Article

Virus reactivation in acute and long COVID-19

https://doi.org/10.1038/s41586-026-10740-z

Received: 19 November 2024

Accepted: 1 June 2026

Published online: 5 August 2026

Open access

Check for updates

Cole Maguire[1], Jing Chen[2,3], Nadine Rouphael[4], Brinkley A. Morse[1,5], Annmarie Hoch[2], Harry Pickering[6], Hoang Van Phan[7], Abigail Glascock[8], Victoria Chu[7], Ravi Dandekar[7], IMPACC Network*, David Corry[9], Farrah Kheradmand[9], Lindsey R. Baden[10], Rafick-Pierre Sekaly[11], Grace A. McComsey[11], Elias K. Haddad[12], Charles B. Cairns[12], Bali Pulendran[13], Ana Fernandez-Sesma[14], Viviana Simon[14], Jordan P. Metcalf[15], Nelson I. Agudelo Higuita[15], William B. Messer[16], Mark M. Davis[13], Kari C. Nadeau[13], Monica Kraft[17], Chris Bime[17], Joanna Schaenman[6], David Erle[7], Carolyn S. Calfee[7], Mark A. Atkinson[18], Scott C. Brakenridge[18], Lauren I. R. Ehrlich[19], Ruth R. Montgomery[20], Albert Shaw[20], Catherine L. Hough[16], David Hafler[20], Alison D. Augustine[21], Patrice M. Becker[21], Bjoern Peters[22], Al Ozonoff[2], Seunghee Kim-Schulze[14], Florian Krammer[14], Steven E. Bosinger[4], Walter Eckalbar[7], Matthew C. Altman[23], Michael Wilson[7], Leying Guan[24], Steven H. Kleinstein[20], Kinga K. Smolen[25,26], Elaine F. Reed[6], Ofer Levy[25,26,27], Holden Maecker[13], Peter Hunt[7], Hanno Steen[25,26], Joann Diray-Arce[2,26], Charles R. Langelier[7] & Esther Melamed[1]✉

Chronic viral infections are ubiquitous in humans, with individuals carrying multiple viruses that can reactivate during physiological stress, including severe illness[1]. Notably, SARS-CoV-2 infection has been shown to reactivate chronic viruses such as Epstein–Barr virus and cytomegalovirus, yet the full extent, temporal dynamics and immunological impact of viral reactivation in COVID-19 remain incompletely understood[2–7]. Here, leveraging multi-omic longitudinal data from 1,154 hospitalized patients with COVID-19 from the Immunophenotyping Assessment in a COVID-19 Cohort (IMPACC) study, we reveal significant reactivation of *Herpesviridae* and *Anelloviridae* during acute COVID-19, with distinct temporal dynamics for different viruses, and demonstrate that reactivation correlates with disease severity, host immune effects and clinical outcomes. Although our results do not establish causation between virus reactivation and clinical outcomes, we highlight the prevalence of chronic viral reactivation during acute COVID-19 and long COVID. Our findings challenge the prevailing view that chronic viral reactivation is primarily a consequence of immunosuppression, demonstrating that reactivations occur frequently in immunocompetent individuals during severe illness and in association with increased systemic inflammation. Additionally, we demonstrate persistence of viral reactivation in convalescence, and report an association of *Anelloviridae* with long COVID. This study provides immune, transcriptomic and metabolomic signatures of viral reactivation that could inform future strategies to prognosticate and treat acute COVID-19 and long COVID.

Viruses use diverse strategies to enhance their persistence and dissemination, including establishing chronic infection[1,8,9]. This strategy is exemplified by human-infecting viruses, particularly members of *Herpesviridae* and *Anelloviridae* families, which establish lifelong infections in a significant portion of the human population[10–12]. Although primary infection typically remains asymptomatic in immunocompetent individuals, some chronic viruses contribute to the development of autoimmune disorders and cancers, among other adverse health outcomes[1,13–15]. These viruses typically remain dormant, but can reactivate during periods of stress, sleep deprivation, surgery, hormonal imbalances or in critical illness[16–19]. The full range of immunological consequences from these viral reactivations remain largely unknown.

Since its emergence, SARS-CoV-2 has resulted in more than 774 million cases of COVID-19 and 7 million deaths[20,21]. Owing to the physiological stress introduced by SARS-CoV-2, underlying viral infections may reactivate and potentially contribute to the immunological consequences of COVID-19. For example, reactivation of *Herpesviridae*, including Epstein–Barr virus (EBV), cytomegalovirus (CMV), human herpesvirus 6 (HHV6) and human herpesvirus 8 (HHV8), is associated with worse clinical outcomes in patients with COVID-19 (refs. 2–6). Reactivation of CMV and EBV in particular have been linked to more

A list of affiliations appears at the end of the paper. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: esther.melamed@austin.utexas.edu

severe outcomes, including increased mortality[2,7]. Additionally, patients with long COVID develop increased EBV antibody titres, raising the possibility that reactivation of these viruses may contribute to long COVID[4,22].

Many of the foundational COVID-19 viral reactivation studies have been limited by small sample sizes, focused on a subset of *Herpesviridae* or relied solely on antibody responses to assess viral reactivation[2–7,22,23], as opposed to measuring transcripts of actively replicating viruses. Thus important gaps remain in our understanding of the dynamics and biology of viral reactivation during acute COVID-19 and their role in long COVID.

To address the knowledge gap in viral reactivation in COVID-19, we leveraged IMPACC, a longitudinal prospective observational study of 1,154 patients who were hospitalized for COVID-19, which evaluated patients during acute hospitalization and for 12 months post-hospitalization. We carried out longitudinal, multi-omic analyses of nasal swabs, peripheral blood mononuclear cells (PBMCs) and endotracheal aspirates and found significant reactivation of chronic viruses, particularly from the *Herpesviridae* and *Anelloviridae* families, associated with COVID-19 severity. By integrating host and viral transcriptomics, cytokine profiling, cellular immunophenotyping, metabolomics and proteomics, we observed distinct viral reactivation dynamics and striking associations between viral reactivation, clinical outcomes, immunologic features and patient demographics, both during acute COVID-19 and during long COVID. Our results provide insights into the endogenous virological landscape of patients with COVID-19, highlighting the complex interplay between SARS-CoV-2 infection, chronic viral reactivation, host immune responses and clinical outcomes.

IMPACC cohort

The Immunophenotyping Assessment in a COVID-19 Cohort (IMPACC) consortium enrolled 1,154 patients who were hospitalized for COVID-19 across 20 US hospitals between May 2020 and March 2021 (Fig. 1a, Extended Data Fig. 1a and Supplementary Table 1). All participants were COVID-19 vaccine-naive at the time of enrolment. To assess COVID-19 severity, participants were assigned to one of five trajectory groups (TG1–TG5) using latent class mixed modelling of respiratory status over the first 28 days[24]. Groups were classified as mild (TG1, length of stay (LOS) approximately 3–5 days, $n = 228$), moderate (TG2, LOS approximately 7–14 days, $n = 263$), severe (TG3, LOS approximately 10–14 days and discharged with limitations, $n = 333$), critical (TG4, critically ill with LOS greater than 28 days, $n = 222$) or fatal within 28 days (TG5, $n = 108$). From each participant, bulk RNA sequencing (RNA-seq) was performed on PBMCs, nasal swabs and for mechanically ventilated patients, endotracheal aspirates, at up to ten visits during one year post-hospital admission (Fig. 1b). Additionally, we assessed whole-blood immune cell populations by cytometry by time of flight (CyTOF), serum EBV and CMV antibody titres, serum cytokine levels by proximity extension assay (PEA), and the plasma proteome and metabolome by mass spectrometry.

RNA-seq of reactivated virome

Upon screening RNA-seq data for any human-infecting virus transcripts, we identified viral RNA in nasal, endotracheal aspirate and PBMC samples for a diverse range of human-infecting viruses beyond SARS-CoV-2, including EBV, CMV, HHV6, human alphaherpesvirus 1 and 2 (HSV1, HSV2) and several *Anelloviridae* and enterovirus species (Fig. 1c,d and Extended Data Fig. 1b–d). Unsurprisingly, SARS-CoV-2 was the most prevalent virus, and was primarily found in nasal and endotracheal aspirate samples (Fig. 1c). We confirmed that SARS-CoV-2 abundance measured by RNA-seq highly correlated with results from reverse transcription with quantitative PCR (RT–qPCR) (Extended Data Fig. 2a,b).

Beyond SARS-CoV-2, most of the other detected viruses were chronically infecting (for example, *Herpesviridae* and *Anelloviridae*); we focused on these owing to their higher prevalence in our cohort (Extended Data Fig. 1b,c). Among *Herpesviridae*, transcripts of HSV1, EBV and CMV were commonly detected across compartments during acute COVID-19 (first 40 days after admission) and were detected less frequently during the convalescent period (2 months or more post-admission) (Fig. 1c). Additionally, we detected a wide range of *Anelloviridae* and acute-infecting enterovirus species (Extended Data Fig. 2c,d), and analysed their collective viral load at the family and genus level, respectively, owing to the sparsity of individual strains. Reactivated viruses were common across recruitment sites (Extended Data Fig. 2e) and viral detection rates were comparable across sequencing cores (Extended Data Fig. 2f), suggesting minimal-to-no enrolment or sequencing site effects. Furthermore, read duplication within batches was low (Extended Data Fig. 2g), with detected duplication likely reflecting specific viral gene expression (Extended Data Fig. 2h). Finally, viral alignment E values were exceedingly low (Extended Data Fig. 2i), providing confidence in the taxonomic alignment.

Notably, each viral species displayed unique temporal dynamics of reactivation relative to hospital admission (Fig. 1d). For example, EBV reactivated early in the disease course, with 24% of participants having detectable transcripts near time of admission (days 1–8, 260 out of 1,080 participants), followed by a gradual decline over time. Unlike EBV, *Anelloviridae* transcript frequency remained constant until day 20 post-admission, followed by a slow decline. By contrast, HSV1 and CMV reactivated later in disease course, with HSV1 detected in 43% of participants with endotracheal aspirate samples (13 out of 30) and 15% of nasal samples (14 out of 90) and CMV detected in 8% of PBMC samples (9 out of 110) at 19–23 days post-admission (Fig. 1d and Extended Data Fig. 3b–e). These temporal dynamics were also reflected in viral reads per million (RPM) across samples (Extended Data Fig. 3a). Of note, as sample composition changed over time owing to participant death, drop-out and discharge, we also assessed viral detection rates by time from symptom onset, which showed similar patterns to the days-from-hospitalization analysis (Extended Data Fig. 3b–i).

Viral detection varied across compartments, with EBV transcripts found to be more common in PBMCs and HSV1 in nasal and endotracheal aspirate samples. Nevertheless, viral transcripts often correlated between compartments (Fig. 1e). Furthermore, among participants with viral reactivation (47.9%, 550 out of 1,148), most had only one virus detected in the acute period (67.6%, 372 out of 550), and co-detection of multiple viruses simultaneously was infrequent (Extended Data Fig. 4a–d).

Finally, we validated the detection of viral transcripts in an external cohort of COVID-19 whole-blood RNA-seq[25,26], observing strikingly similar temporal dynamics for EBV, CMV and *Anelloviridae* over the first 40 days post-symptom onset (Extended Data Fig. 5a–c).

Clinical outcomes of viral reactivation

Next, we evaluated how viral transcript detection within 40 days of hospital admission associated with COVID-19 severity, using IMPACC trajectory groups[24] (Fig. 2a and Extended Data Figs. 5d and 6a). Cumulative link modelling of trajectory groups demonstrated significant associations between COVID-19 severity and *Herpesviridae* and *Anelloviridae* transcripts (Supplementary Table 2). Specifically, we found associations between severity and transcripts from *Anelloviridae* (PBMC adjusted $P = 5.02 \times 10^{-5}$), CMV (nasal adjusted $P = 4.83 \times 10^{-4}$, PBMC adjusted $P = 1.66 \times 10^{-3}$), EBV (nasal adjusted $P = 4.95 \times 10^{-6}$, PBMC adjusted $P = 2.55 \times 10^{-9}$) and HSV1 (nasal adjusted $P = 5.29 \times 10^{-5}$). When limiting to TG4 participants, those with CMV transcripts in any respiratory compartment were more likely to die within 1 year (nasal adjusted $P = 0.039$, endotracheal aspirate adjusted $P = 0.007$). This was also true for TG4 participants with detectable nasal EBV (adjusted

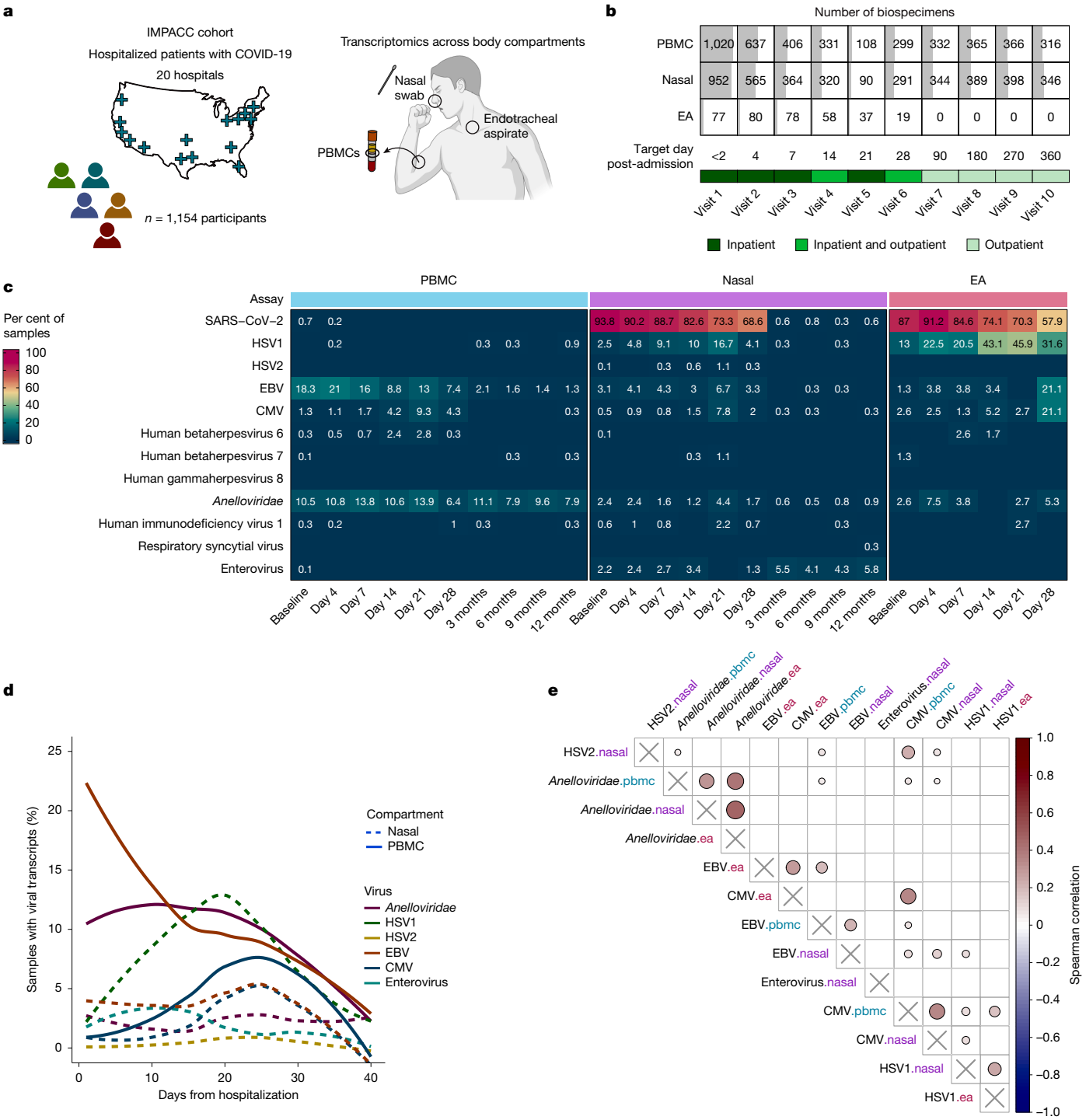


Fig. 1 | The transcriptionally active human virome of the blood, upper airway and lungs in patients who were hospitalized for COVID-19. a, A total of 1,154 participants were recruited across 20 sites for the IMPACC study, and PBMC, nasal and endotracheal aspirate samples were analysed by RNA-seq. Drawing of the coughing person created in BioRender; Maguire, C. https://BioRender.com/glhtw5c (2026). US outline from svgsilh.com (CC0 1.0). **b**, The number of biospecimens for each transcriptomic assay at each time point. The grey background shading of the cell indicates the percentage of total participants who provided a biospecimen for that assay at that time point. EA, endotracheal aspirate. **c**, Heat map showing percentage of samples with detected reads for different viruses for each transcriptomic assay at each time point. **d**, Smoothed curves showing the percentages of total samples that were positive for six common viruses in the nasal and PBMC transcriptomic analyses. Curves were calculated using the percentage of samples that were positive on each day ± 2 days (rolling window approach), followed by a local polynomial regression fitting. **e**, Spearman correlation of viral RPM across transcriptomic assays, with viruses hierarchically clustered. The size of the circle indicates absolute value of the correlation. Assays are denoted with virus type in black and sample type in coloured text. In **e**, only correlations with Benjamini–Hochberg adjusted *P* value ≤ 0.05 are visualized.

P = 0.025) and HSV1 (adjusted *P* = 0.015) (Fig. 2a). Notably, prevalence of chronic viruses was not significantly different between TG4 and TG5 (Supplementary Table 2).

We then evaluated whether viral transcripts varied with age, adjusting for COVID-19 severity (trajectory groups), and found a significant positive association between *Anelloviridae* in PBMCs and

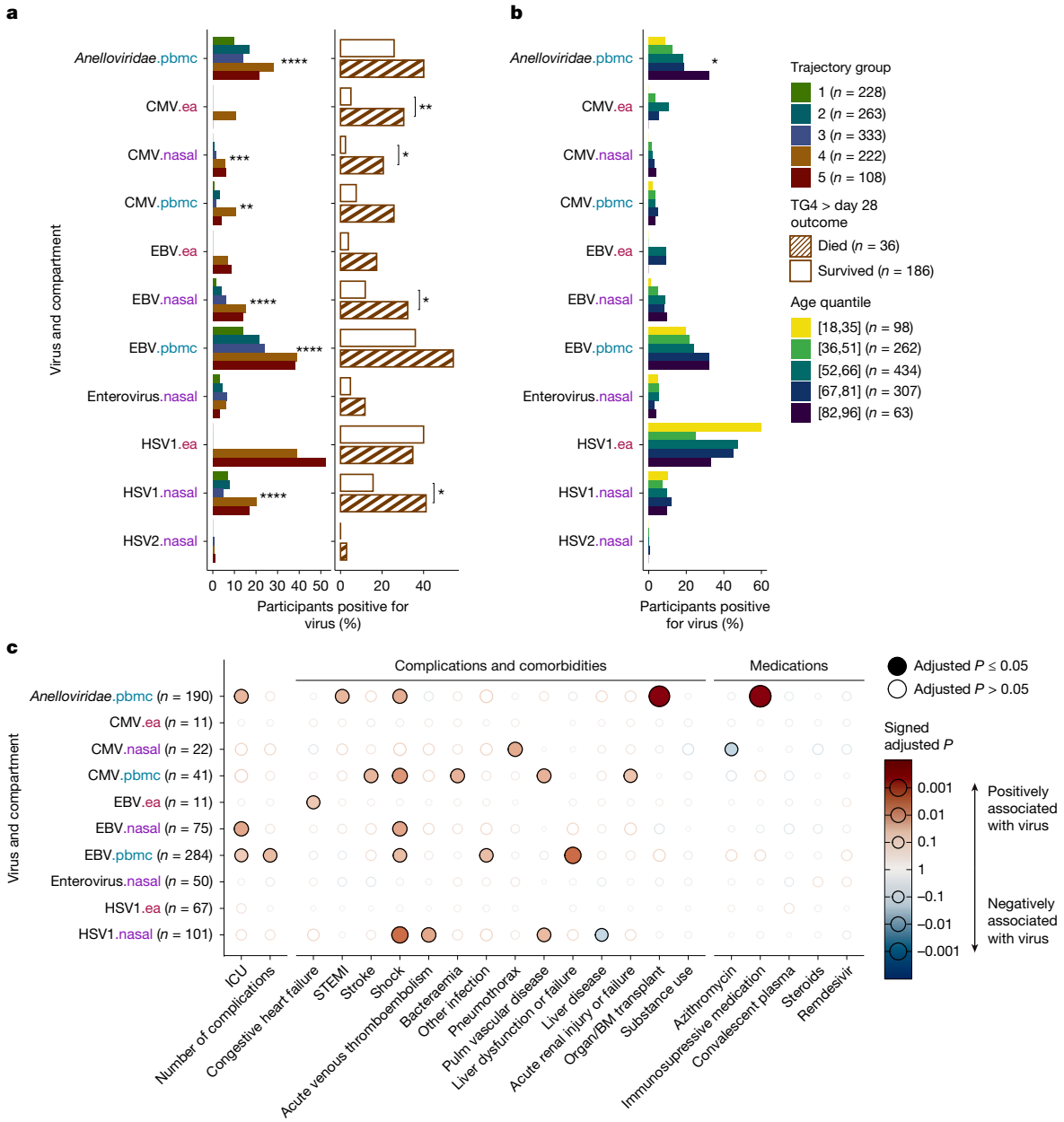


Fig. 2 | Clinical outcomes associated with activation of the human virome in severe COVID-19. a, Percentage of participants in the cohort who had detectable viral reads in at least one sample within 40 days of hospital admission (IMPACC visits 1–6). Participants were split by trajectory group (left), a measure of COVID-19 severity, and participants in TG4 were further subsetted by their long-term mortality outcome (right). For the trajectory group association testing, cumulative link mixed modelling was used to calculate significance, and for the TG4 long-term mortality association testing, a right-censored Cox mixed proportional hazards model was used (Methods). *P* values from both analyses were corrected with the Benjamini–Hochberg procedure. **b**, Percentage of participants in the cohort who had detectable viral reads in at least one sample within 40 days of hospital admission (IMPACC visits 1–6) split by age quintile. Significance was calculated using a cumulative link mixed model that controlled for trajectory group (Methods). **c**, Results from logistic mixed effect modelling of various complications, comorbidities and medication usage evaluating for association with viral transcripts detected within 40 days of hospital admission (IMPACC visits 1–6) while controlling for sex, age quintile and trajectory group. Dot colour indicates directionality of the association, with positive association in red and negative association in blue. Filled dots represent Benjamini–Hochberg adjusted $P \leq 0.05$. ICU, intensive care unit. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$ (Benjamini–Hochberg adjusted *P* values).

increasing age (Fig. 2b and Extended Data Fig. 5e, adjusted $P = 0.04$). Of note, Hispanic ethnicity was significantly associated with both CMV (adjusted $P = 0.02$) and EBV (adjusted $P = 0.02$; Extended Data Fig. 6b). However, no significant association were found between viral transcripts and biological sex, or treatment with remdesivir or steroids (Extended Data Fig. 6c–e).

We next evaluated the association of viral transcripts with comorbidities, medications and clinical complications (Fig. 2c and Supplementary Table 2). *Anelloviridae* transcripts in PBMCs were significantly associated with history of solid organ transplantation, immunosuppressive medications, shock, intensive care unit admission and myocardial infarction. Notably, although medication-mediated immunosuppression was strongly associated with *Anelloviridae*, these participants only comprised 17.4% of *Anelloviridae*-positive participants, demonstrating that presence of *Anelloviridae* is not exclusive to long-term medication-associated immunosuppression (Extended Data Fig. 6f). Furthermore, owing to the correlation between COVID-19 severity, age and immunosuppressive medications, we evaluated all three factors

simultaneously and identified that *Anelloviridae* was independently associated with each ($P = 8.6 \times 10^{-5}$, 5.0×10^{-3} and 2.2×10^{-8}, respectively; Supplementary Table 2).

For *Herpesviridae* (Fig. 2c), CMV in the nasal compartment was linked to pneumothorax, whereas in the PBMCs, CMV was associated with bacteraemia, pulmonary vascular disease, renal complications, shock and stroke. EBV transcripts in the nasal compartment were associated with intensive care unit-level care and shock, whereas EBV in PBMCs correlated with liver failure, concurrent infections, shock and the overall number of complications. HSV1 in the nasal compartment was significantly associated with acute venous thromboembolism and shock, and inversely associated with liver disease.

Anti-viral immune responses

We next leveraged our multi-omic data to validate viral reactivation in COVID-19 and characterize the host immune responses, incorporating serum EBV and CMV antibody levels, immune cell frequencies, serum cytokines, plasma metabolomics and host transcriptomics.

First, we observed that participants with EBV transcripts in PBMCs had persistently elevated EBV IgG and IgA antibody titres, further validating the use of viral transcripts as a meaningful measure of viral activity (Fig. 3a and Supplementary Table 3). Similarly, participants with CMV transcripts had significantly higher CMV seropositivity rates at baseline (Fig. 3b, $P = 0.0018$). Furthermore, using mass spectrometry proteomics, we found that plasma HSV1 proteins were significantly more common in participants with HSV1 transcripts in nasal swabs (Fig. 3c, $P = 0.016$) and in TG4 participants (Extended Data Fig. 7a, $P = 0.0003$).

Using CyTOF, we observed significant associations between EBV transcription in PBMCs and increased proportions of B-cell plasmablasts, the primary host cells of the virus (Fig. 3d). Furthermore, we found that CMV transcripts were associated with a significant increase in CD4 and CD8 central memory T cell frequency, and a reduction in CD27low effector memory CD4 T cells. Of note, detection of both EBV (nasal) and CMV (PBMC) was associated with a higher frequency of activated $CD4^+$ and $CD8^+$ T cells.

To assess whether changes in circulating cell frequencies may explain the association between *Herpesviridae* and *Anelloviridae* transcripts and COVID-19 severity, we repeated our analysis across trajectory groups while controlling for immune cell frequencies, which vary with disease severity[27]. Notably, even after this adjustment, viral transcripts remained significantly associated with COVID-19 severity (Extended Data Fig. 7b).

Serum cytokines in viral reactivation

Next, we investigated whether *Herpesviridae* and *Anelloviridae* transcripts were associated with inflammatory protein changes. Using generalized additive mixed modelling (GAMM), we compared longitudinal cytokine dynamics in patients with or without viral reactivation, controlling for severity (trajectory groups), sex and age (Fig. 4a and Supplementary Table 4, adjusted $P \leq 0.01$). This approach enabled identification of severity-independent cytokine changes associated with reactivation of different viruses.

We found that different viruses had distinct cytokine profiles (Fig. 4a), although several cytokines overlapped between viruses, including CXCL10 (Fig. 4b), CXCL11 (Fig. 4c) and IL-18 (Fig. 4d).

EBV in PBMCs was associated with increases in key cytokines, including IL-6 (Fig. 4e), CCL7, CCL2, IL-10 (Fig. 4f) and CXCL10, all of which are linked with COVID-19 severity[28,29] (Fig. 4a). Of these, only CXCL10 was also associated with EBV in the nasal transcriptomics. EBV in the nasal compartment was additionally correlated with increases in IL-18, CXCL11, CXCL10, IL-18R1, CD274, IL-15RA, IL22RA1, HGF, IFNγ, CCL8 and MMP10 and a decrease in KITLG, suggesting that host immune response to viruses may differ between compartments, with nasal EBV possibly reflecting a more severe state of viral reactivation. It has previously been observed that increased TGFβ preceded EBV reactivation in COVID-19-associated multisystem inflammatory syndrome in children[30], which we did not observe (Extended Data Fig. 7c,d), possibly suggesting that the dynamics of viral reactivation may differ across stages of COVID-19.

Similar to EBV, HSV1 (nasal) and CMV (PBMCs) were associated with a common set of pro-inflammatory serum cytokines (Fig. 4a). HSV1 and CMV both associated with increases in IL-18, CXCL11, CXCL10, IL-18R1, CD274, IL-15RA, CD40, CXCL9, CX3CL1, TNF and CDCP1, and HSV1 also correlated with increased CCL25, SLAMF1, CD5, FGF23, TNFRSF9 and IL-10RB, whereas CMV associated with increased HGF, IFNγ (Fig. 4g), CCL8, CCL7, LIFR, CD8A, ADA and CXCL8 and decreased MMP1 and IL-4. Finally, *Anelloviridae* were only associated with increases in CXCL11 and IL-18 and decreases in TNFSF11.

Plasma metabolites in viral reactivation

We next evaluated the association of viral reactivation with plasma metabolites (Extended Data Fig. 8 and Supplementary Table 5). Viral reactivation was associated with changes in metabolites, particularly those belonging to amino acid and lipid metabolism (Fig. 5a,b). Notably, CMV was associated with the greatest number of metabolome changes, including increased urea and TMAP (*N*,*N*,*N*-trimethyl-L-alanyl-L-proline betaine) levels (Fig. 5a,c and Extended Data Fig. 9a). Additionally, CMV and *Anelloviridae* were associated with increased long chain fatty acids (for example, erucate, arachidate and docosadienoate) (Fig. 5d and Extended Data Fig. 9b,c) and dimethylarginine (symmetric dimethylarginine and asymmetric dimethylarginine, regulators of nitric oxide synthesis) (Extended Data Fig. 9d). We also identified a shared metabolomic signature across multiple viruses, including *Anelloviridae*, HSV1, CMV and EBV, with notable reductions in *S*-methylcysteine sulfoxide and 6-bromotryptophan (Fig. 5e and Extended Data Figs. 8 and 9e).

Host PBMC and nasal transcriptomics

To identify a signature for each chronic virus independent of COVID-19 severity and participant demographics, we evaluated nasal and PBMC transcriptomic data for differentially expressed host genes while controlling for COVID-19 severity, SARS-CoV-2 nasal viral load, sex, age and days from hospital admission (Supplementary Table 6).

In the PBMC transcriptomics, viral detection in either the PBMC or nasal compartments was associated with changes in PBMC gene expression (Extended Data Fig. 9f,g). Gene set enrichment analysis demonstrated that *Anelloviridae* and CMV in the PBMCs, and nasal HSV1 were associated with downregulation of diverse RNA processing and protein translation pathways. Similarly, EBV, CMV and HSV1 were associated with an upregulation of cellular replication pathways. Additionally, *Anelloviridae* and CMV in the PBMCs, and nasal HSV1 and CMV were significantly associated with signatures of neutrophil degranulation, potentially reflecting presence of low-density neutrophils in the PBMCs or involvement of immune genes that are non-specific to neutrophil degranulation. Finally, EBV in the PBMCs was uniquely associated with platelet activation and signalling.

By contrast, nasally reactivated viruses (EBV, CMV and HSV1) had the strongest associations with gene expression changes in the upper airway (Extended Data Fig. 9f,h). Furthermore, CMV, EBV and HSV1 transcripts were all associated with upregulation of interleukin signalling (including IL-10 signalling), lymphocyte immunoregulatory interactions and neutrophil degranulation (Extended Data Fig. 9h). Additionally, upper airway CMV and EBV transcripts were associated with increased anti-parasite and phagocytosis pathways. EBV was also associated with increased expression of genes related to T cell

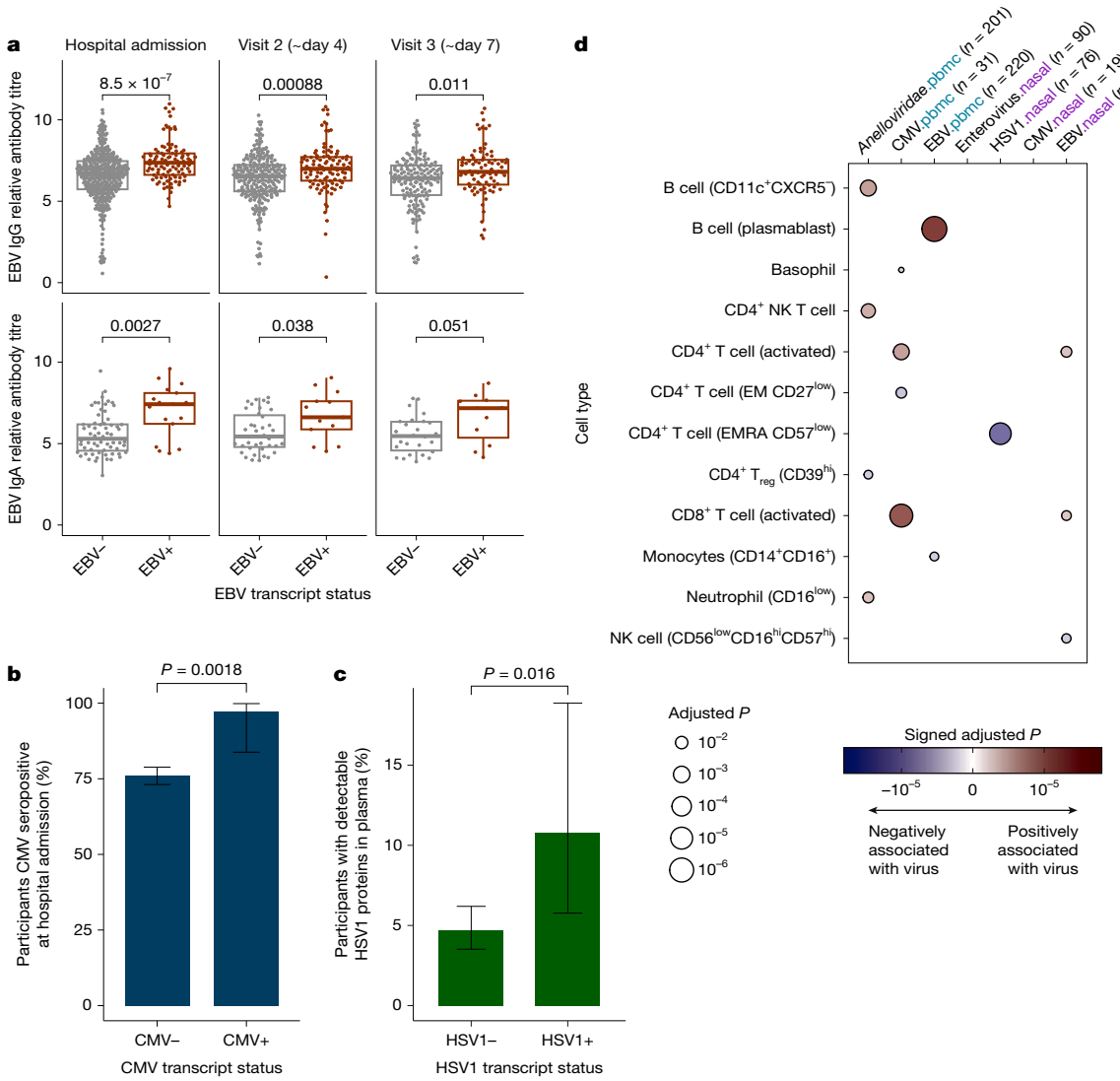


Fig. 3 | Antibody response, proteomics, and circulating cellular immunophenotyping validates chronic viral reactivation in COVID-19. **a**, Relative EBV GP350 IgG antibody titres in participants with detected EBV transcripts in any transcriptomic sample in the first 40 days after hospital admission (EBV+) and participants with no detected transcripts (EBV−). Adjusted *P* values were calculated using a two-sided Wilcoxon rank-sum test with Benjamini–Hochberg correction. Box plots denote the median centre line, interquartile range (box) and 1.5× the interquartile range (whiskers). IgG: EBV− group, $n = 366$ (hospital admission), 239 (visit 2) and 156 (visit 3); EBV+ group, $n = 97$ (admission) 103 (visit 2) and 84 (visit 3). IgA: EBV− group, $n = 64$ (admission), 39 (visit 2) and 26 (visit 3); and EBV+ group. $n = 17$ (admission), 13 (visit 2) and 11 (visit 3). **b**, Percentage of patients who are seropositive for CMV at hospital admission among those with no CMV transcripts in the first 40 days after admission (CMV−; 76.1% (668 out of 878)) versus participants with detectable transcripts in any transcriptomic sample (CMV+; 97.2% (35 out of 36)). Error bars denote 95% confidence interval. CMV−, $n = 878$; CMV+, $n = 36$. **c**, Percentage of participants with detectable HSV1 proteins in the plasma split by participants with HSV1 detectable in the nasal swabs in the first 40 days after admission (HSV1+; 10.8% (11 out of 102)) versus participants with no detectable transcripts (HSV1− (4.7% (49 out of 1,046)). Error bars denote 95% confidence interval. HSV1−, $n = 1,046$; HSV1+, $n = 102$. *P* values in **b**,**c** calculated using a chi-square test of independence. **d**, Results of linear mixed effect modelling identifying whole-blood cell-type frequencies significantly associated with detection of different viruses while controlling for trajectory group, sex and age quintile, with enrolment site and participant as mixed effects. Dot colour indicates directionality of the association, with positive association in red and negative association in blue. *P* values were adjusted using Benjamini–Hochberg correction. Dots show data with Benjamini–Hochberg adjusted $P \leq 0.05$. EM, effector memory; EMRA, effector memory cells re-expressing CD45RA; NK, natural killer; T_{reg}, regulatory T cell.

activation. Finally, *Anelloviridae* in PBMCs was also associated with nasal transcriptome changes, including keratinization and downregulation of proteasome-related proteins.

Viral reactivation in long COVID

Given recent reports linking viral reactivation to long COVID[4,22,31], we assessed viral reactivation across patient-reported outcome (PRO) groups, generated from convalescent survey data[32]. The four PRO groups comprised participants reporting minimal to no-deficits (minimal), physical deficits (physical disability or fatigue), cognitive deficits (cognitive impairment) and global deficits (physical and cognitive).

Upon evaluating viral reactivation during the acute stage of COVID-19, no significant relationship was found with PRO groups (Fig. 6a and Extended Data Fig. 10a,b). However, the sample size was limited owing to the association of chronic viruses with mortality and participant drop-out in the convalescent stage.

We then investigated the correlation between viral detection during the convalescent period and PRO groups. *Anelloviridae* and

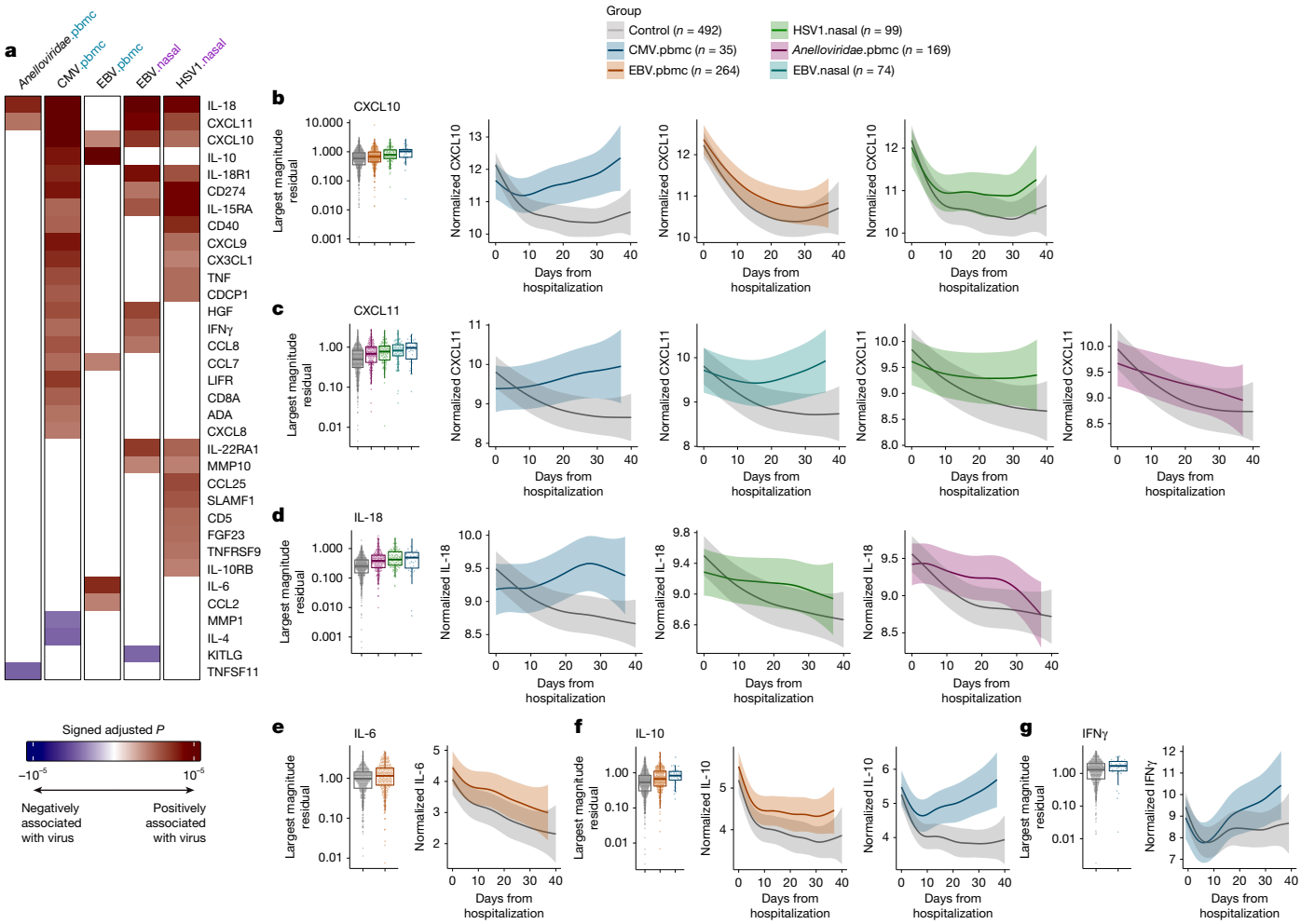


Fig. 4 | Reactivation of chronically infecting viruses in acute COVID-19 correlates with changes in inflammatory cytokine and chemokine expression. a, Summary heat map of Benjamini–Hochberg adjusted *P* values from GAMM evaluating the effects of chronic viral reactivation on cytokine and chemokine dynamics over time. The control group comprised 492 participants without any detected chronic virus transcripts in the acute period (up to 40 days post-hospitalization). The adjusted *P* values are signed and coloured according to direction of the associations of the cytokine or chemokine with the virus. The heat map cell was only coloured when adjusted $P \leq 0.01$ for the main effect or time interaction term for viral reactivation in the model. **b**–**g**, Box plots depicting the largest-magnitude GAMM residuals from the statistical models reported in **a** for each participant by virus, alongside longitudinal GAMM-predicted means and 95% confidence intervals over days from hospitalization for CXCL10 (**b**), CXCL11 (**c**), IL-18 (**d**), IL-6 (**e**), IL-10 (**f**) and IFNγ (**g**). The control group is the model fit for 492 participants without any detected chronic viruses during the first 40 days after hospital admission. Box plots denote median (centre line), interquartile range (box) and 1.5× the interquartile range (whiskers).

enteroviruses were the most frequently detected viruses in convalescent samples (Figs. 1c and 6b and Extended Data Fig. 10c–e). Notably, *Anelloviridae* transcripts were significantly more prevalent in participants from the physical PRO group, which was characterized by high scores on the PRO Measurement Information System (PROMIS) physical function survey, even when controlling for sex, age, immunosuppressive medications and acute COVID-19 severity (adjusted $P = 0.012$). This finding suggests that detection of *Anelloviridae* transcripts may serve as a potential signature for persistent physical disability in patients with long COVID. We further observed that the *Anelloviridae*-associated gene expression signature was consistent between the acute and convalescent periods (Fig. 6c and Extended Data Fig. 7).

Discussion

Chronically infecting viruses are prevalent in the general population, with individuals carrying 8–12 viruses at any given time[1]. These viruses are generally innocuous and typically do not cause infectious symptoms. However, emerging evidence suggests that their reactivation may contribute to autoimmune diseases, malignancies and chronic fatigue syndrome[1,13–15]. Additionally, *Herpesviridae* such as EBV and CMV commonly reactivate in severe infections (for example, malaria or pneumonia) and sepsis[18,33,34] and have more recently been implicated in long COVID[4,22,35]. Thus, there is an urgent need to understand how chronically infecting viruses reactivate in COVID-19 and develop strategies to combat their reactivation.

Here we conducted a prospective, multi-omic analysis of 1,154 patients who were hospitalized due to COVID-19, which revealed widespread reactivation of chronic viruses from the *Herpesviridae* and *Anelloviridae* families. By integrating cellular and cytokine immunophenotyping, metabolomics and transcriptomics, our findings expand on the complex interplay between SARS-CoV-2 and chronic viral reactivation and provide a deeper understanding of the host immune response. Importantly, our findings demonstrate an association of viral reactivation with COVID-19 and long COVID outcomes.

Despite prior studies evaluating viral reactivation in COVID-19 (refs. 2–7,22,23) and other infections[18,33,34], the exact timing and rates of reactivation for different viruses have not been clearly established,

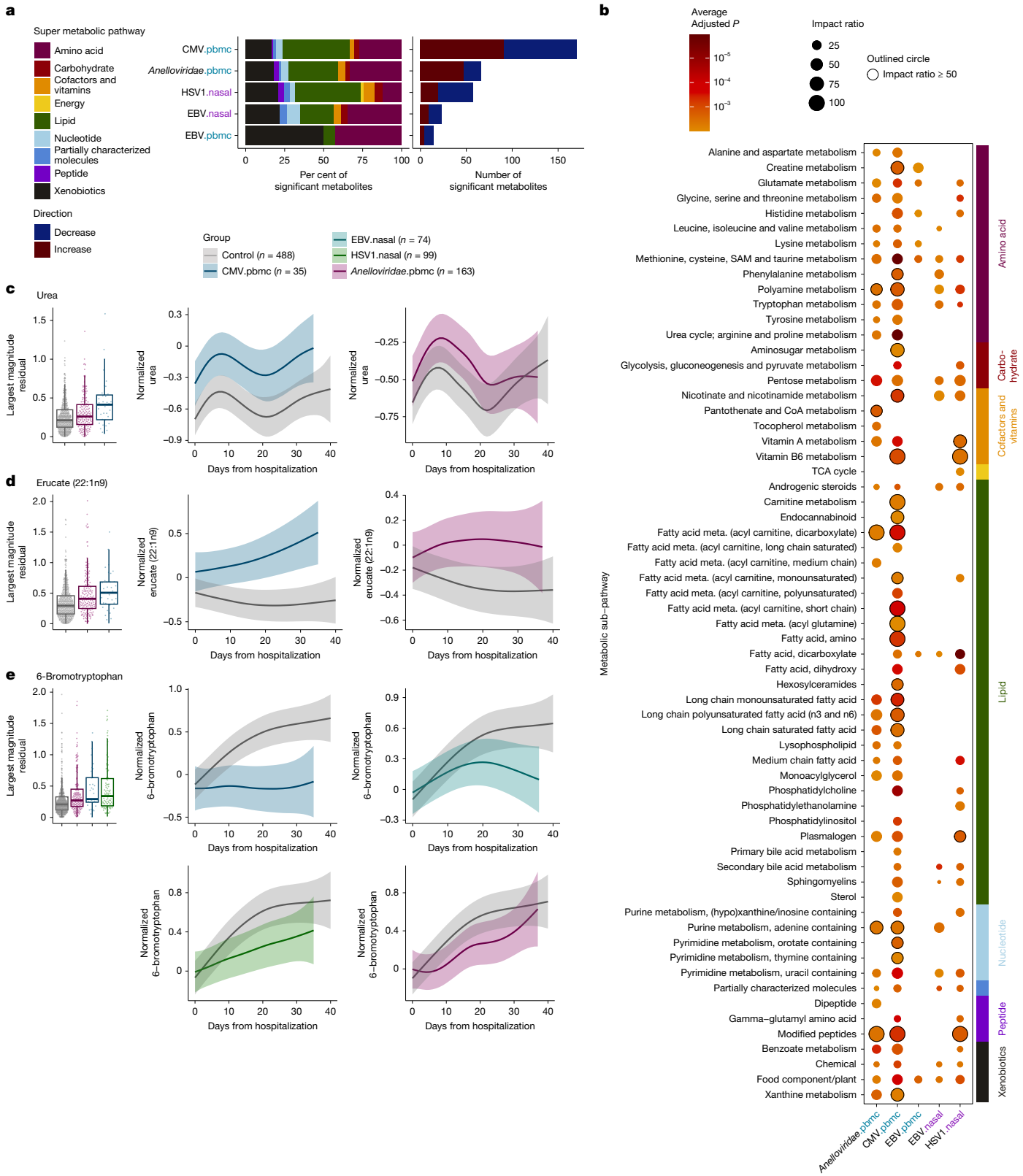


Fig. 5 | Viral reactivation in COVID-19 associated with shifts in the plasma metabolome. a, Number of significant metabolites (Benjamini–Hochberg adjusted *P* value ≤ 0.01) from GAMM evaluating the effects of chronic viral reactivation on metabolite dynamics over time and the percentage of significant metabolites that map to each major branch of metabolism by virus. The control group comprised 488 participants without any detected chronic virus transcripts in the acute period (up to 40 days post-hospitalization). **b**, Dot plot of metabolic sub-pathways containing significantly different metabolites associated with different viral reactivations. Dot size indicates the impact ratio (the per cent of significantly different metabolites from each pathway) and the colour indicates the average adjusted *P* value of significant metabolites in that pathway. CoA, coenzyme A; SAM, *S*-adenosylmethionine; TCA, tricarboxylic acid. **c**–**e**, Box plots depicting the largest-magnitude GAMM residuals for each participant by virus, next to longitudinal GAMM-predicted means and 95% confidence intervals over days from hospitalization for urea (**c**), erucate (**d**) and 6-bromotryptophan (**e**). Box plots denote median (centre line), interquartile range (box) and 1.5× the interquartile range (whiskers).

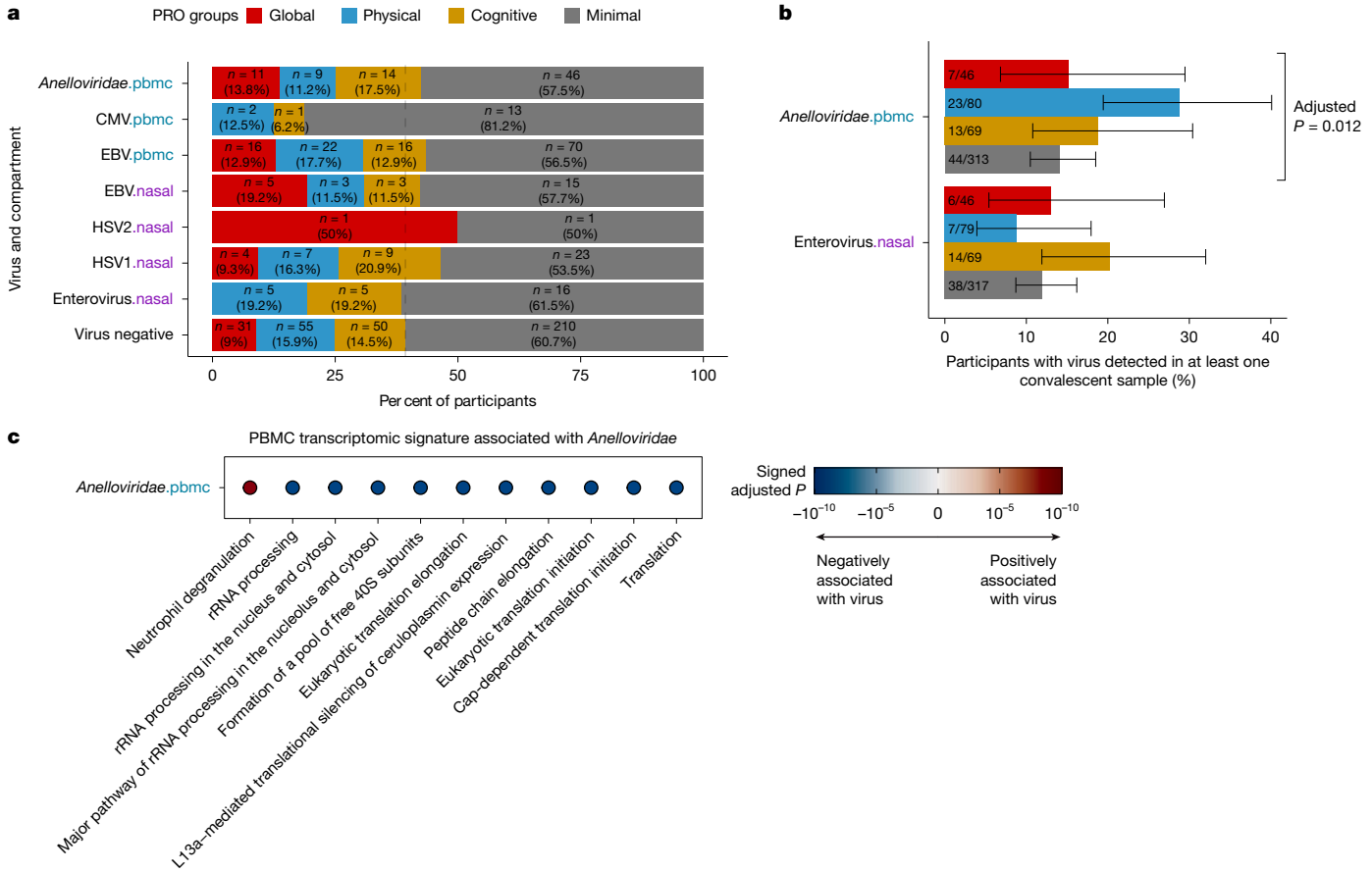


Fig. 6 | Associations between chronic viral reactivation and long COVID. **a**, Percentage of participants belonging to each PRO group for participants grouped by detected viruses in the acute COVID-19 period. The virus-negative group comprises participants who exhibited no viral reactivation for any virus in the acute period, whose rate of total long COVID is indicated by the dashed grey line and serves as the baseline reference. **b**, Percentage of each PRO group that had viral transcripts detected for *Anelloviridae* in the PBMCs or enteroviruses in the upper airway via RNA-seq, in any of their convalescent samples. Error bars denote 95% confidence interval. Statistical significance calculated using a linear mixed effects model with immunosuppressed status due to medications, trajectory group, sex, age quintile and virus detection status included as main effects and enrolment site as a mixed effect; *P* values were adjusted with Benjamini–Hochberg corrections. **c**, Top hypergeometric enriched pathways (determined via lowest adjusted *P* values) from differentially expressed genes associated with *Anelloviridae* in convalescent samples reveals a similar signature of *Anelloviridae* during acute COVID-19 (shown in Extended Data Fig. 9g). Results in **c** were calculated using hypergeometric enrichment of pathways from Reactome, separately on the positive and negative differentially expressed genes. rRNA, ribosomal RNA.

with prior studies limited to single time points in small cohorts or relying on antibody titres to identify viral reactivations. Here, leveraging a large longitudinally sampled multi-institution cohort, we define timing and duration of chronic viral reactivations. Specifically, we found that EBV transcripts in PBMCs peaked early in acute disease following hospital admission and then decreased over time. Similarly, *Anelloviridae* transcripts were most common early during hospitalization and declined several weeks later. By contrast, CMV and HSV1 reactivated later and were primarily in respiratory samples, peaking about 22 days post-hospitalization. Furthermore, we replicated the rates and dynamics of viral reactivations in an external cohort, demonstrating robustness of our findings[25,26]. Of note, patients with EBV transcripts in either the nasal or PBMC compartments at admission had higher EBV IgG antibody titres, suggesting that EBV reactivation may happen prior to hospitalization for some patients. Furthermore, since viruses primarily reactivate in tissues (for example, HSV1 and HSV2 in sensory ganglia, CMV in myeloid or dendritic cells and EBV in lymphoid tissue), our results based on systemic detection (PBMCs) may underestimate tissue-specific reactivations. Collectively, our results provide novel insights into the dynamics of the diverse virological landscape of COVID-19 signifying 'dysvirosis' (a dysregulation of the human virome) and reveal that the time course of reactivation varies for different viruses.

Exactly how and whether viral reactivations contribute to COVID-19 clinical outcomes remains unknown. Although our data cannot causally determine the direct contribution of viral reactivations to acute COVID-19 disease severity, we identified several key observations that implicate viral reactivation in contributing to disease pathology.

First, multiple cytokines and chemokines, including IL-6, IL-10, CXCL10 and CXCL11, which have been previously associated with COVID-19 pathology, were elevated in participants with viral reactivations when controlling for COVID-19 severity. Notably, increased IL-10 associated with EBV and CMV, which are known to upregulate human IL-10 and produce viral IL-10, which may further displace human IL-10 from IL-10R[36,37]. We also identified increased circulating activated $CD4^+$ and $CD8^+$ T cells and increased cellular replication gene expression in PBMCs, suggesting expanding lymphocytes potentially responding to these viral reactivations. Furthermore, we observed that for PBMC viruses, the associated host inflammatory changes were limited to the blood and did not generalize to other compartments where the virus was not detected (such as the nasal compartment). Although it is possible that chronic viral reactivation may be secondary to disease severity, reactivations correlated with expected inflammatory changes of a responding immune system. Additionally, we observed increased EBV IgG titres at time of hospital admission, suggesting that EBV reactivation likely occurs prior to severe COVID-19 symptoms, raising the

suspicion of the potential role of EBV in disease pathophysiology. Thus, the integration of findings across our multi-omics datasets suggests that viral reactivation may not just be epiphenomena but may magnify ongoing inflammation and evoke additional immune responses.

Whereas viral reactivation has frequently been observed in immunosuppressed patients[38–43], we find that viral reactivation occurs frequently in immunocompetent patients with COVID-19. If viral reactivations were only indicative of a dysregulated immune system, we would expect reactivation to associate not only with severity, but also with immunosuppressive status, which we did not observe for the *Herpesviridae* in our cohort. Thus, viral reactivation may represent a broader phenomenon than previously appreciated and should be considered in a wider range of clinical contexts beyond immunosupression.

Further integrating our multi-omic dataset, we observed associations between chronic viral reactivations and plasma metabolites. In particular, levels of methylcysteine sulfoxide and 6-bromotryptophan, which are involved in protection against oxidative stress[44,45], were negatively associated with *Herpesviridae*, suggesting that viral reactivation may contribute to oxidative stress and cellular damage. Furthermore, 6-bromotryptophan has previously been associated with COVID-19 complications[27,46] and impaired kidney function[44], and its persistently low levels in the context of viral reactivation could imply additional risk to kidney health. This finding aligns with sustained increases of urea and TMAP in CMV-infected patients, which is indicative of renal impairment. Additionally, these data raise the possibility that monitoring for CMV reactivation may be beneficial in patients with COVID-19 presenting with acute kidney damage. The association between CMV and renal dysfunction is also interesting, as CMV infection is one of the most common viral complication in kidney transplant recipients[43], suggesting that future studies should evaluate the role of CMV reactivation across different renal disorders. Collectively, these integrated data across our multi-omic study suggest that viral reactivation is associated with specific metabolic perturbations, possibly reflecting strategies by pathogens to manipulate host cell metabolism to their advantage.

Our findings raise important questions about clinical applications of viral reactivation monitoring to guide patient care in acute infections such as COVID-19. Notably, there is already validated infrastructure to evaluate chronically infecting viruses, via quantitative PCR for different *Herpesviridae* and *Anelloviridae* (for example, Torque teno virus (TTV)). Furthermore, the increasing affordability of metagenomic sequencing could further broaden the range of detectable viruses. Additionally, for *Herpesviridae* in particular, there are existing antiviral therapies and several antivirals in clinical trials. Thus, these diagnostic and treatment tools could be readily adapted into broader viral reactivation monitoring and treatment strategies as future work develops specific protocols in patient care.

Importantly, COVID-19 can have ramifications that extend beyond the acute period into persistent sequelae, known as long COVID. Patients with long COVID develop heterogeneous symptoms that may persist for months to years after acute COVID-19 that profoundly affect individuals' overall health, often resulting in disability and loss of income[32,47,48]. Long COVID is estimated to affect 10–30% of individuals after COVID-19 (ref. 49). One of the emerging factors associated with long COVID is chronic viral reactivation, particularly reactivation of EBV[4,22,31].

We demonstrate that EBV reactivation is extensive in severe acute COVID-19 and establish the timing of EBV reactivation relative to SARS-CoV-2 infection and hospitalization. Although we did not find a higher rate of EBV transcripts in the acute phase of COVID-19 for patients who developed long COVID, this result raises the question of why increased EBV antibody titres have been frequently observed in long COVID[4,22] and whether the decoupling of viral load in the acute period and antibody titre in the convalescent period reflects an underlying pathology in the immune response. Of note, the PROs in this study were designed early in the pandemic prior to the definition of long COVID and thus may not fully capture current long COVID phenotypes[50], which may have limited our ability to detect association between EBV reactivation and long COVID. Thus, future work evaluating the dynamics of antibody titres versus viral transcription may continue to reveal the role of EBV in long COVID.

We also report a novel association between the long COVID physical PRO group and *Anelloviridae* transcripts post-hospitalization, even when controlling for age, chronic immunosuppression by medications and acute COVID-19 severity. *Anelloviridae* are a large family of negative-sense single-stranded DNA viruses, and some members of this family (such as TTV) are found in 80–90% of the population[51]. *Anelloviridae* have been previously linked with chronic conditions, such as chronic fatigue syndrome and multiple sclerosis[52,53]. These conditions often present with physical symptoms similar to those reported by patients with long COVID, including fatigue, cognitive dysfunction and post-exertional malaise. More broadly, *Anelloviridae* have also been associated with immunosenescence states[51], suggesting that increased viral transcription may underscore a potential state of immune dysfunction and immunosenescence in long COVID. Previously, the role of *Anelloviridae* in disease has been debated owing to the lack of an identified pathogenic role[51]. However, in our study we observed an association between *Anelloviridae* reactivation and enrichment of genes involved in neutrophil degranulation, suggesting that *Anelloviridae* may contribute to an inflammatory host response. Additionally, TTV has been shown to replicate preferentially in activated T cells, raising a concern for possible pathogenicity rather than just a passive 'passenger' role[54]. Our data highlight the need for future research to dissect the role of *Anelloviridae* in patients with long COVID.

There were several limitations to our study. First, the use of transcripts to identify viral loads is less common than RT–qPCR. However, the extensive sequencing depth of our samples (25–50 million reads) provided sufficient viral identification. Additionally, we only had six time points during acute COVID-19, limiting the ability to identify viral reactivation between sample collection. Nonetheless, with more than 1,000 participants, we were able to calculate the global longitudinal reactivation patterns for viruses. Furthermore, we only sequenced three tissues, which might have missed reactivation in other compartments, and endotracheal aspirate samples were only collected in ventilated patients, limiting our ability to analyse reactivation in this compartment. Similarly, for enteroviruses and *Anelloviridae*, owing to the sparsity of individual virus species, we collapsed reads at the genus and family level, respectively, which may have failed to capture more specific relationships for individual species. Another limitation was participant drop-out during the convalescent stage, particularly in patients with acute viral reactivation, limiting the power of analyses testing the association of acute viral reactivation with long COVID. Finally, IMPACC participants were unvaccinated, hospitalized and primarily exposed to ancestral SARS-CoV-2 strains; thus, future studies should ascertain whether more recent strains similarly drive viral reactivation in acute COVID-19 and long COVID in populations with hybrid immunity or milder disease.

Conclusions

In this study, we integrated clinical, immunologic, virologic and multi-omic data from cellular and cytokine immunophenotyping, metabolomics, proteomics and transcriptomics in a longitudinal cohort of 1,154 patients with COVID-19 to investigate for reactivation of chronic viruses and their association with clinical outcomes in one of the largest COVID-19 studies to date. We found that multiple chronic viruses reactivate during acute COVID-19 infection, particularly from the *Herpesviridae* and *Anelloviridae* families. Of note, immunosuppression did not explain the high rates of viral reactivation, suggesting that viral reactivation may have a larger role in COVID-19, even in immunocompetent individuals. Furthermore, we delineated the dynamics and rates of reactivation for various chronic viruses and

validated these findings in an external cohort. Additionally, we report association of viral reactivation with the host immune response and molecular pathways, as well as acute and chronic clinical sequelae of COVID-19. Notably, our results raise the possibility that viral reactivation may contribute to the development of severe acute COVID-19 and long COVID. Our findings underscore the emerging role of viral reactivations, highlighting the potential contribution to hyperinflammation, and raise the possibility that management of viral reactivations in conditions such as acute COVID-19 and long COVID could improve patient outcomes.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-026-10740-z.

1. Virgin, H. W., Wherry, E. J. & Ahmed, R. Redefining chronic viral infection. *Cell* **138**, 30–50 (2009).
2. Naendrup, J.-H. et al. Reactivation of EBV and CMV in severe COVID-19-epiphenomena or trigger of hyperinflammation in need of treatment? A large case series of critically ill patients. *J. Intensive Care Med.* **37**, 1152–1158 (2022).
3. Shafiee, A. et al. Reactivation of herpesviruses during COVID-19: a systematic review and meta-analysis. *Rev. Med. Virol.* **33**, e2437 (2023).
4. Peluso, M. J. et al. Chronic viral coinfections differentially affect the likelihood of developing long COVID. *J. Clin. Invest.* **133**, e163669 (2023).
5. Le Balc'h, P. et al. Herpes simplex virus and cytomegalovirus reactivations among severe COVID-19 patients. *Crit. Care* **24**, 530 (2020).
6. Simonnet, A. et al. High incidence of Epstein-Barr virus, cytomegalovirus, and human-herpes virus-6 reactivations in critically ill patients with COVID-19. *Infect. Dis. Now* **51**, 296–299 (2021).
7. Mattei, A. et al. Epstein-Barr virus, cytomegalovirus, and herpes simplex-1/2 reactivations in critically ill patients with COVID-19. *Intensive Care Med. Exp.* **12**, 40 (2024).
8. Borrow, P. Mechanisms of viral clearance and persistence. *J. Viral Hepat.* **4**, 16–24 (1997).
9. Deigendesch, N. & Stenzel, W. in *Handbook of Clinical Neurology* Vol. 145 (eds Kovacs, G. G. & Alafuzoff, I.) 227–243 (Elsevier, 2018).
10. Damania, B., Kenney, S. C. & Raab-Traub, N. Epstein-Barr virus: biology and clinical disease. *Cell* **185**, 3652–3670 (2022).
11. Tognarelli, E. I. et al. Herpes simplex virus evasion of early host antiviral responses. *Front. Cell Infect. Microbiol.* **9**, 127 (2019).
12. Okamoto, H. History of discoveries and pathogenicity of TT viruses. *Curr. Top. Microbiol. Immunol.* https://doi.org/10.1007/978-3-540-70972-5_1 (2009).
13. Hussein, H. M. & Rahal, E. A. The role of viral infections in the development of autoimmune diseases. *Crit. Rev. Microbiol.* **45**, 394–412 (2019).
14. Yuwanati, M., Bhatnagar, N. & Mhaske, S. Oncoviruses: an overview of oncogenic and oncolytic viruses. *Oncobiol. Targets* **2**, 4 (2015).
15. Ruiz-Pablos, M., Paiva, B., Montero-Mateo, R., Garcia, N. & Zabaleta, A. Epstein-Barr virus and the origin of myalgic encephalomyelitis or chronic fatigue syndrome. *Front. Immunol.* https://doi.org/10.3389/fimmu.2021.656797 (2021).
16. Kim, S. J. et al. Renal ischemia/reperfusion injury activates the enhancer domain of the human cytomegalovirus major immediate early promoter. *Am. J. Transplant.* **5**, 1606–1613 (2005).
17. de Almeida, S. M. et al. Reactivation of herpes simplex virus-1 following epilepsy surgery. *Epilepsy Behav. Case Rep.* **4**, 76–78 (2015).
18. Hraiech, S. et al. Herpes simplex virus and Cytomegalovirus reactivation among severe ARDS patients under veno-venous ECMO. *Ann. Intensive Care* **9**, 142 (2019).
19. Uchakin, P. N. et al. Fatigue in medical residents leads to reactivation of herpes virus latency. *Interdiscip. Perspect. Infect. Dis.* **2011**, 571340 (2011).
20. Phillips, N. The coronavirus is here to stay—here's what that means. *Nature* **590**, 382–384 (2021).
21. COVID Data Tracker. *Centers for Disease Control and Prevention* https://covid.cdc.gov/covid-data-tracker (2020).
22. Klein, J. et al. Distinguishing features of long COVID identified through immune profiling. *Nature* **623**, 139–148 (2023).
23. Xie, Y. et al. Clinical characteristics and outcomes of critically ill patients with acute COVID-19 with Epstein-Barr virus reactivation. *BMC Infect. Dis.* **21**, 955 (2021).
24. 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).
25. Charney, A. W. et al. Sampling the host response to SARS-CoV-2 in hospitals under siege. *Nat. Med.* **26**, 1157–1158 (2020).
26. Thompson, R. C. et al. Molecular states during acute COVID-19 reveal distinct etiologies of long-term sequelae. *Nat. Med.* **29**, 236–246 (2023).
27. Diray-Arce, J. et al. Multi-omic longitudinal study reveals immune correlates of clinical course among hospitalized COVID-19 patients. *Cell Rep. Med.* **4**, 101079 (2023).
28. Guo, J. et al. Cytokine signature associated with disease severity in COVID-19. *Front. Immunol.* **12**, 681516 (2021).
29. Xu, Z.-S. et al. Temporal profiling of plasma cytokines, chemokines and growth factors from mild, severe and fatal COVID-19 patients. *Signal Transduct. Target. Ther.* **5**, 100 (2020).
30. Goetzke, C. C. et al. TGFβ links EBV to multisystem inflammatory syndrome in children. *Nature* **640**, 762–771 (2025).
31. Cervia-Hasler, C. et al. Persistent complement dysregulation with signs of thromboinflammation in active long Covid. *Science* **383**, eadg7942 (2024).
32. Ozonoff, A. et al. Features of acute COVID-19 associated with post-acute sequelae of SARS-CoV-2 phenotypes: results from the IMPACC study. *Nat. Commun.* **15**, 216 (2024).
33. Goh, C. et al. Epstein-Barr virus reactivation in sepsis due to community-acquired pneumonia is associated with increased morbidity and an immunosuppressed host transcriptomic endotype. *Sci. Rep.* **10**, 9838 (2020).
34. Heininger, A. et al. Cytomegalovirus reactivation and associated outcome of critically ill patients with severe sepsis. *Crit. Care.* **15**, R77 (2011).
35. Gold, J. E., Okyay, R. A., Licht, W. E. & Hurley, D. J. Investigation of long COVID prevalence and its relationship to Epstein-Barr virus reactivation. *Pathogens* **10**, 763 (2021).
36. Jog, N. R., Chakravarty, E. F., Guthridge, J. M. & James, J. A. Epstein Barr virus interleukin 10 suppresses anti-inflammatory phenotype in human monocytes. *Front. Immunol.* **9**, 2198 (2018).
37. Poole, E., Neves, T. C., Oliveira, M. T., Sinclair, J. & da Silva, M. C. C. Human cytomegalovirus interleukin 10 homologs: facing the immune system. *Front. Cell. Infect. Microbiol.* **10**, 245 (2020).
38. Wade, J. C. Viral infections in patients with hematological malignancies. *Hematology Am. Soc. Hematol. Educ. Program* **2006**, 368–374 (2006).
39. Hatayama, Y. et al. Differential reactivation of cytomegalovirus and Epstein-Barr virus in patients with B cell lymphoma. *Viral Immunol.* **36**, 520–525 (2023).
40. Choi, J. & Lim, Y.-S. Characteristics, prevention, and management of hepatitis B virus (HBV) reactivation in HBV-infected patients who require immunosuppressive therapy. *J. Infect. Dis.* **216**, S778–S784 (2017).
41. Haidar, G., Boeckh, M. & Singh, N. Cytomegalovirus infection in solid organ and hematopoietic cell transplantation: state of the evidence. *J. Infect. Dis.* **221**, S23–S31 (2020).
42. Diray-Arce, J. et al. Integrative metabolomics to identify molecular signatures of responses to vaccines and infections. *Metabolites* **10**, 492 (2020).
43. Yang, C.-Y. et al. Risk of cytomegalovirus infection in solid organ transplant recipients: a population-based cross-sectional study. *J. Microbiol. Immunol. Infect.* **58**, 537–544 (2025).
44. Tin, A. et al. Serum 6-bromotryptophan levels identified as a risk factor for CKD progression. *J. Am. Soc. Nephrol.* **29**, 1939–1947 (2018).
45. Lemos, L. I. C. et al. *S*-methyl cysteine sulfoxide mitigates histopathological damage, alleviate oxidative stress and promotes immunomodulation in diabetic rats. *J. Complement. Integr. Med.* **18**, 719–725 (2021).
46. Agamah, F. E. et al. Network-based integrative multi-omics approach reveals biosignatures specific to COVID-19 disease phases. *Front. Mol. Biosci.* **11**, 1393240 (2024).
47. Aziz, R. et al. Clinical characteristics of Long COVID patients presenting to a dedicated academic post-COVID-19 clinic in Central Texas. *Sci. Rep.* **13**, 21971 (2023).
48. Thaweethai, T. et al. Development of a definition of postacute sequelae of SARS-CoV-2 infection. *JAMA* **329**, 1934–1946 (2023).
49. Mandel, H. et al. Long COVID incidence proportion in adults and children between 2020 and 2024: an electronic health record-based study from the RECOVER initiative. *Clin. Infect. Dis.* **80**, 1247–1261 (2025).
50. Geng, L. N. et al. 2024 update of the RECOVER-adult long COVID research index. *JAMA* **333**, 694–700 (2024).
51. Sabbaghian, M., Gheitasi, H., Shekarchi, A. A., Tavakoli, A. & Poortahmasebi, V. The mysterious anelloviruses: investigating its role in human diseases. *BMC Microbiol.* **24**, 40 (2024).
52. Grinde, B. Viruses belonging to Anelloviridae or Circoviridae as a possible cause of chronic fatigue. *J. Transl. Med.* **18**, 485 (2020).
53. Mancuso, R. et al. Torque teno virus (TTV) in multiple sclerosis patients with different patterns of disease. *J. Med. Virol.* **85**, 2176–2183 (2013).
54. Focosi, D., Macera, L., Boggi, U., Nelli, L. C. & Maggi, F. Short-term kinetics of Torque teno virus viraemia after induction immunosuppression confirm T lymphocytes as the main replication-competent cells. *J. Gen. Virol.* **96**, 115–117 (2015).

[1]Dell Medical School, The University of Texas at Austin, Austin, TX, USA. [2]Clinical and Data Coordinating Center (CDCC) Precision Vaccines Program, Boston Children's Hospital, Boston, MA, USA. [3]Research Computing, Boston Children's Hospital, Boston, MA, USA. [4]Emory School of Medicine, Atlanta, GA, USA. [5]Department of Immunology and Microbiology, Baylor College of Medicine, Houston, TX, USA. [6]David Geffen School of Medicine at the University of California

Los Angeles, Los Angeles, CA, USA. [7]University of California San Francisco, San Francisco, CA, USA. [8]Chan Zuckerberg Biohub, San Francisco, CA, USA. [9]Baylor College of Medicine and the Center for Translational Research on Inflammatory Diseases, Houston, TX, USA. [10]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [11]Case Western Reserve University and University Hospitals of Cleveland, Cleveland, OH, USA. [12]Drexel University, Drexel University College of Medicine, Philadelphia, PA, USA. [13]Stanford University School of Medicine, Palo Alto, CA, USA. [14]Icahn School of Medicine at Mount Sinai, New York, NY, USA. [15]Oklahoma University Health Sciences Center, Oklahoma City, OK, USA. [16]Oregon Health Sciences University, Portland, OR, USA. [17]University of Arizona, Tucson, AZ, USA. [18]University of Florida, Gainesville, FL, USA. [19]The University of Texas at Austin, Austin, TX, USA. [20]Yale School of Medicine, New Haven, CT, USA. [21]National Institute of Allergy and Infectious Diseases, National Institute of Health, Bethesda, MD, USA. [22]La Jolla Institute for Immunology, La Jolla, CA, USA. [23]Benaroya Research Institute, University of Washington, Seattle, WA, USA. [24]Yale School of Public Health, New Haven, CT, USA. [25]Precision Vaccines Program, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. [26]Boston Children's Hospital and Harvard Medical School, Boston, MA, USA. [27]Broad Institute of MIT and Harvard, Cambridge, MA, USA.

IMPACC Network

Patrice M. Becker[21], Alison D. Augustine[21], Steven M. Holland[21], Lindsey B. Rosen[21], Serena Lee[21], Tatyana Vaysman[21], Al Ozonoff[2], Joann Diray-Arce[2], Jing Chen[2], Alvin T. Kho[2], Carly E. Milliren[2], Annmarie Hoch[2], Ana C. Chang[2], Kerry McEnaney[2], Caitlin Syphurs[2], Brenda Barton[2], Claudia Lentucci[2], Maimouna D. Murphy[2], Mehmet Saluvan[2], Tanzia Shaheen[2], Shanshan Liu[2], Marisa Albert[2], Arash Nemati Hayati[2], Robert Bryant[2], James Abraham[2], Mitchell Cooney[2], Meagan Karoly[2], Matthew C. Altman[23], Naresh Doni Jayavelu[23], Scott Presnell[23], Bernard Kohr[23], Tomasz Jancsyk[23], Azlann Arnett[23], Bjoern Peters[22], James A. Overton[22], Randi Vita[22], Kerstin Westendorf[22], James A. Overton[28], Ofer Levy[25], Hanno Steen[25], Patrick van Zalm[25], Benoit Fatou[25], Kinga K. Smolen[25], Arthur Viode[25], Simon van Haren[25], Meenakshi Jha[25], David Stevenson[25], Athena N. Nguyen[25], Alec L. Plotkin[25], Sanya Thomas[25], Boryana Petrova[25], Naama Kanarek[25], Lindsey R. Baden[10], Kevin Mendez[10], Jessica Lasky-Su[10], Alexandra Tong[10], Rebecca Rooks[10], Michael Desjardins[10], Amy C. Sherman[10], Stephen R. Walsh[10], Xhoi Mitre[10], Jessica Cauley[10], Xiaofang Li[10], Bethany Evans[10], Christina Montesano[10], Jose Humberto Licona[10], Jonathan Krauss[10], Nicholas C. Issa[10], Jun Bai Park Chang[10], Natalie Izaguirre[10], Scott R. Hutton[29], Greg Michelotti[29], Kari Wong[29], Scott J. Tebbutt[30], Casey P. Shannon[30], Rafick-Pierre Sekaly[11], Slim Fourati[11], Grace A. McComsey[11], Paul Harris[11], Scott Sieg[11], George Yendewa[11], Mary Consolo[11], Heather Tribout[11], Susan Pereira Ribeiro[11], Charles B. Cairns[12], Elias K. Haddad[12], Michele A. Kutzler[12], Mariana Bernui[12], Gina Cusimano[12], Jennifer Connors[12], Kyra Woloszczuk[12], David Joyner[12], Carolyn Edwards[12], Edward Lee[12], Edward Lin[12], Nataliya Melnyk[12], Debra L. Powell[12], James N. Kim[12], I. Michael Goonewardene[12], Brent Simmons[12], Cecilia M. Smith[12], Mark Martens[12], Brett Croen[12], Nicholas C. Semenza[12], Mathew R. Bell[12], Sara Furukawa[12], Renee McLin[12], George P. Tegos[12], Brandon Rogowski[12], Nathan Mege[12], Kristen Ulring[12], Pam Schearer[12], Judie Sheidy[12], Crystal Nagle[12], Vicki Seyfert-Margolis[31], Nadine Rouphael[4], Steven E. Bosinger[4], Arun K. Boddapati[4], Greg K. Tharp[4], Kathryn L. Pellegrini[4], Brandi Johnson[4], Bernadine Panganiban[4], Christopher Huerta[4], Evan J. Anderson[4], Hady Samaha[4], Jonathan E. Sevransky[4], Laurel Bristow[4], Elizabeth Beagle[4], David Cowan[4], Sydney Hamilton[4], Thomas Hodder[4], Amer Bechnak[4], Andrew Cheng[4], Aneesh Mehta[4], Caroline R. Ciric[4], Christine Spainhour[4], Erin Carter[4], Erin M. Scherer[4], Jacob Usher[4], Kieffer Hellmeister[4], Laila Hussaini[4], Lauren Hewitt[4], Nina Mcnair[4], Susan Pereira Ribeiro[4], Sonia Wimalasena[4], Ana Fernandez-Sesma[14], Viviana Simon[14], Florian Krammer[14], Harm van Bakel[14], Seunghee Kim-Schulze[14], Ana Silvia Gonzalez-Reiche[14], Jingjing Qi[14], Brian Lee[14], Juan Manuel Carreño[14], Gagandeep Singh[14], Ariel Raskin[14], Johnstone Tcheou[14], Zain Khalil[14], Adriana van de Guchte[14], Keith Farrugia[14], Zenab Khan[14], Geoffrey Kelly[14], Komal Srivastava[14], Lily Q. Eaker[14], Maria C. Bermúdez-González[14], Lubbertus C. F. Mulder[14], Katherine F. Beach[14], Miti Saksena[14], Deena Altman[14], Erna Kojic[14], Levy A. Sominsky[14], Arman Azad[14], Dominika Bielak[14], Hisaaki Kawabata[14], Temima Yellin[14], Miriam Fried[14], Leeba Sullivan[14], Sara Morris[14], Giulio Kleiner[14], Daniel Stadlbauer[14], Jayeeta Dutta[14], Hui Xie[14], Manishkumar Patel[14], Kai Nie[14], Brian Monahan[14], Adeeb Rahman[32], William B. Messer[16], Catherine L. Hough[16], Sarah A. R. Siegel[16], Peter E. Sullivan[16], Zhengchun Lu[16], Amanda E. Brunton[16], Matthew Strand[16], Zoe L. Lyski[16], Felicity J. Coulter[16], Courtney Micheletti[16], Holden Maecker[13], Bali Pulendran[13], Kari C. Nadeau[13], Yael Rosenberg-Hasson[13], Michael Leipold[13], Natalia Sigal[13], Angela Rogers[13], Andrea Fernandes[13], Monali Manohar[13], Evan Do[13], Iris Chang[13], Alexandra S. Lee[13], Catherine Blish[13], Henna Naz Din[13], Jonasel Roque[13], Linda N. Geng[13], Maja Artandi[13], Mark M. Davis[13], Neera Ahuja[13], Samuel S. Yang[13], Sharon Chinthrajah[13], Thomas Hagan[13], Tyson H. Holmes[13], Koji Abe[13], Elaine F. Reed[6], Joanna Schaenman[6], Ramin Salehi-Rad[6], Adreanne M. Rivera[6], Harry C. Pickering[6], Subha Sen[6], David Elashoff[6], Dawn C. Ward[6], Jenny Brook[6], Estefania Ramires-Sanchez[6], Megan Llamas[6], Claudia Perdomo[6], Clara E. Magyar[6], Jennifer Fulcher[6], David J. Erle[7], Carolyn S. Calfee[7], Carolyn M. Hendrickson[7], Kirsten N. Kangelaris[7], Viet Nguyen[7], Deanna Lee[7], Suzanna Chak[7], Rajani Ghale[7], Ana Gonzalez[7], Alejandra Jauregui[7], Carolyn Leroux[7], Luz Torres Altamirano[7], Ahmad Sadeed Rashid[7], Andrew Willmore[7], Prescott G. Woodruff[7], Matthew F. Krummel[7], Sidney Carrillo[7], Alyssa Ward[7], Charles R. Langelier[7], Ravi Patel[7], Michael Wilson[7], Ravi Dandekar[7], Bonny Alvarenga[7], Jayant Rajan[7], Walter Eckalbar[7], Andrew W. Schroeder[7], Gabriela K. Fragiadakis[7], Alexandra Tsitsiklis[7], Eran Mick[7], Yanedth Sanchez Guerrero[7], Christina Love[7], Lenka Maliskova[7], Michael Adkisson[7], Aleksandra Leligdowicz[7], Alexander Beagle[7], Arjun Rao[7], Austin Sigman[7], Bushra Samad[7], Cindy Curiel[7], Cole Shaw[7], Gayelan Tietje-Ulrich[7], Jeff Milush[7], Jonathan Singer[7], Joshua J. Vasquez[7], Kevin Tang[7], Legna Betancourt[7], Lekshmi Santhosh[7], Logan Pierce[7], Maria Tecero Paz[7], Michael Matthay[7], Neeta Thakur[7], Nicklaus Rodriguez[7], Nicole Sutter[7], Norman Jones[7], Pratik Sinha[7], Priya Prasad[7], Raphael Lota[7], Saurabh Asthana[7], Sharvari Bhide[7], Tasha Lea[7], Yumiko Abe-Jones[7], David A. Hafler[20], Ruth R. Montgomery[20], Albert C. Shaw[20], Steven H. Kleinstein[20], Jeremy P. Gygi[20], Dylan Duchen[20], Shrikant Pawar[20], Anna Konstorum[20], Ernie Chen[20], Chris Cotsapas[20], Xiaomei Wang[20], Charles Dela Cruz[20], Akiko Iwasaki[20], Subhasis Mohanty[20], Allison Nelson[20], Yujiao Zhao[20], Shelli Farhadian[20], Hiromitsu Asashima[20], Omkar Chaudhary[20], Andreas Coppi[20], John Fournier[20], M. Catherine Muenker[20], Khadir Raddassi[20], Michael Rainone[20], William Ruff[20], Syim Salahuddin[20], Wade L. Shulz[20], Pavithra Vijayakumar[20], Haowei Wang[20], Esio Wunder Jr.[20], H. Patrick Young[20], Albert I. Ko[20], Gisela Gabernet[20], Denise Esserman[24], Leying Guan[24], Anderson Brito[24], Jessica Rothman[24], Nathan D. Grubaugh[24], Kexin Wang[24], Leqi Xu[24], David B. Corry[9], Farrah Kheradmand[9], Li-Zhen Song[9], Ebony Nelson[9], Jordan P. Metcalf[15], Nelson I. Agudelo Higuita[15], Lauren A. Sinko[15], J. Leland Booth[15], Douglas A. Drevets[15], Brent R. Brown[15], Monica Kraft[17], Christian Bime[17], Jarrod Mosier[17], Heidi Erickson[17], Ron Schunk[17], Hiroki Kimura[17], Michelle Conway[17], Dave Francisco[17], Allyson Molzahn[17], Connie Cathleen Wilson[17], Ron Schunk[17], Trina Hughes[17], Bianca Sierra[17], Mark A. Atkinson[18], Scott C. Brakenridge[18], Ricardo F. Ungaro[18], Brittany Roth Manning[18], Lyle Moldawer[18], Jordan Oberhaus[33], Faheem W. Guirgis[33], Brittney Borresen[34], Matthew L. Anderson[34], Lauren I. R. Ehrlich[19], Esther Melamed[1], Cole Maguire[1], Dennis Wylie[19], Justin F. Rousseau[1], Kerin C. Hurley[1], Janelle N. Geltman[1], Nadia Siles[1], Jacob E. Rogers[1] & Pablo Guaman Tipan[19]

[28]Knocean, Toronto, Ontario, Canada. [29]Metabolon, Morrisville, NC, USA. [30]Prevention of Organ Failure (PROOF) Centre of Excellence, University of British Columbia, Vancouver, British Columbia, Canada. [31]MyOwnMed, Bethesda, MD, USA. [32]Immunai, New York, NY, USA. [33]University of Florida, Jacksonville, FL, USA. [34]University of South Florida, Tampa, FL, USA.

Methods

Study design and participant recruitment

IMPACC is a prospective longitudinal study that enrolled >1,000 hospitalized patients with COVID-19, as previously described[27,32,55–58]. Participants 18 years and older were recruited from 20 hospitals across 15 academic institutes within the USA (Fig. 1a). All participants were confirmed to be SARS-CoV-2-positive by reverse transcription PCR (RT–PCR) testing and no participants were vaccinated for SARS-CoV-2 at time of enrolment. Nasal swabs, blood, and endotracheal aspirate (for ventilated patients) were collected within 72 h of hospital admission (visit 1) and on days 4, 7, 14, 21 and 28 post-hospital admission in addition to convalescent samples at 3, 6, 9 and 12 months. Additionally, nasal swabs, blood and endotracheal aspirate were also collected (when possible) within 24 h and 96 h of escalation to intensive care unit-level care or when a participant was readmitted to the hospital >48 h after discharge. In the IMPACC dataset, these samples were indicated as 'escalation visits', and in this Article, the escalation visit samples were only used for the longitudinal GAMM and the CyTOF cellular association analyses to minimize effects of more frequent sampling for critically ill participants. Participants were characterized into one of five trajectory groups based on latent class mixed modelling of a seven-point ordinal scale that characterized degree of respiratory illness and reflected acute COVID-19 severity[24]. Similarly, participants were clustered into 4 PRO groups from latent class mixed modelling trajectories of convalescent survey responses collected at 3, 6, 9 and 12 months[32]. The modelling used participant responses from the EQ-5D-5L[59], health recovery score (visual analogue scale of 1–100 to indicate overall physical and mental function compared to pre-COVID function), and the following PROMIS forms (https://commonfund.nih.gov/promis): PROMIS Item Bank v2.0—Physical Function, PROMIS Item Bank v2.0—Cognitive Function[60], PROMIS Scale v1.2—Global Health Mental 2a[61], PROMIS Item Bank v1.0—Psychosocial Illness Impact-Positive—Short Form 8a[61], and PROMIS Pool v1.0—Dyspnea Time Extension[62]. All PROMIS measures were scored and standardized following PROMIS standardized instructions. Clinical characteristics and demographics for the entire cohort are reported in Supplementary Table 1.

Ethics statement

The Department of Health and Human Services Office for Human Research Protections (OHRP) and the National Institute of Allergy and Infectious Diseases (NIAID) concurred that the IMPACC study qualified for public health surveillance exemption. The study protocol was sent for review to each site's institutional review board (IRB), with twelve sites conducting as a public health surveillance study, and three sites integrating the IMPACC study into IRB-approved protocols (The University of Texas at Austin, IRB 2020-04-0117; University of California San Francisco, IRB 20-30497; Case Western Reserve University, IRB STUDY20200573) with participants providing informed consent. Participants enrolled at sites operating as a public health surveillance study were provided information sheets describing the study including the samples to be collected and plans for analysis and data de-identification. Participants who requested not to participate after review of the study plan and information were not enrolled. Participants were not compensated while hospitalized but were subsequently compensated for outpatient visits and surveys. This study was registered at clinicaltrials.gov (NCT04378777) and followed the strengthening the reporting of observational studies in epidemiology (STROBE) guidelines (Extended Data Fig. 1a).

Sample processing and assays

Samples were processed as previously described[27,32,55–58], with the sample protocol extensively documented in the IMPACC study design and protocol paper[55]. In brief, 10 ml of blood and nasal swabs were collected at each visit, with blood processed within 6 h of collection. Blood was collected in both a 2.5 ml Greiner Vacuette CAT Serum Separating Tube (SST) (454243P) for serum and a 7.5 ml Sarstedt Venous blood collection monovette EDTA (NC9453456) for whole blood, PBMCs and plasma. The SST was kept vertical at room temperature for at least 30 min before centrifuging at room temperature for 10 min at 1,000*g*. Serum was then aliquoted at 100 µl for downstream assays.

From the EDTA tube, it was briefly inverted to mix before aliquoting 270 µl of whole blood for CyTOF. The 270 µl CyTOF aliquot was added directly to the Maxpar Direct Immune Profiling Assay tube (MDIPA antibodies listed in Supplementary Table 9) and incubated for 30 min at room temperature. After incubation, 410 µl of Smart Tube Prot1 Stabilizer (SmartTube) was added with a 10 min incubation at room temperature before storage at −80 °C until shipment to the respective processing core. The remaining blood was centrifuged at room temperature for 10 min at 1,000*g* before aliquoting and storing 500 µl of plasma at −80 °C for proteomic and metabolomics. PBMCs were then isolated from the remaining sample using the SepMate and Lymphoprep system (StemCell) following manufacturer protocol and as previously described[55]. PBMCs were then stored at 2.5×10^5 cells in 200 µl of RLT Buffer (Qiagen) and β-mercaptoethanol at −80 °C.

Interior nasal turbinate swabs (herein referred to as nasal swabs) were collected and stored in 1 ml of Zymo-DNA/RNA shield reagent (Zymo Research), before RNA was extracted twice in parallel from 250 µl of sample and purified with the KingFisher Flex sample purification system (ThermoFisher) and the quick DNA-RNA MagBead kit (Zymo Research). The duplicated RNA was pooled and aliquoted at 20 µl for the downstream assays (SARS-CoV-2 RT–qPCR and RNA-seq).

When participants were ventilated, an endotracheal aspirate was also collected in a 40 cm^3 Argyle specimen trap and was processed within 2 h of collection. First, 500 µl of 1:1 diluted endotracheal aspirate with Maxpar PBS (Ca^{2+} and Mg^{2+} free) was mixed with 500 µl of DNA/RNA shield in a Zymo tube with lysis beads and subsequently stored at −80 °C for bulk RNA-seq.

Collected samples were then shipped and processed for nasal, PBMC, and endotracheal aspirate RNA-seq, plasma proteomics, serum cytokine PEA, serum EBV and CMV antibody titres, whole-blood CyTOF, and plasma metabolomics at their respective processing cores as previously described[27,32,55–58]. Each assay is described in brief below with additional technical details in the prior publications[27,32,55–58], and a list of analytes evaluated from the PBMC RNA-seq, nasal RNA-seq, serum cytokines and chemokines, and plasma metabolomics assays is reported in Supplementary Table 8.

Nasal, endotracheal aspirate and PBMC RNA-seq

Nasal transcriptomics data were processed using a workflow managed on the Galaxy platform. RNA extracted from the nasal swabs was DNase-treated and depleted of human ribosomal RNA before amplification with random hexamers (Ribo-Zero Plus kit), except for the convalescent samples (3, 6, 9 and 12 months) which instead used poly-A amplification for library prep (SMART-seq V4). Before sequencing, libraries were quantified on both a Quant-it dsDNA High Sensitivity Assay and Fragment Analyzer (Advaced Analytical; kit ID DNF474), with samples containing adapter dimers as more than 4% of electropherogram area failed before sequencing. Technical controls (K562, Thermo Fisher Scientific, AM7832) were compared across batches to ensure minimal batch variability. Libraries were normalized to 10 nM before base calls were generated on the NovaSeq6000 instrument (RTA v3.1.5) at 100 bp paired-end read length, and demultiplexed unaligned BAM files were produced using Picard's ExtractIlluminaBarcodes and IlluminaBasecallsToSam tools (https://broadinstitute.github.io/picard/). These BAM files were converted to FASTQ format using Samtools bam2fq (v1.4)[63]. Adapter trimming and quality filtering were performed using Trimmomatic (v0.36.5)[64], with reads trimmed by one base at the 3′ end and further trimmed from both ends to ensure

a minimum base quality score of Q30. Adapter sequences were also removed. Trimmed reads were aligned to the GRCh38 human reference genome[65] using STAR (v2.4.2a)[66] with gene annotations from Ensembl release 91 (ref. 67). Gene-level counts were generated using HTSeq-count (v0.4.1)[68]. Quality control metrics were compiled using Picard (v1.134), FASTQC (v0.11.3) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and Samtools (v1.2). Samples with poor quality (defined as median coefficient of variation (CV) in gene coverage >0.8 or aligned counts <1 million) were excluded from downstream analysis.

RNA extracted from endotracheal aspirate specimens was DNase-treated and depleted of human ribosomal RNA. Complementary DNA (cDNA) was synthesized using random hexamers to capture both coding and noncoding transcripts. Libraries were sequenced with 100 bp paired-end reads on the NovaSeq6000 instrument with NovaSeq S4 flow cells (200 cycles) targeting 50 million reads per sample. Human reads were aligned to the GRCh38 reference genome[65] and quality-controlled. Raw counts were normalized across libraries using the TMM method implemented in the edgeR package[69]. To control for batch effects, participant and control samples were co-sequenced within each batch.

For the PBMC transcriptomics, RNA was extracted from the 2.5×10^5 cells stored in RLT Buffer (Qiagen) using the Quick-RNA MagBead Kit (Zymo) with DNase digest. RNA quality was then evaluated with both a Quibit HS RNA assay and Fragment Analyzer (Agilent) before cDNA was prepared with the SMART-Seq v4 Ultra Low Input RNA Kit (Takara Bio) from an input of 10 ng of RNA. After a bead-based clean-up, the Nextera XT DNA Library Preparation kit (Illumina) was used to prepare the sequencing libraries which were validated by capillary electrophoresis with a Fragment Analyzer (Agilent). The resulting libraries were pooled at equimolar concentrations and sequenced on an Illumina NovaSeq6000 at 100 bp paired-end read length with a target of 25 million reads per sample. Adapter trimming and quality filtering were performed using Cutadapt (v1.14). Reads were aligned using STAR (v2.4.2a) to a composite reference genome that included the human genome (GRCh38, Ensembl release 91)[65] and SARS-CoV-2 (NCBI strain MN908947.3)[70]. Gene counts were computed using HTSeq-count internally within the STAR alignment step. Quality control metrics were assessed using FASTQC (v0.11.5), Picard tools (v2.22), and STAR log outputs. QC metrics included average base quality per read (>Q30), per cent and absolute counts of reads uniquely mapped to annotated transcripts, and other alignment-based quality statistics.

Taxonomic alignments for human-infecting viruses from the PBMC, nasal and endotracheal aspirate RNA-seq data were obtained from CZID[71], which removes host reads before aligning remaining reads against the National Center for Biotechnology Information (NCBI) nucleotide and non-redundant databases. A sample was considered positive for a virus if it had at least one read that mapped to both the nucleotide and non-redundant database. At least one water control was included on each plate/batch from all transcriptomics, with no water control samples having detectible reads for the human-infecting viruses evaluated in this Article, supporting the low threshold for positivity. When evaluating for potential batch effects of viruses, we identified one plate from the nasal transcriptomics that had an above average rate of HSV1 positivity (>60% compared to the average batch approximately 10%), which may have been due to cross-contamination from a sample in the batch with an extremely high HSV1 viral load (about 250,000 RPM). As a result, we excluded this one plate from contributing to identifying HSV1 positivity and from all host transcriptomic analyses. Additionally, some samples were sequenced multiple times across batches, in the case of a sample sequenced multiple times, the mean RPM of each virus across the samples was used for that participant event.

In addition to the IMPACC PBMC RNA-seq, RNA-seq data were retrieved from the Mount Sinai COVID-19 Biobank (syn35874390)[25,26]. The retrieved raw fastq were processed as outlined above through CZID to identify reads belonging to human-infecting viruses.

Nasal SARS-CoV-2 RT–qPCR

SARS-CoV-2 viral load was assessed from nasal swab samples using RT–qPCR targeting the N1 and N2 regions of the nucleocapsid gene, following the CDC protocol (https://www.cdc.gov/flu/php/laboratories/influenza-sars-cov-2-multiplex-assay.html). Reactions were performed using Quantabio One-Step RT–qPCR ToughMix on a QuantStudio 5 instrument. Cycle threshold (Ct) values for N1 and N2 were used as the primary readout.

Whole-blood CyTOF

Whole-blood CyTOF prepared samples were thawed according to the SmartTube Prot1 Thaw/Erythocyte lysis protocol before samples were barcoded with the Fluidigm Cell-ID 20-Plex Palladium Barcoding Kit. After barcoding, samples were pooled and additional surface antibody staining was performed (antibodies listed in Supplementary Table 9) at a final concentration of 1 µl antibody per 10 million cells in a volume scaled to cell number (100 µl for every 10 million cells), with a 30 min incubation on ice. After surface staining, samples were washed twice (each 1 ml of CyFACS (1× PBS + 0.2% BSA + 0.05% NaN_3) at 800*g* × 3 min) and subsequently fixed and permeabilized with BD Biosciences Fixation/Permeabilization solution (100 µl per 5 million cells) with a 20 min incubation on ice. Samples were then washed twice with 1× BD Perm Wash buffer (800*g* × 3 min) before staining with an intracellular antibody for GZMB (Supplementary Table 9) at a final concentration of 1 µl antibody per 10 million cells in a volume scaled to cell number (100 µl for every 10 million cells), with a 30 min incubation on ice. Finally, samples were fixed with paraformaldehyde and simultaneous iridium/osmium cell labelling. CyTOF samples were then acquired using the Fluidigm Helios mass cytometer and normalized/concatenated with Fluidigm's CyTOF software. Further cleaning was done using Mt Sinai's internal pipeline, which removed acquisition outliers, EQ beads, and low DNA signal events. Demultiplexing was performed using Pd barcoding and cosine similarity, removing low signal-to-noise cells and acquisition multiplets. Each sample was clustered into 1,000 *k*-means groups. A subset was manually annotated via Clustergrammer2 (https://github.com/ismms-himc/clustergrammer2) to generate a reference matrix. For annotation, cluster similarity to reference cell types was computed, and assignments were made based on highest or consensus similarity. Cell-type labels were then mapped back to single cells for downstream quantification. Antibodies used are listed in the supplemental reporting summary.

Plasma proteomics

Plasma samples underwent protein depletion using perchloric acid to remove the most abundant proteins, enhancing the detection of lower abundance proteins[72–74]. The prepared samples were loaded onto Evotips and analysed using the EVOSEP One system (EVOSEP). The system operated using the 60 samples per day method, which corresponds to a 21 min gradient, optimizing throughput without compromising data quality. The EVOSEP One was coupled to a timsTOF Pro mass spectrometer (Bruker Daltonics) operating in Data-Dependent Acquisition Parallel Accumulation–Serial Fragmentation (DDA-PASEF) mode. HSV1 proteins evaluated included the Swiss-Prot reviewed proteins attributed to human herpesvirus 1 (strain 17, UniProt Proteome ID: UP000009294).

Plasma metabolomics

Plasma metabolite profiling was performed by Metabolon using their in-house standards[75,76]. Samples were randomized, extracted with methanol precipitation, and divided into fractions for analysis via UPLC–MS/MS under both positive and negative ion modes using RP and HILIC chromatography. Analyses were conducted with a Thermo

Q-Exactive mass spectrometer. Metabolites were identified based on retention index, accurate mass (±10 ppm), and MS/MS spectral matching to Metabolon's reference library, following Metabolomics Standards Initiative guidelines[76].

Serum cytokines and chemokines

All samples were analysed using the Olink Inflammation panel (Olink Bioscience), which measures 92 inflammation-related proteins via PEA. In brief, oligonucleotide-labelled antibody pairs bind to target proteins, enabling formation of PCR targets upon proximity. After overnight incubation at 4 °C, PCR amplification was performed, and protein levels were quantified using a microfluidic qPCR system (Biomark, Fluidigm), including built-in controls for quality assurance.

Serum EBV and CMV antibody titre

Serum anti-viral antibodies for EBV (gp350) and CMV (gB) were measured via a Luminex platform as previously described[58]. The gp350 and gB antigens (Sino Biological) were conjugated to barcoded beads as recommended by Luminex. Serum samples were diluted 1:400 in PBS/0.5% Triton X-100 before 25 µl was added to an assay plate containing the antigen-coupled beads and incubated on an orbital shaker at 500–600 rpm for 2 h at room temperature. Each well also had Assay Chex Control beads (Radix Biosolutions) to measure non-specific binding. The plate was then washed with a Bio-Tak Magnetic washer (ELX-405, Bio-Tek) before a secondary antibody goat-anti human IgG (Fc fragment, NC9822979) or Anti-IgA Fc coupled to Phycoerythrin (501941614) were diluted and added to the plate and incubated on an orbital shaker at 500–600 rpm for 30 min at room temperature. The plate was then washed again with again with a Bio-Tak Magnetic washer (ELX-405, Bio-Tek) before 130 µl of wash buffer was added to the wells and read on a Luminex Flex 3D instrument (lower bound of 50 beads per target antigen). This assay was only performed for half of the cohort (n = 479). The data for the EBV titre values were normalized by regressing the values of the four different Assay Chex control beads as well as the batch and plate before analysis.

For an additional 497 participants, CMV serostatus at visit 1 was also measured using a CMV IgG ELISA assay (Aviva, GWB-BQK12C). Serum samples were assessed in duplicate, with an average antibody index value >1.1 considered positive, <0.9 considered negative, and values between 0.9 and 1.1 equivocal per manufacturer specifications. For this study, equivocal status was considered seropositive. The resulting calculated serostatus from both CMV assays were combined to evaluate rate of serostatus at hospital admission between CMV transcript positive and negative patients (Fig. 3b).

Statistics and reproducibility

All IMPACC sites followed a standard protocol[55] for biological sample collection. To mitigate batch effects, samples were randomized to batches while ensuring longitudinal samples from the same individual were run in the same batch. The randomization was stratified by disease severity (mild and moderate versus severe) and age (younger versus older) with representation across batches. Furthermore, race, ethnicity, gender, and enrolment site were then confirmed to be well represented across batches. As this was an observational study, we prioritized biological replicates over technical replicates, except where needed to evaluate for batch effects. Thus, all data used in this study were reflective of data collected as single measurements for each participant at each time point, unless otherwise specified in the methods.

All P values calculated in this manuscript were adjusted using the Benjamini–Hochberg procedure and are indicated as adjusted P values (circumstances where no adjustments were necessary report as only P values). Across all multi-omic analyses, adjustments were corrected for all comparisons or models ran on a per virus basis. The exact statistical approaches used for each analysis are detailed below.

Clinical features and demographics. For testing the association of detection of chronic viral transcripts with trajectory groups and age, we used cumulative link mixed modelling (clmm) from the ordinal (v2019.12-10) R package[77]. Due to both COVID-19 severity and age quantiles being ordinal, cumulative link mixed modelling allowed for this ordinal relationship to be accounted for in addition to including enrolment site as a random effect. However, due to the TG4 and TG5 contributing a majority of the endotracheal aspirate samples, for comparison of the endotracheal aspirate viruses a Fisher's exact test comparing only TG4 and TG5 was used instead of a clmm. The following R formulae were used with the clmm2 function from the ordinal (v 2019.12-10) R package[77]:

```
Trajectory_group ~ virus_status, random = enrollment_site

Admit_age_quintile ~ virus_status + trajectory_group, random
  = enrollment_site
```

For the testing of association of virus positivity status with ethnicity, sex, and steroid and remdesivir administration in Extended Data Fig. 6b–e, a chi-squared test of independence was calculated using the rstatix (v0.7.2) R package[78].

For the testing of virus positivity status with long-term mortality within trajectory group 4, a Cox proportional hazards model was used that included steroid and remdesivir administration as fixed effects and enrolment site as a random effect. All clinical outcomes, including mortality, were captured by clinical staff from the electronic medical record in accordance with protocol and then verified by the IMPACC Clinical and Data Coordinating Center. If death was not ascertained, the model was right censored with the date of participant lost to follow up or completion of the study. The following R formula was used with the coxme function from the coxme R package (v2.2-16)[79]:

```
Surv(time_to_event, right_censoring) ~ ever_steroids
  + ever_remdesivir + virus_status + (1|enrollment_site)
```

For testing of association with other clinical features including complications, comorbidities and medication usage, a logistic mixed effect model was used that also included sex, age quintile and trajectory group as fixed effects and enrolment site as a random effect. Of note, participants positive for a virus in a given tissue were only compared against participants who also had the respective tissue sequenced at least once and were negative at all time points checked. The following R formula was used with the glmer function from the lme4 R package (v 1.1-28)[80]:

```
Clinical_feature ~ virus_status + sex + age_quintile
  + trajectory_group + (1|enrollment_site)
```

To further decouple association of *Anelloviridae* with COVID-19 severity, immunosuppressive medications and age, we conducted an additional logistic mixed effect model that included sex, age quintile, trajectory group and immunosuppressive medication status as fixed effects with enrolment site as a random effect. The following R formula was used with the glmer function from the lme4 R package (v 1.1-28)[80]:

```
Anelloviridae_status ~ ever_immsupp + trajectory_group
  + sex + discretized_admit_age_quantile + (1|enrollment_site)
```

For testing of association of virus detection status with the PASC PRO groups, a linear model was used that also included sex, age quintile, trajectory group and immunosuppressed status by medication as fixed effects in the model. The following R formula was used with the glmer function from the lme4 R package (v 1.1-28)[80]:

PRO_group ~ virus_status + immunosuppressed_status
+ trajectory_group + sex + age_quintile + (1|enrollment_site)

Cellular immunophenotyping. For testing the association of participant events positive for viruses with changes in immunophenotypes of circulating cells, a linear mixed effect model was used that also included trajectory group, sex, age quintile, and visit number as additional main effects in addition to enrolment site and participant as nested random effects in the model. The normalized abundances for cell types were computed by calculating the relative abundance after removing granulocytes from the total before a log1p transformation and scaling. The following R formula was used with the lme function from the nlme (v 3.1-149) R package[81]:

Normalized_celltype_abundance ~ virus_status + trajectory_group
+ sex + discretized_admit_age_quantile + event_type, random
= ~ 1|enrollment_site/participant_id

Serum cytokines and plasma metabolomics. To evaluate both the serum proximity extension array cytokine assay (Olink Inflammation panel) and the plasma metabolomics, GAMM from the gamm4 (v 0.2-6) R package[82] was used to evaluate for differences in individual analytes between patients who had detected transcripts for a chronic virus compared to patients who never had human-infecting viral transcripts detected (other than for SARS-CoV-2). Analytes were modelled against days from admission using cubic regression splines with interactions of both status for a given chronic virus (that is, detected transcript at any collected time point) and trajectory group in addition to fixed effects of status for the chronic virus being evaluated, trajectory group, sex, age at time of admission sorted into quintiles, and SARS-CoV-2 nasal viral RPM at the time of that sample. As chronic viral status is both a main effect and interaction term in the model, we used a lower adjusted *P* value cutoff of 0.01 to account for the fact that a feature was significant if either term was significant. Only participant visits with both PBMC and nasal transcriptomics were used for this analysis. The following R formula was used with the gamm function from the gamm4 (v0.2-6) R package[82]:

Analyte ~ s(days, bs = 'cr') + s(days, bs = 'cr', by = 'virus_status')
+ s(days, bs = 'cr', by = 'trajectory_group') + virus_status
+ sex + trajectory_group + age_quintile
+ sarscovs2_nasal_log_rpm, random
= ~ (1|enrollment_site/participant_id)

Nasal and PBMC RNA-seq. To analyse the signature of host gene expression associated with chronic viruses, the limma (v 3.46.0) R package[83] was used for both the nasal and PBMC RNA-seq to evaluate differential expressed genes associated with participants who had detectable chronic viral transcripts. The following R formula was used:

~age_quintile + sex + visit_number + trajectory_group
+ sarscov2_nasal_log_rpm + virus_status

Upregulated and downregulated differentially expressed genes were then evaluated separately with hypergeometric pathway enrichment using the Reactome database (v95)[84] and the clusterProfiler R package (v 3.18.1)[85].

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 aims to support broad data dissemination while safeguarding participant privacy and data integrity through de-identification and masking of sensitive data fields. All IMPACC data, including those produced as part of this study, have been submitted to the Immunology Database and Analysis Portal (ImmPort), a repository supported by the NIAID Division of Allergy, Immunology and Transplantation under accession code SDY1760, as well as to the NLM's Database of Genotypes and Phenotypes (dbGaP) under accession number phs002686.v2.p2. Both raw and processed datasets are available through restricted access in accordance with the NIH public data sharing policy for IRB-exempted public health surveillance studies. Access may be requested through AccessClinicalData@NIAID (https://accessclinicaldata.niaid.nih.gov/study-viewer/clinical_trials) with typical response times of approximately 2–4 weeks. Further detailed guidance on data access is provided on ImmPort (https://docs.immport.org/home/impaccslides). For alignment of host transcripts from the PBMC, nasal and endotracheal aspirate RNA-seq, the GRCh38 reference genome and Ensembl release 91 were used. The Reactome database (v95) was used for hypergeometric enrichment of differentially expressed genes. The CZID pipeline used the NCBI nucleotide (nt) and NCBI non-redundant (nr) databases. HSV1 proteins for the plasma mass spectrometry were retrieved from UniProt for Swiss-Prot reviewed proteins attributed to human herpesvirus 1 (strain 17, UP000009294). Publicly available data for Extended Data Fig. 5a–c were retrieved from the Mount Sinai COVID-19 Biobank which is available in the data repository Synapse (SynID: syn35874390, https://www.synapse.org/Synapse:syn35874390/wiki/618989). Source data are provided with this paper.

Code availability

All code for this study is publicly available in a Bitbucket repository https://bitbucket.org/kleinstein/impacc-public-code/src/master/chronic_viruses_manuscript/ and deposited in Zenodo (https://doi.org/10.5281/zenodo.19657241)[86].

55. IMPACC. Immunophenotyping assessment in a COVID-19 cohort (IMPACC): A prospective longitudinal study. *Sci. Immunol.* **6**, eabf3733 (2021).
56. Gygi, J. P. et al. Integrated longitudinal multiomics study identifies immune programs associated with acute COVID-19 severity and mortality. *J. Clin. Invest.* **134**, e176640 (2024).
57. Phan, H. V. et al. Host–microbe multiomic profiling reveals age-dependent immune dysregulation associated with COVID-19 immunopathology. *Sci. Transl. Med.* **16**, eadj5154 (2024).
58. Rosenberg-Hasson, Y. et al. Relationship of heterologous virus responses and outcomes in hospitalized COVID-19 patients. *J. Immunol.* **211**, 1224–1231 (2023).
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. 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).
63. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
64. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
65. Schneider, V. A. et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.* **27**, 849–864 (2017).
66. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
67. Cunningham, F. et al. Ensembl 2019. *Nucleic Acids Research* **47**, D745–D751 (2019).
68. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
69. Chen, Y., Chen, L., Lun, A. T. L., Baldoni, P. L. & Smyth, G. K. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Res.* **53**, gkaf018 (2025).
70. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

71. Kalantar, K. L. et al. IDseq-An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *Gigascience* **9**, giaa111 (2020).
72. Viodé, A. et al. Plasma proteomic analysis distinguishes severity outcomes of human ebola virus disease. *mBio* **13**, e0056722 (2022).
73. Viode, A. et al. Longitudinal plasma proteomic analysis of 1117 hospitalized patients with COVID-19 identifies features associated with severity and outcomes. *Sci. Adv.* **10**, eadl5762 (2024).
74. Viode, A. et al. A simple, time- and cost-effective, high-throughput depletion strategy for deep plasma proteomics. *Sci. Adv.* **9**, eadf9717 (2023).
75. Long, T. et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat. Genet.* **49**, 568–578 (2017).
76. Evans, A. M., DeHaven, C. D., Barrett, T., Mitchell, M. & Milgram, E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal. Chem.* **81**, 6656–6667 (2009).
77. Christensen, R. H. B. ordinal: regression models for ordinal data. *R package* https://cran.r-project.org/web/packages (2024).
78. Kassambara, A. rstatix: pipe-friendly framework for basic statistical tests. *R package* https://cran.r-project.org/web/packages/rstatix (2023).
79. Therneau, T. M. coxme: mixed effects Cox models. *R package* https://cran.r-project.org/web/packages/coxme/index.html (2024).
80. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
81. Pinheiro, J. et al. nlme: linear and nonlinear mixed effects models. *R package* https://cran.r-project.org/web/packages/nlme (2024).
82. Wood, S. & Scheipl, F. gamm4: generalized additive mixed models using 'mgcv' and 'lme4'. *R package* https://cran.r-project.org/web/packages/gamm4 (2020).
83. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
84. Gillespie, M. et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **50**, D687–D692 (2022).
85. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
86. Maguire, C., Morse, B. A. & Melamed E. Virus reactivation in acute and long COVID-19 (v 1.0). *Zenodo* https://doi.org/10.5281/zenodo.19657241 (2026).

Acknowledgements We thank the participants of the study for their voluntary enrolment and contribution of samples for this work. Details on the IMPACC Network are provided in the Supplementary Information. We thank S. Thomas, M. Cooney, S. Rao, S. Vignolo, E. Morrocchi, A. Naeim, M. Bernardo, S. Sanchez, S. Intluxay, C. Magyar, J. Brook, E. Ramires-Sanchez, M. Llamas, C. Perdomo, Clara E. Magyar and Jennifer A. Fulcher; members of the UCLA Center for Pathology Research Services and the Pathology Research Portal; M. C. Muenker, D. Duvilaire, M. Kuang, W. Ruff, K. Raddassi, D. Shepherd, H. Wang, O. Chaudhary, S. Salahuddin, J. Fournier, M. Rainone and M. Kuang; and the leadership of Boston Children's Hospital, including W. Chung, G. Fleisher and K. Churchwell for their support for the Precision Vaccines Program. Co-authorship of this report by A.D.A. and P.M.B. does not necessarily represent the official views of the National Institute of Allergy and Infectious Diseases, the National Institutes of Health or any other agency of the United States Government.

Author contributions C.M., C.R.L. and E.M. conceived the study. C.M., J.C., B.A.M. and A.H. conducted formal analysis. C.M., J.C. and B.A.M. designed software. C.M., C.R.L. and E.M. designed the study methodology. The IMPACC Network acquired funding, collected samples, and generated data used in this study. Supervision was provided by N.R., K.K.S., E.F.R., O.L., H.M., P.H., H.S., J.D.-A., C.R.L. and E.M. C.M., J.C., N.R., B.A.M., A.H., H.P., H.V.P., A.G., V.C., R.D., D.C., F. Kheradmand, L.R.B., R.-P.S., G.A.M., E.K.H., C.B.C., B. Pulendran, A.F.-S., V.S., J.P.M., N.I.A.H., W.B.M., M.M.D., K.C.N., M.K., C.B., J.S., D.E., C.S.C., M.A.A., S.C.B., L.I.R.E., R.R.M., A.S., C.L.H., D.H., A.D.A., P.M.B., B. Peters, A.O., S.K.-S., F. Krammer, S.E.B., W.E., M.C.A., M.W., L.G., S.H.K., K.K.S., E.F.R., O.L., H.M., P.H., H.S., J.D.-A., C.R.L. and E.M. edited and reviewed the manuscript.

Funding C.M. discloses funding from National Institutes of Health's National Institute on Drug Abuse (5T32DA018926-18). E.M. and L.I.R.E. disclose funding from National Institutes of Health's National Institute of Allergy and Infectious Diseases (5R01AI104870-07). J.C., A.H., L.R.B., A.O., K.K.S., O.L., H.S. and J.D.-A. disclose funding from National Institutes of Health's National Institute of Allergy and Infectious Diseases (5U19AI118608-04). N.R., R.-P.S. and S.E.B. disclose funding from National Institutes of Health's National Institute of Allergy and Infectious Diseases (4U19AI090023-11). H.C.P., J.S. and E.F.R. disclose funding from National Institutes of Health's National Institute of Allergy and Infectious Diseases (5U19AI128913-03). H.V.P., R.D., D.J.E., C.S.C., W.E., M.W., P.H. and C.R.L. disclose funding from National Institutes of Health's National Institute of Allergy and Infectious Diseases (3U19AI077439-13). D.B.C. and F. Kheradmand disclose funding from National Institutes of Health's National Institute of Allergy and Infectious Diseases (5R01AI135803-03). E.K.H. and C.B.C. disclose funding from National Institutes of Health's National Institute of Allergy and Infectious Diseases (5U19AI128910-04). B. Pulendran, J.P.M., N.I.A.H., W.B.M., M.M.D., K.C.N. and H.M. disclose funding from National Institutes of Health's National Institute of Allergy and Infectious Diseases (5U19AI057229-18). A.F.S., V.S., S.K.-S. and F. Krammer disclose funding from National Institutes of Health's National Institute of Allergy and Infectious Diseases (4U19AI118610-06). C.B. and M.K. disclose funding from National Institutes of Health's National Institute of Allergy and Infectious Diseases (5U19AI125357-05). M.A.A. and S.C.B. disclose funding from National Institutes of Health's National Institute of Allergy and Infectious Diseases (5U54AI142766-03, P01AI042288) and National Institute of Diabetes, Digestive, and Kidney Disease (R01DK130425). R.R.M., A.S., D.A.H., L.G. and S.H.K. disclose funding from National Institutes of Health's National Institute of Allergy and Infectious Diseases (AI089992). C.L.H. discloses funding from National Institutes of Health's National Institute of Allergy and Infectious Diseases (R01AI145835-01A1S1). E.M. discloses funding from National Institutes of Health's National Institute on Alcohol Abuse and Alcoholism (K08 AA027837-05). G.A.M. discloses funding from National Institutes of Health's National Center for Advancing Translational Sciences (UM1TR004528). A.D.A. and P.M.B. disclose funding from the Division of Intramural Research of the National Institute of Allergy and Infectious Diseases. M.C.A. discloses funding from the National Institutes of Health's National Institute of Allergy and Infectious Diseases (5U19AI167891-05 and 5R01AI132774-03). V.C. discloses funding from the National Institutes of Health's National Institute of Allergy and Infectious Diseases (1K23AI185326). The IMPACC Network discloses funding from National Institutes of Health's National Institute of Allergy and Infectious Diseases (5R01AI104870-07, 5R01AI135803-03, 3U19AI077439-13, AI089992, 4U19AI090023-11, 5U19AI118608-04, 4U19AI118610-06, 5U19AI057229-18, 5U19AI125357-05, 3U19AI128913-03, 5U19AI128913-03, 5U19AI128910-04, 5U54AI142766-03 and R01AI145835-01A1S1). B.A.M., A.G. and B. Peters. declare no relevant funding.

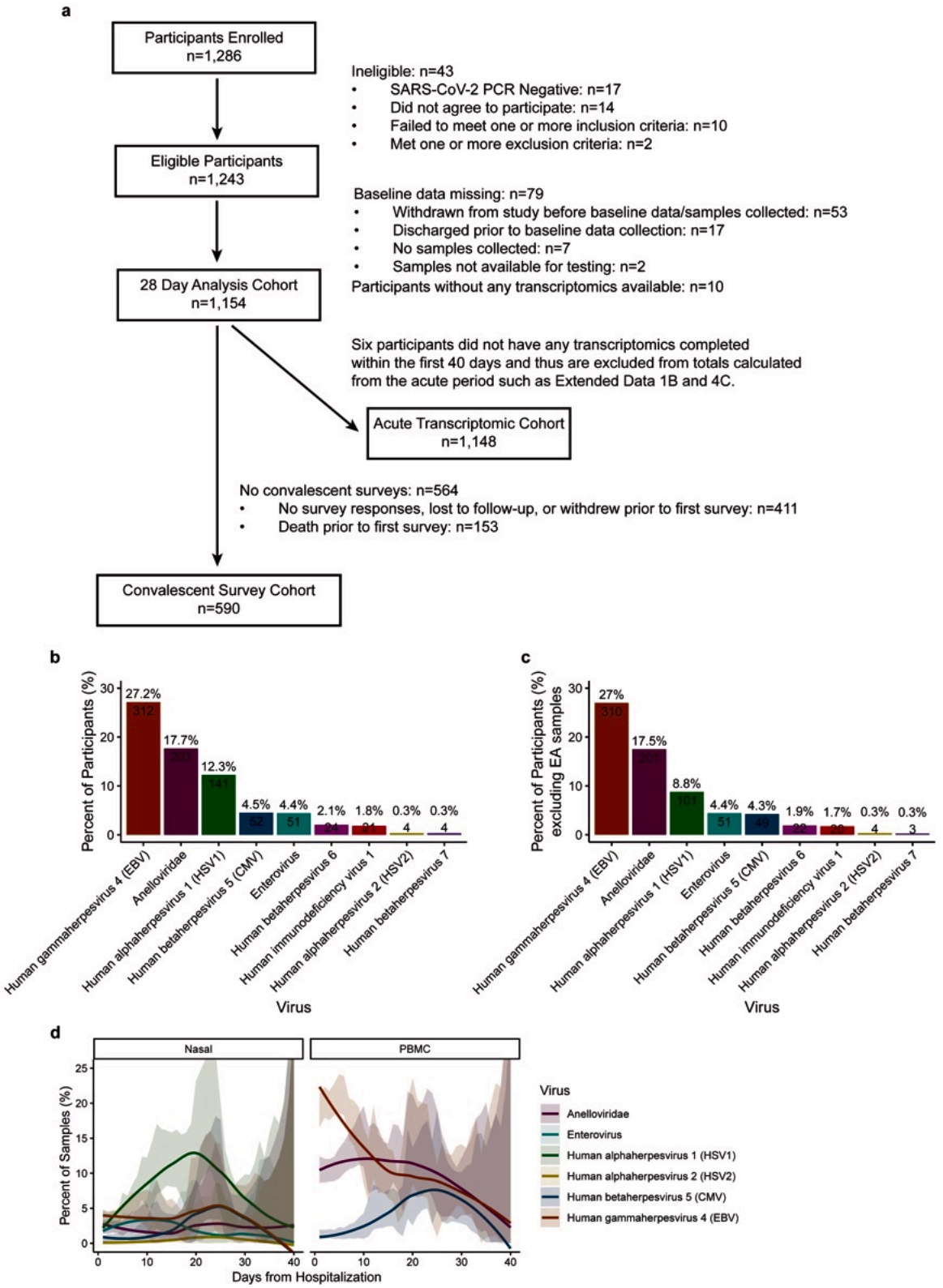
Competing interests The Icahn School of Medicine at Mount Sinai has filed patent applications relating to SARS-CoV-2 serological assays and NDV-based SARS-CoV-2 vaccines which list F. Krammer as co-inventor. Mount Sinai has spun out a company, Kantaro, to market serological tests for SARS-CoV-2. F. Krammer has consulted for Merck and Pfizer (before 2020), and is currently consulting for Pfizer, Seqirus, 3rd Rock Ventures, Merck and Avimex. The Krammer laboratory is also collaborating with Pfizer on animal models of SARS-CoV-2. V. Simon is a co-inventor on a patent filed relating to SARS-CoV-2 serological assays. O.L. is a named inventor on patents held by Boston Children's Hospital relating to vaccine adjuvants and human in vitro platforms that model vaccine action. His laboratory has received research support from, and he is a consultant to, GlaxoSmithKline (GSK). He is a co-founder of and advisor to ARMR Sciences. C.B.C. serves as a consultant to bioMerieux and is funded for a grant from Bill & Melinda Gates Foundation. J.A.O. is a consultant at Knocean Inc. J.L.-S. serves as a scientific advisor of Precion Inc. S.R.H., G.M. and K.W. are employees of Metabolon Inc. V.S.-M. is a current employee of MyOwnMed. N.R. reports grants or contracts with Merck, Sanofi, Pfizer, Vaccine Company, Quidel, Lilly and Immorna, and has participated on data safety monitoring boards for Moderna, Sanofi, Seqirus, Pfizer, EMMES, ICON, BARDA, Imunon, CyanVac and Micron. N.R. has also received support for meetings/travel from Sanofi and Moderna and honoraria from Virology Education. A.R. is a current employee of Immunai Inc. S.K. is a consultant related to ImmPort data repository for Peraton. N.G. is a consultant for Tempus Labs and the National Basketball Association. A.I. is a consultant for 4BIO, Blue Willow Biologics, Revelar Biotherapeutics, RIGImmune, Xanadu Bio and Paratus Sciences. M.K. receives research funds paid to her institution from NIH, ALA; Sanofi, Astra-Zeneca for work in asthma, serves as a consultant for Astra-Zeneca, Sanofi, Chiesi, GSK for severe asthma; and is a co-founder and CMO for RaeSedo, Inc., a company created to develop peptidomimetics for treatment of inflammatory lung disease. E.M. received research funding from Babson Diagnostics and honorarium from Multiple Sclerosis Association of America and has served on the advisory boards of Genentech, Horizon, Teva and Viela Bio. C.S.C. receives research funding from NIH, FDA, DOD, Roche-Genentech and Quantum Leap Healthcare Collaborative as well as consulting services for Janssen, Vasomune, Gen1e Life Sciences, NGMBio and Cellenkos. W.S. was an investigator for a research agreement, through Yale University, from the Shenzhen Center for Health Information for work to advance intelligent disease prevention and health promotion; collaborates with the National Center for Cardiovascular Diseases in Beijing; is a technical consultant to Hugo Health, a personal health information platform; co-founder of Refactor Health, an AI-augmented data management platform for health care; and has received grants from Merck and Regeneron Pharmaceutical for research related to COVID-19. G.A.M. received research grants from Redhill, Cognivue, Pfizer and Genentech, and served as a research consultant for Gilead, Merck, Viiv/GSK and Jenssen. L.N.G. received research funding paid to her institution from Pfizer, Inc. The other authors declare no competing interests.

Additional information

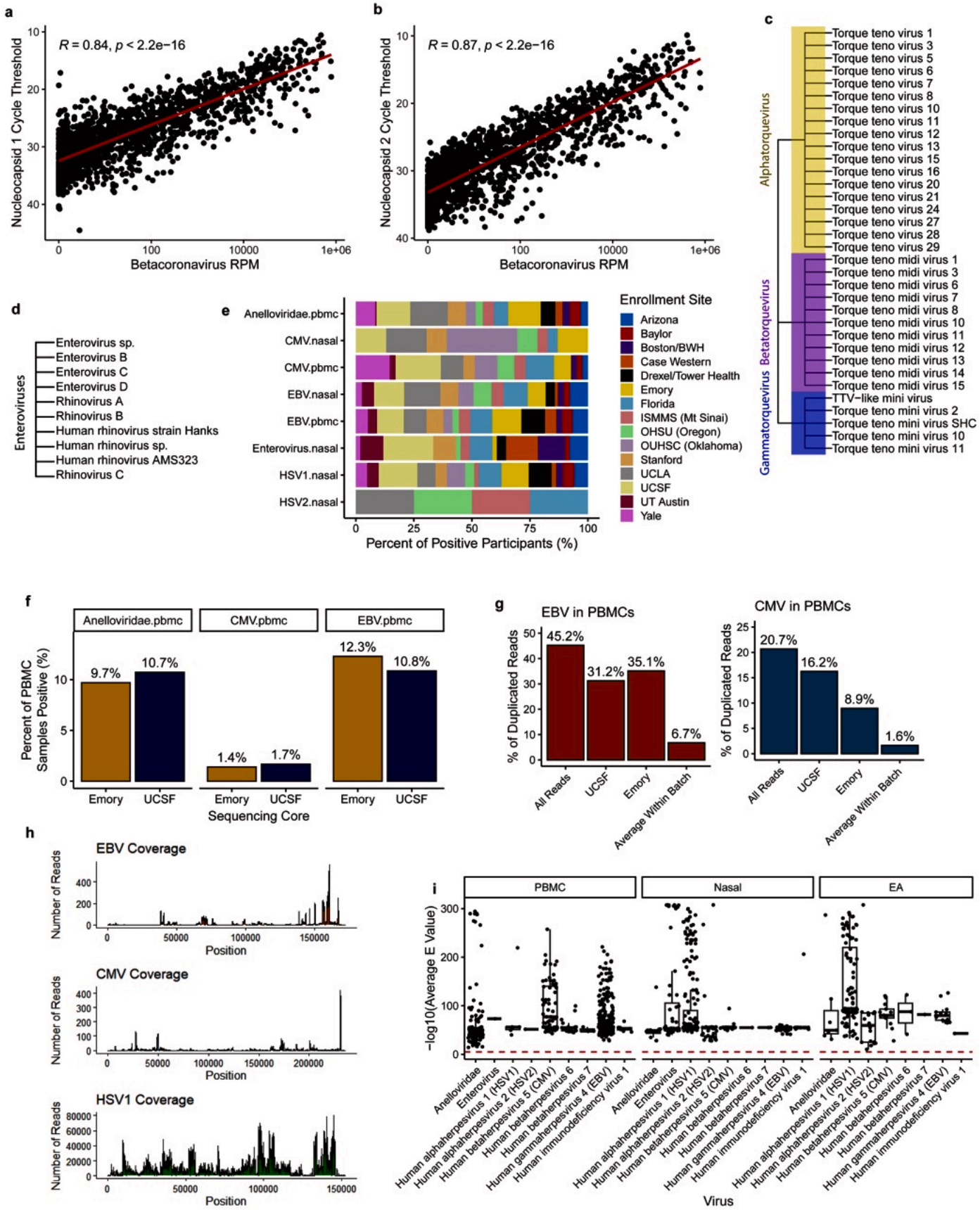
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-026-10740-z.

Correspondence and requests for materials should be addressed to Esther Melamed.

Peer review information *Nature* thanks Emma Thomson who co-reviewed with Shirin Ashraf; Leif Sander, Kendrick Li and the other, anonymous, reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

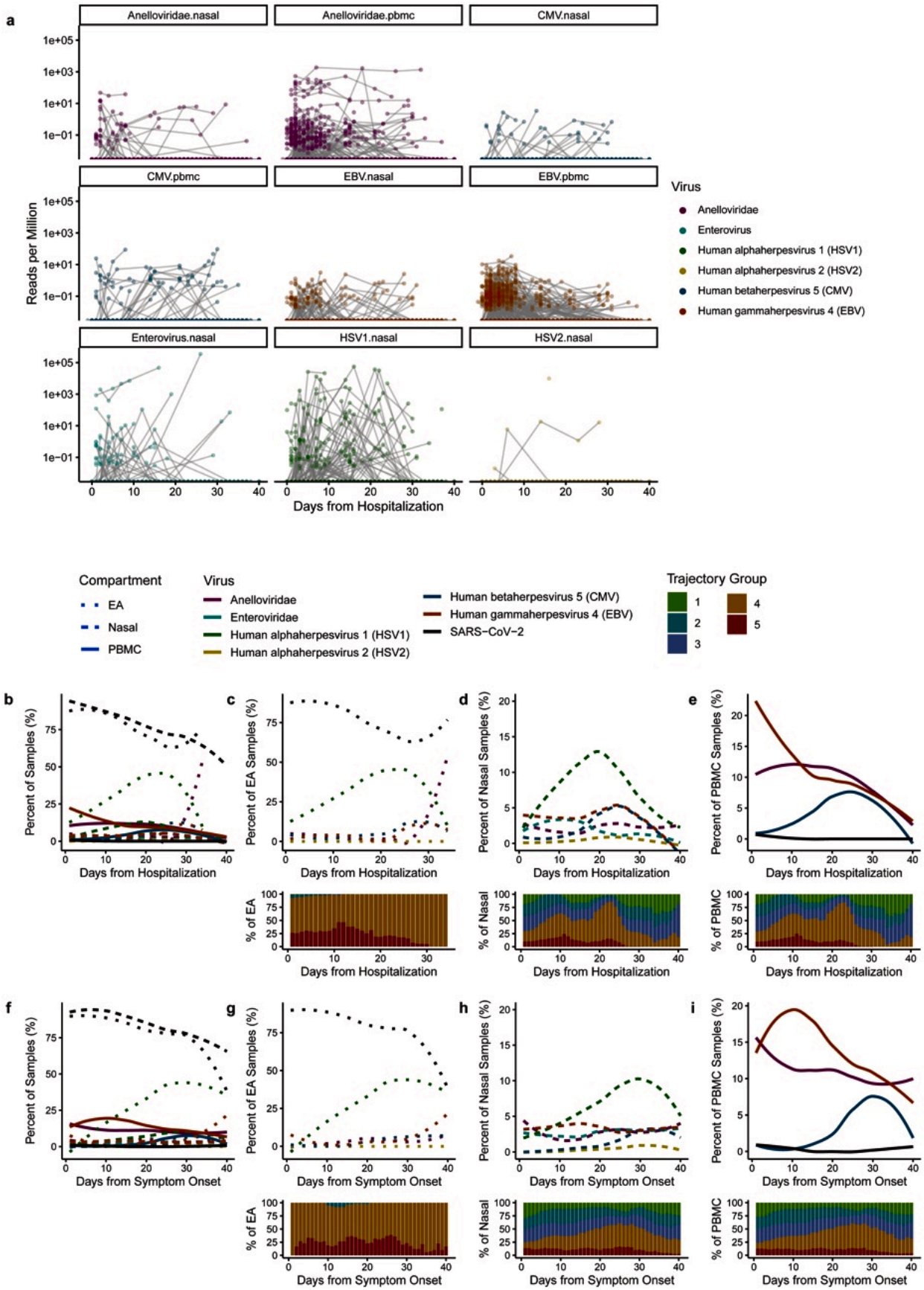


Extended Data Fig. 1 | Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Cohort Diagram and Chronic Virus Prevalence in IMPACC Cohort. a) STROBE cohort diagram modified from Ozonoff A. et al. 2022. Percent of IMPACC participants with detected viruses besides SARS-CoV-2 in **b**) any transcriptomic sample or **c**) in only PBMC or nasal samples (excluding EA samples) ≤ 40 days post-hospitalization (n = 1148). **d**) Smoothed curves demonstrating the percent of total samples that were positive for six common viruses in the nasal and PBMC transcriptomics (a version of Fig. 1d with 95% confidence intervals). Curves were calculated by the percent of samples positive for each day ± two days (a rolling window approach), followed by a local polynomial regression fitting, denoted as the central solid line in the graph. Shaded region denotes 95% confidence interval which were calculated for each day based off the same day ± two days rolling window approach, but without subsequent smoothing.



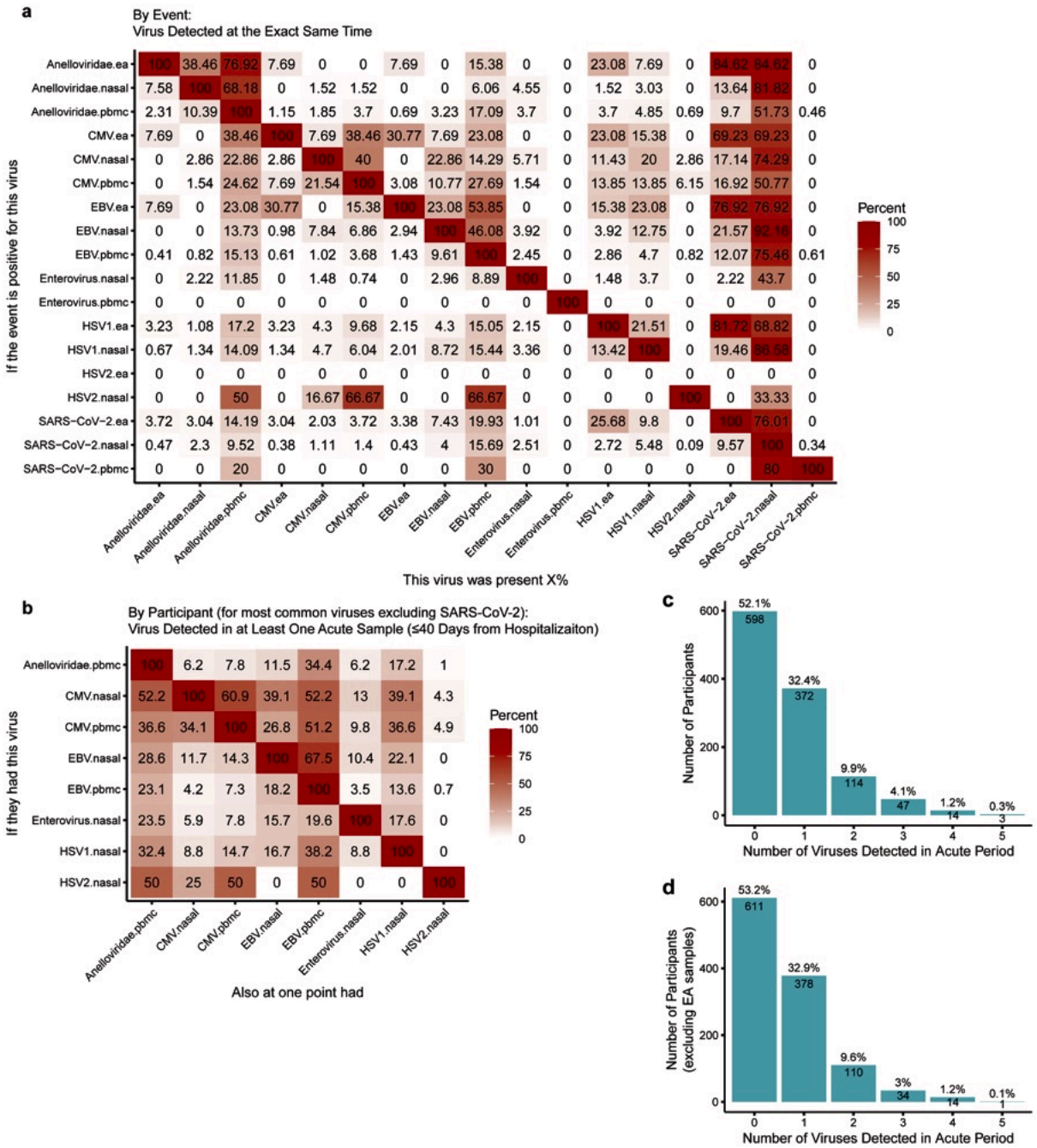
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Diverse Chronic Viral Reactivation Occurs in Acute COVID-19. Pearson correlation (two-sided) of Betacoronavirus reads per million (RPM) from the nasal transcriptomics with **a**) Nucleocapsid 1 cycle threshold and **b**) Nucleocapsid 2 cycle threshold. **c**) Cladogram of the detected *Anelloviridae* in the PBMC transcriptomics (these were collectively collapsed in *Anelloviridae* for all analyses). **d**) Cladogram of the detected enteroviruses in the nasal transcriptomics (these were collectively collapsed in enterovirus for all analyses). **e**) Percent of participants positive for each virus that belonged to each recruitment site. **f**) Percent of PBMC RNA-sequencing samples positive for the most common PBMC viruses between the two sequencing sites of the study demonstrating internal reproducibility. **g**) Percent of reads duplicated across all reads in the study, within either PBMC sequencing site (UCSF or Emory), and the average rate of duplication within each batch. **h**) Read alignment of all reads for EBV and CMV in the PBMC RNA-sequencing and HSV1 in the nasal RNA-sequencing reveals specific gene expression explaining rate of duplicated reads in (**g**). **i**) Average E value of read alignments for a virus in each sample relative to the standard 1e-5 threshold (dashed red line). Of note, E values for HSV1 and HSV2 are shown prior to reassignment based on total ratio of reads found in the sample (see Methods). Boxplots denote median (center line), interquartile range (box), and 1.5x the interquartile range (whiskers). n = 675, 386, and 122 for total number of PBMC, Nasal, and EA samples analyzed and which contributed average E values for one of the plotted viruses in (**i**) respectively.

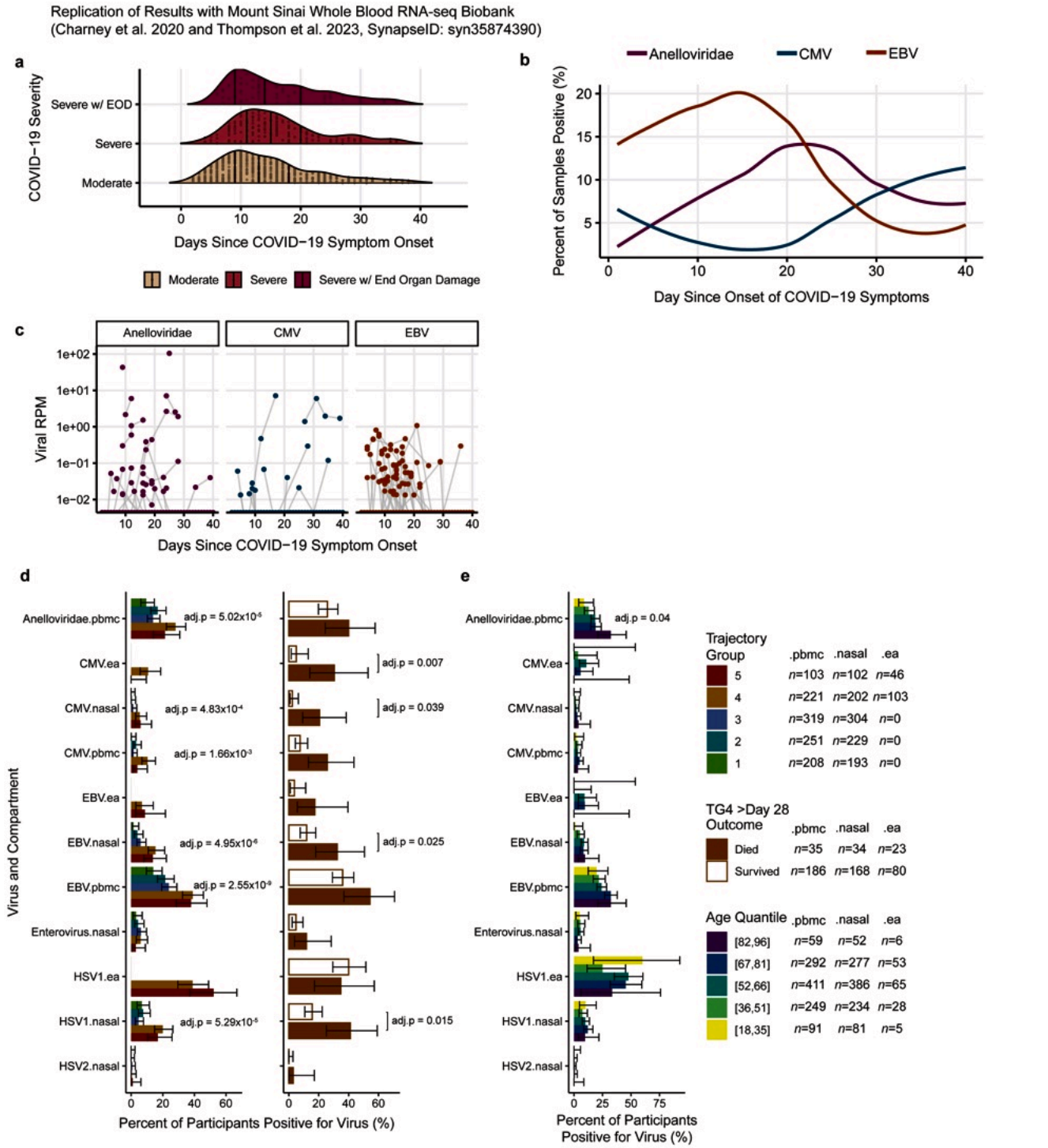


Extended Data Fig. 3 | See next page for caption.

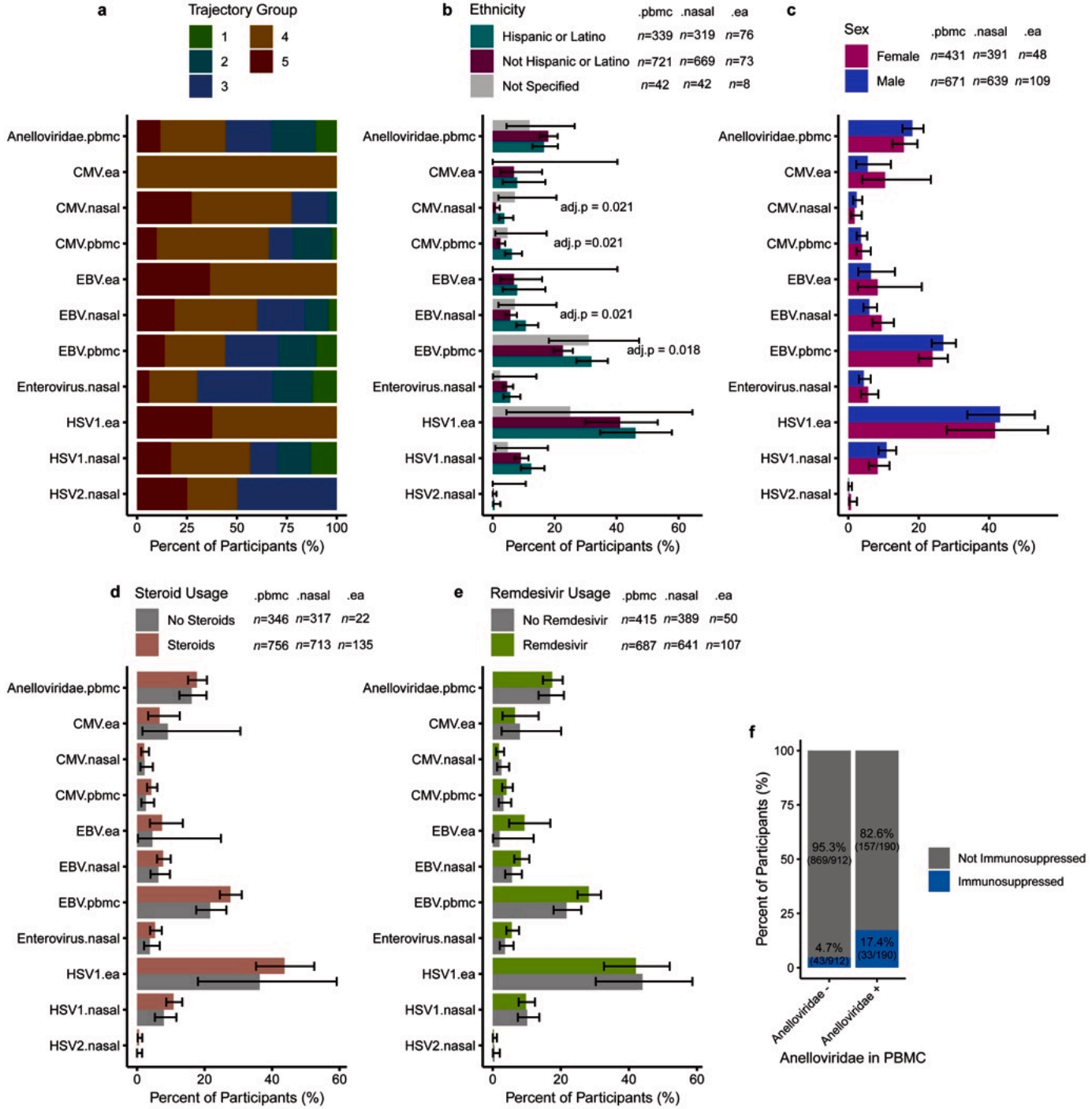
Extended Data Fig. 3 | Dynamics of Viral Reactivation Over Time. a) Reads per million (RPM) of detected viral transcripts overtime with individual participants' samples connected by a gray line. For **b-i**, graphs depict smoothed curves demonstrating the proportion of total samples that were positive for viruses. Curves were calculated by the proportion of samples positive for each day ± two days (a rolling window approach), followed by a local polynomial regression fitting. Below curves in **c-e** and **g-i**, graphs showing the trajectory group of samples contributing to the calculated rate on that day are shown. **b**) The smoothed rate of detection for all viruses in all transcriptomics by *days from hospitalization*, with **c**) subsetted to only the EA samples, **d**) subsetted to only the nasal samples, and **e**) subsetted to only the PBMC samples. **f**) The smoothed rate of detection for all viruses in all transcriptomics by *days from symptom onset*, with **g**) subsetted to only the EA samples, **h)** subsetted to only the nasal samples, and **i**) subsetted to only the PBMC samples.



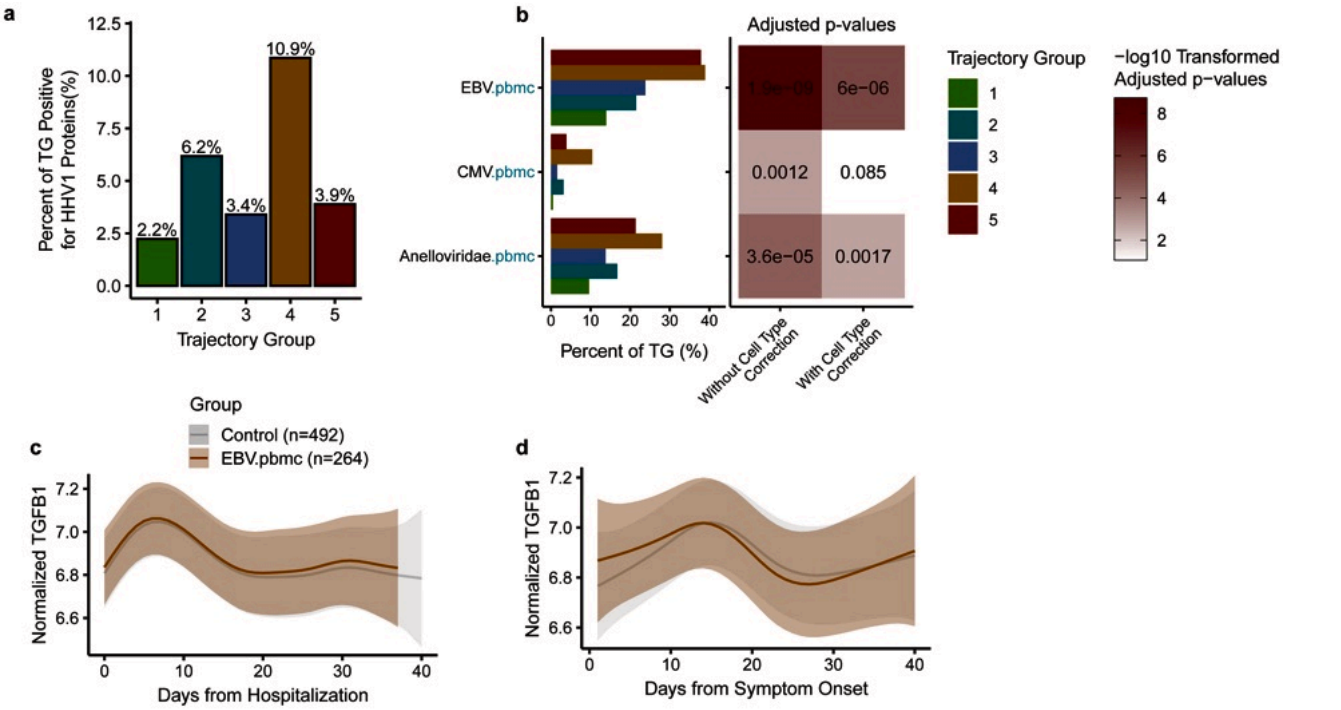
Extended Data Fig. 4 | Conditional Probability Table of Detected Viral Transcripts Within a Visit and Within All of a Participant's Samples. **a**) Heatmap depicting the percent of samples that were simultaneously positive for the virus denoted in the column when the virus in the row was present. **b**) Heatmap depicting the percent of participants that had the virus denoted in the column detected in any of their samples if the virus in the row was also detected in any of their samples. Percent of IMPACC participants with 0, 1, 2, 3, 4, 5, or 6 viruses detected besides SARS-CoV-2 (including *Anelloviridae*, CMV, EBV, HSV1, HSV2, and enteroviruses) in the acute period in **c**) any transcriptomic sample, and **d**) only nasal and PBMC samples (excluding EA samples).



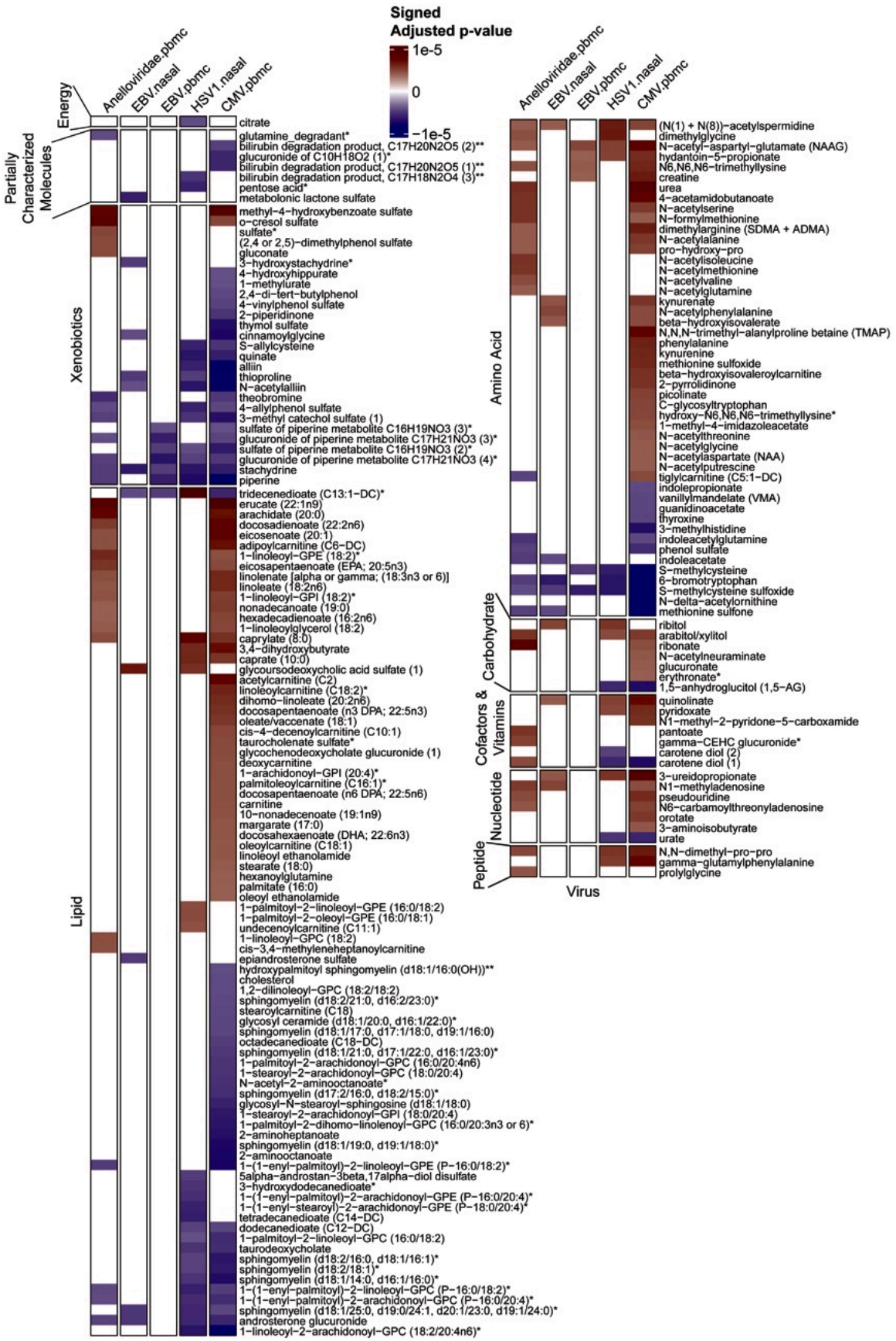
Extended Data Fig. 5 | Validation of Chronic Virus Transcript Detection in an External Hospitalized COVID-19 Cohort (Mount Sinai Biobank). **a**) Distribution of available PBMC RNA-sequencing samples (n = 165) as days since COVID-19 symptom onset from Charney et al. 2020[25] and Thompson et al. 2023[26] by COVID-19 severity. Lines inside the distribution denote the median and interquartile range. **b**) A graph that depicts smoothed curves demonstrating the proportion of total samples that were positive for viruses as a function of days from symptom onset. Curves were calculated by the proportion of samples positive for each day ± two days (a rolling window approach), followed by a local polynomial regression fitting. **c**) Reads per million (RPM) of detected viral transcripts overtime with individual participants' samples connected by a gray line. **d**) Percent of participants in the cohort who had detectable viral reads in at least one sample within 40 days of hospital admission (IMPACC Visits 1-6). Participants were split by each trajectory group (on the left), a measure of COVID-19 severity, and participants in TG4 were further subsetted by their long-term mortality outcome (on the right). Error bars denotes 95% confidence interval. **e**) Percent of participants in the cohort who had detectable viral reads in at least one sample within 40 days of hospital admission (IMPACC Visits 1-6) split by age quantiles. Error bars denotes 95% confidence interval. In (**d**-**e**) Benjamini-Hochberg adjusted p-values of ≤0.05, ≤0.01, ≤0.001, and ≤0.0001 are represented by *, **, ***, and **** respectively. Sample sizes for each group in (**d**-**e**) are reported in the figure legend.



Extended Data Fig. 6 | Relationship of Chronic Viral Reactivation with Additional Clinical Demographics and Medication Usage. a) Percent of patients positive for each virus and compartment within the first 40 days post-hospitalization (only using IMPACC visits 1-6 samples) that are in each trajectory group (TG). Percent of participants positive for each virus in the respective compartments for any sample within the first 40 days post-hospitalization (only using IMPACC visits 1-6 samples) split by **b**) ethnicity, **c**) sex, **d**) administration of steroids during acute COVID-19, **e**) administration of Remdesivir during acute COVID-19. Error bars in (**b-e**) denote 95% confidence interval. Significance calculated using chi-square test of independence with p-value corrections following Benjamini-Hochberg procedure, * indicates adj.p ≤ 0.05. Sample sizes for each group in (**b-e**) are reported in the figure legend.

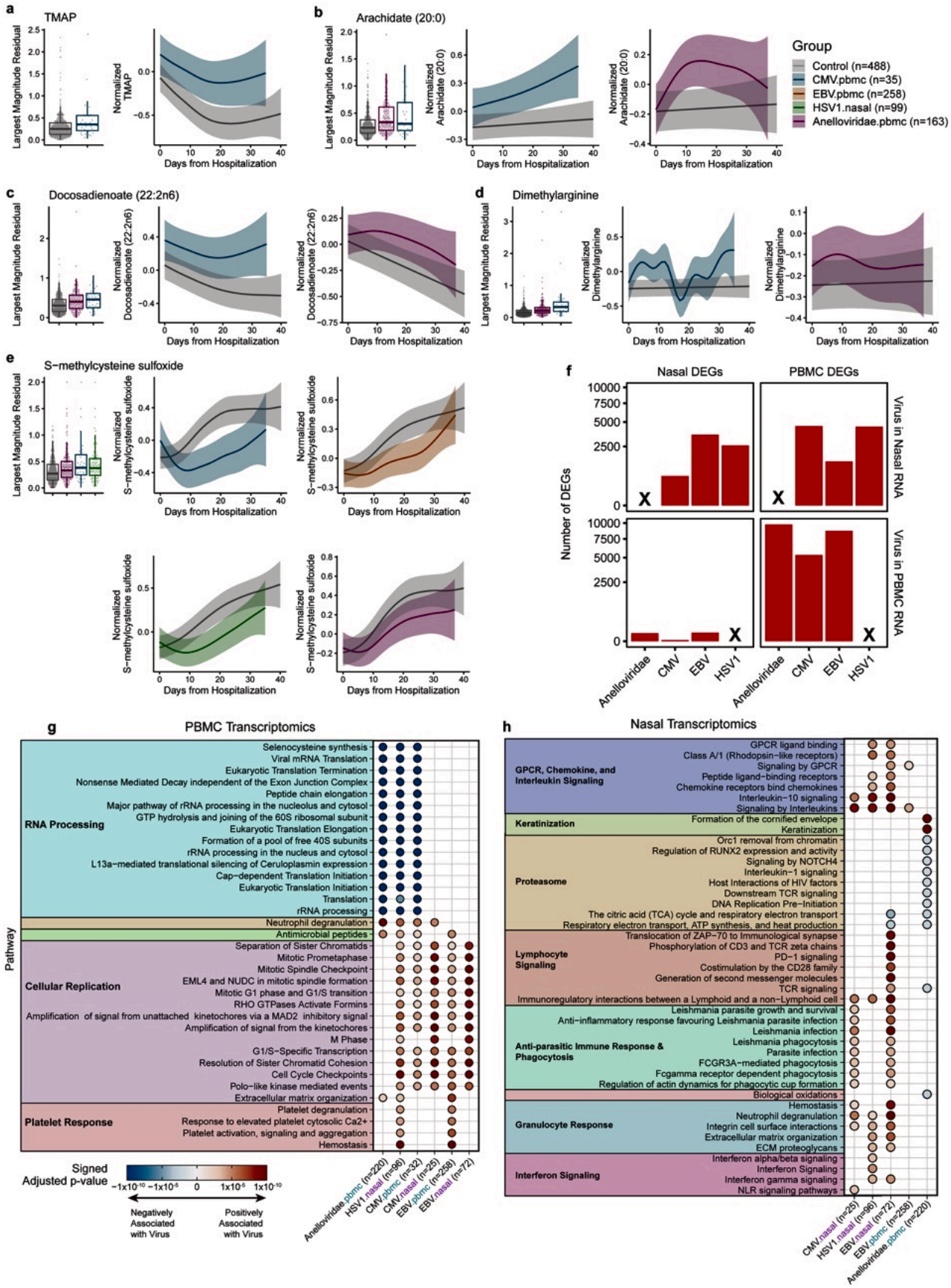


Extended Data Fig. 7 | EBV Associates With COVID-19 Severity Even While Controlling for Changes in Circulating Blood Cell Composition. a) Percent of each trajectory group (TG) with detectable HHV1 proteins in plasma via mass spectrometry at any sample within 40 days of hospital admission. **b**) Percent of participants in the cohort who had detectable viral reads in at least one sample within 40 days of hospital admission for the respective virus in the PBMC split by each TG. On the right, adjusted p-value for cumulative link mixed modeling testing for association of viral prevalence with TG in models with or without correcting for circulating levels of significantly associated cell types (as determined from Fig. 3d) from whole blood CyTOF. The cumulative link mixed model with cell type correction added all significantly associated cell types associated with the virus from the analysis in Fig. 3d as main effects in addition to virus status with enrollment site as a random effect. GAMM model of TGF-β1 as a function of **c**) days from hospitalization and **d**) days from symptom onset for comparison to the findings of Goetzke et al. 2025[30] with EBV reactivation in COVID-19-associated multisystem inflammatory syndrome in children. Shaded interval in (**c**-**d**) denotes 95% confidence interval calculated from GAMM model. The control group in (**c**) and (**d**) was comprised of 492 participants without any detected chronic virus transcripts in the acute period (≤40 days post-hospitalization).



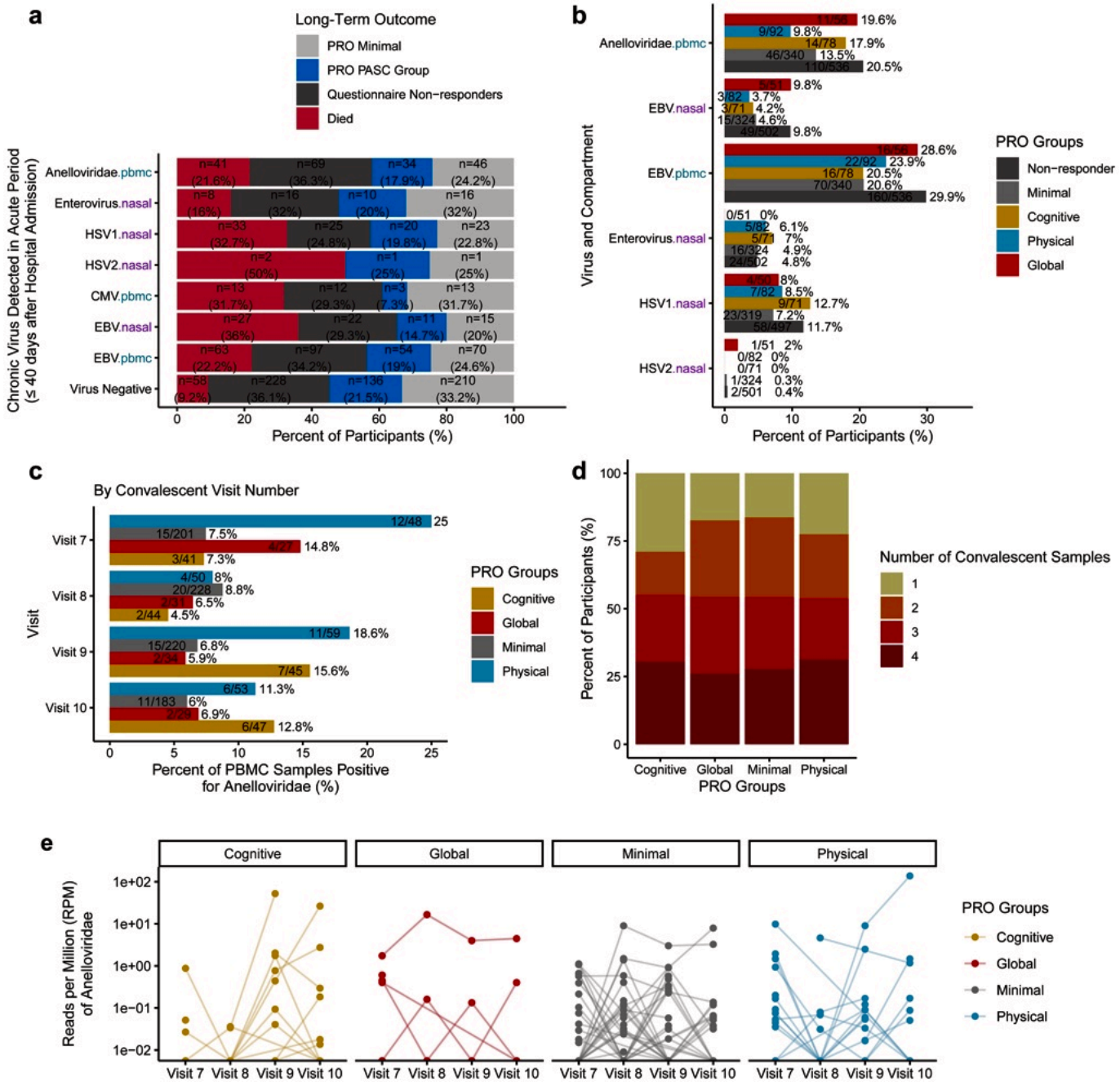
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Chronic Viral Reactivation Correlates with Widespread Changes in the Metabolome. Summary heatmap of Benjamini-Hochberg adjusted p-values from generalized additive mixed models (GAMM) evaluating the effects of chronic viral reactivation on plasma metabolite dynamics over time. The control group was comprised of 488 participants without any detected chronic virus transcripts in the acute period (≤40 days post-hospitalization). The adjusted p-values were signed and colored by direction of the associations of the cytokine/chemokine with the virus. The heatmap cell was only colored if adj.p ≤ 0.01, for either the main effect or time interaction term for viral reactivation in the model. Increasingly red color represents a more significant positive association, and increasingly blue color represents a more significant negative association.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Chronic Viral Reactivation is Associated with Alterations in Host Metabolome and Gene Expression. Boxplots depict the largest-magnitude GAMM residuals for each participant by virus, next to longitudinal GAMM-predicted means and 95% confidence intervals over days from hospitalization for the individual metabolites **a**) TMAP, **b**) Arachidate, **c**) Docosadianoate, **d**) Dimethylarginine, and **e**) S-methylcysteine sulfoxide. Boxplots denote median (center line), interquartile range (box), and 1.5x the interquartile range (whiskers). Results for (**a**-**e**) also shown in Extended Data Fig. 8. **f**) Number of differentially expressed genes (DEGs) for chronic viruses detected (adjusted p-values ≤ 0.05, determined with the limma package, see Methods). The graph is split by the number of DEGs detected for the nasal host gene expression and PBMC host gene expression (columns) as well as by which compartment the virus was detected (rows). X marks denote virus not detected frequently enough in the tissue for analysis (i.e. HSV1 in the PBMCs and *Anelloviridae* in the nasal). This plot demonstrates that viral reactivation in the nasal compartment affects both the nasal and PBMC host gene expression, whereas reactivation in the PBMC only substantially affects gene expression in the PBMCs. **g**) The top enriched pathways (determined via lowest adjusted p-values) from hypergeometric enrichment of differentially expressed genes (adj.p ≤ 0.05) in the PBMC transcriptomics associated with reactivation of chronically infecting viruses. **h**) The top enriched pathways (determined via lowest adjusted p-values) from hypergeometric enrichment of differentially expressed genes (adj.p ≤ 0.05) in the nasal transcriptomics associated with reactivation of chronically infecting viruses. Results for (**g**) and (**h**) calculated using hypergeometric enrichment of pathways from Reactome, separately on the positive and negative differentially expressed genes. Dot only shown for a pathway if adjusted p-value ≤ 0.01.



Extended Data Fig. 10 | Chronic Viral Reactivation Correlates With Lower Responses on Convalescent Surveys and *Anelloviridae* are More Frequent in Convalescent Samples from the Physical PRO Group. a) Percent of participants who either died, did not respond to convalescent symptom surveys (non-responders), or did respond and were grouped with an LC-affiliated patient-reported outcomes (PRO) group or the PRO minimal group displaying minimal deficits by virus positivity status in the acute COVID-19 period (first 40 days after hospitalization, IMPACC Visits 1-6). **b**) Percent of each PRO group and non-responders that had viruses detected in the acute COVID-19 period (first 40 days after hospitalization, IMPACC Visits 1-6). **c**) Percent of PRO group that had *Anelloviridae* detected in the outpatient convalescent samples, split by visit. **d**) Percent of each PRO group that gave 1, 2, 3, or 4 convalescent samples demonstrates no elevated rate in the physical group. **e**) Viral reads per million of *Anelloviridae* across the four convalescent visits (visits 7, 8, 9, and 10) for each of the PRO groups. Participants' visits are connected by a line across the timepoints.

nature portfolio

Corresponding author(s): Esther Melamed

Last updated by author(s): May 5, 2026

Reporting Summary

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

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No open source or custom code was used to collect data in this study.
Data analysis	All code for this study is publicly available in a Bitbucket repository https://bitbucket.org/kleinstein/impacc-public-code/src/master/chronic_viruses_manuscript/ and deposited in Zenodo (https://doi.org/10.5281/zenodo.19657241). Briefly, software and packages used included Samtools (v1.2), Samtools bam2fq (v1.4), Trimmomatic (v0.36.5), STAR (v2.4.2a), HTSeq-count (v0.4.1), Picard (v1.134), Picard tools (v2.22), FASTQC (v0.11.3 and v0.11.5), Cutadapt (v1.14), CZID, R 4.0.3, ordinal (v2019.12-10), rstatix (v0.7.2), coxme (v2.2-16), lme4 (v1.1-28), nlme (v3.1-149), gamm4 (v0.2-6), edgeR (v4), limma (v3.46.0), clusterProfiler (v3.18.1), Clustergrammer2 (https://github.com/ismms-himc/clustergrammer2).

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The IMPACC Data Sharing Plan aims to support broad data dissemination while safeguarding participant privacy and data integrity through de-identification and masking of sensitive data fields. All IMPACC data, including those produced as part of this study, have been submitted to the Immunology Database and Analysis Portal (ImmPort), a repository supported by the NIAID Division of Allergy, Immunology and Transplantation under accession code SDY1760 (https://www.immport.org/shared/study/SDY1760), as well as to the NLM's Database of Genotypes and Phenotypes (dbGaP) under accession number phs002686.v2.p2. Both raw and processed datasets are available through restricted access in accordance with the NIH public data sharing policy for IRB-exempted public health surveillance studies. Access may be requested through AccessClinicalData@NIAID (https://accessclinicaldata.niaid.nih.gov/study-viewer/clinical_trials). Further detailed guidance on data access is provided on ImmPort (https://docs.immport.org/home/impaccslides).

For alignment of host transcripts from the PBMC, Nasal, and EA RNA-seq, the GRCh38 reference genome and Ensembl release 91 were used. The Reactome database (v95) was used for hypergeometric enrichment of differentially expressed genes. The CZID pipeline utilized the NCBI nucleotide (nt) and NCBI non-redundant (nr) databases. HSV1 proteins for the plasma mass spectrometry were retrieved from UniProt for Swiss-Prot reviewed proteins attributed to Human Herpesvirus 1 (strain 17, UP000009294). Publicly available data for Extended Data Figure 5a-c were retrieved from the Mount Sinai COVID-19 Biobank which is available in the data repository Synapse (SynID: syn35874390, https://www.synapse.org/Synapse:syn35874390/wiki/618989).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Patient self-reported biological sex was evaluated for an association with chronic viral transcripts in Supplemental Figure 2 and found no significant association. For the generalized additive mixed modeling analysis used for the serum cytokines and plasma metabolomics as well as for the nasal and PBMC RNA-seq analyses, sex was included as a covariate in the models.
Reporting on race, ethnicity, or other socially relevant groupings	Patient self-reported ethnicity was evaluated for an association with chronic viral transcripts in Supplemental Figure 2.
Population characteristics	Full cohort characteristics are presented in Supplementary Table 1. Briefly, the IMPACC study cohort comprised 1,154 adults hospitalized with COVID-19, with a median age of 59 years. 61% were male and 39% were female. In terms of ethnicity, 48% were Caucasian, 22% Black, and 31% Hispanic. Most common co-morbid conditions included hypertension (58%), diabetes (37%), chronic cardiac disease (27%), chronic lung disease (20%), and chronic kidney disease (15%). At enrollment, participants had varied requirements for respiratory support (23% on no supplemental oxygen, 47% on supplemental oxygen, 18% on non-invasive ventilation, and 12% on invasive mechanical ventilation.) Most common medications during hospitalization included anticoagulation (88%), steroids (68%), remdesivir (62%), and antibiotics (57%). (please refer to Supplementary Table 1 for further details).
Recruitment	Hospitalized COVID-19 patients were recruited as reported in the clinical trial outline (clinicaltrials.gov (NCT0438777)) from 20 hospitals across 15 academic institutes. Patients enrolled at sites operating as a public health surveillance study were provided information sheets describing the study including the samples to be collected and plans for analysis and data de-identification. Patients at sites with IRB-approved protocols obtained participant informed consent. Participants who requested not to participate after review of the study plan and information were not enrolled.
Ethics oversight	The Department of Health and Human Services Office for Human Research Protections (OHRP) and NIAID concurred that the IMPACC study qualified for public health surveillance exemption. The study protocol was sent for review to each site's institutional review board (IRB), with twelve sites conducting as a public health surveillance study, and three sites integrating the IMPACC study into IRB-approved 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 participants providing informed consent.

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

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	Following the scientific objectives and goal of the IMPACC network to generate further hypotheses, the analysis of primary endpoints was planned to be descriptive with additional goals to generate hypotheses for future studies. Therefore, the sample size considerations that were most relevant to the study pertained to precision of estimates and minimum detectable effect sizes. Assuming N=1,000 participants, Type I error rate of 5% i.e. alpha level 0.05, and Type II error rate of 20% (=80% power) i.e. beta level 0.20, estimates of binary endpoints we calculated that the we would have a 95% CI no wider than ± 3.1%.
Data exclusions	All assays in IMPACC were evaluated by a Clinical and Data Coordinating Center (CDCC) which performed quality control on all data. All data that passed quality control were used without any data exclusion.
Replication	Due to the large sample size and variety of assays, no replication experiments were performed.
Randomization	The IMPACC study was a prospective, longitudinal observational cohort study designed to characterize immune response trajectories in adults hospitalized with acute COVID-19 and relate them to immune and clinical outcomes. Randomization was not applicable to this design because IMPACC did not assign an investigational intervention. All participants in the study received standard clinical care.
Blinding	Blinding was not relevant to the IMPACC study as it was an observational study without a drug intervention arm.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

CD8a-146Nd Fluidigm Cat#3146001B; RRID:AB_2687641
Granzyme B Antibody, anti-human/mouse/rat, REAfinity Miltenyi Cat#130-116-486
Goat Anti-Human IgA-UNLB Southern Biotech Cat#2050-01
Purified anti-human IgM Antibody Biolegend Cat#314502
Mouse Anti-Human IgG1 Fc-UNLB Souther Biotech Cat#9054-01
Purified anti-mouse/human CD11b Antibody Biolegend Cat#101202
Purified anti-human/mouse/rat CD278 (ICOS) Antibody Biolegend Cat#313502
Purified anti-human CD39 Antibody Biolegend Cat#328202
Purified anti-human CD169 (Sialoadhesin, Siglec-1) Antibody Biolegend Cat#346002
Purified anti-human CD64 (Maxpar® Ready) Antibody Biolegend Cat#305029
Purified anti-human CD71 Antibody Biolegend Cat#334102
Anti-Human CD279/PD-1 (EH12.2H7)-175Lu Fluidigm Cat#3175008B
Anti-Human CD61 (VI-PL2)-209Bi Fluidigm Cat#3209001B
Anti-Human CD3 (UCHT1)-141Pr antibody Fluidigm Cat#3141019B
Anti-Human HLA-DR (L243)-143Nd antibody Fluidigm Cat#3143013B
Anti-Human CD69 (FN50)-144Nd antibody Fluidigm Cat#3144018B
Anti-Human CD4 (RPA-T4)-145Nd antibody Fluidigm Cat#3145001B
Anti-Human CD8a (RPA-T8)-146Nd antibody Fluidigm Cat#3146001B
Anti-Human CD20 (2H7)-147Sm antibody Fluidigm Cat#3147001B
Anti-Human CD127 (A019D5)-149Sm antibody Fluidigm Cat#3149011B
Anti-Human MIP-1β (D21-1351)-150Nd antibody Fluidigm Cat#3150004B
Anti-Human CD123 (6H6)-151Eu antibody Fluidigm Cat#3151001B
Anti-Human TNFα (Mab11)-152Sm antibody Fluidigm Cat#3152002B
Anti-Human CD62L (DREG-56)-153Eu antibody Fluidigm Cat#3153004B
Anti-Human CD45 (HI30)-154Sm antibody Fluidigm Cat#3154001B
Anti-Human IL-6 (MQ2-13A5)-156Gd antibody Fluidigm Cat#3156011B
Anti-Human IFN-γ (B27)-158Gd antibody Fluidigm Cat#3158017B
Anti-Human CD11c (Bu15)-159Tb antibody Fluidigm Cat#3159001B

Anti-Human CD14 (M5E2)-160Gd antibody Fluidigm Cat#3160001B
Anti-Human CD80/B7.1 (2D10.4)-161Dy antibody Fluidigm Cat#3161023B
Anti-Human CD66b (80H3)-162Dy antibody Fluidigm Cat#3162023B
Anti-Human CD56 (NCAM16.2)-163Dy antibody Fluidigm Cat#3163007B
Anti-Human CD15 (W6D3)-164Dy antibody Fluidigm Cat#3164001B
Anti-Human CD61 (VI-PL2)-165Ho antibody Fluidigm Cat#3165010B
Anti-Human CD11b (ICRF44)-167Er antibody Fluidigm Cat#3167011B
Anti-Human CD206 (15-2)-168Er antibody Fluidigm Cat#3168008B
Anti-Human CD54 (HA58)-170Er antibody Fluidigm Cat#3170014B
Anti-Human CD68 (Y1/82A)-171Yb antibody Fluidigm Cat#3171011B
Anti-Human CD16 (3G8)-209Bi antibody Fluidigm Cat#3209002B
Anti- CoV Nucleocapsid protein (6H3) antibody Abcam Cat#ab273434
Anti-Human Eotaxin (43915) antibody R&D Cat#MAB3201
Anti-Human ACE-2 (535919) antibody NOVUS Cat#MAB9332-100
Anti-Human Cytokeratin (C-11) antibody Biolegend Cat#628602
Anti- CoV Spike protein (1A9) antibody GeneTex Cat#GTX632604
Anti-Human EPX (MM82.2.1) antibody MAYO CLINIC https://www.mayoclinic.org
Anti-Human IL-8 (E8N1) antibody Biolegend Cat#511402
Anti-Human IL-1β (H1b-27) antibody Biolegend Cat#511602
Anti-Human IFN-β (IFNb/A1) antibody Biolegend Cat#514002
Anti-Human Siglec-8 (837535) antibody R&D Cat#MAB7975
Anti-human IgG (Fc specific)-Peroxidase antibody produced in goat Sigma-Aldrich Cat#A0170; RRID: AB_257868
Goat anti-human IgM-HRP SouthernBiotech Cat#2020-05; RRID: AB_2795603
Anti-human IgA (α-chain specific)-Peroxidase antibody produced in goat Sigma-Aldrich Cat#A0295; RRID: AB_257876
Anti-Glial Fibrillary Associated Protein Agilent Cat#Z033429-2
Anti-human IgG (PE) ThermoScientific Cat#12-4998-8
Anti-human pSTAT1 (AF647) BD Cat#612597
Anti-human CD14 (FITC) BD Cat#555397

Validation

Antibodies were used and validated as described by manufacturer and in prior IMPACC publications: Diray-Arce J, et al. Multi-omic longitudinal study reveals immune correlates of clinical course among hospitalized COVID-19 patients. Cell Reports Medicine (2023), and in the study protocol: Immunophenotyping assessment in a COVID-19 cohort (IMPACC): A prospective longitudinal study. Science Immunology (2021).

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration

IMPACC was registered at clinicaltrials.gov with the accession number NCT0438777.

Study protocol

The full trial protocol can be accessed in the prior publication "Immunophenotyping assessment in a COVID-19 cohort (IMPACC): A prospective longitudinal study" at Science Immunology (https://www.science.org/doi/10.1126/sciimmunol.abf3733).

Data collection

Clinical data were captured directly from the electronic health record and stored in the study RedCap. All clinical data were audited after study completion to ensure accuracy.

Outcomes

The two primary objectives of IMPACC were to:
1. Describe the relationship between specific immunologic assessments and severity of illness in hospitalized patients with COVID-19, controlling for time of illness onset, and concurrent participation in clinical trials or off-label use of investigational (or approved) therapeutic agents for COVID-19.'
2. Describe the relationship between burden of disease, assessed by duration of virus shedding in nasal secretions, and severity of illness in hospitalized patients with COVID-19. was Standard individual severity of illness measures were defined s outcomes, and the 7 point ordinal score was exploratory.

The data were collected in case report forms via an electronic database using bedside assessments and data extraction from the electronic medical record. The descriptive clinical outcomes for severity of illness were reported in the first clinical manuscript (Table 1) in eBiomedicine, along with a description of the phenotyping approach used to define the disease trajectories (https://pubmed.ncbi.nlm.nih.gov/35952496/).

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A