Article

Influenza vaccination reveals sex dimorphic imprints of prior mild COVID-19

https://doi.org/10.1038/s41586-022-05670-5

Received: 3 February 2022

Accepted: 19 December 2022

Published online: 4 January 2023

Check for updates

Rachel Sparks[1,14], William W. Lau[1,14], Can Liu[1,2,14], Kyu Lee Han[3], Kiera L. Vrindten[1], Guangping Sun[1,4], Milann Cox[1], Sarah F. Andrews[5], Neha Bansal[1], Laura E. Failla[1], Jody Manischewitz[6], Gabrielle Grubbs[6], Lisa R. King[6], Galina Koroleva[3], Stephanie Leimenstoll[7], LaQuita Snow[7,11], OP11 Clinical Staff*, Jinguo Chen[3], Juanjie Tang[6], Amrita Mukherjee[3], Brian A. Sellers[3], Richard Apps[3], Adrian B. McDermott[5], Andrew J. Martins[1], Evan M. Bloch[8], Hana Golding[6], Surender Khurana[6] & John S. Tsang[1,3,12,13] ✉

Acute viral infections can have durable functional impacts on the immune system long after recovery, but how they affect homeostatic immune states and responses to future perturbations remain poorly understood[1–4]. Here we use systems immunology approaches, including longitudinal multimodal single-cell analysis (surface proteins, transcriptome and V(D)J sequences) to comparatively assess baseline immune statuses and responses to influenza vaccination in 33 healthy individuals after recovery from mild, non-hospitalized COVID-19 (mean, 151 days after diagnosis) and 40 age- and sex-matched control individuals who had never had COVID-19. At the baseline and independent of time after COVID-19, recoverees had elevated T cell activation signatures and lower expression of innate immune genes including Toll-like receptors in monocytes. Male individuals who had recovered from COVID-19 had coordinately higher innate, influenza-specific plasmablast, and antibody responses after vaccination compared with healthy male individuals and female individuals who had recovered from COVID-19, in part because male recoverees had monocytes with higher IL-15 responses early after vaccination coupled with elevated prevaccination frequencies of 'virtual memory'-like $CD8^+$ T cells poised to produce more IFNγ after IL-15 stimulation. Moreover, the expression of the repressed innate immune genes in monocytes increased by day 1 to day 28 after vaccination in recoverees, therefore moving towards the prevaccination baseline of the healthy control individuals. By contrast, these genes decreased on day 1 and returned to the baseline by day 28 in the control individuals. Our study reveals sex-dimorphic effects of previous mild COVID-19 and suggests that viral infections in humans can establish new immunological set-points that affect future immune responses in an antigen-agnostic manner.

Examples of long-term immunological effects of both chronic and resolved viral infections have been described; for example, after recovery from natural acute measles infection, there is marked reduction in humoral immunity and increased susceptibility to non-measles infections for months to years[1]. Live vaccines such as Bacillus Calmette–Guérin (BCG) and measles can impart 'training' effects on innate immune cells such as monocytes and their long-lived progenitors, which could underlie the pathogen non-specific effects of BCG in reducing all-cause mortality in infants[5,6]. COVID-19 can result in persistent clinical sequelae for months after infection, both in hospitalized and mild cases[7]. Although the spectrum of clinical manifestations of post-acute COVID-19 syndrome (also known as long COVID) is expanding, our understanding of the molecular and cellular immunological changes after recovery from SARS-CoV-2 infection is lacking. A better understanding of the functional immune imprints of mild COVID-19 might have particularly important public health implications given that this population constitutes most COVID-19 recoverees. More broadly, the fundamental issues of whether and how homeostatic baseline immune states may have been altered by viral infections, and whether any such alterations may affect responses to future challenges (such as infection or vaccination, with shared or distinct antigens) remain poorly understood.

Here we took advantage of a unique opportunity and epidemiological environment during the early fall of 2020, months after the first wave

[1]Multiscale Systems Biology Section, Laboratory of Immune System Biology, NIAID, NIH, Bethesda, MD, USA. [2]Graduate Program in Biological Sciences, University of Maryland, College Park, MD, USA. [3]NIH Center for Human Immunology, NIAID, NIH, Bethesda, MD, USA. [4]Division of Intramural Research, NIAID, NIH, Bethesda, MD, USA. [5]Vaccine Research Center, NIAID, NIH, Bethesda, MD, USA. [6]Division of Viral Products, Center for Biologics Evaluation and Research (CBER), FDA, Silver Spring, MD, USA. [7]Laboratory of Clinical Immunology and Microbiology, NIAID, NIH, Bethesda, MD, USA. [8]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [11]Present address: Johns Hopkins University, Baltimore, MD, USA. [12]Present address: Yale Center for Systems and Engineering Immunology and Department of Immunobiology, Yale University School of Medicine, New Haven, CT, USA. [14]These authors contributed equally: Rachel Sparks, William W. Lau, Can Liu. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: john.tsang@yale.edu

of COVID-19, when those with mild COVID-19 had recovered clinically, but before they could be reinfected by SARS-CoV-2 or receive COVID-19 vaccination (which was not available until late 2020); moreover, the prevalence of other respiratory infections was extremely low during this time[8]. We enrolled and comparatively assessed healthy individuals who (1) recovered from non-hospitalized, mild cases of COVID-19 and (2) age- and sex-matched controls who never had COVID-19, all from the same geographical region. In addition to assessing the post-COVID-19 immunological statuses, we used influenza vaccination to evaluate the immune responses of these two populations at the serological, transcriptional, proteomic and cellular levels. These analyses reveal basic principles regarding what happens to the immune system after two well-defined immunological encounters in humans: mild COVID-19 as a natural infectious perturbation and influenza vaccination as a controlled and timed intervention with non-SARS-CoV-2 antigens.

Individuals with previous symptomatic SARS-CoV-2 infection ($n = 31$; diagnosed by nasal PCR test) or asymptomatic infection ($n = 2$; diagnosed by antibody test; Methods), and age- and sex-matched healthy control individuals ($n = 40$) with no history of COVID-19 (and negative by antibody test) were recruited from the community during the fall of 2020 and followed longitudinally (Fig. 1a and Methods). The average time after COVID-19 diagnosis was 151 days for recoverees (Extended Data Fig. 1a and Extended Data Table 1) who had clinically mild illness during acute disease that did not require hospitalization (self-reported average length of illness, 16.5 days) and no major medical comorbidities, including infection at the time of enrolment, obesity (body mass index > 30) or autoimmune disease (Fig. 1b). None of the participants was enrolled in COVID-19 vaccine trials, nor did they receive recent vaccination of any kind before administration of the seasonal influenza vaccine in this study. A small number of individuals continued to have mild self-reported sequelae from their illness at study enrolment (3 males and 8 females), the most common being loss of taste and/or smell (Extended Data Table 1). Female participants were more likely to have sequelae (Fisher's exact test, $P = 0.09$ for all participants, $P = 0.03$ for those aged <65 years), at a rate similar to that reported in other large studies[9].

Baseline of mild COVID-19 recoverees

Longitudinal multi-omics profiling was performed using whole-blood transcriptomics (WBT) analysis, single-cell analysis of 138 surface proteins, transcriptome and V(D)J sequence analysis using cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq)[10], serum protein profiling, antibody characterization, peripheral blood immune cell frequencies with haematological parameters from a complete blood count (CBC), as well as clinical and research flow cytometry covering major immune cell lineages and subsets (Fig. 1b and Supplementary Fig. 1). We first assessed the baseline prevaccination differences between the recoverees and the age- and sex-matched healthy control individuals. As sex-dependent immune responses to COVID-19 have been reported[11], our analyses explicitly searched for sex-dependent signatures. Immunological resolution after infection may unfold over time even after symptoms subside, and there were indeed parameters that showed evidence of continued change in our cohort—defined as those that were correlated with time since COVID-19 diagnosis (TSD; Methods and Supplementary Table 1), including, as expected, SARS-CoV-2-neutralizing antibody titres[12] (Extended Data Fig. 1b). However, we were primarily interested in uncovering persistent, TSD-independent post-COVID-19 immune imprints, and we therefore focused on temporally stable immune states associated with previous mild COVID-19 but not correlated with TSD. Thus, we evaluated the differences between (1) female participants who had recovered from COVID-19 (COVR-F) versus healthy control female participants (HC-F); (2) male participants who had recovered from COVID-19 (COVR-M) versus healthy control male participants (HC-M); and (3) COVR-M versus COVR-F after accounting for male–female differences in healthy control individuals (hereafter, sex differences; Supplementary Table 2). The frequencies of myeloid cells such as monocytes and conventional/myeloid dendritic cells (cDCs) tended to be higher in the COVR-M group compared with the HC-M and/or COVR-F groups (Fig. 1c,d and Extended Data Fig. 1c,d), consistent with reports of myeloid cell disruption in COVID-19, particularly in severe, acute disease[13]. Here, a male-specific elevation in monocyte frequencies was detected even months after recovery from mild disease.

WBT data also revealed sex-dependent signatures associated with previous mild COVID-19 (Extended Data Fig. 1e; for example, the monocyte-related M11.0 and M4.0 from the blood transcriptional module (BTM) collection), including metabolic signatures such as oxidative phosphorylation (Supplementary Table 3). WBT differences can be driven by both cell composition and cell-intrinsic transcriptional changes. Indeed, the innate immune, metabolic and T cell-related signatures are driven, at least in part, by the increased circulating monocyte frequencies and correspondingly lower T cell frequencies in the COVR-M group (Fig. 1d and Extended Data Fig. 1f) because these transcriptional enrichment signals became statistically insignificant when monocyte frequencies were taken into account (data not shown).

To assess transcriptional alterations independent of cell frequencies, we used CITE-seq to examine the cell-type-specific contributions underlying the WBT signatures seen above. We clustered single cells and annotated the resulting clusters using surface protein expression profiles (Fig. 1e and Methods). Cell-type-specific transcriptional analysis pointed to both sex-dependent and -independent differences between participants who had recovered from COVID-19 and healthy control individuals (Supplementary Table 4). Among the enriched gene sets from the WBT analysis above (Extended Data Fig. 1e), but now free of cell-frequency confounding, the BTM M11.0/4.0 gene sets exhibit depressed expression in both classical and non-classical monocytes in participants who had recovered from COVID-19 relative to healthy control individuals in both sexes, whereas the converse is true for genes in the T cell activation signature (BTM M7.3) in both $CD8^+$ central memory and effector memory (EM) T cells (Fig. 1f–i, Extended Data Fig. 1g and Supplementary Table 5). The T cell activation signature in $CD8^+$ EMs was particularly pronounced in the COVR-M group (Fig. 1i). The genes driving the monocyte repression enrichment (that is, the leading-edge genes (LEGs)) include numerous surface receptors, such as those encoding pattern recognition receptors (*TLR2*, *TLR4* and *TLR8*), the peptidoglycan-recognizing receptor (*NOD2*), the high-affinity IgE FC receptor (*FCER1G*) and C-type lectin receptor (*CLEC4E*) (Fig. 1f,g). This innate immune receptor (IIR) signature in the monocytes, as well as the T cell activation signature, are predominantly not associated with TSD in both male and female individuals (Extended Data Fig. 1h).

The T cell activation signature probably emerged during and persisted after acute COVID-19 (ref. [14]), but this was less clear for the IIR signature. We therefore examined whether this signature could be linked to gene expression changes seen in acute COVID-19. Using a previously published CITE-seq dataset that we generated from an older, male-biased cohort of individuals from Italy with severe COVID-19 who were hospitalized[15], we noted that, within the classical monocytes, the average expression of the IIR LEGs from above was significantly lower in patients with acute COVID-19 than in healthy control individuals, and was negatively associated with disease severity (Extended Data Fig. 1i). Thus, this depressed IIR signature could have originated from and stably persisted since the acute response to the infection. Previous studies have reported several (potentially overlapping) types of altered monocytes in acute COVID-19, including those with lower antigen presentation, depressed NF-κB/inflammation or myeloid-derived suppressor-cell-like phenotypes[13,16,17]. However, none of these monocyte phenotypes was significantly different in the monocytes of participants who had recovered from COVID-19 compared with healthy control individuals in our cohort at the baseline before influenza vaccination

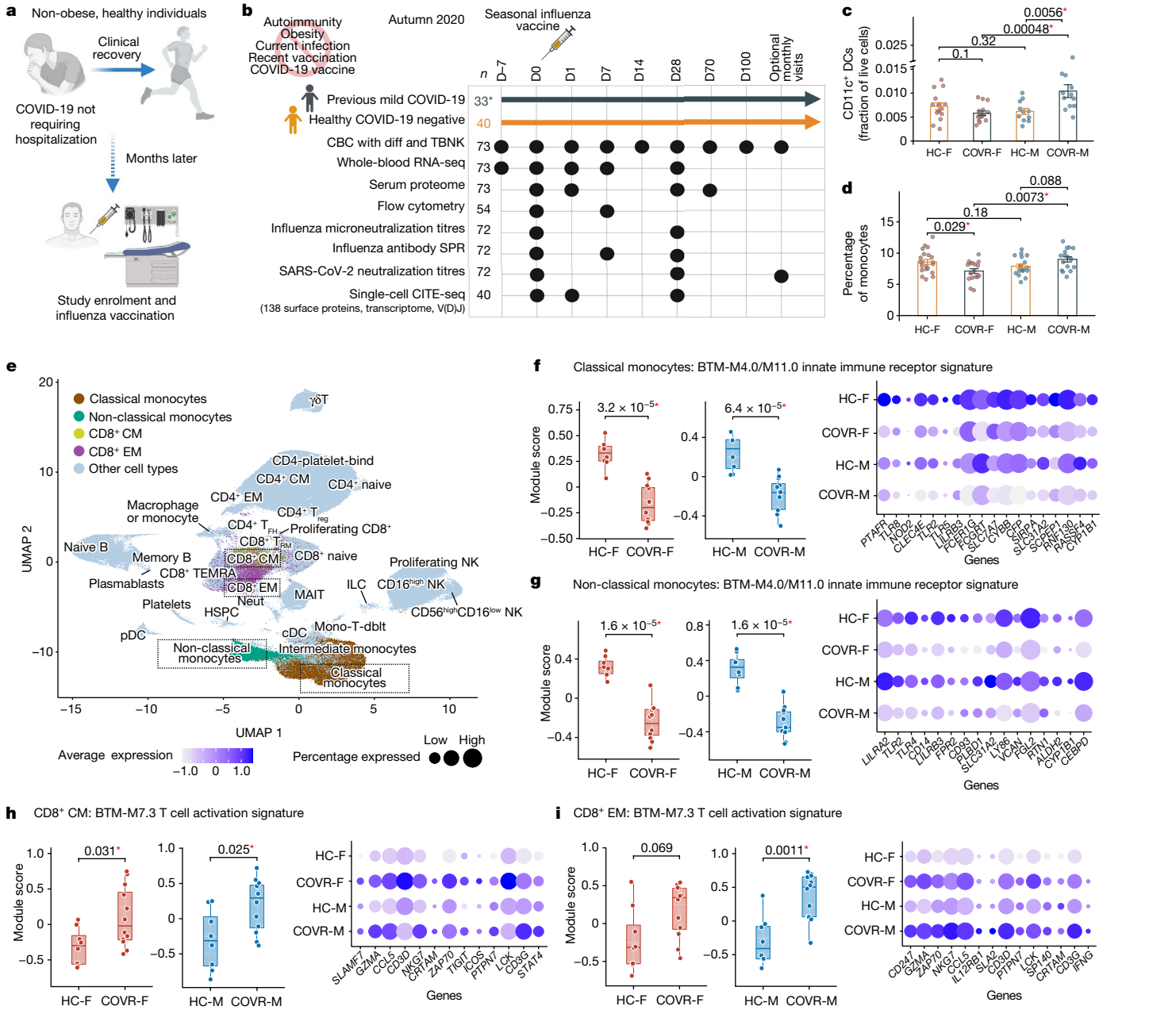


Fig. 1 | Study overview and baseline differences. a, Schematic of the study concept and design. **b**, Data generated in the study. Both participants who had recovered from COVID-19 and healthy control individuals were enrolled at seven days before vaccination (D−7) and were sampled at the indicated timepoints relative to the day of influenza vaccination. The number of participants assayed for each data type is indicated. CBC with diff, complete blood count with differential; SPR, surface plasmon resonance; TBNK, T and B lymphocyte and natural killer cell phenotyping. Where indicated by an asterisk (*), two asymptomatic individuals (based on antibodies) were included. **c**, Comparison of the proportion of CD11c$^+$ DCs (as the fraction of live cells from flow cytometry) between the COVR-F ($n = 15$), HC-F ($n = 16$), COVR-M ($n = 12$) and HC-M ($n = 11$) groups at D0. The error bars indicate the s.e.m. of each group. **d**, Similar to **c**, but for monocytes (from CBC; *y* axis) between the COVR-F ($n = 17$), COVR-M ($n = 16$), HC-F ($n = 21$) and HC-M ($n = 19$) groups at the baseline (average of D−7 and D0). **e**, Uniform manifold approximation and projection (UMAP) analysis of the CITE-seq single-cell data showing clustering of cells on the basis of the expression of cell-surface protein markers (632,100 single cells from all timepoints with CITE-seq data: days 0, 1, 28). The coloured and boxed cell clusters are further examined in **f**–**i**. CD4-platelet-bind: CD4$^+$ T cells with platelet markers; CM, central memory; ILC, innate lymphoid cells; Mono-T-dblt, monocytes and T cell doublets; T_{FH}, T follicular helper cells; T_{reg}, regulatory T cells; T_{RM}, tissue resident memory T cells; HSPC, haematopoietic stem and progenitor cell; Neut, neutrophils; pDC, plasmacytoid dendritic cells; cDC, conventional dendritic cells; MAIT: mucosal-associated invariant T cells. **f**, Comparison of the innate immune receptor (IIR) signature scores (Methods) between the HC-F ($n = 8$) and COVR-F ($n = 12$) (left box) and HC-M ($n = 8$) and COVR-M ($n = 12$) (right box) groups using the CITE-seq classical monocyte pseudobulk expression data at D0 (left). Each point represents a participant. Right, the average gene expression of selected genes, including those in the Gene Ontology (GO) pattern recognition receptor activity and immune receptor activity gene sets. **g**, Similar to **f**, but showing the non-classical monocyte population at D0. **h**, Similar to **f**, but showing the T cell activation (BTM-M7.3) module scores of CD8$^+$ CM T cells at D0. The average gene expression of the selected leading-edge genes shared by male and female from the gene set enrichment analysis (GSEA) is shown (Methods). **i**, Similar to **h**, but showing the CD8$^+$ EM T cell population at D0. All of the box plots show the median (centre line), first and third quantiles (box limits), and max 1.5 × interquartile range (IQR) from box limits in each direction (upper and lower whiskers). Unless otherwise noted, statistical significance of difference between groups was determined using two-tailed Wilcoxon rank-sum tests. Significant ($P < 0.05$) differences are highlighted with a red asterisk (*). The diagrams in **a** and **b** were created using BioRender.

(Supplementary Fig. 2), suggesting that our depressed monocyte gene signature involving pattern recognition and IIR genes is distinct from those identified earlier in acute disease. Together, our findings suggest that, even mild, non-hospitalized SARS-CoV-2 infections may establish new, temporally stable, sex-dependent immunological imprints detectable months after clinical recovery.

To assess whether other natural respiratory viral infections may leave similar unresolved sex-specific immune states, we used a published WBT dataset assessing two independent cohorts of patients with confirmed community influenza A (predominantly pandemic H1N1) infection during two different seasons[18] (2009–2010 and 2010–2011; Extended Data Fig. 2a). By comparing the WBT profiles before and after each season (that is, before infection and after recovery), we found robust post-infection changes consistent between these two independent cohorts in male individuals only (the changes in female individuals were not consistent between these two cohorts/seasons; Extended Data Fig. 2b and Supplementary Table 6). The genes with increased expression after recovery in male individuals were also enriched for genes that were more highly expressed in the COVR-M group compared with the COVR-F group in our cohort (after accounting for the expected sex differences present in healthy participants; Extended Data Fig. 2c). Moreover, the genes with lower expression after recovery from influenza infection in males were enriched for the depressed IIR signature above, including *TLR5* and *VCAN* (Fig. 1f,g and Supplementary Table 6). These observations provide independent support that exposure to a respiratory viral pathogen can lead to persistent immunological imprints that are detectable in the blood, even in healthy individuals with mild disease. However, different viral infections, for example, those with distinct tropisms and inflammatory presentations, are also likely to leave pathogen-dependent imprints with distinct genes and processes; for example, the overlapping signals between post-influenza and post-mild COVID-19 are only a small subset of the sex-specific post-COVID-19 changes that we detected.

Contrasting influenza vaccination responses

We next examined whether previous COVID-19 may impact an individual's response to non-SARS-CoV-2 immunological challenges. The study participants received the seasonal influenza quadrivalent vaccine and were followed longitudinally for up to 100 days, including day 1 (D1), D7 and D28, to assess the vaccine response at the serological, molecular and cellular levels (Figs. 1a,b and 2a). This vaccine was selected in part due to its public health importance—the 2020–2021 influenza season was approaching at the start of our study and it was not clear whether previous COVID-19 infection would affect influenza vaccine responses. Moreover, the responses to seasonal influenza vaccination have been well characterized in healthy adults, including early innate/inflammatory and interferon (IFN) responses on D1 after vaccination and a strong but transient plasmablast peak around D7 culminating in the generation of influenza-specific antibodies[19,20]. Thus, influenza vaccination provides an excellent perturbation to probe the functional impacts of previous mild SARS-CoV-2 infection.

WBT, peripheral immune cell frequency, CITE-seq, influenza-specific B cell and antibody titre analyses (assessing responses on D1, D7 and D28 relative to D0) together pointed to coordinated, sex-specific innate and adaptive response differences to the vaccine, with the COVR-M group generally mounting a more potent response compared with their healthy counterparts and the COVR-F group (Fig. 2b–i, Extended Data Fig. 3a,c,d,g and Supplementary Tables 7 and 8). These include stronger innate/inflammatory and particularly IFN-related transcriptional responses (Fig. 2b and Extended Data Fig. 3a), with corresponding greater increases in circulating IFNγ protein levels in the serum by D1 in the COVR-M group (Fig. 2c). This systemic increase in IFNγ affects diverse cell types expressing the IFNγ signalling components as revealed by single-cell CITE-seq analysis—most peripheral immune cells had higher IFN response signatures on D1 in the COVR-M group compared with the other groups (based on comparing D1 versus D0; Fig. 2d; Fig. 2e shows $CD4^+$ T cells, B cells, monocytes and cDCs as examples). Baseline, prevaccination IFN-related transcriptional activity was largely indistinguishable between the participants who had recovered from COVID-19 and healthy control individuals (Extended Data Fig. 3b). Furthermore, a more robust response was observed for antigen-presentation genes, including both MHC class I and II genes in classical monocytes of the COVR-M group (Fig. 2f). Thus, individuals in the COVR-M group mount a stronger circulating IFNγ and corresponding transcriptional response in both innate and adaptive immune cells by D1 after influenza vaccination.

On the basis of previous studies of influenza vaccination in healthy adults and because heightened innate immune responses elicited by adjuvants are known to enhance adaptive responses[21], we hypothesized that the stronger early inflammatory responses in the COVR-M group would lead to a more robust humoral response. Indeed, we saw increased D7 B cell-related and plasma-cell-related transcriptional signatures in the COVR-M group (Extended Data Fig. 3a,c). Furthermore, the COVR-M group had a greater increase in influenza-specific plasmablasts compared with the HC-M group at D7 (Fig. 2g and Supplementary Fig. 3). Consistent with previous observations in healthy adults[22] and the hypothesis that the stronger early IFN response in the COVR-M group could help to induce a more robust B cell response, we detected a positive correlation between those two parameters, including the extent of influenza-specific plasmablast increases (Extended Data Fig. 3d). Consistently, the COVR-M group also had higher influenza-specific antibody responses compared with the HC-M group across all but one of the vaccine strains at D28 relative to the baseline (Fig. 2h,i, Methods, Extended Data Fig. 3e–g and Supplementary Table 8). Although influenza infection and vaccination history can influence influenza vaccine responses[23], they alone are unlikely to explain the above findings because the COVID-19-recovered and healthy control groups had similar baseline antibody titres (Extended Data Fig. 3e,f), were age/sex-matched and were drawn from the same geographical region with very low influenza infection/transmission during the 2020–2021 season[8]. Moreover, the statistical model used to assess titre response differences incorporated prevaccination influenza titres as a covariate (Methods). The extent of time-dependent immune resolution after COVID-19 was probably not a factor because TSD and D28 titre responses are not correlated in either sex (data not shown). Together, these observations demonstrate that previous mild infection by SARS-CoV-2 can result in sex-dependent, coordinated changes in both innate and adaptive responses to immunization with non-SARS-CoV-2 antigens months after acute disease.

Linking the baseline to innate response

Having established that previous mild COVID-19 is associated with new baseline immune states before influenza vaccination (Fig. 1 and Extended Data Fig. 1) and COVR-M-group-specific responses after vaccination (Fig. 2 and Extended Data Fig. 3), we next attempted to link the two and examined what baseline variables and cellular circuits may contribute to the heightened IFN-related responses in the COVR-M group that could subsequently contribute to their more robust humoral responses (Fig. 3a). Using flow cytometry (Supplementary Fig. 1) and CITE-seq data, we first used a multivariate linear model to identify baseline/prevaccination immune cells of which the frequency predicted the D1 IFN-related responses (D1 versus D0 in serum IFNγ protein levels and IFN transcriptional signature score). A subset of $CD8^+$ T cells with an EM phenotype ($CD45RA^-CCR7^-CD28^+CD27^-$; early effector-like) was a top candidate in the COVR-M group and could therefore be a cellular source of IFNγ after vaccination (Extended Data Fig. 4a,b and Supplementary Fig. 4); the same relationship was not observed in the healthy control individuals (Supplementary Fig. 5a,b).

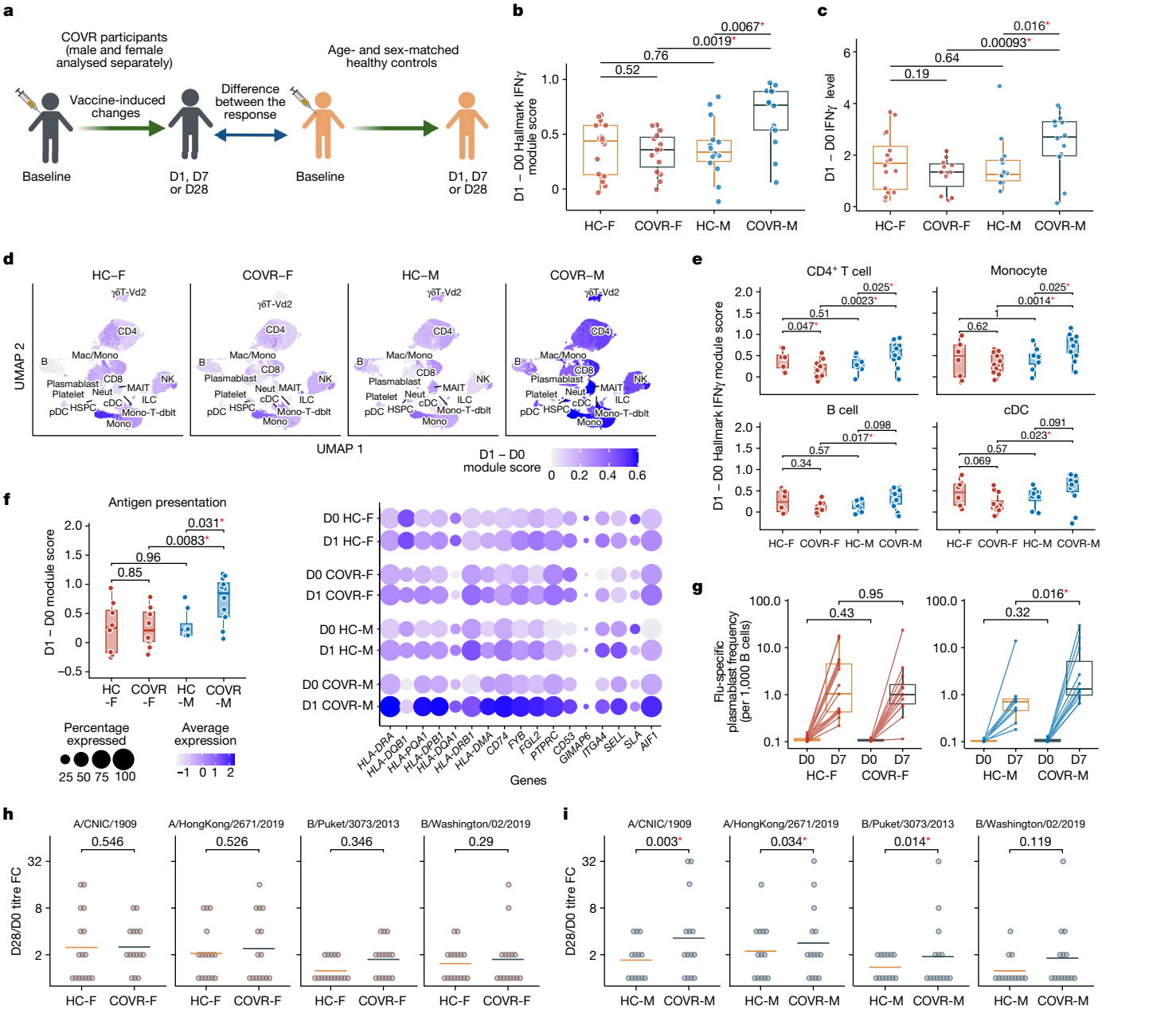


Fig. 2 | Sex-specific response differences to influenza vaccination in individuals who had recovered from COVID-19 and matched control participants. a, Schematic of the sex-specific comparisons of vaccine-induced changes from the baseline at timepoints after vaccination (D1, D7 and D28) between participants who had recovered from COVID-19 and healthy control participants. Analyses were applied to participants aged under 65 years (because older subjects received a higher dose vaccine; Methods). **b**, The D1 whole-blood IFNγ response transcriptional score (D1 – D0, computed using genes from the Hallmark IFNγ response gene set) for the COVR-F ($n = 15$), COVR-M ($n = 14$), HC-F ($n = 16$) and HC-M ($n = 14$) groups. **c**, The D1 response (D1 – D0) of serum IFNγ protein levels for the participants shown in **b**. **d**, Surface-protein-expression-based UMAP analysis (as in Fig. 1e) with cells coloured according to the D1 IFNγ response transcriptional score (D1 – D0; see **b** for the gene set used) within each cell subset for the HC-F ($n = 8$), COVR-F ($n = 12$) and HC-M ($n = 8$) and COVR-M ($n = 12$) groups. Darker colour indicates a greater difference between D1 and D0 for the indicated cell subset. Mac, macrophages; Mono, monocytes. **e**, Similar to **b**, but for the indicated cell subsets (computed using the CITE-seq pseudobulk mRNA expression data for the cell subset) in the HC-F ($n = 8$), COVR-F ($n = 12$), HC-M ($n = 8$) and COVR-M ($n = 12$) groups. **f**, The D1 transcriptional response score (D1 – D0) of the antigen-presentation-related genes in classical monocytes for the same participants in **e** (left) (Methods). Right, the averaged expression of individual LEGs from the antigen-presentation genes (Methods) in classical monocytes. **g**, Influenza-specific plasmablast (PB; all $HA^+CD27^+CD38^+CD20^{low}CD21^{low}$ cells; Methods and Supplementary Fig. 3) frequencies at D0 and D7, plotted separately for the COVR-F ($n = 14$), HC-F ($n = 15$), COVR-M ($n = 11$) and HC-M ($n = 9$) groups. The lines connect data points from the same participant at D0 and D7. **h**, Analysis of the D28/D0 microneutralization titre fold change (FC) for each of the four strains in the seasonal influenza vaccine (columns) in the COVR-F and HC-F groups. Each dot represents one individual. The orange and grey lines indicate the average fold change for the HC-F and COVR-F groups, respectively. Unadjusted *P* values were derived from generalized linear models accounting for age, race, influenza vaccination history and baseline influenza titres (Methods). **i**, Similar to **h**, but for the COVR-M and HC-M groups. All of the box plots show the median (centre line), first and third quantiles (box limits), and max 1.5 × IQR from box limits in each direction (upper and lower whiskers). Unadjusted *P* values are shown. Unless otherwise noted, the statistical significance of the difference between groups was determined using two-tailed Wilcoxon rank-sum tests. Significant ($P < 0.05$) differences are highlighted with a red asterisk (*). The diagram in **a** was created using BioRender.

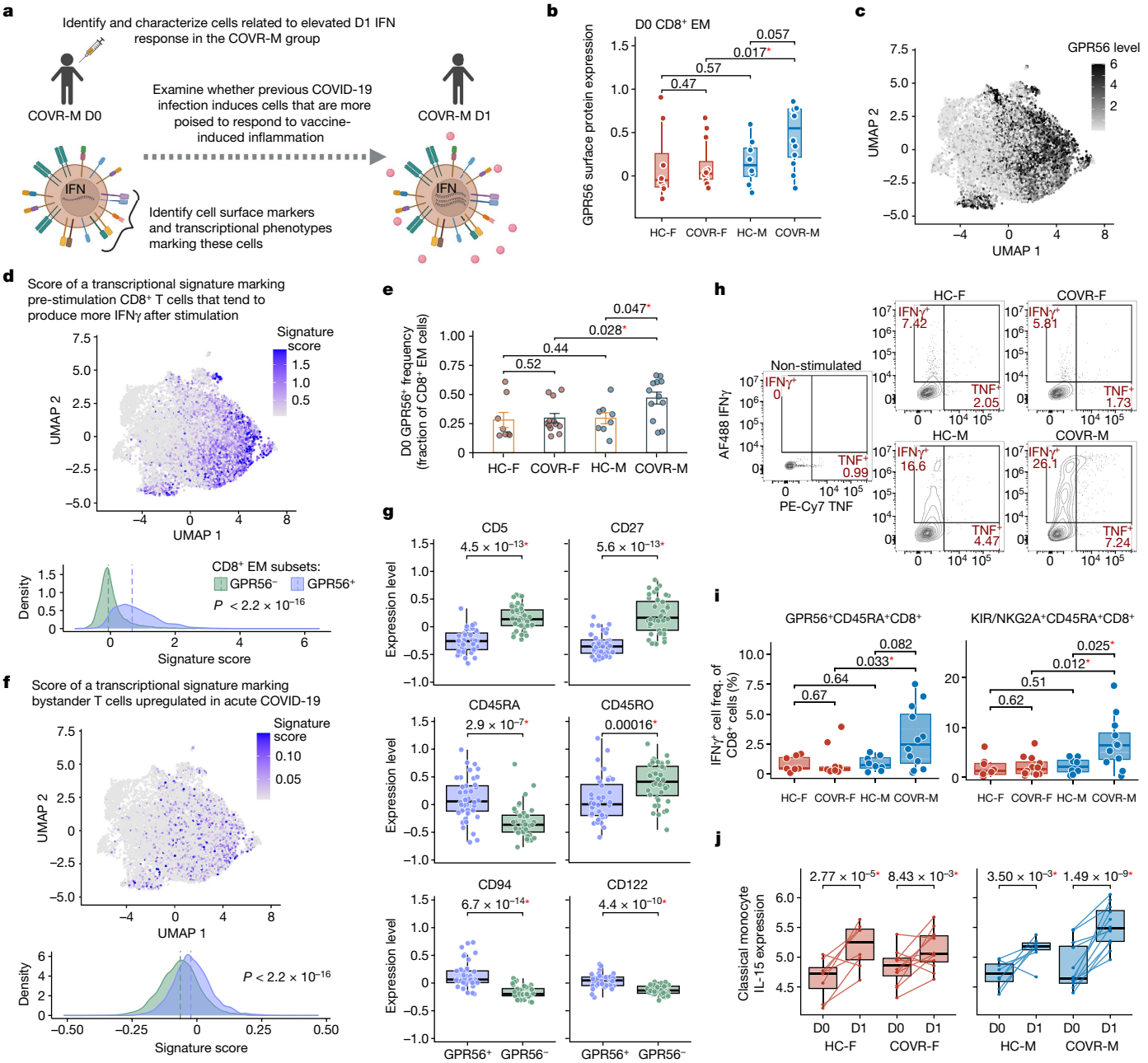


Fig. 3 | Contributors to increased day 1 IFNγ responses in male participants who had recovered from COVID-19. a, Schematic of the approach to assess why the COVR-M group had elevated early IFNγ responses. **b**, Comparison of the sample means of GPR56 surface expression in CD8⁺ EM T cells at D0 for the COVR-F ($n = 12$), HC-F ($n = 8$), COVR-M ($n = 12$) and HC-M ($n = 8$) groups. **c**, UMAP analysis of the D0 surface GPR56 protein expression on CD8⁺ EM cells from all 40 participants with CITE-seq data. The UMAP was derived using the top 60 variable surface proteins within the CD8⁺ EM cells (Methods). **d**, UMAP analysis as described in **c**, but showing the D0 gene-expression signature score computed using genes associated with CD29^{high}CD8⁺ T cells identified earlier in an independent study[25] (Methods) (top). Density plot showing the distribution of the signature score above in the GPR56⁺CD8⁺ and GPR56⁻CD8⁺ EM cells (bottom). The dashed line indicates the median of the distribution. The statistical significance of the signature-score difference between the two cell subsets was determined at the single-cell level. **e**, Comparison of the proportion of GRP56⁺ cells (as fractions of CD8⁺ EM cells in the CITE-seq data) between the same participants as in **b** at D0. The error bars indicate the s.e.m. of each group. **f**, Similar to **d**, but showing the bystander T cell signature score at the baseline (D0) (signature genes originated from refs. [26,27]; Methods). **g**, Comparison of the average expression of the indicated memory cell-surface protein markers for the GPR56⁺CD8⁺ versus GPR56⁻CD8⁺ EM cells at D0 for the same participants as in **b**. Each point represents a participant. **h**, Representative flow cytometry contour plots of IFNγ⁺ and TNF⁺ gates within GPR56⁺CD45RA⁺CD8⁺ T cells after IL-15 stimulation in vitro in the indicated groups. The number shown for each gate denotes the percentage of parent cells (that is, GPR56⁺CD45RA⁺CD8⁺ T cells). **i**, The frequencies of IFNγ⁺GPR56⁺CD45RA⁺ VM-like CD8⁺ T cells (left; as fractions of CD8⁺ T cells) and IFNγ⁺KIR/NKG2A⁺CD45RA⁺CD8⁺ T cells (right; as fractions of CD8⁺ T cells) in the same participants as in **b** after IL-15 stimulation in vitro. **j**, Comparison of D0 and D1 pseudobulk *IL15* mRNA expression (*y* axis) in classical monocytes for the same participants as in **b**. Significance was determined using a linear model accounting for age, race and influenza vaccination history (Methods). All of the box plots show the median (centre line), first and third quantiles (box limits), and max 1.5 × IQR from box limits in each direction (upper and lower whiskers). Unless otherwise noted, the statistical significance of difference between groups was determined using two-tailed Wilcoxon rank-sum tests. Significant ($P < 0.05$) differences are highlighted with a red asterisk (*). The diagram in **a** was created using BioRender.

We next focused on all of the CD8$^+$ T cells from clusters with an EM phenotype (CD8$^+$ EM cells) in the CITE-seq data based on both surface protein markers and mRNA expression (Methods; the top cluster protein markers are shown in Supplementary Table 10). We searched for differences in average surface marker expression of cells in these CD8$^+$ EM clusters across the four participant groups and found that GPR56 was the top differentially expressed marker with increased expression in the COVR-M group relative to the HC-M and COVR-F groups (Fig. 3b,c and Supplementary Table 10). This was intriguing because CD4$^+$ EM and TEMRA (terminally differentiated EM cells re-expressing CD45RA) T cells marked by surface GPR56 expression at the baseline (before stimulation) have been reported to produce increased amounts of IFNγ after stimulation with PMA and ionomycin (PMAI)[24]. Consistent with this, GPR56$^+$CD8$^+$ EM cells in our data are enriched for a transcriptional signature (derived in an independent study[25]) that marks CD8$^+$ EM cells poised to secrete higher levels of IFNγ after PMAI stimulation (Fig. 3d). Thus, GPR56$^+$CD8$^+$ EM cells could be a source of elevated IFNγ production in the COVR-M group after influenza vaccination. Indeed, the frequency of these cells was elevated in the COVR-M group relative to in the HC-M and COVR-F groups before vaccination (Fig. 3e), but was not correlated with the TSD and was therefore temporally stable (assessed by Spearman's correlation: $P = 0.18$ (COVR-F) and $P = 0.51$ (COVR-M)). Moreover, *IFNG* transcripts increased significantly in these cells on D1 after influenza vaccination in the COVR-M group (Extended Data Fig. 4c,d). These data suggest that previous COVID-19 increases the frequency of GPR56$^+$CD8$^+$ EM cells in male individuals and these cells are poised to make more IFNγ early after influenza vaccination, which together contributed to the higher IFNγ production in the COVR-M group; consistent with this hypothesis, this was not observed in GPR56$^-$ cells (Extended Data Fig. 4d and Supplementary Fig. 5c).

Mild, non-hospitalized COVID-19 has been reported to induce bystander activation (non-SARS-CoV-2 specific) of CD8$^+$ T cells[26]. Notably, the GPR56$^+$ cells are also enriched for a transcriptional signature associated with bystander T cell activation[26,27] (Fig. 3f). Moreover, GPR56$^+$CD8$^+$ EM cell frequency is positively correlated with the T cell activation signature score, which was elevated at the baseline in the COVR-M group as shown above (Fig. 1i and Extended Data Fig. 4e). This suggests that some of these cells may have expanded in a bystander manner during the acute phase of the infection. This prompted us to consider whether these GPR56$^+$ cells are similar to bystander-activated virtual memory (VM) CD8$^+$ T cells, a feature of which is their ability to be activated rapidly by inflammatory cytokines alone (for example, IL-12, IL-18 and IL-15) to produce IFNγ without T cell receptor (TCR) stimulation[28,29]. VM CD8$^+$ T cells expand through cytokine stimulation, including IL-15 induced by viral infection (IL-15 concentrations are known to be elevated in patients with acute COVID-19 and are correlated with disease severity[30]), and are characterized by a differentiated EM phenotype expressing CD45RA[28]. We assessed several reported surface markers of these cells[28] in GPR56$^+$ versus GPR56$^-$ cells and found that the GPR56$^+$ cells were indeed phenotypically similar to VM cells (Fig. 3g). For example, GPR56$^+$ cells have higher CD122 but lower CD5 surface expression compared with their GPR56$^-$ counterparts; CD5 surface expression has been linked to the extent of previous IL-15 (or potentially other inflammatory cytokine) encounters[28,31]. Notably, on the basis of the surface levels of CD45RA and CD45RO, the GPR56$^+$ cells appear to situate phenotypically between GPR56$^-$ and TEMRA cells (Extended Data Fig. 4f).

To further test our hypothesis, we performed in vitro stimulation experiments to assess whether GPR56$^+$CD8$^+$ T cells can produce IFNγ in response to several cytokines that are known to be induced by vaccination or infection (Supplementary Fig. 6a). Stimulation with IL-15 showed that GPR56$^+$CD45RA$^+$CD8$^+$ T cells from the COVR-M group produced more IFNγ compared with those from the COVR-F group (Fig. 3h,i). CD8$^+$ VM-like T cells were identified using the surface markers CD45RA$^+$, KIR$^+$ and/or NKG2A$^+$[32,33] and the COVR-M group produced higher levels of IFNγ in these cells (Fig. 3i). Stimulation with IL-12, IL-15 and IL-18 together showed similar trends (Supplementary Fig. 6b). Stimulation with IL-18 alone or IL-12 and IL-18 together also showed similar trends, but these conditions induced less robust IFNγ than IL-15 stimulation (data not shown). We next assessed the cellular source of IL-15 post-influenza vaccination using CITE-seq data and found that classical monocytes from the COVR-M group showed the most significant increases in *IL15* mRNA levels on day 1 after influenza vaccination (Fig. 3j). Together, this suggests that the increased IFNγ response in the COVR-M group after vaccination could be attributed to increased baseline (prevaccination) frequencies of cells that are also intrinsically more responsive to inflammatory stimulation, including classical monocytes that produce elevated IL-15 and CD8$^+$ VM-like T cells that mount a more robust IFNγ response to cytokine stimulation alone.

As VM T cells can be rapidly activated to produce cytokines without clonal, antigen-specific expansion[28], we assessed the clonality of the GPR56$^+$CD8$^+$ EM cells at different timepoints after influenza vaccination using V(D)J/TCR data from CITE-seq. The clonality of both the GPR56$^+$CD8$^+$ EM and TEMRA cells remained stable across D0 (before vaccination), D1 and D28 after influenza vaccination (Extended Data Fig. 4g,h). The frequencies of GPR56$^+$CD8$^+$ EM clones shared across timepoints within individuals were also similar (Extended Data Fig. 4i). Together, these data argue against the notion that the heightened activation of the GPR56$^+$ cells early after influenza vaccination in the COVR-M group was due solely to TCR-dependent T cell activation and clonal expansion. As was shown previously[28,29] and above in our in vitro stimulation data, a more plausible explanation is that these CD8$^+$ VM-like cells were activated to produce IFNγ by the inflammatory cytokines elicited by the influenza vaccine in an antigen-independent manner. Despite their resemblance to VM cells, some of the GPR56$^+$ cells could have developed from naive cells through conventional, non-bystander pathways (for example, some could be developed during acute COVID-19 and are specific for SARS-CoV-2), although none of these cells had a CDR3 sequence that matches a public clone deemed to be specific for SARS-CoV-2 (data not shown). Bona fide, antigen-specific memory CD8$^+$ T cells developed from naive cells through TCR stimulation have also been shown to produce IFNγ in response to inflammatory cytokines alone in mice[34,35].

Our data also revealed other cell types that could have contributed to the increased IFNγ production observed on D1 after vaccination in the COVR-M group (Supplementary Fig. 7a–c). *IFNG* transcripts increased more in the COVR-M group compared with in the HC-M and COVR-F groups on D1 in CD16^{low} natural killer (NK) cells (Supplementary Fig. 7c and Supplementary Table 4). Moreover, the baseline frequency of CD16^{low} NK cells was correlated with the extent of the D1 increase in both *IFNG* expression and serum protein levels (Supplementary Fig. 7b). However, the IFNγ response in total NK cells after IL-15 stimulation in vitro was not significantly higher in the COVR-M group (Extended Data Fig. 4j), probably because CD16^{low} NK cells are a small subset of total NK cells. By contrast, IL-15 stimulation in vitro revealed a higher IFNγ response in MAIT cells in the COVR-M group compared with the COVR-F and HC-M groups (Extended Data Fig. 4j), but the increase in *IFNG* mRNA expression on D1 after influenza vaccination was not statistically significant in the COVR-M group based on CITE-seq data (Supplementary Fig. 7c). CD8$^+$ T cells with a TEMRA (CD45RA$^+$CD45RO$^-$CCR7$^-$) phenotype might also have a role as their IFNγ response after IL-15 stimulation in vitro was higher in the COVR-M group compared with in the COVR-F and HC-M groups (Extended Data Fig. 4j), which is consistent with the CITE-seq data (Supplementary Fig. 7c).

Taken together, we demonstrate a population of CD8$^+$ EM T cells marked by GPR56 expression and VM-like markers with antigen-agnostic pro-inflammatory potential after heterologous vaccination. Importantly, these cells, and potentially CD16^{low} NK, MAIT and CD8$^+$ TEMRA cells (albeit with less support from our CITE-seq data), emerged in otherwise clinically healthy individuals and are especially elevated

and more poised to respond in male individuals who were months recovered from mild SARS-CoV-2 infection, providing additional evidence for sex-specific, functionally relevant immune set points linked to previous mild COVID-19.

Vaccination shifts monocyte imprints

Given the potential for vaccine-induced training effects[6,36,37], we next examined whether influenza vaccination can alter some of the post-COVID-19 transcriptional imprints that we detected earlier (Fig. 4a). We focused on the monocytes owing to the robustly depressed IIR signature reported above (in participants who had recovered from COVID-19 versus healthy control participants; Fig. 1f,g) and because vaccines can potentially induce long-lasting changes in these cells[6,36]. Using the healthy control baseline (D0) as a healthy reference, we used CITE-seq data to assess the average expression of the signature genes (identified above) before and after vaccination in participants who had recovered from COVID-19, separately for classical (Fig. 1f) and non-classical monocytes (Fig. 1g) in male and female individuals (Extended Data Fig. 5a,b). As was observed above, these genes had lower average expression in participants who had recovered from COVID-19 compared with healthy control individuals in both sexes at D0 before vaccination. However, their average expression increased towards that of the healthy control individuals by D1 and persisted until D28 in the COVR-F and COVR-M groups, although the effect appeared to be stronger in the COVR-F group (Extended Data Fig. 5a,b).

Quantifying the average expression (module score) of these sex- and cell-type-dependent gene sets (Fig. 1f,g) within individual participants over time confirmed a similar and significant trend of shift towards the healthy control individuals (Fig. 4b,c). This analysis further revealed that the extent of this change in gene expression was more pronounced in the non-classical than in the classical monocytes (Fig. 4b,c). Notably, the behaviour of these genes was divergent in the healthy control individuals—the gene module score trended lower on D1 and reverted to prevaccination levels by day 28 in the healthy control individuals (Fig. 4b,c). Although the underlying mechanism of this divergence is unclear, the monocytes in healthy control individuals could have responded to the vaccine-induced inflammation by downregulating certain immune receptor genes and associated signalling genes in a negative feedforward mechanism to avoid over-responding, while the 'depressed' monocytes in participants who had recovered from COVID-19 instead responded by increasing the expression of these genes and therefore moving towards the normal (healthy baseline) level.

We next identified the individual genes within these gene sets that moved towards the healthy control baseline (Methods). In both classical and non-classical monocytes, the fraction of reverting genes was significantly higher in female compared with male participants (Fig. 4d,e and Extended Data Fig. 5c), although several TLRs (for example, *TLR2* and *TLR4*) and *NOD2* were significant in both sexes in one or both monocyte subsets. These changes were probably not due to continued immune resolution after infection because the baseline (D0) expression of these genes did not correlate with TSD (Extended Data Fig. 1h), and they increased acutely by D1 after vaccination and persisted to D28. Notably, in contrast to this depressed IIR signature (Fig. 1f,g and Extended Data Fig. 1i), other monocyte-related transcriptional signatures that are known to have lower expression during acute COVID-19—such as genes related to antigen presentation, inflammatory and NF-κB activation and myeloid suppressor cells[13,15–17,38,39]—were similar between participants who had recovered from COVID-19 and healthy control individuals at D0/baseline; vaccination also did not consistently elicit longer-lasting changes in these signatures out to D28, although the COVR-M group tended to have elevated antigen presentation transcriptional responses in non-classical monocytes on D1 that remained mildly elevated by D28 (Extended Data Fig. 5d,e).

Together, CITE-seq analysis revealed that the early (D1) response to influenza vaccination elevates a set of previously (that is, before vaccination) depressed IIR genes in the monocytes of participants who had recovered from COVID-19 out to at least D28 after vaccination. Although the functional relevance of these changes remains to be determined, these results suggest that the early inflammatory responses to influenza vaccination can help to shift the post-COVID-19 immune state of monocytes towards that of healthy individuals, particularly in female recoverees.

Discussion

Although both acute and longer-term immune perturbations in hospitalized patients with COVID-19 have been reported[13,40–43], less is known regarding healthy recovered individuals with previous mild, non-hospitalized SARS-CoV-2 infection months after acute illness, without confounding comorbidities such as obesity, autoimmunity or immunodeficiency. Here we reveal that clinically healthy recoverees of previous non-hospitalized COVID-19 possess sex-specific immune imprints beyond SARS-CoV-2-specific immunity, some of which become apparent only after vaccination with antigens that are distinct from SARS-CoV-2. Our findings are consistent with the sex dimorphic nature of acute responses to SARS-CoV-2 and other immune challenges[11]. Healthy female individuals tend to mount heightened inflammatory responses to infections and vaccines[44]; it was therefore surprising to find the qualitative opposite here in which the COVR-M was found to have a more poised immune status at the baseline and stronger innate and adaptive responses to influenza vaccination. Although persistent immune state changes (over months) in patients with long COVID have been reported[41], most of the individuals in our study reported no or minor post-COVID-19 sequelae. Future research could assess whether some of the sex-specific imprints, including differences in vaccination responses, are associated with long COVID[7].

Our findings suggest that the poised baseline immune states in the COVR-M group helped to establish the more robust IFN, plasmablast and antibody responses on days 1, 7 and 28, respectively, after influenza vaccination. The early IFN responses may be attributed to monocytes with higher *IL15* transcriptional responses early after vaccination coupled with elevated prevaccination frequencies of VM-like $CD8^+$ T cells poised to produce more IFNγ after IL-15 stimulation. The monocyte imprint that we described involving poised *IL15* mRNA production in male recoverees and the transcriptionally depressed innate receptor gene signature in both sexes are consistent with the notion of trained innate immunity[6]. Notably, although the latter signature could be detected in patients with acute COVID-19 with severe disease, it is distinct from the depressed antigen presentation or myeloid-suppressor-cell-like states found in previous studies of acute COVID-19 (refs. [13,15–17,38,39]). As trained innate immunity can be mediated through myriad mechanisms, including chromatin and metabolic changes within cells, future studies could explore these potential mechanisms in monocytes, including the influences of sex/gender, acute disease severity and age among participants with a range of post-COVID-19 clinical sequelae. Given that the half-life of circulating monocytes is relatively short (and can be shorter than 28 days)[45], the partial reversal that we detected is possibly attributable to bone marrow myeloid progenitor cells, as haematopoietic stem and progenitor cells have been shown to exhibit chromatin accessibility changes after SARS-CoV-2 infection[46].

Bystander T cell activation has been reported after natural viral infections[47], including SARS-CoV-2 (ref. [26]). More recently, bystander-activated $CD8^+$ EM T cells have been identified to have an important role in controlling early infection, including VM cells that have no previous antigen exposure or TCR engagement[28,29]. As these cells can emerge after cytokine stimulation alone, it is possible that a stronger or more prolonged cytokine response to SARS-CoV-2 in male relative to female individuals during acute disease may have resulted in the

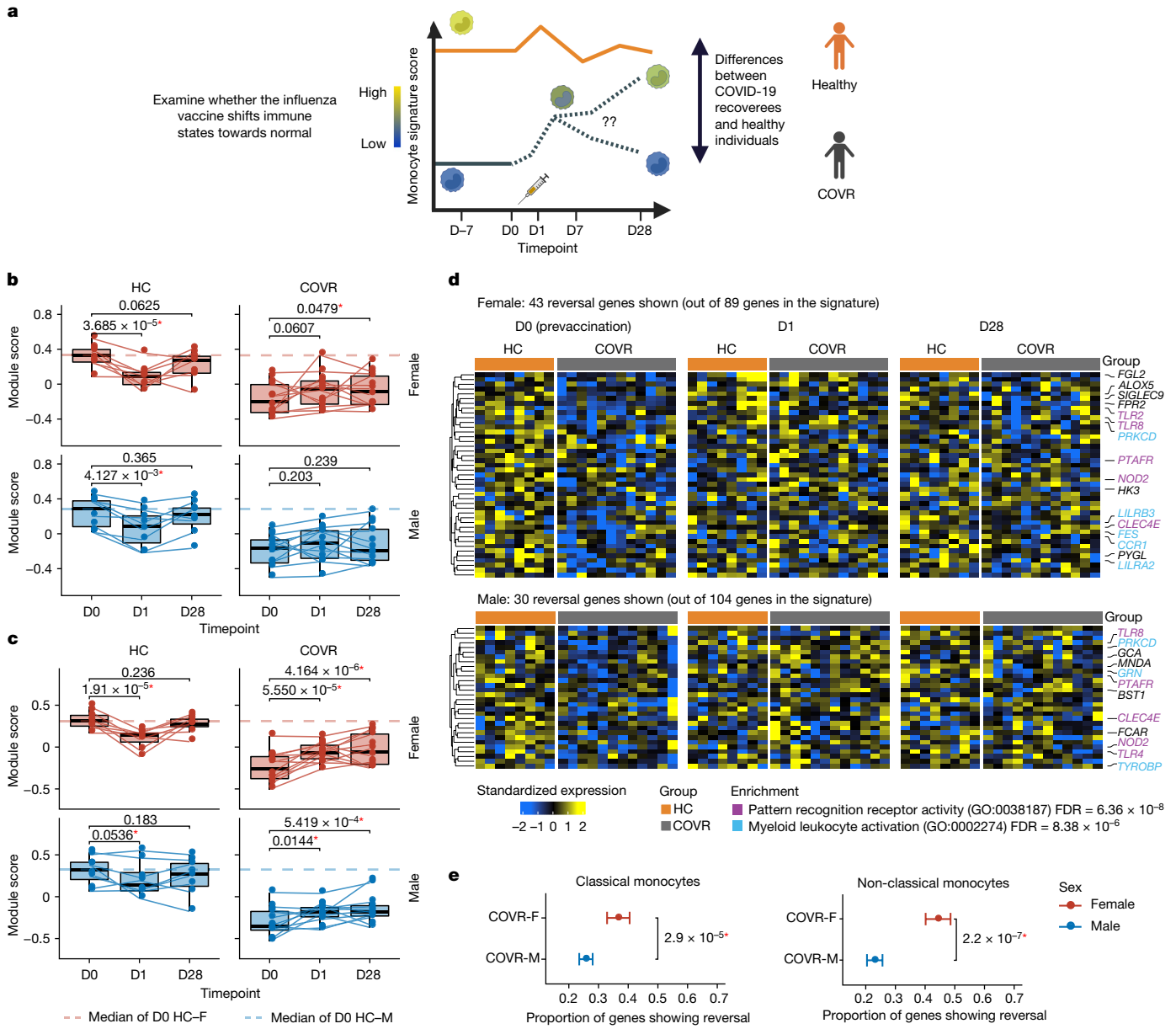


Fig. 4 | Post-mild-COVID-19 gene expression imprints in monocytes shifted by influenza vaccination. a, Schematic of the study questions. **b**, The module scores of the IIR signature (Fig. 1f) in the HC-F ($n = 8$), HC-M ($n = 8$), COVR-F ($n = 12$) and COVR-M ($n = 12$) groups at D0, D1 and D28 using the CITE-seq pseudobulk gene expression data in classical monocytes. The dashed line represents the median D0 score of the healthy control individuals of the same sex. The lines connect data points from the same participant at different timepoints. The statistical significance of differences was determined using a mixed-effects model accounting for age, race and influenza vaccination history (Methods). Unadjusted *P* values are shown. **c**, Similar to **b**, but for non-classical monocytes (Fig. 1g). **d**, Heat map showing the expression of the reversal genes in classical monocytes (row-standardized; see Extended Data Fig. 5c for non-classical monocytes). Reversal genes are defined as those genes in the baseline IIR signature (Fig. 1f) of which the expression in participants who had recovered from COVID-19 at D1 and D28 after vaccination moved towards the baseline (prevaccination) expression of healthy control individuals. The COVR-F (top) and COVR-M (bottom) groups are shown separately; healthy control individuals are also included for comparison. The rows are genes and columns are individual samples (grouped by participant/timepoint) with timepoint and participant group labels shown at the top, including the same participants as in **b** at each timepoint. The names of genes that belong to gene sets of functional interest are shown. False-discovery-rate-corrected enrichment *P* values are shown. **e**, Comparison of the proportion of IIR signature genes (Fig. 1f,g) that show partial reversal in the COVR-F versus COVR-M groups in classical and non-classical monocytes. The mean and 95% confidence intervals (denoted by the bars) were derived from a bootstrapping procedure (Methods). Significance was determined using two-tailed Wilcoxon tests between the bootstrapped samples. All of the box plots show the median (centre line), first and third quantiles (box limits), and max 1.5 × IQR from box limits in each direction (upper and lower whiskers). Significant ($P < 0.05$) differences are highlighted with a red asterisk (*). The diagram in **a** was created using BioRender.

elevated frequencies of the GPR56^{+}CD8^{+} VM-like cells in the COVR-M group. This hypothesis is consistent with reports that male individuals hospitalized with COVID-19 tend to experience greater innate immune activation (as measured by circulating cytokines) compared with female individuals[48,49].

Some of the immune imprints that we observed could be shared among different types of viral infections, but some are probably unique to SARS-CoV-2, as suggested by our comparison with natural influenza infection. Our findings point to the possibility that any infection or immune challenge may change the immune status to establish new

baseline set points encoded by the states of not only a single cell lineage, but also a network of interacting cell types such as VM T cells and monocytes. Moreover, although baseline immune statuses that are predictive of future responses are often different across and temporally stable within individuals over a timescale of months[50,51], our results suggest that such baseline immune states could have been established by past infections and are stable up to the next perturbation. Thus, the baseline immune status of an individual, with the potential to impact future responses in both antigen-specific and antigen-agnostic ways, is shaped by a multitude of previous exposures[2,3]. In addition to revealing underlying principles regarding what happens after two well-defined natural immunological encounters—mild COVID-19 and influenza vaccination in humans—our observations provide a basis for studying more complex scenarios, such as what happens over longer timescales with additional inflammatory encounters. Our research brings forth the concept that even mild viral infections could establish new immunological set-points impacting future immune responses in an antigen-agnostic manner and illustrates how heterologous vaccination could be used as a tool to reveal such functional imprints.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-05670-5.

1. Mina, M. J. et al. Measles virus infection diminishes preexisting antibodies that offer protection from other pathogens. *Science* **366**, 599–606 (2019).
2. Tomalka, J. A., Suthar, M. S., Diamond, M. S. & Sekaly, R. P. Innate antiviral immunity: how prior exposures can guide future responses. *Trends Immunol.* **43**, 696–705 (2022).
3. Tsang, J. S. et al. Improving vaccine-induced immunity: can baseline predict outcome? *Trends Immunol.* **41**, 457–465 (2020).
4. Goodridge, H. S. et al. Harnessing the beneficial heterologous effects of vaccination. *Nat. Rev. Immunol.* **16**, 392–400 (2016).
5. Aaby, P., Netea, M. G. & Benn, C. S. Beneficial non-specific effects of live vaccines against COVID-19 and other unrelated infections. *Lancet Infect. Dis.* **23**, e34–e42 (2022).
6. Netea, M. G. et al. Defining trained immunity and its role in health and disease. *Nat. Rev. Immunol.* **20**, 375–388 (2020).
7. Nalbandian, A. et al. Post-acute COVID-19 syndrome. *Nat. Med.* **27**, 601–615 (2021).
8. Olsen, S. J. et al. Changes in influenza and other respiratory virus activity during the COVID-19 pandemic—United States, 2020–2021. *MMWR Morb. Mortal. Wkly Rep.* **70**, 1013–1019 (2021).
9. Sudre, C. H. et al. Attributes and predictors of long COVID. *Nat. Med.* **27**, 626–631 (2021).
10. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
11. Ursin, R. L. & Klein, S. L. Sex differences in respiratory viral pathogenesis and treatments. *Annu. Rev. Virol.* **8**, 393–414 (2021).
12. Wheatley, A. K. et al. Evolution of immune responses to SARS-CoV-2 in mild-moderate COVID-19. *Nat. Commun.* **12**, 1162 (2021).
13. Schultze, J. L. & Aschenbrenner, A. C. COVID-19 and the human innate immune system. *Cell* **184**, 1671–1692 (2021).
14. Sette, A. & Crotty, S. Adaptive immunity to SARS-CoV-2 and COVID-19. *Cell* **184**, 861–880 (2021).
15. Liu, C. et al. Time-resolved systems immunology reveals a late juncture linked to fatal COVID-19. *Cell* **184**, 1836–1857 (2021).
16. Paludan, S. R. & Mogensen, T. H. Innate immunological pathways in COVID-19 pathogenesis. *Sci. Immunol.* **7**, eabm5505 (2022).
17. Reyes, M. et al. Plasma from patients with bacterial sepsis or severe COVID-19 induces suppressive myeloid cell production from hematopoietic progenitors in vitro. *Sci. Transl. Med.* **13**, eabe9599 (2021).
18. Zhai, Y. et al. Host transcriptional response to influenza and other acute respiratory viral infections—a prospective cohort study. *PLoS Pathog.* **11**, e1004869 (2015).
19. Pulendran, B. Systems vaccinology: probing humanity's diverse immune systems with vaccines. *Proc. Natl Acad. Sci. USA* **111**, 12300–12306 (2014).
20. Tsang, J. S. Utilizing population variation, vaccination, and systems biology to study human immunology. *Trends Immunol.* **36**, 479–493 (2015).
21. Pulendran, B., S. Arunachalam, P. & O'Hagan, D. T. Emerging concepts in the science of vaccine adjuvants. *Nat. Rev. Drug Discov.* **20**, 454–475 (2021).
22. Bucasas, K. L. et al. Early patterns of gene expression correlate with the humoral immune response to influenza vaccination in humans. *J. Infect. Dis.* **203**, 921–929 (2011).
23. Auladell, M. et al. Influenza virus infection history shapes antibody responses to influenza vaccination. *Nat. Med.* **28**, 363–372 (2022).
24. Truong, K.-L. et al. Killer-like receptors and GPR56 progressive expression defines cytokine production of human CD4$^+$ memory T cells. *Nat. Commun.* **10**, 2263 (2019).
25. Nicolet, B. P. et al. CD29 identifies IFN-γ-producing human CD8$^+$ T cells with an increased cytotoxic potential. *Proc. Natl Acad. Sci. USA* **117**, 6686–6696 (2020).
26. Bergamaschi, L. et al. Longitudinal analysis reveals that delayed bystander CD8$^+$ T cell activation and early immune pathology distinguish severe COVID-19 from mild disease. *Immunity* **54**, 1257–1275 (2021).
27. Bangs, S. C. et al. Human CD4$^+$ memory T cells are preferential targets for bystander activation and apoptosis. *J. Immunol.* **182**, 1962–1971 (2009).
28. White, J. T., Cross, E. W. & Kedl, R. M. Antigen-inexperienced memory CD8$^+$ T cells: where they come from and why we need them. *Nat. Rev. Immunol.* **17**, 391–400 (2017).
29. Maurice, N. J., Taber, A. K. & Prlic, M. The ugly duckling turned to swan: a change in perception of bystander-activated memory CD8 T cells. *J. Immunol.* **206**, 455–462 2021).
30. Abers, M. S. et al. An immune-based biomarker signature is associated with mortality in COVID-19 patients. *JCI Insight* **6**, 144455 (2021).
31. Herndler-Brandstetter, D. et al. Post-thymic regulation of CD5 levels in human memory T cells is inversely associated with the strength of responsiveness to interleukin-15. *Hum. Immunol.* **72**, 627–631 (2011).
32. Jacomet, F. et al. Evidence for eomesodermin-expressing innate-like CD8$^+$ KIR/NKG2A$^+$ T cells in human adults and cord blood samples. *Eur. J. Immunol.* **45**, 1926–1933 (2015).
33. Jin, J.-H. et al. Virtual memory CD8$^+$ T cells restrain the viral reservoir in HIV-1-infected patients with antiretroviral therapy through derepressing KIR-mediated inhibition. *Cell. Mol. Immunol.* **17**, 1257–1265 (2020).
34. Yang, J., Zhu, H., Murphy, T. L., Ouyang, W. & Murphy, K. M. IL-18-stimulated GADD45β required in cytokine-induced, but not TCR-induced, IFN-γ production. *Nat. Immunol.* **2**, 157–164 (2001).
35. Kastenmüller, W., Torabi-Parizi, P., Subramanian, N., Lämmermann, T. & Germain, R. N. A spatially-organized multicellular innate immune response in lymph nodes limits systemic pathogen spread. *Cell* **150**, 1235–1248 (2012).
36. Wimmers, F. et al. The single-cell epigenomic and transcriptional landscape of immunity to influenza vaccination. *Cell* **184**, 3915–3935 (2021).
37. Debisarun, P. A. et al. Induction of trained immunity by influenza vaccination—impact on COVID-19. *PLoS Pathog.* **17**, e1009928 (2021).
38. Arunachalam, P. S. et al. Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* **369**, 1210–1220 (2020).
39. Schulte-Schrepping, J. et al. Severe COVID-19 is marked by a dysregulated myeloid cell compartment. *Cell* **182**, 1419–1440 (2020).
40. Pérez-Gómez, A. et al. Dendritic cell deficiencies persist seven months after SARS-CoV-2 infection. *Cell. Mol. Immunol.* **18**, 2128–2139 (2021).
41. 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). 1–7.
42. Utrero-Rico, A. et al. Alterations in circulating monocytes predict COVID-19 severity and include chromatin modifications still detectable six months after recovery. *Biomedicines* **9**, 1253 (2021).
43. You, M. et al. Single-cell epigenomic landscape of peripheral immune cells reveals establishment of trained immunity in individuals convalescing from COVID-19. *Nat. Cell Biol.* **23**, 620–630 (2021).
44. Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626–638 (2016).
45. Patel, A. A. et al. The fate and lifespan of human monocyte subsets in steady state and systemic inflammation. *J. Exp. Med.* **214**, 1913–1923 (2017).
46. Cheong, J.-G. et al. Epigenetic memory of COVID-19 in innate immune cells and their progenitors. Preprint at *bioRxiv* https://doi.org/10.1101/2022.02.09.479588 (2022).
47. Tough, D. F., Borrow, P. & Sprent, J. Induction of bystander T cell proliferation by viruses and type I interferon in vivo. *Science* **272**, 1947–1950 (1996).
48. Takahashi, T. et al. Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature* **588**, 315–320 (2020).
49. Scully, E. P. et al. Sex and gender differences in testing, hospital admission, clinical presentation, and drivers of severe outcomes from COVID-19. *Open Forum Infect. Dis.* **8**, ofab448 (2021).
50. Tsang, J. S. et al. Global analyses of human immune variation reveal baseline predictors of postvaccination responses. *Cell* **157**, 499–513 (2014).
51. Kotliarov, Y. et al. Broad immune activation underlies shared set point signatures for vaccine responsiveness in healthy individuals and disease activity in patients with lupus. *Nat. Med.* **26**, 618–629 (2020).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023

OP11 Clinical Staff

Princess Barber[9], Daly Cantave[9], Anne Carmona[9], Jean Hammer[10], Alaina K. Magnani[9], Valerie Mohammed[9], Cindy Palmer[9] & Deitra Shipman[9]

[9]Clinical Center Nursing Department, NIH, Bethesda, MD, USA. [10]Clinical Research Directorate, Frederick National Laboratory for Cancer Research, Frederick, MD, USA.

Methods

Patient population and sample collection

Participants aged at least 18 years were recruited between August and December 2020 from the local area (Maryland, Virginia and the District of Columbia) and enrolled on National Institutes of Health (NIH) protocol 19-I-0126 (Systems analyses of the immune response to the seasonal influenza vaccine). The study was approved by the NIH Institutional Review Board (ClinicalTrials.gov: NCT04025580) and complied with all relevant ethical regulations. Informed consent was obtained from all of the participants. After informed consent was obtained, a baseline history and physical examination were performed. The participants were asked to characterize any present, persistent symptoms of past SARS-CoV-2 infection. Exclusion criteria included obesity (BMI ≥ 30); history of or suspicion of any autoimmune, autoinflammatory or immunodeficiency disease; history of any vaccine within the past 30 days (live attenuated) or 14 days (non-live attenuated); history of any experimental vaccine; history of a parasitic, amoebic, fungal or mycobacterial infection in the past year; or current infection. The COVID-19 vaccine was not available at the time of the study, and no study participants participated in any COVID-19 vaccine trials. All study visits occurred at the NIH Clinical Center (CC) in Bethesda, MD, USA. Blood samples were collected by phlebotomy staff at the NIH CC. The samples were collected between September 2020 and April 2021. No sample size calculations were done prior to enrolment, in part because there were no reliable effect size estimates related to the impact of prior COVID-19 infection on vaccine responses. The number of subjects in the study was the number that were able to be recruited during the recruitment period.No blinding or randomization was performed.

Samples were collected from participants from three groups: (1) those with a previous history of symptomatic SARS-CoV-2 infection (defined as a history of a positive nasal PCR test and positive Food and Drug Administration (FDA) Emergency Use Authorization (EUA) SARS-CoV-2 antibody test at the time of protocol screening); (2) those with a history of asymptomatic SARS-CoV-2 infection (defined as testing positive using the FDA EUA SARS-CoV-2 antibody test at the time of the protocol exam, but with no history of COVID-like symptoms; no time since COVID-19 infection or diagnosis was identifiable for this group and they were excluded from all TSD analyses); and (3) individuals with no history of SARS-CoV-2 infection (defined as testing negative with the FDA EUA SARS-CoV-2 antibody test at the time of the protocol screening).

Blood for peripheral blood mononuclear cells (PBMCs), serum, whole-blood RNA (Tempus Blood RNA Tube, Thermo Fisher Scientific), complete blood count with differential (CBC) and lymphocyte phenotyping was collected at each of the following timepoints relative to seasonal influenza vaccination (day 0): days −7, 0, 1, 7, 14, 28, 70 and 100. Optional stool samples were collected at days 0, 28 and 100. The participants were provided with Cardinal Health Stool Collection kits (Cardinal Health) and Styrofoam storage containers with ice packs to collect stool samples at home and return in person to the NIH. After day 100, the participants had the option to continue to provide monthly blood samples for PBMCs, serum, whole blood RNA, CBC with differential and lymphocyte phenotyping through August 2021.

At each timepoint after study enrolment, data were collected and managed using the REDCap (v.8.5.27) electronic data capture tools hosted at the NIH[54,55]. REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies, providing (1) an intuitive interface for validated data capture; (2) audit trails for tracking data manipulation and export procedures; (3) automated export procedures for seamless data downloads to common statistical packages; and (4) procedures for data integration and interoperability with external sources. REDCap electronic questionnaires were used to collect information from the participants through two separate IRB-approved surveys. A survey to evaluate vaccine-related adverse events or symptoms was administered on study days 1 and 7 and a separate survey to evaluate for any health changes or new medications was administered at every visit starting on day 0. Surveys were sent by email to the participants and the responses were transferred from the REDCap system to the NIH Clinical Research Information Management System (CRIMSON) system by the study team.

Influenza vaccination

Participants aged between 18 and 64 years were administered the Flucelvax Quadrivalent seasonal influenza vaccine (2020–2021; Seqirus). Participants aged 65 years and older were administered the high-dose Fluzone Quadrivalent seasonal influenza vaccine (2020–2021; Sanofi Pasteur).

Influenza microneutralization titres

Virus-neutralizing titres of pre- and post-vaccination sera were determined in a microneutralization assay based on the methods of the pandemic influenza reference laboratories of the Centers for Disease Control and Prevention (CDC) using low-pathogenicity vaccine viruses and MDCK cells. The X-179A virus is a 5:3 reassortant vaccine containing the *HA*, *NA* and *PB1* genes from A/California/07/2009 (H1N1pdm09) and the five other genes from A/PR/8/34 were donated by the high-growth virus NYMC X-157. Immune sera were also tested for neutralization titres of the seasonal vaccine strains H1N1 A/Brisbane/59/07, H3N2 A/Uruguay/716/07 and B/Brisbane/60/2001. Internal controls in all of the assays were sheep sera generated against the corresponding strains at the Center for Biologics Evaluation and Research, FDA. All individual sera were serially diluted (twofold dilutions starting at 1:10) and were assayed against 100 median tissue culture infectious dose of each strain in duplicates in 96-well plates (1:1 mixtures). The titres represent the highest dilution that completely suppressed virus replication.

SARS-CoV-2 pseudovirus production and neutralization assay

Human codon-optimized cDNA encoding SARS-CoV-2 S glycoprotein (GenBank: NC_045512) was cloned into eukaryotic cell expression vector pcDNA 3.1 between the BamHI and XhoI sites. Pseudovirions were produced by co-transfection of Lenti-X 293T cells with psPAX2(gag/pol), pTrip-luc lentiviral vector and pcDNA 3.1 SARS-CoV-2-spike-deltaC19, using Lipofectamine 3000. The supernatants were collected at 48 h after transfection and filtered through 0.45 μm membranes and titrated using 293T-ACE2 cells (HEK293T cells that express ACE2 protein). The following reagent was obtained through BEI Resources, NIAID, NIH: human embryonic kidney cells (HEK293T) expressing human angiotensin-converting enzyme 2, HEK293T-hACE2 cell line, NR-52511.

For the neutralization assay, 50 μl of SARS-CoV-2 S pseudovirions were pre-incubated with an equal volume of varying dilutions of serum at room temperature for 1 h, then virus–antibody mixtures were added to 293T-ACE2 cells in a 96-well plate. After incubation for 3 h, the inoculum was replaced with fresh medium. After 24 h, cells were lysed and luciferase activity was measured as previously described[56–58]. Controls included a cell-only control, virus without any antibody control and positive control sera. Lenti-X 293T cells were obtained from Takara Bio (Cat. No. 632180). 293T-ACE2 cells were obtained from ATCC. The 293T-ACE2 cells were checked for expression of ACE2 and validated by FACS analysis. Neither of the cell lines was authenticated by karyotyping or other genomic techniques. Both cell lines tested negative for *Mycoplasma*.

SPR-based antibody binding kinetics of human serum

Steady-state equilibrium binding of serum was monitored at 25 °C using the ProteOn Surface Plasmon Resonance (BioRad) system as previously described[59–61]. The purified recombinant SARS-CoV-2 or other proteins were captured to a Ni-NTA sensor chip (BioRad, 176-5031) with 200 resonance units (RU) in the test flow channels. The protein

density on the chip was optimized such as to measure monovalent interactions independent of the antibody isotype. Serial dilutions (10-, 30- and 90-fold) of freshly prepared sample in BSA-PBST buffer (PBS pH 7.4 buffer with Tween-20 and BSA) were injected at a flow rate of 50 µl min^{-1} (120 s contact duration) for association, and disassociation was performed over a 600 s interval. Responses from the protein surface were corrected for the response from a mock surface and for responses from a buffer-only injection. Total antibody binding was calculated using the BioRad ProteOn manager software (v.3.1). All SPR experiments were performed twice, and the researchers performing the assay were blinded to sample identity. Under these optimized SPR conditions, the variation for each sample in duplicate SPR runs was <5%. The maximum resonance units (max RU) data shown in the figures were the RU signal for the tenfold-diluted serum sample.

PBMC isolation

PBMC samples were isolated from blood collected in Vacutainer EDTA tubes (generic laboratory supplier) using SepMate-50 tubes (STEMCELL Technologies) with the following modifications to the manufacturer's protocol: the blood samples were diluted 1:1 with room temperate PBS and mixed by pipetting. The diluted blood was layered on top of a 15 ml Cytiva Ficoll PAQUE-Plus (Cytiva Life Sciences) layer in the SepMate tube. The SepMate tubes were centrifuged at 1,200*g* for 10 min with brake set to 5 at room temperature. After centrifuging, the top plasma layer was removed as much as possible without disturbing the PBMC layer. If there were any cells stuck on the wall of the tube, they were gently scraped from the wall using a pipette so they could be resuspended with the rest of the cells. The cells were poured from the SepMate tube in to a 50 ml conical tube. The tubes containing cells were filled up to 50 ml with cold wash buffer (PBS with 2% FBS) and mixed by inverting. The tubes were centrifuged at 300*g* for 10 min with brake set to 5 at room temperature. After centrifuging, the supernatant was removed without disturbing the cell pellet. After resuspending the pellet with cold wash buffer, the cells were counted using the Guava Muse Cell Analyzer (Luminex). The tubes were again centrifuged at 300*g* for 10 min with brake set to 5 at room temperature. The supernatant was removed without disturbing the cell pellet.

On the basis of the cell count, 6–10 million PBMCs were frozen per vial for each sample. As the cells were counted before the last centrifuging, a 50% cell loss was assumed and accounted for in the calculations from cell count. The cell pellet was resuspended with $n \times 600$ µl (where n is the number of PBMC vials to be frozen) freezing medium (RPMI with 10% FBS) by gentle pipetting. After freezing the medium, $n \times 600$ µl DMSO freeze (FBS with 15% DMSO) was added drop-by-drop while gently shaking the tube. In other words, for each vial of PBMC that was to be frozen, 600 µl of freezing medium and 600 µl of DMSO freeze was added, bringing the total volume for each vial to 1.2 ml. The solution was gently mixed by pipetting before transferring 1.2 ml cell solution to each 1.8 ml cryovial (general laboratory supplier). The cell vials were placed into CoolCell Containers (Thomas Scientific) and the container was placed into a −80 °C freezer. After at least 4 h, the PBMC vials were transferred to liquid nitrogen.

RNA isolation

Blood was drawn directly into the Tempus Blood RNA Tube (Thermo Fisher Scientific) according to the manufacturer's protocol. Two Tempus tubes were collected at each study timepoint. The blood sample from each Tempus tube was aliquoted into two 4.5 ml cryovials (general laboratory supplier). These cryovials were directly stored at −80 °C.

The RNA samples were isolated in groups of 12–22 samples per batch based on careful batching before isolation to reduce confounding factors due to age, gender and patient group.

RNA was isolated from blood in the Tempus tube using the QIAsymphony RNA Kit (Qiagen) using the QIAsymphony SP instrument (Qiagen). Blood samples were thawed on ice before each sample was transferred to a 50 ml conical tube. The total volume of the sample was brought to 12 ml by adding 1× PBS. The tubes were vortexed at full speed for 30 s, followed by centrifugation at 3,500*g* for 1 h at 4 °C. After centrifugation, the supernatant from the tubes was decanted and the tubes were placed upside-down on clean paper towels for 2 min to allow residual liquid to drain. To resuspend the pellet, 800 µl of RLT+ buffer was added to the bottom of each tube and vortexed for few seconds. All 800 µl of each sample was transferred to 2 ml screw cap tubes (Sarstedt). The tubes were placed into #3b adapters (Qiagen) to be loaded onto the QIAsymphony system.

On the QIAsymphony system, the RNA CT 800 protocol was selected and used for RNA isolation. The instrument was set up according to the manufacturer's protocol and the elution volume for RNA samples was set to 100 µl. The final volume of the eluted RNA samples ranged from 65 µl to 95 µl.

RNA yields were determined using the Qubit RNA BR kit or Qubit RNA HS kit (Thermo Fisher Scientific) on the basis of the yield. RNA RIN numbers were measured using RNA ScreenTape (Agilent Technologies). The average RIN was 8.3 and the average yield was 81.3 ng $µl^{-1}$ for the RNA samples.

RNA-seq

RNA-seq libraries were prepared manually using Universal Plus mRNA-Seq with NuQuant, Human Globin AnyDeplete (Tecan Genomics) according to the manufacturer's protocol. For each sample, 500 ng of total RNA was used to isolate mRNA by poly(A) selection. Captured mRNA was washed, fragmented and primed with the mix of random and oligo(dT) primers. After cDNA synthesis, ends were repaired and ligated with unique dual index adaptor pairs. Unwanted abundant transcripts from rRNA, mtRNA and globin were removed using AnyDeplete module. The remaining library was amplified by 14 cycles of PCR and purified with AMPure XP reagent (Beckman Coulter).

Library concentration was determined using the Quant-iT PicoGreen dsDNA Assay kit (Thermo Fisher Scientific) on the BioTek Synergy H1 plate reader (BioTek Instruments) using 2 µl sample. Library size distribution was determined using D1000 ScreenTape (Agilent Technologies) on the 4200 TapeStation System (Agilent Technologies). A total of 32 samples were randomly selected from each plate to measure the library size distribution. To determine fragment size, the region on the electropherogram was set from 200 bp to 700 bp. An average of the fragment sizes was used for the rest of libraries to calculate the molarity.

To create a balanced pool for sequencing, all of the libraries from one plate were diluted to the same molar concentration using the QIAgility liquid handling robot (Qiagen), and equal volumes of normalized samples were pooled. A total of 96 samples were pooled from each plate on plates 1–4 and 35 samples were pooled from plate 5. For an accurate quantification of the pooled libraries, quantitative PCR was performed using the KAPA Library Quantification Kit (Roche).

All of the libraries were sequenced on the NovaSeq 6000 instrument (Illumina) at the Center for Cancer Research Sequencing Facility, National Cancer Institute. The libraries pooled from plates 1–4 were sequenced using one NovaSeq 6000 S4 Reagent Kit (200 cycles) and NovaSeq XP 4-Lane Kit (Illumina) with 100 bp paired-end reads as the sequencing parameter. The library pool from plate 5 was sequenced using the NovaSeq 6000 SP Reagent Kit (300 cycles; Illumina) with 150 bp paired-end reads as the sequencing parameter.

Moreover, after quality control, 11 samples were resequenced as plate 6 on a NextSeq 500 instrument using a NovaSeq 6000 S4 Reagent Kit (200 cycles) with 100 bp paired-end reads as the sequencing parameter. Technical replicates were placed on each plate to control for plate variability.

CITE-seq

Single-cell CITE-seq processing. Frozen PBMC samples were thawed, recovered and washed using RPMI medium with 10% FBS and 10 mg ml^{-1}

DNase I (STEMCELL) and then processed as previously described[15] for CITE-seq staining. In brief, samples from different donors were pooled and different timepoints from the same donor were pooled separately so that each pool contains only one timepoint from one donor. PBMC pools were Fc blocked (Human TruStain FcX, BioLegend) and stained with Totalseq-C human 'hashtag' antibodies (BioLegend), washed with CITE-seq staining buffer (2% BSA in PBS). Hashtagged PBMC pools were then combined, and cells were stained with a cocktail of TotalSeq-C human lyophilized panel (BioLegend) of 137 surface proteins (including 7 isotype controls; Supplementary Table 11) and SARS-CoV-2 S1 protein probe. Cells were then washed, resuspended in PBS and counted before proceeding immediately to the single-cell partition step.

Single-cell CITE-seq library construction and sequencing. PBMC samples were partitioned into single-cell gel-bead in emulsion (GEM) mixed together with the reverse transcription mix using the 10x 5′ Chromium Single Cell Immune Profiling Next GEM v2 chemistry kit (10x Genomics), as previously described[15]. The reverse transcription step was conducted in the Veriti Thermal Cycler (Thermo Fisher Scientific). Single-cell gene expression, cell surface protein, T cell receptor (TCR) and B cell receptor (BCR) libraries were prepared according to the 10x Genomics user guides (https://www.10xgenomics.com/resources/user-guides/). All libraries were quality-controlled using the Bioanalyzer (Agilent) and quantified using Qubit fluorometric quantification (Thermo Fisher Scientific). 10x Genomics 5′ single-cell gene expression, cell surface protein tag, TCR and BCR libraries were pooled and sequenced on the Illumina NovaSeq platform (Illumina) using the following sequencing parameters: read1-100-cycle, i7-10-, i5-10, read2-100.

Serum isolation and protein characterization

Serum was collected directly in serum separator tubes and allowed to clot at room temperature for a minimum of 30 min. Within 2 h of blood collection, the tubes were centrifuged 1,800*g* for 10 min at room temperature. The top (serum) layer was removed using a pipette and stored in individual vials at −80 °C. Serum proteins were analysed using the Olink Target 96 Immuno-Oncology and Olink Target 96 Inflammation panels (Olink Proteomics, Uppsla, Sweden), which comprised 92 proteins each and uses the methodology based on the proximity extension assay. Data are reported as normalized protein expression (NPX) unit.

CBCs and lymphocyte phenotyping

For the participants, standard complete blood counts with differential (CBCs) were performed at the NIH CC in the Department of Laboratory Medicine. Lymphocyte (T cell, B cell, NK cell) flow cytometry quantification was performed using the BD FACSCanto II flow cytometer (BD Biosciences).

PBMC in vitro stimulation

PBMCs were thawed and cultured in PRMI1640 containing 10% fetal bovine serum, 2 mM glutamine, 0.055 mM beta-mercaptoethanol, 1% penicillin–streptomycin, 1 mM sodium pyruvate, 10 mM HEPES and 1% non-essential amino acids, and stimulated under the following conditions: (1) IL-15 (10 ng ml^{-1}), IL-12 (20 ng ml^{-1}), IL-18 (20 ng ml^{-1}) for 48 h; (2) IL-15 (50 ng ml^{-1}) for 48 h; (3) IL-18 (50 ng ml^{-1}) for 48 h; (4) IL-12 (20 ng ml^{-1}), IL-18 (20 ng ml^{-1}) for 48 h; (5) anti-CD3 (1 μg ml^{-1}), anti-CD28 (1 μg ml^{-1}) for 24 h; (6) non-stimulated controls. Protein Transport Inhibitor (BD Biosciences, 554724) and Brefeldin A (BFA, Invitrogen, 00-4506-51) were added 4 h before collection. The following cytokines were purchased from BioLegend: IL-15 (570304), IL-12 (573004) and IL-18 (592104).

Flow cytometry

B cell phenotyping panel including influenza HA probes. Thawed PBMCs were washed in RPMI culture medium containing 50 U ml^{-1} benzonase nuclease and then washed with PBS. Cells were incubated with LIVE/DEAD Fixable Blue Dye (Life Technologies), which was used to exclude dead cells from analysis. Cells were incubated with fluorochrome-conjugated HAs for influenza B (B/Washington/02/2019 and B/Phuket/3073/2013 combined on the same fluorochrome), and influenza A H1 (A/Hawaii/70/2019) and H3 (A/Hongkong/2671/2019) and fluorochrome-conjugated antibodies against IgM, IgA, CD21, CD85J, FCRL5, CD20, IgG, CD38, CD14, CD56, CD3, CD27, CD71, CD19 and IgD for 30 min at 4 °C in the dark. The dyes and detailed information of antibodies in the panel (Sarah Andrews, Vaccine Research Center, National Institute of Allergy and Infectious Diseases, NIH) are summarized in Supplementary Table 12. After incubation with antibodies for 30 min, cells were washed twice with FACS buffer (0.1% BSA/PBS (pH 7.4)) and fixed in 1% paraformaldehyde. A total of 5 million cells were acquired on the Cytek Aurora spectral cytometer (Cytek Biosciences; SpectroFlo (v.2.2.0)). Data were analysed using FlowJo (v.10; BD Biosciences).

General immune phenotyping panel. Thawed PBMCs were washed in RPMI culture medium containing 50 U ml^{-1} benzonase nuclease and then washed with PBS. Cells were incubated with LIVE/DEAD Fixable Blue Dye (Life Technologies), which was used to exclude dead cells from analysis. Cells were washed in FACS staining buffer (1× PBS, 0.5% fetal calf serum, 0.5% normal mouse serum and 0.02% NaN_3) and incubated with Human Fc block reagent (BD Bioscience, 564220) at room temperature for 5 min. Cells were stained at room temperature for 10 min in the dark with fluorochrome-conjugated antibodies against CCR7, CCR6, CXCR5, CXCR3 and TCRgd. Cells were then stained with fluorochrome-conjugated antibodies against CD45RA, CD16, CD11c, CD56, CD8, CD123, CD161, IgD, CD3, CD20, IgM, IgG, CD28, PD-1, CD141, CD57, CD45, CD25, CD4, CD24, CD95, CD27, CD1c, CD127, HLA-DR, CD38, ICOS, CD21, CD19 and CD14 at room temperature for 30 min in the dark. Cells were washed twice with FACS staining buffer (1× PBS, 0.5% fetal calf serum, 0.5% normal mouse serum and 0.02% NaN_3) and fixed in 1% paraformaldehyde. Supplementary Table 13 shows the clones and information of the antibodies used in the phenotyping panel. A total of 1 million PBMCs were acquired using the Cytek Aurora spectral cytometer (Cytek Biosciences; SpectroFlo (v.2.2.0)). The frequency of major populations was analysed using FlowJo (v.10; BD Biosciences) on the basis of previously described manual gating strategies[62–64].

In vitro stimulation T cell panel. In vitro simulated PBMCs were collected and washed in PBS. Cells were incubated with Zombie UV Fixable Viability Dye (BioLegend) in the dark (at room temperature) for 20 min. Cells were then washed and incubated with Human TruStain FcX (BioLegend) for 10 min and subsequently with anti-CCR7 antibodies for 10 min. A cocktail of fluorochrome-conjugated antibodies against CD8, CD4, HLA-DR, CD69, CD45RA, CD11c, CD5, CD3, TCRVa7.2, CD45RO, CD56, CD122, CD158e/k (KIR3DL1/DL2), KIR2D, NKG2A, CD14, CD29 and GPR56 was added and cells were stained for 30 min in the dark. Cells were washed and fixed using the Fixation/Permeablization kit (BD Biosciences). The intracellular proteins IFNγ, TNF and Ki-67 were stained after fixation. The samples were collected using the BD FACSymphony flow cytometer (BD Biosciences) and analysed using FlowJo (v.10). A list of the antibodies used in the panel is provided in Supplementary Table 14.

Data processing and transformation

Bulk RNA-seq data processing. Sequencing reads from plate 5 were adaptor- and quality-trimmed to 100 bp using Trimmomatic (v.0.38.0)[65] to match the read length of the other plates (resulting reads with less than 100 bp were discarded). Reads were then aligned to the human genome hg38 using the STAR (v.2.6.0b) aligner. Duplicate reads from PCR amplification were removed based on unique molecular identifiers using UMI-tools (v.0.5.3). Gene expression quantification was

performed using the featureCounts[66] function from Subread package (v.1.6.2). Samples with less than 5 million assigned reads were resequenced and replaced. Reads were normalized and log transformed using limma voom[67]. Low-expressed genes, defined as having fewer than five samples with >0.5 counts per million reads, were removed. Prevaccination (days −7 and 0) samples from the same healthy control participants were considered to be replicates and were used to estimate latent technical factors using the RUVs function of the RUVSeq[68] R package (v.1.18). Four latent variables were included to derive normalized gene expression values used for visualization and when specifically noted. Variable genes based on intraparticipant variability of prevaccination samples in the healthy control individuals and across technical replicates were filtered out, resulting in a total of 10,017 remaining genes for downstream analyses.

CITE-seq data processing. Single-cell sample demultiplexing and preprocessing. Single-cell sequencing data were demultiplexed, converted to FASTQ format, mapped to the human hg19 reference genome and counted using the CellRanger (10x Genomics) pipeline. The sample-level demultiplex was performed based on two levels as previously described[15]: (1) hashtag antibody staining to distinguish different timepoint samples from a same participant; (2) single-nucleotide polymorphisms (SNPs) called from the whole-blood RNA-seq data to identify different participants. Specifically, CellRanger (v.6.0.1) was used for generating count matrix and the software package demuxlet (v.2, from the popscle software suite)[69] was used to match single-cell gene expression data to each donor and identify empty droplets and doublets.

Single-cell data clustering and cell annotation. Single-cell data were further processed using Seurat (v.4.0.3) running in R v.4.1.1. We removed cells with less than 200 and greater than 5,000 detected genes; greater than 60% of reads mapped to a single gene; greater than 15% mitochondrial reads; cell surface protein tag greater than 20,000; and hashtag antibody counts greater than 20,000. The protein data were normalized and denoised using the DSB method (v.0.3.0)[70]. The following parameters were used in the dsb normalization function: define.pseudocount = TRUE, pseudocount.use = 10, denoise_counts = TRUE, use.isotype.control = TRUE. The DSB-normalized protein data were used to generate the top variable features (n = 100) and principal components (PCs). The shared nearest neighbour graph followed by k-nearest neighbours clustering were then built using the FindNeighbors and FindClusters functions using the first 15 PCs in Seurat (v.4.0.3), respectively. Cell clusters were quality-controlled on the basis of their nearest neighbours and cell surface proteins. Cells were then further clustered within each major cell population using weighted-nearest neighbour (WNN) analysis in Seurat[71] (v.4.1.0) by integrating both cell surface protein and gene expression modalities. WNN FindMultiModalNeighbors was performed using both the top 10 PCs for cell surface protein and RNA of variable features. The WNN clusters were manually annotated and quality-controlled using the surface protein together with gene expression.

CD8$^+$ EM cell annotation for CITE-seq clusters. All CD8$^+$ cells were clustered using WNN as described above. CD8$^+$ clusters were annotated on the basis of their surface markers as reported[72] together with gene expression profile. RNA expression of CD8$^+$ cells was mapped to an external dataset using the Seurat Label transfer method[71,73] (v.4.1.0). Clusters annotated as CD8$^+$ EM are surface CD45RO^{high}, CD45RA^{low}, CD95$^+$, CD62L^{low} and *CCR7* mRNA$^-$ with most cells (around 90%) mapped to CD8$^+$ EM phenotype cells in an external dataset[71,73].

Single-cell TCR data processing. CellRanger (v.6.0.1) was used to assemble V(D)J contigs. The V(D)J assignment and clonotype were from the CellRanger output of the filtered contig_annotations.csv file for each 10x lane. The data were combined for all lanes and paired TCRα and TCRβ chains for each single cell were combined using the scRepertoire R package (v.1.4.0)[74] and integrated with the single-cell CITE-seq Seurat object metadata. Cells annotated as CD8$^+$ T cells and with both α and β chains detected were filtered and analysed. CD8$^+$ subsets and GPR56$^+$CD8$^+$ EM cell clonality were visualized by Circos plots using the Circlize R package (v.0.4.14)[75]. For the purpose of visualization, cells from each subset were downsampled with equal numbers in each subset (a comparison between subsets is shown in Extended Data Fig. 4g) or in each timepoint (a comparison between timepoints is shown in Extended Data Fig. 4h,i). Cells were considered to be the same clone when they had identical CDR3 (both α and β chains). Identical clones were connected within each sample or each participant across timepoints with lines.

OLINK serum proteomics. Missing values were imputed using the k-nearest neighbours approach with k = 10 using the impute R package[76] (v.1.60.0). For each sample, probes targeting the same protein were averaged.

Cytek flow cytometry. Cell frequencies were generated by converting cell counts as fraction of live cells or lymphocytes as specified. The frequency data were $\log_2$-transformed for linear modelling. For populations with zero counts in any of the samples, an offset equal to half of the smallest non-zero value was added across samples.

CBCs and lymphocyte phenotyping. Both absolute and relative counts were $\log_2$-transformed for linear modelling. Missing values were imputed using k-nearest neighbours approach. For parameters with zero values in any of the samples, an offset equal to half of the smallest non-zero value was added across samples.

Statistical analysis

Baseline differential expression analysis. Using the dream[77] function in the variancePartition R package (v.1.16.1), mixed-effects models were applied to determine differential levels of analytes (that is, whole-blood gene expression, serum proteins, cell frequencies, flu titre and SPR, and haematological parameters) between participants who had recovered from COVID-19 and healthy control participants in a sex-specific manner as follows: ~ 0 + group:sex + age + race + batch. effects + (1|participant.ID).

Batch-effect-related covariates were added to specific models depending on the assay type. For bulk RNA-seq, these include the four latent technical factors (see the 'Bulk RNA-seq data processing' section) and the timepoint-matched % neutrophils parameter from the CBC panel. For the Cytek and Olink platforms, sampling batch/plate was included as covariates. In addition to day 0, available samples from day −7 (in the RNA-seq and CBC panel) were included as baseline replicates in the modelling.

Sex-specific group differences were computed from the contrasts covid.Female − healthy.Female and covid.Male − healthy.Male. Overall COVID-19 versus healthy control difference was determined by combining the two contrasts, that is, (covid.Female − healthy.Female)/2 + (covid.Male − healthy.Male)/2. Sex difference linked to SARS-CoV-2 infection was derived from the contrast (covid.male − covid.female) − (healthy.male − healthy.female) to account for normal differences between males and females. P values were adjusted for multiple testing within each assay type and contrast combination using the Benjamini–Hochberg method[78].

Association with TSD. To evaluate whether any of the differences detected at baseline had stabilized or might still be resolving, a linear model was used to test the association of relevant parameters with the time since COVID-19 diagnosis (TSD) among participants who had recovered from COVID-19: ~ 0 + sex + sex:scale(TSD) + age + race + (1|participant.ID).

Two asymptomatic participants without a known TSD were excluded from the model. Association was assessed separately for female and male individuals, and jointly by the combined contrast

(female:TSD + male:TSD)/2. Dependent variables were converted to ranks in the model to reduce the effect of potential outliers.

Using a conservative approach, genes were classified as TSD-associated if they had an unadjusted $P < 0.05$ and were excluded from subsequent analyses as specified. To determine whether any of the baseline differential gene sets were associated with TSD, LEG modules were derived from the union of all LEGs of the same gene set from different contrasts (see the 'Bulk RNA-seq gene set module scores' section). A gene set was considered to be stable if none of three contrasts tested in the association model were significant (using an unadjusted P value threshold of 0.05).

Post-vaccination differential expression analysis. Similar to the workflow used in the baseline differential expression analysis, mixed-effects models were created to evaluate changes and group differences at each available timepoint after vaccination. Participants aged 65 and above were excluded as they received a different type of vaccine. In addition to the baseline covariates, the model also accounts for the participants' flu vaccination history within last 10 years as follows: ~ 0 + visit:group: sex + age + race + flu.vax.count.10yr + batch.effects + (1|participant.ID).

Three types of comparisons were examined using this model:

Timepoint-specific group differences: similar to the contrasts in the baseline model, but for individual timepoints post vaccination (day 1 to day 100).

Vaccine-induced changes in group difference: similar to the timepoint-specific contrasts above, but additionally subtracting off the corresponding baseline contrast to assess the changes relative to the baseline. For example, the differences in vaccine-induced changes for female individuals with COVID-19 versus healthy control individuals at D1 is evaluated with the contrast: (D1.covid.female − D1.healthy.female) − (baseline.covid.female − baseline.healthy.female).

Reversal of COVID-19 versus healthy control difference: instead of using the healthy control participants at the same corresponding timepoints as the reference, post-vaccination samples from the participants who had recovered from COVID-19 were compared to baseline healthy control individuals with the contrasts [timepoint].covid.female − baseline.healthy.female and [timepoint].covid.male − baseline.healthy.male. These contrasts can inform whether any prevaccination differences observed in the participants who had recovered from COVID-19 were reverted towards healthy baseline levels after vaccination. Reversal is defined as having a smaller absolute effect size (using the z.std value from the dream function) at D1 and D28 after vaccination compared with the baseline absolute effect size.

P values were adjusted for multiple testing for each timepoint, assay type and contrast combination using the Benjamini–Hochberg method.

Gene set enrichment of differentially expressed genes. Enriched gene sets were identified using the preranked GSEA algorithm implemented in the clusterProfiler R package (v3.17.0)[79]. Genes were ranked using signed $-\log_{10}$-transformed P values from differential expression models. Enrichment was assessed with gene set lists from MSigDB's Hallmark collection[80], blood transcriptomic modules[81] and cell type gene signatures[53]. Only gene sets with 10 to 300 genes were considered. P values were adjusted per gene set list for each contrast using the Benjamini–Hochberg method and gene sets with FDR-adjusted $P < 0.05$ were considered to be significant. Baseline enriched gene sets were derived by intersecting significant gene sets extracted from differential expression models using samples independently from day −7, day 0, and both days combined. Genes associated with TSD at the baseline (see the 'Association with TSD' section; Supplementary Table 1) were excluded from the post-vaccination enrichment analyses to help to segregate the effect of vaccination from the natural temporal resolution of the SARS-CoV-2 infection.

Pseudobulk differential expression and GSEA. Single cells from a given sample were computationally pooled according to their cell type assignment by summing all reads for a given gene. Pseudobulk libraries that were made up by few cells and were therefore probably not modelled properly using bulk differential expression methods were removed from the analysis for each cell type to remove samples that contained fewer than 4 cells and with less than 35,000 library size after pooling. Low-expressed genes were removed for each cell type individually using the filterByExpr function of edgeR (v.3.26.8)[82] with min.count = 2. Log-transformed counts per million (CPM) of each gene were calculated with scaling factors for library size normalization provided by the calcNormFactors function. Differential expression analysis was performed using the same models described in the 'Post-vaccination differential expression analysis' section without running baseline models separately because the entire CITE-seq cohort was aged under 65 years. Batch assignment and the number of barcodes/cells per sample were included as batch effects in this model.

Similarly, GSEA was performed for each cell type in the same manner as described for the bulk RNA-seq data (see the 'Gene set enrichment of differentially expressed genes' section), which particularly focuses on the baseline enriched gene sets identified by the bulk RNA-seq analysis. The Monaco gene sets were excluded from the single-cell analysis given the cell clusters were annotated and no further cell type demultiplex needed.

Bulk RNA-seq gene set module scores. Gene set module scores were generated from RUVseq (v.1.18) normalized gene expression values (see the 'Bulk RNA-seq data processing' section) using the gene set variation analysis (GSVA) method in the GSVA R package (v.1.30.0)[83]. LEG module scores representing enriched pathway activities were calculated for relevant samples using LEGs identified by GSEA to enhance the signal-to-noise ratio. The average scores between days −7 and 0 were used for calculating post-vaccination changes relative to the baseline.

Pseudobulk gene set module score calculation. Module scores (gene set signature score) representing enriched pathway activities were calculated for each pseudobulk sample of certain cell types. The pseudobulk gene counts were corrected using the removeBatchEffect function in the limma package (v.3.42.2) to remove experimental batch and cell number effects and then normalized with voom[84]. The scores were then generated using the gene set variation analysis (GSVA) method from the GSVA R package (v.1.42.0)[83]. Specifically, for monocyte signatures, LEGs of BTM modules M4.0 and M11.0 were identified by GSEA from the (1) D0.COVR-F versus D0.HC-F and (2) D0.COVR-M versus D0.HC-M models. The union of LEGs was used for the score calculation for female and male samples.

For the BTM-M7.3 T cell activation signature and other signatures from acute COVID-19 data as indicated in the figures, LEGs were used from the indicated comparison groups for the score calculation of female and male individuals separately.

For the monocyte antigen presentation signature, the module score was generated using LEGs from the BTM-M71 enriched in antigen presentation (I) and M95.0 enriched in antigen presentation (II) gene sets of the comparison: D1 − D0 change between the COVR-M versus HC-M groups (Fig. 2f).

For the Hallmark IFNγ response module score, all genes from the gene set were used for calculating module scores in each cell type, so that the differences between cell types could be compared.

Single-cell module score calculation and visualization. To visualize the difference between subject groups in certain gene signatures using single data, the genes from the indicated gene sets were used to calculate the corresponding module score of each single cell. Module scores were calculated using the AddModuleScore function in Seurat

(v.4.1.0) and then visualized in UMAP plots. For D1 versus D0 Hallmark IFNγ response module score difference (D1−D0) shown in UMAP projections (Fig. 2d), cells from the D1.HC-F, D1.COVR-F, D1.HC-M and D1.COVR-M groups were downsampled to the same number of cells. The UMAP embeddings of cells coloured with the average difference (D1−D0) of each high-resolution cell subsets are shown (each of the major cell clusters shown can contain one or more high-resolution cell subsets).

Single-cell module score calculation and test of external acute COVID-19 single-cell CITE-seq data. Single-cell data from the Brescia cohort of ref. [15] were downloaded from the Gene Expression Omnibus (GEO). Single monocyte data were extracted and single-cell data from the Brescia cohort were pooled as described in the 'Pseudobulk differential expression and GSEA' section. The gene set module scores of BTM modules M4.0 and M11.0 for all of the samples were generated using the union LEGs of male and female in the 'Gene set module score calculation' section. The pseudobulk gene counts were normalized using the varianceStabilizingTransformation function of the DEseq2 R package (v.1.34.0)[85]. The scores were then generated using the GSVA method from the GSVA R package (v.1.42.0)[83]. Given there are multiple samples from each participant, the differences between patient groups (healthy control, less severe and more severe, corresponding to HC, DSM-low and DSM-high in ref. [15]) were tested using the Limma (v.3.50.1) linear model, where samples from the same donors were treated as duplicates using duplicateCorrelation. *P* values of *t* statistics from the linear model of the indicated contrasts are shown.

Visualization of gene expression in heat maps. Heat maps showing pseudo-bulk data were generated using the ComplexHeatmap R package (v.2.10.0)[86]. The log[CPM]-normalized expression for each sample for a given cell type was calculated by pooling cells as described in the 'Pseudobulk differential expression and GSEA' section. Heat maps show the *z*-score of the normalized expression for each gene in each sample.

Data visualization. Plots were created using ggplot2 (v.3.3.5) with ggpubr (v.0.4.0) for statistical calculation unless noted.

End-point association. To evaluate the association of relevant parameters, including gene set module scores and cell frequencies, with IFN or antibody titre fold change end points, the following model was applied: end point ~ group:sex + scale(parameter):group:sex + age + race + flu.vax.count.10yr.

The end-point values were converted to rank to reduce the effects of potential outliers. Replicates from the same participants were averaged.

Serology. Influenza antibody titres below the detection limit of 1:20 were set to 1:10. Maximum titre across strains was calculated by normalizing titre levels across all of the samples from both D0 and D28 individually for each of the four strains followed by taking the maximum standardized titre for each sample.

Baseline titre difference analysis. For each of the four strains, a linear model was applied to determine the baseline titre differences between participants who had recovered from COVID-19 and healthy control participants in a sex-specific manner as follows: day 0 titre ~ group:sex + age + race.

Titre values were $\log_{10}$-transformed in the model, and sex-specific group differences were computed from the contrasts covid.Female − healthy.Female and covid.Male − healthy.Male. Participants aged 65 and above were excluded from the analysis.

D28 titre difference analysis. For post-vaccination titre response, influenza vaccination history and baseline titre were included as covariates to partly account for previous exposure, similar to the approach used for influenza vaccine evaluation by the Food and Drug Administration (page 27 of https://www.fda.gov/media/135687/download). Both D28 titre and D28/D0 FC were evaluated as end points to determine group differences between participants who had recovered from COVID-19 and healthy control participants for each of the four strains: endpoint ~ group:sex + age + race + flu.vax.count.10yr + day 0 titre.

For D28 FC, a negative binomial model with log link was applied using the glm.nb function in the MASS R package (v.7.3-53). A linear model was used to fit the D28 titres. Strain-specific titre values were $\log_{10}$-transformed in the model. Group differences were assessed using the same participants and contrasts as in the baseline analysis.

Influenza antibody avidity as measured using SPR was analysed in the same manner as the titre data across HA1 and HA2, with the exception that that a linear model was applied for the fold changes.

Concordance in the natural influenza infection cohorts. A prospective cohort study with participants profiled before and at least 21 days after natural influenza infection in two seasons[18] was used to assess the residual effects of the infection separately in male and female individuals. Gene expression data were downloaded from the GEO (GSE68310). Participants with only influenza A virus infection ($n = 51$ female and $n = 35$ male) were identified and included for this analysis. Low-expressed probes were removed, and the remaining data were converted to gene-based expressions. No additional processing steps were performed as the data were already normalized.

Separately for each season, differential expression analysis between baseline (pre-infection) and spring (long-term post-infection) samples from the same individuals were performed using the dream function in the variancePartition R package (v.1.16.1). A mixed-effects model accounting for flu vaccination history and disease severity (based on fever grade: none, low and high) was constructed as follows: ~ 0 + timepoint:sex + age + num.flu.vaccination + fever.grade + (1|participant.ID).

Differentially expressed genes were identified using the contrasts Spring.F − Baseline.F and Spring.M − Baseline.M for female and male individuals, respectively. Sex difference was evaluated by the contrast (Spring.M − Baseline.M) − (Spring.F − Baseline.F). Concordance of differential expression results between the two seasons was evaluated on the basis of the correlation of effect size across genes (z.std values generated by dream).

Enrichment analysis was performed to determine whether the same set of genes was differentially expressed between pre- and post-influenza infection from this independent cohort and in participants who had recovered from COVID-19 compared with healthy control individuals before vaccination. To better match the age range of participants between the two studies, baseline differential gene analysis was performed again with participants under 65 years of age in the COVID-19 cohort (see the 'Baseline differential expression analysis' section). Given that the male participants showed stronger concordance between the two flu seasons (Extended Data Fig. 2b), COVID-19 differentially expressed genes were ranked by signed $-\log_{10}$-transformed *P* values and tested against a gene set formed by the intersect of differentially expressed ($P < 0.05$) genes in male participants from the flu infection cohort.

Elastic net multivariate predictive modelling. Elastic net models were constructed using the eNetXplorer R package (v.1.1.3)[87] to predict day 1 (D1) INFγ response after influenza vaccination with both CITE-seq and flow cytometry cell frequencies at D0 as predictors. A total of 33 participants (COVR-F = 11, HC-F = 8, COVR-M = 9, HC-M = 5) with both CITE-seq and flow cytometry data were included. On the basis of 20 runs of fivefold cross-validation, a grid of regularization parameters (α and λ) were tested to determine models with best performance and cell subsets with consistent predictive power. Model performance was assessed on the basis of the mean squared error between the

predicted and observed response. The importance of a cell population was determined by the frequency that it was selected by the models (that is, having non-zero coefficient). *P* values of the model performance and feature importance were derived by comparing to null models constructed with permuted response.

TCR diversity metric calculation. Shannon's entropy (*H*′ index) was calculated as a measure of TCR diversity[88,89]. Samples for each $CD8^+$ subset with fewer than 50 cells were filtered out from the calculation. All of the samples were downsampled to 50 cells because the diversity metric can be affected by the sample cell numbers. The process was repeated 1,000 times with random downsampling of 50 cells, and the median Shannon's index was used as an estimate of diversity for a given sample. Differences in the diversity metric between different $CD8^+$ subsets or timepoints were tested using two-tailed Wilcoxon tests.

Reversal genes and bootstrapping to infer the significance of difference in the reversal of monocyte-repressed signature between the COVR-F and COVR-M groups. Reversal genes are defined as those genes of which the COVID-19-recovered versus D0 healthy control absolute effect size (z.std values from dream; see the 'Post-vaccination differential expression analysis' section) are smaller at both D1 and D28 compared with at D0.

Bootstrapping was used to determine the significance of the difference between the COVR-F and COVR-M groups in their proportion of baseline LEGs from the monocyte-depressed signature (BTM M4.0 and M11.0) that moved towards the healthy control baseline. Members from each participant group were randomly sampled with replacement in each round of the bootstrapping and their samples were analysed as described in the 'Post-vaccination differential expression analysis' section. The proportion of LEGs reversed after vaccination was calculated in each round for the COVR-F and COVR-M groups in classical and non-classical monocytes, separately, and the *P* values plotted in Fig. 4e were determined on the basis of 20 rounds of this procedure.

Reporting summary

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

Data availability

Raw and processed data from the whole-blood bulk RNA-seq and single-cell CITE-seq are available from the NCBI GEO under accession numbers GSE194378 and GSE206265, respectively. Additional datasets, including clinical, proteomics, flow cytometry and influenza antibody measurements, are available at Zenodo (https://doi.org/10.5281/zenodo.5935845). The influenza infection dataset that we used was downloaded directly from the GEO (GSE68310).

Code availability

Analysis code, extended patient and sample metadata are available at GitHub (https://github.com/niaid/covid-flu).

52. Nakaya, H. I. et al. Systems biology of vaccination for seasonal influenza in humans. *Nat. Immunol.* **12**, 786–795 (2011).
53. Monaco, G. et al. RNA-seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. *Cell Rep.* **26**, 1627–1640 (2019).
54. Harris, P. A. et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **42**, 377–381 (2009).
55. Harris, P. A. et al. The REDCap consortium: building an international community of software platform partners. *J. Biomed. Inform.* **95**, 103208 (2019).
56. Ravichandran, S. et al. Antibody signature induced by SARS-CoV-2 spike protein immunogens in rabbits. *Sci. Transl. Med.* **12**, eabc3539 (2020).
57. Ravichandran, S. et al. Longitudinal antibody repertoire in 'mild' versus 'severe' COVID-19 patients reveals immune markers associated with disease severity and resolution. *Sci. Adv.* **7**, eabf2467 (2021).
58. Tang, J. et al. Antibody affinity maturation and plasma IgA associate with clinical outcome in hospitalized COVID-19 patients. *Nat. Commun.* **12**, 1221 (2021).
59. Khurana, S. et al. MF59 adjuvant enhances diversity and affinity of antibody-mediated immune response to pandemic influenza vaccines. *Sci. Transl. Med.* **3**, 85ra48 (2011).
60. Khurana, S. et al. Human antibody repertoire after VSV-Ebola vaccination identifies novel targets and virus-neutralizing IgM antibodies. *Nat. Med.* **22**, 1439–1447 (2016).
61. Khurana, S. et al. Repeat vaccination reduces antibody affinity maturation across different influenza vaccine platforms in humans. *Nat. Commun.* **10**, 3338 (2019).
62. De Biasi, S. et al. Marked T cell activation, senescence, exhaustion and skewing towards TH17 in patients with COVID-19 pneumonia. *Nat. Commun.* **11**, 3434 (2020).
63. Heit, A. et al. Vaccination establishes clonal relatives of germinal center T cells in the blood of humans. *J. Exp. Med.* **214**, 2139–2152 (2017).
64. Park, L. M., Lannigan, J. & Jaimes, M. C. OMIP-069: forty-color full spectrum flow cytometry panel for deep immunophenotyping of major cell subsets in human peripheral blood. *Cytometry A* **97**, 1044–1051 (2020).
65. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
66. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
67. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
68. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902 (2014).
69. Kang, H. M. et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018).
70. Mulè, M. P., Martins, A. J. & Tsang, J. S. Normalizing and denoising protein expression data from droplet-based single cell profiling. *Nat. Commun.* **13**, 2099 (2022).
71. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
72. Martin, M. D. & Badovinac, V. P. Defining memory CD8 T cell. *Front. Immunol.* **9**, 2692 (2018).
73. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
74. Borcherding, N., Bormann, N. L. & Kraus, G. scRepertoire: an R-based toolkit for single-cell immune receptor analysis. *F1000Res.* **9**, 47 (2020).
75. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
76. Hastie, T., Tibshirani, R., Narasimhan, B. & Chu, G. impute: imputation for microarray data. R package version 1.72.1 (2022).
77. Hoffman, G. E. & Roussos, P. Dream: powerful differential expression analysis for repeated measures designs. *Bioinformatics* **37**, 192–201 (2021).
78. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
79. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
80. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
81. Li, S. et al. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat. Immunol.* **15**, 195–204 (2014).
82. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
83. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
84. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
85. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
86. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
87. Candia, J. & Tsang, J. S. eNetXplorer: an R package for the quantitative exploration of elastic net families for generalized linear models. *BMC Bioinformatics* **20**, 189 (2019).
88. Rosati, E. et al. Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnol.* **17**, 61 (2017).
89. Morris, E. K. et al. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecol. Evol.* **4**, 3514–3524 (2014).

Acknowledgements We thank the study participants, who helped us to expand our scientific understanding of COVID-19 and viral infections in general; K. Sadtler for assistance with the development of the Cytek 36-colour flow cytometry panel; Q. Xu for assistance with the design of the cytokine stimulation experiments; F. Cheung for computational assistance; M. Mulè for providing software (dsb); the staff at the NCI CCR Sequencing Facility for sequencing support; the NIAID HPC Team for computing support; and R. Germain for reading the manuscript. Figs. 1a,b, 2a, 3a and 4a and Extended Data Fig. 2a were created using BioRender. This research was supported by the Intramural Research Programs of the NIAID and the Intramural Programs of the NIH Institutes supporting the NIH Center for Human Immunology. J.H. was funded in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract no. 75N91019D00024. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

Author contributions Conceptualization: R.S. and J.S.T. Methodology: W.W.L., C.L. and J.S.T. Software and formal analysis: W.W.L. and C.L. Investigation: R.S., W.W.L., C.L., K.L.H., N.B., J.M., G.G., L.R.K., G.K., J.C., J.T. and A.J.M. Resources: R.S., W.W.L., C.L., K.L.V., G.S., M.C., S.F.A., N.B., L.E.F., S.L., L.S., P.B., D.C., A.C., J.H., A.K.M., V.M., C.P., D.S., A.M., B.A.S., R.A., A.B.M., E.M.B., H.G., S.K. and J.S.T. Data curation: R.S., W.W.L., C.L., K.L.V., M.C., N.B. and L.E.F. Writing—original

draft: R.S. and J.S.T. Writing—review and editing: R.S., W.W.L., C.L. and J.S.T. Visualization: R.S., W.W.L. and C.L. Supervision: R.S., H.G., S.K. and J.S.T. Project administration: R.S. and J.S.T. Funding acquisition: R.S. and J.S.T. Human study and sample acquisition: R.S., K.L.V., G.S., M.C., N.B., L.E.F., S.L., L.S., P.B., D.C., A.C., J.H., A.K.M., V.M., C.P., D.S. and A.M. Flow cytometry analysis of PBMCs (including influenza-specific cell analyses): K.L.H. with panel designs, reagents and help from S.F.A. Whole-blood RNA-seq analysis: G.K. and N.B. Influenza microneutralization titres, influenza antibody SPR and SARS-CoV-2 neutralization titres: J.M., G.G., L.R.K. and J.T. CITE-seq data generation: C.L., A.J.M. and J.C. with help from G.K. and K.L.V. PBMC cytokine stimulation experiments: C.L. Integrative analysis of flow cytometry, CITE-seq, blood RNA-seq, circulating protein and antibody data: W.W.L. and C.L. with conceptual contributions from J.S.T.

Competing interests E.M.B. reports personal fees and non-financial support from Terumo BCT, Abbott Laboratories, Tegus and UptoDate, outside of the submitted work; is a member of the US FDA Blood Products Advisory Committee. Any views or opinions that are expressed in this manuscript are those of the author's, based on his own scientific expertise and professional judgement; they do not necessarily represent the views of either the Blood Products Advisory Committee or the formal position of FDA, and do not bind or otherwise obligate or commit either Advisory Committee or the Agency to the views expressed. The other authors declare no competing interests.

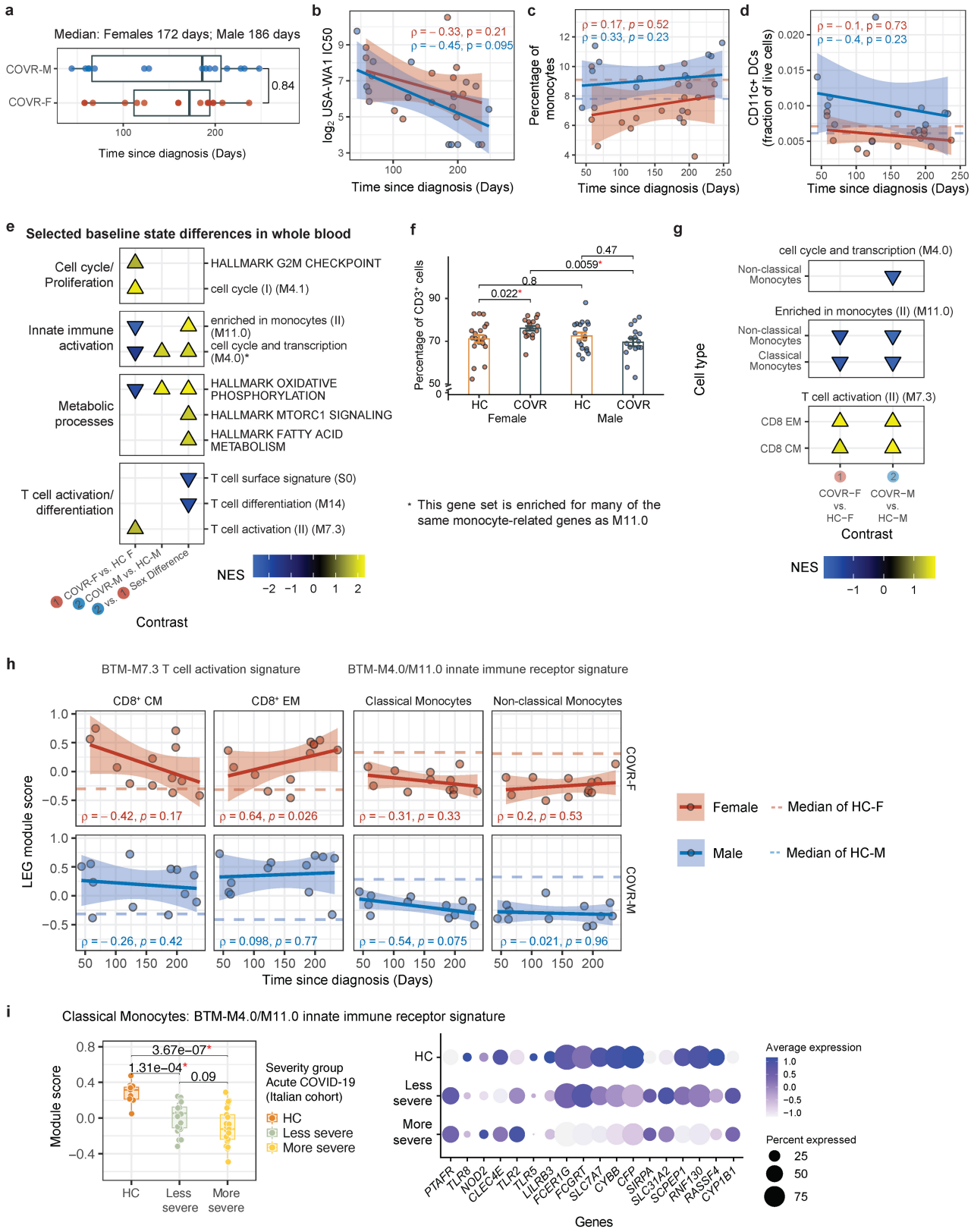
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-05670-5.

Correspondence and requests for materials should be addressed to John S. Tsang.

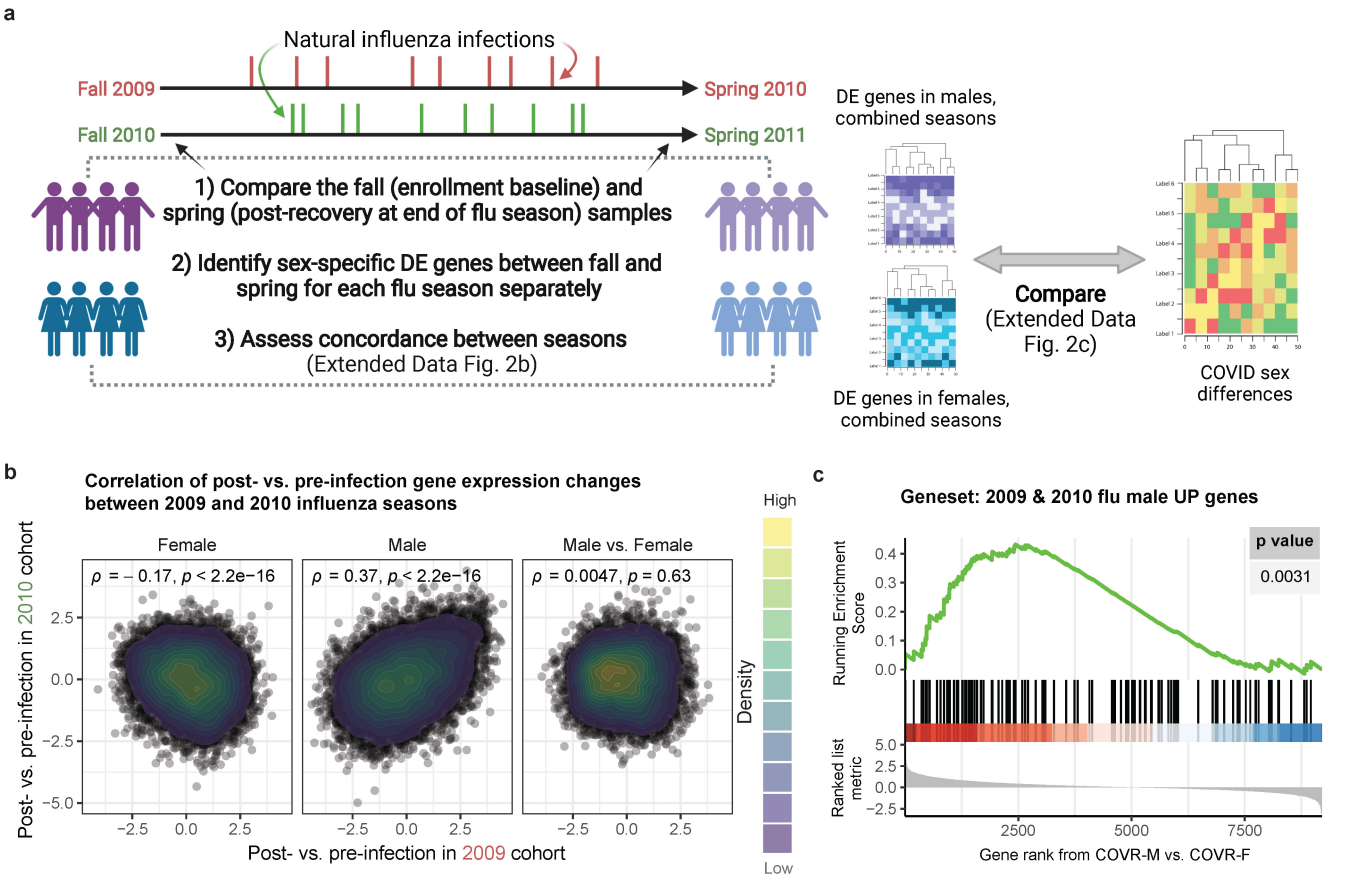
Peer review information *Nature* thanks Andreas Wack and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints.

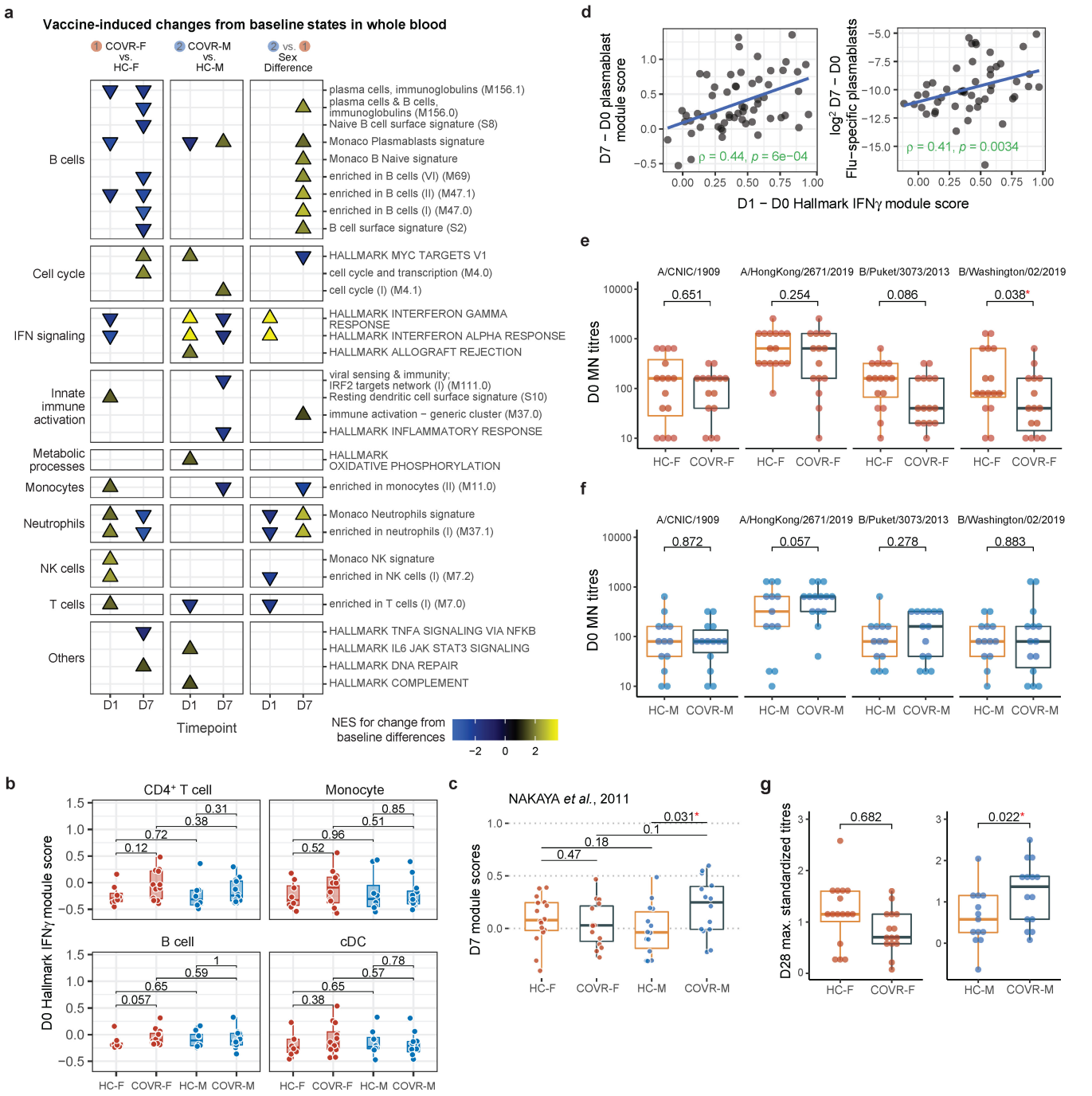


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Baseline differences between COVID-19-recovered participants and matching controls. a, Box plot showing the distribution of time since diagnosis (TSD; x-axis) in COVR-F (n = 16) and COVR-M (n = 15). Two participants with asymptomatic COVID-19 infection and thus unknown TSD are not included. Significance of group difference is determined by two-tailed Wilcoxon test. **b**, Scatterplot showing the correlation between the TSD (x-axis) and the SARS-CoV-2 neutralization titre (WA1 strain; y-axis) for COVR individuals at D0 prior to influenza vaccination. Spearman's rank correlation and unadjusted *P* values are shown. Participants with asymptomatic COVID-19 infection not included in TSD analyses. The shaded area represents 95% confidence interval. **c**, Similar to (**b**), but for the percentage of monocytes in peripheral blood as measured by the complete blood count (y-axis) at D0. Dotted lines represent median level in HC-F and HC-M. **d**, Similar to (**b**), but for the proportion of CD11c+ dendritic cells (DCs; as fraction of live cells; y-axis) as measured by flow cytometry of PBMCs at D0. **e**, Blood transcriptomic analysis of the stable baseline (prevaccination) differences among COVR and HC groups. Enrichment plot shows the normalized enrichment scores (NES) of selected gene sets of the different comparisons (GSEA FDR < 0.05; Methods; see Supplementary Table 3 for all significant gene sets with FDR < 0.05). The NES are plotted separately for COVR-F versus HC females (HC-F), COVR-M versus HC males (HC-M), or the difference between the two sets of comparisons (COVR-M versus COVR-F taking healthy sex differences into account). Positive (negative) NES indicates that gene set scores are higher (lower) in the first group than the second group listed in the comparison. Only gene sets not correlated with TSD across COVR individuals at baseline are considered stable. **f**, Similar to Fig. 1d, but for percent of CD3+ cells (T cells). **g**, Similar to (**e**), but for a subset of monocyte and T cell activation gene sets with significant enrichment ($P < 0.05$) using the D0 CITE-seq pseudobulk expression for the specified cell types (Methods; see Supplementary Table 5 for complete results). **h**, Scatterplots showing the relationship between the TSD and leading-edge gene (LEG) module scores [left two boxes: the T cell activation gene set (BTM-M7.3); right two boxes: the union of the LEGs from gene sets BTM M4.0 and M11.0; Methods] in COVR-F (n = 12) (top row) and COVR-M (n = 12) (bottom row) at D0 using the CITE-seq pseudobulk data of the indicated cell types. Each dot represents a COVR individual. The dotted lines represent the median score for the sex-matched HC group at D0 in the comparison shown. Spearman's rank correlation and *P* values are shown. The shaded area represents 95% confidence interval. **i**, (left) Box plot comparing the classical monocyte pseudobulk module scores of the LEGs used in Fig. 1f (union of female (F) and male (M) gene sets) in an acute COVID-19 CITE-seq dataset from Liu *et al.*[15]. Both M (n = 50) and F (n = 9) individuals are included in all three groups (HC n = 13, less severe n = 21, more severe n = 25). Each dot represents a sample. Unadjusted *P* values from the indicated two-group comparisons are shown. *P* values were generated using the moderated *t* statistics from a linear model in which samples from the same donors were treated as duplicates (Methods). (right) Bubble plot showing expression of the genes in Fig. 1f right panel within the classical monocyte CITE-seq data from Liu *et al.* in the same three patient groups shown in the left panel. All plot elements are the same as indicated in Fig. 1.

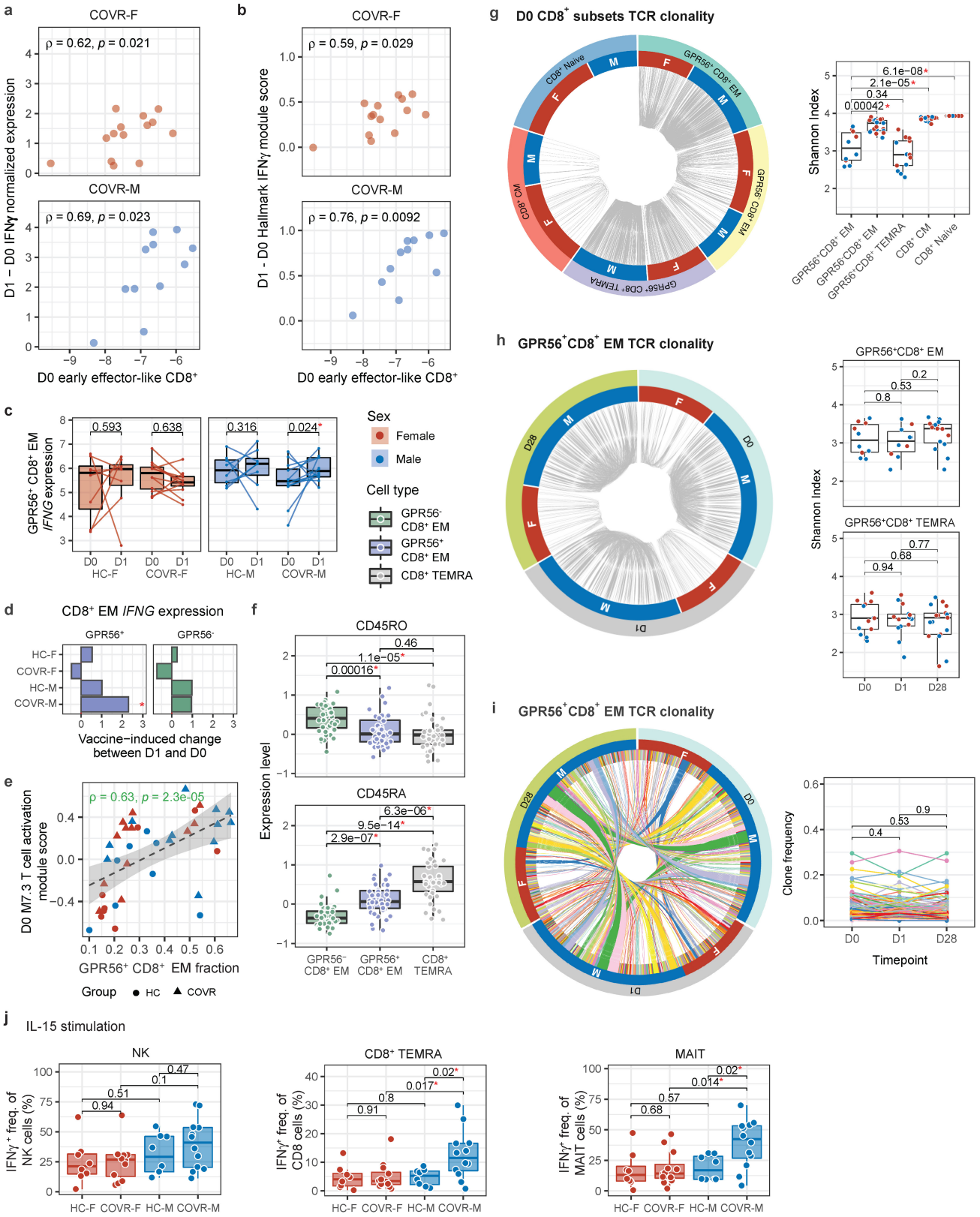


Extended Data Fig. 2 | Persistent post-infection gene expression changes following natural influenza infection. a, Schematic showing the approach used to evaluate changes in blood gene expression between before (pre-infection baseline) and months after natural influenza infection over two distinct seasons published in Zhai *et al.*[18], and how those gene changes may relate to sex-specific differences resulted from prior COVID-19 in this study. **b**, Scatter density plot showing the correlation between the gene expression changes (Supplementary Table 6) before (fall) and after (spring) natural influenza A infection in 2009 (x-axis) and 2010 (y-axis) for females (F; left), males (M; centre), and M vs F contrast (right). Shown are Spearman's rank correlation and unadjusted *P* values. **c**, Gene set enrichment plot of the genes that increased in M between fall (pre-infection) and spring (post-infection) in both 2009–2010 and 2010–2011 seasons. Genes were ranked by the signed -$\log_{10}$-transformed (unadjusted) P values in the COVID-19-recovered (COVR)-M vs COVR-F contrast at baseline using only participants under 65 years of age. The tick marks denote the location of the genes in the influenza gene set. The diagram in **a** was created using BioRender.



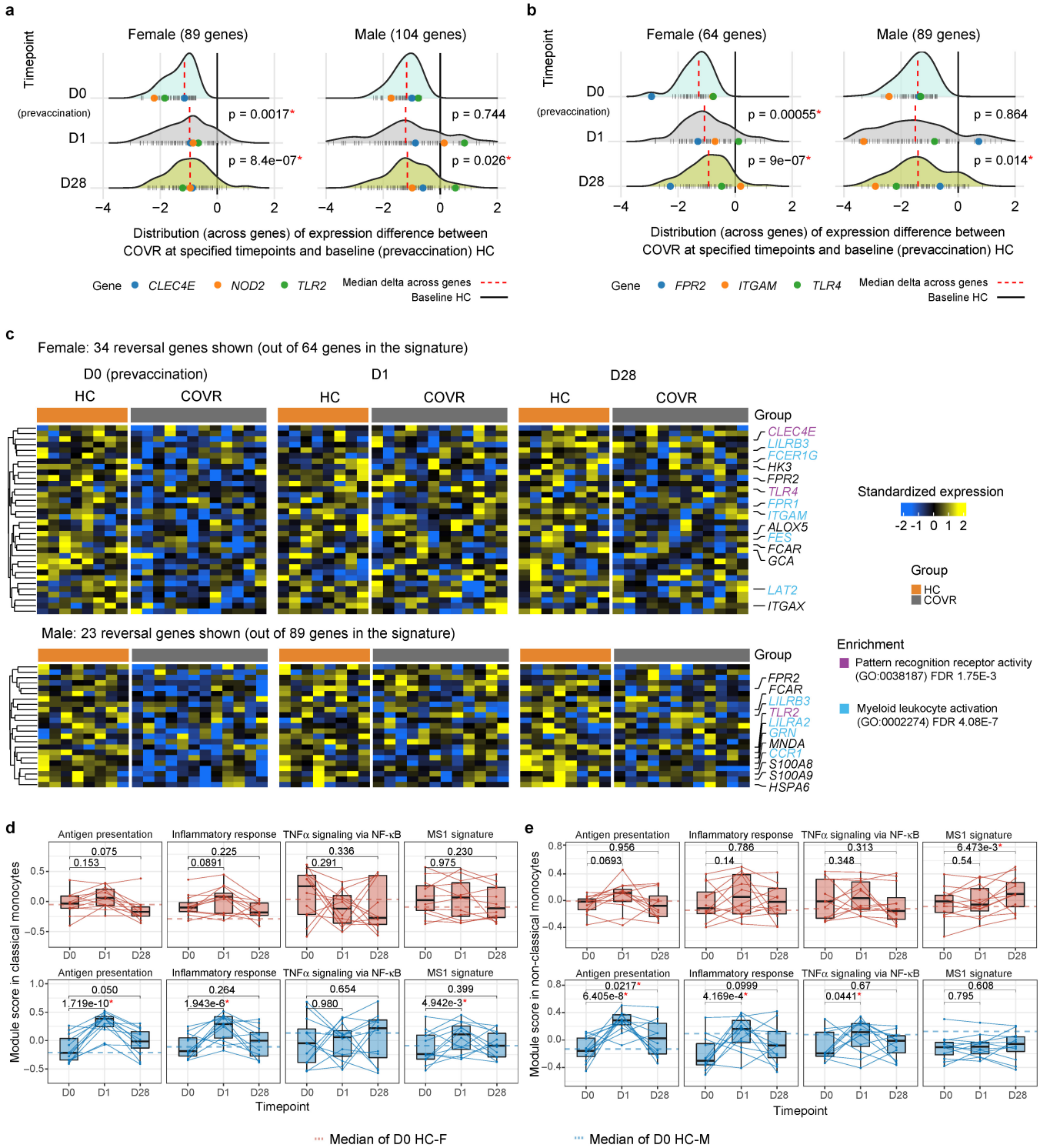
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Sex-specific molecular, cellular, and humoral response differences to influenza vaccination in COVID-19-recovered individuals and matching controls. a, Similar to Extended Data Fig. 1e but here showing enriched gene sets in whole blood comparing the early (D1 and D7) influenza vaccination responses in COVID recovered (COVR) vs. healthy control (HC) participants for females (F; Contrast 1), males (M; Contrast 2), and sex differences (Contrast 2 vs. 1; i.e., COVR-M versus COVR-F taking healthy sex differences into account; see Methods). Plotted are the gene sets that show significant changes from the baseline (D-7 and D0) within each comparison group [e.g., COVR-F and HC-F for D1] and significant differences between the two groups at the indicated timepoints (FDR < 0.05; see Supplementary Table 5). **b**, Similar to Fig. 2e, but showing the D0 Hallmark IFNγ Response module score for the indicated cell types from the CITE-seq pseudobulk expression data. CD4 = $CD4^+$ T cells; cDC = conventional/myeloid dendritic cells; B = B cells. **c**, Box plot showing the D7 whole blood signature score from genes identified in Nakaya *et al.*[52] whose D7/D0 fold-change positively correlated with D28 influenza hemagglutination inhibition titres. Only participants under 65 years of age [COVR-F (n = 15), COVR-M (n = 14), and HC-F (n = 16), and HC-M (n = 14)] are included. Significance of differences is determined by two-tailed Wilcoxon rank-sum test. **d**, Scatter plot showing the correlation of the whole blood D1 – D0 Hallmark Interferon Gamma Response gene set module score (x-axis) to the whole blood D7 – D0 plasmablast (PB) gene set module score (left y-axis; Monaco *et al.*[53]) and D7 – D0 difference of influenza-specific PB (all $HA^+CD27^+CD38^+CD20^{low}CD21^{low}$) frequency from flow cytometry (right y-axis; as fraction of $CD19^+$ B cells). Only study participants < 65 years of age are included. Spearman's rank correlation and unadjusted *P* values are shown. **e**, Box plots showing the D0 (prevaccination) microneutralization titres for each of the four strains in the seasonal influenza vaccine (columns) in females (COVR-F and HC-F) under the age of 65. Unadjusted *P* values are from linear models accounting for age and race (Methods). **f**, Similar to (**e**) but for males (COVR-M and HC-M) under 65 years of age. **g**, Maximum standardized influenza vaccine titre (among the four strains in the vaccine) at D28 after vaccination for females (left) and males (right), respectively. Unadjusted *P* values are from linear regression models accounting for age, race, influenza vaccination history, and baseline influenza titre (Methods). Unadjusted *P* values are shown. All plot elements are the same as indicated in Fig. 2.



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

Extended Data Fig. 4 | GPR56$^+$ virtual memory-like CD8$^+$ T cells contribute to increased day 1 IFNγ response in COVID-19-recovered males. a, Scatterplots showing the correlation between the day 0 (D0) $\log_2$ frequency of early effector-like CD8$^+$ T cells measured by flow cytometry (as fractions of live lymphocytes; see Population 50 in Supplementary Table 9 and Supplementary Fig. 1; x-axis) and the change (D1 – D0) in serum interferon gamma (IFNγ) protein levels measured by the OLINK platform (y-axis) for COVID-19-recovered (COVR) females (COVR-F; top, n = 14) and COVR males (COVR-M; bottom; n = 11). Spearman's rank correlation and *P* values are shown. **b**, Similar to **a** but showing the correlation between the D0 $\log_2$ frequency of early effector-like CD8$^+$ T cells measured by from flow cytometry (as fraction of live lymphocytes; x-axis) and the change (D1 - D0) in the whole blood signature score of the Hallmark IFNγ Response gene set (y-axis). **c**, Box plots comparing D0 and D1 pseudobulk IFNγ gene (*IFNG*) expression (y-axis) in GPR56$^+$ CD8$^+$ EM population for HC-F (n = 8), COVR-F (n = 12), HC-M (n = 8) and COVR-M (n = 12). Significance is determined by a linear model accounting for age, race, and influenza vaccination history (Methods). **d**, Bar plot showing the *t* statistic of the vaccine-induced change (D1 - D0) in *IFNG* expression using CITE-seq pseudobulk data (x-axis) within the GPR56$^+$ and GRP56$^-$ CD8$^+$ EM for HC-F (n = 8), COVR-F (n = 12), HC-M (n = 8), and COVR-M (n = 12). * $P < 0.05$ with exact value shown in **(c)**. **e**, Scatter plot showing the correlation between GPR56$^+$ CD8$^+$ EM cell frequency (as fractions of total CD8$^+$ EM in the CITE-seq data; x-axis) and BTM-M7.3 T cell activation signature score of CD8$^+$ EM cells computed using CITE-seq pseudobulk gene expression data (y-axis). Spearman correlation and *P* values are shown. The shaded area represents the 95% confidence interval. **f**, Related to Fig. 3h but showing CD45RA and CD45RO only with CD8$^+$ TEMRA cells included as an additional comparator. **g**, (left) Circos plot of T cell receptor (TCR) clonality for different CD8$^+$ T cell subsets at D0. Segments in the outer circle represent different CD8$^+$ T cell populations. Segments in inner circle represent male (M) and female (F) for both COVR and HC participants. Grey lines connect clones sharing identical CDR3 sequences within each individual. Cell subsets are downsampled for visualization (Methods). (right) Box plot showing Shannon's entropy index (y-axis) at D0 for each of the indicated CD8$^+$ populations. Significance of differences is determined by two-tailed Wilcoxon test. Shannon's entropy index evaluates the TCR repertoire diversity for each sample. Higher indices indicate higher diversity (i.e., fewer shared clones shown in Circos plot). EM, effector memory; CM, central memory; TEMRA, EM cells re-expressing CD45RA. **h**, (left) Circos plot of TCR clonality for GPR56$^+$ CD8$^+$ EM cells at different timepoints. Segments in the outer circle represent different days in the study (D0, D1, D28). Segments in the inner circle represent males (M) and females (F) for both COVR and HC participants. Grey lines connect clones sharing identical CDR3 sequences within each sample. Timepoints are downsampled for visualization purposes (Methods). (right) Box plot showing Shannon's entropy index (y-axis) of TCR clonality at each of the indicated time points (D0, D1, D28; x-axis) for GPR56$^+$ CD8$^+$ EM T cells (left) and GPR56$^+$ CD8$^+$ TEMRA (right). Significance of differences is determined by two-tailed Wilcoxon test. **i**, (left) Similar to (**h**), but showing the shared clones among different timepoints (segments in the outer circle). Segments in the inner circle represent unique clones for each individual. Clones and lines connecting shared clones are coloured. Samples with less than 30 cells were filtered out for visualization purposes. (right) Line chart showing frequencies of each clone (y-axis) shown in Circos plot (left) at D0, D1 and D28 for each participant. *P*-values of paired Wilcoxon test are shown comparing the clone frequency differences among D0, D1 and D28. **j**, Related to Fig. 3i but showing the frequencies of IFNγ$^+$ NK, IFNγ$^+$CD45RA$^+$CD45RO$^+$ TEMRA CD8$^+$ T cells and IFNγ$^+$ MAIT cells after IL-15 stimulation in vitro. All plot elements are the same as indicated in Fig. 3. Unadjusted *P* values are shown.



Extended Data Fig. 5 | Changes in immune states in COVID-19-recovered individuals following influenza vaccination. a, Distributions of gene-level difference of the innate immune receptor (IIR) signature (see Fig. 1f) in classical monocytes separately for females (F) and males (M) [shown as z-scores, on a per gene level, capturing the average difference between COVID Recovered (COVR) at the indicated timepoint (top to bottom: D0, D1, and D28) and healthy control (HC) at D0; see Methods]. Dashed red vertical lines represent the median of the distribution. Dark tick marks at the bottom represent individual genes and coloured dots highlight specific genes of interest. Significance of differences from D0 is determined by paired two-tailed Wilcoxon rank-sum test. **b**, Similar to (**a**) but for the non-classical monocytes (see Fig. 1g). **c**, Similar to Fig. 4d but for non-classical monocytes (see Fig. 1g for the innate receptor signature in non-classical monocytes). **d**, Similar to Fig. 4b but for COVR-F (red) and COVR-M (blue) only and gene sets shown on top of each plot. Box plots showing the classical monocyte LEG module scores (y-axis) of gene sets from Supplementary Fig. 2: antigen presentation related gene sets, Hallmark Inflammatory response, Hallmark TNF-α signalling via NF-κB, and MS-1 signature from Reyes *et al.*[17]. LEGs from the first three gene sets were found to be repressed in acute COVID-19 patients in Liu *et al.*[15]. **e**, Similar to (**d**), but for non-classical monocytes. All plot elements are the same as indicated in Fig. 4. Unadjusted *P* values are shown.

Extended Data Table 1 | Cohort characteristics

	COVR		Healthy	
	Female	**Male**	**Female**	**Male**
Subject count	17	16	21	19
Age (Years)				
Median	40.2	43.7	52.5	47.6
Mean	44.9	43.7	47.2	47.4
Min	23.4	21.9	22.5	24.0
Max	70.5	67.3	70.4	69.1
Aged > 65	2	2	5	5
Race				
Asian	1	0	2	2
Black	1	0	4	0
Multiple races	0	2	1	3
White	15	14	14	14
Number of Influenza Vaccination in Past 10 Years				
0	0 (0%)	1 (6.25%)	0 (0%)	2 (10.53%)
1	1 (5.88%)	0 (0%)	1 (4.76%)	0 (0%)
2	0 (0%)	1 (6.25%)	2 (9.52%)	2 (10.53%)
3	1 (5.88%)	0 (0%)	1 (4.76%)	0 (0%)
4	0 (0%)	0 (0%)	2 (9.52%)	0 (0%)
5	3 (17.65%)	3 (18.75%)	1 (4.76%)	0 (0%)
6	1 (5.88%)	1 (6.25%)	1 (4.76%)	0 (0%)
7	3 (17.65%)	0 (0%)	1 (4.76%)	1 (5.26%)
8	0 (0%)	1 (6.25%)	1 (4.76%)	3 (15.79%)
9	1 (5.88%)	1 (6.25%)	1 (4.76%)	1 (5.26%)
10	7 (41.18%)	8 (50%)	10 (47.62%)	10 (52.63%)
Experienced side effects after vaccination	16 (94.1%)	9 (56.3%)	17 (90.0%)	15 (78.9%)
COVID-19 Symptoms				
Asymptomatic	1 (5.9%)	1 (6.3%)	-	-
Symptomatic	16 (94.1%)	15 (93.8%)		
Time since COVID-19 Diagnosis (Days) *				
Median	172.0	186.0	-	-
Mean	152.9	149.3	-	-
Min	58.0	44.0	-	-
Max	237.0	248.0	-	-
Duration of acute COVID-19 symptoms (Days) *				
Median	14	10		
Mean	19.62	13.07		
Min	4	1		
Max	87	33		
Experienced COVID-19 residual symptoms at time of screening	8 (47.1%)	3 (18.8%)	-	-
COVID-19 Residual Symptoms				
Brain fog	1 (5.9%)	0 (0%)	-	-
Fatigue	2 (11.8%)	0 (0%)	-	-
Loose stools	0 (0%)	1 (6.3%)	-	-
Reduced sense of taste	1 (5.9%)	1 (6.3%)	-	-
Reduced sense of smell / smell disturbances	5 (29.4%)	2 (12.5%)	-	-
Shortness of breath / Chest pressure	1 (5.9%)	0 (0%)	-	-

* Excluding asymptomatic subjects

Demographics (age, sex, self-reported race), influenza vaccination history, and COVID-19 related information (for recoverees). COVR = COVID-19 recovered.

nature portfolio

Corresponding author(s): John Tsang

Last updated by author(s): 12 December 2022

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	REDCap (Research Electronic Data Capture) v8.5.27
Data analysis	Trimmomatic (v0.38.0), STAR (v2.6.0b), UMI-tools (v0.5.3), subread (v1.6.2), R (v3.6.1, 4.1.0, 4.1.1), variancePartition (v1.16.1), limma (v3.42.2, 3.50.1), clusterProfiler (v3.17.0), ggplot2(v3.3.5), ggpubr (v0.4.0), RUVSeq (v1.18), GSVA (v1.30.0, 1.42.0), edgeR (v3.26.8), impute (v1.60.0), CellRanger (v6.0.1), demuxlet (v2), Seurat (v4.0.3, 4.1.0), dsb (v0.3.0), Circlize (v0.4.14), ComplexHeatmap (v2.10.0), eNetXplorer (v1.1.3), SpectroFlo (v2.2.0), scRepertoire (v1.4.0), DEseq2 (v1.34.0), MASS (v7.3-53), FlowJo (v10)

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 and processed data from the whole blood bulk RNA-seq and single-cell CITE-seq are available from the NCBI Gene Expression Omnibus, accession numbers GEO: GSE194378 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE194378) and GEO:GSE206265 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE206265), respectively (will be released to public at time of publication). Additional datasets, including clinical, proteomics, flow cytometry, CITE-seq, and

influenza antibody measurements, are available at: https://doi.org/10.5281/zenodo.5935845 (will be released to public at time of publication). The influenza infection dataset we utilized was downloaded directly from GEO: GSE68310 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE68310).

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	n = 73 (COVR-F=17; HC-F=21; COVR-M = 16; HC-M = 19). No sample size calculations were done prior to enrollment, in part because there were no reliable effect size estimates related to the impact of prior COVID-19 infection on vaccine responses. The number of subjects in the study was the number that were able to be recruited during the recruitment period.
Data exclusions	None
Replication	Technical replicates were used to assess technical variability of mRNA expression in bulk RNA-seq and single cell CITE-seq, and genes whose technical variance was higher than a predefined threshold were removed from downstream analysis. In addition, the linear models constructed to assess group differences leveraged technical replicates to control for batch and other technical effects.
Randomization	Because this study looked at the cellular and molecular responses to influenza vaccination, there was no indication to randomize individuals to receive influenza vaccination or placebo. The covariates age, self-reported influenza vaccination history (of the past 10 years) and pre-vaccination (baseline) influenza antibody titers were accounted for in our results as they have been included as co-variates in our statistical model for assessing immune response differences among the groups. We model post-vaccination titer as the outcome variable with influenza vaccination history and pre-vaccination titer included as covariates. This approach has also been used for influenza vaccine evaluation by the Food and Drug Administration (e.g., see https://www.fda.gov/media/135687/download page 27).
Blinding	Blinding was not performed as individuals were recruited based on a known history of SARS-CoV-2 infection. Furthermore, as part of the study, an individual's SARS-CoV-2 infection history was confirmed by antibody test in order to allocate them to the correct subject group. Finally, because this study looked at the cellular and molecular responses to influenza vaccination, we did not administer any placebo vaccine, only an FDA-approved influenza vaccine.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used	TotalSeq™-C Human Universal Cocktail, V1.0; Fluorochrome-conjugated antibodies as indicated in Methods Table 1, 2 and 3.
Validation	See Supplementary Tables 11-14 for antibody data, including validation information from the commercial companies who made the products.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Lenti-X- 293T cells were obtained from Takara Bio (Cat. No. 632180). 293T-ACE2 cells were obtained from ATCC.

Authentication	Cell line was checked for expression of ACE2 and validated by FACS analysis. None of the cell lines were authenticated by karyotyping or other genomic techniques.
Mycoplasma contamination	Negative for Mycoplasma
Commonly misidentified lines (See ICLAC register)	No misidentified cell lines were used in the study.

Human research participants

Policy information about studies involving human research participants

Population characteristics	COVID-19-recovered males (n=16; ages 21-67; median time since COVID-19 diagnosis: 186 days). COVID-19-recovered females (n=17; ages 23-70; median time since COVID-19 diagnosis: 172 days). Healthy control males (n=19; ages 24-69). Healthy control females (n=21; ages 22-70).
Recruitment	Subjects were recruited from the community using email listserv messages, newspaper advertisements, and follow-up contact from a COVID-19 convalescent plasma study. All subjects were self-selected based on interest in participating in COVID-19-related research, ability to participate in study visits, and exposure to recruitment materials. As influenza vaccine titers and responses are not routinely measured outside of research settings, it is unlikely that individuals who opted to participate in the study introduced systemic bias as a result of prior knowledge about their individual cellular or molecular responses to the influenza vaccine.
Ethics oversight	Institutional Review Board of the National Institutes of Health. Informed consent was obtained from all study participants prior to the onset of study procedures.

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

Clinical data

Policy information about clinical studies

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

Clinical trial registration	NCT04025580
Study protocol	19-I-0126 (Systems analyses of the immune response to the seasonal influenza vaccine). Trial protocol described on clinicaltrials.gov.
Data collection	Subjects were recruited between August and December 2020. Study visits occurred at the National Institutes of Health (NIH) Clinical Center (CC) in Bethesda, Maryland, USA. Blood samples were collected by phlebotomy staff at the NIH CC. Samples were collected between September 2020 and April 2021.
Outcomes	Primary outcome to be measured was serum microneutralization titers to the four influenza strains in the FDA-approved seasonal influenza vaccines used in the study (Flucelvax and high dose Fluzone).

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	Whole blood was collected from study participants, PBMCs were isolated using a modified Ficoll procedure, and cells were frozen and stored at -80C. Thawed PBMC were washed in RPMI culture medium containing 50U/ml benzonase nuclease and then washed by PBS. Cells were incubated with LIVE/DEAD Fixable Blue Dye, which was used to exclude dead cells from analysis. Cells were incubated with fluorochrome-conjugated B, H1, H3 probes and fluorochrome-conjugated antibodies against IgM, IgA, CD21, CD85J, FCRL5, CD20, IgG, CD38, CD14, CD56, CD3, CD27, CD71, CD19, IgD etc as indicated in the Method for 30 min at 4 C in the dark. After incubation with antibodies for 30 minutes, cells were washed two times with FACS buffer (0.1%BSA/PBS (pH7.4)) and fixed in 1% paraformaldehyde.
Instrument	Cytek Auora (Cytek Biosciences), BD Symphony (BD Biosciences)
Software	SpectroFlo version 2.2.0 (Cytek Biosciences) and BD FACSDiva Software were used for collecting the data and FlowJo version 10 (BD Biosciences) was used for flow cytometry data analysis.

Cell population abundance	Post-sort fractions not evaluated.
Gating strategy	Gating strategies provided in Supplementary Information Figures 1, 3 and 6.

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