*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 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.

*Co-senior authors:

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

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

Alpana Waghmare (awaghmare@fredhutch.org)

SUMMARY

Background

Early host immunity to acute respiratory infections (ARIs) is heterogenous, dynamic, and critical to an individual's infection outcome. Due to limitations in sampling frequency/timepoints, kinetics of early immune dynamics in natural human infections remain poorly understood. In this nationwide prospective cohort study, we leveraged a self-blood collection tool (*home*RNA) to profile detailed kinetics of the pre-symptomatic to convalescence host immunity to contemporaneous respiratory pathogens.

Methods

We enrolled non-symptomatic adults with recent exposure to ARIs who subsequently tested negative (exposed-uninfected) or positive for respiratory pathogens. Participants self-collected blood and nasal swabs daily for seven consecutive days followed by weekly blood collection for up to seven additional weeks. Symptom burden was assessed during each collection. Nasal swabs were tested for SARS-CoV-2 and common respiratory pathogens. 92 longitudinal blood samples spanning the pre-shedding to post-acute phase of eight SARS-CoV-2-infected participants and 40 interval-matched samples from four exposed-uninfected participants were subjected to high-frequency longitudinal profiling of 773 host immune genes.

Findings

Between June 2021 – April 2022, 68 participants across 26 U.S. states completed the study and self-collected a total of 691 and 466 longitudinal blood and nasal swab samples along with 688 symptom surveys. SARS-CoV-2 was detected in 17 out of 22 individuals with study-confirmed respiratory infection. With rapid dissemination of home self-collection kits, two and four COVID-19+ participants started collection prior to viral shedding and symptom onset, respectively, enabling us to profile detailed expression kinetics of the earliest blood transcriptional response to contemporaneous variants of concern. In pre-shedding samples, we observed transient but robust expression of T-cell response signatures, transcription factor complexes, prostaglandin biosynthesis genes, pyrogenic cytokines, and cytotoxic granule genes. This is followed by a rapid induction of many interferon-stimulated genes (ISGs), concurrent to onset of viral shedding and increase in nasal viral load. Finally, we observed increased expression of host defense peptides (HDPs) in exposed-uninfected individuals over the 4-week observational window.

Interpretation

We demonstrated that unsupervised self-collection and stabilization of capillary blood can be applied to natural infection studies to characterize detailed early host immune kinetics at a temporal resolution comparable to that of human challenge studies. The remote (decentralized) study framework enables conduct of large-scale population-wide longitudinal mechanistic studies. Expression of cytotoxic/T-cell signatures in pre-shedding samples preceding expansion of innate ISGs suggests a potential role for T-cell mediated pathogen control during early infection. Elevated expression of HDPs in exposed-uninfected individuals warrants further validation studies to assess their potential role in protective immunity during pathogen exposure.

Funding

This study was funded by R35GM128648 to ABT for in-lab developments of *home*RNA, Packard Fellowship from the David and Lucile Packard Foundation to ABT, and R01AI153087 to AW.

INTRODUCTION

Host innate and acquired immunity is critical to the outcomes of ARIs[1]. Systemic dysregulation of pro-inflammatory molecules are associated with ongoing inflammation and poor outcomes, while effective coordination of early responses limits the severity and duration of symptoms[2]. However, detailed kinetics of early host responses soon after pathogen exposure prior to symptom onset or viral shedding are poorly understood, including those conferring protective immunity in exposed individuals who remained asymptomatic or test negative. As in-person venipuncture presents a significant logistical barrier to longitudinal studies requiring early and frequent samples, prior studies of natural ARIs were largely cross-sectional, were primarily focused on post-symptomatic infections, typically when nasal viral load were already in decline, and were often limited to infrequent sampling across inconsistent collection intervals[3-6].

High-frequency longitudinal profiling of host peripheral immunity, particularly during extremely early pre-symptomatic phases, has largely been conducted in the setting of human challenge studies[7-9]. Controlled human infection models pose limitations to understanding the host response during natural infections[10,11]. 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), thus limiting their broad applicability to contemporaneously circulating variants and to the heterogeneity of population-wide immunologic backgrounds that are shaped by individual susceptibilities, and by prior vaccination and infection[11]. Studies of the host blood transcriptional response during pre-symptomatic phase of natural infections are very limited[12-14] — with the first of such studies published by McClain *et al*. in 2021 that measured the pre-symptomatic host response to pre-COVID-19 respiratory viruses through daily monitoring of close contacts residing in college dormitories[12]. During the COVID-19 pandemic, Gupta *et al*, in 2021 measured the pre-symptomatic host response to SARS-CoV-2 through weekly surveillance of healthcare workers[13]. The first results of a SARS-CoV-2 human challenge study published in March 2022 did not assess the blood transcriptional response to the infection[15]. Based on this study, detectable nasal viral load (viral shedding) occurred between 40-76 hours post challenge followed shortly by onset of symptoms, making the pre-symptomatic phase of SARS-CoV-2 infection from natural exposures particularly hard to capture and characterize[15]. Thus, high-frequency daily measurements of the early host response to natural SARS-CoV-2 infection are unprecedented prior to this study.

A study framework that enables frequent longitudinal measurements, in a natural field setting, of the early host immune response to respiratory infections acquired through community transmissions can be integrated into the study design of future human challenge trials involving natural exposures to pathogens. For vaccine development, challenge by natural exposure to infected person has been proposed as future designs of human challenge trials to circumvent dose-escalation studies, and address limitations surrounding strain selection, diversity, and delays in manufacturing timelines[16]. A rapid dissemination of self-collection kits to recently exposed individuals across a wide geographic area (decentralized sampling) coupled with frequent self-collections at pre-scheduled infection timepoints and immediate on-site stabilization of the transcriptome is integral to conducting population-wide longitudinal mechanistic studies aimed at profiling detailed host immune kinetics during early and highly dynamic stages of natural infections (**Figure 1**). Our team developed *home*RNA that enables self-collection and stabilization of blood[17,18]; we recently demonstrated its capability of capturing SARS-CoV-2 infection signatures in non-hospitalized adults and discriminating distinct infection response between previously vaccinated and unvaccinated individuals[19]. Here, we leveraged *home*RNA to profile, at high temporal resolution, the pre-symptomatic immune kinetics during natural ARIs in recently exposed adults recruited across the United States. To the best of our knowledge, this is the first prospective longitudinal nationwide study to assess the pre-symptomatic blood transcriptional response to a natural respiratory infection at a temporal resolution comparable to that of human challenge studies.

METHODS

Study design

This is a prospective longitudinal study aimed at profiling detailed kinetics of extremely early (pre-symptomatic) host immunity to respiratory pathogens with a focus on SARS-CoV-2 in non-hospitalized adults using self-blood (*home*RNA) and nasal swab collection methodologies. Between June 2021 – April 2022, adult close contacts of individuals with a confirmed (SARS-CoV-2 nucleic acid or antigen test) or suspected (symptomatic) COVID-19 infection who were asymptomatic at the time of screening were recruited from the general population throughout the United States. Each participant self-collected 10-14 blood and seven nasal swab samples over 4-8 weeks. Symptom burden was assessed on each collection day. The study was approved by the University of Washington (UW) IRB (STUDY00012546) and conducted at the UW. All participants provided informed consent. All samples were collected remotely by study participants and online surveys were administered through REDCap[20].

Participants

Participants were recruited via social media (Facebook) and news platforms (WATE-TV), the Institute of Translational Health Sciences Participate-in-Research program, and the UW Husky Coronavirus Testing Study Sites. Adult close contacts were recruited directly via self-reported exposure and via referral from an index case (snowball sampling) (**Figure 1A**). Specifically, index cases (COVID-19+ symptomatic individuals) were enrolled and asked 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 blood and nasal swab samples as described below. In addition, individuals who recently tested positive for SARS-CoV-2 via self-report but remained asymptomatic at the time of screening were enrolled into study. Participants were prospectively enrolled in two waves (June - October 2021; January – April 2022) until at least 15 participants with confirmed ARIs were reached. Longitudinal immune profiles of unexposed-healthy participants were obtained from a previously published study that utilized the same *home*RNA collection and analysis platform[19].

Procedures

A study package containing *home*RNA blood and nasal swab collection kits, prepaid return mailers, an informational packet with all instructions for use, a collection calendar with pre-scheduled collection dates, and specimen return instructions were mailed to study participants, who performed their first collection the day of package receipt. All participants collected blood and nasal swab samples daily for seven consecutive days (study days 1-7) followed by weekly blood collection for three additional weeks (study days 14, 21, 28) (**Figure 1B**). A subset of study-confirmed COVID-19+ participants were selected to extend blood collections for up to four additional weeks (study days 35, 42, 48, 56). Device use and symptom burden surveys were administered on each collection day. The first nasal swab from each participant were clinically tested for SARS-CoV-2 and test results were returned to close contacts with unknown infection status at the time of enrollment. 3-4 additional nasal swabs were selected based on symptom burden at the day of collection for multi-pathogen screening through a lab-developed assay (**Table S1**). Detailed immune kinetics of 773 innate and acquired immune genes (nCounter® Host Response Panel) were profiled using nCounter® direct digital counting of native RNA (**Figure 1D**)[19]. Detailed methods are described in the supplemental methods. Participants were classified based on their self-reported COVID-19 infection status at enrollment and virology status of collected nasal swabs (**Figure S1A**): 1) participants positive for SARS-CoV-2 were classified as "COVID-19+"; 2) participants negative for SARS-CoV-2 but positive for other respiratory pathogens were classified as "other virus"; 3) participants negative for all respiratory pathogens were classified as "exposed-uninfected"; 4) participants negative for SARS-CoV-2 on research samples but self-reported a recent SARS-CoV-2 positive test were classified as "COVID-19 probable". COVID-19 probable participants who initiated their first collection > 14 days after their most recent positive test date were classified as "COVID-19 convalescent".

Statistical analysis

Normalized expression counts are provided in supplemental data 1. Differential gene expression analyses across all pairwise comparisons were performed on ROSALIND® (https://rosalind.bio/). To assess temporal dynamics of host immunity throughout the course of infection, COVID-19+ participant samples were divided into seven temporal groups (G1-G7) relative to the onset of nasal viral shedding (**Table S2**). Pairwise comparisons were performed between each temporal group and all longitudinal samples from unexposed-healthy participants. To identify potential signatures of protective immunity to SARS-CoV-2, immune profiles from exposed-uninfected individuals were compared to both unexposed-healthy and exposed-infected individuals. For all pairwise comparisons, p-values were adjusted for multiple comparisons of all measured transcripts using the Benjamini-Hochberg procedure[21]. An absolute $\log_2$ fold change of > 0.322 (1.25x) and an adjusted p-value of < 0.1 were used to identify differentially expressed transcripts. Significant genes identified from differentially expressed transcripts of all pairwise comparisons and their respective $\log_2$ fold change over the reference group, raw p-values, and adjusted p-values are listed in supplemental data 2. Detailed methods on RNA isolation, gene expression analysis, normalization procedures, and functional enrichment analysis of differentially expressed transcripts are described in supplemental methods.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

RESULTS

Between June 2021 – April 2022, 90 non-symptomatic adults with recent exposure to ARIs were enrolled for high-frequency serial sampling (**Figure 2**). 68 participants across 26 U.S. states completed the study (**Figure 1C**). 691/708 (97.6%)

*home*RNA blood samples, 466/476 (97.9%) nasal swabs, and 688/708 (97.2%) online surveys were completed and returned (**Figure 2**). 22 participants were withdrawn; 59% (13/22) did not initiate sampling procedures (**Figure 2**). The median age (n=68) was 38 (IQR: 28.0 - 47.8; range: 20-74) (**Table 1**). The study enrolled participants from diverse demographics including individuals from understudied, underrepresented, and underreported (U3) populations; 70.6% female, 23.5% Hispanic, and a combined 25% of U3 racial minorities (7.3% Asian; 5.9% Black, 11.8% two or more races) participants completed the study (**Table 1**).

362 swabs from 68 participants were analyzed for respiratory pathogens (**Figure 2; Table S1**). 32.4% (22/68) of participants were positive for respiratory viruses (**Figure 2**). SARS-CoV-2 infection was confirmed in 77.3% (17/22) of infected participants with a secondary attack rate of 11.9% (5/42) observed among close contacts (**Figure 3A**; **Figure S1A**). 16.2% (11/68) of participants reported recent SARS-CoV-2 infection but showed no virological evidence of infection on study swabs (COVID-19 Probable) (**Figure S1A**). 32 exposed close contacts remained test-negative (exposed-uninfected) (**Figure 3A; Figure S1A**). On first sampling, 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 when initiated their first collection while 11.8% (2/17) remained asymptomatic throughout the observational window (**Figure 3B**). 11.8% (2/17) of COVID-19+ participants performed their first collection prior to detectable nasal viral load (pre-shedders), providing us with a unique opportunity to obtain high-resolution immune profiles of the earliest host immunity upon exposure to contemporaneous SARS-CoV-2 variants (**Figure 3B**).

To assess temporal dynamicity of host immune kinetics spanning the infection period, 92 longitudinal blood samples (10-14 timepoints/participant; -4 to 57 days PVS) from eight COVID-19+ participants were subjected to high-resolution immune profiling (**Figure 1D**; **Figure 3C**). Samples were binned into seven temporal groups (**Figure S2A**) based on their relative day PVS ranging from G1 (pre-shedding samples) to G7 (post-acute samples), and each temporal group was compared to unexposed-healthy participants derived from a comparable longitudinal cohort (**Figure S2B-H**)[19]. Early acute phase samples (0-7 days PVS) were further divided into two temporal groups (G2 and G3) to capture rapidly evolving immune signatures during early infection (**Figure S2C-D**). An increase in the number of differentially expressed genes was observed as infection progressed from the pre-shedding (G1) to early acute (G2/G3) phases, peaking at days 4-7 (G3) with 242 differentially expressed genes before a significant decrease during the mid-acute (G4/G5) phases, presumably in conjunction with resolution of infection and inflammation (**Figure S2E-F**). Another increase in number of differentially expressed genes was observed as the host entered the late acute (G6) to post-acute (G7) phases, potentially indicating a pivot from pathogen clearance to restoring immune homeostasis and building of immune memory (**Figure S2G-H**).

Functional enrichment analysis of significantly overexpressed transcripts from each temporal group (**Figure S3A**) showed prominent signatures of T-cell mediated immunity in G1 pre-shedding samples (**Figure 4A** and **4B; Figure S3B**). *CD69*, the natural killer (NK) and T cell early activation marker and its upstream transcriptional regulators (*FOS/JUN/JUNB* AP-1 transcriptional complex) were overexpressed in the G1 pre-shedding phase samples along with T cell exhaustion markers (*LAG3*, *PDRM1*, *TIGIT*, *CTLA4*) (**Figure 4C**). Notably, robust expression of many innate interferon-stimulated genes (ISGs) known to be hallmarks of early host antiviral response were lacking prior to viral shedding (**Figure 4A**). Instead, interferon-γ (IFN-γ), its downstream targets (*GBP1-GBP5* GTPases), and the T-cell specific RANTES chemokine (*CCL5*) and its cognate receptor *CCR5* (known to restrict HIV-1 replication) were enriched in pre-shedding samples. Cytotoxic cell signatures such as the immunoglobulin-like receptor (*KIR2DL3*), killer cell lectin-like receptor (*KLRD1*), and cytotoxic granule proteins such as pore-forming granulysins (*GNLY*) and perforins (*PRF1*), and multiple death-inducing serine proteases (*GZMA, GZMB, GZMH*) were also overexpressed during early infection (**Figure 4A**). Given that IFN-γ is produced predominantly by cytotoxic cells, these results suggest extremely early activation of NK or effector T-cell activity and T-cell mediated viral elimination. Daily measurements during these early phases further demonstrate that expression kinetics of these early host responses can be highly transient as seen by a rapid transcriptional downregulation over the course of a day (**Figure 5**).

With the exception of one asymptomatic participant (Z271) who failed to mount a detectable interferon response (**Figure 5**), progression into the early acute (G2/G3) phases was marked by robust expansion of ISG expression along with *DDX58/IFIH1* mediated induction of IFN-α/β (**Figure S3A, Figure 4A-C**). Viral load and host gene expression kinetics in individual participants further demonstrated that expression of antiviral ISGs was induced upon onset of nasal viral shedding and depict high homogeneity in temporal kinetics suggestive of a strong co-regulation of the transcriptional circuitry among the ISGs (**Figure 5**). The magnitude of ISG response depicts striking alignment with nasal viral load kinetics and dynamics of symptom severity, supporting their role in viral clearance (**Figure 5**). With frequent measurements enabling daily kinetics of early infection response, we observed a decoupling of *IFI27* gene expression kinetics to that of other ISGs in several participants, with *IFI27* displaying distinct temporal trends marked by delayed and prolonged expression. *IFI27*, a known ISG,

was recently shown to outperform other biomarker signature panels in predicting cases from non-cases and proposed to be developed as an early screening biomarker for exposed close contacts to identify pre-symptomatic cases[13]. Recently, we demonstrated that *IFI27* expression kinetics track closely with viral load kinetics in mild-to-moderate SARS-CoV-2 infections[19]. The decoupling of *IFI27* expression kinetics from other ISGs suggests a distinct molecular machinery governing *IFI27*-mediated antiviral mechanism from that of other ISGs.

Evidence of abortive subclinical infections have been observed in individuals exposed to SARS-CoV-2[22]. In human viral 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 did not progress to productive viral replication and clinical disease[8,15]. Here, we aimed to assess if evidence of early protective immunity can be observed in the blood transcriptional response of exposed-uninfected individuals with natural exposures to contemporaneous strains. 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 (**Figure S5A**). As anticipated, a robust interferon response was observed in SARS-CoV-2 infected participants (**Figure S5A**). We found elevated expression of 22 transcripts in exposed-uninfected individuals that were enriched in host defense to bacteria and fungi (**Figure S5B**) including a cluster of neutrophil-associated host defense peptides (**Figure 6A**). To assess if elevated expression of HDPs in exposed-uninfected individuals were reflective of either a transcriptional suppression in exposed-infected individuals or transcriptional activation in uninfected individuals, we compared their levels to unexposed-healthy controls (**Figure S5A**). Except for alkaline phosphatase (*ALPL*), all HDPs (*BPI, DEFA4, LCN2*, *LTF,* and *S100A12*) were significantly overexpressed in exposed-uninfected participants when compared to their levels in unexposed-healthy controls (**Figure 6A and Figure S5A**), suggesting that elevated expression of HDPs in exposed-uninfected individuals may play a role in host protection against progression to clinical disease upon pathogen exposure.

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 (**Figure S5C**). In addition to HDPs, both mast cell-associated receptors and mediators (*CPA3*, *HDC, MS4A2*) (**Figure 6A**) and NK/effector T cell receptors (*KLRB1*, *KLRC1*, *TNFRSF9*) (**Figure 6A**) also displayed elevated expression in exposed-uninfected participants relative to infected participants. 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 was overexpressed in exposed-uninfected participants relative to both infected and healthy participants (**Figure 6A** and **Figure S5A**). Next, we assessed the temporal kinetics of several HDPs with known antimicrobial functions including the iron-sequestering protein lactotransferrin (*LTF*), siderophore-binding protein lipocalin-2 (*LCN2*), bactericidal-permeability protein (*BPI*), and the immunomodulatory neutrophil defensin 4 (*DEFA4*) to assess whether their elevated expression reflects a transient viral exposure-induced response (**Figure 6B**). As shown in **Figure 6B**, expression levels of HDPs were stable and consistently elevated in exposed-uninfected individuals throughout the four-week observation window. On the contrary, the viral exposure-induced expression of *IFI27* depicted high temporal dynamicity with rapid upregulation during early infection followed promptly by downregulation as the infection resolved. This observation warrants further validation of exposed-uninfected individuals in larger cohort sizes to delineate mechanisms underlying host protection during pathogen exposure.

DISCUSSION

Early dynamics of host immunity following pathogen exposure can offer insights into pathways that confer susceptibility or protection against infection, or against symptoms despite infection. The earliest host transcriptional responses to viral infections can be rapid and transient [23-25], making them hard to capture in natural infections when exposure timelines cannot be controlled. In-person venipuncture still presents a logistical barrier to conducting longitudinal studies requiring frequent blood collections; high-frequency immune profiling of early blood transcriptional 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 the natural field setting and in natural exposures to infections acquired through community transmission.

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 use in an observational cohort aimed at profiling extremely early host immunity to natural respiratory infections. To the best of our knowledge, this is the first prospective longitudinal study to perform high-frequency intra-individual measurements spanning the pre-shedding-to-convalescence

disease continuum. Using home self-collection tools, we synced daily kinetics of viral load and host blood transcriptional response in natural infections. Self-collection and the fully remote study design enabled isolation of infected participants, eliminating exposure risks associated with in person studies during early disease when participants are most infectious.

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 (**Figure 1C**). Given that duration of the incubation period varies significantly between individuals and can be as short as several days, a short interval between symptom screening and first sample collection was integral to capturing exposed close contacts who were still in their pre-symptomatic phase. In this study, a small but interesting subset of participants were pre-shedders, who provided us with the unique opportunity to observe the earliest transcriptional changes in the periphery prior to measurable nasal viral load. Two pre-shedders were observed in our cohort: one progressed to moderate symptoms while another remained asymptomatic throughout the duration of the study.

Detailed temporal trajectories of 773 innate and acquired immune genes along with temporally aligned nasal 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 the 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 (**Figure 4**).

Second, elevated expression of HDPs were observed in exposed-uninfected individuals relative to infected and unexposed-healthy participants. 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 and binding of microbial siderophores, an iron-acquisition small molecule produced by many bacteria and fungi. In SARS-CoV-2, defensins have been demonstrated to restrict viral entry in *in vitro* studies[26] and owing to their longstanding reputation as key innate defenders, HDPs were quickly proposed in therapeutic applications against SARS-CoV-2[27] based upon antiviral efficacies observed from *in vitro* cellular assays. Our findings provide evidence in human studies for the potential protective role of HDPs against SARS-CoV-2, which can be validated in a larger cohort and for other ARIs using the decentralized study design introduced in this work. Aside from HDPs, we also observed elevated expression of *TNFRSF9,* a TNF receptor preferentially expressed on antigen-activated T cells, in exposed-uninfected participants relative to infected and healthy participants. Taken together, data from this pilot cohort supports a role for T-cell mediated immunity 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, and during infection, protection against severe disease [22,28,29].

There are several limitations to the study. First, our pilot cohort is limited to a small cohort size, resulting in a limited number of pre-symptomatic and pre-shedding participants observed. However, high temporal dimensionality of the dataset afforded by frequent intra-individual measurements (10-14 timepoints per individual) led to several key observations that can be further validated in larger cohorts. 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 breadth of the biology observed in this pilot cohort. 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 the feasibility of decentralized sampling using home self-collection tools to capture early, transient changes in host response soon after viral exposure in natural infections at high temporal resolution comparable to human challenge studies. Overarchingly, our study provides a technological and study framework—using *home*RNA— to investigate detailed kinetics of host immunity to not just ARIs, but various infectious and non-infectious

diseases with dynamic and rapidly evolving immune states, such as those characterized by recurrences or relapses (e.g., autoimmune disorders and malignancies). Most importantly, providing the autonomy of completing study procedures from a home setting to study participants themselves enables conduct of large-scale population-wide mechanistic studies across diverse communities. By reducing study burden, decentralized sampling has significant potential to augment enrollment of participants most affected by logistical barriers, particularly individuals from underreported, underserved, and understudied communities.

References

1. Ong EZ, Kalimuddin S, Chia WC, 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, Lukassen S, Trump S, 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, Clapham HE, Tan AT, et al. Highly functional virus-specific cellular immune response in asymptomatic SARS-CoV-2 infection. *J Exp Med* 2021; **218**(5).

5. Wright PF, Prevost-Reilly AC, Natarajan H, et al. Longitudinal Systemic and Mucosal Immune Responses to SARS-CoV-2 Infection. *J Infect Dis* 2022; **226**(7): 1204-14.

6. Notarbartolo S, Ranzani V, Bandera A, et al. Integrated longitudinal immunophenotypic, transcriptional and repertoire analyses delineate immune responses in COVID-19 patients. *Sci Immunol* 2021; **6**(62).

7. Zaas AK, Chen M, Varkey J, 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, McClain MT, Chen M, 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, Zaas AK, Rao A, et al. Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza a infection. *PLoS Genet* 2011; **7**(8): e1002234.

10. Choy RKM, Bourgeois AL, Ockenhouse CF, Walker RI, Sheets RL, Flores J. Controlled Human Infection Models To Accelerate Vaccine Development. *Clin Microbiol Rev* 2022; **35**(3): e0000821.

11. Abo YN, Jamrozik E, McCarthy JS, Roestenberg M, Steer AC, Osowicki J. Strategic and scientific contributions of human challenge trials for vaccine development: facts versus fantasy. *Lancet Infect Dis* 2023; **23**(12): e533-e46.

12. McClain MT, Constantine FJ, Nicholson BP, 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, Rosenheim J, Bell LC, 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. Yu S, Di C, Chen S, et al. Distinct immune signatures discriminate between asymptomatic and presymptomatic SARS-CoV-2(pos) subjects. *Cell Res* 2021; **31**(11): 1148-62.

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

16. Eyal N, Lipsitch M, Smith PG. Human Challenge Studies to Accelerate Coronavirus Vaccine Licensure. *J Infect Dis* 2020; **221**(11): 1752-6.

17. 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.

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

19. 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).

20. 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.

21. Benjamini Y, 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.

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

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

24. Zhu H, Cong JP, Yu D, Bresnahan WA, Shenk TE. Inhibition of cyclooxygenase 2 blocks human cytomegalovirus replication. *Proc Natl Acad Sci U S A* 2002; **99**(6): 3932-7.

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

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

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

28. Andreakos E, Abel L, Vinh DC, 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.

29. Pancino G, Saez-Cirion A, Scott-Algara D, Paul P. Natural resistance to HIV infection: lessons learned from HIV-exposed uninfected individuals. *J Infect Dis* 2010; **202 Suppl 3**: S345-50.

ACKNOWLEDGEMENTS

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

We would like to acknowledge Dr. Kimberly Dill-McFarland at the Department of Medicine, University of Washington for discussions surrounding bioinformatic analysis. The REDCap instance used is supported by the Institute of Translational Health Sciences, which is funded by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1TR002319.

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. ABT and EB have ownership in Seabright, LLC, which will advance new tools for diagnostics and clinical research, potentially including the *home*RNA platform. 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, GlaxoSmithKline and Vir; receiving personal fees from Vir; all outside the submitted work. MB has clinical research support from Ansun Biopharma, 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 supplemental data 2. R scripts developed to perform analyses and visualizations are available upon request from the co-senior authors.

FIGURES

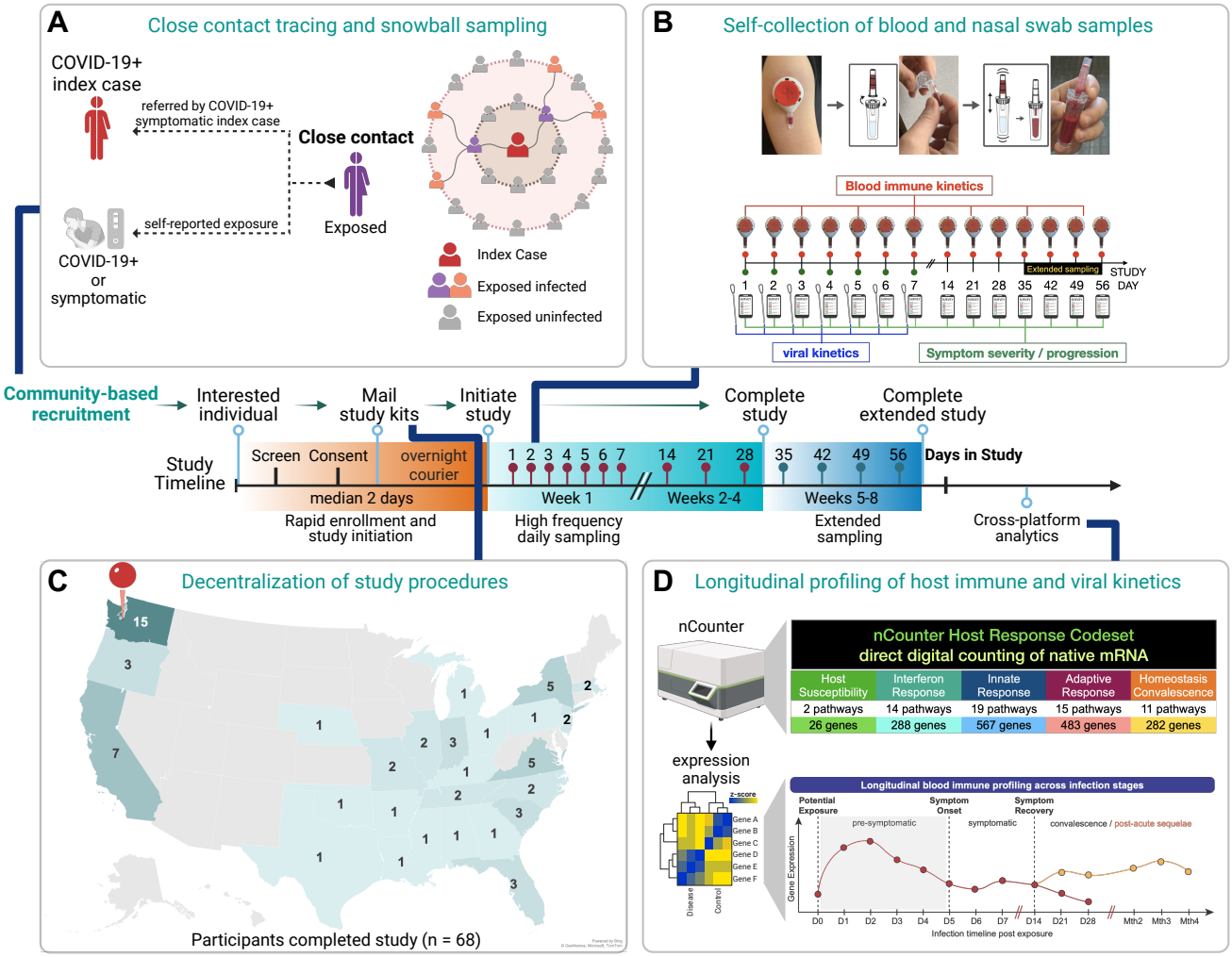

Figure 1. Study framework using home self-collection tools to observe detailed kinetics of early pre-symptomatic host transcriptional response during natural human respiratory infections. Conceptual framework depicting the study design and participation timeline (middle), A) close contact recruitment strategies, B) sample collection details (adapted from Haack and Lim et al, 2021 with permission) , C) geographical representation of study participants, and D) longitudinal gene expression analyses pipeline.

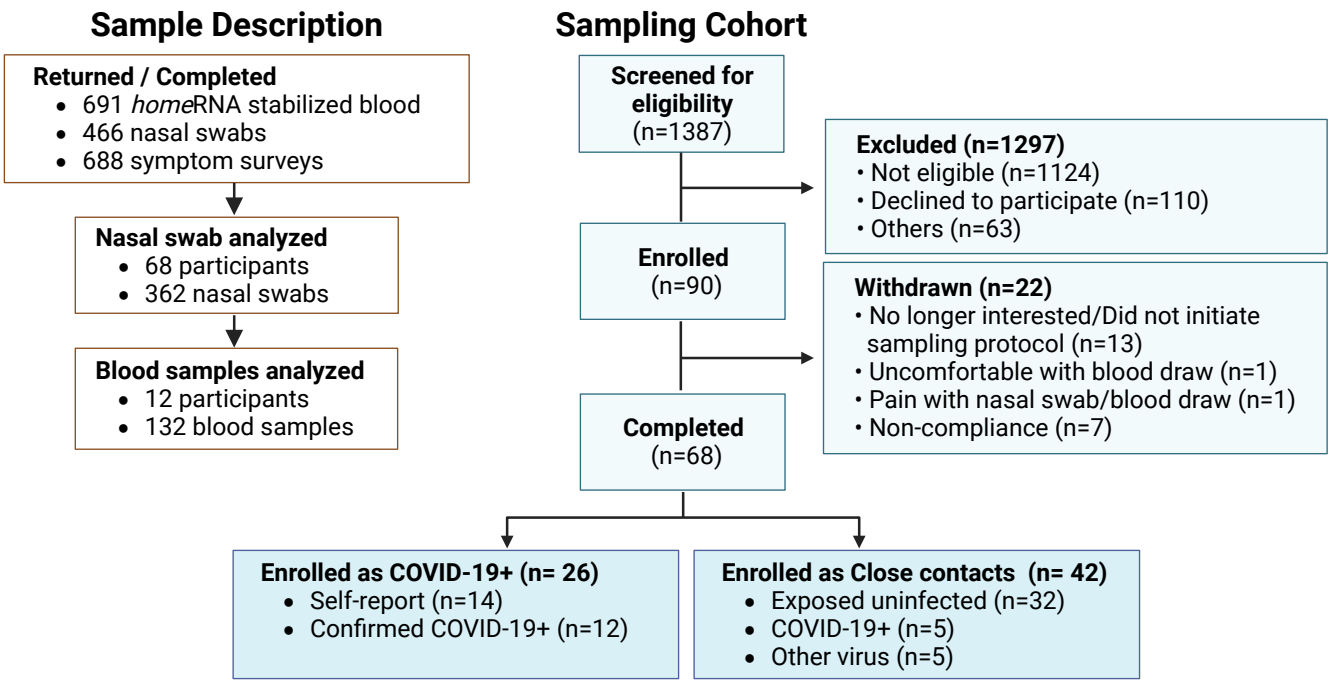

Figure 2. Participant and sample flow chart. Participants enrolled as COVID-19+ self-reported a recent SARS-CoV-2 positive test. Participants enrolled as close contacts had unknown infection status at the time of enrollment.

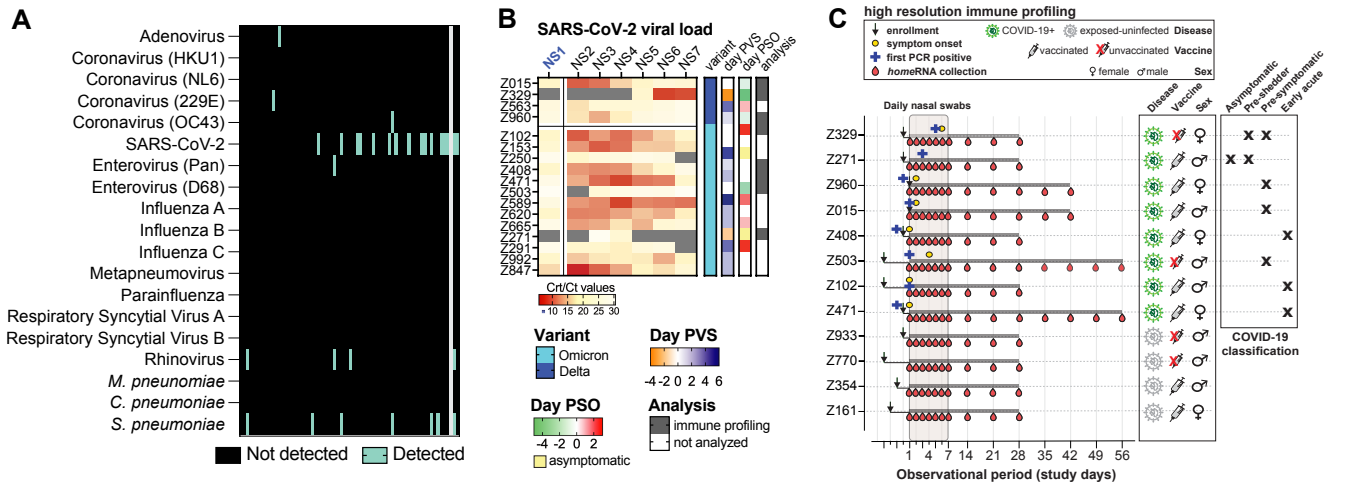

Figure 3. Virology status and classification of participants. A) Heatmap depicting pathogens detected in all 68 participants. Rows and columns represent pathogens and participants, respectively. 3-4 nasal swab per participant were screened for respiratory pathogens. Gray column denotes participant with a single nasal swab assessed for SARS-CoV-2 (not detected); this participant was not included in high-resolution immune profiling. **B)** Heatmap depicting SARS-CoV-2 viral load (n=17). Grey cells denote samples with undetectable viral load. Rows and columns represent participants and nasal swab samples, respectively. NS1 was subjected to clinical SARS-CoV-2 test (Ct values) and NS2-7 were subjected to a lab-developed multi-pathogen screening assay (Crt values). Day 0 denotes onset of symptom (days PSO) and viral shedding (days PVS). **C)** Clinical and sample characteristics of 12 participants (8 COVID-19+ and 4 exposed-uninfected) selected for high-resolution immune profiling.

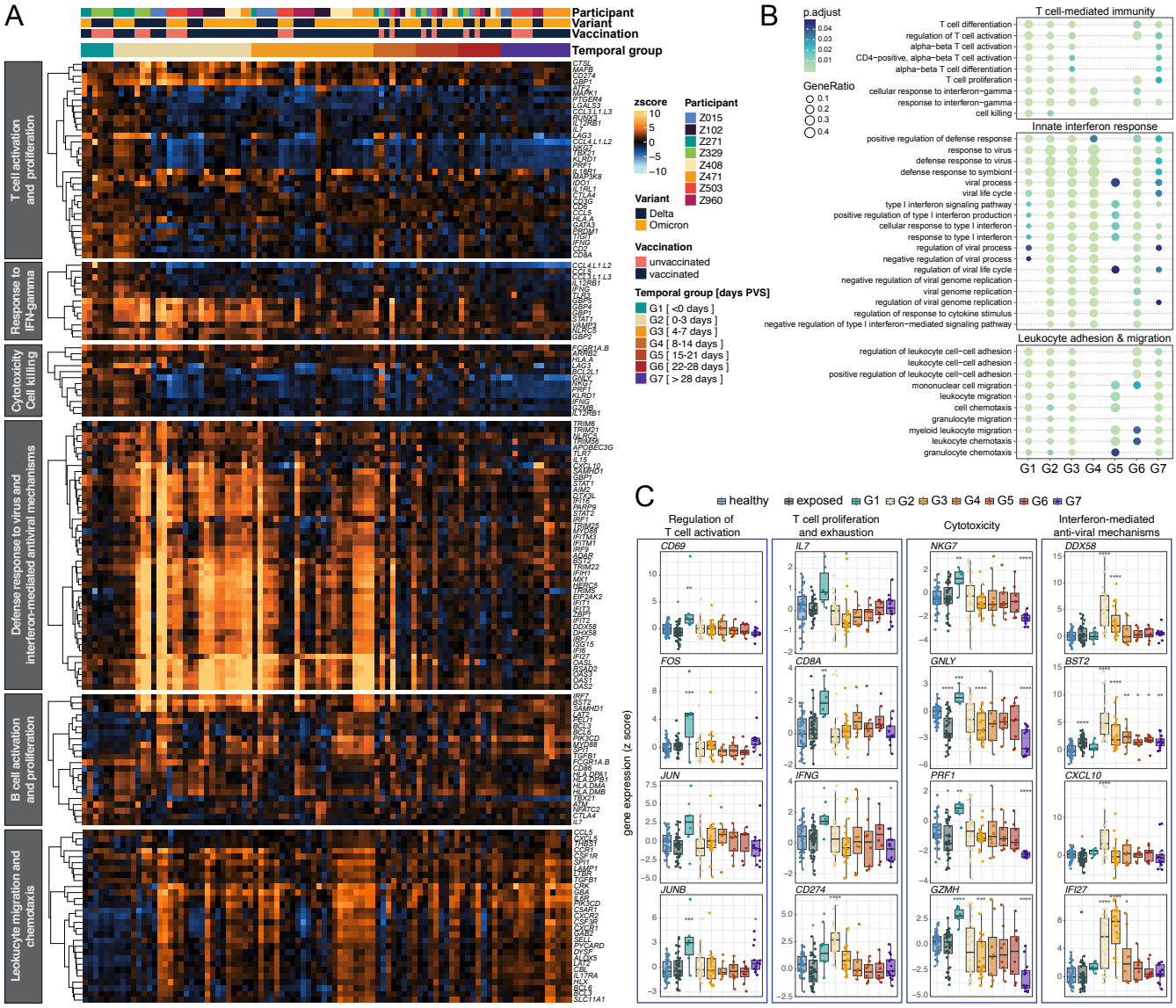

Figure 4. Prominent cytotoxic and T cell signatures observed in early pre-shedding samples. A) Heatmap depicting hierarchical clustering of differentially overexpressed genes (rows) in COVID-19+ participants for each temporal group (G1-G7). Differentially expressed genes were identified through pairwise comparisons between each temporal group and unexposed-healthy reference controls. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg procedure and an absolute $\log_2$ fold change of > 0.322 (1.25x) and an adjusted p-value of < 0.1 were used to identify differential expression. A subset of differentially expressed genes are portrayed in the heatmap and grouped based on membership in select GO biological processes enriched in pre-shedding (G1) to post-acute (G7) temporal groups. Column represents participant samples clustered by temporal group. Gene expression z scores (heatmap color) were computed using the average and standard deviations measured from unexposed-healthy controls ($n_{participants}$ = 5; $n_{samples}$ =39). B) Dot plot displaying enrichment score (gene ratio) of select GO biological processes enriched in pre-shedding, early acute, and late to post-acute infection phases grouped by parent terms (**Figure S4**) across all temporal groups. Dot size represents gene ratio and color represents adjusted p-values. All top 20 enriched GO biological processes from each temporal group are visualized in **Figure S3B**. C) Boxplot depicting gene expression z scores of differentially expressed genes associated with T cell activation and proliferation, cytotoxicity, and innate interferon response grouped by unexposed-uninfected (healthy), exposed-uninfected, and exposed-infected (COVID-19+) participants stratified across G1-G7 temporal groups. Asterisks represent statistical significance determined through pairwise comparison with unexposed-healthy participants, considering multiple comparisons for all genes

using the Benjamini-Hochberg procedure. Significance between groups were labeled based on adjusted p-values as follows: p_{adj}>0.1 (ns), p_{adj} ≤0.1 (*), p_{adj} ≤0.01 (**), p_{adj} ≤0.001 (***), p_{adj} ≤0.0001 (****).

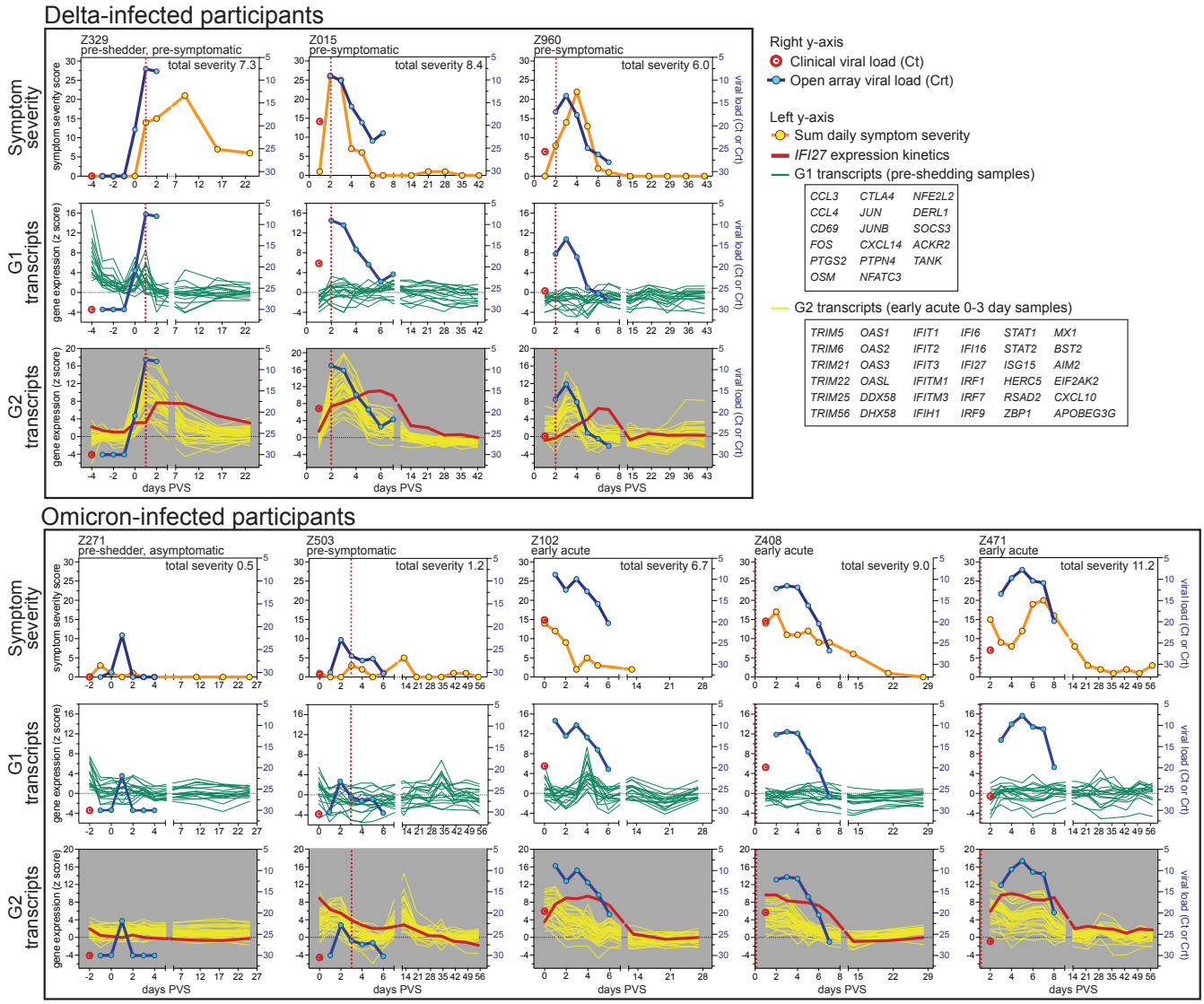

Figure 5. Detailed immune kinetics in individual participants depict transient upregulation of early pre-shedding response transcripts and decoupling of *IFI27* expression kinetics from early acute transcripts in some participants. G1 and G2 transcripts represent a subset of differentially expressed genes identified through pairwise comparison of "G1 vs unexposed-healthy" and "G2 vs unexposed-healthy" comparator group. Red vertical line denotes symptom onset. Symptom onset for early acute participant Z102 outside of plotted temporal window. Viral load is plotted on the right y-axis. Viral load measured using clinical SARS-CoV-2 RT-PCR tests are reported as Ct values while viral load measured using a lab-developed multi-pathogen assay (Open Array) are reported as Crt values. Symptom severity and gene expression are plotted on the left y-axis. Gene expression z scores were computed using the average and standard deviations measured from unexposed-healthy controls ($n_{participants}$ = 5; $n_{samples}$ =39).

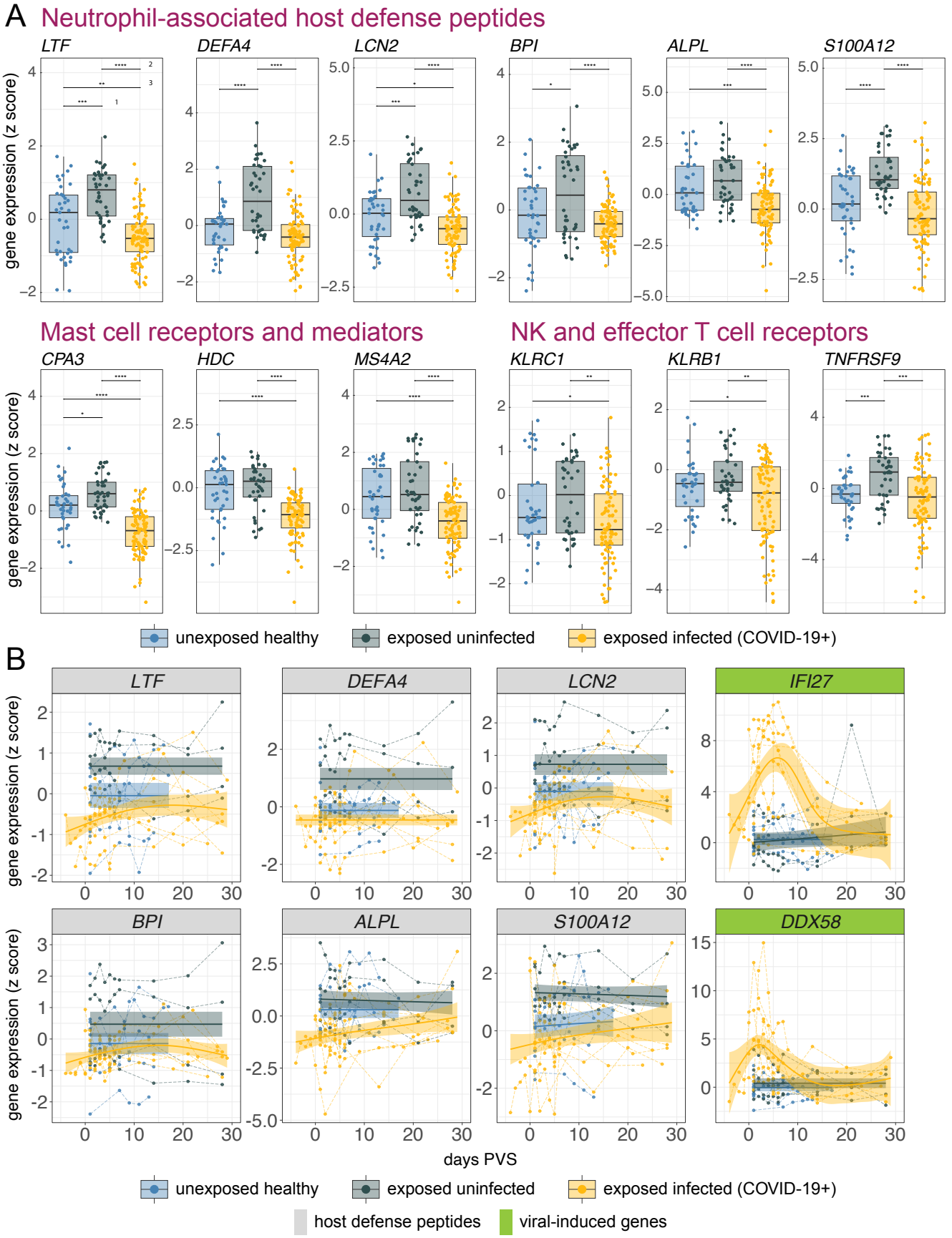
A
Neutrophil-associated host defense peptides
LTF
DEFA4
LCN2
BPI
ALPL
S100A12
gene expression (z score)
Mast cell receptors and mediators
NK and effector T cell receptors
CPA3
HDC
MS4A2
KLRC1
KLRB1
TNFRSF9
unexposed healthy
exposed uninfected
exposed infected (COVID-19+)
B
IFI27
DDX58
days PVS
host defense peptides
viral-induced genes

Figure 6. Exposed-uninfected individuals depict elevated expression levels of host defense peptides relative to unexposed-healthy and COVID-19+ individuals. Box plots depicting gene expression z scores of **A)** neutrophil-associated host defense peptides, mast cell receptors and mediators, and NK and effector T cell receptors for all longitudinal samples stratified by unexposed-uninfected (healthy), exposed-uninfected, and exposed-infected (COVID-19+) participants. Statistical significance between groups were determined through pairwise comparisons of "exposed infected vs exposed uninfected", "exposed-uninfected vs unexposed-healthy", and "exposed infected vs unexposed-healthy", considering multiple comparisons for all genes using the Benjamini-Hochberg procedure. Significance between groups were labeled based on adjusted p-values as follows: $p_{adj}>0.1$ (ns), $p_{adj} \leq 0.1$ (*), $p_{adj} \leq 0.01$ (**), $p_{adj} \leq 0.001$ (***), $p_{adj} \leq 0.0001$ (****). **B)** Generalized additive model (GAM) smooth function plots of host defense peptides and two representative viral-induced genes (*IFI27, DDX58*) stratified by unexposed-uninfected (healthy), exposed-uninfected, and exposed-infected (COVID-19+) participants. Solid lines represent generalized additive model smoothing across all participants within a group and colored shades represent the 95% confidence interval. Each dotted line represents a participant's temporal trajectories for the plotted gene.

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			
First sample (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