nature immunology

Resource

https://doi.org/10.1038/s41590-024-02052-z

Variants and vaccines impact nasal immunity over three waves of SARS-CoV-2

Received: 16 April 2024

Accepted: 5 December 2024

Published online: 20 January 2025

Check for updates

Jaclyn M. L. Walsh[1,2,3,4,24], Vincent N. Miao[4,5,6,7,24], Anna H. Owings[8,9,24], Ying Tang[1,24], Joshua D. Bromley[4,5,6,10,24], Samuel W. Kazer[1,2,4,5], Kyle Kimler[1,4,5,6,11,12], Chelsea Asare[1,4], Carly G. K. Ziegler[4,5,6,7,13], Samira Ibrahim[4,5,6], Tasneem Jivanjee[4,5,6], Micayla George[4,5,6], Andrew W. Navia[4,5,6,11], Riley S. Drake[4,5,6], Adam Parker[9], Benjamin C. Billingsley[8], Paul Dotherow[9], Spurthi Tarugu[8], Sai K. Kota[9], Hannah Laird[8,9], T. Grant Wichman[8], Yesenia T. Davis[9], Neha S. Dhaliwal[8,9], Yilianys Pride[9], Yanglin Guo[14], Michal Senitko[14], Jessie Harvey[8], John T. Bates[15], Gill Diamond[16], Michael R. Garrett[8,15], D. Ashley Robinson[15], I. J. Frame[17], Jonathan J. Lyons[18,19], Tanya O. Robinson[9], Alex K. Shalek[3,4,5,6,7,11,13,20,25], Bruce H. Horwitz[1,3,21,25], Sarah C. Glover[9,15,22,25] & Jose Ordovas-Montanes[1,3,4,5,23,25]

Viral variant and host vaccination status impact infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), yet how these factors shift cellular responses in the human nasal mucosa remains uncharacterized. We performed single-cell RNA sequencing (scRNA-seq) on nasopharyngeal swabs from vaccinated and unvaccinated adults with acute Delta and Omicron SARS-CoV-2 infections and integrated with data from acute infections with ancestral SARS-CoV-2. Patients with Delta and Omicron exhibited greater similarity in nasal cell composition driven by myeloid, T cell and SARS-CoV-2^{hi} cell subsets, which was distinct from that of ancestral cases. Delta-infected samples had a marked increase in viral RNA, and a subset of *PER2*$^{+}$*EGR1*$^{+}$*GDF15*$^{+}$ epithelial cells was enriched in SARS-CoV-2 RNA^{+} cells in all variants. Prior vaccination was associated with increased frequency and activation of nasal macrophages. Expression of interferon-stimulated genes negatively correlated with coronavirus disease 2019 (COVID-19) severity in patients with ancestral and Delta but not Omicron variants. Our study defines nasal cell responses and signatures of disease severity across SARS-CoV-2 variants and vaccination.

The nasal mucosa is a primary site of infection, replication and transmission for SARS-CoV-2 (refs. 1–3). Understanding nasal cellular responses to SARS-CoV-2 and how they differ across variants, vaccination status and disease trajectories is key to developing therapeutic and prophylactic strategies against this virus and others. The emergence of variants of concern, including Delta and Omicron, has contributed to the continued circulation of SARS-CoV-2. Compared to the ancestral strain, Delta is characterized by a substantially elevated viral load, rapid intracellular replication and enhanced fusogenicity[4,5]. Compared to Delta, Omicron exhibits greater transmission rates due to extensive antigenic shift, lower RNA viral loads, slower replication and less severe pathology[6–8]. Investigation of differences in immune responses to each SARS-CoV-2 variant has largely focused on either intracellular mechanisms of innate immune evasion or antigen-specific adaptive immune cells in the blood[9,10]. Whether the differences in variant properties are reflected by differences in

A full list of affiliations appears at the end of the paper. e-mail: jose.ordovas-montanes@childrens.harvard.edu

nasal cell tropism or in the cellular immune response within the nose remains unclear.

SARS-CoV-2 vaccines have dramatically reduced the burden of severe disease and death from COVID-19. Hybrid immunity, the combination of immune memory acquired from both vaccination and infection, can provide even more potent protection than vaccination alone[11]. However, evidence suggests that lack of establishment of durable, effective respiratory mucosal immune memory populations is a limitation of the current vaccination strategies for SARS-CoV-2 (refs. 12–16). Given that vaccination significantly improves peripheral immune responses, respiratory symptoms and disease severity, we sought to understand its impact on the cellular response in the upper airway during acute infection.

Disease outcomes from SARS-CoV-2 infection vary widely, from asymptomatic infections to severe respiratory distress[17]. Studies applying high-dimensional technologies to identify correlates of COVID-19 severity have mostly focused on the blood[18–21]. An interferon (IFN) response was identified as a critical determinant of disease severity in the contexts of anti-IFN autoantibodies[22,23], inborn errors of type I IFN immunity[24] and a muted upper airway epithelial IFN response[25,26] during the first waves of SARS-CoV-2 infections. As the public health context of SARS-CoV-2 continues to grow more complex, it is critical to understand how factors such as variant and vaccination impact correlates of COVID-19 severity.

Prior studies applying scRNA-seq to nasal samples collected during acute SARS-CoV-2 infection have all been conducted in unvaccinated individuals before the emergence of the Delta variant[25,27–30]. Here, we characterized the epithelial and immune cell states in the human nasal mucosa during infection with ancestral, Delta and Omicron SARS-CoV-2. We identified convergent cellular ecosystems in response to Delta and Omicron, distinct from the ancestral-responsive cellular ecosystem. We observed a marked increase in the abundance of SARS-CoV-2 RNA in patients with Delta and unique cell subsets enriched in SARS-CoV-2 RNA^{+} cells within each variant. Comparisons between vaccinated and unvaccinated patients with Delta and Omicron indicated that vaccination was associated with nasal macrophage recruitment and activation. We identified distinct gene signatures and cell states that associated with COVID-19 severity during infection with each variant. Our data provide a comprehensive view of the diversity of nasal cellular responses to SARS-CoV-2.

Results

SARS-CoV-2 induces cellular changes in the nasal mucosa

Nasopharyngeal swabs were collected from a total of 112 adults at the University of Mississippi Medical Center (UMMC), including 26 individuals who were SARS-CoV-2 negative by PCR (hereafter controls) and 86 individuals who were SARS-CoV-2 positive by PCR (Fig. 1a and Table 1). Full clinical metadata of participants are provided (Supplementary Table 1). Of the 112 total samples, 13 controls and 32 SARS-CoV-2 samples (ancestral cases) were collected in April–November 2020 and have been included in a previously published study[25], six controls and 33 SARS-CoV-2 samples (Delta cases) were collected in July–October 2021 and seven controls and 21 SARS-CoV-2 samples (Omicron cases) were collected in January–February 2022 (Fig. 1a). Targeted viral sequencing of Delta and Omicron cases with sufficient viral RNA material indicated that the majority of Delta (22 of 24) and Omicron (12 of 12) cases were infected with the respective variant (Extended Data Fig. 1a,b). Samples with insufficient viral RNA for sequencing were presumed to be infected with the dominant circulating variant at the time of collection.

The median age in years was 51.5 for controls, 58.5 for ancestral, 50 for Delta and 61 for Omicron cases (Table 1). There were no significant differences in biological sex or ethnicity, but self-reported race varied between variant groups (Table 1). Controls had higher rates of diabetes and inflammatory bowel disease, and Omicron cases had higher rates of congestive heart failure than the other groups (Table 1). The number of participants treated with the SARS-CoV-2 antiviral medication remdesivir was highest in Delta cases and lowest in ancestral cases (Table 1). Participants experienced a range of COVID-19 severity quantified by the World Health Organization (WHO) score of required respiratory support (0–8 scale; Methods). COVID-19 severity scores were not significantly different between variant groups (Fig. 1b and Extended Data Fig. 1c). All COVID-19 swabs were collected while participants experienced acute respiratory symptoms, on average, 3.5 days (d) (range, −2 to 12 d) in ancestral, 4.6 d (0–19 d) in Delta and 2.3 d (−3 to 19 d) in Omicron cases after their first SARS-CoV-2-positive PCR test (Extended Data Fig. 1d). Comparison of SARS-CoV-2-specific antibody levels in plasma collected at the same time as the nasopharyngeal swab did not reveal any significant differences in immunoglobulin (Ig)G titers between variant groups (Extended Data Fig. 1e).

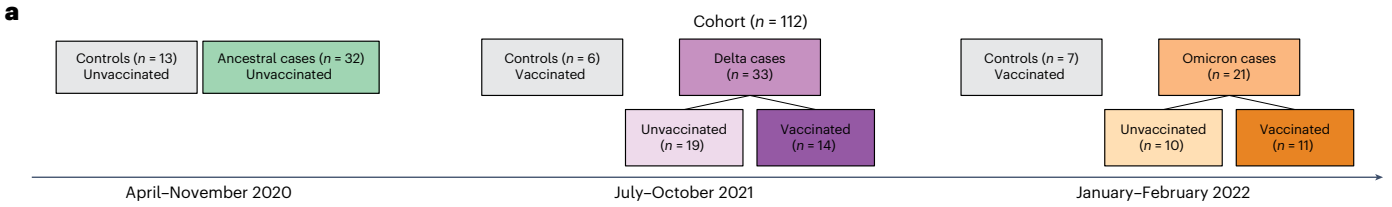
The Delta and Omicron cohorts included participants who had received a SARS-CoV-2 vaccine (14 of 33 Delta cases, 11 of 21 Omicron cases, 13 of 13 controls collected in August 2021–February 2022) (Fig. 1a and Extended Data Table 1). Unless otherwise specified, vaccinated and unvaccinated controls are grouped together hereafter. We included vaccinated participants with severe COVID-19 in our study, such that COVID-19 severity scores did not differ significantly between vaccination groups (Fig. 1c and Extended Data Fig. 1f). Vaccinated Delta cases received either two doses (13 of 14 cases) or one dose (one of 14) of a messenger RNA (mRNA) vaccine 4–8 months (13 of 14) or 1 month (one of 14) before testing positive for SARS-CoV-2 (Extended Data Fig. 1g). Vaccinated Omicron cases received three doses (five of 11) or two doses (five of 11 cases) of an mRNA vaccine or one dose of the J&J vector-based vaccine (one of 11) 9–10 months (four of 11), 1–5 months (five of 11) or an unknown period (two of 11) before testing positive (Extended Data Fig. 1g). Comparisons of demographic and clinical data across vaccination groups are provided (Extended Data Table 1). Within each variant, there was not a significant difference in average time (d) from positive SARS-CoV-2 PCR test to sample collection between vaccinated and unvaccinated cases (Extended Data Fig. 1h). Unvaccinated Delta cases had higher levels of SARS-CoV-2 nucleoprotein (NP)-specific IgG than vaccinated Delta cases (Extended Data Fig. 1i). Due to the timing of sample collection, we were unable to resolve whether NP IgG antibodies were produced in response to a prior infection or the ongoing infection, and thus we did not subdivide the Delta or Omicron cases

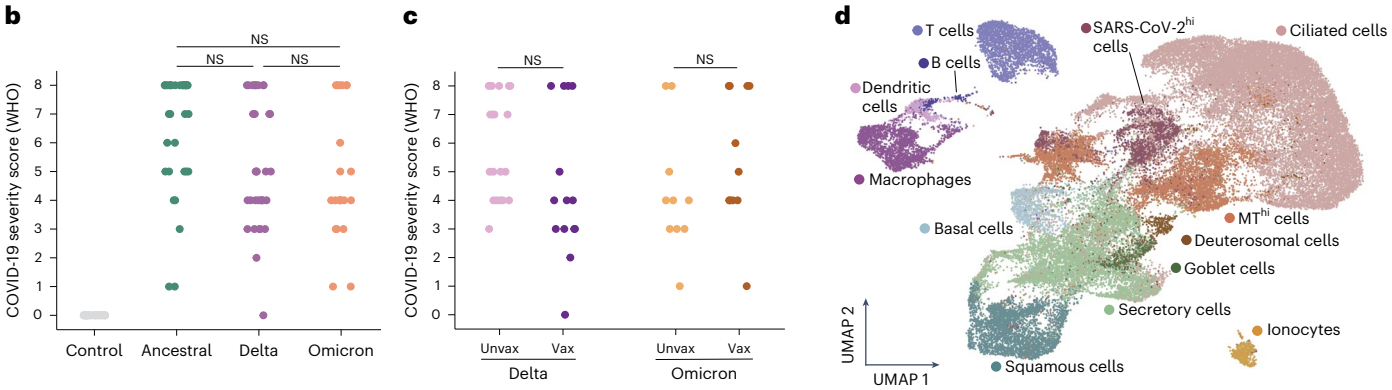
Fig. 1 | SARS-CoV-2 induces changes in the cellular composition of nasal mucosa in patients infected during the ancestral, Delta and Omicron waves. **a**, Schematic showing the distribution of participants across SARS-CoV-2 variant groups (controls, ancestral, Delta and Omicron), vaccination status and collection periods (April–November 2020, July–October 2021, January–February 2022). **b**,**c**, COVID-19 severity scores, defined by the WHO score of respiratory support required at peak of disease (0–8) in controls ($n = 26$) and ancestral ($n = 32$), Delta ($n = 33$) and Omicron ($n = 21$) cases (**b**) and in unvaccinated ('unvax') Delta ($n = 19$), vaccinated ('vax') Delta ($n = 14$), unvaccinated Omicron ($n = 10$) and vaccinated Omicron ($n = 11$) cases (**c**). NS, not significant. **d**, Uniform manifold approximation and projection (UMAP) of 48,730 cells from all participants ($n = 112$) colored by major cell type following iterative Louvain clustering. **e**, Violin plots of marker genes for cell type annotations shown in **d**. SCT, sctransform. **f**, Frequency of secretory cells (ancestral versus control, $P = 4.6 \times 10^{-4}$), deuterosomal cells (ancestral versus control, $P = 5.2 \times 10^{-5}$), ciliated cells (Delta versus control, $P = 7.87 \times 10^{-3}$; Omicron versus control, $P = 0.037$), SARS-CoV-2^{hi} cells (ancestral versus control, $P = 0.011$; Delta versus control, $P = 0.039$; Omicron versus control, $P = 3.54 \times 10^{-6}$), T cells and macrophages out of all cells in the sample from control ($n = 26$) and ancestral ($n = 32$), Delta ($n = 33$) and Omicron ($n = 21$) cases. Box plots represent median (center line), upper and lower quartiles (box limits) and 1.5× the interquartile range (whiskers). **g**, Expression of SARS-CoV-2 entry or host cell processing genes (x axis) in the major cell types defined as in **d** (y axis) in all samples ($n = 112$). **b**,**c**,**f**, Statistical tests are two-sided Kruskal–Wallis tests with Benjamini–Hochberg correction for multiple comparisons. *Dunn's post hoc test, $P < 0.05$.

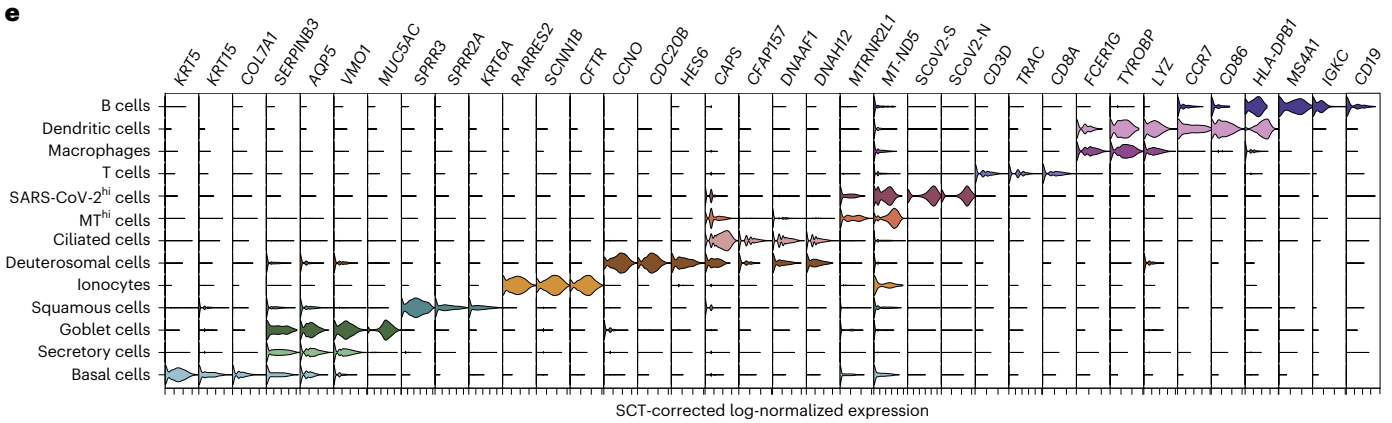
https://doi.org/10.1038/s41590-024-02052-z

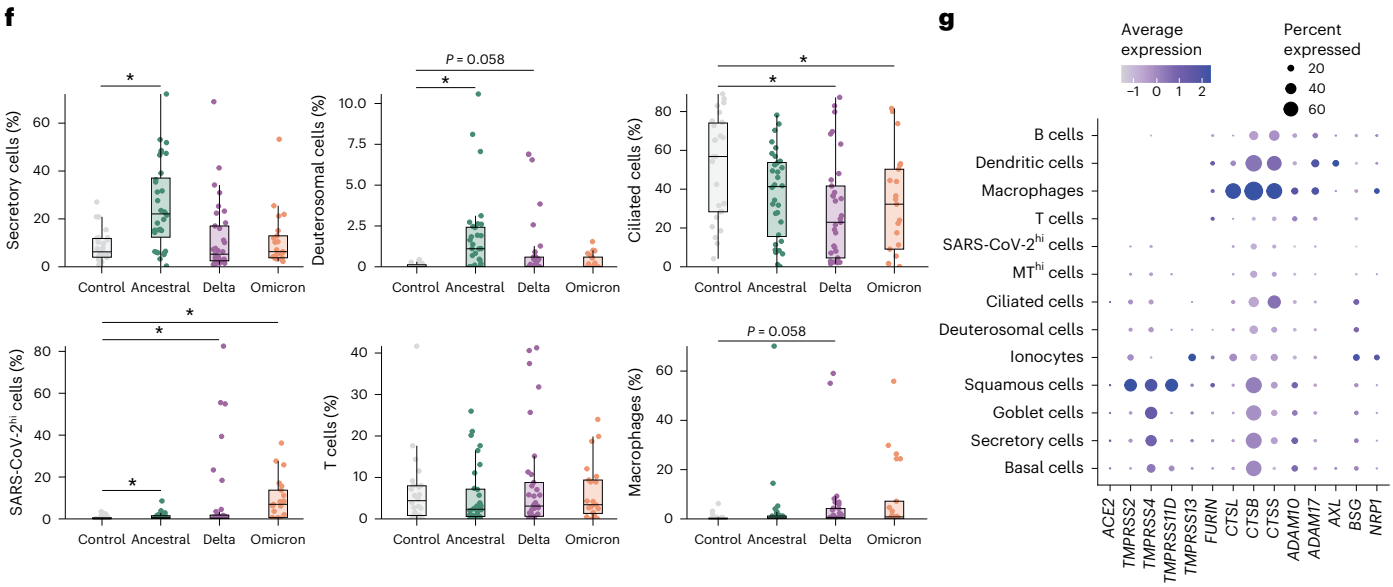
by serology-based exposure history to SARS-CoV-2. Together, our participant cohort encompasses three distinct SARS-CoV-2 variants, vaccinated and unvaccinated individuals and a range of COVID-19 severity.

Single-cell suspensions of viable epithelial and immune cells were isolated from each nasopharyngeal swab (Methods)[25]. scRNA-seq libraries were generated using Seq-Well S^3. After iterative removal of low-quality or low-complexity cells (Supplementary Table 2), we generated a dataset of 48,730 cells (Extended Data Fig. 2a) encompassing diverse epithelial and immune cell types (Fig. 1d). We detected the majority of cell types described in atlases of the airway epithelium[31],









https://doi.org/10.1038/s41590-024-02052-z

Table 1 | Composition of participant cohort based on infection with the ancestral, Delta and Omicron variants of SARS-CoV-2

SARS-CoV-2 variant group	Control	Ancestral	Delta	Omicron
Number of participants	26 (23.2%)	32 (28.6%)	33 (29.5%)	21 (18.8%)
Age (years)*				
Minimum	20	28	22	33
Median (IQR)	51.5 (35)	58.5 (18.25)	50 (25)	61 (9)[a]
Maximum	81	84	83	79
Sex				
Female	65.4% (17/26)	46.9% (15/32)	42.4% (14/33)	28.6% (6/21)
Ethnicity				
Hispanic	0% (0/26)	3.1% (1/32)	0% (0/33)	0% (0/21)
Not Hispanic	100% (26/26)	96.9% (31/32)	100% (33/33)	100% (21/21)
Race**				
White/Caucasian	42.3% (11/26)	25% (8/32)	51.5% (17/33)	33.3% (7/21)
Black/African American	42.3% (11/26)	65.6% (21/32)	48.5% (16/33)	61.9% (13/21)
Asian	15.4% (4/26)[b]	0% (0/32)	0% (0/33)	0% (0/21)
American Indian	0% (0/26)	^9.4% (3/32)	0% (0/33)	4.8% (1/21)
WHO score (swab)***				
Median (IQR)	0 (0)	5 (2.25)[a]	4 (1)[a]	4 (2)[a,c]
WHO score (peak)***				
Median (IQR)	0 (0)	7.5 (3)[a]	5 (4)[a]	4 (4)[a]
Vaccination status***				
Vaccinated	50% (13/26)	0% (0/32)[b]	42.4% (14/33)	52.4% (11/21)
Unvaccinated	50% (13/26)	100% (32/32)[b]	57.6% (19/33)	47.6% (10/21)
BMI				
Median (IQR)	28.6 (14.7)	31.4 (12.3)	33.1 (12.0)	30.7 (15.3)
Pre-existing conditions				
Diabetes**	19.2% (5/26)[b]	59.4% (19/32)	39.4% (13/33)	61.9% (13/21)
Chronic kidney disease	3.8% (1/26)	15.6% (5/32)	15.2% (5/33)	28.6% (6/21)
Congestive heart failure*	11.5% (3/26)	3.1% (1/32)	9.1% (3/33)	28.6% (6/21)[b]
Lung disorder	11.5% (3/26)	34.4% (11/32)	21.2% (7/33)	19% (4/21)
Hypertension	50% (13/26)	71.9% (23/32)	60.6% (20/33)	52.4% (11/21)
IBD**	19.2% (5/26)[b]	0% (0/32)	0% (0/33)	4.8% (1/21)
Treatment				
Steroids	N/A	59.4% (19/32)	84.8% (28/33)	71.4% (15/21)
Remdesivir***	N/A	9.4% (3/32)[b]	72.7% (24/33)[b]	57.1% (12/21)
Mortality**	0% (0/26)[b]	50% (16/32)[b]	27.3% (9/33)	28.6% (6/21)

Participants were assigned as control (SARS-CoV-2 negative by PCR, April–November 2020, $n=13$; July–October 2021, $n=6$; January–February 2022, $n=7$), ancestral (SARS-CoV-2$^+$ by PCR, April–November 2020, $n=32$), Delta (SARS-CoV-2$^+$ by PCR, July–September 2021, $n=33$) or Omicron (SARS-CoV-2$^+$ by PCR, January–February 2022, $n=21$). Age (Omicron, $P=0.044$), biological sex, ethnicity, self-reported race ($P=5.2\times10^{-3}$), COVID-19 severity (WHO scores at time of sample collection (Omicron versus ancestral, $P=0.042$) and at peak disease (all $P<5\times10^{-5}$)), vaccination status ($P=2.91\times10^{-5}$), body mass index (BMI), pre-existing conditions (diabetes, $P=5.8\times10^{-3}$; congestive heart failure, $P=0.042$; inflammatory bowel disease (IBD), $P=3.9\times10^{-3}$), COVID-19 treatments (remdesivir, $P=8.9\times10^{-7}$) and mortality rates ($P=2.3\times10^{-3}$) were compared across variant groups. For continuous variables, median and interquartile range (IQR) are presented, and a two-sided Kruskal–Wallis test with Benjamini–Hochberg correction for multiple comparisons was used for statistical comparison. *$P<0.05$, **$P<0.01$, ***$P<0.001$, otherwise nonsignificant. [a]Dunn post hoc test relative to control group, $P<0.05$. [b]Dunn post hoc test relative to ancestral group, $P<0.05$. For categorical variables, percent and number of participants in the variant group belonging to the listed subgroup are presented, and a χ^2 test was used for comparison. [c]χ^2 standardized Pearson residual > 2.

including ciliated cells (*CAPS*, *CFAP157*, *DNAH12*), which are responsible for mucociliary clearance; secretory and goblet cells (*AQP5*, *SERPINB3*, *VMO1*, *MUC5AC*), the main producers of mucus in the airway; basal cells (*KRT5*, *KRT15*, *COL7A1*), the main stem cells of the airway; squamous cells (*SPRR3*, *SPRR2A*, *KRT6A*), which are important in barrier function; ionocytes (*CFTR*, *RARRES2*, *SCNN1B*), rare cells for which the function is likely related to ion transport and fluid regulation[32,33]; and deuterosomal cells (*CCNO*, *CDC20B*, *HES6*), specialized ciliated precursors characterized by genes involved in centriole amplification[34,35] (Fig. 1d,e).

Within epithelial cells, we also recovered clusters of cells defined by specific transcriptional patterns, such as high amounts of SARS-CoV-2 RNA (hereafter SARS-CoV-2^{hi} cells) or mitochondrial transcripts (MT^{hi} cells), rather than cell type signatures (Fig. 1d,e). These clusters had low expression of ciliated genes (for example, *CAPS*), formed distinct clusters and persisted during iterative rounds of removal of low-quality cells (Supplementary Table 2) and as such were retained for further analysis. We detected multiple immune cell types with important roles in viral infection, including macrophages (*FCER1G*, *TYROBP*,

https://doi.org/10.1038/s41590-024-02052-z

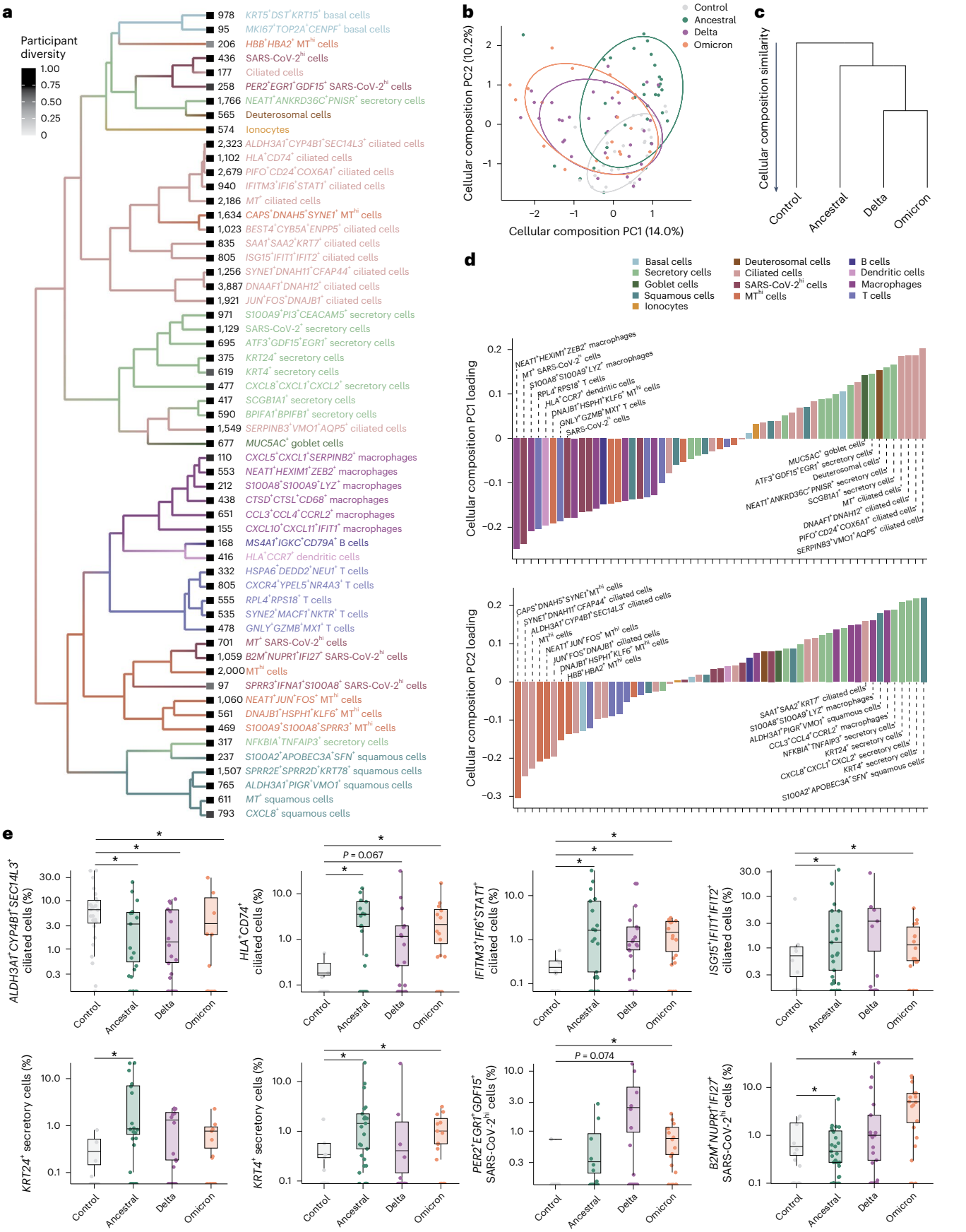


Fig. 2 | Delta and Omicron cases share nasal cell ecosystems that are distinct from those of ancestral cases. a, Cell lineage tree generated using ARBOL depicting the hierarchical clustering of the 57 epithelial and immune cell subsets identified by iterative clustering (Extended Data Figs. 3 and 4). Subset labels are colored by major cell type, and branches are colored by cell type majority. Squares represent participant diversity (Shannon's diversity index) for each subset. Cell numbers for each subset are displayed to the left of the label. *PIFO*, official gene symbol *CIMAP3*. **b**, PCA of cellular composition in the nasal mucosa in all participants. Each point represents an individual participant, and the distance reflects differences in the abundance of cell subsets as in **a**. Ellipses for control, ancestral, Delta and Omicron groups represent centroids. **c**, Hierarchical clustering of all cellular composition PCs (average of participants within variant group). **d**, Ranked loading of epithelial and immune cell subsets for cellular composition PC1 and PC2 as in **b**. Top eight positive and negative loadings are labeled for each PC. **e**, Frequency of *ALDH3A1*$^{+}$*CYP4B1*$^{+}$*SEC14L3*$^{+}$ ciliated cells (ancestral versus control, $P = 1.0 \times 10^{-4}$; Delta versus control, $P = 4.56 \times 10^{-5}$; Omicron versus control, 2.29×10^{-5}), *HLA*$^{+}$*CD74*$^{+}$ ciliated cells (ancestral versus control, $P = 0.019$; Omicron versus control, $P = 3.94 \times 10^{-3}$), *IFITM3*$^{+}$*IFI6*$^{+}$*STAT1*$^{+}$ ciliated cells (ancestral versus control, $P = 0.018$; Delta versus control, $P = 0.022$; Omicron versus control, $P = 3.56 \times 10^{-3}$), *ISG15*$^{+}$*IFIT1*$^{+}$*IFIT2*$^{+}$ ciliated cells (ancestral versus control, $P = 0.01$; Omicron versus control, $P = 0.015$), *KRT24*$^{+}$ secretory cells (ancestral versus control, $P = 9.83 \times 10^{-3}$), *KRT4*$^{+}$ secretory cells (ancestral versus control, $P = 3.9 \times 10^{-4}$; Omicron versus control, $P = 0.018$), *PER2*$^{+}$*EGR1*$^{+}$*GDF15*$^{+}$ SARS-CoV-2^{hi} cells (Omicron versus control, $P = 3.88 \times 10^{-4}$) and *B2M*$^{+}$*NUPR1*$^{+}$*IFI27*$^{+}$ SARS-CoV-2^{hi} cells (ancestral versus control, $P = 0.046$; Omicron versus control, $P = 1.58 \times 10^{-4}$) out of all cells in the sample in all controls ($n = 26$) and ancestral ($n = 32$), Delta ($n = 33$) and Omicron ($n = 21$) cases. Percentage is shown on a log scale. Box plots represent median (center line), upper and lower quartiles (box limits) and 1.5× the interquartile range (whiskers). Statistical tests are two-sided Kruskal–Wallis tests with Benjamini–Hochberg correction for multiple comparisons. *Dunn's post hoc test, $P < 0.05$.

LYZ), dendritic cells (*CCR7*, *CD86*), T cells (*CD3D*, *TRAC*, *CD8A*) and B cells (*MS4A1*, *IGKC*, *CD19*) (Fig. 1d,e). We refer to this level of annotation as 'cell types'.

To evaluate the SARS-CoV-2-induced changes in the nasal mucosa, we compared the frequency of major epithelial and immune cell types across controls and ancestral, Delta and Omicron cases. We observed significantly higher proportions of secretory (Cohen's d effect size = 1.2, $P = 4.6 \times 10^{-4}$) and deuterosomal ($d = 0.92$, $P = 5.2 \times 10^{-5}$) cells in ancestral cases than in controls (Fig. 1f). We found a reduction in the frequency of ciliated cells in both Delta ($d = 0.84$, $P = 7.8 \times 10^{-3}$) and Omicron ($d = 0.72$, $P = 0.037$) cases compared to controls (Fig. 1f) and an increase in the frequency of SARS-CoV-2^{hi} cells in all three variants, although to a greater extent in the Delta and Omicron cases compared to controls (Fig. 1f). We observed limited changes in the frequency of macrophages or T cells in each variant group compared to controls (Fig. 1f). The same trends were detected when we compared the frequencies of each major epithelial or immune cell type among all epithelial cells or all immune cells (Extended Data Fig. 2d). We next evaluated expression of the SARS-CoV-2 receptor encoded by *ACE2* as well as proteases and cofactors important for viral processing[2,36]. Expression of *ACE2* was low overall, with the highest expression in secretory, goblet, squamous and ciliated epithelial cells (Fig. 1g). Notably, SARS-CoV-2^{hi} cells had low expression of *ACE2* (Fig. 1g). The protease encoded by *TMPRSS2* was highly expressed in squamous cells (Fig. 1g). *TMPRSS4* and *CTSB* were the most highly expressed proteases and were present in multiple cell types, such as basal, secretory, goblet and squamous cells (Fig. 1g). Thus, our expanded atlas of nasal cells during SARS-CoV-2 infection revealed largely consistent cell type changes across variants, with the exception of increases in secretory and deuterosomal cells in the ancestral cases.

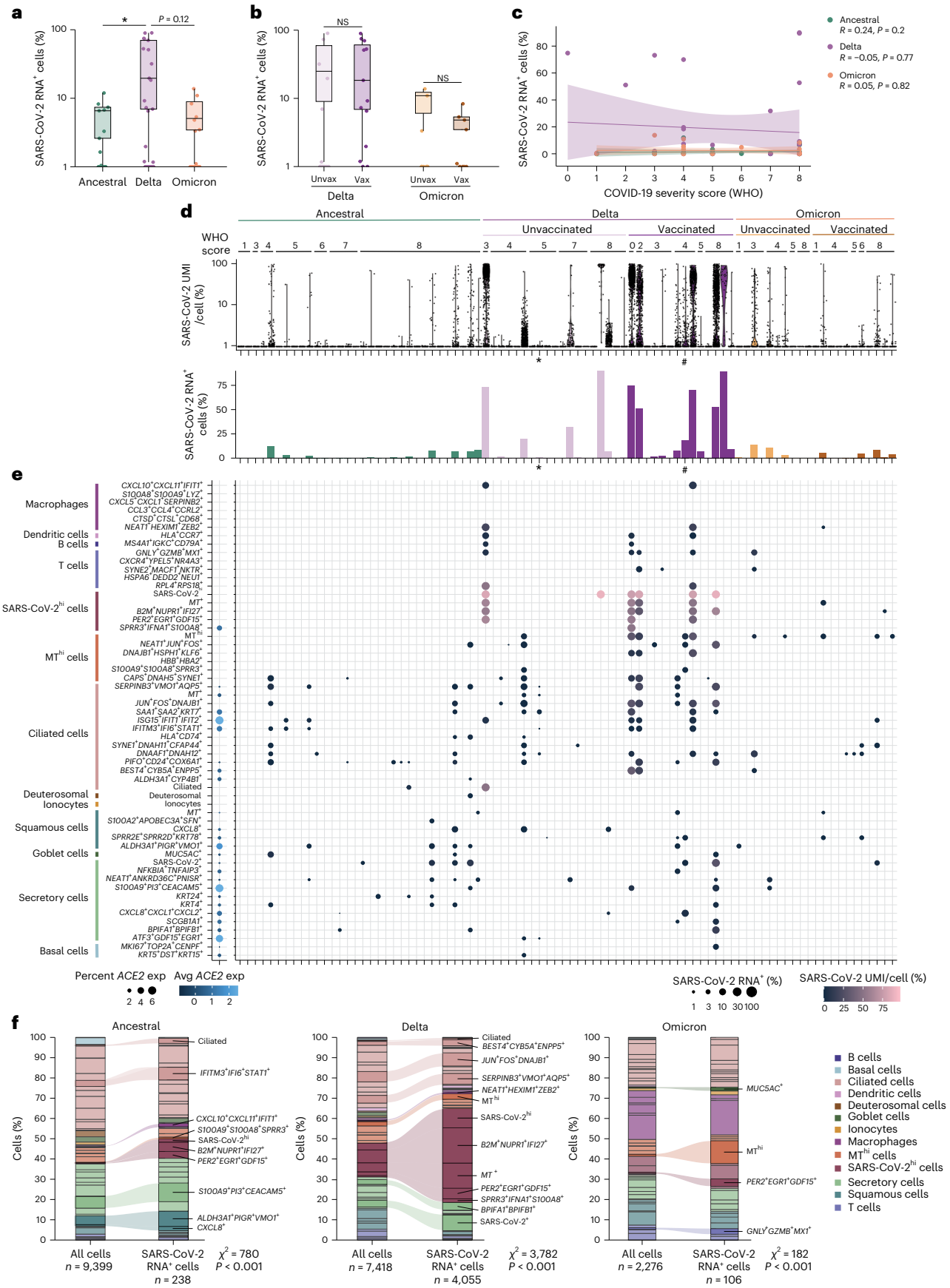
The response to SARS-CoV-2 includes diverse nasal cell subsets

Next, we interrogated the diversity of epithelial and immune cell states in the nasal mucosa during SARS-CoV-2 infection. For each major epithelial cell type with at least 1,000 cells (basal or secretory, squamous, ciliated, MT^{hi} and SARS-CoV-2^{hi}), we separated the cells, re-normalized the data and performed a new clustering analysis (Extended Data Fig. 3a–h and Supplementary Tables 3 and 4). Within secretory epithelial cells, clusters were marked by genes such as those encoding lineage markers (for example, club cell marker *SCGB1A1*, goblet cell marker *MUC5AC*), keratins (*KRT4*, *KRT24*) and neutrophil-recruitment chemokines (*CXCL1*, *CXCL2* and *CXCL8)* (Extended Data Fig. 3b,d). Analysis of the squamous epithelial cells revealed a distinct *CXCL8*^{hi} cluster (Extended Data Fig. 3c,d). The ciliated epithelial cell clusters included two subsets with high expression of cilia-forming genes (*DNAAF1*, *DNAH12*) and three subsets marked by IFN response genes (*IFITM3*, *IFI6*, *ISG15*, *HLA-A*, *HLA-B*) (Extended Data Fig. 3e,h). One of the SARS-CoV-2^{hi} clusters uniquely expressed the circadian regulators encoded by *PER1* and *PER2*, the hormone encoded by *GDF15* and the transcription factor encoded by *EGR1* (Extended Data Fig. 3g,h), consistent with a set of genes for which expression correlated with high amounts of SARS-CoV-2 RNA within nasopharyngeal ciliated cells in a SARS-CoV-2 human challenge study[30].

The same subclustering approach in immune cell types with at least 1,000 cells identified multiple subsets of T cells and myeloid cells (Extended Data Fig. 4a–d and Supplementary Table 5). The strongest signature within T cells was cytotoxicity (*GZMB*, *GNLY*), and one cluster was defined by high expression of *CXCR4* (Extended Data Fig. 4b,d). We identified multiple macrophage clusters that expressed unique patterns of chemokines: a *CCL3*$^{+}$*CCL4*$^{+}$ subset, a *CXCL5*$^{+}$*CXCL1*$^{+}$ subset, a *CXCL10*$^{+}$*CXCL11*$^{+}$ subset and a *LYZ*^{hi} monocyte-like subset (Extended

Fig. 3 | SARS-CoV-2 variants have unique patterns of viral RNA distribution among nasal cells. a,**b**, Frequency of SARS-CoV-2 RNA^{+} cells out of all cells in the sample in ancestral ($n = 32$), Delta ($n = 33$) and Omicron ($n = 21$) cases (Delta versus ancestral, $P = 0.036$) (**a**) and unvaccinated Delta ($n = 19$), vaccinated Delta ($n = 14$), unvaccinated Omicron ($n = 10$) and vaccinated Omicron ($n = 11$) cases (**b**). Box plots represent median (center line), upper and lower quartiles (box limits) and 1.5× the interquartile range (whiskers). Statistical tests are two-sided Kruskal–Wallis tests with Benjamini–Hochberg correction for multiple comparisons. *Dunn's post hoc test, $P < 0.05$. **c**, Spearman correlation between COVID-19 severity score (WHO range 0–8) and the frequency of SARS-CoV-2 RNA^{+} cells out of all cells in the sample in ancestral ($n = 32$), Delta ($n = 33$) and Omicron ($n = 21$) cases. Error bands represent 95% confidence intervals. **d**, Violin plot showing the percent of SARS-CoV-2 transcripts out of total RNA transcripts (each violin represents one participant) and bar plot showing the percentage of cells from each participant that were assigned as SARS-CoV-2 RNA^{+} cells (each bar represents one participant). The variant group (ancestral, Delta and Omicron), vaccination status and COVID-19 severity score (WHO range 0–8) are shown above the plot. Symbols indicate samples in which variants other than Delta were identified from viral sequencing: *Alpha (B.1.1.7 like), #Gamma (P.1 like). UMI, unique molecular identifier. **e**, Dot plot of *ACE2* expression (exp) across cell subsets in all participants (left) and of SARS-CoV-2 RNA^{+} cells (right) for each cell subset (rows) and participant (columns), ordered as in **d**. Dot size represents the percentage of *ACE2*$^{+}$ cells out of all cells in that subset (left) or the percentage of SARS-CoV-2 RNA^{+} cells out of cells in that subset for that participant (right). Gradient represents average (avg) scaled expression of *ACE2* (left) or the average percent of SARS-CoV-2 transcripts out of total transcripts in each cell (right). Right: dots shown are for combinations of participant and cell subset with at least ten cells. **f**, Distribution of cell subsets among all cells and SARS-CoV-2 RNA^{+} cells for a subset of ancestral ($n = 17$), Delta ($n = 19$) and Omicron ($n = 10$) cases. Participants with SARS-CoV-2 RNA^{+} cells among ancestral cases, confirmed Delta (B.1.617.2-like) and Omicron (BA.1-like) cases or probable Omicron (BA.1-like) cases identified by viral sequencing (Extended Data Fig. 1a,b) were included. A one-sided χ^2 test (ancestral, $\chi^2 = 779.88$, $P < 2.2 \times 10^{-16}$; Delta, $\chi^2 = 3781.6$, $P < 2.2 \times 10^{-16}$; Omicron, $\chi^2 = 182.3$, $P < 1.5 \times 10^{-15}$) was performed on each variant separately. Labeled detailed cell types represent those with χ^2 standardized residual > 2.

https://doi.org/10.1038/s41590-024-02052-z



Data Fig. 4c,d). Across the entire dataset, our final annotation included 57 detailed epithelial and immune clusters, hereafter referred to as 'cell subsets'. These observations highlighted the diversity of cell states in the nasal mucosa during the antiviral response to SARS-CoV-2.

Delta and Omicron drive similar cellular ecosystems

To visualize the relationships between the 57 epithelial and immune cell subsets, we generated a host cell phylogenetic 'tree'. The centroids of each cell subset were hierarchically clustered based on gene expression distances, such that transcriptionally similar subsets would have the shortest distances. As expected, subsets derived from the same major cell types mostly clustered together (Fig. 2a). Exceptions included *NFKBIA*$^+$*TNFAIP3*$^+$ secretory cells, which clustered closer to squamous cells, and *SERPINB3*$^+$*VMO1*$^+$*AQP5*$^+$ ciliated cells, which grouped with other secretory clusters (Fig. 2a), indicating cases where convergent end state transcriptional responses overshadowed original cell type identity. In some cell types, such as T cells, the subsets were all closely related to each other, while, in others (for example, ciliated or secretory cells), there was a greater amount of diversity as reflected by the hierarchical distances (Fig. 2a). Additionally, participant diversity was maintained in the cell subsets (Fig. 2a).

We next used cellular composition analysis to assess the 'cellular ecosystem' of each sample. For every participant, we calculated the abundance of all 57 immune and epithelial cell subsets normalized to 500 cells (the average total cell number across participants) (Supplementary Table 6). Principal-component analysis (PCA) indicated that control samples had less variation in nasal cellular composition and that ancestral cases had a cellular composition distinct from those of Delta and Omicron cases, which were highly overlapping (Fig. 2b). Hierarchical clustering of centroids for each variant group across all compositional principal components (PCs) indicated that Delta and Omicron cases had the greatest cellular composition similarity (Fig. 2c).

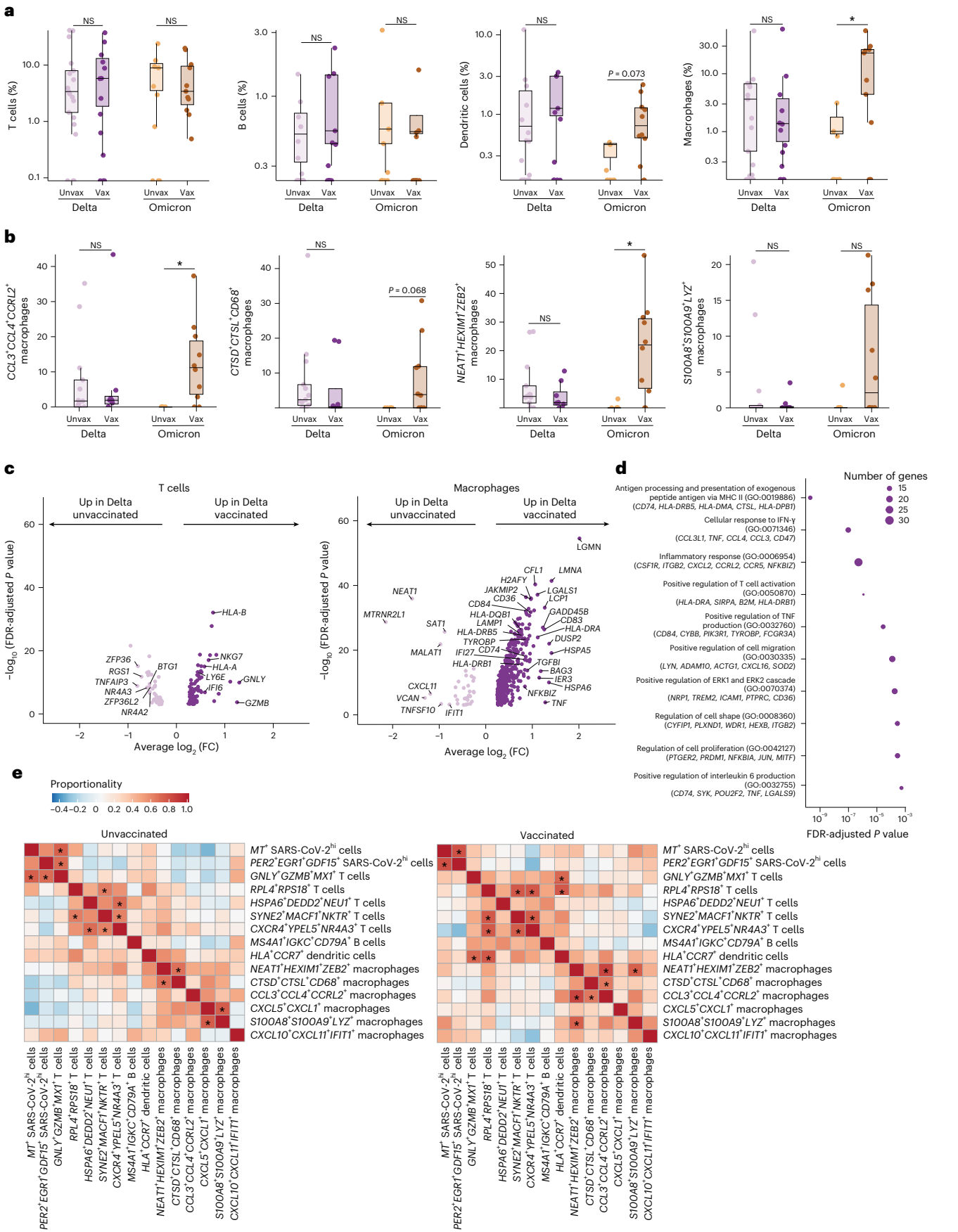
Because the first two compositional PCs defined the greatest amount of variation, we investigated which cell subsets contributed the most to each component. Ranking the subsets by PC loading indicated that the cellular composition of controls (positive PC1, negative PC2) was mostly defined by ciliated and MT^{hi} cell subsets, while the unique cellular composition of ancestral cases (positive PC1, positive PC2) was most heavily influenced by specific secretory cell, squamous cell and macrophage subsets (Fig. 2d). The strongest contributors to the shared ecosystem of Delta and Omicron cohorts (negative PC1) were macrophage, dendritic cell, T cell and SARS-CoV-2^{hi} cell subsets (Fig. 2d).

We then quantified the frequency of each cell subset across the variant groups. Controls had the highest frequency of ciliated cells with an *ALDH3A1*$^+$*CYP4B1*$^+$*SEC14L3*$^+$ resting or metabolic phenotype (Fig. 2e). There was an increase in the frequency of two IFN-responsive ciliated subsets (*IFITM3*$^+$*IFI6*$^+$*STAT1*$^+$ and *ISG15*$^+$*IFIT1*$^+$*IFIT2*$^+$) as well as an *HLA*$^+$*CD74*$^+$ (*HLA*$^+$ refers to multiple HLA genes, including *HLA-A*, *HLA-B*, and *HLA-DRB1*) cell subset in the ancestral, Delta and Omicron groups compared to controls (Fig. 2e), indicating a conserved shift in ciliated cell phenotype during SARS-CoV-2 infection. Two distinct secretory cell subsets marked by *KRT4* and *KRT24*, respectively, contributed positively to cellular composition PC2 (Fig. 2d) and were enriched in the ancestral group (*KRT24*$^+$, $d = 0.6$, $P = 9.8 \times 10^{-3}$; *KRT4*$^+$, $d = 0.54$, $P = 3.9 \times 10^{-4}$) (Fig. 2e). Delta and Omicron cases shared an increase in the frequency of *PER2*$^+$*EGR1*$^+$*GDF15*$^+$ SARS-CoV-2^{hi} epithelial cells (Delta, $d = 0.46$, $P = 0.074$; Omicron, $d = 1.12$, $P = 3.9 \times 10^{-4}$) (Fig. 2e), while a *B2M*$^+$*NUPR1*$^+$*IFI27*$^+$ SARS-CoV-2^{hi} cell subset emerged in ancestral and Omicron cases (Fig. 2e). These data suggested that, while some epithelial cell subsets, such as ciliated cells expressing IFN response and antigen presentation genes, were consistently induced in all three SARS-CoV-2 variant groups, others, such as *KRT24*$^+$ secretory cells in ancestral cases, were unique to specific variants.

To confirm the existence of variant-specific transcriptional states using an orthogonal approach, we applied non-negative matrix factorization (NMF) to the two most abundant cell types in our dataset, ciliated cells and secretory cells. Within ciliated cells, we identified 25 'factors' that represented coordinated gene programs (Supplementary Table 7). Notably, multiple factors overlapped highly with the transcriptional programs used to identify the ciliated cell subsets, such as a resting or metabolic phenotype (ciliated factor 8: *CAPS*, *ALDH1A1*, *SEC14L3*), cilia formation (ciliated factor 10: *DNAH12*, *DNAH5*, *SYNE1*), antigen presentation (ciliated factor 13: *HLA-B*, *HLA-A*, *B2M*) and IFN stimulation (ciliated factor 21: *ISG15*, *MX1*, *IFIT1*) (Extended Data Fig. 5a). We computed the average factor scores within pseudobulk ciliated cells for each participant to compare the use of the gene programs across variant groups. We observed the same trends as in the cluster-based analysis, namely, consistent decreases in genes involved in ciliogenesis and an increase in IFN-responsive programs (Fig. 2e and Extended Data Fig. 5b). NMF analysis in basal and secretory cells identified 23 factors representing coordinated gene programs (Supplementary Table 7). These included secretory factor 6 (*KRT4*, *KRT7*, *AQP5*) and secretory factor 11 (*S100A2*, *KRT24*, *KRT19*), which overlapped with the *KRT4*$^+$ and *KRT24*$^+$ subsets; secretory factor 5, which consisted of basal cell genes (*KRT5*, *KRT15*, *DST*); and factor 20, which resembled a goblet cell program (*MUC5AC*, *BPIFB1*) (Extended Data Fig. 5c). Comparison between variant groups of average scores for secretory factors 6 and 11 demonstrated an ancestral-specific increase in *KRT24* expression (Extended Data Fig. 5d), similar to the cluster-based frequency analysis (Fig. 2e). We evaluated the use of these factors across basal and secretory cell subsets and found that factor 5 was most highly expressed in the basal cell subset and factor 20 expression was highest in the goblet cell subset (Extended Data Fig. 5e). As such, the NMF analysis demonstrated that the clustering approach used here identified robust transcriptional states that could be used to compare gene programs within major cell types.

Fig. 4 | Prior vaccination is associated with nasal macrophage recruitment and activation during SARS-CoV-2 infection. a, Frequency of T cells, B cells, dendritic cells and macrophages (Omicron vaccinated versus Omicron unvaccinated, $P = 0.0168$) out of all cells in the sample in all unvaccinated Delta ($n = 19$), vaccinated Delta ($n = 14$), unvaccinated Omicron ($n = 10$) and vaccinated Omicron ($n = 11$) cases. Percentage is shown on a log scale. **b**, Frequency of *CCL3*$^+$*CCL4*$^+$*CCRL2*$^+$ (Omicron vaccinated versus Omicron unvaccinated, $P = 0.05$), *CTSD*$^+$*CTSL*$^+$*CD68*$^+$, *NEAT1*$^+$*HEXIM1*$^+$*ZEB2*$^+$ (Omicron vaccinated versus Omicron unvaccinated, $P = 0.023$) and *S100A8*$^+$*S100A9*$^+$*LYZ*$^+$ macrophage subsets as a percentage of all immune cells in the sample in Delta and Omicron cases with at least ten immune cells (Delta unvaccinated, $n = 13$; Delta vaccinated, $n = 8$; Omicron unvaccinated, $n = 5$; Omicron vaccinated, $n = 10$). **c**, Volcano plot showing genes that were DE in T cells (Delta unvaccinated, $n = 17$ cases, 507 cells; Delta vaccinated, $n = 11$ cases, 642 cells) and macrophages (Delta unvaccinated, $n = 13$ cases, 353 cells; Delta vaccinated, $n = 10$ cases, 496 cells) between unvaccinated and vaccinated Delta cases. DE genes were identified by Wilcoxon rank-sum test with Bonferonni correction for multiple comparisons and considered DE if log (fold change (FC)) > 0.25 and adjusted P value < 0.001. A full list of DE genes in T cells and macrophages between unvaccinated and vaccinated Delta cases is provided in Supplementary Table 8. *H2AFY*, official gene symbol *MACROH2A1*. **d**, GO terms in genes identified as upregulated in macrophages from vaccinated Delta cases compared to unvaccinated Delta cases (as in **c**). False discovery rate (FDR)-adjusted P value represents results from Benjamini correction for multiple comparisons. Select GO terms from the top 20 most significant are shown. A full list of GO terms in genes upregulated in macrophages from vaccinated Delta cases compared to unvaccinated Delta cases is provided in Supplementary Table 9. ERK, extracellular signal-regulated kinase. **e**, Heatmaps of proportionality between pairs of cell subsets including at least one immune cell subset from unvaccinated ($n = 29$) and vaccinated ($n = 25$) cases (Delta and Omicron grouped together). *FDR < 0.05. **a**,**b**, Box plots represent median (center line), upper and lower quartiles (box limits) and 1.5× the interquartile range (whiskers). Statistical tests are two-sided Kruskal–Wallis tests with Benjamini–Hochberg correction for multiple comparison. *Dunn's post hoc test, $P < 0.05$.

https://doi.org/10.1038/s41590-024-02052-z



SARS-CoV-2 variants have distinct viral RNA distribution

Because SARS-CoV-2 variants exhibit distinct replication dynamics, symptoms and transmission rates[9], we tested whether some of this variation was reflected in the type of nasopharynx cells targeted by the virus and the amount of viral RNA that accumulated in these cells. We quantified viral transcripts in each cell by aligning to a joint human and SARS-CoV-2 genome and assigned cells as SARS-CoV-2 RNA$^+$ cells if the amount of viral RNA was above the expected ambient amount (Extended Data Fig. 6a; Methods)[25]. The frequency of SARS-CoV-2 RNA$^+$ cells was higher in Delta samples than in ancestral ($d = 0.79$, $P = 0.036$) or Omicron ($d = 0.68$, $P = 0.12$) cases (Fig. 3a). Frequency of SARS-CoV-2 RNA$^+$ cells was not associated with vaccination status (Fig. 3b) or COVID-19 severity (Fig. 3c).

Analysis of total SARS-CoV-2 transcripts per cell and the percent of SARS-CoV-2 RNA$^+$ cells in each sample indicated consistent detection of SARS-CoV-2 transcripts across participants and higher amounts of viral RNA (both per cell and in percent of cells) in Delta cases (Fig. 3d). To check that the higher amount of SARS-CoV-2 RNA in Delta samples was not due to sampling at earlier time points after infection, we compared total viral transcripts per cell across time from symptom onset. We detected a higher amount of SARS-CoV-2 RNA per cell in Delta cases than in both ancestral and Omicron cases at matched time points (Extended Data Fig. 6b). In addition, the total transcripts captured per cell correlated weakly with total viral transcripts per cell (Extended Data Fig. 6c).

ISG15$^+$*IFIT1*$^+$*IFIT2*$^+$ and *IFITM3*$^+$*IFI6*$^+$*STAT1*$^+$ IFN-responsive ciliated cells, *S100A9*$^+$*PI3*$^+$*CEACAM5*$^+$ and *ATF3*$^+$*GDF15*$^+$*EGR1*$^+$ secretory cells and *ALDH3A1*$^+$*PIGR*$^+$*VMO1*$^+$ squamous cells had the highest expression of *ACE2* (Fig. 3e). Individual participants tended to have SARS-CoV-2 RNA$^+$ cells in multiple distinct epithelial cell subsets (Fig. 3e), indicating broad cellular targeting. SARS-CoV-2 transcripts were mostly restricted to ciliated cells, secretory cells and squamous cells in ancestral cases (Fig. 3e). In Delta cases, SARS-CoV-2 transcripts were also detected in immune, MT^{hi} and SARS-CoV-2^{hi} cells (Fig. 3e). In Omicron cases, SARS-CoV-2 transcripts were restricted to a few subsets, including *DNAAF1*$^+$*DNAH12*$^+$ and *SYNE1*$^+$*DNAH11*$^+$*CFAP44*$^+$ ciliated cells and *SPRR 2E*$^+$*SPRR2D*$^+$*KRT78*$^+$ squamous cells (Fig. 3e).

The cell subsets enriched in the SARS-CoV-2 RNA$^+$ fraction were highly variable across the SARS-CoV-2 variant groups (Fig. 3f). In ancestral cases, the enriched clusters included *IFITM3*$^+$*IFI6*$^+$*STAT1*$^+$ ciliated cells, *S100A9*$^+$*PI3*$^+$*CEACAM5*$^+$ secretory cells, *ALDH3A1*$^+$*PIGR*$^+$ *VMO1*$^+$ squamous cells and select SARS-CoV-2^{hi} subsets (Fig. 3f). Delta cases had the highest number of subsets enriched within SARS-CoV-2 RNA$^+$ cells (Fig. 3f), with *BEST4*$^+$*CYB5A*$^+$*ENPP5*$^+$ ciliated cells, *SERPINB3*$^+$*VMO1*$^+$*AQP5*$^+$ ciliated cells and *BPIFA1*$^+$*BPIFB1*$^+$ secretory cells overrepresented in SARS-CoV-2 RNA$^+$ cells only in Delta cases (Fig. 3f). Omicron cases had the fewest number of cell subsets enriched, limited to MT^{hi} cells, *MUC5AC*$^+$ goblet cells, *PER1*$^+$*EGR1*$^+$*GDF15*$^+$ SARS-CoV-2^{hi} cells and *GNLY*$^+$*GZMB*$^+$*MX1*$^+$ T cells (Fig. 3f). The only cells that were consistently enriched in SARS-CoV-2 RNA$^+$ cells in all three variant groups were the *PER1*$^+$*EGR1*$^+$*GDF15*$^+$ SARS-CoV-2^{hi} epithelial cells (Fig. 3f). Together, our data revealed patterns of viral RNA abundance and distribution that were mostly unique to each variant, with the exception of a *PER1*$^+$*EGR1*$^+$*GDF15*$^+$ SARSCoV2^{hi} subset that may be a conserved transcriptional response.

Vaccination is associated with nasal macrophage activation

We next assessed the impact of prior vaccination on nasal immune responses to SARS-CoV-2. Delta and Omicron groups were divided into unvaccinated and vaccinated groups (Extended Data Table 1 and Extended Data Fig. 1f–h). There were no significant changes in the frequency of T cells or B cells in vaccinated compared to unvaccinated cases in either variant group (Fig. 4a). Within Omicron cases, vaccinated participants had higher frequencies of dendritic cells ($d = 1.36$, $P = 0.073$) and macrophages ($d = 1.2$, $P = 0.016$) than unvaccinated participants (Fig. 4a). Although not statistically significant, we observed trends toward higher frequencies of dendritic cells, macrophages and T cells in vaccinated controls than in unvaccinated controls (Extended Data Fig. 7a). We did not observe significant differences in the frequency of major epithelial cell types in vaccinated compared to unvaccinated controls or cases (Extended Data Fig. 7b). There was no correlation between the frequency of any major immune cell type and time from vaccination (Extended Data Fig. 7c). Among macrophages in vaccinated Omicron samples (Fig. 4a), we observed increases in the frequency of multiple macrophage subsets out of all immune cells, including a *CCL3*$^+$*CCL4*$^+$*CCRL2*$^+$ subset, *a CTSD*$^+$*CTSL*$^+$*CD68*$^+$ subset and a *NEAT1*$^+$*HEXIM1*$^+$*ZEB1*$^+$ subset (Fig. 4b).

We did not observe an increase in the frequency of dendritic cells or macrophages in vaccinated compared to unvaccinated Delta cases (Fig. 4a). Differential gene expression within immune cell types in the Delta cases (the number of immune cells in unvaccinated Omicron samples was too low for this analysis) found very few differentially expressed (DE) genes in T cells, among them cytotoxicity genes such as *GZMB*, *GNLY* and *NKG7*, which were upregulated in vaccinated compared to unvaccinated cases (Fig. 4c and Supplementary Table 8). However, there was a marked phenotypic shift in macrophages in vaccinated compared to unvaccinated Delta cases (Fig. 4c and Supplementary Table 8). Gene ontology (GO) analysis of the genes upregulated in macrophages from vaccinated Delta cases compared to unvaccinated Delta cases revealed pathways including major histocompatibility complex (MHC) class II antigen presentation, IFN-γ response, tumor necrosis factor (TNF) production and cell migration (Fig. 4d and Supplementary Table 9). There were not enough genes upregulated in macrophages from unvaccinated Delta cases compared to vaccinated Delta cases to perform GO analysis. Together, these data suggest that prior mRNA vaccination is associated with macrophage recruitment and activation in the nasal mucosa during breakthrough SARS-CoV-2 infection.

To gain a holistic understanding of how vaccination may shift immune networks in the nasal response to SARS-CoV-2, we performed proportionality analysis to assess relationships between immune cell subsets and determine whether specific subsets were changing frequency in tandem. Evaluation all of pairs of subsets in the dataset that included at least one immune cell subset indicated that the majority of proportional relationships were between cell subsets within major cell types (Fig. 4e). As such, T cell subsets were more likely to change with other T cell subsets than with macrophage subsets (Fig. 4e). We observed proportional relationships between *GNLY*$^+$*GZMB*$^+$*MX1*$^+$ cytotoxic T cells and two SARS-CoV-2^{hi} cell subsets (*PER2*$^+$*EGR1*$^+$*GDF15*$^+$ and *MT*$^+$) (*MT*$^+$ refers to multiple mitochondrial genes, for example *MT-CO3*, *MT-ND5*) in unvaccinated Delta and Omicron cases grouped together (Fig. 4e). Dendritic cells varied proportionally with *GNLY*$^+$*GZMB*$^+$*MX1*$^+$ cytotoxic T cells and *RPL4*$^+$*RPS18*$^+$ T cells with a naive or memory phenotype in vaccinated Delta and Omicron cases grouped together (Fig. 4e). Thus, although more DC–T cell interactions were detected in vaccinated individuals, we did not detect marked overall changes in the nasal immune networks in vaccinated compared to unvaccinated cases.

Severe COVID-19 correlates with variant-specific phenotypes

An impaired IFN response in the upper respiratory tract has been associated with severe COVID-19 (refs. 25,26). To identify cellular and transcriptional signatures of COVID-19 severity across the SARS-CoV-2 variants, we divided participants by the amount of respiratory support required at the peak of their disease course (WHO score range 0–8). To evaluate a respiratory epithelial-specific IFN response signature, we scored cells based on a module of genes expressed in human nasal basal cells upon treatment with IFN-α (for example, *ISG15*, *IFI6*, *MX1*, *IFIT1*, *IFITM1*) or IFN-γ (for example, *CXCL10*, *CXCL11*, *WARS1*, *GBP1*, *CD74*) (Supplementary Table 10). We restricted the analysis to cell types with sufficient numbers across multiple disease severity scores (0–8) in all three SARS-CoV-2 variant cohorts, which included ciliated cells,

https://doi.org/10.1038/s41590-024-02052-z

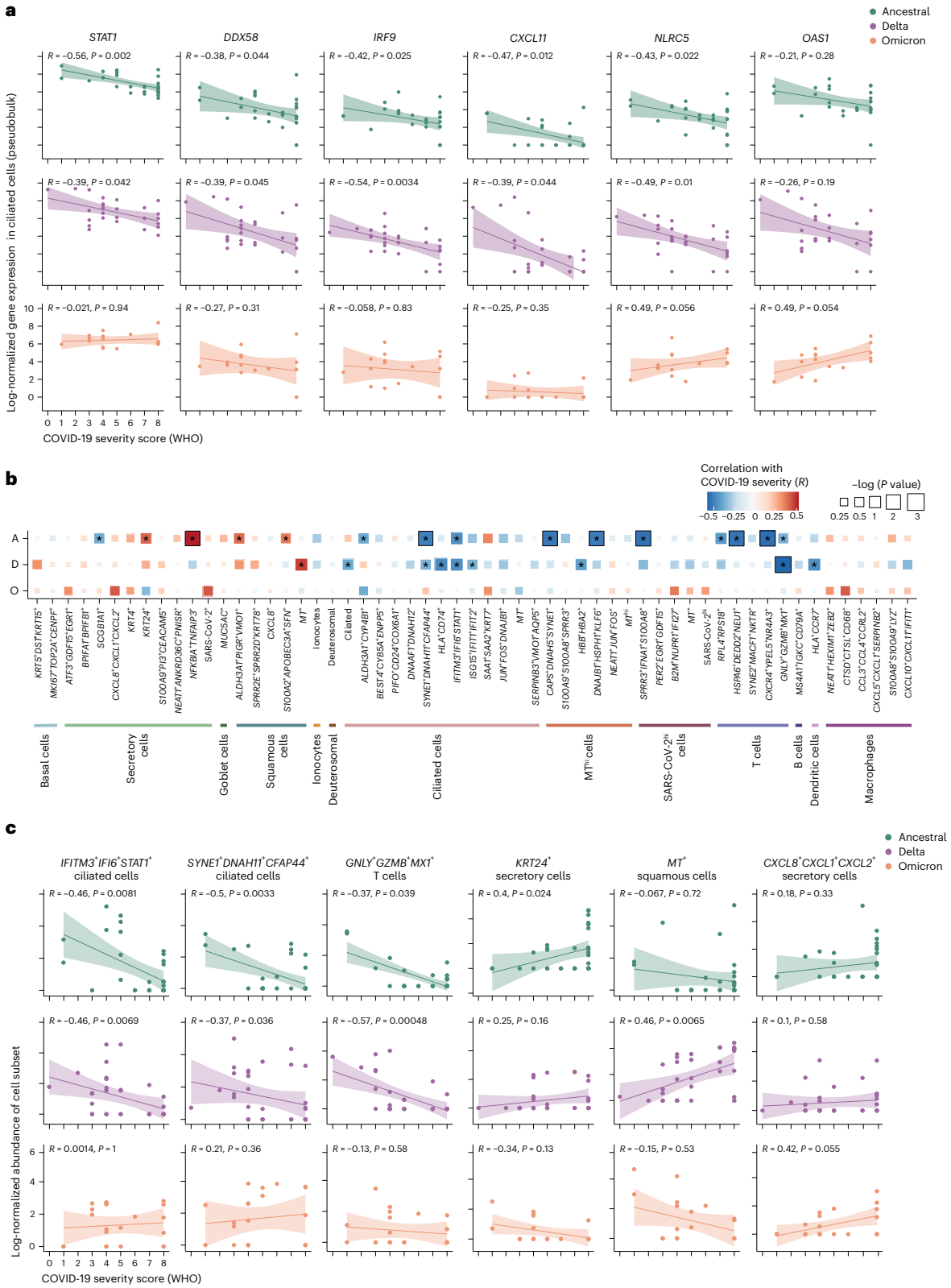


Fig. 5 | COVID-19 severity correlates with cellular and transcriptional phenotypes in a variant-specific manner. a, Spearman correlation (two-tailed test) between COVID-19 severity score (WHO range 0–8) and log-normalized expression of ISGs *STAT1*, *DDX58*, *IRF9*, *CXCL11*, *NLRC5* and *OAS1* in ciliated cells (pseudobulk per participant) from all ancestral ($n = 32$), Delta ($n = 33$) and Omicron ($n = 21$) cases. **b**, Heatmap of Spearman correlation (two-tailed test) between COVID-19 severity score (WHO range 0–8) and log-normalized abundance of cell subsets in all ancestral (A, $n = 32$), Delta (D, $n = 33$) and Omicron (O, $n = 21$) cases. Color indicates Spearman R; square size indicates −log (P value); border indicates FDR < 0.05, *$P < 0.05$. **c**, Spearman correlation as in **b** for *IFITM3*$^+$*IFI6*$^+$*STAT1*$^+$ ciliated cells, *SYNE1*$^+$*DNAH11*$^+$*CFAP44*$^+$ ciliated cells, *GNLY*$^+$*GZMB*$^+$*MX1*$^+$ T cells, *KRT24*$^+$ secretory cells, *MT*$^+$ squamous cells and *CXCL8*$^+$*CXCL1*$^+$*CXCL2*$^+$ secretory cells in all ancestral ($n = 32$), Delta ($n = 33$) and Omicron ($n = 21$) cases. **a**,**c**, Each point represents one participant. Error bands represent 95% confidence intervals.

secretory cells and T cells. In ancestral and Delta cases, the highest IFN response module score was detected in mild cases (WHO score 0–1) and the lowest module score was found in the most severe cases (WHO score 8) (Extended Data Fig. 8). Within Omicron cases, IFN response module scores were lower than in ancestral or Delta cases and increased slightly with COVID-19 severity (Extended Data Fig. 8).

Next, we used cell type-level pseudobulk analysis to assess whether IFN-stimulated gene (ISG) expression within a particular cell type correlated with COVID-19 severity. We used ciliated cells, which were best represented in the largest number of participants across severity scores in ancestral, Delta and Omicron cases. The expression of ISGs *STAT1*, *DDX58 (RIGI)*, *IRF9*, *CXCL1* and *NLRC5* in ciliated cells negatively correlated with COVID-19 severity in both ancestral and Delta cases (Fig. 5a). By contrast, in Omicron cases, the expression of ISGs in ciliated cells either did not vary (*STAT1*, *DDX58*, *IRF9*, *CXCL11*) or correlated positively (*NLRC5*, *OAS1*) with COVID-19 severity (Fig. 5a), suggesting that a robust nasal epithelial IFN response may not be protective against severe COVID-19 or may even be detrimental in Omicron infections.

Because the pattern of impaired IFN responses in severe COVID-19 was not consistent across variant groups, we performed a more unbiased analysis of severity correlates. Evaluation of the relationships between the frequencies of major epithelial and immune cell types and SARS-CoV-2 RNA$^+$ cells with all metadata variables, including COVID-19 severity score, within the entire dataset showed that the frequency of secretory cells and SARS-CoV-2 RNA$^+$ cells positively correlated with COVID-19 severity while the frequency of ciliated cells negatively correlated with severity (Extended Data Fig. 9 and Supplementary Table 11). The frequencies of secretory cells, SARS-CoV-2 RNA$^+$ cells and squamous cells positively correlated with the frequency of neutrophils within white blood cells in the peripheral blood (Extended Data Fig. 9), a known sign of severe lung disease[37].

We also evaluated whether the abundance of any specific immune or epithelial cell subset correlated with the severity score. Multiple ciliated cell subsets, including an *IFITM3*$^+$*IFI6*$^+$*STAT1*$^+$ IFN-responsive cell subset and a *SYNE1*$^+$*DNAH11*$^+$*CFAP44*$^+$ cilia^{hi} cell subset, negatively correlated with severity in ancestral and Delta cases but not in Omicron cases (Fig. 5b,c). The abundance of *GNLY*$^+$*GZMB*$^+$*MX1*$^+$ cytotoxic T cells negatively correlated with severity in ancestral and Delta samples, while other T cell subsets, such as *CXCR4*$^+$*YPEL5*$^+$*NR4A3*$^+$ T cells, negatively correlated with severity in ancestral cases only (Fig. 5b,c). We also detected variant-specific severity signatures. In ancestral cases only, *KRT24*$^+$ and *NFKBIA*$^+$*TNFAIP3*$^+$ secretory cells positively correlated with severity, while a *SCGB1A1*$^+$ secretory subset resembling club cells negatively correlated with severity (Fig. 5b,c). Two subsets of squamous cells (*ALDH3A1*$^+$*PIGR*$^+$*VMO1*$^+$ and *S100A2*$^+$*APOBEC3A*$^+$*SFN*$^+$) positively correlated with severity in ancestral cases only, while *MT*$^+$ squamous cells positively correlated with severity in Delta cases only (Fig. 5b,c). Fewer cell subsets significantly correlated with severity in either direction in Omicron cases (Fig. 5b), with the strongest positive correlation found in *CXCL8*$^+$*CXCL1*$^+$*CXCL2*$^+$ secretory cells, SARS-CoV-2 RNA$^+$ secretory cells and *CTSD*$^+$*CTSL*$^+$*CD68*$^+$ macrophages (Fig. 5b,c). These observations indicated that distinct nasal cell subsets were associated with disease severity during infection with each SARS-CoV-2 variant.

Discussion

Here, we present a detailed characterization of the human nasal cellular response to SARS-CoV-2 infection across three viral variants, vaccination status and disease severity. By defining the diverse cell types, subsets and states present in the human nasal mucosa during acute infection with SARS-CoV-2, we discovered convergent cellular ecosystems in Delta and Omicron cases that were distinct from those in ancestral cases. Characterization of SARS-CoV-2 RNA across variants revealed a *PER2*$^+$*EGR1*$^+$*GDF15*$^+$ epithelial subset that was consistently enriched in SARS-CoV-2 RNA$^+$ cells. Comparison of responses between vaccinated and unvaccinated individuals showed that prior vaccination was associated with nasal macrophage activation during acute infection with Delta and Omicron. Assessment of correlates of severity across variants indicated that an impaired nasal IFN response was no longer associated with severe COVID-19 during infection with Omicron and that each SARS-CoV-2 variant had distinct cellular subsets arising in the most severe cases. Our work identified unique responses in the nasal mucosa and underscores the importance of studying viral immunity in native human barrier tissues.

Despite the divergent phylogeny of Delta and Omicron SARS-CoV-2 variants, which underlies distinct properties including replication rate, transmission kinetics and immune evasion[9], we found an overlapping nasal cellular ecosystem in Delta and Omicron cases that was driven by immune and SARS-CoV-2^{hi} cell subsets. The contribution of macrophage, dendritic cell and T cell subsets to this ecosystem could reflect increased immune infiltration due to vaccination, higher rates of prior exposure or combinations of these factors. The cellular composition of ancestral cases was distinct and was defined in part by *KRT4*$^+$ and *KRT24*$^+$ secretory cell subsets and deuterosomal cells. *KRT4* is a marker of suprabasal cells and differentiating basal cells[38,39], while deuterosomal cells are known as a precursor population for ciliated cells[34,35]. Considering that the ancestral cohort had not been vaccinated or infected with SARS-CoV-2, these observations could suggest that, in the absence of adaptive immune memory, unique epithelial differentiation pathways emerge during the antiviral response.

We found higher overall levels of viral RNA in Delta cases and unique subsets enriched in the SARS-CoV-2 RNA$^+$ fraction in each variant. Notably, only the *PER2*$^+$*EGR1*$^+$*GDF15*$^+$ SARS-CoV-2^{hi} epithelial cell subset was consistently enriched across all three SARS-CoV-2 variants. This gene signature is consistent with a set of genes identified as highly correlated with SARS-CoV-2 RNA in a SARS-CoV-2 human challenge study[30]. In macrophages, the circadian regulator encoded by *PER2* has been reported to modulate the expression and activity of Toll-like receptor 9 (TLR9)[40], while its cooperating protein PER1 has an anti-inflammatory role[41,42]. A relationship between circadian rhythms and inflammation in airway epithelial cells was described in club cells from murine lung explants[43]. The metabolic hormone GDF15 has been described as a host tolerance factor during acute inflammation[44] and was upregulated in the nasal mucosa in a proteomic study in individuals with long COVID-19 (ref. 45). EGR1 was reported as a host restriction factor for SARS-CoV-2 replication[46]. The conserved *PER2*$^+$*EGR1*$^+$*GDF15*$^+$ SARS-CoV-2^{hi} epithelial cell subset may thus reflect an acute host cell resilience program with long-term detrimental consequences.

Given the evidence that infection-induced mucosal immunity can provide stronger protection than peripheral immunity induced by vaccination alone[14,15], we expected to see a subtle phenotype in nasal adaptive immune cells in vaccinated individuals. While the frequencies and gene expression profiles of adaptive immune cells were similar between unvaccinated and vaccinated participants, we found evidence of macrophage activation in vaccinated individuals compared to unvaccinated individuals. This phenotype could arise from an increase in Fc receptor-mediated activation from vaccine-elicited antibodies or could be a response to cytokines from vaccine-induced memory T cells, but more data are necessary to make these connections. Evidence of trained immunity via epigenetic alterations in monocytes and macrophages has been reported following both SARS-CoV-2 vaccination and infection[47–50]. Our data raise the possibility that trained immunity (arising from vaccination and/or prior infection) in peripheral monocytes could result in enhanced macrophage recruitment and activation in the respiratory tract during acute infection.

During the initial waves of SARS-CoV-2 infection, interruption or impairment of the IFN response was associated with severe COVID-19 (refs. 22–26). We found that the expression of ISGs and abundance of IFN-responsive cell subsets in the nasal mucosa associated with mild symptoms in ancestral and Delta cases but not Omicron cases. Instead, the expression of ISGs such as *OAS1* and *NLRC5* in ciliated cells

positively correlated with COVID-19 severity in Omicron cases. We also identified epithelial cell subsets that correlated with severity in a variant-specific manner. These results highlight the need to continue to investigate how immune and inflammatory responses to SARS-CoV-2 relate to disease outcome as novel variants and subvariants emerge and as individuals are frequently reinfected and boosted with a rapidly evolving viral species.

Our study has several limitations. First, the inability to cleanly resolve prior exposure by serology prevented us from distinguishing between vaccine-induced, infection-induced and hybrid immune memory. Second, as we have sampled only one time point per individual, we were unable to comment on the dynamics of the immune response. Third, because we used frozen nasal swab samples with low cell yields, our data originate from lower cell numbers and lower-quality cells than in samples such as peripheral blood mononuclear cells. The low number of cells impacted our ability to perform particular analyses, including pseudobulk differential expression between variant cohorts. Fourth, the analysis based on single-cell cluster identification is dependent on upstream processing steps and in part on operator bias, which may impact reproducibility and robustness. Finally, detection of SARS-CoV-2 RNA within a cell may not accurately reflect productive infection, and we were not able to perform additional assays to confirm the location and identity of virally infected cells. Together, our study provides a comprehensive view of how the ecosystem of cells in the nasal mucosa responding to SARS-CoV-2 infection changes across viral variants, vaccination status and COVID-19 severity. This analysis sheds light on the complexity and variability of the antiviral response in the nasal mucosa and may help the development of future vaccines and therapeutics for SARS-CoV-2 and other respiratory viruses.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41590-024-02052-z.

References

1. Gallo, O., Locatello, L. G., Mazzoni, A., Novelli, L. & Annunziato, F. The central role of the nasal microenvironment in the transmission, modulation, and clinical progression of SARS-CoV-2 infection. *Mucosal Immunol.* **14**, 305–316 (2021).
2. Sungnak, W. et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* **26**, 681–687 (2020).
3. Ahn, J. H. et al. Nasal ciliated cells are primary targets for SARS-CoV-2 replication in the early stage of COVID-19. *J. Clin. Invest.* **131**, e148517 (2021).
4. Mlcochova, P. et al. SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* **599**, 114–119 (2021).
5. Puhach, O., Meyer, B. & Eckerle, I. SARS-CoV-2 viral load and shedding kinetics. *Nat. Rev. Microbiol.* **21**, 147–161 (2023).
6. Shuai, H. et al. Attenuated replication and pathogenicity of SARS-CoV-2 B.1.1.529 Omicron. *Nature* **603**, 693–699 (2022).
7. Suzuki, R. et al. Attenuated fusogenicity and pathogenicity of SARS-CoV-2 Omicron variant. *Nature* **603**, 700–705 (2022).
8. Willett, B. J. et al. SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. *Nat. Microbiol.* **7**, 1161–1179 (2022).
9. Carabelli, A. M. et al. SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat. Rev. Microbiol.* **21**, 162–177 (2023).
10. Minkoff, J. M. & tenOever, B. Innate immune evasion strategies of SARS-CoV-2. *Nat. Rev. Microbiol.* **21**, 178–194 (2023).
11. Sette, A. & Crotty, S. Immunological memory to SARS-CoV-2 infection and COVID-19 vaccines. *Immunol. Rev.* **310**, 27–46 (2022).
12. Ssemaganda, A. et al. Expansion of cytotoxic tissue-resident $CD8^+$ T cells and $CCR6^+CD161^+$ $CD4^+$ T cells in the nasal mucosa following mRNA COVID-19 vaccination. *Nat. Commun.* **13**, 3357 (2022).
13. Puhach, O. et al. SARS-CoV-2 convalescence and hybrid immunity elicits mucosal immune responses. *EBioMedicine* **98**, 104893 (2023).
14. Pieren, D. K. J. et al. Limited induction of polyfunctional lung-resident memory T cells against SARS-CoV-2 by mRNA vaccination compared to infection. *Nat. Commun.* **14**, 1887 (2023).
15. Mitsi, E. et al. Respiratory mucosal immune memory to SARS-CoV-2 after infection and vaccination. *Nat. Commun.* **14**, 6815 (2023).
16. Morens, D. M., Taubenberger, J. K. & Fauci, A. S. Rethinking next-generation vaccines for coronaviruses, influenzaviruses, and other respiratory viruses. *Cell Host Microbe* **31**, 146–157 (2023).
17. Meyerowitz, E. A., Scott, J., Richterman, A., Male, V. & Cevik, M. Clinical course and management of COVID-19 in the era of widespread population immunity. *Nat. Rev. Microbiol.* **22**, 75–88 (2024).
18. Lucas, C. et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature* **584**, 463–469 (2020).
19. Arunachalam, P. S. et al. Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* **369**, 1210–1220 (2020).
20. Unterman, A. et al. Single-cell multi-omics reveals dyssynchrony of the innate and adaptive immune system in progressive COVID-19. *Nat. Commun.* **13**, 440 (2022).
21. Diray-Arce, J. et al. Multi-omic longitudinal study reveals immune correlates of clinical course among hospitalized COVID-19 patients. *Cell Rep. Med.* **4**, 101079 (2023).
22. Bastard, P. et al. Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* **370**, eabd4585 (2020).
23. van der Wijst, M. G. P. et al. Type I interferon autoantibodies are associated with systemic immune alterations in patients with COVID-19. *Sci. Transl. Med.* **13**, eabh2624 (2021).
24. Zhang, Q. et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **370**, eabd4570 (2020).
25. Ziegler, C. G. K. et al. Impaired local intrinsic immunity to SARS-CoV-2 infection in severe COVID-19. *Cell* **184**, 4713–4733 (2021).
26. Sposito, B. et al. The interferon landscape along the respiratory tract impacts the severity of COVID-19. *Cell* **184**, 4953–4968 (2021).
27. Chua, R. L. et al. COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. *Nat. Biotechnol.* **38**, 970–979 (2020).
28. Loske, J. et al. Pre-activated antiviral innate immunity in the upper airways controls early SARS-CoV-2 infection in children. *Nat. Biotechnol.* **40**, 319–324 (2021).
29. Yoshida, M. et al. Local and systemic responses to SARS-CoV-2 infection in children and adults. *Nature* **602**, 321–327 (2022).
30. Lindeboom, R. G. H. et al. Human SARS-CoV-2 challenge uncovers local and systemic response dynamics. *Nature* **631**, 189–198 (2024).
31. Hewitt, R. J. & Lloyd, C. M. Regulation of immune responses by the airway epithelial cell landscape. *Nat. Rev. Immunol.* **21**, 347–362 (2021).
32. Montoro, D. T. et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324 (2018).

33. Plasschaert, L. W. et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377–381 (2018).
34. Ruiz García, S. et al. Novel dynamics of human mucociliary differentiation revealed by single-cell RNA sequencing of nasal epithelial cultures. *Development* **146**, dev177428 (2019).
35. Revinski, D. R. et al. CDC20B is required for deuterosome-mediated centriole production in multiciliated cells. *Nat. Commun.* **9**, 4668 (2018).
36. Jackson, C. B., Farzan, M., Chen, B. & Choe, H. Mechanisms of SARS-CoV-2 entry into cells. *Nat. Rev. Mol. Cell Biol.* **23**, 3–20 (2022).
37. Schulte-Schrepping, J. et al. Severe COVID-19 is marked by a dysregulated myeloid cell compartment. *Cell* **182**, 1419–1440 (2020).
38. Cumplido-Laso, G., Benitez, D. A., Mulero-Navarro, S. & Carvajal-Gonzalez, J. M. Transcriptional regulation of airway epithelial cell differentiation: insights into the Notch pathway and beyond. *Int. J. Mol. Sci.* **24**, 14789 (2023).
39. Zhou, Y. et al. Airway basal cells show regionally distinct potential to undergo metaplastic differentiation. *eLife* **11**, e80083 (2022).
40. Silver, A. C., Arjona, A., Walker, W. E. & Fikrig, E. The circadian clock controls Toll-like receptor 9-mediated innate and adaptive immunity. *Immunity* **36**, 251–261 (2012).
41. Katoku-Kikyo, N. et al. The circadian regulator PER1 promotes cell reprogramming by inhibiting inflammatory signaling from macrophages. *PLoS Biol.* **21**, e3002419 (2023).
42. Wang, T. et al. PER1 prevents excessive innate immune response during endotoxin-induced liver injury through regulation of macrophage recruitment in mice. *Cell Death Dis.* **7**, e2176 (2016).
43. Gibbs, J. et al. An epithelial circadian clock controls pulmonary inflammation and glucocorticoid action. *Nat. Med.* **20**, 919–926 (2014).
44. Luan, H. H. et al. GDF15 is an inflammation-induced central mediator of tissue tolerance. *Cell* **178**, 1231–1244 (2019).
45. Liew, F. et al. Large-scale phenotyping of patients with long COVID post-hospitalization reveals mechanistic subtypes of disease. *Nat. Immunol.* **25**, 607–621 (2024).
46. Zhao, Y. et al. EGR1 functions as a new host restriction factor for SARS-CoV-2 to inhibit virus replication through the E3 ubiquitin ligase MARCH8. *J. Virol.* **97**, e0102823 (2023).
47. Arunachalam, P. S. et al. Systems vaccinology of the BNT162b2 mRNA vaccine in humans. *Nature* **596**, 410–416 (2021).
48. Tahtinen, S. et al. IL-1 and IL-1ra are key regulators of the inflammatory response to RNA vaccines. *Nat. Immunol.* **23**, 532–542 (2022).
49. Yamaguchi, Y. et al. Consecutive BNT162b2 mRNA vaccination induces short-term epigenetic memory in innate immune cells. *JCI Insight* **7**, e163347 (2022).
50. Cheong, J.-G. et al. Epigenetic memory of coronavirus infection in innate immune cells and their progenitors. *Cell* **186**, 3882–3902 (2023).

[1]Division of Gastroenterology, Hepatology and Nutrition, Boston Children's Hospital, Boston, MA, USA. [2]Department of Immunology, Harvard Medical School, Boston, MA, USA. [3]Program in Immunology, Harvard Medical School, Boston, MA, USA. [4]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [5]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA. [6]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [7]Program in Health Sciences and Technology, Harvard Medical School and MIT, Boston, MA, USA. [8]Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. [9]Division of Digestive Diseases, University of Mississippi Medical Center, Jackson, MS, USA. [10]Graduate Program in Microbiology, Massachusetts Institute of Technology, Cambridge, MA, USA. [11]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA. [12]Division of Pediatric Hematology/Oncology, Boston Children's Hospital, Boston, MA, USA. [13]Harvard Graduate Program in Biophysics, Cambridge, MA, USA. [14]Division of Pulmonary, Critical Care and Sleep Medicine, University of Mississippi Medical Center, Jackson, MS, USA. [15]Department of Cell and Molecular Biology, University of Mississippi Medical Center, Jackson, MS, USA. [16]Department of Oral Immunology and Infectious Diseases, University of Louisville School of Dentistry, Louisville, KY, USA. [17]Department of Pathology, University of Mississippi Medical Center, Jackson, MS, USA. [18]Division of Allergy and Immunology, Department of Medicine, University of California San Diego, La Jolla, CA, USA. [19]Veterans Affairs San Diego Healthcare System, La Jolla, CA, USA. [20]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA. [21]Division of Emergency Medicine, Boston Children's Hospital, Boston, MA, USA. [22]Department of Medicine, Section of Gastroenterology and Hepatology, Tulane University, New Orleans, LA, USA. [23]Harvard Stem Cell Institute, Cambridge, MA, USA. [24]These authors contributed equally: Jaclyn M. L. Walsh, Vincent N. Miao, Anna H. Owings, Ying Tang, Joshua D. Bromley. [25]These authors jointly supervised this work: Alex K. Shalek, Bruce H. Horwitz, Sarah C. Glover, Jose Ordovas-Montanes. ✉e-mail: jose.ordovas-montanes@childrens.harvard.edu

Methods

Participant details

Eligible participants were recruited from outpatient clinics, medical surgical units, intensive care units or endoscopy units at the UMMC between April 2020 and September 2022. The UMMC Institutional Review Board approved the study under IRB 2020-0065, and all research complies with the relevant ethical regulations. All participants or their legally authorized representative provided written informed consent. Participants did not receive compensation. Participants were eligible for inclusion in the SARS-CoV-2 infected cohorts if they were at least 18 years old, had a positive nasopharyngeal swab for SARS-CoV-2 by PCR, had COVID-19-related symptoms including fever, chills, cough, shortness of breath and sore throat and weighed more than 110 lb. Participants were eligible for the control group if they were at least 18 years old, had a current negative SARS-CoV-2 test (PCR or rapid antigen test) and weighed more than 110 lb. Exclusion criteria for the cohort included a history of blood transfusion within 4 weeks of sample collection and an inconclusive SARS-CoV-2 infection diagnosis based on nucleic acid testing. A total of 112 individuals were included in the study: 32 individuals sampled during the ancestral wave of COVID-19 (April–November 2020; 15 female, 17 male, median age of 58.5 years), 33 individuals during the Delta wave (July–September 2021; 14 female, 19 male, median age of 50 years), 21 during the Omicron wave (January–February 2022; six female, 15 male, median age of 61 years) and 26 healthy control participants (13 collected April–November 2020, six collected August–October 2021, seven collected January–February 2022; 17 female, ten male, median age of 51.5 years). Participants were assigned to the vaccinated group if they had received an mRNA or vector-based vaccine for COVID-19 before sample collection and to the unvaccinated group if they had not received any COVID-19 vaccine. COVID-19 cases were classified according to the eight-level ordinal scale proposed by the WHO representing the amount of respiratory support required at the peak of their disease (0–8 scale: 0, asymptomatic; 1, no limitation of activities; 2, limitation of activities; 3, hospitalized, no oxygen therapy; 4, oxygen by mask or nasal prongs; 5, non-invasive ventilation or high-flow oxygen; 6, intubation and mechanical ventilation; 7, ventilation and additional organ support; 8, death)[51]. Full clinical metadata of all participants are provided (Supplementary Table 1). No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to or larger than those reported in previous publications[25,27–30]. Data collection and analysis were not performed blind to the conditions of the experiments.

Inclusion and ethics

We worked to ensure gender balance in the recruitment of human participants. We worked to ensure ethnic or other types of diversity in the recruitment of human participants. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis and/or interpretation of the work.

Sample collection

Nasopharyngeal samples were collected by a trained healthcare provider using FLOQSwabs (Copan flocked swabs) following the manufacturer's instructions. Collectors would don personal protective equipment. The nasopharyngeal swabs were collected from patients in the standard clinical fashion. The swab was then placed into a cryogenic vial with 900 µl of heat-inactivated FBS and 100 µl dimethylsulfoxide. Vials were placed into a Mr. Frosty Freezing Container (Thermo Fisher Scientific) for optimal cell preservation. A Mr. Frosty containing the vials was placed in a cooler with dry ice for transportation from patient areas to the laboratory. Once in the laboratory, the Mr. Frosty was placed into a −80 °C freezer overnight, and, on the next day, the vials were moved to liquid nitrogen storage containers.

Detection of SARS-CoV-2-specific antibodies

Plasma from a subset of participants (14 control, 24 ancestral, 30 Delta, 17 Omicron) was collected on the same day as the nasopharyngeal swab. SARS-CoV-2-specific antibodies were detected using an ELISA with recombinant SARS-CoV-2 proteins. Briefly, 384-well MaxiSorp ELISA plates (Thermo Fisher Nunc) were coated with recombinantly expressed protein at a concentration of 3 µg ml^{-1} and incubated overnight at 4 °C. Plates were washed 4× with a BioTek 405 TS plate washer and blocked with 1% powdered milk solution in PBS for 1 h at room temperature. Samples were serially diluted in 1% powdered milk–PBS–Tween and applied to the plate. Plates were incubated for 2 h at room temperature before washing 4×. Horseradish peroxidase-conjugated anti-human IgG or IgM (SouthernBiotech) was applied to the plates and incubated for 1 h at room temperature. Plates were washed 6× and developed with tetramethylbenzidine (Sigma) for 30 min. The reaction was stopped by the addition of 1 M H_2SO_4, and absorbance was read at 450 nm. The endpoint titer was calculated as the inverse of the sample dilution that produced an absorbance of 0.2 absorbance units above background. All assays were performed in triplicate.

Dissociation and collection of viable single cells from nasopharyngeal swabs

Swabs in freezing medium (90% FBS, 10% dimethylsulfoxide) were stored in liquid nitrogen until immediately before dissociation. A detailed sample protocol can be found here: https://protocols.io/view/human-nasopharyngeal-swab-processing-for-viable-si-bjhkkj4w.html (ref. 52). Briefly, nasal swabs in freezing medium were thawed, and each swab was rinsed with RPMI before incubation in 1 ml RPMI, 10 mM DTT (Sigma) for 15 min at 37 °C with agitation. Next, the nasal swab was incubated in 1 ml Accutase (Innovative Cell Technologies) for 30 min at 37 °C with agitation. The 1 ml of RPMI with 10 mM DTT from the nasal swab incubation was centrifuged at 400*g* for 5 min at 4 °C to pellet cells, the supernatant was discarded, and the cell pellet was resuspended in 1 ml Accutase and incubated for 30 min at 37 °C with agitation. The original cryovial containing the freezing medium and the original swab washes were combined and centrifuged at 400*g* for 5 min at 4 °C. The cell pellet was then resuspended in RPMI with 10 mM DTT and incubated for 15 min at 37 °C with agitation and centrifuged as above, the supernatant was aspirated, and the cell pellet was resuspended in 1 ml Accutase and incubated for 30 min at 37 °C with agitation. All cells were combined following Accutase digestion and filtered using a 70-µm nylon strainer. The filter and the swab were washed with RPMI with 10% FBS and 4 mM EDTA, and all washes were combined. Dissociated, filtered cells were centrifuged at 400*g* for 10 min at 4 °C and resuspended in 200 µl RPMI with 10% FBS for counting. Cells were diluted to 20,000 cells in 200 µl for scRNA-seq. If fewer than 20,000 cells in total were recovered, all cells were input into scRNA-seq.

Viral variant sequencing

RNA was isolated from supernatants collected during swab dissociation using the QIAamp MinElute Virus Spin Kit (Qiagen, 57704) for RNA extraction. The input volume was 100 µl of supernatant, and the final elution volume was 50 µl. Moreover, 10 µl of each sample was used for PCR detection of SARS-CoV-2 genes. Samples with a CT value < 28 were sent to the UMMC Molecular and Genomics Core Facility, and sequencing was performed using the Illumina COVIDSeq Test and platform. COVIDSeq is an amplicon-based next-generation sequencing test based on 2019-nCoV primers designed to detect RNA from the SARS-CoV-2 virus (based on ARTIC multiplex PCR protocol version 4, with 98 amplicons), along with an internal control consisting of 11 human mRNA targets. Individual samples were processed and barcoded using the COVIDSeq library prep kit (according to manufacturer instructions) using an automated PerkinElmer Zephyr G3 next-generation sequencing instrument. Subsequently, samples were pooled (96–384 samples) and run on a NextSeq 500 instrument using a MO flow cell (150 cycle)

or a HO flow cell (150 cycle) depending on the number of samples in the pooled library. The DRAGEN COVIDSeq Test App (via the Illumina BaseSpace Cloud Computing Platform) was used for alignment of reads to confirm positive samples as well as to identify variants and clade.

Single-cell RNA sequencing

Seq-Well S^3 was used for scRNA-seq[25,53–57]. Briefly, a maximum of 20,000 single cells were deposited onto Seq-Well arrays preloaded with a single barcoded mRNA capture bead (ChemGenes) per well[58]. Cells were allowed to settle by gravity into wells for 10 min, after which the arrays were washed with PBS and RPMI, sealed with a semi-permeable membrane for 30 min and incubated in lysis buffer (5 M guanidinium thiocyanate, 1 mM EDTA, 1% BME, 0.5% sarkosyl) for 20 min. Arrays were then incubated in a hybridization buffer (2 M NaCl, 8% (vol/vol) PEG 8000) for 40 min, and then the beads were removed from the arrays and collected in 1.5-ml tubes in wash buffer (2 M NaCl, 3 mM $MgCl_2$, 20 mM Tris-HCl, 8% (vol/vol) PEG 8000). Beads were resuspended in a reverse transcription master mix, and reverse transcription, exonuclease digestion, second-strand synthesis and whole-transcriptome amplification were carried out as previously described. Libraries were generated using Illumina Nextera XT Library Prep Kits and sequenced using NovaSeq S4 kits at the Broad Institute Sequencing Core: read 1, 21 (cell barcode, UMI); read 2, 50 (digital gene expression); index 1, 8 (N700 barcode).

Data preprocessing

Libraries were aligned using STAR within the Drop-seq Computational Protocol (https://github.com/broadinstitute/Drop-seq) and implemented in Cumulus (https://cumulus.readthedocs.io/en/latest/drop_seq.html, snapshot 11, default parameters)[58]. A custom reference of the combined human GRCh38 (from Cell Ranger version 3.0.0, Ensembl 93) and SARS-CoV-2 RNA genomes was used for alignment[25]. Aligned cell-by-gene matrices for each array were merged into one Seurat object across all participants. After CellBender correction, cells were filtered to remove those with fewer than 200 UMIs, fewer than 150 unique genes and more than 50% mitochondrial reads. This resulted in a dataset of 32,994 genes and 55,319 cells across 112 study participants (26 SARS-CoV-2$^-$, 86 SARS-CoV-2$^+$).

SARS-CoV-2 RNA$^+$ assignment

We first aligned to a custom reference genome containing both human and SARS-CoV-2 genes and quantified the total number of SARS-CoV-2 UMIs in each cell[25,59]. Next, we determined the proportion of ambient RNA in each cell by comparing the raw matrices with CellBender-corrected matrices. The CellBender (version 0.2.0) remove-background function was run with default parameters and epochs = 50, expected_cells = 500, hardware_boot_disk_size_GB = 100, hardware_disk_size_GB = 100, low_count_threshold = 5 and total_droplets_included = 10,000. Including the top 10,000 barcodes enabled quantification of viral reads in empty wells, because at least 70% of the cell barcodes recovered were of low complexity. Finally, we tested whether the abundance of viral RNA in each cell was significantly above the abundance in empty wells given the ambient RNA profile of each array using an exact binomial test (binom.test in R). We performed Benjamini–Hochberg correction to determine the FDR (p.adjust, method = 'BH' in R), and cells with FDR < 0.01 were assigned as SARS-CoV-2 RNA$^+$.

Cell clustering and annotation

After SARS-CoV-2 RNA$^+$ assignment, CellBender-corrected matrices were used for further analysis. All analysis was carried out using the Seurat (version 4.3.0) package in R (version 4.2.3). Data were normalized using sctransform (version 0.3.5)[60]. PCA was run using the top 3,000 most variable genes with SARS-CoV-2 genes excluded. Harmony (version 0.1.1) integration was used to account for variant-specific batch effects[61]. The RunHarmony() function was run with default parameters, reduction = 'pca', group.by.vars = 'Cohort' and assay.use = 'SCT', where 'Cohort' was either ancestral, Delta or Omicron. The output dimensionality reduction from Harmony was then used to generate nearest-neighbor graphs and UMAPs. Cells were clustered using Louvain clustering. Unique marker genes for each cluster were identified using Seurat's FindAllMarkers() function with cutoffs min.pct = 0.25 and logfc.threshold = 0.25. Of note, the application of CellBender for ambient RNA removal and Harmony for integration could result in excessive filtering of transcripts or loss of biologically derived variation in gene expression, respectively.

Before annotation, cells were iteratively clustered to remove low-quality, low-complexity cells as well as doublets (Supplementary Table 2). Low-quality and low-complexity cells were identified by high expression of mitochondrial genes and little-to-no expression of other cell type marker genes. Doublets were identified by expression of marker genes for multiple distinct cell types and high feature counts. The first round of clustering the entire dataset (dims = 1:28, resolution = 0.8) resulted in removal of 4,005 cells. Upon a second round of clustering the entire dataset (dims = 1:27, resolution = 0.8), all mitochondrial-high clusters expressed some levels of ciliated cell markers; so these were not removed. Clusters were then assigned a lineage of either 'epithelial' (44,952 cells) or 'immune' (6,362 cells).

Epithelial cells were then isolated, normalized with sctransform, integrated with Harmony and reclustered (dims = 1:30, resolution = 1). Preliminary coarse annotations were then assigned using marker genes from the literature[25,31,34,53,62]. We identified nine major epithelial cell types, including basal, secretory, goblet, squamous, ciliated and deuterosomal cells, ionocytes, an MT^{hi} cluster and a SARS-CoV-2^{hi} cluster. Each major epithelial cell type was first isolated and reclustered to ensure that the coarse annotations were correct and that no low-quality cells or doublets remained (Supplementary Table 3). This resulted in removal of an additional 1,730 epithelial cells. In the second round of epithelial subclustering, final annotations were assigned (Extended Data Fig. 3 and Supplementary Table 4). Clustering parameters were selected such that stable end clusters were identified, each cluster had a minimum of 95 cells and each cluster had at least ten significantly DE genes (FDR < 0.001). Cell subsets were labeled by their cell type and two to three distinctive marker genes (for example, *ISG15*$^+$*IFIT1*$^+$*IFIT2*$^+$ ciliated cells). In the case that there were not unique genes marking clusters beyond the major cell type marker genes, the cell type-level annotation was maintained.

Subclustering basal and secretory cells together revealed two basal subsets, nine secretory subsets and one goblet cell cluster. Basal cells were divided into those expressing classical basal cell markers (*KRT5*, *KRT15*, *COL7A1*, *DST*) and a proliferating cluster (*MKI67*, *TOP2A*, *CENPF*). The goblet cell cluster uniquely expressed high levels of *MUC5AC*. Marker genes of the nine secretory clusters included keratin genes (*KRT4*, *KRT24*), club cell-like genes (*SCGB1A1*), antimicrobial secretory genes (*BPIFA1*, *BPIFB1*, *S100A9*, *PI3*), chemokine genes (*CXCL8*, *CXCL1*, *CXCL2*), growth response genes (*GDF15*, *ATF3*, *EGR1*), NF-κB response genes (*NFKBIA*, *TNFAIP3*) and SARS-CoV-2 genes. We identified five squamous clusters marked by expression of small proline-rich protein-encoding genes (*SPRR2E*, *SPRR2D*), chemokine genes (*CXCL8*), secretory genes (*PIGR*, *VMO1*), mitochondrial genes or the combination of *S100A2* and *APOBEC3A*. Ciliated cells had the greatest amount of diversity with 13 distinct subsets. Three ciliated clusters had higher expression of ciliogenesis genes (*PIFO (official gene symbol CIMAP3)*, *DNAH11*, *CFAP44*, *DNAAF1*, *DNAH12*). Three other ciliated clusters expressed IFN response genes (*ISG15*, *IFIT1*, *IFIT2*, *IFITM3*, *IFI6*, *STAT1*) or high amounts of antigen presentation genes (*HLA-A*, *HLA-B*, *HLA-DRB1*, *CD74*). We also identified ciliated clusters marked by cytochrome genes (*CYP4B1*, *CYPB5A*), alarmin genes (*SAA1*, *SAA2*), stress response genes (*JUN*, *FOS*) and secretory genes (*VMO1*, *AQP5*). Within the MT^{hi} cells, there were clusters defined by high expression of ciliogenesis genes,

calprotectin subunit genes (*S100A8*, *S100A9*), hemoglobin genes (*HBB*, *HBA2*) and additional stress response genes (*HSPH1*, *NEAT1*). SARS-CoV-2^{hi} clusters were marked by *IFNA1*, circadian clock and growth response genes (*PER1*, *PER2*, *EGR1*, *GDF15*) and IFN response genes (*B2M*, *IFI27*). We did not further subcluster deuterosomal cells or ionocytes due to low cell numbers.

We applied the same approach to identify diverse subsets within the 6,362 immune cells in our dataset (Extended Data Fig. 4 and Supplementary Table 5). Removal of immune–epithelial doublets and a low-quality immune cluster resulted in a final immune object with 5,408 cells. The majority of immune cells were either T cells or macrophages, with small populations of B cells and dendritic cells. We further subclustered the T cells and identified clusters marked by ribosomal genes (indicative of quiescence), stress response genes (*HSPA6*, *DEDD2*), the cytoskeletal gene *MACF1* and the natural killer cell-triggering receptor encoded by *NKTR*, the chemokine receptor encoded by *CXCR5* and cytotoxicity genes (*GNLY*, *GZMB*). We also observed diversity within the macrophage and dendritic cell compartment. Genes identifying myeloid clusters included the long noncoding RNA encoded by *NEAT1*, cysteine cathepsin genes (*CTSD*, *CTSL*), chemokine genes (one cluster with *CCL3*, *CCL4* and *CCRL2*, one with *CXCL5* and *CXCL1* and one with *CXCL10* and *CXCL11*) and antimicrobial genes (*S100A8*, *S100A9*, *LYZ*). B cells were not subclustered due to low cell numbers.

Cellular composition analyses

We used ARBOL (version 4.0.0) to understand relationships between our end clusters and their distribution across participant groups[63]. To create the binary tree, we ran ARBOLcentroidTaxonomy() with default parameters, tree_reduction = 'harmony', centroid_method = 'mean', distance_method = 'cosine', hclust_method = 'complete' and nboot = 100. ggraph (version 2.1.0) was then used to plot the phylogenetic tree, where line and text colors represent the cell type and the participant diversity and number of cells is shown for each subset (Fig. 2a).

We then created an abundance matrix for further cellular compositional analysis[64,65]. Abundance was calculated by deriving the frequency of each subset in each participant and scaling to 500 cells per participant (Supplementary Table 6). PCA on the log-normalized abundance matrix was carried out using the prcomp() function in R. For hierarchical clustering, we calculated the distance between the centroid for each variant group and used the dendro_data() function to extract and plot the dendrogram.

Non-negative matrix factorization

We applied NMF as an orthogonal method to evaluate gene programs within major cell types[66]. NMF was run using the Singlet package (version 0.99.6), a Seurat-compatible adaptation of RcppML (version 0.5.6) (with dependencies limma (version 03.54.2) and fgsea (version 1.24.0))[67]. The runNMF() function was applied to cell type-specific objects with default parameters and assay = 'SCT'. We ran this analysis on ciliated and secretory cells, the two most abundant cell types in our dataset. VizDimLoadings() from the Seurat package was used to visualize gene loadings of each factor (reduction = 'nmf'). Complete lists of identified factors and the top 50 genes contributing to each factor can be found in Supplementary Table 7. To determine the factor scores on a per-participant basis, the cell embedding values for each factor were extracted and averaged across all cells within the cell type (ciliated or secretory) for each participant.

Differential gene expression by participant group

To compare gene expression between cells from distinct participant groups, we used a Wilcoxon rank-sum test. Cells from each cell type belonging to either unvaccinated or vaccinated participants were compared in a pairwise manner, implemented using the Seurat FindMarkers function (Supplementary Table 8). We considered genes as DE with an FDR-adjusted P value < 0.001 and log (fold change) > 0.25. GO analysis was run using the Database for Annotation, Visualization, and Integrated Discovery (Supplementary Table 9)[68].

Proportionality analysis

Proportionality analysis was performed using the propr package (version 4.3.0)[69] on non-log-transformed immune cluster abundances across Delta and Omicron participants. Proportionality was calculated separately for vaccinated and unvaccinated participants. Investigating differential proportionality using the propd() function revealed no clusters with significantly different relationships between unvaccinated and vaccinated groups.

IFN gene module scores

Gene lists corresponding to 'IFN-α response' or 'IFN-γ response' are derived from previously published population RNA-seq data from nasal epithelial basal cells treated in vitro with 0.1–10 ng ml^{-1} IFN-α or IFN-γ for 12 h (Supplementary Table 10)[53,70]. Module scores were calculated using the Seurat function AddModuleScore() with default inputs.

Pseudobulk analysis

Pseudobulk count matrices for each combination of participant and major cell type were generated using the Seurat AggregateExpression() function. Analysis was restricted to cell types where at least three distinct COVID-19 severity scores had at least three participants with at least ten cells across all three variants, which included only ciliated cells. A DESeq2 (package version 1.38.3) object was generated using the aggregated ciliated cell count matrix and the COVID-19 severity score plus variant group as design conditions. Logarithmic normalization of the DESeq2 count matrix was performed before correlation analysis.

Statistical testing and data visualization

All statistical tests were implemented in R (version 4.2.3). Comparisons between cell type frequencies by disease group were tested using a two-sided Kruskal–Wallis test with FDR correction, implemented in R using the dunn.test package (version 1.3.5). Effect size was calculated using the cohen.d() function from the effsize package (version 0.8.1). Spearman correlation was used where appropriate, implemented using the cor.test function in R. All box plots were generated using the geom_boxplot() function in the ggplot2 package (version 3.4.2) with the following statistics represented: center line, median; box limits, upper and lower quartiles; whiskers, 1.5× the interquartile range. P values, n and all summary statistics are provided either in the Results, figure legends, figure panels or supplementary tables. R packages dplyr (version 1.1.1), tidyverse (version 2.0.0), forcats (version 1.0.0), stats (version 4.2.3) and Matrix.utils (version 0.9.7) were used for data wrangling. ggplot2 (version 3.4.2), ggraph (version 2.1.0), cowplot (version 1.1.1), ggdendro (version 0.1.23), ggalluvial (version 0.12.5), stringr (version 1.5.0), ggrepel (version 0.9.3), ggnewscale (version 0.4.9), RColor brewer (version 1.1-3), ggbiplot (version 0.55), ggpubr (version 0.6.0), gridExtra (version 2.3), pheatmap (version 1.0.12), ggplotify (version 0.1.1) and EnhancedVolcano (version 1.16.0) were used for data visualization. Color palettes were generated using iWantHue (https://medialab.github.io/iwanthue/).

Reporting summary

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

Data availability

scRNA-seq data are publicly available for download and visualization via the Broad Institute Single Cell Portal (accession code SCP2593). Custom reference FASTA and GTF for SARS-CoV-2 are available for download: https://github.com/ShalekLab/SARSCoV2-genome-reference.

Code availability

Code was written using R version 4.2.3 and publicly available packages. No new algorithms or functions were created, and code used in-built functions in listed packages available on CRAN. All code generated and used to analyze the data reported in this paper is available on GitHub in the jo-m-lab repository: https://github.com/jo-m-lab/SARSCoV2-Variants-Vaccines-Paper.

References

51. *WHO R&D Blueprint Novel Coronavirus COVID-19 Therapeutic Trial Synopsis* (World Health Organization, 2020); https://www.who.int/publications/i/item/covid-19-therapeutic-trial-synopsis
52. Tang, Y. et al. Human nasopharyngeal swab processing for viable single-cell suspension. *protocols.io;* https://doi.org/10.17504/protocols.io.bjhkkj4w (2020).
53. Ordovas-Montanes, J. et al. Allergic inflammatory memory in human respiratory epithelial progenitor cells. *Nature* **560**, 649–654 (2018).
54. Aicher, T. P. et al. Seq-Well: a sample-efficient, portable picowell platform for massively parallel single-cell RNA sequencing. In *Single Cell Methods: Sequencing and Proteomics* (ed. Proserpio, V.) 111–132 (Springer, 2019).
55. Gierahn, T. M. et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).
56. Drake, R. S. et al. in *Single Cell Transcriptomics: Methods and Protocols* (eds Calogero, R. A. & Benes, V.) 57–104 (Springer, 2023).
57. Hughes, T. K. et al. Second-strand synthesis-based massively parallel scRNA-seq reveals cellular states and molecular features of human inflammatory skin pathologies. *Immunity* **53**, 878–894 (2020).
58. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
59. Kim, D. et al. The architecture of SARS-CoV-2 transcriptome. *Cell* **181**, 914–921 (2020).
60. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
61. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
62. Deprez, M. et al. A single-cell atlas of the human healthy airways. *Am. J. Respir. Crit. Care Med.* **202**, 1636–1645 (2020).
63. Zheng, H. B. et al. Concerted changes in the pediatric single-cell intestinal ecosystem before and after anti-TNF blockade. *eLife* **12**, RP91792 (2023).
64. Kazer, S. W. et al. Primary nasal influenza infection rewires tissue-scale memory response dynamics. *Immunity* **57**, 1955–1974 (2024).
65. Quinn, T. P., Erb, I., Richardson, M. F. & Crowley, T. M. Understanding sequencing data as compositions: an outlook and review. *Bioinformatics* **34**, 2870–2878 (2018).
66. Kotliar, D. et al. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-seq. *eLife* **8**, e43803 (2019).
67. DeBruine, Z. J., Andrew Pospisilik, J. & Triche, T. J. Fast and interpretable non-negative matrix factorization for atlas-scale single cell data. Preprint at *bioRxiv* https://doi.org/10.1101/2021.09.01.458620 (2024).
68. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
69. Quinn, T. P., Richardson, M. F., Lovell, D. & Crowley, T. M. propr: an R-package for identifying proportionally abundant features using compositional data analysis. *Sci. Rep.* **7**, 16252 (2017).
70. Ziegler, C. G. K. et al. SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. *Cell* **181**, 1016–1035 (2020).

Acknowledgements

We thank the study participants and their families for enabling this research, the clinical support staff at the UMMC for assistance with sample collection and members of the Ordovas-Montanes laboratory (Boston Children's Hospital, Harvard Medical School, Broad Institute, Ragon Institute), the Shalek laboratory (MIT, Broad Institute, Ragon Institute), the Horwitz laboratory (Boston Children's Hospital) and the Glover laboratory (UMMC, Tulane University) for thoughtful discussion and feedback. We sincerely thank A. Sharpe and all members of the Sharpe laboratory (Harvard Medical School) for insightful discussion and feedback as well. We thank J. Gould, K. Siddle, B. Li, S. Fleming and the Broad Institute viral-ngs and Cumulus teams for assistance with computational pipelines. This project was made possible in part by grant number 2020-216949 from the Chan Zuckerberg Initiative DAF, an advised fund of the Silicon Valley Community Foundation to A.K.S. and J.O.-M. Work performed through the UMMC Molecular and Genomics Facility is supported, in part, by funds from the NIGMS, including the Molecular Center of Health and Disease (P20GM144041, M.R.G.), the Mississippi INBRE (P20GM103476, A. Flynt) and Obesity, Cardiorenal and Metabolic Diseases—COBRE (P30GM149404, J. Hall). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. J.O-.M. is a New York Stem Cell Foundation—Robertson Investigator. J.O.-M. was supported by the AbbVie–Harvard Medical School Alliance, the Richard and Susan Smith Family Foundation, the AGA Research Foundation's AGA–Takeda Pharmaceuticals Research Scholar Award in IBD (AGA2020-13-01), the HDDC Pilot and Feasibility (P30 DK034854), the Leona M. and Harry B. Helmsley Charitable Trust, the Pew Charitable Trusts Biomedical Scholars, the Broad Next Generation Award, the Chan Zuckerberg Initiative Pediatric Networks, the Mathers Foundation, the New York Stem Cell Foundation, NIH R01 HL162642, NIH R01 DE031928 and the Cell Discovery Network, a collaborative funded by the Manton Foundation and the Warren Alpert Foundation at Boston Children's Hospital. J.M.L.W. was supported by NIH Training Grant 5TL1TR002543, S.W.K. was supported by the Cancer Research Institute's Irvington Postdoctoral Fellowship, and A.K.S. was supported by the Bill and Melinda Gates Foundation, a Sloan Fellowship in Chemistry, the NIH (5U24AI118672) and the Ragon Institute of MGH, MIT and Harvard.

Author contributions

Conceptualization, J.O.-M., A.K.S., S.C.G., B.H.H.; methodology, J.M.L.W., V.N.M., A.H.O., Y.T., J.D.B., S.W.K., K.K., C.A., C.G.K.Z., A.W.N., S.C.G., J.O.-M., A.K.S., B.H.H.; software, K.K., S.W.K.; formal analysis, J.M.L.W., V.N.M., A.H.O., Y.T., J.D.B., S.W.K., K.K; investigation, J.M.L.W., V.N.M., A.H.O., Y.T., J.D.B., C.G.K.Z., A.W.N., S.I., T.J., M.G., R.S.D., J.T.B., B.C.B., D.A.R.; resources, S.C.G., A.H.O.; data curation, A.H.O, Y.T., N.S.D., T.O.R., J.M.L.W.; writing (original draft), J.M.L.W., J.O.-M., A.K.S., S.C.G., B.H.H.; writing (review and editing), J.M.L.W., V.N.M., A.H.O., Y.T., J.D.B., S.W.K., K.K., C.A., C.G.K.Z., S.I., T.J., M.G., A.W.N., R.S.D., A.P., B.C.B., P.D., S.T., S.K.K., H.L., T.G.W., Y.T.D., N.S.D., Y.P., Y.G., M.S., J.H., J.T.B., G.D., M.R.G., D.A.R., I.J.F., J.J.L., T.O.R., J.O.-M., A.K.S., S.C.G., B.H.H.; visualization, J.M.L.W., Y.T.; supervision, J.O.-M., A.K.S., S.C.G., B.H.H., T.O.R.; project administration, Y.P., T.O.R.; funding acquisition, J.O.-M., A.K.S., S.C.G., B.H.H.

Competing interests

V.N.M. reports compensation from MPM Capital and RA Capital Management unrelated to this work. S.W.K. reports compensation for consulting services with Monopteros Therapeutics, Flagship Pioneering and Radera Biosciences. M.S. reports compensation for consulting services with GE Precision Healthcare. J.J.L. reports compensation for

consulting services with Blueprint Medicines and Human Immunology Biosciences. J.O.-M. reports compensation for consulting services with Cellarity, Tessel Biosciences and Radera Biotherapeutics. A.K.S. reports compensation for consulting and/or SAB membership from Merck, Honeycomb Biotechnologies, Cellarity, Hovione, Ochre Bio, Third Rock Ventures, Relation Therapeutics, Dahlia Biosciences, Bio-Rad Laboratories, IntrECate Biotherapeutics, FogPharma and Passkey Therapeutics. C.G.K.Z., V.N.M., A.H.O., A.W.N., Y.T., J.D.B., A.K.S., S.C.G., B.H.H. and J.O.-M. are co-inventors on a provisional patent application relating to methods of stratifying and treating viral infections. The other authors declare no competing interests.

Additional information

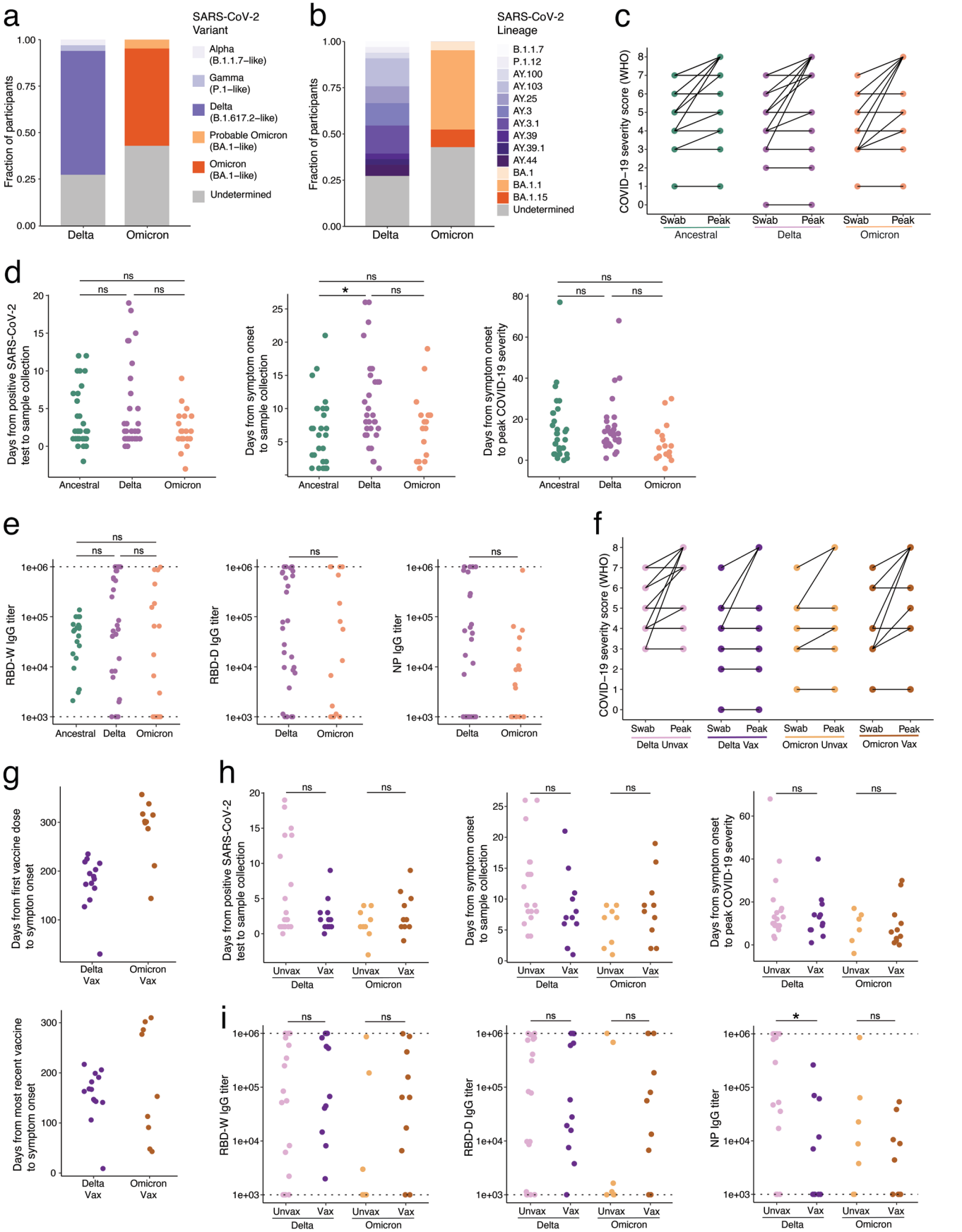
Extended data is available for this paper at https://doi.org/10.1038/s41590-024-02052-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41590-024-02052-z.

Correspondence and requests for materials should be addressed to Jose Ordovas-Montanes.

Peer review information *Nature Immunology* thanks Daniel Altmann and Martijn Nawijn for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Ioana Staicu, in collaboration with the *Nature Immunology* team.

https://doi.org/10.1038/s41590-024-02052-z

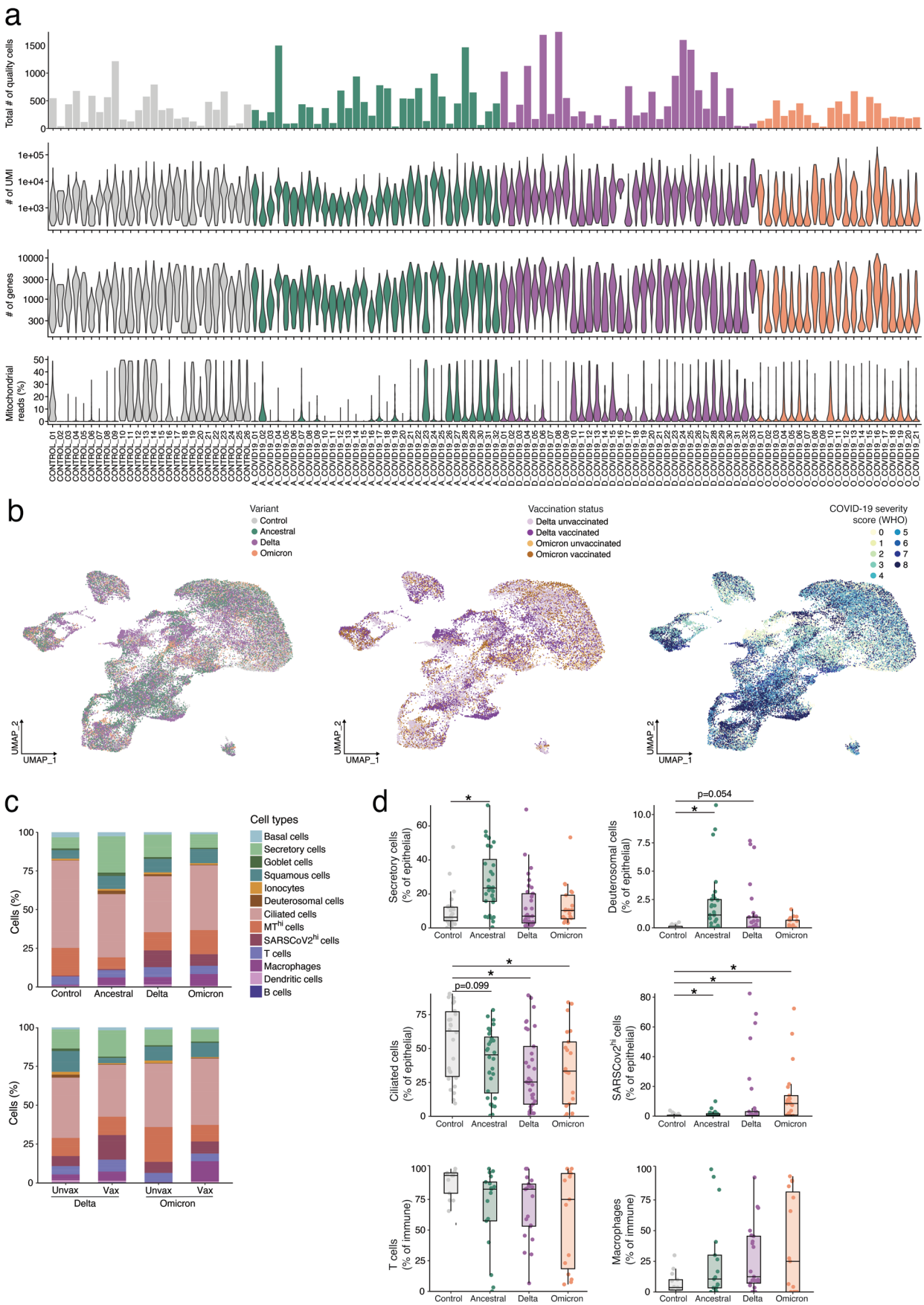


Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s41590-024-02052-z

Extended Data Fig. 1 | Participant metadata across variant and vaccination groups. a-b) Stacked bar charts showing fraction of Delta or Omicron cases in SARS-CoV-2 variant (a) and lineage (b) groups as identified by targeted sequencing of viral RNA from supernatant collected during nasopharyngeal swab processing (restricted to participants with SARS-CoV-2 S CT value < 30, Delta: n = 22, Omicron: n = 12). 'Undetermined': CT values > 30, sample not sequenced (Delta: n = 8, Omicron: n = 9). SARS-CoV-2 variant (a) and lineages (b) were determined using DRAGEN COVIDSeq Test software (see Methods, 'Viral Variant Sequencing'). **c**) Comparison of COVID-19 severity score (WHO score of respiratory support) at time of nasopharyngeal swab collection ('Swab') and peak disease severity ('Peak') in all ancestral (n = 32), Delta (n = 33), and Omicron (n = 21) cases. **d**) Length of time in days from positive SARS-CoV-2 PCR test to nasopharyngeal swab collection (all cases), from reported symptom onset to nasopharyngeal swab collection (Ancestral: n = 30, Delta: n = 31, Omicron: n = 18, Delta vs. Ancestral: p = 0.031), and from reported symptom onset to peak COVID-19 severity score (Ancestral: n = 30, Delta: n = 31, Omicron: n = 18). **e**) SARS-CoV-2 plasma serology across variant groups. Plasma samples were taken from a subset of participants (Ancestral: n = 24, Delta: n = 30, Omicron: n = 17) on the same day as the nasopharyngeal swab used for scRNA-seq. Dotted lines indicate limits of detection. RBD-W: original Washington strain receptor binding domain; RBD-D: Delta strain receptor binding domain; NP: nucleoprotein. **f**) Comparison of COVID-19 severity score as in (c) across vaccination groups (Delta unvaccinated ('Unvax'): n = 19, Delta vaccinated ('Vax'): n = 14, Omicron unvaccinated: n = 10, Omicron vaccinated: n = 11). **g**) Length of time in days between first vaccine dose or most recent vaccine dose and positive SARS-CoV-2 PCR test in vaccinated Delta (n = 14) and Omicron (n = 10) cases. **h**) Disease course timing as in (d) across vaccination groups (All cases for PCR test to sample collection, for symptom onset data Delta unvaccinated: n = 19, Delta vaccinated: n = 12, Omicron unvaccinated: n = 7, Omicron vaccinated: n = 10. **i**) SARS-CoV-2 plasma serology as in (e) across vaccination groups (Delta unvaccinated: n = 18, Delta vaccinated: n = 12, Omicron unvaccinated: n = 7, Omicron vaccinated: n = 10, Delta vaccinated vs Delta unvaccinated NP IgG: p = 0.049). **d**,**e**,**h**,**i**, statistical test represents two-sided Kruskal-Wallis test results with Benjamini-Hochberg correction for multiple comparison. *Dunn post hoc test, p < 0.05.

https://doi.org/10.1038/s41590-024-02052-z

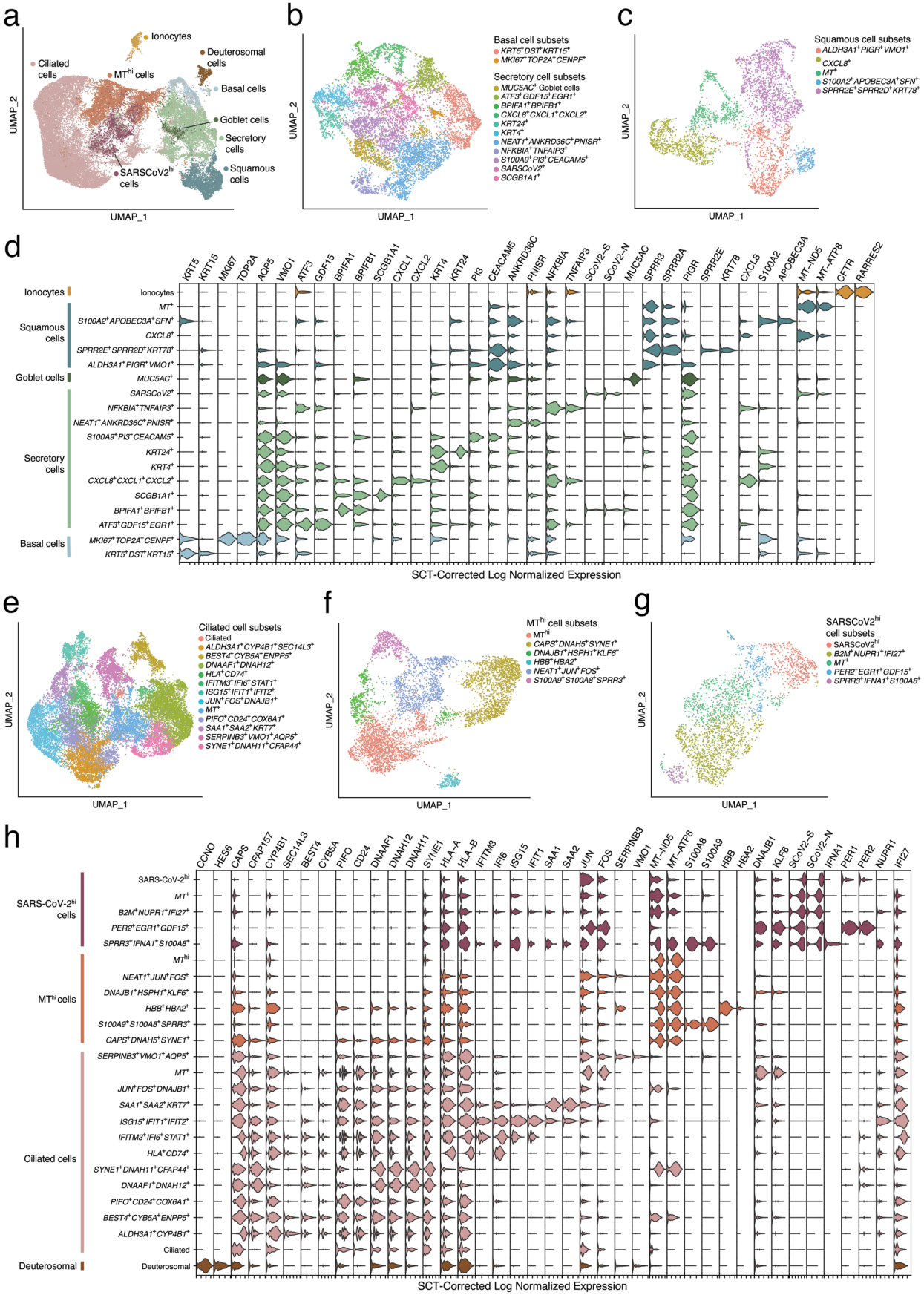


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41590-024-02052-z

Extended Data Fig. 2 | scRNA-seq Quality Metrics and Cell Type Distributions. **a**) Single cell quality metrics (total number of cells, number of unique UMI (unique molecular identifiers) per cell, number of unique genes per cell, and percent mitochondrial transcripts) for each participant (bars or violins) after filtering for low quality cells. **b**) UMAP of 48,730 cells from all participants (n = 112) colored by variant group (left), UMAP of 23,987 cells from Delta and Omicron participants (n = 54) colored by vaccination status (center), and UMAP of 48,730 cells from all participants (n = 112) colored by WHO peak respiratory support score (right). **c**) Stacked bar chart showing frequency of major cell types across variant (control: n = 26, ancestral, n = 32, Delta, n = 33, Omicron, n = 21) and vaccination (Delta unvaccinated ('Unvax'): n = 19, Delta vaccinated ('Vax'): n = 14, Omicron unvaccinated: n = 10, Omicron vaccinated: n = 11) groups. **d**) Frequency of secretory cells (ancestral vs. control: $p = 4.8*10^{-4}$), deuterosomal cells (ancestral vs. control: $p = 5.6*10^{-6}$, Delta vs. control: p = 0.054), ciliated cells (Delta vs. control: p = 0.011, Omicron vs. control: p = 0.031), SARS-CoV-2^{hi} cells (ancestral vs. control: p = 0.013, Delta vs. control: p = 0.03, Omicron vs. control: $p = 5*10^{-3}$), T cells, and macrophages as a fraction of all cells in that lineage (epithelial or immune) in that sample from all controls (n = 26) and all ancestral (n = 32), Delta (n = 33), and Omicron (n = 21) cases. Box plots represent median (center line), upper and lower quartiles (box limits); and 1.5x the inter-quartile range (whiskers). Compare to Fig. 1f. Statistical test represents two-sided Kruskal-Wallis test with Benjamini-Hochberg correction for multiple comparisons. *Dunn's post hoc test, $p < 0.05$.

https://doi.org/10.1038/s41590-024-02052-z

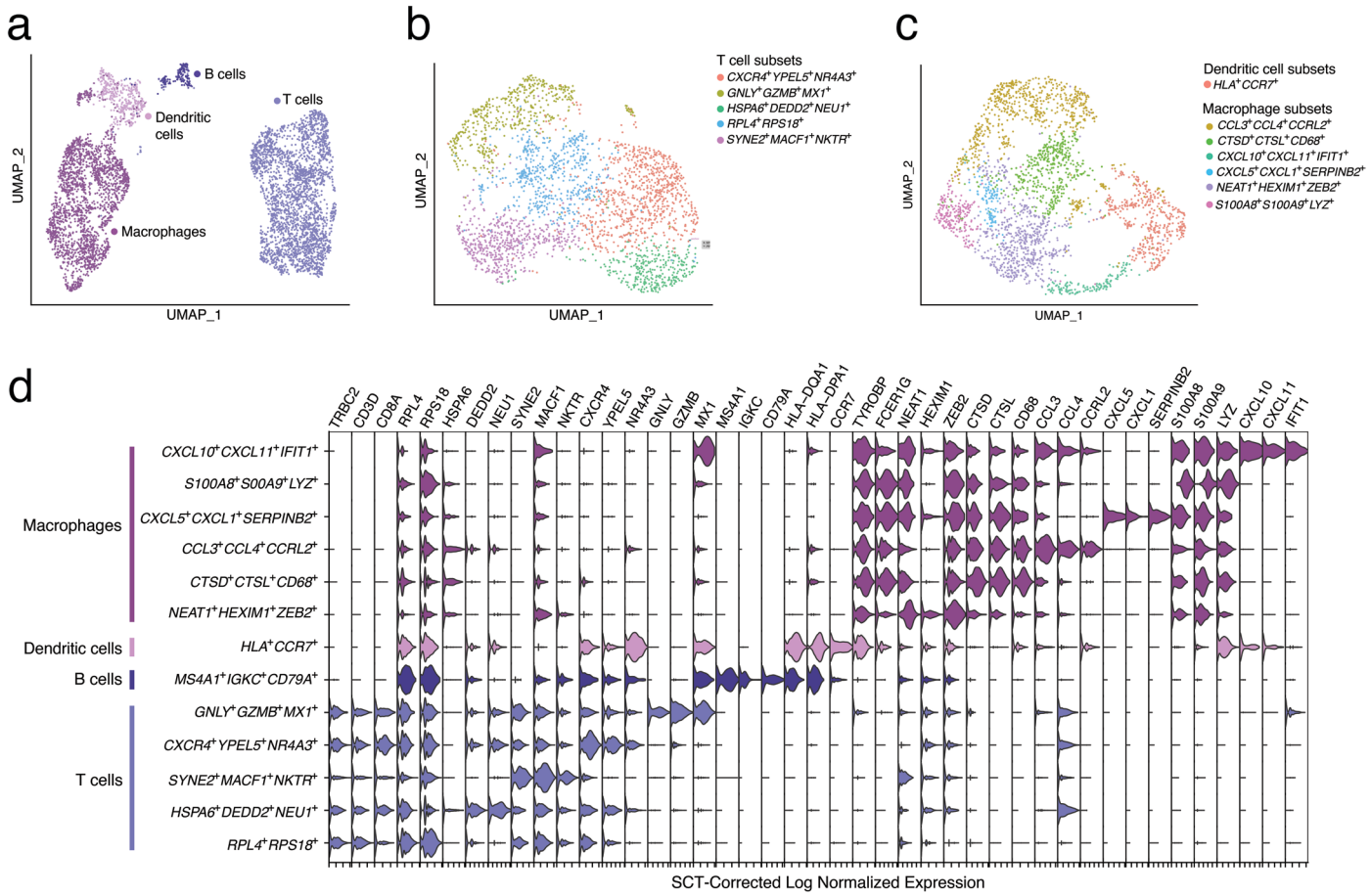


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41590-024-02052-z

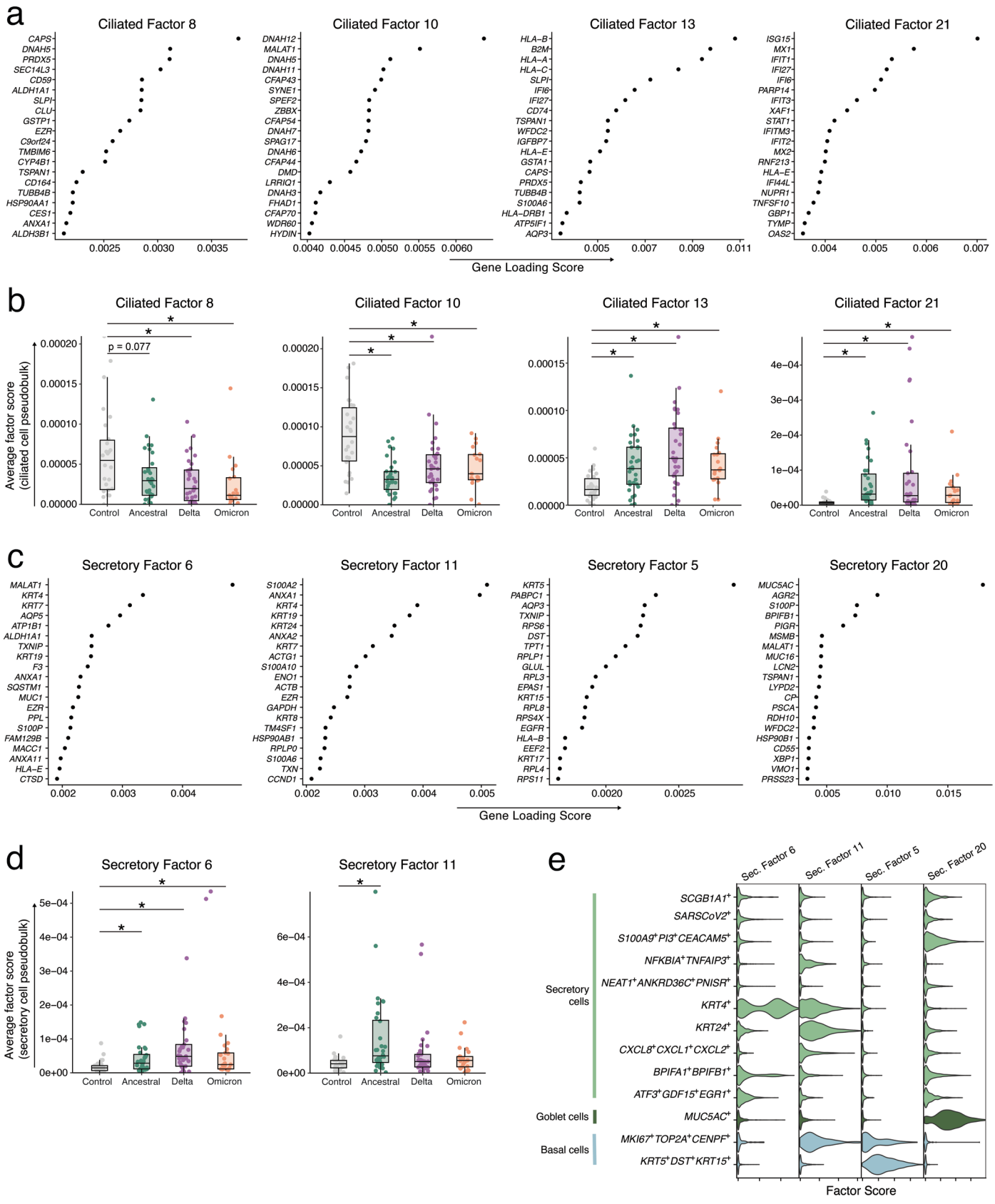
Extended Data Fig. 3 | Epithelial diversity in the nasal mucosa during SARS-CoV-2 infection. a) UMAP of 43,322 epithelial cells from all participants (n = 112) colored by major epithelial cell type as identified in Fig. 1d. **b-c**) UMAPs of detailed subsets within (b) basal/secretory cells (n = 112 participants, 9,106 cells) and (c) squamous cells (n = 112 participants, 3,913 cells). Cells from all participants belonging to that cell type were sub-clustered with louvain clustering. See Supplementary Tables 3, 4 for full lists of subset marker genes. **d**) Violin plot of marker genes for basal, secretory, goblet, squamous, and ionocyte subsets. **e**–**g**) UMAPs of detailed subsets within (e) ciliated cells (n = 112 participants, 20,683 cells), (f) MT^{hi} cells (n = 112 participants, 5,930 cells) and (g) SARS-CoV-2 RNA^{hi} cells (n = 112 participants, 2,551 cells). **h**) Violin plot of marker genes for ciliated, deuterosomal, MT^{hi}, and SARS-CoV-2 RNA^{hi} subsets.

https://doi.org/10.1038/s41590-024-02052-z



Extended Data Fig. 4 | Immune cell diversity in the nasal mucosa during SARS-CoV-2 infection. a) UMAP of 5,408 immune cells from all participants colored by major immune cell type as identified in Fig. 1d. **b-c**) UMAPs of detailed subclusters within (b) T cells (n = 112 participants, 2,705 cells) and (c) myeloid cells (n = 112 participants, 2,535 cells). Cells from all participants belonging to that cell type were sub-clustered with louvain clustering. See Supplementary Table 5 for full lists of subset marker genes. **d**) Violin plot of marker genes for each immune subset.

https://doi.org/10.1038/s41590-024-02052-z

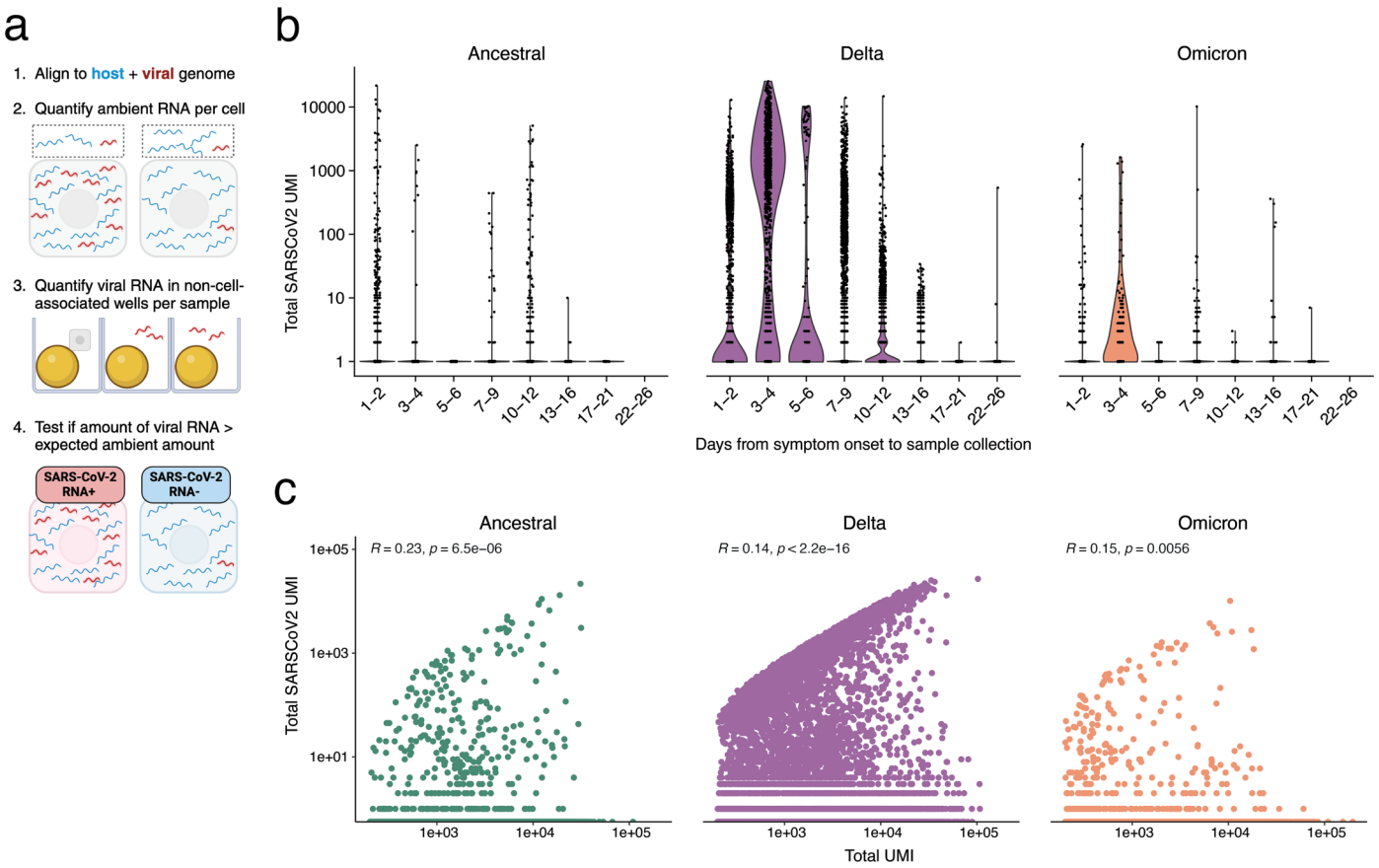


Extended Data Fig. 5 | See next page for caption.

https://doi.org/10.1038/s41590-024-02052-z

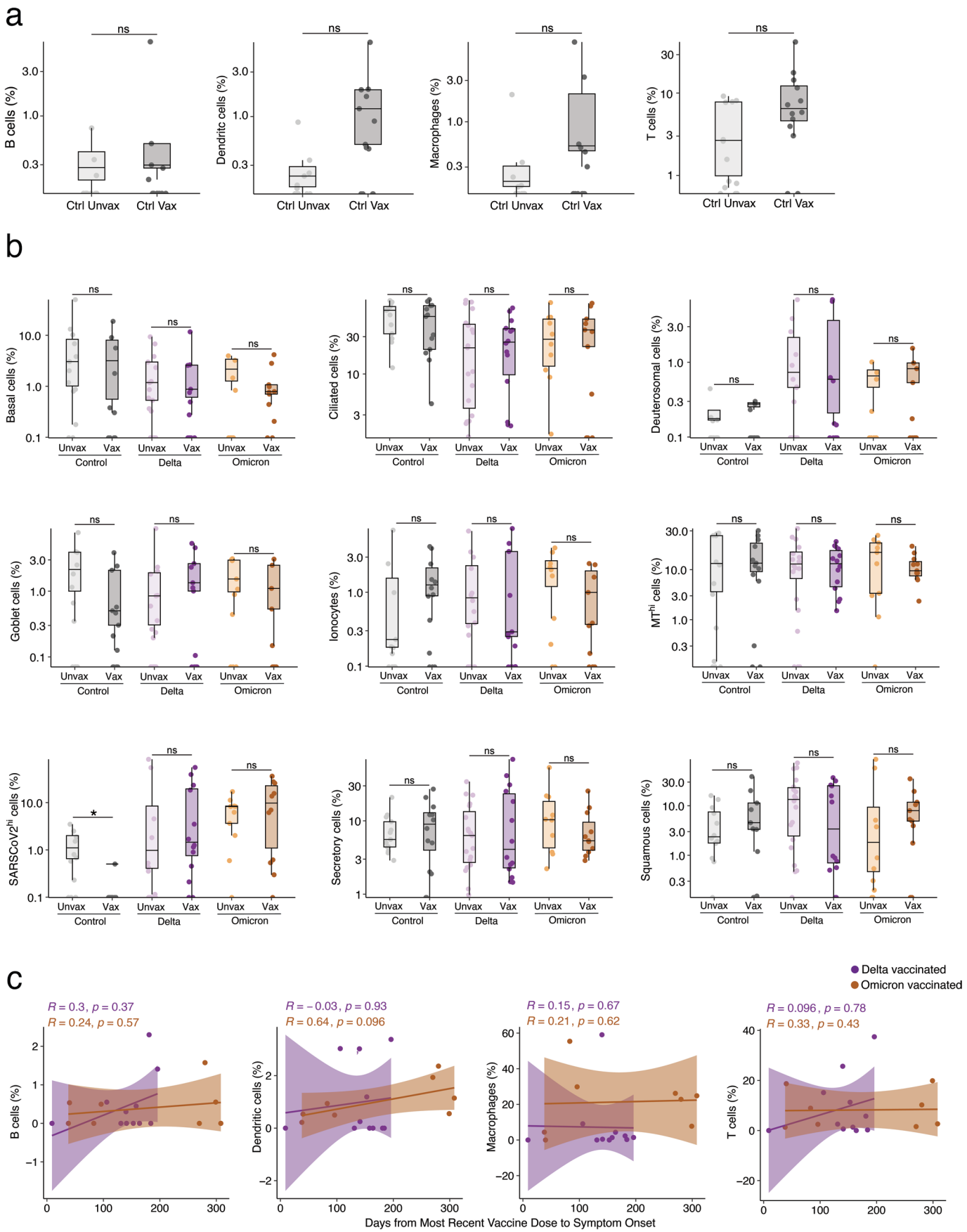
Extended Data Fig. 5 | Non-negative matrix factorization (NMF) Analysis. **a**) Gene loadings of ciliated factors 8, 10, 13, and 21 identified by running NMF on all ciliated cells (n = 20,683) from all participants (n = 112). Complete list of genes contributing to all 25 factors is provided in Supplementary Table 7. **b**) Average score of ciliated factor 8 (Delta vs. control: p = 0.012, Omicron vs. control: $p = 8.46*10^{-3}$), ciliated factor 10 (ancestral vs. control: $p = 7.54*10^{-6}$, Delta vs. control, $p = 3.32*10^{-3}$, Omicron vs. control: p = 0.011), ciliated factor 13 (ancestral vs. control: $p = 3.89*10^{-3}$, Delta vs. control: $p = 1.98*10^{-5}$, Omicron vs. control: $p = 3*10^{-3}$), and ciliated factor 21 (ancestral vs. control: $p = 1.08*10^{-4}$, Delta vs. control, $p = 4.51*10^{-4}$, Omicron vs. control: $p = 4.35*10^{-3}$) within pseudobulk ciliated cells for all controls (n = 26) and ancestral (n = 32), Delta (n = 33), and Omicron (n = 21) cases. **c**) Gene loadings of secretory factors 6, 11, 5, and 20 identified by running NMF on all basal and secretory cells (n = 9,106) from all participants (n = 112). Complete list of genes contributing to all 23 factors is provided in Supplementary Table 7. **d**) Average factor score of secretory factor 6 (ancestral vs. control: p = 0.033, Delta vs. control, $p = 5.1*10^{-4}$, Omicron vs. control: p = 0.035) and secretory factor 11 (ancestral vs. control: $p = 2.72*10^{-3}$) within pseudobulk secretory cells for all controls (n = 26) and ancestral (n = 32), Delta (n = 33), and Omicron (n = 21) cases. **e**) Violin plot of scores for secretory factors 6, 11, 5, and 20 across all basal, goblet, and secretory subsets in all participants (n = 112). b,d Box plots represent median (center line), upper and lower quartiles (box limits), and 1.5x the inter-quartile range (whiskers). Statistical tests represent two-sided Kruskal-Wallis test with Benjamini-Hochberg correction for multiple comparisons. *Dunn's post hoc test, $p < 0.05$.

https://doi.org/10.1038/s41590-024-02052-z



Extended Data Fig. 6 | Detection and assignment of SARS-CoV-2 RNA+ cells. **a**) Schematic of method for SARS-CoV-2 RNA+ assignment. scRNA-seq data was aligned to a joint human and SARS-CoV-2 genome. The amount of ambient RNA in each cell was estimated using Cellbender. Cells were assigned SARS-CoV-2 RNA+ if the amount of viral RNA was greater than the expected ambient amount (see Methods). **b**) Total SARS-CoV-2 transcripts per cell in a subset of ancestral (n = 30), Delta (n = 31), and Omicron (n = 18) cases, divided by time from symptom onset. **c**) Spearman correlation (two-tailed test) between the total number of transcripts (UMI) per cell and the total number of SARS-CoV-2 transcripts per cell in all ancestral (n = 32), Delta (n = 33), and Omicron (n = 21) cases. Schematic in **a** created with BioRender.com.

https://doi.org/10.1038/s41590-024-02052-z

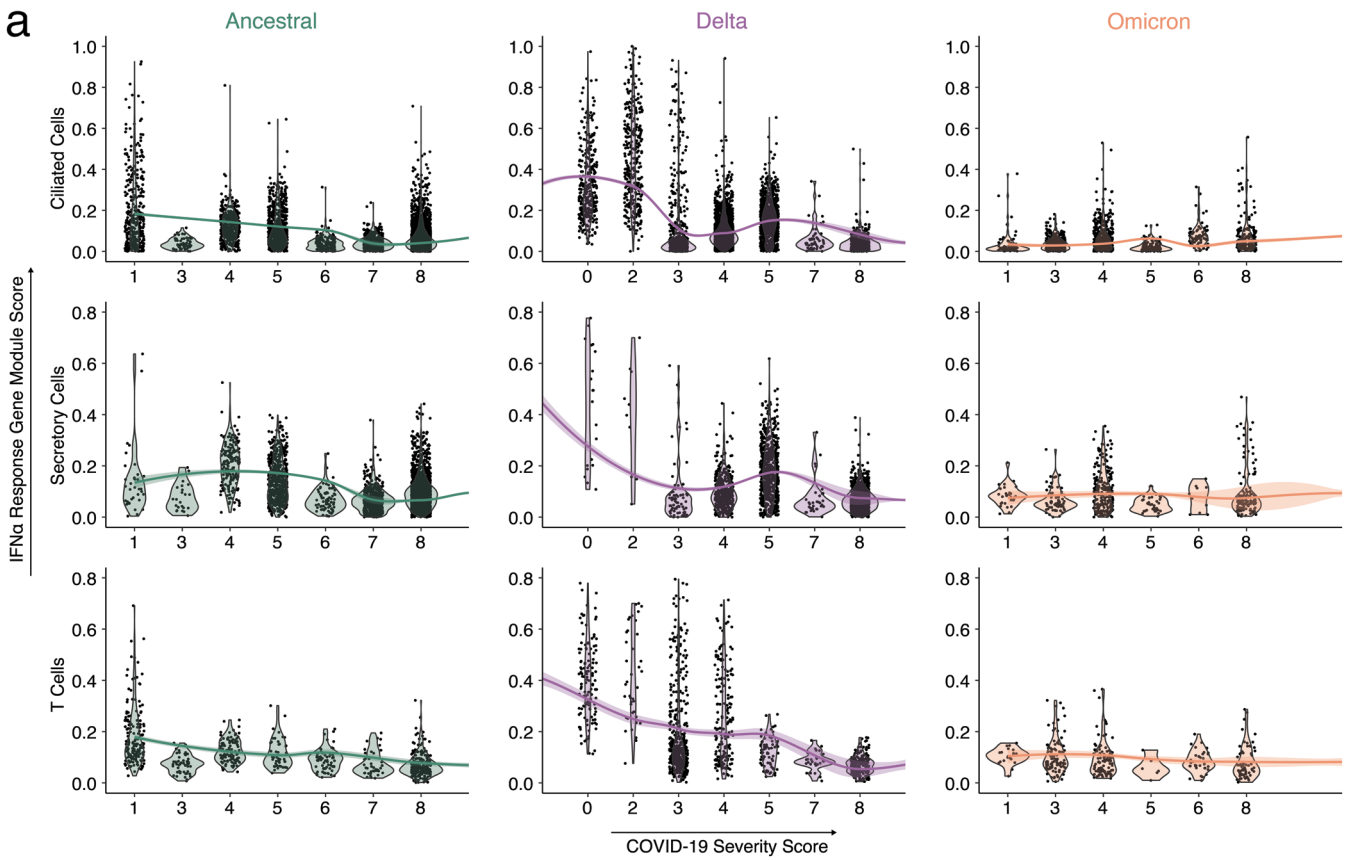


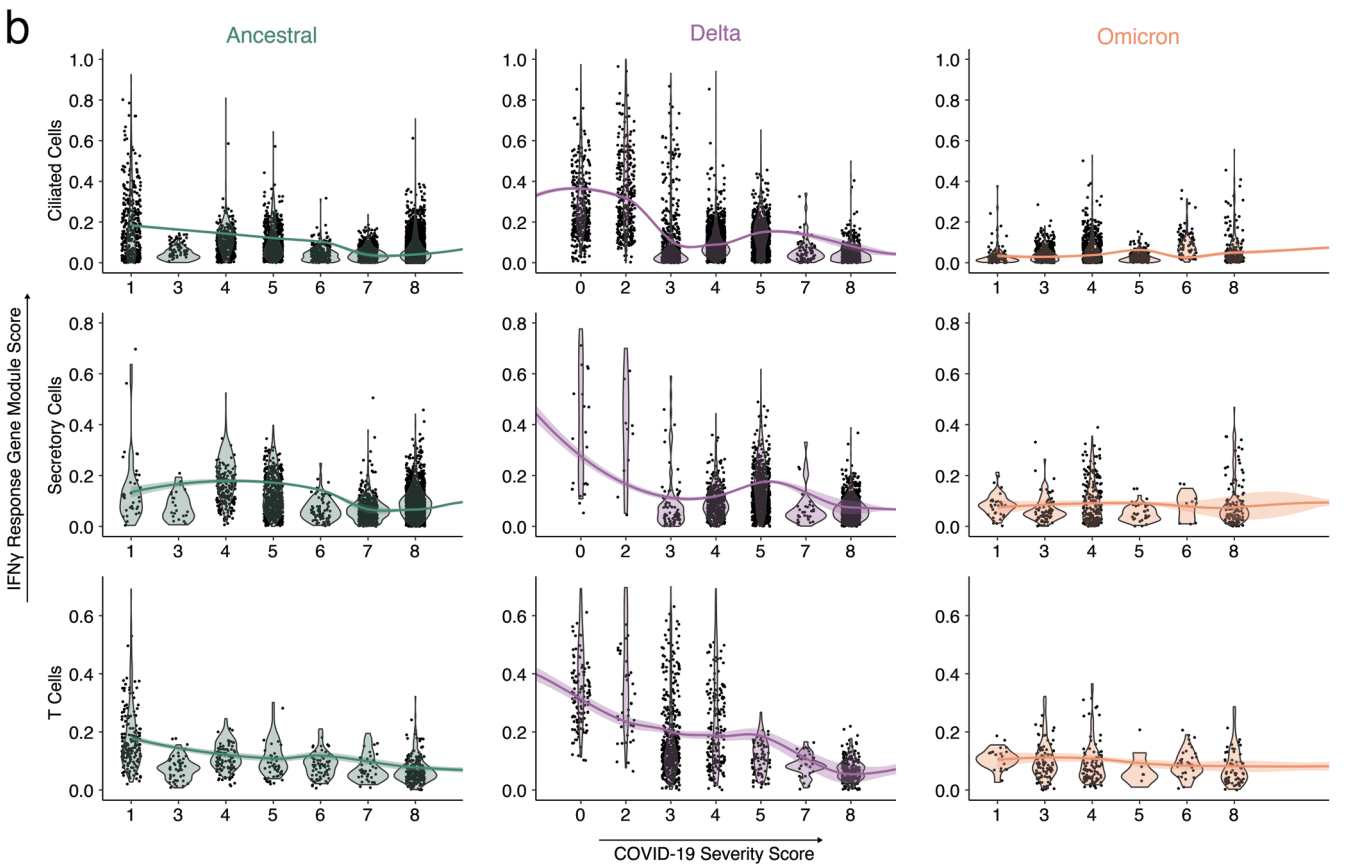
Extended Data Fig. 7 | See next page for caption.

https://doi.org/10.1038/s41590-024-02052-z

Extended Data Fig. 7 | Frequency of Major Cell Types Across Vaccination Groups. a) Frequency of B cells, dendritic cells, macrophages, and T cells as a percentage of all cells in the sample in control participants. Control ('Ctrl') unvaccinated ('Unvax') (n = 13) samples were collected during the Ancestral wave, and control vaccinated ('Vax') (n = 13) samples were collected during the Delta and Omicron waves. **b**) Frequency of epithelial cell types (basal cells, ciliated cells, deuterosomal cells, goblet cells, ionocytes, MT^{hi} cells, SARS-CoV-2^{hi} cells (Ctrl Vax vs Ctrl Unvax: p = 0.041), secretory cells, and squamous cells) as a percentage of all cells in the sample in unvaccinated ('Unvax') controls (n = 13), vaccinated ('Vax') controls (n = 13), unvaccinated Delta (n = 19), vaccinated Delta (n = 14), unvaccinated Omicron (n = 10), and vaccinated Omicron (n = 11) cases. **c**) Spearman correlation (two-tailed test) between time from most recent SARS-CoV-2 vaccine dose and frequency of immune cell types (B cells, dendritic cells, macrophages, and T cells) in all vaccinated Delta (n = 14) and Omicron (n = 11) cases. Error bands represent 95% confidence interval. a,b Box plots represent median (center line), upper and lower quartiles (box limits); and 1.5x the inter-quartile range (whiskers). Statistical tests represent two-sided Kruskal-Wallis test with Benjamini-Hochberg correction for multiple comparisons. *Dunn's post hoc test, $p < 0.05$.

https://doi.org/10.1038/s41590-024-02052-z



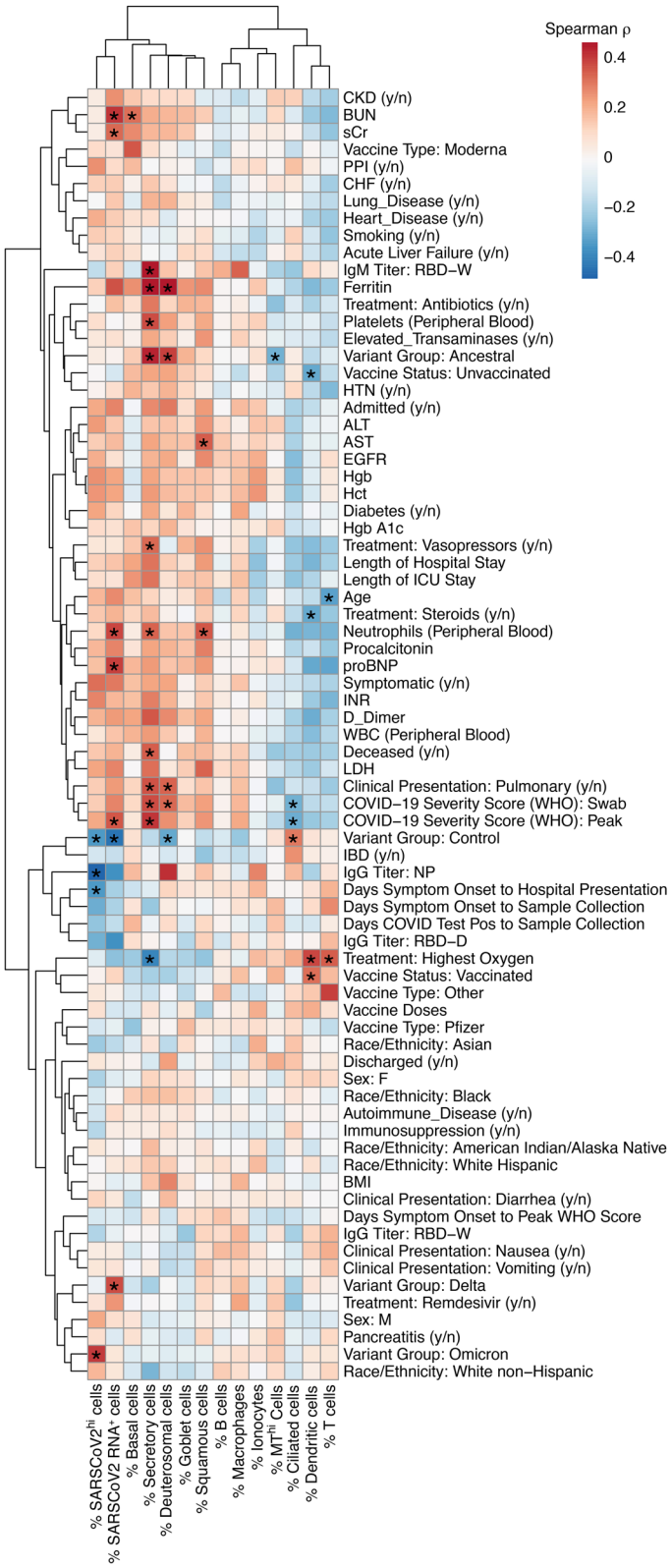


Extended Data Fig. 8 | See next page for caption.

https://doi.org/10.1038/s41590-024-02052-z

Extended Data Fig. 8 | Expression of ISG Modules Across COVID-19 Severity. **a-b**) Violin plots of IFN alpha (a) and IFN gamma (b) response gene module scores in ciliated (n = 15,180 cells), secretory (n = 6,656), and T cells (n = 2,181) from all ancestral (n = 32), Delta (n = 33), and Omicron (n = 21) cases. Cells are separated by COVID-19 severity score (WHO range 0–8). Genes uses for module score were derived from previously published results of stimulating human nasal basal cells with IFN alpha or IFN gamma (Supplementary Table 10)[70]. Lines represent local regression using locally estimated scatterplot smoothing. Error bands represent normal confidence interval.

https://doi.org/10.1038/s41590-024-02052-z



Extended Data Fig. 9 | Association of clinical metadata and nasal cell type frequencies. Heatmap of spearman correlation (two-tailed test) between clinical parameters and demographic information (rows) and frequency of major cell types (as identified in Fig. 1) and SARS-CoV-2 RNA+ cells (as identified in Fig. 3) (columns) in all participants (n = 112). Color indicates spearman R, *FDR < 0.05.

https://doi.org/10.1038/s41590-024-02052-z

Extended Data Table 1 | Composition of participant vaccination groups

Vaccination Group	Delta unvaccinated	Delta vaccinated	Omicron unvaccinated	Omicron vaccinated
Number of Participants	19	14	10	11
Age (years)***				
Minimum	23	22	33	59
Median (IQR)	47 (19)	#66 (30.25)	54 (14.5)	#65 (9)
Maximum	75	83	62	79
Sex				
Female	52.6% (10/19)	28.6% (4/14)	40% (4/10)	18.2% (2/11)
Ethnicity				
Hispanic	0% (0/19)	0% (0/14)	0% (0/10)	0% (0/11)
Not Hispanic	100% (19/19)	100% (14/14)	100% (10/10)	100% (11/11)
Race				
White/Caucasian	(8/19)	(9/14)	(4/10)	(3/11)
Black/African American	(11/19)	(5/14)	(6/10)	(7/11)
Asian	(0/19)	(0/14)	(0/10)	(0/11)
American Indian	(0/19)	(0/14)	(0/10)	(1/11)
WHO score (swab)				
Median (IQR)	5 (2)	4 (1.75)	3.5 (1.75)	4 (2)
WHO score (peak)				
Median (IQR)	5 (3.5)	4 (4.25)	4 (1.75)	5 (4)
BMI				
Median (IQR)	33.45 (12.3)	32.1 (7.9)	28 (14.2)	30.7 (15.8)
Pre-existing conditions				
Diabetes	36.8% (7/19)	42.8% (6/14)	50% (5/10)	72.7% (8/11)
Chronic kidney disease	10.5% (2/19)	21.4% (3/14)	20% (2/10)	36.4% (4/11)
Congestive heart failure	0% (0/19)	21.4% (3/14)	30% (3/10)	27.3% (3/11)
Lung disorder	15.8% (3/19)	28.5% (4/14)	20% (2/10)	18.2% (2/11)
Hypertension	57.9% (11/19)	64.3% (9/14)	70% (7/10)	36.4% (4/11)
IBD	0% (0/19)	0% (0/13)	10% (1/10)	0% (0/11)
Treatment				
Steroids	^94.7% (18/19)	71.4% (10/14)	60% (6/10)	81.8% (9/11)
Remdesivir*	^89.5% (17/19)	50% (7/14)	40% (4/10)	72.7% (8/11)
Mortality	26.3% (5/19)	28.5% (4/14)	20% (2/10)	36.4% (4/11)

Participants within the Delta and Omicron variant groups (total n=54) were assigned as unvaccinated or vaccinated if they had previously received a SARS-CoV-2 vaccine. Participants were grouped by variant and vaccination status (Delta unvaccinated: n=19, Delta vaccinated: n-14, Omicron unvaccinated: n=10, Omicron vaccinated: n=11). Age (Delta: p=0.006, Omicron: p=0.0022), biological sex, ethnicity, self-reported race, COVID-19 severity (WHO scores at time of sample collection and at peak disease), body mass index (BMI), pre-existing conditions (IBD: inflammatory bowel disease), COVID-19 treatments (Remdesivir: p=0.023), and mortality rates were compared across vaccination groups within variants. For continuous variables, median and interquartile range (IQR) are presented and two-sided Kruskal-Wallis test with Benjamini-Hochberg correction for multiple comparison was used for statistical comparison. #Dunn post-hoc test relative to unvaccinated group within same variant, $p<0.05$; For categorical variables, percent and number of participants in variant group belonging to listed sub-group are presented and a chi-squared test was used for comparison. ^chi-square standardized Pearson residual >2. *$p<0.05$, **$p<0.01$, ***$p<0.001$, otherwise non-significant.

nature portfolio

Corresponding author(s): Jose Ordovas-Montanes

Last updated by author(s): 12/2/2024

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection: No software was used for data collection.

Data analysis	Code was written using R v4.2.3 and publicly available packages. No new algorithms or functions were created and code used in-built functions in listed packages available on CRAN. All code generated and used to analyze the data reported in this paper is availaible on GitHub in the jo-m-lab repository: https://github.com/jo-m-lab/SARSCoV2-Variants-Vaccines-Paper Software: STAR aligner (v2.7.10), Drop-Seq Computational Protocol (https://github.com/broadinstitute/ Drop-seq) implemented on Cumulus (https://cumulus.readthedocs.io/en/latest/drop_seq.html, snapshot 11, default parameters) CellRanger v(3.0.0) CellBender (v0.2.0) R (v4.2.3) R packages: Seurat (v4.3.0), SCTransform (v0.3.5), Harmony (v0.1.1), ARBOL (v4.0.0, Singlet (v0.99.6), RcppML (v0.5.6), limma(v03.54.2), fgsea(v1.24.0), propr (v4.3.0), DESeq2(v1.38.3), dunn.test (v1.3.5), effsize (v0.8.1), ,dplyr (v1.1.1), tidyverse (v2.0.0), forcats (v1.0.0), stats (v4.2.3), Matrix.utils (v0.9.7), ggplot2 (v3.4.2), ggraph(v2.2.0), cowplot (v1.1.1), ggdendro (v0.1.23), ggalluvial (v0.12.5), stringr (v1.5.0), ggrepel (v0.9.3), ggnewscale (v0.4.9), RColor brewer (v1.1-3), ggbiplot (v0.55), ggpubr v(0.6.0), gridExtra (v2.3), pheatmap (v1.0.12), ggplotify (v0.1.1), EnhancedVolcano (v1.16.0), Database for Annotation, Visualization, and Integrated Discovery (DAVID) (https://david.ncifcrf.gov) DRAGEN COVIDseq Test App (via Illumina BaseSpace Cloud Computing Platform)

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

Single-cell RNA-seq data is publicly available for download and visualization via the Broad Institute Single Cell Portal (Accession Code SCP2593): https://singlecell.broadinstitute.org/single_cell/study/SCP2593. Custom reference FASTA and GTF for SARS-CoV-2 is available for download: https://github.com/ShalekLab/SARSCoV2-genome-reference.

Research involving human participants, their data, or biological material

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

Reporting on sex and gender	Sex was collected as a variable and is reported in Table 1, and is also a metadata category on the Single Cell Portal
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	Participants were eligible for inclusion in the COVID-19 group if they were at least 18 years old, had a positive nasopharyngeal swab for SARS-CoV-2 by PCR, had COVID-19 related symptoms including fever, chills, cough, shortness of breath, and sore throat, and weighed more than 110 lb. Participants were eligible for the Control group if they were at least 18 years old, had a current negative SARS-CoV-2 test (PCR or rapid antigen test), and weighed more than 110 lb. Exclusion criteria for the cohort included a history of blood transfusion within 4 weeks and subjects who could not be assigned a definitive COVID-19 diagnosis from nucleic acid testing. 112 individuals were included in the study: 32 individuals sampled during the Ancestral wave of COVID-19 (April-November 2020; 15 female, 17 male), 33 individuals during the Delta wave (July-September 2021; 14 female, 19 male), 21 during the Omicron wave (January-February 2022; 6 female, 15 male) and 26 Control participants (17 female, 10 male). The median age of Ancestral strain participants was 58.5, of Delta participants was 50, of Omicron was 61 and of Control participants was 53 years old. Participants were assigned to the vaccinated group if they had received at least two doses of an mRNA vaccine for COVID-19 prior to sample collection, and to the unvaccinated group if they had not received any COVID-19 vaccine. COVID-19 participants were classified according to the 8-level ordinal scale proposed by the WHO representing the amount of respiratory support required at the peak of their disease.
Recruitment	Eligible participants were recruited from outpatient clinics, medical surgical units, intensive care units (ICU), or endoscopy units at the University of Mississippi Medical Center (UMMC) between April 2020 and September 2022. Our participants may be biased towards those likely to volunteer to participate in a research study.
Ethics oversight	The UMMC Institutional Review Board approved the study under IRB#2020-0065.

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	Sample size was determined based on our prior single-cell work in adult SARS-CoV-2 infection (Ziegler et al. Cell 2021) as well as other nasal scRNA-seq studies of COVID-19 (Chua et al. Nat. Biotech 2020, Loske et al. Nat. Biotech 2021, Yoshida et al. Nature 2022, Lindeboom et al. Nature 2024).
Data exclusions	Cells were excluded based on pre-determined scRNA-seq quality metrics (>50% mitochondrial reads, <200 unique transcripts, <150 unique genes) consistent with metrics applied in our prior work (Ziegler et al., Cell 2021). Samples were excluded if they had less than 30 cells remaining after filtering on these quality control metrics.
Replication	Clustering of data was done repeatedly with minor variations in parameters (number of principal components and resolution). Clusters presented in this paper represent those that were identified consistently. Plots and statistical results were replicated successfully by re-running code. Replication in independent samples was not possible due to the nature of the participant cohort.
Randomization	No intervention occurred in this observational study, and therefore no randomization was done.
Blinding	Blinding was not possible in this study. Samples were collected over a period of two years and were grouped based on the major SARS-CoV-2 variant that was circulating at the time of collection, which was known to the investigators. Prior to sample collection, participants were screened for SARS-CoV-2 infection (PCR test) and vaccination status. During sample processing, investigators were aware of sample identity and the number of samples being processed in each batch belonging to each participant group. Identification of cell clusters was done agnostic to sample identity, but all other analysis was done by comparing participant groups.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*