Immunology and Inflammation

Self-inhibiting percolation and viral spreading in epithelial tissue

Reviewed Preprint
v2 • May 22, 2024
Revised by authors

Reviewed Preprint
v1 • January 24, 2024

Xiaochan Xu, Bjarke Frost Nielsen, Kim Sneppen

Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark • Novo Nordisk Foundation Center for Stem Cell Medicine, reNEW, University of Copenhagen, 2200 Copenhagen Ø, Denmark • PandemiX Center, Department of Science and Environment, Roskilde University, 4000 Roskilde, Denmark • High Meadows Environmental Institute, Princeton University, Princeton, New Jersey 08544, USA

Abstract

SARS-CoV-2 induces delayed type-I/III interferon production, allowing it to escape the early innate immune response. The delay has been attributed to a deficiency in the ability of cells to sense viral replication upon infection, which in turn hampers activation of the antiviral state in bystander cells. Here, we introduce a cellular automaton model to investigate the spatiotemporal spreading of viral infection as a function of virus and host-dependent parameters. The model suggests that the considerable person-to-person heterogeneity in SARS-CoV-2 infections is a consequence of high sensitivity to slight variations in biological parameters near a critical threshold. It further suggests that within-host viral proliferation can be curtailed by the presence of remarkably few cells that are primed for IFN production. Thus the observed heterogeneity in defense readiness of cells reflects a remarkably cost-efficient strategy for protection.

eLife assessment

This study presents a cellular automaton model to study the dynamics of virus-induced signalling and innate host defense against viruses such as SARS-CoV-2 in epithelial tissue. The simulations and data analysis are **convincing** and represent a **valuable** contribution that would be of interest to researchers studying the dynamics of viral propagation.

https://doi.org/10.7554/eLife.94056.2.sa3

Introduction

Adaptive immune responses are relatively slow since they require pathogen-specific priming of immune cells (***Sette and Crotty, 2021***). For example, the time required for the body to activate adaptive immunity against the SARS-CoV-2 virus upon initial infection is around 10 days, comparable to the delay of immunization against SARS-CoV-2 after vaccination (***Polack et al., 2020***). Instead, the earliest infection dynamics are largely governed locally, by infected cells and their neighbor-hood. The innate responses including both interferon (IFN) mediated intercellular

communication and expression of antiviral genes (ISGs) are determinants for confining the viral spread in the respiratory tract. Here, we address the spread of viruses within epithelial tissue, using SARS-CoV-2 as a model pathogen. The overall considerations are similar for other viruses, but the parameters governing infection may vary considerably due to the specific countermeasures of the virus in question, affecting its ability to bypass human antiviral defenses.

In terms of countermeasures, insufficient type I and III interferon secretion upon infection is a main immune signature feature of SARS-CoV-2 infection (***Blanco-Melo et al., 2020***; ***Hatton et al., 2021***; ***Stanifer et al., 2020***; ***Minkoff and tenOever, 2023***). The failure to activate immediate antiviral responses with IFNs is also a pathogenic aspect of other viruses including Ebola (***Mohamadzadeh et al., 2007***), Marburg (***He et al., 2019***) and Herpes simplex (***Barreca and O'Hare, 2004***). Secretion of IFN relies on the cell's ability to sense viral products during its replication. Despite the presence of sensors for DNA and RNA viruses in cells, many species of viruses partially evade detection. The SARS-CoV-2 virus is such a case: Only two of 16 putative RNA virus sensors, IFIH1 (MDA5) and DHX58 (LGP2) from the RIG-I-like receptor (RLR) family, play roles in inducing IFN upon SARS-CoV-2 infection (***Yin et al., 2021***) and IFIH1 is antagonized by SARS-CoV-2 (***Liu et al., 2021***).

Intriguingly, evidence shows that pre-activated innate immune states help combat the SARS-CoV-2 infection. The higher basal expression of viral sensors, IFIH1 and DDX58 (also from the RLR family), in the upper airway of children (relative to adults), reduces the severity of COVID (***Loske et al., 2022***). Furthermore, well-differentiated primary nasal epithelial cells derived from a donor with pre-activated IFNγ show resistance to SARS-CoV-2 infection (***Broadbent et al., 2022***). Thus, the extent to which innate immunity contributes to the observed heterogeneity in responses to SARS-CoV-2 between hosts (***Schaller et al., 2021***; ***Desai et al., 2020***) is a compelling subject for investigation.

To address this question, we reanalyze single-cell RNAseq data (***Fiege et al., 2021***; ***Ravindra et al., 2021***) providing gene expression profiles of virus sensors and antiviral genes in host cells during early SARS-CoV-2 infection. We propose a cellular automaton model based on a few transition rules suggested by observed cell states, to explain the heterogeneity in early disease progression as a consequence of criticality in the virus-host interaction system.

Cell states during infection

Directly observing cell responses and cell state transitions in a patient's body upon viral infection is virtually impossible. Human bronchial epithelial cells (HBECs) mimic the airway epithelium and have been used as a representative model for investigating the consequences of the viral invasion (***de Jong et al., 1993***, ***1994***; ***Davis et al., 2015***). Single-cell RNAseq provides snapshots of the states of individual cells indicated by high-dimensional gene expression profiles at the mRNA level and can uncover the heterogeneity of cell responses obscured by aggregate measurement. Thus, by combining HBECs as a model and single-cell RNAseq data, one can in principle infer cell state transitions following viral infection. More importantly, single-cell RNAseq also captures copies of viral genes during sequencing, which allows us to estimate viral replication inside cells simultaneously.

To reconstruct the trajectory of cell state transitions during early SARS-CoV-2 infection, we re-analyze single-cell RNAseq data from experiments where HBECs are sampled from different conditions: Mock (corresponding to state before infection, 0 h), 24 and 48 hours post-viral infection (hpi) (***Fiege et al., 2021***). We focus on genes associated with antiviral responses, interferon genes from the host cells, and detected viral genes. We project high-dimensional gene expression data onto a 2D plane using Uniform Manifold Approximation and Projection (UMAP) and obtain a low-dimensional visualization of single-cell expression patterns (**Fig. 1a**). As a dimension

reduction algorithm, UMAP is a manifold learning technique that favors the preservation of local distances over global distances (***McInnes et al., 2018***; ***Becht et al., 2019***). It constructs a weighted graph from the data points and optimizes the graph layout in the low-dimensional space. On the UMAP plane (**Fig. 1a**), each dot represents a cell sample and the distance between dots correlates with the level of similarity of cellular states. The cells are not divided absolutely into discrete clusters and rather show continuous trajectories. We cluster the cells with the principal components analysis (PCA) results from their gene expression. With the first 16 principal components, we calculate k-nearest neighbors and construct the shared nearest neighbor graph of the cells then optimize the modularity function to determine clusters. We present the cluster information on the UMAP plane and use the same UMAP coordinates for all the plots in this paper hereafter.

Different clusters on the UMAP indicate distinct cellular states during the progression of infection. For instance, there are three sub-clusters of susceptible cells (O_1, O_2, O_3). Neither viral genes nor IFNs are detected in these cells and only a few antiviral genes are expressed. The viral sensors (DHX58, DDX58, and IFIH1) are at their lowest level (**Fig. 1b**, Fig. S1). We refer to all of these cells as *O* cells due to their relatively similar gene expression profiles in terms of viral replication genes. The proportion of *O* cells decreases over time as the infection spreads (**Fig. 1c**).

We also observe three infected cell clusters where viral genes are primarily detected, V^i, V^r, and V. With the increasing counts of viral genes, we infer that the V^i cluster is the earliest state after an *O* cell has been infected and the virus begins replicating. Some but not all antiviral genes are activated in the V^i cells (IFIT1/2/3 and OAS1/2/3) (**Fig. 1b** and Fig. S1), indicating that these cells are still vulnerable to viral invasion. This cluster is followed by two subsequent clusters, the V^r cluster with pronounced viral replication and *A* cluster with barely any viral replication.

In the *V* cluster, the viral genes reach their highest level, and antiviral genes are strongly inhibited, indicating that the virus has fully hijacked the cell. The antiviral genes are expressed most strongly in the *A* cluster and partially in the *N* cluster, indicating that the antiviral capability of the *N* cluster is weaker than the full antiviral state. Although the *N* cluster also shows a high level of viral genes, it severely lacks one of the viral genes (cov.E, Fig. S2) compared with the most highly expressed viral genes of the *V* cluster. This observation implies that viral replication and activation of the antiviral state coexist in the IFN-secreting cells (*N* cluster). We note the existence of a small subgroup of the V^r cluster, close to the *A* cluster, that exhibits relatively high levels of both antiviral genes and viral genes but no appreciable IFN (**Fig. 1d–f**). As in the *N* cluster, the viral gene E is barely detected in these cells, indicating incomplete viral replication. However, in contrast to the *N* cluster, the antiviral genes are expressed to their full extent (Fig. S1 and S2). Thus, these cells are more likely to sustain the antiviral state.

At 24 hpi, some cells have switched from the pre-infection state (*O*) to other states. At 48 hpi, almost all cells have transitioned to other states and only a few cells remain in the *O* state (**Fig. 1c** and **g**). The aggregated gene expression of representative antiviral genes and detected viral genes indicates the cells move from the *O* state towards the three remaining terminal states on the considered timescale of 2 days: Antiviral state (*A*, **Fig. 1d**), Virus-conquered state (*V*, **Fig. 1e**), and IFN producing state (*N*, **Fig. 1f**). Central for the overall defense is the relatively few cells that reach the IFN-producing state (*N*). These cells also express *A* and *V* genes.

When IFN is not expressed, the antiviral genes and viral genes exclude each other (**Fig. 1d** and **e**), except for a few cells around (UMAP1, UMAP2) $\sim(-2.5, 7.5)$ (green cells at hpi = 48, **Fig. 1g**). They represent cells where the virus succeeded in stopping IFN secretion, but could not fully hijack the cell. We still regard these cells as antiviral cells in our model.

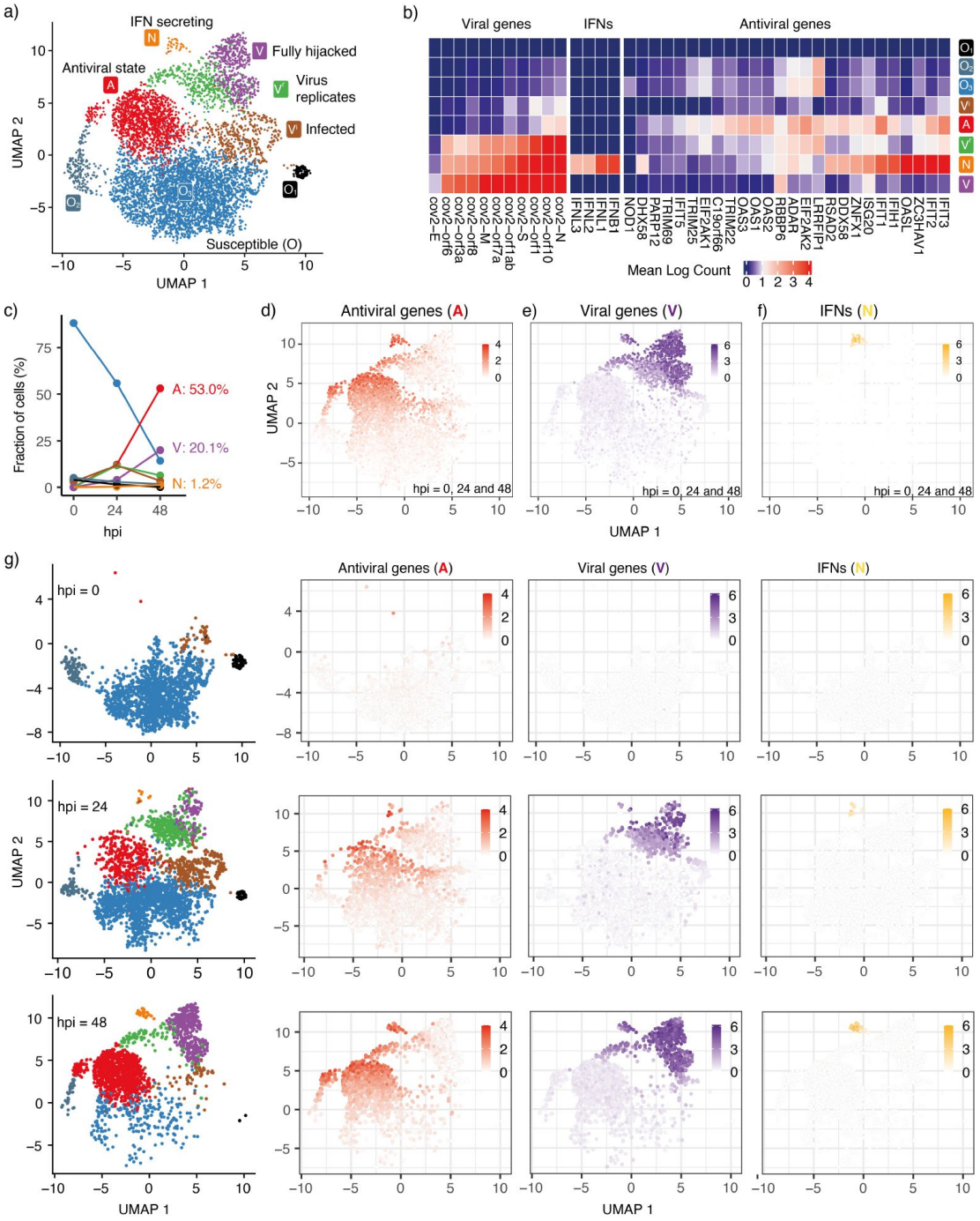


Figure 1.

Cell states during SARS-CoV-2 infection in human tracheal/bronchial epithelial cells. **a)** 6162 cells (***Fiege et al., 2021***) covering samples of mock-infected (0 h), 24 hpi (hours post-infection), and 48 hpi visualized with UMAP. **b)** Average expression of representative viral genes, IFNs, and antiviral genes in cells within each cluster (state). **c)** Cell proportions of clusters at different time points (hpi = 0, 24, and 48). The same colors are used for the lines as for the cluster (state) in **a). d)** Average expression of antiviral genes (IFIT1, IFIT2, IFIT3, IFIT5, IFIH1, OAS1, OAS2, OAS3, OASL, DDX58) in each cell. **e)** Average expression of viral genes (cov.orf1ab, cov.S, cov.orf3a, cov.orf6, cov.M, cov.N) in each cell. **f)** Average expression of interferon genes (IFNB1, IFNL1, IFNL2, IFNL3) in each cell. 103 cells (1.7%) are IFN-positive in each cell. **g)** Progression of viral infection as indicated by changes in cell proportions of different states. Cells are shown separately at each time point in the leftmost column. The right columns show the average expression of antiviral genes, viral genes, and IFNs in each of these cells. Colorkeys indicate the gene expression level from low (white) to high (red, purple, or yellow).

The *N* state is associated with both high levels of virus sensors and viral genes, in agreement with the observation that IFN production is initiated after exposure to the virus (***Lei et al., 2020***) and that IFN can induce an antiviral state inside the same cell (***Sanceau et al., 1987***). Expression of the key SARS-CoV-2 sensitive sensors (IFIH1, DDX58, DHX58) is sparse in the *O* state (Fig. S1), indicating that a small fraction of cells have virus-sensing capacity prior to infection and are ready to mount a defense. This cell population increases with IFN tissue diffusion.

Model

We introduce a cellular automaton model to capture the cell state dynamics during the early stages of SARS-CoV-2 infection in a sheet of epithelial tissue. At each simulation, we seed an infection site on a 2D square lattice and study how the infection spreads as the sites on the lattice switch between cell states following a set of simple rules derived from the observations of the single-cell RNAseq data.

In addition to the states corresponding to the dominant clusters observed in the single-cell data (**Fig. 1a**) (*O, A, V* and *N* states corresponding to *O, A, V* and *N* clusters), we introduce a transient pre-antiviral state (*a*) that can switch to the *N* state rapidly upon viral exposure, considering the heterogeneity of viral sensing ability in susceptible cells.

It follows from this description, that those RNA viruses that can be sensed by a larger repertoire of sensors should be modeled with a larger fraction of cells in the *a* state.

The model is initialized with cells predominantly in the *O* state and a small fraction, p_a, in the pre-antiviral state *a*.

Alternatively, one could formulate an equivalent model in which the initial state consisted entirely of *O* cells (and an infection seed), and the parameter p_a would instead be understood as the probability for an *O* cell to switch to the *N* or *A* state when exposed to the virus or IFNs, respectively. This would be functionally equivalent to our model, and as such, the value of p_a must depend on both host and virus. In particular, a virus that can effectively interfere with the defense and signaling of host cells will be modeled by a low p_a value.

It is worth noting that the proportion of cells in the *a* state before the onset of SARS-CoV-2 infection is expected to be higher in hosts with pre-activated antiviral innate immunity (***Loske et al., 2022***; ***Broadbent et al., 2022***), meaning that the value of p_a will, in general, depend on the exposure history of the host.

The cell state transitions triggered by IFN signaling or viral replication are known in viral infection, but how exactly the transitions are orchestrated for specific infections is poorly understood. The UMAP cell state distribution hints at possible preferred transitions between states. The closer two cell states are on the UMAP, the more likely transitions between them are, all else being equal. For instance, the antiviral state (*A*) is easily established from a susceptible cell (*O*), but not from the fully virus-hijacked cell (*V*). The IFN-secreting cell state (*N*) requires the co-presence of the viral and antiviral genes and thus the cell cluster is located between the antiviral state (*A*) and virus-infected state (*V*) but distant from the susceptible cells (*O*).

Inspired by the UMAP data visualization (**Fig. 1a**), we propose the following transitions between five main discrete cell states (**Fig. 2a**):

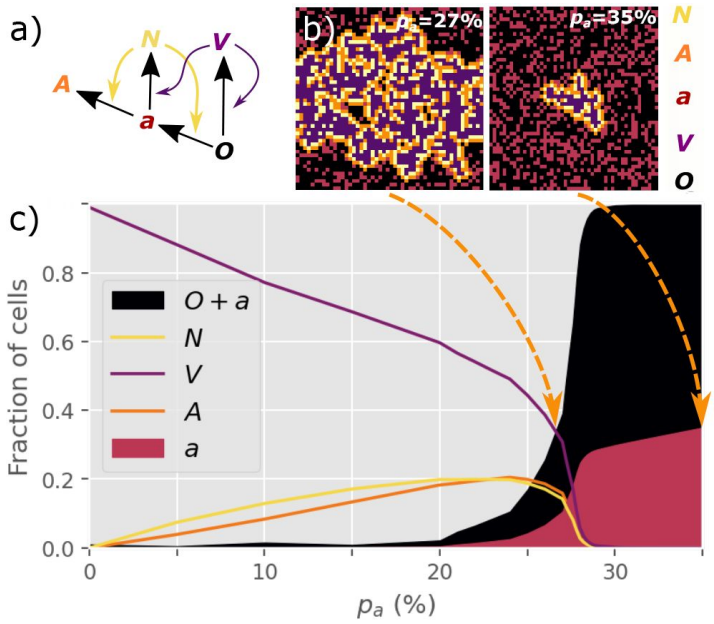


Figure 2.

NOVAa model. **a)** The cell state transitions are included in the NOVAa model. The straight black arrows indicate transitions between cell states. The curved yellow arrows indicate the effects of IFNs on activating antiviral states. The curved purple arrows indicate viral spread to cells with *O* and *a* states. The simpler OVA model bypasses the *a* and *N* states, by allowing *V* to induce direct $O \rightarrow A$ transitions (Fig. S3). **b)** Final states of a small lattice (50 × 50) simulations at two different values of p_a (both at IFN spreading radius R = 1). **c)** The fraction of cells in each state in the final frozen configuration as a function of p_a. A critical transition is observed at $p_a = p_c \sim 27.8\%$. At lower values of p_a, most cells terminate in the *V* state, representing an aggressive tissue infection. Simulations were performed on a lattice with linear dimension $L = 1000$.

- *N*, IFN-secreting cells. These arise from pre-antiviral cells (*a* state) that become infected (but not infectious). Here, we assume that the secretion of IFNs by the *N* cells is a faster process than possible apoptosis (***Wen et al., 1997***; ***Tesfaigzi, 2006***) of these cells and that the diffusion of IFNs to the neighborhood is not significantly affected by apoptosis.
- *O*, unaffected (susceptible) cells.
- *V*, infected and virus-producing cells. This state arises when a susceptible (*O*) cell is exposed to a virus from another *V* cell.
- *a*, pre-antiviral state. It develops into either the *A* or *N* state upon exposure to signals from *N* cells or virus from *V* cells.
- *A*, antiviral state immune to infection. It is achieved when a pre-antiviral (*a*) cell is exposed to IFN. We do not consider the decay of the antiviral state as it may last more than 72 h (***Gaajetaan et al., 2013***).

The dynamics are defined in terms of discrete time steps, representing the characteristic timescales of cellular viral infection. We explore the model for an extended time, keeping in mind that in reality other immune cells such as natural killer (NK) cells and macrophages may migrate to the infected site and reduce viral spread (***McNab et al., 2015***).

The four rules of the model are (**Fig. 2a**):

$$N(a) = A, \quad N(O) = a, \quad V(a) = N, \quad V(O) = V$$

where the notation $X(Y) = Z$ denotes a cell in state X acting on a cell in state Y and changing it to state Z in one time-step. Thus, cells in states *0*, *a*, and *A* are unable to influence their neighbors. The *V* state is the only directly self-replicating state.

Each site of the $L \times L$ lattice is assigned to either the *O* (probability: $1 - p_a$) or the *a* state (probability: p_a). Infection is initiated by a single *V* cell, and we explore the percolation of the infection to larger scales. A time step consists of L^2 updates, in which a random site *i* is selected. If a *V* cell is selected, it interacts with its 4 nearest neighbors according to the rules $V(O) = V$ and $V(a) = N$. If an *N* cell is selected, it interacts with all cells within a radius R, according to the rules $N(O) = a$ and $N(a) = A$. The radius R thus quantifies the diffusion range of IFNs relative to the virus. We explore a version of the dynamics where interactions only happen with a reduced probability $p_{conv} < 1$ rather than being deterministically applied to all neighbours (Fig. S5). We find that this does not affect the critical behavior of the model. Periodic boundary conditions are imposed in the model throughout.

Results

At R = 1, the final number of infected cells depends strongly on the value of p_a. At a low p_a of 0.27, infections typically spread to the entire system, while at a higher p_a of 0.35, the propagation of the *V* state is inhibited (**Fig. 2b**).

We observe a threshold-like behavior of the final attack rate of the virus when the initial p_a changes continuously (**Fig 2c**). The virus spreads macroscopically for $p_a < p_c \approx 27.8\%$. At higher p_a, cells are sufficiently prone to convert to the antiviral state to prevent the infection from percolating.

The size distribution $P(s)$ of infection clusters (defined as the number of *N* and *V* cells in a cluster) around the critical value of $p_a = 27.8\%$ obeys power-law decay (**Fig. 3a**). In **Fig. 3b**, the distribution $P(s) \propto 1/s^{\tau}$ is further explored by re-scaling and the cluster size exponent is confirmed

as τ = 1.83 ± 0.03 when p_a = 28%. Notably, this exponent is below the equilibrium 2D percolation yielding τ = 2.05 (***Stauffer and Aharony, 2018***). Further, our exponent $\tau \sim 1.8$ is above the percolation-inspired cluster growth model for virus spread of ref. (***Gönci et al., 2010***) which has an exponent between τ = 1.58 and τ = 1.64 depending on the distribution of individual cells' predefined ability to become infected. Meanwhile, the propagation for the different states could be accelerated by the smaller value of p_a with the same R (Fig. S4).

The actual critical value of p_a depends strongly on the choice of neighborhood. In particular, at R = 1, the *V* and *N* states have the same range in the tissue (a proxy for diffusivity), while a more realistic scenario is to allow IFNs to diffuse faster in the tissue (R > 1), facilitating the initiation of the antiviral state. The critical percolation threshold p_c decreases almost exponentially with the value of R (**Fig. 4b**), and viral propagation can be stopped for p_a as low as 0.4% when R ≥ 5. Such a small fraction of initial cells in the *a* state is consistent with the remarkably few *N* cells observed in experiments (**Fig. 1b**). Thus, a higher diffusivity of IFN provides a more than proportional decrease in the required number of antiviral cells. As revealed by the reanalysis of RNAseq data in **Fig. 1**, the fraction of IFN-positive cells is relatively low – around 1.7%. Comparing with simulations near the critical point, we find that, at R = 5, the ratio of *N* cells to all affected cells (*N* + *A* + *V*) in the final state, $\lim_{t\to\infty} N/(N+A+V) \approx 2\%$, i.e. it is of comparable magnitude to the experimental value. This holds in a wide range around the critical point, $p_a \sim p_c$.

The exponents for the cluster size distribution are the same at *R* = 1 and *R* = 5, while the structures of the clusters are different (**Fig. 4a** and **c**). Greater R leads to a different microscopic structure with fewer *A* and *N* cells in the final state (**Fig. 4c**).

To put the above findings in perspective we further explore a simplified version of our model with only 3 states (supplementary Fig. S3), the OVA model, which may be seen as a rephrasing of models for induced antiviral states suggested by ***Howat et al***. (***2006***); ***Segredo-Otero and Sanjuán*** (***2020***); ***Michael Lavigne et al***. (***2021***). In the OVA model, p_a is the probability that an infected cell produces interferons to warn neighbor cells within radius ***R***. In the OVA model, one update consists of selecting a random cell. If the cell is in the *V* state then its neighbor cells may change by exposure to the virus, provided that they are susceptible (*O*). Each of the four neighbors is now chosen in random order, and if a neighbor cell *i* is in the *O* state, a random number $ran_i \in [0, 1]$ is drawn. If $ran_i \geq P_a$ the neighbor is flipped to the *V* state. If, on the other hand, $ran_i < p_a$, all *O* cells within a radius R around the neighbor *i* are converted to the *A* state. Thus, for large ***R*** and moderate p_a, the spread of infection will be mitigated. We find that the OVA model has an "outbreak size" exponent $\tau \sim 1.8$, similar to the NOVAa model. However, the change in microstructure as a function of the IFN range ***R*** observed in the NOVAa model (compare **Fig. 4** panel a and c) is not observed in the OVA model (Fig. S3), where the features instead scale proportionally with ***R***. We also simulated standard percolation by randomly adding disks of radius R of blocking ("antiviral") cells and checking for percolation of the infected state. While the critical behavior of the standard percolation model approximately resembles that of the OVA model (**Fig. 4b**), the antiviral state of the OVA model is somewhat less effective at blocking the spread (reflected in a higher threshold p_c).

Finally, we examined a version of the model where the discrete idealization of *N* cells acting at all cells within a specific radius ***R*** is replaced by a probabilistic conversion with a diffusion-like profile. The algorithm for this is described in the supplement, with results in Fig. S6 to be compared to **Fig. 4**. We find that the probabilistic spreading of IFN is more effective, in terms of demanding lower ***R*** for obtaining similar limits on spreading of the infection. This is likely due to the existence of long-range interactions (however rare), when neighbours are selected according to a Gaussian profile.

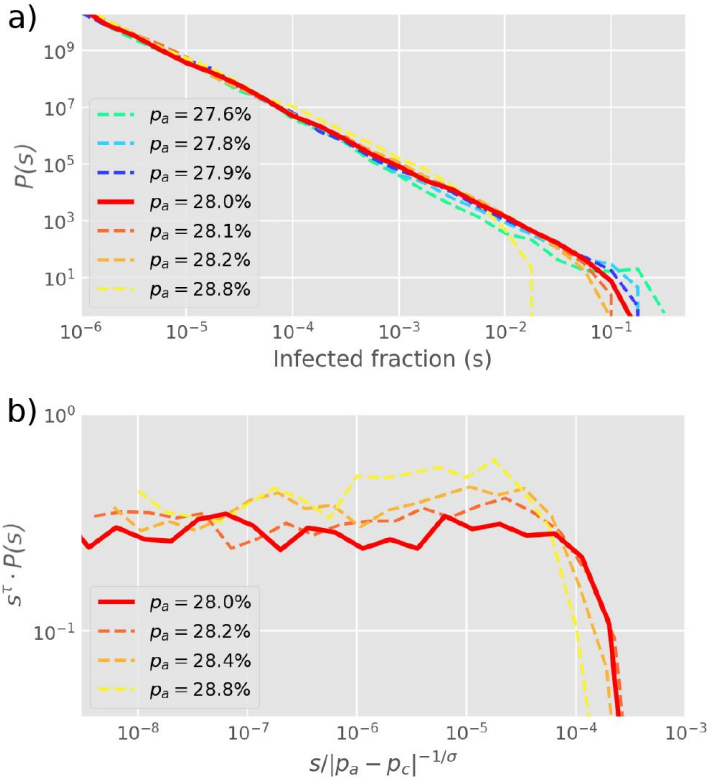


Figure 3.

Cluster size distribution. **a)** The distribution $P(s)$ of cluster sizes of infected cells ($s = (N + V)L^2$) for different values of p_a, simulated by starting with one infected cell in a 2D square lattice of linear extent $L = 2000$. **b)** The exponents from a) are extracted by re-scaling $P(s)$ as shown on the y-axis, yielding $\tau = 1.83$. The cut-off exponent is estimated as $\sigma \sim 1$. Simulations plot the final outbreak sizes from 10,000 initial infections of one cell. The histogram is log-binned with 5 bins per decade. The critical point at $p_a = p_c = 0.28$ is determined as the value with the longest scaling regime.

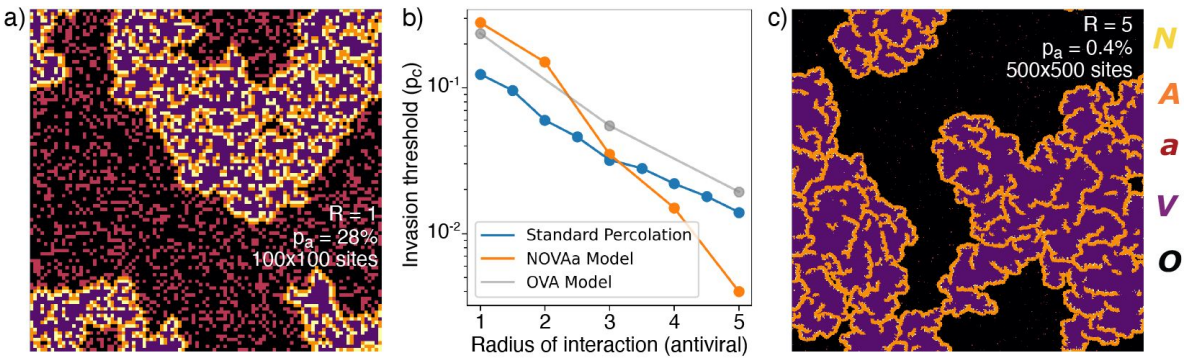


Figure 4.

Range of IFN. **a)** Typical cluster for an R = 1 simulation at $p_a \sim p_c = 0.278$. **b)** The dependence of p_c with R, approximately reproduced by a fit $p_c \sim 3^{-R}$. For comparison, the OVA model as well as percolation has $p_c \sim 3^{-R/2}$. In all cases, when p_a is above p_c then the virus is prevented from spreading. **c)** Cluster distribution for R = 5 at $p_a \sim p_c = 0.004$, at a 5 times larger linear scale than a).

Discussion

There are some preexisting models of the interplay between virus, host cells and triggered immune responses, with an antiviral state triggered by IFN signaling from neighbor cells (***Graw and Perelson, 2016***). Cellular automaton models of infection dynamics in epithelial tissue were explored by (***Howat et al., 2006***; ***Segredo-Otero and Sanjuán, 2020***; ***Michael Lavigne et al., 2021***), with the overall result that spreading depends on competition between the virus and an induced antiviral cell state. This competition is recapitulated in our model in terms of the two effective parameters p_a and $\boldsymbol{R}$. Our model emphasizes the threshold dynamics, with a critical transition between effective confinement and unhindered spread that depends sensitively on the details of the relevant cell states. In particular, the presence of the specialized IFN-producing N cells allows for disease confinement at a much lower concentration of pre-antiviral cells (lower value of p_a) in the NOVAa model, than in the OVA model which lacks the N state (**Fig. 4b**). As a consequence of low p_a, the number of final N-state cells is also much lower.

The low concentration of ready-to-fight cells may seem perplexing, leading one to surmise that the organism could easily fight off an infection by only slightly increasing its investment in these primed cells. However, one should keep in mind that e.g. the human organism does indeed have ready-to-fight cells that are able to eliminate most foreign RNA, and only leave a few truly infectious viruses. As highlighted in the introduction, these select viruses often employ strategies to lower the p_a, for example by only being sensed by a small fraction of the RNA virus-sensitive receptors of our cells.

The parameter p_a can be interpreted as the probability that a cell is sufficiently antiviral to convert to a N state upon infection with a given virus. The relevant value of p_a will depend on the virus considered (and will be small for viruses that inhibit cell responses to infection) as well as on the host (e.g. on age (***Kissler et al., 2020***) and recent infection history). Dysregulated IFN responses are characteristic of the effective immunomodulatory strategies used by betacoronaviruses (***Channappanavar et al., 2019***; ***Acharya et al., 2020***).

The parameter $\boldsymbol{R}$ reflects the signaling efficiency of an interferon-producing cell. Since $\boldsymbol{R}$ is measured in units of the typical distance that the virus spreads, it depends on viral properties, including its burst size, diffusion, and adsorption to host cells, with higher adsorption being associated with larger $\boldsymbol{R}$ values. For SARS-CoV-2 this suggests that lower ACE2 receptor counts would result in less adsorption to nearby cells, in turn allowing the virus to spread to more distant tissues (***Bastolla, 2021***) suggesting a lower value of $\boldsymbol{R}$.

For viruses that do not delay the production of IFNs, p_a would be higher than for SARS-CoV-2, allowing neighbor cells around an infected site to form a kind of “ring vaccination” as the antiviral state dominates. In this sense, our model is consistent with the previous modeling of the roles of autocrine and paracrine interferon signaling suppression of viral infection (see e.g. (***Michael Lavigne et al., 2021***) for parallels between IFN response and “ring vaccination”).

We do not consider viral particles which enter the bloodstream and seed new infections non-locally. This may allow the virus to spread in the tissue at what would otherwise constitute subcritical conditions in our model. Further, there may be tissue-specific variations in both p_a and $\boldsymbol{R}$, adding larger-scale heterogeneity to the overall spreading. As the disease progresses one would expect additional heterogeneity to emerge, associated with variability in later host responses including macrophage activation and adaptive immunity (***Wang et al., 2021***).

The remarkable heterogeneity of disease progression in COVID-19, in the form of widely variable symptoms (**Tabata et al., 2020**) and transmission risk (**Nielsen et al., 2021**; **Kirkegaard and Sneppen, 2021**), has been widely observed. For instance, among university students, just 2% of SARS-CoV-2 positive hosts provided 90% of total respiratory viral load (**Yang et al., 2021**). In our formalism, we would understand such variability in terms of a p_a that is comparable to the critical value, but varying between hosts. A slight change of p_a then results in dramatic fluctuations in the outcome of an infection.

To be more quantitative, for SARS-CoV-2 the detected virus count on average grows by a factor of 3.5 (**Kissler et al., 2020**) in one infection generation of 8 hours (not to be confused with the between-host "generation time" of the infection). This within-host reproductive number is far below the number of viruses produced from a cell, indicating severe restrictions from the innate immune system. On the other hand, 3.5 is still above the threshold for spreading, indicating that within-host amplification is super-critical. However, the measured amplification includes viruses that "jump" to other spots in an infected person, thereby suggesting a local spreading that is closer to the critical value than an amplification of 3.5 would suggest.

Our study finally compared the NOVAa model with the simpler OVA scenario that recapitulates earlier modeling of induced antiviral states (**Howat et al., 2006**; **Segredo-Otero and Sanjuán, 2020**; **Michael Lavigne et al., 2021**). These papers all build on a more homogeneous role of infected cells, each inducing some immunization of surrounding cells. They emphasize the larger range of IFN signals compared to viral diffusion (**Howat et al., 2006**), focus on the competition between viruses with different abilities to suppress IFN signaling (**Segredo-Otero and Sanjuán, 2020**), or introduce a cellular automaton approach where the antiviral state leads to a type of ring vaccination that prevents the virus from spreading when more IFN is produced (**Michael Lavigne et al., 2021**). Our OVA model may be seen as a simplified and more stochastic version of the last model. The NOVAa model then adds the additional benefits associated with the experimentally observed but low-abundance *N* state cells, which by their rarity adds to predicted randomness between the fate of individual infection centers during an early viral infection.

Acknowledgements

BFN acknowledges financial support from the Carlsberg Foundation in the form of an Internationalisation Fellowship (grant no. CF23-0173) and through the Carlsberg Foundation's Semper Ardens programme (grant no. CF20-0046), as well as from the Danish National Research Foundation (grant no. DNRF170). Novo Nordisk Foundation Center for Stem Cell Medicine is supported by Novo Nordisk Foundation grant NNF21CC0073729.

References

Acharya D., Liu G., Gack M. U. (2020) **Dysregulation of type i interferon responses in covid-19** *Nature Reviews Immunology* **20**:397–398

Barreca C., O'Hare P. (2004) **Suppression of herpes simplex virus 1 in mdbk cells via the interferon pathway** *Journal of virology* **78**:8641–8653

Bastolla U. (2021) **Mathematical model of sars-cov-2 propagation versus ace2 fits covid-19 lethality across age and sex and predicts that of sars** *Frontiers in Molecular Biosciences* **8**

Becht E., McInnes L., Healy J., Dutertre C.-A., Kwok I. W., Ng L. G., Ginhoux F., Newell E. W. (2019) **Dimensionality reduction for visualizing single-cell data using umap** *Nature biotechnology* **37**:38–44

Blanco-Melo D. *et al.* (2020) **Imbalanced host response to sars-cov-2 drives development of covid-19** *Cell* **181**:1036–1045

Broadbent L., Bamford C. G., Lopez Campos G., Manzoor S., Courtney D., Ali A., Touzelet O., McCaughey C., Mills K., Power U. F. (2022) **An endogenously activated antiviral state restricts sars-cov-2 infection in differentiated primary airway epithelial cells** *Plos one* **17**

Channappanavar R. *et al.* (2019) **Ifn-i response timing relative to virus replication determines mers coronavirus infection outcomes** *The Journal of clinical investigation* **129**:3625–3639

Davis A. S., Chertow D. S., Moyer J. E., Suzich J., Sandouk A., Dorward D. W., Logun C., Shelhamer J. H., Taubenberger J. K. (2015) **Validation of normal human bronchial epithelial cells as a model for influenza a infections in human distal trachea** *Journal of Histochemistry & Cytochemistry* **63**:312–328

de Jong P. M., van Sterkenburg M. A., Hesseling S. C., Kempenaar J. A., Mulder A. A., Mommaas A. M., Dijkman J. H., Ponec M. (1994) **Ciliogenesis in human bronchial epithelial cells cultured at the air-liquid interface** *American journal of respiratory cell and molecular biology* **10**:271–277

de Jong P. M., van Sterkenburg M. A., Kempenaar J. A., Dijkman J. H., Ponec M. (1993) **Serial culturing of human bronchial epithelial cells derived from biopsies** *In Vitro Cellular & Developmental Biology-Animal* **29**:379–387

Desai N. *et al.* (2020) **Temporal and spatial heterogeneity of host response to sars-cov-2 pulmonary infection** *Nature communications* **11**

Fiege J. K. *et al.* (2021) **Single cell resolution of sars-cov-2 tropism, antiviral responses, and susceptibility to therapies in primary human airway epithelium** *PLoS pathogens* **17**

Gaajetaan G. R., Geelen T. H., Vernooy J. H., Dentener M. A., Reynaert N. L., Rohde G. G., Beuken E. V., Grauls G. E., Bruggeman C. A., Stassen F. R. (2013) **Interferon-β induces a long-lasting antiviral state in human respiratory epithelial cells** *Journal of Infection* **66**:163–169

Gönci B., Németh V., Balogh E., Szabó B., Dénes Á., Környei Z., Vicsek T. (2010) **Viral epidemics in a cell culture: novel high resolution data and their interpretation by a percolation theory based model** *PloS one* **5**

Graw F., Perelson A. S. (2016) **Modeling viral spread** *Annual review of virology* **3**:555–572

Hatton C. F. *et al.* (2021) **Delayed induction of type i and iii interferons mediates nasal epithelial cell permissiveness to sars-cov-2** *Nature Communications* **12**

He F. B., Melén K., Kakkola L., Julkunen I. (2019) **Interaction of ebola virus with the innate immune system** *Emerging Challenges in Filovirus Infections*

Howat T. J., Barreca C., O'Hare P., Gog J. R., Grenfell B. T. (2006) **Modelling dynamics of the type i interferon response to in vitro viral infection** *Journal of The Royal Society Interface* **3**:699–709

Kirkegaard J. B., Sneppen K. (2021) **Superspreading quantified from bursty epidemic trajectories** *Scientific reports* **11**:1–7

Kissler S. M. *et al.* (2020) **Sars-cov-2 viral dynamics in acute infections** *MedRxiv* :2020–10

Lei X. *et al.* (2020) **Activation and evasion of type i interferon responses by sars-cov-2** *Nature communications* **11**

Liu G. *et al.* (2021) **Isg15-dependent activation of the sensor mda5 is antagonized by the sars-cov-2 papain-like protease to evade host innate immunity** *Nature microbiology* **6**:467–478

Loske J. *et al.* (2022) **Pre-activated antiviral innate immunity in the upper airways controls early sars-cov-2 infection in children** *Nature biotechnology* **40**:319–324

McInnes L., Healy J., Melville J. (2018) **Umap: Uniform manifold approximation and projection for dimension reduction** *arXiv preprint*

McNab F., Mayer-Barber K., Sher A., Wack A., O'garra A. (2015) **Type i interferons in infectious disease** *Nature Reviews Immunology* **15**:87–103

Michael Lavigne G., Russell H., Sherry B., Ke R. (2021) **Autocrine and paracrine interferon signalling as 'ring vaccination'and 'contact tracing'strategies to suppress virus infection in a host** *Proceedings of the Royal Society B* **288**

Minkoff J. M., tenOever B. (2023) **Innate immune evasion strategies of sars-cov-2** *Nature Reviews Microbiology* :1–17

Mohamadzadeh M., Chen L., Schmaljohn A. L. (2007) **How ebola and marburg viruses battle the immune system** *Nature Reviews Immunology* **7**:556–567

Nielsen B. F., Simonsen L., Sneppen K. (2021) **Covid-19 superspreading suggests mitigation by social network modulation** *Physical Review Letters* **126**

Polack F. P. *et al.* (2020) **Safety and efficacy of the bnt162b2 mrna covid-19 vaccine** *New England journal of medicine* **383**:2603–2615

Ravindra N. G. *et al.* (2021) **Single-cell longitudinal analysis of sars-cov-2 infection in human airway epithelium identifies target cells, alterations in gene expression, and cell state changes** *PLoS biology* **19**

Sanceau J., Sondermeyer P., Beranger F., Falcoff R., Vaquero C. (1987) **Intracellular human gammainterferon triggers an antiviral state in transformed murine l cells** *Proceedings of the National Academy of Sciences* **84**:2906–2910

Schaller M. A. *et al.* (2021) **Ex vivo sars-cov-2 infection of human lung reveals heterogeneous host defense and therapeutic responses** *JCI insight* **6**

Segredo-Otero E., Sanjuán R. (2020) **The role of spatial structure in the evolution of viral innate immunity evasion: A diffusion-reaction cellular automaton model** *PLoS Computational Biology* **16**

Sette A., Crotty S. (2021) **Adaptive immunity to sars-cov-2 and covid-19** *Cell* **184**:861–880

Stanifer M. L., Kee C., Cortese M., Zumaran C. M., Triana S., Mukenhirn M., Kraeusslich H.-G., Alexandrov T., Bartenschlager R., Boulant S. (2020) **Critical role of type iii interferon in controlling sars-cov-2 infection in human intestinal epithelial cells** *Cell reports* **32**

Stauffer D., Aharony A. (2018) **Introduction to percolation theory**

Tabata S. *et al.* (2020) **Clinical characteristics of covid-19 in 104 people with sars-cov-2 infection on the diamond princess cruise ship: a retrospective analysis** *The Lancet Infectious Diseases* **20**:1043–1050

Tesfaigzi Y. (2006) **Roles of apoptosis in airway epithelia** *American journal of respiratory cell and molecular biology* **34**:537–547

Wang S., Hao M., Pan Z., Lei J., Zou X. (2021) **Data-driven multi-scale mathematical modeling of sars-cov-2 infection reveals heterogeneity among covid-19 patients** *PLoS Computational Biology* **17**

Wen L.-P., Madani K., Fahrni J. A., Duncan S. R., Rosen G. D. (1997) **Dexamethasone inhibits lung epithelial cell apoptosis induced by ifn-Y and fas** *American Journal of Physiology-Lung Cellular and Molecular Physiology* **273**:L921–L929

Yang Q. *et al.* (2021) **Just 2% of sars-cov-2-positive individuals carry 90% of the virus circulating in communities** *Proceedings of the National Academy of Sciences* **118**

Yin X. *et al.* (2021) **Mda5 governs the innate immune response to sars-cov-2 in lung epithelial cells** *Cell reports* **34**

Editors

Reviewer #1 (Public Review):

Summary:

The manuscript describes a model based on 5-state cellular automata of development of an infection. The model is motivated and qualitatively justified by time-resolved measurements of expression levels of viral, interferon-producing, and antiviral genes. The model is set up in such a way that the crucial difference in outcomes (infection spreading vs. confinement) depends on the initial fraction of special virus-sensing cells. Those cells (denoted as 'type a') cannot be infected and do not support the propagation of infection, but rather inhibit it in a somewhat autocatalytic way. Presumably, such feedback makes the transition between two outcomes very sharp: a minor variation in concentration of 'a' cells results in qualitative change from one outcome to another. As in any percolation-like system, the transition between propagation and inhibition of infection goes through a critical state with all its attributes, including a power-law distribution of the cluster size (corresponding to the fraction of infected cells) with a fairly universal exponent and a cutoff at the upper limit of this distribution.

Strengths:

The proposed model suggests a well-justified explanation for the frequently observed yet puzzling diversity of outcomes of viral infections such as COVID.

Weaknesses:

None.

https://doi.org/10.7554/eLife.94056.2.sa2

Reviewer #2 (Public Review):

Xu et al. introduce a cellular automaton model to investigate the spatiotemporal spreading of viral infection. In this study, the author first analyzes the single-cell RNA sequencing data from experiments and identifies four clusters of cells at 48 hours post-viral infection, including susceptible cells (O), infected cells (V), IFN-secreting cells (N), and antiviral cells (A). Next, a cellular automaton model (NOVAa model) is introduced by assuming the existence of a transient pre-antiviral state (a). The model consists of an LxL lattice; each site represents one cell. The cells change their state following the rules depending on the interaction of neighboring cells. The model introduces a key parameter, p_a, representing the fraction of pre-antiviral state cells. Cell apoptosis is omitted in the model. Model simulations show a threshold-like behavior of the final attack rate of the virus when p_a changes continuously. There is a critical value p_c, so that when p_a < p_c, infections typically spread to the entire system, while at a higher p_a > p_c, the propagation of the infected state is inhibited. Moreover, the radius R that quantifies the diffusion range of N cells may affect the critical value p_c; a larger R yields a smaller value of the critical value p_c. The authors further examine the result with stochastic version dynamics, and the main findings are unchanged upon stochastic dynamics. The structure of clusters is different for different values of R; greater R leads to a different microscopic structure with fewer A and N cells in the final state. Compared with the single-cell RNA seq data, which implies a low fraction of IFN-positive cells of around 1.7%, the model simulation suggests R=5. The authors also explored a simplified version of the model, the OVA model, with only three states. The OVA model also has an outbreak size. The OVA model shows dynamics similar to the NOVAa model. However, the change in microstructure as a function of the IFN range R observed in the NOVAa model is not observed in the OVA model.

https://doi.org/10.7554/eLife.94056.2.sa1

Reviewer #3 (Public Review):

Summary:

This study considers how to model distinct host cell states that correspond to different stages of a viral infection: from naïve and susceptible cells to infected cells and a minority of important interferon-secreting cells that are the first line of defense against viral spread. The study first considers the distinct host cell states by analyzing previously published single-cell RNAseq data. Then an agent-based model on a square lattice is used to probe the dependence of the system on various parameters. Finally, a simplified version of the model is explored, and shown to have some similarity with the more complex model, yet lacks the dependence on the interferon range. By exploring these models one gains an intuitive understanding of the system, and the model may be used to generate hypotheses that could be tested experimentally, telling us "when to be surprised" if the biological system deviates from the model predictions.

Strengths:

- Clear presentation of the experimental findings and a clear logical progression from these experimental findings to the modeling.
- The modeling results are easy to understand, revealing interesting behavior and percolation-like features.
- The scaling results presented span several decades and are therefore compelling.
- The results presented suggest several interesting directions for theoretical follow-up work, as well as possible experiments to probe the system (e.g. by stimulating or blocking IFN secretion).

Weaknesses:

- The fixed time-step of the agent-based modeling may introduce biases. I would consider simulating the system with Gillespie dynamics where the reaction rates depend on the ambient system parameters.
- Single-cell RNAseq data requires careful handling or it may generate false leads. The strength of the RNAseq evidence presented is not clear.

Two places where the manuscript could be extended:

- Since the "range" of IFN is an important parameter, it makes sense to consider other lattice geometries other than the square lattice, which is somewhat pathological. Perhaps a hexagonal lattice would generalize better.
- Tissues are typically three-dimensional, not two-dimensional. (Epithelium is an exception). It would be interesting to see how the modeling translates to the three-dimensional case. Percolations transitions are known to be very sensitive to the dimensionality of the system.

Justification of claims and conclusions:

The claims and conclusions are well justified.

https://doi.org/10.7554/eLife.94056.2.sa0

Author response:

The following is the authors' response to the original reviews.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public Review):***
>
> *Summary:*
>
> *The manuscript "Self-inhibiting percolation and viral spreading in epithelial tissue" describes a model based on 5-state cellular automata of development of an infection. The model is motivated and qualitatively justified by time-resolved measurements of expression levels of viral, interferon-producing, and antiviral genes. The model is set up in such a way that the crucial difference in outcomes (infection spreading vs. confinement) depends on the initial fraction of special virus-sensing cells. Those cells (denoted as 'type a') cannot be infected and do not support the propagation of infection, but rather inhibit it in a somewhat autocatalytic way. Presumably, such feedback makes the transition between two outcomes very sharp: a minor variation in concentration of à' cells results in qualitative change from one outcome to another. As in any percolation-like system, the transition between propagation and inhibition of infection goes through a critical state with all its attributes. A power-law distribution of the cluster size (corresponding to the fraction of infected cells) with a fairly universal exponent and a cutoff at the upper limit of this distribution.*
>
> *Strengths:*
>
> *The proposed model suggests an explanation for the apparent diversity of outcomes of viral infections such as COVID.*

Author response: We thank the referee for the concise and accurate summary of our work.

> *Weaknesses:*
>
> *Those are not real points of weakness, though I think addressing them would substantially improve the manuscript.*

Author response: Below we will address these point by point.

> *The key point in the manuscript is the reduction of actual biochemical processes to the NOVAa rules. I think more could be said about it, be it referring to a set of well-known connections between expression states of cells and their reaction to infection or justifying it as an educated guess.*

Author response: We have now improved this part in the model section. We have added a few sentences explaining how the cell state transitions are motivated by the UMAP results:

“The cell state transitions triggered by IFN signaling or viral replication are known in viral infection, but how exactly the transitions are orchestrated for specific infections is poorly understood. The UMAP cell state distribution hints at possible preferred transitions between states. The closer two cell states are on the UMAP, the more likely transitions between them are, all else being equal. For instance, the antiviral state (A) is easily established from a susceptible cell (O), but not from the fully virus-hijacked cell (V). The IFN-secreting cell state (N) requires the co-presence of the viral and antiviral genes and thus the cell cluster is located between the antiviral state (A) and virus-infected state (V) but distant from the susceptible cells (O).

Inspired by the UMAP data visualization (Fig. 1a), we propose the following transitions between five main discrete cell states”

> *Another aspect where the manuscript could be improved would be to look a little beyond the strange and 'not-so-relevant for a biomedical audience' focus on the percolation critical state. While the presented calculation of the precise percolation threshold and the critical exponent confirm the numerical skills of the authors, the probability that an actual infected tissue is right at the threshold is negligible. So in addition to the critical properties, it would be interesting to learn about the system not exactly at the threshold: For example, how the speed of propagation of infection depends on subcritical p_a and what is the cluster size distribution for supercritical p_a.*

Author response: We agree that further exploring the model away from the critical threshold is worthwhile. While our main focus has been on explaining the large degree of heterogeneity in outcomes – readily explained as a consequence of the sharp threshold-like behavior – we now include plots of the time-evolution of the infection (as well as the remaining states) over time for subcritical values of *pa*. The plots can be found in Figure S4 of the supplement.

> ***Reviewer #2 (Public Review):***
>
> *Xu et al. introduce a cellular automaton model to investigate the spatiotemporal spreading of viral infection. In this study, the author first analyzes the single-cell RNA sequencing data from experiments and identifies four clusters of cells at 48 hours post-viral infection, including susceptible cells (O), infected cells (V), IFN-secreting cells (N), and antiviral cells (A). Next, a cellular automaton model (NOVAa model) is introduced by assuming the existence of a transient pre-antiviral state (a). The model consists of an LxL lattice; each site represents one cell. The cells change their state following the rules depending on the interaction of neighboring cells. The model introduces a key parameter, p_a, representing the fraction of pre-antiviral state cells. Cell apoptosis is omitted in the model. Model simulations show a threshold-like behavior of the final attack rate of the virus when p_a changes continuously. There is a critical value p_c, so that when p_a < p_c, infections typically spread to the entire system, while at a higher p_a > p_c, the propagation of the infected state is inhibited. Moreover, the radius R that quantifies the diffusion range of N cells may affect the critical value p_c; a larger R yields a smaller value of the critical value p_c. The structure of clusters is different for different values of R; greater R leads to a different microscopic structure with fewer A and N cells in the final state. Compared with the single-cell RNA seq data, which implies a low fraction of IFN-positive cells - around 1.7% - the model simulation suggests R=5. The authors also explored a simplified version of the model, the OVA model, with only three states. The OVA model also has an outbreak size. The OVA model shows dynamics similar to the NOVAa model. However, the change in microstructure as a function of the IFN range R observed in the NOVAa model is not observed in the OVA model.*

Author response: We thank the referee for the comprehensive summary of our work.

> *Data and model simulation mainly support the conclusions of this paper, but some weaknesses should be considered or clarified.*

Author response: Thank you - we will address these point by point below.

> *(1) In the automaton model, the authors introduce a parameter p_a, representing the fraction of pre-antiviral state cells. The authors wrote: The parameter p_a can also be understood as the probability that an O cell will switch to the N or A state when exposed to the virus of IFNs, respectively.' Nevertheless, biologically, the fraction of pre-antiviral state cells does not mean the same value as the probability that an O cell switches to the*

> *N or A state. Moreover, in the numerical scheme, the cell state changes according to the deterministic role N(O)=a and N(a)=A. Hence, the probability p_a did not apply to the model simulation. It may need to clarify the exact meaning of the parameter p_a.*

Author response: We acknowledge that this was an imprecise formulation, and have now changed it.

What we tried to convey with that comment was that, alternatively to having a certain fraction of cells be in the *a* state initially, one could instead have devised a model in which We should note that even the current model has a level of stochasticity, since we choose the cells to be updated with a constant probability rate - we choose N cells to update in each timestep, with replacement.

However, based on your suggestion, we simulated a version of the dynamics which included stochastic conversion, i.e. each action of a cell on a nearby cell happens only with a probability p_conv (and the original model is recovered as the p_conv=1 scenario). Of course, this slows down the dynamics (or effectively rescales time by a factor p_conv), but crucially we find that it does not appreciably affect the location of the threshold p_c. Below we include a parameter scan across p_a values for R=1 and p_conv=0.5, which shows that the threshold continues to appear at around p_a=27%. each O-state cell simply had a probability to *act* as an a-state cell upon exposure to the virus or to interferons, i.e. to switch to an N state (if exposed to virus) or to the A state (if exposed to interferons). In this simplified model, there would be no functional difference, since it would simply amount to whether each cell had a probability to be designated an a-cell initially (as in our model), or upon exposure. So our remark mainly served to explain that the role of the *p_a* parameter is simply to encode that a certain fraction of virus-naive cells behave this way (whether predetermined or not).

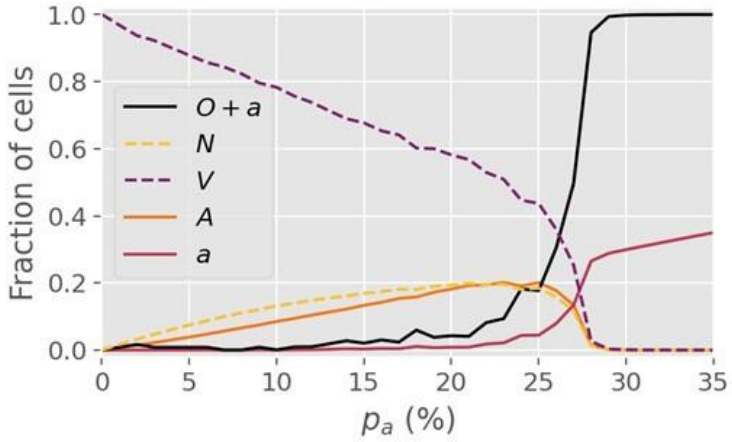
> *(2) The current model is deterministic. However, biologically, considering the probabilistic model may be more realistic. Are the results valid when the probability update strategy is considered? By the probability model, the cells change their state randomly to the state of the neighbor cells. The probability of cell state changes may be relevant for the threshold of p_a. It is interesting to know how the random response of cells may affect the main results and the critical value of p_a.*

Author response: This is a good point - we are firm believers in the importance of stochasticity. We should note that even the current model has a level of stochasticity, since we choose the cells to be updated with a constant probability rate - we choose N cells to update in each timestep, with replacement.

However, based on your suggestion, we simulated a version of the dynamics which included *stochastic conversion*, i.e. each action of a cell on a nearby cell happens only with a probability p_conv (and the original model is recovered as the p_conv=1 scenario). Of course, this slows down the dynamics (or effectively rescales time by a factor p_conv), but crucially we find that it does not appreciably affect the location of the threshold p_c. Below we include a parameter scan across p_a values for R=1 and p_conv=0.5, which shows that the threshold continues to appear at around p_a=27%.

We now discuss these findings in the supplement and include the figure below as Fig. S5.

Author response image 1.



> *(3) Figure 2 shows a critical value p_c = 27.8% following a simulation on a lattice with dimension L = 1000. However, it is unclear if dimension changes may affect the critical value.*

Author response: Re-running the simulations on a lattice 4x as large (i.e. L=2000) yields a similar critical value of 27-28% for *R=1*, so we are confident that finite size effects do not play a major role at L=1000 and beyond. For R=5, however, we find that a minimum lattice size greater than L=1000 is necessary to determine the critical threshold. Concretely, we find that the threshold value *pc* for R=5 changes somewhat when the lattice size is increased from 1000 to 2000, but is invariant under a change from 2000 to 3000, so we conclude that L=2000 is sufficient for *R=5*. The *pc* value for *R=5* cited in the manuscript (~0.4%) was determined from simulations at L=2000.

> ***Reviewer #3 (Public Review):***
>
> *Summary:*
>
> *This study considers how to model distinct host cell states that correspond to different stages of a viral infection: from naïve and susceptible cells to infected cells and a minority of important interferon-secreting cells that are the first line of defense against viral spread. The study first considers the distinct host cell states by analyzing previously published single-cell RNAseq data. Then an agent-based model on a square lattice is used to probe the dependence of the system on various parameters. Finally, a simplified version of the model is explored, and shown to have some similarity with the more complex model, yet lacks the dependence on the interferon range. By exploring these models one gains an intuitive understanding of the system, and the model may be used to generate hypotheses that could be tested experimentally, telling us "when to be surprised" if the biological system deviates from the model predictions.*

Author response: Thank you for the summary! We agree with the role that you describe for a model such as this one.

> *Strengths:*
>
> *- Clear presentation of the experimental findings and a clear logical progression from these experimental findings to the modeling.*

> *- The modeling results are easy to understand, revealing interesting behavior and percolation-like features.*
>
> *- The scaling results presented span several decades and are therefore compelling. - The results presented suggest several interesting directions for theoretical follow-up work, as well as possible experiments to probe the system (e.g. by stimulating or blocking IFN secretion).*
>
> *Weaknesses:*
>
> *- Since the "range" of IFN is an important parameter, it makes sense to consider lattice geometries other than the square lattice, which is somewhat pathological. Perhaps a hexagonal lattice would generalize better.*
>
> *- Tissues are typically three-dimensional, not two-dimensional. (Epithelium is an exception). It would be interesting to see how the modeling translates to the three-dimensional case. Percolation transitions are known to be very sensitive to the dimensionality of the system.*

Author response: We agree that probing different lattice geometries (2- and 3-dimensional alike) would be interesting and worthwhile. However, for this manuscript, we prefer to confine the analysis to the current, simple case. We do agree, however, that an extensive exploration of the role of geometry is an interesting future possibility.

> *- The fixed time-step of the agent-based modeling may introduce biases. I would consider simulating the system with Gillespie dynamics where the reaction rates depend on the ambient system parameters.*
>
> *- Single-cell RNAseq data typically involves data imputation due to the high sparsity of the measured gene expression. More information could be provided on this crucial data processing step since it may significantly alter the experimental findings.*
>
> *Justification of claims and conclusions:*
>
> *The claims and conclusions are well justified.*
>
> ***Recommendations for the authors:***
>
> ***Reviewer #1 (Recommendations For The Authors):***
>
> *It is necessary to explain what UMAP does. Is clustering done in the space of twenty-something original dimensions or 2D? How UMAP1 and UMAP2 are selected and are those the same in all plots?*

Author response: We have now added a few sentences to clarify the point raised above - the second snippet explains how clustering is performed:

“As a dimension reduction algorithm, UMAP is a manifold learning technique that favors the preservation of local distances over global distances (McInnes et al., 2018; Becht et al., 2019). It constructs a weighted graph from the data points and optimizes the graph layout in the low-dimensional space.”

“We cluster the cells with the principal components analysis (PCA) results from their gene expression. With the first 16 principal components, we calculate k-nearest neighbors and construct the shared nearest neighbor graph of the cells then optimize the modularity function to determine clusters. We present the cluster information on the UMAP plane and use the same UMAP coordinates for all the plots in this paper hereafter.”

> *Figure 1, what do bars in the upper right corners of panels d,e,f, and g indicate? Averaged' refers to time average? Something is missing in Cell proportions are labeled with corresponding colors in a)' .*

Author response: Thank you - we have now modified the figure caption. The bars in the upper right corners of panels d, e, f are color keys for gene expression, the brighter the color is, the higher the gene expression is.

“Averaged” gene expression refers to the mean expression of that particular gene across the cells within each indicated cluster.

The lines in c) correspond to cell proportions in different states at different time points. The same state in 1) and c) is shown in the same color.

> *Line 46, However' does not sound right in this context. Would Also' be better?*

Author response: We agree and have corrected it in the revised manuscript.

> *Line 96The viral genes are also partially expressed in these cells, but different from the N cluster, the antiviral genes are fully expressed (Fig. S1 and S2).' The sentence needs to be rephrased.*

Author response: We have rephrased the sentence: “As in the *N* cluster, the viral gene E is barely detected in these cells, indicating incomplete viral replication. However, in contrast to the *N* cluster, the antiviral genes are expressed to their full extent (Fig. S1 and S2).”

> *Line 126, missing "be", large' -> larger'.*

Author response: Thank you, we have now corrected these typos.

> *Line 139-140 The logical link between ignoring apoptosis and the diffusion of IFN is unclear.*

Author response: We modified the sentence as “Here, we assume that the secretion of IFNs by the N cells is a faster process than possible apoptosis (Wen et al., 1997; Tesfaigzi, 2006) of these cells and that the diffusion of IFNs to the neighborhood is not significantly affected by apoptosis.”

> *Fig. 2a Do the yellow arrows show the effect of IFN and the purple arrows the propagation of viral infection?*

Author response: That is correct. We have added this information to the figure caption: “The straight black arrows indicate transitions between cell states. The curved yellow arrows indicate the effects of IFNs on activating antiviral states. The curved purple arrows indicate viral spread to cells with O and a states.”

> *Fig. 3, n(s) as the axis label vs P(s) in the text? How do the curves in panel a) look when the p_a is well above or below p_c?*

Author response: Thank you. We have edited the labels in the figure to reflect the symbols used in the text.

> *Boundary conditions? From Fig. 4, apparently periodic?*

Author response: Yes, we use periodic boundary conditions in the model. We clarify it in the model section now (last sentence).

> *It will be good to see a plot with time dependences of all cell types for a couple of values of p_a, illustrating propagation and cessation of the infection.*

Author response: We agree, and have added a Figure S4 in the supplement which explores exactly that. Thank you for the suggestion.

A verbal qualitative description of why p_a has such importance and how the infection is terminated for large p_a would help.

> ***Reviewer #2 (Recommendations For The Authors):***
>
> *Below are two minor comments:*
>
> *(1) In the single-cell RNA sequencing data analysis, the authors describe the cell clusters O, V, A, and N. However, showing how the clusters are identified from the data might be more straightforward.*

Author response: Technically, we cluster the cells using principal components analysis (PCA) results of their gene expression. With the first 16 principal components, we calculate *k*-nearest neighbors and construct the shared nearest neighbor graph of the cells and then optimize the modularity function to determine clusters. We manually annotate the clusters with O, V, A, and N based on the detected abundance of viral genes, antiviral genes, and IFNs.

> *(2) In Figure 3, what does n(s) mean in Figure 3a? And what is the meaning of the distribution P(s) of infection clusters? It may be stated clearly.*

Author response: The use of *n(s)* was inconsistent, and we have now edited the figure to instead say *P(s)*, to harmonize it with the text. *P(s)* is the distribution of cluster sizes, *s*, expressed as a fraction of the whole system. In other words, once a cluster has reached its final size, we record *s=(N+V)/L^2* where *N* and *V* are the number of N and V state cells in the cluster (note that, by design, each simulation leads to a single cluster, since we seed the infection in one lattice point). We now indicate more clearly in the caption and the main text what exactly P(*s*) and *s* refer to.

> ***Reviewer #3 (Recommendations For The Authors):***
>
> *- Would the authors kindly share the simulation code with the community? Also, the data analysis code should be shared to follow current best practices. This needs to be standard practice in all publications. I would go as far as to say that in 2024 publishing a data analysis / simulation study without sharing the relevant code should be ostracized by the community.*

Author response: We absolutely agree and have created a GitHub repository in which we share the C++ source code for the simulations and a Python notebook for plotting. The public repository can be found at https://github.com/BjarkeFN/ViralPercolation. We add this information in supplement under section “Code availability”.

> *- I would avoid the use of the wording "critical" threshold since this is almost guaranteed to infuriate a certain type of reader.*

> *- Line 265 has a curious use of " ... " which should be replaced with something more appropriate.*

Author response: Thank you for pointing it out! We have checked the typos.

https://doi.org/10.7554/eLife.94056.2.sa4