713 **Supplementary materials:**

714 Materials and Methods

715 Figures S1-S4

716 Tables S1-S8

717

Materials and Methods

Participants and samples. We prospectively investigated hospitalized COVID-19 patients between April 2020 and December 2021 who initially presented with a symptomatic infection and positive SARS-CoV-2 nasopharyngeal swab polymerase chain reaction. All participants were admitted to the Centre Hospitalier de l'Université de Montréal (CHUM) and recruited into the Biobanque Québécoise de la COVID-19 (BQC19)[56]. Patients had no known prior exposure to SARS-CoV-2 (i.e., all infections were primary infections), were not vaccinated at the time of primary sampling (days after symptom onset [DSO] $\leq$ 20), and did not undergo plasma transfer therapy. Blood draws were performed during the acute phase of SARS-CoV-2 infection (defined as DSO $\leq$ 20 days, mean DSO = 12.1 days, DSO range = 6 - 20 days, n = 63 samples) and during various convalescent follow-up time points (defined as DSO > 20 days, mean DSO = 128.8 days, DSO range = 31 - 370 days) for a subset of individuals sampled during the acute phase (n = 39 samples). Additionally, PBMCs collected prior to the COVID-19 pandemic from healthy control individuals living in Montréal, Canada (n = 18 samples) were processed for single-cell data collection in parallel with infected patient samples. We also computationally integrated a set of publicly available healthy controls (n = 90 individuals) described in Randolph et al. (2021)[4], which is detailed below ("Single-cell RNA-sequencing data processing and integration"). The study was approved by the respective IRBs (multicentric protocol: MP-02-2020-8929 for BQC19 participants; CHUM protocol 19.387 for control individuals) and written, informed consent was obtained from all participants or, when incapacitated, their legal guardian before enrollment and sample collection.

DNA sequencing and imputation. DNA was extracted from whole blood using the Chemagic™ DNA Blood 400 H96 kit (Perkin Elmer, CMG-1091). SNP genotyping was conducted using the Axiom™ Precision Medicine Research Array from Applied Biosystems (Applied Biosystems, 902981) per the manufacturer's instructions. The array was processed using the GeneTitan™ Multi-Channel instrument (Applied Biosystems). All samples were grouped with the Axiom Analysis Suite 5.1.1 software, and the "Best Practice Workflow" was performed using the following high-quality call rate parameters: Axiom_PMRA.r3 library and threshold configuration Human.v5 with minimum call rate of 97.0%. Marker quality control tests were performed on a subset of ancestrally homogeneous participants, who were determined via comparison to 2,504 individuals across 5 super populations from the 1000 Genomes Project Phase 3 data [57]. Batch effect quality control and replicate discordance checks were performed, and variants that failed either test were removed. Only single nucleotide variants with single character allele-codes (A, C, G, or T) (PLINK --snps-only 'just-acgt' option) were retained. Additionally, variants with low allele frequencies (minor allele frequency [MAF] < 0.001), low genotyping call rates (marker-wise missingness < 0.01), a deviation from Hardy-Weinberg equilibrium (HWE) (p-value < $1x10^{-6}$), and positioned in regions of high link disequilibrium (LD) were removed.

Sample quality filtering was performed considering the set of filtered genotypes described above. Outlier samples with a high genotype missingness rate (overall missing genotype rate > 0.04) or high/low principal component corrected heterozygosity rate on autosomal chromosomes (> ±3SD, respectively) were considered low quality and removed. Sex chromosome composition was determined by estimating X chromosome marker heterozygosity using PLINK (--check-sex 0.4 0.7). Individuals with discordant self-reported sex and genetic sex were removed prior to genotype imputation. All other samples that passed quality control filters were used for imputation.

763 Genotype phasing and imputation was performed using the Michigan Imputation Server[58] with the
764 TOPMed reference panel[59]. After imputation, variants with a posterior genotype probability (GP)
765 < 90% were set to missing within each individual using QCTOOL (v2.0.7, -threshold 0.9 filter).
766
767 **Whole blood processing.** At the time of sampling, whole blood was collected in up to three tubes
768 containing acid citrate dextrose (ACD) and processed within 6 hours of collection. Blood from the
769 same donor was pooled and centrifuged at 400 g for 10 min at room temperature (RT). After
770 centrifugation, plasma was collected, aliquoted, and stored at -80°C. The remaining blood was
771 topped up to 30 ml with HBSS medium at RT. Ficoll-Paque separation was then used to isolate
772 PBMCs. PBMCs were washed with R+ (RPMI 1640 + 0.1M HEPES + 20 U/ml Penicillin-
773 Streptomycin), resuspended in 5 ml R+ with 10% fetal bovine serum (FBS), and counted with
774 Trypan blue. Cells were spun down at 400 g for 10 min at 4°C and resuspended in cold FBS at 20
775 M/ml. A freezing solution of FBS with 20% DMSO was added drop-by-drop to the cell suspension
776 while the tube was continuously agitated. Cell suspensions were transferred into cryovials (1
777 ml/vial), immediately placed into Mr. Frosty Freezing Containers, and stored at -80°C. The
778 following day, PBMCs were transferred to liquid nitrogen for long-term storage.
779
780 **Sample processing for single-cell RNA-sequencing.** PBMCs were thawed in groups of 3 to 4
781 samples (processing batch 1) or 16 to 19 samples (processing batch 2), rested for 2 hours in RPMI
782 1640 supplemented with 10% FBS (Corning, MT35015CV), 2 mM L-glutamine (ThermoFisher
783 Scientific, 25-030-081), and 10 ug/ml gentamicin (ThermoFisher Scientific, 15710064), and
784 subsequently processed for single-cell collection. Cells from different samples were pooled per
785 processing batch for a total of 29 multiplexed sample batches (n = 124 samples). For each

multiplexed cell pool, 12,000 cells were targeted for collection using the Chromium Next GEM Single Cell 3' Reagent (v3.1 Dual Index chemistry) kit (10x Genomics, 1000268). After GEM generation, the reverse transcription (RT) reaction was performed in a thermal cycler as described (53°C for 45 min, 85°C for 5 min), and post-RT products were stored at -20°C for up to one week until downstream processing.

Single-cell RNA-sequencing library preparation and sequencing. Post-RT reaction cleanup, cDNA amplification and sequencing library preparation were performed as described in the Single Cell 3' Reagent Kits v3.1 (Dual Index) User Guide (10x Genomics). Briefly, cDNA was cleaned with DynaBeads MyOne SILANE beads (ThermoFisher Scientific, 37002D) and amplified in a thermal cycler using the following program: 98°C for 3 min, [98°C for 15 s, 63°C for 20 s, 72°C for 1 min] x 11 cycles, 72°C 1 min. After cleanup with the SPRIselect reagent kit (Beckman Coulter, B23317), libraries were constructed by performing the following steps: fragmentation, end-repair, A-tailing, double-sided SPRIselect cleanup, adaptor ligation, SPRIselect cleanup, sample index PCR (98°C for 45 s, [98°C for 20 s, 54°C for 30 s, 72°C for 20 s] x 14 cycles, 72°C 1 min), and double-sided SPRIselect size selection. Prior to sequencing, all multiplexed single-cell libraries were quantified using the KAPA Library Quantification Kit for Illumina Platforms (Roche, 50-196-5234). For each processing batch (n = 2), libraries were pooled in an equimolar ratio and sequenced 100 base pair paired-end on an Illumina NovaSeq 6000 (processing batch 1 average mean reads per cell = 48,613, average median genes detected per cell = 1,627; processing batch 2 average mean reads per cell = 59,246, average median genes detected per cell = 2,007).

Single-cell RNA-sequencing data processing and integration. FASTQ files from each multiplexed capture library were mapped to the pre-built GRCh38 human reference transcriptome (downloaded 10x Genomics) using the cellranger (v6.0.2) count function[60]. souporcell (v2.0, Singularity v3.4.0)[61] in --skip_remap mode was used to demultiplex cells into samples based on genotypes from a common variants file (1000 Genomes Project samples filtered to SNPs with ≥ 2% allele frequency in the population, downloaded from https://github.com/wheaton5/souporcell). For each sample batch, hierarchical clustering of the known genotypes obtained from DNA-sequencing and cluster genotypes estimated by souporcell was used to assign individuals to souporcell cell clusters. All samples except for three were successfully demultiplexed; samples unable to be confidently assigned to a set of cells were removed (n samples retained = 121). After demultiplexing, Seurat (v4.3.0, R v4.0.3)[62] was used to perform cell-level quality control filtering. One sample was removed due to a very low number of cells captured (n = 20 cells total), leaving a total of 120 samples. High-quality cells were retained for downstream analysis if they had: 1) a "singlet" status called by souporcell, 2) between 500 – 4000 genes detected (nFeature_RNA), 3) a mitochondrial UMI percentage < 20%, and 4) less than 25,000 total molecules (nCount_RNA), leaving 236,143 cells. Gene filtering was performed using the CreateSeuratObject min.cells parameter, in which only genes present in at least five cells were kept (n = 30,986 genes).

Due to the large discrepancy between the number of cells assayed in healthy control individuals (n = 38,663) versus acute and convalescent samples (n = 197,480) in our dataset, we integrated a publicly available set of high-quality cells derived from control, non-infected individuals (n = 124,976 cells, 90 samples) described in Randolph et al., (2021)[4], hereafter referred to as the "non-infected IAV controls". First, we removed IAV-derived transcripts (n = 10 genes) from the raw count matrix of the non-infected IAV controls. Next, we merged all datasets, split

the resulting Seurat object by dataset ("COVID batch1", "COVID batch2" or "IAV controls"), and ran SCTransform[63] to normalize and scale the UMI counts within dataset. We simultaneously regressed out variables corresponding to experiment batch, percent mitochondrial UMIs per cell, and individual label in all datasets, and additionally, regressed out sampling time point (e.g., control, acute, follow-up) in the COVID data. We then integrated the three datasets together using the SelectIntegrationFeatures, PrepSCTIntegration, FindIntegrationAnchors, and IntegrateData framework[62]. After integration, dimensionality reduction was performed via UMAP (RunUMAP function, dims = 1:30) and PCA (RunPCA function, npcs = 30). A Shared Nearest Neighbor Graph was constructed using the FindNeighbors function (dims = 1:20, all other parameters set to default), and clusters were subsequently called using the FindClusters algorithm (resolution = 0.5, all other parameters set to default)[62]. In total, our integrated dataset consisted of 361,119 high-quality cells across all samples (n = 236,143 from the combined COVID datasets, n = 124,976 from the non-infected IAV dataset, n = 208 samples altogether).

Cell type assignment. We performed cell type annotation via label transfer to map cell type information onto our data. To perform the label transfer, we downloaded a multimodal human PBMC reference dataset derived from scRNA-seq paired with CITE-seq as described in Hao et al.[15]. We followed the Seurat v4 Reference Mapping workflow, consisting of the FindTransferAnchors and MapQuery functions, with the Hao et al. reference dataset used as our reference UMAP and the following parameters: normalization.method = "SCT" and reference.reduction = "spca". These fine-scale populations were then collapsed into the following broad super populations encompassing the six major cell types found in PBMCs using the predicted.celltype.l2 definitions derived from Hao et al.: $CD4^+$ T cells = c("CD4 CTL", "CD4

Naive", "CD4 Proliferating", "CD4 TCM", "CD4 TEM", "Treg"), $CD8^+$ T cells = c("CD8 Naive", "CD8 Proliferating", "CD8 TCM", "CD8 TEM"), NK cells = c("NK", "NK Proliferating", "NK_CD56bright"), $CD14^+$ monocytes = "CD14_monocytes", $CD16^+$ monocytes = "CD16_monocytes", and B cells = c("B intermediate", "B memory", "B naive"). In total, we annotated 342,127 high-quality cells falling into the major PBMC populations across all individuals and conditions (n $CD4^+$ T cells = 153,479, $CD8^+$ T cells = 53,562, $CD14^+$ monocytes = 70,060, $CD16^+$ monocytes = 5,446, B cells = 34,805, NK cells = 24,775).

Calculation of pseudobulk estimates. Pseudobulk estimates were used to summarize single-cell expression values into bulk-like expression estimates within samples. This was performed for all six major cell types ($CD4^+$ T cells, $CD8^+$ T cells, B cells, $CD14^+$ monocytes, $CD16^+$ monocytes, NK cells). Within each cell type cluster for each sample, raw UMI counts were summed across all cells assigned to that sample for each gene using the sparse_Sums function in textTinyR (v1.1.3) (https://cran.r-project.org/web/packages/textTinyR/textTinyR.pdf), yielding an n x m expression matrix, where n is the number of samples included in the study (n = 208) and m is the number of genes detected in the single-cell analysis (m = 30,986) for each of the 6 clusters.

Calculation of residuals for modeling. For each cell type, lowly-expressed genes were filtered using cell type-specific cutoffs (removed if they had a median logCPM < 1.0 in $CD14^+$ monocytes, < 1.5 in $CD4^+$ T cells, < 2.0 in B cells and $CD8^+$ T cells, < 2.5 in $CD16^+$ monocytes, and < 3.0 in NK cells), leaving the following number of genes per cell type: $CD4^+$ T cells = 10,337, $CD8^+$ T cells = 10,036, B cells = 10,179, $CD14^+$ monocytes = 10,882, $CD16^+$ monocytes = 9,398, and NK cells = 9,882. Within each cell type, only samples with $\geq$ 5 cells per sample were kept for

877 downstream modeling. Further, three samples were removed for downstream analysis because they
878 consistently clustered as outliers on gene expression PCAs for multiple cell types (one COVID-19
879 patient at the acute infection time point and two non-infected IAV controls), leaving the following
880 number of samples per cell type:

Cell type	N healthy controls	N patients	N follow-ups
B	106	63	38
$CD4^+$ T	106	63	39
$CD8^+$ T	106	63	39
$CD14^+$ monocytes	106	63	39
$CD16^+$ monocytes	47	44	39
NK	63	63	39

881

882 After removing lowly-expressed genes, normalization factors to scale the raw library sizes were
883 calculated using calcNormFactors in edgeR (v 3.26.8)[64]. The voom function in limma (v3.40.6)[65]
884 was used to apply these size factors, estimate the mean-variance relationship, and convert raw
885 pseudocounts to logCPM values. The inverse variance weights calculated by voom were obtained
886 and included in the respective lmFit call for all downstream models unless otherwise noted[65].

887

888 **Calculation of per-individual ssGSEA scores.** To construct the ssGSEA Hallmark pathway
889 scores, we calculated single sample Gene Set Enrichment Analysis (ssGSEA) scores from the
890 pseudobulk COVID-19 patient logCPM gene expression estimates corrected for age, sex, dataset,
891 and the number of cells for a given cell type collected per sample using the Gene Set Variation
892 Analysis (GSVA, v1.32.0) package in R with default parameters and method = "ssgsea"[66]. ssGSEA

is a method that allows you to summarize gene expression patterns for any desired target gene set, and for each sample, it will return a score representative of that gene set. These scores were calculated per cell type, and for each of the pathway-specific ssGSEA scores, the input gene set was derived from either a Hallmark or Gene Ontology (GO) Biological Process gene set[22]. The following gene sets were used to define the per-sample pathway scores: (1) inflammatory response score – Hallmark inflammatory response pathway, (2) TNF-α score – Hallmark TNF-α signaling via NF-κB pathway, (3) oxidative phosphorylation score – Hallmark Oxidative phosphorylation pathway, and (4) antigen processing score – GO Biological Process antigen processing and presentation pathway.

Modeling SARS-CoV-2 infection effects. Only healthy controls and COVID-19 patients sampled during the primary infection time point were retained for modeling of infection effects (i.e., follow-up samples were excluded). The following linear model was used to identify genes differentially expressed between healthy control individuals and COVID-19 patients:

$$E(i,j) \sim \begin{cases} \beta_0(i) + \beta_{age}(i) \cdot age(j) + \beta_{sex}(i) \cdot sex(j) + \beta_{dataset}(i) \cdot dataset(j) + \\ \beta_{counts}(i) \cdot counts(j) + \varepsilon^{ctl}(i,j) \text{ if condition} = ctl \\ \\ \beta_0(i) + \beta_{COVID}(i) + \beta_{age}(i) \cdot age(j) + \beta_{sex}(i) \cdot sex(j) + \\ \beta_{dataset}(i) \cdot dataset(j) + \beta_{counts}(i) \cdot counts(j) + \varepsilon^{COVID}(i,j) \text{ if condition} = COVID \end{cases}$$

Here, $E(i,j)$ represents the expression estimate of gene i for individual j, $\beta_0(i)$ is the global intercept accounting for the expected expression of gene i in a non-infected female measured in the COVID batch 1 dataset, and $\beta_{COVID}(i)$ represents the global estimate of the effect of SARS-CoV-2 infection in patients per gene. Age represents the mean-centered, scaled (mean = 0, sd = 1) age per

individual, with $\beta_{age}(i)$ being the effect of age on expression levels, sex represents the self-identified sex for each individual (factor levels = “Female”, “Male”), with $\beta_{sex}(i)$ capturing the effect of sex on expression, dataset represents the dataset in which the sample was obtained (factor levels = "COVID batch 1", "COVID batch 2", "IAV controls"), with $\beta_{dataset}(i)$ capturing the dataset effect, and counts represents the number of cells captured within that cell type for sample *j*, with $\beta_{counts}(i)$ capturing the effect of cell number on expression. Finally, ε^{cdt} represents the residuals for each respective condition (control or COVID) for each gene *i*, individual *j* pair. The model was fit using the lmFit and eBayes functions in limma[65], and the estimates of the global infection effect $\beta_{COVID}(i)$ (i.e., the differential expression effects due to SARS-CoV-2 infection) were extracted across all genes along with their corresponding p-values. We controlled for false discovery rates (FDR) using an approach analogous to that of Storey and Tibshirani[2,67], which derives the distribution of the null model empirically. To obtain a null, we performed 10 permutations, where infection status label (i.e., control/COVID) was permuted across individuals. We considered genes significantly differentially expressed upon infection if they had β_{COVID} $|\log_2 FC| > 0.5$ and an FDR < 0.05.

Modeling COVID-19 disease severity effects within patients. To model the effect of COVID-19 disease severity on gene expression, we restricted our analyses to COVID-19 patients sampled during the primary infection time point for which we had information about disease severity (n = 63). Disease severity was assessed using a five-point scale of respiratory support needed at the time of patient sampling that includes the following categories: 0-Moderate = no supplemental oxygen (n = 16); 1-Severe = nasal cannula (n = 17); 2-Critical = non-invasive ventilation (n = 9); 3-Critical = intubation (n = 20); 4-Critical = extracorporeal membrane oxygenation (ECMO) (n =

1). The following model was used to evaluate the effect of severity at the time of patient sampling on expression:

$$E(i,j) \sim \beta_0(i) + \beta_{severity}(i) \cdot severity(j) + \beta_{age}(i) \cdot age(j) + \beta_{sex}(i) \cdot sex(j) + \beta_{BMI}(i) \cdot BMI(j) + \beta_{dataset}(i) \cdot dataset(j) + \beta_{counts}(i) \cdot counts(j) + \varepsilon(i,j)$$

Here, $E(i,j)$ represents the expression estimate of gene i for individual j, $\beta_0(i)$ is the global intercept accounting for the expected expression of gene i in a female COVID-19 patient, and $\beta_{severity}(i)$ indicates the effect of severity on gene i during the primary sampling time point. Severity ($severity(j)$) represents respiratory support score per individual and was treated as a numeric variable. Body mass index (BMI) represents the mean-centered, scaled (mean = 0, sd = 1) BMI per individual, with $\beta_{BMI}(i)$ being the effect of BMI on expression levels. If BMI was not reported for an individual (n missing = 26), this missing data was filled with the average BMI across patients. All other terms in the model are equivalent to that described in "Modeling SARS-CoV-2 infection effects". The model was fit using the lmFit and eBayes functions in limma[65], and the estimates of $\beta_{severity}(i)$ were extracted across all genes along with their corresponding p-values. We again controlled for false discovery rates (FDR) by empirically deriving the null distribution. To obtain a null, we performed 10 permutations, where respiratory support score (i.e., 0 - 5) was permuted across patients. We considered genes significantly correlated with disease severity if they had an FDR < 0.05.

Gene set enrichment analyses. The R package fgsea (v1.10.1)[68] was used to perform gene set enrichment analysis for the severity effects using the H hallmark gene sets[23]. Ranked t-statistics for each cell type were obtained directly from the topTable function in limma[65], and the background set for a cell type was the set of genes sufficiently expressed (i.e., passed the lowly-

960 expressed gene filter threshold) for that cell type. Pre-ranked t-statistics were used to perform the
961 enrichment using fgsea with the following parameters: minSize = 15, maxSize = 500, nperm =
962 100,000. Normalized enrichments scores (NES) and Benjamini-Hochberg adjusted p-values
963 output by fgsea were collected for each analysis.
964
965 **eQTL mapping and integration with mashr.** eQTL mapping was performed for each cell type
966 using the pseudobulk expression data. A linear regression model was used to ascertain associations
967 between SNP genotypes and expression levels. Input expression matrices were quantile-
968 normalized within each set of disease state samples (i.e., healthy controls, acute COVID-19
969 patients, and follow-ups) prior to association testing. eQTL were mapped separately for each
970 disease state using the R package MatrixEQTL (v2.3)[69]. Prior to mapping, SNPs were filtered
971 using the following criteria in our COVID-19 dataset and the Randolph et al. dataset separately:
972 1) keep those with a minor allele frequency > 5% across all individuals, 2) exclude those with >
973 10% of missing data, and 3) exclude those that deviate from Hardy-Weinberg equilibrium at $p < 10^{-5}$
974 (--maf 0.05 --geno 0.10 --hwe 0.00001 PLINK v1.9 filters)[70]. Only SNPs that passed these
975 filters and were present in both datasets were retained and merged across datasets (n = 4,194,100
976 SNPs kept). Local associations (i.e., putative *cis*-eQTL) were tested against all SNPs located
977 within the gene body and 100 kilobases upstream and downstream of the transcription start site
978 (TSS) and transcription end site (TES) for each gene tested.
979 Within our follow-up samples, some individuals were sampled multiple times during the
980 convalescent period. To avoid counting these genetically duplicate samples more than once when
981 eQTL mapping, we downsampled the follow-ups to include only a single sample with DSO > 20
982 per individual. For each individual with multiple follow-up time points, we chose to keep the

sample with the maximum DSO, which dropped our sample size from n = 39 to n = 26. This duplicate sampling structure was not present in the healthy control or acute COVID-19 samples, so the full sample set was used to map eQTL for these disease states.

We accounted for unmeasured surrogate confounders by performing PCA on a correlation matrix based on the gene expression data. Subsequently, up to 15 principal components (PCs) were regressed out prior to performing the association analysis for each gene. A specific number of PCs to regress in each cell type-disease state pair, corresponding to the number of PCs that led to the detection of the largest number of eQTL in each condition, was then chosen empirically (Table S8). To avoid spurious associations resulting from population structure, the first two eigenvectors obtained from a PCA on the genotype data using SNPRelate (v1.20.1, gdsfmt v1.22.0)[71] were included in the linear model. Other covariates included were age (mean-centered, scaled), sex, number of cells detected per sample, and dataset.

To gain power to detect *cis*-eQTL effects, we implemented mashr[25], which leverages sharing information across cell types and disease states. We considered a set of shared genes that were expressed across all cell types (n = 7,646). For each of these genes, we chose the single top *cis*-SNP, defined as the SNP with the lowest FDR across all cell types (n = 6) in the acute COVID-19 patient condition, to input into mashr. We extracted the effect sizes and computed the standard errors of these betas from the Matrix eQTL outputs for each gene-SNP pair across cell types and conditions. We defined a set of strong tests (i.e., the 7,646 top gene-SNP associations) as well as a set of random tests, which we obtained from randomly sampling 200,000 rows of a matrix containing all gene-SNP pairs tested merged across conditions. The mashr workflow was as follows: i) the correlation structure among the null tests was learned using the random test subset, ii) the data-driven covariance matrices were learned using the strong test subset (from 5 PCs), iii)

the mash model was fit to the random test subset using canonical and data-driven covariance matrices, and iv) the posterior summaries were computed for the strong test subset. We used the local false sign rate (lfsr) to assess significance of our posterior eQTL effects and considered a gene-SNP pair to have a significant eQTL effect if the lfsr was < 0.10.

Calculation of functional cell state scores per cell. To obtain the cell state scores used for modeling cell state-dependent single-cell eQTL, first, the raw single-cell UMI counts across all samples were obtained per cell type. All subsequent processing steps were performed for each cell type independently. Raw cell counts in the form of a Seurat object were split by dataset, and SCTransform was used to normalize and scale the UMI counts within dataset, regressing the effects of experiment batch, percent mitochondrial UMIs per cell, and age in all datasets, and additionally, sex in the COVID batch 1 and batch 2 datasets. The SelectIntegrationFeatures, PrepSCTIntegration, FindIntegrationAnchors, and IntegrateData pipeline was then used to integrate cells, returning all features following integration (features.to.integrate = all_features)[60]. The scaled data matrix (@scale.data slot) of the integrated data, which holds the residuals of the corrected log-normalized integrated counts, was obtained, and these values were used to calculate ssGSEA scores (using the same parameters described above in "Calculation of per-individual ssGSEA scores") per cell for our pathways of interest. Here, we applied ssGSEA to the full scaled SCTransform gene *x* cell matrix, allowing us to generate cell state scores for each single cell in the dataset. Our pathways of interest included the following immune-related and metabolism-related pathways in the MSigDB Hallmark gene sets (n = 6)[22]: Apoptosis, Inflammatory response, Interferon-α response, Interferon-γ response, Oxidative phosphorylation, and TNF-α signaling via NF-κB.

Modeling cell state-genotype interaction effects. We used a poisson mixed effects model to test for cell state-dependent eQTL because this model has previously been used to detect significant cell state-genotype interaction effects in single-cell data[7]. Only COVID-19 patients sampled during the primary infection time point were included in these analyses (n = 63). Single-cell eQTL modeling was performed independently in each cell type; for each cell type, we tested the gene-SNP pairs for which we had evidence of a significant eQTL (lfsr < 0.10) within patients in the pseudobulk eQTL analysis (n genes: B cells = 1,395, $CD4^+$ T cells = 1,804, $CD8^+$ T cells = 1,508, $CD14^+$ monocytes = 2,084, $CD16^+$ monocytes = 1,410, NK cells = 1,523). For $CD4^+$ T cells, we downsampled the number of cells prior to constructing the model inputs to 60,000 cells due to vector size constraints in R. To control for genetic background and latent confounders, we included both genotype and expression PCs in our cell state eQTL models. We computed genotype PCs using the same approach as above in "eQTL mapping and integration with mashr". Expression PCs were calculated from non-batch corrected integrated and scaled counts using the same method as described in "Calculation of functional state scores per cell," but omitting the batch correction step (i.e., no variables were regressed in the SCTransform call). PCA was run on the cell *x* gene matrix of non-corrected integrated and scaled counts subset on the top 3,000 variable features using the prcomp_irlba function in the R package irlba (v2.3.5.1)[72].

To test for interactions with cell state, we used the following poisson mixed effects interaction model, where each gene's UMI counts were modeled as a function of genotype as well as additional donor-level and cell-level covariates. For each gene:

$$\log(E_i) \sim \beta_0 + \beta_G X_{d,G} + \beta_{dataset} X_{d,dataset} + \beta_{age} X_{d,age} + \beta_{sex} X_{d,sex} + \beta_{nUMI} \log(X_{i,nUMI})$$

$$+ \beta_{MT} X_{i,MT} + \sum_{k=1}^{3} \beta_{gPC_k} X_{d,gPC_k} + \sum_{k=1}^{5} \beta_{ePC_k} X_{i,ePC_k} + \beta_{cell\ state} X_{i,cell\ state}$$

$$+ \beta_{G\,x\,cell\ state} X_{d,G} X_{i,cell\ state} + (\phi_d \mid d) + (\kappa_b \mid b) + \varepsilon$$

Here, E is the expression of the gene in cell i, β_0 is the intercept, and ε represents the residuals. All other βs represent fixed effects for various covariates in cell i, donor d, or experimental batch b as follows: G = genotype at the eQTL variant, *dataset* = dataset from which sample originates, *age* = scaled age of donor, *sex* = sex of donor, *nUMI* = number of UMI per cell (accounts for sequencing depth), *MT* = percent of mitochondrial UMIs per cell, *gPC* = genotype PCs, *ePC* = single-cell expression PCs prior to batch correction, and *cell state* = functional cell state score per cell (described above). Donor was modeled as a random individual effect ($\phi_d \mid d$) to account for the fact that multiple cells were sampled per individual, and experimental batch was also modeled as a random effect ($\kappa_b \mid b$). Finally, $\beta_{G\,x\,cell\ state} X_{d,G} X_{i,cell\ state}$ represents the cell state x genotype interaction term of interest.

Single-cell poisson mixed interaction models were fit using the glmer function in the lme4 R package (v 1.1-29) with the following parameters: family = "poisson", nAGQ = 0, and control = glmerControl(optimizer = "nloptwrap")[73]. To determine significance, we used a likelihood ratio test (LRT) comparing two models, one with and one without the cell state interaction term and calculated a p-value for the test statistic against the Chi-squared distribution with one degree of freedom. To correct for multiple hypothesis testing, we performed one permutation in which cell state scores were permuted across all cells per pathway tested, and we obtained a null LRT p-value distribution using the same framework as above with our permuted data. We then calculated q-

values for the cell state-genotype interaction term using the empirical p-value distribution across all tested eQTL using the empPvals and qvalue functions from the qvalue package (v2.16.0)[74].

Colocalization of GWAS and eQTL signals. Specifically for colocalization analyses, eQTL were remapped in each cell type-disease state pair with Matrix eQTL[69] using a 1 megabase (Mb) *cis*-window, with all other modeling parameters kept constant, to broaden our search space and increase our probability of detecting colocalized variants. We assessed colocalization between our identified eQTLs in each cell type-disease state pair and the COVID-19 GWAS meta-analyses of European-ancestry subjects from the COVID-19 Host Genetics Initiative (HGI)[11] release 7 (https://www.covid19hg.org/results/r7/). We tested two outcomes: "critical illness" and "hospitalization" (named A2 and B2, respectively by the COVID-19 HGI). A Bayesian analysis was implemented using the coloc (v5.1.0.1)[75] R package with default settings to analyze all variants in the 1 Mb genomic locus centered on the lead eQTL in the single-cell data. We only considered GWAS loci with associations below 1×10^{-4}. We defined colocalization as PP4 > 0.8, where PP4 corresponds to the posterior probability of colocalization between eQTL and GWAS signals. Colocalization was visualized using the R package LocusCompareR (v1.0.0)[76] with default parameters, except for the genome parameter which was set to "hg38". LD r^2 with the lead SNP was calculated using the default "EUR" population.

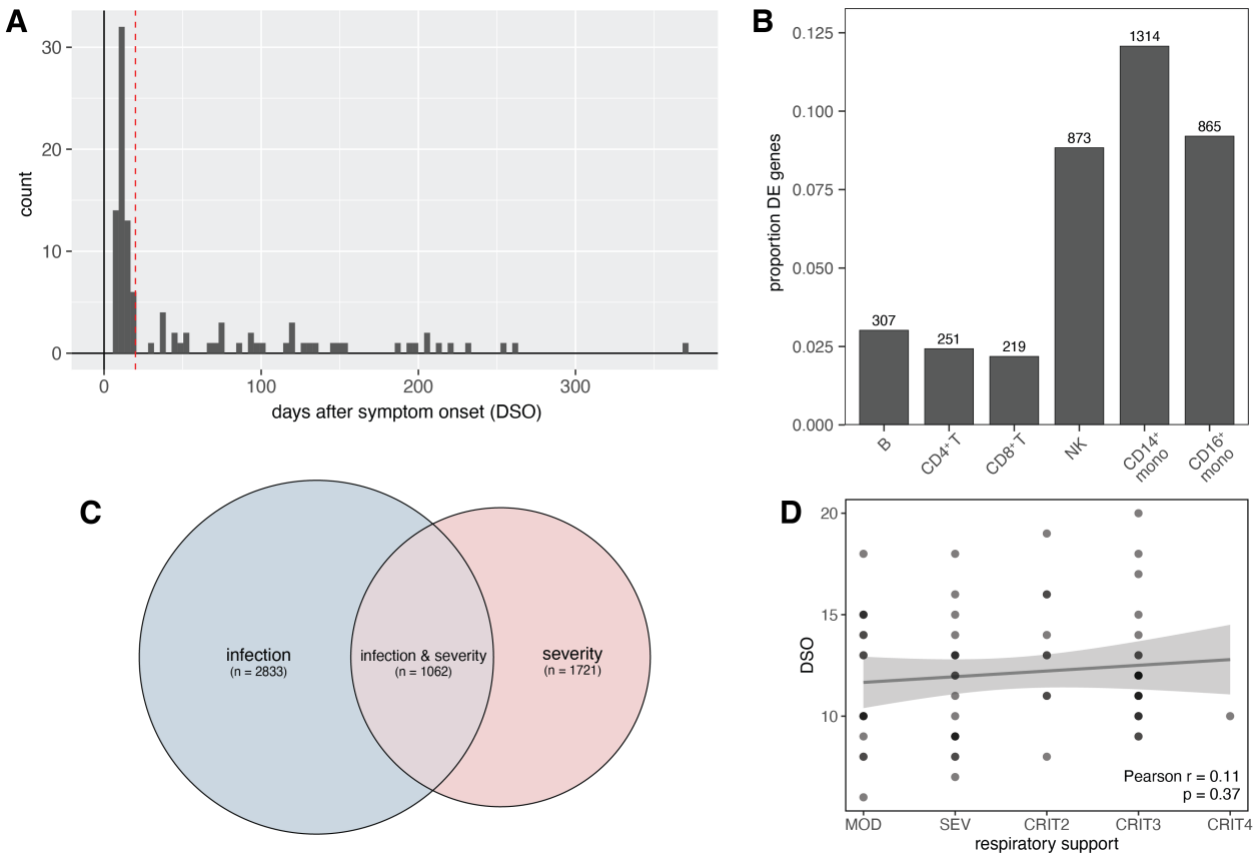

Fig. S1. Sampling time points and global SARS-CoV-2 infection effects. (A) Distribution of days since symptom onset (DSO) at the time of sample collection across acute and convalescent COVID-19 patients in our cohort. Samples were considered to be in the acute phase of infection if DSO ≤ 20 (red line), and samples with DSO > 20 were considered follow-ups. **(B)** Numbers and proportions (y-axis) of genes significantly differentially expressed ($|\log_2 FC| > 0.5$, FDR < 0.05) in COVID-19 patients compared to healthy controls. **(C)** Overlap between the set of significantly differentially expressed genes upon infection (blue circle, left) and the set of genes significantly correlated with disease severity (red circle, right). **(D)** Correlation between respiratory support score and days since symptom onset (DSO). P-value and best-fit slope were determined from a linear regression model correcting for dataset.

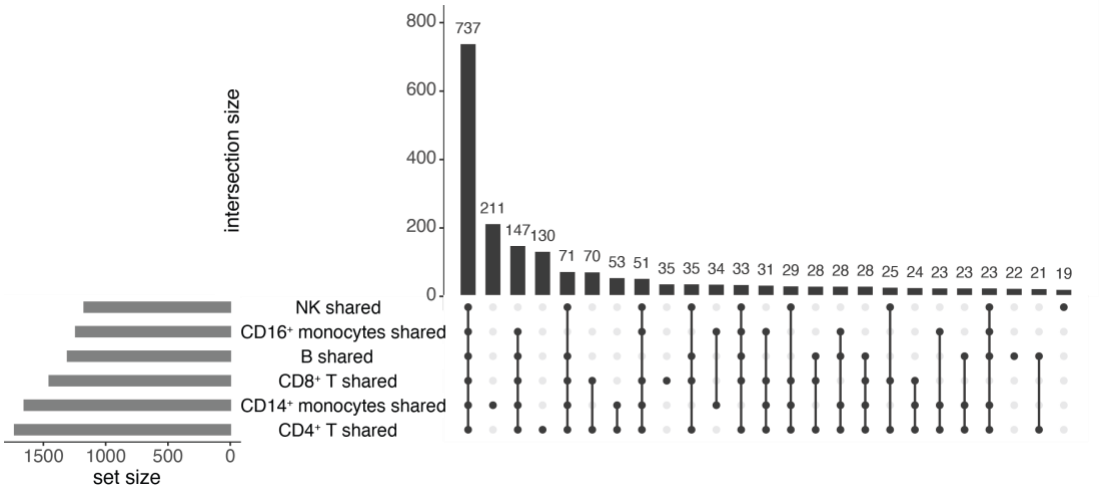

Fig. S2. **Sharing patterns among disease-state-shared eGenes.** Significant eGene sharing patterns among disease-state-shared eGenes ($lfsr_{CTL} < 0.1$ and $lfsr_{COVID} < 0.3$ or $lfsr_{COVID} < 0.1$ and $lfsr_{CTL} < 0.3$) in healthy controls and COVID-19 patients across cell types.

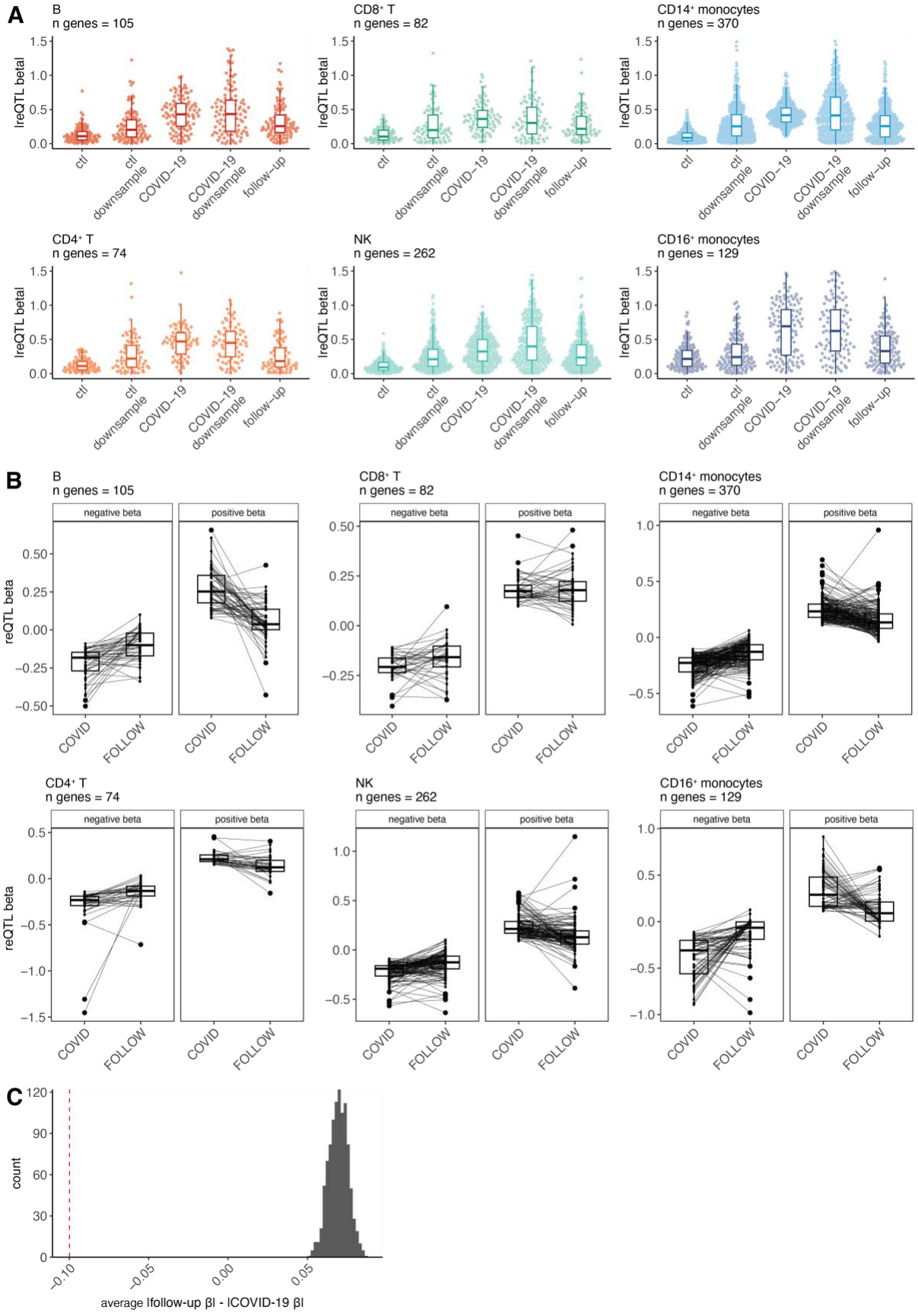

Fig. S3. Cell type-specific response eQTL patterns. (A) Distribution of effect sizes for the cell type-specific reQTL sets plotted across cell types in healthy controls ("ctl"), patients ("COVID-19"), and follow-ups ("follow-up") for the full sample set, as well as a downsampled set in the control ("ctl downsample") and patient ("COVID-19 downsample") groups. Downsampled sets mirrored the follow-up data structure (n = 26 samples) and were derived as follows: i) for controls, 26 individuals were randomly sampled from the control group, and ii) for patients, the 21 follow-up individuals with a corresponding acute infection time point sample were included. Here, all eQTL effect sizes are taken directly from Matrix eQTL (i.e., prior to running mash). **(B)** Paired reQTL effect sizes in COVID-19 patients ("COVID") and follow-ups ("FOLLOW") across cell types. The change in effect size for each gene from patient to follow-up samples is plotted as a black line. **(C)** The observed mean Δ response magnitude across the 370 $CD14^{+}$ monocyte-specific reQTL (red dotted line) compared to the null expectation when permuting random sets of shared eGenes of the same size (n = 370) and computing their mean (n permutations = 1,000, null shown in gray). The observed mean is significantly lower ($p < 0.001$) than random expectation.

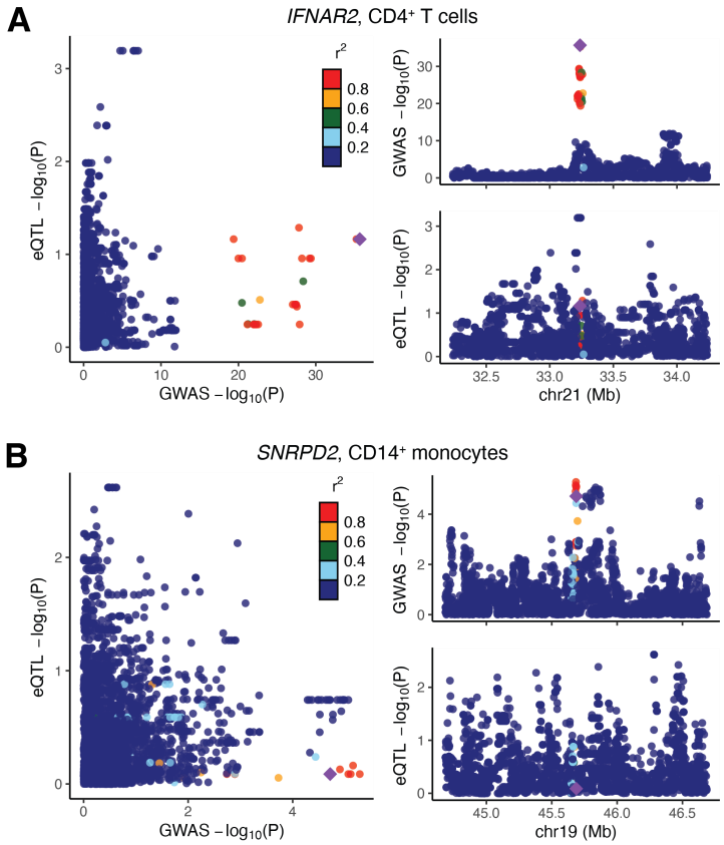

Fig. S4. Colocalization patterns in COVID-19 follow-up samples. (A) The colocalization signal for the lead SNP rs9636867 (*IFNAR2*, $CD4^+$ T cells, GWAS: hospitalization due to severe COVID-19) is absent in follow-ups. **(B)** The colocalization signal for the lead SNP rs7246757 (*SNRPD2*, $CD14^+$ monocytes, GWAS: hospitalization due to severe COVID-19) is absent in follow-ups. For both **(A)** and **(B)**, the larger plot on the left shows the correlation between GWAS p-values (x-axis) and eQTL p-values (y-axis) in follow-ups. Smaller plots on the right show Manhattan plots for the GWAS signal (top) and the eQTL signal in follow-ups (bottom). The lead SNP is depicted as a purple diamond.

Table S8. Gene expression principal components (PCs) regressed in the pseudobulk eQTL analysis. PCs regressed and number of significant eQTL per cell type and disease state are reported.

Cell type	**N Regressed PCs**			**N genes < 0.10 FDR, Matrix eQTL**		
	Control	*COVID-19*	*Follow-up*	*Control*	*COVID-19*	*Follow-up*
CD14$^+$ monocytes	1 to 3	1 to 14	1 to 2	430	1286	56
CD16$^+$ monocytes	1	1	1	10	49	6
CD4$^+$ T	1 to 10	1 to 4	1 to 2	1665	730	77
CD8$^+$ T	1 to 12	1 to 13	1 to 3	424	274	25
B	1 to 5	1 to 8	1	285	192	9
NK	1 to 13	1 to 6	1 to 2	74	230	9