The raw data for this paper is listed in 29 files:

1. Fig_1_panelA_Participant_Swab_data.xlsx
2. Fig_1_panelA_Swab_CalibrationCurvedata.csv
3. Fig_1_panelB_Saliva_CalibrationCurvedata.xlsx
4. Fig_1_panelB_Saliva_Participant_Data.csv
5. Fig_2A_data.xlsx
6. Fig_2B_data.xlsx
7. Fig_2C_data.xlsx
8. Fig_2D_data.xlsx
9. Fig_2E_data.xlsx
10. Fig_2F_data.xlsx
11. Fig_2G_data.xlsx
12. Fig_2_SequencingVariant_GISAID_Table.xlsx
13. Fig_3A_data.xlsx
14. Fig_3B_data.xlsx
15. Fig_3C_data.xlsx
16. Fig_3D_data.xlsx
17. SI_Fig_1A_saliva_LOD.csv
18. SI_Fig_1B_swab_LOD.xlsx
19. SI_Fig_2_CalibrationCurve_data_plots.csv
20. SI_Fig_3A_data.xlsx
21. SI_Fig_3B_data_saliva.xlsx
22. SI_Fig_3B_data_swab.csv
23. SI_Fig_4A_data.xlsx
24. SI_Fig_4B_data.xlsx
25. SI_Fig_4C_data.xlsx
26. SI_Fig_4D_data.xlsx
27. SI_Fig_4E_data.xlsx
28. SI_Fig_4F_data.xlsx
29. SI_Fig_4G_data.xlsx

The file Fig_1_panelA_Participant_Swab_data.xlsx provides the data of the participant nasal swab samples run with RT-qPCR and RT-ddPCR protocols for panel A in Figure 1.

1. “SAMPLE ID” = Caltech-internal sample ID code to participant sample.
2. “N1 Cq” = mean of RT-qPCR technical duplicate wells of the SARS-CoV-2 N1 gene target
3. “RT-qPCR N1 Conc. Cp/ml” = the calculated viral load concentration of the SARS-CoV-2 N1 gene in copies/mL (cp/ml) from the calibration curves equation provided in the paper methods and the mean Cq reported in the column “N1 Cq”
4. “ddPCR Concentration in Spectrum (cp/ml)” = the calculated viral load concentration of the SARS-CoV-2 N1 gene quantified with RT-ddPCR and dilution-corrected to obtain the concentration in the original sample
5. “Geometric Mean” = geometric mean of the concentration (copies/mL) from the RT-qPCR method reported in column “RT-qPCR N1 Conc. Cp/ml” and RT-ddPCR method reported in column “ddPCR Concentration in Spectrum (cp/ml)”

The file Fig_1_panelA_Swab_CalibrationCurvedata.csv provides the data of the nasal swab calibration curve samples included in Figure 1 panel A; details of the methods are provided in the paper

1. "Protocol" = Which extraction protocol was run ("Swab" or "Saliva")
2. "Marker" = Which SARS-CoV-2 gene target the measurements correspond to SARS-CoV-2 "N1"
3. "Theoretical SARS-CoV-2 Concentration (input into KF extraction)" = theoretical concentration of SARS-CoV-2 heat-inactivated particles, in units of copies/mL, based on dilutions from the concentration reported in the COA, in each of the samples input into the KingFisher (KF) extraction protocol
4. "Avg_Cq" = mean of three PCR technical replicate Cq values with the RT-qPCR protocol; individual replicate Cq data are not shown
5. "Replicate" = Technical extraction replicate sample number at each concentration, three replicates are denoted as "Rep1", "Rep2", or "Rep3"
6. "Calculated Concentration (cp/ml)" = the calculated viral load concentration of the SARS-CoV-2 N1 gene in copies/mL (cp/ml) from the calibration curves N1 s equation provided in the paper methods and the mean Cq reported in the column "Avg_Cq"

The file Fig_1_panelB_Saliva_CalibrationCurvedata.csv provides the data of the saliva calibration curve samples included in Figure 1 panel B; details of the methods are provided in the paper

1. "Protocol" = Which extraction protocol was run ("Swab" or "Saliva")
2. "Gene" = Which SARS-CoV-2 gene target the measurements correspond to SARS-CoV-2 "N1"
3. "Estimated Input (cp/ml)= theoretical concentration of SARS-CoV-2 heat-inactivated particles, in units of copies/mL (cp/ml), based on dilutions from the concentration reported in the COA, in each of the samples input into the extraction protocol
4. "Replicate" = Technical extraction replicate sample number at each concentration, three replicates are denoted as "Rep 1", "Rep 2", or "Rep 3"
5. "ExtractionRep1" = mean of duplicate PCR technical replicate Cq values with the RT-qPCR protocol; individual replicate Cq data are not shown
6. "Calculated Concentration (cp/ml)" = the calculated viral load concentration of the SARS-CoV-2 N1 gene in copies/mL (cp/ml) from the calibration curves N1 saliva equation provided in the paper methods and the mean Cq reported in the column "ExtractionRep1"

The file Fig_1_panelB_Saliva_Participant_Data.xlsx provides the data of the participant saliva samples run with RT-qPCR and RT-ddPCR protocols for panel B in Figure 1.

1. "SAMPLE ID" = Caltech-internal sample ID code to participant sample.
2. "N1 Cq" = mean of RT-qPCR technical duplicate wells of the SARS-CoV-2 N1 gene target
3. "RT-qPCR N1 Conc. Cp/ml" = the calculated viral load concentration of the SARS-CoV-2 N1 gene in copies/mL (cp/ml) from the calibration curves equation provided in the paper methods and the mean Cq reported in the column "N1 Cq"
4. "ddPCR Concentration in Spectrum (cp/ml)" = the calculated viral load concentration of the SARS-CoV-2 N1 gene quantified with RT-ddPCR and converted to the concentration in the original sample
5. "Geometric Mean" = geometric mean of the concentration (copies/mL) from the RT-qPCR method reported in column "RT-qPCR N1 Conc. Cp/ml" and RT-ddPCR method reported in column "ddPCR Concentration in Spectrum (cp/ml)"

The files Fig_2A_data.xlsx, Fig_2B_data.xlsx, Fig_2C_data.xlsx, Fig_2D_data.xlsx, Fig_2E_data.xlsx, Fig_2F_data.xlsx, and Fig_2G_data.xlsx provides a line listing of information about each sample provided by

each participant. These data are plotted in Figure 2 of the linked paper. For each sample, the following data columns exist:

1. “Participant” = which figure panel these data correspond to, same code is used in the supplemental table for participant demographic data
2. “Days Post-Enrollment” = Collection time relative to enrollment in days. “0.00” is 12am on the day of enrollment
3. “Sample Type” = which sample type “saliva” or “Saliva” or “Nasal Swab” these measurements correspond to. (“saliva” and “Saliva” are equivalent)
4. “Viral Load N1 (copies/mL)” = the calculated viral load concentration of the SARS-CoV-2 N1 gene in units copies/mL from the calibration curves equation provided in the paper methods and the mean Cq reported in the column “RTqPCR (N1) mean Cq”. If results are categorized as “Not Detected” or “Indeterminate” based off the methods, this is reported instead of a calculation of viral load.
5. “Viral Load N2 (copies/mL)” = the calculated viral load concentration of the SARS-CoV-2 N2 gene in units copies/mL from the calibration curves equation provided in the paper methods and the mean Cq reported in the column “RTqPCR (N2) mean Cq”. If results are categorized as “Not Detected” or “Indeterminate” based off the methods, this is reported instead of a calculation of viral load.
6. “RTqPCR (N1) mean Cq” = mean of two PCR technical replicate Cq values of the N1 gene; from the two columns “RTqPCR (N1) Cq Replicate 1” and “RTqPCR (N1) Cq Replicate 2”
7. “RTqPCR (N2) mean Cq” = mean of two PCR technical replicate Cq values of the N2 gene; from the two columns “RTqPCR (N2) Cq Replicate 1” and “RTqPCR (N2) Cq Replicate 2”
8. “RTqPCR (RNaseP) mean Cq” = mean of the two RT-qPCR technical replicate Cq values for the human RNAse P target; from the two columns “RTqPCR (RNaseP) Cq Replicate 1” and “RTqPCR (RNaseP) Cq Replicate 2”
9. “RTqPCR (N1) Cq Replicate 1” = Cq value for the SARS-CoV-2 N1 Gene Target in the first technical replicate well. Values listed as 45 are note measured values, but a proxy for “not detected” by the PCR instrument.
10. “RTqPCR (N1) Cq Replicate 2” = Cq value for the SARS-CoV-2 N1 Gene Target in the second technical replicate well. Values listed as 45 are note measured values, but a proxy for “not detected” by the PCR instrument.
11. “RTqPCR (N2) Cq Replicate 1” = Cq value for the SARS-CoV-2 N2 Gene Target in the first technical replicate well. Values listed as 45 are note measured values, but a proxy for “not detected” by the PCR instrument.
12. “RTqPCR (N2) Cq Replicate 2” = Cq value for the SARS-CoV-2 N2 Gene Target in the second technical replicate well. Values listed as 45 are note measured values, but a proxy for “not detected” by the PCR instrument.
13. “RTqPCR (RNaseP) Cq Replicate 1” = Cq value for the human RNaseP Gene Target in the first technical replicate well. Values listed as 45 are note measured values, but a proxy for “not detected” by the PCR instrument.
14. “RTqPCR (RNaseP) Cq Replicate 2” = Cq value for the human RNaseP Gene Target in the first technical replicate well. Values listed as 45 are note measured values, but a proxy for “not detected” by the PCR instrument.
15. “Cough” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “cough”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “Severe”

16. “Shortness of Breath” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “shortness of breath”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “Severe”
17. “Congestion” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “congestion”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “Severe”
18. “Runny Nose” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “runny nose”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “Severe”
19. “Change in Taste/Smell” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “change in taste/smell”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “Severe”
20. “Sore Throat” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “sore throat”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “Severe”
21. “Nausea” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “nausea”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “Severe”
22. “Vomiting” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “vomiting”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “Severe”
23. “Diarrhea” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “diarrhea”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “Severe”
24. “Fever” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “fever”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “Severe”
25. “Headache” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “headache”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “Severe”
26. “Muscle Aches” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “muscle aches”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “Severe”
27. “Other Symptoms” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for a field left for any participant write-in symptoms

The File Fig_2_SequencingVariant_GISAID_Table.xlsx provides the documentation of the publicly available viral sequence genomes files to determine the variants of each of the samples.

1. "Figure 2 Panel" = which panel in Figure 2 these data correspond to
2. "Variant" = SARS-CoV-2 viral variant identified
3. "Sample Type" = which sample type "Saliva" or "Nasal Swab" these measurements correspond to
4. "GISAID_VirusName" = GISAID viral name identifying SARS-CoV-2 viral consequence genome sequence.
5. "Accession ID" = GISAID accessionID for SARS-CoV-2 viral consequence genome sequence.

The file Fig_3A_data.xlsx describes the percent of positive participants by either high sensitivity saliva (positive here meaning positive result from our assay with high analytical sensitivity) or inferred positivity with a low sensitivity nasal swab sample (based on SARS-CoV-2 *N1* gene viral load > 1.0e5 copies/mL ), relative to the 0.5 Day window of infection (aligned to first positive sample by any timepoint).

1. "Days from First Positive Result in Either Sample Type" = number of days from the first positive result by our high-sensitivity assay (see methods) in either sample type (saliva or nasal swab) for all participants
2. "Total Number of Participants Contributing Saliva Samples at Timepoint (N)" = total number of participants with saliva samples passing quality control evaluations (see methods) for safety and Human *RNaseP* gene Cq threshold at the corresponding aligned timepoint in column "Days from First Positive Result in Either Sample Type". Maximum number of 7, corresponding to the number of participants in the study; each participant has a maximum of one sample per timepoint. Numbers may vary before day 0 as each participant had a variable number of negative test results before first detected SARS-CoV-2 RNA
3. "Number of Participants Positive By Saliva in High Sensitivity Assay (n)" = number of participants with detectable SARS-CoV-2 RNA (see methods) in saliva sample type using an assay with high analytical sensitivity
4. "Percent of Participants Positive by a High Sensitivity Saliva Assay at Timepoint (%)" = percent (100%*n/N) of total participants contributing saliva samples that have detectable SARS-CoV-2 RNA in saliva using an assay with high analytical sensitivity at the corresponding timepoint (column "Number of Participants Positive By Saliva in High Sensitivity Assay (n)" out of the total number of participant saliva samples incolumn "Total Number of Participants Contributing Saliva Samples at Timepoint (N)")
5. "Number of Participants With Viral Load > 1.0e5cp/mL in Saliva, for Inferred Positive by Low Sensitivity Assay (n)" = number of participants with a measured SARS-CoV-2 *N1* gene viral load of great er than 1.0e5 copies/mL (cp/mL) in saliva samples, which would predict a positive result based on a low analytical sensitivity assay with an LOD of SARS-CoV-2 *N1* gene viral load of 1.0e5 copies/mL
6. "Percent of Participants Inferred to be Positive by a Low Sensitivity Saliva Assay at Timepoint (%)" = percent (100%*n/N) of total participants contributing saliva samples that have a viral load of SARS-CoV-2 in saliva that would be detectable using a low sensitivity assay at the corresponding timepoint (column "Number of Participants With Viral Load > 1.0e5cp/mL in Saliva, for Inferred Positive by Low Sensitivity Assay (n)" out of the total number of participant nasal swab samples (column "Total Number of Participants Contributing Saliva Samples at Timepoint (N)")
7. "Total Number of Participants Contributing Nasal Swab Samples at Timepoint (N)" = total number of participants with nasal swab samples passing quality control criteria (see methods) for safety and RNaseP Cq threshold at the corresponding aligned timepoint in column "Days from First Positive Result

in Either Sample Type." Maximum number of 7, corresponding to the number of participants in the study, each participant has a maximum of one sample per timepoint. Numbers may vary before day 0 as each participant had a variable number of negative test results before first detected SARS-CoV-2 RNA

8. "Number of Participants Positive By Nasal Swab in High Sensitivity Assay (n)" = number of participants with detectable SARS-CoV-2 RNA (see methods) in nasal swab sample type using a assay with high analytical sensitivity
9. "Percent of Participants Positive by a High Sensitivity Nasal Swab Assay at Timepoint (%)" = percent (100%*n/N) of total participants contributing nasal swab samples that have detectable SARS-CoV-2 RNA in nasal swabs using an assay with high analytical sensitivity at the corresponding timepoint (column "Number of Participants Positive By Nasal Swab in High Sensitivity Assay (n)" out of the total number of participant saliva samples in column "Total Number of Participants Contributing Nasal Swab Samples at Timepoint (N)")
10. "Number of Participants With Viral Load >1.0e5cp/mL in Nasal Swab, for Inferred Positive by Low Sensitivity Assay (n)" = number of participants with a measured SARS-CoV-2 *N1* gene viral load of greater than 1.0e5 copies/mL (cp/mL) in nasal swab samples, which would predict a positive result based on a low sensitivity assay with an LOD of SARS-CoV-2 *N1* gene viral load of 1.0e5 copies/mL.
11. "Percent of Participants Inferred to be Positive by a Low Sensitivity Nasal Swab Assay at Timepoint (%)" = percent (100%*n/N) of total participants contributing nasal swab samples that have a viral load of SARS-CoV-2 in nasal swabs that would be detectable using a low sensitivity assay at the corresponding timepoint (column "Number of Participants With Viral Load >1.0e5cp/mL in Nasal Swab, for Inferred Positive by Low Sensitivity Assay (n)") out of the total number of participant nasal swab samples (column Total Number of Participants Contributing Nasal Swab Samples at Timepoint (N)")

The file Fig_3B_data.xlsx compares the viral load of SARS-CoV-2 RNA in saliva versus in nasal swabs during different intervals of infection

1. "Participant (Shown in Figure 2 Panel)" = the participant referred to in the listed panel of figure 2
2. "Days of Detectable Infection (Days from First Positive Sample by either Sample Type" = number of days from the first positive result by our high-sensitivity assay in either sample type (saliva or nasal swab)
3. "Saliva Viral Load (cp/mL Saliva)" = viral load of SARS-CoV-2 RNA *N1* gene copies/mL (cp/mL) detected in the saliva sample. If results are categorized as "Not Detected" or "Indeterminate" based off the methods, this is reported instead of a calculation of viral load.
4. "Nasal Swab Viral Load (cp/mL lysis buffer)" = viral load of SARS-CoV-2 RNA *N1* gene copies/mL (cp/mL of lysis buffer) detected in the nasal swab sample. If results are categorized as "Not Detected" or "Indeterminate" based off the methods, this is reported instead of a calculation of viral load.
5. "Window of Infection" = Participants were enrolled and became positive at different times. In order to plot all measurements from the same approximate time point (first detectable viral load), we constructed windows of infection (0 to 2 days, 2 to 4 days, 4 to 6 days, 6 to 8 days, 8 to 12 days, and 12 to 16 days. A window of "Days 0-2" include time points from infection of 0-1.5, as shown in the column "Timepoint from Infection". A window of "Days 2-4" include time points from infection of 2-3.5, as shown in the column "Days of Detectable Infection (Days from First Positive Sample by either Sample Type". A window of "Days 4-6" include time points from infection of 4-5.5, as shown in the column "Timepoint from Infection". A window of "Days 6-8" include time points from infection of 6-7.5, as shown in the column "Timepoint from Infection". A window of "Days 8-12" include time points from infection of 8-11.5, as shown in the column "Timepoint from Infection". A window of "Days 12-16"

include time points from infection of 12-16 as shown in the column "Timepoint from Infection". All times are relative to the first positive result for each participant using our high-sensitivity assays [see Methods] in either sample type [saliva or nasal swab]). The first positive sample for each participant is assigned to time 0. Measurements were then assigned to a window of infection, listed here and shown on Figure 3.

The file Fig_3C_data.xlsx shows the peak viral loads from each participant, in each sample type, as well as how many days post enrollment and how many days from the first positive sample the peak viral load was measured in each sample type

1. "Participant (Shown in Figure 2 Panel)" = the study participant referred to in the listed panel of Figure 2
2. "Peak Saliva Viral Load (cp/mL Saliva)" = the highest measurement of SARS-CoV-2 RNA *N1* gene copies/mL (cp/mL) viral load quantified in saliva for each participant throughout their enrollment period
3. "Days From Enrollment – Saliva" = number of days each participant has been enrolled in the study at the corresponding peak saliva viral load measurement reported in column "Peak Saliva Viral Load (cp/mL Saliva)"
4. "Days from First Positive Saliva Sample" = number of days elapsed from the peak viral load, reported in column "Peak Saliva Viral Load (cp/mL Saliva)," and the first positive result measured in saliva for each participant
5. "Peak Nasal Swab Viral Load (cp/mL lysis buffer)" = the highest measurement of SARS-CoV-2 RNA *N1* gene copies/mL (cp/mL) viral load quantified in nasal swabs for each participant
6. "Days From Enrollment – Nasal Swab" = number of days each participant has been enrolled in the study at the corresponding peak nasal swab viral load measurement reported in column "Peak Nasal Swab Viral Load (cp/mL lysis buffer)"
7. "Days from First Positive Nasal Swab Sample" = number of days elapsed from the peak viral load, reported in column "Peak Nasal Swab Viral Load (cp/mL lysis buffer)," and the first positive result measured in nasal swabs for each participant

The file Fig_3D_data.xlsx shows the percent of saliva or nasal swab samples positive by a high sensitivity test (positive in our assay) or a low sensitivity assay (viral load >1.0e5cp/mL) for the first ten days of infection, separated by whether ANY symptom was reported at the time of sample collection.

1. "Sensitivity" = indicates whether the data in the row is the result of observed positivity either measured by our high-sensitivity assay (when the value is 'High', see Methods) or inferred positivity for a low-sensitivity test if the viral load of the sample was >1.0e5 cp/mL low-sensitivity test (when the value is 'Low')
2. "Symptomatic" = indicates whether the corresponding samples were collected while the participant reported at least one COVID-19 like symptom (see Methods); this column has "TRUE" indicating the row is for samples collected while symptomatic, and "FALSE" indicating the row is for sample collected when no symptoms were reported
3. "Total Saliva Samples" = total number of saliva samples assessed at the corresponding test sensitivity and symptom criteria
4. "Positive Saliva Samples" = number of saliva samples considered positive for SARS-CoV-2 (see methods) at the corresponding test sensitivity and symptom criteria
5. "Percent Positive – Saliva" = percent of saliva samples considered positive for SARS-CoV-2 at the corresponding test sensitivity and symptom criteria; calculated as a ratio of columns "Positive Saliva Samples" and "Total Saliva Samples"

6. “Percent False Negative – Saliva” = the total number of saliva samples (“Total Saliva Samples”) minus the number of positive saliva samples (“Positive Saliva Samples”), divided by the total number of samples (“Total Saliva Samples”) at the corresponding test sensitivity and symptom criteria. Samples in this dataset are from individuals with laboratory-confirmed SARS-CoV-2 infection, so the expected result for all samples would be positive, hence why those that did not result positive are considered falsely negative.
7. “Total Nasal Swab Samples” = total number of nasal swab samples assessed at the corresponding test sensitivity and symptom criteria
8. “Positive Nasal Swab Samples” = number of nasal swab samples considered positive for SARS-CoV-2 (see methods) at the corresponding test sensitivity and symptom criteria
9. “Percent Positive – Nasal Swabs” = percent of nasal swab samples considered positive for SARS-CoV-2 at the corresponding test sensitivity and symptom criteria
10. “Percent False Negative – Nasal Swabs” = the total number of nasal swab samples (“Total Nasal Swab Samples”) minus the number of positive nasal swab samples (“Positive Nasal Swab Samples”), all divided by the total number of samples (“Total Nasal Swab Samples”) at the corresponding test sensitivity and symptom criteria. Samples in this dataset are from individuals with laboratory-confirmed SARS-CoV-2 infection, so the expected result for all samples would be positive, hence why those that did not result positive are considered falsely negative.

The file SI_Fig_1A_saliva_LOD.csvprovides the data for the saliva limit of detection in Figure S2 panel A; see methods in the paper for details

1. “rep1” = Cq value for the Gene Target in the first technical replicate well
2. “rep2” = Cq value for the Gene Target in the second technical replicate well
3. “Marker” = Gene target (SARS-CoV-2 “N1”, SARS-CoV-2 “N2”, or human “RNaseP”)
4. “Sample_Name” = Replicate samples for the 1000 copies/mL SARS-CoV-2 heat-inactivated particles replicates (rep_1 – rep_20) and 0 copies/mL SARS-CoV-2 heat-inactivated particles replicates (ExNeg1 – ExNeg3)
5. “Avg” = mean of the two PCR technical replicate Cq values of the Gene Target; from the two columns “rep1” and “rep2”
6. “Std Dev” = standard deviation of the two PCR technical replicate Cq values of the Gene Target; from the two columns “rep1” and “rep2”

The file SI_Fig_1B_swab_LOD.xlsx provides the data for the nasal swab limit of detection in Figure S2 panel B; see methods in the paper for details.

1. “Sample_Name” = Replicate samples for the 1000 copies/mL SARS-CoV-2 heat-inactivated particles replicates (1K_1 – 1K_20), 0 copies/mL SARS-CoV-2 heat-inactivated particles replicates (ExNeg1 – ExNeg3), RTqPCR positive control “qPCRpos”, RTqPCR negative control “qPCRNeg”, and Extraction positive control at 7,500 copies/mL SARS-CoV-2 heat-inactivated particles “ExPos”
2. “N1 rep 1” = Cq value for the SARS-CoV-2 N1 Gene Target in the first technical replicate well
3. “N1 rep 2” = Cq value for the SARS-CoV-2 N1 Gene Target in the second technical replicate well
4. “N1_AVG” = mean of two PCR technical replicate Cq values of the N1 gene; from the two columns “N1 rep 1” and “N1 rep 2”
5. “N2 rep 1” = Cq value for the SARS-CoV-2 N2 Gene Target in the first technical replicate well
6. “N2 rep 2” = Cq value for the SARS-CoV-2 N2 Gene Target in the second technical replicate well
7. “N2_AVG” = mean of two PCR technical replicate Cq values of the N2 gene; from the two columns “N2 rep 1” and “N2 rep 2”

8. “RNASEP rep 1” = Cq value for the human RNaseP Gene Target in the first technical replicate well
9. “RNASEP rep 2” = Cq value for the human RNaseP Gene Target in the second technical replicate well
10. “RNASEP_AVG” = mean of two PCR technical replicate Cq values of the human RNaseP gene; from the two columns “RNASEP rep 1” and “RNASEP rep 2”

The file SI_Fig_2_CalibrationCurve_data_plots.csv provides the RTqPCR data for Figure S1, the Calibration curve of SARS-CoV-2 inactivated particles to establish viral load conversion equations; see methods in the paper for details

1. “Sample Type” = which sample type “Saliva” or “Nasal Swab” these protocols (RNA extraction protocols) correspond to
2. “theoretical input based on COA (cp/mL)” = theoretical concentration of SARS-CoV-2 heat-inactivated particles, in units of copies/mL (cp/mL), based on dilutions from the concentration reported in the COA, in each of the samples input into the extraction protocol
3. “Gene” = Gene target of either SARS-CoV-2 “N1” or SARS-CoV-2 “N2”
4. “Replicate” = Technical extraction replicate sample number at each concentration, three replicates are denoted as “Rep 1”, “Rep 2”, or “Rep 3”
5. “Cq” = mean of triplicate (swab) or duplicate (saliva) PCR technical replicate Cq values with the RT-qPCR protocol; individual replicate Cq data are not shown

The file SI_Fig_3A_data.xlsx provides the data for the distribution of extraction positive control viral load values in Figure S3 panel A; see methods in the paper for details. For each entry, the following data columns exist:

1. “Extraction Batch ID” = Caltech-internal unique identifier for each unique extraction batch performed
2. “Sample Type” = the type of contrived, positive-control, sample “Saliva” or “Nasal Swab” extracted in the corresponding extraction batch. Positive controls were prepared as described in the supplementary methods.
3. “Mean SARS-CoV-2 N1 Cq Value” = the mean Cq value of the SARS-CoV-2 N1 gene target from duplicate RT-qPCR reactions performed on the extraction positive control sample included in each extraction batch

The files SI_Fig_3B_data_saliva.xlsx provides the data used to calculate the RNA stability of participant saliva samples by the half-life equation, and the data used to assemble the ECDF; see methods in the paper for details. These data are plotted in Supplementary Figure 3 panel B of the linked paper. For each entry, the following data columns exist:

1. “Sample_type” = which sample type “saliva” or “swab” these measurements correspond to
2. “Sample ID” = Caltech-internal sample ID code to participant sample.
3. “Date Extracted” = longhand calendar date (%Y%m%d) in which the sample was extracted
4. “Date Collected” = longhand calendar date (%Y%m%d) in which the sample was self-collected by the participant
5. “N1 Replicate 1” = Cq value for the SARS-CoV-2 N1 Gene Target in the first technical replicate well. Values listed as 45 or are left blank are not measured values, but a proxy for “not detected” by the PCR instrument and indicate that this well not detected.
6. “N1 Replicate 2” = Cq value for the SARS-CoV-2 N2 Gene Target in the first technical replicate well. Values listed as 45 or are left blank are not measured values, but a proxy for “not detected” by the PCR instrument and indicate that this well not detected.

7. “N1_mean” = mean of two PCR technical replicate Cq values of the SARS-CoV-2 N1 gene target; from the two columns “N1 Replicate 1” and “N1 Replicate 2”
8. “VL_cpmL” = the calculated viral load concentration of the SARS-CoV-2 N1 gene in units copies/mL from the calibration curves equation provided in the paper methods and the mean Cq reported in the column “N1_mean”.
9. “RNaseP” = mean of the two RT-qPCR technical replicate Cq values for the human RNAseP gene target
10. “Extraction Protocol” = denotes which extraction protocol was used for this replicate. “KF_screen” refers to a KingFisher extraction protocol for screening positivity of samples, varying slightly from the time series protocols described in the paper. “KF_tseries” refers to the KingFisher extraction protocols
11. “Extraction Batch” = Caltech-internal unique identifier for each unique extraction batch performed
12. “Replicates of Identical protocol” = number denoting how many replicates of the identical protocol “KF_tseries” or “KF_screen” was performed on this sample
13. “ex_replicates” = which replicate, of any extraction protocol “KF_tseries” or “KF_screen”
14. “days_4C (x)” = denotes the number of days between when the samples were self-collected form the participants and stored at 4°C in the RNA stabilization buffer prior to extraction
15. “Half-Life (days)” = half-life, representing 1 Cq increase (2-fold decrease) in detected *N1* gene RNA, calculated as described in the Supplementary methods equation 5.

SI_Fig_3B_data_swab.csv provides the data used to calculate the RNA stability of participant nasal swab samples by the half-life equation, and the data used to assemble the ECDF; see methods in the paper for details. These data are plotted in Supplementary Figure 3 panel B of the linked paper. For each sample, the following data columns exist:

1. “Sample_type” = which sample type “saliva” or “swab” these measurements correspond to
2. “Sample ID” = Caltech-internal sample ID code to participant sample.
3. “Date Archived” = longhand calendar date in which the physical nasal swab was removed from the RNA stabilization solution. The sample, material liberated from the swab, was returned to storage at 4°C in the RNA stabilization solution without the swab.
4. “Date Extracted” = longhand calendar date (%Y%m%d) in which the sample was extracted
5. “Date Collected” = longhand calendar date (%Y%m%d) in which the sample was self-collected by the participant
6. “N1 Replicate 1” = Cq value for the SARS-CoV-2 N1 Gene Target in the first technical replicate well. Values listed as 45 or are left blank are not measured values, but a proxy for “not detected” by the PCR instrument and indicate that this well not detected.
7. “N1 Replicate 2” = Cq value for the SARS-CoV-2 N2 Gene Target in the first technical replicate well. Values listed as 45 or are left blank are not measured values, but a proxy for “not detected” by the PCR instrument and indicate that this well not detected.
8. “N1_mean” = mean of two PCR technical replicate Cq values of the SARS-CoV-2 N1 gene target; from the two columns “N1 Replicate 1” and “N1 Replicate 2”
9. “VL_cpmL” = the calculated viral load concentration of the SARS-CoV-2 N1 gene in units copies/mL from the calibration curves equation provided in the paper methods and the mean Cq reported in the column “N1_mean”.
10. “RNaseP” = mean of the two RT-qPCR technical replicate Cq values for the human RNAseP gene target
11. “Extraction Protocol” = denotes which extraction protocol was used for this replicate. “KF_screen” refers to a KingFisher extraction protocol for screening positivity of samples, varying slightly from the time series protocols described in the paper. “KF_tseries” refers to the KingFisher extraction protocols
12. “Extraction Batch” = Caltech-internal unique identifier for each unique extraction batch performed

13. “ex_replicates” = which replicate, of any extraction protocol “KF_tseries” or “KF_screen”
14. “days_4C (x)” = denotes the number of days between when the samples were self-collected form the participants and stored at 4°C in the RNA stabilization buffer prior to extraction
15. “day_archive” = denotes the number of days the sample was stored at 4°C with the nasal swab, the sample was stored at 4°C without the nasal swab for the remaining duration of time
16. “Half-Life (days)” = half-life, representing 1 Cq increase (2-fold decrease) in detected *N1* gene RNA, calculated as described in the Supplementary methods equation 5.

The files SI_Fig_4A_data.xlsx, SI_Fig_4B_data.xlsx, SI_Fig_4C_data.xlsx, SI_Fig_4D_data.xlsx, SI_Fig_4E_data.xlsx, SI_Fig_4F_data.xlsx, SI_Fig_4G_data.xlsx, provides a line listing of information about each sample provided by each participant. These data are plotted in Supplementary Figure 4 of the linked paper. For each sample, the following data columns exist:

1. “Participant” = which figure panel these data correspond to, same code is used in the supplemental table for participant demographic data
2. “Days Post-Enrollment” = Collection time relative to enrollment in days. “0.00” is 12am on the day of enrollment
3. “Sample Type” = which sample type “saliva” or “Nasal Swab” these measurements correspond to. which sample type “saliva” or “Saliva” or “Nasal Swab” these measurements correspond to. (“saliva” and “Saliva” are equivalent)
4. “Viral Load N1 (copies/mL)” = the calculated viral load concentration of the SARS-CoV-2 N1 gene in units copies/mL from the calibration curves equation provided in the paper methods and the mean Cq reported in the column RTqPCR (N1) mean Cq”. If results are categorized as “Not Detected” or “Indeterminate” based off the methods, this is reported instead of a calculation of viral load.
5. “Viral Load N2 (copies/mL)” = the calculated viral load concentration of the SARS-CoV-2 N2 gene in units copies/mL from the calibration curves equation provided in the paper methods and the mean Cq reported in the column “RTqPCR (N2) mean Cq”. If results are categorized as “Not Detected” or “Indeterminate” based off the methods, this is reported instead of a calculation of viral load.
6. “Adjusted Viral Load N1 (copies/mL)” = Viral load of the SARS-CoV-2 *N1* gene adjusted for possible RNA degradation, according to equation 6 in the Supplemental methods. Calculations use the columns “Viral Load N1 (copies/mL)”, the median half-life for the sample type (“saliva” or “swab”) as calculated in Fig S3, and “Storage time at 4°C [days]”.
7. “Adjusted Viral Load N2 (copies/mL)” = Viral load of the SARS-CoV-2 *N1* gene adjusted for possible RNA degradation, according to equation 6 in the Supplemental methods. Calculations use the columns “Viral Load N2 (copies/mL)” , the median half-life for the sample type (“saliva” or “swab”) as calculated in Fig S3, and “Storage time at 4°C [days]”.
8. “Storage time at 4°C [days]” = denotes the number of days between when the samples were self-collected form the participants and stored at 4°C in the RNA stabilization buffer prior to extraction. Date only was recorded for the extraction (and not time of day) so anything with a very fast turnaround that calculated a negative time at 4°C were set to 0 days because the samples were extracted right after arrival in the lab.
9. “RTqPCR (N1) mean Cq” = mean of two PCR technical replicate Cq values of the N1 gene; from the two columns “RTqPCR (N1) Cq Replicate 1” and “RTqPCR (N1) Cq Replicate 2”
10. “RTqPCR (N2) mean Cq” = mean of two PCR technical replicate Cq values of the N2 gene; from the two columns “RTqPCR (N2) Cq Replicate 1” and “RTqPCR (N2) Cq Replicate 2”

11. “RTqPCR (RNaseP) mean Cq” = mean of the two RT-qPCR technical replicate Cq values for the human RNAse P target; from the two columns “RTqPCR (RNaseP) Cq Replicate 1” and “RTqPCR (RNaseP) Cq Replicate 2”
12. “RTqPCR (N1) Cq Replicate 1” = Cq value for the SARS-CoV-2 N1 Gene Target in the first technical replicate well. Values listed as 45 are note measured values, but a proxy for “not detected” by the PCR instrument.
13. “RTqPCR (N1) Cq Replicate 2” = Cq value for the SARS-CoV-2 N1 Gene Target in the second technical replicate well. Values listed as 45 are note measured values, but a proxy for “not detected” by the PCR instrument.
14. “RTqPCR (N2) Cq Replicate 1” = Cq value for the SARS-CoV-2 N2 Gene Target in the first technical replicate well. Values listed as 45 are note measured values, but a proxy for “not detected” by the PCR instrument.
15. “RTqPCR (N2) Cq Replicate 2” = Cq value for the SARS-CoV-2 N2 Gene Target in the second technical replicate well. Values listed as 45 are note measured values, but a proxy for “not detected” by the PCR instrument.
16. “RTqPCR (RNaseP) Cq Replicate 1” = Cq value for the human RNaseP Gene Target in the first technical replicate well. Values listed as 45 are note measured values, but a proxy for “not detected” by the PCR instrument.
17. “RTqPCR (RNaseP) Cq Replicate 2” = Cq value for the human RNaseP Gene Target in the first technical replicate well. Values listed as 45 are note measured values, but a proxy for “not detected” by the PCR instrument.
18. “Cough” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “cough”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “severe”
19. “Shortness of Breath” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “shortness of breath”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “severe”
20. “Congestion” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “congestion”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “severe”
21. “Runny Nose” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “runny nose”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “severe”
22. “Change in Taste/Smell” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “change in taste/smell”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “severe”
23. “Sore Throat” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “sore throat”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “severe”

24. “Nausea” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “nausea”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “severe”
25. “Vomiting” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “vomiting”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “severe”
26. “Diarrhea” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “diarrhea”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “severe”
27. “Fever” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “fever”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “severe”
28. “Headache” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “headache”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “severe”
29. “Muscle Aches” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for the symptom “muscle aches”. Early versions of the survey has “TRUE” and “FALSE” categorization; updated surveys have “I have not experienced this symptom”, “Mild” or “severe”
30. “Other Symptoms” = Participant symptom survey results for the symptom tracker card filled out at the time of collection of this sample for a field left for any participant write-in symptoms