



**HAL**  
open science

## Number of infection events per cell during HIV-1 cell-free infection

Yusuke Ito, Azaria Remion, Alexandra Tauzin, Keisuke Ejima, Shinji Nakaoka, Yoh Iwasa, Shingo Iwami, Fabrizio Mammano

► **To cite this version:**

Yusuke Ito, Azaria Remion, Alexandra Tauzin, Keisuke Ejima, Shinji Nakaoka, et al.. Number of infection events per cell during HIV-1 cell-free infection. *Scientific Reports*, 2017, 7 (1), pp.6559. 10.1038/s41598-017-03954-9 . inserm-03451921

**HAL Id: inserm-03451921**

**<https://inserm.hal.science/inserm-03451921>**

Submitted on 26 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# SCIENTIFIC REPORTS



OPEN

## Number of infection events per cell during HIV-1 cell-free infection

Yusuke Ito<sup>1</sup>, Azaria Remion<sup>2,3,4</sup>, Alexandra Tauzin<sup>2,3,4</sup>, Keisuke Ejima<sup>5</sup>, Shinji Nakaoka<sup>6,7</sup>, Yoh Iwasa<sup>1</sup>, Shingo Iwami<sup>1,6,8</sup> & Fabrizio Mammano<sup>2,3,4</sup>

Received: 18 October 2016

Accepted: 9 May 2017

Published online: 26 July 2017

HIV-1 accumulates changes in its genome through both recombination and mutation during the course of infection. For recombination to occur, a single cell must be infected by two HIV strains. These coinfection events were experimentally demonstrated to occur more frequently than would be expected for independent infection events and do not follow a random distribution. Previous mathematical modeling approaches demonstrated that differences in target cell susceptibility can explain the non-randomness, both in the context of direct cell-to-cell transmission, and in the context of free virus transmission (Q. Dang *et al.*, Proc. Natl. Acad. Sci. USA 101:632-7, 2004; K. M. Law *et al.*, Cell reports 15:2711-83, 2016). Here, we build on these notions and provide a more detailed and extensive quantitative framework. We developed a novel mathematical model explicitly considering the heterogeneity of target cells and analysed datasets of cell-free HIV-1 single and double infection experiments in cell culture. Particularly, in contrast to the previous studies, we took into account the different susceptibility of the target cells as a continuous distribution. Interestingly, we showed that the number of infection events per cell during cell-free HIV-1 infection follows a negative-binomial distribution, and our model reproduces these datasets.

The Human Immunodeficiency Virus type-1 (HIV-1) population in an infected individual is characterized by high genetic diversity that allows rapid adaptation to the changing environment, such as the development of an immune response or the initiation of an antiretroviral therapy. Genetic recombination events participate in the continuous production of these viral variants. For recombination to take place, distinct viruses must infect the same cell, and then different genomes must be packaged into a single virion so that the reverse transcription process can generate a chimeric viral genome by template switching (reviewed in ref. 1). By mixing the viral genomes, in one step, recombination creates new variants whose adaptation to the environment may exceed those of the parental viruses<sup>2,3</sup>. This process could participate in the unfavourable prognosis of patients infected by two strains of HIV, known as double infection<sup>4</sup>. Although the majority of HIV-infected lymphocytes in the peripheral blood of patients carry only one viral genome copy<sup>5,6</sup>, the epidemiologic spread of circulating recombinant forms (CRF) of HIV-1 demonstrates that recombination, and thus double infections, take place in infected patients<sup>7</sup>. It was also shown that recombination in HIV-infected patients may rescue defective viral genomes that carry drug resistance mutations<sup>8,9</sup>.

The frequency of cells carrying multiple viral genomes is influenced by the virus transmission route. HIV-1 infection can spread either by cell-free virus particles or by a cell-associated process, in which viral particles and cellular receptors converge at the donor- and target-cell contact sites<sup>10,11</sup>. Previous work has established that cell-associated HIV-1 transmission leads to frequent multiple infection events, while the majority of cells infected by free virions carry a single genome<sup>12</sup>. The genomes transmitted by one infected cell via the cell-mediated pathway, however, are expected to be very similar, thus reducing the likelihood that recombination will produce chimeric variants with new properties following this transmission method. Interestingly, despite the difference in efficacy, both virus transmission pathways result in a higher frequency of double-infected cells than would be expected for independent transmission events, showing that these infections do not follow a random distribution<sup>13-17</sup>. Two previous studies proposed that differences in cell susceptibilities could justify the experimental

<sup>1</sup>Department of Biology, Faculty of Sciences, Kyushu University, Fukuoka, 819-0395, Japan. <sup>2</sup>INSERM, U941, 75010, Paris, France. <sup>3</sup>Univ Paris Diderot, Sorbonne Paris Cité, IUH, 75010, Paris, France. <sup>4</sup>Institut Universitaire d'Hématologie, Hôpital Saint-Louis, 75010, Paris, France. <sup>5</sup>School of Public Health, University of Alabama at Birmingham, Alabama, 35294-0011, USA. <sup>6</sup>PRESTO, JST, Saitama, 332-0012, Japan. <sup>7</sup>Institute of Industrial Science, The University of Tokyo, Meguro-ku, Tokyo, 153-8505, Japan. <sup>8</sup>CREST, JST, Saitama, 332-0012, Japan. Yusuke Ito, Azaria Remion, Shingo Iwami and Fabrizio Mammano contributed equally to this work. Correspondence and requests for materials should be addressed to S.I. (email: [siwami@kyushu-u.org](mailto:siwami@kyushu-u.org)) or F.M. (email: [fabrizio.mammano@inserm.fr](mailto:fabrizio.mammano@inserm.fr))

observation of double infection frequencies that are higher than expected for independent infection events<sup>13,18</sup>. In one study, the authors considered only either “susceptible” or “unsusceptible” cell populations, and using a mathematical model found that heterogeneity in target cell susceptibility could account for the observation that more double infections occur *in vivo* than predicted by random models<sup>18</sup>. In that report, the percentage of susceptible cells in a target cell population was estimated to be 2.76%. In the other study, the cell population was considered as composed of a discrete number of subpopulations characterized by distinct susceptibility levels, and for simplicity 5 subpopulations of the same size were considered<sup>13</sup>. Here, by considering susceptibility as a continuous variable, we expand on those original reports, and provide a more detailed quantitative framework. We describe a novel mathematical model that explicitly considers the heterogeneity of target cells as a continuous variable. By fitting the model to experimental datasets of cell-free HIV-1 single and double infections, we show that the number of infection events per cell follows a negative-binomial distribution. We also quantified the increase in the double and multiple infection events as a function of the amount of inoculated virus, and we found that a significant proportion of cells can be infected by multiple genomes following cell-free HIV-1 exposure. Together, our results re-evaluate the potential impact of cell-free HIV-1 infection on HIV-1 genetic recombination.

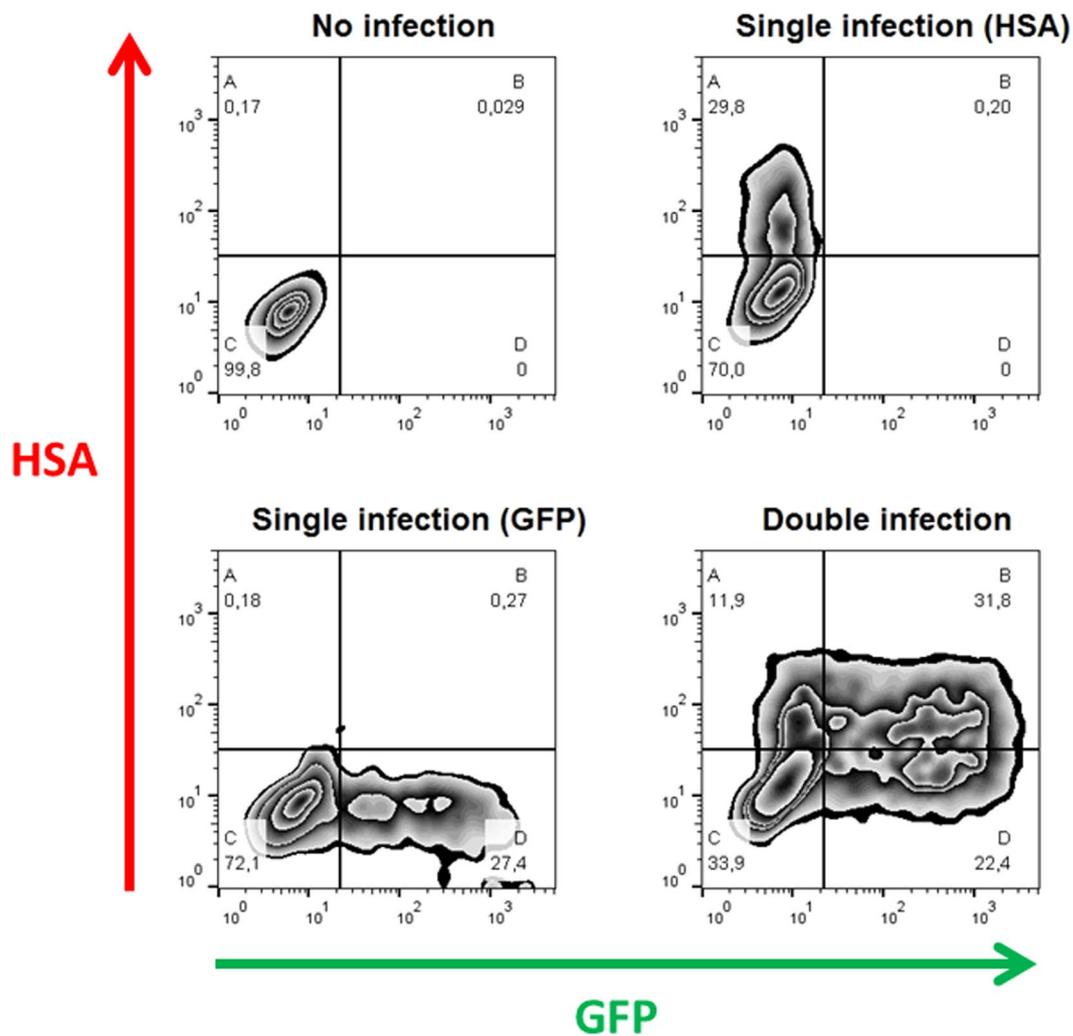
## Materials and Methods

**Cells and proviral plasmids.** HEK293T cells were maintained in Dulbecco modified Eagle’s medium (DMEM) supplemented with 10% heat-inactivated foetal calf serum (FCS) and antibiotics (100 IU/ml penicillin and 100 µg/ml streptomycin). MT4R5 cells<sup>19</sup> were grown in RPMI-1640 medium supplemented with 10% heat-inactivated FCS, 100 IU/ml of penicillin, 100 µg/ml of streptomycin, and 0.25 µg/ml of amphotericin B. All cultures were maintained at 37 °C in a humidified atmosphere with 5% CO<sub>2</sub>.

The proviral constructs used here were derived from previously published plasmids based on the pNL4-3 construct and each carried a sequence coding for either green fluorescent protein (GFP) or heat stable antigen (HSA) reporter proteins cloned before the *nef* gene, with an IRES sequence allowing concomitant expression of the viral and reporter proteins<sup>20,21</sup>. To prevent virus spread in culture, we have modified these constructs by deleting 1.3 kb of the *env* gene (between the *KpnI* and *BglIII* sites<sup>15</sup>). To complement these proviral constructs, we used an HIV-1 Env-expresser plasmid in which HIV-1 (pNL4-3) Env, (as well as Tat and Rev) expression is under the control of a chimeric SR $\alpha$  promoter (SV40-early promoter and LTR from HTLV-1)<sup>15</sup>.

**Preparation of virus stocks, infection, and datasets.** Stocks of viruses expressing either GFP or HSA were prepared by transfecting sub-confluent 293-T cells in T75 flasks by JetPei (Polyplus Inc. Illkirch, France), following the manufacturer’s instructions. Medium was changed 16 h later, and the virus-containing supernatant was collected 40 h post-transfection and overlaid on a 20% sucrose cushion in a Beckman SW32 tube, after which particles were pelleted by centrifugation (98,000 g, 4 °C) for 90 min. Viral pellets were re-suspended in RPMI medium with FCS to obtain a 10-fold concentration as compared with the initial amount present in the culture supernatant, separated into several aliquots, and frozen at –80 °C. One day before infection,  $2.0 \times 10^5$  MT4R5 cells per well were seeded in a 96-well plate. Cells were then exposed to two-fold dilutions of one virus (single infection) or of both viruses at the same time (coinfection), using the indicated combinations of each virus input amount. Two hours after infection, cells were washed to eliminate excess virus and cultured for a total of 48 h. The percentage of GFP-positive cells was measured after cell fixation with 2% para-formaldehyde (PFA). HSA expression on the surface of infected cells was detected using a rat anti-HSA PerCPy5.5 antibody (Pharmingen, Le Pont-de Claix, France) before fixation in PFA. Flow-cytometry data were acquired using a FACSCalibur instrument (Becton Dickinson, Le Pont-de Claix, France) with CellQuest software and were analysed using FlowJo software (Treestar, Ashland, OR, USA). Cell viability was measured in parallel and found to be stable for at least 72 h post-infection (data not shown).

**Calculation of the frequencies in different FACS quadrants.** Cells in a FACS graph can be divided into four quadrants, A, B, C, and D, based on their expression of GFP and/or HSA (Fig. 1). Cells in quadrant A express HSA only, those in quadrant B are positive both for HSA and GFP, cells in quadrant C are uninfected (i.e., negative both for HSA and GFP), and cells in quadrant D are positive for GFP only. Note that multiple infection events by viruses expressing the same reporter gene in quadrants A and D cannot be distinguished using FACS. Since for a given susceptibility parameter,  $s$ , the probability of a target cell being uninfected (i.e., no virus) is  $e^{-\beta s V}$ , the probability of a target cell being infected is  $1 - e^{-\beta s V}$ <sup>16</sup>. We assumed that the susceptibility parameter,  $s$ , obeys the Gamma distribution with the scale parameter,  $p$ , and the rate parameter,  $q$  (see Results for detailed calculations). Because two fluorescent proteins (i.e., HSA and GFP) are used, the term  $V$  is divided into  $\bar{V}_{\text{HSA}}$  and  $\bar{V}_{\text{GFP}}$ , which represent the amount of effective HIV-1 expressing HSA and GFP, respectively. Additionally, to consider the case of each inoculated HIV-1 dataset (see Preparation of virus stocks and infection), we assumed  $\bar{V}_{\text{HSA}} = V_{\text{HSA}}$  and  $\bar{V}_{\text{GFP}} = V_{\text{GFP}}$  for 3.12 µl of inoculated HIV-1 expressing HSA and GFP, respectively, and  $\bar{V}_{\text{HSA}} = V_{\text{HSA}} \times r_*$  and  $\bar{V}_{\text{GFP}} = V_{\text{GFP}} \times r_*$  for the other amounts of inoculated HIV-1. We estimated the distribution of the following 11 parameters ( $\theta$ ): the shape parameter,  $p$ , the composed parameters,  $\beta V_{\text{HSA}}/q$  and  $\beta V_{\text{GFP}}/q$ , for single HSA and GFP HIV-1 experiments, respectively, and the scaling parameters  $r_*$  for 6.25, 12.5, 25, 37, 50, 75, 100, and 200 µl of the inoculated HIV-1 (Table 1) (i.e.,  $\theta = \{p, \beta V_{\text{HSA}}/q, \beta V_{\text{GFP}}/q, r_{6.25}, r_{12.5}, r_{25}, r_{37}, r_{50}, r_{75}, r_{100}, r_{200}\}$ ). Therefore, under the assumptions, the theoretically predicted frequency of quadrant A, B, C, and D, respectively, in double HIV-1 infection experiments is calculated as follows:



**Figure 1.** Flow cytometry analysis of single and double HIV-1 infection. Panels represent the following conditions, clockwise starting from the upper left panel: no infection; infection by the HSA virus; coinfection with HSA and GFP viruses, infection by the GFP virus. In each panel, the quadrants correspond to HSA<sup>+</sup> (A); HSA<sup>+</sup>GFP<sup>+</sup> (B); uninfected cells (C); and GFP<sup>+</sup> (D). The percentage of cells in each quadrant is indicated under the letter identifying the quadrant.

Parameters	Estimated values (mean)	95% CI
$p$	1.176	0.878–1.512
$\beta\bar{V}_{\text{HSA}}/q$	$1.484 \times 10^{-2}$	$1.034 \times 10^{-2}$ – $1.963 \times 10^{-2}$
$\beta\bar{V}_{\text{GFP}}/q$	$2.115 \times 10^{-2}$	$1.486 \times 10^{-2}$ – $2.796 \times 10^{-2}$
$r_{6.25}$	2.417	1.487–3.344
$r_{12.5}$	4.231	2.588–5.812
$r_{25}$	4.352	2.683–6.026
$r_{37}$	11.09	7.100–15.16
$r_{50}$	8.115	5.322–10.97
$r_{75}$	12.79	8.485–17.42
$r_{100}$	16.75	11.82–22.16
$r_{200}$	26.87	17.49–36.81

**Table 1.** Estimated parameters in the mathematical model of cell-free infection.

$$\begin{aligned}
 F_{A,co}(\theta) &= \int_0^\infty (1 - e^{-\beta\bar{V}_{HSA}})e^{-\beta\bar{V}_{GFP}} \frac{q^p s^{p-1}}{\Gamma(p)} e^{-qs} ds, \\
 F_{B,co}(\theta) &= \int_0^\infty (1 - e^{-\beta\bar{V}_{HSA}})(1 - e^{-\beta\bar{V}_{GFP}}) \frac{q^p s^{p-1}}{\Gamma(p)} e^{-qs} ds, \\
 F_{C,co}(\theta) &= \int_0^\infty e^{-\beta\bar{V}_{HSA}} e^{-\beta\bar{V}_{GFP}} \frac{q^p s^{p-1}}{\Gamma(p)} e^{-qs} ds, \\
 F_{D,co}(\theta) &= \int_0^\infty e^{-\beta\bar{V}_{HSA}}(1 - e^{-\beta\bar{V}_{GFP}}) \frac{q^p s^{p-1}}{\Gamma(p)} e^{-qs} ds.
 \end{aligned}$$

We further calculated and simplified those equations as follows:

$$F_{A,co}(\theta) = \frac{1}{\left(1 + \frac{\beta}{q}\bar{V}_{GFP}\right)^p} - \frac{1}{\left(1 + \frac{\beta}{q}(\bar{V}_{GFP} + \bar{V}_{HSA})\right)^p}, \tag{1}$$

$$F_{B,co}(\theta) = 1 - \frac{1}{\left(1 + \frac{\beta}{q}\bar{V}_{HSA}\right)^p} - \frac{1}{\left(1 + \frac{\beta}{q}\bar{V}_{GFP}\right)^p} + \frac{1}{\left(1 + \frac{\beta}{q}(\bar{V}_{HSA} + \bar{V}_{GFP})\right)^p}, \tag{2}$$

$$F_{C,co}(\theta) = \frac{1}{\left(1 + \frac{\beta}{q}(\bar{V}_{HSA} + \bar{V}_{GFP})\right)^p}, \tag{3}$$

$$F_{D,co}(\theta) = \frac{1}{\left(1 + \frac{\beta}{q}\bar{V}_{HSA}\right)^p} - \frac{1}{\left(1 + \frac{\beta}{q}(\bar{V}_{HSA} + \bar{V}_{GFP})\right)^p}. \tag{4}$$

Notably, in the experiments with only HSA-expressing HIV-1 (i.e., single HSA HIV-1 experiments), we derived  $F_{A,r}(\theta) = 1 - 1/(1 + \beta\bar{V}_{HSA}/q)^p$ ,  $F_{B,r}(\theta) = 0$ ,  $F_{C,r}(\theta) = 1/(1 + \beta\bar{V}_{HSA}/q)^p$ , and  $F_{D,r}(\theta) = 0$ . Furthermore, in the single GFP HIV-1 experiments, we derived  $F_{A,g}(\theta) = 0$ ,  $F_{B,g}(\theta) = 0$ ,  $F_{C,g}(\theta) = 1/(1 + \beta\bar{V}_{GFP}/q)^p$ , and  $F_{D,g}(\theta) = 1 - 1/(1 + \beta\bar{V}_{GFP}/q)^p$  (see Data fitting, concerning the meaning of the index  $r, g, co$ ).

**Data fitting.** To fit the predicted frequency of each quadrant (i.e., Eqs (1–4)) with the experimental measurements by FACS analyses, we employed likelihood estimation to obtain an optimal set of parameter values. If we assume that the datasets used follow the Gaussian distribution with the mean  $\mu_i$  and the variance  $\sigma^2$  (i.e.,  $N(\mu_i, \sigma^2)$ ), the corresponding likelihood function is given as follows:

$$L(\rho; x) = \prod \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\mu-x)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\sum(\mu-x)^2}.$$

Note that  $\rho$  and  $x$  represent the set of parameters and the measurements, respectively. The specific form of sum of squared residuals (SSR) is given by

$$\begin{aligned}
 SSR(\theta) &= \sum_{r=1}^{10} \{(F_{A,r} - \tilde{F}_{A,r})^2 + (F_{C,r} - \tilde{F}_{C,r})^2\} + \sum_{g=1}^{10} \{(F_{C,g} - \tilde{F}_{C,g})^2 + (F_{D,g} - \tilde{F}_{D,g})^2\} \\
 &+ \sum_{co=1}^{18} \{(F_{A,co} - \tilde{F}_{A,co})^2 + (F_{B,co} - \tilde{F}_{B,co})^2 + (F_{C,co} - \tilde{F}_{C,co})^2 \\
 &+ (F_{D,co} - \tilde{F}_{D,co})^2\}.
 \end{aligned}$$

Here,  $\theta$  is the set of parameters needed to estimate.  $F_{quadrant,r,g,co}$  and  $\tilde{F}_{quadrant,r,g,co}$  ( $quadrant = \{A, B, C, D\}$ ) represent the predicted frequencies and the measurements by FACS analyses, respectively. The index  $r, g, co$  represent experiments with different amounts of inoculated HIV-1 (i.e.,  $r = 1, \dots, 10$  correspond to 3.12, 6.25, 12.5, 25, 37, 50, 75, 100, 100 and 200  $\mu$ l, respectively, of single HSA HIV-1 experiments;  $g = 1, \dots, 10$  correspond to 3.12, 6.25, 12.5, 25, 50, 50, 75, 100, 100 and 200  $\mu$ l, respectively, of single GFP HIV-1 experiments;  $co = 1, \dots, 18$  correspond to 25 and 25, 25 and 50, 50 and 25, 37 and 50, 50 and 50, 50 and 37 and 75, 75 and 50, 75 and 25 and 100, 100 and 25, 37 and 100, 50 and 100, 100 and 50, 100 and 50, 75 and 100, 100 and 75, 100 and 100  $\mu$ l and 100 and 100  $\mu$ l, respectively, of double HSA and GFP HIV-1 experiments). If the variance  $\sigma^2$  is constant, then likelihood estimation is completely determined by the SSR between theoretical values and measurements (i.e.,  $\sum(\mu - x)^2$ ). However, to take into account for possible variations, we assume that the variance is not constant, and employed Bayesian inference approach with the Markov Chain Monte Carlo (MCMC) method to estimate distribution of parameters.

**Bayesian inference method for the parameter estimation.** The R package FME enables one to perform MCMC sampling by “delayed rejection and adaptive Metropolis algorithm”<sup>22</sup>. In the framework of the FME package to perform Bayesian inference, the error between observations and model predictions is assumed to follow Gaussian distribution with the mean 0 and the variance  $\sigma^2$ . Moreover, the reciprocal of the variance (i.e.,  $1/\sigma^2$ ) follows a Gamma distribution, while the prior distribution of all parameters is a Gaussian distribution<sup>22</sup>. In this study, 120,000 MCMC samples were generated, and the first 20,000 chains were discarded as burn-in samples. Convergence of the Markov chain was manually checked by the output of the ‘traceplot’ and the histogram of the posterior distribution. As shown in Table 1, the 95% CI (credible interval) represents the range from 2.5% to 97.5% in each estimated distribution, and the mean value represents the one of each posterior distribution, and also the estimated posterior distributions are obtained by using the same random seed. The fits of Eqs (1–4) to the experimental data with different amounts of inoculated HIV-1 are shown in Fig. 2 and Fig. S1. Our estimated parameter values are summarized in Table 1.

## Results

### The number of infection events during cell-free infection follows a negative-binomial distribution.

It was previously demonstrated that cells infected by more than one virus occur at a frequency higher than that expected by Poisson distribution<sup>12, 13, 15–17</sup>. Here, we developed a novel mathematical model explicitly considering the heterogeneity of target cells susceptibility to infection. In our model, we define the following parameters:  $V$  is the amount of effective virus for infection events,  $\beta$  is the infection rate of HIV-1, and  $s$  is the susceptibility of the target cells to HIV-1 infection. The probability of a target cell being infected (i.e., carrying and expressing an integrated HIV genome) by  $n$  viruses can be determined by Poisson distribution as previously described<sup>12, 16, 17</sup>:

$$f(X = n; \beta s V) = \frac{(\beta s V)^n e^{-\beta s V}}{n!}. \quad (5)$$

Additionally, to consider the heterogeneity of target cell susceptibility<sup>13, 15</sup>, we assumed that the susceptibility parameter,  $s$ , obeys the following Gamma distribution<sup>23</sup>:

$$g(s; p, q) = \frac{q^p s^{p-1} e^{-sq}}{\Gamma(p)}. \quad (6)$$

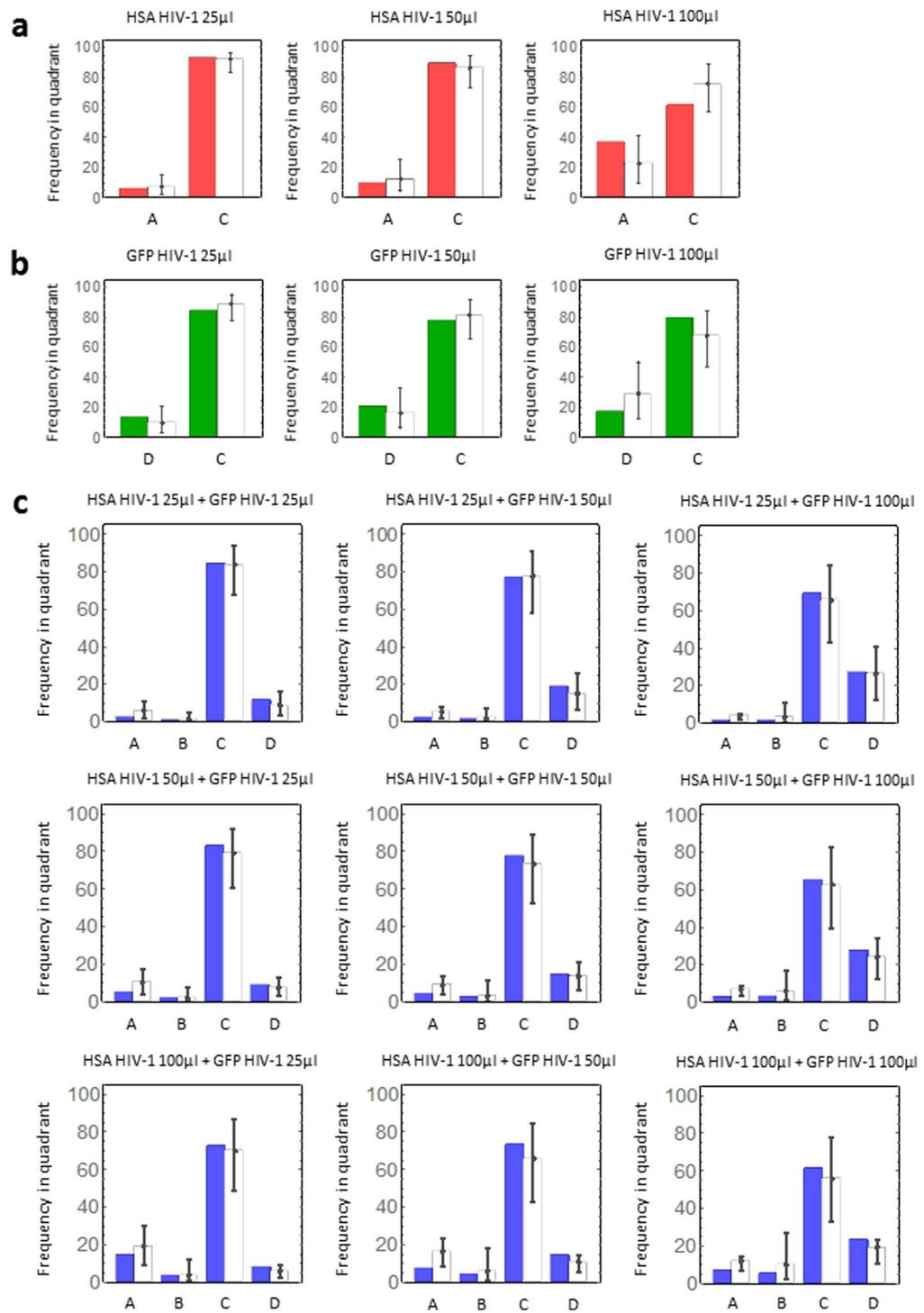
It is well known that Gamma distribution can approximate any one-peak distribution and reproduce a variety of biological phenomena<sup>23–25</sup>. Here  $p > 0$  and  $q > 0$  are the shape and rate parameters, and  $\Gamma(\cdot)$  is the gamma function. In a previous study<sup>13</sup>, it was artificially assumed that there are a finite number of subpopulations of cells with different susceptibilities to infection (i.e., five discrete susceptible populations). We extended that assumption to a continuous range of susceptible populations allowing our model to reflect, for example, the level of expression of CD4 and/or co-receptors on target cells, which are widely but continuously distributed<sup>14, 26, 27</sup>. Using Eqs (5, 6), we calculated the probability density function for a cell being infected by  $n$  HIV-1 in a heterogeneous target cell population:

$$\begin{aligned} Pr \left( X = n; p, 1/\left(1 + \frac{\beta V}{q}\right) \right) &= \int_0^\infty \frac{(\beta s V)^n e^{-(\beta s V)}}{n!} \frac{q^p s^{p-1}}{\Gamma(p)} e^{-qs} ds, \\ &= (\beta V)^n q^p \frac{1}{n! \Gamma(p)} \int_0^\infty s^{n+p-1} e^{-(\beta V + q)s} ds, \\ &= \frac{\Gamma(n+p)}{n! \Gamma(p)} \left( \frac{q}{\beta V + q} \right)^p \left( \frac{\beta V}{\beta V + q} \right)^n, \\ &= \binom{n+p-1}{n} \left( \frac{1}{1 + \frac{\beta V}{q}} \right)^p \left( \frac{\frac{\beta V}{q}}{1 + \frac{\beta V}{q}} \right)^n. \end{aligned} \quad (7)$$

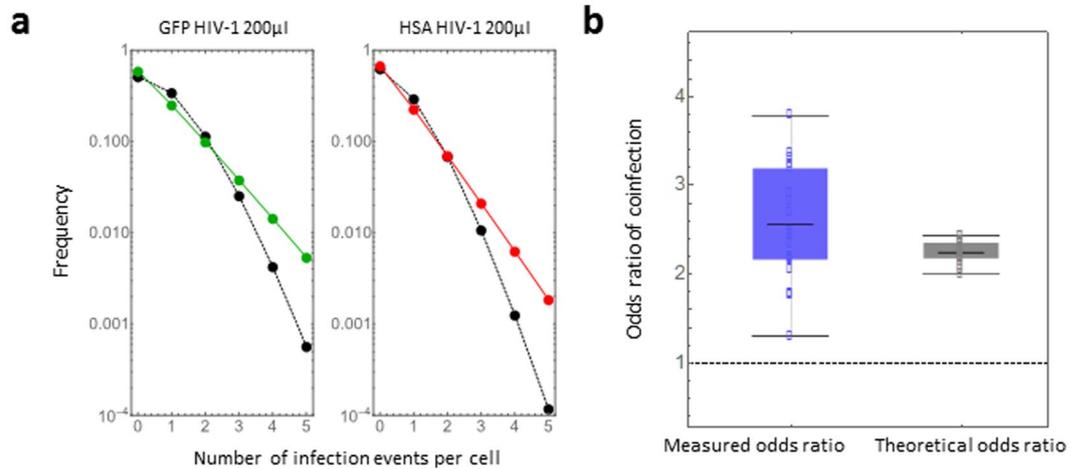
Therefore, the number of HIV-1 infection events per cell during cell-free HIV-1 infection (i.e.,  $Pr(X = n)$ ) follows a negative-binomial distribution of mean  $\beta p V/q$  and variance  $\beta p V/q(1 + \beta V/q)$ .

We estimated the parameters in Eq. (7) by fitting Eqs (1–4) to the experimental measured frequencies of quadrants A, B, C, and D in our FACS analyses (see MATERIALS AND METHODS), and these values are summarized in Table 1. A set of representative analyses is shown in Fig. 2. In these panels, the coloured and white bars represent experimental measurements and theoretical predictions, respectively. In both single (Fig. 2a,b) and double HIV-1 infection experiments (Fig. 2c), our mathematical model reproduces all experimental datasets well. An independent set of data and the corresponding analysis are shown in Supplemental Figure 1, both for single and double infection experiments.

The expected negative-binomial distributions of the number of infection events per cell in 200  $\mu$ l for GFP and HSA HIV-1 single experiments are shown in green and red curves, respectively, as examples in Fig. 3a. The black curves represent the expected Poisson distribution with the mean of the Gamma distributed susceptibility parameter,  $s$  (i.e., the target cell population is assumed to have homogeneous susceptibility). While the mean and variance of a Poisson distribution are the same, the variance of a negative-binomial distribution is larger than its mean. This property of negative-binomial distribution explains that the more susceptible one cell is, the more effectively it will be infected by HIV-1<sup>13, 14</sup>.



**Figure 2.** Frequency of single infection and coinfection: **(a)** The experimental and theoretical frequencies of quadrants A (i.e., HSA-positive) and C (i.e., HSA-negative) in three independent experiments using only HSA HIV-1 are shown by red and white bars, respectively. **(b)** The experimental and theoretical frequencies of quadrants D (i.e., GFP-positive) and C (i.e., GFP-negative) in single GFP HIV-1 experiments are shown by green and white bars, respectively. **(c)** The experimental and theoretical frequencies of quadrants A (i.e., HSA-positive), B (i.e., positive both for HSA and GFP), C (i.e., negative both for HSA and GFP), and D (i.e., GFP-positive) in double HIV-1 experiments are shown by blue and white bars, respectively. Two independent experiments were run with the nine indicated combinations of HSA and GFP HIV-1 corresponding to all possible combinations of the three different amounts of each virus used in single experiments. Note that each error bar represents the 95% credible interval obtained from Markov Chain Monte Carlo (MCMC) parameter inferences (Table 1).



**Figure 3.** Frequency of multiple infection events per cell: (a) The expected negative-binomial distributions of the number of infection events per cell in 200  $\mu\text{l}$  in GFP and HSA HIV-1 single experiments are shown in green and red curves, respectively. These curves were drawn using the mean derived from MCMC parameter inferences (Table 1). The black curves represent the expected Poisson distribution with the mean of the Gamma-distributed susceptibility parameter,  $s$  (i.e., the target cell population is assumed to have homogeneous susceptibility). (b) The experimental odds ratio ( $OR_E$ ) and theoretical odds ratio ( $OR_M$ , calculated by our estimated parameters) are shown in blue and black box plots, respectively. The dotted line corresponds to the odds ratio of 1 predicted by a Poisson distribution (i.e.,  $OR_M = 1$ : random HIV-1 infection).

**Calculation of the odds ratio in a cell-free HIV-1 infection.** The frequency of co-infected cells with HIV-1 expressing HSA and GFP has previously been quantified by calculating the odds ratios<sup>13–15</sup>. The odds of HSA-positive cells being GFP-positive can be calculated by

$$\frac{[\tilde{F}_B/(\tilde{F}_A + \tilde{F}_B)]/\{1 - [\tilde{F}_B/(\tilde{F}_A + \tilde{F}_B)]\}}{[\tilde{F}_B/\tilde{F}_A]} = \tilde{F}_B/\tilde{F}_A,$$

while the odds of HSA-negative cells being GFP-positive can be calculated by  $\tilde{F}_D/\tilde{F}_C$ . If the coinfection were random (i.e., independent events), then the  $\tilde{F}_B/\tilde{F}_A$  and  $\tilde{F}_D/\tilde{F}_C$  would be expected to be equal to 1, that is, the experimental odds ratio

$$OR_E = (\tilde{F}_B/\tilde{F}_A)/(\tilde{F}_D/\tilde{F}_C) = \tilde{F}_B\tilde{F}_C/\tilde{F}_A\tilde{F}_D = 1.$$

If coinfection occurred more or less frequently than that expected from random events, then the expected odds ratio would be  $OR_E > 1$  and  $OR_E < 1$ , respectively. A higher frequency of coinfection has been experimentally confirmed in independent reports<sup>13–15</sup>. To study whether or not our novel model quantitatively reproduces this important property of HIV-1 coinfection, we derived the theoretical odds ratio,  $OR_M$ , from Eqs (1–4) as follows:

$$OR_M = \frac{F_B(\theta)F_C(\theta)}{F_A(\theta)F_D(\theta)} = 1 + \frac{\frac{1}{\left(1 + \frac{\beta}{q}(\bar{V}_{HSA} + \bar{V}_{GFP})\right)^p} - \frac{1}{\left(1 + \frac{\beta}{q}\bar{V}_{HSA}\right)^p \left(1 + \frac{\beta}{q}\bar{V}_{GFP}\right)^p}}{\left\{ \frac{1}{\left(1 + \frac{\beta}{q}\bar{V}_{GFP}\right)^p} - \frac{1}{\left(1 + \frac{\beta}{q}(\bar{V}_{HSA} + \bar{V}_{GFP})\right)^p} \right\} \left\{ \frac{1}{\left(1 + \frac{\beta}{q}\bar{V}_{HSA}\right)^p} - \frac{1}{\left(1 + \frac{\beta}{q}(\bar{V}_{HSA} + \bar{V}_{GFP})\right)^p} \right\}}.$$

The second term is always positive (see Supplementary Note), which implies that  $OR_M > 1$  for all arbitrary parameter values. Therefore, under the condition of heterogeneous target cell susceptibility, our model always predicts that non-random HIV-1 coinfection occurs more frequently than would be expected for independent infection events (i.e.,  $OR_M > 1$ ). Notably, if the target cells have homogeneous susceptibilities (that is, the susceptibility parameter,  $s$ , is not distributed but fixed), then our model converges to a Poisson distribution and  $OR_M = 1$ . In Fig. 3b, we compared the odds ratio measured by our experiments,  $OR_E$ , with the theoretical odds ratio,  $OR_M$ , which was calculated by our estimated parameters (Table 1). Thus, our novel model quantitatively reproduces the odds ratio and captures the known property of coinfection during cell-free HIV-1 infection *in vitro*.

**Quantification of infection events during cell-free HIV-1 infection.** As discussed in previous work<sup>12,16</sup>, some multiple infection events cannot be detected by FACS analyses because cells that are infected with one copy of HIV-1 expressing HSA and those carrying two copies of the same viral genome are similarly

HSA-positive. Therefore, FACS analysis usually underestimates the true frequency of multiple infection<sup>12</sup>. As we derived  $F_{quadrant,r,g,co}(r = 1, \dots, 10)(g = 1, \dots, 10)(co = 1, \dots, 18)$  in the MATERIALS AND METHODS, we calculated the number of infection events during cell-free HIV-1 infection, using Eqs (1–4) and our estimated parameters in Table 1. The frequency of cells infected by multiple HSA or GFP virions was calculated by the following equation, in which  $n_{HSA}$  and  $n_{GFP}$  are the number of infection events with HSA and GFP virions, respectively:

$$\int_0^\infty \frac{(\beta s V_{HSA})^{n_{HSA}} e^{-\beta s V_{HSA}}}{n_{HSA}!} \frac{(\beta s V_{GFP})^{n_{GFP}} e^{-\beta s V_{GFP}}}{n_{GFP}!} \frac{q^p s^{p-1}}{\Gamma(p)} e^{-qs} ds.$$

The estimated multiple infection frequencies in double HIV-1 infection experiments are shown in Fig. 4a (for other combinations of multiple infection, the frequency is less than 0.001). For example, in the experiment with 100  $\mu$ l each of HSA and GFP HIV-1, although the experimentally measured frequency of coinfection was 5.47% (Fig. 2c, lower right panel, blue bar in B), our model revealed that 18.0% of the target cells were multiply infected (i.e.,  $1 - (0.57 + 0.10 + 0.15) = 0.18$ ). Our estimated value during cell-free infection is smaller than the values previously estimated (21% in ref. 12) during cell-to-cell HIV-1 infection in cell culture. This analysis further supports that cell-to-cell infection enhances multiple infection events as compared with cell-free infection. In Fig. 4b, we calculated the mean frequency of multiple infection events following incubation with different amounts of HIV-1 in both single and double infection experiments. The marks  $\blacktriangle$ ,  $\blacklozenge$ ,  $\bullet$ , and  $\blacksquare$  show the mean frequencies of zero, one, two, and three infection events per cell, respectively (data not shown for four or more infections events). As the amount of inoculated virus increases, the frequencies of multiple events consistently increase, and those of uninfected cells decrease. Interestingly, for these numbers of infection events, the frequencies reach steady state values around 150  $\mu$ l of inoculated HIV-1. Thus, our quantitative analyses reveal the true frequency of multiple infection events and demonstrate that cell-free HIV-1 infection by itself induces multiple infection events. Furthermore, taking advantage of our modelling approach, we could estimate the mean number of infection events per infected cell during cell-free HIV-1 infection. In Fig. 4c, we found that the number increases from 1.02 to 1.65 as the amount of inoculated HIV-1 increases. Our estimated range is consistent with the previous observation of proviral copy number in infected cells measured by fluorescence *in situ* hybridization<sup>12</sup>.

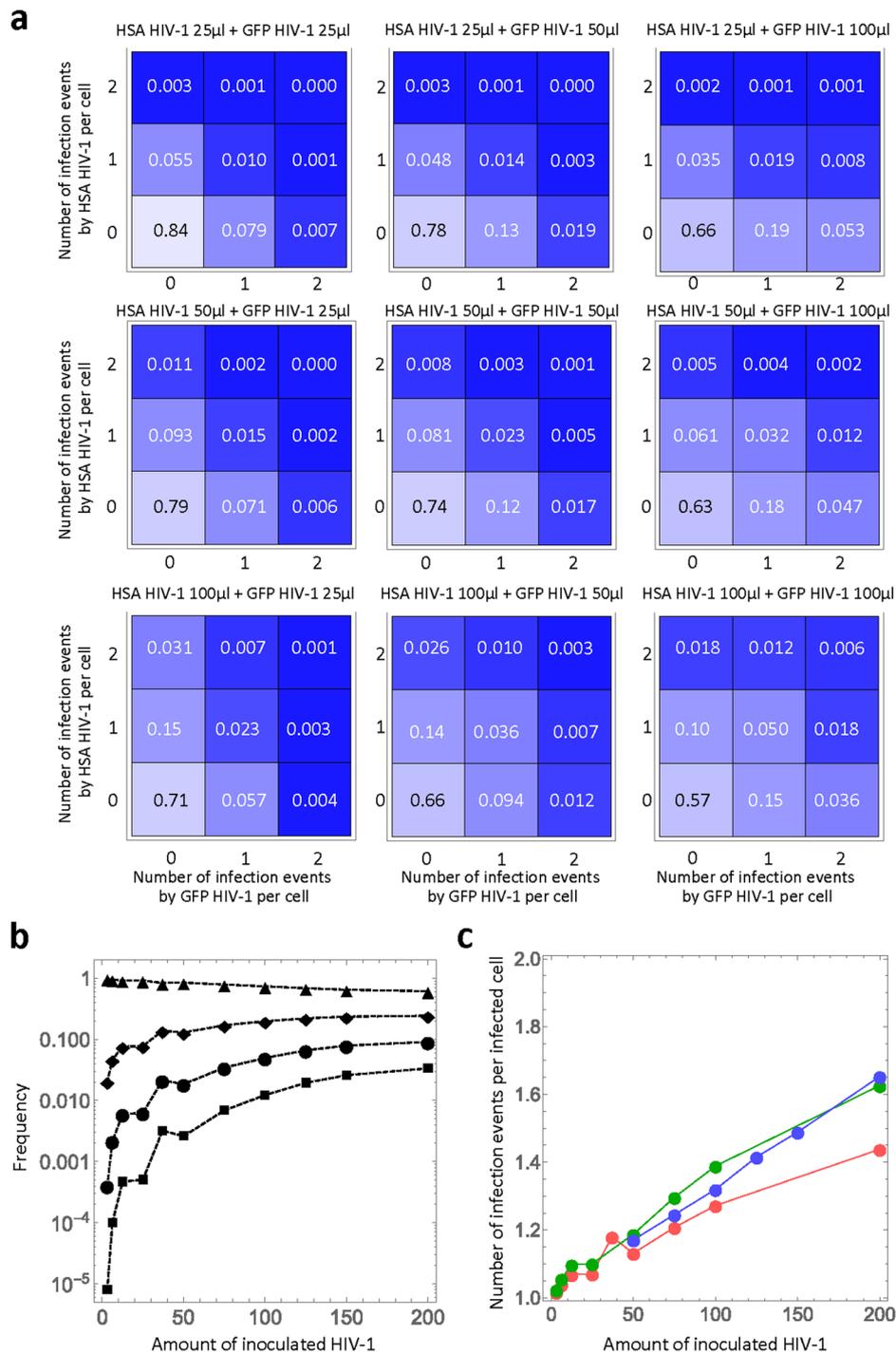
## Discussion

In this study, we modelled the distribution of HIV infection events during cell-free infection *in vitro*, taking into account differences in susceptibility within the target cell population. Our model fits well with our collected experimental data for both single and double infections, indicating that the assumptions and the mathematical formulation successfully capture the biological processes underlying the distribution of HIV-1 infection events. More importantly, our mathematical model describes the hypothesis that variation in target cell susceptibility could account for non-random co-infection more accurately than previous work.

Two previous reports suggested that differences in cell susceptibilities could be responsible for the observed higher frequency of double infections, as compared to predictions based on the assumption of independent infection events<sup>13, 18</sup>. In those reports, cell susceptibility was either considered as a binary feature (cells were either susceptible or non-susceptible)<sup>18</sup>, or a limited number of subpopulations characterized by distinct susceptibility levels were considered<sup>13</sup>. Here, by considering susceptibility as a continuous variable, we expand on those original reports. The idea that the susceptibility of target cells is continuously distributed is intuitive because, even in cultured cells, it should be slightly different due to, for example, the change in the expression of (co-)receptors on each cell membrane over time. Employing this condition, we showed that the theoretical odds ratio ( $OR_M$ ) is always greater than 1 (see Fig. 3 and Supplementary Note). This result is consistent with the findings of previous work<sup>13–15</sup>. Hence, considering susceptibility as a continuous variable seems to be a more appropriate assumption for describing the cooperative nature of the HIV-1 infection process, for which several quantitative and qualitative parameters (number of receptors, availability of nucleotide pool, phase of the cell cycle, etc.) participate in defining the susceptibility of each cell within a population.

The approach described here may also be customized to describe other biological situations that display similar properties, for instance a distribution of eclipse period of virus-infected cells<sup>24, 28</sup>. Note that our model does not restrict the analysis to a situation in which the two events (infection by the GFP and the HSA virus, in our study) have the same efficiency. Indeed, for each virus we estimated a composed parameter ( $\beta \bar{V}_{HSA}/q$  or  $\beta \bar{V}_{GFP}/q$  in Table 1) to express the effective virus dose for a given volume of different viruses (e.g., 3.12  $\mu$ l of inoculated HIV-1 expressing HSA or GFP). This choice increases the flexibility of our approach and allows an extension of its range of potential applications. Furthermore, we performed the experiments under conditions compatible with a large proportion of cells remaining uninfected to prevent saturation of the system and the potential associated biases. Of note, MT4 cells and their derivatives are highly susceptible to HIV infection. A gradual increase in the percentage of infected cells is observed when they are exposed to increasing virus doses, including conditions in which the majority of cells are infected. This feature allowed testing a wide range of experimental conditions, assisting the development of our model. Additionally, the model can be customized to work with less susceptible cells by experimentally determining the values of the parameters,  $p$  and  $q$ , because these two parameters are strongly associated with the susceptibility in a target cell population.

Our model allows estimation of the frequency of single and multiple infection events for individual virus inputs. As shown in Fig. 3a (for 200  $\mu$ l of each virus), our model predicts values that differ from the Poisson distribution; in particular, it predicts higher frequencies of multiple infection events. Also, our model predicts an  $OR_M$  of coinfection that is always  $>1$ , reproducing the experimental observations from different groups<sup>13–17</sup>, while the Poisson distribution predicts an  $OR_M = 1$ . Finally, having derived the values of the relevant parameters,



**Figure 4.** Quantitative analyses of multiple infection: **(a)** The distribution of the number of infection events per cell in double HIV-1 infection experiments with nine different combinations of virus amounts are shown. The number in each square is the estimated frequency of the corresponding infection events. **(b)** The mean frequency of the multiple infection events per cell inoculated with each different amount of HIV-1 is calculated. For double infection experiments, the amount is defined as the total inoculation of HSA and GFP HIV-1. The marks  $\blacktriangle$ ,  $\blacklozenge$ ,  $\bullet$ , and  $\blacksquare$  show the mean frequencies of zero, one, two, and three infection events per cell, respectively. **(c)** The estimated mean number of infection events per infected cell with different amounts of HIV-1 is calculated. The red, green, and blue curves correspond to the experiments with HSA and GFP HIV-1 single infections, and double infections, respectively. In **(a)**, **(b)**, and **(c)**, these calculations were all performed using the mean value obtained from MCMC parameter inferences (Table 1).

our model allows the estimation of the frequency of multiple infections with the same virus, which are events that could not be experimentally determined in our system. Indeed, infection with one or more viruses carrying the same tag will produce similar distributions in the resulting FACS plots, preventing their experimental discrimination. In Fig. 4b, we quantified the frequency of multiple infection events for different amounts of inoculated HIV-1. As the inoculated amount increases, a significant proportion of cells can be infected by multiple viruses (e.g. ●/■). This demonstrates that cell-free HIV-1 infection by itself may have an important impact on driving the recombination of viruses. The range of infection events per infected cell predicted by our model in Fig. 4c fits the range of previously experimentally measured multiple infection events produced by cell-free HIV-1 infection using *in situ* hybridization<sup>12</sup>. Taken together, our findings and predictions lead to a more detailed understanding of the link between co-infection events and recombination.

An alternative mechanism was previously proposed to explain the disproportionate frequency of double-infected cells observed during co-infection experiments. The authors proposed that otherwise silent infection events were detected as a consequence of the additional Tat expression induced by the second virus<sup>17</sup>. We have previously demonstrated that in our experimental system this potential artefact did not play a role, since the use of lentiviral vectors that only expressed GFP resulted in similarly high frequencies of double-infected cells as those obtained using vectors that also encode Tat<sup>15</sup>.

In addition to these *in vitro* experimental and theoretical analyses, other mathematical models and computer simulations have been proposed to explain the observation of multiple infections *in vivo*. For instance, it was described that CD4<sup>+</sup> T cells from the spleen of HIV-infected individuals carry on average 3.2 HIV proviral copies<sup>29,30</sup>. Also, the number of multiply infected cells was found to correlate with the square of the overall number of infected cells<sup>21,31,32</sup> in homogeneous target cell populations.

In agreement with previous reports<sup>12</sup>, we show here that multiple infection events take place with measurable frequencies following incubation with cell-free HIV-1 particles in a heterogeneous target cell population. Although the alternative pathway of infection that relies on cell contact-mediated infection is a more powerful transmission means, the impact of cell-free virus coinfection on HIV-1 genetic diversity may be expected to be more substantial because of the likelihood that spatially separated cells harbour genetically distinct variants. The impact on HIV-1 diversity, and consequently on the potential to adapt to a changing environment, are expected to correlate with the genetic distance between the parental variants. In the absence of antiretroviral treatment, the recurrent exposure of cells to infectious virions over the course of several years during this chronic infection creates numerous occasions for coinfections to happen. In view of these considerations, infection by cell-free virions emerges as a relevant means of HIV-1 diversification.

## References

- Burke, D. S. Recombination in HIV: an important viral evolutionary strategy. *Emerg Infect Dis* **3**, 253–259, doi:10.3201/eid0303.970301 (1997).
- Charpentier, C., Nora, T., Tenaillon, O., Clavel, F. & Hance, A. J. Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients. *J Virol* **80**, 2472–2482, doi:10.1128/jvi.80.5.2472-2482.2006 (2006).
- Nora, T. *et al.* Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. *J Virol* **81**, 7620–7628, doi:10.1128/jvi.00083-07 (2007).
- Sagar, M. *et al.* Infection with multiple human immunodeficiency virus type 1 variants is associated with faster disease progression. *J Virol* **77**, 12921–12926 (2003).
- Josefsson, L. *et al.* Majority of CD4<sup>+</sup> T cells from peripheral blood of HIV-1-infected individuals contain only one HIV DNA molecule. *Proc Natl Acad Sci USA* **108**, 11199–11204, doi:10.1073/pnas.1107729108 (2011).
- Josefsson, L. *et al.* Single cell analysis of lymph node tissue from HIV-1 infected patients reveals that the majority of CD4<sup>+</sup> T-cells contain one HIV-1 DNA molecule. *PLoS Pathog* **9**, e1003432, doi:10.1371/journal.ppat.1003432 (2013).
- Allen, T. M. & Altfeld, M. HIV-1 superinfection. *J Allergy Clin Immunol* **112**, 829–835, quiz 836, 10.1016/j.jaci.2003.08.037 (2003).
- Donahue, D. A., Bastarache, S. M., Sloan, R. D. & Wainberg, M. A. Latent HIV-1 can be reactivated by cellular superinfection in a Tat-dependent manner, which can lead to the emergence of multidrug-resistant recombinant viruses. *J Virol* **87**, 9620–9632, doi:10.1128/jvi.01165-13 (2013).
- Quan, Y., Liang, C., Brenner, B. G. & Wainberg, M. A. Multidrug-resistant variants of HIV type 1 (HIV-1) can exist in cells as defective quasispecies and be rescued by superinfection with other defective HIV-1 variants. *J Infect Dis* **200**, 1479–1483, doi:10.1086/606117 (2009).
- Jolly, C. & Sattentau, Q. J. Retroviral spread by induction of virological synapses. *Traffic* **5**, 643–650, doi:10.1111/j.1600-0854.2004.00209.x (2004).
- Chen, P., Hubner, W., Spinelli, M. A. & Chen, B. K. Predominant mode of human immunodeficiency virus transfer between T cells is mediated by sustained Env-dependent neutralization-resistant virological synapses. *J Virol* **81**, 12582–12595, doi:10.1128/jvi.00381-07 (2007).
- Del Portillo, A. *et al.* Multiploid inheritance of HIV-1 during cell-to-cell infection. *J Virol* **85**, 7169–7176, doi:10.1128/jvi.00231-11 (2011).
- Dang, Q. *et al.* Nonrandom HIV-1 infection and double infection via direct and cell-mediated pathways. *Proc Natl Acad Sci USA* **101**, 632–637, doi:10.1073/pnas.0307636100 (2004).
- Chen, J. *et al.* Mechanisms of nonrandom human immunodeficiency virus type 1 infection and double infection: preference in virus entry is important but is not the sole factor. *J Virol* **79**, 4140–4149, doi:10.1128/jvi.79.7.4140-4149.2005 (2005).
- Remion, A., Delord, M., Hance, A. J., Saragosti, S. & Mammano, F. Kinetics of the establishment of HIV-1 viral interference and comprehensive analysis of the contribution of viral genes. *Virology* **487**, 59–67, doi:10.1016/j.virol.2015.09.028 (2016).
- Haqqani, A. A. *et al.* Central memory CD4<sup>+</sup> T cells are preferential targets of double infection by HIV-1. *Virol J* **12**, 184, doi:10.1186/s12985-015-0415-0 (2015).
- Bregnard, C., Pacini, G., Danos, O. & Basmaciogullari, S. Suboptimal provirus expression explains apparent nonrandom cell coinfection with HIV-1. *J Virol* **86**, 8810–8820, doi:10.1128/jvi.00831-12 (2012).
- Law, K. M. *et al.* *In Vivo* HIV-1 Cell-to-Cell Transmission Promotes Multicopy Micro-compartmentalized Infection. *Cell reports* **15**, 2771–2783, doi:10.1016/j.celrep.2016.05.059 (2016).
- Amara, A. *et al.* G protein-dependent CCR5 signaling is not required for efficient infection of primary T lymphocytes and macrophages by R5 human immunodeficiency virus type 1 isolates. *J Virol* **77**, 2550–2558 (2003).

20. Imbeault, M., Lodge, R., Ouellet, M. & Tremblay, M. J. Efficient magnetic bead-based separation of HIV-1-infected cells using an improved reporter virus system reveals that p53 up-regulation occurs exclusively in the virus-expressing cell population. *Virology* **393**, 160–167, doi:10.1016/j.virol.2009.07.009 (2009).
21. Levy, D. N., Aldrovandi, G. M., Kutsch, O. & Shaw, G. M. Dynamics of HIV-1 recombination in its natural target cells. *Proc Natl Acad Sci USA* **101**, 4204–4209, doi:10.1073/pnas.0306764101 (2004).
22. Soetaert, K. & Petzoldt, T. Inverse modelling, sensitivity and monte carlo analysis in R using package FME. *Journal of Statistical Software* **33**, 1–28 (2010).
23. MacDonald, N., Cannings, C. & Hoppensteadt, F. C. *Biological delay systems: linear stability theory*. (Cambridge University Press, 2008).
24. Kakizoe, Y. *et al.* A method to determine the duration of the eclipse phase for *in vitro* infection with a highly pathogenic SHIV strain. *Sci Rep* **5**, 10371, doi:10.1038/srep10371 (2015).
25. Bliss, C. I. & Fisher, R. A. Fitting the negative binomial distribution to biological data. *Biometrics* **9**, 176–200 (1953).
26. Kabat, D., Kozak, S. L., Wehrly, K. & Chesebro, B. Differences in CD4 dependence for infectivity of laboratory-adapted and primary patient isolates of human immunodeficiency virus type 1. *J Virol* **68**, 2570–2577 (1994).
27. Platt, E. J., Wehrly, K., Kuhmann, S. E., Chesebro, B. & Kabat, D. Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. *J Virol* **72**, 2855–2864 (1998).
28. Pinilla, L. T., Holder, B. P., Abed, Y., Boivin, G. & Beauchemin, C. A. The H275Y neuraminidase mutation of the pandemic A/H1N1 influenza virus lengthens the eclipse phase and reduces viral output of infected cells, potentially compromising fitness in ferrets. *J Virol* **86**, 10651–10660, doi:10.1128/jvi.07244-11 (2012).
29. Jung, A. *et al.* Recombination: Multiply infected spleen cells in HIV patients. *Nature* **418**, 144, doi:10.1038/418144a (2002).
30. Dixit, N. M. & Perelson, A. S. Multiplicity of human immunodeficiency virus infections in lymphoid tissue. *J Virol* **78**, 8942–8945, doi:10.1128/jvi.78.16.8942-8945.2004 (2004).
31. Wodarz, D. & Levy, D. N. Effect of different modes of viral spread on the dynamics of multiply infected cells in human immunodeficiency virus infection. *J R Soc Interface* **8**, 289–300, doi:10.1098/rsif.2010.0266 (2011).
32. Dixit, N. M. & Perelson, A. S. HIV dynamics with multiple infections of target cells. *Proc Natl Acad Sci USA* **102**, 8198–8203, doi:10.1073/pnas.0407498102 (2005).

## Acknowledgements

We thank Michel Tremblay and David Levy for providing the HIV-1 proviral plasmids expressing HSA and GFP, respectively. This work was supported in part by the Japan Science and Technology Agency (JST) PRESTO program (to S.I.); JST RISTEX program (to S.I.); JST CREST program (to S.I.); the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers 15KT0107, 26287025, 16KT0111, 16H04845, and 16K13777 (to S.I.); the Japan Agency for Medical Research and Development, AMED (to S.I.); Mitsui Life Social Welfare Foundation (to S.I.); The Shin-Nihon of Advanced Medical Research (to S.I.); GSK Japan Research Grant 2016 (to S.I.); a grant from Agence Nationale de Recherches sur le Sida et les Hépatites Virales (ANRS) (to F.M.); A.R. was supported by a doctoral fellowship from “Région Martinique”. Finally, we are grateful to the editor and reviewers for their valuable comments and suggestions that improved the presentation of this paper.

## Author Contributions

S.I. and F.M. designed the experiments. A.R. and A.T. conducted the experiments. Y.I., S.N., K.E., Y.I., and S.I. carried out the computational analysis. S.I. and F.M. supervised the project. Y.I., Y.I., S.I., and F.M. wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-03954-9

**Competing Interests:** The authors declare that they have no competing interests.

**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017