Supplementary Information for

Kinetics of SARS-CoV-2 infection in the human upper and lower respiratory tracts and their relationship with infectiousness

Ruian Ke, Carolin Zitzmann, Ruy M. Ribeiro, Alan S. Perelson*.

*Correspondence should be addressed to:
Alan S. Perelson
Email: asp@lanl.gov
Telephone: 505-667-68297

This PDF file includes:

Supplementary Methods

1. Choice of fixed parameter values

Total target cell numbers in the URT and the LRT in the absence of infection ($T_{1,0}$ and $T_{2,0}$, respectively)

In the URT, it has been estimated that there are 4×10^8 epithelial cells (1). Approximately 1% of these cells express both angiotensin-converting enzyme 2 (ACE2) and the type II transmembrane serine protease TMPRSS2 based on single-cell RNAseq experiments (2, 3). These two proteins are the receptor and the protease needed for SARS-CoV-2 binding and fusion with the cell membrane leading to viral entry into the cell (4). Thus, we assume the initial number of target cells in the URT, $T_{1,0} = 4\times10^6$.

In the LRT, cells that express both ACE2 and TMPRSS2 are mostly lung epithelial cells. The fraction of lung epithelial cells expressing ACE2 and TMPRSS2 is also about 1% (2, 3). It is estimated that there are approximately 1×10^{10} bronchial epithelial cells (5), 4×10^{10} alveolar type I cells, and 7×10^{10} alveolar type II cells in a human adult (6, 7). We thus calculate the total number of lung epithelial cells expressing ACE2 and TMPRSS2 as $1\%\times(1+4+7)\times10^{10}=1.2\times10^9$ cells. There are 5 lobes in the two lungs and the trachea is connected to two lobes, i.e. the upper lobes. We assume that initially only target cells in the upper lobes are available for infection. Thus, we approximate the initial number of target cells in the LRT, $T_{2,0}$, as $T_{2,0}=2/5\times1.2\times10^9=4.8\times10^8$ cells. Note that in the extended target cell model and the full model, we allow more target cells to become available for infection as infection proceeds.

Hou et al. recently reported that the number of cells that are susceptible to SARS-CoV-2 infection may be larger than that estimated based on gene expression measured using single-cell RNAseq, and that there exists a decreasing gradient of susceptible cells from the URT to the LRT (8). Therefore, there exists uncertainty in the number of target cells estimated above. However, because for a standard viral dynamic model the number of initial target cells and the virus production rate are unidentifiable and only their product is identifiable (9), an increase (decrease) in the initial number of target cells will lead to corresponding decrease (increase) in the estimate of the virus production rate but not in the estimate of other parameters such as R_0 (see Eqns [4] and [5]).

Initial number of infected cells, $I_{1,0}$

Evidence strongly suggests that the URT is the initial site of infection (8, 10). Thus, we assume that one cell in the URT is infected at the start of infection, i.e. $I_{1,0} = 1$ cell. This approach is similar to that in Ref. (11) as we use it because in our model, we do not track the number of viruses in the URT (instead, we track viruses in a swab, V_T). As shown in Smith et al. (11), this assumption does not change the dynamics of the model significantly as any initial viral particles that succeed in initiating infection must infect one or more cells (rapidly) before being cleared. In sensitivity analysis, we test $I_{1,0} = 10$ cells.

Virus clearance rate, c

We set c=10/day, because *in vivo* viral clearance is usually fast in many infections, including for respiratory infections such as influenza viruses (11) (1, 12). We have used this value of *c* in a

previous model of URT infection by SARS-CoV-2 (13). In sensitivity analyses, we set c= 5 or 20/day.

Eclipse period, $1/k_1$ and $1/k_2$

For simplicity, we assume that the mean duration of the eclipse period of infected cells is the same for cells in the URT and the LRT, i.e. $k_1 = k_2$, because this should be a property of intracellular infection. We set $k_1 = k_2 = 4$ /day (i.e. $\frac{1}{k_1} = \frac{1}{k_2}$ =6 hours) according to cell culture experiments suggesting infected cells start to produce virus between 4-8 hours post infection (14). In sensitivity analyses, we set $\frac{1}{k_1} = \frac{1}{k_2}$ =4 or 8 hours, with the 4 hour choice motivated by the experiments in Hou et al. (8) that showed viral titers of 10^4 PFU/ml in vitro at 6 hour post infection, the earliest time point sampled.

The rate of virus transport from the URT to the LRT, Γ

Hou et al. provided evidence supporting the idea that virus is seeded to the LRT through the oral-lung aspiration axis (8). Since aspiration is generally an infrequent event, we assume the average rate of virus transport from the URT and the LRT is low (also see next section for further experimental support), although the precise value has not been quantified. We set $\Gamma = 0.01$/day. In sensitivity analyses, we vary this parameter by two orders of magnitude, letting $\Gamma = 0.001$ or 0.1/day.

2. Simplification of the target cell limited model

We reduced the number of parameters in the target cell limited model by making biologically reasonable simplifications. We first derive the ODEs for V_T and V_S to replace the expressions for V_1 and V_2 as

$$\begin{aligned} \frac{dV_T}{dt} &= f_1 \frac{dV_1}{dt} = f_1 p_1 I_1 - (c + g_{12})V_T + \frac{f_1}{f_2} g_{21} V_S \\ \frac{dV_S}{dt} &= f_2 \frac{dV_2}{dt} = f_2 p_2 I_2 - (c + g_{21})V_S + \frac{f_2}{f_1} g_{12} V_T \end{aligned} \qquad \text{[S1]}$$

In what follows, we assume that $g_{12} \ll c$. This is a reasonable assumption because we expect viral transfer to the LRT to be slow (15, 16), and viral clearance is typically fast. Further, in previous studies (10, 17), full genome sequencing was done for all patients. In one patient a G4664A mutation was observed in a throat swab, while a sputum sample from the same day still showed the original allele 4664G, consistent with slow transport into the LRT. Under this assumption, we have $c + g_{12} \approx c$. Thus, $\frac{dV_T}{dt} \approx f_1 p_1 I_1 - cV_T + \frac{f_1}{f_2} g_{21} V_S$. There is no evidence the viral load dynamics in the LRT strongly affects dynamics in the URT. For example, in general viral load in the URT continues to decrease after its peak even when the viral load in the LRT is at its peak and stays elevated. If we set $g_{21} = 0$, the ODEs for V_T and V_S become

$$\begin{aligned} \frac{dV_T}{dt} &= f_1 p_1 I_1 - cV_T \\ \frac{dV_S}{dt} &= f_2 p_2 I_2 - cV_S + \frac{f_2}{f_1} g_{12} V_T \end{aligned} \qquad \text{[S2]}$$

To formally test the validity of assuming $g_{21} = 0$, we fitted a model that does not assume $g_{21} = 0$ to the data collected before day 14 post infection in Wolfel et al. (10). This estimation indicated g_{21} to be approximately 10^{-5}, justifying the assumption that g_{21}=0. In addition, this model is not preferred over the model with $g_{21} = 0$ based on its AIC score (388.6 vs. 388.0; see Table S8).

For consistency of units, we also let $\pi_T = f_1 p_1$, $\pi_S = f_2 p_2$, $\beta_T = \frac{\beta_1}{f_1}$, $\beta_S = \frac{\beta_2}{f_2}$ and $\Gamma = \frac{f_2}{f_1} g_{12}$, leading to the ODEs shown in the main text.

3. Models describing long term dynamics

In our analyses of the long-term data, i.e. data beyond 14 days post infection, we consider four alternative processes in addition to target cell limitation. In these models, we add in turn an innate immunity component, proliferation of target cells and the possibility of extra target cells in the lungs. Because adaptive immune responses are likely to develop and become effective by about 2 weeks post infection, we incorporated the impact of adaptive immunity into all of these models as described below.

Modeling the adaptive immune response

We model the impact of adaptive immune responses similar to Pawelek et al. (12). Adaptive immune responses, such as T cell responses, are likely to develop and become effective by about 2 weeks post infection (18, 19). We assumed that the rates of infected cell killing in both the URT and the LRT increase after day 14 post infection and thus use the equations below describing $\delta_1(t)$ and $\delta_2(t)$:

$$\delta_i(t) = \begin{cases} \delta_{i,0} & t < 14 \text{ days} \\ \delta_{i,0} e^{w(t-14)} & t \geq 14 \text{ days} \end{cases}, i = 1, 2, \qquad [S3]$$

where w is the rate of exponential increase in the killing rate of infected cells. In reality, the death rate does not continue to increase exponentially, but we assume that it is a good approximation for the time frame of our data set. Antibody responses are also expected to develop over the same time scale (20) and could contribute to increased infected cell death and to viral neutralization, a process that also would lead to fewer infected cells. Here due to the lack of explicit data on the immune response in the set of infected individuals we study (10), the total adaptive immune response is summarized by the change in δ_i given above.

Innate immunity model

The type-I interferon response can play an important role in protecting target cells from becoming infected (21). We extend the TCL model with adaptive immunity by including a type-I interferon response (a major part of the innate immune response) following the framework presented in previous models for influenza infection dynamics (1, 12, 22). In addition to the compartments in the TCL model, this model keeps track of type I interferon (F) and cells refractory to infection (R). Interferon is produced from infected cells and binds to interferon receptors on target cells stimulating an antiviral response that can make cells refractory to viral infection (R). Such cells are said to be in an antiviral state (21).

The ODEs for target cells, refractory cells and interferon in the innate immunity model are

$$\begin{aligned}
\frac{dT_1}{dt} &= -\beta_T V_T T_1 - \phi_1 F_1 T_1 + \rho_1 R_1 \\
\frac{dR}{dt} &= \phi_1 F_1 T_1 - \rho_1 R_1 \\
\frac{dF_1}{dt} &= s_1 I_1 - \mu_1 F_1 \\
\frac{dT_2}{dt} &= -\beta_S V_S T_2 - \phi_2 F_2 T_2 + \rho_2 R_2 \\
\frac{dR_2}{dt} &= \phi_2 F_2 T_2 - \rho_2 R_2 \\
\frac{dF_2}{dt} &= s_2 I_2 - \mu_2 F_2
\end{aligned} \qquad \text{[S4]}$$

ODEs for other compartments are the same as in the TCL model, with $\delta_i(t)$ as above.

In this model, the impact of the innate immune response is to convert target cells into refractory cells at rate $\phi_1 F_1 T_1$ in the URT, where ϕ_1 is a rate constant. Refractory cells can become target cells again at rate ρ_1. Interferon is produced and cleared at rates s_1 and μ_1, respectively. The innate immune response in the LRT follows the same structure as in the URT.

To minimize the number of unknown parameters, we simplify the model by making the quasi-steady-state assumption that the interferon dynamics are much faster than the dynamics of infected cells and assume that $\frac{dF_1}{dt} = 0$. Thus $s_1 I_1 = \mu_1 F_1$ or $F_1 = \frac{s_1}{\mu_1} I_1$. Making the same assumption for the interferon dynamics in the LRT, we get $F_2 = \frac{s_2}{\mu_2} I_2$. This allows us to use a model that does not explicitly keep track of F_1 and F_2, for which there is no information.

Let $\Phi_1 = \phi \frac{s_1}{\mu_1}$ and $\Phi_2 = \phi_2 \frac{s_2}{\mu_2}$, so that the ODEs for the innate immunity model become, with $\delta_i(t)$ given by [S3] above:

$$\begin{aligned}
\frac{dT_1}{dt} &= -\beta_T V_T T_1 - \Phi_1 I_1 T_1 + \rho_1 R_1 \\
\frac{dR}{dt} &= \Phi_1 I_1 T_1 - \rho_1 R_1 \\
\frac{dE_1}{dt} &= \beta_T V_T T_1 - k_1 E_1 \\
\frac{dI_1}{dt} &= k_1 E_1 - \delta_1(t) I_1 \\
\frac{dV_T}{dt} &= \pi_T I_1 - c V_T \\
\frac{dT_2}{dt} &= -\beta_S V_S T_2 - \Phi_2 I_2 T_2 + \rho_2 R_2 \\
\frac{dR_2}{dt} &= \Phi_2 I_2 T_2 - \rho_2 R_2 \\
\frac{dE_2}{dt} &= \beta_S V_S T_2 - k_2 E_2 \\
\frac{dI_2}{dt} &= k_2 E_2 - \delta_2(t) I_2
\end{aligned} \qquad \text{[S5]}$$

$$\frac{dV_S}{dt} = \pi_S I_2 - cV_S + \Gamma V_T$$

Target cell proliferation model

In this model, we assume that cells in the lungs can proliferate. We model proliferation of target cells in the LRT by adding a term $q\left(1 - \frac{T_2+E_2+I_2}{T_{2,0}}\right)T_2$, where q is a rate constant for proliferation. Then the ODE for T_2 becomes

$$\frac{dT_2}{dt} = q\left(1 - \frac{T_2+E_2+I_2}{T_{2,0}}\right)T_2 - \beta_S V_S T_2, \qquad [S6]$$

Extended target cell model

This model has the same structure as the TCL model with adaptive immune response, except that we assume new target cells (T_N) appear in the LRT at time t_T, i.e. we increase T_2 by T_N at time t_T in the model. This can arise from the spatial spread of infection in the lungs. As infection progresses, spread of the virus into new areas of the lungs can cause new target cells to become available for infection. Here because of limited data we chose to model this by a simple increase in the number of target cells, i.e. in the variable T_2, at time t_T rather than by employing an explicitly spatial model.

The combined model

In the combined model, we combine the innate immunity model and the extended target cells model.

4. Parameter fitting

We take two approaches for model fitting. First, when fitting models to data collected up to day 14 post infection, we use a population approach, based on non-linear mixed effect modeling, to fit the model to data from 8 patients in URT and LRT, simultaneously. We allow random effects on the fitted parameters. We tested the incubation period τ as a continuous covariate for each fitted parameter. More specifically, for a lognormally distributed population parameter θ the median value for this parameter for individuals with incubation period τ is given by $\bar{\theta} = \theta_{pop} e^{\sigma_\theta \tau}$ where the population parameter value θ_{pop} and the covariate coefficient σ_θ, as well as the variance of θ_{pop}, are to be estimated. All estimations were performed using Monolix (Monolix Suite 2018R1, Antony, France: Lixoft SAS, 2018. lixoft.com/products/monolix/).

Second, when fitting models to long-term time series involving all data collected up to day 31 post infection, we fit the model to data from each patient separately. This approach is taken due to the substantial variability in the long-term viral dynamics among individuals, which precludes us to assume they are a sample from a homogeneous population. To estimate parameter values, we first calculate the residual sum of squares (RSS) between the simulated viral load and the measured viral load, both on a $\log_{10}$ scale. Using log for viral load data is standard, because the error is typically multiplicative. For censored data points, i.e. data points that are below the limit of detection (LoD; as defined in (10)), we calculate the RSS in the same way as in Ref. (23). If the simulated viral load at a particular time point is below the LoD, the error for this data point is 0, i.e. both simulated viral load and data are below LoD; if the simulated viral load is above LoD,

we calculate the error between the simulated viral load and the LoD. Parameter estimations were performed by minimizing the RSS. The fitting is performed using the 'optim' package in R (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/).

5. Uncertainty analysis

Uncertainty analysis of the best model describing the long-term time series, i.e. the extended target cell model, has been assessed by analyzing structural and practical parameter identifiability. For that we calculated for every model parameter 95% confidence intervals with the profile likelihood method (PLE) in the Data2Dynamics environment in Matlab Release 2016b (The Mathworks). A parameter is identifiable if the 95% confidence interval (or the parameter profile) is finite (24, 25). The results of this analysis are presented in Table S7.

6. Modeling the impact of therapeutics

We consider two classes of therapeutics, 1) those that reduce virus production from infected cells, such as the nucleotide analog remdesivir (26), and 2) those that inhibit/block virus entry into target cells, such as entry inhibitors and neutralizing antibody-based therapies. For the first type of antivirals, we modify the ODEs by replacing the virus production parameters, π_T and π_S by $(1-\epsilon_p)\pi_T$ and $(1-\epsilon_p)\pi_S$, respectively, where ϵ_p is the drug efficacy and ranges between 0 and 1. For the second type of antiviral, we modify the ODEs by replacing the parameters β_T and β_S in the target cell compartments and eclipse cell compartments by $(1-\epsilon_\beta)\beta_T$and $(1-\epsilon_\beta)\beta_S$, respectively, where ϵ_β is the drug efficacy and ranges between 0 and 1. If the two types of drugs are given together as combination therapy, then both modifications are simultaneously introduced. Within this framework, we quantify the effect of therapy on the viral dynamics in the URT and specifically on the infectiousness of the person, using our model in the main text. We also analyze the effect of therapy in the LRT.

SI References:

1. Baccam P, Beauchemin C, Macken CA, Hayden FG, & Perelson AS (2006) Kinetics of influenza A virus infection in humans. *J Virol* 80(15):7590-7599.
2. Muus C*, et al.* (2020) Integrated analyses of single-cell atlases reveal age, gender, and smoking status associations with cell type-specific expression of mediators of SARS-CoV-2 viral entry and highlights inflammatory programs in putative target cells. *bioRxiv*:DOI: 10.1101/2020.1104.1119.049254.
3. Sungnak W*, et al.* (2020) SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat Med* 26(5):681-687.
4. Hoffmann M*, et al.* (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181(2):271-280 e278.
5. Mercer RR, Russell ML, Roggli VL, & Crapo JD (1994) Cell number and distribution in human and rat airways. *Am J Respir Cell Mol Biol* 10(6):613-624.
6. Bianconi E*, et al.* (2013) An estimation of the number of cells in the human body. *Ann Hum Biol* 40(6):463-471.
7. Crapo JD, Barry BE, Gehr P, Bachofen M, & Weibel ER (1982) Cell number and cell characteristics of the normal human lung. *Am Rev Respir Dis* 126(2):332-337.
8. Hou YJ*, et al.* (2020) SARS-CoV-2 reverse genetics reveals a variable infection gradient in the respiratory tract. *Cell* 182(2):429-446 e414.
9. Miao H, Xia X, Perelson AS, & Wu H (2011) On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev Soc Ind Appl Math* 53(1):3-39.
10. Wolfel R*, et al.* (2020) Virological assessment of hospitalized patients with COVID-2019. *Nature* 581(7809):465-469.
11. Smith AP, Moquin DJ, Bernhauerova V, & Smith AM (2018) Influenza Virus Infection Model With Density Dependence Supports Biphasic Viral Decay. *Front Microbiol* 9:1554.
12. Pawelek KA*, et al.* (2012) Modeling within-host dynamics of influenza virus infection including immune responses. *PLoS Comput Biol* 8(6):e1002588.
13. Gonçalves A*, et al.* (2020) Timing of antiviral treatment initiation is critical to reduce SARS-CoV-2 viral load. *CPT: Pharmacometrics & Systems Pharmacology* (in press):DOI: 10.1101/2020.1104.1104.20047886.
14. Ogando NS*, et al.* (2020) SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. *J Gen Virol*:2020.2004.2020.049924.
15. Gonzalez-Parra G & Dobrovolny HM (2019) The rate of viral transfer between upper and lower respiratory tracts determines RSV illness duration. *J Math Biol* 79(2):467-483.
16. Quirouette C, Younis NP, Reddy MB, & Beauchemin CAA (2020) A mathematical model describing the localization and spread of influenza A virus infection within the human respiratory tract. *PLoS Comput Biol* 16(4):e1007705.
17. Bohmer MM*, et al.* (2020) Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. *Lancet Infect Dis*:DOI: 10.1016/S1473-3099(1020)30314-30315.
18. Zhang X*, et al.* (2020) Viral and host factors related to the clinical outcome of COVID-19. *Nature* 583(7816):437-440.
19. Weiskopf D*, et al.* (2020) Phenotype and kinetics of SARS-CoV-2-specific T cells in COVID-19 patients with acute respiratory distress syndrome. *Sci Immunol* 5(48).

20. To KK*, 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(5):565-574.
21. Garcia-Sastre A & Biron CA (2006) Type 1 interferons and the virus-host relationship: a lesson in detente. *Science* 312(5775):879-882.
22. Saenz RA*, et al.* (2010) Dynamics of influenza virus infection and pathology. *J Virol* 84(8):3974-3983.
23. Ke R, Cong ME, Li D, Garcia-Lerma JG, & Perelson AS (2017) On the death rate of abortively infected cells: estimation from simian-human immunodeficiency virus infection. *J Virol* 91(18):e00352-00317.
24. Raue A, Becker V, Klingmuller U, & Timmer J (2010) Identifiability and observability analysis for experimental design in nonlinear dynamical models. *Chaos* 20(4):045105.
25. Raue A*, et al.* (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25(15):1923-1929.
26. Agostini ML*, et al.* (2018) Coronavirus susceptibility to the antiviral remdesivir (GS-5734) is mediated by the viral polymerase and the proofreading exoribonuclease. *mBio* 9(2).
27. Leung NHL*, et al.* (2020) Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* 26(5):676-680.

Supplementary Figures

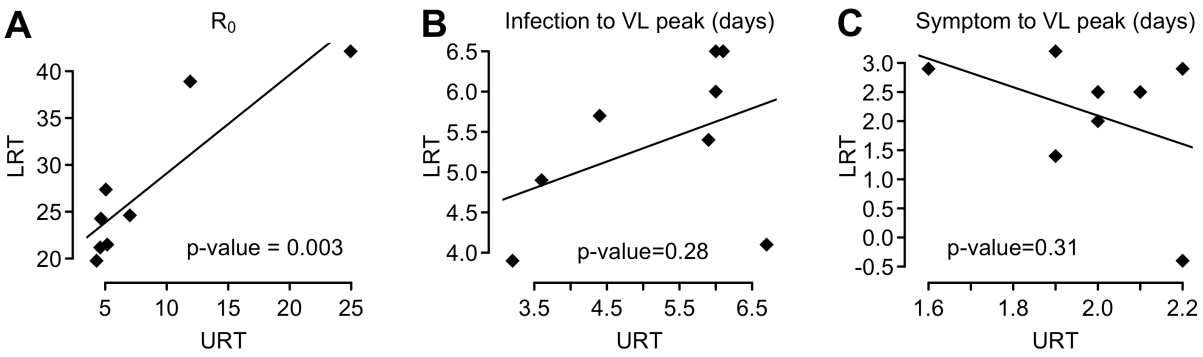

Figure S1. Correlations between viral kinetics characteristics in the URT (x-axis) and the LRT (y-axis). The characteristics shown are **(A)** R_0, **(B)** time from infection to viral load (VL) peak and **(C)** time from symptom onset to viral load peak.

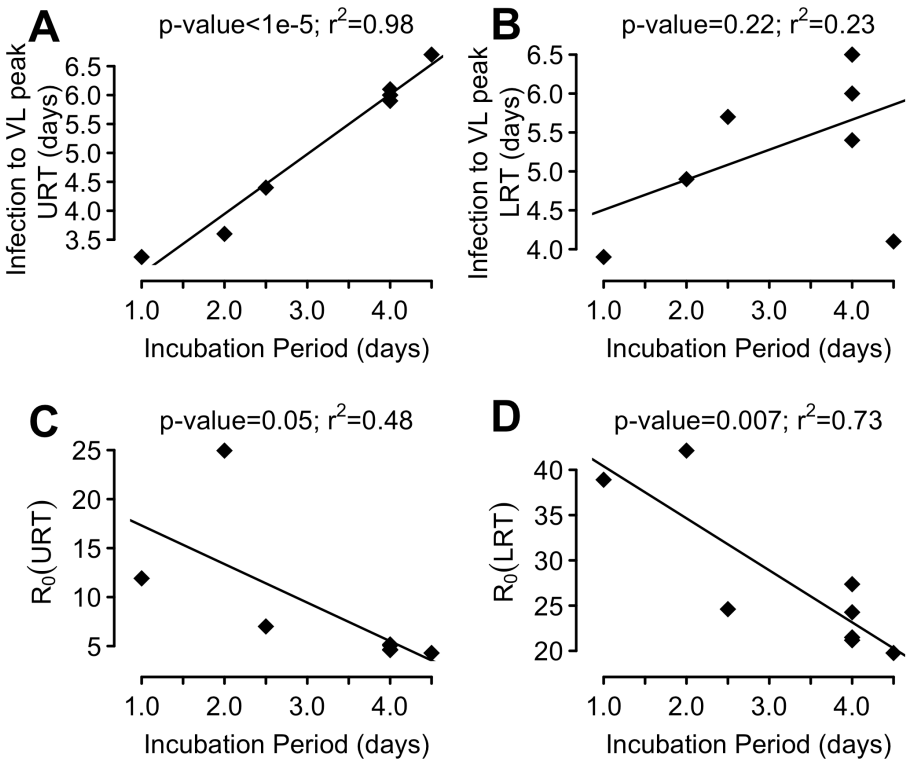

Figure S2. Regression analyses of the relationship between the duration of the incubation period and estimated parameters characterizing viral dynamics. The estimated parameters include the times from infection to viral peak in the URT and the LRT (A and B, respectively), and R_0 values in the URT and LRT (C and D, respectively).

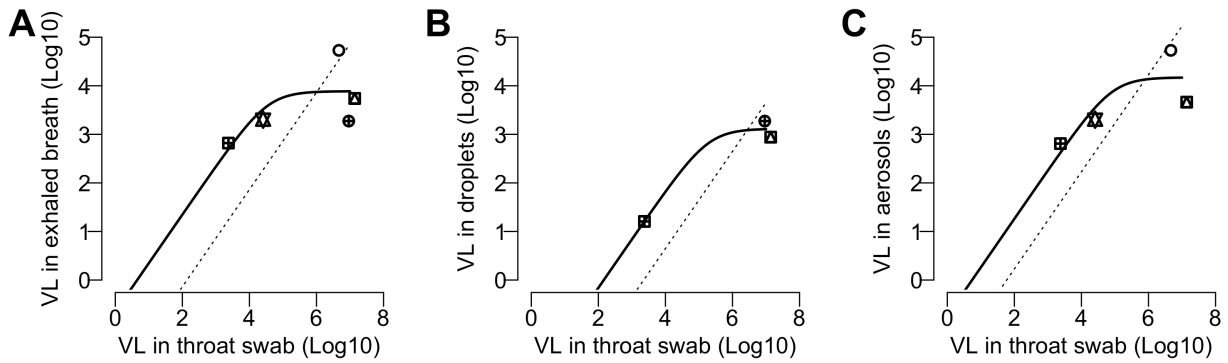

Figure S3. The relationship between the number of seasonal coronaviruses in exhaled breath and the viral load in throat swabs. All datasets are from Ref. (27). The number of total viruses (i.e. viruses in respiratory droplets plus aerosols) in exhaled breath (panel A), in respiratory droplets (panel B) and in aerosols (panel C) are plotted against the viral load in throat swabs. Symbols denote individuals from whom samples were taken. The data points below the limit of detection were not included. Solid lines show the best-fit using a Michaelis-Menten function, i.e. $y = V_m \frac{x}{x+K_m}$, where x and y denote the viral load shown on the x- and y-axis, respectively and V_m and K_m are constants; whereas the dotted lines show the best-fit using a linear function, i.e. $y = ax$. Across the three panels, the Michaelis-Menten function fits the data significantly better. Note that the conclusion that the number of exhaled viruses saturates when viral load is high remains the same when we include data points below the limit of detection in our fitting.

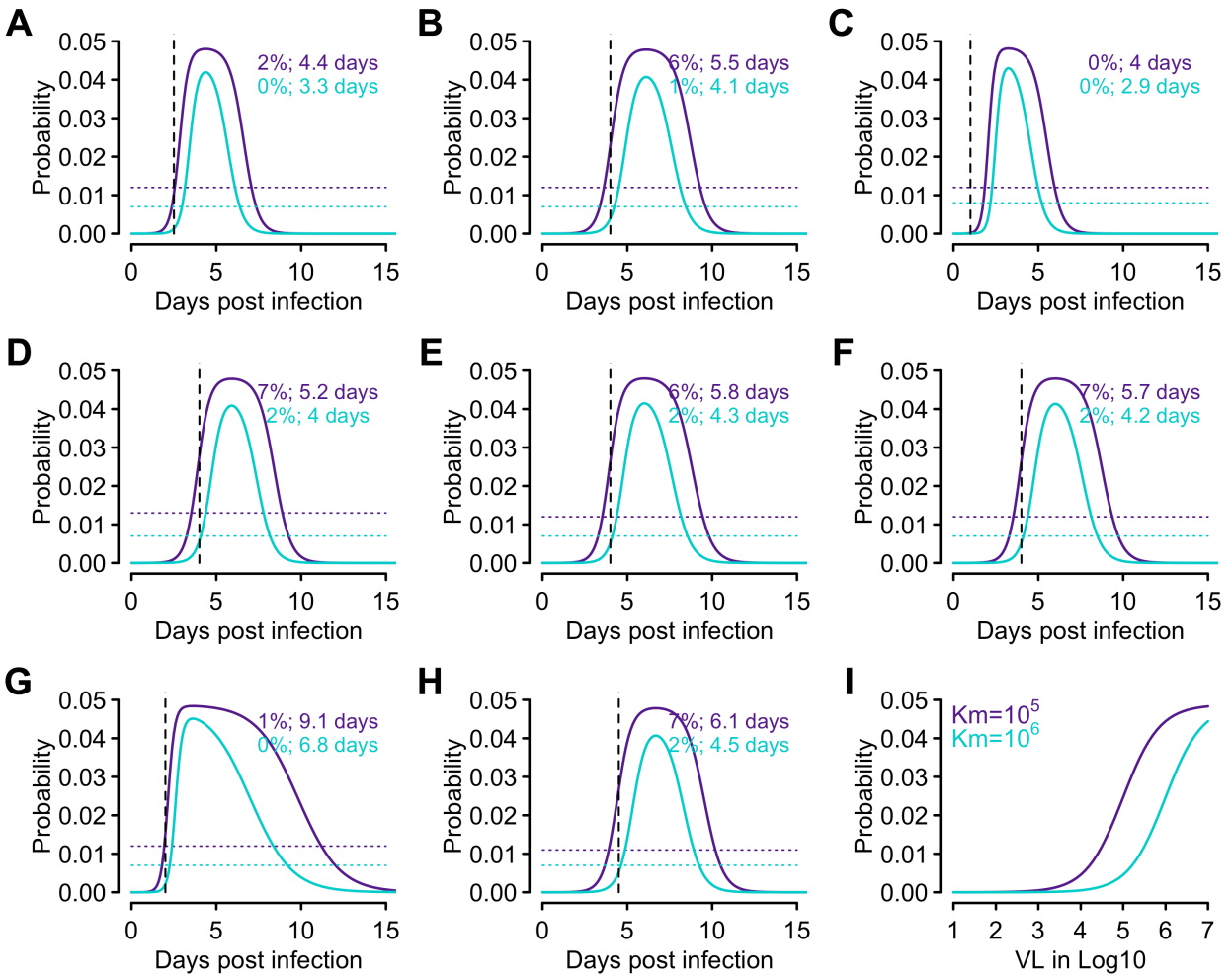

Figure S4. The probability of transmission for a typical contact over time predicted for each individual using the probabilistic model linking viral load to infectiousness. We set $\theta = 0.05$ and $K_m = 10^5$ (purple) or 10^6 RNA copies (cyan). 95% of the total infectiousness probability occurs above the dashed horizontal lines. The mean durations of infectiousness across all individuals are 5.7 days and 4.3 days for $K_m = 10^3$ and 10^4 RNA copies, respectively. The infectious periods last on average 5.6 and 4.6 days post symptom onset for $K_m = 10^3$ and 10^4 RNA copies, respectively. Notation is the same as in Fig. 3 in the main text.

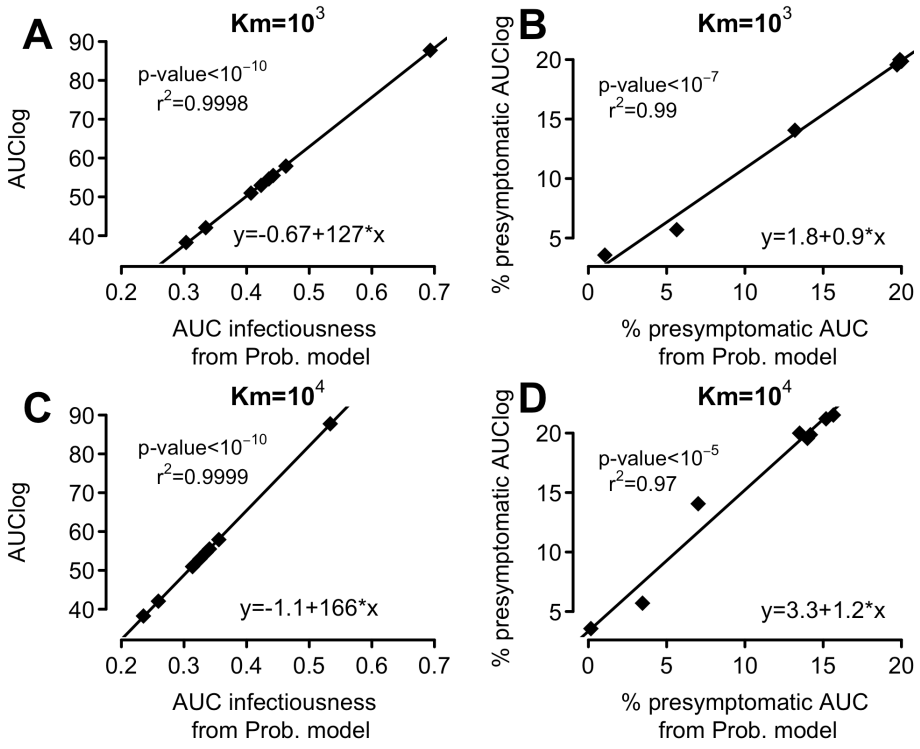

Figure S5. Consistency between the area under the logarithm of the viral load curve, i.e. AUClog, and the probability model using $K_m=10^3$ (A and B) or $K_m=10^4$ RNA copies (C and D). (A and C) Regression of and correlations between the area under the curve of infectiousness from the probability model, $p(t)$, and AUClog calculated from the model fit to the URT viral load for the 8 individuals shown in Fig. 1. **(B and D)** Regression of and correlations between the percentage of presymptomatic transmission predicted by AUClog calculated from the model fit and by the AUC of the infectiousness curve. The formula defining the regression line is shown in the bottom-right corner of the plots.

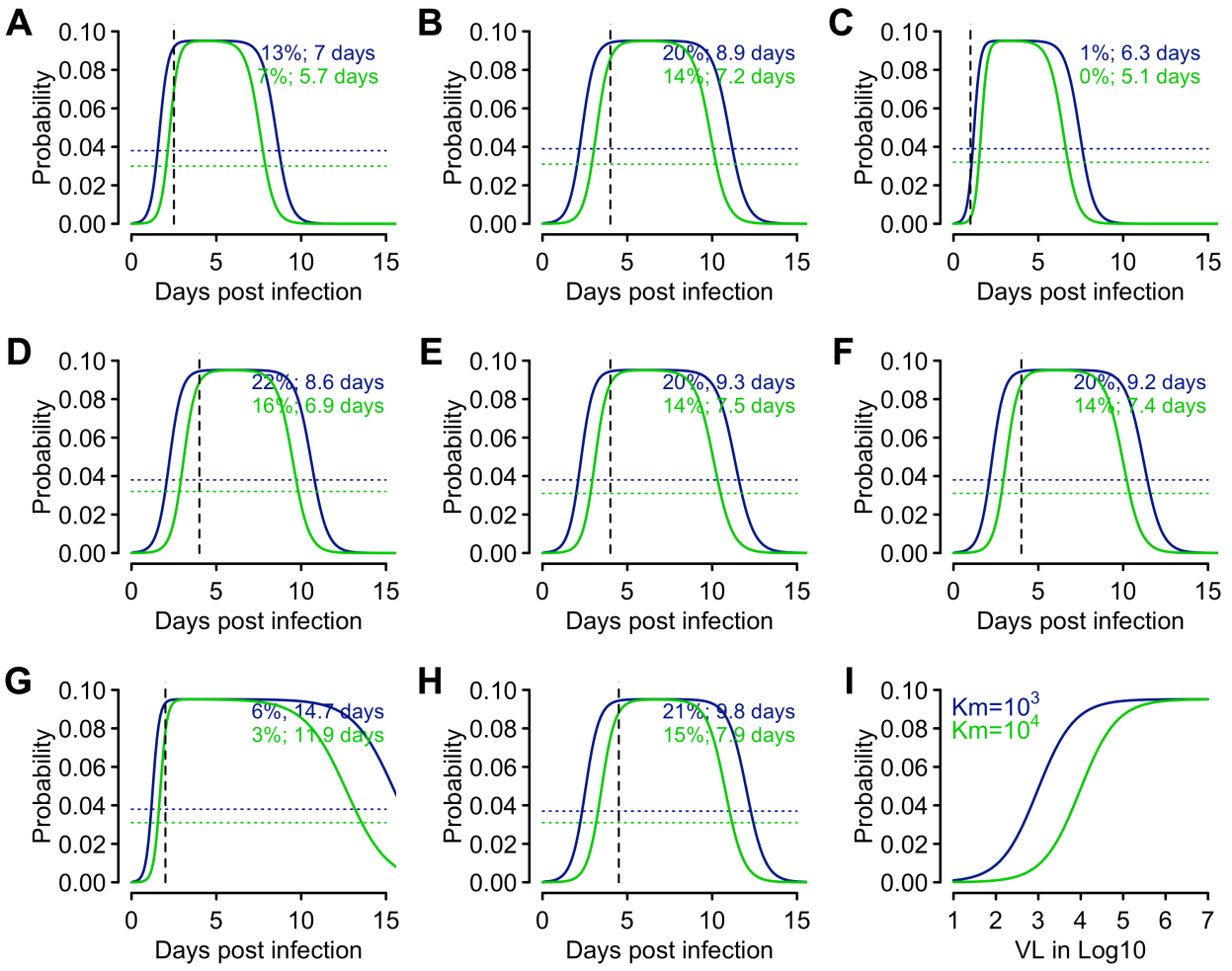

Figure S6. The probability of transmission for a typical contact over time predicted for each individual using the model linking viral load to infectiousness. We set $\theta = 0.1$ (instead of 0.05 as in the main text) and $K_m = 10^3$ (blue) or 10^4 (green) RNA copies. As in Fig. 3 in the main text, 95% of the total infectiousness probability occurs above the dashed horizontal lines. The mean durations of infectiousness across all individuals are 9.2 days and 7.5 days for $K_m = 10^3$ and 10^4 RNA copies, respectively. The infectious periods last on average 7.8 and 6.7 days post symptom onset for $K_m = 10^3$ and 10^4 RNA copies, respectively. Other notation is the same as in Fig. 3 in the main text.

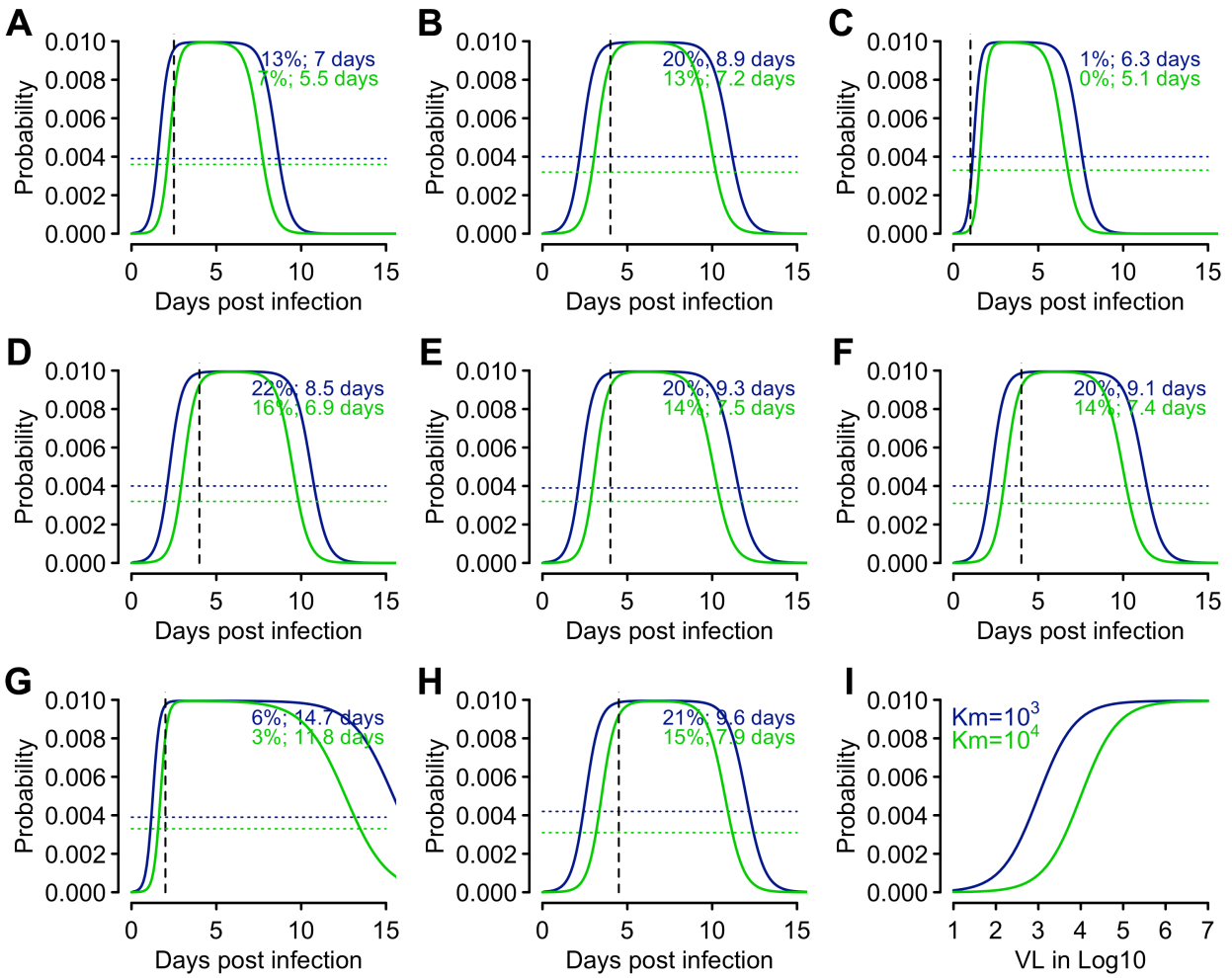

Figure S7. The probability of transmission for a typical contact over time predicted for each individual using the model linking viral load to infectiousness. We set $\theta = 0.01$ (instead of 0.05 as in the main text) and $K_m = 10^3$ (blue) or 10^4 (green) RNA copies. As in Fig. 3 in the main text, 95% of the total infectiousness probability occurs above the dashed horizontal lines. The mean durations of infectiousness across all individuals are 9.2 days and 7.4 days for $K_m = 10^3$ and 10^4 RNA copies, respectively. The infectious periods last on average 7.8 and 6.7 days post symptom onset for $K_m = 10^3$ and 10^4 RNA copies, respectively. Other notation is the same as in Fig. 3 in the main text.

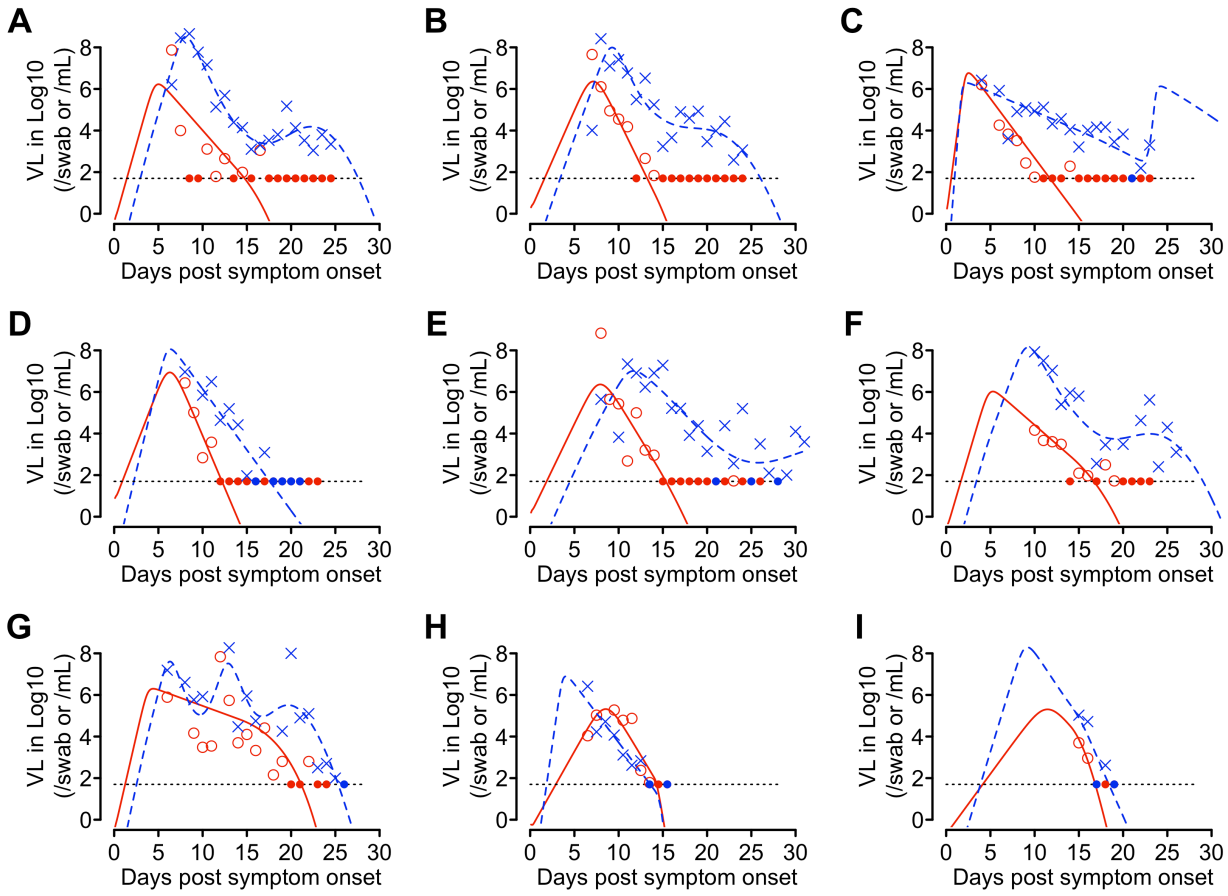

Figure S8. Data (symbols) and simulations (lines) of the target cell proliferation model using best-fit parameter values for each individual. Red and blue denote viral load kinetics in the URT and the LRT, respectively. Symbols show the data from throat swabs (red squares) and the sputum samples (blue 'x's) as reported in Ref. (10). The dashed black lines show the limit of detection, and closed dots indicate data below the limit of detection.

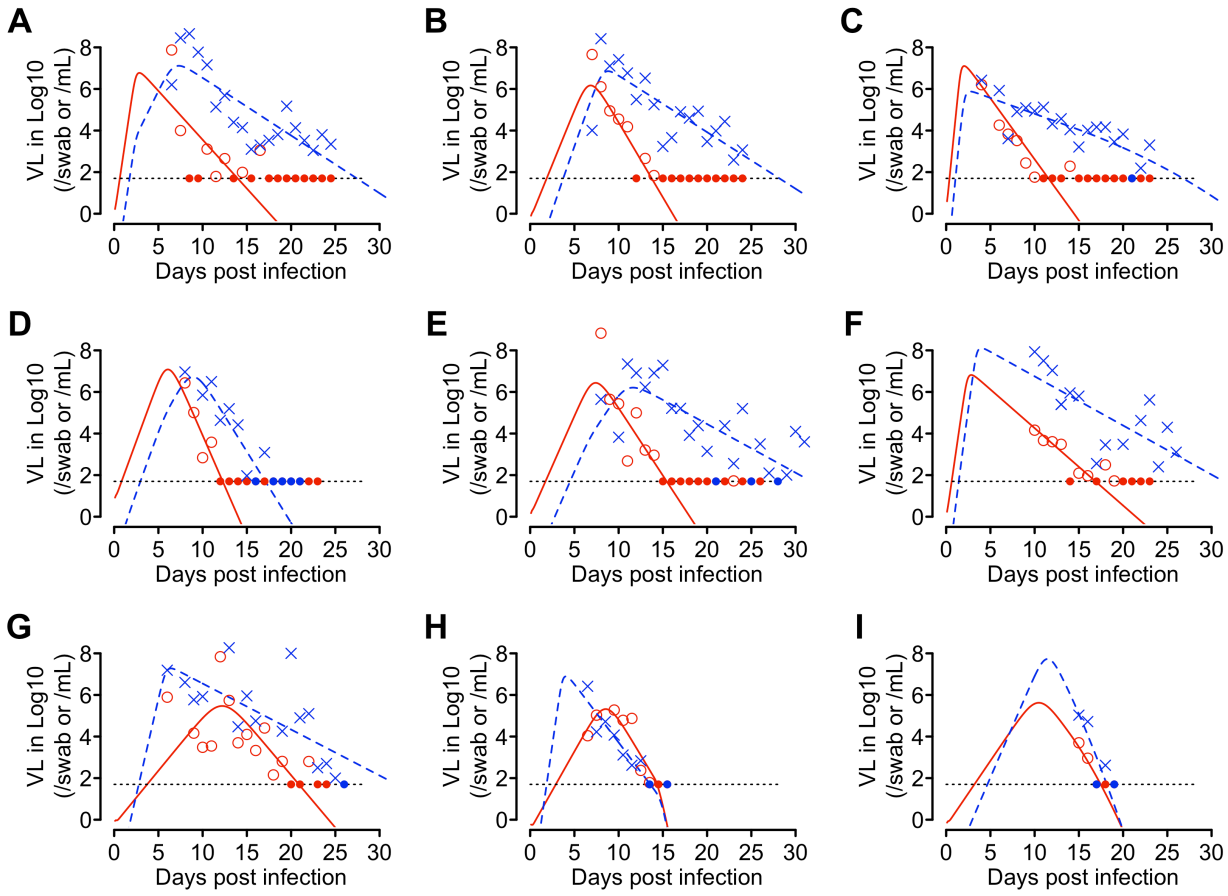

Figure S9. Data (symbols) and simulations (lines) of the target cell limited (TCL) model using best-fit parameter values for each individual. Red and blue denote viral load kinetics in the URT and the LRT, respectively. Symbols show the data from throat swabs (red squares) and the sputum samples (blue 'x's) as reported in Ref. (10). The dashed black lines show the limit of detection, and closed dots indicate data below the limit of detection.

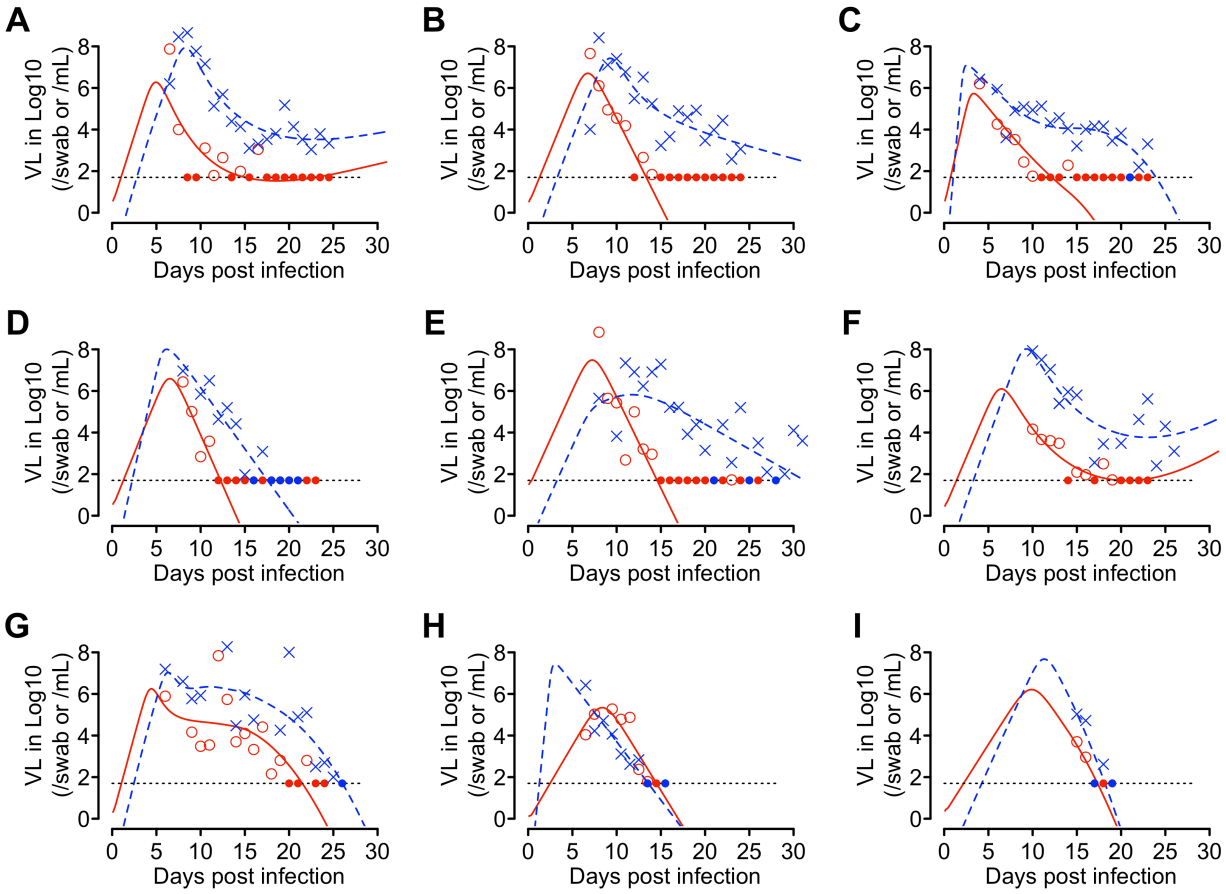

Figure S10. Data (symbols) and simulations (lines) of the innate immunity model using best-fit parameter values for each individual. Red and blue denote viral load kinetics in the URT and the LRT, respectively. Symbols show the data from throat swabs (red squares) and the sputum samples (blue 'x's) as reported in Ref. (10). The dashed black lines show the limit of detection, and closed dots indicate data below the limit of detection.

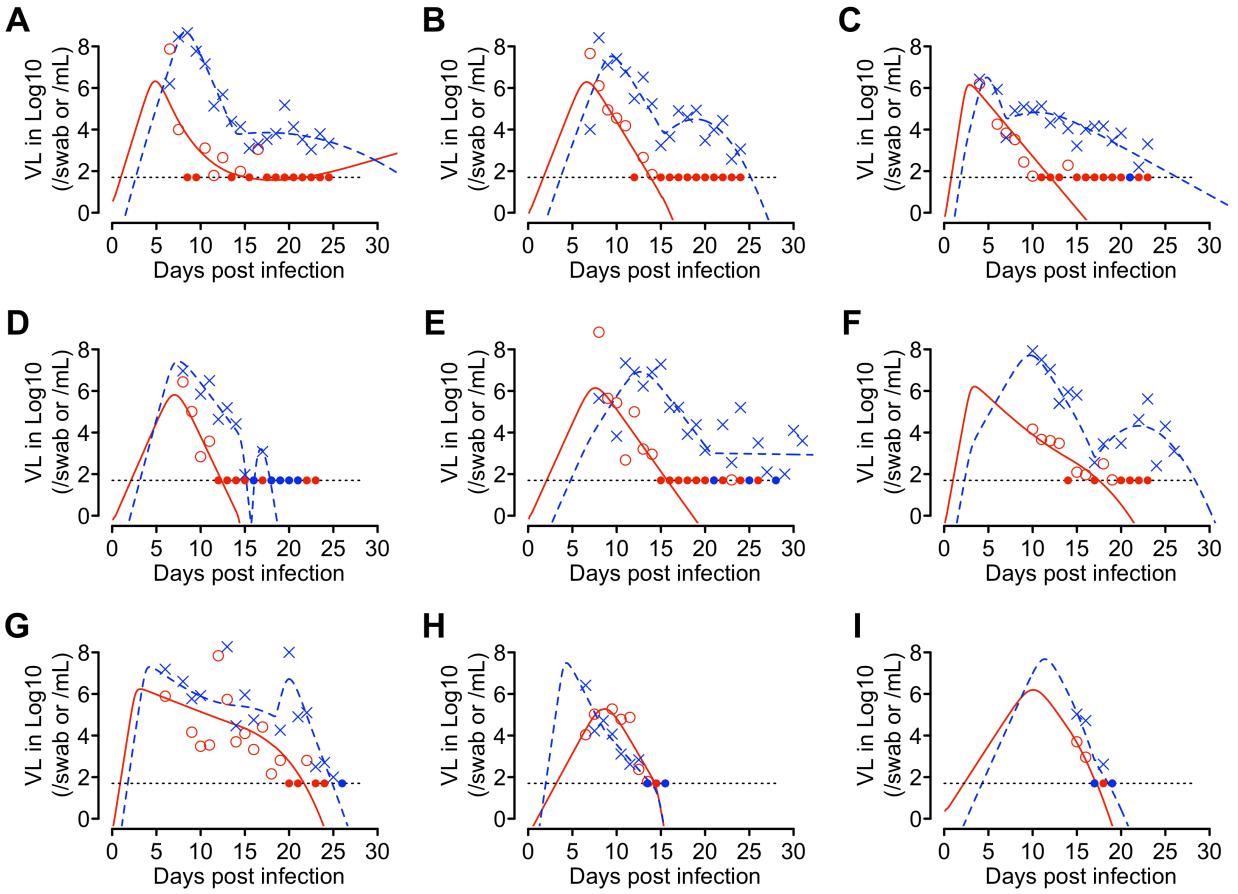

Figure S11. Data (symbols) and simulations (lines) of the combined model using best-fit parameter values for each individual. Red and blue denote viral load kinetics in the URT and the LRT, respectively. Symbols show the data from throat swabs (red squares) and the sputum samples (blue 'x's) as reported in Ref. (10). The dashed black lines show the limit of detection, and closed dots indicate data below the limit of detection.

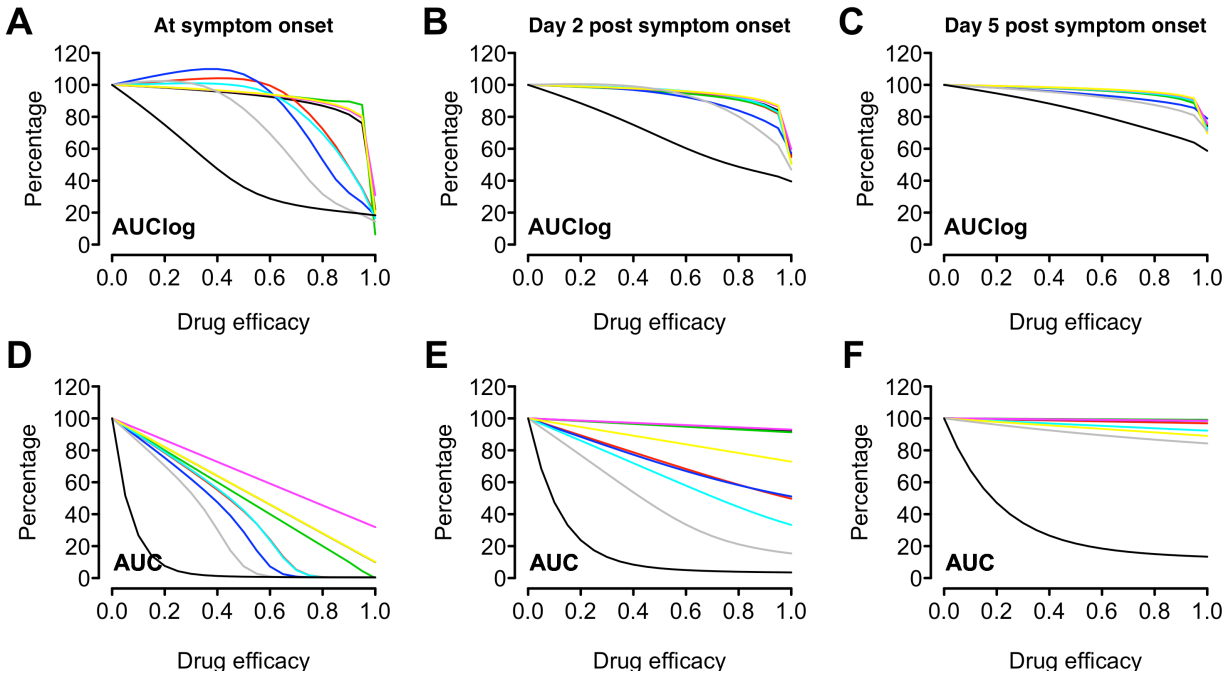

Figure S12. Simulated impact on AUClog or AUC in the URT of a potential therapeutic that blocks viral production. The area under the log viral load curve (AUClog; panels A, B and C) and the area under the viral load curve (AUC; panels D, E and F) are used to summarize the viral dynamics. AUClog or AUC values are normalized to the values obtained assuming drug efficacy of 0. We assumed that a therapeutic that blocks viral production is administered at the day of symptom onset (left panels), day 2 post symptom onset (middle panels) and day 5 post symptom onset (right panels). Each line represents simulations of the extended target cell model for an individual using their best-fit parameter values.

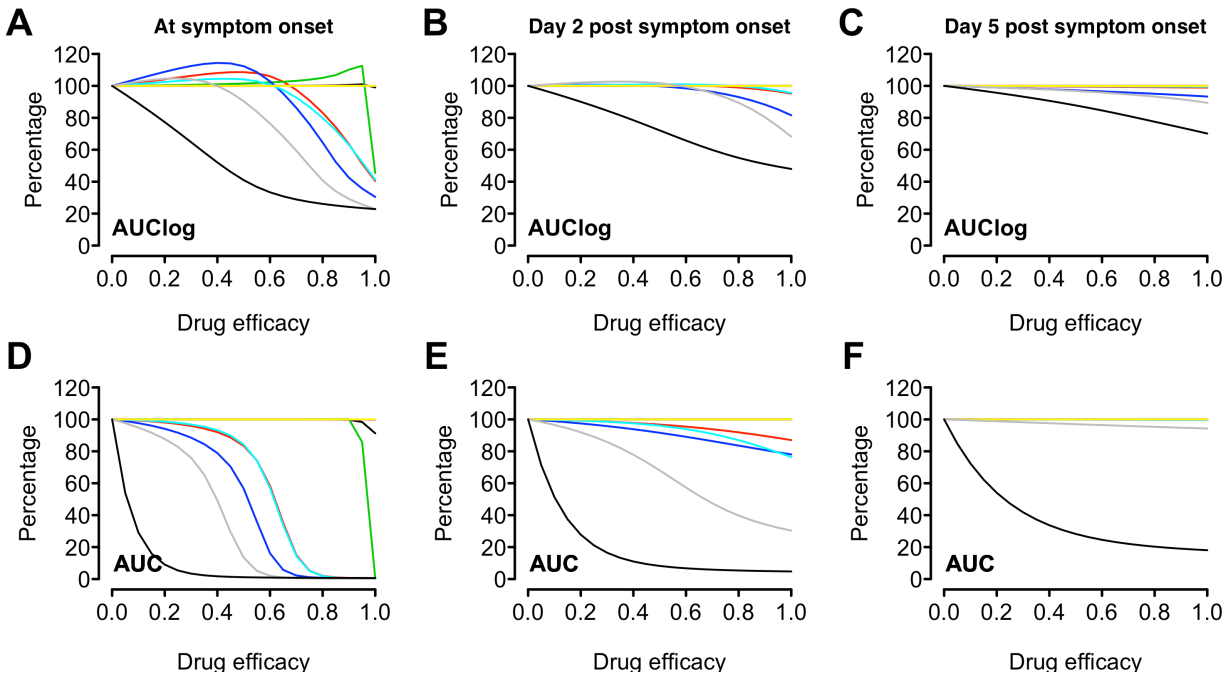

Figure S13. Simulated impact on AUClog or AUC in the URT of a potential therapeutic that blocks viral entry. The area under the log viral load curve (AUClog; panels A, B and C) and the area under the viral load curve (AUC; panels D, E and F) are used to summarize the viral dynamics. AUClog or AUC values are normalized to the values obtained assuming drug efficacy of 0. We assumed that a therapeutic that blocks viral production is administered at the day of symptom onset (left panels), day 2 post symptom onset (middle panels) and day 5 post symptom onset (right panels). Each line represents simulations of the extended target cell model for an individual using their best-fit parameter values.

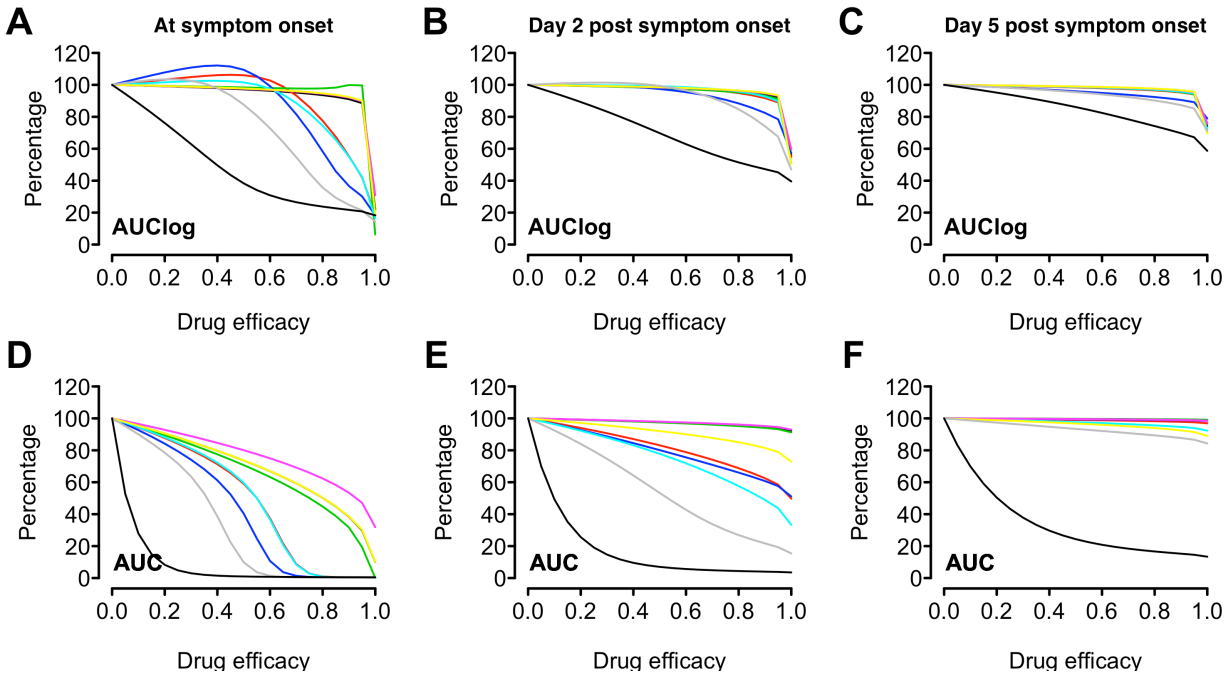

Figure S14. Simulated impact on AUClog or AUC in the URT of a potential combination therapy. The area under the log viral load curve (AUClog; panels A, B and C) and the area under the viral load curve (AUC; panels D, E and F) are used to summarize the viral dynamics. AUClog or AUC values are normalized to the values obtained assuming drug efficacy of 0. We assumed that a therapeutic that blocks viral production is administered at the day of symptom onset (left panels), day 2 post symptom onset (middle panels) and day 5 post symptom onset (right panels). Each line represents simulations of the extended target cell model for an individual using their best-fit parameter values.

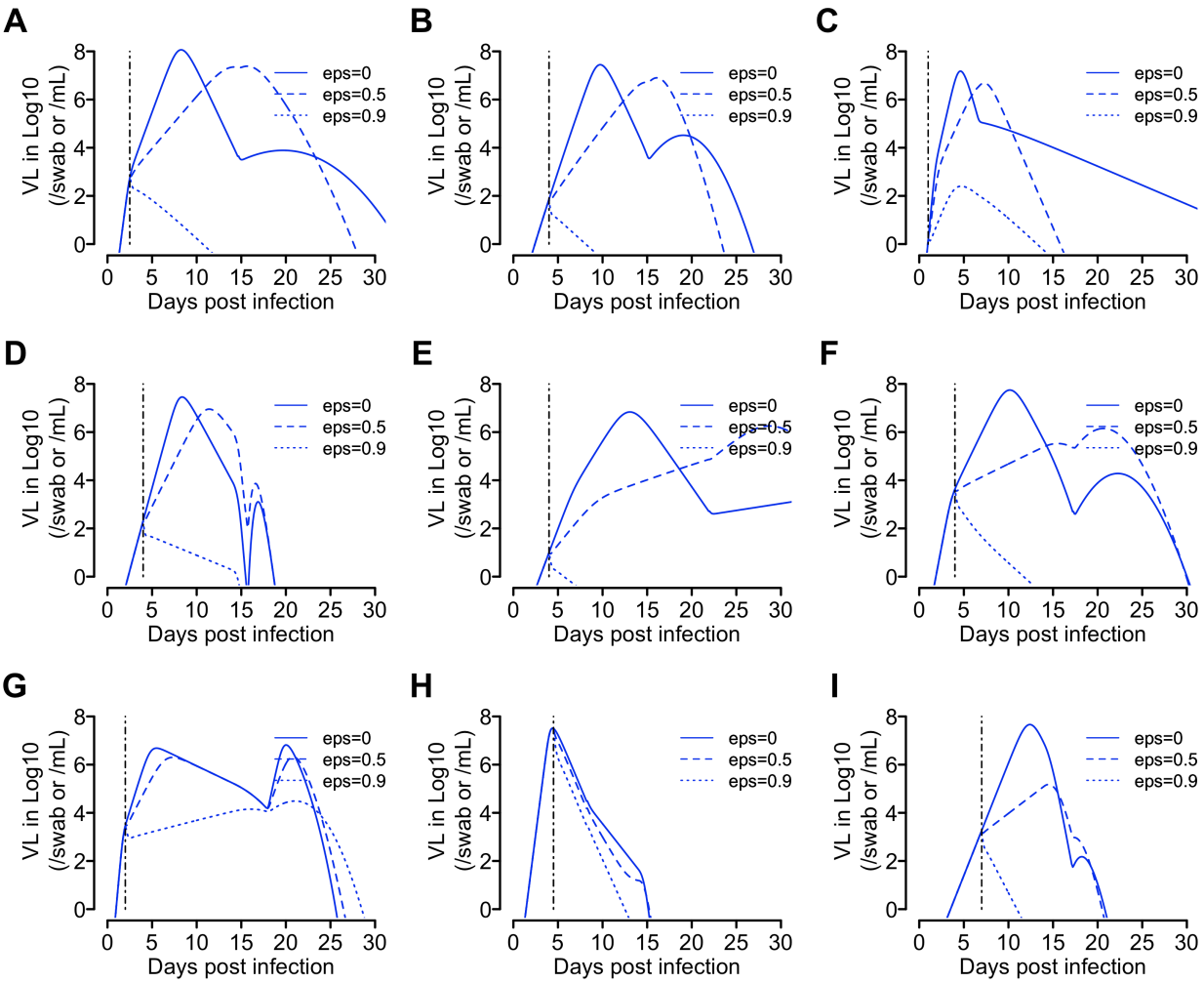

Figure S15. The impact of therapeutics on the viral load dynamics in the LRT when it is administered at symptom onset. Time course simulations of the extended target cell model using best-fit parameters for each patient. We assumed that the efficacy of the therapeutic to be 0 (solid line), 50% (dashed line) and 90% (dotted line).

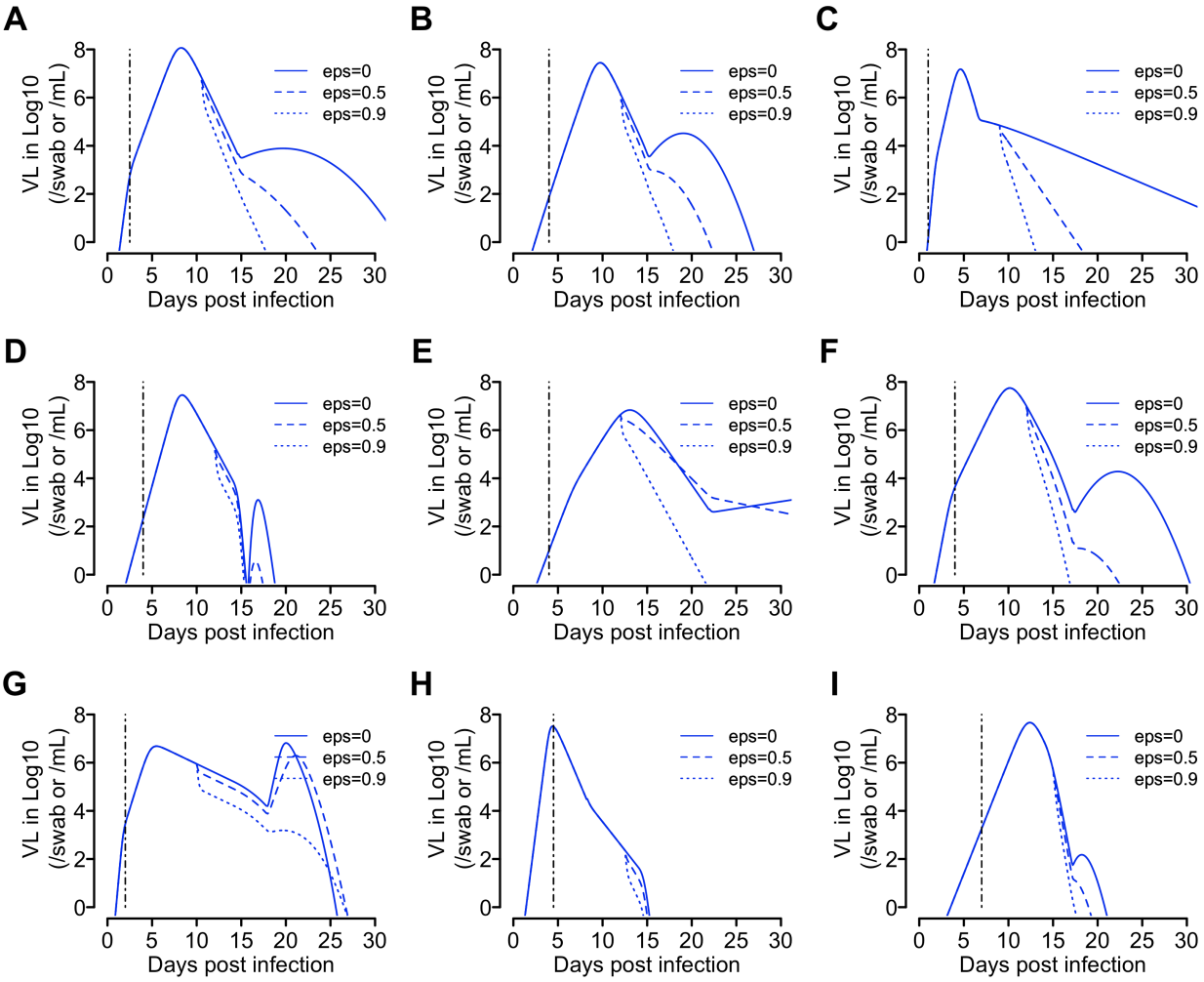

Figure S16. The impact of therapeutics on the viral load dynamics in the LRT when it is administered at day 8 post symptom onset. Time course simulations of the extended target cell model using best-fit parameters for each patient. We assumed that the efficacy of the therapeutic to be 0 (solid line), 50% (dashed line) and 90% (dotted line).

Supplementary Tables

Table S1. Comparison of models including "the incubation period" as a continuous covariate for model parameters using AIC scores. The parameter that covaries with the incubation period is listed in the model column.

Model (parameter that covaries with time to symptom onset)	AIC
No covariate	399.6
β_T	**388.0***
δ	392.7
π_T	**388.5***
β_S	400.4
δ_2	400.1
π_S	400.7

*Two models that are significantly better at explaining early data points (i.e. before 14 days of infection) than other models based on their AIC scores.

Table S2. Best-fit individual parameter values of the target cell limited model with "the incubation period" as a continuous covariate for parameter β_T. SD denotes standard deviation.

Panel (ID)*	β_T (10^{-8}/swab/day)	δ_1 (/day)	π_T (/day)	β_S (10^{-8}/mL/day)	δ_2 (/day)	π_S (/day)
A (1)	85.15	2.48	50.93	69.78	0.55	0.41
B (2)	51.95	2.31	50.99	68.86	0.54	0.35
C (3)	138.61	2.38	51.07	179.73	0.81	0.37
D (4)	53.08	2.44	53.34	95.9	0.71	0.37
E (7)	51.35	1.98	49.8	70.67	0.53	0.34
F (8)	51.77	2.06	50.15	75.3	0.49	0.37
G (10)	100.01	0.82	50.84	96.11	0.42	0.38
H (14)	44.46	2.15	51.83	168.57	1.62	0.4
Mean	**72**	**2.1**	**51.1**	**103.1**	**0.7**	**0.4**
SD	**33**	**0.5**	**1.1**	**45.3**	**0.4**	**0.02**

* Panels correspond to the panels shown in Fig. 1. ID numbers are according to the ID of the patients reported in Ref. (17).

Table S3. Estimated population parameters by fitting the TCL model with different fixed parameter values.

Model*	β_T (10^{-7} swab /copy/day)	σ_β	δ (/day)	π_T (/swab/day)	β_S (10^{-7} sample /copy/day)	δ_2 (/day)	π_S (/sample/day)
Baseline Model	19.3	-0.3	1.9	6.6	65.5	0.6	0.4
$I_0 = 10$ cells	18.7	-0.3	1.8	5.7	56.5	0.6	0.4
$c = 5$/day	49.8	-0.4	1.8	8.5	85	0.6	0.2

$c = 20$/day	13.7	-0.3	1.9	7.9	79.4	0.5	0.7
$k = 3$/day	40.1	-0.4	1.9	7.2	71.8	0.7	0.4
$k = 6$/day	19.4	-0.3	1.8	6.9	68.7	0.6	0.3
$\Gamma = 0.001$ /day	26.1	-0.3	1.8	6.6	65.9	0.7	0.6
$\Gamma = 0.1$ /day	16.5	-0.3	2	0.7	6.7	1.6	1.6

* In each model, the fixed parameter values are set to the same values as in the baseline model except for the parameters stated.

Table S4. Model outputs from sensitivity analyses of the TCL model. The times to viral peak (VL) and the reproductive numbers are the mean values of the values calculated using estimated individual parameter values in the population fitting approach.

Model	Time from infection to VL peak (days)		Time from symptom onset to VL peak (days)		$R_{0,URT}$	$R_{0,LRT}$	AIC score
	URT	LRT	URT	LRT			
Baseline Model	5.2	5.4	1.9	2	8.5	27.5	388
$I_0 = 10$ cells	5	5.6	1.6	2.2	7	21.8	389.8
$c = 5$/day	5.5	5.8	2.1	2.4	5.4	17.9	389
$c = 20$/day	5.1	4.7	1.7	1.4	14.6	68.1	389.6
$k = 3$/day	5.2	5.7	1.8	2.3	11.1	30.5	391.8
$k = 6$/day	5.3	5.5	1.9	2.1	6.9	23.4	387.6
$\Gamma = 0.001$/day	5.3	5.3	1.9	2	8.3	42.8	390
$\Gamma = 0.1$/day	5.4	7	2	3.7	7.9	5.3	389.9

Table S5. Model selection using AIC scores. The lower the AIC score the better the model.

Panel (ID)	TCL model	Innate immunity model	Proliferation model	Extended target cell model	Combined model
A (1)	22.6	3.6	15.8	9.3	3.2
B (2)	-6.3	-7	-10.5	-11.2	-7.3
C (3)	-22.5	-21.7	-22.1	-26.1	-21.3
D (4)	-21.5	-15.5	-19.8	-34.4	-29.8
E (7)	18.6	24.3	12.3	14.7	18.4
F (8)	6.8	1	-2.9	-6.9	-3.5
G (10)	32.1	29.2	27.0	28	30.5
H (14)	-12.1	-5	-10.1	-10.1	-4
I (16)	1.5	7.3	2.8	-1.5	8.2
Total	19.1	16.2	-7.5	**-38.3***	-5.6

*The extended target cell model has the lowest overall AIC scores among all the models.

Table S6. Parameter values of the best fit of the extended target cell model to the full dataset. SD denotes standard deviation.

Panel (ID)*	β_1 (10^{-6} /swab/day)	$\delta_{1,0}$ (/day)	π_1 (/day)	β_2 (10^{-7} /mL/day)	$\delta_{2,0}$ (/day)	π_2 (/day)	t_T (day)	Log_{10} T_N (cells)	w (/day)
A (1)	21.45	0.86	3.68	0.17	2.2	10.89	14.7	8.21	0.06

B (2)	1.31	1.82	15.53	0.8	2.18	2.46	15	8.44	0.18
C (3)	13.35	1.16	11.61	2.63	4.17	1.67	6.5	7.92	0
D (4)	2.4	3.55	11.53	1.35	1.6	1.7	15.7	10.99	2.4
E (7)	1.41	1.42	12.47	1.06	2.17	1.08	22	8.21	0
F (8)	6.94	0.76	5.89	0.17	3.33	10.34	17.3	8.79	0.15
G (10)	18.21	0.38	8.74	9.19	0.41	0.15	17.85	9	0.22
H (14)	5.12	3.53	4.5	4.9	2.04	1.64	8.3	6.89	1.89
I (16)	1.53	4.06	9.65	0.29	3.96	8.15	17.11	9.47	0.66
Mean	**7.97**	**1.95**	**9.29**	**2.28**	**2.45**	**4.23**	**14.94**	**8.66**	**0.62**
SD	**7.78**	**1.39**	**3.97**	**3.00**	**1.19**	**4.28**	**4.80**	**1.14**	**0.9**

* Panels correspond to the panels shown in Fig. 4. ID numbers are according to the ID of the patients reported in Ref. (17).

Table S7. Uncertainty analysis for the extended target cell model. The two numbers in each entry are the lower and upper bounds of the confidence intervals for the best-fit parameters. '-' denotes parameters that cannot be reliably identified.

Panel (ID)	$\boldsymbol{\beta_1}$ **(10^{-6} /swab/day)**	$\boldsymbol{\delta_{1,0}}$ **(/day)**	$\boldsymbol{\pi_1}$ **(/day)**	$\boldsymbol{\beta_2}$ **(10^{-7} /mL/day)**	$\boldsymbol{\delta_{2,0}}$ **(/day)**	$\boldsymbol{\pi_2}$ **(/day)**	$\boldsymbol{t_T}$ **(day)**	$\boldsymbol{\mathrm{Log}_{10}}$ $\boldsymbol{T_N}$ **(cells)**	$\boldsymbol{w}$ **(/day)**
A (1)	5.3, 206.5	0.5, 1.2	0.4, 5.7	0.02, 1.2	1.2, 4.6	2.0, 145.9	12.2, 16.6	7.6, 8.7	0, 0.3
B (2)	0.03, 4.5	1.1, 10.0	4.1, 1021	0.1, 2.8	1.0, 7.4	0.5, 15.1	13.0, 16.4	7.7, 8.9	0, 0.3
C (3)	2.3, -	0.8, 1.5	0.2, 227.8	1.6, 78.6	1.0, 10.0	0.2, 7.5	4.2, 7.6	6.1, 8.3	0, 0.1
D (4)	0.003, 29.0	0.7, 10.0	1.5, -	0.03, 4.1	0.8, 1.8	-	14.9, 15.9	9.8, 13.9	0.5, 10
E (7)	0.2, 9.8	0.6, 1.3	1.6, 113.1	0.7, 2.0	1.1, 6.5	0.8, 1.4	15.9, 29.1	7.8, 8.3	0, 0.1
F (8)	0.9, -	0.5, 1.0	0.1, 111.3	0.06, 1.1	1.2, 10	-	15.5, 18.9	8.1, 9.0	0, 0.3
G (10)	-	0.3, 0.5	-	-	0.3, 0.7	0.1, 75.0	18.5, 19.7	5.4, 10.1	0.1, 0.25
H (14)	0.9, 25.9	1.1, 36.7	0.6, 56.2	-	0.9, 10	0.3, 10.6	3.6, 30	0, 8.0	0.2, 10
I (16)*	-	-	-	-	-	-	-	-	-

* Parameters for patient 16 cannot be identified due to the small number of data points.

Table S8. Model fitting suggests g_{21} is approximately 0. Estimated population parameter values and summary of model outputs by fitting the TCL model (without setting $g_{21} = 0$) to data up to 14 days of infection. The AIC score of the model is 388.6.

Parameter	**Description**	**Estimated value (population estimate)**
β_T (random effects)	Infectivity parameter (URT)	1.92e-6 swab/copy/day
σ_β	Covariate coefficient of time to symptom onset on β_T	-0.36
δ (random effects)	Death rate of infected cells (URT)	2.0 /day

π_T (random effects)	Composite parameter for virus production and sampling (URT)	62.9 /swab/day
β_S (random effects)	Infectivity parameter (LRT)	8.1e-7 sample/copy/day
δ_2 (random effects)	Death rate of infected cells (LRT)	0.56 /day
π_S (random effects)	Composite parameter for virus production and sampling (LRT)	0.33 /sample/day
g_{21} (fixed effect)	Transport rate from the LRT to the URT	7.8e-5 /day