Med, Volume *3*

Supplemental information

Gastrointestinal symptoms and fecal shedding of SARS-CoV-2 RNA suggest prolonged gastrointestinal infection

Aravind Natarajan, Soumaya Zlitni, Erin F. Brooks, Summer E. Vance, Alex Dahlen, Haley Hedlin, Ryan M. Park, Alvin Han, Danica T. Schmidtke, Renu Verma, Karen B. Jacobson, Julie Parsonnet, Hector F. Bonilla, Upinder Singh, Benjamin A. Pinsky, Jason R. Andrews, Prasanna Jagannathan, and Ami S. Bhatt

Supplementary figures

Gastrointestinal symptoms and fecal shedding of SARS-CoV-2 suggest prolonged gastrointestinal infection

Aravind Natarajan*[1,2], Soumaya Zlitni*[1,2], Erin F. Brooks*[2], Summer E. Vance*[2], Alex Dahlen[3], Haley Hedlin[3], Ryan M. Park, Alvin Han[4], Danica T. Schmidtke[4], Renu Verma[5], Karen B. Jacobson[5], Julie Parsonnet[6,7], Hector F. Bonilla[6], Upinder Singh[5], Benjamin A. Pinsky[5,8], Jason Andrews[5], Prasanna Jagannathan[4,6], Ami S. Bhatt^[1,2]

[1] Department of Genetics, Stanford University, Stanford, CA, USA.

[2] Department of Medicine (Hematology, Blood and Marrow Transplantation), Stanford University, Stanford, CA, USA.

[3] Quantitative Science Unit, Stanford University, Stanford, CA, USA.

[4] Department of Microbiology and Immunology, Stanford University, Stanford, CA, USA.

[5] Department of Medicine (Infectious Diseases and Geographic Medicine), Stanford University, Stanford, CA, USA.

[6] Department of Medicine (Infectious Diseases), Stanford University, Stanford, CA, USA.

[7] Department of Medicine (Epidemiology and Population Health), Stanford University, Stanford, CA, USA.

[8] Department of Pathology, Stanford University, Stanford, CA, USA

*These authors contributed equally to this work

^ Address correspondences to: Ami S. Bhatt, 269 Campus Dr, CCSR 1155b, Stanford University, Palo Alto, CA 94305. Tel: (650) 498-4438; Email: asbhatt@stanford.edu.

Contents

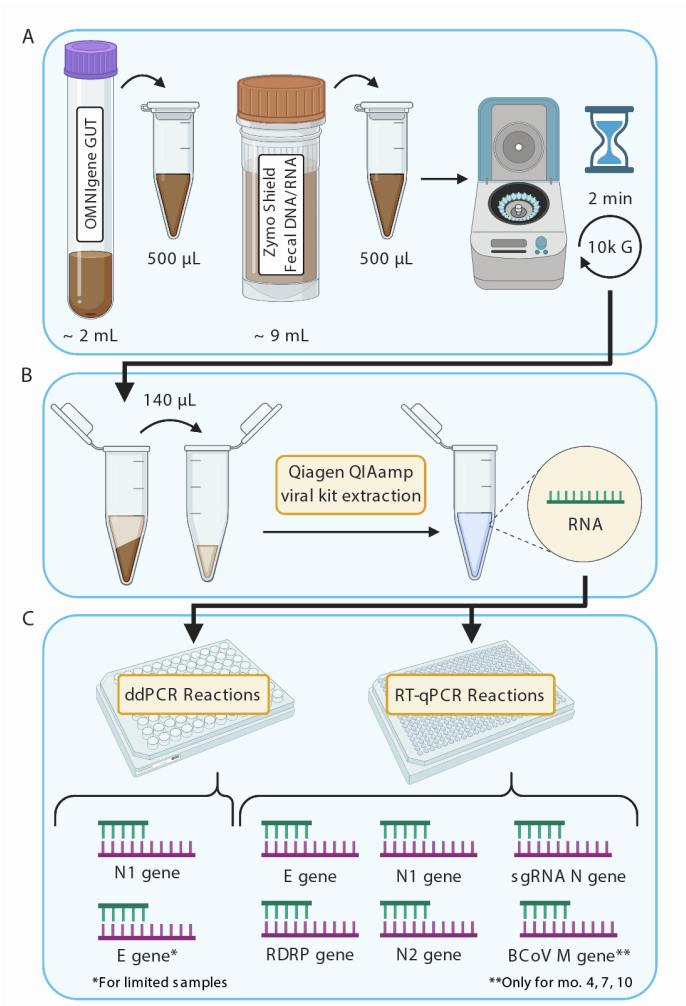

Figure S1. Related to Figure 2, Schematic illustration of RNA extraction and PCR-based assay methodology

A. Sample preparation. Stool samples collected in the Omnigene (OG) and Zymo DNA/RNA shield (ZY) kit were homogenized by vortexing for 30 seconds. 500 µL of the well-mixed samples was transferred to a 1.5 mL microcentrifuge tube. Samples were then centrifuged at 10,000 X g for two minutes to pellet solid waste and separate the aqueous contents. **B.** 140 µL of the aqueous supernatant was transferred to a fresh 1.5 mL microcentrifuge tube and carried through a standard QIAamp viral RNA extraction protocol (Qiagen) as per manufacturer's instructions. In the final step, RNA was eluted in two rounds using 50 µL of elution buffer. Eluted viral RNA was then arrayed in 96-well plates and stored at -80°C. **C.** Viral RNA arrayed in 96-well plates were quantified using two orthogonal methods, droplet digital PCR (ddPCR) and reverse-transcriptase quantitative PCR (RT-qPCR). Out of 673 extracted RNA samples, duplexed ddPCR reactions detecting the E and N1 genes were carried out on 278 samples, and single reactions quantifying the N1 were performed on 395 samples. In RT-qPCR, all 673 RNA extracts were assayed for the E, N1, N2 and RdRP genes in duplicate reactions. Additionally, a subset of samples collected up to 4 months were assayed for sgRNA targeting the N1 gene. Finally, stool samples collected at the later time points, in month 4, 7 and 10 samples, had been spiked with 10 µl attenuated BCoV vaccine as an extraction control. These samples were also assayed for the BCoV M gene.

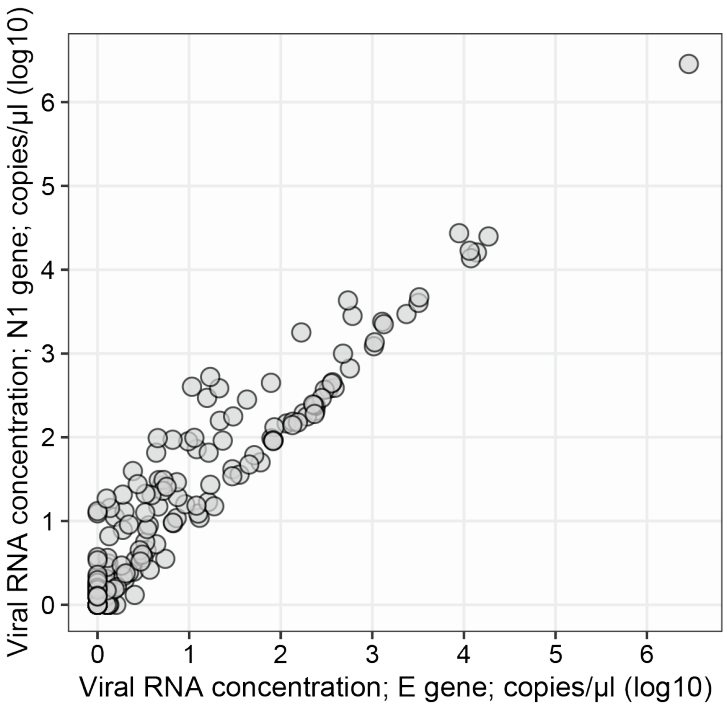

Figure S2. Related to Figure 2, correlation between measurement of gRNA targeting the E and N1 genes using ddPCR

Fecal viral RNA concentrations determined using droplet digital PCR (ddPCR) with primers/probes targeting the E and N1 gene in the SARS-CoV-2 genome from 278 samples. The x-axis lists viral RNA concentration in $\log_{10}$ copies per µL measured by targeting the E gene, while the y-axis lists RNA concentration measured by targeting the N1 gene. Each data point refers to viral RNA derived from 278 samples collected predominantly in the first month of the study. The linear association between these measurements by two orthogonal genes is evaluated by Pearson's correlation, $R = 0.96$, $P<0.0001$. Note that all viral RNA concentrations are expressed on a logarithmic scale by applying the transformation $\log_{10}$(viral RNA concentration+1).

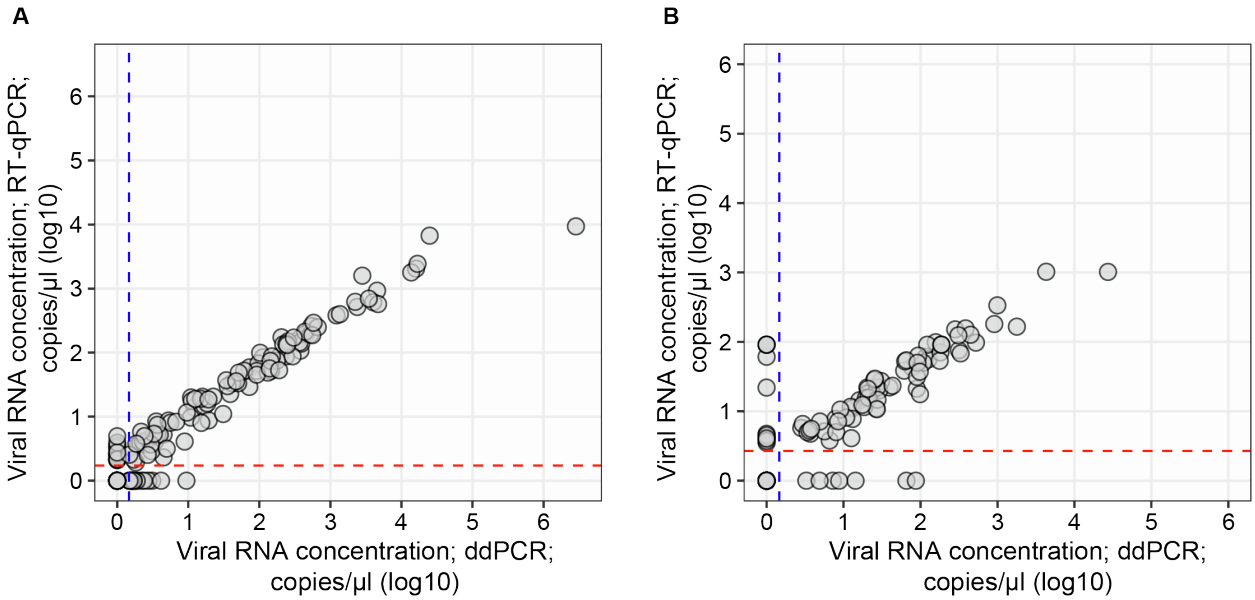

Figure S3. Related to Figure 2, RT-qPCR and ddPCR measurements of SARS-CoV-2 RNA targeting the N1 gene are concordant across both sample collection methods

Association between paired concentrations of fecal SARS-CoV-2 RNA from clinical stool samples measured using orthogonal methods, ddPCR and RT-qPCR, targeting the N1 gene. The x-axis represents viral RNA concentration determined using ddPCR in $\log_{10}$ copies per µL and the y-axis represents viral RNA concentration determined using RT-qPCR in $\log_{10}$ copies per µL. The linear association between the two measurements is evaluated by Pearson's correlation. **A.** Stool samples collected from 104 subjects in the Zymo DNA/RNA shield (ZY) kit. R = 0.98, P<0.0001. **B.** Stool samples collected from 108 subjects in the Omnigene (OG) kit. R = 0.90, P<0.0001. The limit of blank for each assay is indicated by the red and blue dashed lines for RT-qPCR and ddPCR, respectively. Note that all viral RNA concentrations are expressed on a logarithmic scale by applying the transformation $\log_{10}$(viral RNA concentration+1).

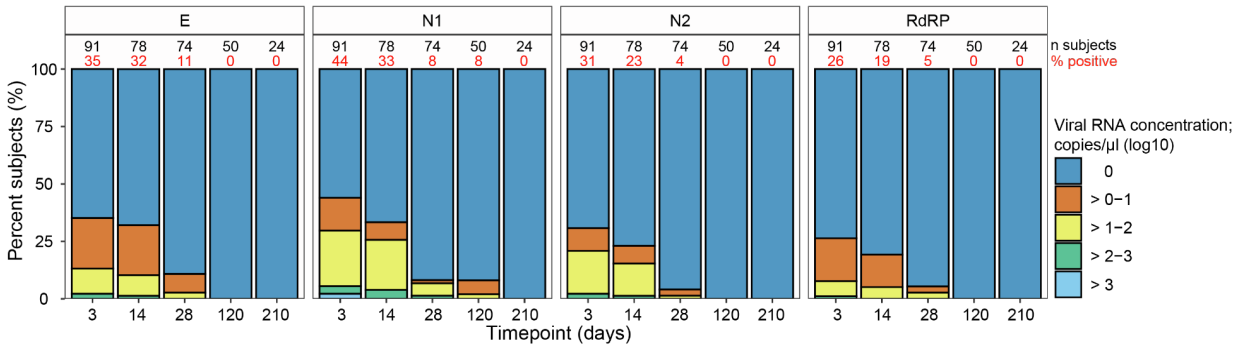

Figure S4. Related to Figure 2, SARS-CoV-2 viral RNA concentration in stool samples collected in the OG kit

Similar to Figure 2c, which is the same plot for ZY kits. Fecal viral RNA concentration was determined using RT-qPCR with primers/probes targeting the E, N1, N2, RdRP genes in the SARS-CoV-2 genome as indicated in the tab at the top of every panel. The x-axis lists time point categories since enrollment as days 3 (range 0 - 7), 14 (8 - 21), 28 (22 - 35), 120 (75 - 165), 210 (166 - 255) and 300 (>255). The y-axis lists the percentage of subjects that bear a given viral RNA concentration as indicated by the color scheme in the stacked bar plot; dark blue refers to samples with no detectable viral RNA, orange to viral RNA concentrations between 0 and 1 $\log_{10}$ copies per µL, yellow between 1 and 2 $\log_{10}$ copies per µL, green between 2 and 3 $\log_{10}$ copies per µL, and light blue over 3 $\log_{10}$ copies per µL. Number of subjects and percentage of subjects that provided a positive stool sample are listed above each stacked bar in black and red fonts respectively. Note that all viral RNA concentrations are expressed on a logarithmic scale by applying the transformation $\log_{10}$(viral RNA concentration+1).

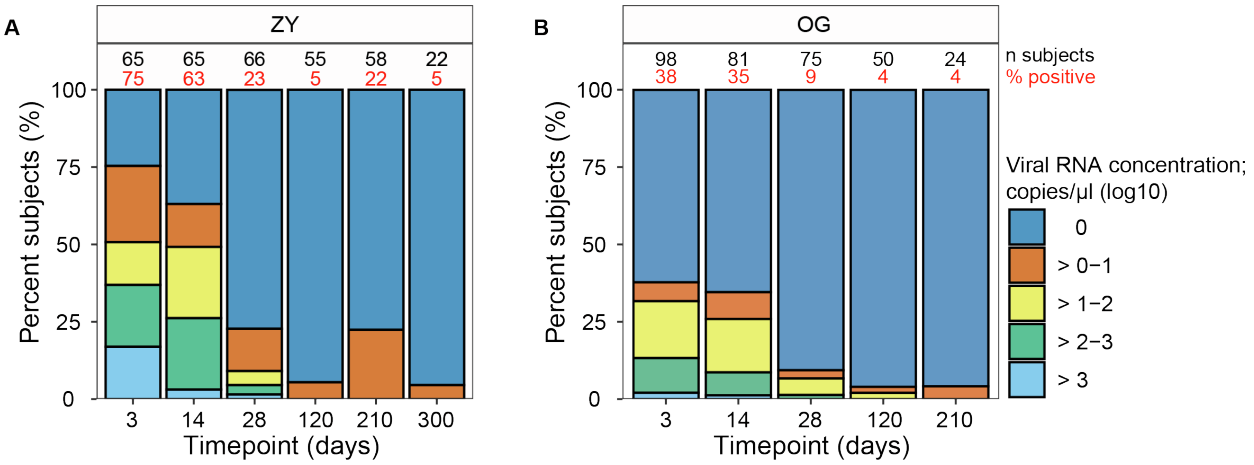

Figure S5. Related to Figure 2, SARS-CoV-2 viral RNA concentration in stool samples quantified by ddPCR across both sample collection methods

Fecal viral RNA concentrations determined using droplet digital PCR (ddPCR) with primers/probes targeting the N1 gene in the SARS-CoV-2 genome. Viral RNA was derived from stool samples preserved in the Omnigene (OG) or Zymo DNA/RNA shield (ZY) kits as indicated in the tab at the top of the panels. The x-axis lists time point categories since enrollment as days 3 (range 0 - 7), 14 (8 - 21), 28 (22 - 35), 120 (75 - 165), 210 (166 - 255) and 300 (>255). The y-axis lists the percentage of subjects that bear a given viral RNA concentration as indicated by the color scheme in the stacked bar plot; dark blue refers to samples with no detectable viral RNA, orange to viral RNA concentrations between 0 and 1 $\log_{10}$ copies per µL, yellow between 1 and 2 $\log_{10}$ copies per µL, green between 2 and 3 $\log_{10}$ copies per µL, and light blue over 3 $\log_{10}$ copies per µL. Number of subjects and percentage of subjects that provided a positive stool sample are listed above each stacked bar in black and red fonts respectively. Note that all viral RNA concentrations are expressed on a logarithmic scale by applying the transformation $\log_{10}$(viral RNA concentration+1).

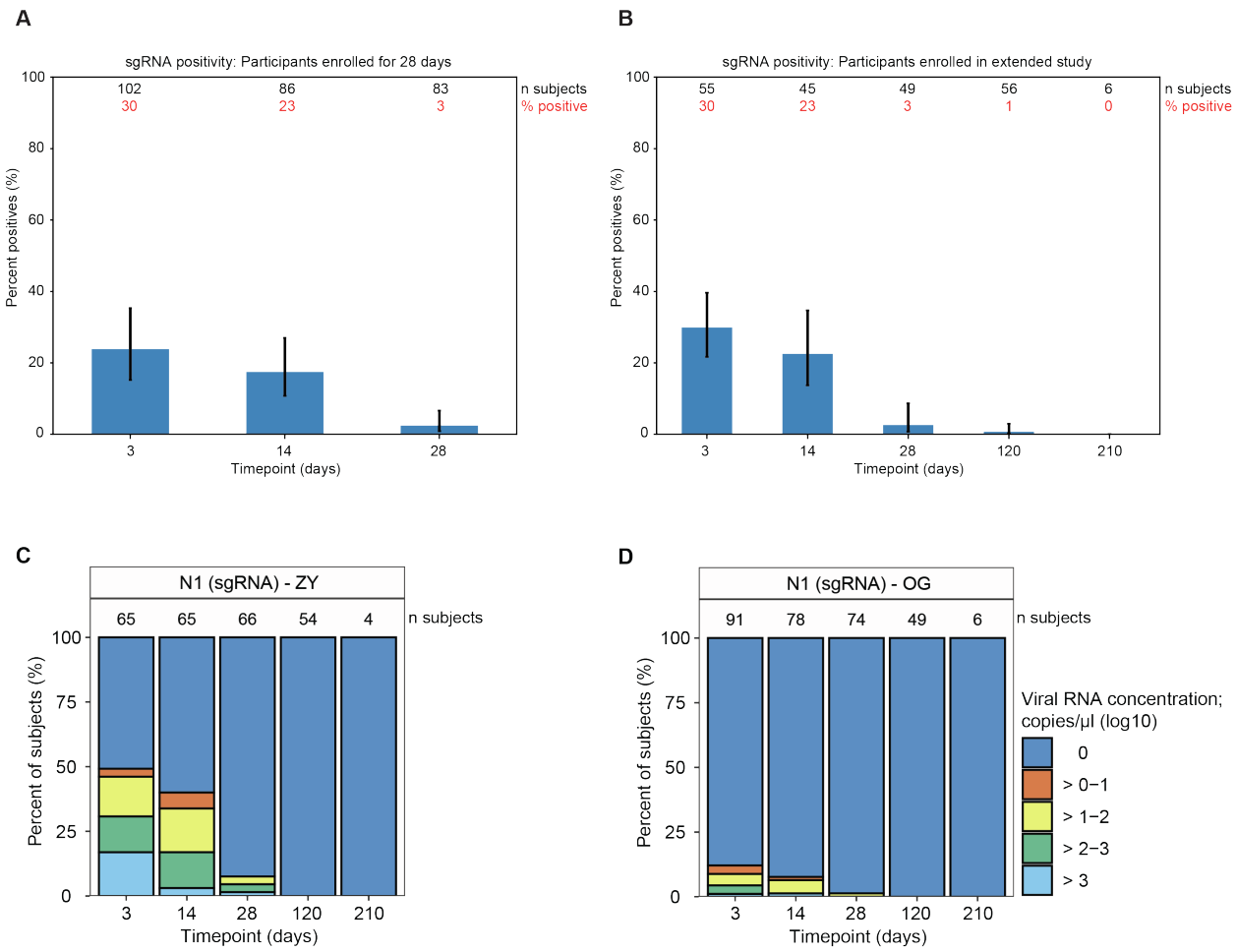

Figure S6. Related to Figure 2, fecal viral sgRNA measurements over time

Fecal subgenomic RNA (sgRNA) concentrations determined using RT-qPCR with primers/probes targeting the N1 gene with a canonical 5' untranslated region. **A.** Summary of viral subgenomic RNA (sgRNA) concentrations derived from fecal samples acquired from participants enrolled in the study. The x-axis lists time point categories since enrollment as days 3 (range 0 - 7), 14 (8 - 21) and 28 (22 - 35). The y-axis lists the percentage of participants with positive fecal samples at each of the time points. Fecal positivity rates are evaluated using the logistic GEE model described in the statistical methods section, which averages over all of the sample collection methods, and technical replicates. The bars indicate 95% confidence intervals. Number of participants and percent positive are listed as numbers at the top of the plot in black and red fonts, respectively. **B.** Same as panel a, except restricted to the subset of those who participated in the extended study, and following them through all 6 time points. As before, the x-axis lists time point categories since enrollment: day 3 (range 0 - 7), 14 (8 - 21), 28 (22 - 35), 120 (75 - 165), 210 (166 - 255) and 300 (>255), and the y-axis lists the percentage of participants with positive fecal samples at each of the time points, with 95% confidence intervals. Number of participants and percent positive are listed in black and red fonts. **C,D.** Viral sgRNA concentrations derived from stool samples preserved in the Zymo DNA/RNA shield fecal collection tube (ZY) (**C**) and OMNIgene GUT collection tube (OG) (**D**). The x-axis lists time point categories since enrollment as days 3 (range 0 - 7), 14 (8 - 21), 28 (22 - 35), 120 (75 - 165) and 210 (166 - 255). The y-axis lists the percentage of subjects with a given concentration of sgRNA as indicated by the color scheme in the stacked bar plot; dark

blue refers to samples with no detectable sgRNA RNA, orange to sgRNA concentrations between 0 and 1 $\log_{10}$ copies per µL, yellow between 1 and 2 $\log_{10}$ copies per µL, green between 2 and 3 $\log_{10}$ copies per µL, and light blue over 3 $\log_{10}$ copies per µL. Number of subjects and percentage of subjects that provided a positive stool sample are listed above each stacked bar in black and red fonts respectively. Note that all viral RNA concentrations are expressed on a logarithmic scale by applying the transformation $\log_{10}$(viral RNA concentration+1).

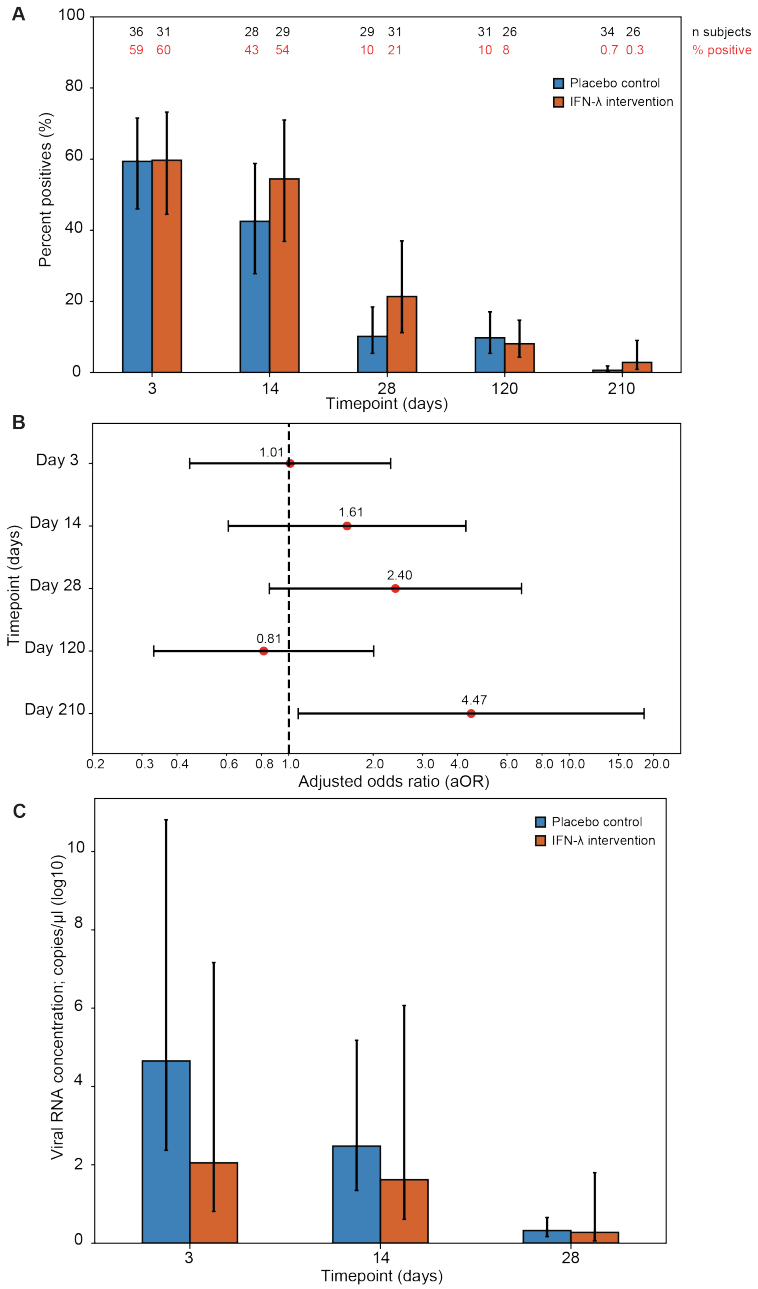

Figure S7. Related to Figure 3, sensitivity analysis of the effect of IFN-λ on fecal SARS-CoV-2 RNA

A. Percentage of participants with detectable fecal SARS-CoV-2 RNA across each of the study arms, as evaluated using the logistic GEE model which averages over all of the sample collection methods, gene types, and technical replicates as described in the statistical methods section. The plot includes the subset of those who participated in the extended study, following them through all 6 time points. The x-axis lists time point categories since enrollment as days 3 (range 0 - 7), 14 (8 - 21), 28 (22 - 35), 120 (75 - 165), and 210 (166 - 255). The y-axis indicates the percentage of participants with detectable fecal SARS-CoV-2 RNA. The blue bar corresponds to participants in the placebo control arm, and the orange bar corresponds to participants in the IFN-λ intervention arm. Each bar also marks the 95% confidence interval. Number of participants and percentage of participants that provided a positive stool sample are listed above each stacked bar in black and

red fonts, respectively. **B.** Odds ratio comparing detectable fecal SARS-CoV-2 RNA shedding in the IFN-λ intervention arm to the placebo arm at each of the 6 time points for the subset of those who participated in the extended study. The x-axis marks the odds ratio adjusted for age, sex, collection kit type (OG or ZY) and target gene (E, N1, N2, or RdRP) (aOR). The y-axis marks the time point in the study as days 3 (range 0 - 7), 14 (8 - 21), 28 (22 - 35), 120 (75 - 165), and 210 (166 - 255). The point denotes the aOR, flanked by lines denoting the 95% confidence intervals. The red dashed vertical line at aOR = 1.0 denotes no association. **C.** Fecal viral RNA concentration in stool samples collected across each of the study arms from all participants over the first month of enrollment, as evaluated using a negative binomial GEE model. Like the primary statistical model, this one averages over sample collection kits, genes, and replicates, with fixed effects to adjust for those features, as well as age and sex; we also transformed the viral RNA concentration by $\log_{10}$(viral RNA concentration+1) to reduce dispersion. The x-axis lists time point categories since enrollment as days 3 (range 0 - 7), 14 (8 - 21), and 28 (22 - 35). The y-axis indicates viral RNA concentration in $\log_{10}$ copies per μL. The blue bar corresponds to participants in the placebo control arm, and the orange bar corresponds to participants in the IFN-λ intervention arm. Each bar also marks the 95% confidence interval.