



OPEN

## Modeling poliovirus replication dynamics from live time-lapse single-cell imaging data

Ashley I. Teufel<sup>1,2</sup>✉, Wu Liu<sup>3</sup>, Jeremy A. Draghi<sup>4</sup>, Craig E. Cameron<sup>5</sup> & Claus O. Wilke<sup>2</sup>

Viruses experience selective pressure on the timing and order of events during infection to maximize the number of viable offspring they produce. Additionally, they may experience variability in cellular environments encountered, as individual eukaryotic cells can display variation in gene expression among cells. This leads to a dynamic phenotypic landscape that viruses must face to replicate. To examine replication dynamics displayed by viruses faced with this variable landscape, we have developed a method for fitting a stochastic mechanistic model of viral infection to time-lapse imaging data from high-throughput single-cell poliovirus infection experiments. The model's mechanistic parameters provide estimates of several aspects associated with the virus's intracellular dynamics. We examine distributions of parameter estimates and assess their variability to gain insight into the root causes of variability in viral growth dynamics. We also fit our model to experiments performed under various drug treatments and examine which parameters differ under these conditions. We find that parameters associated with translation and early stage viral replication processes are essential for the model to capture experimentally observed dynamics. In aggregate, our results suggest that differences in viral growth data generated under different treatments can largely be captured by steps that occur early in the replication process.

Populations of RNA viruses are composed of distinct genetic variants that compete within host cells<sup>1–4</sup>. The host cells themselves display a significant amount of variability in gene expression<sup>5–7</sup>, causing the competition of RNA variants to occur within a dynamic system. Populations of viruses display high cell-to-cell variation in their production of viral transcripts<sup>8–11</sup>, due to genetic variability in viral populations coupled with variability in host cell conditions. The stochasticity of the replication process has been well documented in many viral systems<sup>8,10,12,13</sup> and has been observed in high-throughput single-cell poliovirus (PV) infections experiments<sup>14–16</sup>. By measuring viral replication across the infection process under various conditions, these types of high-throughput experiments can gather a wealth of metrics about the timing and progression of infection. These metrics can be examined and compared across experimental conditions to enable an in-depth look at the process of viral replication.

Knowledge of the replication processes and mutation rates can yield insight into managing drug resistance, immune escape, vaccination, pathogenesis, and the emergence of new diseases<sup>17</sup>. Data generated from high-throughput single-cell experiments provide a rich source of information for measuring changes in the replication process under various treatments, but they lack the capacity to describe the underlying intracellular dynamics leading to the observed differences<sup>18</sup>. A quantitative understanding of why intracellular dynamics differ across treatments can lead to insight about specific viral features that could be of interest in medical applications. For example, if a drug treatment increases the length of time it takes before an infected cell lyses, this may have downstream effects on other aspects of the viral replication process, such as an undesirable increase in mutational opportunity. However, measuring how a treatment specifically impacts intracellular dynamics in-vivo is a challenging task<sup>19</sup>.

To examine the intracellular dynamics of viral infection, we have developed a method for fitting population-level viral replication data from high-throughput single-cell time lapse live cell imaging experiments to a mechanistic model of poliovirus replication. Our method simulates time-course viral replication trajectories in single cells for each member of a viral population in order to generate population-level data. We estimate the parameters in the mechanistic model by minimizing the difference between distributions of four metrics that

<sup>1</sup>Santa Fe Institute, Santa Fe, NM 87505, USA. <sup>2</sup>Department of Integrative Biology, The University of Texas at Austin, Austin, TX 78712, USA. <sup>3</sup>Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA. <sup>4</sup>Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061, USA. <sup>5</sup>Department of Microbiology and Immunology, University of North Carolina, Chapel Hill, NC 27599, USA. ✉email: ateufel@utexas.edu

describe the dynamics observed in populations of infected cells. We then examine the distributions of mechanistic parameter estimates to assess which of these parameters are informative about intercellular dynamics. We find that parameters associated with translation and early stage replication processes are essential for describing viral growth. To further assess which aspects of our model are important for fitting experimentally generated viral growth data, we use our model fitting method to fit growth data from experiments under three different drug treatments. These drug treatments include a 3C protease inhibitor (rupintrivir), a polymerase inhibitor (2'-C-meA), and a Hsp90 inhibitor (ganetespib). We again find that the same parameters appear to control our model's ability to capture realistic growth dynamics. Our results suggest that rates of some specific aspects of the viral replication process can be estimated using this modeling approach, and that these rates can inform us about aspects of the replication process that are critical for generating variation in population-level viral growth data.

## Results

**Characterizing population-level infection dynamics from single-cell data.** Time-lapse live-cell imaging of infections can offer unique insight into the timing and magnitude of viral replication within a cell. While transcriptomics approaches have been used to generate time-course data along with a wealth of other data, these approaches are commonly employed to examine dynamics across a small number of time points<sup>4,8,9,20</sup>. By contrast, time-lapse imaging offers the ability to capture temporal dynamics across a large number of time points, allowing for a closer examination of gene expression, growth trajectories, and their variation among cells<sup>15,16,21</sup>.

To capture single-cell time-lapse imaging of viral growth trajectories, Guo et al.<sup>15</sup> introduced a microfluidics device that uses cell density to control single-cell occupancy of wells on the device. This platform was later modified<sup>16</sup> to include a trapping structure on the rear wall of each well, which improved the device's ability to capture cells. The modified device contains 5700 wells, and due to the trapping structure under optimal conditions 86.1% of the wells contain single cells. The ability to capture this large number of infection trajectories provides an adequate sample size for statistical analysis. Experimental data published with this device includes infection trajectories of poliovirus (PV) encoding GFP (GFP-PV), infecting individual HeLa S3 cells in the presence of several antiviral agents<sup>16</sup>.

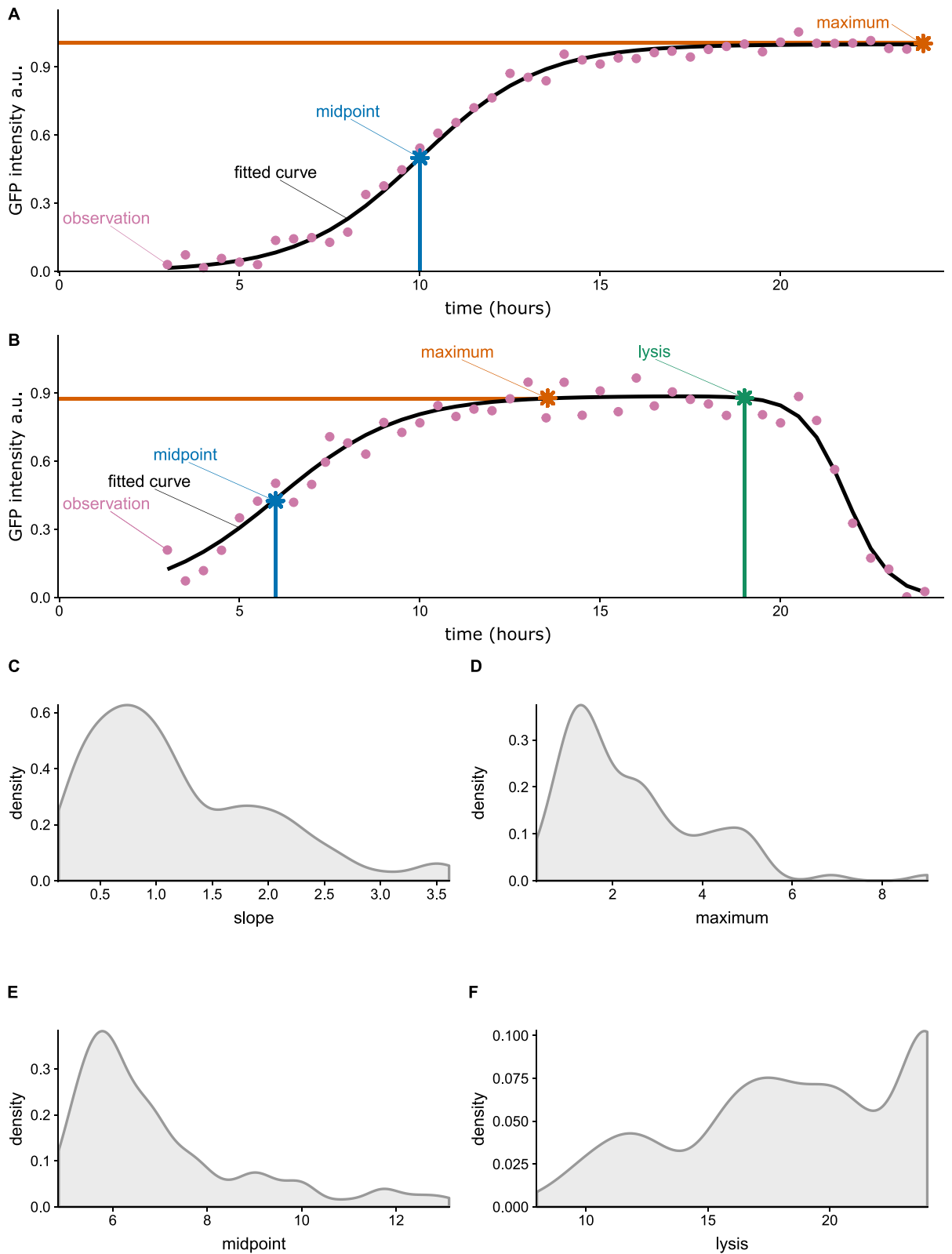
Here, we make use of these previously published time-lapse live-cell imaging data sets for GFP-PV. GFP-PV was created by inserting a GFP coding sequence between the capsid-coding P1 region of the PV genome and the non-structural protein-coding P2 region. Using the microfluidic device described above, infection trajectories in single cells were measured via fluorescence detection, where changes in fluorescence were monitored over time with time-lapse microscopy<sup>16</sup>. These time-lapse data were then analyzed by fitting either a sigmoidal function (modeling no lysis, Fig. 1A) or a double sigmoidal function that captures the diffusion of GFP when the cell lyses (modeling lysis, Fig. 1B) to each of the infection events fluorescence profiles<sup>22</sup>.

The sigmoidal model is determined by three fitted parameters, the maximum intensity observed, the time at which intensity has reach half of its maximum, and the slope. The double sigmoidal model is determined similarly, and considers the maximum intensity observed, the final intensity reached, the time at which half of the maximum intensity is reached, the time at which intensity has decayed halfway to its final value, the slope of increasing intensity, and the slope of decreasing intensity. The time of lysis is calculated by finding the point that a line which passes through the decay midpoint and its corresponding slope intersects with the maximum intensity.

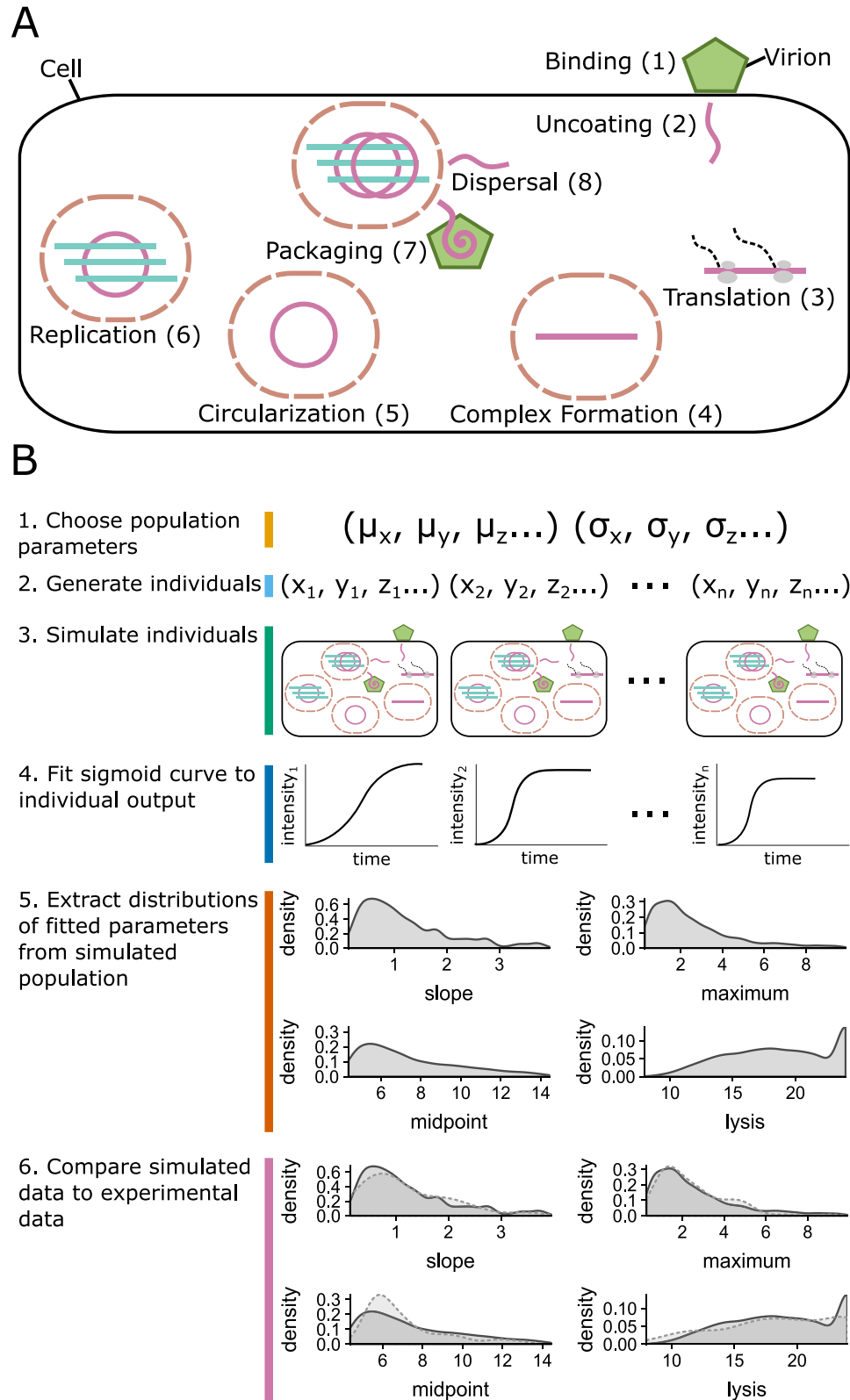
For each infection, one or the other of these two fitted functions is selected as the best fit, based on the AIC model-fit criterion<sup>22</sup>. This fitting procedure allows us to quantify the distributions of parameters: the slope of the first sigmoidal function at the midpoint (slope), the maximum amount of GFP intensity reached (maximum), the midpoint at which half of the maximum GFP intensity is reached (midpoint), and the lysis time (lysis) (Fig. 1A,B). By aggregating these parameters over many individual infections, we can estimate the distributions of each of these parameters (Fig. 1C–F) and their relationships with each other (Fig. S1). These distributions can then be used to compare population-level infection dynamics generated under different experimental conditions. The variability observed in these parameters is due to a combination of both cellular and viral processes. Previous work<sup>15</sup> demonstrated that the maximum, slope, and length of infection-time are primarily determined by the virus, while the starting and midpoint parameters seem to be determined by cellular factors. What these factors are is not known, however. Cell-cycle was specifically excluded as a potential cause of cell-to-cell variation.

**Model of intercellular replication.** Schulte et al.<sup>23</sup> previously introduced a stochastic model that simulates individual reactions and tracks abundances of PV molecular species within a cell. We make use of this model<sup>23</sup> to generate populations of viral growth trajectories that are fitted to the distributions of parameters collected from single-cell experimental data generated by Liu et al.<sup>16</sup>. The model introduced by Schulte et al.<sup>23</sup> consists of eight distinct steps in the viral replication process (Fig. 2A). First, infecting virions bind a cell, their genome is uncoated, and translation begins. Once enough protein product is built up this allows for replication complexes to form. Within each replication complex, the genome then circularizes and replication begins. Newly replicated genomes can either be packaged into new virion particles or disperse into the cell. The full mathematical description of this model is presented in Schulte et al.<sup>23</sup> and is summarized in Table 1. Here, we briefly elaborate on the the biological meaning of each step and the parameters used to describe them.

First, infecting virions bind to and infect a cell (Fig. 2A, step 1). The binding process is based on a population's average multiplicity of infection (MOI) and the number of infecting virions is drawn from a Poisson distribution with a mean equal to the MOI. Uncoating occurs after a gamma-distributed delay occurs<sup>24</sup>, resulting in positive-sense genomes being introduced into the cell (Fig. 2A, step 2). Once positive-sense genomes are present translation begins (Fig. 2A, step 3), dependent on a parameter that describes translation ( $c_{trans}$ ) and the amount of RNA genomes present.



**Figure 1.** Sigmoidal functions fit to experimental data and distributions of fitted parameters when the sigmoidal functions are fit to a population of infection events generated by Liu et al.<sup>16</sup> without drug treatment. (A) Example of sigmoidal function fit to experimentally data of GFP intensity measured in arbitrary units (a.u.). A sigmoidal curve (black line) is fitted to intensity observations from a single infection event (purple dots), allowing for the estimation of the midpoint (blue) and the maximum intensity value (orange). (B) Example of double-sigmoidal function fit to a single infection event where lysis is observed. A double sigmoidal curve (black line) is fitted to intensity observations from a single infection event (purple dots), allowing for the estimation of the midpoint (blue), the maximum intensity value (orange), and the time point of lysis (green). (C) Distribution of slopes calculated at the midpoint. The slope is related to the viral replication rate. (D) Distribution of the maximum intensity. (E) Distribution of the amount of time until the half of the maximum intensity is reached. (F) Distribution of the length of infection time.



**Figure 2.** Illustration of PV replication cycle and parameter estimation procedure. (A) The replication cycle of PV as represented in the model of Schulte et al.<sup>23</sup>. This figure is adapted from that work. Numbered steps correspond to individually modeled reactions given in Table 1. (B) Computational procedure to compare the output of the mechanistic model of a single PV infection to the experimental data obtained from populations of single-cell infections.

Reaction	Name	Equation	Reactants and products
1	Binding	$RNA_{initial}^+ \sim Poisson(MOI)$	infecting viron $\rightarrow$ coated genome
2	Uncoating	$t_{uncoat} \sim Gamma(0.02, 0.678)$	coated genome $\rightarrow$ positive sense genome
3	Translation	$r_{trans} = c_{trans}[RNA_{lin}^+]$	positive sense genome $\rightarrow$ 3A + CAP
4	Complex formation	$r_{com} = c_{com}(1 - \frac{[com]}{com_{max}})[3A]$	positive sense genome + 3A $\rightarrow$ complex + 3CD + positive sense genome
5	Circularization	$r_{circ}^i = c_{circ}[3CD]_i$	positive sense genome + complex + 3CD $\rightarrow$ circularized genome
6	Replication	$r_{rep+} = c_{rep\ pos}[RNA_{circ}^+](1 - \frac{rep}{rep_{max}})$ $r_{rep-} = c_{rep\ neg}[RNA_{circ}^-](1 - \frac{rep}{rep_{max}})$	circularized genome $\rightarrow$ negative sense genome negative sense genome $\rightarrow$ positive sense genome
7	Packaging	$p_{pack} = 1 - e^{-c_{pack}[CAP]}$	positive sense genome + CAP $\rightarrow$ packaged genome
8	Dispersal	$p_{stay} = c_{stay}(1 - p_{pack})$	positive sense genome $\rightarrow$ genome dispersed in cell

**Table 1.** The reactions that describe the replication cycle of PV. Numbered steps correspond to individually modeled reactions as described in Fig. 2. See Schulte et al.<sup>23</sup> for a full mathematical description of the model.

After a sufficient amount of protein products are present, the positive sense strands form replication complexes (Fig. 2A, step 4). Complex formation is modeled as a first-order reaction that is a function of a parameter that describes complex formation ( $c_{com}$ ), the amount of the protein product bounded by the maximum amount of possible complexes that could be formed ( $com_{max}$ ), and the amount of a protein product (3A) that facilitates this complex formation. We assume that this reaction consumes 3A, and we represent this consumption by the parameter  $c_{3A}$ . The genomes can then circularize to become competent for replication (Fig. 2A, step 5), and this process is modeled as a function of a parameter that describes circularization ( $c_{circ}$ ) and the amount of a protein (3CD) needed for the genome to circularize within each replication complex (indexed by  $i$ ).

Replication of negative-sense then positive-sense genomes occurs (Fig. 2A, step 6) after a positive-sense genome has circularized. The rates of production of newly synthesized negative and positive sense genomes are a function of parameters that describe the rate of positive and negative strand synthesis ( $c_{rep-}$  and  $c_{rep+}$ , respectively). This replication process depends on the amount of positive and negative sense circularized genomes, and is bounded by the maximum possible number of replication cycles ( $rep_{max}$ ). The  $rep_{max}$  parameter reflects resource limitations within the cell. The equation that describes the rate of positive strand replication accounts for the rate at which replication is initiated on positive-sense templates, leading to the productions of negative-sense strands. Similarly, the equation that describes the rate of negative strand replication accounts the rate of positive-sense production (Table 1 reaction 6).

After newly synthesized positive-sense genomes are produced, they can either be packaged into new virion particles (Fig. 2A, step 7), diffuse into the cytoplasm and begin translation (Fig. 2A, step 8), or remain in the current complex to continue replicating. The probability of packaging a newly synthesized positive-sense genome (Fig. 2A, step 7) is a function of a parameter that describes the rate of packaging ( $c_{pack}$ ) and the amount of available capsule proteins (CAP). The probability of a newly synthesized positive strand dispersing in the cell (Fig. 2A, step 8) is a function of the probability that it stays within the replication complex ( $c_{stay}$ ) or that it is packaged ( $p_{pack}$ ).

As in Schulte et al.<sup>23</sup>, ten parameters are fit based on a limited set of observations; here, the slope, max, mid-point, and lysis in the simple sigmoidal model, or seven parameters in the double-sigmoidal model. The central motivation for this complexity is the need to capture the effects of stochasticity in infection dynamics, which we expect to be substantial and non-linear based on first principles and prior experimental and modeling work. Viral infections typically begin with as few as one copy of the genome; early random variation in the processes of replication, translation and decay is then amplifies as the infection proceeds<sup>13</sup>. The small numbers of molecules and positive feedbacks (e.g., each copy of the genome can be translated into more proteins, leading to faster genome replication) make the effects of this randomness non-linear. This nonlinearity prevents us from accurately representing noise as, for example, a Gaussian perturbation to mean behavior predicted by an ODE model. Instead, we use a kinetic model to present biologically inferred steps and processes, as described in Schulte et al.<sup>23</sup>. This kinetic model allows for variation in single-cell data to be ascribed to cell-to-cell heterogeneity in specific processes.

**Fitting the model of intercellular replication dynamics to population-level data.** Schulte et al.<sup>23</sup> fitted their stochastic model of PV infection to population-level quantitative RT-PCR data of positive-sense and negative sense strands gathered from replicate experiments at six time points and three different multiplicities of infection (MOI). The fitting was performed via Approximate Bayesian Computation (ABC). Here, we modify this approach in two crucial aspects. First, instead of fitting the model to RT-PCR data (Schulte et al.<sup>23</sup>'s approach), we fit the model to fluorescence time courses generated during infection. Second, instead of fitting the model to population-level data, we fit it to data representing individual infections in single cells. Since we have thousands of individual infection time courses that display substantial variation among them due to the stochastic nature of both infection dynamics and cellular environments, we cannot fit an individual set of model

parameters. Instead, we have to assume that each parameter is represented by a normal distribution characterized by a mean and a variance, and we fit these distributions to the population of infection time courses. In effect, we are converting Schulte et al.<sup>23</sup>'s fixed-effects model into a random-effects model to capture the variation among individual infection time courses due to variation in internal state among different infected cells. Because this random-effects model presents a much increased parameter search space we employ a genetic algorithm (GA) for initial parameter optimization, followed by multiple rounds of the ABC procedure outlined by Schulte et al.<sup>23</sup> to estimate posterior distributions of parameter values.

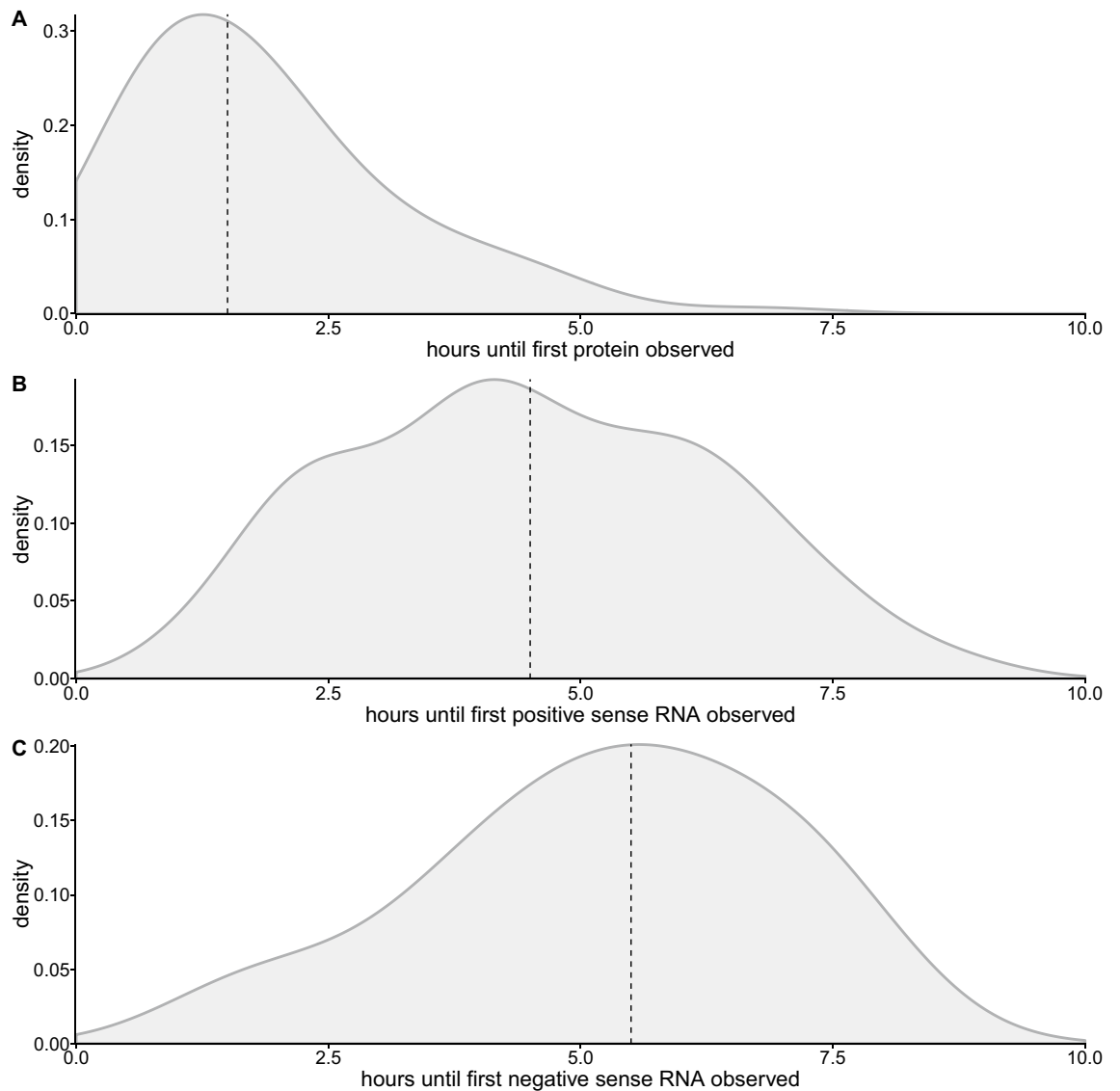
Fitting the stochastic model of intercellular replication dynamics to a population of infection time courses is implemented as follows: We first randomly generate a set of model parameters that represent the viral population that is to be simulated (Fig. 2B, step 1). This set of model parameters consists of a mean ( $\mu_i$ ) and a variance ( $\sigma_i$ ) parameter for each of the model's mechanistic variables. We then generate a population of individuals by randomly drawing sets of variables from the distributions specified by  $\mu_i$  and  $\sigma_i$  (Fig. 2B, step 2) and simulating the infection dynamics for each individual cell (Fig. 2B, step 3) using the model introduced by Schulte et al.<sup>23</sup>. Each individual simulation produces an infection time course from which we extract the slope, maximum, midpoint, and time of lysis parameters as described by Caglar et al.<sup>22</sup> (Fig. 2B, step 4). The estimates of these parameters are then pooled among individuals to create distributions that describe the population of infections (Fig. 2B, step 5). We then compare each of these four simulated parameter distributions to the experimentally observed distributions with a K-S test (Fig. 2B, step 6). The K-S tests produce a  $D$  statistic that represents how similar each of the four simulated parameter distributions are to the experimental observations. The smaller the value of  $D$  the more similar the simulated and experimental distributions are. We sum over the  $D$  values for each of four distribution comparisons to create a score for the population parameters (Fig. 2B, step 6). We then employ GA followed by multiple rounds of ABC using this scoring procedure to minimize the difference between the simulated and experimental parameter distributions to estimate the mechanistic parameters in our model of PV infection (for details, see Methods: Model Fitting).

**Estimates of inter-cellular replication dynamics.** As the first application of our simulation model, we fit its output using a combination of GA and ABC to the distributions of the slope, maximum, midpoint, and lysis parameters extracted from single-cell infection time courses performed without any drug treatment, and we visually assess that our fitted distributions correspond to the experimental observations (Fig. S2A–D). We also examine Q–Q plots that compare the simulated and experimentally observed data to confirm the quality of fit (Fig. S3A–D). To assess whether our model behaves as a proxy for viral replication, we track when in time we first observe protein and RNA production in our model. We find the median amount of time until first protein production is close to 1 hour, while it takes approximately 4.5 hours for positive sense RNA production to be observed, and 5.5 hours until negative sense RNA production observed (Fig. 3). The temporal order and timing of these events is consistent with the known biology of PV replication<sup>25</sup>.

Next, we examine the posterior distributions of fitted parameters from our mechanistic model to discern which parameters are informative for describing experimentally observed population-level replication dynamics. We find that several of our posterior distributions appear to have fairly peaked distributions, implying that these parameters have converged around a central value (Fig. 4). Parameters whose posterior distributions converge include our estimates of the parameters that describe the processes of translation ( $c_{trans}$ ) (Fig. 4A), complex formation ( $c_{com}$ ) (Fig. 4B), circularization ( $c_{circ}$ ) (Fig. 4C), and positive strand replication ( $c_{rep+}$ ) (Fig. 4D). Estimates of the posterior distributions of these parameters were observed to converge by Schulte et al.<sup>23</sup> as well. Our observation that these distributions appear to converge around a centralized value suggests that individuals within a viral population behave similarly in regards to translation, compartmentalization, circularization, and positive strand replication. Further, appropriately capturing these dynamics appears to be important for our model to describe experimentally generated population-level growth data.

Posterior distributions for parameters associated with negative strand replication, packaging, and the maximum number of compartments (Fig. 4E–G) have large variance. This variation in posterior estimates suggests that PV may exhibit variability in how the respective processes are being carried out within individual cells, or that these aspects of the model are less essential for capturing dynamics observed in the experimental data. We observe skewed posterior distributions of our estimates of the maximum number of replication cycles ( $rep_{max}$ ) (Fig. 4H), consumption of the protein product 3A (Fig. 4I), and the probability of a newly synthesized positive strand to stay within the replication complex (Fig. 4J). These skewed distributions may suggest that while variation exists, an increased maximum number of replication cycles, a reduced amount of protein consumption, and a high probability of a new strand to stay in the replication complex occur more often.

**Informative parameters.** To further examine which parameters in our model are informative for describing experimentally observed growth data we perform principal component analysis on the posterior estimates to reduce the dimensionality of our parameter space. We find that the first principal component explains nearly 18% of the observed variance, and that the first six principal components explain a cumulative 73% of the observed variance (Fig. 5A). Examining the relative contribution of features in each of the principal components (Fig. 5B–G), we find that the parameters that describe complex formation and circularization have the largest amount of relative contribution in the first principal component (Fig. 5B). These two parameters are expected to be related since circularization occurs within complexes (Fig. 2B steps 4 and 5). Considering these two features in the first principal components are also parameters we observed to converge (Fig. 4) suggests that these parameters are an important aspect of our model for generating viral growth data. We also observe that the signs of the relative contribution of these parameters differ suggesting that these two parameters are negatively correlated. The large positive loading on  $c_{com}$  and a large negative loading on  $c_{circ}$  means that the biggest difference among



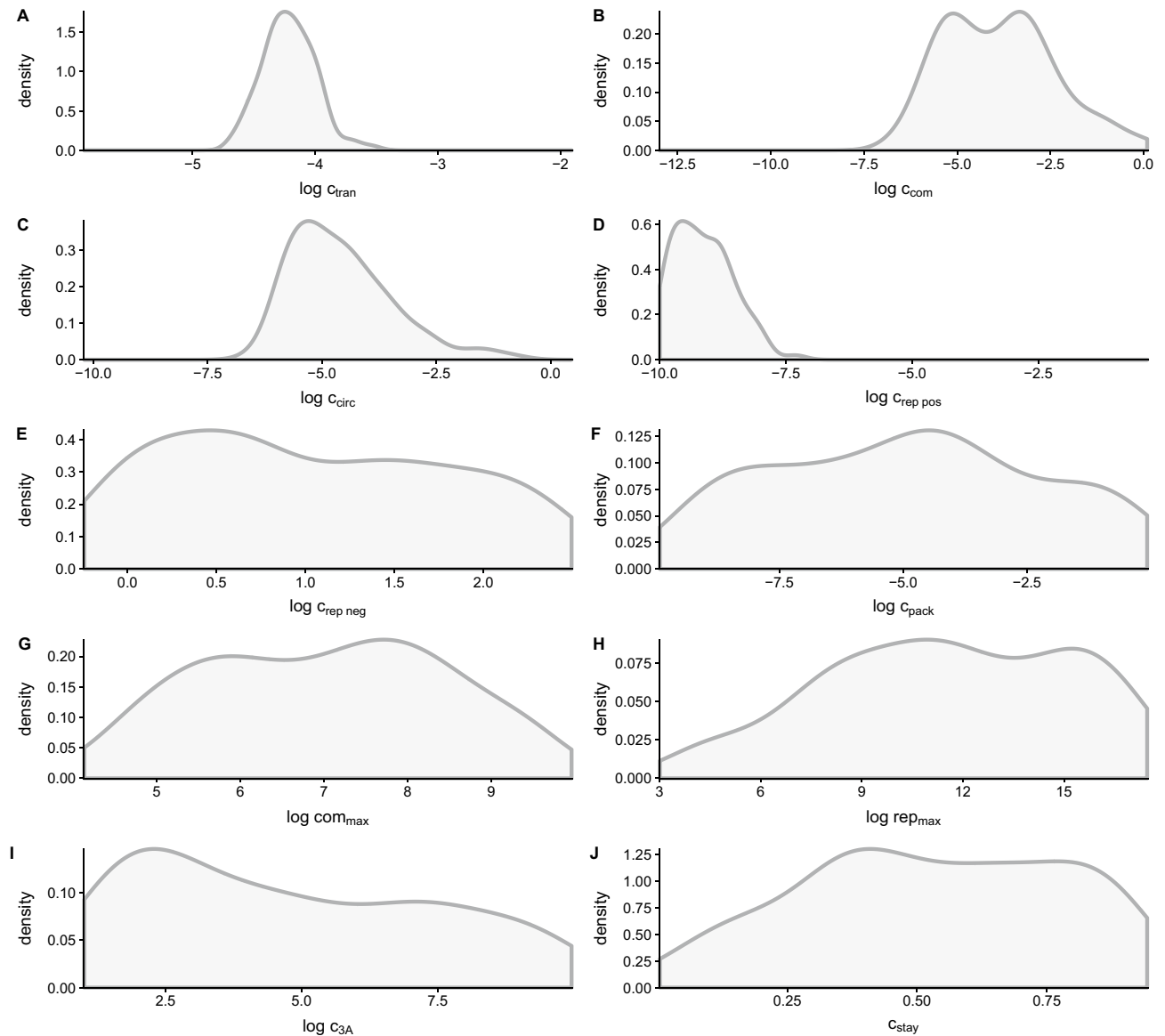
**Figure 3.** Time until first occurrence of events as estimated from our model of PV infection when the model is fit to the experimental data shown in Fig. 1C–F. **(A)** Hours until protein is first produced. **(B)** Hours until the first production of positive sense RNA. **(C)** Hours until the first production of negative sense RNA.

fitted parameter sets is the trade-off between these two parameters. Hence, increasing  $c_{com}$  relative to  $c_{circ}$  or vice versa would have the largest effect on the fitted trajectories.

The maximum number of replication events ( $rep_{max}$ ) and the consumption of the protein product 3A have the largest amount of relative contribution in the second principal component (Fig. 5C). This means that on the axis orthogonal to PC 1  $rep_{max}$  and  $c_{3A}$  account for the greatest amount of variation in terms of the spread of parameters relative to PC 1. In the model the maximum number of replication events represents an upper bound on replication, and it may be influential on the maximum amount of GFP produced in our simulations. The protein product 3A is needed for complex formation, suggesting that accounting for the use of resources produced during translation may also be an important aspect of modeling viral growth data.

Additionally, to investigate how parameters interact within our model we examine the correlations between parameters (Fig. S4). We find that the strongest relationship exists between parameters associated with compartmentalization and circularization. Interestingly, these parameters are found to have the largest amount of relative contributions in the first principal component (Fig. 5B). Other parameter estimates are less correlated, though there appears to be a relationship between parameters associated with the maximum number of replication cycles and the probability that a new strand will stay in its replication complex, as well as translation and complex formation.

**The effect of drug treatments on parameter estimates.** To examine if the mechanisms found to be informative about viral growth dynamics in our initial data set—without drug treatment—also describe the variability observed in experiments with drug treatment, we expand our analysis to three data sets generated

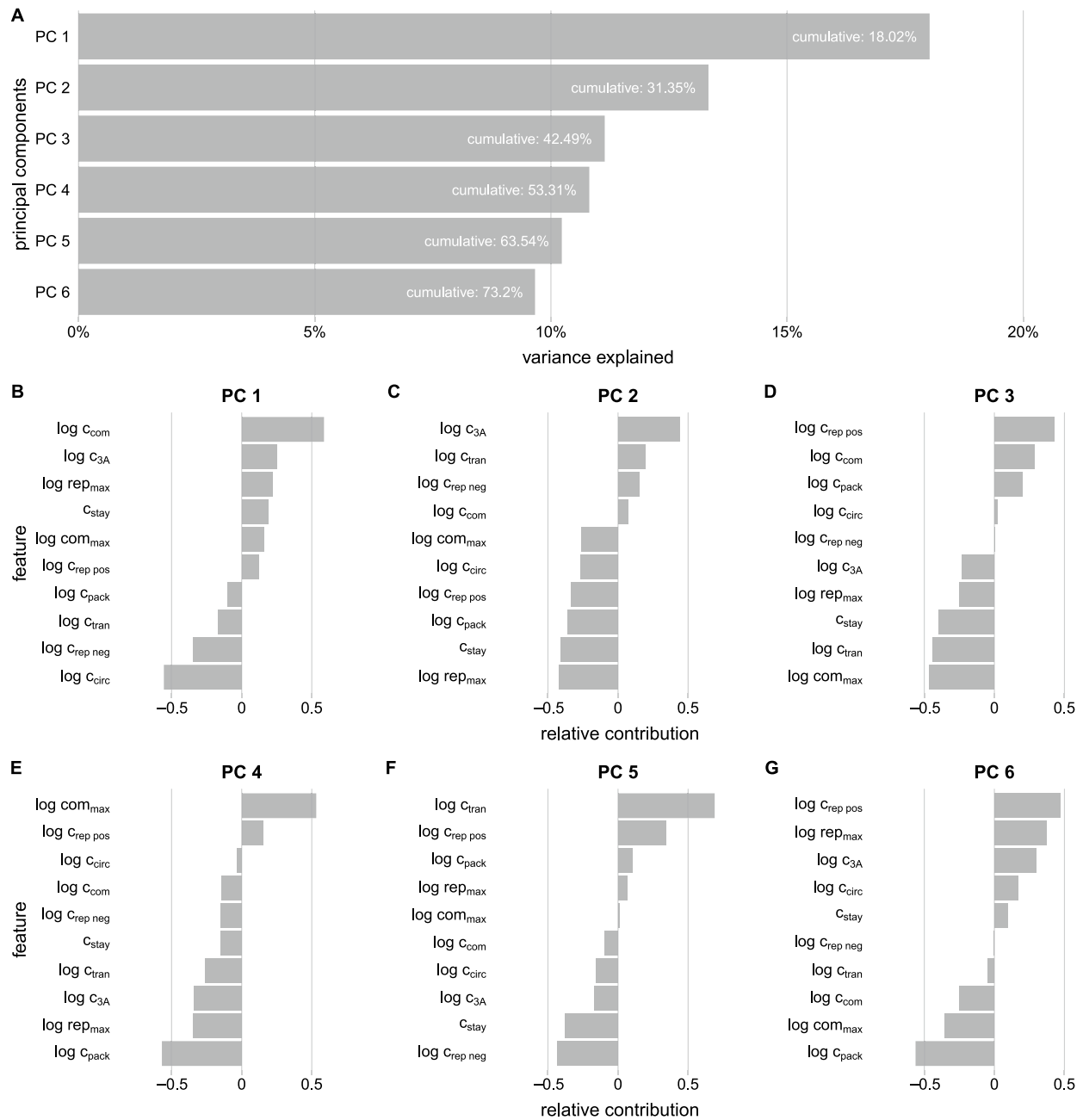


**Figure 4.** Estimates of mechanistic parameter posterior distributions in our model of PV infection from fitting the model to experimental data generated without drug treatment (Fig. 1C–F). Parameters are estimated by fitting the model described by the equations in Table 1 and illustrated in Fig. 2A. The parameters shown correspond to those labeled in each reaction. (A) Translation, which occurs in step 3 of the model. (B) Compartmentalization, a part of step 4. (C) Circularization, step 5. (D, E) Replication of positive and negative sense RNA, step 6. (F) Packaging, step 7. (G) The maximum number of compartments possible, considered in step 6. (H) The maximum number of replication cycles permitted by cellular resources, a limiting factor in step 4. (I) Consumption of the protein product 3A, step 4. (J) The probability for a newly synthesized genome to stay in the replication complex, step 8.

using different classes of drugs (Fig. S2E–P, Fig. S3E–P). We find that the estimates of several of the distribution of means of our mechanistic parameters differ (Fig. 6A–D, H–J) while some do not (Fig. 6E–G) when fitting our model to these additional three data sets. Treatment with the drug rupintrivir, a 3C protease inhibitor, affects our estimates of parameters associated with translation, compartmentalization, circularization, and replication of positive strands (Fig. 6A–D). Rupintrivir increases the range of parameter estimates we find to be informative when fitting data generated without drug treatment. However, when fitting the model to experimental data generated under treatment with rupintrivir we observe poor convergence (Fig. S2E–H, Fig. S3E–H), though it is notable that only the posterior distributions of the four parameters that clearly converge in the no-drug case differ.

Fitting our model to experimental observations under treatment with 2'-C-meA, a polymerase inhibitor, we find that our estimates of parameters associated with translation, compartmentalization, circularization, consumption of protein products, and the probability of a new strand staying in its replication complex differ (Fig. 6A–C, I, J) (see also Fig. S2I–L, Fig. S3I–L). 2'-C-meA decreases the translation (Fig. 6A) and circularization (Fig. 6C) parameters, while increasing the parameters associated with compartmentalization (Fig. 6B) and the

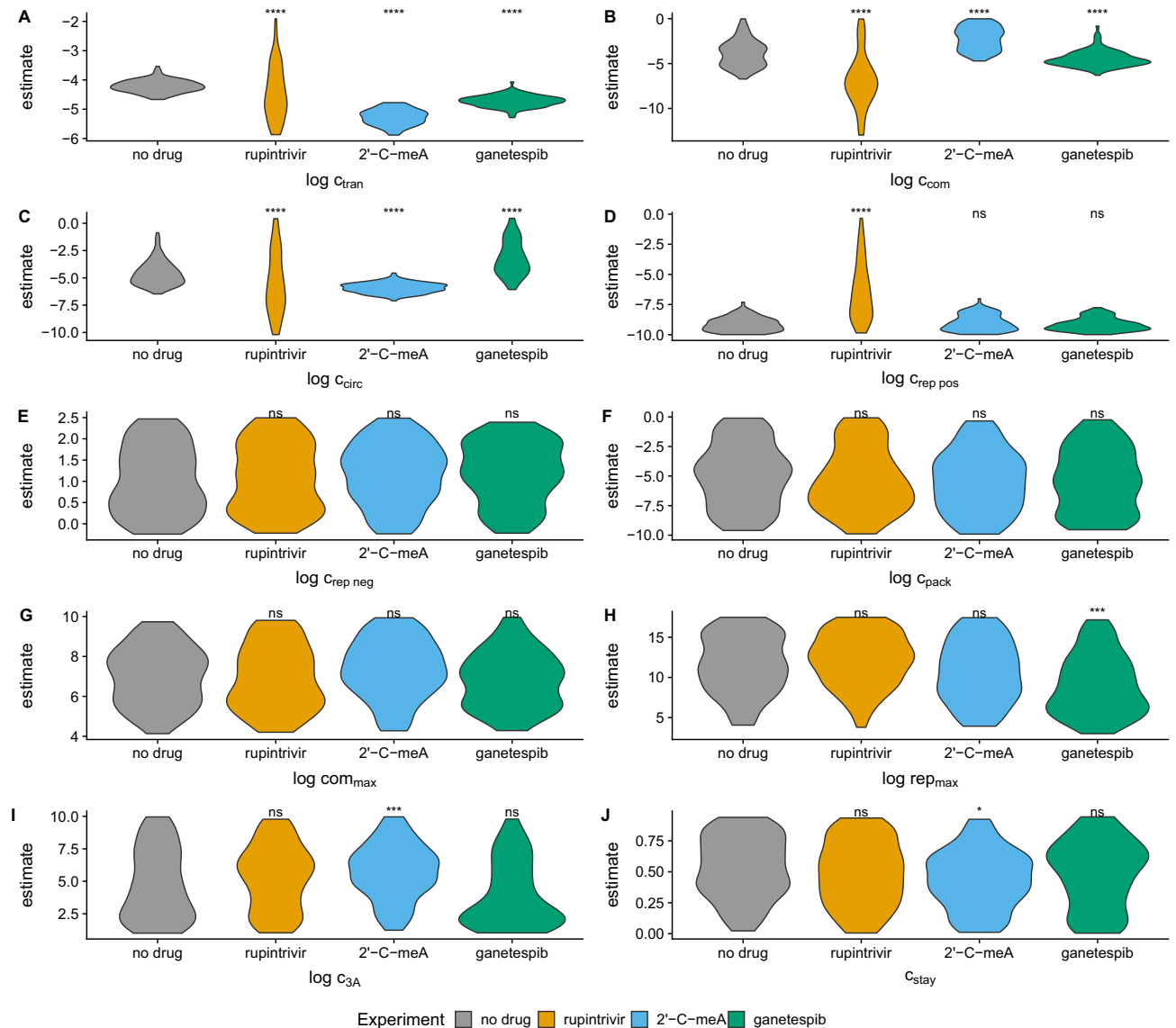




**Figure 5.** Principal component analysis of parameter distributions estimated from fitting our mechanistic model of PV infection to experimental data (Fig. 1C–F). (A) The amount of variance explained by the first 6 principal components. The white labeling provides the cumulative amount of variance explained by including all components up to and including the labeled one. (B–G) Relative contribution of features for each of the first six principal component axes.

consumption of 3A (Fig. 6I). Notably, our estimate of the probability of a new strand to stay in the replication complex is decreased under this drug treatment (Fig. 6J). 2'-C-meA is the only drug treatment where we find a substantial change in the probability of a strand to stay in its replication complex, suggesting that this treatment impacts the viral replication process differently than the other two treatments.

Fitting our model to experimental observations under treatment with ganetespib, an Hsp90 inhibitor, impacts our estimates of translation (Fig. 6A), compartmentalization (Fig. 6B), circularization (Fig. 6C), maximum number of replication cycles (Fig. 6H), and the probability of a new strand staying in its replication complex (Fig. 6A–C, H, J) (see also Fig. S2M–P, Fig. S3M–P). Treatment with ganetespib appears to similarly impact our estimates of some aspect of the protein production and utilization process as the other drug treatments. The most notable change to parameter estimates under treatment with ganetespib is the decrease in the maximum number of replication cycles (Fig. 6H). This treatment is the only out of the three to do so. In aggregate, we find



**Figure 6.** Comparison of posterior parameter distributions between the no drug treatment and drug treatments. Asterisks indicate distributions that differ significantly from the no drug treatment after Bonferroni correction for multiple testing. Non significant (ns) corresponds to  $p > 0.05$ , a single asterisk corresponds to  $p \leq 0.05$ , two asterisks correspond to  $p \leq 0.01$ , three correspond to  $p \leq 0.001$ , and four correspond to  $p \leq 0.0001$  from a K-S test. Parameters are estimated by fitting the model described by the equations in Table 1 and illustrated in Fig. 2A. Parameters correspond to those labeled in each reaction. (A) Translation, which occurs in step 3 of the model. (B) Compartmentalization, a part of step 4. (C) Circularization, step 5. (D, E) Replication of positive and negative sense RNA, step 6. (F) Packaging, step 7. (G) The maximum number of compartments possible, considered in step 6. (H) The maximum number of replication cycles permitted by cellular resources, a limiting factor in step 4. (I) Consumption of the protein product 3A, step 4. (J) The probability for a newly synthesized genome to stay in the replication complex, step 8.

that each of the drug treatments affects different parameter estimates. This finding suggests that treatment with each of the drugs impacts the viral replication process differently.

The drug treatment data also result in estimates of parameter variances that differ from the case of no-drug treatment in several cases (Fig. S5). Each of the drug treatments appears to impact variation in translation (Fig. S5A) and in the maximum number of replication cycles (Fig. S5H). Rupintrivir also affects variation in the probability that a new strand will stay in its replication compartment (Fig. S5J), while ganetespi affects variation in the maximum number of compartments (Fig. S5G).

To examine which parameters are informative for estimating experimentally observed growth data under drug treatments we again perform principal component analysis. We find that the first 6 principal components of each drug treatment explain around 75% of the variance (Fig. S6A,D,G), similar to that found in the no drug case. Examining the features of the first and second principal components under treatment with rupintrivir, we find that parameters associated with compartmentalization, strand replication, and packaging have the largest

relative contributions (Fig. S6B,C). The first and second principal components under treatment with 2'-C-meA show that parameters associated with consumption of the protein 3A, circularization, the maximum number of compartments and replication cycles have the largest relative contributions (Fig. S6E,F). When fitting our model to experimental data treated with ganetespib we again find parameters associated with circularization, compartmentalization, strand replication, and packaging also have the largest relative contributions (Fig. S6H,I).

Notably, many of the parameters found to have a larger amount of relative contribution in the first two principal components of each of the drug treatments are found to have larger relative contributions in the no drug case. Considering that parameters associated with compartmentalization and circularization have a larger amount of relative contribution in the first two principal components in many of our experiments suggests that these parameters are important for our model to recapitulate experimental observations.

Additionally, drug treatments influence the correlations between parameters. Treatment with rupintrivir results in a negative correlation between compartmentalization and the replication of positive strands (Fig. S7). Treatment with the other two drugs 2'-C-meA (Fig. S8) and ganetespib (Fig. S9) results in a correlation between compartmentalization and circularization, similar to that observed in the no drug case (Fig. S4).

**Estimates of the number of generations between infecting and packaged virions.** Schulte et al.<sup>23</sup> used their model to estimate the mean number of generations between infecting and packaged virions. We repeat this analysis on our four data sets to assess if our parameter estimates generate similar generation estimates (Fig. S10). Examining the distributions of mean number of generations between infecting and packaged virions, we find that these distributions are slightly right skewed and two of the drug treatments generate distributions that are significantly different from the no drug case (Fig. S10B,C). The long tails of these distribution suggest that a small fraction of packaged virions undergo an increased number of generations. The median estimate of the mean number of generations in the treatment without the drug is 4.75, while treatment with rupintrivir results in 5.77 generations, treatment with 2'-C-meA results in 4.67 generations, and treatment with ganetespib results in 4 generations (Fig. S10). We find that rupintrivir (Fig. S10B) and 2'-C-meA (Fig. S10C) both increase the median of the distribution of generations for a packaged virion compared to the no drug case, while treatment with ganetespib decreases the median number of generations (Fig. S10D).

The number of replication cycles in PV has previously been estimated to be 2.1–4.6<sup>26</sup>, and these estimates were based on measurements of the initial and final virus titers combined with burst size measurements taken from previous studies<sup>2,27–29</sup>. Schulte et al.<sup>23</sup> estimated a slightly higher number of approximately 5 replication cycles. Our estimate of between 4–5.77 replication cycles seems to correspond well to the estimates from these other studies, suggesting that, similarly to our findings regarding the timing of events (Fig. 3), our model is behaving as a reasonable proxy of biological reality.

## Discussion

The use of microfluidic platforms combined with live cell imaging has expanded rapidly over the last five years, presenting new avenues and challenges for elucidating mechanisms that contribute variability in viral growth<sup>30</sup>. Here, we make use of a single-cell live-imaging data set<sup>16</sup> that offers a high-resolution view of the infection process by measuring viral growth at many time points to examine variation within and among populations. By measuring kinetic parameters of growth trajectories from individual members of a population of infections, the aggregate of these data can be used to examine the sources of variability and allow for comparison across treatments.

To better understand variability in viral growth we have developed a model-fitting framework to estimate parameters of viral replication by fitting a stochastic replication model to a population of infection time-courses measured in individual infected cells. We confirmed that the model captures several biologically relevant aspects of viral replication not obviously visible in the input data, such as the timing of intercellular events and the number of generations between infecting and packaged virions. Using this framework, we have identified the features of the model that are critical for capturing viral growth dynamics. We find that accounting for translation and for early stage processes such as complex formation and circularization are the most essential aspects of the model. Estimates of the parameters representing these processes appear to converge around a central value and are found to be important features in PCA analysis. Other parameters in the model such as the maximum number of compartments appear to have high variation. The variation in these parameters maybe due to high stochasticity in these aspects of the replication process or simply because capturing these dynamics are less essential for our model to generate dynamics similar to those observed in experimental data.

Using three experimental data sets generated under various drug treatments, we have observed substantially different estimates of critical parameters compared to the no-drug case. The differences in the posterior distributions under drug treatment may be due to the fact that our model does not include a mechanistic description of these drug effects. Hence, our estimates of parameters from these experiments may not necessarily reflect how specifically drug treatments impact aspects of viral replication, but rather they may highlight which parameters in our model are capable of capturing the variation in growth dynamics under various experimental conditions. Additionally, we have found that two of the drug treatments, rupintrivir and 2'-C-meA, lead to an increased median number of generations between infecting and packaged virions. This observation is most likely due to the fact that drug treatments lead to a longer time until cell lysis, increasing the opportunity for more generations to occur.

The observation that some drug treatments increase the length of time until cell lysis generates the interesting hypothesis that some drug treatments may increase the number of replication cycles inside the cell. The increased number of replication cycles would indicate a greater effective mutation rate under some drug treatment. Due to the relationship between the number of replication cycles, mutation accumulation, and drug resistance, examining the number of replication cycles is critical for developing effective and safe antiviral treatment regimens, and

therefore many studies attempt to estimate it<sup>17,31–33</sup>. Drug treatments have been shown to affect genetic diversity in various other systems. Treatment with 3-azido-3-deoxythymidine (AZT) has been found to influence the *in vivo* mutation rate of HIV-1 and can increase the rate of mutation by a factor of 7 in a single round of replication<sup>34</sup>. Treatment with favipiravir has been shown to result in an effective increase in the viral mutation rate of influenza A virus<sup>35</sup>. The broad-spectrum antiviral drug ribavirin is an RNA virus mutagen that can cause a 9.7-fold increase in mutagenesis in poliovirus resulting in error catastrophe<sup>36</sup>. Drug treatments have also been shown to increase viral replication. Corticosteroids increase viral replication in the adenovirus type 5/New Zealand rabbit ocular model<sup>37</sup>. Valproic acid A has been shown to increase viral replication in HIV<sup>38</sup>. Treatment with rapamycin or everolimus facilitates hepatitis E virus replication<sup>39</sup>. Extending the model used here to account for the mechanisms of a drug's action to examine how drug treatments impact the PV replication process could be a useful step for studying the relationship between drug treatments and viral evolution.

Our observations of which aspects of the viral replication process are important for capturing experimentally observed viral dynamics are based on a specific model, and our findings are only sound if the model is indeed capturing biological reality. Considering that our prediction of the timing of events and approximately 4–5.77 replication cycles align with the previous literature, we believe that the model correctly captures at least some aspects of the viral replication process. To further improve the model, a concise picture of how the replication process occurs is needed.

Results from this work also highlight the need for more quantitative experimental approaches to enhance model fitting. Notably, many of our estimated parameters appear not to converge around a central value, suggesting that some parameters in this model can not be uniquely identified. Models that exhibited large uncertainty of some parameter estimates when fit to data are common in many systems biology models and often referred to as 'sloppy'<sup>40</sup>. Further computationally intensive work is needed to examine parameter identifiability and the relationship between data, stochastic modeling, and stochastic model fitting procedures.

Including additional metrics in the fitting process could help to resolve some of the issues with parameters that appear not to converge around any central value. For example, if the formation of replication complexes could be microscopically tracked, including this data in the fitting process could be beneficial. To explore intracellular dynamics in greater detail this model could be extended to include more detailed description of each of the reactions in the model along with parameters that account for intracellular selection, recombination, and mechanisms of the effects of drug treatments. However, extending the model does have some drawbacks. Accounting for further levels of biological reality would require an increase in the number of parameters in the model, which could lead to overfitting. It would also likely increase the amount of computational time necessary to fit the model.

Fitting stochastic models to population-level time-course data is a relatively new field of research. One issue that consistently arises in this field is that fitting by comparison of experimental and simulation distributions has the distinct drawback of high computational execution time<sup>41–44</sup>. We have encountered this issue in our study as well. The ABC runs can take several weeks to converge for a single set of experimental data, even when the fitting code takes advantage of parallel model evaluations running on over 100 CPUs at the same time. In other words, convergence of a single ABC run takes on the order of 20,000 to 50,000 CPU hours. This high computational cost arises because we need to execute hundreds of independent simulation runs to evaluate any one given set of parameter values. Because the model is already so computationally costly, expanding it to incorporate further levels of biological reality would undoubtedly increase the run time even further. Thus, irrespective of the specific analysis we have carried out here, we see a need to develop more efficient algorithms for fitting stochastic models to population level data.

Infection at the single-cell level is highly heterogeneous and a number of differing experimental approaches have been developed to study these dynamics<sup>45–50</sup>. In this study we made use of data from time-lapse live cell imaging experiments<sup>16</sup>. Similar experimental approaches have been employed to study the dynamics of other viruses<sup>45,46</sup>, though transcriptomics based approaches have also proved a valuable tool for examining heterogeneity during infection<sup>47–50</sup>. The technology behind both transcriptomic and live-cell imaging has developed rapidly over the last several years<sup>30</sup> and recent studies are now capable of combining live-cell imaging with ultrasensitive quantification of proteins and mRNA in single cells<sup>51</sup>. As both of these technologies evolve they allow for an increasingly detailed look in the viral replication process; presenting growing challenges and emergent opportunities for the development of computational methods capable of processing and interpreting the data they generate<sup>52</sup>.

The results presented here represent a stepping stone in the pursuit of using mechanistic modeling approaches to tease apart details of high-throughput single-cell viral growth dynamics that may be unobservable with experimental approaches alone. Additionally, this work introduces a framework for fitting models designed to describe single time-course trajectories to population level dynamics. Models like these can inform researchers about which aspects of the viral replication process are more variable or stable than others. Identifying which aspects of intercellular dynamics lead to the variation observed in viral growth data could lead to insight relevant to drug development efforts. Though our findings are limited by the biological reality that our model is able to capture, they suggest that factors that occur early in the viral replication process are a key aspect of describing viral growth data.

## Methods

**Processing input data.** Using previously published data<sup>16</sup> from single-cell PV growth experiments generated under four conditions, no drug, 2'-C-meA treatment, rupintrivir treatment, and ganetespiib treatment, we compile several metrics prior to model fitting. First, for each of the experimental conditions, we only retain infection time courses that can be classified as either sigmoidal or double-sigmoidal according to the sicegar R package<sup>22</sup>. Using the estimates of slope, midpoint, maximum, and lysis time produced by sicegar, we generate

the parameter distributions that describe each of the experimental conditions. In cases where the cell fails to lyse within 24 hours, we record the lysis time as 24 hours.

**Model fitting.** Our model fitting and analysis procedure is implemented in the R package *spire* (Single-cell Population Infection Replication Estimation) (see Code and Data Availability below), which builds on PV simulation code originally developed by Schulte et al.<sup>23</sup>.

Our model of PV replication consists of 25 parameters. These parameters include 10 mechanistic parameters and their variances that describe infection dynamics within a cell, a fluorescence scaling term, the shape and rate parameters of a Gamma distribution of the time until lysis, and a lag time and its variance until fluorescence. We then sample from each of these distributions of 25 parameters to create a parameter vector, which we refer to as the instantiating vector. For each instantiating vector we generate 100 variations of the vector, by sampling from distributions determined by the instantiating vector's parameter values. For each of the 100 variant parameter sets, we then simulate a PV infection time course using the model introduced by Schulte et al.<sup>23</sup>. Each time course data produced by the model is then fit using the R package *sicegar*<sup>22</sup>, which estimates the time of lysis, the maximum intensity, the midpoint, and the slope. We aggregate the estimates from each of 100 time courses to produce distributions of time of lysis, maximum value, midpoint, and slope. Finally, we compare our estimates of these distributions to the observed experimental distributions with a scoring function comprised of summing over the *D* values of Kolmogorov-Smirnov tests. The smaller the sum of the score, the better the instantiating vector describes the data.

To speed up convergence of our fitting procedure, several checks are run on a proposed parameter set before the PV simulation is run. If the bounds between the estimated lag and lysis time are outside of those experimentally observed, the parameter set is considered invalid. To prevent running PV simulations with excessively long run times, a timeout function is also implemented. If a single instance of the PV simulation takes longer than 5 seconds to run, the parameter set that spawned that instance of the simulation is considered invalid. When a time course is returned from the PV simulation, if *sicegar* is unable to fit the data then the parameter set is considered invalid. Additionally, if the estimates returned from *sicegar* of the slope, maximum, midpoint, and lysis are above or below the maximum and minimum experimentally derived values by more than 10% the parameter set is considered invalid.

To estimate the model parameters we first use a genetic algorithm (GA)<sup>53</sup> to quickly sample the range of valid parameter values. This GA is run for 100 generations, and the population of instantiating solutions of the final generation is then used to initialize a subsequent approximate Bayesian computation (ABC) fit. We employ the ABC parameter inference method introduced by Schulte et al.<sup>23</sup> and run the ABC for numerous rounds. The completion of a round is determined by the algorithm accepting 100 instantiating parameter sets. At each round of the ABC we calculate the mean of the score function for the 100 accepted parameter sets and use the mean as an acceptance criterion during the next round of the ABC. This process of accepting 100 parameter sets and adjusting the acceptance criteria is repeated until the acceptance criterion drops below a value of 0.55, which takes approximately 15 rounds of ABC.

To record the timing of events in the PV simulation, we record the timing of each individual reaction in a simulation using the converged parameter set. From these timing records, we can identify when protein and RNA production takes place. Similarly, we can estimate the mean number of generations between infection and producing packaged virions.

**Data analysis.** Once the ABC acceptance criterion has dropped below the cut off we visually compare our estimated distributions of slope, maximum, midpoint, and lysis time to the experimental results (Fig. S2). To assess the correspondence of our estimated distributions to the experimental results we also examine quantile-quantile plots comparing the observed and estimated distributions (Fig. S3). From visual inspection and from examining the quantile-quantile plots, it appears that our estimated distributions correspond reasonably well to the experimental data, with the exception of fitting to data treated with rupintrivir which failed to converge after several weeks of running the fitting algorithm.

To examine how parameters within the model relate to one another, we calculate their correlation coefficients and cluster these correlations using hierarchical clustering. This clustering allows us to visualize related coefficients closer to one another. To examine which parameters are informative for estimating experimentally observed growth data we perform principal component analysis using the R package *DataExplorer*<sup>54</sup>. We compare the parameter estimates from the experiments under the no drug treatment to those estimated using drug treatments with a K-S test. We use a Bonferroni correction to correct for multiple testing. We apply the correction separately for each drug treatment and correct for  $m = 10$  independent hypotheses corresponding to the 10 different parameter distributions tested.

### Data availability

The software, results, and analysis tools, including the R package *spire*, are available at [https://github.com/ateufel/modeling\\_single\\_cell\\_virology/](https://github.com/ateufel/modeling_single_cell_virology/). The repository has also been archived on Zenodo at <https://doi.org/10.5281/zenodo.3695913>.

Received: 10 October 2020; Accepted: 15 March 2021

Published online: 05 May 2021

## References

- Sobrinho, F., Dávila, M., Ortín, J. & Domingo, E. Multiple genetic variants arise in the course of replication of foot-and-mouth disease virus in cell culture. *Virology* **128**, 310–318 (1983).
- Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. & Andino, R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**, 344 (2006).
- Lauring, A. S. & Andino, R. Quasispecies theory and the behavior of RNA viruses. *PLoS Pathogens* **6**, e1001005 (2010).
- Andino, R. & Domingo, E. Viral quasispecies. *Virology* **479**, 46–51 (2015).
- Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
- Raser, J. M. & O’Shea, E. K. Noise in gene expression: origins, consequences, and control. *Science* **309**, 2010–2013 (2005).
- Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008).
- Russell, A. B., Trapnell, C. & Bloom, J. D. Extreme heterogeneity of influenza virus infection in single cells. *Elife* **7**, e32303 (2018).
- O’Neal, J. T. *et al.* West Nile virus-inclusive single-cell RNA sequencing reveals heterogeneity in the type I interferon response within single cells. *J. Virol.* **93**, e01778-18 (2019).
- Russell, A. B., Elshina, E., Kowalsky, J. R., Te Velthuis, A. J. & Bloom, J. D. Single-cell virus sequencing of influenza infections that trigger innate immunity. *J. Virol.* **93**, e00500-19 (2019).
- Brooke, C. B. *et al.* Most influenza A virions fail to express at least one essential viral protein. *J. Virol.* **87**, 3155–3162 (2013).
- Delbrück, M. The burst size distribution in the growth of bacterial viruses (bacteriophages). *J. Bacteriol.* **50**, 131 (1945).
- Heldt, F. S., Kupke, S. Y., Dorl, S., Reichl, U. & Frensing, T. Single-cell analysis and stochastic modelling unveil large cell-to-cell variability in influenza A virus infection. *Nat. Commun.* **6**, 1–12 (2015).
- Schulte, M. B. & Andino, R. Single-cell analysis uncovers extensive biological noise in poliovirus replication. *J. Virol.* **88**, 6205–6212 (2014).
- Guo, F. *et al.* Single-cell virology: on-chip investigation of viral infection dynamics. *Cell Rep.* **21**, 1692–1704 (2017).
- Liu, W. *et al.* More than efficacy revealed by single-cell analysis of antiviral therapeutics. *Sci. Adv.* **5**, eaax4761 (2019).
- Sanjuán, R. & Domingo-Calap, P. Mechanisms of viral mutation. *Cell. Mol. Life Sci.* **73**, 4433–4448 (2016).
- Iwami, S., Koizumi, Y., Ikeda, H. & Kakizoe, Y. Quantification of viral infection dynamics in animal experiments. *Front. Microbiol.* **4**, 264 (2013).
- Dolan, P. T., Whitfield, Z. J. & Andino, R. Mapping the evolutionary potential of RNA viruses. *Cell Host Microbe* **23**, 435–446 (2018).
- Zanini, F., Pu, S.-Y., Bekerman, E., Einav, S. & Quake, S. R. Single-cell transcriptional dynamics of flavivirus infection. *Elife* **7**, e32942 (2018).
- Shao, Q. *et al.* Coupling of DNA replication and negative feedback controls gene expression for cell-fate decisions. *Iscience* **6**, 1–12 (2018).
- Caglar, M. U., Teufel, A. I. & Wilke, C. O. Sicegar: R package for sigmoidal and double-sigmoidal curve fitting. *PeerJ* **6**, e4251 (2018).
- Schulte, M. B., Draghi, J. A., Plotkin, J. B. & Andino, R. Experimentally guided models reveal replication principles that shape the mutation distribution of RNA viruses. *eLife* **4**, e03753 (2015).
- Brandenburg, B. *et al.* Imaging poliovirus entry in live cells. *PLoS Biol.* **5**, e183 (2007).
- Racaniello, V. R. Picornaviridae, the virus and their replication. *Fields Virol.* 795–838 (2007).
- Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M. & Belshaw, R. Viral mutation rates. *J. Virol.* **84**, 9733–9748 (2010).
- De La Torre, J. C., Giachetti, C., Semler, B. L. & Holland, J. J. High frequency of single-base transitions and extreme frequency of precise multiple-base reversion mutations in poliovirus. *Proc. Natl. Acad. Sci. USA* **89**, 2531–2535 (1992).
- Drake, J. W. & Holland, J. J. Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. USA* **96**, 13910–13913 (1999).
- De la Torre, J. C., Wimmer, E. & Holland, J. J. Very high frequency of reversion to guanidine resistance in clonal pools of guanidine-dependent type 1 poliovirus. *J. Virol.* **64**, 664–671 (1990).
- Liu, W., He, H. & Zheng, S.-Y. Microfluidics in single-cell virology: Technologies and applications. *Trends Biotechnol.* **38**(12), 1360–1372 (2020).
- Turgeon, M. *Immunology & Serology in Laboratory Medicine—E-Book* (Elsevier Health Sciences, 2013).
- Coffin, J. HIV population dynamics in vivo: Implications for genetic variation, pathogenesis, and therapy. *Science* **267**, 483–489 (1995).
- Martínez, J. L. & Baquero, F. Mutation frequencies and antibiotic resistance. *Antimicrob. Agents Chemother.* **44**, 1771–1777 (2000).
- Mansky, L. M. & Bernard, L. C. 3-Azido-3-deoxythymidine (AZT) and AZT-resistant reverse transcriptase can increase the in vivo mutation rate of human immunodeficiency virus type 1. *J. Virol.* **74**, 9532–9539 (2000).
- Bank, C. *et al.* An experimental evaluation of drug-induced mutational meltdown as an antiviral treatment strategy. *Evolution* **70**, 2470–2484 (2016).
- Crotty, S., Cameron, C. E. & Andino, R. RNA virus error catastrophe: Direct molecular test by using ribavirin. *Proc. Natl. Acad. Sci. USA* **98**, 6895–6900 (2001).
- Romanowski, E., Roba, L., Wiley, L., Araullo-Cruz, T. & Gordon, Y. The effects of corticosteroids on adenoviral replication. *Arch. Ophthalmol.* **114**, 581–585. <https://doi.org/10.1001/archophth.1996.01100130573014> (1996).
- Moog, C., Kuntz-Simon, G., Caussin-Schwemling, C. & Obert, G. Sodium valproate, an anticonvulsant drug, stimulates human immunodeficiency virus type 1 replication independently of glutathione levels. *J. Gen. Virol.* **77**, 1993–1999 (1996).
- Zhou, X. *et al.* Rapamycin and everolimus facilitate hepatitis E virus replication: Revealing a basal defense mechanism of PI3K-PKB-mTOR pathway. *J. Hepatol.* **61**, 746–754 (2014).
- Gutenkunst, R. N. *et al.* Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* **3**, e189 (2007).
- Poovathingal, S. K. & Gunawan, R. Global parameter estimation methods for stochastic biochemical systems. *BMC Bioinf.* **11**, 414 (2010).
- Lillacci, G. & Khammash, M. The signal within the noise: Efficient inference of stochastic gene regulation models using fluorescence histograms and stochastic simulations. *Bioinformatics* **29**, 2311–2319 (2013).
- Aguilera, L. U., Zimmer, C. & Kummer, U. A new efficient approach to fit stochastic models on the basis of high-throughput experimental data using a model of IRF7 gene expression as case study. *BMC Syst. Biol.* **11**, 26 (2017).
- McKinley, T. J. *et al.* Approximate Bayesian computation and simulation-based inference for complex stochastic epidemic models. *Stat. Sci.* **33**, 4–18 (2018).
- Ramji, R., Wong, V. C., Chavali, A. K., Gearhart, L. M. & Miller-Jensen, K. A passive-flow microfluidic device for imaging latent HIV activation dynamics in single t cells. *Integr. Biol.* **7**, 998–1010 (2015).
- Akpinar, F., Timm, A. & Yin, J. High-throughput single-cell kinetics of virus infections in the presence of defective interfering particles. *J. Virol.* **90**, 1599–1612 (2016).
- Rosenwasser, S. *et al.* Unmasking cellular response of a bloom-forming alga to viral infection by resolving expression profiles at a single-cell level. *PLoS Pathogens* **15**, e1007708 (2019).
- Nowakowski, T. J. *et al.* Expression analysis highlights AXL as a candidate Zika virus entry receptor in neural stem cells. *Cell Stem Cell* **18**, 591–596 (2016).

49. Golumbeanu, M. *et al.* Single-cell RNA-Seq reveals transcriptional heterogeneity in latent and reactivated Hiv-infected cells. *Cell reports* **23**, 942–950 (2018).
50. Rato, S., Rausell, A., Muñoz, M., Telenti, A. & Ciuffi, A. Single-cell analysis identifies cellular markers of the HIV permissive cell. *PLoS Pathog.* **13**, e1006678 (2017).
51. Lin, J. *et al.* Ultra-sensitive digital quantification of proteins and mRNA in single cells. *Nat. Commun.* **10**, 1–10 (2019).
52. Riordon, J., Sovilj, D., Sanner, S., Sinton, D. & Young, E. W. Deep learning with microfluidics for biotechnology. *Trends Biotechnol.* **37**, 310–324 (2019).
53. Mullen, K., Ardia, D., Gil, D. L., Windover, D. & Cline, J. Deoptim: An R package for global optimization by differential evolution. *J. Stat. Softw.* **40**, 1–26 (2011).
54. Boxuan, C. Dataexplorer: Data explorer. *R package* (2018). R package version 0.7.0.

## Acknowledgements

This work was supported by the National Institutes of Health Grant R01 AI120560 to C.E.C. and C.O.W., and by the National Institutes of Health Grant R01 GM088344 to C.O.W. A.I.T was funded by SFI.

## Author contributions

C.E.C and C.O.W were responsible for the funding and the conception of the project. W.L. and C.E.C were responsible for the acquisition of the data. A.I.T and C.O.W designed and conducted the computational analysis. A.I.T, J.A.D, and C.O.W wrote the paper and prepared the figures.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-87694-x>.

**Correspondence** and requests for materials should be addressed to A.I.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021