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

AUTHORS:

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

AFFILIATIONS:

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

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 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

49 samples from 144 mildly symptomatic participants of two different but complementary
50 cohorts[5,11,12]. We successfully identified three groups with significantly different durations of viral
51 shedding (i.e., the mean durations were 11.5 days, 17.4 days, and 30.0 days, respectively),
52 implying a large inter-participant heterogeneity in virus infection dynamics. However, when we
53 analyzed a total of 47 variables, including basic demographic information, daily clinical symptoms,
54 results of blood tests, and vital signs, none explained the stratified grouping.

55 We also retrospectively explored whether salivary micro-RNAs were associated with the
56 stratification by using stored residual saliva specimens. Micro-RNAs are non-coding RNAs that
57 regulate numerous cellular processes by modulating protein levels through direct binding to
58 mRNA (coding-RNA), thereby influencing translation efficiency or mRNA abundance. Various
59 microRNAs are recognized for their impact on the viral replication ability and immune response
60 in viruses like EBV, HCV, and HIV etc[13]. Many studies are underway to investigate micro-RNAs
61 as potential targets for virus diagnosis and treatment. However, the relationship between micro-
62 RNAs and the patterns of virus shedding of the SARS-CoV-2 virus in body remains unknown. We
63 quantified the expression levels of 92 micro-RNAs and found that no micro-RNA significantly
64 explained the stratified groups, although the mir-1846 level may have been weakly correlated with
65 the peak viral load. Our findings provide important insights into the complexities of viral shedding
66 patterns in saliva and suggest that predicting the heterogeneity of viral dynamics using basic
67 clinical and micro-RNA data may be a challenging task. These insights are critical for developing
68 accurate diagnostic tools, effective treatments, and prevention strategies for COVID-19.

69 # Results

70 ## Description of cohort and study design

71 We used longitudinal saliva viral load data obtained from cohorts of the nelfinavir (NFV)
72 clinical trial (jRCT2071200023[11,12]) and the University of Illinois at Urbana-Champaign[5]. All
73 infections were either mild or asymptomatic. All participants in these cohorts confirmed that they
74 had never been previously infected with SARS-CoV-2, and none were vaccinated against SARS-
75 CoV-2 at the time of enrolment. Of 182 participants from these two studies, 144 symptomatic
76 participants, excluding 22 asymptomatic participants and 16 participants with incomplete
77 observational data, were considered in our analysis (**Fig 1A**). In addition, we annotated the saliva
78 viral load results of 90 participants from the NFV clinical trial with their sex, age, daily symptoms,
79 blood test results, and vital signs, as summarized in **Fig 1B**, **Table 1**, and **Supplementary Table**
80 **1**. The data from the two studies complemented one another: the longitudinal data from the
81 University of Illinois contained data on the early phase of infection (before symptom onset), while
82 the data from the NFV clinical trial contained data on the late phase of viral RNA load (more than
83 14 days after symptom onset) (**Fig 1C**). The time-series viral load in the saliva samples for all
84 individual used in the analysis are plotted in **Fig 1D**. We also analyzed 60 stored saliva specimens
85 from 30 participants of the NFV clinical trial for micro-RNA analysis.

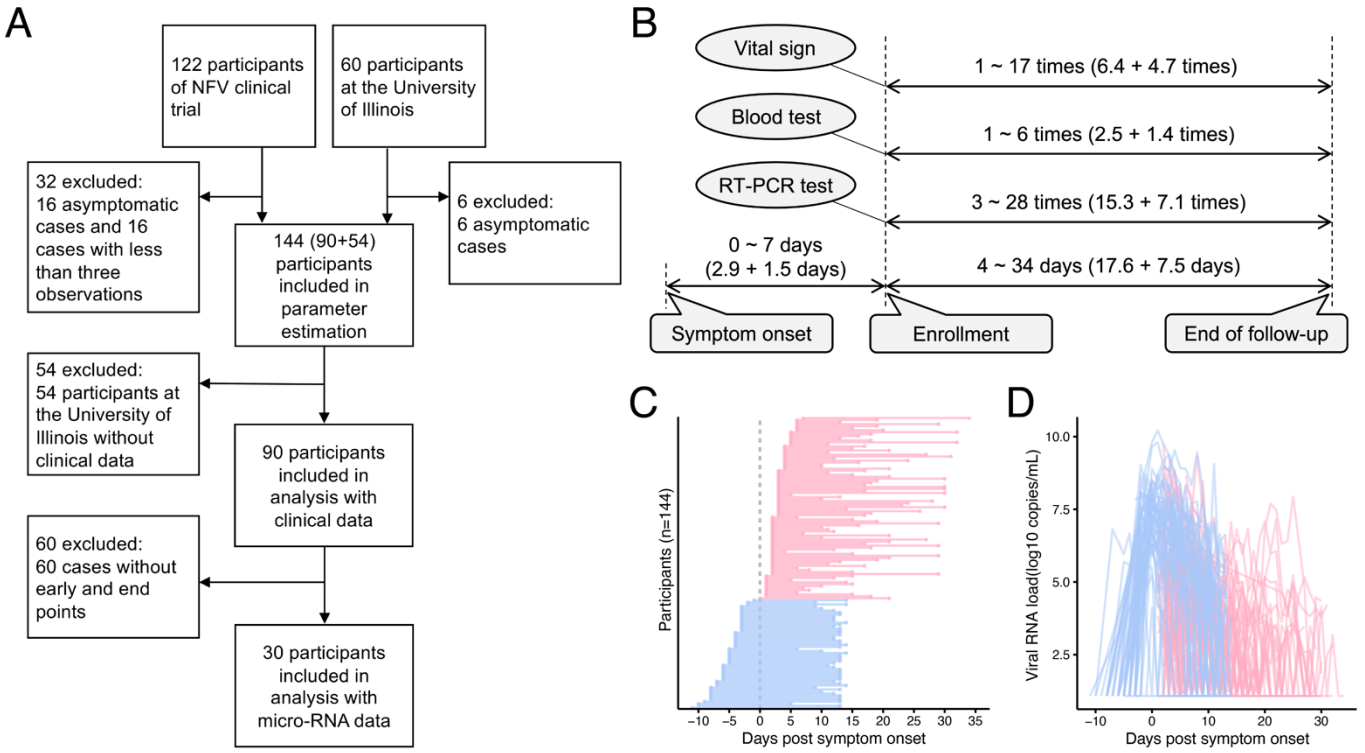


Figure 1 | Characteristics of cohorts from the NFV clinical trial (jRCT2071200023[11,12]) and the University of Illinois[5]: (A) Flowchart of cohorts from the NFV clinical trial and the University of Illinois along with the number of participants and inclusion criteria for our analysis are described. **(B)** Data collection schedule of viral load, blood test, and vital signs from participants in the NFV clinical trial is described. **(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.

95 **Table 1 | Clinical data of the overall cohort from the NFV clinical trial and in groups**

96 **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)
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)

97 * Mean (standard deviation) and **median [Interquartile range].

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

99 We employed a previously developed mathematical model describing SARS-CoV-2
100 infection dynamics [i.e., Eqs.(1-2)] to evaluate interparticipant heterogeneity (details are provided
101 in **Methods**), and reconstructed the best-fit virus dynamics in saliva of 144 symptomatic
102 participants (**Extended Data Fig 1** and **Supplementary Table 2**). We also applied the more
103 realistic mathematical model used in Ke et al.[5] [i.e., Eqs.(3-6)], but the fitting by these two
104 mathematical models reconstructed our dataset equally (**Supplementary Fig 1** and
105 **Supplementary Table 3**). This is because the longitudinal data from the NFV clinical trial included
106 the late phase of the viral RNA load, which was not included in the data from the University of
107 Illinois; therefore, Eqs.(1-2) are sufficient for the purpose of reconstruction of viral load. To avoid
108 complexities and assumptions on parameters of Eqs.(3-6) as described in Ke et al.[5], we used our
109 simple model of Eqs.(1-2) for further analysis.

110 Next, to stratify the time-course pattern of viral shedding, we first applied unsupervised
111 random forest clustering to the individual "reconstructed" virus dynamics of 144 participants (e.g.,
112 **Extended Data Fig 1**). However, this analysis failed to divide the time-course pattern into different
113 clusters (data not shown). To overcome this problem, we quantified the peak, duration, up-slope
114 (i.e., growth rate), and down-slope (i.e., decay rate) of the reconstructed dynamics as "features"
115 of the virus dynamics (**Supplementary Table 4**). Interestingly, the unsupervised random forest
116 clustering based on these features identified 3 groups (i.e., G1: N=46, G2: N=61, and G3: N=37)
117 in which the time-course patterns were clearly discriminated. This finding suggested that there is
118 a heterogeneity of virus infection dynamics in saliva (see **Methods**). **Fig 2A** indicates a two-
119 dimensional Uniform Manifold Approximation and Projection (UMAP) embedding of the three
120 stratified groups. Using a different color for each group (gray for G1, magenta for G2, and blue
121 for G3), we also plotted the estimated individual viral load (**Fig 2B**), and the highlighted time-
122 course pattern of each group by the Partial Least-Squares Discriminant Analysis (PLS-DA) (**Fig**
123 **2C**).

124 The distributions of the four features are described in **Fig 2D**; a statistically significant
125 between-group difference was found in the duration of viral shedding. The mean durations were
126 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,
127 G2, and G3, respectively. In our previous report[10], we consistently confirmed that there were at
128 least 3 groups showing different durations of viral shedding in upper respiratory specimens.

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

136 Another possibility that explains the different viral duration observed here may be a
137 difference in VOC genotypes. To test this, we used data from 55 participants in our NFV clinical
138 trial who had been characterized according to which VOCs [i.e., B.1.1.7 (Alpha), B.1.672.2 or
139 AY.29 (Delta), and other variants] they had been infected with (see **Methods**), in addition to data
140 from 54 participants of the University of Illinois cohort. However, we observed no trend in the
141 number of individuals belonging to each group among the VOCs (p=0.728 by the Fisher's exact
142 test: **Fig 2F**), which is consistent with the conclusion in Ke et al.[5].

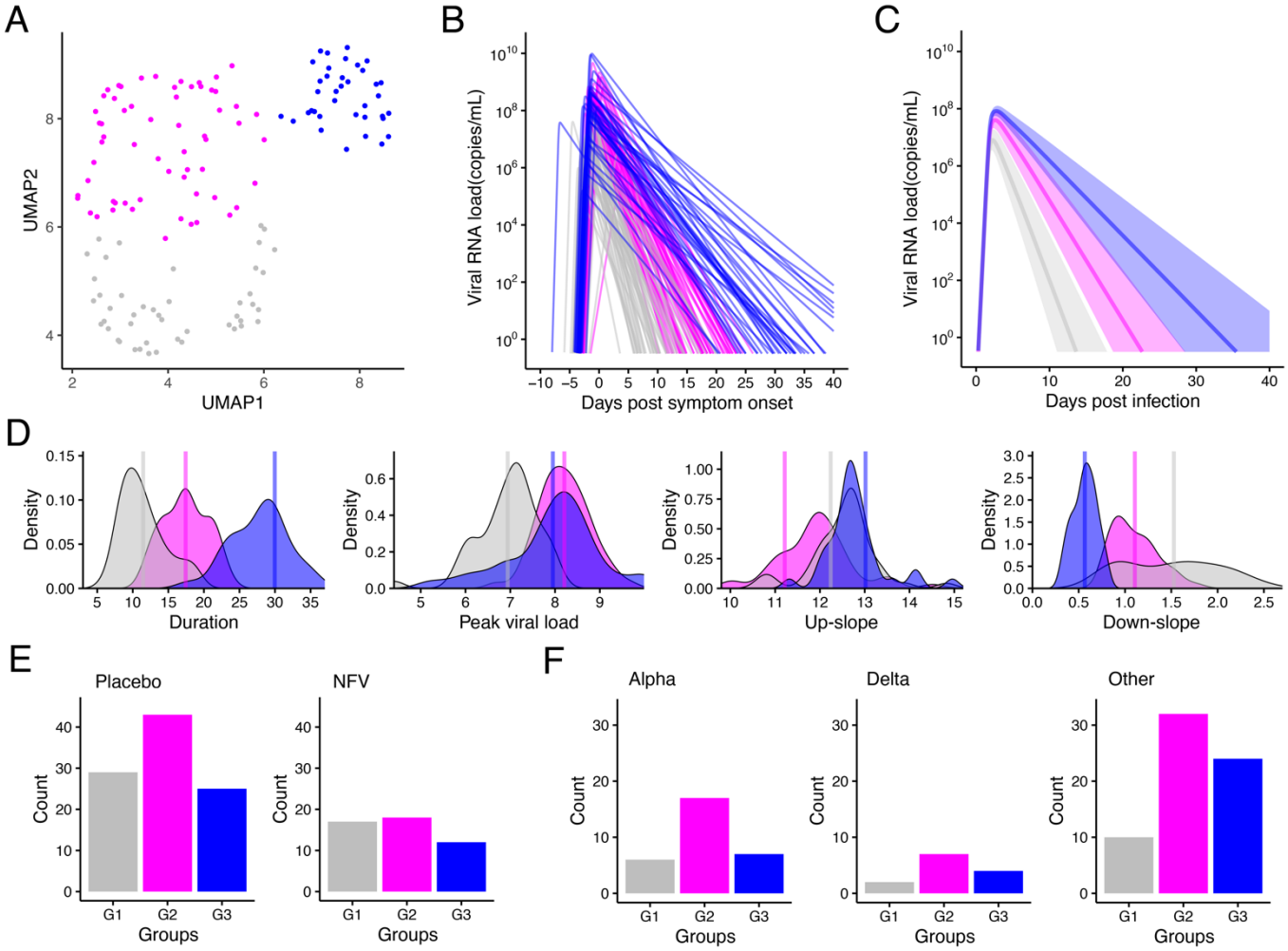


Figure 2 | Stratification of individual SARS-CoV-2 viral dynamics in saliva: (A) UMAP of stratified viral RNA load based on the extracted features from the reconstructed individual-level viral dynamics is shown. **(B)** The reconstructed individual viral RNA load is shown. Colors for individual-level 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.

156 **Basic clinical data may not explain heterogeneity in individual viral shedding**

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

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

174 Next, we asked whether the stratification of the study population is associated with clinical
175 symptoms of COVID-19 that could be caused by active replication of SARS-CoV-2. In general,
176 the clinical symptoms of COVID-19 are cough, fever, shortness of breath, muscle pain, sore throat,
177 confusion, chest pain, headache, rhinorrhea, diarrhea, and nausea and vomiting. In our study,
178 individual-level symptom data were available as 8 categories in the cohorts from both the NFV
179 clinical trial and the University of Illinois (**Supplementary Table 1**). We tried to use a random
180 forest classifier to investigate whether symptom data could predict the stratified groups and
181 obtained ROC-AUCs of 57%, 54%, and 48% for predicting G1, G2, and G3, respectively (**Fig 3C**).

182 In fact, SARS-CoV-2 human challenge clearly showed no quantitative correlation between the
183 individuals' time-series pattern of viral load and symptoms[15].

184 Additionally, we investigated the relationship between each feature of viral dynamics (i.e.,
185 duration of viral shedding, peak viral load, up-slope, and down-slope) and the clinical data by
186 using the Pearson's correlation coefficient (**Extended Data Table 1**). Overall, correlation
187 coefficients were low (0.06 on average) with high p-values, which suggests that no feature was
188 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

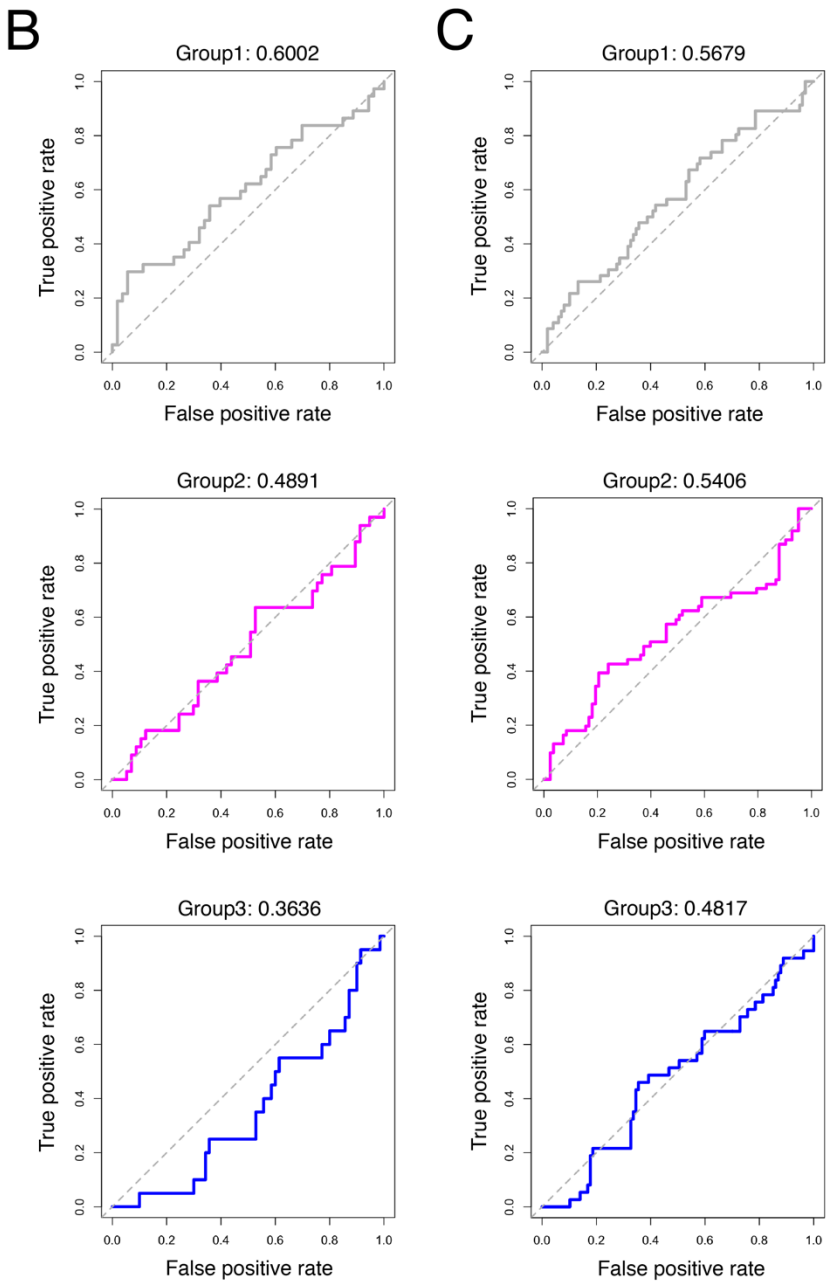


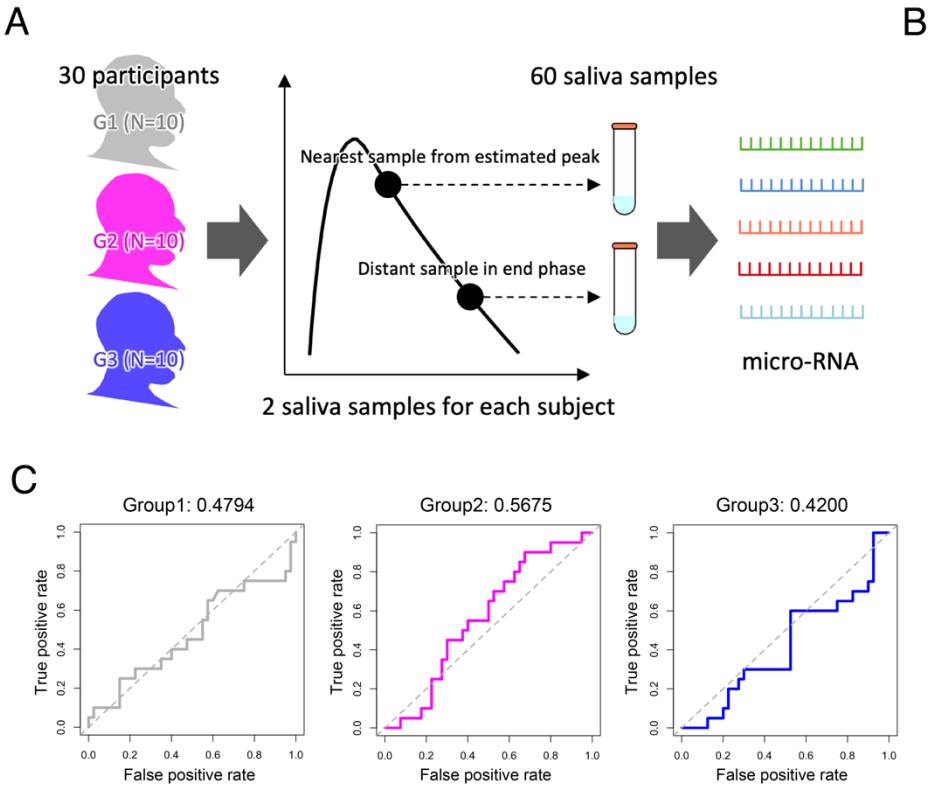
Figure 3 | Correlation between clinical data and viral shedding patterns: (A) P-values of ANOVA corrected by the FDR to compare clinical data among the three stratified groups are shown. Clinical data are listed in reverse order of p-values. **(B)** and **(C)** show ROC curves of random forest classifiers trained on predicting each group by using data for 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.

196 **Relationship between salivary microRNAs and viral shedding patterns in COVID-19**
197 **patients**

198 Various proteins in saliva have antiviral effects. It is also expected that some micro-RNAs
199 in saliva may impair SARS-CoV-2 replication[1] and could thus act as a biomarker to predict
200 COVID-19 disease progression[16]. We here used the stored residual saliva specimens from the
201 NFV clinical trial to identify micro-RNA(s) associated with the stratified groups (i.e., G1, G2 and
202 G3). We note that, because all residual saliva specimens are annotated with the individual
203 participant and we know which participants belong to which stratified group, we can select and
204 compare saliva specimens from G1, G2, and G3 without bias. This implies that we can impartially
205 select participants in equal numbers from each group, unaffected by other factors. Specifically,
206 we collected 60 stored saliva specimens from the NFV clinical trial to perform micro-RNA analysis
207 for 30 participants. We picked two samples for each participant to evaluate the role of micro-RNAs
208 during both the peak and the late phase (i.e., 30 samples for each phase): the nearest sample
209 from the estimated peak and the most distant sample above the detection limit in the late phase
210 (**Fig 4A**). We normalized micro-RNA expression among the samples by using the DESeq2
211 Bioconductor package. We summarize the information on the micro-RNAs we obtained from
212 saliva specimens in **Supplementary Table 5**.

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

222 Furthermore, we investigated the relationship between the 4 features of viral dynamics
223 and micro-RNA levels. Here we used the Spearman's correlation coefficient (**Extended Data**
224 **Table 2**). Overall, we did not find strong correlations between micro-RNA levels and features
225 (Spearman's correlation coefficients on average of 0.002, 0.024, -0.001, and -0.001 for duration
226 of viral shedding, peak viral load, up-slope, and down-slope, respectively, for the 60 total samples).
227 Only the mir-1846 level exhibited a weak negative correlation (-0.53 Spearman's correlation
228 coefficient with 0.01 p-value) with the peak viral load (**Extended Data Figure 2**). We confirmed
229 similar trends even when we analyzed the micro-RNA level for the peak and late phases
230 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.

241 ## Discussion

242 Being able to quickly and efficiently diagnose COVID-19 is essential in monitoring the
243 pandemic. Because the sampling process for saliva is noninvasive, and because it is inexpensive
244 and minimizes the risk for transmissions to health care workers[1], saliva sampling has excellent
245 potential and advantages over other sampling methods from biological specimens such as the
246 lower and upper respiratory tract[2,17]. Given the significant individual heterogeneity in the saliva
247 viral shedding[5,18], identifying biomarker(s) for viral shedding patterns will be crucial for improving
248 public health interventions in the era of living with COVID-19. To improve our understanding of
249 SARS-CoV-2 infection dynamics in saliva to enable application of saliva testing in the fight against
250 COVID-19, we quantified and stratified longitudinal virus dynamics in saliva samples from 144
251 mildly symptomatic individuals from the cohorts of the NFV clinical trial[11] and the University of
252 Illinois at Urbana-Champaign[5]. In addition to the large heterogeneity in virus infection dynamics,
253 we identified three groups (i.e., G1, G2 and G3) with different viral shedding patterns (**Fig 2D**).

254 Immunocompromised patients have been reported to have a prolonged duration of viral
255 shedding, lasting over three months, underscoring the critical role of host immune responses in
256 controlling viral infections[19-22]. Although oral immune responses remain poorly understood, Huang
257 et al. recently confirmed by using single-cell RNA sequencing of the human minor salivary glands
258 and gingiva that SARS-CoV-2 infection can trigger sustained, localized immune responses in
259 saliva[3]. In this study, we observed significant differences in the down-slopes of viral shedding in
260 saliva among participants in different groups, with a more rapid decline in G1. This decline is likely
261 attributed to a stronger immune response to SARS-CoV-2 in G1 participants than in participants
262 in G2 and G3, as reflected in the death rate of infected cells due to the immune response (**Fig.
263 2D**). Lower levels of viral replication have also been observed among infected participants with
264 high baseline levels of mucosal IgA (but not IgG), as reported elsewhere[23]. Recently, we
265 demonstrated that rapid anti-spike secretory IgA antibody responses can contribute to reducing
266 viral shedding durations and amounts in nasopharyngeal mucosa[24]. These findings highlight the
267 importance of biomarkers that directly reflect an individual's immune response, such as antibody

268 induction, in predicting viral shedding patterns. Therefore, quantifying the time-series pattern of
269 mucosal IgA and its correlation with saliva viral load may provide crucial insights into the
270 stratification of SARS-CoV-2 infection dynamics.

271 For the purpose of predicting viral shedding patterns during the early stage of infection,
272 we first explored the association of 39 basic clinical variables, 8 daily symptoms, and the levels
273 of 92 micro-RNAs with the stratified groups. However, none of the factors were significant (**Table**
274 **1**, **Fig 3A**, **Fig 4B, Supplementary Table 1, Supplementary Table 2** and **Supplementary Fig**
275 **2**). In contrast, we showed that mir-1846, which is an exogenous micro-RNA that is specifically
276 classified as an Oryza sativa micro-RNA (osa-microRNA)[25], may exhibit a weak negative
277 correlation. Exogenous micro-RNAs enter the human body primarily through food and can affect
278 human metabolism by interacting and binding with human genes. mir-1846 is reported to interact
279 with two human genes[25] that are known to be associated with the progression of melanoma,
280 various cancers, and leukemia. This suggests that mir-1846 levels may be linked to human
281 immunity. Few studies have investigated the role of mir-1846 in humans, but our findings suggest
282 the need for further investigations into the impact of this micro-RNA level on human immunity.
283 Our research sheds light on the intricate patterns of viral shedding in saliva.

284 Our approach has several limitations that must be considered in our next study: First, our
285 analysis was limited to participants with symptomatic infection and excluded those with
286 asymptomatic infection (22 asymptomatic individuals out of a total of 182 individuals, i.e., 12% of
287 participants) because we integrated datasets with different time scales from different cohorts.
288 Although our data do not include participants infected with omicron variants, others have reported
289 that the omicron variant may cause a higher proportion of asymptomatic infection[26]. Thus,
290 evaluating the effect of asymptomatic infection will be important to update our stratification,
291 especially for recent (or future emerging) VOCs. Second, micro-RNAs participate in the post-
292 transcriptional regulation of gene expression; however, they do not provide direct insights into
293 immune cell dynamics. Given the reported association between the duration of viral shedding and
294 mucosal immunity as discussed above, it appears imperative to analyze modalities that are

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

304 In conclusion, our study revealed that the dynamics of SARS-CoV-2 infection in saliva
305 can be classified into three groups based mainly on the duration of viral shedding. However,
306 accurately predicting the variability in viral dynamics remains a challenging task, because it
307 requires a more comprehensive understanding of the complex shedding patterns in saliva, as well
308 as detailed clinical and molecular data. The identification of a sensitive, simple, and rapid
309 biomarker for saliva viral shedding will be imperative for future COVID-19 outbreak control.
310

311 # METHODS

312 ## Ethics statement

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

318

319 ## Saliva viral load data

320 Longitudinal saliva viral load data of participants with symptomatic and asymptomatic
321 COVID-19 (122 cases) were obtained from the NFV clinical trial[11]. Briefly, the NFV clinical trial
322 was a prospective, randomized, open-label, blinded-endpoint, parallel-group trial conducted
323 between July 2020 and October 2021 at 11 university and teaching hospitals in Japan. This study
324 consisted of a 14-day treatment period and a 14-day follow-up period, with no significant
325 differences in the time to viral clearance between patients who received standard-of-care plus
326 NFV administration and those who had the standard-of-care alone[11,12]. Therefore, the participants
327 with COVID-19 were analyzed together here. In addition, we obtained similar saliva viral load data
328 (60 cases) from the cohort of the University of Illinois at Urbana-Champaign[5]. This cohort
329 contained all faculty, staff, and students at the University of Illinois at Urbana-Champaign, who
330 undergo at least twice weekly quantitative PCR-RT testing during fall of 2020 and spring of 2021.
331 Among those 182 cases, we focused only on symptomatic participants. Also, we excluded the
332 participants who had less than three measured viral loads that were not limit detections (i.e., 90
333 cases from the NFV clinical trial and 54 cases from the University of Illinois were used in this
334 study). This decision was based on the understanding that reasonable estimates cannot be
335 derived when the number of data points less than the number of parameters in the mathematical
336 model. The limit of detection for viral load data from the University of Illinois is 1.08 copies/ml.

337 However, the limit of detection for viral load data from NFV clinical trial was unclear. Considering
338 this, we assumed 1.08 copies/ml as the limit of detection for all viral load data.
339
340 **Viral genome sequencing**
341 The cDNA had been synthesized from RNA of SARS-CoV-2–positive saliva samples.
342 Reverse transcription, multiplex PCR reaction, and Illumina library prep were conducted using a
343 protocol published previously[27]. The pooled library was first purified by AMPureXP at 0.8 x
344 concentration and then again at 1.2 x concentration. The purified library was sequenced for 151
345 cycles at both paired-ends in Illumina iSeq100. Sequence analysis was performed using the nf-
346 core/viralcon pipeline (10.5281/zenodo.3901628).
347
348 **Quantifying biomarkers in saliva**
349 Total RNA from saliva was extracted with MagMAX CORE Nucleic Acid Purification kits
350 (Applied Biosystems, Foster City, CA). Micro-RNAs were detected using Illumina Hiseq x Ten
351 (Illumina, Inc, San Diego, CA) with data processing by ribodepletion (Genewiz-Azenta, South
352 Plainfield, NJ). To remove technical sequences, the pass filter data in the fastq format were
353 processed by Trimmomatic (v0.30) to be high-quality clean data. Following quality trimming,
354 micro-RNAs were identified and checked using miRDeep2[28]. Normalization of micro-RNA
355 expression among samples and differential expression analysis were carried out using the
356 DESeq2 Bioconductor package.
357
358 **Basic clinical data**
359 Basic clinical data including basic demographic characteristics of the study participants,
360 symptoms, and findings of physical examinations and laboratory tests were obtained according
361 to the study protocol[11]. We here used information on age, daily symptoms, results of blood tests,
362 and vital signs of the symptomatic participants in the NFV clinical trial (summarized in **Table 1**
363 and **Supplementary Table 1**).

364

365 **Mathematical modeling**

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

368 $$\frac{df(t)}{dt} = -\beta f(t)V(t), \qquad (1)$$

369 $$\frac{dV(t)}{dt} = \gamma f(t)V(t) - \delta V(t). \qquad (2)$$

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

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

376 $$\frac{dT(t)}{dt} = -\beta T(t)V(t), \qquad (3)$$

377 $$\frac{dE(t)}{dt} = \beta T(t)V(t) - kE(t), \qquad (4)$$

378 $$\frac{dI(t)}{dt} = kE(t) - \delta(t)I(t), \qquad (5)$$

379 $$\frac{dV(t)}{dt} = \pi I(t) - cV(t). \qquad (6)$$

380 The variables $T(t)$, $E(t)$, and $I(t)$ are the total number uninfected target cells, cells in the
381 eclipse phase of infection, and productively infected cells, respectively. The parameters $1/k$, π,
382 and c are the average duration of the eclipse phase, the virus production rate, and the clearance
383 rate of viruses, respectively. The death rate of infected cells is assumed to be time-dependent to
384 mimic the killing of infected cells by immune effector cells: $\delta(t) = \delta_1$ for $t < t_1$ and $\delta(t) = \delta_1 +$
385 δ_2 for $t_1 \leq t$. For a detailed explanation of the immune effector cell model, the reader is referred
386 to Ke et al.[5].

387

388 **Parameter estimation**

389 A nonlinear mixed-effects modelling approach incorporates fixed effects as well as random
390 effects that describe the inter-participant variability in parameters. Including random effects
391 amounts to a partial pooling of the data of all participants to improve estimates of the parameters
392 applicable across the cases. In fact, in our study, the saliva specimens of cases from the cohort
393 of the University of Illinois and the NFV clinical trial were obtained mainly during the "early" (i.e.,
394 from the date of infection to no more than 14 days post symptom onset) and "late" (i.e., no more
395 than 28 days from around symptom onset) phases of SARS-CoV-2 infection, respectively.
396 Therefore, cohorts from these two studies can reveal the whole time-course pattern of SARS-
397 CoV-2 infection dynamics in a complementary manner through the random effects of a nonlinear
398 mixed-effects modelling.

399 In our analyses, the variable $V(t)$ in Eq.(2) corresponds to the viral load in saliva for
400 SARS-CoV-2. To fit the patient's viral load data, we used a program MONOLIX 2021R2
401 (www.lixoft.com), implement maximum likelihood estimation of parameters in nonlinear mixed
402 effect model. The nonlinear mixed effect model allows a fixed effect as well as interpatient
403 variability. This method estimates each parameter $\theta_i (= \theta \times e^{\eta_i})$ for each individual where θ is a
404 fixed effect, and η_i is a random effect, and which obeys a Gaussian distribution with mean 0 and
405 standard deviation Ω. Here we used lognormal distributions as prior distributions of parameters
406 to guarantee the positiveness (i.e., negative values do not biologically make sense). In parameter
407 estimation, as time 0 in the original dataset represents the time of symptom onset, we also
408 estimated the time from infection to symptom onset (corresponding to τ in **Supplementary Table**
409 **2** and **3**) along with other parameters. The fixed effect parameters and random effect parameters
410 were estimated by use of the stochastic approximation Expectation/Maximization (SAEM)
411 algorithm and empirical Bayes method, respectively. A right-truncated normal distribution was
412 used in the likelihood function to account for the left censoring of the viral load data (i.e., when

413 the viral load is not detectable)[29]. We changed the initial values multiple times to avoid a local
414 minimum of AIC and confirmed the robustness of parameter estimation.
415
416 **Unsupervised clustering and stratification of SARS-CoV-2 infection dynamics**
417 Unsupervised random forest clustering was performed on the selected features of the
418 virus infection dynamics, that is, the peak viral load, duration of viral shedding, up-slope, and
419 down-slope (rfUtilities package in R). The use of random forest allows us to avoid overfitting by
420 bootstrap aggregating (bagging) and to achieve better generalization performance[30]. After a
421 random forest dissimilarity (i.e., the distance matrix between all pairs of samples) was obtained,
422 it was visualized with Uniform Manifold Approximation and Projection (UMAP) in a two-
423 dimensional plane and was stratified with spectral clustering (Python scikit-learn). The optimal
424 number of clusters was determined by the eigengap heuristic method.
425
426 **Random forest classifiers for characterizing stratified groups**
427 Random forest classifiers were trained to predict either of the three stratified groups (G1-
428 G3) using randomForest packages in R. The receiver operating characteristic (ROC) curve for
429 each classifier was drawn from out-of-bag (OOB) samples using the pROC package in R. For
430 example, the ROC for G1 is for the random forest classifier predicting "G1" versus "G2 or G3".
431 We list the variables for the supervised random forest in **Table 1**, **Supplementary Table 1**, and
432 **Supplementary Table 2.**
433
434 **Statistical analysis**
435 When necessary, the variables were compared among different groups using Fisher's
436 exact test (for categorical variables), analysis of variance (ANOVA, for numerical variables from
437 clinical data with more than two groups), and Kruskal-Wallis ANOVA (for variables from micro-
438 RNA data with more than two groups). Corrections of p-values for multiple testing were performed
439 by the False Discovery Rate (FDR). Also, the variables were investigated for their relationship

440 with features of viral load using the Pearson's correlation coefficient (for variables from clinical
441 data) and the Spearman's correlation coefficient (for variables from micro-RNA data). All
442 statistical analyses were performed using R (version 4.1.3).

443 ## LIST OF SUPPLEMENTARY MATERIALS

444 **Supplementary Figure 1** | Comparison of three model fits to viral load in saliva specimens for
445 individual participants

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

447 **Supplementary Table 2** | Summary of parameter estimation by the model described in Eqs.(1-
448 2)

449 **Supplementary Table 3** | Summary of parameter estimation by the model described in Eqs.(3-
450 6)

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

452 **Supplementary Table 5** | Salivary micro-RNAs obtained from 60 specimens of 30 participants

453 **REFERENCES**

454 1 Baghizadeh Fini, M. Oral saliva and COVID-19. *Oral Oncol* **108**, 104821 (2020).
455 https://doi.org:10.1016/j.oraloncology.2020.104821
456 2 Wyllie, A. L. *et al.* Saliva or Nasopharyngeal Swab Specimens for Detection of SARS-CoV-2. *N*
457 *Engl J Med* **383**, 1283-1286 (2020). https://doi.org:10.1056/NEJMc2016359
458 3 Huang, N. *et al.* SARS-CoV-2 infection of the oral cavity and saliva. *Nat Med* **27**, 892-903
459 (2021). https://doi.org:10.1038/s41591-021-01296-8
460 4 Kim, K. S. *et al.* A quantitative model used to compare within-host SARS-CoV-2, MERS-CoV,
461 and SARS-CoV dynamics provides insights into the pathogenesis and treatment of SARS-CoV-2.
462 *PLoS Biol* **19**, e3001128 (2021). https://doi.org:10.1371/journal.pbio.3001128
463 5 Ke, R. *et al.* Daily longitudinal sampling of SARS-CoV-2 infection reveals substantial
464 heterogeneity in infectiousness. *Nat Microbiol* **7**, 640-652 (2022). https://doi.org:10.1038/s41564-
465 022-01105-z
466 6 Jeong, Y. D. *et al.* Designing isolation guidelines for COVID-19 patients with rapid antigen tests.
467 *Nat Commun* **13**, 4910 (2022). https://doi.org:10.1038/s41467-022-32663-9
468 7 Goyal, A., Cardozo-Ojeda, E. F. & Schiffer, J. T. Potency and timing of antiviral therapy as
469 determinants of duration of SARS-CoV-2 shedding and intensity of inflammatory response. *Sci*
470 *Adv* **6** (2020). https://doi.org:10.1126/sciadv.abc7112
471 8 Marc, A. *et al.* Quantifying the relationship between SARS-CoV-2 viral load and infectiousness.
472 *Elife* **10** (2021). https://doi.org:10.7554/eLife.69302
473 9 Néant, N. *et al.* Modeling SARS-CoV-2 viral kinetics and association with mortality in
474 hospitalized patients from the French COVID cohort. *Proc Natl Acad Sci U S A* **118** (2021).
475 https://doi.org:10.1073/pnas.2017962118
476 10 Iwanami, S. *et al.* Detection of significant antiviral drug effects on COVID-19 with reasonable
477 sample sizes in randomized controlled trials: A modeling study. *PLoS Med* **18**, e1003660 (2021).
478 https://doi.org:10.1371/journal.pmed.1003660
479 11 Hosogaya, N. *et al.* Efficacy and safety of nelfinavir in asymptomatic and mild COVID-19
480 patients: a structured summary of a study protocol for a multicenter, randomized controlled trial.
481 *Trials* **22**, 309 (2021). https://doi.org:10.1186/s13063-021-05282-w
482 12 Miyazaki, T. *et al.* A Multicenter Randomized Controlled Trial To Evaluate the Efficacy and
483 Safety of Nelfinavir in Patients with Mild COVID-19. *Microbiol Spectr* **11**, e0431122 (2023).
484 https://doi.org:10.1128/spectrum.04311-22
485 13 Trobaugh, D. W. & Klimstra, W. B. MicroRNA Regulation of RNA Virus Replication and
486 Pathogenesis. *Trends Mol Med* **23**, 80-93 (2017). https://doi.org:10.1016/j.molmed.2016.11.003
487 14 Grinsztajn, L., Oyallon, E. & Varoquaux, G. Why do tree-based models still outperform deep
488 learning on tabular data? *arXiv preprint arXiv:2207.08815* (2022).
489 15 Killingley, B. *et al.* Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in
490 young adults. *Nat Med* **28**, 1031-1041 (2022). https://doi.org:10.1038/s41591-022-01780-9
491 16 Pimenta, R. *et al.* MiR-200c-3p expression may be associated with worsening of the clinical
492 course of patients with COVID-19. *Mol Biol Res Commun* **10**, 141-147 (2021).
493 https://doi.org:10.22099/mbrc.2021.40555.1631
494 17 To, K. K. *et al.* Temporal profiles of viral load in posterior oropharyngeal saliva samples and
495 serum antibody responses during infection by SARS-CoV-2: an observational cohort study.
496 *Lancet Infect Dis* **20**, 565-574 (2020). https://doi.org:10.1016/s1473-3099(20)30196-1
497 18 Hay, J. A. *et al.* Quantifying the impact of immune history and variant on SARS-CoV-2 viral
498 kinetics and infection rebound: a retrospective cohort study. *medRxiv*, 2022.2001.2013.22269257
499 (2022). https://doi.org:10.1101/2022.01.13.22269257
500 19 Leung, W. F. *et al.* COVID-19 in an immunocompromised host: persistent shedding of viable
501 SARS-CoV-2 and emergence of multiple mutations: a case report. *Int J Infect Dis* **114**, 178-182
502 (2022). https://doi.org:10.1016/j.ijid.2021.10.045

20 Niyonkuru, M. *et al.* Prolonged viral shedding of SARS-CoV-2 in two immunocompromised patients, a case report. *BMC Infect Dis* **21**, 743 (2021). https://doi.org:10.1186/s12879-021-06429-5

21 Nakajima, Y. *et al.* Prolonged viral shedding of SARS-CoV-2 in an immunocompromised patient. *J Infect Chemother* **27**, 387-389 (2021). https://doi.org:10.1016/j.jiac.2020.12.001

22 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). https://doi.org:10.21037/apm-20-996

23 Havervall, S. *et al.* Anti-Spike Mucosal IgA Protection against SARS-CoV-2 Omicron Infection. *N Engl J Med* **387**, 1333-1336 (2022). https://doi.org:10.1056/NEJMc2209651

24 Miyamoto, S. *et al.* 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). https://doi.org:10.1073/pnas.2314808120

25 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). https://doi.org:10.1016/j.compbiolchem.2020.107305

26 Garrett, N. *et al.* High Rate of Asymptomatic Carriage Associated with Variant Strain Omicron. *medRxiv* (2022). https://doi.org:10.1101/2021.12.20.21268130

27 Kentaro Itokawa, T. S., Masanori Hashino, Rina Tanaka, Makoto Kuroda. nCoV-2019 sequencing protocol for illumina. *protocols.io* (2020). https://doi.org:10.17504/protocols.io.bnn7mdhn

28 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). https://doi.org:10.1093/nar/gkr688

29 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).

30 Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2 (Springer, 2009).

ACKNOWLEDGMENTS

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

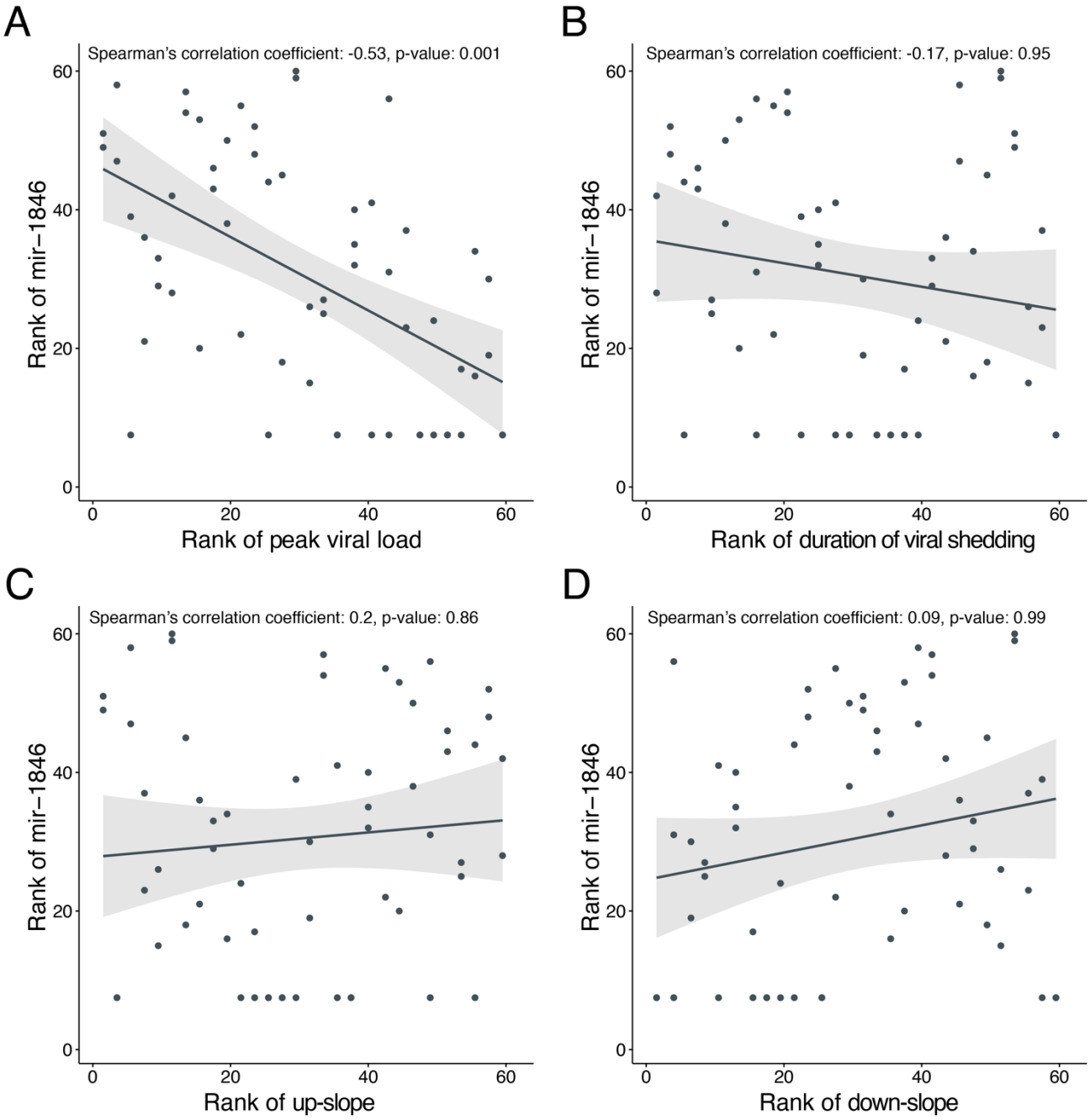
The authors declare no conflicts of interest associated with this manuscript.

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-cell-limited 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.

573 **Extended Data Table 1 | Pearson's correlation coefficients between clinical data and features of viral dynamics.**

Clinical data	Duration of viral shedding		Peak viral load		Up-slope		Down-slope	
	Pearson's r	p-value	Pearson's r	p-value	Pearson's r	p-value	Pearson's r	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

575 **Extended Data Table 2 | Spearman's correlation coefficients between micro-RNA 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

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

576

Supplementary Information

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

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

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

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 (**Expended 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].

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

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 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}	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, ϑ_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)

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, ϑ_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 4 | Features of reconstructed individual viral dynamics for each group

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 5 | Salivary micro-RNAs obtained from 60 specimens of 30 participants

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)

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)

Reference

1 Ke, R. *et al.* Daily longitudinal sampling of SARS-CoV-2 infection reveals substantial heterogeneity in infectiousness. *Nat Microbiol* **7**, 640-652 (2022). https://doi.org:10.1038/s41564-022-01105-z