Spatially distributed infection increases viral load in a computational model of SARS-CoV-2 lung infection

Melanie E. Moses, Steven Hofmeyr, Judy L. Cannon,
Akil Andrews, Rebekah Gridley, Monica Hinga, Kirtus Leyba,
Abigail Pribisova, Vanessa Surdidijaja, Humayra Tasnim,
and Stephanie Forrest

A key question in SARS-CoV-2 infection is why viral loads and patient outcomes vary dramatically across individuals. Because spatial-temporal dynamics of viral spread and immune response are challenging to study in vivo, we developed Spatial Immune Model of Coronavirus (SIMCoV), a scalable computational model that simulates billions of lung cells, including respiratory epithelial cells and T cells. SIMCoV replicates viral growth dynamics observed in patients and shows that spatially dispersed infections lead to increased viral loads. The model shows how the timing and strength of the T cell response can affect viral persistence, oscillations, and control. When the branching airway structure is explicitly represented, we find that virus spreads faster than in a 2D layer of epithelial cells, but much more slowly than in an undifferentiated 3D grid or in a well-mixed ODE model. These results illustrate how realistic spatially explicit computational models can improve understanding of within-host dynamics of SARS-CoV-2 infection.

1 Introduction

Reducing the spread, severity, and mortality caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection is an urgent global priority. A key question is why viral loads and patient outcomes vary so dramatically among infected individuals. While severe disease is correlated with some known risk factors, such as age, co-morbidities, and some immune response characteristics (*1, 2, 3*), it remains challenging to understand why some patients develop high viral loads while others do not. We developed a scalable computational model to study the role of spatial effects in determining the time course of viral load within patients, which has been shown to affect disease severity (*4*) and transmission (*5*).

Most prior work uses Ordinary Differential Equation (ODE) models to represent within-host virus dynamics, e.g., (*6, 7, 8, 9, 10, 11, 12*). Such models are useful for studying the onset and duration of the infective period (*13*) and the effect of various therapeutics given at different times (*14, 15, 16*). However, they have limited ability to fully account for dynamics in the large and complex space of the lung (*17, 12*).

The Spatial Immune Model of Coronavirus (SIMCoV) agent-based model (ABM) simulates infection dynamics and $CD8^+$ T cell responses in millions to billions of spatially organized epithelial cells. We highlight the importance of the spatial dispersion of virus in the lung in determining peak viral load. We focus on the T cell component of immune response because it is a key component of immunity to SARS-CoV-2. Earlier studies report that T cell levels are correlated with disease severity (*18*), and there is mounting evidence that T cells are critical for protection from severe illness and for preventing subsequent infection (*19, 20, 21, 22*). These studies collectively suggest that an early and strong T cell response is correlated with less severe disease, but they do not provide a quantitative understanding of how T cell response impacts viral dynamics.

Our simulations replicate viral dynamics reported from infected patients and explain differences among those patients in terms of the initial number of Foci of Infection (FOI) and the magnitude and timing of the $CD8^+$ T cell response. A SIMCoV simulation also mimics the spread of tissue damage observed in CT scans from an infected patient. SIMCoV demonstrates how the spatial structure of the lung can affect viral spread in the tissue and contribute to the wide diversity of outcomes and infection dynamics observed in COVID-19 patients.

Results

SIMCoV Model Overview

SIMCoV explicitly represents virus, epithelial cells, $CD8^+$ cytotoxic T cells, and the inflammatory signals that cause T cells to extravasate into lung tissue (Fig. 1). We initialize SIMCoV with SARS-CoV-2 virus that can infect lung epithelial cells. Epithelial cells are modeled as a 2D grid, where each grid point represents a 5 x 5 x 5 μm^3 volume. Upon infection, epithelial cells incubate virus in what is often called an eclipse phase, and then they express virus until they die. Expressing cells produce both virus and an inflammatory signal, which is an abstraction of the cytokines that influence T cell extravasation into tissue. The virus and inflammatory signal both diffuse through tissue. After seven days, we create an abstract pool representing T cells that are activated in lymph nodes and then circulate in the vasculature. When T cells reach the lung and encounter a concentration of inflammatory signal above a threshold, they extravasate and then are explicitly represented as mobile agents in SIMCoV. We adjust the probability of T cell extravasation according to the fraction of the lung that is modeled in the simulation. T cells that have extravasated move at random across the 2D grid until they either die or encounter an epithelial cell incubating or expressing virus, in which case the T cells cause apoptosis of the infected epithelial cell with some probability. Each transition in this process is parameterized with estimates or probabilities detailed in Materials and Methods and in configuration files provided with

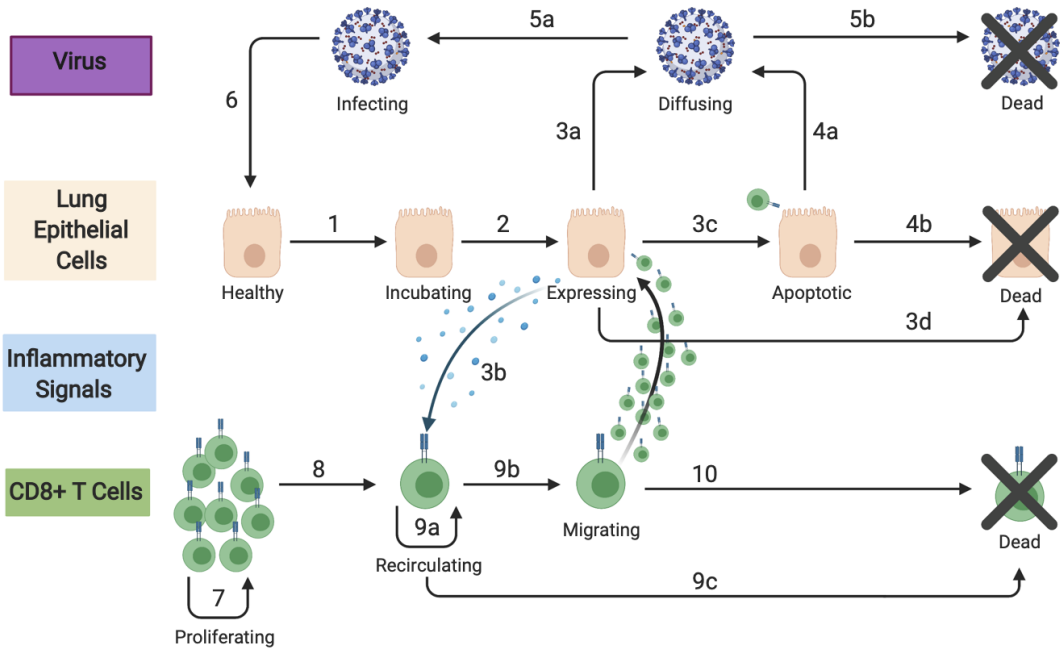

Figure 1: SIMCoV model components and their interactions. Epithelial and T cells are represented as agents; virions and inflammatory signals are represented as concentrations. Numbered transitions are described in Materials and Methods (Table 2).

the SIMCoV code in `https://github.com/AdaptiveComputationLab/simcov`.

SIMCoV most closely resembles an ABM developed for influenza (*23, 24, 25*), extending the earlier work in several ways: (1) scaling up from a two-dimensional monolayer to a high-performance computing (HPC) implementation of billions of cells in a three-dimensional branching lung structure; (2) including T cell circulation and migration from the vasculature to the site of infection; and (3) parameterizing the model for SARS-CoV-2.

Fig. 2 shows the results of a run of SIMCoV with the default configuration, representing infection spreading over a 15 x 15 mm^2 layer of lung epithelial tissue (5 μm deep). The top row shows the state of epithelial cells (left), virion concentration (center), and inflammatory

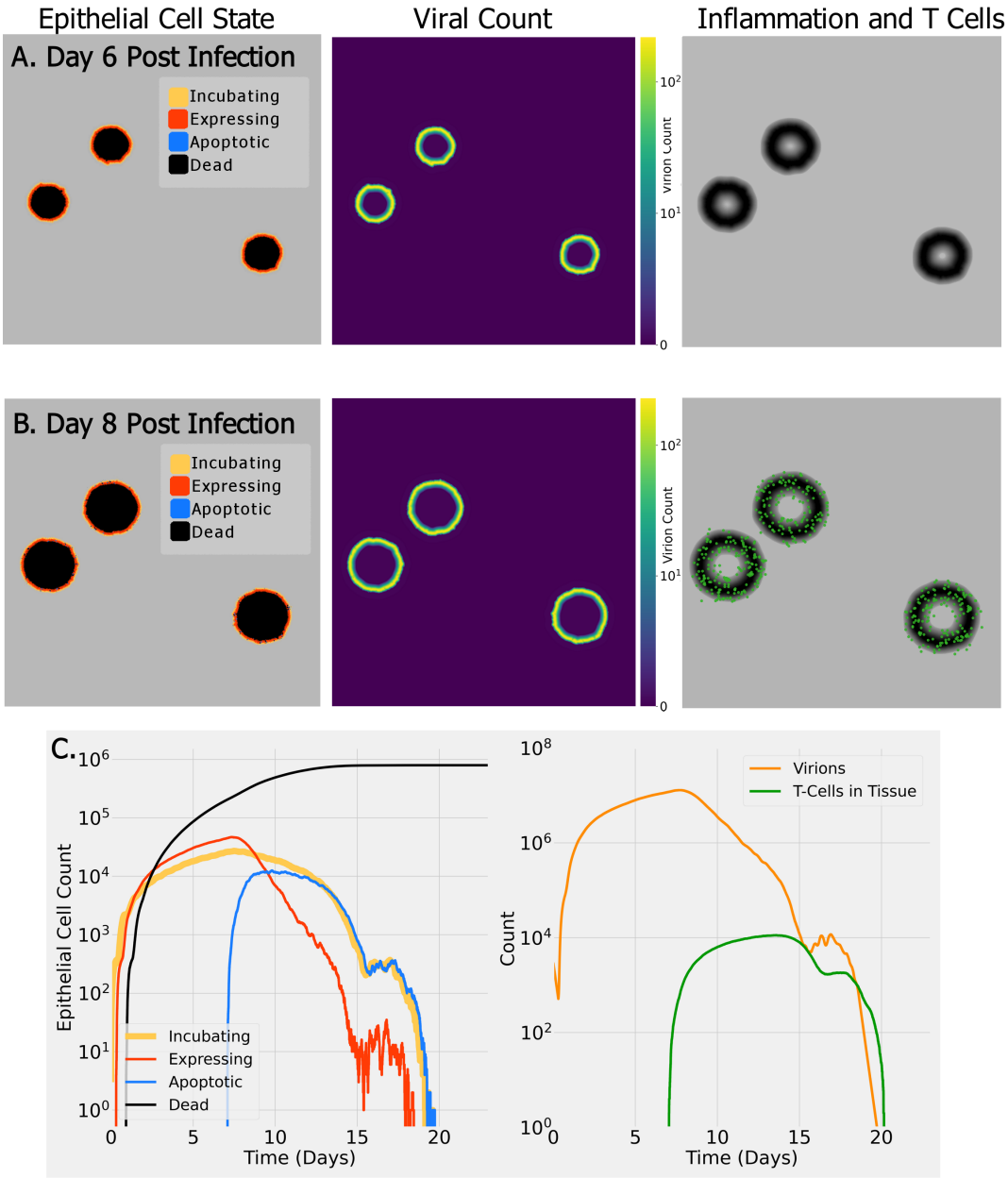

Figure 2: Example SIMCoV simulation on a 15 x 15 mm^2 layer of epithelial tissue. (a) 6 days post infection (dpi); (b) 8 dpi. Color indicates epithelial cell state (see legend), virus concentration (see scale bar), and presence of inflammatory signal (black) and T cells (green); (c) (left panel) shows the dynamics of epithelial cell state, and (right panel) shows virus and T cell dynamics. A video demonstration of this simulation can be viewed at `https://www.youtube.com/watch?v=5K4ZSeCn-Ks`.

signal (right) at six days post infection (dpi), and the second row (Fig. 2b) shows the same including extravasated T cells at eight dpi. The dynamics of the epithelial cells, the count of virions, and the number of T cells moving through tissue are shown in the bottom row (Fig. 2c). The number of epithelial cells incubating and expressing virus (Fig. 2c, left) and viral load (Fig. 2c, right, orange line) all rise quickly after infection and then increase more slowly until seven dpi. The $CD8^+$ T cell response begins at seven dpi, and when T cells encounter infected cells, the infected cells become apoptotic. This reduces both the incubating and expressing epithelial cells and the viral load, and the cumulative number of dead cells levels off. In this example, at approximately 15 dpi, the T cells induce apoptosis in nearly all expressing epithelial cells, which leads to a reduction in produced inflammatory signals and consequently fewer T cells entering the lung. This causes the virus level to stabilize temporarily. However, once incubating and newly infected cells express more virus, more T cells enter the tissue to seek out these expressing cells and reinstate control over the infection.

Peak viral load is proportional to the number of initial FOI

We used SIMCoV to test how spatially differentiated locations of initial infection affect viral dynamics. Images of the lungs of SARS-CoV-2 patients and experimental animal models suggest that there may be multiple sites of viral infection (*26, 27*). We tested the effect of seeding SIMCoV with multiple FOI on viral dynamics. We considered one, four, and sixteen spatially dispersed initial sites of infection. We find that each four-fold increase in the number of FOI increases the peak viral load and number of extravasated T cells by a factor of approximately four (Fig. 3). This remains true when the absolute number of initially infected cells is held constant across all experiments, i.e., by creating a single FOI with a cluster of sixteen adjacent infected cells and four FOI with four adjacent infected cells, so that all three experiments contain sixteen initially infected cells but with varying number of FOI. The spatially dispersed

sites act independently, so virus spreads more quickly with higher FOI because there are more infectable cells surrounding each site of infection. T cells respond by entering the tissue in greater numbers when there is more inflammatory signal produced at more sites of infection.

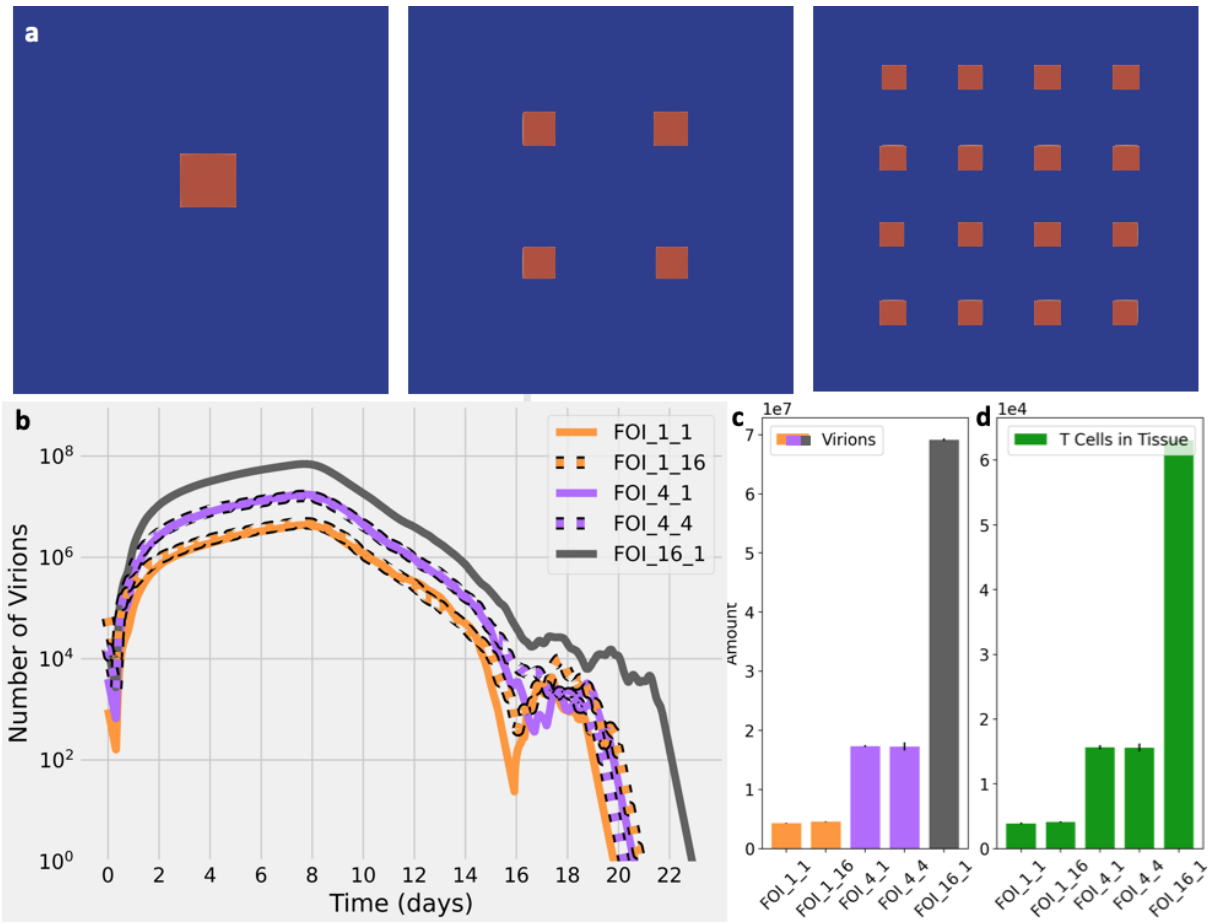

Figure 3: Effect of FOI on virus and T cell dynamics. **(a)** Initial spatial distribution in SIMCoV for 1 FOI (left), 4 FOI (center), and 16 FOI (right). **(b)** Viral load over time given different FOI. **(c)** Peak number of virions for different FOI with error bars showing the standard deviation for 10 replicates. There is a 3.90-fold increase from the 1 FOI configuration to the 4 FOI configuration and a further 3.99 increase for the 16 FOI configuration. **(d)** Peak number of T cells that extravasate into lung tissue for the different FOI. Similar to panel **(c)**, the increase is 3.95-fold for each 4-fold increase in FOI.

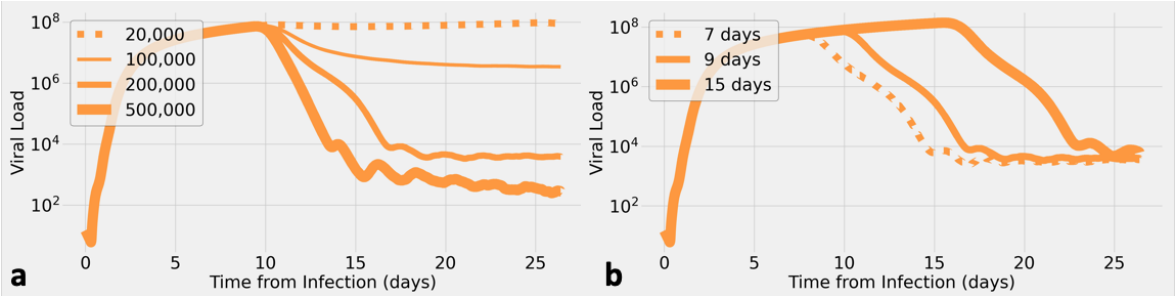

Figure 4: Effect of varying key T cell simulation parameters. (a) Effect of varying T cell generation rate when T cell delay is 9 dpi; (b) effect of varying T cell delay when T cell generation rate is 200,000 T cells/min.

Effect of CD8$^+$ T cell response on viral clearance

CD8$^+$ T cells that kill virally infected cells are another key factor that varies among patients (*19, 22*). The CD8$^+$ T cell response is dependent on interactions in physical space: CD8$^+$ T cells are produced in lymph nodes, circulate in the blood and migrate to the lung, then extravasate at locations in lung tissue with inflammatory signal, and then move within infected tissue until they directly interact with infected cells to cause cell death (*28,29,30,31*). T cells are likely to play an important role in controlling SARS-CoV-2 infection as elderly individuals and patients suffering from severe COVID-19 disease have fewer circulating CD8$^+$ T cells (*32, 19*). However, due to technical difficulties assessing T cell responses in patients, relatively little is known about how T cell responses in the infected lung might impact viral dynamics and control.

We use SIMCoV to test how the magnitude of the T cell response affects viral control. We find that varying numbers of circulating T cells in the blood dramatically impacts viral control (Fig 4a). Even a 2-fold decrease in the production rate of circulating T cells can delay viral control (compare solid orange lines) or even differentiate between viral control versus continued viral persistence (compare dashed orange line where virus-responding T cells are produced at a rate of 50,000 per minute to thin solid orange line where it is 100,000 per minute).

Later CD8$^+$ T cell entry into the infected lung also delays the time it takes to reach peak

viral load and allows a slightly higher peak load (Fig. 4b). Delayed T cell arrival also increases the time taken for eventual control of the infection, but it does not dramatically change the rate of reduction of viral load. Even when T cell entry into infected lung is delayed from 7 to 15 days, viral control can eventually be achieved, albeit later in the infection. The delay in viral control allows a longer period of elevated viral load, which may increase the window of transmission as well as tissue damage from the virus and the immune response. Consequently, delays in T cell responses may lead to more severe disease.

SIMCoV viral loads compared to patient data

Having shown how the number of initial FOI and the timing and number of responding $CD8^+$ T cells affect viral dynamics, we next consider how these factors affect SIMCoV fits to time courses of viral load in COVID-19 patients. Following previous ODE models, e.g. (*7, 8, 14, 12*), we use data from sputum samples reported in Wolfel et. al. (*33*). Most of these earlier models used a target cell limited (TCL) framework, in which viral load peaks when newly infectable "target" epithelial cells are depleted by previous infection. These models often predict viral loads that peak within the first four dpi. However, SARS-CoV-2 peak viral load occurs later in the lung than in the upper respiratory tract (*34*), and patient data show complex dynamics in the lung that can last 15 to 30 dpi (Fig. 5). Simple TCL models may not be as suitable for describing infection dynamics in the larger area of the lung where the epithelial lining of the branching airways and alveoli forms a large and complex surface area of up to 80 m^2 (*35*). Although target cell limitation may be reached earlier in the upper respiratory tract, in the lung, it may not be achieved before the adaptive immune response begins (*17*).

We tested whether SIMCoV can explain the substantial variation in patient viral load data by varying only the initial FOI, T cell response, and viral clearance, which includes other immune activity. We started with the default model parameters (Materials and Methods - Table 2 and

Patient	T Cell Generation Rate (1000 T cells/min)	T Cell Onset (days)	Initial Sites of Infection	Viral Clearance	SIMCoV RMSLE	ODE RMLSE
A	200	8	220	0.003	1.83	1.79
B	160	7	150	0.003	2.82	2.55
C	90	5	1	0.003	2.01	2.05
D	120	6	40	0.005	2.32	1.62
E	120	10	10	0.003	3.74	3.40
F	150	9	80	0.003	2.34	1.92
G	90	6.5	50	0.004	3.58	2.88
H	160	6	3	0.004	1.31	1.82

Table 1: SIMCoV parameter settings used for each patient.

Fig. 2) and varied the number of initial sites of infection from one to 220 sites. We also varied the first day that T cells arrive in lung tissue (from 5 to 10 days), the T cell generation rate (from 90,000 to 200,000 cells per minute), and the viral clearance term (from 0.003 to 0.005) to account for variations in other components of the immune response.

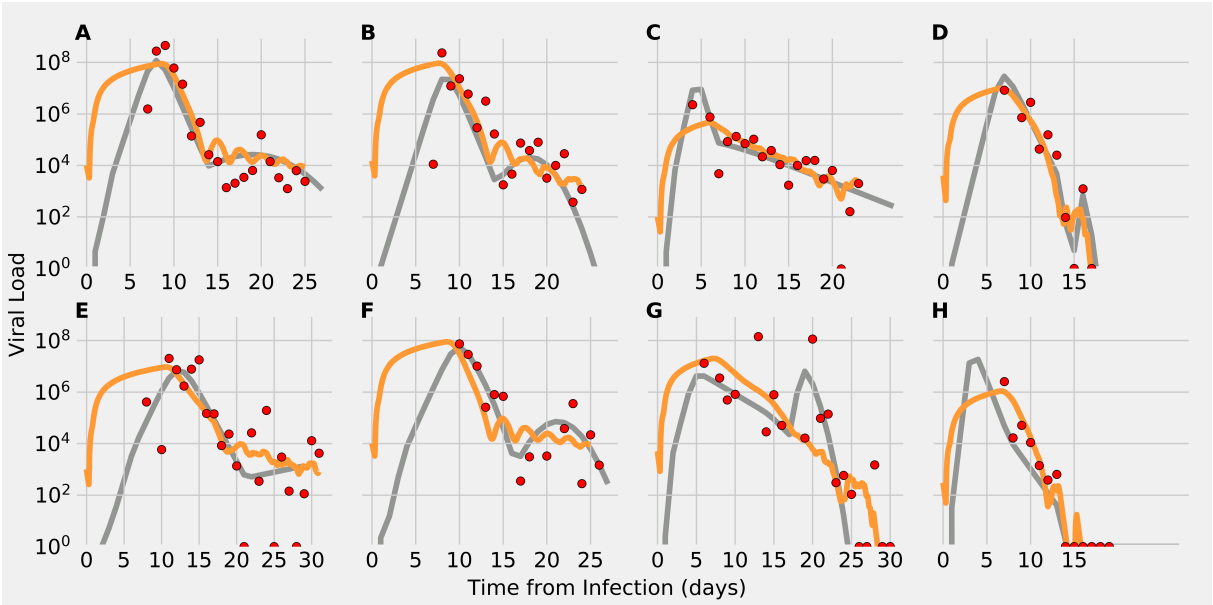

Figure 5: Viral load predictions from SIMCoV simulations (orange) compared to the extended ODE in (*12*) (gray), and patient data from (*33*) (red points).

We find that SIMCoV captures the dynamics of patient viral load over the full 25-30 day

time courses with fits that are as good as those of the extended ODE described in Ke et al. (*12*) (Fig. 5). Table 1 shows that the Root Mean Log Squared Error (RMLSE) is similar between SIMCoV and the ODE; the RMLSE is a goodness-of-fit measure that is suitable for curves with a wide-range of values. By varying the initial FOI, we replicate how peak viral load varies by several orders of magnitude across the eight patients. Compared to the ODE model, SIMCoV produces a less pronounced peak with higher viral loads early in infection that persist for several days. The SIMCoV curve is consistent with the observation that people are most infectious early in infection before symptom onset. SIMCoV also predicts a longer duration of elevated viral titers, which matches patient viral load data.

Varying the T cell generation rate and first day of arrival captures the unique shapes of viral load trajectories after seven dpi. SIMCoV exhibits both oscillations in viral load and second peaks that are seen in some patients (*33*). Our results support the hypothesis that decreased T cell numbers can account for viral persistence and inability to control viral load in specific individuals (*19, 20*).

Previous TCL models have replicated the data in Wolfel et al. (*33*) by expanding the availability of target cells later in infection and varying many parameters in each patient (*7, 12*). For example, the extended ODE model in (*12*) achieves good fits by varying infectivity, infected cell death rate, virion production rate, number and timing of new target cells introduced to represent newly seeded infections, and the exponential increase of the infected cell death rate to account for the adaptive immune response.

By incorporating spatial interactions, SIMCoV provides a more parsimonious explanation for the dramatically different viral load trajectories among patients by varying only the number of FOI in the initial infection and the magnitude and timing of the immune response, particularly the T cell response. These factors are consistent with hypothesized explanations for highly variable patient outcomes. Elderly individuals have lower T cell numbers, and our model suggests

that fewer T cells along with slower T cell response could help explain why elderly individuals have increased disease severity (*19, 22*). More speculatively, the number of FOI could increase with a larger inhaled viral innoculum, a factor that Gandhi et. al. (*36*) suggest would increase disease severity. Variation in FOI and T cell response may also account for the long infectious period in some patients (*6, 4, 5*).

SIMCoV reflects spatial features of CT scans

Computed tomography (CT) scans of COVID-19 patients have been valuable tools for assessing and studying SARS-CoV-2 infection severity since the beginning of the COVID-19 pandemic. A hallmark of SARS-CoV-2 infection in CT images is the presence of lesions called Ground Glass Opacities (GGOs) and consolidations, likely indicating tissue damage caused by inflammatory cell infiltration (*37, 38*). CTs have been used as a diagnostic for specific COVID pneumonia, and CT scans often show multi-focal distribution of both GGOs and consolidations, suggesting spatial distribution of infection as an important feature of SARS-CoV-2 infection. Although their clinical value in prognosis and diagnosis is still unclear (*27, 39, 40*), we can test how well SIMCoV replicates the spread of infection seen in CTs from COVID-19 patients.

As a proof-of-concept, we considered a time series of three chest CT slices taken at a similar depth from a 36-year-old female patient from Hubei Province, China (*41*). Images were taken at 2, 7, and 11 days after hospital admission. Because symptoms are estimated to begin on average 4 days post-infection (*42,43*), we estimate that these images are at 6, 11, and 15 dpi. The original images, with GGOs and consolidations labeled in green, are shown on the top row of Fig. 6. Because GGOs and consolidations are associated with inflammation, which is an indicator of immune response induced at sites of viral infection, we used GGO and consolidation locations to seed a 2D simulation in SIMCoV.

For each of the CT images, we identified areas exhibiting traits of GGOs and consolidations,

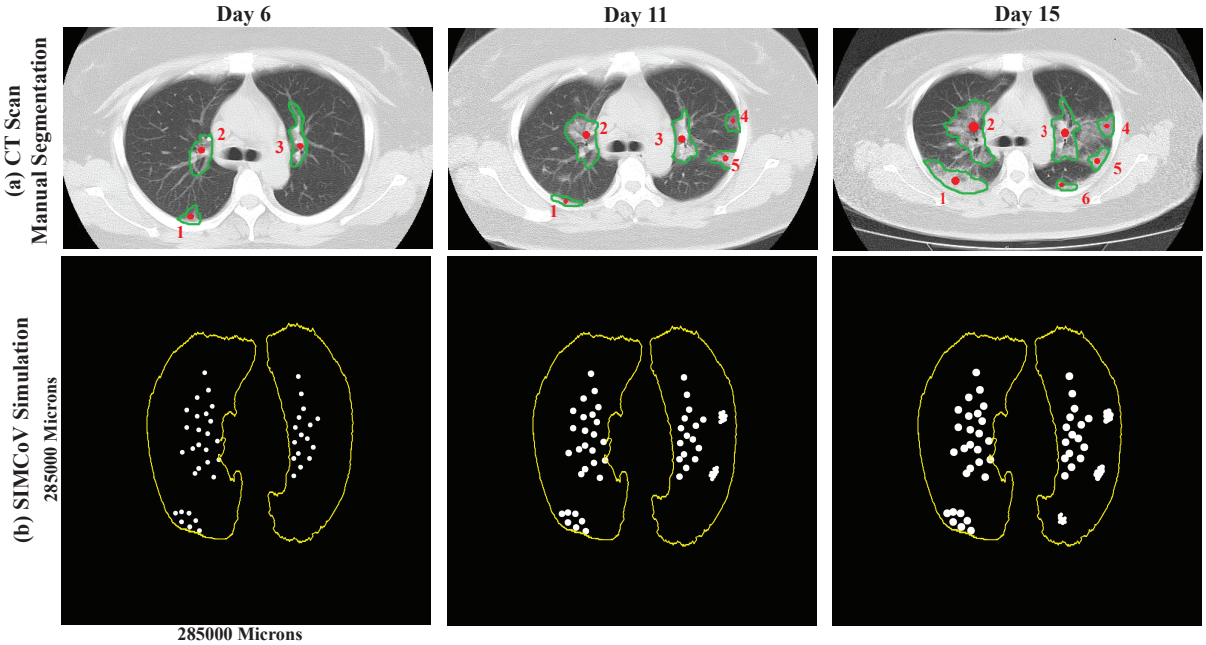

Figure 6: Comparing multi-focal viral spread in a single patient CT scan to a SIMCoV simulation. **(a)** CT scans on estimated 6, 11, and 15 dpi; GGOs and consolidations are outlined in green, and centroids are shown in red. **(b)** SIMCoV experiments; white regions indicate the accumulation of inflammatory signals and non-healthy (incubating, apoptotic, expressing, and dead) epithelial cells. The yellow line outlines the lung. In the SIMCoV simulation, infection sites 1, 2, and 3 were seeded at 0 dpi, sites 4 and 5 were seeded at 7 dpi and site 6 was seeded at 12 dpi. Visibly damaged regions in the CT scans resemble damaged regions in the SIMCoV simulation.

as indicated by green outlines in Fig. 6, and labeled the centroids of each area with red dots. We seeded SIMCoV with clusters of SARS-CoV-2 infected cells surrounding those locations. The model was initialized with infection at sites 1, 2 and 3. Sites 4 and 5 were seeded at 7 dpi and site 6 was seeded at 12 dpi to reflect new sites of infection that appear in the CT scan.

The simulation was run on a 57000 x 57000 2D grid which is equivalent to 285000 x 285000 x 5 μm^3. The yellow outlined lung area in Fig. 6b approximately matches the dimension of an average 2D slice of human lung (*44*). SIMCoV was run for 15 simulation days. The infected and damaged areas containing non-healthy epithelial cells and inflammatory signals from the SIMCoV experiment are shown in white in Fig. 6b. These areas of tissue damage at 6, 11, and

15 dpi in SIMCoV resemble the images of tissue damage present in the CT images in Fig. 6a, suggesting that SIMCoV can replicate spatial features of viral spread as reflected in CT scans.

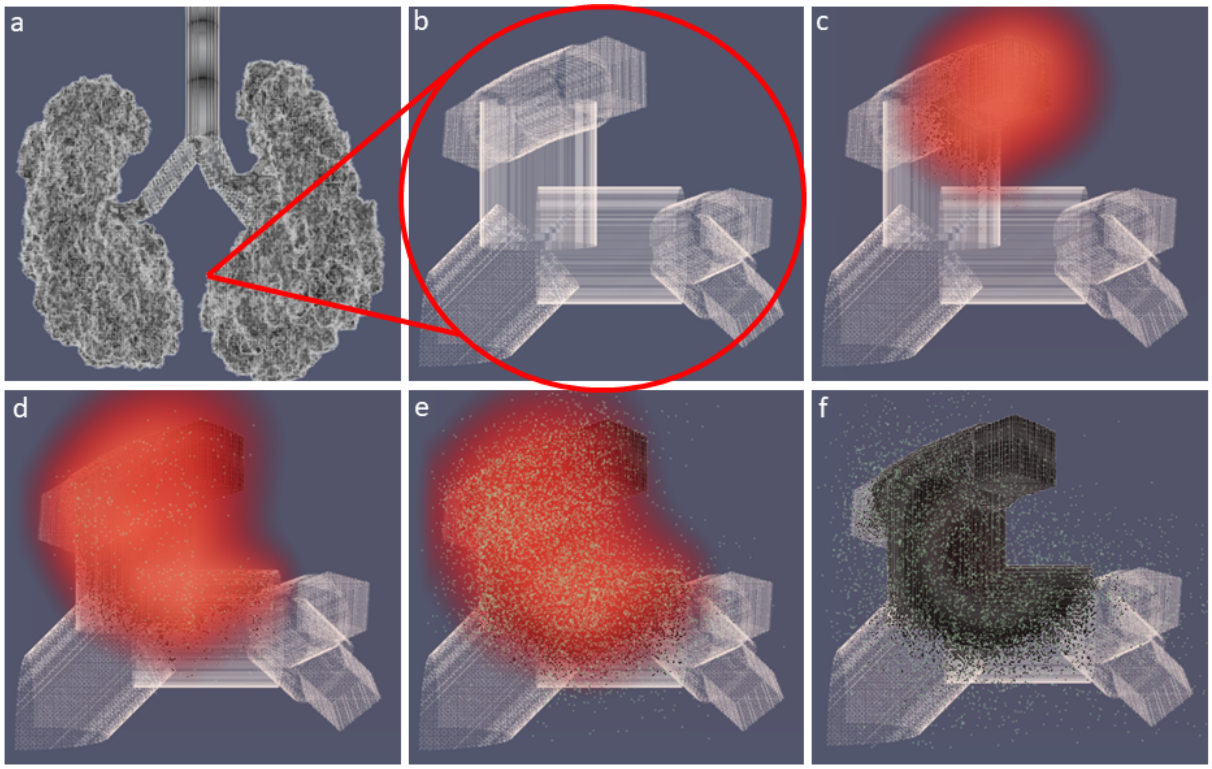

Figure 7: Simulated 3D branching airway with a single FOI. **(a)** shows the full lung model; **(b)** - **(f)** show T cells (green), virions (red) and a small region of epithelial cells (gray) at **(b)** 0 dpi, **(c)** 4.2 dpi, **(d)** 6.3 dpi, **(e)** 6.9 dpi, and **(f)** 8.3 dpi. Dead epithelial cells are indicated in black. A video corresponding to these images is available at `https://youtu.be/WbiLH9NRfbA`.

Modeling viral spread over branching airways

The SIMCoV experiments described above simulate a single layer of epithelial cells (Figs. 2, 3, 4, 5, 6). However, the lung is a highly complex structure of branching airways descending from large bronchioles down to alveoli lined by epithelial cells, with much of the volume consumed by alveolar space for gas exchange (*35*). To investigate whether the lung airway structure impacts viral spread and T cell control, we developed a bifurcating fractal branching algorithm

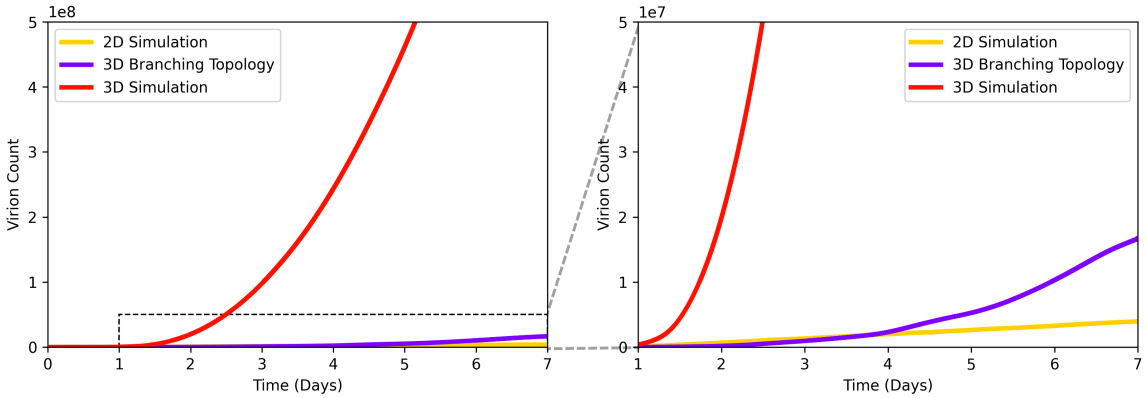

Figure 8: Viral spread over different model topologies. Virion count is plotted vs. time for three different SIMCoV configurations: 2D grid (yellow); 3D grid (red); branching model (purple). Each simulation was run with approximately 225 million epithelial cells.

and integrated it with SIMCoV. Following Yeh et. al (*45*), we model 26 branching generations (from the trachea to alveolar sacs) in all five lobes of the lung. The 3D branching model incorporates trachea, bronchi, bronchioles, and ducts, represented as cylinders of epithelial cells with length, diameter, branching angle, and gravity angle calculated at each level of the branching hierarchy, and each terminal alveolar sac is represented as a hollow 200 x 200 x 200 μm^3 cube of epithelial cells. Fig. 7a shows the full model of the branching structure, and Figs. 7b-f show how virus spreads over a small region of bronchioles and alveoli in a 1500 x 1500 x 1500 μm^3 volume.

Fig. 8 shows how the topology of epithelial cells modeled in SIMCoV affects the predicted rate of viral spread in the first seven dpi and before the T cell response. Viral spread over the branching network (indicated with purple in Fig. 8) shows four times greater growth than the 2D layer (indicated with yellow) that was modeled earliler. In contrast, an undifferentiated 3D grid of cells generates 66 times more infected cells at the end of the first week of infection. This suggests that the computationally less expensive 2D model provides a better approximation of the dynamics of the branching epithelium than the 3D non-branching model. Future imple-

mentations of SIMCoV will incorporate different diffusion and T cell movement rates in the airway, epithelium, and interstitial space as well as details about the percentages and locations of infectable epithelial cells.

Discussion

SIMCoV models the spread of SARS-CoV-2 through the lung and subsequent immune control by T cells. Results from our spatial model complement earlier well-mixed ODE models but make different predictions. For example, the earlier models predict a sharp early peak in viral load in the first few days of infection, while SIMCoV predicts a broader peak which rises rapidly for the first few days post-infection and then continues to increase slowly until the T cell response begins between day 7 and 10 (*46*). Viral levels then decline gradually but continue to oscillate as some virus is produced and T cells discover pockets of hidden infection. The viral load dynamics predicted by SIMCoV are consistent with observations that the viral load in the lung peaks later than in the upper respiratory tract and can persist for much longer (*47, 16, 17*). SIMCoV simulations suggest that variable spatial seeding of virus in multiple FOI is a key factor accounting for the highly heterogeneous viral loads observed in patients.

ABMs, such as SIMCoV, are appealing because they can represent and visualize the complex heterogeneous movements of individual cells and diffusing fields of small particles (i.e. virions or cytokines) (*48*). Our prior results showed how key factors such as movement of infection, molecules, and cells affect search efficiency, timing of immune response initiation, and successful clearance of infection, and that interactions that depend on movement and physical cell-cell contact require a spatially-explicit modeling framework (*49, 25*).

Some earlier SARS-CoV-2 ODE TCL models included an implicit immune response incorporated in the cell clearance term and an explicit adaptive immune response, e.g. (*8, 14, 50, 16, 15*), similar to the approach we took with SIMCoV. These models were tuned to fit the same

lung viral load data from (*33*) that we fit here; however, because SIMCoV explicitly represents cell and virus movement, it mimics the highly variable dynamics by varying only a few key parameters. Specifically, varying T cell arrival time by only a few days (from five to 10 dpi, consistent with (*46*)) and the magnitude of the T cell and other immune responses by two-fold accounts for the variability and oscillations in viral clearance of the patients from Wolfel et al. (Fig. 5). This is consistent with studies that show $CD8^+$ T cells in particular are critical for clearance and control of SARS-CoV-2 infection (*22, 19, 51, 18, 52*) and that T cell killing of infected cells requires direct cell to cell contact (*28, 30, 53*). Because SIMCoV represents spatial positioning and T cell movement explicitly, their physical co-location with infected cells arises naturally in the model. This modeling strategy provides a mechanistic explanation, which shows how even a small change in the number of responding T cells and the timing of their entry into the infected lung can affect viral clearance. Although we do not explicitly model other important aspects of immune response to SARS-CoV-2, for example, Type I interferons, other innate responses, and antibody responses, SIMCoV accounts for these effects in the viral clearance term. Future studies using SIMCoV could explicitly represent these additional immune parameters.

In contrast with T cells, peak viral load depends on a much larger range of variation in FOI, from one to 220. In our analysis, the number of FOI within the lung is the key driver of peak viral load, and accounting for it requires a spatially explicit modelling framework. Spatial dispersion of virus may be particularly important for SARS-CoV-2 infection and other lung infections because the epithelial surface area of the lung is so large. In contrast to the smaller surface areas of the nasopharynx, where TCL can be reached very quickly regardless of how much initial virus is inhaled, viral spread through the larger area of the lung takes more time (*17*). In SIMCoV, virus diffuses locally from each infected location, and viral spread is constrained by the number of nearby cells that can become infected. In contrast to ODE models,

distant cells are out of reach of the infection. Thus, each additional FOI causes a linear increase in viral load. We find that the number of initial FOI in the lung has greater impact on viral dynamics than the actual number of virions (Fig. 4). Because SIMCoV can be seeded with multiple FOI, it can reproduce the patchy spread of tissue damage seen in CT scans (Fig. 6). High variability in how much virus individual patients inhale and how dispersed that virus is in the lung may be an important contributor to the large variation in viral load and disease outcome in demographically similar patients.

Spatial simulations in SIMCoV also allowed us to test how the branching topology of the lung might impact viral spread. Our results show that the geometry of lung tissue has a substantial effect on the rate of viral spread (Fig. 8). Surprisingly, of the models we considered, the 2D model is closer to the more realistic branching model than the non-branching 3D model. The non-branching 3D model and well-mixed ODE models assume that virus can contact many more cells at each point in time, greatly overestimating the rate of viral spread, while the 2D model slightly underestimates spread compared to a topological branching model of the lung.

Previous spatially explicit models were limited both by data on infection dynamics and the computational power to explicitly represent spatial interactions among billions of cells (*48*). SIMCoV overcomes these limitations by leveraging the data shared in response to the COVID-19 pandemic and an underlying architecture that takes advantage of HPC capabilities. The SIMCoV platform could be used to simulate other spatial interactions such as predator prey dynamics between immune and infected cells or collective action dynamics (*11*), like the collective search strategies of T cells (*54*).

To enable open and reproducible science, SIMCoV is freely available under an open-source license, and it was designed to be easily extensible. Future extensions of the model could investigate the effects of mucus and the complex dynamics of airflow in the respiratory tract (*55*) with more detailed topological models of the airways and alveoli and the fraction and distribu-

tion of epithelial cells with receptors necessary for infection. SIMCoV could also incorporate additional aspects of the innate and adaptive immune response. Even without these features and with limited parameter tuning, SIMCoV reflects the patchiness of tissue damage in patient CT scans (Fig. 6) and predicts the time course of patient viral load (Fig. 5). In future work, we will initialize SIMCoV simulations with locations of tissue damage from CT scans to predict viral load, and potentially disease severity, in individual patients. By predicting the time course of viral loads within individuals, SIMCoV can help to identify factors that determine windows of transmission between individuals and thereby improve understanding of epidemic spread.

Materials and Methods

Model Description

SIMCoV[1] is an Agent Based Model (ABM) where space is represented by a discrete Cartesian grid, which can be either two or three-dimensional. For clarity, we describe the model with reference to a 2D layer of epithelial cells, although the mechanisms apply equally to other topologies such as 3D simulations that include the branching structure of lung airways.

The grid points are spaced five microns apart (roughly the diameter of a T cell), and components of the model only exist at these discrete locations. The model is run as a discrete-time simulation, where each time step represents one minute, approximately the time it takes for a T cell to move five microns (one grid point) (*31, 56*). At each time step, the model components are updated according to their states and interactions.

There are four main components of SIMCoV: epithelial cells, $CD8^+$ T cells, virions, and inflammatory signals, representing the subset of cytokines that cause T cell extravasation into the lung tissue. The interactions between these components are shown in Figure 1. Each ep-

[1]The source code for SIMCoV is publicly available at `https://github.com/AdaptiveComputationLab/simcov`.

ithelial cell has a fixed grid location with no more than one cell per location. However, some locations can be empty, representing airways (as can be seen in Figure 7). In addition, a single T cell may also occupy a grid point, and every grid point can have a concentration of virions and inflammatory signals. The number of virions is represented explicitly and is capped at 125,000 virions, which is the maximum number estimated to fit within five μm^3. The concentration of inflammatory signal is represented as a floating point number between 1×10^{-6} and 1.

Each simulation run begins with a number of sites of infection (parameter **b** in Table 2), which are grid points with an initial level of virions (parameter **c**). At each time step, the virions will diffuse uniformly at a given rate (parameter **j**) to neighboring grid points, both those with and without epithelial cells, and some fraction (parameter **i**) of the virions will be cleared, reducing the concentration.

Initially, epithelial cells are *healthy* (not infected), but all are able to be infected. At the end of a time step, if there are virions in the same location as a healthy epithelial cell, that cell will become infected (transition **1** in Figure 1) with a probability proportional to the virion concentration, multiplied by the *infectivity* (parameter **g**). An infected epithelial cell is initially in an *incubating* state, in which it remains for a period of time defined by the mean of a Poisson distribution (parameter **d**)[2], after which it transitions to an *expressing* state (transition **2**). An epithelial cell that is *expressing* will produce virions at a constant rate per time step (parameter **h**, transition **3a**), and those will be added to the local concentration and diffuse outwards (transition **4a**) and be cleared (transition **4b**). In the absence of a T cell response, after a further time period (parameter **e**), an *expressing* cell will transition to the *dead* state (transition **3c**).

When epithelial cells are *expressing* (and *apoptotic* - more on that later), they will produce inflammatory signals at a constant rate (parameter **k**) per time step (transition **3b**). Those signals will diffuse to neighboring grid points and decay at constant rates (parameters **m** and **l**, respec-

[2]Parameters that are time periods are all Poisson distributions specified by the mean.

Parameter	Description	Default	Reference
a) Dimensions	Simulation size in x,y, and z dimensions	15000x15000x1	*
b) Initial Infections	Number of grid points with initial infections	1	*
c) Initial Virions	Number of virions at initial infection locations	1000	*
d) Incubation Period	Average minutes until an infected cell starts expressing virions	480	(*57*)
e) Expressing Period	Average minutes after expressing starts until cell death	900	(*10*)
f) Apoptosis Period	Average minutes after apoptosis is induced until cell death	180	(*58*)
g) Infectivity	Probability of one virion infecting one cell per minute	0.001	(*25*)
h) Virion Production	Number of virions produced by expressing cell per minute	1.1	(*59*)
i) Virion Clearance	Fraction by which virion count drops per minute	0.004	(*25*)
j) Virion Diffusion	Fraction of virions that diffuse into all neighboring grid points per minute	0.15	(*25*)
k) Inflammatory Signal Production	Concentration of inflammatory signal produced by expressing cells per minute	1	*
l) Inflammatory Signal Decay	Fraction by which inflammatory signal concentration drops per minute	0.01	*
m) Inflammatory Signal Diffusion	Fraction of inflammatory signal that diffuses into all neighboring grid points per minute	1	(*25*)
n) T Cell Production	Number of T cells generated per minute into circulation	105000	(*25, 60*)
o) T Cell Initial Delay	Average minutes before T cells start to be produced	10080	(*61, 62, 46*)
p) T Cell Vascular Period	Average minutes before death for a T cell in the vasculature	5760	(*25*)
q) T Cell Tissue Period	Average minutes before death for a T cell after it extravasates	1440	(*63*)
r) T Cell Binding Period	Average minutes T cell is bound to an epithelial cell	10	(*30*)

Table 2: Table of SIMCoV parameters. The default values represent the COVID-19 parameters we collected from a variety of sources. Time periods are Poisson distributions, defined by the average. One time step is one minute, and grid points are five microns apart. Unless stated otherwise, the default parameters were used for all experiments. Parameters marked with (*) were not taken from the literature—see Parameter Derivation.

tively). The concentration of inflammatory signal is represented as a floating point number with a minimum value (1×10^{-6}), so that, eventually, it will be completely eliminated if there is no more being added by infected cells. In the current implementation of SIMCoV, the inflammatory signal serves only as a trigger for T cell extravasation, but it can easily be extended to represent other chemokines and cytokines in future work.

It is assumed that after an initial delay (parameter **o**) activated antigen-specific $CD8^+$ T cells start to enter the blood from lymph nodes (transition **7**) at a constant rate (parameter **n**). T cells in the vasculature extravasate into lung tissue when they detect inflammation (transition **9b**). T cell movement through the vasculature is not explicitly modeled; instead, a random location is chosen for each T cell at every time step and if that location contains an inflammatory signal, then the T cell extravasates. The random location is chosen from among all of the possible locations for a full-sized lung, so if only a fraction of the lung is being modeled, then the probability of extravasation is reduced accordingly. Only one T cell can occupy a location in the tissue at a time, so a T cell will not extravasate to a location already occupied by another T cell. T cells that do not extravasate continue to circulate in the vasculature (transition **9a**), and those that fail to extravasate before the vascular T cell lifespan (parameter **p**) will die and be removed from the simulation (transition **9c**).

A T cell that is in the tissue can potentially bind to an epithelial cell that is at the same or a neighboring location (transition **10a**). The probability of binding is proportional to the period the cell has been incubating, with a maximum of 1 if the cell is expressing. Multiple T cells can bind to a single infected cell, providing they are neighbors or colocated with the infected cell. When a T cell binds to an epithelial cell, the latter transitions to the *apoptotic* state (transition **3d**) and after a given time period (parameter **f**), it dies (transition **6b**). Apoptotic epithelial cells continue to produce virions (transition **6a**) and inflammatory signals at the same rates as expressing cells. T cells remain bound for a fixed time period (parameter **r**), after which

they move around in the tissue, at every time step moving to a randomly selected neighboring location that does not already contain a T cell. Hence, T cells move at maximum rate of five microns per minute (*31*). If after moving, a T cell finds itself a neighbor of or colocated with an infected cell, it once again can bind. This process of random movement and possible binding continues until the T cell dies (transition **10b**) after a certain time period (parameter **q**). T cells in the tissue never return to the vasculature.

Parameter Derivation

Where possible, parameter settings were derived from a close investigation of available data, as described here:

a. *Dimensions*: in general, set to a sufficiently large size so that a simulation will not induce target cell limitation in the first week of infection; for the default 2D simulation, the z-dimension will be 1.

b. *Initial Infections*: set to one to represent the minimum number of infected locations in the lower respiratory tract.

c. *Initial Virions*: set according to the *infectivity* parameter (**g**) in order to successfully infect at least one cell initially.

d. *Incubation Period*: roughly the eclipse period; estimated to be 7-8 hours or 480 time steps in (*57*).

e. *Expressing Period*: estimated to be about 15 hours or 900 minutes in (*10*), which is similar to the expressing period of the influenza virus (about 1000 minutes (*25*)).

f. *Apoptosis Period*: estimated to range between 48 minutes and 168 minutes in mice (*58*); we estimate a slightly longer period, 180 minutes, for humans.

g. *Infectivity*: set to 0.001, which is the probability of infection per cell per minute, i.e. it will take on average 1,000 minutes for a virion to infect a cell. This is within the range of 12 to 1,200 minutes estimated for influenza (*25*). This is a largely unknown value that will vary between diseases and disease strains.

h. *Virion Production*: calculated by dividing the burst size by the expressing period (**e**); the burst size is the total virions produced by an infected cell, which was found to be 1,000 virions (*59*).

i. *Virion Clearance*: set to 0.004 virions per minute, which is within prior calculations of the virion decay rate of the influenza virus, ranging from 7e-7 to 0.07 virions per minute (*25*). We vary this parameter from 0.003 to 0.005 to fit patient data.

j. *Virion Diffusion*: this is calculated from the Einstein-Smoluchowski equation for particle movement, $D \approx \alpha.x^2/(2t)$, where x is the mean distance diffused in one direction along one axis in time t, and α is the SIMCoV parameter. Hence, for a mean distance of five microns (one grid point), a time of one minute (one simulation time step), and a virion diffusion parameter of 0.15, the diffusion coefficient is $0.0312\mu m^2/s$. This diffusion coefficient is similar to the diffusion coefficient calculated for influenza (*25*), which is $0.0318\mu m^2/s$.

k. *Inflammatory Signal Production*: the concentration of inflammatory signal produced per minute is not explicitly expressed in literature; therefore, we provide an estimate and modify it to best fit published viral data.

l. *Inflammatory Signal Decay*: there are no explicit measurements of this value in the literature; we estimate a half life of 70 minutes, which is similar to the half life of 30 minutes estimated as plausible for influenza (*25*).

m. *Inflammatory Signal Diffusion*: calculated from the Einstein-Smoluchowski equation for particle movement, $D \approx \alpha.x^2/(2t)$, where x is the mean distance diffused in one direction along one axis in time t, and α is the SIMCoV parameter. Hence, for a mean distance of five microns (one grid point), a time of one minute, and a inflammatory signal diffusion parameter of 1, the diffusion coefficient is $0.210\mu m^2/s$, which is within the range of the chemokine diffusion coefficient calculated in (*25*) (from $0.00318\mu m^2/s$ to $318\mu m^2/s$).

n. *T Cell Production*: the number of T cells produced by replication in the mediastinal lymph nodes of the lung is calculated by scaling up the T cell production rate in Levin et al. (*25*), which was estimated from mouse models (*60*). That previous calculation assumed that a single virus-specific T cell encounters an antigen presenting cell and begins to replicate on day 0 post infection, doubling every 8 hours. After 5 days, those activated CD8+ T cells could generate approximately 30,000 new T cells per day (or 21 per minute) into circulation. In this current SIMCoV model, we assume a linear 5,000-fold increase from mouse to human, which reflects the approximate increase in lung and blood volume and gives a rate of 105,000 T cells per minute entering circulation.

o. *T Cell Initial Delay*: estimated from (*46*) for SARS-CoV-2. Our default time is 7 dpi, and we vary from 5 to 10 dpi to fit patient data. We note that if T cells double every 8 hours and replication in humans lasts 2.3 days longer than in mice (7 versus 5 dpi), this would cause a 128-fold increase in T cell production. With 40 times greater SARS-CoV-2 specific T cell precursors, this would generate 5,000 times greater T cell production in humans. Alternatively, exponential replication of a single T cell precursor for 9 days would generate 5,000 times greater production in humans.

p. *T Cell Vascular Period*: equivalent to the T cell age in the blood for a model of T cell responses to influenza (*25*).

q. *T Cell Tissue Period*: largely unknown, but estimates were provided in Keating et al. (*63*).

r. *T Cell Binding*: taken from Halle et al. (*30*).

Model Implementation

The SIMCoV model is implemented in UPC++ (*64*), which is a high performance language for parallel computation that runs on a variety of platforms, including distributed memory supercomputers and networked clusters. Consequently, the SIMCoV application can run on systems from laptops up to high performance supercomputers without requiring any code changes. However, because of the computational cost of the simulation, only small areas can be simulated on smaller systems with one or a few compute nodes. Simulations of significant fractions of the lung requires hundreds or thousands of compute nodes.

Massively Parallel Simulation

SIMCoV is designed by the composition of highly parallelizable functions to make a highly scalable simulation framework. Running immense agent-based simulations of this type would not be computationally feasible without the application of high performance computing techniques (HPC). For SIMCoV, we rely on the HPC primitives and multi-process communication techniques provided by UPC++, a library extension to the C++ programming language with state-of-the-art HPC capabilities (*64*). We reduce the communication and computational work load of running large SIMCoV simulations with different approaches. To mitigate expensive cross process calculation, we maintain three-dimensional or two-dimensional blocks of closely spaced grid points in contiguous single process memory. This lowers the chance that for any given update of agents in the simulation that two processes have to share memory. To prevent unnecessary computation costs, we maintain an *active list* of agents and grid points that will be updated at each iteration of the simulation. Some grid points can be predicted to remain

constant from one iteration to the next, so there is no need to update their behavior for that step.

SIMCoV can be used to run large simulations in minutes on suitable supercomputers. For example, for the CT-scan comparisons (Section 1), we simulated multiple initial FOI over a full lung slice, comprising over three billion epithelial cells. On the NERSC Cori supercomputer, it took 15 minutes to complete a two week simulation on 32 nodes.

Acknowledgements

We thank Ruian Ke for sharing insights and an early draft of his ODE model; Matthew Fricke, UNM CARC, and ASU RC for HPC resources and support; and Greg Forest for many helpful discussions. This project was funded by NSF 2030037 and 2029696 which includes Coronavirus Aid, Relief, and Economic Security (CARES) Act funding. DARPA provided partial funding through AFRL FA-8650-18-C-6898 (JLC, MM, HT and AA).This work is also partly supported by the Advanced Scientific Computing Research (ASCR) program within the Office of Science of the DOE under contract number DE-AC02-05CH11231 (SH); Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Energy (SH). We used resources of the NERSC supported by the Office of Science of the DOE under Contract No. DEAC02-05CH11231. **Author Contributions:** M.E.M., S.F., J.L.C. and S.H. conceived the study. S.H. designed and coded the model, S.H., A.A., R.G., M.H., K.L., A.P. V.S. and H.T. ran the models, assembled data, performed calculations, and analyzed output. M.E.M, S.F., J.L.C. and S.H. wrote the manuscript with contributions from all other authors.

References

1. C. R. Moderbacher, *et al.*, *Cell* **183**, 996 (2020).

2. P. S. Arunachalam, *et al.*, *Science* **369**, 1210 (2020).

3. J. Yang, *et al.*, *International Journal of Infectious Diseases* **94**, 91 (2020).

4. J. Fajnzylber, *et al.*, *Nature Communications* **11**, 1 (2020).

5. M. Marks, *et al.*, *The Lancet Infectious Diseases* (2021).

6. N. Néant, *et al.*, *Proceedings of the National Academy of Sciences* **118** (2021).

7. S. Wang, *et al.*, *Mathematical biosciences* **328**, 108438 (2020).

8. E. A. Hernandez-Vargas, J. X. Velasco-Hernandez, *Annual reviews in control* (2020).

9. A. Gonçalves, *et al.*, *CPT: pharmacometrics & systems pharmacology* **9**, 509 (2020).

10. K. S. Kim, *et al.*, Modelling SARS-CoV-2 dynamics: Implications for therapy (2020).

11. C. J. E. Metcalf, B. T. Grenfell, A. L. Graham, *PLoS pathogens* **16**, e1009105 (2020).

12. R. Ke, C. Zitzmann, R. M. Ribeiro, A. S. Perelson, *medRxiv* (2020).

13. K. Ejima, *et al.*, *Epidemics* **35**, 100454 (2021).

14. A. Goyal, E. F. Cardozo-Ojeda, J. T. Schiffer, *Science advances* **6**, eabc7112 (2020).

15. K. S. Kim, *et al.*, *PLoS biology* **19**, e3001128 (2021).

16. A. S. Perelson, R. Ke, *Clinical Pharmacology & Therapeutics* (2020).

17. S. Q. Du, W. Yuan, *Journal of medical virology* **92**, 1615 (2020).

18. X. Zhang, *et al.*, *Nature* **583**, 437 (2020).

19. A. Sette, S. Crotty, *Cell* (2021).

20. A. T. Tan, *et al.*, *Cell reports* **34**, 108728 (2021).

21. A. Grifoni, *et al.*, *Cell* **181**, 1489 (2020).

22. A. Bertoletti, A. T. Tan, N. Le Bert, *Oxford Open Immunology* **2** (2021).

23. C. Warrender, S. Forrest, F. Koster, *Bulletin of mathematical biology* **68**, 2233 (2006).

24. H. Mitchell, *et al.*, *Journal of virology* **85**, 1125 (2011).

25. D. Levin, *et al.*, *Journal of Theoretical Biology* **398** (2016).

26. M. Imai, *et al.*, *Proceedings of the National Academy of Sciences* **117**, 16587 (2020).

27. K. Zhang, *et al.*, *Cell* **181**, 1423 (2020).

28. T. H. Harris, *et al.*, *Nature* **486**, 545 (2012).

29. S. Ariotti, *et al.*, *The Journal of Immunology* **195**, 5285 (2015).

30. S. Halle, *et al.*, *Immunity* **44**, 233 (2016).

31. P. Mrass, *et al.*, *Nature communications* **8**, 1 (2017).

32. X. Zhang, *et al.*, *Nature* (2020).

33. R. Wölfel, *et al.*, *Nature* (2020).

34. J. Regan, *et al.*, *Open Forum Infectious Diseases* (2021).

35. B. M. Wiebe, H. Laursen, *Apmis* **106**, 651 (1998).

36. M. Gandhi, C. Beyrer, E. Goosby, *Journal of general internal medicine* **35**, 3063 (2020).

37. S. A. Harmon, *et al.*, *Nature Communications* **11**, 1 (2020).

38. Z. Cheng, *et al.*, *Radiology of Infectious Diseases* pp. 1–7 (2020).

39. M. Roberts, *et al.*, *Nature Machine Intelligence* **3**, 199 (2021).

40. N. Varble, *et al.*, *European Radiology* (2020).

41. Q. Cai, *et al.*, *Allergy: European Journal of Allergy and Clinical Immunology* **75**, 1742 (2020).

42. R.-H. Du, *et al.*, *Annals of the American Thoracic Society* **17** (2020).

43. C. Faes, *et al.*, *International Journal of Environmental Research and Public Health* **17** (2020).

44. J. D. Crapo, B. E. Barry, P. Gehr, M. Bachofen, E. R. Weibel, *American Review of Respiratory Disease* **126**, 332 (1982).

45. H.-C. Yeh, G. Schum, *Bulletin of Mathematical Biology* **42**, 461 (1980).

46. S. Kalimuddin, *et al.*, *Med* (2021).

47. M. Cevik, *et al.*, *The Lancet Microbe* (2020).

48. A. L. Bauer, C. A. Beauchemin, A. S. Perelson, *Information sciences* **179**, 1379 (2009).

49. G. M. Fricke, K. A. Letendre, M. E. Moses, J. L. Cannon, *PLoS Computational Biology* **12**, e1004818 (2016).

50. M. P. Xavier, R. F. Reis, R. W. dos Santos, M. Lobosco, *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (IEEE, 2020), pp. 1311–1317.

51. L. Ni, *et al.*, *Frontiers in immunology* **12** (2021).

52. C. Lucas, *et al.*, *Nature* **584**, 463 (2020).

53. H. D. Hickman, *et al.*, *Immunity* **42**, 524 (2015).

54. M. E. Moses, J. L. Cannon, D. M. Gordon, S. Forrest, *Frontiers in immunology* **10**, 1357 (2019).

55. R. Levy, D. B. Hill, M. G. Forest, J. B. Grotberg, Pulmonary fluid flow challenges for experimental and mathematical modeling (2014).

56. M. R. Looney, *et al.*, *Nature Methods* **8**, 91 (2011).

57. M. Schneider, *et al.*, *Journal of Virology* **86**, 10112 (2012).

58. V. V. Ganusov, R. J. De Boer, *Journal of Virology* **82**, 11749 (2008).

59. Y. M. Bar-On, A. Flamholz, R. Phillips, R. Milo **9**, e57309 (2020).

60. H. Miao, *et al.*, *J Virol* **84**, 6687 (2010).

61. K. G.-I. Mohn, *et al.*, *The Journal of Infectious Diseases* **221**, 1528 (2020).

62. T. Powell, *et al.*, *PloS one* **8**, e62778 (2013).

63. R. Keating, *et al.*, *Immunology & Cell Biology* **96**, 1104 (2018).

64. Y. Zheng, A. Kamil, M. B. Driscoll, H. Shan, K. Yelick pp. 1105–1114 (2014).