Stratification of viral shedding patterns in saliva of COVID-19 patients

AUTHORS:

Hyeongki Park[1,2], Raiki Yoshimura[1], Shoya Iwanami[1], Kwang Su Kim[1,3,4], Keisuke Ejima[5,6], Naotoshi Nakamura[1], Kazuyuki Aihara[7], Yoshitsugu Miyazaki[8], Takashi Umeyama[8], Ken Miyazawa[8], Takeshi Morita[9], Koichi Watashi[9], Christopher B. Brooke[10,11], Ruian Ke[12], Shingo Iwami[1,7, 13,14,15,16,17,†,*] and Taiga Miyazaki[18,†,*]

AFFILIATIONS:

[1]interdisciplinary Biology Laboratory (iBLab), Division of Natural Science, Graduate School of Science, Nagoya University, Nagoya, Japan. [2]School of Biomedical Convergence Engineering, Pusan National University, Yangsan, South Korea. [3]Department of Science System Simulation, Pukyong National University, Busan, South Korea. [4]Department of Mathematics, Pusan National University, Busan, South Korea. [5]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. [6]The Tokyo Foundation for Policy Research, Tokyo, Japan. [7]International Research Center for Neurointelligence, The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Tokyo, Japan. [8]Department of Chemotherapy and Mycoses, National Institute of Infectious Diseases, Tokyo, Japan. [9]Research Center for Drug and Vaccine Development, National Institute of Infectious Diseases, Tokyo, Japan. [10]Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [11]Department of Statistics, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [12]Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM, USA. [13]Institute of Mathematics for Industry, Kyushu University, Fukuoka, Japan. [14]Institute for the Advanced Study of Human Biology (ASHBi), Kyoto University, Kyoto, Japan. [15]Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS), RIKEN, Saitama, Japan. [16]NEXT-Ganken Program, Japanese Foundation for Cancer Research (JFCR), Tokyo, Japan. [17]Science Groove Inc., Fukuoka, Japan. [18]Division of Respirology, Rheumatology, Infectious Diseases, and Neurology, Department of Internal Medicine, Faculty of Medicine, University of Miyazaki, Miyazaki, Japan.

†These authors contributed equally to this study.

[*]To whom correspondence may be addressed.

Email: iwami.iblab@bio.nagoya-u.ac.jp (S.I.) and taiga_miyazaki@med.miyazaki-u.ac.jp (T.M.).

Abstract

Living with COVID-19 requires continued vigilance against the spread and emergence of variants of concern (VOCs). Rapid and accurate saliva diagnostic testing, alongside basic public health responses, is a viable option contributing to effective transmission control. Nevertheless, our knowledge regarding the dynamics of SARS-CoV-2 infection in saliva is not as advanced as our understanding of the respiratory tract. Here we analyzed longitudinal viral load data of SARS-CoV-2 in saliva samples from 144 patients with mild COVID-19 (a combination of our collected data and published data). Using a mathematical model, we quantified individual-level viral dynamics and stratified them into three groups using a clustering approach. Notably, the three groups exhibited distinct differences viral RNA detection durations: 11.5 days (95% CI: 10.6 to 12.4), 17.4 days (16.6 to 18.2), and 30.0 days (28.1 to 31.8), respectively. Surprisingly, this stratified grouping remained unexplained despite our analysis of 47 types of clinical data, including basic demographic information, clinical symptoms, results of blood tests, and vital signs. Additionally, we quantified the expression levels of 92 micro-RNAs in a subset of saliva samples, but these also failed to explain the observed stratification, although the mir-1846 level may have been weakly correlated with peak viral load. Our study provides insights into SARS-CoV-2 infection dynamics in saliva, highlighting the challenges in predicting the duration of viral RNA detection without indicators that directly reflect an individual's immune response, such as antibody induction. Given the significant individual heterogeneity in the kinetics of saliva viral shedding, identifying biomarker(s) for viral shedding patterns will be crucial for improving public health interventions in the era of living with COVID-19.

Main Text

Coronavirus disease 2019 (COVID-19) vaccinations, which are effective in preventing infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and severe COVID-19 illness, have enabled the gradual and safe removal of COVID-19 restrictions on everyday life over the past year. However, we still face the emergence of variants of concerns (VOCs), which is a major worry in the era of "living with COVID-19". As a result, to prevent major outbreaks, basic public health responses such as testing, isolation, and quarantine are demanded and important. In particular, rapid and accurate diagnostic tests are essential for controlling ongoing transmission. Salivary diagnostic testing is a convenient tool for early and efficient diagnosis of COVID-19 because it is easy for health care professionals and patients to administer[1,2].

The oral cavity is an important target for SARS-CoV-2[3], and viral particles in the lower and upper respiratory tract can reach the oral cavity through liquid droplets[1]. Saliva droplets are thus a potential route of SARS-CoV-2 transmission[1,3]. Although SARS-CoV-2 infection dynamics within the upper respiratory tract are well characterized thanks to data from oropharyngeal and nasopharyngeal swabs[4-9], infection dynamics in saliva are poorly understood. One potential explanation is the practical challenges inherent in using saliva as a specimen type. Handling steps such as tube opening, pipetting, and vortexing can generate infectious aerosols, requiring strict biosafety precautions (e.g., work in a biosafety cabinet). Also, salivary RNases and the abundant oral microbiota may degrade viral RNA or increase background noise, particularly in low viral load samples.

Recent reports[3,5] have suggested that SARS-CoV-2 infection dynamics differ qualitatively across tissues. Salivary gland tissue (parotid, submandibular, and sublingual) and other tissues are highly compartmentalized, and saliva represents the corresponding biofluid of the oral compartment. However, although individual-level heterogeneity in virus dynamics (especially individual infectiousness) has been evaluated, it remains unknown how individual-level viral shedding patterns in saliva are stratified, and which factor(s) determine the

patterns[5,10]. Since saliva is a route of direct viral transmission, salivary viral load provides particularly valuable information for assessing a patient's immediate infectiousness compared with measurements from other anatomical sites. Accordingly, to shape a country's early response to future VOCs, such as through isolation and screening guidelines based on salivary diagnostic testing, it is critical to address these points with the use of a high quality and quantity of saliva specimens annotated with basic clinical patient data.

So far, our ability to understand and characterize whole SARS-CoV-2 infection dynamics has been hindered by several limitations of the clinical data that made it impossible to capture either the early or the late phase of infection or to annotate individual-level clinical information. To overcome these limitations, here we quantified and stratified longitudinal virus dynamics in saliva samples from 144 mildly symptomatic participants of two different but complementary cohorts[5,11,12]. We successfully identified three groups with distinct viral shedding patterns. Notably, the groups showed clear differences in the duration of viral RNA detection (the mean durations were 11.5 days, 17.4 days, and 30.0 days, respectively). These findings imply a large inter-participant heterogeneity in virus infection dynamics. However, when we analyzed a total of 47 variables, including basic demographic information, daily clinical symptoms, results of blood tests, and vital signs, none explained the stratified grouping.

We also retrospectively explored whether salivary micro-RNAs were associated with the stratification by using stored residual saliva specimens. Micro-RNAs are non-coding RNAs that regulate numerous cellular processes by modulating protein levels through direct binding to mRNA (coding-RNA), thereby influencing translation efficiency or mRNA abundance. Various microRNAs are recognized for their impact on the viral replication ability and immune response in viruses like EBV, HCV, and HIV etc[13]. Many studies are underway to investigate micro-RNAs as potential targets for virus diagnosis and treatment. However, the relationship between micro-RNAs and the patterns of virus shedding of the SARS-CoV-2 virus in body remains unknown. We quantified the expression levels of 92 micro-RNAs and found that no micro-RNA significantly explained the stratified groups, although the mir-1846 level may have been weakly

correlated with the peak viral load. Our findings provide important insights into the complexities of viral shedding patterns in saliva and suggest that predicting the heterogeneity of viral dynamics using basic clinical and micro-RNA data may be a challenging task. These insights are critical for developing accurate diagnostic tools, effective treatments, and prevention strategies for COVID-19.

Results

Description of cohort and study design

We used longitudinal saliva viral load data obtained from cohorts of the nelfinavir (NFV) clinical trial (jRCT2071200023[11,12]) and the University of Illinois at Urbana-Champaign[5]. All infections were either mild or asymptomatic. All participants in these cohorts reported through surveys (self-declaration) that they had never been previously infected with SARS-CoV-2, and none were vaccinated against SARS-CoV-2 at the time of enrolment (see **Methods** for details, including the measurement dates). Of 182 participants from these two studies, 144 symptomatic participants, excluding 22 asymptomatic participants and 16 participants with incomplete observational data, were considered in our analysis (**Fig 1A**). In addition, we annotated the saliva viral load results of 90 participants from the NFV clinical trial with their sex, age, daily symptoms, blood test results, and vital signs. These clinical data were collected longitudinally for each patient (**Fig 1B**). While all patients had clinical measurements at the enrollment time point (a mean of 2.9 days after symptom onset), the frequency of subsequent measurements varied among individuals (**Fig 1B**). Therefore, our analysis focused on the clinical data obtained at enrollment, which were consistently available across all patients. This approach is also appropriate given the study's aim of identifying early predictors of viral dynamics (see later).

Table 1 and **Supplementary Table 1** present the mean values of clinical information and daily symptom data, respectively. The clinical data described in **Table 1** were collected only for the NFV cohort. Therefore, we first used both cohorts to estimate viral dynamics, but in the subsequent analyses examining the relationship between the clinical data and estimated viral dynamics, we relied only on the NFV cohort. Because daily symptom data described in **Supplementary Table 1** were collected for both cohorts, both cohorts were included in the analyses relating daily symptom data to viral dynamics.

The data from the two studies complemented one another: the longitudinal data from the University of Illinois contained data on the early phase of infection (before symptom onset), while the data from the NFV clinical trial contained data on the late phase of viral RNA load (more than

14 days after symptom onset) (**Fig 1C**). Viral load data were collected an average of 15.2 times per participant (SD, 5.7). Pre-symptom onset measurements averaged 0 (SD, 0) in the NFV cohort and 3 (SD, 1) in the Illinois cohort. At 14 or more days after symptom onset, the corresponding averages were 5.5 (SD, 6.1) and 0.1 (SD, 0.3), respectively. The time-series viral load in the saliva samples for all individual used in the analysis are plotted in **Fig 1D**. We also analyzed 60 stored saliva specimens from 30 participants of the NFV clinical trial for micro-RNA analysis.

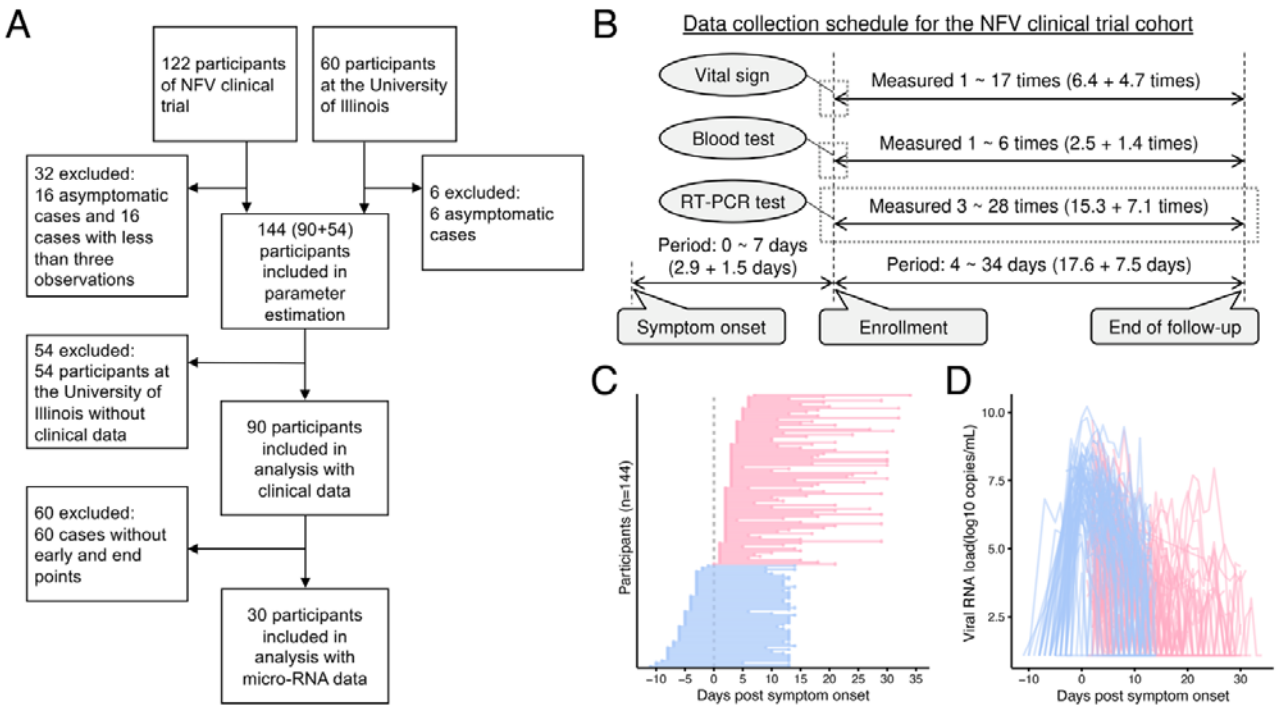


Figure 1 | Characteristics of cohorts from the NFV clinical trial (jRCT2071200023[11,12]) and the University of Illinois[5]: (A) Flowchart of cohorts from the NFV clinical trial and the University of Illinois along with the number of participants and inclusion criteria for our analysis are described. **(B)** Data collection schedule of viral load, blood test, and vital signs from participants in the NFV clinical trial is described. The grey box highlights the measurement dates for the observations included in the analysis. **(C)** and **(D)** show, for each participant (N=144 participants, 2191 samples), the timeline of sample collection and the captured SARS-CoV-2 viral RNA load for saliva RT-qPCR, respectively. The red and blue colors indicate samples for cohorts from the NFV clinical trial and the University of Illinois, respectively.

Table 1 | Clinical data of the overall cohort from the NFV clinical trial and in groups stratified by longitudinal virus dynamics

Clinical data	Unit	Group1 (N=33)	Group2 (N=37)	Group3 (N=20)	Overall (N=90)
Basic demographic information					
Age	years	47 (13.9)*	40.1 (12.6)	43.4 (13.7)	43.7 (13.6)
Sex***	-	46%	67%	70%	59%
Vital signs					
Systolic blood pressure	$mmHg$	116 (16.3)	119.8 (14.4)	122.2 (20.3)	118.8 (16.6)
Diastolic blood pressure	$mmHg$	77.5 (14.3)	75.9 (10.9)	78.5 (14.8)	77.1 (13.2)
Pulse rate	bpm	83.5 (16.9)	86.9 (14.3)	79.1 (13.1)	83.8 (15.3)
SpO_2	%	97.5 (1)	97.7 (0.8)	97.6 (0.9)	97.6 (0.9)
Respiratory rate	bpm	17 (3.9)	16.5 (3.4)	16.8 (2.4)	16.8 (3.4)
Blood test results					
White blood cell count	$10^9/L$	5.1 [30.2]**	5.3 [31.9]	5.9 [34]	5.3 [32.8]
Neutrophil	%	62.4 (10.3)	60 (11.1)	59.8 (11.4)	61 (10.8)
Eosinophil	%	1 (1.8)	0.9 (1.5)	1 (1.2)	1 (1.6)
Basophil	%	0.3 (0.2)	0.4 (0.3)	0.4 (0.3)	0.3 (0.3)
Lymphocytes	%	26.1 (8.4)	28.1 (10)	28 (9.2)	27.3 (9.1)
Monocyte	%	10.1 (3.7)	10.6 (3.5)	10.8 (4.2)	10.4 (3.7)
Red blood cell count	$10^{12}/L$	5 [450.3]	5.5 [464.1]	5.4 [450.9]	5.3 [453]
Amount of hemoglobin	g/dL	14 (1.4)	15.1 (1.1)	14.8 (1.3)	14.6 (1.4)
Hematocrit	%	41.6 (3.6)	44.1 (3.2)	43.8 (3.6)	43 (3.6)
Platelet count	$10^9/L$	121.6 (88.7)	127.8 (98)	136.6 (102.4)	127.2 (94.4)
CRP	mg/dL	1.3 (2.6)	1.3 (1.9)	1.1 (1.8)	1.3 (2.2)
Protein	g/dL	7.4 (0.4)	7.4 (0.4)	7.4 (0.3)	7.4 (0.4)
Albumin	g/dL	4.2 (0.3)	4.3 (0.3)	4.4 (0.3)	4.3 (0.3)
ALT	U/L	20.2 (11.9)	29.9 (23.9)	31.3 (22)	26.2 (19.8)
AST	U/L	21.9 (7.8)	28.3 (14)	28.7 (13.4)	25.8 (12)
γ-GTP	U/L	30.9 (25.6)	42 (41.8)	47.8 (35)	38.7 (34.7)
ALP	U/L	69.9 (20.9)	72 (20.3)	72.8 (20.6)	71.3 (20.4)
LDH	U/L	172.1 (34.1)	184.1 (43.8)	186.6 (29.5)	179.7 (37.2)
Bilirubin	mg/dL	0.5 (0.2)	0.5 (0.2)	0.6 (0.2)	0.5 (0.2)
CK	U/L	85.2 (48.2)	126.7 (146.5)	118.2 (133)	107.7 (113.1)
Na	mEq/L	139.3 (2.9)	139.9 (2.2)	139.8 (2.8)	139.6 (2.6)
K	mEq/L	3.9 (0.3)	3.9 (0.3)	3.9 (0.2)	3.9 (0.3)
Cl	mEq/L	102.6 (3.3)	102.7 (3.6)	102.8 (2.7)	102.6 (3.3)
BUN	mg/dL	11.4 (3)	11.3 (2.5)	12.5 (3.1)	11.6 (2.9)
CRE	mg/dL	0.8 (0.2)	0.8 (0.2)	0.9 (0.2)	0.8 (0.2)
Blood glucose	mg/dL	105.9 (19.9)	102.4 (17.1)	109 (15.7)	105.3 (18)
HbA1c	%	5.6 (0.3)	5.5 (0.4)	5.6 (0.3)	5.6 (0.3)
Procalcitonin	$\mu g/L$	0.1 (0.2)	0.1 (0.1)	0.1 (0.04)	0.1 (0.1)
Fibrinogen	mg/dL	371 (94.8)	365.3 (101.1)	349. 2(73)	364.1 (92.3)
PT	%	103.1 (20.5)	105.5 (17.5)	109.3 (15.5)	105.4 (18.4)
APTT	sec	31.4 (3.7)	31.1 (5.3)	30.7 (3.2)	31.1 (4.2)
PT-INR	-	1 (0.1)	1 (0.1)	1 (0.1)	1 (0.1)
D-dimer	mg/L	0.8 (1.3)	0.5 (0.2)	0.5 (0.1)	0.6 (0.8)

* Mean (standard deviation) and **median [Interquartile range]. ***Sex was reported as the proportion of males.

Quantifying and stratifying SARS-CoV-2 infection dynamics in saliva

We employed a previously developed mathematical model describing SARS-CoV-2 infection dynamics [i.e., Eqs.(1-2)] to evaluate interparticipant heterogeneity (details are provided in **Methods**), and reconstructed the best-fit virus dynamics in saliva of 144 symptomatic participants (**Supplementary Fig 1** and **Supplementary Table 2**). We also applied a mechanistically more realistic mathematical model Eqs.(3-6) that we had previously used with the Illinois dataset[5] (**Supplementary Table 3**). Both models showed stable results in the visual predictive checks (VPC) and convergence assessments and adequately described the data (**Supplementary Fig 2** and **Supplementary Fig 3**). The fit with Eqs.(3-6) yielded a lower Akaike Information Criterion (AIC) of 6,720, compared with 6,834 obtained with Eqs.(1-2). Nevertheless, the reconstructed viral dynamics in the NFV clinical trial dataset were highly similar between the two models, whereas somewhat larger differences were observed for the Illinois dataset (**Supplementary Fig 4**). This may reflect the inclusion of late- and end-phase viral load measurements in the NFV trial, which were not available in the Illinois dataset. It should be noted that the primary aim of applying mathematical modeling in this study was not to dissect viral mechanisms in detail, but to reconstruct viral dynamics for stratification purposes. Accordingly, we adopted the simpler model [i.e., Eqs.(1-2)], which yields comparable dynamics while being more practical and sufficient for reconstructing viral load in this context. This approach also avoids the additional complexity and parameter assumptions required by Eqs.(3-6), as described in [5].

Next, to stratify the time-course pattern of viral shedding, we first applied unsupervised random forest clustering to the individual “reconstructed” virus dynamics of 144 participants (e.g., **Supplementary Fig 1**). However, this analysis failed to divide the time-course pattern into different clusters (data not shown). To overcome this problem, we quantified the peak, duration, up-slope (i.e., growth rate), and down-slope (i.e., decay rate) of the reconstructed dynamics as “features” of the virus dynamics (**Supplementary Table 4**). Interestingly, the unsupervised random forest clustering based on these features identified 3 groups (i.e., G1: N=46, G2: N=61,

and G3: N=37) in which the time-course patterns were clearly discriminated. This finding suggested that there is a heterogeneity of virus infection dynamics in saliva (see **Methods**). **Fig 2A** indicates a two-dimensional Uniform Manifold Approximation and Projection (UMAP) embedding of the three stratified groups. Using a different color for each group (gray for G1, magenta for G2, and blue for G3), we also plotted the estimated individual viral load (**Fig 2B**), and the highlighted time-course pattern of each group by the Partial Least-Squares Discriminant Analysis (PLS-DA) (**Fig 2C**). In addition, we evaluated the impact of statistical uncertainty in the estimated viral dynamics on the stratification by comparing the uncertainty-adjusted and original distance matrices using the Mantel test (see **Methods**). The two matrices were strongly correlated (Mantel $r = 0.72$, $p < 0.001$), indicating that the stratification is robust to parameter uncertainty.

The distributions of the four features are described in **Fig 2D**; a statistically significant between-group difference was found in the duration of viral RNA detection. The mean durations were 11.5 days (95% CI: 10.6 to 12.4), 17.4 days (16.6 to 18.2), and 30.0 days (28.1 to 31.8) for G1, G2, and G3, respectively. In our previous report[10], we consistently confirmed that there were at least 3 groups showing different duration of viral RNA detection in upper respiratory specimens.

Because of previous work concluding that there are no significant differences in virus infection dynamics between NFV-treated and untreated participants[12], we analyzed all data together regardless of treatment (see **Methods**). To further confirm whether NFV affects the stratification of the time-course pattern of viral shedding, we compared the number of individuals belonging to each group (i.e., G1, G2 and G3) between NFV-treated and untreated participants (including the members of the University of Illinois cohort) and found no trend for the stratification ($p=0.784$ by the Fisher's exact test: **Fig 2E**).

Another possibility that explains the different viral duration observed here may be a difference in VOC genotypes. To test this, we used data from 55 participants in our NFV clinical trial who had been characterized according to which VOCs [i.e., B.1.1.7 (Alpha), B.1.672.2 or

AY.29 (Delta), and other variants] they had been infected with (see **Methods**), in addition to data from 54 participants of the University of Illinois cohort. However, we observed no trend in the number of individuals belonging to each group among the VOCs (p=0.728 by the Fisher's exact test: **Fig 2F**), which is consistent with the conclusion in Ke et al.[5].

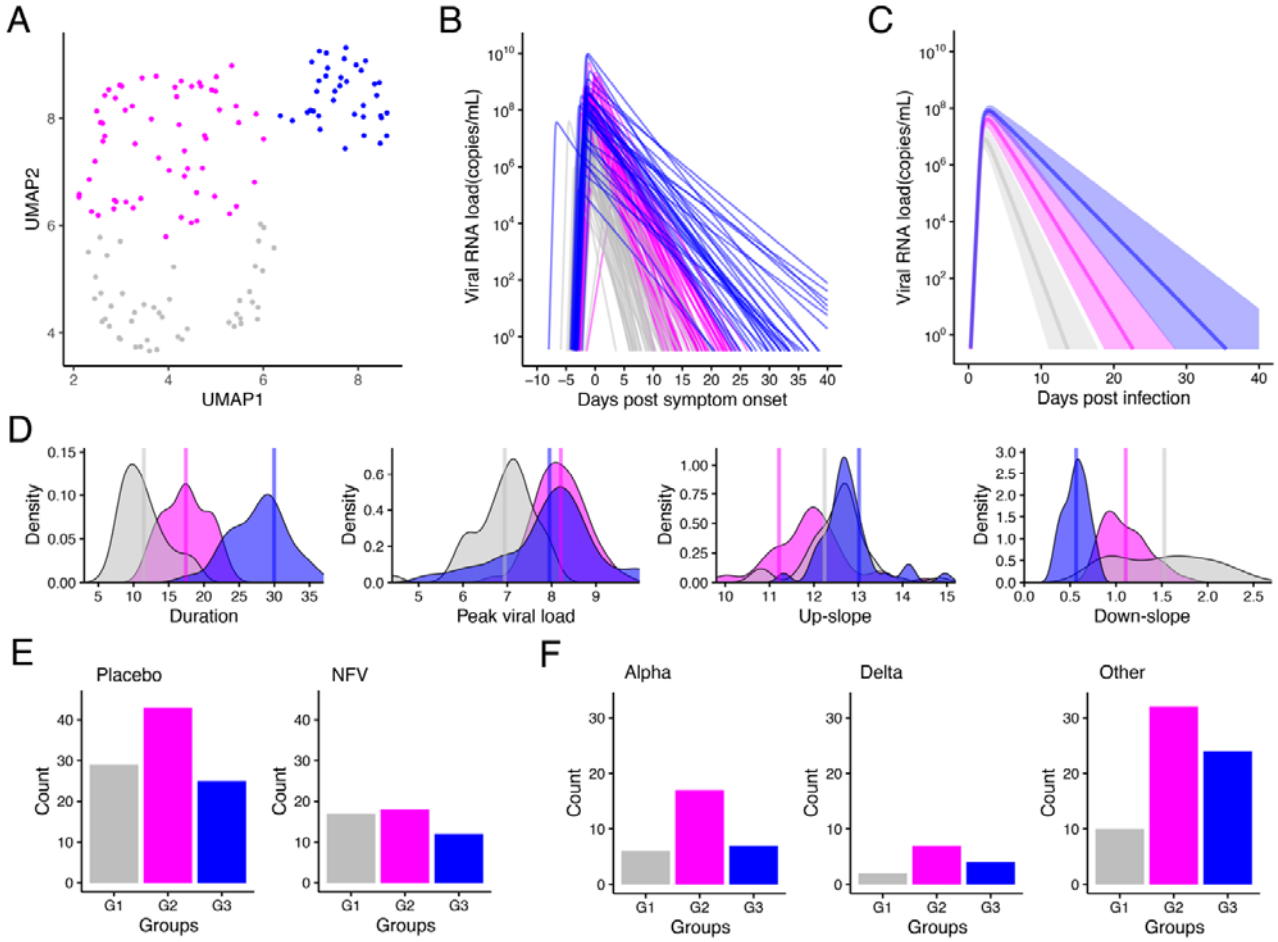


Figure 2 | Stratification of individual SARS-CoV-2 viral dynamics in saliva: (A) UMAP of stratified viral RNA load based on the extracted features from the reconstructed individual-level viral dynamics is shown. **(B)** The reconstructed individual viral RNA load is shown. Colors for individual-level viral dynamics correspond to the colors of the dots in the UMAP described in **(A)**. **(C)** The time-course patterns of each group highlighted by the Partial Least-Squares Discriminant Analysis (PLS-DA). **(D)** Distributions between groups of each feature used for stratification of viral shedding patterns are shown. The p-values of ANOVA for the difference in each feature among stratified group are all less than 0.05. **(E)** Distributions of the number of individuals in each stratified group for the standard-of-care alone (left, n=97) and standard-of-care plus NFV administration (right, n=47) participants are shown. **(F)** Distributions of the number of individuals in each stratified group for Alpha variants (left, n=30), Delta variants (middle, n=13), and other variants (right, n=66) of SARS-CoV-2 are shown.

Basic clinical data may not explain heterogeneity in individual viral shedding

Using data from the NFV clinical trial, we annotated the saliva viral loads of 90 participants with basic demographic information, daily symptoms, blood test results, and vital signs (**Fig 1A, Table 1** and **Supplementary Table 1)**. We also annotated the saliva viral loads of 52 participants from the University of Illinois with daily symptoms (**Supplementary Table 1**).

To identify factors that were significantly correlated with the viral shedding patterns in saliva specimens obtained from the NFV clinical trial, we first examined the 39 variables summarized in **Table 1**. Each factor was compared between the three groups by ANOVA and the p-values were corrected by the False Discovery Rate (FDR). However, we found no clinical data that differed significantly (i.e., corrected p-value of ANOVA of less than 0.05) among the stratified groups (**Fig 3A**). To avoid overfitting by bootstrap aggregating (bagging), we also trained a random forest classifier (see **Methods**), a tree-based machine learning algorithm suitable for tabular data[14], to predict the group from the clinical data of 90 individuals in the NFV clinical trial cohort and obtained ROC-AUCs of 60%, 49%, and 36% for predicting G1, G2, and G3, respectively (**Fig 3B**). We were not able to achieve a high ROC-AUC for predicting the shedding patterns based on the basic clinical data. We also attempted to make prediction based on clinical data which exhibited relatively low p-values in the ANOVA analysis. However, we were unable to achieve a high prediction accuracy with this approach (data not shown).

Next, we asked whether the stratification of the study population is associated with clinical symptoms of COVID-19 that could be caused by active replication of SARS-CoV-2. In general, the clinical symptoms of COVID-19 are cough, fever, shortness of breath, muscle pain, sore throat, confusion, chest pain, headache, rhinorrhea, diarrhea, and nausea and vomiting. In our study, individual-level symptom data were available as 8 categories in the cohorts from both the NFV clinical trial and the University of Illinois (**Supplementary Table 1**). Symptom data were obtained via participant self-report and encoded as categorical indicators (presence/absence) for each symptom. We tried to use a random forest classifier to investigate whether symptom data could predict the stratified groups and obtained ROC-AUCs of 57%, 54%, and 48% for predicting

G1, G2, and G3, respectively (**Fig 3C**). In fact, SARS-CoV-2 human challenge clearly showed no quantitative correlation between the individuals' time-series pattern of viral load and symptoms[15].

Additionally, we investigated the relationship between each feature of viral dynamics (i.e., duration of viral RNA detection, peak viral load, up-slope, and down-slope) and the clinical data by using the Pearson's correlation coefficient (**Supplementary Table 5**). Overall, correlation coefficients were low (0.06 on average) with high p-values, which suggests that no feature was likely to be explained by these clinical data.

A

Clinical data	p-value
Amount of hemoglobin	0.0546
Hematocrit	0.1209
Albumin	0.3016
AST	0.3783
ALT	0.3931
Age	0.6535
Bilirubin	0.6535
γ-GTP	0.7041
D-dimer	0.7041
Pulse rate	0.7041
PT-INR	0.7041
LDH	0.7696
Basophil	0.7696
CK	0.7696
BUN	0.7696
CRE	0.8255
Systolic blood pressure	0.8255
Procalcitonin	0.8255
Blood glucose	0.8676
PT	0.9289
White blood cell count	0.9289
Neutrophil	0.9864
Lymphocytes	0.9864
Na	0.9864
Fibrinogen	0.9864
SpO_2	0.9864
Monocyte	0.9864
HbA1c	0.9864
Diastolic blood pressure	0.9864
K	0.9864
Respiratory rate	0.9864
APTT	0.9864
Platelet count	0.9864
ALP	0.9864
CRP	0.997
Protein	0.997
Eosinophil	0.997
Cl	0.997
Red blood cell count	0.997

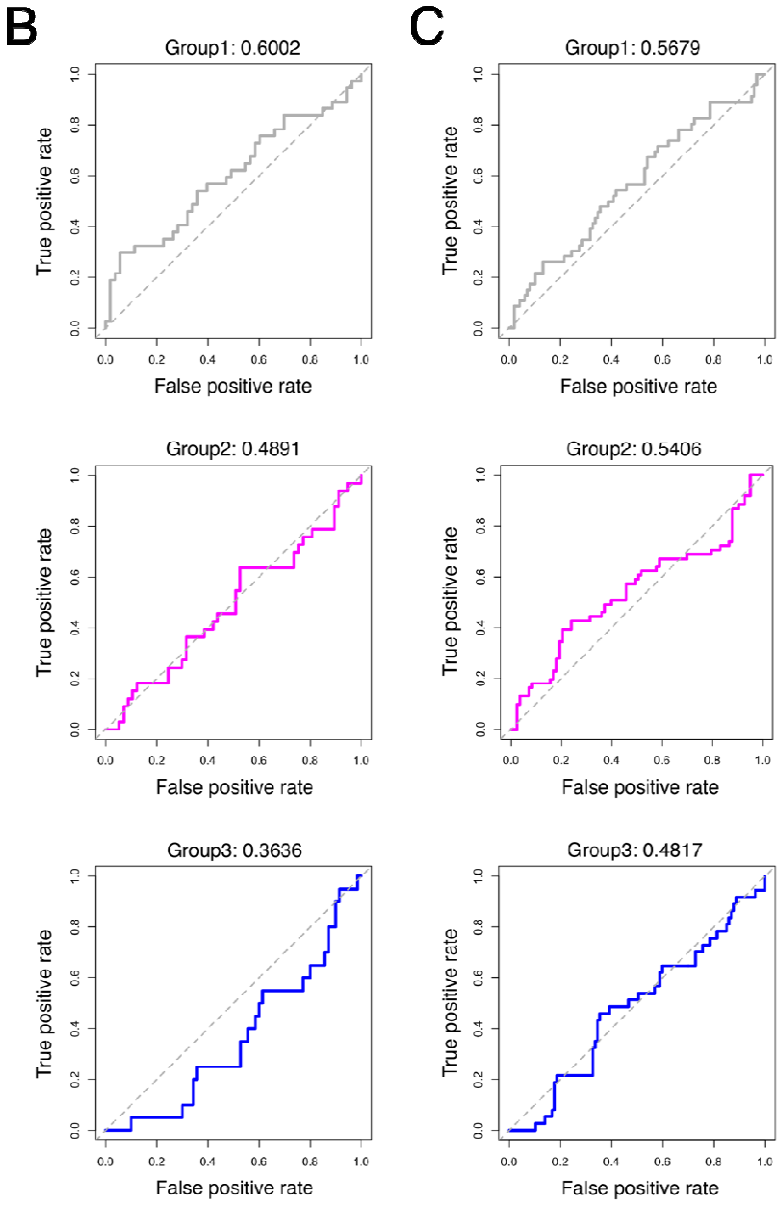


Figure 3 | Correlation between clinical data and viral shedding patterns: (A) P-values of ANOVA corrected by the FDR to compare clinical data among the three stratified groups are shown. Clinical data are listed in reverse order of p-values. **(B)** and **(C)** show ROC curves of random forest classifiers trained on predicting each group by using data for **(B)** clinical values described in **Table 1** and **(C)** symptom onset data described in **Supplementary Table 1**, respectively. The corresponding AUC (area under curve) value of each ROC curve is shown on the top of each panel.

Relationship between salivary micro-RNAs and viral shedding patterns in COVID-19 patients

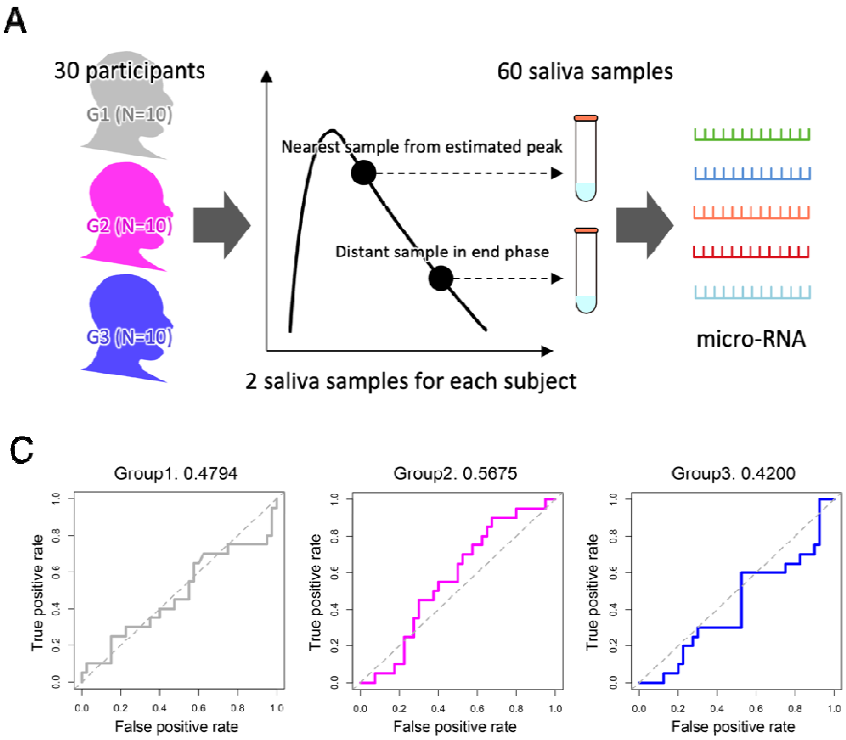
Various proteins in saliva have antiviral effects. It is also expected that some micro-RNAs in saliva may impair SARS-CoV-2 replication[1]. In addition, saliva-based micro-RNAs have already been reported to be associated with various diseases[16-21], including COVID-19[22]. Based on this, we anticipated that among the biomarkers obtainable from saliva, micro-RNAs could potentially serve as non-invasive predictors of viral dynamics. Accordingly, despite being less readily obtainable than clinical information, micro-RNAs are accessible from saliva sample and were anticipated to be closely linked to viral dynamics, so we incorporated them into our analysis.

We here used the stored residual saliva specimens from the NFV clinical trial to identify micro-RNA(s) associated with the stratified groups (i.e., G1, G2 and G3). We note that, because all residual saliva specimens are annotated with the individual participant and we know which participants belong to which stratified group, we can select and compare saliva specimens from G1, G2, and G3 without bias. This implies that we can impartially select participants in equal numbers from each group, unaffected by other factors. Specifically, we collected 60 stored saliva specimens from the NFV clinical trial to perform micro-RNA analysis for 30 participants. We picked two samples for each participant to evaluate the role of micro-RNAs during both the peak and the late phase (i.e., 30 samples for each phase): the nearest sample from the estimated peak and the most distant sample above the detection limit in the late phase (**Fig 4A**). We normalized micro-RNA expression among the samples by using the DESeq2 Bioconductor package. We summarize the information on the micro-RNAs we obtained from saliva specimens in **Supplementary Table 6**.

First, we compared micro-RNA levels between two phases using pairwise t-tests (and Mann-Whitney U tests) with FDR correction. As a result, no micro-RNA showed a statistically significant difference. This suggests that micro-RNA levels remain relatively stable during the course of infection and may therefore serve as a biomarker independent of sampling time.

Next, similar to the analysis using clinical data, we compared the expression levels of 92 micro-RNAs between the three stratified groups. Because the micro-RNA data were non-parametric, we used the Kruskal-Wallis ANOVA for analysis and corrected the p-values by FDR. However, we failed to find micro-RNA levels that differentiated the stratified groups (i.e., with a corrected p-value of Kruskal-Wallis ANOVA of less than 0.05) in the three trials using the data from the peak phase, the late phase, and both phases (e.g., **Fig 3B** for the total 60 samples). We also trained a random forest classifier to predict each group from the micro-RNA levels for the 60 total samples, and obtained ROC-AUCs of 48%, 57%, and 42%, respectively. Again, we did not obtain enough ROC-AUCs to predict stratified groups by using the collected micro-RNA data.

Furthermore, we investigated the relationship between the 4 features of viral dynamics and micro-RNA levels. Here we used the Spearman's correlation coefficient (**Supplementary Table 7**). Overall, we did not find strong correlations between micro-RNA levels and features (Spearman's correlation coefficients on average of 0.002, 0.024, -0.001, and -0.001 for duration of viral RNA detection, peak viral load, up-slope, and down-slope, respectively, for the 60 total samples). Only the mir-1846 level exhibited a weak negative correlation (-0.53 Spearman's correlation coefficient with 0.01 p-value) with the peak viral load (**Supplementary Fig 5**). We confirmed similar trends even when we analyzed the micro-RNA level for the peak and late phases separately.



B

Micro-RNA	p-value
mir-1846	0.2760
mir-1255	0.7909
mir-449	0.7909
mir-941	0.7909
mir-944	0.7909
mir-34	0.7909
mir-28	0.7909
mir-210	0.7909
mir-650	0.7909
mir-938	0.7909
mir-10	0.7909
mir-582	0.8014
mir-671	0.8014
mir-765	0.8014
mir-598	0.8014
amir-148	0.8014
mir-628	0.8014
mir-811	0.8255
mir-19	0.8255
mir-204	0.8255
mir-185	0.8255
mir-551	0.8255
mir-128	0.8255
mir-562	0.8255

Figure 4 | Correlation between micro-RNA data and viral shedding patterns: (A) The strategy of micro-RNA data collection from saliva samples in the NFV clinical trial is described. We picked a total of 30 participants by choosing 10 participants from each group. We chose two samples (the nearest to estimated peak and the most distant but above the detection limit in the late phase) from each participant for quantifying micro-RNA. **(B)** P-values of Kruskal-Wallis ANOVA corrected by the FDR for each micro-RNA level are shown. Micro-RNA levels are listed by reverse order of p-values. Only the 24 micro-RNA levels with the lowest p-values are shown. **(C)** ROC curves of random forest classifiers trained on predicting each group by using levels of 92 micro-RNAs are shown. The corresponding AUC value of each ROC curve is presented on the top of each panel.

Discussion

Being able to quickly and efficiently diagnose COVID-19 is essential in monitoring the pandemic. Because the sampling process for saliva is noninvasive, and because it is inexpensive and minimizes the risk for transmissions to health care workers[1], saliva sampling has excellent potential and advantages over other sampling methods from biological specimens such as the lower and upper respiratory tract[2,23]. Given the significant individual heterogeneity in the saliva viral shedding[5,24], identifying biomarker(s) for viral shedding patterns will be crucial for improving public health interventions in the era of living with COVID-19.

To improve our understanding of SARS-CoV-2 infection dynamics in saliva to enable application of saliva testing in the fight against COVID-19, we quantified and stratified longitudinal virus dynamics in saliva samples from 144 mildly symptomatic individuals from the cohorts of the NFV clinical trial[11] and the University of Illinois at Urbana-Champaign[5], and we uniquely analyzed the relationships among viral dynamics, clinical data, and micro-RNAs. Our mathematical modeling analysis indicates that viral dynamics in saliva may exhibit distinct patterns compared to those in the upper respiratory tract. We estimated that viral replication in saliva is characterized by a relatively rapid early growth phase, with a mean (standard deviation) doubling time of 1.44 (0.49) hours. Compared with prior studies analyzing viral dynamics in the upper respiratory tract using similar models, which reported doubling times of 2-4 hours[5,25,26], our findings suggest that viral replication in saliva proceeds faster than in the upper respiratory tract. Multiple previous studies have also shown that viral loads in saliva rise more rapidly than in the nasal cavity, are detected with higher sensitivity early in infection, and reach their peak earlier[5,27-30].

In addition to the large heterogeneity in virus infection dynamics, we identified three groups (i.e., G1, G2 and G3) with different viral shedding patterns (**Fig 2D**). Immunocompromised patients have been reported to have a prolonged duration of viral RNA detection, lasting over three months, underscoring the critical role of host immune responses in controlling viral infections[31-34]. Although oral immune responses remain poorly understood,

Huang et al. recently confirmed by using single-cell RNA sequencing of the human minor salivary glands and gingiva that SARS-CoV-2 infection can trigger sustained, localized immune responses in saliva[3]. In this study, we observed significant differences in the down-slopes of viral shedding in saliva among participants in different groups, with a more rapid decline in G1. This decline is likely attributed to a stronger immune response to SARS-CoV-2 in G1 participants than in participants in G2 and G3, as reflected in the death rate of infected cells due to the immune response (**Fig 2D**). Lower levels of viral replication have also been observed among infected participants with high baseline levels of mucosal IgA (but not IgG), as reported elsewhere[35]. Recently, we demonstrated that rapid anti-spike secretory IgA antibody responses can contribute to reducing duration of viral RNA detection and amounts in nasopharyngeal mucosa[36]. These findings highlight the importance of biomarkers that directly reflect an individual's immune response, such as virus-specific antibody induction and T cell levels etc., in predicting viral shedding patterns. Therefore, quantifying the time-series pattern of mucosal IgA and its correlation with saliva viral load may provide crucial insights into the stratification of SARS-CoV-2 infection dynamics.

For the purpose of predicting viral shedding patterns during the early stage of infection, we first explored the association of 39 basic clinical variables, 8 daily symptoms, and the levels of 92 micro-RNAs with the stratified groups. However, none of the factors were significant (**Table 1**, **Fig 3A**, **Fig 4B, Supplementary Table 1** and **Supplementary Table**). On the other hand, it is noteworthy that all infection cases were mild and that most participants had clinical indicators within normal ranges. This lack of clinical heterogeneity within the cohort may have limited the ability to fully capture the diversity of infection dynamics. Moreover, the clinical parameters analyzed in this study are, a priori, unlikely to exhibit strong correlations with virologic outcomes. In contrast, we showed that mir-1846, which is an exogenous micro-RNA that is specifically classified as an Oryza sativa micro-RNA (osa-microRNA)[37], may exhibit a weak negative correlation. Exogenous micro-RNAs enter the human body primarily through food and can affect human metabolism by interacting and binding with human genes. mir-1846 is reported to interact

with two human genes[37] that are known to be associated with the progression of melanoma, various cancers, and leukemia. This suggests that mir-1846 levels may be linked to human immunity. Few studies have investigated the role of mir-1846 in humans, but our findings suggest the need for further investigations into the impact of this micro-RNA level on human immunity. Our research sheds light on the intricate patterns of viral shedding in saliva.

Our approach has several limitations that must be considered in our next study: First, our analysis was limited to participants with symptomatic infection and excluded those with asymptomatic infection (22 asymptomatic individuals out of a total of 182 individuals, i.e., 12% of participants) because we integrated datasets with different time scales from different cohorts. Although our data do not include participants infected with omicron variants, others have reported that the omicron variant may cause a higher proportion of asymptomatic infection[38]. Thus, evaluating the effect of asymptomatic infection will be important to update our stratification, especially for recent (or future emerging) VOCs. Second, potential viral rebound was neither prespecified nor systematically assessed. A subset of participants exhibited patterns consistent with possible rebound (e.g., S01-16 and S01-43 in **Supplementary Fig 1**), which could affect estimates of viral RNA detection duration. Future studies should predefine an operational criterion for viral rebound and explicitly incorporate it into the modeling framework to strengthen robustness. Third, micro-RNAs participate in the post-transcriptional regulation of gene expression; however, they do not provide direct insights into immune cell dynamics. Given the reported association between the duration of viral RNA detection and mucosal immunity as discussed above, it appears imperative to analyze modalities that are directly linked to the immune response in the future. Fourth, some of our results may have limited relevance to the current COVID-19 situation, as most people have now either been infected or vaccinated. Nevertheless, investigating the relationship between viral shedding patterns in saliva and various clinical and microRNA data, and developing a method to do so, remains important. Such research can offer valuable insights into the early response to emerging infectious viruses in the future.

Another potential limitation of this study is the timing of saliva specimen sampling, although we took great care to select and compare specimens from G1, G2, and G3 without bias. As a result of our clinical trial design (jRCT2071200023[11,12]), participants were enrolled after the onset of symptoms, thereby restricting saliva specimen collection exclusively to the post-symptom phase. Unfortunately, we lack samples from the pre-infection, pre-symptomatic, and early infection phases. Consequently, the absence of individual-level baseline values for micro-RNA means that inter-participant heterogeneity in micro-RNA levels may obscure signals related to distinct viral infection dynamics in saliva.

In conclusion, our study revealed that the dynamics of SARS-CoV-2 infection in saliva can be classified into three groups based mainly on the duration of viral RNA detection. However, accurately predicting the variability in viral dynamics remains a challenging task, because it requires a more comprehensive understanding of the complex shedding patterns in saliva, as well as detailed clinical and molecular data. The identification of a sensitive, simple, and rapid biomarker for saliva viral shedding will be imperative for future COVID-19 outbreak control.

METHODS

Ethics statement

The NFV clinical trial was approved by the institutional review board of Nagasaki University Hospital (approval number: I20-001) and is registered with the Japan Registry of Clinical Trials (jRCT2071200023). All participants provided written, informed consent for secondary use of clinical information and samples. The present study was approved by the ethics committee of Nagoya University (approval number: hc 22-01).

Saliva viral load data

Longitudinal saliva viral load data of participants with symptomatic and asymptomatic COVID-19 (122 cases) were obtained from the NFV clinical trial[11]. Briefly, the NFV clinical trial was a prospective, randomized, open-label, blinded-endpoint, parallel-group trial conducted between July 2020 and October 2021 at 11 university and teaching hospitals in Japan. This study consisted of a 14-day treatment period and a 14-day follow-up period, with no significant differences in the time to viral clearance between patients who received standard-of-care plus NFV administration and those who had the standard-of-care alone[11,12]. Therefore, the participants with COVID-19 were analyzed together here. In addition, we obtained similar saliva viral load data (60 cases) from the cohort of the University of Illinois at Urbana-Champaign[5]. This cohort contained all faculty, staff, and students at the University of Illinois at Urbana-Champaign, who undergo at least twice weekly quantitative PCR-RT testing during fall of 2020 and spring of 2021.

The viral load data for Illinois cohort was calculated based on the linear relationship between viral load (copies/mL) and Ct values presented in[5], specific to the measurement method used in this study. Among those 182 cases, we focused only on symptomatic participants. Asymptomatic individuals were excluded since they lack a definable date of symptom onset, which prevents us from establishing a clear baseline for the time axis in parameter estimation. Also, we excluded the participants who had less than three measured viral loads that were not limit detections (i.e., 90 cases from the NFV clinical trial and 54 cases from the University of

Illinois were used in this study). Because the trial design discontinued follow-up after two consecutive results below the detection limit, most of these participants already had viral loads near the detection limit at baseline, making reliable model fitting infeasible. The limit of detection for viral load data from the University of Illinois is 1.08 copies/ml. However, the limit of detection for viral load data from NFV clinical trial was unclear. Considering this, we assumed 1.08 copies/ml as the limit of detection for all viral load data. To examine the impact of assumptions regarding the detection limit, we performed a sensitivity analysis. Specifically, for the NFV cohort, we compared the estimated viral dynamics when the detection limit was set to 0 and 2, instead of 1.08. Overall, the results were largely consistent across these scenarios (**Supplementary Fig 6**).

Viral genome sequencing

The cDNA had been synthesized from RNA of SARS-CoV-2–positive saliva samples. Reverse transcription, multiplex PCR reaction, and Illumina library prep were conducted using a protocol published previously[39]. The pooled library was first purified by AMPureXP at 0.8 x concentration and then again at 1.2 x concentration. The purified library was sequenced for 151 cycles at both paired-ends in Illumina iSeq100. Sequence analysis was performed using the nf-core/viralcon pipeline (10.5281/zenodo.3901628).

Quantifying biomarkers in saliva

Total RNA from saliva was extracted with MagMAX CORE Nucleic Acid Purification kits (Applied Biosystems, Foster City, CA). Micro-RNAs were detected using Illumina Hiseq x Ten (Illumina, Inc, San Diego, CA) with data processing by ribodepletion (Genewiz-Azenta, South Plainfield, NJ). To remove technical sequences, the pass filter data in the fastq format were processed by Trimmomatic (v0.30) to be high-quality clean data. Following quality trimming, micro-RNAs were identified and checked using miRDeep2[40]. Normalization of micro-RNA expression among samples and differential expression analysis were carried out using the DESeq2 Bioconductor package.

Basic clinical data

Basic clinical data including basic demographic characteristics of the study participants, symptoms, and findings of physical examinations and laboratory tests were obtained according to the study protocol[11]. We here used information on age, daily symptoms, results of blood tests, and vital signs of the symptomatic participants in the NFV clinical trial (summarized in **Table 1** and **Supplementary Table 1**).

Mathematical modeling

To describe SARS-CoV-2 infection dynamics in saliva specimens, we here mainly used the following mathematical model developed in our recent studies[4,6,10]:

$$\frac{df(t)}{dt} = -\beta f(t)V(t), \qquad (1)$$

$$\frac{dV(t)}{dt} = \gamma f(t)V(t) - \delta V(t). \qquad (2)$$

The variables $f(t)$ and $V(t)$ are the fraction of uninfected target cells and the total amount of virus, respectively, and the parameters β, γ, and δ are the rate constant for virus infection, the maximum rate constant for viral replication, and the death rate of infected cells, respectively.

In addition to comparing the simple model [i.e., a target-cell-limited model; Eqs.(1-2)], we also used the following "immune effector cell model" developed in Ke et al.[5] for the saliva viral load (see **Supplementary Fig 4**):

$$\frac{dT(t)}{dt} = -\beta T(t)V(t), \qquad (3)$$

$$\frac{dE(t)}{dt} = \beta T(t)V(t) - kE(t), \qquad (4)$$

$$\frac{dI(t)}{dt} = kE(t) - \delta(t)I(t), \qquad (5)$$

$$\frac{dV(t)}{dt} = \pi I(t) - cV(t). \qquad (6)$$

The variables $T(t)$, $E(t)$, and $I(t)$ are the total number uninfected target cells, cells in the eclipse phase of infection, and productively infected cells, respectively. The parameters $1/k$, π, and c are the average duration of the eclipse phase, the virus production rate, and the clearance rate of viruses, respectively. The death rate of infected cells is assumed to be time-dependent to mimic the killing of infected cells by immune effector cells: $\delta(t) = \delta_1$ for $t < t_1$ and $\delta(t) = \delta_1 + \delta_2$ for $t_1 \leq t$. For a detailed explanation of the immune effector cell model, the reader is referred to Ke et al.[5].

Parameter estimation

A nonlinear mixed-effects modelling approach incorporates fixed effects as well as random effects that describe the inter-participant variability in parameters. Including random effects amounts to a partial pooling of the data of all participants to improve estimates of the parameters applicable across the cases. In this study, viral load data from the Illinois cohort were collected primarily during the early to middle phase of SARS-CoV-2 infection, whereas those from the NFV cohort were obtained mainly during the middle to late phase. In particular, the Illinois cohort provides pre-symptomatic data, while the NFV cohort includes data beyond 14 days after symptom onset, with each cohort lacking the phase represented in the other. Although the NFV cohort lacks pre-symptomatic data, the nonlinear mixed-effects model first derives population-level parameters from all participants and then accounts for individual variability through random effects. This allows pre-symptomatic information from the Illinois cohort to inform the inferred viral dynamics of participants in NFV cohort. Conversely, for the late phase, the NFV cohort serves the complementary role. By jointly analyzing both cohorts with a nonlinear mixed-effects model to estimate individual-level parameters, we can therefore capture the complete time-course pattern of SARS-CoV-2 infection dynamics across all participants, to the extent possible.

In our analyses, the variable $V(t)$ in Eq.(2) corresponds to the viral load in saliva for SARS-CoV-2. To fit the patient's viral load data, we used a program MONOLIX 2021R2

(www.lixoft.com), implement maximum likelihood estimation of parameters in nonlinear mixed effect model. The nonlinear mixed effect model allows a fixed effect as well as interpatient variability. This method estimates each parameter $\theta_i (= \theta \times e^{\eta_i})$ for each individual where θ is a fixed effect, and η_i is a random effect, and which obeys a Gaussian distribution with mean 0 and standard deviation Ω. Here we used lognormal distributions as prior distributions of parameters to guarantee the positiveness (i.e., negative values do not biologically make sense). In parameter estimation, as time 0 in the original dataset represents the time of symptom onset, we also estimated the time from infection to symptom onset (corresponding to τ in **Supplementary Table 2** and **Supplementary Table 3**) along with other parameters. The fixed effect parameters and random effect parameters were estimated by use of the stochastic approximation Expectation/Maximization (SAEM) algorithm and empirical Bayes method, respectively. A right-truncated normal distribution was used in the likelihood function to account for the left censoring of the viral load data (i.e., when the viral load is not detectable)[41]. The standard errors were obtained from the Fisher Information Matrix using the linearization method (**Supplementary Fig 3C**). We changed the initial values multiple times to avoid a local minimum of AIC and confirmed the robustness of parameter estimation.

Unsupervised clustering and stratification of SARS-CoV-2 infection dynamics

Unsupervised random forest clustering was performed on the selected features of the virus infection dynamics, that is, the peak viral load, duration of viral RNA detection, up-slope, and down-slope (rfUtilities package in R). The use of random forest allows us to avoid overfitting by bootstrap aggregating (bagging) and to achieve better generalization performance[42]. After a random forest dissimilarity (i.e., the distance matrix between all pairs of samples) was obtained, it was visualized with Uniform Manifold Approximation and Projection (UMAP) in a two-dimensional plane and was stratified with spectral clustering (Python scikit-learn). The optimal number of clusters was determined by the eigengap heuristic method.

Furthermore, to evaluate how statistical uncertainty in the estimated viral dynamics might influence the stratification, we propagated parameter uncertainty as follows. For each participant, we resampled the parameters of the mathematical model 100 times within their 95% credible intervals and calculated the corresponding features of reconstructed viral dynamics. We then averaged the features across the 100 replicates to obtain an uncertainty-adjusted feature vector for each participant. Using these vectors, we re-calculated the distance matrix as described above. Then agreement between the uncertainty-adjusted and original distance matrix was quantified with a permutation-based Mantel test.

Random forest classifiers for characterizing stratified groups

Random forest classifiers were trained to predict either of the three stratified groups (G1-G3) using randomForest packages in R. The receiver operating characteristic (ROC) curve for each classifier was drawn from out-of-bag (OOB) samples using the pROC package in R. For example, the ROC for G1 is for the random forest classifier predicting "G1" versus "G2 or G3". We list the variables for the supervised random forest in **Table 1**, **Supplementary Table 1**, and **Supplementary Table 2.**

Statistical analysis

When necessary, the variables were compared among different groups using Fisher's exact test (for categorical variables), analysis of variance (ANOVA, for numerical variables from clinical data with more than two groups), and Kruskal-Wallis ANOVA (for variables from micro-RNA data with more than two groups). Corrections of p-values for multiple testing were performed by the False Discovery Rate (FDR). Also, the variables were investigated for their relationship with features of viral load using the Pearson’s correlation coefficient (for variables from clinical data) and the Spearman’s correlation coefficient (for variables from micro-RNA data). All statistical analyses were performed using R (version 4.1.3).

Data availability

The datasets (NFV clinical trial) analyzed in this study are not publicly available owing to privacy and ethical considerations. They are, however, available from the corresponding author upon reasonable request, subject to ethics approval from all participating institutions.

LIST OF SUPPLEMENTARY MATERIALS

Supplementary Figure 1 | Reconstructed viral dynamics in saliva samples for individual participants

Supplementary Figure 2 | Visual predictive checks (VPC) for the viral load models

Supplementary Figure 3 | Convergence diagnostics of the SAEM algorithm

Supplementary Figure 4 | Comparison of three model fits to viral load in saliva specimens for individual participants

Supplementary Figure 5 | Correlation between mir-1846 level and the features of infection dynamics

Supplementary Figure 6 | Sensitivity analysis of the detection limit for viral load data in the NFV cohort

Supplementary Table 1 | Daily symptom data for whole cohorts for each group

Supplementary Table 2 | Summary of parameter estimation by the model described in Eqs.(1-2)

Supplementary Table 3 | Summary of parameter estimation by the model described in Eqs.(3-6)

Supplementary Table 4 | Features of reconstructed individual viral dynamics for each group

Supplementary Table 5 | Pearson's correlation coefficients between clinical data and features of viral dynamics

Supplementary Table 6 | Salivary micro-RNAs obtained from 60 specimens of 30 participants

Supplementary Table 7 | Spearman's correlation coefficients between micro-RNA data and features of viral dynamics

REFERENCES

1. Baghizadeh Fini, M. Oral saliva and COVID-19. *Oral Oncol* **108**, 104821 (2020), DOI: 10.1016/j.oraloncology.2020.104821.
2. Wyllie, A.L., Fournier, J., Casanovas-Massana, A., Campbell, M., Tokuyama, M., Vijayakumar, P., Warren, J.L., Geng, B., Muenker, M.C., Moore, A.J., Vogels, C.B.F., Petrone, M.E., Ott, I.M., Lu, P., Venkataraman, A., Lu-Culligan, A., Klein, J., Earnest, R., Simonov, M., Datta, R., Handoko, R., Naushad, N., Sewanan, L.R., Valdez, J., White, E.B., Lapidus, S., Kalinich, C.C., Jiang, X., Kim, D.J., Kudo, E., Linehan, M., Mao, T., Moriyama, M., Oh, J.E., Park, A., Silva, J., Song, E., Takahashi, T., Taura, M., Weizman, O.E., Wong, P., Yang, Y., Bermejo, S., Odio, C.D., Omer, S.B., Dela Cruz, C.S., Farhadian, S., Martinello, R.A., Iwasaki, A., Grubaugh, N.D. & Ko, A.I. Saliva or Nasopharyngeal Swab Specimens for Detection of SARS-CoV-2. *N Engl J Med* **383**, 1283-1286 (2020), DOI: 10.1056/NEJMc2016359.
3. Huang, N., Pérez, P., Kato, T., Mikami, Y., Okuda, K., Gilmore, R.C., Conde, C.D., Gasmi, B., Stein, S., Beach, M., Pelayo, E., Maldonado, J.O., Lafont, B.A., Jang, S.I., Nasir, N., Padilla, R.J., Murrah, V.A., Maile, R., Lovell, W., Wallet, S.M., Bowman, N.M., Meinig, S.L., Wolfgang, M.C., Choudhury, S.N., Novotny, M., Aevermann, B.D., Scheuermann, R.H., Cannon, G., Anderson, C.W., Lee, R.E., Marchesan, J.T., Bush, M., Freire, M., Kimple, A.J., Herr, D.L., Rabin, J., Grazioli, A., Das, S., French, B.N., Pranzatelli, T., Chiorini, J.A., Kleiner, D.E., Pittaluga, S., Hewitt, S.M., Burbelo, P.D., Chertow, D., Frank, K., Lee, J., Boucher, R.C., Teichmann, S.A., Warner, B.M. & Byrd, K.M. SARS-CoV-2 infection of the oral cavity and saliva. *Nat Med* **27**, 892-903 (2021), DOI: 10.1038/s41591-021-01296-8.
4. Kim, K.S., Ejima, K., Iwanami, S., Fujita, Y., Ohashi, H., Koizumi, Y., Asai, Y., Nakaoka, S., Watashi, K., Aihara, K., Thompson, R.N., Ke, R., Perelson, A.S. & Iwami, S. A quantitative model used to compare within-host SARS-CoV-2, MERS-CoV, and SARS-CoV dynamics provides insights into the pathogenesis and treatment of SARS-CoV-2. *PLoS Biol* **19**, e3001128 (2021), DOI: 10.1371/journal.pbio.3001128.
5. Ke, R., Martinez, P.P., Smith, R.L., Gibson, L.L., Mirza, A., Conte, M., Gallagher, N., Luo, C.H., Jarrett, J., Zhou, R., Conte, A., Liu, T., Farjo, M., Walden, K.K.O., Rendon, G., Fields, C.J., Wang, L., Fredrickson, R., Edmonson, D.C., Baughman, M.E., Chiu, K.K., Choi, H., Scardina, K.R., Bradley, S., Gloss, S.L., Reinhart, C., Yedetore, J., Quicksall, J., Owens, A.N., Broach, J., Barton, B., Lazar, P., Heetderks, W.J., Robinson, M.L., Mostafa, H.H., Manabe, Y.C., Pekosz, A., McManus, D.D. & Brooke, C.B. Daily longitudinal sampling of SARS-CoV-2 infection reveals substantial heterogeneity in infectiousness. *Nat Microbiol* **7**, 640-652 (2022), DOI: 10.1038/s41564-022-01105-z.
6. Jeong, Y.D., Ejima, K., Kim, K.S., Joohyeon, W., Iwanami, S., Fujita, Y., Jung, I.H., Aihara, K., Shibuya, K., Iwami, S., Bento, A.I. & Ajelli, M. Designing isolation guidelines for COVID-19 patients with rapid antigen tests. *Nat Commun* **13**, 4910 (2022), DOI: 10.1038/s41467-022-32663-9.
7. Goyal, A., Cardozo-Ojeda, E.F. & Schiffer, J.T. Potency and timing of antiviral therapy as determinants of duration of SARS-CoV-2 shedding and intensity of inflammatory response. *Sci Adv* **6**(2020), DOI: 10.1126/sciadv.abc7112.
8. Marc, A., Kerioui, M., Blanquart, F., Bertrand, J., Mitjà, O., Corbacho-Monné, M., Marks, M. & Guedj, J. Quantifying the relationship between SARS-CoV-2 viral load and infectiousness. *Elife* **10**(2021), DOI: 10.7554/eLife.69302.
9. Néant, N., Lingas, G., Le Hingrat, Q., Ghosn, J., Engelmann, I., Lepiller, Q., Gaymard, A., Ferré, V., Hartard, C., Plantier, J.C., Thibault, V., Marlet, J., Montes, B., Bouiller, K., Lescure, F.X., Timsit, J.F., Faure, E., Poissy, J., Chidiac, C., Raffi, F., Kimmoun, A., Etienne, M., Richard, J.C., Tattevin, P., Garot, D., Le Moing, V., Bachelet, D., Tardivon, C., Duval, X., Yazdanpanah, Y., Mentré, F., Laouénan, C., Visseaux, B. & Guedj, J. Modeling SARS-CoV-2 viral kinetics and association with mortality in hospitalized patients from the French COVID cohort. *Proc Natl Acad Sci U S A* **118**(2021), DOI: 10.1073/pnas.2017962118.

10. Iwanami, S., Ejima, K., Kim, K.S., Noshita, K., Fujita, Y., Miyazaki, T., Kohno, S., Miyazaki, Y., Morimoto, S., Nakaoka, S., Koizumi, Y., Asai, Y., Aihara, K., Watashi, K., Thompson, R.N., Shibuya, K., Fujiu, K., Perelson, A.S., Iwami, S. & Wakita, T. Detection of significant antiviral drug effects on COVID-19 with reasonable sample sizes in randomized controlled trials: A modeling study. *PLoS Med* **18**, e1003660 (2021), DOI: 10.1371/journal.pmed.1003660.
11. Hosogaya, N., Miyazaki, T., Fukushige, Y., Takemori, S., Morimoto, S., Yamamoto, H., Hori, M., Kurokawa, T., Kawasaki, Y., Hanawa, M., Fujii, Y., Hanaoka, H., Iwami, S., Watashi, K., Yamagoe, S., Miyazaki, Y., Wakita, T., Izumikawa, K., Yanagihara, K., Mukae, H. & Kohno, S. Efficacy and safety of nelfinavir in asymptomatic and mild COVID-19 patients: a structured summary of a study protocol for a multicenter, randomized controlled trial. *Trials* **22**, 309 (2021), DOI: 10.1186/s13063-021-05282-w.
12. Miyazaki, T., Hosogaya, N., Fukushige, Y., Takemori, S., Morimoto, S., Yamamoto, H., Hori, M., Ozawa, Y., Shiko, Y., Inaba, Y., Kurokawa, T., Hanaoka, H., Iwanami, S., Kim, K., Iwami, S., Watashi, K., Miyazawa, K., Umeyama, T., Yamagoe, S., Miyazaki, Y., Wakita, T., Sumiyoshi, M., Hirayama, T., Izumikawa, K., Yanagihara, K., Mukae, H., Kawasuji, H., Yamamoto, Y., Tarumoto, N., Ishii, H., Ohno, H., Yatera, K., Kakeya, H., Kichikawa, Y., Kato, Y., Matsumoto, T., Saito, M., Yotsuyanagi, H. & Kohno, S. A Multicenter Randomized Controlled Trial To Evaluate the Efficacy and Safety of Nelfinavir in Patients with Mild COVID-19. *Microbiol Spectr* **11**, e0431122 (2023), DOI: 10.1128/spectrum.04311-22.
13. Trobaugh, D.W. & Klimstra, W.B. MicroRNA Regulation of RNA Virus Replication and Pathogenesis. *Trends Mol Med* **23**, 80-93 (2017), DOI: 10.1016/j.molmed.2016.11.003.
14. Grinsztajn, L., Oyallon, E. & Varoquaux, G. Why do tree-based models still outperform deep learning on tabular data? *arXiv preprint arXiv:2207.08815* (2022), DOI.
15. Killingley, B., Mann, A.J., Kalinova, M., Boyers, A., Goonawardane, N., Zhou, J., Lindsell, K., Hare, S.S., Brown, J., Frise, R., Smith, E., Hopkins, C., Noulin, N., Löndt, B., Wilkinson, T., Harden, S., McShane, H., Baillet, M., Gilbert, A., Jacobs, M., Charman, C., Mande, P., Nguyen-Van-Tam, J.S., Semple, M.G., Read, R.C., Ferguson, N.M., Openshaw, P.J., Rapeport, G., Barclay, W.S., Catchpole, A.P. & Chiu, C. Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. *Nat Med* **28**, 1031-1041 (2022), DOI: 10.1038/s41591-022-01780-9.
16. Alizadeh, N., Zahedi, H., Koopaie, M., Fatahzadeh, M., Mousavi, R. & Kolahdooz, S. Diagnosis of lung cancer using salivary miRNAs expression and clinical characteristics. *BMC Pulm Med* **25**, 41 (2025), DOI: 10.1186/s12890-025-03502-6.
17. Dominguez-de-Barros, A., Sirvent-Blanco, C., Garcia-Perez, O., Gajate-Arenas, M., Garcia-Ramos, A., Migliazzo, C., Pinero, J.E., Lorenzo-Morales, J. & Cordoba-Lanus, E. Telomere Length, Oxidative Stress Markers, and Related miRNAs in Non-Invasive Samples of Mild COVID-19 Cases. *Int J Mol Sci* **26**(2025), DOI: 10.3390/ijms26104934.
18. Setti, G., Pezzi, M.E., Viani, M.V., Pertinhez, T.A., Cassi, D., Magnoni, C., Bellini, P., Musolino, A., Vescovi, P. & Meleti, M. Salivary MicroRNA for Diagnosis of Cancer and Systemic Diseases: A Systematic Review. *Int J Mol Sci* **21**(2020), DOI: 10.3390/ijms21030907.
19. Xie, Z., Yin, X., Gong, B., Nie, W., Wu, B., Zhang, X., Huang, J., Zhang, P., Zhou, Z. & Li, Z. Salivary microRNAs show potential as a noninvasive biomarker for detecting resectable pancreatic cancer. *Cancer Prev Res (Phila)* **8**, 165-173 (2015), DOI: 10.1158/1940-6207.CAPR-14-0192.
20. Zhang, Z., Liu, T., Dong, M., Ahamed, M.A. & Guan, W. Sample-to-answer salivary miRNA testing: New frontiers in point-of-care diagnostic technologies. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **16**, e1969 (2024), DOI: 10.1002/wnan.1969.
21. Pimenta, R., Viana, N.I., Dos Santos, G.A., Candido, P., Guimarães, V.R., Romão, P., Silva, I.A., de Camargo, J.A., Hatanaka, D.M., Queiroz, P.G.S., Teruya, A., Echenique, L., Besen, B., Leite, K.R.M., Srougi, V., Srougi, M. & Reis, S.T. MiR-200c-3p expression may be associated with worsening of the clinical course of patients with COVID-19. *Mol Biol Res Commun* **10**, 141-147 (2021), DOI: 10.22099/mbrc.2021.40555.1631.

22. Ahamed, M.A., Zhang, Z., Kshirsagar, A., Politza, A.J., Sethuraman, U., Suresh, S., Hicks, S., Guo, F. & Guan, W. Multiplexed Salivary miRNA Quantification for Predicting Severe COVID-19 Symptoms in Children Using Ligation-RPA Amplification Assay. *medRxiv* (2025), DOI: 10.1101/2025.03.10.25323665.
23. To, K.K., Tsang, O.T., Leung, W.S., Tam, A.R., Wu, T.C., Lung, D.C., Yip, C.C., Cai, J.P., Chan, J.M., Chik, T.S., Lau, D.P., Choi, C.Y., Chen, L.L., Chan, W.M., Chan, K.H., Ip, J.D., Ng, A.C., Poon, R.W., Luo, C.T., Cheng, V.C., Chan, J.F., Hung, I.F., Chen, Z., Chen, H. & Yuen, K.Y. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect Dis* **20**, 565-574 (2020), DOI: 10.1016/s1473-3099(20)30196-1.
24. Hay, J.A., Kissler, S.M., Fauver, J.R., Mack, C., Tai, C.G., Samant, R.M., Connolly, S., Anderson, D.J., Khullar, G., MacKay, M., Patel, M., Kelly, S., Manhertz, A., Eiter, I., Salgado, D., Baker, T., Howard, B., Dudley, J.T., Mason, C.E., Nair, M., Huang, Y., DiFiori, J., Ho, D.D., Grubaugh, N.D. & Grad, Y.H. Quantifying the impact of immune history and variant on SARS-CoV-2 viral kinetics and infection rebound: a retrospective cohort study. *medRxiv*, 2022.2001.2013.22269257 (2022), DOI: 10.1101/2022.01.13.22269257.
25. Gunawardana, M., Webster, S., Rivera, S., Cortez, J.M., Jr., Breslin, J., Pinales, C., Buser, C., Ibarrondo, F.J., Yang, O.O., Bobardt, M., Gallay, P.A., Adler, A.P., Ramirez, C.M., Anton, P.A. & Baum, M.M. Early SARS-CoV-2 dynamics and immune responses in unvaccinated participants of an intensely sampled longitudinal surveillance study. *Commun Med (Lond)* **2**, 129 (2022), DOI: 10.1038/s43856-022-00195-4.
26. Iyaniwura, S.A., Ribeiro, R.M., Zitzmann, C., Phan, T., Ke, R. & Perelson, A.S. The kinetics of SARS-CoV-2 infection based on a human challenge study. *Proc Natl Acad Sci U S A* **121**, e2406303121 (2024), DOI: 10.1073/pnas.2406303121.
27. Migueres, M., Mansuy, J.M., Vasseur, S., Claverie, N., Lougarre, C., Soulier, F., Tremeaux, P. & Izopet, J. Omicron Wave SARS-CoV-2 Diagnosis: Evaluation of Saliva, Anterior Nasal, and Nasopharyngeal Swab Samples. *Microbiol Spectr* **10**, e0252122 (2022), DOI: 10.1128/spectrum.02521-22.
28. Puhach, O., Meyer, B. & Eckerle, I. SARS-CoV-2 viral load and shedding kinetics. *Nat Rev Microbiol* **21**, 147-161 (2023), DOI: 10.1038/s41579-022-00822-w.
29. Savela, E.S., Viloria Winnett, A., Romano, A.E., Porter, M.K., Shelby, N., Akana, R., Ji, J., Cooper, M.M., Schlenker, N.W., Reyes, J.A., Carter, A.M., Barlow, J.T., Tognazzini, C., Feaster, M., Goh, Y.Y. & Ismagilov, R.F. Quantitative SARS-CoV-2 Viral-Load Curves in Paired Saliva Samples and Nasal Swabs Inform Appropriate Respiratory Sampling Site and Analytical Test Sensitivity Required for Earliest Viral Detection. *J Clin Microbiol* **60**, e0178521 (2022), DOI: 10.1128/JCM.01785-21.
30. Smith, R.L., Gibson, L.L., Martinez, P.P., Ke, R., Mirza, A., Conte, M., Gallagher, N., Conte, A., Wang, L., Fredrickson, R., Edmonson, D.C., Baughman, M.E., Chiu, K.K., Choi, H., Jensen, T.W., Scardina, K.R., Bradley, S., Gloss, S.L., Reinhart, C., Yedetore, J., Owens, A.N., Broach, J., Barton, B., Lazar, P., Henness, D., Young, T., Dunnett, A., Robinson, M.L., Mostafa, H.H., Pekosz, A., Manabe, Y.C., Heetderks, W.J., McManus, D.D. & Brooke, C.B. Longitudinal Assessment of Diagnostic Test Performance Over the Course of Acute SARS-CoV-2 Infection. *J Infect Dis* **224**, 976-982 (2021), DOI: 10.1093/infdis/jiab337.
31. Leung, W.F., Chorlton, S., Tyson, J., Al-Rawahi, G.N., Jassem, A.N., Prystajecky, N., Masud, S., Deans, G.D., Chapman, M.G., Mirzanejad, Y., Murray, M.C.M. & Wong, P.H.P. COVID-19 in an immunocompromised host: persistent shedding of viable SARS-CoV-2 and emergence of multiple mutations: a case report. *Int J Infect Dis* **114**, 178-182 (2022), DOI: 10.1016/j.ijid.2021.10.045.
32. Niyonkuru, M., Pedersen, R.M., Assing, K., Andersen, T.E., Skov, M.N., Johansen, I.S. & Madsen, L.W. Prolonged viral shedding of SARS-CoV-2 in two immunocompromised patients, a case report. *BMC Infect Dis* **21**, 743 (2021), DOI: 10.1186/s12879-021-06429-5.
33. Nakajima, Y., Ogai, A., Furukawa, K., Arai, R., Anan, R., Nakano, Y., Kurihara, Y., Shimizu, H., Misaki, T. & Okabe, N. Prolonged viral shedding of SARS-CoV-2 in an immunocompromised patient. *J Infect Chemother* **27**, 387-389 (2021), DOI: 10.1016/j.jiac.2020.12.001.

34. Wei, L., Liu, B., Zhao, Y. & Chen, Z. Prolonged shedding of SARS-CoV-2 in an elderly liver transplant patient infected by COVID-19: a case report. *Ann Palliat Med* **10**, 7003-7007 (2021), DOI: 10.21037/apm-20-996.
35. Havervall, S., Marking, U., Svensson, J., Greilert-Norin, N., Bacchus, P., Nilsson, P., Hober, S., Gordon, M., Blom, K., Klingström, J., Åberg, M., Smed-Sörensen, A. & Thålin, C. Anti-Spike Mucosal IgA Protection against SARS-CoV-2 Omicron Infection. *N Engl J Med* **387**, 1333-1336 (2022), DOI: 10.1056/NEJMc2209651.
36. Miyamoto, S., Nishiyama, T., Ueno, A., Park, H., Kanno, T., Nakamura, N., Ozono, S., Aihara, K., Takahashi, K., Tsuchihashi, Y., Ishikane, M., Arashiro, T., Saito, S., Ainai, A., Hirata, Y., Iida, S., Katano, H., Tobiume, M., Tokunaga, K., Fujimoto, T., Suzuki, M., Nagashima, M., Nakagawa, H., Narita, M., Kato, Y., Igari, H., Fujita, K., Kato, T., Hiyama, K., Shindou, K., Adachi, T., Fukushima, K., Nakamura-Uchiyama, F., Hase, R., Yoshimura, Y., Yamato, M., Nozaki, Y., Ohmagari, N., Suzuki, M., Saito, T., Iwami, S. & Suzuki, T. Infectious virus shedding duration reflects secretory IgA antibody response latency after SARS-CoV-2 infection. *Proc Natl Acad Sci U S A* **120**, e2314808120 (2023), DOI: 10.1073/pnas.2314808120.
37. Rakhmetullina, A., Pyrkova, A., Aisina, D. & Ivashchenko, A. In silico prediction of human genes as potential targets for rice miRNAs. *Comput Biol Chem* **87**, 107305 (2020), DOI: 10.1016/j.compbiolchem.2020.107305.
38. Garrett, N., Tapley, A., Andriesen, J., Seocharan, I., Fisher, L.H., Bunts, L., Espy, N., Wallis, C.L., Randhawa, A.K., Ketter, N., Yacovone, M., Goga, A., Bekker, L.G., Gray, G.E. & Corey, L. High Rate of Asymptomatic Carriage Associated with Variant Strain Omicron. *medRxiv* (2022), DOI: 10.1101/2021.12.20.21268130.
39. Kentaro Itokawa, T.S., Masanori Hashino, Rina Tanaka, Makoto Kuroda. nCoV-2019 sequencing protocol for illumina. *protocols.io* (2020), DOI: 10.17504/protocols.io.bnn7mdhn.
40. Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W. & Rajewsky, N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res* **40**, 37-52 (2012), DOI: 10.1093/nar/gkr688.
41. Samson, A., Lavielle, M. & Mentré, F. Extension of the SAEM algorithm to left-censored data in nonlinear mixed-effects model: Application to HIV dynamics model. *Computational Statistics & Data Analysis* **51**, 1562-1574 (2006), DOI.
42. Hastie, T., Tibshirani, R., Friedman, J.H. & Friedman, J.H. *The elements of statistical learning: data mining, inference, and prediction*, (Springer, 2009).

ACKNOWLEDGMENTS

This study was supported in part by AMED Research Program 24fk0210154h0001 (to. H.P.); National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2022R1C1C2003637) (to K.S.K.); Grant-in-Aid for JSPS Scientific Research (KAKENHI) B 23H03497 (to S.I.); Grant-in-Aid for Transformative Research Areas 22H05215 (to S.I.); Grant-in-Aid for Challenging Research (Exploratory) 22K19829 (to S.I.); AMED CREST 19gm1310002 (to S.I.); AMED Research Program on Emerging and Re-emerging Infectious Diseases 22fk0108509 (to S.I.), 23fk0108684 (to S.I.), 23fk0108685 (to S.I.); AMED Research Program on HIV/AIDS 22fk0410052 (to S.I.); AMED Program for Basic and Clinical Research on Hepatitis 22fk0210094 (to S.I.); AMED Program on the Innovative Development and the Application of New Drugs for Hepatitis B 22fk0310504h0501 (to S.I.); AMED Strategic Research Program for Brain Sciences 22wm0425011s0302; AMED JP22dm0307009 (to K.A.); JST MIRAI JPMJMI22G1 (to S.I.); Moonshot R&D JPMJMS2021 (to K.A. and S.I.) and JPMJMS2025 (to S.I.); JST CREST JPMJCR25Q6 (to S.I.); the Cabinet Agency for Infectious Disease Crisis Management (to S.I.); the JIHS Intramural Research Fund (24 rin 002) (to S.I.); Institute of AI and Beyond at the University of Tokyo (to K.A.); Shin-Nihon of Advanced Medical Research (to S.I.); The Japan Prize Foundation (to S.I.).

AUTHOR CONTRIBUTIONS

YM, KW, SI, and TM designed the research. HP, RY, KSK, KE, NN, SIwanami, KA, TU, KM, TMorita, CBB, RK and SI carried out the computational analysis. SI and TM supervised the project. All authors contributed to writing the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no conflicts of interest associated with this manuscript.

INSTITUTIONAL REVIEW BOARD STATEMENT

This study was approved by the ethics committees of Nagoya University (hc22-01).