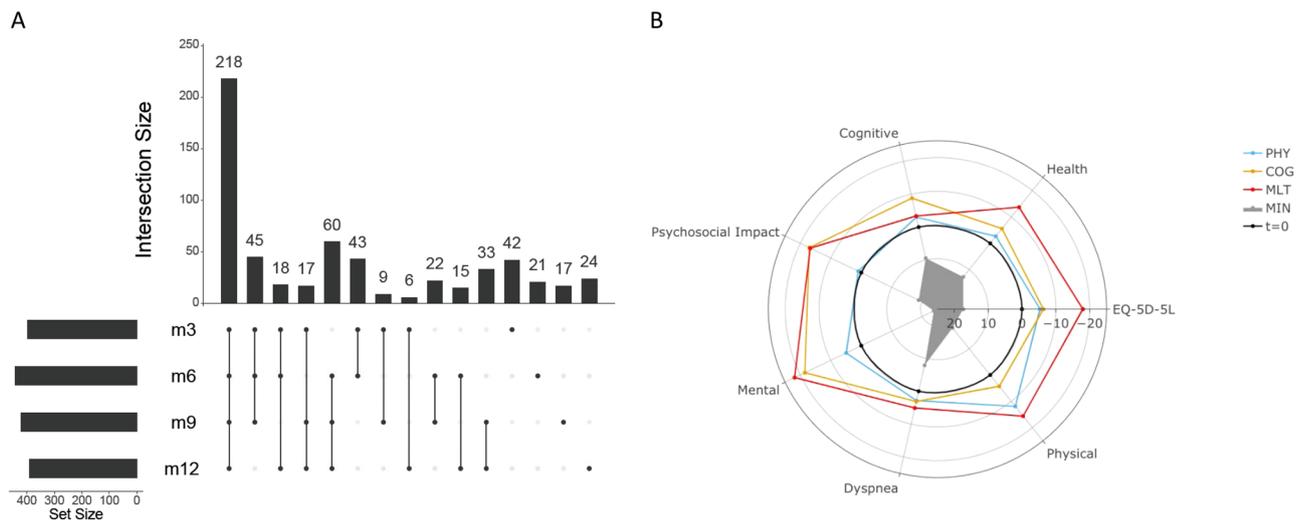


Fig. 1 | Survey completion and clustering of participant-reported outcomes after hospital discharge ($N = 590$). **A** Upset plot depicting the number of participants completing surveys at 3 (m3), 6 (m6), 9 (m9), and 12 months (m12) after hospital discharge. **B** Radar plot showing relative deficit for each of four different clusters across several participant-reported outcomes: EQ-5D-5L Health Recovery Score (Health), PROMIS Cognitive Function Score (Cognitive), PROMIS Psychosocial Illness Impact Positive Score (Psychosocial), PROMIS Global Mental Health

Score (Mental), PROMIS Dyspnea Score (Dyspnea), and PROMIS Physical Function Score (Physical). The radial axis denotes a t -statistic comparing the within-cluster mean to the remaining sample, with $t = 0$ denoting the overall sample mean and negative values denoting a deficit. The 4 clusters are: solid gray, minimal deficit (MIN); blue line, physical predominant deficit (PHY); yellow line, mental/cognitive predominant deficit (COG); and red line, multidomain deficit (MLT). PROMIS Patient-Reported Outcomes Measurement Information System.

of the entire cohort but were less likely to have prolonged hospitalization or discharge limitations compared with non-responders (42% versus 57%)⁸. Demographics, clinical characteristics, baseline radiographic and laboratory findings, as well as main outcomes during the COVID-19 hospital stay are provided for the survey respondent cohort in Table 1. The median age was 57 years (IQR 19), and 360 (61%) were men. 131 (22%) were Black/African American, and 189 (32%) were Hispanic/Latinx. 94% of participants had at least one comorbidity, most commonly hypertension (327; 55%) and diabetes (200; 34%). The median body mass index (BMI) was 31.8 kg/m² (IQR 27.4–37.0). Two hundred and forty-five (52%) had an elevated baseline C-reactive protein (CRP) (≥ 10 mg/L) and 297 (50%) had an abnormal baseline D-dimer (> 0.5 mg/L) upon hospital admission. Only 137 (25%) had no infiltrate on chest imaging upon hospital admission. One hundred and forty-one (24%) did not receive any oxygen therapy and 159 (27%) received ICU level care while inpatient. Four hundred (68%) received steroids and 377 (64%) received remdesivir. The median length of stay was 6 days (IQR 4–10) and most study participants had at least one complication (490; 83%) while inpatient. None of the participants had received a COVID-19 vaccine prior to admission; after discharge, 62% reported receiving the primary vaccine series and 36% a single booster dose. Three hundred and five (52%) out of 590 reported at least one symptom during the quarterly surveys, most commonly dyspnea (29%), followed by muscle aches/myalgia (21%), cough (20%), headache (19%), and fatigue/malaise (18%) (Table 1) (Figs. 1S and 2S). Thirty-two percent had symptoms affecting more than one organ system.

Latent class and cluster analysis

Fitting latent class mixed models (LCMMs) derived from the longitudinal patterns to each of eight participant-reported outcomes (PROs), we selected quadratic models with three groupings for EQ-5D-5L, Health Recovery Score, and PROMIS Dyspnea Score, and a linear model with three groupings for PROMIS Cognitive Function Score. There were no distinct groupings for the PROMIS Physical Function Score, PROMIS Global Mental Health Score, and PROMIS Psychosocial

Illness Impact Positive Score. Median (IQR) values by visit for each PRO grouping are shown in Fig. 3SA–G.

We identified the Ward algorithm with six clusters as the optimal model. Fitting statistics across the five algorithms are shown in Fig. 3SH. Comparison of t -statistics across the PROs identified no associated deficit with certain clusters, which we then collapsed into a single cluster which we labeled minimal deficit (cluster MIN, 358 participants, 60.7%, Table 1A–S). Based on associations with specific PROs, clinical phenotypes were defined by labeling the remaining three clusters as physical predominant (cluster PHY, 92 participants, 15.6%), mental/cognitive predominant (cluster COG, 82 participants, 13.9%), and multi/pan domain deficit (cluster MLT, 58 participants, 9.8%). Table 1B–S and Fig. 1B, respectively, show a table and radar plot with the t -statistics for each PRO across these four clusters, which were named based on the predominant deficit for a specific PRO.

Selected demographic characteristics, key comorbidities, and laboratory findings were significantly associated with the four PRO clusters by bivariate analysis (Table 1). Acute phase disease severity, whether defined by respiratory score at admission, SOFA score at admission, ICU utilization, mechanical ventilation, or inpatient overall clinical trajectory, was not associated with PRO cluster assignment (Table 1). Any use of remdesivir and steroids in the inpatient period was not associated with a decrease in PASC prevalence (Table 1). Adjusted multinomial logistic regression analyses comparing to participants with the minimal deficit cluster indicated that participants in the PHY cluster were more likely to have reported comorbidities of chronic pulmonary disease (OR 2.46 95% CI 1.41–4.29) or chronic neurologic disorder (OR 2.13 95% CI 1.20–3.78) and less likely to be males (OR 0.55; 95% CI 0.35–0.87) and non-white race (OR 0.66; 95% CI 0.47–0.93). Relative to participants in the MIN cluster, participants with mental/cognitive predominant deficit were less likely to be 65 years or older (OR 0.41; 95% CI 0.18–0.94), less likely to be males (OR 0.54; 95% CI 0.36–0.82), more likely to have chronic cardiac disease (OR 1.72; 95% CI 1.02–2.88) and had longer acute hospitalization (OR per week 1.28; 95% CI 1.12–1.46). Participants with MLT deficit were more likely to have chronic pulmonary disease (OR 1.78; 95% CI 1.01–3.13) or chronic neurologic disorder (OR 4.37; 95% CI 2.14–8.94),

Table 1 | Baseline characteristics of the survey respondent cohort

	Overall (n = 590)	Cluster MIN (n = 358, 60.7%)	Cluster PHY (n = 92, 15.6%)	Cluster COG (n = 82, 13.9%)	Cluster MLT (n = 58, 9.8%)	Over- all P value
Demographic characteristics						
Age at enrollment (years), median (IQR)	57.0 (47.0–66.0)	57.0 (45.0–66.0)	61.0 (50.5–69.0)	54.5 (46.0–62.0)	60.0 (52.0–71.0)	0.006
Sex at birth, no. (%)	360 (61)	238 (66)	49 (53)	42 (51)	31 (53)	0.009
Female	230 (39)	120 (34)	43 (47)	40 (49)	27 (47)	
White	289 (49)	175 (49)	55 (60)	27 (33)	32 (55)	0.011
Black	131 (22)	82 (23)	16 (17)	20 (24)	13 (22)	
Non-Hispanic	384 (65)	233 (65)	57 (62)	49 (60)	45 (78)	0.313
Hispanic	189 (32)	113 (32)	32 (35)	32 (39)	12 (21)	
Comorbidities, no. (%)						
Hypertension	327 (55)	183 (51)	56 (61)	50 (61)	38 (66)	0.066
Diabetes	200 (34)	108 (30)	37 (40)	33 (40)	22 (38)	0.124
Chronic respiratory (not asthma)	104 (18)	48 (13)	27 (29)	14 (17)	15 (26)	0.001
Asthma	101 (17)	56 (16)	20 (22)	16 (20)	9 (16)	0.499
Chronic cardiac disease	142 (24)	70 (20)	29 (32)	22 (27)	21 (36)	0.008
Chronic kidney disease	69 (12)	44 (12)	9 (10)	12 (15)	4 (7)	0.488
Malignant neoplasm	49 (8)	32 (9)	9 (10)	5 (6)	3 (5)	0.631
Chronic neurological disorder	61 (10)	22 (6)	14 (15)	9 (11)	16 (28)	<0.001
Liver disease	28 (5)	15 (4)	6 (7)	4 (5)	3 (5)	0.822
History of SOT or BMT	31 (5)	16 (4)	4 (4)	7 (9)	4 (7)	0.443
Current or former smoking and/or vaping	182 (31)	101 (28)	33 (36)	24 (29)	24 (41)	0.147
Substance use (drugs, alcohol, and/or cannabis)	40 (7)	22 (6)	5 (5)	5 (6)	8 (14)	0.167
BMI category in Kg/m ² , no. (%) ^b						
Normal weight (25.1–29.9)	65 (11)	41 (11)	5 (5)	12 (15)	7 (12)	0.021
Overweight (30–39.9)	162 (27)	108 (30)	19 (21)	20 (24)	15 (26)	
Class 1–2 Obesity (40–49.9)	256 (43)	151 (42)	43 (47)	36 (44)	26 (45)	
Class 3 Obesity (≥50)	90 (15)	43 (12)	25 (27)	13 (16)	9 (16)	

Table 1 (continued) | Baseline characteristics of the survey respondent cohort

	Overall (n = 590)	Cluster MIN (n = 358, 60.7%)	Cluster PHY (n = 92, 15.6%)	Cluster COG (n = 82, 13.9%)	Cluster MLT (n = 58, 9.8%)	Over-all P value
Number of comorbidities, median (IQR)	(n = 590) 3.0 (2.0–4.0)	3.0 (1.0–4.0)	4.0 (2.5–5.0)	3.0 (2.0–5.0)	3.0 (2.0–6.0)	<0.001
None	35 (6)	30 (8)	0 (0)	3 (4)	2 (3)	0.006
1	92 (16)	68 (19)	9 (10)	11 (13)	4 (7)	
2	116 (20)	75 (21)	14 (15)	13 (16)	14 (24)	
3	118 (20)	65 (18)	21 (23)	22 (27)	10 (17)	
4	83 (14)	45 (13)	17 (18)	12 (15)	9 (16)	
5 or more	146 (25)	75 (21)	31 (34)	21 (26)	19 (33)	
Baseline visit						
Infiltrates on chest X-ray or chest tomography						0.645
No infiltrates	137 (25)	80 (24)	27 (30)	17 (23)	13 (23)	
Unilateral infiltrates	50 (9)	32 (9)	8 (9)	3 (4)	7 (13)	
Bilateral infiltrates	369 (66)	225 (66)	53 (60)	55 (73)	36 (64)	
Level of respiratory support						
Mechanically ventilated or ECMO	41 (7)	22 (6)	6 (7)	9 (11)	4 (7)	0.075
Non-invasive ventilation or high-flow oxygen	124 (21)	86 (24)	12 (13)	19 (23)	7 (12)	
Supplemental oxygen (not high flow)	284 (48)	173 (48)	52 (57)	32 (39)	27 (47)	
None	141 (24)	77 (22)	22 (24)	22 (27)	20 (34)	
235 or lower	100 (17)	67 (19)	9 (10)	16 (20)	8 (14)	0.537
236–315	101 (17)	58 (16)	19 (21)	16 (20)	8 (14)	
315 or higher	355 (60)	214 (60)	56 (61)	46 (56)	39 (67)	
(n = 590)	0.0 (0–2.0)	0.0 (0–2.0)	0.0 (0–2.0)	0.5 (0–3.0)	0.0 (0.0–2.0)	0.435
Baseline labs						
Lymphocyte count (1000 s/microliter), median (IQR) ^d	(n = 495) 1.1 (0.7–1.7)	1.0 (0.7–1.6)	1.2 (0.7–1.8)	1.1 (0.5–1.6)	1.2 (0.7–1.9)	0.371
Platelet count (1000 s/microliter), median (IQR) ^e	(n = 560) 238.5 (178–303)	241.0 (189–313)	218.5 (149–282)	247.0 (180–292)	222.0 (166–282)	0.017
ALT (Units/L), median (IQR) ^e	(n = 515) 33.0 (22–57)	34.0 (23–59)	30.0 (19–48)	33.5 (21–56)	31.0 (22–50)	0.378
Creatinine (mg/dL), median (IQR) ^b	(n = 568) 0.9 (0.7–1.1)	0.9 (0.7–1.1)	0.9 (0.7–1.1)	0.8 (0.7–1.1)	0.8 (0.7–1.1)	0.991
CRP (mg/L), median (IQR) ^f	(n = 418) 14.5 (6.0–74.7)	14.5 (6.7–76.5)	12.4 (5.0–82.2)	26.0 (6.8–93.0)	9.0 (4.5–60.4)	0.048
D-dimer (mg/L), median (IQR) ^g	(n = 416) 0.7 (0.5–1.3)	0.8 (0.5–1.4)	0.5 (0.4–0.8)	0.9 (0.4–1.8)	0.7 (0.5–1.3)	0.008
Troponin (ng/mL), median (IQR) ^h	(n = 199) 0.0 (0.01–0.06)	0.0 (0.01–0.06)	0.0 (0.01–2.9)	0.0 (0.01–0.05)	0.0 (0.02–0.04)	0.252
Severity & utilization during acute hospitalization						
Length of stay (days), median (IQR)	(n = 590) 5.0 (4.0–10.0)	5.0 (3.0–9.0)	5.0 (3.0–9.0)	6.0 (5.0–12.0)	6.0 (5.0–12.0)	0.021
ICU at any time during acute hospitalization	159 (27)	92 (26)	23 (25)	25 (30)	19 (33)	0.577
Acute trajectory group	145 (25)	97 (27)	24 (26)	15 (18)	9 (16)	0.327

Table 1 (continued) | Baseline characteristics of the survey respondent cohort

	Overall (n = 590)	Cluster MIN (n = 358, 60.7%)	Cluster PHY (n = 92, 15.6%)	Cluster COG (n = 82, 13.9%)	Cluster MLT (n = 58, 9.8%)	Over- all P value
TG2	195 (33)	109 (30)	27 (29)	33 (40)	26 (45)	
TG3	151 (26)	93 (26)	26 (28)	18 (22)	14 (24)	
TG4	99 (17)	59 (16)	15 (16)	16 (20)	9 (16)	
Acute complications reported						
Any complications	490 (83)	308 (86)	72 (78)	66 (80)	44 (76)	0.099
Number of complications, median (IQR)	2.0 (1.0–3.0)	2.0 (1.0–3.0)	2.0 (1.0–3.0)	2.0 (1.0–3.0)	2.0 (1.0–3.0)	0.95
Acute renal injury/failure	94 (16)	55 (15)	20 (22)	13 (16)	6 (10)	0.289
Liver dysfunction/failure	68 (12)	47 (13)	7 (8)	5 (6)	9 (16)	0.135
Anemia	62 (11)	33 (9)	10 (11)	16 (20)	3 (5)	0.024
Shock (use of vasopressors)	45 (8)	20 (6)	8 (9)	8 (10)	9 (16)	0.048
Bacteremia	39 (7)	20 (6)	8 (9)	6 (7)	5 (9)	0.637
Atrial fibrillation	32 (5)	16 (4)	6 (7)	6 (7)	4 (7)	0.644
Acute venous thromboembolism	29 (5)	16 (4)	5 (5)	6 (7)	2 (3)	0.686
Congestive heart failure (CHF)/cardiomyopathy	26 (4)	14 (4)	6 (7)	4 (5)	2 (3)	0.716
Hyperglycemia	24 (4)	14 (4)	4 (4)	3 (4)	3 (5)	0.968
Medications during acute hospitalization						
Steroids	400 (68)	251 (70)	63 (68)	55 (67)	31 (53)	0.094
Remdesivir	377 (64)	236 (66)	60 (65)	47 (57)	34 (59)	0.4
Convalescent symptoms						
Any symptom reported during convalescence	305 (52)	149 (42)	58 (64)	55 (68)	43 (74)	<0.001
Any upper respiratory (sore throat, conjunctivitis)	106 (18)	38 (11)	18 (20)	25 (31)	25 (43)	<0.001
Conjunctivitis/red eyes	79 (13)	26 (7)	14 (16)	17 (21)	22 (38)	<0.001
Sore throat	43 (7)	14 (4)	7 (8)	10 (12)	12 (21)	<0.001
Any cardiopulmonary (cough, dyspnea)						
Dyspnea	171 (29)	73 (20)	31 (34)	33 (41)	34 (59)	<0.001
Cough	116 (20)	44 (12)	26 (29)	21 (26)	25 (43)	<0.001
Any systemic (fever, fatigue, myalgia, chills)						
Fatigue	107 (18)	33 (9)	17 (19)	29 (36)	28 (48)	<0.001
Myalgia	121 (21)	39 (11)	24 (27)	29 (36)	29 (50)	<0.001
Fever	27 (5)	9 (3)	5 (6)	8 (10)	5 (9)	0.012
Chills	15 (3)	6 (2)	4 (4)	4 (5)	1 (2)	0.223

Table 1 (continued) | Baseline characteristics of the survey respondent cohort

	Overall (n = 590)	Cluster MIN (n = 358, 60.7%)	Cluster PHY (n = 92, 15.6%)	Cluster COG (n = 82, 13.9%)	Cluster MLT (n = 58, 9.8%)	Over- all P value
Any neurologic (headache, anosmia)	162 (28)	64 (18)	27 (30)	39 (48)	32 (55)	<0.001
Head-ache	112 (19)	40 (11)	20 (22)	27 (33)	25 (43)	<0.001
Anos-mia	83 (14)	29 (8)	13 (14)	21 (26)	20 (34)	<0.001
Any gastrointestinal (nausea/vomiting)	41 (7)	11 (3)	11 (12)	9 (11)	10 (17)	<0.001

Cluster MIN: minimal deficit, cluster PHY: deficit, physical predominant, cluster COG: deficit, mental/cognitive predominant and cluster MLT: deficit, multidomain. SOT solid organ transplantation, BMT bone marrow transplantation, ECMO extracorporeal membrane oxygenation, SpO₂/FIO₂ saturation of oxygen, SOFA sequential organ failure assessment score, ICU intensive care unit, ALT alanine transaminase, CRP: C-reactive protein.

Two-sided *p* values without adjustment for multiple comparisons are reported from Wilcoxon rank-sum test for continuous age, χ^2 test for categorical variables.

[†]28% other/declined/unknown/missing.

[‡]Less than 5% missing data.

[§]Less than 10% missing data.

[¶]Less than 20% missing data.

^{**}Less than 15% missing data.

^{***}Less than 30% missing data.

^{††}66% did not have troponin level on admission.

received less oxygen supplementation OR = 0.54 (95% CI: 0.34, 0.87) and had longer acute hospitalization (OR per week 1.44; 95% CI 1.19–1.75) relative to those in the MIN cluster (Fig. 2).

Laboratory assays by PRO cluster

N1 gene SARS-CoV-2 PCR cycle threshold (Ct) values up to 28 days since admission differed significantly among the four PRO clusters (Fig. 3 and Fig. 4S). Specifically, the viral RNA levels from the respiratory tract were significantly higher (lower Ct values) throughout those 28 days in participants within both the PHY and MLT clusters compared the MIN cluster in the longitudinal GAM model (Fig. 4S, adj. *p* = 0.015). There was no difference in Ct values between clusters with MIN and COG deficits. Viral levels declined over time in all four PRO clusters and the rate of viral clearance did not differ significantly among the PRO clusters (Fig. 4S). Similar trends were also observed with N2 SARS-CoV-2 PCR Ct values (Fig. 4S, adj. *p* = 0.013).

Acute phase anti-SARS-CoV-2 RBD IgG and Spike IgG titers, were also significantly associated with PRO clusters (Fig. 3 and Fig. 4S). Both anti-RBD IgG and anti-Spike IgG levels were significantly lower in the MLT compared to the MIN and COG clusters by the GAM model (Fig. 4S, adj. *p* = 0.023) and in the PHY compared to the MIN and COG clusters by the GLM model (Figs. 3 and 4S, adj. *p* = 0.014). There was a swift rise in antibody (Ab) levels during first 7 days followed by a modest increase between 7 and 20 days, reaching a plateau at day 28 in all PRO clusters. We also observed significantly faster rise in Ab levels between 7 and 20 days for clusters PHY and MLT compared to the other 2 clusters (Fig. 4S; shape, anti-RBD IgG adj. *p* = 0.005 and anti-Spike IgG adj. *p* = 0.0017). The ratio of anti-RBD IgG to N1 Ct followed similar patterns (Fig. 4S).

We next considered whether circulating leukocyte subset frequencies during the acute phase correlated with PRO clusters and carried out deep immunophenotyping using CyTOF to quantify cell frequencies in whole blood. Our analysis of cell subsets revealed that in the first 28 days of hospitalization, differences in the frequency of circulating B lymphocytes were significantly associated with PRO clusters (Fig. 4, adj. *p* = 0.0191, Participants with MLT deficits (adj. *p* = 0.0005) and COG deficits (adj. *p* = 0.025) had a significantly lower frequency of circulating B cells compared to those with MIN deficits (Fig. 4). Notably, within B cells, participants with MLT deficits showed a lower frequency of naive B cells compared to MIN cluster, although the difference did not remain significant after multiple-comparison correction for all B cell subsets (Fig. 5S). Other immune cell subtypes were not significantly associated with PRO cluster. The ratios of SARS-CoV-2 PCR Ct values and antibody titers to B cell numbers are shown in Fig. 6S, indicating that the lower frequency of circulating B cells may be relevant to outcome in the context of distinct kinetics of Ab production and viral clearance.

Autoantibodies with blocking activity against type I interferons (alpha, beta, and omega) were observed at the onset of hospitalization in 4.3% participants (24 out of 563) across four PRO clusters (Table 2S). Anti-interferon autoantibodies were detected in 5.3% of males and 2.7% female, (*P* = 0.14) and were more common in individuals older than 65 years of age (8.9% older vs 2.8% younger, *P* = 0.039). We observed a proportionately larger fraction of individuals with positive autoantibodies against type I IFNs from PHY and MLT PRO clusters (PHY cluster = 6.9%, MLT cluster = 7.3%) compared to the other 2 clusters (MIN cluster = 3.2%, COG cluster = 3.8%) (*p* = 0.2). In a matched case-control (1:3) based on age and sex comparison, individuals with IFN autoantibodies had significantly higher viral loads than individuals without IFN autoantibodies (Fig. 7S, N1 Ct *P* = 0.012 and N2 Ct *P* = 0.006). We did not see significant differences in Ab titers against either SARS-CoV-2 RBD IgG or spike IgG levels between these matched groups.

Using a panel of 92 inflammatory markers, we analyzed serum samples using a proximity extension assay (Olink) across the period of

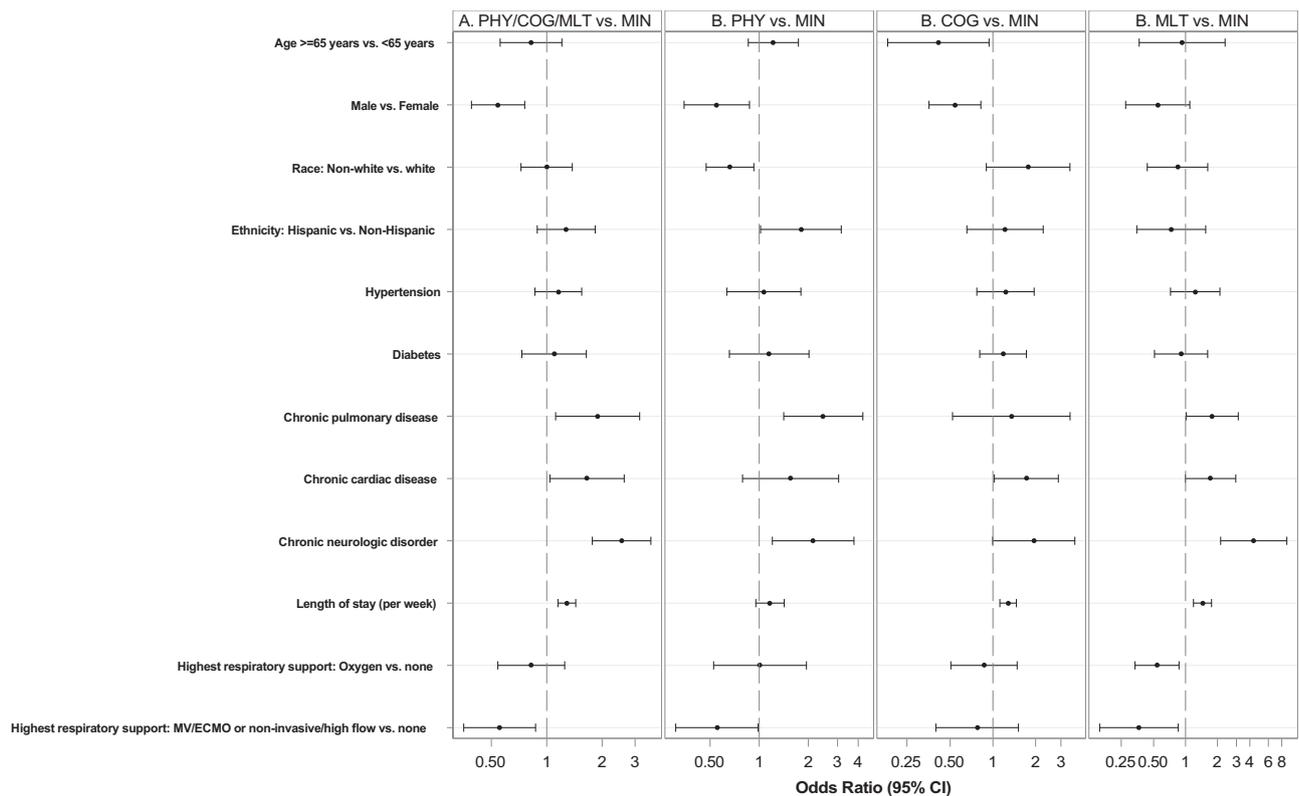


Fig. 2 | Forest Plot showing adjusted odds ratios (ORs) for factors associated with patient-reported outcome (PRO) clusters with more deficits compared to minimal deficit, using multivariable multinomial logistic regression ($N = 590$). A Comparison of PRO clusters PHY, COG, and MLT with PRO cluster MIN.

B Comparison of PRO clusters PHY and MIN (left); clusters COG and MIN (middle); clusters MLT and MIN (right). MIN minimal deficit, PHY physical predominant deficit, COG mental/cognitive predominant deficit, MLT multidomain deficit, MV Mechanical ventilation, ECMO Extracorporeal membrane oxygenation.

acute illness; one protein, fibroblast growth factor 21 (FGF21), was significantly elevated in the COG cluster (adj. $p = 0.0025$) as well as the MLT cluster (adj. $p = 0.00033$), relative to the MIN cluster. The highest mean FGF21 values were in the MLT cluster (Fig. 5, Fig. 8S).

Analyzing 658 serum metabolites, 27 modules were identified from a weighted gene correlation network analysis (WGCNA) which corresponded to each metabolite feature. We observed a significant difference in shape (referred to as the smoothing term in the γ_{mm4} documentation) for methylhistidine metabolism (global metabolomics module 3) and acylcarnitine metabolism (global metabolomics module 18) among the PRO clusters (Fig. 9SA, SB). Notably, significantly lower levels of metabolites related to methylhistidine metabolism were observed for participants in the PHY and the MLT clusters, compared to the MIN cluster (shape adj. $p = 0.049$). Further, significantly higher levels of metabolites related to acylcarnitine metabolism were observed for participants in the PHY cluster, compared to the MIN cluster (adj. $p = 0.049$).

Discussion

In this large prospective study that followed participants from the time of acute COVID-19 hospitalization, more than half of the participants hospitalized with COVID-19 had persistent symptoms lasting 3 or more months after discharge, consistent with other studies⁹. The clustering of Patient-Reported Outcomes (PROs) by predominant deficit (physical predominant, mental/cognitive predominant, and multidomain deficits) supports PASC as a heterogeneous clinical entity with distinct sub-phenotypes associated with unique perturbations of the immune system in the acute phase of the illness^{10,11}. This cluster-based analysis also revealed a specific participant phenotype associated with persistent mental and cognitive impairments that were distinct from phenotypes associated with broader physical dysfunction. These findings

suggest tailored approaches will be needed in the management of PASC¹²⁻¹⁴.

Overall demographic and clinical risk factors for PASC in our cohort include female sex, comorbidities^{13,15,16} such as chronic heart, lung, or neurologic disease, as well as longer length of hospital stay¹⁷. The biological basis for why females may be more susceptible to PASC than males has yet to be defined, though several models have been proposed^{18,19}. One potential mechanism is autoimmunity, though we did not find any difference between males and females in autoantibodies against type I interferons. Hormonal factors may also play a role in perpetuating the hyperinflammatory status of the acute disease phase even after initial recovery²⁰. While certain comorbidities are associated with PASC (e.g., chronic pulmonary diseases in the physical predominant deficit cluster), it is unclear based on our data if some of the PASC disease burden could be misattributed to COVID-19 or that COVID-19 accentuates these pre-existing conditions. Consistent with prior reports in non-hospitalized patients, we found no association between PASC and acute COVID-19 disease severity²¹⁻²⁴, but a longer length of stay was associated with all PRO deficit clusters when compared to those with minimal functional deficits, similar to findings from a population-based cohort study²⁵. Interestingly, our results suggest that supportive interventions such as oxygen therapy may be associated with a lower likelihood of being in the multidomain deficit PRO cluster, supporting the notion that early acuity-based interventions may positively influence clinical outcomes²⁶.

Our data demonstrate that higher SARS-CoV-2 viral burden and lower Ab titers during the acute phase are associated with both the physical predominant deficit as well as the multidomain deficit PRO clusters. Of note, these virologic and serologic findings do not distinguish participants with minimal deficits from those in the mental/

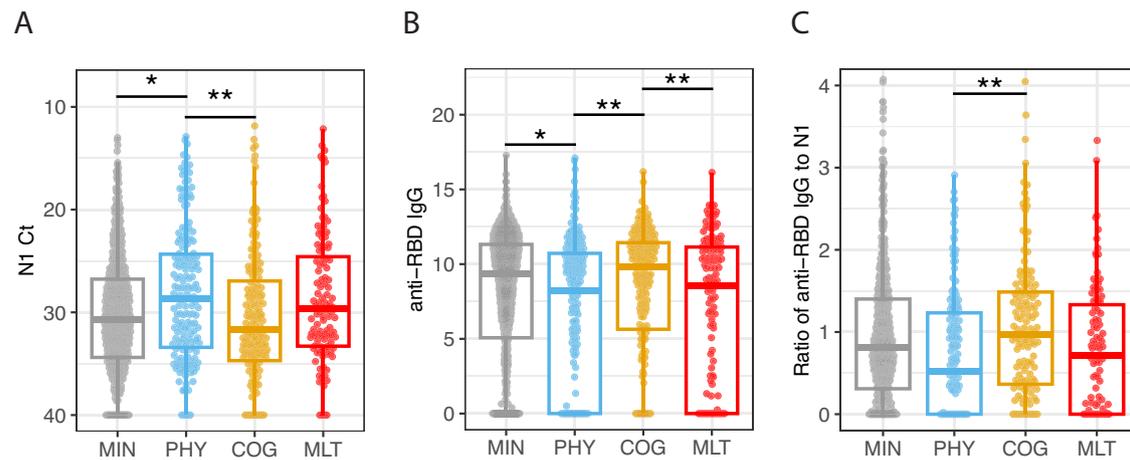


Fig. 3 | SARS-CoV-2 viral RNA levels and antibody responses. **A** N1 Ct values: shown are SARS-CoV-2 N1 gene PCR cycle threshold (Ct) values (viral loads) measured from samples collected during the first 28 days of hospital admission by four PRO clusters, minimal deficit (MIN, $n = 657$), physical predominant (PHY, $n = 174$), deficit, mental/cognitive predominant (COG, $n = 172$ and deficit, multidomain (MLT, $n = 112$). Shown are median values (horizontal lines), interquartile ranges (boxes), and 1.5 IQR (whiskers), as well as all individual points. Because lower Ct values indicate higher viral loads, the y axis is reversed. The viral loads were significantly (adj. $p = 0.03$) associated with the PRO clusters. **B** anti-RBD IgG values: Shown are anti-RBD IgG values measured from samples collected during the first 28 days of hospital admission by four PRO clusters, minimal deficit (MIN, $n = 907$), physical predominant (PHY, $n = 221$), deficit, mental/cognitive predominant (COG, $n = 230$) and deficit, multidomain (MLT, $n = 149$). Shown are median values of area under the curve (AUC) (horizontal lines), interquartile ranges (boxes), and 1.5 IQR (whiskers), as well as all individual points. The titers were significantly (adj. $p = 0.014$) associated with the PRO clusters. **C** Ratio of anti-RBD IgG to N1 values:

shown are scaled ratio of anti-RBD IgG to SARS-CoV-2 viral loads (N1 gene) values from samples collected during the first 28 days of hospital admission by four PRO clusters, minimal deficit (MIN, $n = 560$), physical predominant (PHY, $n = 156$), deficit, mental/cognitive predominant (COG, $n = 141$) and deficit, multidomain (MLT, $n = 99$). Shown are median values (horizontal lines), interquartile ranges (boxes), and 1.5 IQR (whiskers), as well as all individual points. The ratio of titers to viral loads was also significantly (adj. $p = 0.05$) associated with the PRO clusters. The four PRO clusters are the following in gray: minimal deficit (MIN), in blue: deficit, physical predominant (PHY), in yellow: deficit, mental/cognitive predominant (COG), and in red: deficit, multidomain (MLT). The lines and asterisks on top of the figure denote pairwise statistical significance, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Statistical differences were determined from generalized linear mixed effects models adjusting for age, sex, participant, and enrollment site. P values were adjusted using the Benjamini-Hochberg method to account for multiple comparisons. See Methods for more details.

cognitive predominant deficit PRO cluster, suggesting that different factors could lead to the development of this particular cluster. The described calculated ratio (IgG/Ct value) is unique not only in the acute phase to determine the trajectory of acute disease course but also associates with PRO clustering in the convalescent phase⁸, and thus may represent a practical approach for patient risk stratification for both early mortality and subsequent morbidity from PASC. Our data confirm findings from other studies suggesting that PASC is associated with initial high SARS-CoV-2 RNA levels²⁷ and a suboptimal serological response^{28–30} consistent with a reduced number of circulating B cells mostly in the multidomain deficit cluster. The therapeutic and protective effects of immunoglobulins were the basis for the pre-Omicron use of convalescent plasma and monoclonal antibodies (mAbs) and are the basis for the continued use of vaccination in the treatment and prevention of COVID-19. Recent observations of PASC resolution after SARS-CoV-2 vaccination raise the possibility of depletion of persisting viral reservoirs³¹. Since none of our study participants (enrolled in 2020 and early 2021) were vaccinated prior to their illness and only 6% received convalescent plasma, it was not possible to determine the impact of early Ab therapy or vaccination on the likelihood of PASC. Regarding other COVID-19-directed interventions, use of remdesivir and steroids in the inpatient period was not associated with a decrease in PASC prevalence. However, a recent study reported a protective effect against PASC when another antiviral, nirmatrelvir/ritonavir, was used in the acute phase of COVID-19, consistent with our finding of high viral load being associated with the development of PASC^{32,33}. PASC has been associated with the detection of different autoantibodies early in the disease course^{6,27,34–36}. In our study, the presence of IFN-specific autoantibodies is associated with viral load and severity of acute COVID³⁷. Autoantibodies neutralizing IFN α , IFN β , and/or IFN ω result in a persistent dampening of IFN responses, likely leading to

insufficient viral clearance (as seen in our study in our case-control subanalysis) and tissue damage.

While numerous cytokines have been associated with PASC in other studies, in this study, we found that only fibroblast growth factor 21 (FGF21) was significantly associated with the PRO clusters^{38,39}. FGF21 is a cytokine known to regulate systemic glucose and lipid metabolism that is secreted from muscle in response to stress⁴⁰ or even infection, particularly mitochondrial myopathy⁴¹ supporting a potential catabolic role of FGF21 on human muscle health⁴². In our metabolomics data, we found a significant association between global metabolomics module 3 (methylhistidine metabolism) and the PRO clusters, consistent with the inverse relationship between 3-methylhistidine (3MH) and FGF21, thought to be mediated by insulin sensitivity⁴³. This observation is also consistent with the relationship between 3-methylhistidine and muscle cells, in which 3MH is a potential biomarker for muscle atrophy and skeletal muscle toxicity. Whether the association we observed with FGF21 reflects underlying mitochondrial dysfunction as the pathobiological basis for PASC is unclear and deserves further investigation⁴³. Elevation of plasma FGF21 has also been noted in patients with myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS), an entity with clinical features that overlaps with PASC including fatigue, post-exertional malaise, sleep disturbance, and brain fog⁴⁴. Interestingly, acylcarnitines also noted to be significant comparing PHY and MIN cluster, are known to have an essential role in metabolism and breaking down fatty acids for energy production^{45,46}. In addition, acyl carnitine substrates have been investigated as surrogate biomarkers for ME/CFS^{47,48} and impaired metabolic health⁴⁹, suggesting its possible role in reduced physical function after COVID-19 hospitalization.

Overall, participants in the MLT cluster with the most severe functional deficits during the year following hospitalization for COVID-

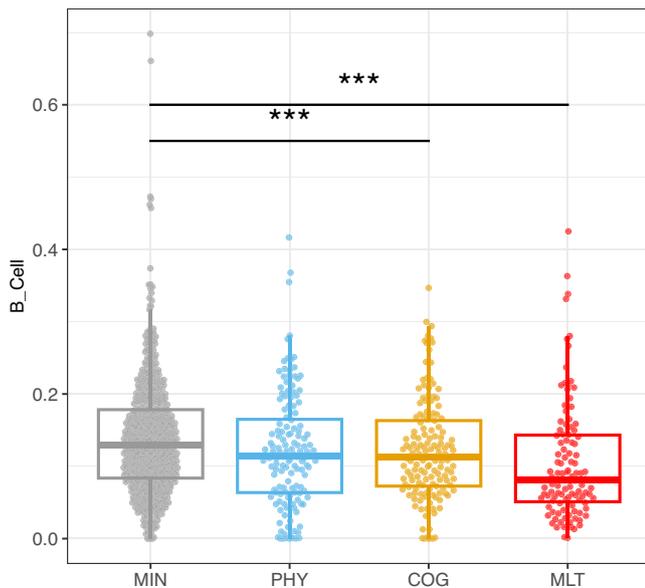


Fig. 4 | B cell to non-granulocyte frequency. Shown are B cell to non-granulocyte frequency values from samples collected during the first 28 days of hospital admission by four PRO clusters, minimal deficit (MIN, $n = 584$), physical predominant (PHY, $n = 140$), deficit, mental/cognitive predominant (COG, $n = 145$) and deficit, multidomain (MLT, $n = 107$). Shown are median values (horizontal lines), interquartile ranges (boxes), and 1.5 IQR (whiskers), as well as all individual points. The repeated-measurement model identified significant differences of B cell to non-granulocyte frequency in association with convalescent clusters (adj. $p = 0.0191$). The 4 clusters are the following in gray: minimal deficit (MIN), in blue: deficit, physical predominant (PHY), in yellow: deficit, mental/cognitive predominant (COG), and in red: deficit, multidomain (MLT). The lines and asterisks on top of the figure denote pairwise statistical significance, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Statistical differences were determined from generalized linear mixed effects models adjusting for age, sex, participant, and enrollment site. P values were adjusted using the Benjamini–Hochberg method to account for multiple comparisons. See Methods for more details.

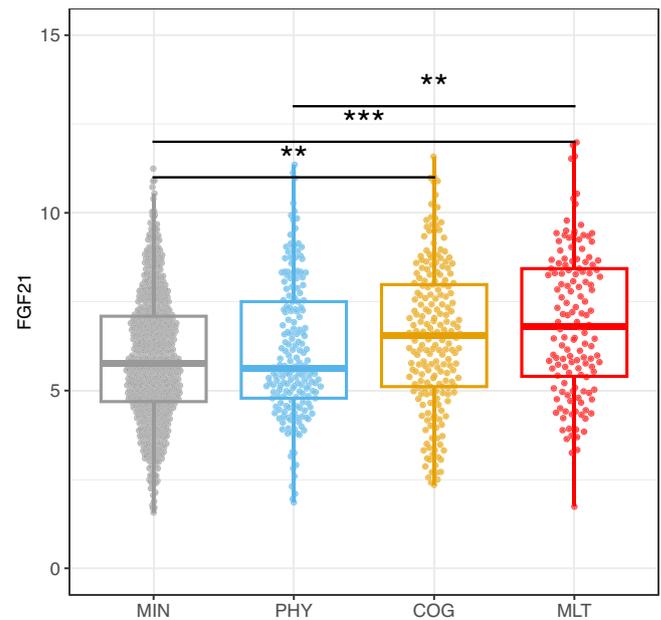


Fig. 5 | Circulating fibroblast growth factor 21 expression. Circulating fibroblast growth factor 21 (FGF21) NPX (Normalized protein expression): Shown are FGF21 NPX values from samples collected during the first 28 days of hospital admission by four PRO clusters, minimal deficit (MIN, $n = 716$), physical predominant (PHY, $n = 189$), deficit, mental/cognitive predominant (COG, $n = 210$) and deficit, multidomain (MLT, $n = 139$). Shown are median values (horizontal lines), interquartile ranges (boxes), and 1.5 IQR (whiskers), as well as all individual points. The generalized additive model (GAM) identified a significant difference in FGF21 expression level in association with convalescent cluster groups (adj. $p = 0.0135$). The four clusters are the following in gray: minimal deficit (MIN), in blue: deficit, physical predominant (PHY), in yellow: deficit, mental/cognitive predominant (COG), and in red: deficit, multidomain (MLT). Statistical differences were determined from generalized linear mixed effects models adjusting for age, sex, participant, and enrollment site. P values were adjusted using the Benjamini–Hochberg method to account for multiple comparisons. See Methods for more details.

19 had a suboptimal serological response likely driven by low circulating B cells potentially leading to high viral replication in the acute phase of the disease. In a case-control analysis, the highest percentage of autoantibodies against interferon (IFN) was noted in the MLT cluster, and the presence of these autoantibodies correlated with a high viral load. In addition, we noted an elevation in FGF21 levels in the MLT cluster similar to the COG cluster indicating possible muscle injury/stress that could explain common PASC symptoms such as fatigue and malaise. Interestingly, the COG cluster was not associated with a high viral load nor suboptimal serological response, suggesting the possibility that early immunologic parameters associated with mental and cognitive deficits following COVID-19 may be distinct from those leading to other functional limitations.

This study has several strengths, including enrollment of a diverse population from a wide variety of geographically dispersed hospitals and detailed clinical and biological phenotyping, as well as several limitations. Due to the timing of our recruitment window, very few study participants were infected with variants of concern or variants of interest or vaccinated prior to hospital admission⁸; thus, symptom persistence in our cohort may not be representative of patients infected with more recent emerging SARS-CoV-2 variants or with breakthrough infections⁵⁰, but does provide an important characterization of post-acute disease follow-up in a virus-naïve population.

Along the same lines, certain symptoms that are now frequently linked to PASC (e.g., ‘brain fog’, sleep disturbance, dysautonomia) were not recorded as the surveys were designed prospectively early in the

pandemic (March 2020) when PASC had not yet been reported. However, our use of standardized PROs that targeted cognitive, mental, and psychosocial functions enabled the identification of a specific cluster with predominant mental and cognitive deficits. Although symptoms at hospital admission were captured, pre-COVID symptomatology was not recorded, limiting estimation of the proportion of persistent symptoms directly attributable to PASC versus part of a pre-existing comorbid condition. In addition, we did not attempt to identify alternative causes of persistent or new symptoms. However, PRO measures were chosen to attempt to mitigate this limitation by including a comparison to pre-illness baseline or some other appropriate recall period when possible. Our incidence of new onset or worsening health-related quality measures is in line with other published studies with 31% impairment in one study⁵¹, and 15.4% with poor physical component, and 32.6% with poor mental component in another study⁵².

Other investigators observed that certain occupations and socioeconomic status are associated with PASC^{15,16}; we were unable to assess such associations due to a lack of occupational and socioeconomic data in our study.

Exclusion of non-hospitalized patients also affects the generalizability of our findings to patients with COVID-19 not requiring hospitalization. Also, our study did not include control groups: (1) who did not have COVID, (2) hospitalized for a non-COVID-19 respiratory viral infection, and/or (3) hospitalized for elective procedures where the length of stay is similar to COVID-19.

T-cell dysfunction has been described in PASC⁵³, but our study did not include this assessment. We also did not fully explore other

autoantibodies beyond those against type I IFNs (e.g., Ro/SS-A, La/SS-B, UI-snRNP, Jo-1, and PI⁵⁴ to β 2 adrenoceptor, muscarinic M2 receptor, angiotensin II AT1 receptor, and angiotensin 1–7 MAS receptor)⁶. Similarly, EBV reactivation has been reported as potentially associated with PASC and was not investigated here.

We did not have an independent validation cohort, however, enrollment at multiple sites may have decreased selection biases.

While we largely report immunologic results observed in the analysis of blood, we also evaluated the acute immune response in the upper airway in non-intubated participants and in the lower airway in participants receiving mechanical ventilation; studies that did not detect significant associations with convalescent phenotypes. Acknowledging the broad organotropism of SARS-CoV-2, future research should explore other compartments serving as potential viral reservoirs⁵⁵.

In addition to preventing and treating acute infections, there is a dire need to better understand and develop treatments for individuals with PASC. Our study represents a large multi-site prospective cohort with extensive clinical data capture and 12 months of follow-up after discharge, as well as intensive immunophenotyping efforts that employed a variety of innovative assays with rigorous data management and a standardized analysis pipeline. Our findings suggest that a functional antiviral Ab immune response contributes to viral clearance and may decrease the occurrence of PASC presenting as significant physical and multidomain deficits. Our results also highlight the benefit of measuring immune responses during the acute phase for the early identification of patients at high risk for PASC, which may facilitate testing and monitoring of targeted PASC prevention and treatment.

Methods

Ethics

NIAID staff conferred with the Department of Health and Human Services Office for Human Research Protections (OHRP) regarding the potential applicability of the public health surveillance exception [45CFR46.102(l) (2)] to the IMPACC study protocol. OHRP concurred that the study satisfied criteria for the public health surveillance exception, and the IMPACC study team sent the study protocol, and participant information sheet for review and assessment to institutional review boards (IRBs) at participating institutions. Twelve institutions elected to conduct the study as public health surveillance, while 3 sites with prior IRB-approved biobanking protocols elected to integrate and conduct IMPACC under their institutional protocols (University of Texas at Austin, IRB 2020-04-0117; University of California San Francisco, IRB 20-30497; Case Western Reserve University, IRB STUDY20200573) with informed consent requirements. Participants enrolled under the public health surveillance exclusion were provided information sheets describing the study, samples to be collected, and plans for data de-identification and use. Those who requested not to participate after reviewing the information sheet were not enrolled. In addition, participants did not receive compensation for study participation while inpatient, and subsequently were offered compensation during outpatient follow-ups.

Study design and setting

The study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for reporting observational studies⁵⁶. This study was registered at clinicaltrials.gov (NCT0438777).

Study participants

Patients 18 years and older admitted to 20 US hospitals (affiliated with 15 academic institutions) between May 2020 and March 2021 were enrolled within 72 h of hospital admission for COVID-19 infection. Only confirmed positive SARS-CoV-2 PCR and

symptomatic cases attributable to COVID-19 infection were followed longitudinally⁵⁷. Participants were provided compensation on an outpatient basis.

Data collection, study variables, and biologic samples

Specific data elements were acquired via a review of electronic medical records during the inpatient period⁸. The study was designed to enroll participants of both sexes, and sex at birth was collected based on self-report or caregiver report. Length of hospital stay, complications, mortality, and other protocol-defined outcomes were assessed over 28 days. In addition, self-reported symptoms, reinfections, SARS-CoV-2 vaccination, rehospitalizations, and standardized patient-reported outcome surveys were assessed quarterly for the duration of the study up to 12 months after initial hospital discharge.

Biologic samples collected consisted of blood and mid-turbinate nasal swabs (self, or staff collected). The timepoints were as follows: enrollment (Day 1), and Days 4, 7, 14, 21, and 28 post hospital admission (and if feasible, for discharged participants, Days 14 and 28).

Patient-reported data was collected using a comprehensive digital remote monitoring tool, in the form of a mobile application developed by *My Own Med, Inc.* Along with the mobile application, an administrative portal was developed to collect information by study personnel during site visits or via telephone interviews with a study coordinator to ensure real-time electronic data capture.

The surveys administered at these remote visits:⁵⁸

- Upper respiratory symptoms: sore throat, conjunctivitis/red eyes
- Cardiopulmonary symptoms: shortness of breath (dyspnea), cough
- Systemic symptoms: fever, chills, fatigue/malaise, muscle aches (myalgia)
- Neurologic symptoms: loss of smell/taste (anosmia/ageusia), headache
- Gastrointestinal symptoms: nausea/vomiting

In addition, the functional assessments of general health and the evaluation of deficits in specific health domains were conducted using validated Patient-Reported Outcome (PRO) measures, including:

- EQ-5D-5L, a standardized, self-administered instrument that describes and quantifies health-related quality of life⁵⁹
- Patient-Reported Outcomes Measurement Information System (PROMIS). The PROMIS measures administered included:
 - *PROMIS® Item Bank v2.0 - Physical Function and PROMIS Item Bank v2.0 - Cognitive Function*, two computer adaptive surveys with tailored questionnaires based on item response theory⁶⁰.
 - *PROMIS Scale v1.2 - Global Health Mental 2a and PROMIS Item Bank v1.0 - Psychosocial Illness Impact-Positive - Short Form 8a*, two surveys with fixed questions^{61–64}.
 - *PROMIS Pool v1.0 - Dyspnea Time Extension* computer adaptive instrument for participants who reported shortness of breath^{65–67}. This 7-item questionnaire assesses whether there has been a meaningful increase or decrease in the duration of time needed by an adult to perform a given task in the past 7 days compared to 3 months ago due to shortness of breath.

For all PROMIS measures, scoring was based on PROMIS standardized instructions and conversion to a t-statistic⁶⁸.

- Health Recovery Score: Overall health was also assessed by a health recovery score utilizing a Visual Analog Scale of 1–100 to indicate overall physical and mental function compared to pre-COVID function.

All data were reviewed centrally to ensure accuracy and consistency. Any data concerns were resolved by querying the site.

The full study data collection forms for the quarterly outpatient surveys are provided in the Supplementary Information (Surveys Administered).

Assays

- SARS-CoV-2 viral load was assessed by a central laboratory from nasal swab samples at each defined time point by RT-PCR of the viral N1, and N2 genes⁶⁹(Supplementary Methods).
- Anti-SARS-CoV-2 spike (S), and receptor binding domain (RBD) antibodies were quantitated by enzyme-linked immunosorbent assay (ELISA) in serum specimens⁷⁰(Supplementary Methods).
- Autoantibodies: blocking antibodies against type I IFNs (IFN α , IFN β , and IFN ω) were assessed in a multiplex, particle-based assay (Supplementary Methods).
- Blood CyTOF 65 cell subsets were identified using a panel of 43 antibodies to cell surface markers expressed by distinct lineages and intracellular markers of functional status. A semi-automated gating strategy was used⁷¹ (Supplementary Methods).
- Proximity Extension Assay (O-Link) multiplex assay inflammatory panel (Olink Bioscience, Uppsala, Sweden) includes 92 different proteins associated with human inflammatory conditions (Supplementary Methods).
- Plasma global metabolomics was assessed using liquid chromatography-mass spectrometry technique as described (Supplementary Methods).

Statistics

Convalescent clinical outcome assessment

Overall analytic approach. Because we focused on how longitudinal patterns in PROs might define clinical phenotypes relevant to PASC, we modeled each PRO using an approach that assumes the population is composed of distinct groups, each of which follows a different underlying and unobserved trajectory, with individual-level variation around that trajectory. Our approach to the high-dimensional data problem presented by multiple PROs captured across multiple time-points was first to preserve longitudinal patterns within each PRO using LCMMs. We then reduced further the dimensionality of the data by clustering the resulting groupings across multiple PROs using five clustering algorithms and four diagnostic statistics. This approach mirrors a recent study⁷² with similar aims to cluster high-dimensional data across several clinical outcomes. We compared the mean values of each PRO within each cluster to the remaining sample to interpret each cluster with respect to the contributions of each PRO. Finally, we compared clinical, demographic, and laboratory assay variables across each of the clusters thus determined. Convalescent clinical outcomes were analyzed using R Statistical Software version 4.2.1.

Latent class analysis. We considered the PROs collected at quarterly intervals and modeled longitudinally using LCMM, a family of models of which a well-known and commonly used example is the group-based trajectory model^{73,74}, implemented by R package “lcm”. Outcome variables were the EQ-5D-5L, global Health Recovery Score, PROMIS Cognitive Function Score, PROMIS Physical Function Score, PROMIS Dyspnea Score, PROMIS Global Mental Health Score, and PROMIS Psychosocial Illness Impact Positive Score. We evaluated linear and quadratic models with number of groupings ranging from 1 to 5, and specified the model based on convergence criteria and goodness-of-fit using Bayesian Information Criteria (BIC). For each outcome, we selected the model that converged and had the lowest BIC.

Cluster analysis. Using the assigned groups from the LCMM step for each PRO, we then applied cluster analysis to group participants with similar PRO longitudinal patterns. For those PROs with no distinct

longitudinal clusters, we assigned to each participant the within-participant mean for that PRO. We calculated inter-participant similarity using Gower distance implemented by R package “CluMix”. We applied five clustering algorithms (Ward, McQuitty, Average, PAM, and Complete) to the distance matrix to identify the optimal number of clusters, and selected the optimal model based on four cluster fitting statistics (within-cluster SS, average silhouette width, Dunn index, and ratio of within-to-between SS). We then excluded cluster solutions with degenerate clusters (e.g., those with only one participant).

Thus, the best model performed well on the four fitting statistics overall and had a clinically interpretable number of clusters. We further excluded solutions with clusters of size $n = 5$ or smaller. We generated cluster assignments using R package “cluster” with fit statistics implemented by package “fpc”. To estimate the strength of association of each PRO with particular clusters, we calculated a t-statistic comparing the mean value of each PRO within each cluster versus the mean value of that PRO across the remaining clusters. The t-statistics were recoded such that negative values indicated a greater degree of patient-reported deficit, while positive values indicated no reported deficit.

Statistical analysis—demographic & clinical variables. We report median (interquartile range, IQR) for continuous variables and frequency (percent) for categorical variables. We examined bivariate associations between demographic and clinical factors and the PRO clusters using the Wilcoxon rank-sum test for continuous variables and chi-square test for categorical variables. Multinomial logistic regression was used to examine the adjusted associations between demographic and clinical factors and cluster membership, comparing the likelihood of being in each of the deficit clusters relative to MIN deficit cluster. $P < 0.05$ was considered statistically significant.

Analysis of laboratory assays. To identify modules of correlated features from high-dimensional ‘omics data, we utilized Weighted Gene Co-expression Network Analysis (WGCNA) v1.71. We specified the module “value” as the first principal component of features in the module to summarize each group of assay readouts for subsequent analysis. For interpretation, the features in each module were annotated to biological processes by performing an enrichment analysis leveraging biological knowledge bases, including MSigDB Hallmark gene sets, SMPDB metabolites and pathways. To identify the associations between different immune measurements and the four PRO clusters, we used two complementary approaches that each account for repeated measures per individual. The first was the use of generalized linear mixed effects models (GLMs) including a random effect for individual but not accounting for the timing of sample collection, the second was generalized additive mixed effects models (GAMs) that do account for timing of sample collection and allowed us to investigate longitudinal patterns. For both approaches, we utilized the measurements from samples collected within 28 days of hospital admission (with up to 6 samples per participant). In the GLM approach, we ignored the time of sample collection and identified features with different mean values from the aggregated timepoints among the PRO clusters. In the GAM approach, we investigated whether there were either differences in the average values over time or differences in the temporal patterns of features among PRO clusters. Each model is adjusted for fixed effects of participant age, sex, and random effects for participant and enrollment site. We used R packages, “lme4” for the GLM approach and “gam4” for GAM approach. Significant associations were defined at false discovery rate (FDR) < 0.05 using the Benjamini-Hochberg method to account for multiple comparisons. For both approaches, significant features were tested by post-hoc pairwise comparisons to identify the differences between each pair of PRO clusters to facilitate interpretation. Features for which the aggregated mean values in the GLM, the average over time (referred to as intercept

in the gamm4 documentation), or the shape (referred to as the smoothing term in the gamm4 documentation) differed among PRO clusters at FDR < 5% were considered significant.

Case-control analysis: anti-IFN antibodies. To determine whether IFN autoantibodies were associated with viral burden, we performed a case-control analysis, and identified age and sex-matched controls for the 24 individuals who tested positive for IFN autoantibodies with blocking activity at their earliest hospital visit (3:1 ratio of controls to cases). SARS-CoV-2 viral load (N1 Ct and N2 Ct values), SARS-CoV-2 RBD, and Spike binding IgG titers were compared between cases and controls. Significant differences in median levels between the two groups were assessed using Wilcoxon rank-sum test.

Reporting summary

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

Data availability

The IMPACC Data Sharing Plan is designed to enable the widest dissemination of data, while also protecting the privacy of the participants and the utility of the data by de-identifying and masking potentially sensitive data elements. All IMPACC data, including those generated in this study, have been deposited in the Immunology Database and Analysis Portal (ImmPort), a NIAID Division of Allergy, Immunology and Transplantation funded data repository under accession code [SDY1760](#). All raw and processed data are available under restricted access to comply with the NIH public data sharing policy for IRB-exempted public health surveillance studies, access can be obtained via AccessClinicalData@NIAID (https://accessclinicaldata.niaid.nih.gov/study-viewer/clinical_trials). Additional guidelines for access are outlined on ImmPort (<https://docs.immport.org/home/impaccslides>). In addition, raw LC-MS data for metabolomics are submitted to Metabolights repository under accession number [MTBLS850](#).

Code availability

All codes for the analyses and tables generated by this study have been deposited in the Bitbucket repository https://bitbucket.org/kleinstein/impacc-public-code/src/master/convalescent_manuscript/ and are publicly available as of the date of publication.

References

- COVID-19 rapid guideline: managing the long-term effects of COVID-19. <https://www.nice.org.uk/guidance/ng188>. (2022).
- National Institute for Health and Care Excellence: Clinical Guidelines. COVID-19 rapid guideline: managing the long-term effects of COVID-19. London: National Institute for Health and Care Excellence (NICE) Copyright © NICE 2020. (2020).
- Admon, A. J. et al. Assessment of symptom, disability, and financial trajectories in patients hospitalized for COVID-19 at 6 months. *JAMA Netw. Open* **6**, e2255795 (2023).
- Groff, D. et al. Short-term and long-term rates of postacute sequelae of SARS-CoV-2 infection: a systematic review. *JAMA Netw. Open* **4**, e2128568 (2021).
- Merad, M., Blish, C. A., Sallusto, F. & Iwasaki, A. The immunology and immunopathology of COVID-19. *Science* **375**, 1122–1127 (2022).
- Wallukat, G. et al. Functional autoantibodies against G-protein coupled receptors in patients with persistent long-COVID-19 symptoms. *J. Transl. Autoimmun.* **4**, 100100 (2021).
- Klein, J. et al. Distinguishing features of long COVID identified through immune profiling. *Nature* <https://www.medrxiv.org/content/10.1101/2022.08.09.22278592v1> (2023).
- Ozonoff, A. et al. Phenotypes of disease severity in a cohort of hospitalized COVID-19 patients: results from the IMPACC study. *EBioMedicine* **83**, 104208 (2022).
- Huang, C. et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet* **397**, 220–232 (2021).
- Zhang, H. et al. Data-driven identification of post-acute SARS-CoV-2 infection subphenotypes. *Nat. Med.* **29**, 226–235 (2023).
- Reese, J.T. et al. Generalizable long COVID subtypes: findings from the NIH N3C and RECOVER Programs. *medRxiv* <https://www.medrxiv.org/content/10.1101/2022.05.24.22275398v1> (2022).
- Huang, Y. et al. COVID symptoms, symptom clusters, and predictors for becoming a long-hauler: looking for clarity in the haze of the pandemic. *medRxiv* <https://www.medrxiv.org/content/10.1101/2021.03.03.21252086v1> (2021).
- Sudre, C. H. et al. Attributes and predictors of long COVID. *Nat. Med.* **27**, 626–631 (2021).
- Whitaker, M. et al. Persistent COVID-19 symptoms in a community study of 606,434 people in England. *Nat. Commun.* **13**, 1957 (2022).
- Subramanian, A. et al. Symptoms and risk factors for long COVID in non-hospitalized adults. *Nat. Med.* **28**, 1706–1714 (2022).
- Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK: 3 November 2022. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/prevalenceofongoingsymptomsfollowingcoronaviruscovid19infectionintheuk/3november2022#:~:text=An%20estimated%202.1%20million%20people,of%201%20October%202022%3B%20these> (2022).
- Tleyjeh, I. M. et al. Prevalence and predictors of Post-Acute COVID-19 Syndrome (PACS) after hospital discharge: a cohort study with 4 months median follow-up. *PLoS One* **16**, e0260568 (2021).
- Ortona, E. & Malorni, W. Long COVID: to investigate immunological mechanisms and sex/gender related aspects as fundamental steps for tailored therapy. *Eur. Respir. J.* **59**, 2102245 (2022).
- Bai, F. et al. Female gender is associated with long COVID syndrome: a prospective cohort study. *Clin. Microbiol. Infect.* **28**, 611.e9–e16 (2022).
- Gubbels Bupp, M. R., Potluri, T., Fink, A. L. & Klein, S. L. The confluence of sex hormones and aging on immunity. *Front Immunol.* **9**, 1269 (2018).
- Chan Sui Ko, A. et al. Number of initial symptoms is more related to long COVID-19 than acute severity of infection: a prospective cohort of hospitalized patients. *Int. J. Infect. Dis.* **118**, 220–223 (2022).
- Townsend, L. et al. Persistent fatigue following SARS-CoV-2 infection is common and independent of severity of initial infection. *PLoS One* **15**, e0240784 (2020).
- Pérez-González, A. et al. Long COVID in hospitalized and non-hospitalized patients in a large cohort in Northwest Spain, a prospective cohort study. *Sci. Rep.* **12**, 3369 (2022).
- Bell, M. L. et al. Post-acute sequelae of COVID-19 in a non-hospitalized cohort: results from the Arizona CoVHORT. *PLoS One* **16**, e0254347 (2021).
- Hedberg, P. et al. Post COVID-19 condition diagnosis: a population-based cohort study of occurrence, associated factors, and health-care use by severity of acute infection. *J. Intern. Med.* **293**, 246–258 (2023).
- Mosier J. M. et al. A target for increased mortality risk in critically ill patients: the concept of perpetuity. *J. Clin. Med.* **10**, 3971 (2021).
- Su, Y. et al. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell* **185**, 881–895.e20 (2022).
- García-Abellán, J. et al. Antibody response to SARS-CoV-2 is associated with long-term clinical outcome in patients with COVID-19: a longitudinal study. *J. Clin. Immunol.* **41**, 1490–1501 (2021).

29. Jia, X. et al. Anti-nucleocapsid antibody levels and pulmonary comorbid conditions are linked to post-COVID-19 syndrome. *JCI Insight* **7**, e156713 (2022).
30. Thompson, R. C. et al. Molecular states during acute COVID-19 reveal distinct etiologies of long-term sequelae. *Nat. Med.* **29**, 236–246 (2023).
31. Arnold, D. et al. Are vaccines safe in patients with Long COVID? A prospective observational study. *medRxiv* <https://www.medrxiv.org/content/10.1101/2021.03.11.21253225v3> (2021).
32. Boglione, L. et al. Risk factors and incidence of long-COVID syndrome in hospitalized patients: does remdesivir have a protective effect? *QJM* **114**, 865–871 (2022).
33. Xie Y., Choi T. & Al-Ally Z. Nirmatrelvir and the risk of post-acute sequelae of COVID-19. *medRxiv* <https://www.medrxiv.org/content/10.1101/2022.11.03.22281783v1> (2022).
34. Seeßle, J. et al. Persistent symptoms in adult patients 1 year after coronavirus disease 2019 (COVID-19): a prospective cohort study. *Clin. Infect. Dis.* **74**, 1191–1198 (2022).
35. L’Huillier, A. G. et al. Autoantibodies against apolipoprotein A-1 after COVID-19 predict symptoms persistence. *Eur. J. Clin. Invest.* **52**, e13818 (2022).
36. Chang, S. E. et al. New-onset IgG autoantibodies in hospitalized patients with COVID-19. *Nat. Commun.* **12**, 5417 (2021).
37. Zhang, Q., Bastard, P., Cobat, A. & Casanova, J. L. Human genetic and immunological determinants of critical COVID-19 pneumonia. *Nature* **603**, 587–598 (2022).
38. Ajaz, S. et al. Mitochondrial metabolic manipulation by SARS-CoV-2 in peripheral blood mononuclear cells of patients with COVID-19. *Am. J. Physiol. Cell Physiol.* **320**, C57–c65 (2021).
39. Kukla, M. et al. Fetuin-A deficiency but not pentraxin 3, FGF-21, or irisin, predisposes to more serious COVID-19 course. *Biomolecules* **11**, 1422 (2021).
40. Sun, H., Sherrier, M. & Li, H. Skeletal muscle and bone - emerging targets of fibroblast growth factor-21. *Front. Physiol.* **12**, 625287 (2021).
41. Keipert, S. et al. Skeletal muscle mitochondrial uncoupling drives endocrine cross-talk through the induction of FGF21 as a myokine. *Am. J. Physiol. Endocrinol. Metab.* **306**, E469–E482 (2014).
42. Jung, H. W. et al. Association between serum FGF21 level and sarcopenia in older adults. *Bone* **145**, 115877 (2021).
43. Marchesini, G., Forlani, G., Zoli, M., Vannini, P. & Pisi, E. Muscle protein breakdown in uncontrolled diabetes as assessed by urinary 3-methylhistidine excretion. *Diabetologia* **23**, 456–458 (1982).
44. Domingo, J. C. et al. Are circulating fibroblast growth factor 21 and N-terminal prohormone of brain natriuretic peptide promising novel biomarkers in myalgic encephalomyelitis/chronic fatigue syndrome? *Antioxid. Redox Signal.* **34**, 1420–1427 (2021).
45. Tarasenko, T. N., Cusmano-Ozog, K. & McGuire, P. J. Tissue acylcarnitine status in a mouse model of mitochondrial β -oxidation deficiency during metabolic decompensation due to influenza virus infection. *Mol. Genet. Metab.* **125**, 144–152 (2018).
46. Makrecka-Kuka, M. et al. Plasma acylcarnitine concentrations reflect the acylcarnitine profile in cardiac tissues. *Sci. Rep.* **7**, 17528 (2017).
47. Jinushi, R. et al. Comparison of serum acylcarnitine levels in patients with myalgic encephalomyelitis/chronic fatigue syndrome and healthy controls: a systematic review and meta-analysis. *J. Transl. Med.* **21**, 398 (2023).
48. Fluge, Ø. et al. Metabolic profiling indicates impaired pyruvate dehydrogenase function in myalgic encephalopathy/chronic fatigue syndrome. *JCI Insight* **1**, e89376 (2016).
49. Kawczynska-Drozd, A. et al. Ghrelin inhibits vascular superoxide production in spontaneously hypertensive rats. *Am. J. Hypertens.* **19**, 764–767 (2006).
50. Antonelli, M., Pujol, J. C., Spector, T. D., Ourselin, S. & Steves, C. J. Risk of long COVID associated with delta versus omicron variants of SARS-CoV-2. *Lancet* **399**, 2263–2264 (2022).
51. Rass, V. et al. Neurological outcome and quality of life 3 months after COVID-19: a prospective observational cohort study. *Eur. J. Neurol.* **28**, 3348–3359 (2021).
52. Qu, G. et al. Health-related quality of life of COVID-19 patients after discharge: a multicenter follow-up study. *J. Clin. Nurs.* **30**, 1742–1750 (2021).
53. Visvabharathy, L. et al. Neuro-COVID long-haulers exhibit broad dysfunction in T cell memory generation and responses to vaccination. *medRxiv* <https://www.medrxiv.org/content/10.1101/2021.08.08.21261763v1.full> (2021).
54. Son, K. et al. Circulating anti-nuclear autoantibodies in COVID-19 survivors predict long-COVID symptoms. *Eur. Respir. J.* **61**, 2200970 (2022).
55. Stein, S. R. et al. SARS-CoV-2 infection and persistence in the human body and brain at autopsy. *Nature* **612**, 758–763 (2022).
56. von Elm, E. et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* **370**, 1453–1457 (2007).
57. IMPACC Manuscript Writing Team; IMPACC Network Steering Committee. Immunophenotyping assessment in a COVID-19 cohort (IMPACC): a prospective longitudinal study. *Sci. Immunol.* **6**, eabf3733 (2021).
58. Symptoms of COVID-19. <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>. (2022).
59. Devlin, N. J. & Brooks, R. EQ-5D and the EuroQol group: past, present and future. *Appl. Health Econ. Health Policy* **15**, 127–137 (2017).
60. Lai, J. S., Wagner, L. I., Jacobsen, P. B. & Cella, D. Self-reported cognitive concerns and abilities: two sides of one coin? *Psychooncology* **23**, 1133–1141 (2014).
61. Salsman, J. M. et al. Assessing psychological well-being: self-report instruments for the NIH Toolbox. *Qual. Life Res.* **23**, 205–215 (2014).
62. Lai, J. S., Garcia, S. F., Salsman, J. M., Rosenbloom, S. & Cella, D. The psychosocial impact of cancer: evidence in support of independent general positive and negative components. *Qual. Life Res.* **21**, 195–207 (2012).
63. National Institutes of Health. Patient-Reported Outcomes Measurement Information System (PROMIS®). January 29, 2019 2019. <https://commonfund.nih.gov/promis/index> (accessed 5 May 2021 2021).
64. Hays, R. D., Bjorner, J. B., Revicki, D. A., Spritzer, K. L. & Cella, D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Qual. Life Res.* **18**, 873–880 (2009).
65. Choi, S. W., Victorson, D. E., Yount, S., Anton, S. & Cella, D. Development of a conceptual framework and calibrated item banks to measure patient-reported dyspnea severity and related functional limitations. *Value Health* **14**, 291–306 (2011).
66. Yount, S. E. et al. Brief, valid measures of dyspnea and related functional limitations in chronic obstructive pulmonary disease (COPD). *Value Health* **14**, 307–315 (2011).
67. Irwin, D. E. et al. Correlation of PROMIS scales and clinical measures among chronic obstructive pulmonary disease patients with and without exacerbations. *Qual. Life Res.* **24**, 999–1009 (2015).
68. PROMIS. Dyspnea scoring manual. https://staging.healthmeasures.net/images/PROMIS/manuals/Scoring_Manuals/PROMIS_Dyspnea_Scoring_Manual.pdf. (2021).
69. CDC’s Influenza SARS-CoV-2 multiplex assay. <https://www.cdc.gov/coronavirus/2019-ncov/lab/multiplex.html> (2022).
70. Amanat, F. et al. A serological assay to detect SARS-CoV-2 seroconversion in humans. *medRxiv* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7239062/> (2020).

71. Team, I. M. W. & Committee, I. N. S. Immunophenotyping assessment in a COVID-19 cohort (IMPACC): a prospective longitudinal study. *Sci. Immunol.* **6**, eabf3733 (2021).
72. Bacharier, L. B. et al. Longitudinal phenotypes of respiratory health in a high-risk urban birth cohort. *Am. J. Respir. Crit. Care Med.* **199**, 71–82 (2019).
73. Hetherington, E., Plamondon, A. & Williamson, T. Trajectory modeling with latent groups: potentials and pitfalls. *Curr. Epidemiol. Rep.* **7**, 171–178 (2020).
74. Herle, M. et al. Identifying typical trajectories in longitudinal data: modelling strategies and interpretations. *Eur. J. Epidemiol.* **35**, 205–222 (2020).

Acknowledgements

The study was funded by supplements to project grants and cooperative agreements awarded by the National Institute of Allergy and Infectious Disease (NIAID), including 5R01AI135803-03 (D.B.C., F.K.), 5U19AI118608-04 (O.L., H.S., L.R.B., J.D.A.), 5U19AI128910-04 (E.H., C.B.C., R.P.S., G.A.M.), 4U19AI090023-11 (B.P., N.R., S.B.), 4U19AI118610-06 (V.S., A.F.S., F.K., H.V.B., S.K.S.), R01AI145835-01A1S1 (W.B.M., C.L.H.), 5U19AI062629-17 (J.P.M., N.I.A.H.), 5U19AI057229-17 (M.M.D., K.C.N., H.M.), 5U19AI125357-05 (M.K., C.B.), 5U19AI128913-03 (J.S., F.R.), 3U19AI077439-13 (D.E., C.C., C.R.L., W.E.), 5U54AI142766-03 (M.A.A., S.C.B.), 5R01AI104870-07 (L.E., E.M.), 3U19AI089992-09 (D.H., R.R.M., A.C.S.), and 5R01AI132774-03 (M.C.A.). NIAID project scientists participated collaboratively in study design, data analyses, interpretation, and writing of the report. The manuscript content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author contributions

A.O., P.B., M.C.A., N.R., A.A., L.B., C.S.C., G.M., C.B.C., V.S., C.H., K.C.N., J.S., A.S., D.C., N.I.A.H., C.B., S.C.B., E.M., and V.S.M. were involved in the study concept and design. A.O., C.M., N.D.J., H.V.B., M.C.A., D.E., S.L., F.K.R., J.D.-A., P.B., and N.R. contributed to the study methodology. C.M., N.D.J., H.V.B., S.L., F.K.R., L.B.R., S.K.S., C.R.L., W.E., S.E.B., H.T.M. were involved in data curation. L.B., C.S.C., G.M., C.B.C., N.R., V.S., C.H., K.C.N., J.M., A.S., D.C., N.I.A.H., C.B., S.C.B., and E.M. contributed to the investigation. C.M., N.D.J., H.V.B., S.L., A.O., J.-D.A., J.M., and V.S.M. provided software. A.O., C.M., N.D.J., H.V.B., M.C.A., D.E., S.L., F.K.R., J.Q., C.S., R.D., and C.P.S. were involved in formal analysis. A.O., C.M., N.D.J., H.V.B., M.C.A., D.E., S.L., and F.K.R. were involved in data validation, visualization, or replication of analyses. A.O., C.M., M.C.A., J.S., N.R., and P.B. drafted the manuscript. A.O., J.S., N.D.J., C.M., C.S., C.B.C., M.K., L.B., A.S., F.K.R., H.V.B., D.E., S.L., A.F.S., V.S., D.H., R.M., S.K., O.L., C.B., E.H., D.J.E., B.P.U., K.C.N., M.D., C.H., W.M., N.I.A.H., J.M., M.A.A., S.C.B., D.C., F.K.H., L.I.E., E.M., G.M., R.S., J.D.-A., B.P.E., A.A., E.R., M.C.A., P.B., H.S., L.N.G., N.R., and C.P.S. critically reviewed and revised the manuscript. P.B., A.A., A.O., S.K., B.P.E., N.R., E.R., J.-D.-A., and J.M. were involved in project administration, and A.O., J.S., N.D.J., C.M., C.S.C., C.B.C., M.K., L.B., A.S., F.K.R., H.V.B., D.E., S.L., A.F.S., V.S., D.H., R.M., S.K., O.L., C.B., E.H., D.E., B.P.U., K.C.N., M.D., C.H., W.M., N.I.A.H., J.M., M.A.A., S.C.B., D.C., F.K.H., L.I.E., E.M., G.M., R.S., J.D.-A., B.P.E., A.A., E.R., M.C.A., P.B., and N.R. provided study resources. O.L., R.S., E.H., A.F.S., W.M., M.D., B.P.U., E.F., D.J.E., R.M., D.H., F.K.H., J.M., M.K., M.A.A., and L.I.E. obtained funding. P.B., A.O., M.C.A., and N.R. provided supervision.

Competing interests

M.C.A. has received grant support from NIH-R01AI32774 for this project funded through this R01 and travel fees from NIAID for travel to the American Thoracic Society 2022 to present data related to this study. M.A.A. has received funding from NIH, NIAID-5U54AI142766-03 through institution. L.R.B. has received grant support from NIH NIAID, through institution. P.M.B. and A.D.A. federal employees serving as a project scientist for this project but had no role in funding decisions or oversight

of relevant grants. B.P. has received funding from NIH/NIAID. C.B.C. has received funding from NIAID with payments to institution (Drexel University), Bill & Melinda Gates Foundation for COVID-19 work paid to institution, consulting fees from bioMerieux on clinical biomarkers, serves as DSMB, Advisory board for Convalescent Plasma COVID-19 study for the National Heart, Lung and Blood Institute (NHLBI), and is acting Leadership as President Board of Directors for the National Foundation of Emergency Medicine (NFEM), a non-profit supporting emergency medicine research and researchers. C.C. has received funding from NHLBI, grants from Bayer, Roche-Genentech, Quantum Leap Healthcare Collaborative, and consulting fees from Vasomune, GenLife Sciences, Cellenkos, Janssen. L.E. has received grant funding from NIH R01AI104870-S1. D.E. has received NIH grants awarded to institution (UCSF). C.L.H. has received funding from NIH and the American Lung Association, travel support from Stanford, Harvard, Critical Care Clinical Trialists, Critical Care Reviews, and the University of Michigan, serves as DSMB Quantum Health for iSPY COVID, and paid participation as member, Board of Directors for American Thoracic Society. J.P.M. has received funding from NIH Grant # 3U19AI0626629-17S2. S.H.K. has received paid consulting fees from Peraton for personal consulting related to Immport data repository. M.K. has received grants support from NIAID through institution (University of Arizona). F.K.R. has received funding from NIAID Collaborative Influenza Vaccine Innovation Centers (CIVIC) contract 75N93019C00051, JPB Foundation and the Open Philanthropy Project (research grant 2020-215611, 5384), National Cancer Institute, NIH contract no.75N91019D00024, Task order no. 75N91020F00003, research funding from Pfizer for development of animal models for SARS-CoV-2, consulting fees from Pfizer, Seqirus, Avimex, Third Rock Ventures, paid lecture fees, and patents filed at the Icahn School of Medicine at Mount Sinai relating to SARS-CoV-2 serological assays (the “Serology Assays” and NDV-based SARS-CoV-2 vaccines which list F.K.R. as co-inventor). G.A.M. has received consulting fees from Gilead. E.M. has received funding from the NIH IMPACC R01AI104870-S1, grants from Babson Diagnostics, K0826-1616-11 through Institution Dell Medical School at UT Austin, paid speaker fees from MS Association of America, and serves as DSMB for Advisory boards of Genentech, Horizon, Teva and Viela Bio. W.B.M. has received funding from NIH NIAID R01AI14583. R.R.M. has contracts and grants for IMPACC study from NIAID AI 089992 aid and a Leadership Councilor role 2018-2021 for Society of Leukocyte Biology. K.C.N. has research funding from the National Heart, Lung, and Blood Institute (NHLBI), National Institute of Environmental Health Sciences (NIEHS), Food Allergy Research & Education (FARE), Director of the World Allergy Organization center of Excellence for Stanford and National Institute of Allergy and Infectious Diseases (NIAID) through institution, paid participation for service on Data Safety Monitoring Board of Director from World Allergy Organization Center of Excellence for Stanford, earns stocks as co-founder at Seed Health, IgGenix, ClostraBio, ImmunelD and financial interests as advisor and co-founder for Cour Pharma, Before Brands, Alladapt, and Latitude, is a national scientific committee member for Network ITN, NIH clinical research centers, and has listed Patents: Mixed allergen composition and methods for using the same Granulocyte-based methods for detecting and monitoring immune system disorders, licensee: (Alladapt and Before Brands no: US15/048,609); Application number: US12/610,94 Methods and Assays for Detecting and Quantifying Pure Subpopulations of White blood cells in immune system disorders. O.L. has received NIH/NIAID grants through institution for Human Immunology Project Consortium Funding (U19) 1-U19-AI118608-01A1 as PI Role and support as a speaker for presentation regarding the Coronavirus pandemic from Midsized Bank Coalition of America (MBCA) and Moody’s Analytics. E.F.R. has received grants supported by NIAID U19AI12891303. N.G.R. has research grants from NIH, Pfizer, Merck, Sanofi, Quidel, and Lilly, and serves on safety committees for ICON and EMMES and the advisory boards of Moderna and Sanofi. Her institution has also received funding from NIH to conduct clinical trials of COVID-19

vaccines. J.S. received a grant from NIAID U19 for study implementation. A.C.S. has financial support from NIH U19 AI089992, NIH K24 AG042489. V.S. filed Patents at the Icahn School of Medicine at Mount Sinai relating to SARS-CoV-2 serological assays (the “Serology Assays” and NDV-based SARS-CoV-2 vaccines, which list V.S. as co-inventor. H.B. has received funding from NIH (Dengue Human Immunology Project Consortium - Mount Sinai IMPACC COVID-19 Cores), U19 AI118610 S1, NIH, CEIRR,75N93021C00014. SEB All funding sources are from NIH U19 AI090023-1S1, paid honoraria for serving on SAB for NIH NIAID P01AI174856-01, travel support from the University of Manitoba, and received TAK-242 from TAKEDA for pre-clinical testing. H.S. has received funding from 3 U19 AI 118608-05S3 through institution and travel support from NIAID. L.G. has received funding from Sean N. Parker Center for Allergy & Asthma Research, grants from Pfizer, and consulting fees from UnitedHealth Group. H.M. has received funding from NIH grant 2U19AI057229. R.D. has received grant support from UCSF COVID-19 Immunophenotyping Clinical Study and Core Laboratories (Grant number U19AI077439). V.S.M. has grants paid consulting as Vice Chair for EveryLife Foundation, Advisory board for Gemini labs, and stocks at My Own Med, Inc., d/b/a Respond Health. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-44090-5>.

Correspondence and requests for materials should be addressed to Nadine Rouphael.

Peer review information *Nature Communications* thanks Yamilé López-Hernández, Alvin D Jeffery, Honghan Wu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <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 license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Al Ozonoff^{1,22}, Naresh Doni Jayavelu^{2,22}, Shanshan Liu^{1,22}, Esther Melamed^{3,22}, Carly E. Milliren^{1,22}, Jingjing Qi^{4,22}, Linda N. Geng^{5,22}, Grace A. McComsey⁶, Charles B. Cairns⁷, Lindsey R. Baden⁸, Joanna Schaefer⁹, Albert C. Shaw¹⁰, Hady Samaha¹¹, Vicki Seyfert-Margolis¹², Florian Krammer¹³, Lindsey B. Rosen¹³, Hanno Steen¹⁴, Caitlin Syphurs¹, Ravi Dandekar¹⁴, Casey P. Shannon¹⁵, Rafick P. Sekaly¹⁶, Lauren I. R. Ehrlich¹⁷, David B. Corry¹⁶, Farrah Kheradmand¹⁶, Mark A. Atkinson¹⁷, Scott C. Brakenridge¹⁷, Nelson I. Agudelo Higuera¹⁸, Jordan P. Metcalf¹⁸, Catherine L. Hough¹⁹, William B. Messer¹⁹, Bali Pulendran²⁰, Kari C. Nadeau²⁰, Mark M. Davis²⁰, Ana Fernandez Sesma²¹, Viviana Simon²¹, Harm van Bakel²¹, Seunghee Kim-Schulze²¹, David A. Hafler²¹, Ofer Levy²², Monica Kraft²⁰, Chris Bime²⁰, Elias K. Haddad⁷, Carolyn S. Calfee¹⁴, David J. Erle¹⁴, Charles R. Langley¹⁴, Walter Eckalbar¹⁴, Steven E. Bosinger¹¹, IMPACC Network*, Bjoern Peters²¹, Steven H. Kleinstein¹⁰, Elaine F. Reed⁹, Alison D. Augustine¹³, Joann Diray-Arce¹, Holden T. Maecker⁵, Matthew C. Altman^{2,23}, Ruth R. Montgomery^{10,23}, Patrice M. Becker^{13,23} & Nadine Rouphael^{11,23} ✉

¹Clinical & Data Coordinating Center (CDCC), Precision Vaccines Program, Boston Children’s Hospital, Boston, MA, USA. ²Benaroya Research Institute, Seattle, WA, USA. ³The University of Texas at Austin, Austin, TX, USA. ⁴Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁵Stanford University, Stanford, CA, USA. ⁶Case Western Reserve University and University Hospitals of Cleveland, Cleveland, OH, USA. ⁷Drexel University/Tower Health Hospital, Philadelphia, PA, USA. ⁸Boston Clinical Site: Precision Vaccines Program, Boston Children’s Hospital, Brigham and Women’s Hospital, and Harvard Medical School, Boston, MA, USA. ⁹David Geffen School of Medicine at the University of California Los Angeles, Los Angeles, CA, USA. ¹⁰Yale School of Medicine, and Yale School of Public Health, New Haven, CT, USA. ¹¹Emory University, Atlanta, GA, USA. ¹²MyOwnMed, Inc, Bethesda, MD, USA. ¹³National Institute of Allergy and Infectious Diseases/National Institutes of Health, Bethesda, MD, USA. ¹⁴University of California San Francisco School of Medicine, San Francisco, CA, USA. ¹⁵Centre for Heart Lung Innovation, Providence Research, St. Paul’s Hospital, and the PROOF Centre of Excellence, Vancouver, BC, Canada. ¹⁶Baylor College of Medicine, and the Center for Translational Research on Inflammatory Diseases, Michael E. DeBakey VA Medical Center, Houston, TX, USA. ¹⁷University of Florida/University of South Florida, Tampa, FL, USA. ¹⁸Oklahoma University Health Sciences Center, Oklahoma City, OK, USA. ¹⁹Oregon Health & Science University, Portland, OR, USA. ²⁰University of Arizona, Tucson, AZ, USA. ²¹La Jolla Institute for Immunology, La Jolla, CA, USA. ²²These authors contributed equally: Al Ozonoff, Naresh Doni Jayavelu, Shanshan Liu, Esther Melamed, Carly E. Milliren, Jingjing Qi, Linda N. Geng. ²³These authors jointly supervised this work: Matthew C. Altman, Ruth R. Montgomery, Patrice M. Becker, Nadine Rouphael. ✉ e-mail: nroupha@emory.edu

IMPACC Network

IMPACC Steering Committee Al Ozonoff ^{1,22}, Joann Diray-Arce¹, Matthew C. Altman ^{2,23}, Lauren I. R. Ehrlich ³, Esther Melamed^{3,22}, Ana Fernandez Sesma ⁴, Viviana Simon ⁴, Bali Pulendran ⁵, Kari C. Nadeau ⁵, Mark M. Davis ⁵, Grace A. McComsey⁶, Rafick P. Sekaly ⁶, Charles B. Cairns ⁷, Elias K. Haddad ⁷, Lindsey R. Baden⁸, Ofer Levy ⁸, Joanna Schaeenman⁹, Elaine F. Reed⁹, Albert C. Shaw¹⁰, David A. Hafler ¹⁰, Ruth R. Montgomery ^{10,23}, Steven H. Kleinstein ¹⁰, Nadine Rouphael ^{11,23}  , Patrice M. Becker ^{13,23}, Alison D. Augustine¹³, Carolyn S. Calfee¹⁴, David J. Erle ¹⁴, David B. Corry¹⁶, Farrah Kheradmand ¹⁶, Mark A. Atkinson¹⁷, Scott C. Brakenridge¹⁷, Nelson I. Agudelo Higueta¹⁸, Jordan P. Metcalf¹⁸, Catherine L. Hough¹⁹, William B. Messer ¹⁹, Monica Kraft²⁰, Chris Bime²⁰ & Bjoern Peters ²¹

Clinical & Data Coordinating Center (CDCC) Al Ozonoff ^{1,22}, Carly E. Milliren^{1,22}, Joann Diray-Arce¹, Caitlin Syphurs¹, Kerry McEnaney¹, Brenda Barton¹, Claudia Lentucci¹, Mehmet Saluvan¹, Ana C. Chang¹, Annmarie Hoch¹, Marisa Albert¹, Tanzia Shaheen¹, Alvin T. Kho¹, Shanshan Liu^{1,22}, Sanya Thomas¹, Jing Chen¹, Maimouna D. Murphy¹, Mitchell Cooney¹, Arash Nemati Hayati¹, Robert Bryant¹ & James Abraham¹

IMPACC Data Analysis Group Al Ozonoff ^{1,22}, Joann Diray-Arce¹, Naresh Doni Jayavelu^{2,22}, Matthew C. Altman ^{2,23}, Scott Presnell², Tomasz Jancsyk², Cole Maguire³, Jingjing Qi^{4,22}, Brian Lee⁴, Slim Fourati⁶, Charles B. Cairns ⁷, Denise A. Esserman¹⁰, Leying Guan¹⁰, Steven H. Kleinstein ¹⁰, Jeremy Gygi¹⁰, Shrikant Pawar¹⁰, Anderson Brito¹⁰, Gabriela K. Fragiadakis¹⁴, Ravi Patel¹⁴, Casey P. Shannon ¹⁵, Scott J. Tebbutt¹⁵, Bjoern Peters ²¹, James A. Overton²¹, Randi Vita²¹ & Kerstin Westendorf²¹

IMPACC Site Investigators Lauren I. R. Ehrlich ³, Esther Melamed^{3,22}, Rama V. Thyagarajan³, Justin F. Rousseau³, Dennis Wylie³, Todd A. Triplett³, Erna Kojic⁴, Viviana Simon ⁴, Kari C. Nadeau ⁵, Sharon Chinthrajah⁵, Neera Ahuja⁵, Angela J. Rogers⁵, Maja Artandi⁵, Linda N. Geng^{5,22}, Grace A. McComsey⁶, George Yendewa⁶, Charles B. Cairns ⁷, Debra L. Powell⁷, James N. Kim⁷, Brent Simmons⁷, I. Michael Goonewardene⁷, Cecilia M. Smith⁷, Mark Martens⁷, Lindsey R. Baden⁸, Amy C. Sherman⁸, Stephen R. Walsh⁸, Nicolas C. Issa⁸, Joanna Schaeenman⁹, Ramin Salehi-Rad⁹, Albert C. Shaw¹⁰, Charles Dela Cruz¹⁰, Shelli Farhadian¹⁰, Akiko Iwasaki¹⁰, Albert I. Ko¹⁰, Nadine Rouphael ^{11,23}  , Evan J. Anderson¹¹, Aneesh K. Mehta¹¹, Jonathan E. Sevransky¹¹, Vicki Seyfert-Margolis¹², Carolyn S. Calfee¹⁴, David J. Erle ¹⁴, Aleksandra Leligdowicz¹⁴, Michael A. Matthay¹⁴, Jonathan P. Singer¹⁴, Kirsten N. Kangelaris¹⁴, Carolyn M. Hendrickson¹⁴, Matthew F. Krummel¹⁴, Charles R. Langelier ¹⁴, Prescott G. Woodruff¹⁴, David B. Corry¹⁶, Farrah Kheradmand ¹⁶, Scott C. Brakenridge¹⁷, Matthew L. Anderson¹⁷, Faheem W. Guirgis¹⁷, Nelson I. Agudelo Higueta¹⁸, Jordan P. Metcalf¹⁸, Douglas A. Drevets¹⁸, Brent R. Brown¹⁸, William B. Messer ¹⁹, Sarah A. R. Siegel¹⁹, Zhengchun Lu¹⁹, Monica Kraft²⁰, Chris Bime²⁰, Jarrod Mosier²⁰ & Hiroki Kimura²⁰

IMPACC Core Laboratory Joann Diray-Arce¹, Matthew C. Altman ^{2,23}, Bernard Khor², Florian Krammer ⁴, Harm van Bakel ⁴, Adeeb Rahman⁴, Daniel Stadlbauer⁴, Jayeeta Dutta⁴, Seunghee Kim-Schulze⁴, Ana Silvia Gonzalez-Reiche⁴, Adriana van de Guchte⁴, Juan Manuel Carreño⁴, Gagandeep Singh⁴, Ariel Raskin⁴, Johnstone Tcheou⁴, Dominika Bielak⁴, Hisaaki Kawabata⁴, Brian Lee⁴, Hui Xie⁴, Geoffrey Kelly⁴, Manishkumar Patel⁴, Kai Nie⁴, Temima Yellin⁴, Miriam Fried⁴, Leeba Sullivan⁴, Sara Morris⁴, Holden T. Maecker ⁵, Scott Sieg⁶, Hanno Steen ⁸, Patrick van Zalm⁸, Benoit Fatou⁸, Kevin Mendez⁸, Jessica Lasky-Su⁸, Scott R. Hutton⁸, Greg Michelotti⁸, Kari Wong⁸, Meenakshi Jha⁸, Arthur Viode⁸, Naama Kanarek⁸, Boryana Petrova⁸, Albert C. Shaw¹⁰, Yujiao Zhao¹⁰, Charles Dela Cruz¹⁰, Ruth R. Montgomery ^{10,23}, Steven E. Bosinger ¹¹, Arun K. Boddapati¹¹, Greg K. Tharp¹¹, Kathryn L. Pellegrini¹¹, Elizabeth Beagle¹¹, David Cowan¹¹, Sydney Hamilton¹¹, Susan Pereira Ribeiro¹¹, Thomas Hodder¹¹, Lindsey B. Rosen¹³, Serena Lee¹³, Charles R. Langelier ¹⁴, Michael R. Wilson¹⁴, Ravi Dandekar¹⁴, Bonny Alvarenga¹⁴, Jayant Rajan¹⁴, Walter Eckalbar¹⁴, Andrew W. Schroeder¹⁴, Alexandra Tsitsiklis¹⁴, Eran Mick¹⁴, Yanedth Sanchez Guerrero¹⁴, Christina Love¹⁴, Lenka Maliskova¹⁴ & Michael Adkisson¹⁴

IMPACC Clinical Study Team Cole Maguire³, Nadia Siles³, Janelle Geltman⁴, Kerin Hurley³, Miti Saksena⁴, Deena Altman⁴, Erna Kojic⁴, Komal Srivastava⁴, Lily Q. Eaker⁴, Maria C. Bermúdez-González⁴, Katherine F. Beach⁴, Levy A. Sominsky⁴, Arman R. Azad⁴, Lubbertus C. F. Mulder⁴, Giulio Kleiner⁴, Alexandra S. Lee⁵, Evan Do⁵, Andrea Fernandes⁵, Monali Manohar⁵, Thomas Hagan⁵, Catherine A. Blish⁵, Hena Naz Din⁵, Jonasel Roque⁵, Samuel Yang⁵, Natalia Sigal⁵, Iris Chang⁵, Heather Tribut⁶, Paul Harris⁶, Mary Consolo⁶, Jennifer Connors⁷, Mariana Bernui⁷, Michele A. Kutzler⁷, Carolyn Edwards⁷, Edward Lee⁷, Edward Lin⁷, Brett Croen⁷, Nicholas C. Semenza⁷, Brandon Rogowski⁷, Nataliya Melnyk⁷, Kyra Woloszczuk⁷, Gina Cusimano⁷, Mathew R. Bell⁷, Sara Furukawa⁷, Renee McLin⁷, Pamela Schearer⁷, Julie Sheidy⁷, George P. Tegos⁷

Crystal Nagle⁷, Ofer Levy⁸, Kinga Smolen⁸, Michael Desjardins⁸, Simon van Haren⁸, Xhoi Mitre⁸, Jessica Cauley⁸, Xiaofang Li⁸, Alexandra Tong⁸, Bethany Evans⁸, Christina Montesano⁸, Jose Humberto Licona⁸, Jonathan Krauss⁸, Jun Bai Park Chang⁸, Natalie Izaguirre⁸, Rebecca Rooks⁸, David Elashoff⁹, Jenny Brook⁹, Estefania Ramirez-Sanchez⁹, Megan Llamas⁹, Adreanne Rivera⁹, Claudia Perdomo⁹, Dawn C. Ward⁹, Clara E. Magyar⁹, Jennifer A. Fulcher⁹, Harry C. Pickering⁹, Subha Sen⁹, Omkar Chaudhary¹⁰, Andreas Coppi¹⁰, John Fournier¹⁰, Subhasis Mohanty¹⁰, M. Catherine Muenker¹⁰, Allison Nelson¹⁰, Khadir Raddassi¹⁰, Michael Rainone¹⁰, William E. Ruff¹⁰, Syim Salahuddin¹⁰, Wade L. Schulz¹⁰, Pavithra Vijayakumar¹⁰, Haowei Wang¹⁰, Elsie Wunder Jr.¹⁰, H. Patrick Young¹⁰, Yujiao Zhao¹⁰, Jessica Rothman¹⁰, Anna Konstorum¹⁰, Ernie Chen¹⁰, Chris Cotsapas¹⁰, Nathan D. Grubaugh¹⁰, Xiaomei Wang¹⁰, Leqi Xu¹⁰, Hiromitsu Asashima¹⁰, Laurel Bristow¹¹, Laila Hussaini¹¹, Kieffer Hellmeister¹¹, Hady Samaha¹¹, Sonia Tandon Wimalasena¹¹, Andrew Cheng¹¹, Christine Spainhour¹¹, Erin M. Scherer¹¹, Brandi Johnson¹¹, Amer Bechnak¹¹, Caroline R. Ciric¹¹, Lauren Hewitt¹¹, Erin Carter¹¹, Nina Mcnair¹¹, Bernadine Panganiban¹¹, Christopher Huerta¹¹, Jacob Usher¹¹, Vicki Seyfert-Margolis¹², Tatyana Vaysman¹³, Steven M. Holland¹³, Yumiko Abe-Jones¹⁴, Saurabh Asthana¹⁴, Alexander Beagle¹⁴, Sharvari Bhide¹⁴, Sidney A. Carrillo¹⁴, Suzanna Chak¹⁴, Gabriela K. Fragiadakis¹⁴, Rajani Ghale¹⁴, Ana Gonzalez¹⁴, Alejandra Jauregui¹⁴, Norman Jones¹⁴, Tasha Lea¹⁴, Deanna Lee¹⁴, Raphael Lota¹⁴, Jeff Milush¹⁴, Viet Nguyen¹⁴, Logan Pierce¹⁴, Priya A. Prasad¹⁴, Arjun Rao¹⁴, Bushra Samad¹⁴, Cole Shaw¹⁴, Austin Sigman¹⁴, Pratik Sinha¹⁴, Alyssa Ward¹⁴, Andrew Willmore¹⁴, Jenny Zhan¹⁴, Sadeed Rashid¹⁴, Nicklaus Rodriguez¹⁴, Kevin Tang¹⁴, Luz Torres Altamirano¹⁴, Legna Betancourt¹⁴, Cindy Curiel¹⁴, Nicole Sutter¹⁴, Maria Tercero Paz¹⁴, Gayelan Tietje-Ulrich¹⁴, Carolyn Leroux¹⁴, Ravi Patel¹⁴, Neeta Thakur¹⁴, Joshua J. Vasquez¹⁴, Lekshmi Santhosh¹⁴, Li-Zhen Song¹⁶, Ebony Nelson¹⁶, Lyle L. Moldawer¹⁷, Brittany Borresen¹⁷, Brittney Roth-Manning¹⁷, Ricardo F. Ungaro¹⁷, Jordan Oberhaus¹⁷, J. Leland Booth¹⁸, Lauren A. Sinko¹⁸, Amanda Brunton¹⁹, Peter E. Sullivan¹⁹, Matthew Strnad¹⁹, Zoe L. Lyski¹⁹, Felicity J. Coulter¹⁹, Courtney Micheletti¹⁹, Michelle Conway²⁰, Dave Francisco²⁰, Allyson Molzahn²⁰, Heidi Erickson²⁰, Connie Cathleen Wilson²⁰, Ron Schunk²⁰, Bianca Sierra²⁰ & Trina Hughes²⁰