

***home*RNA self-blood collection enables high-frequency temporal profiling of pre-symptomatic host immune kinetics to respiratory viral infection: a prospective cohort study.**

Authors: Fang Yun Lim ^{1,2}, Hannah G. Lea^{2,3}, Ashley Dostie², Tammi van Neel², Grant Hassan², Meg G. Takezawa², Lea M. Starita^{4,5}, Karen Adams^{2,6}, Michael Boeckh^{1,7}, Joshua T. Schiffer^{1,7}, Alpana Waghmare^{*1,8,9}, Erwin Berthier^{*2}, Ashleigh B. Theberge^{*2,10}

Affiliations:

¹Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Center; Seattle, WA, U.S.A.

²Department of Chemistry, University of Washington; Seattle, WA, U.S.A.

³Department of Therapeutic Radiology, Yale University School of Medicine; New Haven, CT, U.S.A.

⁴Brotman Baty Institute, University of Washington; Seattle, Washington

⁵Department of Genome Sciences, University of Washington, Seattle, Washington, U.S.A.

⁶Institute of Translational Health Sciences, School of Medicine, University of Washington, Seattle, WA, U.S.A.

⁷Department of Medicine, University of Washington; Seattle, Washington, U.S.A.

⁸Department of Pediatrics, University of Washington; Seattle, Washington, U.S.A.

⁹Seattle Children's Research Institute; Seattle, Washington, U.S.A.

¹⁰Department of Urology, University of Washington; Seattle, Washington, U.S.A.

*Co-senior authors:

Ashleigh B. Theberge (abt1@uw.edu),

Erwin Berthier (erwin.berthier@gmail.com),

Alpana Waghmare (awaghmare@fredhutch.org)

Supplemental Methods

homeRNA blood collection

Sampling kit assembly and device fabrication: Tasso-SST blood collection devices were purchased from Tasso, Inc. The stabilizer tube components were injection molded out of polycarbonate (PC: Makrolon 2407) by Protolabs, Inc (Maple Plain, MN). Fabrication of the stabilizer tube and components of the *homeRNA* blood collection kit has been previously described¹.

Blood collection: Blood collection was performed at the residential homes or other temporary locations (if traveling during study) on pre-scheduled sampling days. Detailed description of the *homeRNA* collection procedures has been previously published². Briefly, after sanitizing and warming the collection site (upper arm), Tasso-SST blood collection device was adhered, activated, and up to 0.5 mL of liquid capillary blood was collected. Immediately after collection, participants connected the blood collection tube to the stabilizer tube containing RNAlater and to mix and complete stabilization (**Figure 1B**). Stabilized blood was packaged according to the return instructions, stored at ambient temperature, and returned to the lab via overnight courier services. Samples collected on weekdays were mailed back to the University of Washington within 24 hours of collection using Next Day Courier Service (overnight shipping) and delivered directly to a secure -20°C freezer. When collection was scheduled on a weekend, participants were asked to store stabilized samples in a cool location until the first available courier pick up service the following Monday. Samples were transferred to -80°C for long term storage until further sample processing.

Respiratory specimen collection, viral load kinetics, and SARS-CoV-2 viral sequencing

Participants were asked to collect nasal swab samples from both nostrils using a sterile polyester-flocked swab (Puritan PurFlock Ultra; #25-3306-U) and transfer swabs into universal transport medium (Copan Diagnostics) immediately after collection. The first nasal swab (NS1) was returned to a CLIA-certified lab for SARS-CoV-2 RT-PCR testing and the remaining swabs (NS2 - NS7) were returned to the University of Washington for multi-pathogen array analysis. Nasal swabs were delivered directly to a secure -20°C freezer and transferred to -80°C for long term storage. Nucleic acid extraction and pathogen quantification using the Multi-pathogen OpenArray platform for non-clinical testing of respiratory pathogens was performed as previously described². For each participant, 3-4 nasal swabs were tested for SARS-CoV-2 and 18 common respiratory pathogens as previously described. For participants with positive SARS-CoV-2 results, one representative swab with the lowest C_{rt} value (highest viral load) was selected for SARS-CoV-2 sequencing.

Online surveys

All surveys were administered online via REDCap³. A Welcome Survey was administered immediately upon screening and consenting. The survey collected information on recent SARS-CoV-2 test, COVID-19 vaccination history, clinical traits (age, sex, weight), ethnicity and race, living arrangements, and pre-existing health conditions. Daily Use Surveys were administered on each scheduled collection timepoints and collected symptom burden information and usage parameters associated with the self-sampling devices. Two general surveys intended to query general user experience pertaining to the decentralized study framework such as timeliness of courier services and convenience of the specimen return logistics were administered at the end of week 1 and again, at the completion of all study procedures.

Gene expression analysis

RNA isolation, cleanup, and concentration

Total RNA was isolated using the Ribopure™ Blood RNA Isolation Kit (Thermo Fisher #AM1928) according to the manufacturer-recommended alternate protocol for enrichment of small RNAs (e.g. miRNA, tRNA, 5S/5.8SrRNAs). Briefly, stabilized blood cells were pelleted, lysed using 800 µL of lysis solution and 50 µL of acetic acid, and RNA was extracted using 500 µL of acid-phenol:chloroform (PCI). The RNA-containing aqueous phase was collected and denatured with 1 mL Denaturation Solution (Ambion #AM8540G) and 1.25 v/v of 100% ethanol. Nuclease-free water was added in 300 µL increments until solution turned clear. For the final purification of total RNA enriched with small RNAs, extracted RNA was bound to a silica filter cartridge, washed once with 70:30 (v/v) ethanol: denaturation, followed by two washes of the final wash solution (80% ethanol/50mM sodium chloride), and eluted twice in 100 µL of total elution volume. RNA yield was quantified on the Cytation5 Take3 plate. RNA quantification for nCounter gene expression analysis was performed on the Qubit 4 fluorometer using the Qubit RNA HS Assay Kit (Thermo Fisher). RNA quality was measured on the Bioanalyzer 2100 (Agilent Technologies). RIN values for RNA samples with < 5 ng/µL were assayed using the RNA 6000 Pico Kit (Agilent Technologies #5067-1513) and > 5 ng/µL were assayed using RNA 6000 Nano kit (Agilent Technologies #5067-1511). RNA samples were stored at -80 °C until ready for nCounter gene expression analysis.

nCounter gene expression analysis: The nCounter Pro Analysis System (nanoString) was used to perform direct detection and digital counting of native RNA transcripts. 50-120 ng of total RNA from each participant sample were hybridized to the nCounter Host Response Panel codeset version v1.1 (nanoString) to generate the expression dataset. The nCounter Host

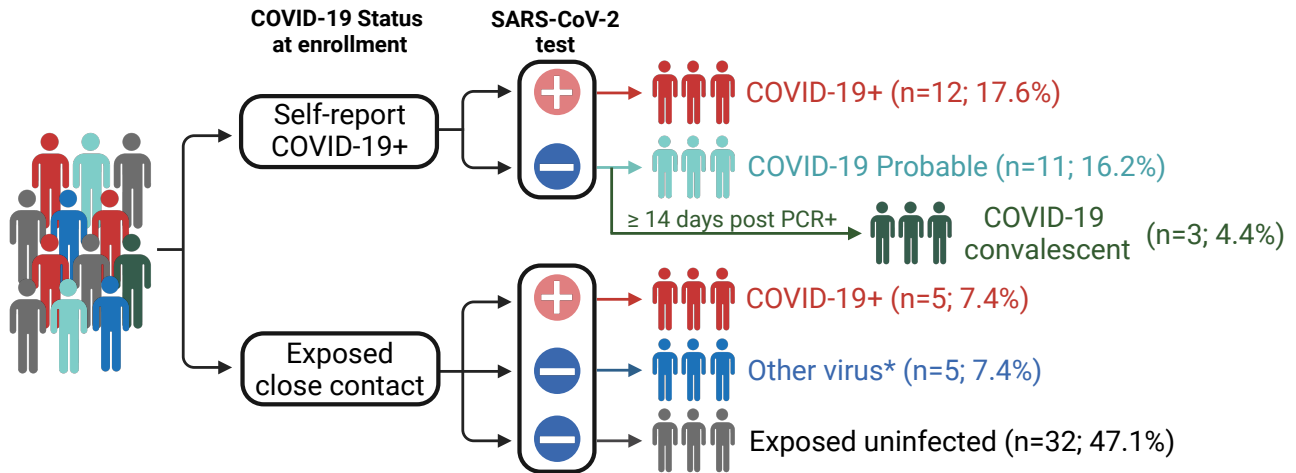
Response codeset targets 773 genes associated with the immune response to infectious disease along with 12 candidate reference (housekeeping) genes. The target-probe hybrids were immobilized on a cartridge, aligned, and digitally counted on the nCounter Pro digital analyzer.

nCounter data quality control and normalization: Raw nCounter expression counts were normalized using the nSolver™ software (nanoString). Quality control (QC) and normalization procedures were performed as previously described². All samples passed QC metrics for imaging (> 75% field of view), binding density (0.1 – 2.25 spots/square micron), and positive control linearity (> 0.95). Samples were subjected to i) positive control normalization and ii) codeset content normalization. Five reference genes (*GUSB*, *HRPT1*, *MRPS7*, *NMT1*, *PGK1*) were used in codeset content normalization. Normalized expression counts are provided in supplemental data 1.

Functional enrichment analysis: Functional enrichment analyses were performed using the `gseGO()` function of the `clusterProfiler` package^{4,5}. Significantly overexpressed and underexpressed transcripts identified from differential gene expression analysis were analyzed separately. Enrichment of all Gene Ontology (GO) terms (biological processes, cellular component, and molecular functions) were performed for significant transcripts identified from each pairwise comparisons. GO terms with an adjusted p-value < 0.05 following the Benjamini-Hochberg procedure were classified as significant. For figures 4B and S3B, GO biological process terms were ranked based on adjusted p-values and the top 20 enriched terms associated with transcripts overexpressed in the G1-G7 time-series analyses were compared using the `compareCluster()` function of the `clusterProfiler` package and visualized as dot plots using the `dotplot()` function of the `enrichplot` package. The top 30 enriched terms were visualized in **Figure S5B**. To facilitate interpretation of the enriched GO terms and reduce term redundancies, semantic similarity between pairs of GO terms were calculated and a matrix of dissimilarities was computed using the `calculateSimMatrix()` function of the `rrvgo` package⁶. The terms were clustered using complete linkage and reduced using the `reduceSimMatrix()` function at a threshold of 0.7. Reduced terms were visualized as treemaps and scatterplots (**Figure S4**).

Other data visualization: Volcano plots and box plots were generated in R Statistical Software (v4.2.1, R Core Team 2022)⁷. Study workflow and conceptual diagrams were constructed in Biorender and figures were compiled in Adobe illustrator.

A Assignment of comparator groups



B COVID-19+ participant classification

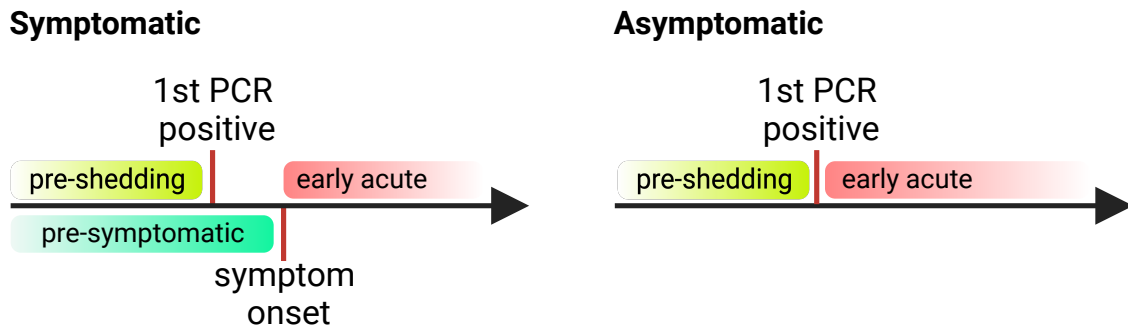


Figure S1. Participant classification. A) Comparator group assignment based on nasal swab test results B) Classification of symptomatic and asymptomatic COVID-19+ participants based on timing of their first sample collection day (colored boxes).

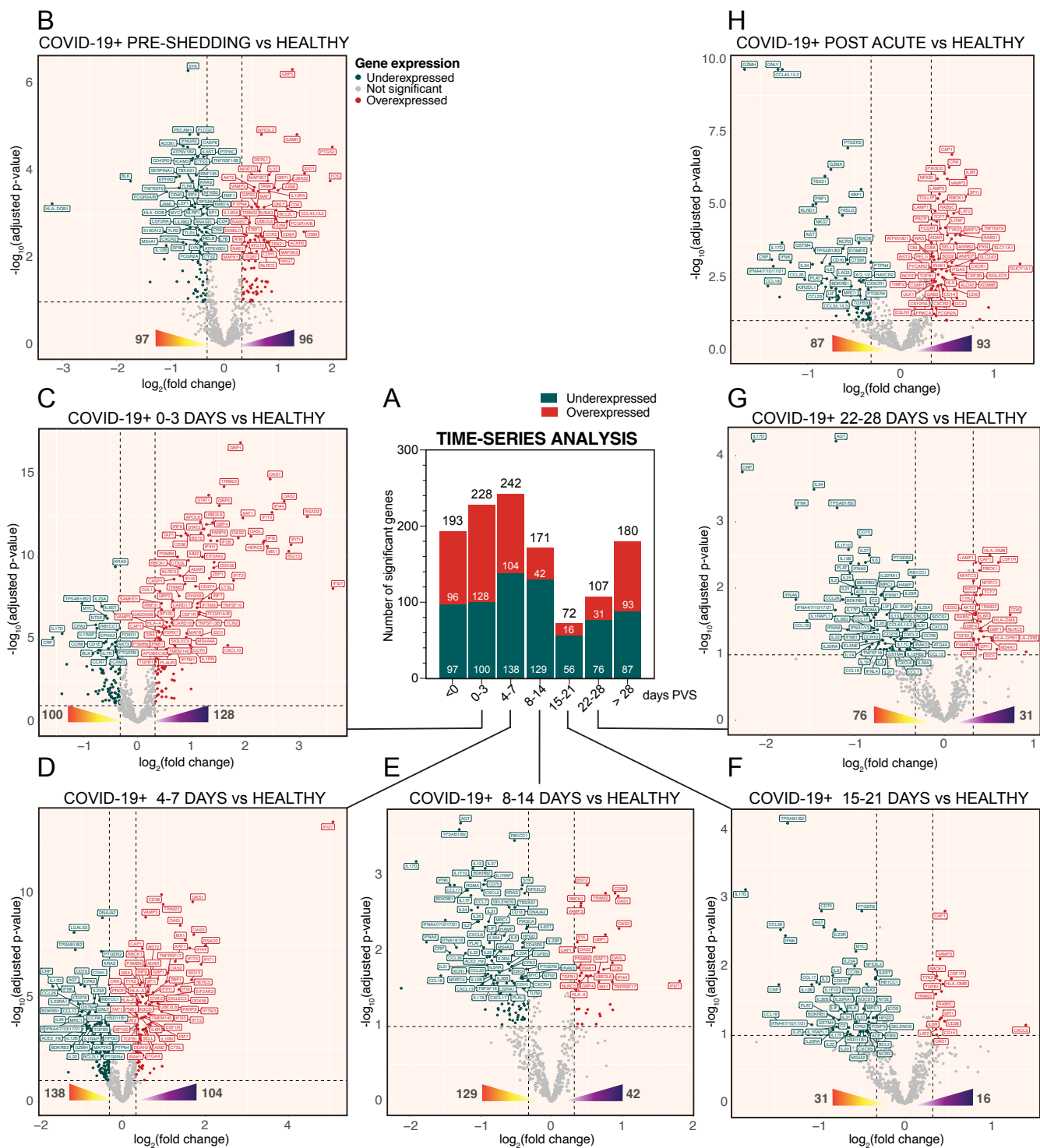
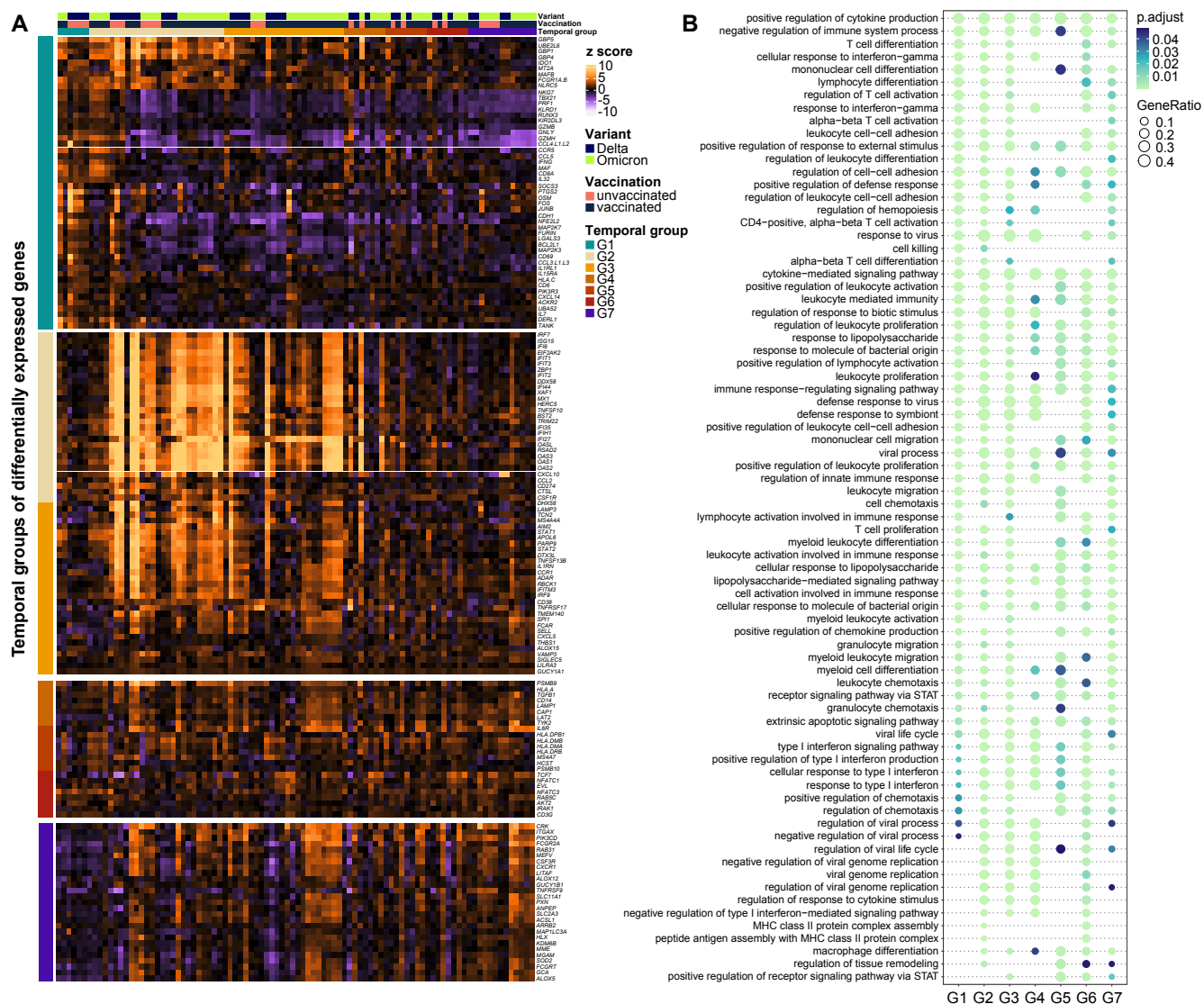
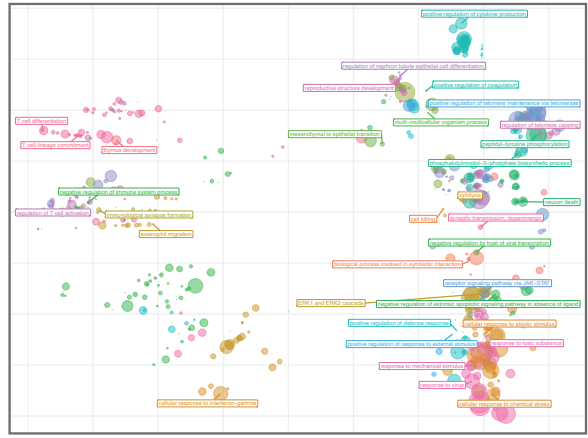


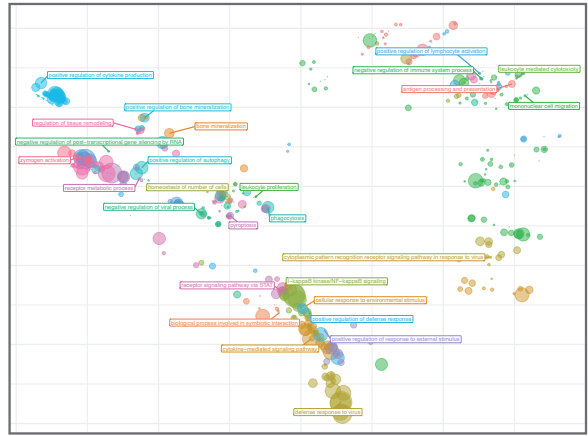
Figure S2. Dynamics of host immunity to SARS-CoV-2 spanning the pre-shedding (G1) to post-acute (G7) phase of infection. A) Differentially expressed genes (DEGs) identified in each of the seven COVID-19+ temporal groups (G1-G7) compared to unexposed-healthy participants. Volcano plots depict DEGs in B) G1 pre-shedding (<0 days PVS), C) G2 early acute (0-3 days PVS), D) G3 early acute (4-7 days PVS), E) G4 mid acute (8-14 days PVS), F) G5 mid acute (15-21 days PVS), G) G6 late acute (22-28 days PVS), and H) G7 post-acute (>28 days PVS) temporal groups.



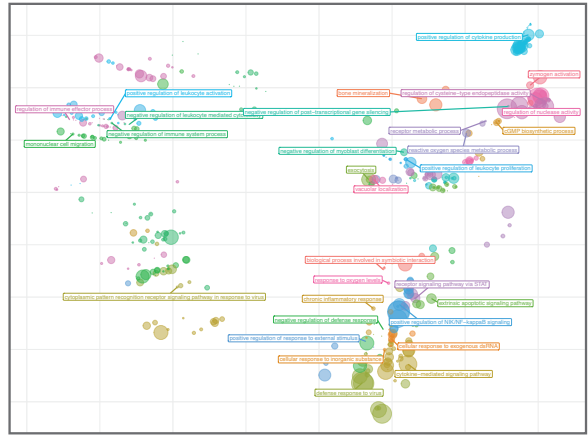
G1



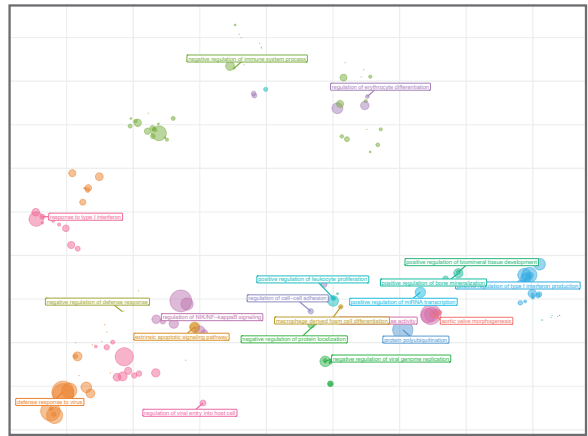
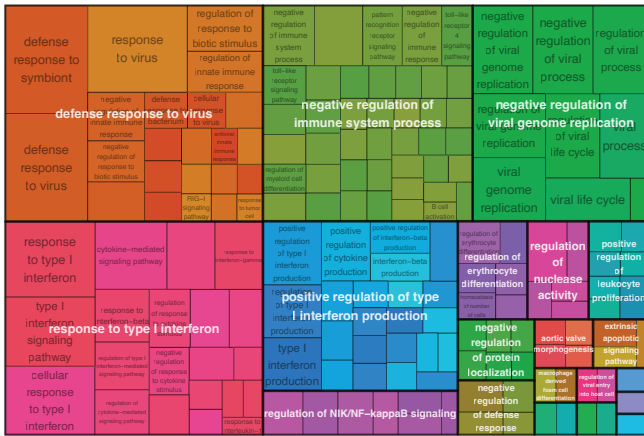
G2



G3



G4



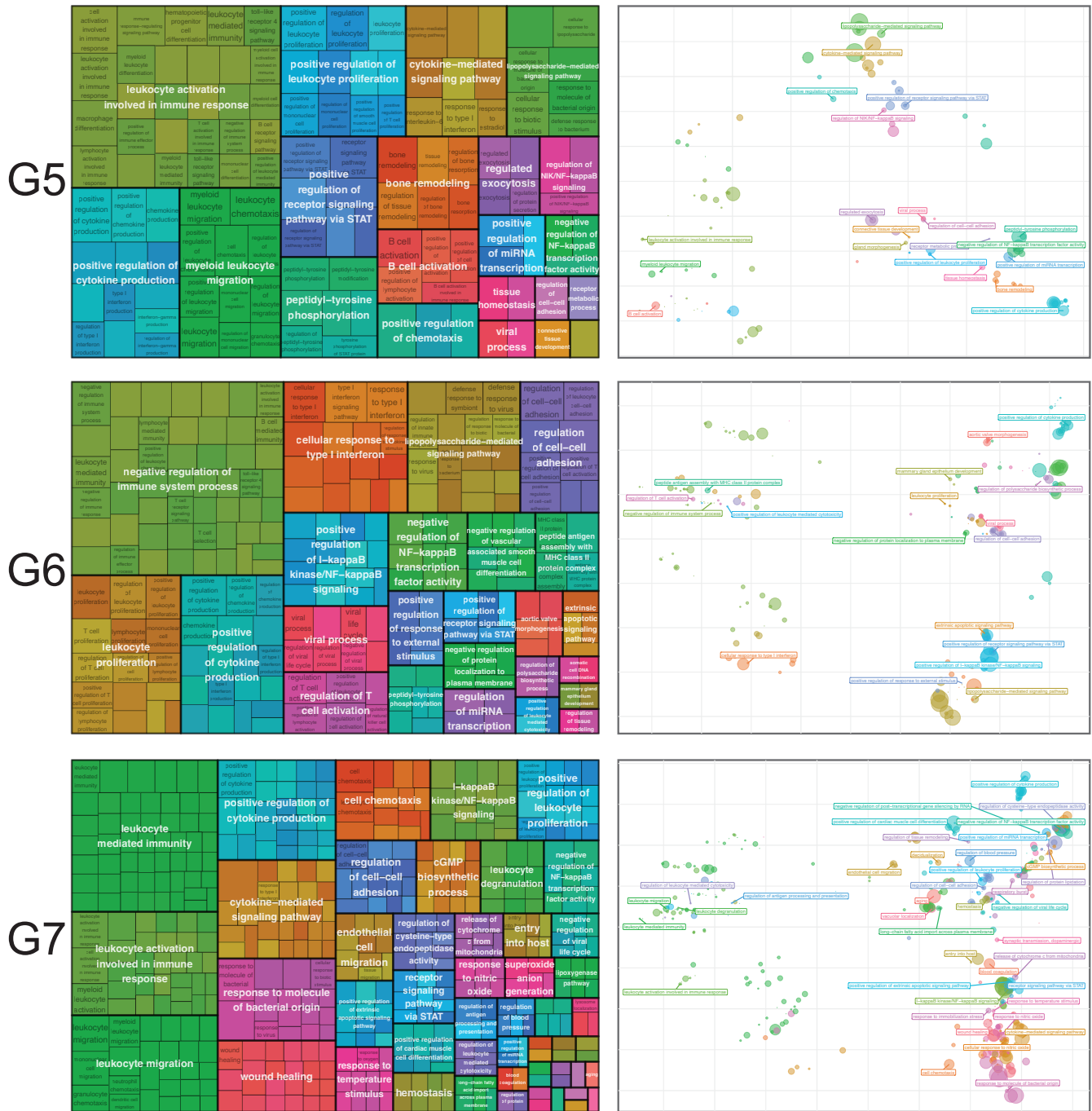


Figure S4. Gene Ontology (GO) enrichment analysis of genes overexpressed in COVID-19+ participants across G1-G7 temporal groups. Treemaps (left) and scatterplots (right) of reduced GO terms (see supplemental methods for detailed description of GO term reductions).

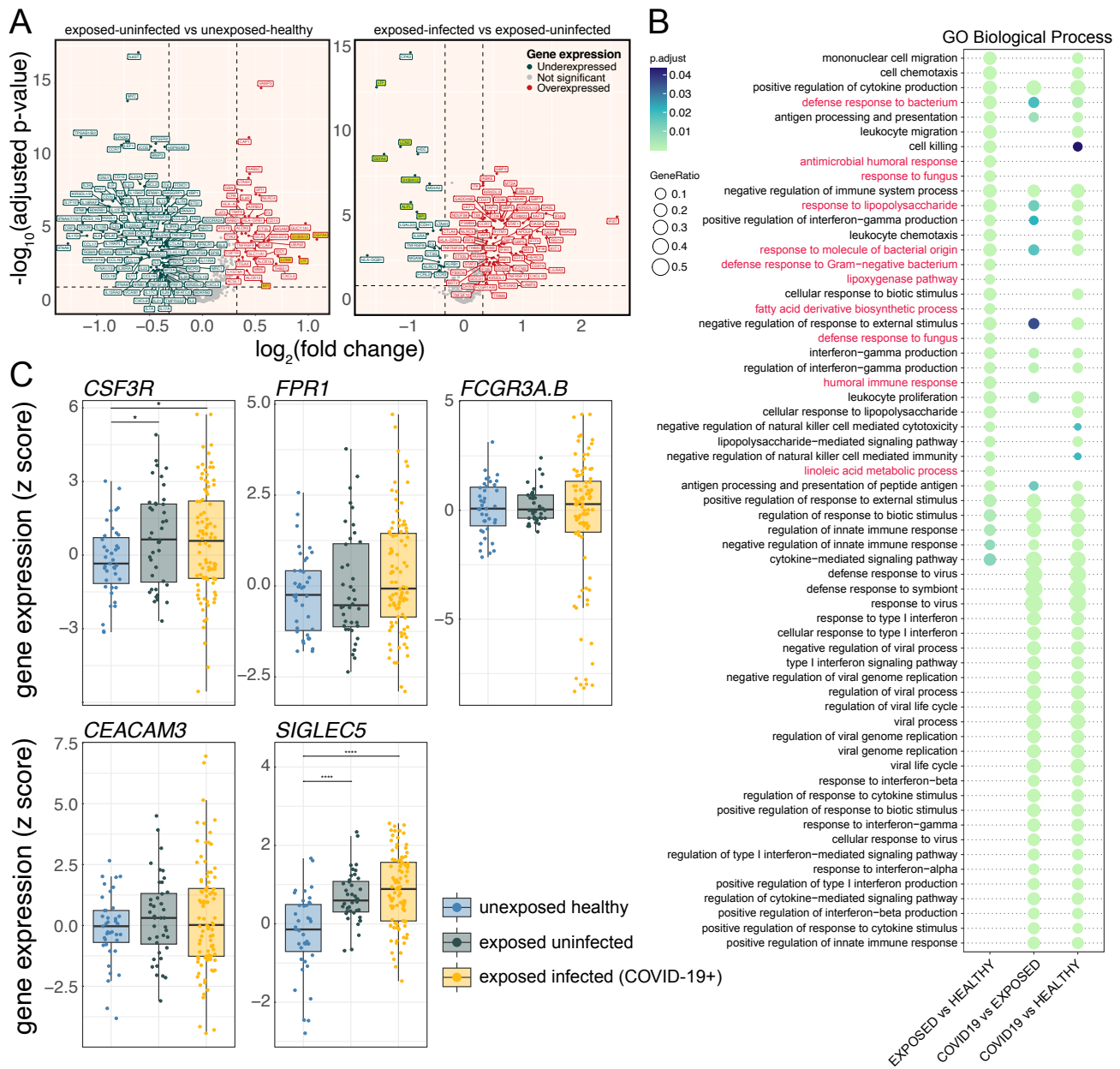


Figure S5. Enrichment of host antimicrobial defense pathways in exposed-uninfected individuals. Volcano plots depict significant genes from pairwise comparisons of **A**) COVID-19+ versus exposed-uninfected and exposed-uninfected versus unexposed-healthy participants. Significantly overexpressed and underexpressed genes are colored in red and green respectively. Top 150 significant genes based on adjusted p-value are labeled on the plot. Neutrophil-associated HDPs, most cell markers, and T cell markers are highlighted in yellow. **B**) Dot plot displaying the top 30 overexpressed GO biological processes from [exposed-uninfected vs unexposed-healthy], [exposed-infected vs exposed-uninfected], and [exposed-infected vs unexposed-healthy] pairwise comparisons ranked by adjusted p values. Dot size represents gene ratio and color represents adjusted p-values. Pathways associated with host bacterial and fungal defenses are highlighted in red. **C**) Box plots depicting gene expression z scores of neutrophil-associated receptors and surface proteins for all longitudinal samples stratified by unexposed-healthy, exposed-uninfected, and exposed-infected (COVID-19+) participants. Statistical significance between groups were determined using the Wilcoxon signed rank test with Benjamini-Hochberg multiple testing correction procedure and significance between groups were labeled as follows: $p > 0.05$ (ns), $p \leq 0.05$ (*), $p \leq 0.01$ (**), $p \leq 0.001$ (***), $p \leq 0.0001$ (****).

SUPPLEMENTAL TABLES

Table S1. List of respiratory pathogens screened

Pathogen Category	Strains
Influenza	Pan Flu A, Flu A (H3N2), Flu A (H1N1), Pan Flu B, Pan Flu C
Parainfluenza	Human Parainfluenza 1 - 4
Enterovirus	Pan enterovirus and Enterovirus D68
Rhinovirus	Rhinovirus 1 and 2
Adenovirus	Adenovirus 1 and 2
non-SARS coronavirus	HKU1, NL63, 229E, OC43
SARS coronavirus	SARS-CoV-2
Respiratory syncytial virus	RSV A and RSV B
Metapneumovirus	hMPV
Bacterial pneumonia	<i>Streptococcus pneumoniae</i> , <i>Mycoplasma pneumoniae</i> , <i>Chlamydia pneumoniae</i>

Table S2: Temporal group of COVID-19+ participant samples

Temporal group	Days PVS	Infection stage
G1	< 0	Pre-shedding
G2	0-3	Early acute
G3	4-7	Early acute
G4	8-14	Mid-acute
G5	15-21	Mid-acute
G6	22-28	Late acute
G7	> 28	Post-acute

References and Notes

- Haack AJ, Lim FY, Kennedy DS, et al. homeRNA: A Self-Sampling Kit for the Collection of Peripheral Blood and Stabilization of RNA. *Anal Chem* 2021; **93**(39): 13196-203.
- Lim FY, Kim SY, Kulkarni KN, et al. High-frequency home self-collection of capillary blood correlates IFI27 expression kinetics with SARS-CoV-2 viral clearance. *J Clin Invest* 2023; **133**(23).
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; **42**(2): 377-81.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012; **16**(5): 284-7.
- Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2021; **2**(3): 100141.
- Sayols S. rrvgo: a Bioconductor package for interpreting lists of Gene Ontology terms. *MicroPubl Biol* 2023; **2023**.
- R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2022.