

# Identification of soluble biomarkers that associate with distinct manifestations of long COVID

Received: 23 May 2024

Accepted: 14 March 2025

Published online: 30 April 2025

 Check for updates

Yu Gao<sup>1,13</sup>, Curtis Cai<sup>1,13</sup>, Sarah Adamo<sup>1,2,13</sup>, Elsa Biteus<sup>3,4</sup>, Habiba Kamal<sup>3</sup>, Lena Dager<sup>5</sup>, Kelly L. Miners<sup>6</sup>, Sian Llewellyn-Lacey<sup>6</sup>, Kristin Ladell<sup>6</sup>, Pragati S. Amratia<sup>6</sup>, Kirsten Bentley<sup>6</sup>, Simon Kollnberger<sup>6</sup>, Jinghua Wu<sup>1</sup>, Mily Akhirunnesa<sup>1</sup>, Samantha A. Jones<sup>7</sup>, Per Julin<sup>8,9</sup>, Christer Lidman<sup>3</sup>, Richard J. Stanton<sup>6</sup>, Paul A. Goepfert<sup>10</sup>, Michael J. Peluso<sup>11</sup>, Steven G. Deeks<sup>11</sup>, Helen E. Davies<sup>7,13</sup>, Soo Aleman<sup>3,5,13</sup>, Marcus Buggert<sup>1,13</sup> & David A. Price<sup>6,12,13</sup> ✉

Long coronavirus disease (COVID) is a heterogeneous clinical condition of uncertain etiology triggered by infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Here we used ultrasensitive approaches to profile the immune system and the plasma proteome in healthy convalescent individuals and individuals with long COVID, spanning geographically independent cohorts from Sweden and the United Kingdom. Symptomatic disease was not consistently associated with quantitative differences in immune cell lineage composition or antiviral T cell immunity. Healthy convalescent individuals nonetheless exhibited higher titers of neutralizing antibodies against SARS-CoV-2 than individuals with long COVID, and extensive phenotypic analyses revealed a subtle increase in the expression of some co-inhibitory receptors, most notably PD-1 and TIM-3, among SARS-CoV-2 nonspike-specific CD8<sup>+</sup> T cells in individuals with long COVID. We further identified a shared plasma biomarker signature of disease linking breathlessness with apoptotic inflammatory networks centered on various proteins, including CCL3, CD40, IKBKG, IL-18 and IRAK1, and dysregulated pathways associated with cell cycle progression, lung injury and platelet activation, which could potentially inform the diagnosis and treatment of long COVID.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has left a pernicious legacy of global ill health, commonly known as long coronavirus disease (COVID)<sup>1</sup>. The etiology of this heterogeneous condition remains obscure, but common symptoms include breathlessness, cognitive impairment, often described as ‘brain fog’, fatigue and pain, alongside a host of other clinical manifestations indicating the involvement of different organ systems in the body<sup>2,3</sup>. Several hypotheses have been proposed to account for such diverse and persistent

symptomatology, including immune dysregulation, ongoing inflammation and tissue damage, and viral persistence<sup>3–9</sup>. The reactivation of latent herpesviruses may also contribute to the pathogenesis of long COVID<sup>3,10,11</sup>.

A handful of ‘omics approaches have been used to probe the molecular intricacies of long COVID. For example, affinity proteomics studies have identified distinct inflammatory phenotypes and enrichment of the NF-κB and type 2 interferon (IFN) signaling pathways as correlates

A full list of affiliations appears at the end of the paper. ✉ e-mail: [marcus.buggert@ki.se](mailto:marcus.buggert@ki.se); [priced6@cardiff.ac.uk](mailto:priced6@cardiff.ac.uk)

of disease, highlighting associations with various soluble biomarkers, such as IFN $\gamma$ , interleukin-1 (IL-1), IL-6 and tumor necrosis factor (TNF), which are typically upregulated in acute COVID-19 (refs. 12–14). Similar findings have been described using conventional approaches to cytokine quantification<sup>15</sup>. A longitudinal multiomics study further reported that various autoantibodies, altered cytomegalovirus (CMV)-specific and SARS-CoV-2-specific CD8<sup>+</sup> T cell dynamics, and Epstein–Barr virus (EBV) and SARS-CoV-2 viremia were associated with the emergence of particular subtypes of long COVID<sup>10</sup>. More recently, another multiomics study found that elevated herpesvirus-specific antibody titers, immune cell perturbations, and decreased cortisol levels were distinguishing features of persistent illness after infection with SARS-CoV-2 (ref. 16). By contrast, hypothesis-driven approaches and low-resolution proteomics have identified dysregulation of the complement system, which is known to drive inflammation, as a consistent feature of long COVID<sup>17,18</sup>. These observations suggest that multiple factors could be associated with the development of discrete symptom complexes and patterns of disease onset within the clinically diverse spectrum of long COVID.

In this study, we used a variety of multidimensional approaches and integrative data analysis pipelines to profile the immune system and the plasma proteome in healthy convalescent individuals with a molecularly confirmed history of infection with SARS-CoV-2 and individuals with long COVID, spanning geographically independent cohorts from Sweden and the United Kingdom. We found higher titers of SARS-CoV-2 spike-specific neutralizing antibodies in healthy convalescent individuals than in individuals with long COVID. By contrast, minimal intergroup differences were apparent in immune cell lineage composition and virus-specific CD4<sup>+</sup> and CD8<sup>+</sup> T cell immunity, although some co-inhibitory receptors, especially PD-1 and TIM-3, were relatively overexpressed among SARS-CoV-2 nonspike-specific CD8<sup>+</sup> T cells in individuals with long COVID. We also detected a unique array of soluble biomarkers in the plasma proteome that correlated directly with the clinical manifestation of breathlessness in individuals with long COVID. Network and pathway analyses linked these biomarker signatures with apoptotic processes and inflammation, highlighting key roles for signaling cascades involving ceramide, FAS, NF- $\kappa$ B and TNF. Moreover, core network components, including CCL3, CD40 and IL-18, were identified as potential contributors to persistent inflammation in individuals with long COVID. These results provide a mechanistic framework to unravel the complex etiology and pathogenesis of ongoing symptomatic disease triggered by infection with SARS-CoV-2.

## Results

### Clinical characterization

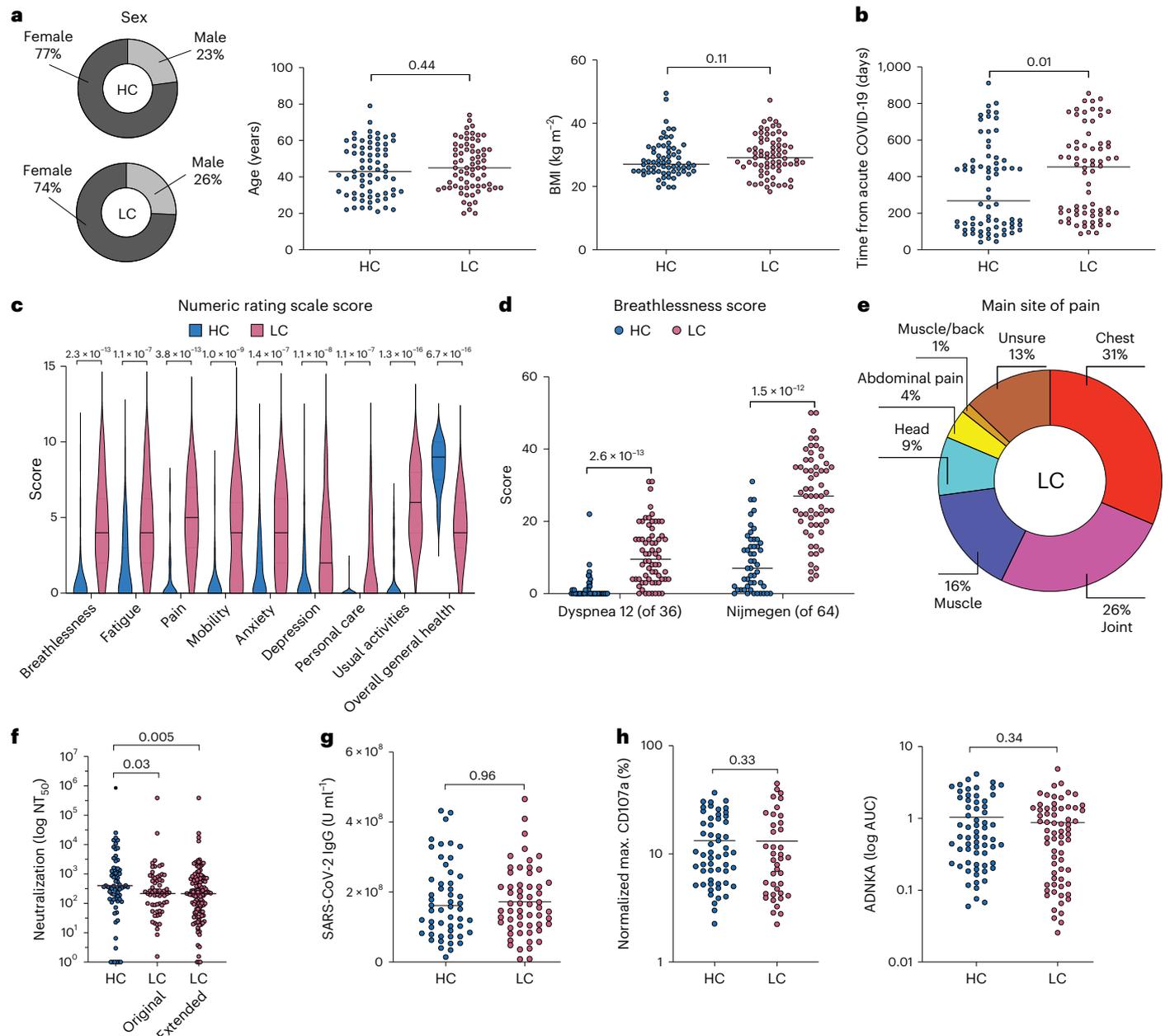
The primary cohort included healthy convalescent individuals (controls;  $n = 70$ ) and individuals with long COVID (cases;  $n = 70$ ) recruited from University Hospital Llandough (Table 1 and Supplementary Table 1). All participants had a clearly defined episode of symptomatically mild acute COVID-19 confirmed via direct molecular evidence of infection with SARS-CoV-2. Intergroup comparisons revealed largely equivalent distributions for age (cases, median = 45 years; controls, median = 43 years), body mass index (BMI; cases, median = 29.8 kg m<sup>-2</sup>; controls, median = 28.5 kg m<sup>-2</sup>), race (cases, White = 88.6%; controls, White = 82.9%), sex (cases, female = 74.3%; controls, female = 77.2%), time since initial reported infection (cases, median = 416 days; controls, median = 268 days), and vaccination against SARS-CoV-2 (cases, median number of vaccinations = 3; controls, median number of vaccinations = 3; Fig. 1a,b and Table 1). All baseline medical evaluations were normal in individuals with long COVID. Symptom scores are depicted in Fig. 1c. Breathlessness was further assessed using the Dyspnea-12 questionnaire, scored out of 36, and the Nijmegen questionnaire, scored out of 64, which provide a metric for hyperventilation (Fig. 1d). Pain was most commonly localized to the chest (31%), joints (26%) and muscles (16%) in individuals with long COVID (Fig. 1e). The secondary

**Table 1 | Key features of the primary cohort recruited from the United Kingdom**

Characteristic	Controls ( $n=70$ )	Cases ( $n=70$ )
Age (years), median (range)	43 (21–79)	45 (20–74)
Female (%)	54 (77.1)	52 (74.3)
Race (%)		
White	58 (82.9)	62 (88.6)
Black	0 (0.0)	1 (1.4)
Asian	9 (12.9)	4 (5.7)
Mixed	2 (2.9)	2 (2.9)
Other	1 (1.4)	1 (1.4)
BMI > 30 kg m <sup>-2</sup> (%)	21/66 (31.8)	32/69 (46.3)
Coexisting conditions (%)		
Yes	13 (18.6)	35 (50.0)
No	57 (81.4)	35 (50.0)
Employment status (%)		
Pre-COVID-19		
Employed	45/49 (91.8)	59/66 (89.4)
Employed, altered duties	0/49 (0.0)	0/66 (0.0)
Employed, sick leave	0/49 (0.0)	0/66 (0.0)
Unemployed	0/49 (0.0)	1/66 (1.5)
Retired	1/49 (2.0)	4/66 (6.1)
Student	3/49 (6.1)	2/66 (3.0)
Post-COVID-19		
Employed	46/49 (93.9)	35/66 (53.0)
Employed, altered duties	0/49 (0.0)	11/66 (16.7)
Employed, sick leave	0/49 (0.0)	11/66 (16.7)
Unemployed	0/49 (0.0)	2/66 (3.0)
Retired	1/49 (2.0)	5/66 (7.6)
Student	2/49 (4.1)	2/66 (3.0)
Date of infection with SARS-CoV-2 (%)		
March 2020 to August 2020	11 (15.7)	14 (20.0)
September 2020 to June 2021	23 (32.9)	25 (35.7)
July 2021 to October 2022	36 (51.4)	31 (44.3)
COVID-19 vaccination status, median (IQR)		
Number of vaccinations before infection	1.5 (0–3)	0 (0–2)
Total number of vaccinations	3 (3–3)	3 (3–3)
Symptoms, median (IQR) <sup>a</sup>		
Breathlessness	0 (0–0)	3 (2–6)
Fatigue	0 (0–2)	6 (4–8)
Musculoskeletal	0 (0–0)	3.5 (0–6)
Neuropsychiatric	0 (0–0)	4 (1–6)
Pain	0 (0–0)	4 (2–5)
Ability to maintain self-care, median (IQR)	0 (0–0)	0 (0–3)
Ability to maintain daily tasks, median (IQR)	0 (0–0)	5 (3–8)
Overall general health, median (IQR) <sup>b</sup>	0 (–1 to 0)	–4 (–2 to –6)

IQR, interquartile range. <sup>a</sup>Difference in numeric rating scale score (0 = no symptom, 10 = worst possible symptom) before versus after COVID-19. <sup>b</sup>Difference in numeric rating scale score (0 = worst possible, 10 = best possible) before versus after COVID-19.

cohort included healthy convalescent individuals (controls;  $n = 30$ ) and individuals with long COVID (cases;  $n = 95$ ) recruited from Karolinska University Hospital (Table 2).



**Fig. 1 | Cohort overview and humoral immunity in healthy convalescent individuals and individuals with long COVID recruited from the United Kingdom.** **a**, Ring charts and scatter dot plots showing sex, age and BMI for healthy convalescent individuals (HC,  $n = 70$ ) and individuals with long COVID (LC,  $n = 70$ ). **b**, Scatter dot plots showing the corresponding time to sampling from the initial diagnosis of acute COVID-19. **c**, Violin plots showing the corresponding distribution of clinical symptom numeric rating scale scores. **d**, Scatter dot plots showing breathlessness scores as assessed using the Dyspnea-12 and Nijmegen questionnaires (HC,  $n = 49$  and  $n = 49$ , respectively; LC,  $n = 66$  and  $n = 62$ , respectively). **e**, Ring chart highlighting the anatomical distribution of pain experienced by individuals with long COVID. **f**, Scatter dot plot showing SARS-CoV-2-specific neutralization activity quantified as the

highest plasma dilution that achieved a 50% reduction in plaque formation (NT<sub>50</sub>; HC,  $n = 70$ ; LC, original  $n = 70$  and extended  $n = 146$ ). **g**, Scatter dot plot showing total SARS-CoV-2 spike-specific immunoglobulin titers (HC,  $n = 52$ ; LC,  $n = 57$ ). **h**, Scatter dot plots showing maximum CD107a mobilization (left) quantified as percentage values relative to the corresponding positive controls and normalized ADNKA (right) quantified as a function of degranulation (CD107a<sup>+</sup>) among viable NK cells (Aqua<sup>-</sup>CD3<sup>+</sup>CD56<sup>+</sup>) with potent cytotoxic activity (CD57<sup>+</sup>; HC,  $n = 55$  and  $n = 66$ , respectively; LC,  $n = 40$  and  $n = 66$ , respectively); AUC, area under the curve. Horizontal bars represent median values (**a–d** and **f–h**). Significance was evaluated using a two-tailed Mann–Whitney *U*-test (**a–d** and **f–h**).

### Neutralizing antibody titers are suboptimal in long COVID

To evaluate the humoral immune system, we measured total SARS-CoV-2 spike-specific immunoglobulin titers, virus neutralization activity, and antibody-dependent natural killer (NK) cell activation (ADNKA) in plasma samples obtained from donors in the United Kingdom. Healthy convalescent individuals exhibited substantially better neutralization

activity in standard plaque reduction assays than individuals with long COVID (Fig. 1f), despite equivalent overall titers of antibodies targeting the spike protein of SARS-CoV-2 (Fig. 1g). This finding was confirmed across a larger number of donors from the same cohort<sup>17</sup>, achieving even greater significance (Fig. 1f). By contrast, no such intergroup differences were apparent for ADNKA measured as a cumulative metric

**Table 2 | Key features of individuals with long COVID recruited from Sweden**

Characteristic	Borg CR10 $\leq$ 2 (n=60)	Borg CR10 $>$ 2 (n=35)
Age (years), median (range)	46 (21–64)	50 (23–66)
Female (%)	88.7%	75.6%
BMI $>$ 30 kg m $^{-2}$ (%)	15/41 (22.9)	19/45 (27.2)

against all expressed viral target proteins using healthy donor cell preparations with a surrogate marker of potential cytotoxicity (Fig. 1h), namely CD57 (ref. 19).

Collectively, these findings identify a qualitative deficit in the humoral immune response against SARS-CoV-2, specifically impacting neutralization activity in individuals with long COVID.

### Immune cell perturbations are limited in long COVID

To evaluate the cellular immune system, we first conducted a multi-dimensional flow cytometric analysis of the major lineages typically present among peripheral blood mononuclear cells (PBMCs), focusing initially on donors recruited from the United Kingdom (Fig. 2a). Using dimensionality reduction and Gaussian mixture models, we identified clusters that corresponded to the major lineages of monocytes, B cells, NK cells and T cells (Fig. 2b and Extended Data Fig. 1), but these analyses were unable to differentiate between some other immune cell subsets, such as basophils and plasmacytoid dendritic cells (pDCs; Fig. 2b), and were also unable to differentiate between healthy convalescent individuals and individuals with long COVID (Fig. 2c). We therefore interrogated the data conventionally using a manual flow cytometric gating strategy (Extended Data Fig. 2a).

In the adaptive lymphocyte compartment, similar proportions of naive B cells, total B cells, naive T cells, total T cells, naive CD4 $^{+}$  T cells, total CD4 $^{+}$  T cells, naive CD8 $^{+}$  T cells and total CD8 $^{+}$  T cells were identified in healthy convalescent individuals and individuals with long COVID (Fig. 2d), and in the innate lymphocyte compartment, similar proportions of immature NK cells (CD16 $^{-}$ CD56 $^{\text{bright}}$ ), mature NK cells (CD16 $^{+}$ CD56 $^{\text{dim}}$ ), total NK cells (including CD16 $^{-}$ CD56 $^{\text{dim}}$ ) and total innate lymphoid cells (CD127 $^{+}$ ) were identified in healthy convalescent individuals and individuals with long COVID (Fig. 2e). A comparable pattern was observed for classical monocytes (CD14 $^{+}$ ), intermediate monocytes (CD14 $^{+}$ CD16 $^{+}$ ) and conventional DCs (CD11c $^{+}$ CD123 $^{-}$ ) in the myeloid cell lineage, whereas the proportions of nonclassical monocytes (CD16 $^{+}$ ) were relatively increased in healthy convalescent individuals (Fig. 2f), and the proportions of basophils (CD123 $^{+}$ HLA-DR $^{-}$ ) and pDCs (CD123 $^{+}$ HLA-DR $^{+}$ ) were relatively increased in individuals with long COVID (Fig. 2g). Hierarchical clustering confirmed these differences within an otherwise rather uniform immune cell landscape (Fig. 2h,i). By contrast, no such perturbations were apparent in the secondary cohort of donors recruited from Sweden, although the proportions of classical monocytes were relatively decreased and the proportions of intermediate monocytes were relatively increased in individuals with long COVID (Extended Data Fig. 3a–d).

Collectively, these data indicate that immune cell perturbations are quantitatively subtle and, despite intercohort variability, generally confined to the myeloid compartment in individuals with long COVID.

### T cell immunity remains largely unaltered in long COVID

To extend these findings, we quantified CD4 $^{+}$  and CD8 $^{+}$  memory T cell responses against SARS-CoV-2 and the persistent herpesviruses CMV and EBV, exposure to which has been differentially linked with the development of long COVID<sup>10,16,20,21</sup>. We used activation-induced marker (AIM) assays for this purpose, enumerating functional antigen-specific CD4 $^{+}$  T cells by assessing the upregulation of CD69 and CD40L (CD154) and functional antigen-specific CD8 $^{+}$  T cells by assessing the upregulation

of CD69 and 4-1BB (CD137) after peptide stimulation directly ex vivo<sup>22,23</sup>. PBMCs were stimulated with individual peptide pools spanning the major immunogenic proteins from SARS-CoV-2 (spike, nucleocapsid, combined membrane and envelope, ORF1a, ORF1b and ORF3–ORF10) and selected immunogenic proteins from CMV (IE-1, IE-2 and pp65) and EBV, the latter segregated according to lytic (BRLF1, BZLF1, BMLF1 and BARF1) and latent phases (EBNA1, EBNA2, EBNA3A, EBNA3B, EBNA3C and LMP2) of the viral life cycle (Extended Data Fig. 2b). The frequencies of antiviral CD4 $^{+}$  and CD8 $^{+}$  T cells were statistically indistinguishable across all of these specificities in healthy convalescent individuals and individuals with long COVID recruited from the United Kingdom (Fig. 3a). By contrast, the frequencies of CD4 $^{+}$  T cells targeting the SARS-CoV-2 nucleocapsid protein and the EBV latent proteins and the frequencies of CD8 $^{+}$  T cells targeting the SARS-CoV-2 spike protein, the CMV proteins and the EBV lytic proteins were higher in individuals with long COVID than in healthy convalescent individuals recruited from Sweden (Extended Data Fig. 3e).

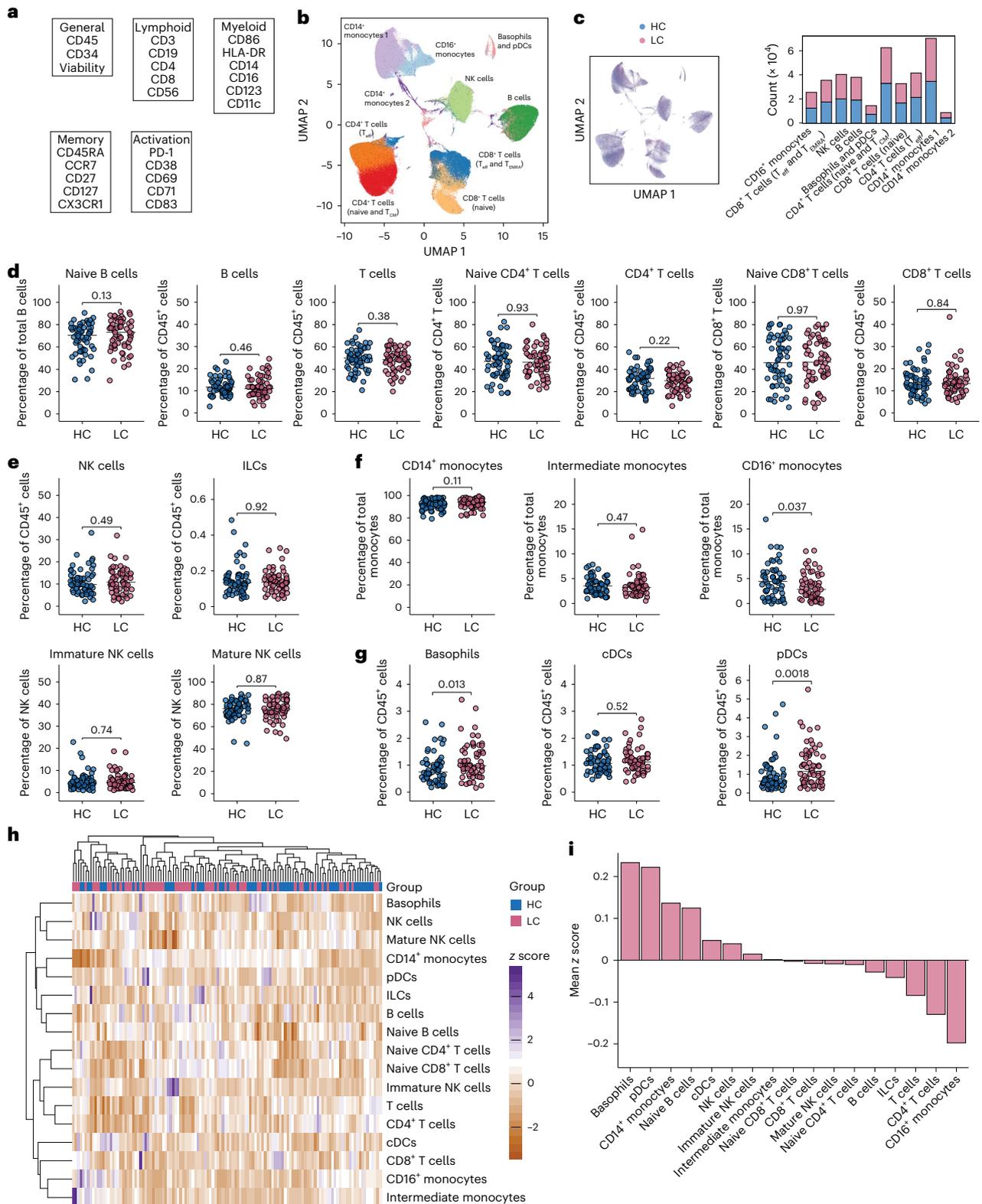
In further experiments, we measured the expression of immunophenotypic markers related to activation, memory, effector function and exhaustion among CD4 $^{+}$  and CD8 $^{+}$  T cells targeting defined proteins from SARS-CoV-2, CMV or EBV. In the primary cohort, no significant intergroup differences in expression intensity were observed for CD28, CD39, CD71, CD95, CX3CR1 or PD-1, but some markers of activation (CD38 and HLA-DR), exhaustion (TIGIT) and stemness (CD127) were variably downregulated among some antiviral CD4 $^{+}$  and CD8 $^{+}$  T cell populations in the context of long COVID (Fig. 3b). More profound differences were apparent in the secondary cohort, potentially reflecting the limited number of healthy convalescent individuals relative to the number of individuals with long COVID (Extended Data Fig. 3f). Lineage analysis further revealed comparable expression of CD38, CD69, HLA-DR and PD-1 among global CD4 $^{+}$  and CD8 $^{+}$  effector memory T cells and terminally differentiated effector memory T cells in healthy convalescent individuals and individuals with long COVID recruited from the United Kingdom (Fig. 3c).

Collectively, these results demonstrate that experimental findings are not necessarily transferable across geographically distinct cohorts of individuals with long COVID, likely reflecting differences in clinical characterization and sample size. Our findings nonetheless align with the notion that circulating antiviral CD4 $^{+}$  and CD8 $^{+}$  T cell populations are largely equivalent in healthy convalescent individuals and individuals with long COVID<sup>24</sup>.

### SARS-CoV-2-specific CD8 $^{+}$ T cells are phenotypically variable

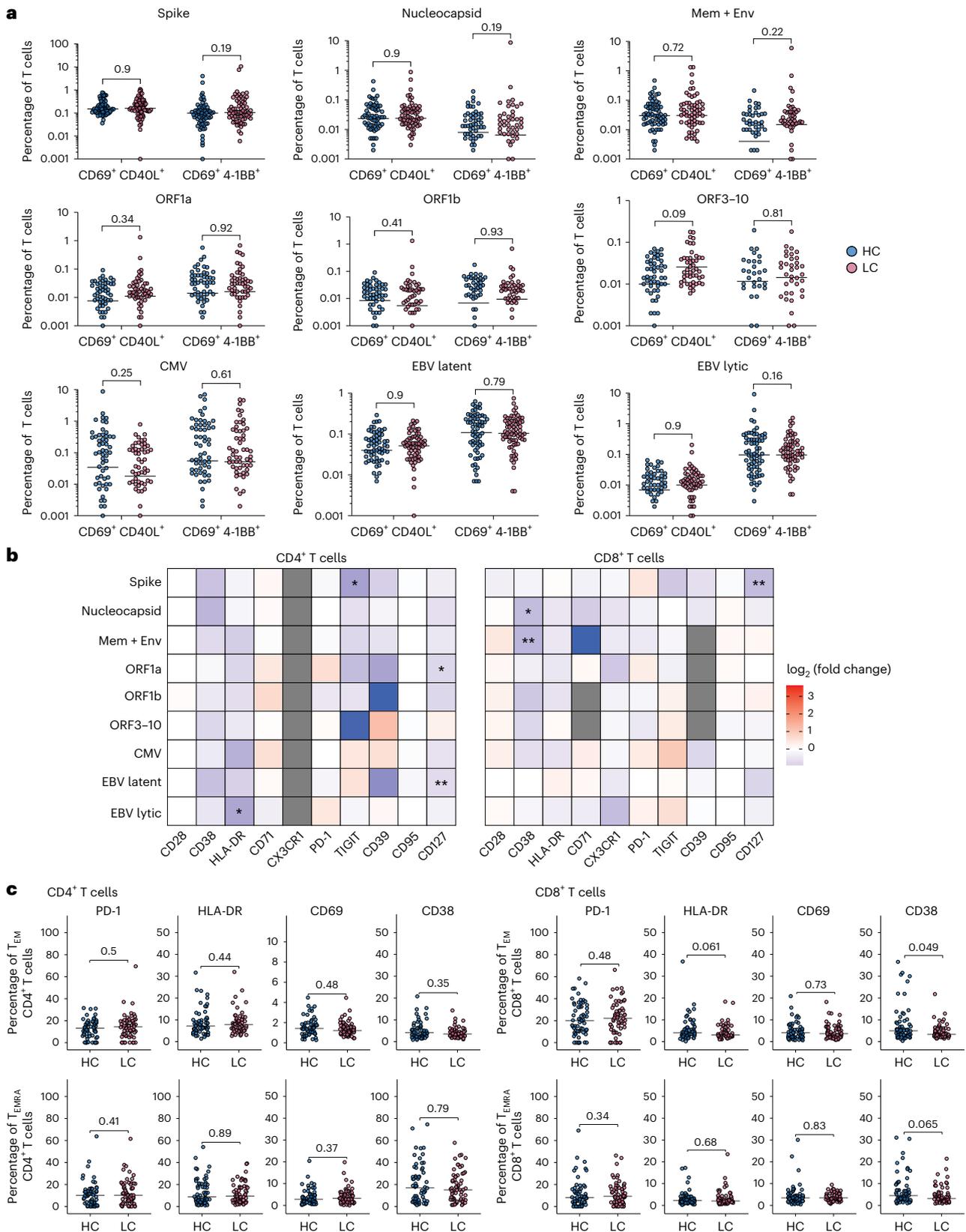
To refine our phenotypic analyses, which were potentially confounded by alterations in surface marker expression arising as a consequence of antigen-induced activation, we used peptide–HLA class I tetramers directly ex vivo to identify and characterize unperturbed CD8 $^{+}$  T cells targeting specific epitopes from SARS-CoV-2, CMV, EBV or influenza A virus (IAV)<sup>25,26</sup>. For this purpose, we selected healthy convalescent individuals ( $n = 17$ ) and individuals with long COVID ( $n = 15$ ) from the primary cohort based on the expression of HLA-A\*02:01 and/or HLA-B\*07:02. As a means to calibrate our findings against CD8 $^{+}$  T cells with known features of exhaustion<sup>27</sup>, we also performed similar analyses using samples from untreated individuals infected with human immunodeficiency virus type 1 (HIV-1), extending the range of specificities to include epitopes restricted by HLA-A\*24:02, HLA-B\*08:01 and HLA-B\*57:01 (Fig. 4a).

CD8 $^{+}$  T cells targeting spike epitopes from SARS-CoV-2 expressed CD38 and HLA-DR more frequently than CD8 $^{+}$  T cells targeting non-spike epitopes from SARS-CoV-2 (Fig. 4b and Extended Data Fig. 4a), likely as a consequence of repeated subunit vaccination. Similarly, CD8 $^{+}$  T cells targeting viral epitopes associated with persistent (CMV, EBV and HIV-1) or recurrent antigen exposure (IAV) expressed CD38 and HLA-DR more frequently than CD8 $^{+}$  T cells targeting non-spike epitopes from SARS-CoV-2, and a comparable pattern was observed



**Fig. 2 | Immune cell lineages in healthy convalescent individuals and individuals with long COVID recruited from the United Kingdom. a**, List of surface markers used to characterize immune cell lineages in the periphery. **b**, Uniform manifold approximation and projection (UMAP) representation of immune cell lineages identified via dimensionality reduction of marker expression values; T<sub>eff</sub>, effector T cells; T<sub>EMRA</sub>, terminally differentiated effector memory T cells; T<sub>CMR</sub>, central memory T cells. **c**, Distribution of cells by group of origin in UMAP space (left) or within UMAP clusters (right). **d**, Scatter dot plots showing the frequencies of naive and total B and T cells gated manually. **e**, Scatter dot plots showing the frequencies

of innate lymphocytes gated manually; ILCs, innate lymphoid cells. **f**, Scatter dot plots showing the frequencies of monocytes gated manually. **g**, Scatter dot plots showing the frequencies of basophils and DCs gated manually; cDCs, conventional DCs. **h**, Heat map showing hierarchically clustered z scores derived from the frequencies of immune cell subsets gated manually. **i**, Bar plot showing mean z scores for each immune cell subset gated manually for individuals with long COVID (healthy convalescent individuals, *n* ≤ 70; individuals with long COVID, *n* ≤ 70; **b–i**). Horizontal bars represent median values (**d–g**). Significance was evaluated using a two-tailed Mann–Whitney *U*-test (**d–g**).



**Fig. 3 | T cell immunity in healthy convalescent individuals and individuals with long COVID recruited from the United Kingdom. a**, Scatter dot plots showing the frequencies of functional CD4<sup>+</sup> and CD8<sup>+</sup> T cells targeting defined proteins from SARS-CoV-2, CMV or EBV; Mem, membrane; Env, envelope. **b**, Heat map summarizing the phenotypic attributes of functional CD4<sup>+</sup> and CD8<sup>+</sup> T cells targeting defined proteins from SARS-CoV-2, CMV or EBV. Data are shown for each marker as the log<sub>2</sub>-transformed fold change in percent positive for each

population among individuals with long COVID versus healthy convalescent individuals; \*P < 0.05 and \*\*P < 0.01. **c**, Scatter dot plots showing the frequencies of functional CD4<sup>+</sup> and CD8<sup>+</sup> effector memory T (T<sub>EM</sub>; top) cells and terminally differentiated effector memory T (T<sub>EMRA</sub>; bottom) cells expressing the indicated activation markers; healthy convalescent individuals, n ≤ 70; individuals with long COVID, n ≤ 70 (a–c). Horizontal bars represent median values (a and c). Significance was evaluated using a two-tailed Mann–Whitney U-test (a–c).

for expression of the cytotoxic serine protease granzyme B (Fig. 4b). Co-inhibitory receptor expression also varied as a function of viral specificity, typically paralleling the likely frequency of antigen exposure (Fig. 4c). Of particular note, we found that CD8<sup>+</sup> T cells targeting spike epitopes from SARS-CoV-2 expressed co-inhibitory receptors more intensely than CD8<sup>+</sup> T cells targeting nonspike epitopes from SARS-CoV-2, based on a combined score for PD-1, TIM-3, LAG-3 and TIGIT (Fig. 4d). No such differences were observed for the transcription factors TCF-1, T-BET or EOMES (Fig. 4e). However, CD8<sup>+</sup> T cells targeting epitopes from CMV or HIV-1 expressed T-BET more intensely than CD8<sup>+</sup> T cells targeting nonspike epitopes from SARS-CoV-2, and CD8<sup>+</sup> T cells targeting epitopes from CMV, EBV or HIV-1 expressed EOMES more intensely than CD8<sup>+</sup> T cells targeting nonspike epitopes from SARS-CoV-2 (Fig. 4e).

Collectively, these observations support the premise that antigen exposure drives the expression of activation markers and co-inhibitory receptors as a function of viral specificity and further suggest that such encounters are not sufficiently frequent in the convalescent phase to induce exhaustion among CD8<sup>+</sup> T cells targeting nonspike epitopes from SARS-CoV-2, irrespective of progression to long COVID.

### SARS-CoV-2-specific CD8<sup>+</sup> T cell phenotypes in long COVID

To determine if any of these phenotypic attributes segregated with disease, we visualized our flow cytometry data using the dimensionality reduction technique uniform manifold approximation and projection (UMAP), focusing on CD8<sup>+</sup> T cells targeting nonspike epitopes from SARS-CoV-2. A largely overlapping distribution was observed for healthy convalescent individuals and individuals with long COVID (Fig. 4f). Phenograph analysis further revealed seven clusters, most of which displayed an even representation (Fig. 4g). However, clusters 3 and 7 were more obviously represented among healthy convalescent individuals, and cluster 5 was more obviously represented among individuals with long COVID (Fig. 4g). Of note, cluster 5 exhibited the highest expression intensities of co-inhibitory receptors, including PD-1 (Fig. 4g and Extended Data Fig. 4b). In line with this observation, we found that CD8<sup>+</sup> T cells targeting nonspike epitopes from SARS-CoV-2 expressed co-inhibitory receptors more intensely in individuals with long COVID than in healthy convalescent individuals, reaching significance for TIM-3 (Fig. 4h). No such differences were observed for CD8<sup>+</sup> T cells targeting spike epitopes from SARS-CoV-2 (Fig. 4h). Moreover, CD8<sup>+</sup> T cells targeting nonspike epitopes from SARS-CoV-2 displayed higher co-inhibitory scores in individuals with long COVID than in healthy convalescent individuals, suggesting a link between antigen exposure and disease (Fig. 4i). It was also notable that co-inhibitory scores varied across specificities within the nonspike repertoire (Fig. 4j).

In further analyses, we found that CD8<sup>+</sup> T cells targeting nonspike or spike epitopes from SARS-CoV-2 expressed TCF-1, a key determinant of memory formation, more intensely in healthy convalescent

individuals than in individuals with long COVID (Fig. 4k). Moreover, CD8<sup>+</sup> T cells targeting lytic epitopes from EBV expressed CXCR3 more frequently, granzyme B less frequently, and TCF-1 more intensely in healthy convalescent individuals than in individuals with long COVID (Fig. 4l and Extended Data Fig. 4c). No such differences were observed for CD8<sup>+</sup> T cells targeting epitopes from CMV or CD8<sup>+</sup> T cells targeting latent epitopes from EBV (Extended Data Fig. 4d,e).

Collectively, these findings suggest a possible role for cumulative viral antigen exposure in the pathogenesis of long COVID, potentially accompanied by suboptimal immune control of EBV.

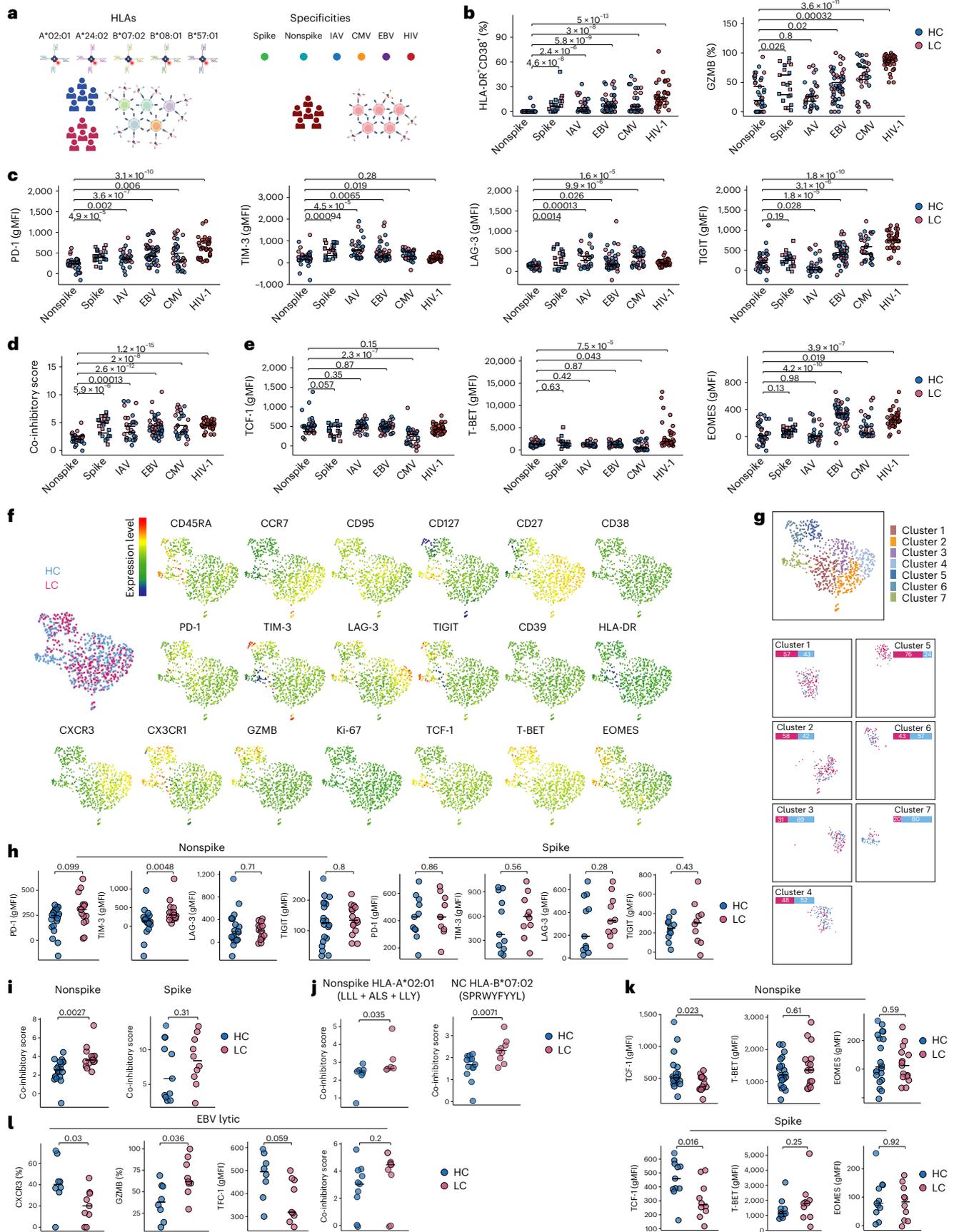
### Plasma proteomic signatures of breathlessness in long COVID

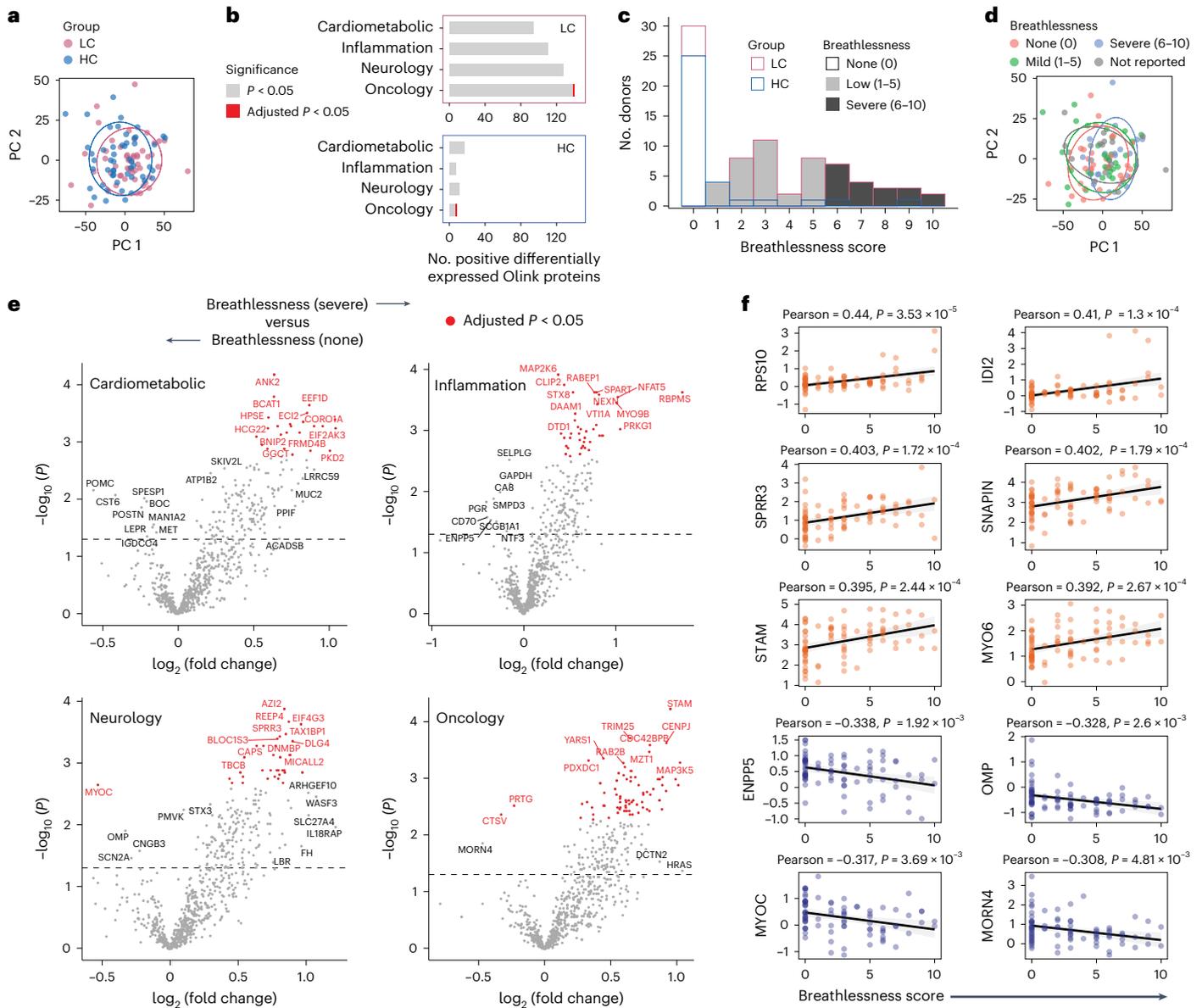
To explore disease pathogenesis more systematically, we used a data-driven approach to select healthy convalescent individuals ( $n = 51$ ) and individuals with long COVID ( $n = 51$ ) from the primary cohort for plasma proteome characterization using a Proximity Extension Assay (Olink Explore 3072). Briefly, immune cell subset proportions were summarized via principal component analysis (PCA), and outlier samples were excluded based on the greatest deviation from the origin along PC1 to PC4. Target proteins were grouped into eight panels under the following broad themes: cardiometabolic ( $n = 2$ ), inflammation ( $n = 2$ ), neurology ( $n = 2$ ) and oncology ( $n = 2$ ). PCA revealed that donors could not be separated by disease status (Fig. 5a) but could be separated to some extent by BMI (Extended Data Fig. 5a). We then performed a differential expression analysis, which revealed a skewed upregulation of many proteins in individuals with long COVID (Supplementary Table 2), although most fell below the threshold for significance after multiple-hypothesis correction (Fig. 5b and Extended Data Fig. 5b). A gene set enrichment analysis (GSEA) further showed that several pathways, including those related to ceramide, platelet-derived growth factor receptor- $\beta$  (PDGFRB) and HIV-1 Nef, were associated with this proteomic signature of long COVID (Extended Data Fig. 5c).

To extend our analyses beyond a simple binary classification, we stratified donors into three groups for each clinical symptom, irrespective of the initial categorization as healthy convalescent individuals or individuals with long COVID (Fig. 5c and Extended Data Fig. 5d). Using this approach, we found that breathlessness was strongly associated with differential protein expression (Extended Data Fig. 5d). Donors with severe breathlessness (score of 6–10) segregated from donors with no (score of 0) or mild breathlessness (score of 1–5) via PCA (Fig. 5d) and exhibited distinct patterns of protein upregulation (Fig. 5e and Supplementary Table 3). Moreover, GSEA confirmed that severe breathlessness was associated with the enrichment of several pathways that characterized the proteomic signature of long COVID, including those related to ceramide and HIV-1 Nef (Extended Data Fig. 5e). Of note, breathlessness and other symptom scores were largely independent of age and correlated only weakly with BMI (Extended Data Fig. 5f,g), which is known to impact the plasma proteome<sup>28</sup>.

**Fig. 4 | Phenotypic characteristics of virus-specific CD8<sup>+</sup> T cells in healthy convalescent individuals and individuals with long COVID recruited from the United Kingdom.** **a**, Schematic representation of the experimental design. **b**, Scatter dot plots showing the expression frequencies of HLA-DR and CD38 or granzyme B (GZMB) among tetramer<sup>+</sup>CD8<sup>+</sup> T cells. **c**, Scatter dot plots showing the expression intensities of co-inhibitory receptors among tetramer<sup>+</sup>CD8<sup>+</sup> T cells. **d**, Scatter dot plot showing co-inhibitory scores, calculated as the cumulative normalized expression intensities of the co-inhibitory receptors shown in **c**, among tetramer<sup>+</sup>CD8<sup>+</sup> T cells. **e**, Scatter dot plots showing the expression intensities of transcription factors among tetramer<sup>+</sup>CD8<sup>+</sup> T cells. **f**, UMAP visualization summarizing the phenotypic characteristics of tetramer<sup>+</sup>CD8<sup>+</sup> T cells targeting nonspike epitopes from SARS-CoV-2. Individual marker representations are colored by expression intensity. **g**, Phenograph clustering (top) and cluster distribution of tetramer<sup>+</sup>CD8<sup>+</sup> T cells (bottom). **h**, Scatter dot plots showing the expression

intensities of co-inhibitory receptors among tetramer<sup>+</sup>CD8<sup>+</sup> T cells targeting nonspike or spike epitopes from SARS-CoV-2. **i**, Scatter dot plots showing co-inhibitory scores among tetramer<sup>+</sup>CD8<sup>+</sup> T cells targeting nonspike or spike epitopes from SARS-CoV-2. **j**, Scatter dot plots showing co-inhibitory scores among tetramer<sup>+</sup>CD8<sup>+</sup> T cells targeting nonspike epitopes from SARS-CoV-2 restricted by HLA-A\*02:01 or HLA-B\*07:02; NC, nucleocapsid. **k**, Scatter dot plots showing the expression intensities of transcription factors among tetramer<sup>+</sup>CD8<sup>+</sup> T cells targeting nonspike or spike epitopes from SARS-CoV-2. **l**, Scatter dot plots showing the phenotypic characteristics of tetramer<sup>+</sup>CD8<sup>+</sup> T cells targeting lytic epitopes from EBV; healthy convalescent individuals,  $n = 17$ ; individuals with long COVID,  $n = 15$ ; untreated individuals infected with HIV-1,  $n = 14$  (**b–l**). Horizontal bars represent median values (**b–e** and **h–l**). Significance was evaluated using a two-tailed Mann–Whitney  $U$ -test (**b–e** and **h–l**); gMFI, geometric mean fluorescence intensity.





**Fig. 5 | Dysregulation of the plasma proteome associated with breathlessness in healthy convalescent individuals and individuals with long COVID recruited from the United Kingdom.** **a**, PCA of plasma protein concentrations colored by donor group for healthy convalescent individuals ( $n = 51$ ) and individuals with long COVID ( $n = 51$ ). **b**, Bar plots showing the corresponding numbers of differentially upregulated plasma proteins from each panel. Significance was evaluated using a two-tailed Mann–Whitney  $U$ -test with (red) or without (gray) Benjamini–Hochberg correction. **c**, Stacked histogram showing the distribution of breathlessness scores for healthy convalescent individuals ( $n = 34$ ) and individuals with long COVID ( $n = 48$ ). **d**, PCA of plasma protein concentrations colored by breathlessness score tiers for healthy convalescent individuals ( $n = 51$ ) and individuals with long COVID ( $n = 51$ ), irrespective of clinical assignment. **e**, Volcano plots showing the corresponding differentially

expressed plasma proteins from each panel versus the highest and lowest breathlessness score tiers, irrespective of clinical assignment. Significance was evaluated using a two-tailed Mann–Whitney  $U$ -test with (red) or without (gray) Benjamini–Hochberg correction. The dashed line indicates  $P = 0.05$ . **f**, Correlation dot plots showing the highest ( $n = 6$ ) and lowest ranked plasma proteins ( $n = 4$ ) in terms of normalized expression versus breathlessness scores for healthy convalescent individuals ( $n = 51$ ) and individuals with long COVID ( $n = 51$ ), irrespective of clinical assignment; IDI2, isopentenyl-diphosphate  $\delta$ -isomerase-2; SPRR3, small proline-rich protein 3; ENPP5, ectonucleotide pyrophosphatase/phosphodiesterase family member 5; OMP, olfactory marker protein. Significance was evaluated using the two-tailed Pearson coefficient. Shading indicates the 95% confidence interval for each regression line.

To identify specific proteins associated with breathlessness, we performed a correlation analysis without prior stratification based on symptom severity. The concentrations of almost all plasma proteins were skewed toward a positive correlation with breathlessness score (Extended Data Fig. 6a and Supplementary Table 4). The most positively correlated proteins included isopentenyl-diphosphate  $\delta$ -isomerase 2 and small proline-rich protein 3, and the most negatively correlated proteins included ectonucleotide pyrophosphatase/

phosphodiesterase family member 5 and olfactory marker protein (Fig. 5f). We then used the list of proteins ranked according to correlation with breathlessness to perform a GSEA, which showed that dysregulation of the plasma proteome was associated with related phenotypes, such as atelectasis (lung collapse) and tachypnea (rapid breathing), and further revealed an enrichment for pathways linked to cell cycle progression (for example, RhoA), inflammation (for example, TNF) and platelet activation (for example, PDGFRB and thromboxane

A2 (TXA2); Extended Data Fig. 6b). In a further step designed to identify proteins with outsized roles in the breathlessness signatures associated with inflammation, we performed network analyses using Cytoscape. The output highlighted a complex protein network centered around CD40 in a module that also included CCL3, CCL4, IKBKG and IL-18 (Extended Data Fig. 6c).

To validate these findings, we performed similar analyses using plasma samples from donors in the secondary cohort, focusing on individuals with long COVID classified as low (0–2;  $n = 60$ ) or moderate (3–7;  $n = 35$ ) according to the Borg CR10 scale, which measures perceived exertion during physical activity (Fig. 6a)<sup>29</sup>. Differential expression analysis revealed upregulated protein expression among individuals with a moderate score, although no markers achieved significance after multiple-hypothesis correction (Fig. 6b and Supplementary Table 5). GSEA of the ranked list of proteins nonetheless identified enrichment of signaling pathways observed in the primary cohort (Extended Data Fig. 5e), including MET and PI3K/AKT/MTOR (Fig. 6c). Unbiased analyses further revealed that high Borg CR10 scores were correlated with pathways enriched among individuals in the primary cohort with severe breathlessness, including those associated with ceramide, syndecan-4 and TXA2 (Extended Data Fig. 7a,b and Supplementary Table 6).

To unify these data, we correlated protein expression across the primary and secondary cohorts as a function of symptom severity. A total of 275 proteins were differentially expressed among individuals with severe breathlessness and individuals with a moderate Borg CR10 score (Fig. 6d and Supplementary Table 7). Of these, all but three were consistently upregulated in both cohorts among donors with greater symptom severity, suggesting a shared signature of plasma proteome dysregulation in individuals with long COVID. Network analysis of differentially expressed inflammatory markers in the secondary cohort further identified a major hub centered around CD40 (Extended Data Fig. 7c), reminiscent of the primary cohort pattern (Extended Data Fig. 6c). A similar analysis of significantly upregulated proteins from the inflammation panel ( $n = 82$ ) spanning both cohorts revealed that the most confident network connections were centered around CCL3, CD40, IKBKG, IL-18 and IRAK1 (Fig. 6e). Many of these proteins are associated with the NF- $\kappa$ B pathway. In addition, overrepresentation analysis of upregulated proteins spanning all panels across both cohorts identified apoptosis as the most significant hit among all gene sets in the Hallmark Collection and ceramide and FAS as highly significant hits in the Pathway Interaction Database (Fig. 6f). Of note, proteins associated with these pathways, including apoptosis-inducing factor mitochondria-associated 1, caspase-3, caspase-7 and IL-18, were specifically upregulated among donors with severe breathlessness recruited from the United Kingdom (Fig. 6g).

Collectively, these results identify dysregulated plasma proteins that could serve as biomarkers of persistent breathlessness after infection with SARS-CoV-2, potentially facilitating the diagnosis and treatment of long COVID.

## Discussion

Long COVID continues to pose medical challenges with unmet diagnostic and therapeutic needs that reflect the elusive mechanistic nature of a symptomatically heterogeneous disease. In this study, we used high-dimensional flow cytometry and plasma proteomics to seek biomarkers that could inform the pathogenesis of long COVID. Quantitative differences in immune cell lineage composition and virus-specific CD4<sup>+</sup> and CD8<sup>+</sup> T cell immunity were minimal and nonreproducible across two geographically distinct cohorts in direct comparisons of healthy convalescent individuals and individuals with long COVID. Antibody neutralization activity was nonetheless significantly higher in healthy convalescent individuals than in individuals with long COVID, despite comparable SARS-CoV-2 spike-specific IgG titers and equivalent levels of ADNKA, and some co-inhibitory receptors, especially

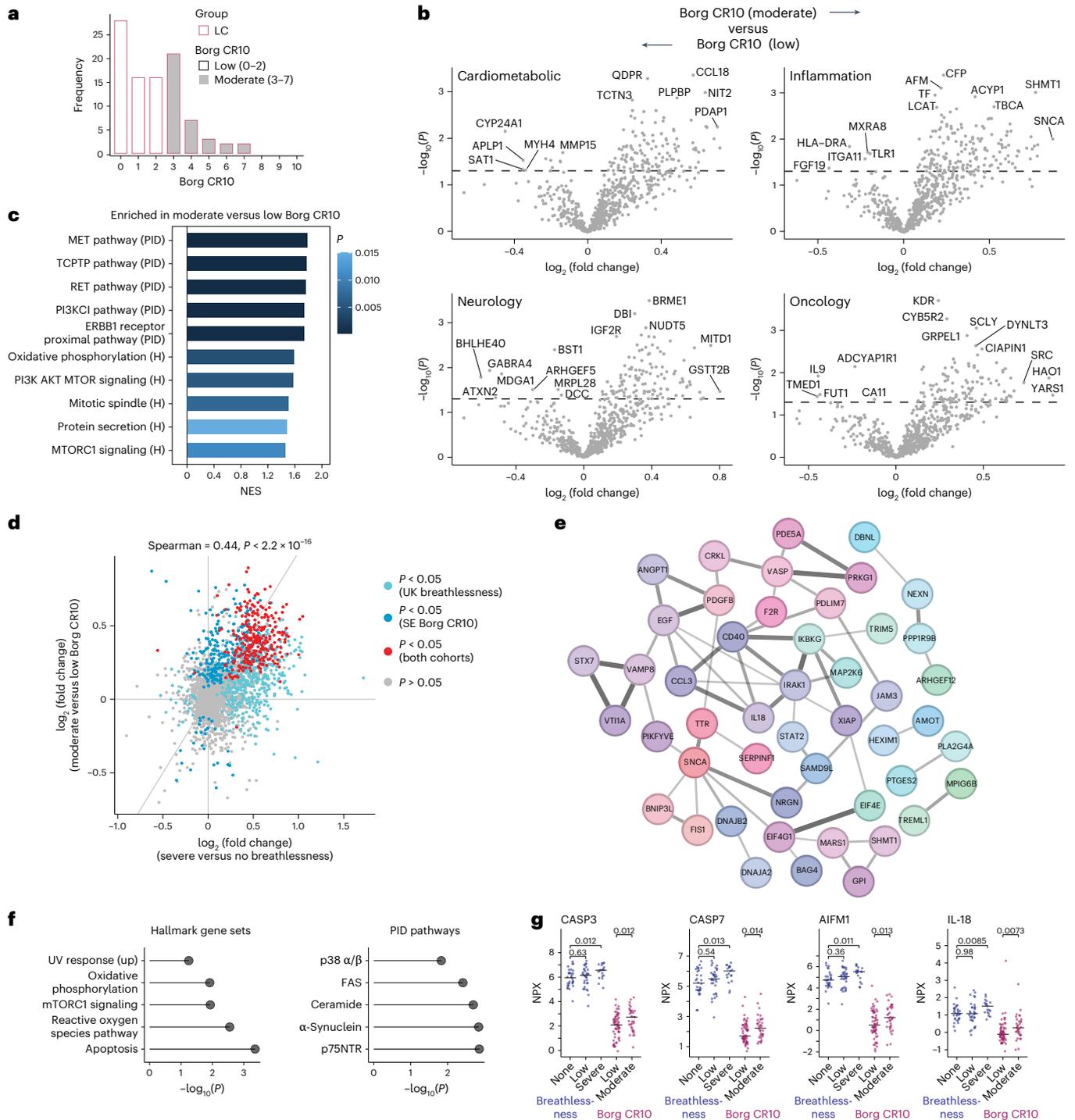
PD-1 and TIM-3, were relatively overexpressed among SARS-CoV-2 nonspike-specific CD8<sup>+</sup> T cells in individuals with long COVID. Our data also revealed an informative plasma biomarker signature linking persistent respiratory symptoms with apoptotic inflammatory networks and pathway dysregulation indicative of cell cycle progression, lung injury and platelet activation in individuals with long COVID.

Donor groups in our primary cohort were carefully matched for age, BMI, race, sex, time since infection, and vaccination against SARS-CoV-2, thereby minimizing the impact of confounding factors that could potentially bias comparative analyses of healthy convalescent individuals and individuals with long COVID. Women were overrepresented as a consequence<sup>30–32</sup>. The predominant symptoms were breathlessness, fatigue, pain, mobility issues, anxiety and depression, which align with the known clinical spectrum of long COVID<sup>3</sup>. Pain was localized primarily to the chest, joints and muscles, again consistent with distributions reported in other individuals with long COVID<sup>33,34</sup>. In contrast to influenza virus and other acute respiratory pathogens, which predominantly exacerbate localized symptoms during and after infection, these diverse postacute sequelae likely reflect an underlying etiological complexity, which mandates a systematic approach to the diagnosis and management of individuals with long COVID<sup>35</sup>.

In contrast to a recent study<sup>16</sup>, we found that healthy convalescent individuals were better able to neutralize SARS-CoV-2 than individuals with long COVID. This observation suggests a qualitative difference in antibody induction, potentially reflecting the fact that healthy convalescent individuals were vaccinated more frequently before infection than individuals with long COVID, which could help mitigate the risk of persistent disease<sup>36</sup>. No such differences were detected with respect to overall SARS-CoV-2 spike-specific IgG titers or ADNKA. This latter finding could be explained by the functional equivalence of antibodies targeting the spike protein<sup>37</sup> and/or by the availability of nonspike targets expressed on the cell surface after infection with SARS-CoV-2 (ref. 38).

Systemic immune perturbations are thought to play a role in the pathogenesis of long COVID<sup>3,21</sup>. For example, innate immune cell activation and a paucity of naive B and T cells have been described in one cohort of individuals with long COVID<sup>8</sup>, whereas a relative abundance of highly cytotoxic CD8<sup>+</sup> T cells and NK cells has been described in another cohort of individuals with long COVID<sup>39</sup>. We found only nuanced differences between the immune cell lineage profiles of healthy convalescent individuals and individuals with long COVID. In the primary cohort, these differences were limited to nonclassical monocytes, which were relatively overrepresented in healthy convalescent individuals, and basophils and pDCs, which were relatively overrepresented in individuals with long COVID, whereas in the secondary cohort, these differences were limited to classical monocytes, which were relatively overrepresented in healthy convalescent individuals, and intermediate monocytes, which were relatively overrepresented in individuals with long COVID. Such inconsistencies likely reflect a number of factors, including comorbidities and disease heterogeneity, and underscore the importance of cross-validation in studies of long COVID<sup>3,10,40,41</sup>.

SARS-CoV-2 proteins can be detected in many tissues long after the acute infectious event and could potentially engender a state of chronic immune activation linked with the development of long COVID<sup>6,42,43</sup>. It is also known that CD4<sup>+</sup> and CD8<sup>+</sup> memory T cells provide durable immunity against SARS-CoV-2 (refs. 22,23,25,44). These cells are exquisitely poised to mount anamnestic responses and would likely proliferate and shift to an activated and/or exhausted phenotype under conditions of recurrent antigen stimulation associated with a failure to clear residual viral products and/or ongoing viral replication, thereby feasibly becoming immunopathogenic rather than protective in the context of long COVID<sup>10,45</sup>. In line with this notion, one study reported sustained SARS-CoV-2-specific CD4<sup>+</sup> T cell responses during late recovery in individuals with long COVID<sup>46</sup>, and another study reported enhanced expression of the exhaustion markers CTLA-4 and PD-1 among SARS-CoV-2-specific CD8<sup>+</sup> T cells



**Fig. 6 | Shared signatures of plasma proteome dysregulation associated with respiratory symptoms in healthy convalescent individuals and individuals with long COVID. a**, Bar plot showing the distribution of Borg CR10 scores for individuals in the secondary cohort with long COVID ( $n = 95$ ). **b**, Volcano plots showing differentially expressed plasma proteins from each panel versus Borg CR10 score tiers for individuals in the secondary cohort with long COVID ( $n = 95$ ). Significance was evaluated using a two-tailed Mann-Whitney  $U$ -test. The dashed line indicates  $P = 0.05$ . No proteins achieved adjusted  $P < 0.05$  after Benjamini-Hochberg correction. **c**, GSEA showing differentially expressed plasma proteins by rank versus the Borg CR10 score tiers for individuals with long COVID ( $n = 95$ ), irrespective of clinical assignment. The top five terms from the Hallmark (H) and Pathway Interaction Database (PID) gene sets (Molecular Signatures Database (MSigDB) Collections) are shown. Significance was evaluated using the GSEA method without correction; NES, normalized enrichment score. **d**, Scatter dot plot showing individual plasma protein fold change across breathlessness score

tiers (primary cohort, x axis) and Borg CR10 score tiers (secondary cohort, y axis). Significance was evaluated using a two-tailed Spearman rank test. Proteins are colored according to significance without Benjamini-Hochberg correction. **e**, Network analysis showing differentially expressed plasma proteins from the inflammation panel across both cohorts depicted using Cytoscape. Nodes and edges represent proteins and functional relevance, respectively. Edge thickness represents the level of confidence. **f**, Overrepresentation analysis of significantly upregulated plasma proteins across both cohorts showing the five top terms from the Hallmark Collection and the Pathway Interaction Database. Significance was evaluated using a hypergeometric test. **g**, Comparison of plasma protein concentrations between cohorts split by symptom severity. Horizontal bars represent median values. Significance was evaluated using a two-tailed Mann-Whitney  $U$ -test; NPX, normalized protein expression; AIFM1, apoptosis-inducing factor mitochondria-associated 1; CASP, caspase.

in individuals with long COVID<sup>14</sup>. The converse scenario in terms of response magnitude has also been described for IFN $\gamma$ -producing CD8<sup>+</sup> T cells targeting the nucleocapsid protein of SARS-CoV-2 (ref. 47). In our primary cohort, SARS-CoV-2-specific CD4<sup>+</sup> and CD8<sup>+</sup> T cell responses were comparable in magnitude across the entire viral proteome in healthy convalescent individuals and individuals with long COVID, whereas in our secondary cohort, relatively elevated frequencies of SARS-CoV-2 nucleocapsid-specific CD4<sup>+</sup> T cells and SARS-CoV-2 spike-specific CD8<sup>+</sup> T cells were observed in individuals with long COVID. However, more refined analyses of the primary cohort revealed altered memory profiles and enhanced co-inhibitory scores among SARS-CoV-2 nonspike-specific CD8<sup>+</sup> T cells in individuals with long COVID, indicating a relatively greater cumulative history of exposure to antigens derived from SARS-CoV-2. An alternative possibility is that immune exhaustion facilitates viral persistence, but further studies are required to determine the protective versus reactive properties of SARS-CoV-2-specific CD4<sup>+</sup> and CD8<sup>+</sup> T cells in relation to the pathogenesis of long COVID.

Many factors can affect immune responses against SARS-CoV-2, including genetic background, infection history and vaccination status<sup>7,48–50</sup>, and many factors beyond immune responses against SARS-CoV-2 have been linked with the pathogenesis of long COVID<sup>3,51</sup>, including reactivation of the herpesviruses CMV and/or EBV<sup>3,10,11,52,53</sup>. Most of these latter associations have been defined serologically<sup>16,20,54</sup>. We addressed the same issue by interrogating CD4<sup>+</sup> and CD8<sup>+</sup> T cells targeting immunodominant regions of CMV or EBV. In the primary cohort, no intergroup differences in response magnitude were detected for any specificity, whereas in the secondary cohort, the frequencies of CD4<sup>+</sup> T cells targeting EBV latent proteins and the frequencies of CD8<sup>+</sup> T cells targeting CMV proteins or EBV lytic proteins were relatively elevated in individuals with long COVID. Phenotypic analyses focused on the primary cohort further revealed high co-inhibitory scores among CD8<sup>+</sup> T cells targeting epitopes from CMV, which overexpressed PD-1, and a terminally differentiated profile among CD8<sup>+</sup> T cells targeting lytic epitopes from EBV in individuals with long COVID. In line with these observations, we found that SARS-CoV-2 spike-specific CD8<sup>+</sup> T cells overexpressed various activation markers, including CD38 and HLA-DR, and various co-inhibitory receptors spanning PD-1, TIM-3, LAG-3 and TIGIT, likely reflecting recurrent antigen exposure as a consequence of repeated subunit vaccination<sup>55</sup>. Accordingly, our data align with the notion that bystander viral reactivation frequently accompanies the development of persistent disease after infection with SARS-CoV-2 (refs. 10,54,56).

Plasma proteomics has emerged as a useful strategy to help decipher the molecular basis of various diseases via the identification of systemic biomarkers indicative of tissue-localized pathology<sup>12,57–59</sup>. Using a high-throughput platform in conjunction with a symptom-targeted approach, we found that severe breathlessness was associated with extensive dysregulation of the plasma proteome. It should be noted that our approach was focused on a curated panel of proteins spanning a targeted fraction of the entire proteome, such that we potentially failed to identify some biomarkers and pathways characteristic of long COVID. Our findings nonetheless align broadly with other strands of evidence indicating that chronic inflammation is a cornerstone of long COVID<sup>8,12,14,15,60</sup>. Moreover, network analyses identified connections centered around CD40, incorporating various caspases (CASP2 and CASP7) and kinases (IKBK and MAP2K6), collectively linking breathlessness with inflammatory apoptosis and/or cell death, which could feasibly reflect ongoing exposure to antigens derived from SARS-CoV-2 (refs. 61,62). Similar inflammatory profiles have been identified previously in individuals with long COVID<sup>12,13</sup>. Pathway analyses further identified dysregulated proteins associated with cell cycle progression (for example, RhoA) and platelet activation (for example, PDGFRB and TXA2). In sum, these observations fit with a dynamic process of lung damage and remodeling attributable

primarily to hypercoagulability and thromboinflammation<sup>18</sup>, potentially accompanied by amyloid fibrin microclot deposition<sup>63</sup>, endothelial dysfunction<sup>64</sup> and vasculoproliferation<sup>65</sup>, which collectively impair oxygen exchange and lead to the sensation of breathlessness in individuals with long COVID.

One limitation of our study was that the matching process for the primary cohort was not entirely accurate, such that healthy convalescent individuals were sampled earlier after infection (median = 268 days) than individuals with long COVID (median = 416 days). This discrepancy combined with a preferential loss of functionally optimal antibodies could partially explain the relative paucity of neutralization activity in individuals with long COVID. A minority of healthy convalescent individuals also reported breathlessness as a symptom, likely attributable to other pathologies affecting the respiratory system, which were not assessed clinically. Moreover, our approach was limited to samples acquired from the vascular circulation, which emerging evidence suggests is a highly specialized immunological niche<sup>66</sup>. In addition, the origins and roles of proteins detected in plasma samples are open to interpretation, providing only indirect evidence for any given underlying pathology. Comparative analyses of disease-relevant tissue samples will therefore be required to validate the localized pathology associated with our reported systemic cellular and molecular signatures of long COVID.

In summary, our findings suggest that lung damage associated with the canonical symptom of breathlessness can be identified via the systemic upregulation of multiple apoptotic, cardiovascular and inflammatory biomarkers in the presence of a largely unperturbed cellular immune system, indicative of localized tissue pathology and ongoing but minimal exposure to viral antigens potentially facilitated by suboptimal humoral immunity in individuals with long COVID.

## 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/s41590-025-02135-5>.

## References

1. Nalbandian, A. et al. Post-acute COVID-19 syndrome. *Nat. Med.* **27**, 601–615 (2021).
2. Subramanian, A. et al. Symptoms and risk factors for long COVID in non-hospitalized adults. *Nat. Med.* **28**, 1706–1714 (2022).
3. Davis, H. E., McCorkell, L., Vogel, J. M. & Topol, E. J. Long COVID: major findings, mechanisms and recommendations. *Nat. Rev. Microbiol.* **21**, 133–146 (2023).
4. Chen, B., Julg, B., Mohandas, S. & Bradfute, S. B. RECOVER Mechanistic Pathways Task Force. Viral persistence, reactivation, and mechanisms of long COVID. *eLife* **12**, e86015 (2023).
5. Zadeh, F. H., Wilson, D. R. & Agrawal, D. K. Long COVID: complications, underlying mechanisms, and treatment strategies. *Arch. Microbiol. Immunol.* **7**, 36–61 (2023).
6. Castanares-Zapatero, D. et al. Pathophysiology and mechanism of long COVID: a comprehensive review. *Ann. Med.* **54**, 1473–1487 (2022).
7. Stein, S. R. et al. SARS-CoV-2 infection and persistence in the human body and brain at autopsy. *Nature* **612**, 758–763 (2022).
8. Phetsouphanh, C. et al. Immunological dysfunction persists for 8 months following initial mild-to-moderate SARS-CoV-2 infection. *Nat. Immunol.* **23**, 210–216 (2022).
9. Proal, A. D. et al. SARS-CoV-2 reservoir in post-acute sequelae of COVID-19 (PASC). *Nat. Immunol.* **24**, 1616–1627 (2023).
10. Su, Y. et al. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell* **185**, 881–895 (2022).

11. Sherif, Z. A. et al. Pathogenic mechanisms of post-acute sequelae of SARS-CoV-2 infection (PASC). *eLife* **12**, e86002 (2023).
12. Talla, A. et al. Persistent serum protein signatures define an inflammatory subcategory of long COVID. *Nat. Commun.* **14**, 3417 (2023).
13. Suhre, K. et al. Identification of robust protein associations with COVID-19 disease based on five clinical studies. *Front. Immunol.* **12**, 781100 (2021).
14. Yin, K. et al. Long COVID manifests with T cell dysregulation, inflammation and an uncoordinated adaptive immune response to SARS-CoV-2. *Nat. Immunol.* **25**, 218–225 (2024).
15. Schultheiss, C. et al. The IL-1 $\beta$ , IL-6, and TNF cytokine triad is associated with post-acute sequelae of COVID-19. *Cell Rep. Med.* **3**, 100663 (2022).
16. Klein, J. et al. Distinguishing features of long COVID identified through immune profiling. *Nature* **623**, 139–148 (2023).
17. Baillie, K. et al. Complement dysregulation is a prevalent and therapeutically amenable feature of long COVID. *Med.* **5**, 239–253 (2024).
18. Cervia-Hasler, C. et al. Persistent complement dysregulation with signs of thromboinflammation in active long COVID. *Science* **383**, eadg7942 (2024).
19. Chattopadhyay, P. K. et al. The cytolytic enzymes granzyme A, granzyme B, and perforin: expression patterns, cell distribution, and their relationship to cell maturity and bright CD57 expression. *J. Leukoc. Biol.* **85**, 88–97 (2009).
20. Peluso, M. J. et al. Chronic viral coinfections differentially affect the likelihood of developing long COVID. *J. Clin. Invest.* **133**, e163669 (2023).
21. Altmann, D. M., Whettlock, E. M., Liu, S., Arachchilage, D. J. & Boyton, R. J. The immunology of long COVID. *Nat. Rev. Immunol.* **23**, 618–634 (2023).
22. Gao, Y. et al. Immunodeficiency syndromes differentially impact the functional profile of SARS-CoV-2-specific T cells elicited by mRNA vaccination. *Immunity* **55**, 1732–1746 (2022).
23. Gao, Y. et al. Ancestral SARS-CoV-2-specific T cells cross-recognize the Omicron variant. *Nat. Med.* **28**, 472–476 (2022).
24. Altmann, D. M. et al. Persistent symptoms after COVID-19 are not associated with differential SARS-CoV-2 antibody or T cell immunity. *Nat. Commun.* **14**, 5139 (2023).
25. Sekine, T. et al. Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. *Cell* **183**, 158–168 (2020).
26. Adamo, S. et al. Memory profiles distinguish cross-reactive and virus-specific T cell immunity to mpox. *Cell Host Microbe* **31**, 928–936 (2023).
27. Buggert, M. et al. T-bet and Eomes are differentially linked to the exhausted phenotype of CD8<sup>+</sup> T cells in HIV infection. *PLoS Pathog.* **10**, e1004251 (2014).
28. Zaghlool, S. B. et al. Revealing the role of the human blood plasma proteome in obesity using genetic drivers. *Nat. Commun.* **12**, 1279 (2021).
29. Williams, N. The Borg Rating of Perceived Exertion (RPE) scale. *Occup. Med.* **67**, 404–405 (2017).
30. Takahashi, T. et al. Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature* **588**, 315–320 (2020).
31. Ganesh, R. et al. The female-predominant persistent immune dysregulation of the post-COVID syndrome. *Mayo Clin. Proc.* **97**, 454–464 (2022).
32. Bai, F. et al. Female gender is associated with long COVID syndrome: a prospective cohort study. *Clin. Microbiol. Infect.* **28**, 611.e9–611.e16 (2022).
33. Choutka, J., Jansari, V., Hornig, M. & Iwasaki, A. Unexplained post-acute infection syndromes. *Nat. Med.* **28**, 911–923 (2022).
34. Al-Aly, Z., Bowe, B. & Xie, Y. Long COVID after breakthrough SARS-CoV-2 infection. *Nat. Med.* **28**, 1461–1467 (2022).
35. Xie, Y., Choi, T. & Al-Aly, Z. Long-term outcomes following hospital admission for COVID-19 versus seasonal influenza: a cohort study. *Lancet Infect. Dis.* **24**, 239–255 (2024).
36. Catala, M. et al. The effectiveness of COVID-19 vaccines to prevent long COVID symptoms: staggered cohort study of data from the UK, Spain, and Estonia. *Lancet Respir. Med.* **12**, 225–236 (2024).
37. Grant, M. D. et al. Combined anti-S1 and anti-S2 antibodies from hybrid immunity elicit potent cross-variant ADCC against SARS-CoV-2. *JCI Insight* **8**, e170681 (2023).
38. Fielding, C. A. et al. SARS-CoV-2 host-shutoff impacts innate NK cell functions, but antibody-dependent NK activity is strongly activated through non-spike antibodies. *eLife* **11**, e74489 (2022).
39. Galan, M. et al. Persistent overactive cytotoxic immune response in a Spanish cohort of individuals with long-COVID: identification of diagnostic biomarkers. *Front. Immunol.* **13**, 848886 (2022).
40. Taquet, M. et al. Incidence, co-occurrence, and evolution of long-COVID features: a 6-month retrospective cohort study of 273,618 survivors of COVID-19. *PLoS Med.* **18**, e1003773 (2021).
41. Davis, H. E. et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *EClinicalMedicine* **38**, 101019 (2021).
42. Raveendran, A. V., Jayadevan, R. & Sashidharan, S. Long COVID: an overview. *Diabetes Metab. Syndr.* **15**, 869–875 (2021).
43. Crook, H., Raza, S., Nowell, J., Young, M. & Edison, P. Long COVID—mechanisms, risk factors, and management. *BMJ* **374**, n1648 (2021).
44. Tarke, A. et al. SARS-CoV-2 vaccination induces immunological T cell memory able to cross-recognize variants from Alpha to Omicron. *Cell* **185**, 847–859 (2022).
45. Wong, L. R. & Perlman, S. Immune dysregulation and immunopathology induced by SARS-CoV-2 and related coronaviruses—are we our own worst enemy? *Nat. Rev. Immunol.* **22**, 47–56 (2022).
46. Files, J. K. et al. Duration of post-COVID-19 symptoms is associated with sustained SARS-CoV-2-specific immune responses. *JCI Insight* **6**, 151544 (2021).
47. Peluso, M. J. et al. Long-term SARS-CoV-2-specific immune and inflammatory responses in individuals recovering from COVID-19 with and without post-acute symptoms. *Cell Rep.* **36**, 109518 (2021).
48. Silva Andrade, B. et al. Long-COVID and post-COVID health complications: an up-to-date review on clinical conditions and their possible molecular mechanisms. *Viruses* **13**, 700 (2021).
49. Cheung, C. C. L. et al. Residual SARS-CoV-2 viral antigens detected in GI and hepatic tissues from five recovered patients with COVID-19. *Gut* **71**, 226–229 (2022).
50. Alwan, N. A. The road to addressing long COVID. *Science* **373**, 491–493 (2021).
51. Astin, R. et al. Long COVID: mechanisms, risk factors and recovery. *Exp. Physiol.* **108**, 12–27 (2023).
52. Ahamed, J. & Laurence, J. Long COVID endotheliopathy: hypothesized mechanisms and potential therapeutic approaches. *J. Clin. Invest.* **132**, e161167 (2022).
53. Vojdani, A., Vojdani, E., Saidara, E. & Maes, M. Persistent SARS-CoV-2 infection, EBV, HHV-6 and other factors may contribute to inflammation and autoimmunity in long COVID. *Viruses* **15**, 400 (2023).
54. 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).
55. Cai, C. et al. SARS-CoV-2 vaccination enhances the effector qualities of spike-specific T cells induced by COVID-19. *Sci. Immunol.* **8**, eadh0687 (2023).

56. Peluso, M. J. & Deeks, S. G. Early clues regarding the pathogenesis of long-COVID. *Trends Immunol.* **43**, 268–270 (2022).
57. Shu, T. et al. Plasma proteomics identify biomarkers and pathogenesis of COVID-19. *Immunity* **53**, 1108–1122 (2020).
58. Al-Nesf, M. A. Y. et al. Prognostic tools and candidate drugs based on plasma proteomics of patients with severe COVID-19 complications. *Nat. Commun.* **13**, 946 (2022).
59. Gu, X. et al. Probing long COVID through a proteomic lens: a comprehensive two-year longitudinal cohort study of hospitalised survivors. *EBioMedicine* **98**, 104851 (2023).
60. Peluso, M. J. et al. Markers of immune activation and inflammation in individuals with postacute sequelae of severe acute respiratory syndrome coronavirus 2 infection. *J. Infect. Dis.* **224**, 1839–1848 (2021).
61. Krishnamachary, B. et al. Extracellular vesicle-mediated endothelial apoptosis and EV-associated proteins correlate with COVID-19 disease severity. *J. Extracell. Vesicles* **10**, e12117 (2021).
62. Andre, S. et al. T cell apoptosis characterizes severe COVID-19 disease. *Cell Death Differ.* **29**, 1486–1499 (2022).
63. Kell, D. B., Laubscher, G. J. & Pretorius, E. A central role for amyloid fibrin microclots in long COVID/PASC: origins and therapeutic implications. *Biochem. J.* **479**, 537–559 (2022).
64. Boccatonda, A., Campello, E., Simion, C. & Simioni, P. Long-term hypercoagulability, endotheliopathy and inflammation following acute SARS-CoV-2 infection. *Expert Rev. Hematol.* **16**, 1035–1048 (2023).
65. Iosef, C. et al. Plasma proteome of long-COVID patients indicates HIF-mediated vasculo-proliferative disease with impact on brain and heart function. *J. Transl. Med.* **21**, 377 (2023).
66. Buggert, M., Price, D. A., Mackay, L. K. & Betts, M. R. Human circulating and tissue-resident memory CD8<sup>+</sup> T cells. *Nat. Immunol.* **24**, 1076–1086 (2023).

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

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

© The Author(s) 2025

<sup>1</sup>Center for Infectious Medicine, Department of Medicine Huddinge, Karolinska Institutet, Stockholm, Sweden. <sup>2</sup>Laboratory of Translational Immuno-Oncology, Department of Biomedicine, University and University Hospital Basel, Basel, Switzerland. <sup>3</sup>Division of Infectious Diseases and Dermatology, Department of Medicine Huddinge, Karolinska Institutet, Stockholm, Sweden. <sup>4</sup>Center for Clinical Research Sörmland, Uppsala University, Uppsala, Sweden. <sup>5</sup>Department of Infectious Diseases, Karolinska University Hospital, Stockholm, Sweden. <sup>6</sup>Division of Infection and Immunity, Cardiff University School of Medicine, University Hospital of Wales, Cardiff, UK. <sup>7</sup>Department of Respiratory Medicine, University Hospital Llandough, Penarth, UK. <sup>8</sup>Post-COVID Polyclinic, Karolinska University Hospital, Stockholm, Sweden. <sup>9</sup>Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden. <sup>10</sup>Division of Infectious Diseases, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. <sup>11</sup>Department of Medicine, University of California, San Francisco, San Francisco, CA, USA. <sup>12</sup>Systems Immunity Research Institute, Cardiff University School of Medicine, University Hospital of Wales, Cardiff, UK. <sup>13</sup>These authors contributed equally: Yu Gao, Curtis Cai, Sarah Adamo, Helen E. Davies, Soo Aleman, Marcus Buggert, David A. Price. ✉ e-mail: [marcus.buggert@ki.se](mailto:marcus.buggert@ki.se); [priced6@cardiff.ac.uk](mailto:priced6@cardiff.ac.uk)

## Methods

### Study design

The objective of this study was to characterize the immunological and proteomic features of long COVID. SARS-CoV-2 spike-specific antibody titers were measured using an enzyme-linked immunosorbent assay. Antibody neutralization activity and ADNKA were quantified against the England-2 strain of SARS-CoV-2. Immune cell lineages were profiled via multidimensional flow cytometry. Antigen-specific CD4<sup>+</sup> and CD8<sup>+</sup> T cells were enumerated functionally using a flow cytometric AIM assay. Antigen-specific CD8<sup>+</sup> T cells were further identified physically using peptide–HLA class I tetramers to enable detailed phenotypic analyses via multidimensional flow cytometry. Plasma proteomes were analyzed using a targeted affinity platform. Clinical symptoms were integrated with the frequencies and phenotypic attributes of immune cells to delineate plasma biomarkers and signaling pathways associated with long COVID.

### Donors

The primary cohort included healthy convalescent individuals (controls;  $n = 70$ ) and individuals with long COVID (cases;  $n = 70$ ) recruited from University Hospital Llandough (Table 1 and Supplementary Table 1). All participants had a clearly defined episode of symptomatically mild acute COVID-19 confirmed via direct molecular evidence of infection with SARS-CoV-2. None required hospitalization. Cases were diagnosed according to the National Institute for Health and Care Excellence guideline NG188 (<https://www.nice.org.uk/guidance/ng188>). Groups were matched as closely as possible for age, BMI, race, sex, time since infection, and vaccination against SARS-CoV-2 (Fig. 1a,b and Table 1). Eligible individuals were men and nonpregnant women over the age of 18 years with no alternative explanatory disease and symptoms that persisted for at least 12 weeks after the initial diagnosis of acute COVID-19. One persistent symptom was sufficient for the diagnosis of long COVID. All individuals underwent a comprehensive medical evaluation, including chest radiography, electrocardiography, lung function tests (spirometry with gas transfer as indicated and measurement of exhaled nitric oxide), and standard blood tests (autoantibody screens; bone, liver and kidney function; coagulation screens; full blood count; markers of nutrition). Symptoms were scored individually using a numeric self-rating scale from 0 (no symptom) to 10 (worst possible symptom). Overall general health was scored similarly on an inverse scale from 0 (worst possible) to 10 (best possible). The secondary cohort included healthy convalescent individuals (controls;  $n = 30$ ) and individuals with long COVID (cases;  $n = 95$ ) recruited from the Karolinska University Hospital (Table 2). All participants in the primary cohort were recruited between March and August 2022, and all participants in the secondary cohort were recruited between June and October 2022. PBMCs from donors with untreated chronic HIV-1 infection ( $n = 14$ ) were obtained from the University of Alabama at Birmingham or the University of California, San Francisco.

### Samples

PBMCs were isolated via standard density gradient centrifugation and cryopreserved in fetal bovine serum (Thermo Fisher Scientific) containing 10% dimethyl sulfoxide (DMSO; Sigma-Aldrich). EDTA plasma samples were stored at  $-80^{\circ}\text{C}$ .

### Ethics

All participants provided written informed consent in accordance with the principles of the Declaration of Helsinki (2013). The primary study was approved by the Cardiff University School of Medicine Research Ethics Committee (21/55) and the Health Research Authority and Health and Care Research Wales (20/NW/0240), and the secondary study was approved by the Swedish Ethical Review Authority (2022-00100-01).

### Cells and viruses

A549 and VeroE6 cells expressing human angiotensin-converting enzyme 2 (ACE2) and transmembrane serine protease 2 (TMPRSS2) were used to support viral entry and propagation<sup>67</sup>. Antibody functionality assays were performed using the England-2 strain of SARS-CoV-2 (ref. 38).

### Peptides

SARS-CoV-2 peptides were manufactured as 15-mers overlapping by 11 amino acids spanning the spike protein (Peptides & Elephants) or as 20-mers overlapping by 10 amino acids spanning the nucleocapsid, combined membrane and envelope, ORF1a, ORF1b and ORF3–ORF10 proteins (Sigma-Aldrich). EBV peptides were manufactured as 15-mers overlapping by 11 amino acids spanning the BRLF1, BZLF1, BMLF1 and BARF1 proteins (lytic pool) and the EBNA1, EBNA2, EBNA3A, EBNA3B, EBNA3C and LMP2 proteins (latent pool; JPT Peptide Technologies). CMV peptides were manufactured as 15-mers overlapping by 11 amino acids spanning the combined IE-1, IE-2 and pp65 proteins (JPT Peptide Technologies). Lyophilized peptides were reconstituted at a stock concentration of  $10\text{ mg ml}^{-1}$  in DMSO and further diluted to  $100\text{ }\mu\text{g ml}^{-1}$  in phosphate-buffered saline (PBS).

### Tetramers

Peptide–HLA class I complexes were generated and tetramerized with fluorescent tags as described previously<sup>68,69</sup>. The following specificities were used in this study: CMV pp65 HLA-A\*02:01 NLVPMVATV (BV421), CMV pp65 HLA-B\*07:02 TPRVTGGGAM (PE), EBV BMLF1 (lytic) HLA-A\*02:01 GLCTLVAML (PE), EBV EBNA3A (latent) HLA-B\*07:02 RPPIFIRRL (BV421), HIV-1 p2p7p1p6 Gag HLA-A\*02:01 FLGKIWPESHK (PE), HIV-1 p17 Gag HLA-A\*02:01 SLYNTVATL (BV421), HIV-1 Pol HLA-A\*02:01 ILKEPVHGV (PE), HIV-1 p17 Gag HLA-A\*24:02 KYKLHIVW (BV421), HIV-1 Nef HLA-A\*24:02 RYPLTFGW (PE), HIV-1 p24 Gag HLA-B\*07:02 GPGHKARVL (BV421), HIV-1 p24 Gag HLA-B\*08:01 EIYKRWII (PE), HIV-1 p24 Gag HLA-B\*57:01 KAFSPEVIPMF (PE), HIV-1 p24 Gag HLA-B\*57:01 QASQEVKNW (BV421), IAV matrix protein M1 HLA-A\*02:01 GILGFVFTL (BV421), IAV nucleoprotein HLA-B\*07:02 LPFDKTTVM (BV421), SARS-CoV-2 spike HLA-A\*02:01 YLQPRTFLL (BV421), SARS-CoV-2 nucleocapsid HLA-A\*02:01 LLLDRLNQL (PE), SARS-CoV-2 ORF3 HLA-A\*02:01 ALSKGVHFV (PE), SARS-CoV-2 ORF3 HLA-A\*02:01 LLYDANYFL (PE) and SARS-CoV-2 nucleocapsid HLA-B\*07:02 SPRWYFYLL (PE).

### Antibody quantification

SARS-CoV-2 spike-specific antibody titers were measured using a SARS-CoV-2 Spike (Trimer) Ig Total ELISA Kit (Thermo Fisher Scientific). Samples were assayed in duplicate and calibrated against a standard curve. Data were analyzed using Prism version 9.5.0 (GraphPad).

### Neutralization assay

Antibody neutralization activity was quantified as described previously<sup>38</sup>. Briefly, serial dilutions of plasma were mixed in duplicate with 600 plaque-forming units of England-2, incubated for 1 h at  $37^{\circ}\text{C}$ , and added to VeroE6 cells expressing ACE2 and TMPRSS2. After 48 h, cell monolayers were fixed in 4% paraformaldehyde (Thermo Fisher Scientific), permeabilized with 0.5% NP-40 (Merck), and blocked with PBS containing 0.1% Tween-20 (PBST) and 3% nonfat milk for 1 h at room temperature (RT). The primary antibody (anti-SARS-CoV-2 nucleocapsid protein, clone 1C7, Stratech Scientific) was diluted 1:500 in PBST containing 1% nonfat milk and added to the cell monolayers for 1 h at RT. Cells were then washed with PBST. The secondary antibody (anti-mouse IgG-HRP, polyclonal, Jackson ImmunoResearch) was diluted 1:3,000 in PBST containing 1% nonfat milk and added to the cell monolayers for 1 h at RT. Cells were then washed again with PBST. Assays were developed using SIGMAFAST OPD (Sigma-Aldrich) and analyzed at an optical density of 450 nm using a CLARIOstar Plus Microplate Reader

(BMG Labtech). Control wells contained no sample, a standardized sample with moderate neutralization activity, or no SARS-CoV-2. The neutralization titer for each sample was calculated as the highest plasma dilution that achieved a 50% reduction in plaque formation ( $NT_{50}$ ).

### ADNKA

ADNKA was quantified as described previously<sup>38,70</sup>. Briefly, target A549 cells expressing ACE2 and TMPRSS2 were infected overnight with England-2 (multiplicity of infection = 5), collected using TrypLE Express Enzyme (Thermo Fisher Scientific), mixed with healthy donor PBMCs at a ratio of 1:10, and incubated with serial dilutions of plasma in the presence of anti-CD107a-FITC (clone H4A3, BioLegend) and GolgiStop (0.7  $\mu$ l ml<sup>-1</sup>; BD Biosciences) for 5 h at 37 °C. Cells were then washed with cold PBS, stained with anti-CD3-PE-Cy7 (clone UCHT1, BioLegend), anti-CD56-BV605 (clone 5.H11, BioLegend), anti-CD57-APC (clone HNK-1, BioLegend) and LIVE/DEAD Fixable Aqua (Thermo Fisher Scientific) for 30 min at 4 °C, washed again with cold PBS, and fixed in 4% paraformaldehyde (Thermo Fisher Scientific). Control wells contained a seronegative sample, uninfected target cells, or a standardized sample that elicited moderate ADNKA. Data were acquired using an Attune NxT Flow Cytometer (Thermo Fisher Scientific). Activation was quantified as a function of degranulation (CD107a<sup>+</sup>) among viable NK cells (Aqua<sup>-</sup>CD3<sup>+</sup>CD56<sup>+</sup>) with potent cytotoxic activity (CD57<sup>+</sup>) using FlowJo version 10.9.0 (FlowJo) and normalized to the standardized sample via area under the curve analyses in Prism version 9.5.0 (GraphPad).

### Immune cell lineage analysis

PBMCs were thawed quickly, resuspended in RPMI 1640 Complete Medium (Sigma-Aldrich) supplemented with DNase I (10 U ml<sup>-1</sup>; Sigma-Aldrich), and seeded at  $1 \times 10^6$  cells per well in 96-well U-bottom plates (Corning). Cells were incubated first with Human TruStain FcX (BioLegend) for 10 min at RT and then with LIVE/DEAD Fixable Aqua (Thermo Fisher Scientific) for 10 min at RT. Anti-CCR7-APC-Cy7 (clone G043H7, BioLegend) and anti-CX3CR1-PE (clone 2A9-1, BioLegend) were added for 15 min at 37 °C. Cells were then stained with anti-CD3-BV650 (clone OKT3, BioLegend), anti-CD4-PE-Cy5.5 (clone S3.5, Thermo Fisher Scientific), anti-CD8-BUV396 (clone RPA-T8, BD Biosciences), anti-CD11c-BB515 (clone B-ly6, BD Biosciences), anti-CD14-PE-Cy5 (clone 61D3, Thermo Fisher Scientific), anti-CD16-BUV496 (clone 3G8, BD Biosciences), anti-CD19-BUV563 (clone HIB19, BD Biosciences), anti-CD27-BV786 (clone O323, BioLegend), anti-CD34-BB660 (clone 581, BD Biosciences), anti-CD38-APC (clone HB7, BD Biosciences), anti-CD45-BUV805 (clone HI30, BD Biosciences), anti-CD45RA-BV570 (clone HI100, BioLegend), anti-CD56-BUV615 (clone NCAM16.2, BD Biosciences), anti-CD69-BUV737 (clone FN50, BD Biosciences), anti-CD71-BUV661 (clone M-A712, BD Biosciences), anti-CD83-BB790 (clone HB15e, BD Biosciences), anti-CD86-BB630 (clone 2331 (FUN-1), BD Biosciences), anti-CD123-PE-Cy7 (clone 7G3, BD Biosciences), anti-CD127-BV421 (clone A019D5, BioLegend), anti-HLA-DR-BV605 (clone G46-6, BD Biosciences) and anti-PD-1-R718 (clone EH12.1, BD Biosciences) for 30 min at RT (Supplementary Table 8). Stained cells were washed twice with FACS buffer (PBS containing 2% fetal bovine serum and 2 mM EDTA), fixed in Cytofix Fixation Buffer (BD Biosciences), and acquired using a FACSymphony A3 (BD Biosciences). Data were analyzed using FlowJo version 10.9.0 (FlowJo).

### AIM assay

PBMCs were thawed quickly, resuspended in RPMI 1640 Complete Medium (Sigma-Aldrich) supplemented with DNase I (10 U ml<sup>-1</sup>; Sigma-Aldrich), and rested at  $1 \times 10^6$  cells per well in 96-well U-bottom plates (Corning) for 3 h at 37 °C. The medium was then supplemented with unconjugated anti-CD40 (clone HB14, Miltenyi Biotec) and anti-CXCR5-BB515 (clone RF8B2, BD Biosciences), followed 15 min later by the relevant peptides (each at 0.5  $\mu$ g ml<sup>-1</sup>), and the cultures were

incubated for 12 h at 37 °C. Negative-control wells contained equivalent DMSO. After incubation, cells were washed with PBS, labeled with LIVE/DEAD Fixable Aqua (Thermo Fisher Scientific) for 10 min at RT, washed with FACS buffer, and stained with anti-CCR4-BB700 (clone 1G1, BD Biosciences), anti-CCR6-BUV737 (clone 11A9, BD Biosciences), anti-CCR7-APC-Cy7 (clone G043H7, BioLegend), anti-CX3CR1-PE (clone 2A9-1, BioLegend) and anti-CXCR3-AF647 (clone G025H7, BioLegend) for 10 min at 37 °C. Cells were then stained further with anti-CD3-BUV805 (clone UCHT1, BD Biosciences), anti-CD4-BUV496 (clone SK3, BD Biosciences), anti-CD8-BUV395 (clone RPA-T8, BD Biosciences), anti-CD14-BV510 (clone M5E2, BioLegend), anti-CD19-BV510 (clone HIB19, BioLegend), anti-CD28-BUV563 (clone CD28.2, BD Biosciences), anti-CD38-APC-R700 (clone HIT2, BD Biosciences), anti-CD39-BV711 (clone A1, BioLegend), anti-CD45RA-BV570 (clone HI100, BioLegend), anti-CD69-BV650 (clone FN50, BioLegend), anti-CD71-BUV661 (clone M-A712, BD Biosciences), anti-CD95-PE-Dazzle594 (clone DX2, BioLegend), anti-CD127-PE-Cy5 (clone A019D5, BioLegend), anti-CD137-PE-Cy7 (clone 4B4-1, BioLegend), anti-CD154-BV421 (clone 24-31, BioLegend), anti-HLA-DR-BV605 (clone G46-6, BD Biosciences), anti-PD-1-BUV615 (clone EH12.1, BD Biosciences) and anti-TIGIT-BV786 (clone 741182, BD Biosciences) for 30 min at RT in the presence of Brilliant Stain Buffer Plus (BD Biosciences; Supplementary Table 9). Stained cells were washed twice with FACS buffer, fixed in Cytofix Fixation Buffer (BD Biosciences), and acquired using a FACSymphony A5 (BD Biosciences). Data were analyzed using FlowJo version 10.9.0 (FlowJo).

### Tetramer staining and phenotypic analysis

PBMCs were thawed quickly, resuspended in RPMI 1640 Complete Medium (Sigma-Aldrich) supplemented with DNase I (10 U ml<sup>-1</sup>; Sigma-Aldrich), and seeded at  $2 \times 10^6$  cells per well in 96-well U-bottom plates (Corning). Cells were incubated first with dasatinib (50  $\mu$ M; STEMCELL Technologies) for 10 min at RT and then with the relevant peptide-HLA class I tetramers (each at 1  $\mu$ g per stain) for 20 min at RT (Supplementary Table 10). After incubation, cells were washed with PBS, labeled with LIVE/DEAD Fixable Aqua (Thermo Fisher Scientific) for 10 min at RT, washed with FACS buffer, and stained with anti-CCR7-APC-Cy7 (clone G043H7, BioLegend), anti-CX3CR1-BUV615 (clone 2A9-1, BD Biosciences) and anti-CXCR3-PE-Cy5 (clone G025H7, BioLegend) for 10 min at 37 °C. Cells were then stained further with anti-CD3-BUV805 (clone UCHT1, BD Biosciences), anti-CD4-PE-Cy5.5 (clone RM4-5, Thermo Fisher Scientific), anti-CD8-BUV395 (clone RPA-T8, BioLegend), anti-CD14-BV510 (clone M5E2, BioLegend), anti-CD19-BV510 (clone HIB19, BioLegend), anti-CD27-BV786 (clone O323, BioLegend), anti-CD38-BUV496 (clone HIT2, BD Biosciences), anti-CD39-BV711 (clone A1, BioLegend), anti-CD45RA-BV570 (clone HI100, BioLegend), anti-CD95-BB700 (clone DX2, BD Biosciences), anti-CD127-BB630 (clone HIL-7R-M21, BD Biosciences), anti-HLA-DR-BV650 (clone G46-6, BD Biosciences), anti-LAG-3-BUV661 (clone 3DS223H, Thermo Fisher Scientific), anti-PD-1-BUV737 (clone EH12.1, BD Biosciences), anti-TIGIT-PE-Dazzle594 (clone A15153G, BioLegend) and anti-TIM-3-BV605 (clone F38-2E2, BioLegend) for 20 min at RT, washed twice with FACS buffer, fixed/permeabilized using a FoxP3 Transcription Factor Staining Buffer Set (Thermo Fisher Scientific), and stained intracellularly with anti-EOMES-EF660 (clone WD1928, eBioscience), anti-granzyme B-BB790 (clone GB11, BD Biosciences), anti-Ki67-AF700 (clone B56, BD Biosciences), anti-T-BET-PE-Cy7 (clone 4B10, eBioscience) and anti-TCF-1-AF488 (clone C63D9, Cell Signaling Technology) for 30 min at RT (Supplementary Table 11). Stained cells were washed twice with FACS buffer and acquired using a FACSymphony A3 (BD Biosciences). Data were analyzed using FlowJo software version 10.9.0 (FlowJo).

### Plasma proteomics

A data-driven approach was used to select healthy convalescent individuals ( $n = 51$ ) and individuals with long COVID ( $n = 51$ ) for

plasma proteome characterization via a Proximity Extension Assay (Olink Proteomics). Immune cell subset proportions were summarized using a PCA. Outlier samples were excluded based on the greatest deviation from the origin along PC1 to PC4. Plasma samples were analyzed in two batches using Explore 3072 (Olink Proteomics). Sixteen bridge samples were included for quality control purposes in each batch.

### General statistics

Differences between groups were assessed using a two-tailed Mann–Whitney *U*-test. Raw *P* values are shown. Correlations were evaluated using the two-tailed Pearson coefficient or a two-tailed Spearman rank test. Significance was assigned at  $P < 0.05$ . Basic statistical analyses were performed using Prism version 9.5.0 (GraphPad).

### Flow cytometry data analysis

Samples acquired for immune cell lineage analysis were gated to the single-cell/viable/CD45<sup>+</sup> population and subsequently exported to contain only 3,000 events using the FlowJo Plugin DownSample version 3. Exported fcs files were loaded into R using flowCore version 2.6.0. All data were concatenated into a single matrix with compensated markers (excluding viability, CD34 and CD45). Data for each marker were scaled and centered for analysis using umap version 0.2.10.0. Clustering was performed using a Gaussian mixture model (maxNumComponents = 10) implemented in mclust version 6.0.0. Data were visualized using ggplot2 version 3.4.2. Antigen-specific CD4<sup>+</sup> and CD8<sup>+</sup> T cell frequencies assessed via the AIM assay were calculated after background subtraction. Samples acquired for detailed phenotypic characterization were excluded below a threshold of five tetramer<sup>+</sup> CD8<sup>+</sup> T cells per specificity. The expression of each marker was then normalized to the average geometric mean fluorescence intensity across all samples and specificities and used to calculate the co-inhibitory score, representing the summed data for PD-1, TIM-3, LAG-3 and TIGIT. Statistical analyses were performed using R version 4.2.1.

### Plasma proteome data analysis

Bridge sample data were normalized using the olink\_normalization function implemented in OlinkAnalyze version 3.4.1. Differential expression analyses were performed using a Wilcoxon rank-sum/Mann–Whitney *U*-test with Benjamini–Hochberg correction implemented via the olink\_wilcox function in OlinkAnalyze version 3.4.1. GSEA was performed using fgsea version 1.20.0 incorporating lists of all analyzed proteins ordered by correlation coefficient or fold change. Gene sets were downloaded from the MSigDB using msigdb version 7.5.1. Overrepresentation analysis was performed using the fora function implemented in fgsea version 1.20.0 incorporating all measured proteins as the ‘universe’. At least five proteins were required in each gene set for consideration. Significance was evaluated using a hypergeometric test. Correlations were calculated using the cor.test function implemented in stats version 4.1.3. PCAs were performed using the prcomp function implemented in stats version 4.1.3. Data were visualized using ggplot2 version 3.4.2 and pheatmap version 1.0.12. All analyses were performed using R version 4.2.1. Network analyses of plasma proteins that were differentially expressed as a function of symptom severity were performed using the stringApp in Cytoscape version 3.10.3 (ref. 71).

### Reporting summary

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

### Data availability

Raw proteomics data are available via Zenodo (<https://doi.org/10.5281/zenodo.14772494>)<sup>72</sup>. Any additional information required to reanalyze the data reported in this paper is available from the corresponding authors upon reasonable request.

## References

- Rihn, S. J. et al. A plasmid DNA-launched SARS-CoV-2 reverse genetics system and coronavirus toolkit for COVID-19 research. *PLoS Biol.* **19**, e3001091 (2021).
- Melenhorst, J. J. et al. Detection of low avidity CD8<sup>+</sup> T cell populations with coreceptor-enhanced peptide-major histocompatibility complex class I tetramers. *J. Immunol. Methods* **338**, 31–39 (2008).
- Price, D. A. et al. Avidity for antigen shapes clonal dominance in CD8<sup>+</sup> T cell populations specific for persistent DNA viruses. *J. Exp. Med.* **202**, 1349–1361 (2005).
- Vlahava, V. M. et al. Monoclonal antibodies targeting nonstructural viral antigens can activate ADCC against human cytomegalovirus. *J. Clin. Invest.* **131**, e139296 (2021).
- Doncheva, N. T., Morris, J. H., Gorodkin, J. & Jensen, L. J. Cytoscape StringApp: network analysis and visualization of proteomics data. *J. Proteome Res.* **18**, 623–632 (2019).
- Cai, C., Gao, Y. & Buggert, M. Olink data for ‘Identification of soluble biomarkers that associate with distinct manifestations of long COVID’. Zenodo <https://doi.org/10.5281/zenodo.14772494> (2025).

## Acknowledgements

We express our gratitude to all donors, health care personnel, study coordinators, administrators and laboratory managers involved in this work. The findings reported here represent independent research funded in part by the National Institute for Health and Care Research in response to the emergence of long COVID (COV-LT2-0041). The views expressed in this publication are those of the authors and not necessarily those of the National Institute for Health and Care Research or The Department of Health and Social Care (United Kingdom). Additional support was provided by the SciLifeLab/KAW National COVID-19 Research Program, the Swedish Research Council (2021-06534) and the PolyBio Research Foundation (Balvi B43). Y.G. was supported by the Åke Wibergs Stiftelse (M22-0099) and the Magnus Bergvalls Stiftelse (2022-307). C.C. was supported by the Swedish Society for Medical Research (PG-22-0432-H-01). S. Adamo was supported by a Swiss National Science Foundation Postdoc Mobility Grant (P500PB\_211069). S. Aleman was supported by the Swedish Research Council (2021-06534). M.B. was supported by the Swedish Research Council (2018-02330, 2020-06121, 2021-01141, 2021-04779 and 2022-01313), the Knut and Alice Wallenberg Foundation (2021.0136), the European Research Council (101041484), the Swedish Society for Medical Research (CG-22 0009), the Swedish Cancer Society (22 2237 Pj), the Karolinska Institutet (2019-00969, 2021-00513 and 2022-01719), the Åke Wibergs Stiftelse (M20-0190) and the Center for Innovative Medicine (FoUI-988204).

## Author contributions

Y.G., C.C., S. Adamo, H.E.D., S. Aleman, M.B. and D.A.P. conceptualized the study. Y.G., C.C., S. Adamo, S.L.-L., P.S.A., K.B. and J.W. performed experiments. Y.G., C.C. and S. Adamo analyzed and visualized data. Y.G., E.B., H.K., L.D., K.L.M., S.L.-L., K.L., S.K., M.A., S.A.J., P.J., C.L., P.A.G., M.J.P., S.G.D., H.E.D., S. Aleman, M.B. and D.A.P. provided resources and/or collected samples. H.E.D., S. Aleman, M.B. and D.A.P. acquired funding for the project. R.J.S., H.E.D., S. Aleman, M.B. and D.A.P. supervised the work. Y.G., C.C., S. Adamo, S. Aleman, M.B. and D.A.P. wrote the paper. M.B. and D.A.P. are joint senior authors and codirected the study. All authors approved the final draft of the paper and concurred with the decision to submit for publication.

## Funding

Open access funding provided by Karolinska Institute.

## Competing interests

S. Aleman has received honoraria for educational events and lectures unrelated to this work from Gilead, AbbVie, Biogen and MSD and

reports grants from Gilead and AbbVie. M.B. is a consultant for Bristol Myers Squibb, Mabtech, Pfizer, Oxford Immunotec and MSD. The other authors declare no competing interests.

### Additional information

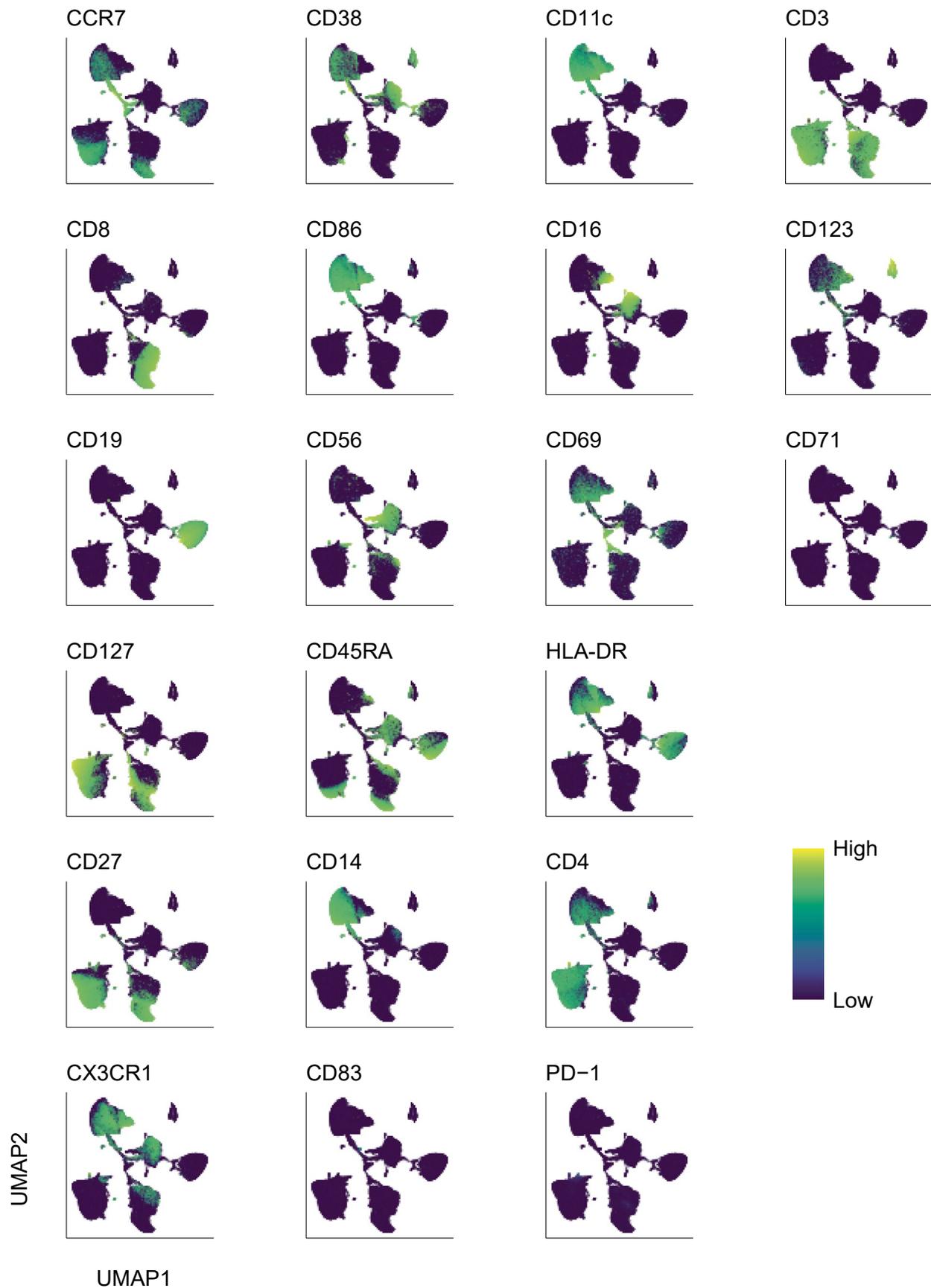
**Extended data** is available for this paper at <https://doi.org/10.1038/s41590-025-02135-5>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41590-025-02135-5>.

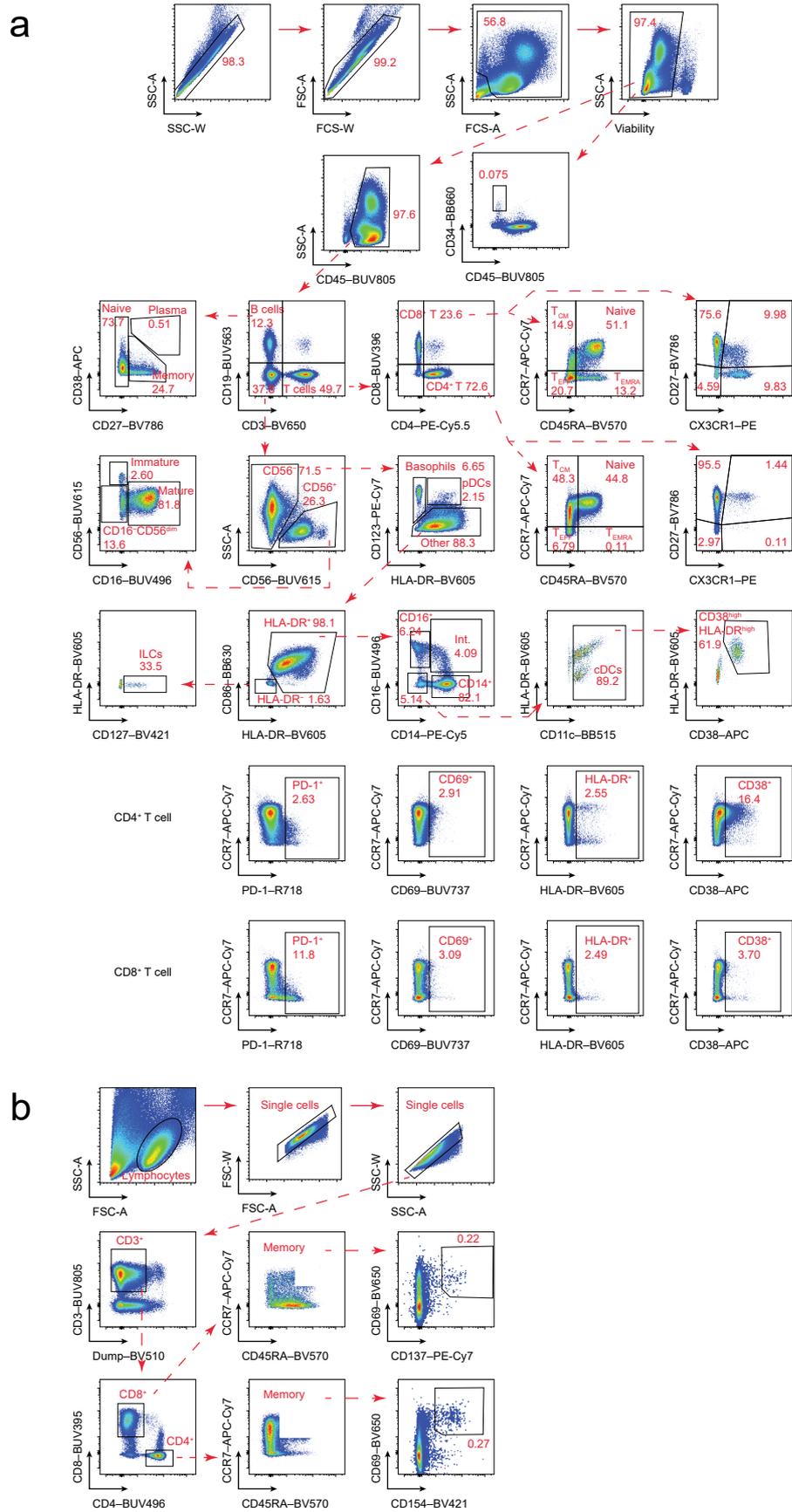
**Correspondence and requests for materials** should be addressed to Marcus Buggert or David A. Price.

**Peer review information** *Nature Immunology* thanks Randy Cron and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: P. Jauregui, in collaboration with the *Nature Immunology* team.

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



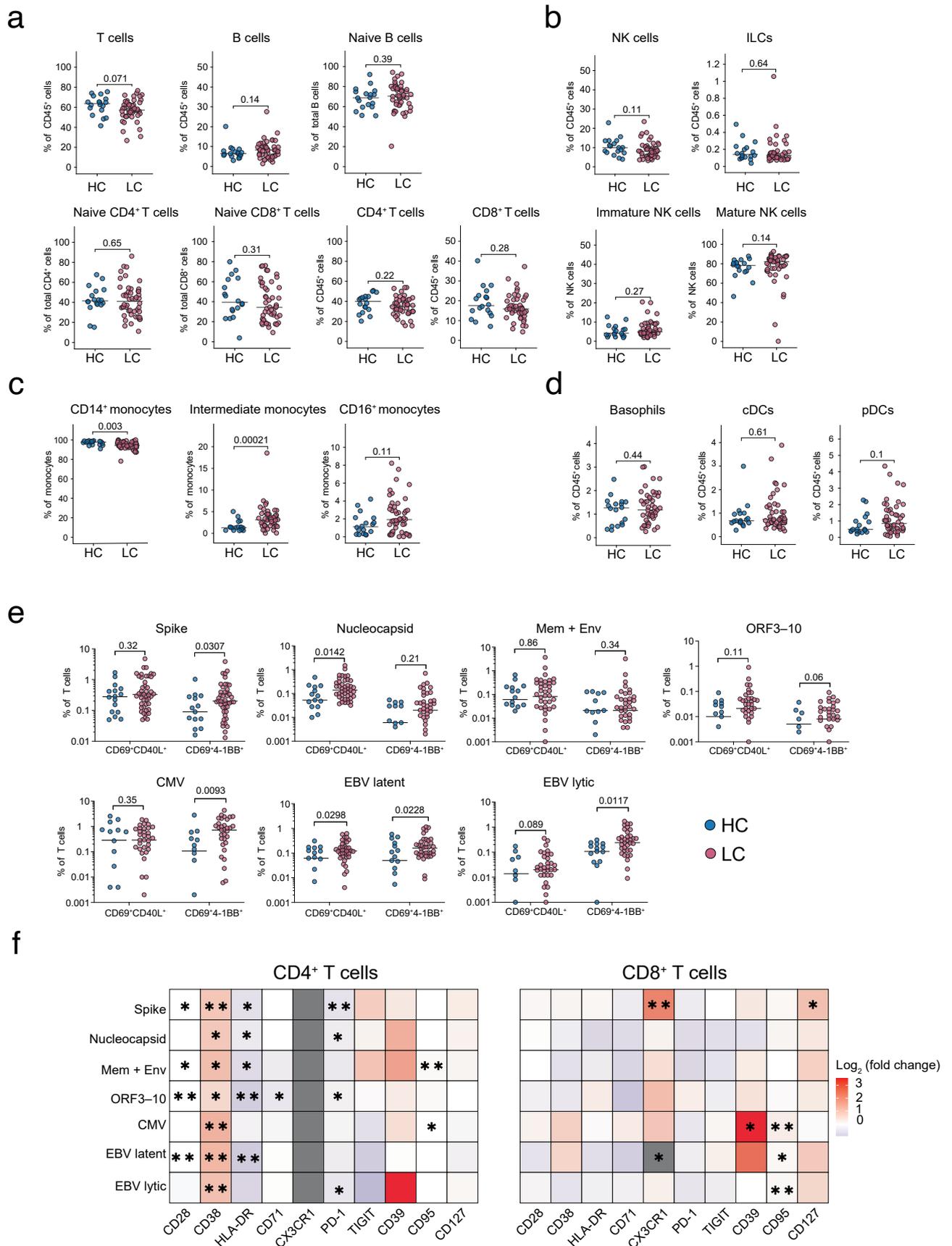
**Extended Data Fig. 1 | Expression of immune cell lineage markers measured via flow cytometry.** UMAP representation of individual immune cell lineage markers among peripheral blood mononuclear cells after dimensionality reduction.



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

**Extended Data Fig. 2 | Flow cytometric gating strategies for the identification of immune cell lineages and activation-induced marker upregulation.** (a) Flow cytometric gating strategy for immune cell lineage characterization. Numbers indicate percentages in the drawn gates. (b) Flow cytometric gating strategy

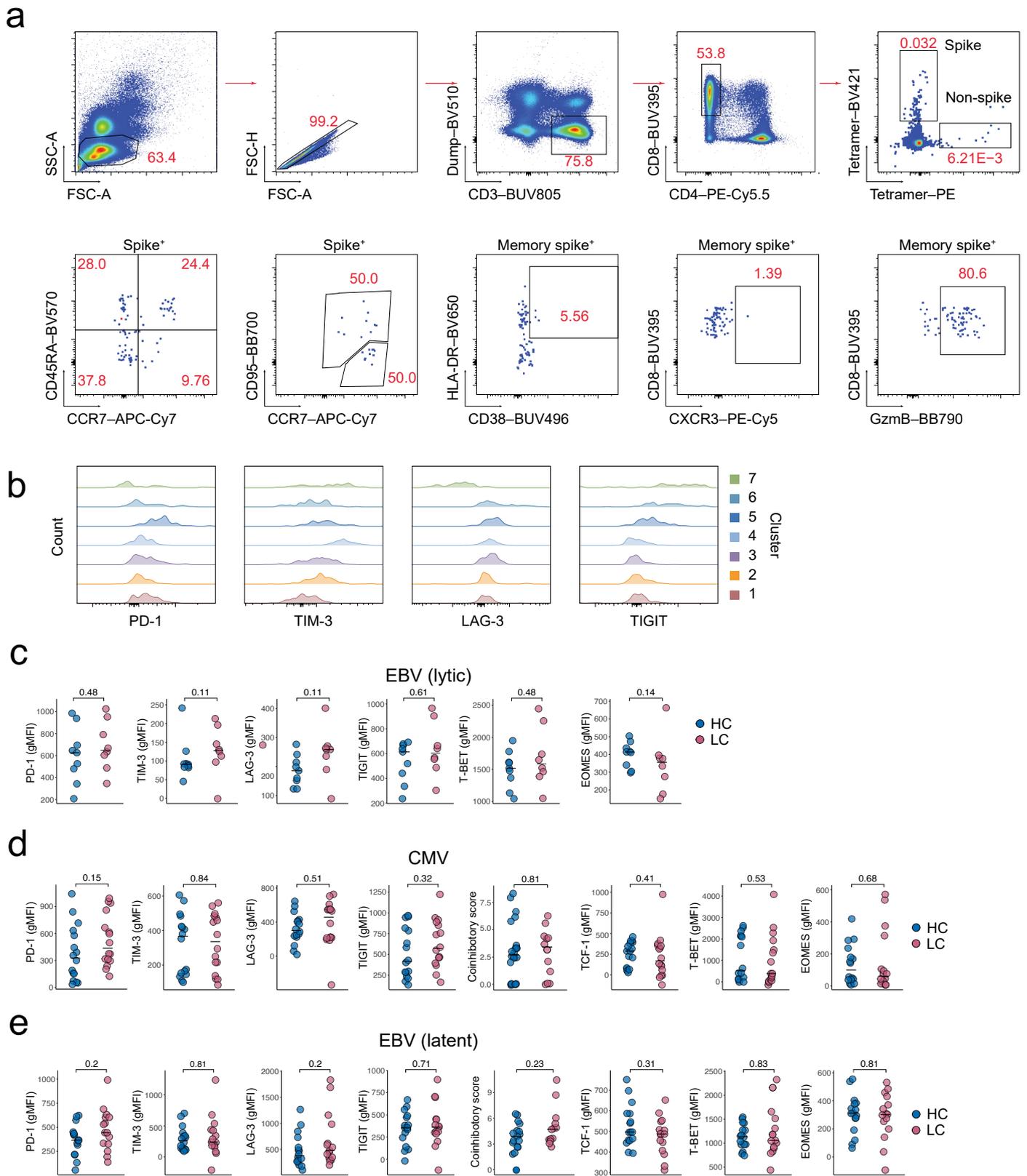
for the identification of antigen-specific CD4<sup>+</sup> and CD8<sup>+</sup> T cells via upregulation of the activation-induced markers CD69 and CD154 or CD69 and CD137, respectively. Numbers indicate percentages in the drawn gates.



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

**Extended Data Fig. 3 | Immune cell lineages and T cell immunity in healthy convalescent individuals and patients with long COVID recruited from Sweden.** (a) Scatter dot plots showing the frequencies of naive and total B and T cells gated manually. (b) Scatter dot plots showing the frequencies of innate lymphocytes gated manually. (c) Scatter dot plots showing the frequencies of monocytes gated manually. (d) Scatter dot plots showing the frequencies of basophils and dendritic cells gated manually. (e) Scatter dot plots showing the frequencies of functional CD4<sup>+</sup> and CD8<sup>+</sup> T cells targeting defined proteins from

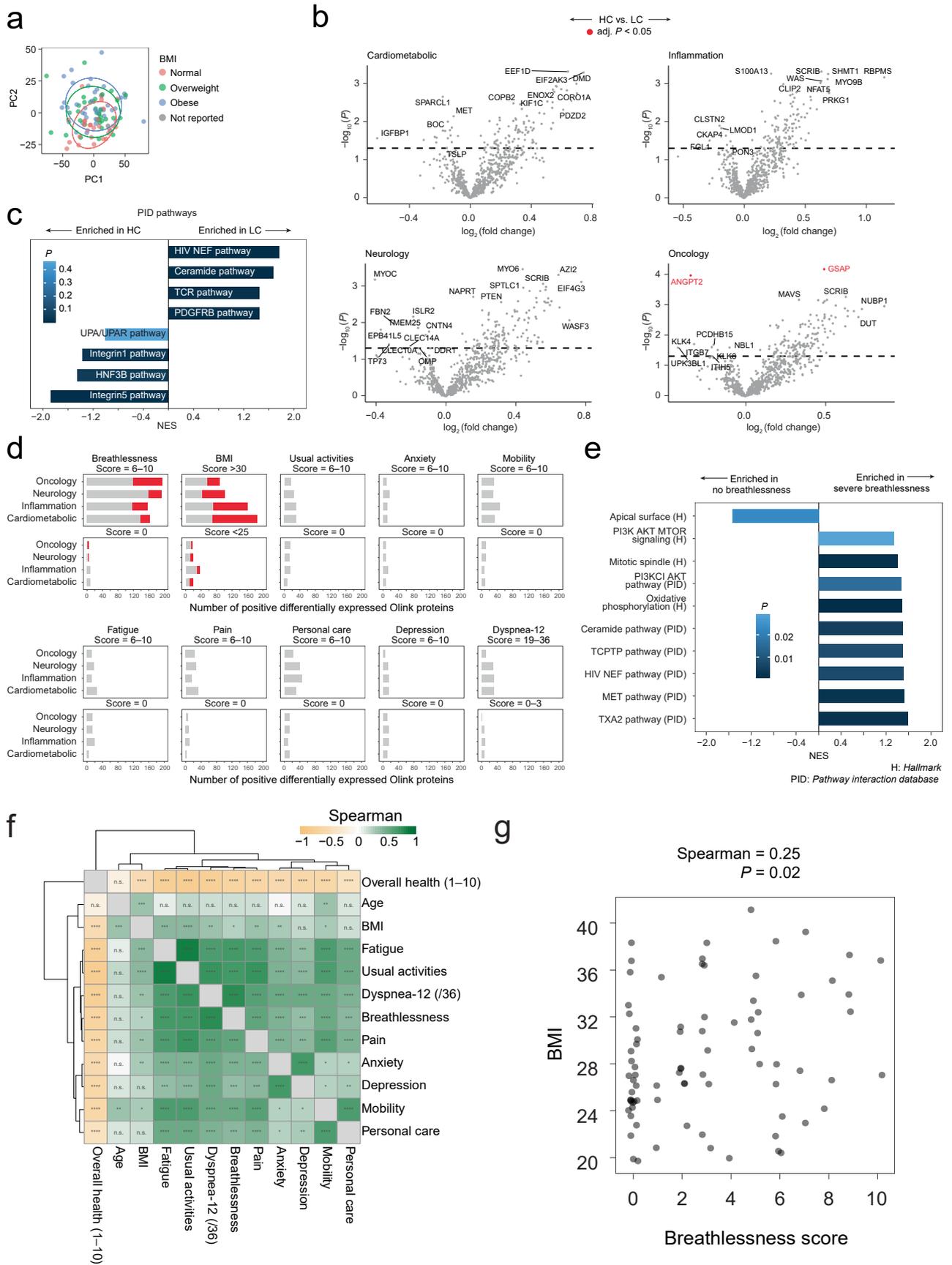
SARS-CoV-2, CMV, or EBV. (f) Heatmap summarizing the phenotypic attributes of functional CD4<sup>+</sup> and CD8<sup>+</sup> T cells targeting defined proteins from SARS-CoV-2, CMV, or EBV. Data are shown for each marker as the log<sub>2</sub>-transformed fold change in percent positive for each population among patients with long COVID (LC) versus healthy convalescent individuals (HC). \* $P < 0.05$ , \*\* $P < 0.01$ . HC,  $n \leq 20$ ; LC,  $n \leq 56$  (a, b, c, d, e and f). Horizontal bars represent median values (a, b, c, d and e). Significance was evaluated using a two-tailed Mann-Whitney U test (a, b, c, d, e and f).



Extended Data Fig. 4 | See next page for caption.

**Extended Data Fig. 4 | Additional phenotypic characteristics of virus-specific CD8<sup>+</sup> T cells in healthy convalescent individuals and patients with long COVID recruited from the UK.** (a) Flow cytometric gating strategy for the identification of tetramer<sup>+</sup> CD8<sup>+</sup> T cells directly *ex vivo*. Numbers indicate percentages in the drawn gates. (b) Flow cytometry histograms showing the expression patterns of coinhibitory receptors among clusters of tetramer<sup>+</sup> CD8<sup>+</sup> T cells targeting nonspike epitopes from SARS-CoV-2 identified using Phenograph. (c) Scatter dot plots showing the expression intensities of selected markers among tetramer<sup>+</sup>

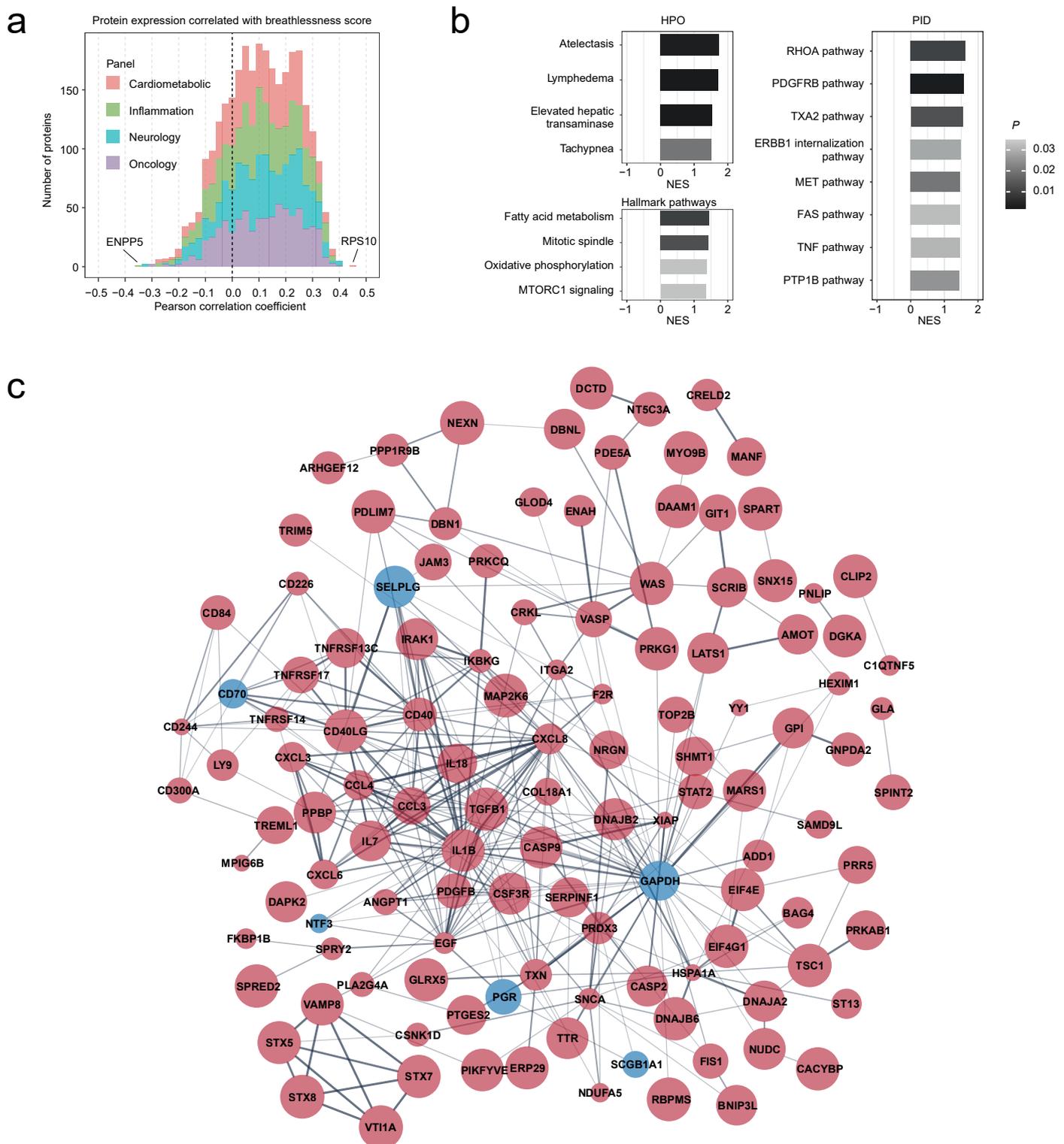
CD8<sup>+</sup> T cells targeting lytic epitopes from EBV. (d) Scatter dot plots showing the expression intensities of selected markers among tetramer<sup>+</sup> CD8<sup>+</sup> T cells targeting epitopes from CMV. (e) Scatter dot plots showing the expression intensities of selected markers among tetramer<sup>+</sup> CD8<sup>+</sup> T cells targeting latent epitopes from EBV. Healthy convalescent individuals (HC),  $n = 17$ ; patients with long COVID (LC),  $n = 15$  (b, c, d and e). Horizontal bars represent median values (c, d and e). Significance was evaluated using a two-tailed Mann-Whitney U test (c, d and e). gMFI, geometric mean fluorescence intensity.



Extended Data Fig. 5 | See next page for caption.

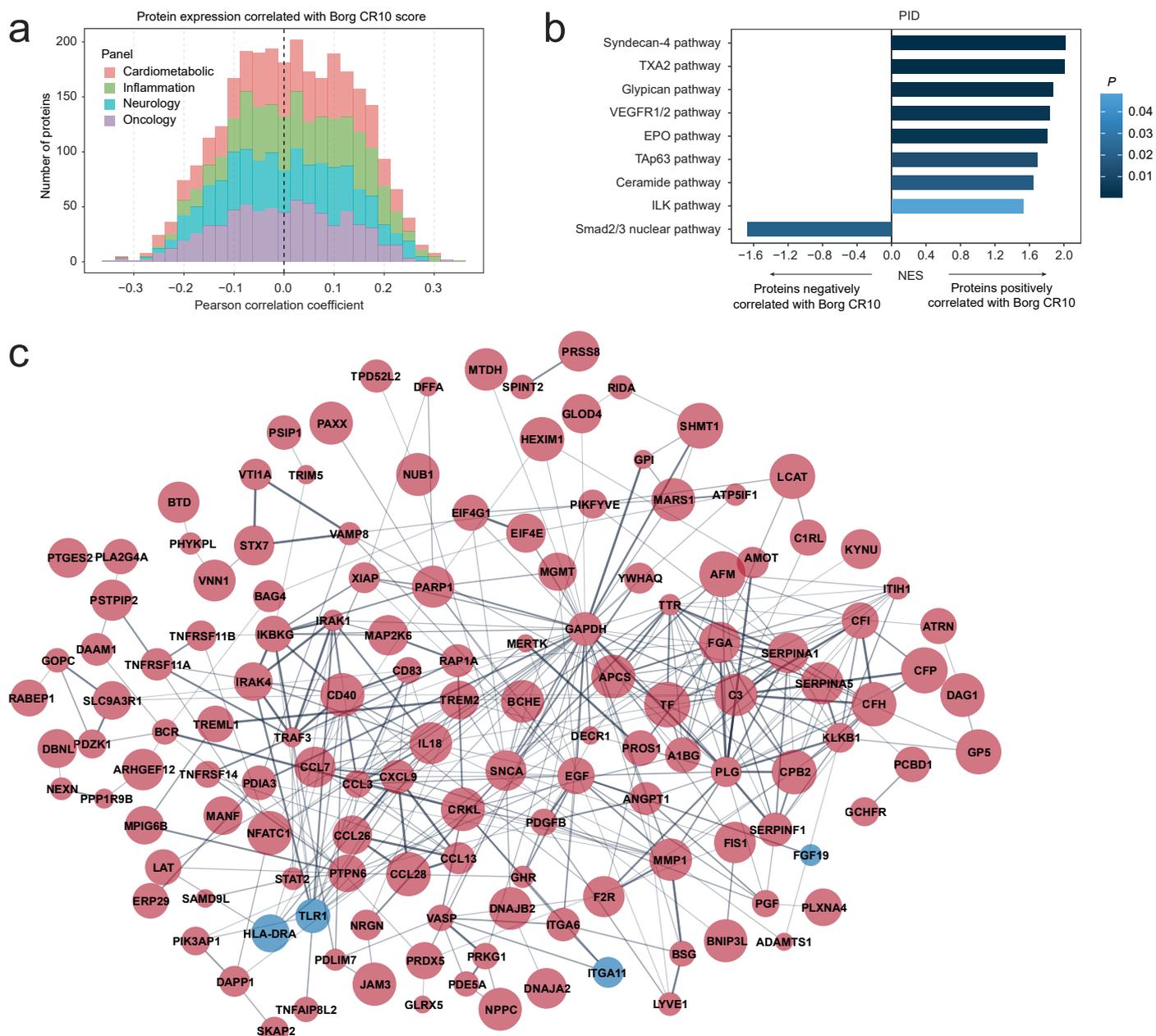
**Extended Data Fig. 5 | Dysregulation of the plasma proteome associated with clinical assignment and symptomatology in healthy convalescent individuals and patients with long COVID recruited from the UK.** (a) Principal component analysis of plasma protein concentrations colored by body mass index (BMI) for healthy convalescent individuals ( $n = 51$ ) and patients with long COVID ( $n = 51$ ). (b) Volcano plots showing the corresponding differentially expressed plasma proteins from each panel versus clinical assignment. Significance was evaluated using a two-tailed Mann–Whitney U test with (red) or without Benjamini–Hochberg correction (gray). The dashed line indicates  $P = 0.05$ . (c) Gene set enrichment analysis (GSEA) showing the corresponding differentially expressed plasma proteins by rank versus clinical assignment. (d) Bar plots showing the numbers of differentially upregulated plasma proteins from each panel versus the highest and lowest symptom score tiers for healthy convalescent

individuals ( $n = 34$ ) and patients with long COVID ( $n = 48$ ), irrespective of clinical assignment. Significance was evaluated using a two-tailed Mann–Whitney U test with (red) or without Benjamini–Hochberg correction (gray). (e) GSEA showing the corresponding differentially expressed plasma proteins by rank versus the highest and lowest breathlessness score tiers, irrespective of clinical assignment. (f) Heatmap showing correlations among clinical scores for healthy convalescent individuals ( $n = 51$ ) and patients with long COVID ( $n = 51$ ), irrespective of clinical assignment. (g) Correlation dot plot showing the corresponding breathlessness scores versus BMI, irrespective of clinical assignment. Significance was evaluated using the GSEA method without correction (c and e) or a two-tailed Spearman rank test (f and g). H, Hallmark; NES, normalized enrichment score; PID, Pathway Interaction Database.



**Extended Data Fig. 6 | Dysregulation of the plasma proteome associated with breathlessness in healthy convalescent individuals and patients with long COVID recruited from the UK.** (a) Stacked histogram showing the distribution of correlation coefficients from pairwise comparisons of plasma protein concentrations versus breathlessness scores for healthy convalescent individuals ( $n = 34$ ) and patients with long COVID ( $n = 48$ ), irrespective of clinical assignment. Significance was evaluated using the two-tailed Pearson coefficient. (b) Gene set enrichment analysis (GSEA) showing selected terms from Human Phenotype Ontology (HPO), the Pathway Interaction Database (PID), and the Hallmark Collection. Plasma protein concentrations were ranked

by correlation with breathlessness scores for healthy convalescent individuals ( $n = 34$ ) and patients with long COVID ( $n = 48$ ), irrespective of clinical assignment. Significance was evaluated using the GSEA method without correction. NES, normalized enrichment score. (c) Network analysis showing the corresponding differentially expressed plasma proteins from the inflammation panel depicted using Cytoscape. Each node represents a protein. Node color indicates protein concentration, and node size indicates significance. Red denotes overexpression in patients with long COVID, and blue denotes underexpression in patients with long COVID. Each edge represents the functional relevance between a pair of proteins, and line thickness represents the level of confidence.



**Extended Data Fig. 7 | Dysregulation of the plasma proteome associated with perceived exertion in patients with long COVID recruited from Sweden.**

(a) Stacked histogram showing the distribution of correlation coefficients from pairwise comparisons of plasma protein concentrations versus Borg CR10 score for patients with long COVID ( $n = 95$ ). Significance was evaluated using the two-tailed Pearson coefficient. (b) Gene set enrichment analysis (GSEA) showing plasma protein concentrations ranked by correlation with Borg CR10 scores for patients with long COVID ( $n = 95$ ). Significance was evaluated using the GSEA

method without correction. NES, normalized enrichment score. (c) Network analysis showing the corresponding differentially expressed plasma proteins from the inflammation panel depicted using Cytoscape. Each node represents a protein. Node color indicates protein concentration, and node size indicates significance. Red denotes overexpression in patients with a moderate Borg CR10 score, and blue denotes underexpression in patients with a moderate Borg CR10 score. Each edge represents the functional relevance between a pair of proteins, and line thickness represents the level of confidence.

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

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

### Software and code

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

Data collection	BD FACSDiva version 9.1 Olink Explore version 3.0.5
Data analysis	R version 4.2.1 FlowJo version 10.9.0 Prism version 10 CytoScape version 3.10.3

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](#)

Raw proteomics data are available via Zenodo ([doi.org/10.5281/zenodo.14772494](https://doi.org/10.5281/zenodo.14772494)). This paper does not report original code. Data from the Molecular Signatures Database (MSigDB) used version 7.5.1. Any additional information required to reanalyze the data reported in this paper is available from the corresponding authors upon request.

## 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	Biological sex is reported in Figure 1a, Table 1 and 2.
Reporting on race, ethnicity, or other socially relevant groupings	Ethnicity/race is reported in Table 1 and 2.
Population characteristics	Population characteristics are summarized Table 1 and 2.
Recruitment	Participants were recruited from two distinct sites as healthy convalescent individuals or patients with long COVID.
Ethics oversight	Cardiff University School of Medicine Research Ethics Committee Health Research Authority and Health and Care Research Wales Swedish Ethical Review Authority

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](https://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	No formal sample size calculation was performed. The sample size was determined based on feasibility, ethical considerations, and prior studies investigating similar research questions.
Data exclusions	No data were excluded from the flow cytometry analyses. A few samples were not selected for proteomic analyses due to cost limitations. Samples not selected for proteomic analyses were considered to be outliers based on circulating immune cell lineage composition (Figure 2).
Replication	Replication was not performed for individual samples due to limited material availability.
Randomization	Randomization was applied where relevant in this study. Specifically, for the OLINK Proteomics analysis, sample preparation followed the manufacturer's recommendations, which included a randomization step during sample loading onto the 96-well plate to minimize potential batch effects.
Blinding	Investigators were not blinded. All samples were treated identically. Each experimental protocol was performed identically for each sample.

## 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 &amp; experimental systems

n/a	Involvement	Included
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Plants

## Methods

n/a	Involvement	Included
<input checked="" type="checkbox"/>	<input type="checkbox"/>	ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/>	MRI-based neuroimaging

## Antibodies

## Antibodies used

CCR7 APC-Cy7 353212 G043H7 BioLegend  
 CD38 APC 567144 HB7 BD Biosciences  
 PD-1 R718 566974 EH12.1 BD Biosciences  
 CD11c BB515 564491 B-ly6 BD Biosciences  
 CD86 BB630 Custom 2331(FUN-1) BD Biosciences  
 CD34 BB660 Custom 581 BD Biosciences  
 CD83 BB790 Custom HB15e BD Biosciences  
 CD16 BUV496 612945 3G8 BD Biosciences  
 CD19 BUV563 741361 HIB19 BD Biosciences  
 CD56 BUV615 613002 BD Biosciences  
 CD69 BUV737 612817 FN50 BD Biosciences  
 CD45 BUV805 612891 HI30 BD Biosciences  
 CD127 BV421 351310 A019D5 BioLegend  
 CD3 BV650 317323 OKT3 BioLegend  
 CD27 BV786 302832 O323 BioLegend  
 CD14 PE-Cy5 15-0149-41 61D3 Thermo Fisher Scientific  
 CD4 PE-Cy5.5 MHCD0418 S3.5 Thermo Fisher Scientific  
 CD123 PE-Cy7 560826 7G3 BD Biosciences  
 CD38 APC-R700 564979 HIT2 BD Biosciences  
 CXCR3 AF647 353712 G025H7 BioLegend  
 CCR4 BB700 566475 1G1 BD Biosciences  
 CXCR5 BB515 564624 RF8B2 BD Biosciences  
 CD137 PE-Cy7 309818 4B4-1 BioLegend  
 CD127 PE-Cy5 351324 A019D5 BioLegend  
 CD95 PE-Dazzle594 305634 DX2 BioLegend  
 CX3CR1 PE 341604 2A9-1 BioLegend  
 TIGIT BV786 747838 741182 BD Biosciences  
 CD39 BV711 328228 A1 BioLegend  
 CD69 BV650 310934 FN50 BioLegend  
 HLA-DR BV605 562845 G46-6 BD Biosciences  
 CD45RA BV570 304132 HI100 BioLegend  
 CD14 BV510 301842 M5E2 BioLegend  
 CD19 BV510 302242 HIB19 BioLegend  
 LIVE/DEAD Fixable Aqua L34966 Thermo Fisher Scientific  
 CD154 BV421 310824 24-31 BioLegend  
 CD3 BUV805 612895 UCHT1 BD Biosciences  
 CCR6 BUV737 612780 11A9 BD Biosciences  
 CD71 BUV661 750651 M-A712 BD Biosciences  
 PD-1 BUV615 612991 EH12.1 BD Biosciences  
 CD28 BUV563 741392 CD28.2 BD Biosciences  
 CD4 BUV496 612936 SK3 BD Biosciences  
 CD8 BUV395 563795 RPA-T8 BD Biosciences  
 CX3CR1 BUV615 751513 2A9-1 BD Biosciences  
 CXCR3 PE-Cy5 353756 G025H7 BioLegend  
 PD-1 BUV737 612791 EH12.1 BD Biosciences  
 LAG-3 BUV661 376-2239-42 3DS223H Thermo Fisher Scientific  
 CD38 BUV496 612946 HIT2 BD Biosciences  
 CD27 BV786 302832 O323 BioLegend  
 HLA-DR BV650 564231 G46-6 BD Biosciences  
 TIM-3 BV605 345018 F38-2E2 BioLegend  
 CD95 BB700 566543 DX2 BD Biosciences  
 CD127 BB630 Custom HIL-7R-M21 BD Biosciences  
 CD4 PE Cy5.5 35-0042-82 RMA-4.5 Thermo Fisher Scientific  
 TIGIT PE-Dazzle 372716 A15153G BioLegend  
 Granzyme B BB790 Custom GB11 BD Biosciences  
 TCF-1 AF488 6444S C63D9 Cell Signaling  
 T-BET PE-Cy7 25-5825-82 4B10 eBioscience  
 KI67 AF700 561277 B56 BD Biosciences

EOMES EF660 50-4877-42 WD1928 eBioscience  
 Anti-mouse IgG HRP 115-005-003 polyclonal Jackson ImmunoResearch  
 Anti-SARS-CoV-2 nucleocapsid protein BSM-41411M 1C7 Stratech Scientific

## Validation

All primary antibodies used in this study were commercially purchased and routinely tested by the respective manufacturers for validation in flow cytometry.

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

## Cell line source(s)

A549 (ATCC#CCL-185) and VeroE6 cells (ATCC#CRL-1586) expressing human angiotensin-converting enzyme 2 (ACE2) and transmembrane serine protease 2 (TMPRSS2).

## Authentication

The A549 (ATCC# CCL-185) and VeroE6 (ATCC# CRL-1586) cell lines were authenticated through morphology assessment, STR (short tandem repeat) profiling. No additional authentication was performed beyond the manufacturer's validation. Cells were used at low passage numbers and routinely monitored for morphology and growth characteristics.

## Mycoplasma contamination

None detected testing routinely.

Commonly misidentified lines  
(See [ICLAC](#) register)

No commonly misidentified cell lines were used in this study.

## Plants

## Seed stocks

Not applicable.

## Novel plant genotypes

Not applicable.

## Authentication

Not applicable.

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

## Sample preparation

PBMCs were isolated via standard density gradient centrifugation and cryopreserved in fetal bovine serum (Thermo Fisher Scientific) containing 10% dimethyl sulfoxide (Sigma-Aldrich).

## Instrument

BD FACSymphony™ A3 or BD FACSymphony™ A5.

## Software

BD FACSDiva version 9.1  
 FlowJo version 10.9.0

## Cell population abundance

No cell sorting experiments were performed in this study.

## Gating strategy

Gating strategies are shown in Extended Data Figures 2a, 2b, 4b.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.