*home*RNA self-blood collection by exposed close contacts enables high-frequency temporal profiling of the pre-symptomatic host immune kinetics to respiratory viral infection

Authors: Fang Yun Lim [1,2], Hannah G. Lea[2,3], Ashley Dostie[2], Tammi van Neel[2], Grant Hassan[2], Meg G. Takezawa[2], 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:

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

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

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

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

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

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

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

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

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

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

*Corresponding authors: Ashleigh B. Theberge (abt1@uw.edu), Erwin Berthier (erwin.berthier@gmail.com), Alpana Waghmare (awaghmare@fredhutch.org)

ABSTRACT

Background

Host immunity is critical in determining outcomes of acute respiratory viral infections (ARVIs). However, detailed kinetics of host immune responses following natural exposures are poorly understood. Investigating the host response during the pre-symptomatic phase of viral infection is challenging, and prior work has largely relied on human challenge studies. In this prospective longitudinal study, we utilized a self-blood collection tool (*hom*eRNA) to profile the host response during pre-symptomatic ARVIs in recently exposed adults and present a study framework for the conduct of large-scale longitudinal mechanistic studies.

Methods

We prospectively recruited non-symptomatic adults with recent exposure to ARVIs who subsequently tested negative (exposed uninfected) and positive for respiratory pathogens. Study participants performed self-collection of blood and nasal swabs across a 4-week observation window. Daily monitoring of symptoms, viral load, and blood transcriptional responses was performed for the first week followed by weekly monitoring of blood transcriptional responses and symptoms. Nasal swabs were assayed for respiratory pathogens including SARS-CoV-2. Immune kinetics from 132 longitudinal blood samples (8 SARS-CoV-2 infected and 4 exposed uninfected) were profiled at high temporal resolution for 773 host response genes.

Findings

68 participants across 26 U.S. states completed the study between June 2021 – April 2022, with 97.6% of scheduled longitudinal blood collections (n=691), 97.9% of nasal swabs (n=466) and 97.2% of symptom surveys (n=688) returned. SARS-CoV-2 infection was confirmed in 25% of the participants (n=17) Expression of host immediate early genes (IEGs) involved in AP-1 transcriptional complex and prostaglandin biosynthesis along with genes encoding the early T-cell activation antigen (*CD69*), pyrogenic cytokines (IL-6, MIP-1β, and IFN-γ), cytotoxic cell receptors and granule proteins, and interferon-induced GTPases were detected in the periphery prior to onset of

viral shedding in the nasal passage. Upon onset of viral shedding, robust induction of interferon stimulated genes (ISGs) were observed. We also observed elevated expression of the host defense peptides *DEFA4*, *LCN2*, *LTF*, *BPI* (HDPs) in exposed uninfected individuals.

Interpretation

Signatures of T-cell responses prior to nasal viral shedding followed by robust induction of innate ISGs upon onset of viral shedding suggests that T-cell derived immune memory may play a role in pathogen control during early phases of the infection. Elevated levels of HDPs in exposed uninfected individuals suggest a potential role for neutrophil-mediated immunity in host defense during pathogen exposure. Finally, we demonstrated that unsupervised self-collection and stabilization of blood using *home*RNA can be used to study early host immune kinetics to natural ARVIs at a temporal resolution comparable to that of human challenge studies.

INTRODUCTION

Innate and acquired immune responses (host defenses) are critical in determining the outcomes of acute respiratory infections (1). Systemic dysregulation and excessive expression of host defenses are associated with uncontrolled inflammation and poor outcomes, while early effective responses limit the severity and duration of symptoms (2). However, detailed kinetics surrounding mechanisms of early host defenses soon after pathogen exposure and prior to symptom onset or viral shedding are poorly understood, especially in exposed individuals who remained asymptomatic or test negative. As in-person venipuncture still presents a significant logistical barrier to longitudinal studies requiring early and frequent samples, most longitudinal studies of natural ARVIs to date have focused on post-symptomatic infection, typically when viral load is already in decline (3-6). Current studies of natural ARVIs are often limited to cross-sectional cohorts, while longitudinal studies are often limited to infrequent collections across variable collection intervals.

High-resolution immune profiling of the host peripheral immunity after pathogen exposure, particularly during the pre-symptomatic phase, has largely been limited to human challenge studies (7-9). Controlled human infection models pose several limitations to understanding the host response during natural infections (10). Restricted by ethical and regulatory considerations, these models often employ use of a limited set of challenge strains, a uniform viral challenge dose, and are performed on low-risk, healthy volunteers (young adults) with strict selection criteria, thus limiting their broad applicability to naturally circulating variants of concern and to the heterogeneity of population-wide immunologic backgrounds, shaped by prior vaccination and infection. The only SARS-CoV-2 human challenge study published to-date did not assess the blood transcriptional response to the infection (11). Based on this study, onset of nasal viral shedding ranged between 40-76 hours post challenge, making the pre-symptomatic phase of SARS-CoV-2 infection from natural exposures particularly hard to capture and characterize (11).

Studies of the host peripheral transcriptional response during pre-symptomatic phase of natural infections are very limited (12, 13) — with the first of such studies published by McClain et al. in 2021, a five-year prospective

study involving recruitment of a large cohort of Duke university students, following them daily for symptoms, and sending study personnel to the dorms to collect blood and nasopharyngeal swabs daily for five consecutive days (12). This is followed by the COVIDsortium study that assessed healthcare workers weekly for 16 weeks during the pandemic (13). The targeted recruitment at dormitories and healthcare centers enabled both studies to capture the pre-symptomatic phase of ARVIs but neither study assessed intra-individual immune kinetics spanning the period of disease at high temporal resolution.

A study framework that enables rapid dissemination of specimen collection kits to recently exposed individuals across a wide geographic area (decentralized sampling) coupled with frequent self-collections at pre-determined intervals and immediate on-site stabilization of the transcriptome is integral to conducting population-wide longitudinal mechanistic studies aimed at profiling the host immune kinetics during early phase of natural infections (**Fig. 1**). Our team developed a home blood collection kit (*home*RNA) that enables self-sampling and stabilization of liquid capillary blood (14, 15) and demonstrated its capability of capturing host response to SARS-CoV-2 in non-hospitalized adults in addition to distinguishing differential infection response between previously vaccinated and unvaccinated adults (16). Here, we leveraged *home*RNA to profile, at high temporal resolution, the pre-symptomatic host immunity to natural ARVIs in exposed close contacts (CCs) (**Fig. 1**). To the best of our knowledge, this is the first prospective longitudinal study to profile the pre-symptomatic blood transcriptional response to natural SARS-CoV-2 infections in the general population at a temporal resolution comparable to that of human challenge studies.

RESULTS

Study design and participant characteristics

Asymptomatic CCs and SARS-CoV-2 infected (COVID-19+) individuals were recruited from the general population across the United States through referral-based index-cluster snowball sampling (12) and self-report exposures for high-frequency serial sampling (**Fig. 1A – 1C**). Upon informed consent, study kits were mailed to study participants using overnight courier service. All participants collected blood and nasal swab samples for seven

consecutive days followed by three weekly blood collections (**Fig. 1B**). Symptom surveys were administered on each collection day to track symptom progression. The first nasal swabs (NS1) were sent for clinical SARS-CoV-2 testing while the remaining swabs (NS2-NS7) were banked and assayed for the presence of SARS-CoV-2 and other common respiratory pathogens at the conclusion of the study. A select number of participants with positive SARS-CoV-2 clinical test results were invited to extend weekly blood collection for up to four additional weeks (**Fig. 1B**).

Between June 2021 – April 2022, we enrolled a total of 90 asymptomatic COVID-19+ and exposed CCs during the Delta and Omicron pandemic waves for high-frequency serial sampling (**Fig. S1A**). 68 participants (42 CCs and 26 COVID-19+ at enrollment) across 26 U.S. states completed the study (**Fig. 1C**); a total of 691 *home*RNA blood samples (97.6% of scheduled collections; n=708) and 466 nasal swabs (97.9% of scheduled collections; n=476) were returned to our lab and a total of 688 online surveys (97.2%; n=708) were completed (**Fig. S1A**). 22 participants were withdrawn; 59% (13/22) of whom did not initiate sampling procedures (**Fig. S1A**). The median age of participants (n=68) was 38 (IQR: 28.0 - 47.8; range: 20-74) (**Table 1**). The study enrolled participants from diverse demographics including individuals from underrepresented, underreported, and understudied (U3) populations; 70.6% female, 23.5% Hispanic, and a combined 25% of underrepresented racial minorities (7.3% Asian; 5.9% Black, 11.8% two or more races) participants completed the study (**Table 1**).

Infection rates of exposed close contacts and comparator group assignment

A total of 362 nasal swabs from 68 participants were analyzed for SARS-CoV-2 and 18 other respiratory pathogens (**Fig. S1A**). 25% of participants (12 COVID-19+ and 5 CCs) were positive for SARS-CoV-2 with a secondary attack rate of 11.9% (5/42) observed among CCs (**Fig. S1B**). 16.2% of participants reported recent SARS-CoV-2 infection but showed no virological evidence of infection (COVID-19 Probable) (**Fig. S1B**) while other respiratory viruses were detected in 7.4% of study participants (**Fig. S1B** and **Fig. 2A**). 32 CCs remained test-negative (exposed uninfected) (**Fig. S1B**). The median days post onset of viral shedding (days PVS) was 1 (range: -4 – 6) and the median days post symptom onset (days PSO) was 0 (range: -5 – 3) for all COVID-19+ participants (**Table 1**). 23.5% (4/17) of COVID-19+ participants were pre-symptomatic on their first collection while 11.8%

(2/17) remained asymptomatic throughout the observational window (**Fig. 2B**). 11.8% (2/17) of COVID-19+ participants performed their first collection prior to detectable viral shedding (pre-shedders), providing us with a unique opportunity to profile the earliest host immunity upon exposure to naturally circulating SARS-CoV-2 variants at an infection timeline and temporal resolution comparable to that of human challenge studies (**Fig. 2B**).

Elevated expression of immediate early genes and T-cell-associated responses detected in the periphery prior to airway viral shedding.

A total of 132 longitudinal blood samples (10-12 timepoints/participant) from 12 participants (8 early COVID-19+ and 4 exposed uninfected) were subjected to high resolution immune profiling (**Fig. 1D**, **Fig. 2C** and **Fig. S1A**). Pre-shedding and pre-symptomatic phase samples were collected for five of the COVID-19+ participants (**Fig. 2C**). Relative to the first sample collection, the median days PVS was 0 (range: -4 – 2) and the median days PSO was -1 (range: -5 – 3) (**Table 1**). When longitudinal samples from COVID-19+ participants ($n_{samples}$=92; $n_{participants}$=8) were compared to those of unexposed uninfected (healthy) participants ($n_{samples}$=32; $n_{participants}$=5) derived from a similar dataset (16), a total of 206 genes were differentially expressed (**Fig. 3A**). 89 genes were over expressed in COVID-19+ participants including various innate interferon stimulated genes (ISGs) and the known pre-symptomatic marker *IFI27* (13) (**Fig. 3A** and **Fig. 3B**). Notably, we observed a larger number of under expressed genes in this early infection cohort including those encoding the class II major histocompatibility complex (MHC) involved in peptide presentation (*HLA-DQB1*) and the IL-17 family cytokine (*IL17D*) known to suppress CD8 T cell responses via suppression of dendritic cell activity (17) (**Fig. 3B**).

To further assess dynamicity of host immune kinetics throughout the infection period, longitudinal samples from COVID-19+ participants were binned into seven temporal groups based on their relative day PVS (**Fig. 3C**). Smaller bin sizes were assigned to samples collected between -5 – 7 days PVS to achieve greater granularity of transcriptional changes during these early timepoints (**Fig. 3C**). Each temporal bin was compared to healthy participants (**Fig. 3C**). An increase in transcriptional response can be observed as infection progressed from the pre-shedding to acute post-shedding and symptomatic phase of the infection, peaking at 7 days PVS with 242

differentially expressed genes (DEGs) before a significant decrease in transcriptional response during the mid-acute phase (8–21-day PVS), presumably in conjunction with resolution of infection and inflammation (**Fig. 3C**). Another wave of transcriptional response showing significant increase in DEGs can be observed as the host entered the late acute (> 22 days PVS) to post-acute (>28 days PVS) phase of the infection, potentially indicative of immune pivoting from pathogen clearance to restoring immune homeostasis and building immune memory (**Fig. 3C**).

During the pre-shedding phase, increased expression of host immediate-early genes (IEGs) was observed including subunits of the AP-1 transcription factor complex (*FOS*, *JUN, JUNB*) and the cyclooxygenase responsible for prostaglandin biosynthesis (COX-2; *PTGS2*) (**Fig. 3C**). Marked increase in the type II C-lectin receptor (*CD69*), a natural killer (NK) and T lymphocyte early activation marker and a target of the AP-1 transcriptional complex, was also observed (**Fig. 3C**) (18). Genes encoding inflammatory cytokines and chemokines such as the IL-6 cytokine Oncostatin M (*OSM*) and the macrophage inflammatory proteins MIP1-α (*CCL3*/*CCL3L1*/*CCL3L3*) and MIP1-β (*CCL4/CCL4L1/CCR4L2*) were also upregulated prior to viral shedding (**Fig. 3C**). Notably, robust expression of many ISGs known to be hallmarks of host early antiviral response were lacking prior to viral shedding (**Fig. 3C**). Instead, interferon-γ (IFN-γ) and the IFN-γ induced GTPases (*GBP1*, *GBP5*) were enriched in the pre-shedding phase along with cytotoxic cell signatures such as the immunoglobulin-like receptors (*KIR2DL3*, *KIR3DL1*), killer cell lectin-like receptor (*KLRD1*), along with cytotoxic granule proteins such as the pore-forming granulysins (*GNLY*) and perforins (*PRF1*), and multiple death-inducing serine proteases (*GZMA, GZMB, GZMH*) (**Fig. 3C**).) Given that IFN-γ is produced predominantly by cytotoxic cells (e.g., natural killer cells and CD8+ T cells), these results suggest extremely early activation of NK or T cell activity prior to detectable viral RNA.

Host defense peptides potentially confer protective immunity in individuals exposed to SARS-CoV-2.

Evidence of abortive subclinical infections have been observed in individuals exposed to SARS-CoV-2 (19). In human challenge studies, only a subset of experimentally infected participants developed clinical infections, suggesting the presence of early protective immunity in individuals who were exposed but showed no detectable

virus shedding (8, 11). Here, we aimed to assess if evidence of early protective immunity can be observed within the host blood transcriptional response in exposed uninfected individuals. To achieve this, we performed a nested case-control pairwise comparison between exposed infected ($n_{samples}$=92; $n_{participants}$=8) and exposed uninfected ($n_{samples}$=40; $n_{participants}$=4) participants (**Fig. 2C** and **Fig. 4A**). As anticipated, robust signatures of interferon response were observed in infected participants (**Fig. 4A**). 22 genes were over expressed in individuals exposed to ARVIs but remained test-negative that were enriched in neutrophil-associated host defense peptides (HDPs) (*BPI, DEFA4, LCN2*, *LTF,* and *S100A12*) (**Fig. 4A**), reflecting either a transcriptional suppression in infected individuals or transcriptional activation in uninfected individuals. To determine this, transcriptional profiles from both exposed uninfected and infected participants were compared to that of healthy controls. As shown in **Fig. 4B**, all HDPs except for alkaline phosphatase (*ALPL*) were significantly over expressed in exposed uninfected individuals relative to both healthy and infected individuals, suggesting that elevated expression of HDPs in exposed uninfected individuals may play a role in early pathogen clearance and protection against progression to clinical disease. When expression levels of key neutrophil-associated surface receptors and markers (*CEACAM3, CSF3R, FPR1, FCGR3A, SIGLEC5*) were assessed, we did not observe a similar increase in expression in exposed uninfected relative to healthy or infected individuals, suggesting that cell type composition alone cannot account for the elevated levels of HDPs observed in exposed uninfected individuals (**Fig. S2**). In addition to HDPs, both mast cell associated receptors and mediators (*CPA3*, *HDC, MS4A2*) and NK/effector T cell receptors (*KLRB1*, *KLRC1*, *TNFRSF9*) also displayed elevated expression in exposed uninfected participants (**Fig. 4D**). Notably, the CD137 TNF receptor (*TNFRSF9*) preferentially expressed on the surface of T cells activated by antigen presenting cells and known to contribute to clonal expansion and development of CD4+ and CD8+ T cells were overexpressed in exposed uninfected participants relative to both infected and healthy participants (**Fig. 4A, Fig. 4B**, **and Fig. 4D**). Taken together, our results show preliminary evidence of neutrophil-mediated protective immunity in the blood transcriptional signatures of individuals exposed to respiratory viruses including SARS-CoV-2, warranting further evaluation of this cohort to delineate mechanisms underlying host protection against development of clinical infections.

DISCUSSION

Early pre-symptomatic host responses following pathogen exposure can provide valuable insights into immune signatures that confer susceptibility or protection against infection, or against symptoms despite infection. The earliest host transcriptional response to viral infections can be rapid and transient (20-22), making them hard to capture in natural infections when exposure or infection timelines cannot be controlled. In-person venipuncture still presents a logistical barrier to conducting longitudinal studies requiring frequent blood collections. Thus, high frequency transcriptional profiling of the pre-symptomatic host response is largely limited to controlled human challenge studies and there is currently an unmet need for a study framework that enables conduct of such studies in natural infections. In two hallmark studies on the pre-symptomatic phase of natural ARVIs (12, 13). McClain *et al*. performed serial sampling with daily observations of close contacts (college students) exposed to various seasonal respiratory viruses across a 5-day period prior to the COVID-19 pandemic. Gupta *et al*. performed serial sampling with weekly observations of healthcare workers over 16-weeks during the COVID-19 pandemic, but the study did not obtain sufficient timepoints during the pre-symptomatic and early acute phase to provide mechanistic insights into the kinetics of key immune players during early infection.

Here, we present a fully remote study framework that utilized self-collection of both blood and nasal swabs to characterize the dynamicity of a rapidly evolving immune state and report its findings in a pilot cohort aimed at profiling the host pre-symptomatic immune kinetics to natural respiratory infections. During the recruitment window spanning June 2021- April 2022 recruitment window, participants were predominantly exposed to SARS-CoV-2 Delta and Omicron variants. To our knowledge, this is the first prospective longitudinal study to perform high-frequency transcriptional profiling of the host immune kinetics to natural SARS-CoV-2 infections spanning the duration of illness. Incorporation of cross-platform home self-sampling methodologies enabled collection of both localized (airway) and systemic (blood) time-matched samples at pre-determined intervals across the disease continuum (from pre-symptomatic to convalescence) and thus, provided a unique opportunity to track trajectories of innate and acquired immunity at a temporal resolution comparable to that of human challenge studies. As collection of both blood and nasal swab samples were performed by study participants remotely, isolation of

infected participants can be fully maintained during study participation and risks of viral exposure and transmission associated with repeated in-person study visits during their most infectious phase was essentially non-existent.

A key strength of the study lies in the rapid dissemination of collection kits to participants over wide geographical areas, enabling large-scale recruitment across diverse communities throughout the United States (**Fig. 1C**). Given that duration of the incubation period varies significantly between individuals and can be as short as several days, a rapid initiation between enrollment and first sample collection was integral to capturing exposed close contacts (29.4% of total COVID-19+) who were still in their pre-symptomatic phase. A small but interesting subset of participants consists of pre-shedders. As viral RNA can oftentimes be detected prior to onset of symptoms (23), these pre-shedding participants provide us with the unique opportunity to observe the earliest transcriptional changes in the periphery prior to measurable viral load in the airway. Two pre-shedders were observed in our pilot cohort: one progressed to moderate symptoms while another remained asymptomatic throughout the duration of the study.

Detailed temporal trajectories of over 700 innate and acquired immune genes along with temporally aligned viral load profiled for each individual participant led us to several key findings. Notably, robust expression of many ISGs known to drive the early inflammatory response to respiratory viruses were lacking during the pre-shedding phase of infection. Instead, we identified a group of transient early response genes that were induced in both pre-shedders regardless of their symptom outcomes and found that early transcriptional response in the periphery consists of immediate early genes associated with the AP-1 transcriptional complex and prostaglandin biosynthesis, pyrogenic cytokines involved in inflammatory signaling, the type II interferon (IFN-γ) expressed pre-dominantly by T cells, and robust signatures of NK cells and cytolytic T lymphocytes (**Fig. 3C**). The increase in cytolytic T lymphocytes signatures was further supported by under expression of the IL-17D cytokine previously shown to suppress CD8+ T cell activity in infected individuals.

Second, elevated expression of neutrophil associated HDPs were observed in exposed close contacts who remained test negative compared to those who became infected or those who were not exposed (**Fig. 4**). The antimicrobial properties of HDPs that confer protective immunity against various bacterial, fungal, and viral pathogens are multifaceted ranging from indirect immunomodulation of the inflammatory cascades to direct nutritional immunity through metal sequestration. In SARS-CoV-2, defensins have been demonstrated to restrict viral entry in *in vitro* studies (24) and owing to their longstanding reputation as key innate defenders, HDPs were quickly proposed in therapeutic applications against SARS-CoV-2 (25) based upon antiviral efficacies observed from *in vitro* cellular assays. Our findings provided the first clinical evidence for the potential protective role of HDPs against SARS-CoV-2. Aside from HDPs, we also observed elevated expression of *TNFRSF9*, a TNF receptor preferentially expressed on antigen-activated T cells in exposed uninfected participants. Taken together, our data supports a role for T-cell mediated responses in early host transcriptional response to pathogen exposure, both during the pre-shedding/pre-symptomatic phase of infection as well as in exposed uninfected individuals. Further evaluation of this population can offer valuable insights into host immunity that contributes to abortive infections or protection against development of clinical disease (19, 26, 27).

There are several limitations to the study. First, our pilot cohort is limited to a small sample size, resulting in a limited number of pre-symptomatic and pre-shedding participants observed. However, frequent intra-individual measurements led to several key findings that can be further validated in larger studies. Second, unsupervised self-sampling methodologies are inherently susceptible to user-associated variations that can impact sensitive molecular assay readouts such as gene expression and SARS-CoV-2 viral load. This can result in underestimation of SARS-CoV-2 viral load that potentially sets the shedding onset timeline later than it is. Third, a significant subset of close contacts reported onset of symptoms by their first sample collection despite being asymptomatic the day prior. These participants, although not truly pre-symptomatic, offer a unique window into the prodromal phase and still enable us to study early immune events in addition to key mediators during the transition between pre-symptomatic to onset of illness. Lastly, the use of a small target response panel in addition to lack of evaluation of the localized host response in the airway limits the breath of the biology observed in this pilot cohort. Therefore, future studies

using genome-wide transcriptome coupled with evaluation of host mucosal response will enable us to correlate systemic responses to that in the mucosa and provide a comprehensive understanding of the host immunity soon after pathogen exposure. In sum, we successfully demonstrated that decentralized sampling using home self-collection tools can capture early, transient changes in host response soon after viral exposure in natural infections at high temporal resolution comparable to human challenge studies. This study framework will enable conduct of large-scale population-wide mechanistic studies requiring detailed immune kinetics during rapidly evolving immune states.

METHODS AND MATERIALS

Study design and participant recruitment

The study is aimed at profiling the pre-symptomatic host transcriptional response to ARVIs with a focus on SARS-CoV-2 in non-hospitalized adults using *home*RNA self-collected capillary blood samples. To achieve this, adult individuals (close contacts) who were recently in close contact with someone with a confirmed (SARS-CoV-2 nucleic acid or antigen test) or suspected (symptomatic) COVID-19 infection but were asymptomatic at the time of screening were recruited from the general population. Close contacts were recruited directly via self-reported exposure and indirectly via snowball sampling method (index case referral). Specifically, eligible index cases (COVID-19+ symptomatic individuals) who provided informed consent were invited to refer their adult close contacts who gave permission. Eligible close contacts who provided informed consent were enrolled into the study for serial sampling of both nasal and blood samples as described below. In addition, adults who self-reported a recent COVID-19+ test result but remained asymptomatic at the time of screening were also invited to participate in the study. Eligible individuals were prospectively enrolled until at least 15 participants with confirmed ARVIs were reached.

Participants were recruited via social media and news platforms across the United States, the Institute of Translational Health Sciences (ITHS) Participate-in-Research program, and the University of Washington Husky Coronavirus Testing Study Sites. Following screening and informed consent, a study package containing ten

*home*RNA blood collection kits, seven nasal swab kits, prepaid return mailers, additional supplies, and an informational packet with instructions for use, a collection calendar with pre-determined collection dates, and specimen return instructions were mailed to the participants. Participants initiated sampling procedures (Day 1) the day of receipt (median 2 days post screening). All participants collected both blood and nasal swab samples daily for 7 days (study days 1-7) followed by just blood samples weekly for 3 additional weeks (study days 14, 21, 28). A select number of study-confirmed COVID-19+ participants were invited to participate in an extended sampling phase for up to 4 additional weeks (study days 35, 42, 48, 56) (**Fig. 1**). In addition, device use and symptom burden surveys were distributed for each sampling day. The first nasal swab from each participant was clinically tested for SARS-CoV-2 while a select number of the remaining nasal swabs (based on reported symptoms) were subjected to pathogen screening through a lab-developed multi-pathogen array assay for presence of SARS-CoV-2 and other common respiratory viral and bacterial pathogens (28). For close contacts, clinical SARS-CoV-2 results from their first nasal swab were returned to them while clinical results for participants with known COVID-19+ status at the time of enrollment were not returned to the participants in order to mitigate possibility of a negative result from the study interfering with their standard of care and isolation procedures.

Setting

This study was conducted at the University of Washington, Seattle WA and approved by the UW IRB under protocol STUDY00012546. All participants provided informed consent. All sample collections were performed remotely by study participants and online surveys were administered through REDCap (29). Participants were recruited to the study between June 2021- April 2022.

***home*RNA 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 *home*RNA blood collection kit has been previously described (28).

Blood collection: Blood collection was performed at the residential homes or other temporary locations on pre-determined sampling days (if traveling during study). Detailed description of the *home*RNA collection procedures has been previously published (14). 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 connect the blood collection tube to the stabilizer tube containing RNAlater and mix thoroughly to complete stabilization (**Fig. 1B**). Stabilized blood was packaged according to the return instructions 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 -20ºC freezer. When collection was scheduled on a weekend, participants were asked to store stabilized samples in a cool location prior to the first available courier pick up service the following Monday. Samples were transferred to the -80ºC freezer for long term storage until further 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 immediately transfer swabs into universal transport medium (Copan Diagnostics) immediately after collection. The first nasal swab (NS1) was returned to a clinical 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. Similar to *home*RNA-stabilized blood, nasal swabs were delivered directly to a secure -20ºC freezer and transferred to a -80ºC freezer 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 described (28). For each participant, 3-4 nasal swabs were tested for SARS-CoV-2 and 18 common respiratory pathogens as previously described (28). For participants with positive SARS-CoV-2 results, one representative swab with the lowest Crt value (highest viral load) was selected for SARS-CoV-2 sequencing as previously described (28).

Online surveys

All surveys were administered online via REDCap (29). A Welcome Survey was administered immediately upon enrollment. 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 co-morbidities. 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, prior to study exit.

Comparator group assignment

We assigned participants into five major comparator groups based on their self-reported COVID-19 infection status at enrollment and viral detection status from both clinical SARS-CoV-2 testing and multi-pathogen OpenArray assay as follows: 1) participants with positive SARS-CoV-2 in at least one tested swab sample were assigned to the "COVID-19+" group; 2) participants with negative SARS-CoV-2 in all tested swab samples and no evidence of other respiratory pathogens tested were assigned to the "Exposed Uninfected" group; 3) participants with negative SARS-CoV-2 in all tested swab samples and but showed evidence of other respiratory pathogens were assigned to the "Other virus" group; 4) participants with negative SARS-CoV-2 in all tested swab samples and self-reported a recent SARS-CoV-2 positive test were assigned to the "COVID-19 probable" group; and 5) a subset of COVID-19 probable participants whose self-reported test date exceeded 14 days on their first collection date were assigned as "COVID-19 convalescent".

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 ºC 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 small RNA-enriched 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 host 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 described (28). 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. Selection of reference genes was performed using

ROSALIND® (https://rosalind.bio/) based on manufacturer's recommendation. Five reference genes (*GUSB, HRPT1, MRPS7, NMT1, PGK1*) were used in codeset content normalization.

Differential gene expression analysis and data visualization: Differential expression analyses were performed on ROSALIND® (https://rosalind.bio/). For time-series analysis, COVID-19+ participants were divided into seven temporal groups based on the date from their first collection to their respective onset of viral shedding. Pairwise comparisons were performed for each temporal group using all longitudinal samples from the healthy participants as a reference group. P-values were adjusted for multiple comparisons by controlling the false discovery rate (FDR) using the Benjamini-Hochberg (BH) procedure (30). Absolute $\log_2$ fold change of >0.322 corresponding to a 1.25x fold change and an adjusted p-value of < 0.1 were used to identify differentially expressed transcripts. Volcano plots were generated in R Statistical Software (v4.2.1, R Core Team 2022) (31). Boxplots were generated in R Statistical Software (v4.2.1, R Core Team 2022) and statistical analyses performed using the ggpubr R package. Study workflow and conceptual diagrams were constructed in Biorender and construction of final figures were done in Adobe illustrator.

References and Notes

1. Ong EZ, et al. Temporal dynamics of the host molecular responses underlying severe COVID-19 progression and disease resolution. *EBioMedicine.* 2021;65:103262.
2. Brodin P. Immune determinants of COVID-19 disease presentation and severity. *Nat Med.* 2021;27(1):28-33.
3. Chua RL, et al. COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat Biotechnol.* 2020;38(8):970-9.
4. Le Bert N, et al. Highly functional virus-specific cellular immune response in asymptomatic SARS-CoV-2 infection. *J Exp Med.* 2021;218(5).
5. Wright PF, et al. Longitudinal Systemic and Mucosal Immune Responses to SARS-CoV-2 Infection. *J Infect Dis.* 2022;226(7):1204-14.
6. Notarbartolo S, et al. Integrated longitudinal immunophenotypic, transcriptional and repertoire analyses delineate immune responses in COVID-19 patients. *Sci Immunol.* 2021;6(62).

7. Zaas AK, et al. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. *Cell Host Microbe.* 2009;6(3):207-17.

8. Woods CW, et al. A host transcriptional signature for presymptomatic detection of infection in humans exposed to influenza H1N1 or H3N2. *PLoS One.* 2013;8(1):e52198.

9. Huang Y, et al. Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza a infection. *PLoS Genet.* 2011;7(8):e1002234.

10. Choy RKM, et al. Controlled Human Infection Models To Accelerate Vaccine Development. *Clin Microbiol Rev.* 2022;35(3):e0000821.

11. Killingley B, et al. Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. *Nat Med.* 2022;28(5):1031-41.

12. McClain MT, et al. A blood-based host gene expression assay for early detection of respiratory viral infection: an index-cluster prospective cohort study. *Lancet Infect Dis.* 2021;21(3):396-404.

13. Gupta RK, et al. Blood transcriptional biomarkers of acute viral infection for detection of pre-symptomatic SARS-CoV-2 infection: a nested, case-control diagnostic accuracy study. *Lancet Microbe.* 2021;2(10):e508-e17.

14. Haack AJ, et al. homeRNA: A Self-Sampling Kit for the Collection of Peripheral Blood and Stabilization of RNA. *Anal Chem.* 2021;93(39):13196-203.

15. Brown LG, et al. At-home blood collection and stabilization in high temperature climates using homeRNA. *Front Digit Health.* 2022;4:903153.

16. Lim FY, et al. Longitudinal home self-collection of capillary blood using homeRNA correlates interferon and innate viral defense pathways with SARS-CoV-2 viral clearance. *medRxiv.* 2023:2023.01.24.23284913.

17. Lee Y, et al. Interleukin-17D Promotes Pathogenicity During Infection by Suppressing CD8 T Cell Activity. *Front Immunol.* 2019;10:1172.

18. Castellanos MC, et al. Expression of the leukocyte early activation antigen CD69 is regulated by the transcription factor AP-1. *J Immunol.* 1997;159(11):5463-73.

19. Swadling L, et al. Pre-existing polymerase-specific T cells expand in abortive seronegative SARS-CoV-2. *Nature.* 2022;601(7891):110-7.

20. Marcinowski L, et al. Real-time transcriptional profiling of cellular and viral gene expression during lytic cytomegalovirus infection. *PLoS Pathog.* 2012;8(9):e1002908.

21. Zhu H, et al. Inhibition of cyclooxygenase 2 blocks human cytomegalovirus replication. *Proc Natl Acad Sci U S A.* 2002;99(6):3932-7.

22. Cheemarla NR, et al. Dynamic innate immune response determines susceptibility to SARS-CoV-2 infection and early replication kinetics. *J Exp Med.* 2021;218(8).

23. Puhach O, et al. SARS-CoV-2 viral load and shedding kinetics. *Nat Rev Microbiol.* 2023;21(3):147-61.

24. Xu C, et al. Human Defensins Inhibit SARS-CoV-2 Infection by Blocking Viral Entry. *Viruses.* 2021;13(7).

25. Mabrouk DM. Antimicrobial peptides: features, applications and the potential use against covid-19. *Mol Biol Rep.* 2022;49(10):10039-50.

26. Andreakos E, et al. A global effort to dissect the human genetic basis of resistance to SARS-CoV-2 infection. *Nat Immunol.* 2022;23(2):159-64.

27. Pancino G, et al. Natural resistance to HIV infection: lessons learned from HIV-exposed uninfected individuals. *J Infect Dis.* 2010;202 Suppl 3:S345-50.

28. Lim FY, et al. Longitudinal home self-collection of capillary blood using homeRNA correlates interferon and innate viral defense pathways with SARS-CoV-2 viral clearance. *medRxiv.* 2023:2023.01.24.23284913.

29. Harris PA, et al. 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.

30. Benjamini Y, and Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological).* 1995;57(1):289-300.

31. R Development Core Team. Vienna, Austria: R Foundation for Statistical Computing; 2022.

ACKNOWLEDGEMENTS

Funding:

- R35GM128648 (ABT, for in-lab developments of *home*RNA)
- Packard Research Fellowship from the David and Lucile Packard Foundation (ABT)
- R01AI153087 (AW)

Author contributions:

Conceptualization: FYL, KA, ABT, EB, AW

Methodology: FYL, HGL, AD, LS

Investigation: FYL, HGL, AD, LS

Analysis and visualization: FYL, HGL, TVN

Funding acquisition: ABT, AW

Project administration: FYL, GH, AD, MT, KA, HGL, ABT

Supervision: ABT, EB, AW, FYL

Writing – original draft: FYL, HGL, ABT, EB

Writing – review & editing: FYL, HGL, TVN, AW, JTS, MB, LS, ABT, EB

Declaration of Interest:

FYL, EB, and ABT filed patent 17/361,322 (Publication Number: US20210402406A1) through the University of Washington on *home*RNA. ABT reports filing multiple patents through the University of Washington and receiving a gift to support research outside the submitted work from Ionis Pharmaceuticals. EB has ownership in Salus Discovery, LLC, and Tasso, Inc. that develops blood collection systems used in this manuscript, and is employed by Tasso, Inc. Technologies from Salus Discovery, LLC are not included in this publication. He is an inventor on multiple patents filed by Tasso, Inc., the University of Washington, and the University of Wisconsin. The terms of

this arrangement have been reviewed and approved by the University of Washington in accordance with its policies governing outside work and financial conflicts of interest in research. AW reports receiving clinical trial support to their institution from Pfizer, Ansun Biopharma, Allovir, and GlaxoSmithKline/Vir; receiving personal fees from Kyorin Pharmaceuticals; and receiving grants from Amazon outside the submitted work. MB has clinical research support from Ansun Biopharma, Amazon, GSK, Vir Biotechnology, and Merck and receives personal fees from Allovir, Moderna, and Merck. JTS has clinical trial support from Aicuris and receives personal fees from Glaxo Smith Kline and Pfizer.

Data and material availability:

All data are available in the main text or the supplementary materials. nCounter gene expression data available as normalized counts in data file S1. All statistical analyses were conducted using R and Rosalind Bio. Packages used to perform analyses are specified in the statistical section. R scripts developed to perform analyses and visualizations are available upon request.

FIGURES

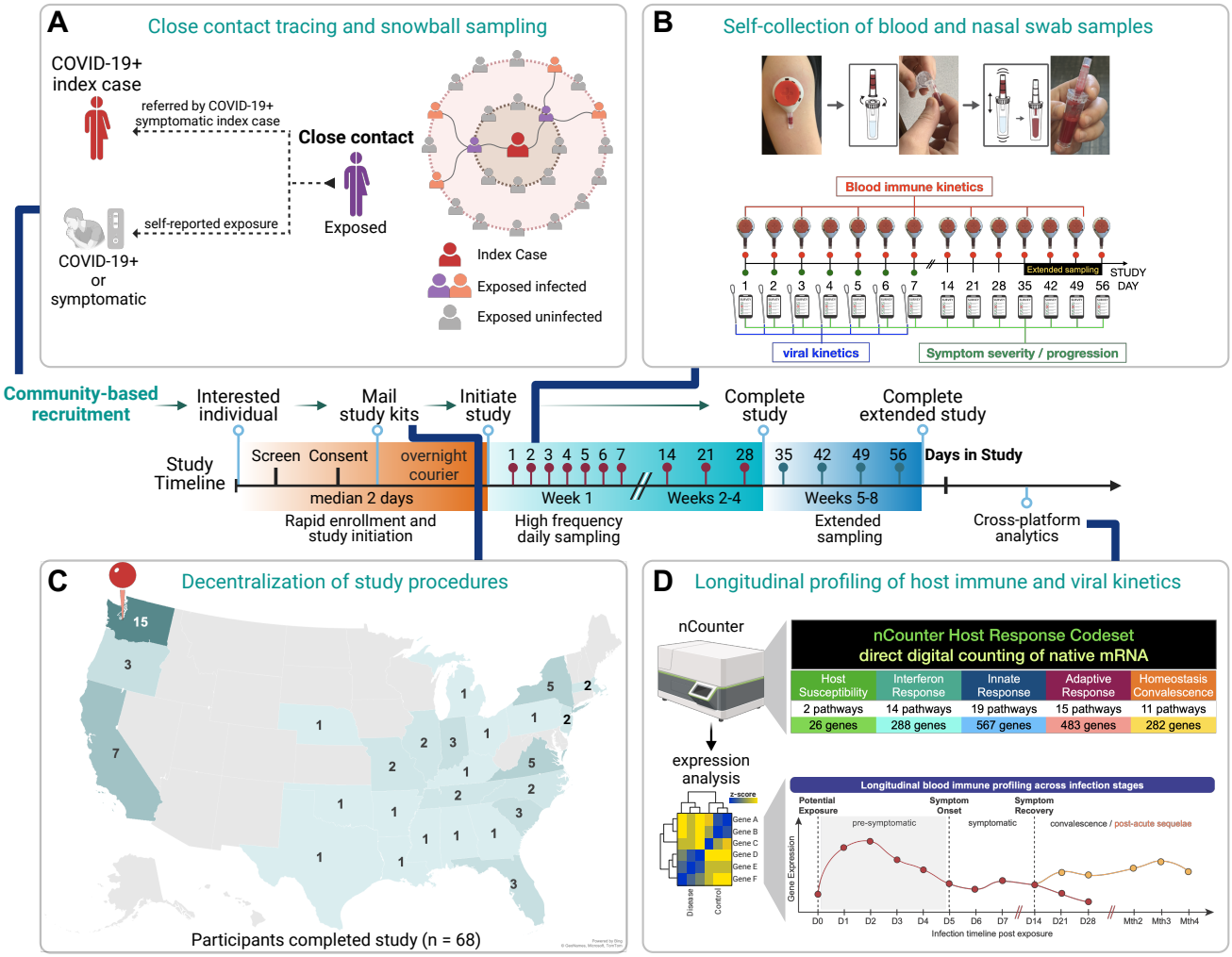


Figure 1. Study framework to capture the pre-symptomatic host transcriptional response to naturally acquired SARS-CoV-2 infection. Conceptual framework depicting the study design and participation timeline (middle), A) close contact recruitment strategies, B) sample collection details, C) geographical representation of study participants, and D) longitudinal gene expression analyses pipeline.

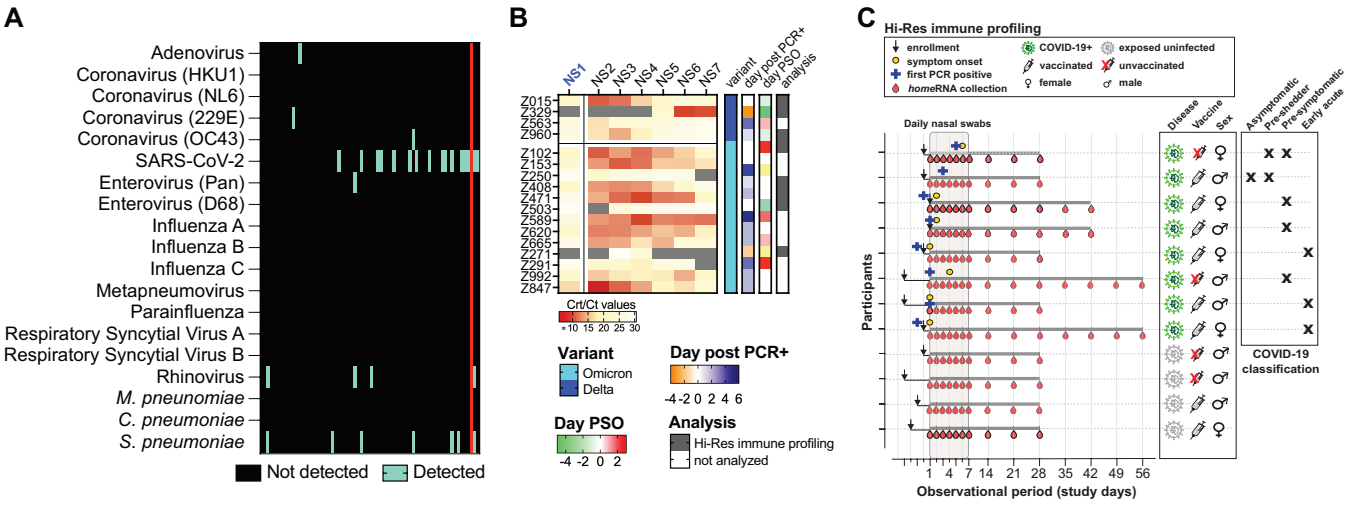


Figure 2. SARS-CoV-2 infection rate and classification of participants. A) Heatmap depicting pathogens detected in all 68 participants. Pathogens are represented in rows while participants are represented in columns. 3-4 nasal swabs were assessed for respiratory pathogens including SARS-CoV-2. Red line denotes one participant with a single nasal swab assessed. **B)** Heatmap depicting SARS-CoV-2 viral kinetics in 17 COVID-19+ participants. Grey cells denote participants samples with undetectable viral load. Rows represent participants while columns represent nasal swab samples. **C)** Clinical and sample characteristics of 12 participants selected for high-resolution immune profiling.

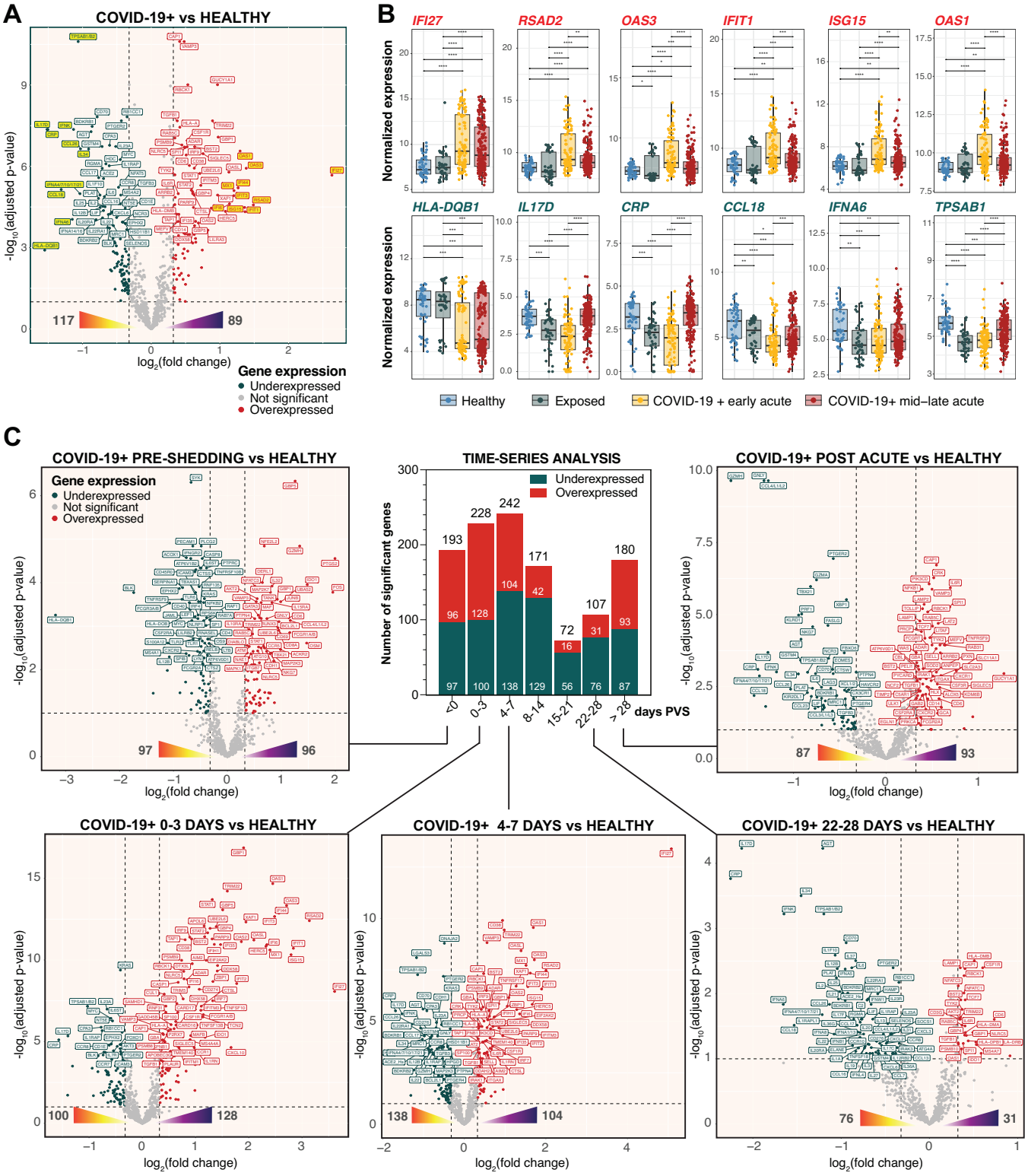
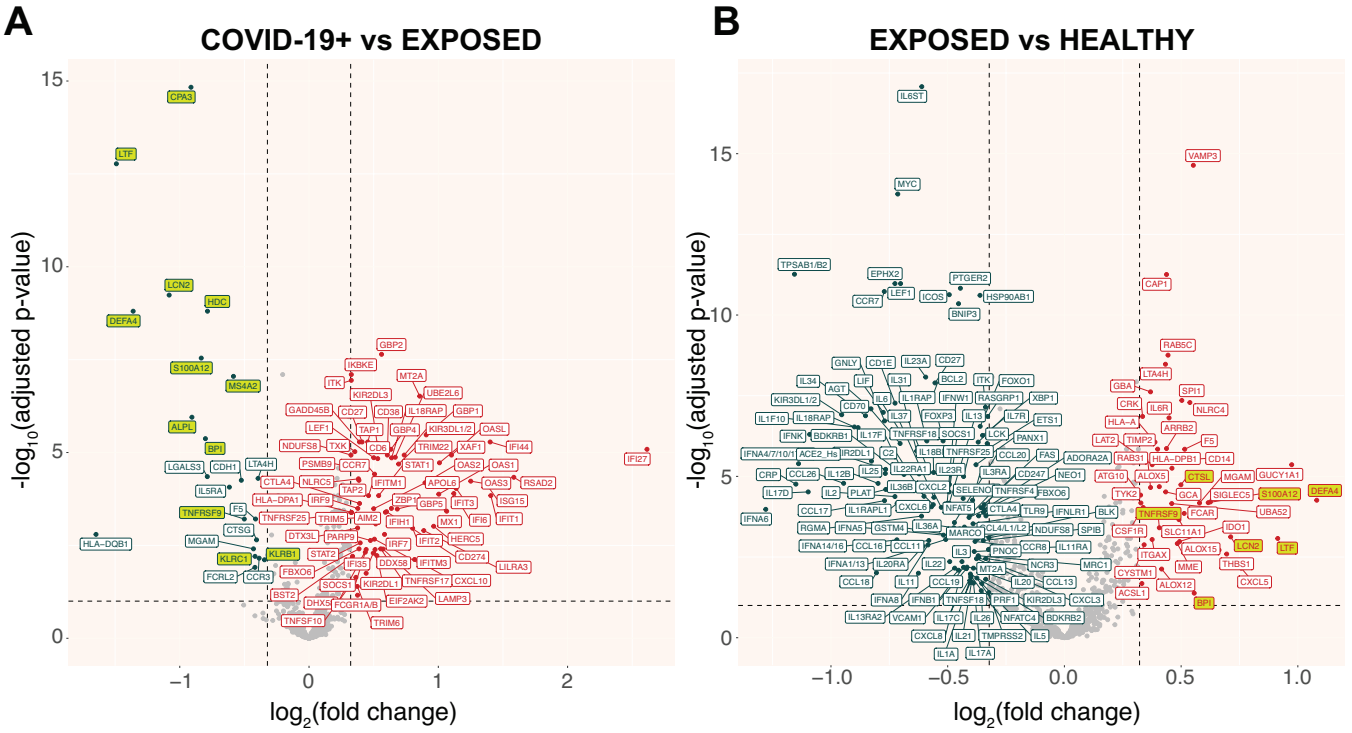

A
COVID-19+ vs HEALTHY
-log10(adjusted p-value)
log2(fold change)
117
89
Gene expression
Underexpressed
Not significant
Overexpressed
B
IFI27
RSAD2
OAS3
IFIT1
ISG15
OAS1
HLA-DQB1
IL17D
CRP
CCL18
IFNA6
TPSAB1
Normalized expression
Healthy
Exposed
COVID-19 + early acute
COVID-19+ mid–late acute
C
COVID-19+ PRE-SHEDDING vs HEALTHY
97
96
TIME-SERIES ANALYSIS
Number of significant genes
Underexpressed
Overexpressed
193
228
242
171
72
107
180
96
128
104
42
16
31
93
97
100
138
129
56
76
87
<0
0-3
4-7
8-14
15-21
22-28
> 28
days PVS
COVID-19+ POST ACUTE vs HEALTHY
87
93
COVID-19+ 0-3 DAYS vs HEALTHY
100
128
COVID-19+ 4-7 DAYS vs HEALTHY
138
104
COVID-19+ 22-28 DAYS vs HEALTHY
76
31

Figure 3. Elevated expression of immediate early genes (IEGs) and T-cell responses observed in the periphery prior to viral shedding. A) Volcano plot depicts significant genes from pairwise comparisons of all longitudinal samples obtained from COVID-19+ vs healthy participants. Significantly overexpressed and underexpressed genes are colored in red and green respectively. Top 100 significant genes based on adjusted p-value are labeled on the plot and top 10 differentially expressed genes based on log2(fold change) highlighted in yellow. **B)** Box plots depicting the log2 normalized expression of top 6 over expressed and under expressed genes stratified by unexposed uninfected (healthy), exposed uninfected, COVID-19+ early, and COVID-19+ mid-late acute participants from (16). 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), $p \leq 0.05$ (*), $p \leq 0.01$ (**), $p \leq 0.001$ (***), $p \leq 0.0001$ (****). **C)** Time series analysis of early COVID-19+ host transcriptional response. Volcano plots showing DEGs for pairwise comparisons between pre-shedding (<0 days PVS), 0-3 days PVS, 4-7 days PVS, 22-28 days PVS, and post-acute (>28 days PVS) temporal groups and healthy participants.


A
COVID-19+ vs EXPOSED
B
EXPOSED vs HEALTHY
-log10(adjusted p-value)
log2(fold change)

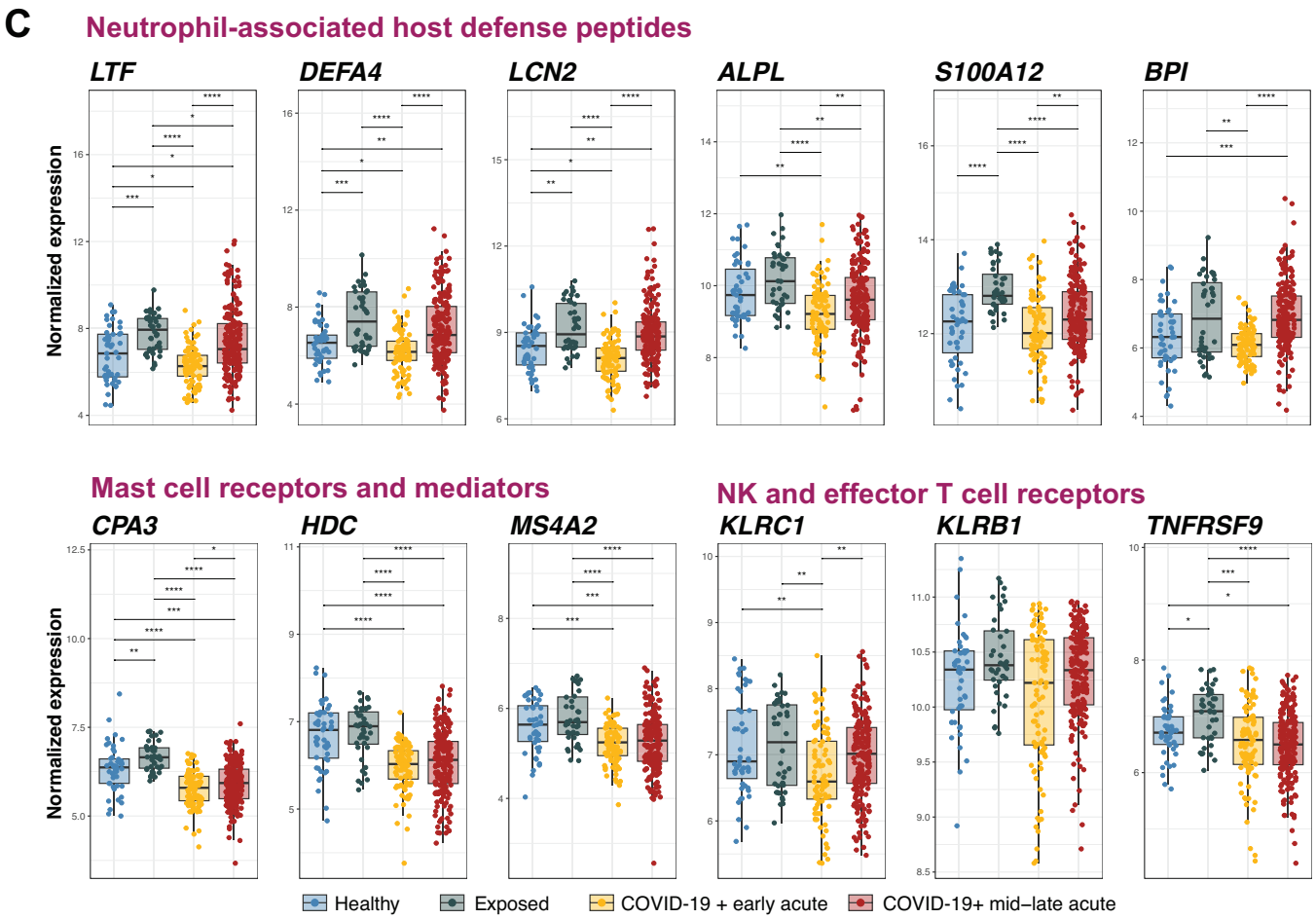

C
Neutrophil-associated host defense peptides
LTF
DEFA4
LCN2
ALPL
S100A12
BPI
Mast cell receptors and mediators
CPA3
HDC
MS4A2
NK and effector T cell receptors
KLRC1
KLRB1
TNFRSF9
Normalized expression
Healthy
Exposed
COVID-19 + early acute
COVID-19+ mid–late acute

Figure 4. Increased expression of host defense peptides in exposed uninfected adults. Volcano plots depict significant genes from pairwise comparisons of **A)** COVID-19+ versus exposed uninfected and **B)** exposed uninfected versus 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, mast cell markers, and T cell markers are highlighted in yellow. **C)** Box plots depicting the log2 normalized expression of genes for all longitudinal samples stratified by unexposed uninfected (healthy), exposed uninfected, COVID-19+ early, and COVID-19+ mid-late acute participants from (16). 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), $p \leq 0.05$ (*), $p \leq 0.01$ (**), $p \leq 0.001$ (***), $p \leq 0.0001$ (****).

TABLES

Table 1. Demographics and clinical characteristics of participants who completed the study.

	All participants (n=68)	Hi-Res immune profiling (n=12)	
		COVID-19+ (n=8)	Exposed Uninfected (n=4)
Demographics			
Sex, (%)			
Female	70.60%	50%	25%
Male	29.40%	50%	75%
Age, (years), Median IQR	38 (28-47.75)	38(28.5-54)	40 (38-45)
Ethnicity, (%)			
Hispanic	23.50%	12.50%	25%
Non-hispanic	76.50%	87.50%	75%
Race, (%)			
Asian	7.30%	0%	0%
Black/African American	5.90%	12.50%	0%
White	70.60%	87.50%	75%
Two or more races	11.80%	0%	25%
Unknown	4.40%	0%	0%
Clinical characteristics			
Weight (lbs)			
Median (IQR)	160 (141.3-188.8)	160 (147.8-178.3)	210 (183.8-228.8)
Height (inches)			
Median (IQR)	66 (64-69)	68.5 (66.25-70.75)	70.50 (68.50-71)
COVID-19 vaccination, (%)			
Unvaccinated	67.6%	25%	50%
Vaccinated	32.4%	75%	50%
Infection timeline			
Days, Median (IQR, range)			
post viral shedding	1(0-2, -4-6)	0(-1.5-1.0, -4-2)	na
post symptom onset	0(-1-1, -5-3)	-1(-3-0, -5-3)	na