



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

Hyeongki Park, Raiki Yoshimura, Shoya Iwanami, Kwang Su Kim, Keisuke Ejima, Naotoshi Nakamura, Kazuyuki Aihara, Yoshitsugu Miyazaki, Takashi Umeyama, Ken Miyazawa, Takeshi Morita, Koichi Watashi, Christopher B. Brooke, Ruian Ke, Shingo Iwami , Taiga Miyazaki 

Reviewed Preprint

Published from the original preprint after peer review and assessment by eLife.

About eLife's process

Reviewed preprint version 1

April 11, 2024 (this version)

Posted to preprint server

February 2, 2024

Sent for peer review

January 30, 2024

interdisciplinary Biology Laboratory (iBLab), Division of Natural Science, Graduate School of Science, Nagoya University, Nagoya, Japan • Department of Science System Simulation, Pukyong National University, Busan, South Korea • Department of Mathematics, Pusan National University, Busan, South Korea • Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore • The Tokyo Foundation for Policy Research, Tokyo, Japan • International Research Center for Neurointelligence, The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Tokyo, Japan • Department of Chemotherapy and Mycoses, National Institute of Infectious Diseases, Tokyo, Japan • Research Center for Drug and Vaccine Development, National Institute of Infectious Diseases, Tokyo, Japan • Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL, USA • Department of Statistics, University of Illinois at Urbana-Champaign, Urbana, IL, USA • Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM, USA • Institute of Mathematics for Industry, Kyushu University, Fukuoka, Japan • Institute for the Advanced Study of Human Biology (ASHBi), Kyoto University, Kyoto, Japan • Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS), RIKEN, Saitama, Japan • NEXT-Ganken Program, Japanese Foundation for Cancer Research (JFCR), Tokyo, Japan • Science Groove Inc., Fukuoka, Japan • Division of Respiratory, Rheumatology, Infectious Diseases, and Neurology, Department of Internal Medicine, Faculty of Medicine, University of Miyazaki, Miyazaki, Japan

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

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 successfully stratified infection dynamics into three distinct groups with clear patterns of viral shedding: viral shedding durations in the three groups 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), 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 shedding 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.

eLife assessment

This **useful** work by Park attempts to use machine learning algorithms to identify correlates of different observed shedding patterns in two COVID-19 cohorts. The evidence supporting the study conclusions is **incomplete** due to the lack of uniformity in assays between the 2 cohorts, relevant metadata (previous infection/vaccination status, viral variant), early viral load data in the cohorts, and **incomplete** statistical analyses. With a strengthened analysis, the work may be of interest to virologists, clinicians, and public health scientists.

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-2³, and viral particles in the lower and upper respiratory tract can reach the oral cavity through liquid droplets¹. Saliva droplets are thus a potential route of SARS-CoV-2 transmission^{1,3}. Although SARS-CoV-2 infection dynamics within the respiratory tract are well characterized thanks to data from oropharyngeal and nasopharyngeal swabs^{4–9}, infection dynamics in saliva are poorly understood. Recent reports^{3,5} have suggested that saliva and other tissues are highly compartmentalized and that SARS-CoV-2 infection dynamics are qualitatively different in different tissues. 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}. Thus, 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 3 samples from 144 mildly symptomatic participants of two different but complementary cohorts^{5,11,12}. We successfully identified three groups with significantly different durations of viral shedding (i.e., the mean durations were 11.5 days, 17.4 days, and 30.0 days, respectively), implying 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-Champaign⁵. All infections were either mild or asymptomatic. All participants in these cohorts confirmed that they had never been previously infected with SARS-CoV-2, and none were vaccinated against SARS-CoV-2 at the time of enrolment. 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, as summarized in **Fig 1B**, **Table 1**, and **Supplementary Table 1**. 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**). 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.

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 (**Extended Data Fig 1** and **Supplementary Table 2**). We also applied the more realistic mathematical model used in Ke et al.⁵ [i.e., Eqs.(3-6)], but the fitting by these two mathematical models reconstructed our dataset equally (**Supplementary Fig 1** and **Supplementary Table 3**). This is because the longitudinal data from the NFV clinical trial included the late phase of the viral RNA load, which was not included in the data from the University of Illinois; therefore, Eqs.(1-2) are sufficient for the purpose of reconstruction of viral load. To avoid complexities and assumptions on parameters of Eqs.(3-6) as described in Ke et al.⁵, we used our simple model of Eqs.(1-2) for further analysis.

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., **Extended Data 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

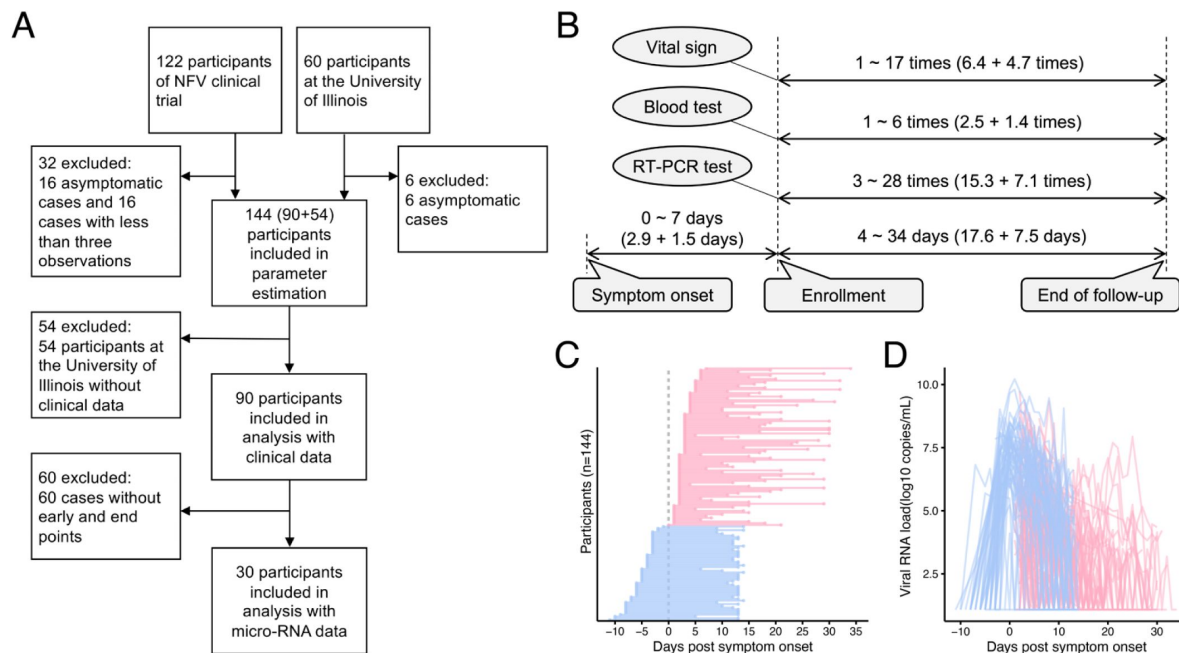


Figure 1

Characteristics of cohorts from the NFV clinical trial (jRCT2071200023¹¹[12](#)) and the University of Illinois⁵[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. (**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.

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)
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	μg/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].

Table 1

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

“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 twodimensional 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 timecourse pattern of each group by the Partial Least-Squares Discriminant Analysis (PLS-DA) (**Fig 2C** [↗](#)).

The distributions of the four features are described in **Fig 2D** [↗](#); a statistically significant between-group difference was found in the duration of viral shedding. 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 report¹⁰ [↗](#), we consistently confirmed that there were at least 3 groups showing different durations of viral shedding in upper respiratory specimens.

Because of previous work concluding that there are no significant differences in virus infection dynamics between NFV-treated and untreated participants¹² [↗](#), 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 [↗](#).

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 data¹⁴ [↗](#), 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).

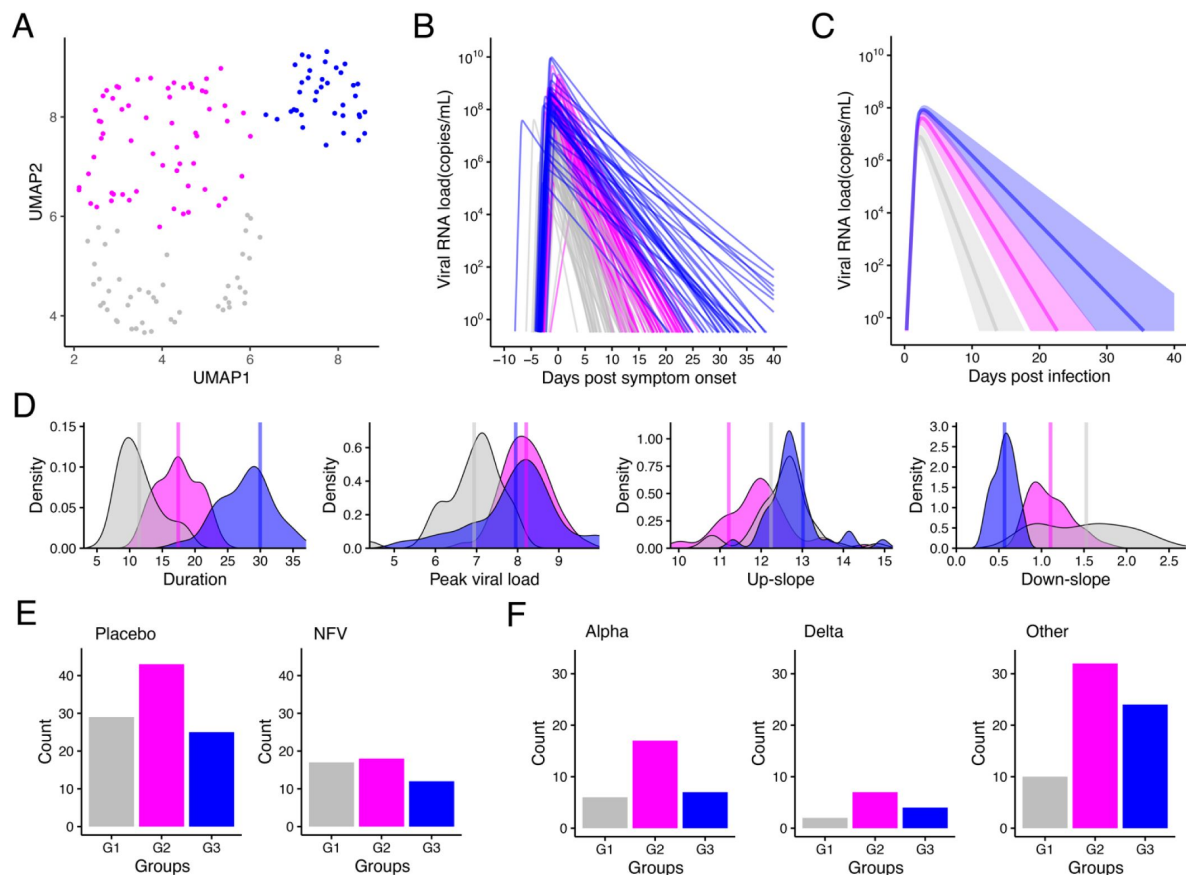


Figure 2

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.

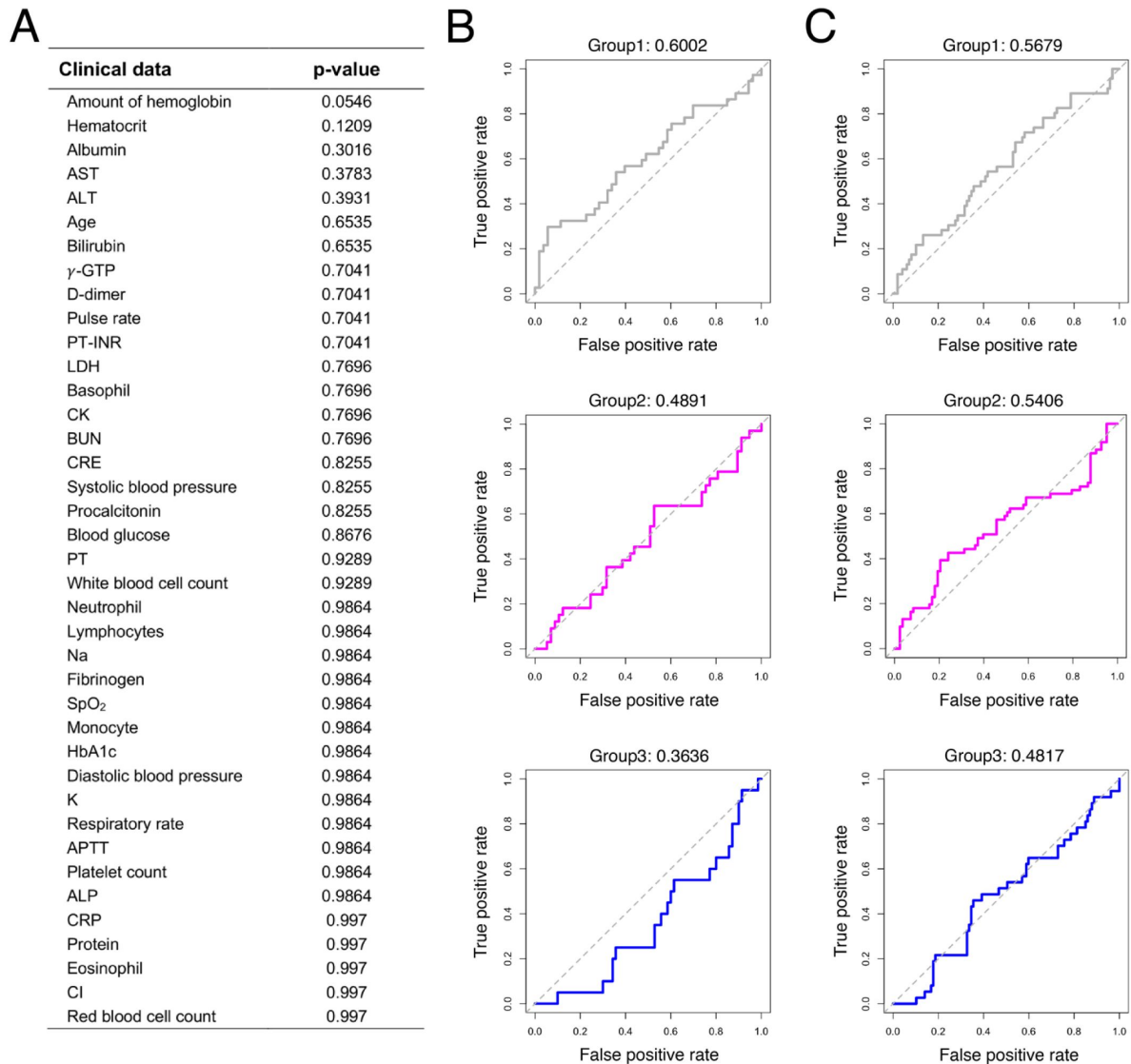


Figure 3

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 39 clinical values and 8 daily symptoms, respectively. The corresponding AUC (area under curve) value of each ROC curve is shown on the top of each panel.

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** [↗](#)). 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 symptoms¹⁵ [↗](#).

Additionally, we investigated the relationship between each feature of viral dynamics (i.e., duration of viral shedding, peak viral load, up-slope, and down-slope) and the clinical data by using the Pearson's correlation coefficient (**Extended Data Table 1** [↗](#)). 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.

Relationship between salivary microRNAs 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 replication¹⁷ [↗](#) and could thus act as a biomarker to predict COVID-19 disease progression¹⁶ [↗](#). 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 5** [↗](#).

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 (**Extended Data Table 2** [↗](#)). 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 shedding, 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

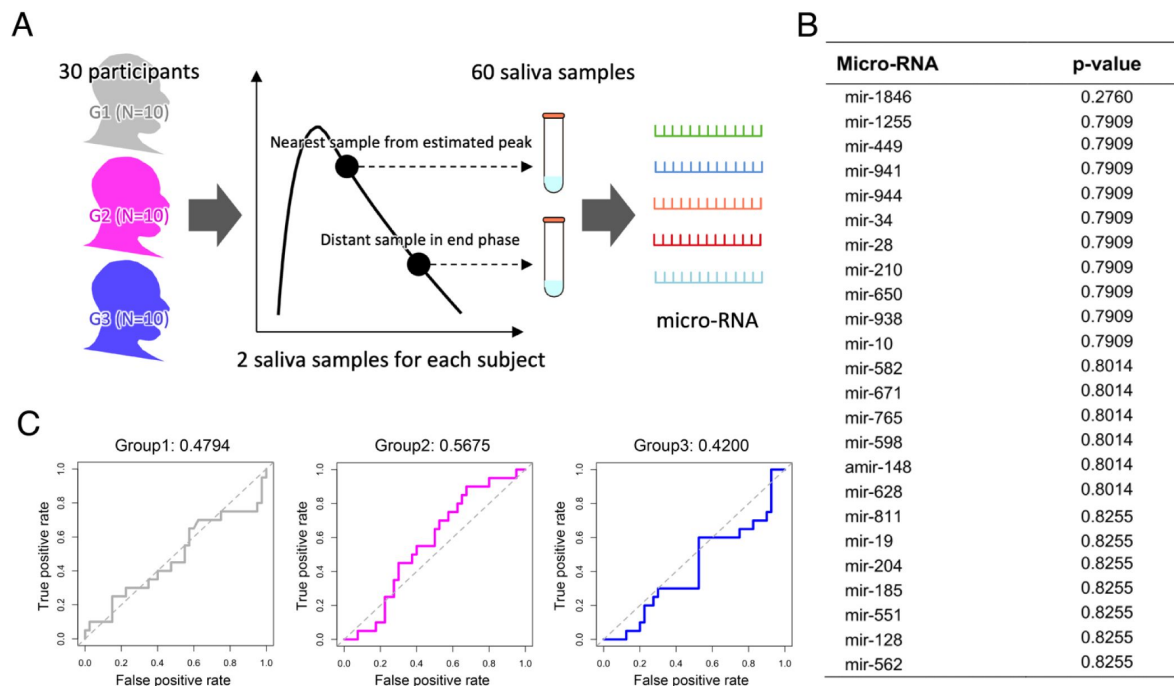


Figure 4

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.

coefficient with 0.01 p-value) with the peak viral load (**Extended Data Figure 2**). We confirmed similar trends even when we analyzed the micro-RNA level for the peak and late phases separately.

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 workers¹, saliva sampling has excellent potential and advantages over other sampling methods from biological specimens such as the lower and upper respiratory tract^{2,17}. Given the significant individual heterogeneity in the saliva viral shedding^{5,18}, 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 trial¹¹ and the University of Illinois at Urbana-Champaign⁵. 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 shedding, lasting over three months, underscoring the critical role of host immune responses in controlling viral infections^{19–22}. 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 saliva³. 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 elsewhere²³. Recently, we demonstrated that rapid anti-spike secretory IgA antibody responses can contribute to reducing viral shedding durations and amounts in nasopharyngeal mucosa²⁴. These findings highlight the importance of biomarkers that directly reflect an individual's immune response, such as antibody induction, 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**, **Supplementary Table 2** and **Supplementary Fig 2**). 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)²⁵, 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 genes²⁵ 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 infection²⁶. Thus, evaluating the effect of asymptomatic infection will be important to update our stratification, especially for recent (or future emerging) VOCs. Second, 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 shedding and mucosal immunity as discussed above, it appears imperative to analyze modalities that are directly linked to the immune response 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 shedding. 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 trial¹¹. 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-Champaign⁵. 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. Among those 182 cases, we focused only on symptomatic participants. 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). This decision was based on the understanding that reasonable estimates cannot be derived when the number of data

points less than the number of parameters in the mathematical model. 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.

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 previously²⁷. 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 miRDeep2²⁸. 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 protocol¹¹. 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}:

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

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

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 1**):

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

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

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

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

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 $t_1 < t$. 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 fact, in our study, the saliva specimens of cases from the cohort of the University of Illinois and the NFV clinical trial were obtained mainly during the “early” (i.e., from the date of infection to no more than 14 days post symptom onset) and “late” (i.e., no more than 28 days from around symptom onset) phases of SARS-CoV-2 infection, respectively. Therefore, cohorts from these two studies can reveal the whole time-course pattern of SARS-CoV-2 infection dynamics in a complementary manner through the random effects of a nonlinear mixed-effects modelling.

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 **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)²⁹. 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 shedding, 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 performance³⁰. 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 twodimensional plane and was stratified with spectral clustering (Python scikit-learn). The optimal number of clusters was determined by the eigengap heuristic method.

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).

Acknowledgements

This study was supported in part by a National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2022R1C1C2003637) (to K.S.K.); 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.); Institute of AI and Beyond at the University of Tokyo (to K.A.); Shin-Nihon of Advanced Medical Research (to S.I.); SECOM Science and Technology Foundation (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.

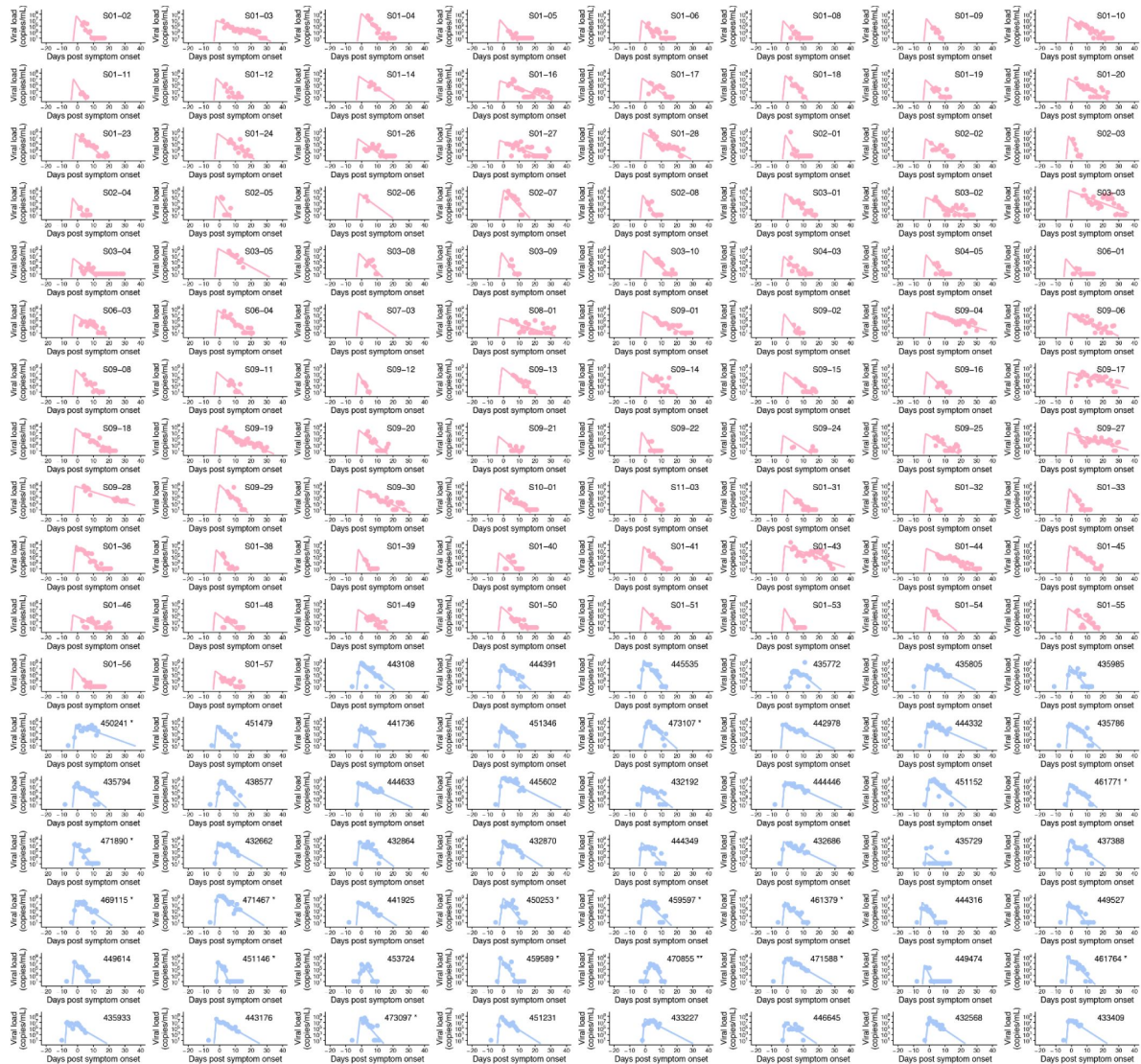
Competing financial interests

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).

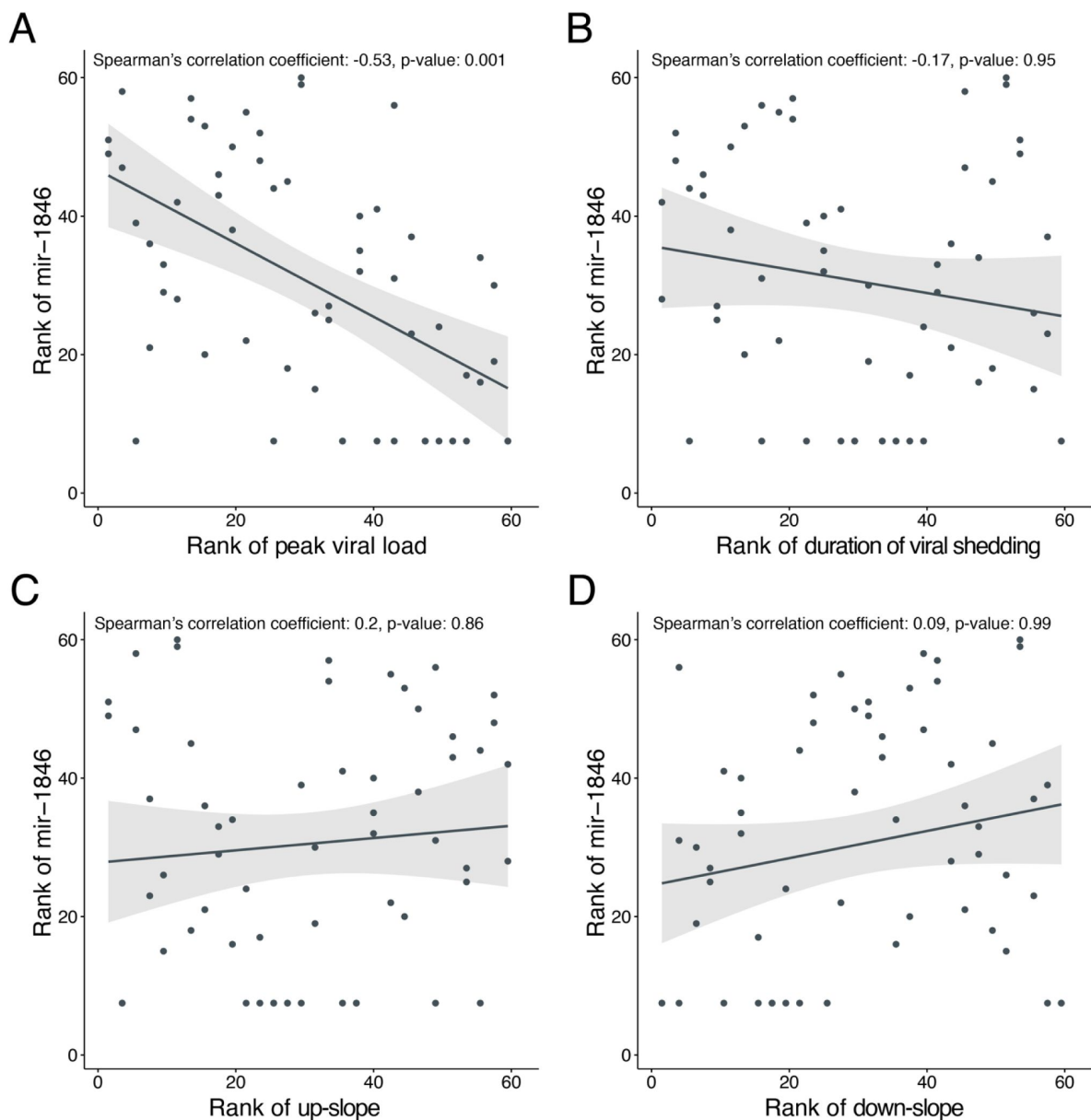
Extended data figures



Extended Data Figure 1.

Reconstructed viral dynamics in saliva samples for individual participants:

The individual-level model fits to saliva RT-qPCR results using the target-celllimited model described in [Eqs.\(1-2\)](#), for the same cohorts described in [Fig. 1](#), are shown. The closed dots and the solid curves indicate the measured data and the estimated viral dynamics, respectively. Individuals from the NFV clinical trial and the University of Illinois are shown in red and blue, respectively.



Extended Data Figure 2

Correlation between mir-1846 level and the features of infection dynamics:

The correlation between the rank of mir-1846 and the rank of (A) the peak viral load, (B) the duration of viral shedding, (C) up-slope and (D) down-slope are shown, respectively. The Spearman's correlation coefficients and p-values are presented in the top of each panel. The black solid line and the gray shaded area indicate results of the linear regression and their 95% confidence levels, respectively.

Clinical data	Duration of viral shedding		Peak viral load		Up-slope		Down-slope	
	Pearson's <i>r</i>	p-value	Pearson's <i>r</i>	p-value	Pearson's <i>r</i>	p-value	Pearson's <i>r</i>	p-value
Age	-0.08	0.8359	-0.16	0.5783	0.15	0.5880	0.08	0.8141
Systolic blood pressure	0.08	0.8359	-0.04	0.9388	0.12	0.6448	-0.16	0.4519
Diastolic blood pressure	-0.06	0.8359	-0.2	0.5152	0.14	0.5880	-0.03	0.9204
Pulse rate	-0.13	0.6307	-0.05	0.9388	0	0.9710	0.01	0.9445
SpO ₂	-0.04	0.8359	-0.06	0.9388	0.07	0.9178	0.01	0.9512
Respiratory rate	0	0.9770	0.02	0.9765	0.08	0.9178	0.04	0.8681
White blood cell count	-0.07	0.8359	0	0.9868	-0.01	0.9710	0.07	0.8610
Neutrophil	-0.04	0.8359	-0.02	0.9765	0.15	0.5880	0.05	0.8681
Eosinophil	-0.04	0.8359	-0.24	0.4083	0.16	0.5880	-0.08	0.8386
Basophil	0.09	0.8359	-0.05	0.9388	-0.15	0.5880	-0.14	0.4933
Lymphocytes	0.01	0.9770	0.04	0.9388	-0.15	0.5880	-0.02	0.9231
Monocyte	0.11	0.6993	0.08	0.8939	-0.09	0.9166	-0.06	0.8681
Red blood cell count	0.08	0.8359	0.07	0.8939	0.01	0.9710	0.01	0.9445
Amount of hemoglobin	0.26	0.1688	0.26	0.4083	0	0.9710	-0.2	0.3194
Hematocrit	0.25	0.1688	0.18	0.5209	0.07	0.9178	-0.2	0.3194
Platelet count	-0.05	0.8359	-0.13	0.6790	0.03	0.9710	-0.04	0.8681
CRP	-0.06	0.8359	0.01	0.9765	0.02	0.9710	0.13	0.4933
Protein	0	0.9770	-0.1	0.8684	0.08	0.9178	-0.03	0.9200
Albumin	0.22	0.1688	0.21	0.5152	-0.15	0.5880	-0.17	0.3756
ALT	0.24	0.1688	0.04	0.9388	0.14	0.5880	-0.24	0.2570
AST	0.26	0.1688	0.16	0.5783	0.1	0.8732	-0.23	0.2570
γ-GTP	0.16	0.4469	-0.07	0.8939	0.18	0.5880	-0.21	0.3194
ALP	0.21	0.1791	-0.07	0.8939	0.26	0.4708	-0.27	0.2570
LDH	0.12	0.6676	-0.01	0.9765	0.06	0.9178	-0.15	0.4661
Bilirubin	0.22	0.1688	0.15	0.5783	0.03	0.9710	-0.19	0.3194
CK	0.08	0.8359	0.01	0.9765	-0.04	0.9710	-0.12	0.4971
Na	-0.02	0.9374	-0.12	0.7050	0.02	0.9710	-0.05	0.8681
K	-0.02	0.9374	-0.08	0.8939	0.06	0.9178	0.07	0.8610
Cl	-0.07	0.8359	-0.19	0.5209	0.06	0.9178	-0.04	0.8681
BUN	0.22	0.1688	-0.03	0.9388	0.13	0.6448	-0.25	0.2570
CRE	0.22	0.1688	0.18	0.5209	0.03	0.9710	-0.14	0.4933
Blood glucose	0.04	0.8359	-0.12	0.7050	0.15	0.5880	-0.06	0.8681
HbA1c	-0.04	0.8359	-0.13	0.6790	0.07	0.9178	0.04	0.8681
Procalcitonin	-0.14	0.5327	0.05	0.9388	-0.18	0.5880	0.19	0.3194
Fibrinogen	-0.08	0.8359	0.02	0.9765	0.04	0.9710	0.13	0.4933
PT	0.16	0.4469	0.03	0.9388	-0.02	0.9710	-0.12	0.4971
APTT	-0.05	0.8359	0.08	0.8939	0.01	0.9710	0.13	0.4971
PT-INR	-0.22	0.1688	-0.01	0.9765	-0.01	0.9710	0.18	0.3655
D-dimer	-0.06	0.8359	-0.13	0.6790	0.13	0.6448	-0.01	0.9512

Extended Data Table 1

Pearson's correlation coefficients between clinical data and features of viral dynamics.

Micro-RNA	Duration of viral shedding		Peak viral load		Up-slope		Down-slope	
	Spearman's ρ	p-value	Spearman's ρ	p-value	Spearman's ρ	p-value	Spearman's ρ	p-value
mir-10	-0.16	0.9507	-0.31	0.2323	0.06	0.944	0.07	0.9867
mir-103	0.14	0.9507	0.06	0.8198	0.07	0.9321	-0.16	0.9867
mir-1180	-0.08	0.9982	0.2	0.597	-0.16	0.8616	0.12	0.9867
mir-1226	-0.03	0.9982	0.09	0.7441	0	0.993	0.03	0.9867
mir-1253	-0.16	0.9507	0.02	0.9244	-0.17	0.8616	0.17	0.9867
mir-1255	-0.33	0.6991	-0.09	0.7441	-0.08	0.9247	0.34	0.7121
mir-128	0.04	0.9982	0.12	0.6894	-0.09	0.9247	-0.01	0.9867
mir-1296	-0.01	0.9982	0.17	0.6225	0.02	0.9497	0.02	0.9867
mir-130	0.13	0.9507	0.18	0.597	0.02	0.9497	-0.07	0.9867
mir-145	0.12	0.9553	0.01	0.9822	0.08	0.9247	-0.12	0.9867
mir-146	0.17	0.9507	0.16	0.635	0.09	0.9247	-0.15	0.9867
mir-148	-0.02	0.9982	0.25	0.4697	-0.11	0.9029	0.04	0.9867
mir-154	-0.03	0.9982	0.14	0.635	-0.18	0.8616	0.07	0.9867
mir-16	0.25	0.7309	0.07	0.7731	0.18	0.8616	-0.27	0.9509
mir-17	0.02	0.9982	0.04	0.87	-0.04	0.9497	-0.04	0.9867
mir-181	0.08	0.9982	0.06	0.8198	0	0.9925	-0.07	0.9867
mir-185	-0.07	0.9982	-0.19	0.597	0.08	0.9247	0	0.9918
mir-19	-0.09	0.9982	0.25	0.4697	-0.16	0.8616	0.16	0.9867
mir-191	-0.01	0.9982	-0.14	0.635	0.12	0.9029	-0.04	0.9867
mir-192	-0.03	0.9982	-0.06	0.8198	0.19	0.8616	0.01	0.9867
mir-193	-0.17	0.9507	-0.1	0.7441	0.02	0.9497	0.16	0.9867
mir-204	-0.02	0.9982	0.28	0.3167	-0.25	0.8616	0.06	0.9867
mir-210	-0.04	0.9982	0.21	0.597	-0.21	0.8616	0.06	0.9867
mir-214	0.14	0.9507	-0.02	0.9244	0.14	0.9029	-0.13	0.9867
mir-218	-0.16	0.9507	0.02	0.9244	-0.17	0.8616	0.17	0.9867
mir-224	0.22	0.8213	0.16	0.635	0.22	0.8616	-0.2	0.9867
mir-23	-0.02	0.9982	0.08	0.7719	0.11	0.9029	0.03	0.9867
mir-24	-0.1	0.9982	-0.08	0.7731	0.05	0.944	0.11	0.9867
mir-26	0.07	0.9982	-0.06	0.8198	0.01	0.9736	-0.09	0.9867
mir-28	-0.27	0.6991	-0.11	0.6894	-0.09	0.9247	0.23	0.9867
mir-29	-0.07	0.9982	0	0.9926	-0.12	0.9029	0.06	0.9867
mir-290	0.2	0.9175	0.11	0.6894	0.19	0.8616	-0.17	0.9867
mir-296	-0.02	0.9982	0.08	0.7731	-0.15	0.8892	0.04	0.9867
mir-3180	-0.02	0.9982	-0.2	0.597	0.12	0.9029	-0.01	0.9867
mir-320	-0.01	0.9982	-0.12	0.6615	-0.03	0.9497	0	0.9979
mir-33	0	0.9982	-0.05	0.8198	0.13	0.9029	0.01	0.9867
mir-331	-0.03	0.9982	-0.05	0.8373	-0.05	0.944	-0.01	0.9867
mir-338	0	0.9982	0.16	0.635	0.08	0.9247	0.02	0.9867
mir-34	0.27	0.6991	0.12	0.6615	0.12	0.9029	-0.24	0.9867
mir-340	-0.13	0.9507	-0.11	0.6894	0.05	0.944	0.11	0.9867
mir-342	0.01	0.9982	0.2	0.597	-0.18	0.8616	0.01	0.9867
mir-345	-0.15	0.9507	-0.13	0.6517	-0.08	0.9247	0.11	0.9867
mir-365	0.06	0.9982	0	0.9926	0.09	0.9247	-0.09	0.9867
mir-375	0.13	0.9507	0.05	0.8198	0.07	0.9321	-0.09	0.9867
mir-425	-0.11	0.985	-0.14	0.635	0.01	0.9724	0.06	0.9867
mir-449	0.29	0.6991	0.14	0.635	0.17	0.8616	-0.27	0.9509
mir-489	-0.11	0.985	0.09	0.7441	-0.2	0.8616	0.14	0.9867
mir-491	0.07	0.9982	0.14	0.635	-0.08	0.9247	-0.07	0.9867
mir-492	0.16	0.9507	-0.01	0.9867	0.17	0.8616	-0.14	0.9867
mir-500	-0.03	0.9982	0.16	0.635	-0.08	0.9247	0.07	0.9867
mir-505	0	0.9982	-0.22	0.597	0.16	0.8616	-0.07	0.9867
mir-544	0.11	0.985	-0.05	0.8198	-0.02	0.9497	-0.13	0.9867
mir-548	0.03	0.9982	-0.11	0.6894	0.02	0.9497	-0.04	0.9867
mir-550	-0.1	0.9982	-0.02	0.9244	-0.03	0.9497	0.14	0.9867
mir-551	-0.05	0.9982	0.17	0.6225	-0.17	0.8616	0.06	0.9867
mir-552	-0.09	0.9982	-0.07	0.8012	-0.02	0.9497	0.09	0.9867
mir-562	-0.09	0.9982	-0.17	0.6207	-0.08	0.9247	0.06	0.9867
mir-572	0.04	0.9982	0.13	0.6605	-0.1	0.9247	-0.02	0.9867
mir-574	-0.14	0.9507	-0.18	0.597	-0.02	0.9497	0.08	0.9867
mir-576	0	0.9982	0.09	0.7511	-0.07	0.9331	0.02	0.9867
mir-581	0.17	0.9507	0.14	0.635	0.13	0.9029	-0.17	0.9867
mir-582	0.07	0.9982	0.2	0.597	-0.17	0.8616	-0.01	0.9867
mir-584	0.15	0.9507	0.04	0.8503	-0.07	0.9331	-0.16	0.9867
mir-589	0.06	0.9982	0.13	0.6517	0.06	0.944	-0.03	0.9867
mir-597	-0.03	0.9982	-0.1	0.724	0.05	0.944	0.03	0.9867
mir-598	-0.08	0.9982	-0.32	0.2323	0.11	0.9029	0	0.9918
mir-612	-0.07	0.9982	0.1	0.724	-0.17	0.8616	0.1	0.9867
mir-616	0.13	0.9507	0.19	0.597	-0.07	0.936	-0.08	0.9867
mir-618	-0.13	0.9507	0.02	0.9353	-0.05	0.944	0.13	0.9867

Extended Data Table 2

Spearman's correlation coefficients between micro-RNA data and features of viral dynamics.

mir-624	0.03	0.9982	0.18	0.597	-0.13	0.9029	0.01	0.9867
mir-628	0.03	0.9982	0.18	0.597	-0.08	0.9247	0.01	0.9867
mir-632	0.08	0.9982	0.29	0.3167	-0.13	0.9029	-0.02	0.9867
mir-636	0.15	0.9507	0.09	0.7511	0.04	0.9497	-0.14	0.9867
mir-639	-0.06	0.9982	-0.19	0.597	0.2	0.8616	-0.01	0.9867
mir-642	-0.02	0.9982	0.09	0.7511	0.05	0.944	0.03	0.9867
mir-643	-0.02	0.9982	0.24	0.4955	-0.17	0.8616	0.05	0.9867
mir-650	-0.01	0.9982	0.07	0.8012	-0.12	0.9029	0.03	0.9867
mir-651	-0.01	0.9982	0.09	0.7511	-0.04	0.9497	0.05	0.9867
mir-671	-0.13	0.9507	-0.05	0.8198	0.04	0.944	0.1	0.9867
mir-672	0.01	0.9982	0.2	0.597	-0.18	0.8616	0.01	0.9867
mir-675	-0.22	0.8213	-0.14	0.635	0.1	0.9247	0.22	0.9867
mir-692	-0.03	0.9982	0.05	0.8198	-0.2	0.8616	0.05	0.9867
mir-720	0.12	0.9553	0.14	0.635	0	0.993	-0.1	0.9867
mir-744	-0.05	0.9982	-0.2	0.597	0.06	0.944	-0.03	0.9867
mir-765	0.25	0.7309	-0.08	0.7559	0.38	0.2349	-0.26	0.9509
mir-885	-0.07	0.9982	0	0.9926	0.01	0.9724	0.09	0.9867
mir-887	0.09	0.9982	0.14	0.635	0.03	0.9497	-0.06	0.9867
mir-938	0.22	0.8213	0.14	0.635	0.14	0.9029	-0.17	0.9867
mir-941	-0.22	0.8213	-0.37	0.1235	0.12	0.9029	0.08	0.9867
mir-944	0.28	0.6991	0.37	0.1235	-0.01	0.9925	-0.2	0.9867
mir-1846	-0.17	0.9507	-0.53	0.0011	0.2	0.8616	0.09	0.9867
mir-811	-0.13	0.9507	-0.31	0.2323	0.05	0.944	0.06	0.9867

Extended Data Table 2 (continued)

Supplementary Information



Supplementary Figure 1.

Comparison of three model fits to viral load in saliva samples for individual participants:

Three different individual-level model fits to saliva RT-qPCR results using the target-cell-limited model and the immune effector model described in [Eqs.\(1-2\)](#) and [Eqs.\(3-6\)](#), respectively, are shown. The black closed dots and solid curves indicate the measured viral load and the expected viral dynamics using the target-cell-limited model, respectively (**Expanded Data Fig 1**). The red and blue dashed curves indicate estimated viral dynamics using the immune effector model for all the 144 individuals described in [Fig 1](#) and for the 54 individuals in the cohort from the University of Illinois, respectively. Namely, the blue dashed curves are the model fits presented in [1](#).

Daily symptom	Group1 (N=46)	Group2 (N=61)	Group3 (N=37)	Overall (N=144)
Cough	84.8%	75.4%	64.9%	75.7%
Dyspnea	32.6%	24.6%	37.8%	30.6%
Nasal discharge	58.7%	67.2%	59.5%	62.5%
Sore throat	63.0%	65.6%	59.5%	63.2%
Nausea/vomiting/diarrhea	52.2%	68.9%	54.1%	59.7%
Myalgia	39.1%	57.4%	45.9%	48.6%
Olfaction abnormal	41.3%	52.5%	59.5%	50.7%
Fever	45.7%	24.6%	32.4%	33.3%

Supplementary Table 1

Daily symptom data for whole cohorts for each group

Parameters	Description	Unit	ϑ : Fixed effect (SE)*	Ω : SD of random effect (SE)*
β	Rate constant for virus infection	(RNA copies/ml) ⁻¹ day ⁻¹	2.08×10^{-7} (6.79×10^{-8})	2.79 (0.26)
γ	Maximum rate constant for viral replication	Day ⁻¹	13.2 (1.44)	0.45 (0.08)
δ	Death rate of virus-producing cells	Day ⁻¹	0.97 (0.05)	0.54 (0.05)
τ	Days from infection to symptom onset	Days	3.73 (0.21)	0.31 (0.05)

* The parameter for patient k , $\vartheta_i (= \vartheta \times e^{\pi_k})$ is represented as a product of ϑ (a fixed effect) and e^{π_k} (a random effect). π_k follows the normal distribution with mean 0 and standard deviation Ω . SE: standard error.

Supplementary Table 2

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

Parameters	Description	Unit	ϑ : Fixed effect (SE)*	Ω : SD of random effect (SE)*
β	Rate constant for virus infection	(RNA copies/ml) ⁻¹ day ⁻¹	7.74×10^{-6} (2.59×10^{-6})	2.04 (0.27)
δ_1	Death rate of virus-producing cells	Day ⁻¹	0.58 (0.05)	0.52 (0.06)
δ_2	Death rate of virus-producing cells by immune effector cells	Day ⁻¹	1.09 (0.13)	0.6 (0.11)
t_1	Days from symptom onset when immune effector cells start to work	Days	10.5 (2.61)	1.44 (0.24)
π	Rate constant for virus production	Day ⁻¹	1.73 (0.59)	2.61 (0.31)
τ	Days from infection to symptom onset	Days	4.13 (0.28)	0.44 (0.06)

* The parameter for patient k , $\vartheta_i (= \vartheta \times e^{\pi_k})$ is represented as a product of ϑ (a fixed effect) and e^{π_k} (a random effect). π_k follows the normal distribution with mean 0 and standard deviation Ω . SE: standard error.

Supplementary Table 3

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

Feature	Group1 (N=46)	Group2 (N=61)	Group3 (N=37)	Overall (N=144)	p-value
Duration of virus shedding (days)	11.5 (3.16)*	17.4 (3.14)	30.0 (5.8)	18.8 (8.13)	<0.001
Peak viral load (RNA copies/ml)	6.94 (0.69)	8.21 (0.56)	7.95 (1.03)	7.74 (0.93)	0.020
Up-slope (day ⁻¹)	12.2 (1.65)	11.2 (1.85)	13 (1.18)	12 (1.79)	<0.001
Down-slope (day ⁻¹)	1.53 (0.55)	1.11 (0.24)	0.57 (0.13)	1.1 (0.51)	<0.001

* Mean (standard deviation)

Supplementary Table 4

Features of reconstructed individual viral dynamics for each group

Micro-RNA	Group1 (N=10)	Group2 (N=10)	Group3 (N=10)	Overall (N=30)
mir-10	45.4 (28.98)	27.5 (11.41)	34.25 (23.71)	35.72 (23.43)
mir-103	0.12 (0.27)	0.42 (0.85)	0.11 (0.23)	0.22 (0.54)
mir-1180	2.16 (4.39)	0.94 (1.3)	0.81 (1.45)	1.3 (2.79)
mir-1226	0.85 (1.3)	0.63 (0.85)	0.44 (0.74)	0.64 (0.99)
mir-1253	0.04 (0.16)	0 (0)	0 (0)	0.01 (0.09)
mir-1255	0.09 (0.23)	0 (0)	0 (0)	0.03 (0.14)
mir-128	0 (0)	0.03 (0.12)	0.01 (0.05)	0.01 (0.07)
mir-1296	1.25 (2.41)	0.41 (0.63)	0.46 (0.78)	0.71 (1.53)
mir-130	1.02 (1.98)	0.74 (0.91)	1.04 (1.12)	0.93 (1.4)
mir-145	0 (0)	0.01 (0.06)	0.02 (0.1)	0.01 (0.07)
mir-146	9.72 (4.52)	11.41 (6.49)	10.98 (6.22)	10.71 (5.76)
mir-148	28.7 (5.9)	31.47 (6.88)	29.53 (10.38)	29.9 (7.91)
mir-154	0.6 (1.95)	1 (1.68)	0.82 (1.88)	0.81 (1.82)
mir-16	174.2 (34.41)	178.55 (46.78)	184.2 (37.33)	178.98 (39.39)
mir-17	0 (0)	0.02 (0.11)	0 (0)	0.01 (0.06)
mir-181	0.93 (1.36)	0.99 (1.1)	1.28 (2.15)	1.07 (1.58)
mir-185	7.39 (7.39)	5.04 (2.31)	7.87 (13.71)	6.77 (9.02)
mir-19	14.13 (8.99)	17.23 (12.16)	11.49 (6.35)	14.28 (9.6)
mir-191	2.87 (5.2)	1.22 (1.54)	1.51 (1.68)	1.87 (3.3)
mir-192	0.14 (0.46)	0.07 (0.3)	0.04 (0.13)	0.08 (0.32)
mir-193	0.16 (0.7)	0 (0)	0 (0)	0.05 (0.4)
mir-204	0.21 (0.6)	0.61 (1.66)	0.18 (0.41)	0.33 (1.05)
mir-210	0.41 (1.5)	1.49 (3.46)	0.14 (0.4)	0.68 (2.23)
mir-214	0 (0)	0 (0)	0.01 (0.05)	0 (0.03)
mir-218	0.03 (0.15)	0 (0)	0 (0)	0.01 (0.09)
mir-224	0 (0)	0 (0)	0.04 (0.17)	0.01 (0.1)
mir-23	0.42 (0.97)	0.26 (0.6)	0.22 (0.41)	0.3 (0.69)
mir-24	0.21 (0.44)	0.13 (0.3)	0.1 (0.22)	0.15 (0.33)
mir-26	33.95 (15.04)	35.26 (13.11)	32.99 (7.73)	34.07 (12.18)
mir-28	8.2 (3.78)	7.68 (4.77)	10.76 (23.76)	8.88 (13.99)
mir-29	5.66 (3.95)	5.51 (2.56)	4.93 (2.81)	5.37 (3.12)
mir-290	0 (0)	0 (0)	0.03 (0.14)	0.01 (0.08)
mir-296	0 (0)	0.11 (0.49)	0 (0)	0.04 (0.28)
mir-3180	0.02 (0.11)	0 (0)	0.06 (0.28)	0.03 (0.17)
mir-320	0.74 (1.15)	0.47 (0.78)	0.82 (1.34)	0.68 (1.11)

Supplementary Table 5

Salivary micro-RNAs obtained from 60 specimens of 30 participants

mir-33	0.01 (0.06)	0.04 (0.17)	0.02 (0.1)	0.02 (0.12)
mir-331	5.34 (5.31)	4.07 (3.57)	3.35 (3.53)	4.25 (4.23)
mir-338	1.45 (1.66)	1.55 (2.17)	1.07 (1.21)	1.36 (1.71)
mir-34	0.51 (1.77)	0.49 (1.31)	1.83 (3.75)	0.94 (2.55)
mir-340	0.01 (0.05)	0 (0)	0 (0)	0 (0.03)
mir-342	0 (0)	0.02 (0.1)	0 (0)	0.01 (0.06)
mir-345	6.98 (7.32)	4.14 (3.53)	5.2 (5.86)	5.44 (5.81)
mir-365	0.44 (1.5)	0.33 (0.6)	0.35 (0.95)	0.37 (1.06)
mir-375	0.14 (0.32)	0.18 (0.38)	1.03 (3.51)	0.45 (2.06)
mir-425	6.94 (2.85)	5.84 (2.51)	7 (4.82)	6.59 (3.52)
mir-449	0.07 (0.29)	0.04 (0.12)	0.77 (2.29)	0.29 (1.36)
mir-489	0 (0)	0.03 (0.15)	0 (0)	0.01 (0.09)
mir-491	0 (0)	0.13 (0.58)	0 (0)	0.04 (0.34)
mir-492	0 (0)	0 (0)	1.59 (7.1)	0.53 (4.1)
mir-500	2.22 (2.69)	2.12 (1.59)	1.79 (1.9)	2.04 (2.08)
mir-505	3.22 (2.83)	2.68 (3.79)	2.91 (3.43)	2.94 (3.32)
mir-544	0.3 (1.34)	0.08 (0.25)	0.19 (0.68)	0.19 (0.87)
mir-548	0.32 (0.69)	0.31 (0.86)	0.81 (2.26)	0.48 (1.45)
mir-550	0.43 (1.1)	0.61 (1.47)	0.36 (0.84)	0.46 (1.15)
mir-551	0.01 (0.05)	0.05 (0.17)	0 (0)	0.02 (0.1)
mir-552	0.01 (0.06)	0 (0)	0 (0)	0 (0.04)
mir-562	0.05 (0.16)	0 (0)	0.6 (2.68)	0.22 (1.55)
mir-572	0 (0)	0.05 (0.23)	0 (0)	0.02 (0.13)
mir-574	20.93 (14.57)	15.29 (10.63)	16.91 (15.11)	17.71 (13.56)
mir-576	2.81 (5.95)	1.47 (1.5)	2.71 (4.5)	2.33 (4.36)
mir-581	0 (0)	0.01 (0.06)	0.04 (0.17)	0.02 (0.1)
mir-582	0.23 (0.45)	0.8 (1.24)	0.31 (0.41)	0.45 (0.82)
mir-584	1.27 (1.63)	2.77 (6.44)	2.41 (2.84)	2.15 (4.15)
mir-589	0.37 (0.73)	0.37 (0.68)	0.3 (0.52)	0.35 (0.64)
mir-597	0.04 (0.12)	0.08 (0.3)	0.19 (0.6)	0.1 (0.39)
mir-598	14.96 (15.14)	7.66 (6.68)	11.74 (10.09)	11.46 (11.4)
mir-612	0.03 (0.13)	0.02 (0.08)	0 (0)	0.02 (0.09)
mir-616	0.05 (0.17)	0.2 (0.4)	0.16 (0.41)	0.14 (0.34)
mir-618	1.2 (1.61)	1.32 (1.54)	0.66 (0.95)	1.06 (1.41)
mir-624	0.19 (0.39)	0.52 (0.95)	0.27 (0.5)	0.33 (0.67)
mir-628	0.9 (1.29)	2.33 (3.55)	0.67 (0.77)	1.3 (2.31)
mir-632	0 (0)	0.05 (0.23)	0.01 (0.06)	0.02 (0.14)
mir-636	0 (0)	0.08 (0.36)	0.03 (0.15)	0.04 (0.23)
mir-639	0.1 (0.46)	0 (0)	0 (0)	0.03 (0.26)
mir-642	1.05 (1.88)	1.27 (2.29)	0.46 (0.61)	0.93 (1.75)
mir-643	0.06 (0.25)	0.1 (0.27)	0.01 (0.06)	0.06 (0.21)
mir-650	0 (0)	0.23 (0.91)	0 (0)	0.08 (0.53)
mir-651	0.39 (1.08)	0.39 (1.31)	0.15 (0.33)	0.31 (0.99)
mir-671	2.89 (3.29)	1.47 (1.67)	1.71 (3.6)	2.02 (2.99)
mir-672	0 (0)	0.04 (0.2)	0 (0)	0.01 (0.12)
mir-675	0.01 (0.06)	0 (0)	0 (0)	0 (0.03)
mir-692	0.01 (0.06)	0.16 (0.4)	0.3 (1.34)	0.16 (0.8)
mir-720	0 (0)	0.02 (0.1)	3.17 (14.19)	1.07 (8.19)
mir-744	18.41 (20.9)	13.73 (12.2)	14.16 (16.82)	15.43 (16.86)
mir-765	0.13 (0.46)	0.09 (0.36)	0.17 (0.33)	0.13 (0.38)
mir-885	1.23 (1.96)	1.15 (2.34)	0.72 (0.98)	1.03 (1.83)
mir-887	0.02 (0.11)	0.02 (0.1)	0.02 (0.08)	0.02 (0.1)
mir-938	0 (0)	0 (0)	0.05 (0.19)	0.02 (0.11)
mir-941	13.61 (10.91)	7.37 (4.13)	7.17 (4.16)	9.38 (7.65)
mir-944	0.19 (0.45)	0.6 (0.96)	0.91 (1.27)	0.57 (0.99)
mir-1846	4.96 (6.31)	1.67 (5.15)	11.6 (27.8)	6.08 (16.96)
mir-811	18.1 (17.8)	12.64 (23.67)	39.74 (78.61)	23.49 (49.11)

* mean (standard deviation)

Supplementary Table 5 (continued)

References

1. Baghizadeh Fini M. (2020) **Oral saliva and COVID-19** *Oral Oncol* **108** <https://doi.org/10.1016/j.oraloncology.2020.104821>
2. Wyllie A. L., et al. (2020) **Saliva or Nasopharyngeal Swab Specimens for Detection of SARS-CoV-2** *N Engl J Med* **383**:1283–1286 <https://doi.org/10.1056/NEJMc2016359>
3. Huang N., et al. (2021) **SARS-CoV-2 infection of the oral cavity and saliva** *Nat Med* **27**:892–903 <https://doi.org/10.1038/s41591-021-01105-z>
4. Kim K. S., et al. (2021) **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** <https://doi.org/10.1371/journal.pbio.3001128>
5. Ke R., et al. (2022) **Daily longitudinal sampling of SARS-CoV-2 infection reveals substantial heterogeneity in infectiousness** *Nat Microbiol* **7**:640–652 <https://doi.org/10.1038/s41564-022-01105-z>
6. Jeong Y. D., et al. (2022) **Designing isolation guidelines for COVID-19 patients with rapid antigen tests** *Nat Commun* **13** <https://doi.org/10.1038/s41467-022-32663-9>
7. Goyal A., Cardozo-Ojeda E. F., Schiffer J. T (2020) **Potency and timing of antiviral therapy as determinants of duration of SARS-CoV-2 shedding and intensity of inflammatory response** *Sci Adv* **6** <https://doi.org/10.1126/sciadv.abc7112>
8. Marc A., et al. (2021) **Quantifying the relationship between SARS-CoV-2 viral load and infectiousness** *Elife* **10** <https://doi.org/10.7554/eLife.69302>
9. Neant N., et al. (2021) **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** <https://doi.org/10.1073/pnas.2017962118>
10. Iwanami S., et al. (2021) **Detection of significant antiviral drug effects on COVID-19 with reasonable sample sizes in randomized controlled trials: A modeling study** *PLoS Med* **18** <https://doi.org/10.1371/journal.pmed.1003660>
11. Hosogaya N., et al. (2021) **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** <https://doi.org/10.1186/s13063-021-05282-w>
12. Miyazaki T., et al. (2023) **A Multicenter Randomized Controlled Trial To Evaluate the Efficacy and Safety of Nelfinavir in Patients with Mild COVID-19** *Microbiol Spectr* **11** <https://doi.org/10.1128/spectrum.04311-22>
13. Trobaugh D. W., Klimstra W. B (2017) **MicroRNA Regulation of RNA Virus Replication and Pathogenesis** *Trends Mol Med* **23**:80–93 <https://doi.org/10.1016/j.molmed.2016.11.003>
14. Grinsztajn L., Oyallon E., Varoquaux G. (2022) **Why do tree-based models still outperform deep learning on tabular data?** *arXiv preprint arXiv:2207.08815*

15. Killingley B., et al. (2022) **Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults** *Nat Med* **28**:1031–1041 <https://doi.org/10.1038/s41591-022-01780-9>
16. Pimenta R., et al. (2021) **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 <https://doi.org/10.22099/mbrc.2021.40555.1631>
17. To K. K., et al. (2020) **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 [https://doi.org/10.1016/s1473-3099\(20\)30196-1](https://doi.org/10.1016/s1473-3099(20)30196-1)
18. Hay J. A., et al. (2022) **Quantifying the impact of immune history and variant on SARS-CoV-2 viral kinetics and infection rebound: a retrospective cohort study** *medRxiv* <https://doi.org/10.1101/2022.01.13.22269257>
19. Leung W. F., et al. (2022) **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 <https://doi.org/10.1016/j.ijid.2021.10.045>
20. Niyonkuru M., et al. (2021) **Prolonged viral shedding of SARS-CoV-2 in two immunocompromised patients, a case report** *BMC Infect Dis* **21** <https://doi.org/10.1186/s12879-021-06429-5>
21. Nakajima Y., et al. (2021) **Prolonged viral shedding of SARS-CoV-2 in an immunocompromised patient** *J Infect Chemother* **27**:387–389 <https://doi.org/10.1016/j.jiac.2020.12.001>
22. Wei L., Liu B., Zhao Y., Chen Z (2021) **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 <https://doi.org/10.21037/apm-20-996>
23. Havervall S., et al. (2022) **Anti-Spike Mucosal IgA Protection against SARS-CoV-2 Omicron Infection** *N Engl J Med* **387**:1333–1336 <https://doi.org/10.1056/NEJMc2209651>
24. Miyamoto S., et al. (2023) **Infectious virus shedding duration reflects secretory IgA antibody response latency after SARS-CoV-2 infection** *Proc Natl Acad Sci U S A* **120** <https://doi.org/10.1073/pnas.2314808120>
25. Rakhmetullina A., Pyrkova A., Aisina D., Ivashchenko A (2020) **In silico prediction of human genes as potential targets for rice miRNAs** *Comput Biol Chem* **87** <https://doi.org/10.1016/j.compbiolchem.2020.107305>
26. Garrett N., et al. (2022) **High Rate of Asymptomatic Carriage Associated with Variant Strain Omicron** *medRxiv* <https://doi.org/10.1101/2021.12.20.21268130>
27. Kentaro Itokawa T. S., Hashino Masanori, Tanaka Rina, Kuroda Makoto (2020) **nCoV-2019 sequencing protocol for illumina** *protocols.io* <https://doi.org/10.17504/protocols.io.bnn7mdhn>
28. Friedlander M. R., Mackowiak S. D., Li N., Chen W., Rajewsky N (2012) **miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades** *Nucleic Acids Res* **40**:37–52 <https://doi.org/10.1093/nar/gkr688>

29. Samson A., Lavielle M., Mentre F (2006) **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
30. Hastie T., Tibshirani R., Friedman J. H., Friedman J. H (2009) **The elements of statistical learning: data mining, inference, and prediction**
1. Ke R., et al. (2022) **Daily longitudinal sampling of SARS-CoV-2 infection reveals substantial heterogeneity in infectiousness** *Nat Microbiol* **7**:640–652 <https://doi.org/10.1038/s41564-022-01105-z>

Article and author information

Hyeongki Park

interdisciplinary Biology Laboratory (iBLab), Division of Natural Science, Graduate School of Science, Nagoya University, Nagoya, Japan

Raiki Yoshimura

interdisciplinary Biology Laboratory (iBLab), Division of Natural Science, Graduate School of Science, Nagoya University, Nagoya, Japan

Shoya Iwanami

interdisciplinary Biology Laboratory (iBLab), Division of Natural Science, Graduate School of Science, Nagoya University, Nagoya, Japan
ORCID iD: [0000-0003-1474-5348](https://orcid.org/0000-0003-1474-5348)

Kwang Su Kim

interdisciplinary Biology Laboratory (iBLab), Division of Natural Science, Graduate School of Science, Nagoya University, Nagoya, Japan, Department of Science System Simulation, Pukyong National University, Busan, South Korea, Department of Mathematics, Pusan National University, Busan, South Korea

Keisuke Ejima

Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore, The Tokyo Foundation for Policy Research, Tokyo, Japan

Naotoshi Nakamura

interdisciplinary Biology Laboratory (iBLab), Division of Natural Science, Graduate School of Science, Nagoya University, Nagoya, Japan
ORCID iD: [0000-0002-4798-3660](https://orcid.org/0000-0002-4798-3660)

Kazuyuki Aihara

International Research Center for Neurointelligence, The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Tokyo, Japan
ORCID iD: [0000-0002-4602-9816](https://orcid.org/0000-0002-4602-9816)

Yoshitsugu Miyazaki

Department of Chemotherapy and Mycoses, National Institute of Infectious Diseases, Tokyo, Japan

Takashi Umeyama

Department of Chemotherapy and Mycoses, National Institute of Infectious Diseases, Tokyo, Japan

Ken Miyazawa

Department of Chemotherapy and Mycoses, National Institute of Infectious Diseases, Tokyo, Japan

Takeshi Morita

Research Center for Drug and Vaccine Development, National Institute of Infectious Diseases, Tokyo, Japan

Koichi Watashi

Research Center for Drug and Vaccine Development, National Institute of Infectious Diseases, Tokyo, Japan

Christopher B. Brooke

Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL, USA,
Department of Statistics, University of Illinois at Urbana-Champaign, Urbana, IL, USA
ORCID iD: [0000-0002-6815-1193](https://orcid.org/0000-0002-6815-1193)

Ruian Ke

Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM, USA
ORCID iD: [0000-0001-5307-8934](https://orcid.org/0000-0001-5307-8934)

Shingo Iwami

interdisciplinary Biology Laboratory (iBLab), Division of Natural Science, Graduate School of Science, Nagoya University, Nagoya, Japan, International Research Center for Neurointelligence, The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Tokyo, Japan, Institute of Mathematics for Industry, Kyushu University, Fukuoka, Japan, Institute for the Advanced Study of Human Biology (ASHBi), Kyoto University, Kyoto, Japan, Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS), RIKEN, Saitama, Japan, NEXT-Ganken Program, Japanese Foundation for Cancer Research (JFCR), Tokyo, Japan, Science Groove Inc., Fukuoka, Japan

For correspondence: iwami.iblab@bio.nagoya-u.ac.jp

ORCID iD: [0000-0002-1780-350X](https://orcid.org/0000-0002-1780-350X)

Taiga Miyazaki

Division of Respiriology, Rheumatology, Infectious Diseases, and Neurology, Department of Internal Medicine, Faculty of Medicine, University of Miyazaki, Miyazaki, Japan

For correspondence: taiga_miyazaki@med.miyazakiu.ac.jp

Copyright

© 2024, Park et al.

This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editors

Reviewing Editor

Joshua Schiffer

Fred Hutchinson Cancer Research Center, Seattle, United States of America

Senior Editor

Eduardo Franco

McGill University, Montreal, Canada

Reviewer #1 (Public Review):

Summary:

This study by Park and colleagues uses longitudinal saliva viral load data from two cohorts (one in the US and one in Japan from a clinical trial) in the pre-vaccine era to subset viral shedding kinetics and then use machine learning to attempt to identify clinical correlates of different shedding patterns. The stratification method identifies three separate shedding patterns discriminated by peak viral load, shedding duration, and clearance slope. The authors also assess micro-RNAs as potential biomarkers of severity but do not identify any clear relationships with viral kinetics.

Strengths:

The cohorts are well developed, the mathematical model appears to capture shedding kinetics fairly well, the clustering seems generally appropriate, and the machine learning analysis is a sensible, albeit exploratory approach. The micro-RNA analysis is interesting and novel.

Weaknesses:

The conclusions of the paper are somewhat supported by the data but there are certain limitations that are notable and make the study's findings of only limited relevance to current COVID-19 epidemiology and clinical conditions.

(1) The study only included previously uninfected, unvaccinated individuals without the omicron variant. It has been well documented that vaccination and prior infection both predict shorter duration shedding. Therefore, the study results are no longer relevant to current COVID-19 conditions. This is not at all the authors' fault but rather a difficult reality of much retrospective COVID research.

(2) The target cell model, which appears to fit the data fairly well, has clear mechanistic limitations. Specifically, if such a high proportion of cells were to get infected, then the disease would be extremely severe in all cases. The authors could specify that this model was selected for ease of use and to allow clustering, rather than to provide mechanistic insight. It would be useful to list the AIC scores of this model when compared to the model by Ke.

(3) Line 104: I don't follow why including both datasets would allow one model to work better than the other. This requires more explanation. I am also not convinced that non-linear mixed effects approaches can really be used to infer early model kinetics in individuals from one cohort by using late viral load kinetics in another (and vice versa). The approach seems better for making population-level estimates when there is such a high amount of missing data.

(4) Along these lines, the three clusters appear to show uniform expansion slopes whereas the NBA cohort, a much larger cohort that captured early and late viral loads in most individuals,

shows substantial variability in viral expansion slopes. In Figure 2D: the upslope seems extraordinarily rapid relative to other cohorts. I calculate a viral doubling time of roughly 1.5 hours. It would be helpful to understand how reliable of an estimate this is and also how much variability was observed among individuals.

(5) A key issue is that a lack of heterogeneity in the cohort may be driving a lack of differences between the groups. Table 1 shows that SpO2 values and lab values that all look normal. All infections were mild. This may make identifying biomarkers quite challenging.

(6) Figure 3A: many of the clinical variables such as basophil count, Cl, and protein have very low pre-test probability of correlating with virologic outcome.

(7) A key omission appears to be microRNA from pre and early-infection time points. It would be helpful to understand whether microRNA levels at least differed between the two collection timepoints and whether certain microRNAs are dynamic during infection.

(8) The discussion could use a more thorough description of how viral kinetics differ in saliva versus nasal swabs and how this work complements other modeling studies in the field.

(9) The most predictive potential variables of shedding heterogeneity which pertain to the innate and adaptive immune responses (virus-specific antibody and T cell levels) are not measured or modeled.

(10) I am curious whether the models infer different peak viral loads, duration, expansion, and clearance slopes between the 2 cohorts based on fitting to different infection stage data.

<https://doi.org/10.7554/eLife.96032.1.sa2>

Reviewer #2 (Public Review):

Summary:

This study argues it has found that it has stratified viral kinetics for saliva specimens into three groups by the duration of "viral shedding"; the authors could not identify clinical data or microRNAs that correlate with these three groups.

Strengths:

The question of whether there is a stratification of viral kinetics is interesting.

Weaknesses:

The data underlying this work are not treated rigorously. The work in this manuscript is based on PCR data from two studies, with most of the data coming from a trial of nelfinavir (NFV) that showed no effect on the duration of SARS-CoV-2 PCR positivity. This study had no PCR data before symptom onset, and thus exclusively evaluated viral kinetics at or after peak viral loads. The second study is from the University of Illinois; this data set had sampling prior to infection, so has some ability to report the rate of "upswing." Problems in the analysis here include:

-- The PCR Ct data from each study is treated as equivalent and referred to as viral load, without any reports of calibration of platforms or across platforms. Can the authors provide calibration data and justify the direct comparison as well as the use of "viral load" rather than "Ct value"? Can the authors also explain on what basis they treat Ct values in the two studies as identical?

-- The limit of detection for the NFV PCR data was unclear, so the authors assumed it was the same as the University of Illinois study. This seems a big assumption, as PCR platforms can differ substantially. Could the authors do sensitivity analyses around this assumption?

-- The authors refer to PCR positivity as viral shedding, but it is viral RNA detection (very different from shedding live/culturable virus, as shown in the Ke et al. paper). I suggest updating the language throughout the manuscript to be precise on this point.

-- Eyeballing extended data in Figure 1, a number of the putative long-duration infections appear to be likely cases of viral RNA rebound (for examples, see S01-16 and S01-27). What happens if all the samples that look like rebound are reanalyzed to exclude the late PCR detectable time points that appear after negative PCRs?

-- There's no report of uncertainty in the model fits. Given the paucity of data for the upslope, there must be large uncertainty in the up-slope and likely in the peak, too, for the NFV data. This uncertainty is ignored in the subsequent analyses. This calls into question the efforts to stratify by the components of the viral kinetics. Could the authors please include analyses of uncertainty in their model fits and propagate this uncertainty through their analyses?

-- The clinical data are reported as a mean across the course of an infection; presumably vital signs and blood test results vary substantially, too, over this duration, so taking a mean without considering the timing of the tests or the dynamics of their results is perplexing. I'm not sure what to recommend here, as the timing and variation in the acquisition of these clinical data are not clear, and I do not have a strong understanding of the basis for the hypothesis the authors are testing.

It's unclear why microRNAs matter. It would be helpful if the authors could provide more support for their claims that (1) microRNAs play such a substantial role in determining the kinetics of other viruses and (2) they play such an important role in modulating COVID-19 that it's worth exploring the impact of microRNAs on SARS-CoV-2 kinetics. A link to a single review paper seems insufficient justification. What strong experimental evidence is there to support this line of research?

<https://doi.org/10.7554/eLife.96032.1.sa1>

Reviewer #3 (Public Review):

The article presents a comprehensive study on the stratification of viral shedding patterns in saliva among COVID-19 patients. The authors analyze longitudinal viral load data from 144 mildly symptomatic patients using a mathematical model, identifying three distinct groups based on the duration of viral shedding. Despite analyzing a wide range of clinical data and micro-RNA expression levels, the study could not find significant predictors for the stratified shedding patterns, highlighting the complexity of SARS-CoV-2 dynamics in saliva. The research underscores the need for identifying biomarkers to improve public health interventions and acknowledges several limitations, including the lack of consideration of recent variants, the sparsity of information before symptom onset, and the focus on symptomatic infections.

The manuscript is well-written, with the potential for enhanced clarity in explaining statistical methodologies. This work could inform public health strategies and diagnostic testing approaches. However, there is a thorough development of new statistical analysis needed, with major revisions to address the following points:

(1) Patient characterization & selection: Patient immunological status at inclusion (and if it was accessible at the time of infection) may be the strongest predictor for viral shedding in

saliva. The authors state that the patients were not previously infected by SARS-COV-2. Was Anti-N antibody testing performed? Were other humoral measurements performed or did everything rely on declaration? From Figure 1A, I do not understand the rationale for excluding asymptomatic patients. Moreover, the mechanistic model can handle patients with only three observations, why are they not included? Finally, the 54 patients without clinical data can be used for the viral dynamics fitting and then discarded for the descriptive analysis. Excluding them can create a bias. All the discarded patients can help the virus dynamics analysis as it is a population approach. Please clarify. In Table 1 the absence of sex covariate is surprising.

(2) Exact study timReviewer #3 (Public Review):eline for explanatory covariates: I understand the idea of finding « early predictors » of long-lasting viral shedding. I believe it is key and a great question. However, some samples (Figure 4A) seem to be taken at the end of the viral shedding. I am not sure it is really easier to micro-RNA saliva samples than a PCR. So I need to be better convinced of the impact of the possible findings. Generally, the timeline of explanatory covariate is not described in a satisfactory manner in the actual manuscript. Also, the evaluation and inclusion of the daily symptoms in the analysis are unclear to me.

(3) Early Trajectory Differentiation: The model struggles to differentiate between patients' viral load trajectories in the early phase, with overlapping slopes and indistinguishable viral load peaks observed in Figures 2B, 2C, and 2D. The question arises whether this issue stems from the data, the nature of Covid-19, or the model itself. The authors discuss the scarcity of pre-symptom data, primarily relying on Illinois patients who underwent testing before symptom onset. This contrasts earlier statements on pages 5-6 & 23, where they claim the data captures the full infection dynamics, suggesting sufficient early data for pre-symptom kinetics estimation. The authors need to provide detailed information on the number or timing of patient sample collections during each period.

(4) Conditioning on the future: Conditioning on the future in statistics refers to the problematic situation where an analysis inadvertently relies on information that would not have been available at the time decisions were made or data were collected. This seems to be the case when the authors create micro-RNA data (Figure 4A). First, when the sampling times are is something that needs to be clarified by the authors (for clinical outcomes as well). Second, proper causal inference relies on the assumption that the cause precedes the effect. This conditioning on the future may result in overestimating the model's accuracy. This happens because the model has been exposed to the outcome it's supposed to predict. This could question the - already weak - relation with mir-1846 level.

(5) Mathematical Model Choice Justification and Performance: The paper lacks mention of the practical identifiability of the model (especially for tau regarding the lack of early data information). Moreover, it is expected that the immune effector model will be more useful at the beginning of the infection (for which data are the more parsimonious). Please provide AIC for comparison, saying that they have "equal performance" is not enough. Can you provide at least in a point-by-point response the VPC & convergence assessments?

(6) Selected features of viral shedding: I wonder to what extent the viral shedding area under the curve (AUC) and normalized AUC should be added as selected features.

(7) Two-step nature of the analysis: First you fit a mechanistic model, then you use the predictions of this model to perform clustering and prediction of groups (unsupervised then supervised). Thus you do not propagate the uncertainty intrinsic to your first estimation through the second step, ie. all the viral load selected features actually have a confidence bound which is ignored. Did you consider a one-step analysis in which your covariates of interest play a direct role in the parameters of the mechanistic model as covariates? To pursue this type of analysis SCM (Johnson et al. Pharm. Res. 1998), COSSAC (Ayrat et al. 2021

CPT PsP), or SAMBA (Prague et al. CPT PsP 2021) methods can be used. Did you consider sampling on the posterior distribution rather than using EBE to avoid shrinkage?

(8) Need for advanced statistical methods: The analysis is characterized by a lack of power. This can indeed come from the sample size that is characterized by the number of data available in the study. However, I believe the power could be increased using more advanced statistical methods. At least it is worth a try. First considering the unsupervised clustering, summarizing the viral shedding trajectories with features collapses longitudinal information. I wonder if the R package « LongituRF » (and associated method) could help, see Capitaine et al. 2020 SMMR. Another interesting tool to investigate could be latent class models R package « lmm » (and associated method), see Proust-Lima et al. 2017 J. Stat. Softwares. But the latter may be more far-reaching.

(9) Study intrinsic limitation: All the results cannot be extended to asymptomatic patients and patients infected with recent VOCs. It definitively limits the impact of results and their applicability to public health. However, for me, the novelty of the data analysis techniques used should also be taken into consideration.

Strengths are:

- Unique data and comprehensive analysis.
- Novel results on viral shedding.

Weaknesses are:

- Limitation of study design.
- The need for advanced statistical methodology.

<https://doi.org/10.7554/eLife.96032.1.sa0>